Cloudera Data Warehouse Public Cloud

Data Warehouse Release Notes

Date published: 2024-01-01
Date modified: 2024-12-05

https://docs.cloudera.com/

Legal Notice

© Cloudera Inc. 2025. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

Version mapping 6
Version mapping for supported releases 6
Version mapping for unsupported releases 7
Impala shell and Impyla versions 8
December 18, 2024 - Hotfix 8
Fixed issues 8
December 5, 2024 - GA 8
What's new 8
Known issues 10
Fixed issues 24
Deprecation notices 26
Behavior changes 26
October 4, 2024 - Hotfix 28
Fixed issues 28
Known issues 28
August 15, 2024 - Hotfix 29
Fixed issues 29
Known issues 30
Behavior changes 30
July 26, 2024 - GA 31
What's new 31
Preview features in Cloudera Data Warehouse on Public Cloud 33
Known issues in Cloudera Data Warehouse on Public Cloud 35
Fixed issues in Cloudera Data Warehouse on Public Cloud 48
Behavior changes in Cloudera Data Warehouse on Public Cloud 49
May 2, 2024 - Hotfix 51
Fixed issues 51
Behavior changes 51
Unsupported releases 52
March 26, 2024 - Hotfix 52
Fixed issues 53
Known issues 53
March 4, 2024 - Hotfix 69
Fixed issues 69
Known issues 69
Behavior changes 86
February 29, 2024 - GA 86
What's new in Cloudera Data Warehouse on Public Cloud 86
Preview features in Cloudera Data Warehouse on Public Cloud 90
Known issues in Cloudera Data Warehouse on Public Cloud 91
Fixed issues in Cloudera Data Warehouse on Public Cloud 109
Behavior changes in Cloudera Data Warehouse on Public Cloud 109
What's new (unsupported releases) 110
January 18, 2024 110
January 10, 2024 110
December 19, 2023 110
December 1, 2023 111
November 20, 2023 111
October 12, 2023 115
October 5, 2023 115

August 30, 2023....116
June 29, 2023....120
May 30, 2023....120
May 5, 2023....121
February 14, 2023....126
February 7, 2023....126
December 13, 2022....128
November 21, 2022....128
September 15, 2022....130
August 4, 2022....131
June 22, 2022....133
May 10, 2022....134
April 27, 2022....134
March 8, 2022....136
January 19, 2022....137
December 22, 2021....137
November 20, 2021....138
October 29, 2021....138
October 22, 2021....139
September 28, 2021....139
August 27, 2021....139
August 9, 2021....141
July 29, 2021....141
July 13, 2021....141
June 14, 2021....142
June 1, 2021....142
May 21, 2021....142
May 3, 2021....143
April 9, 2021....143
March 31, 2021....143
March 23, 2021....144
March 12, 2021....144
February 23, 2021....144
February 10, 2021....145
January 29, 2021....145
December 21, 2020....145
November 3, 2020....145
September 10, 2020....147
August 27, 2020....148
August 14, 2020....148
July 30, 2020....149
June 26, 2020....149
May 14, 2020....149
April 7, 2020....151
March 13, 2020....151
Known issues (unsupported releases)....152
January 18, 2024....152
January 10, 2024....168
December 1, 2023....184
December 19, 2023....199
November 20, 2023....215
October 16, 2023....231
October 12, 2023....246
October 5, 2023....260
August 30, 2023....275
June 29, 2023....289

May 30, 2023....304
May 5, 2023....319
February 7, 2023....334
February 14, 2023....346
December 13, 2022....346
November 21, 2022....359
September 15, 2022....372
August 4, 2022....384
June 22, 2022....395
General known issues on public clouds....404
Known issues in CDW running on AWS environments....405
Known issues in CDW running on Azure environments....408
Known issues in Data Analytics Studio on public clouds....408
Known issues in Hue on public clouds....409
Known issues in Database Catalog on public clouds....411
Known issues in Hive Virtual Warehouses on public clouds....411
Known issues in Impala Virtual Warehouses on public clouds....413
Behavior changes (unsupported releases)....415
November 17, 2023....415
August 30, 2023....417
February 7, 2023....417
Fixed issues (unsupported releases)....418
January 18, 2024 Fixed issues....418
December 19, 2023 Fixed issues....418
December 1, 2023 Fixed issues....420
November 20, 2023 Fixed issues....420
October 5, 2023 Fixed issues....421
August 30, 2023 Fixed issues....421
May 5, 2023 fixed issues....421
Feb 7, 2023 fixed issues....422
December 13, 2022 fixed issues....422
November 22, 2022 fixed issues....422
September 15, 2022 fixed issues....422
August 4, 2022 fixed issues....422
April 27, 2022 fixed issues....423
CVE-2021-44228 remediation for CDW....423

Version information for Cloudera Data Warehouse on Public Cloud

Cloudera Data Warehouse (CDW) includes components covered in the Cloudera Runtime documentation. A mapping of the CDW versions to its Hive, Impala, Hue versions and CDP CLI counterparts provides the path you follow to select the correct CLI functions for CDW. CDW Runtime version mapping to Apache Iceberg and to Impala client versions is also covered.

Review the Amazon EKS Kubernetes release calendar and AKS Kubernetes release calendar for support information.

Review the Support Lifecycle Policy page for information about CDW releases that have reached end of support (unsupported releases).

Related Information

Cloudera Data Warehouse Runtime Changelog

CDP CLI Reference

CDP Control Plane API Changelog

Version mapping for supported releases

Review the release date, Cloudera Data Warehouse version, supported AKS and EKS versions, runtime versions (this includes Hive, Impala, Hue, and Iceberg), Cloudera Data Visualization version, and CDP CLI version for the supported releases.

Data Warehouse component version information

Version mapping information is provided for the following Data Warehouse components:

- CDW: The Cloudera Data Warehouse Control Plane (CRUD application) version and its release date
- AKS/EKS: Azure Kubernetes Service or Amazon Elastic Kubernetes version of newly-activated environments
- Runtime: The CDW Runtime version that includes Hive, Impala, Iceberg, and Hue
- Apache Iceberg: The Apache Iceberg version
- CDV: The version of Cloudera Data Visualization
- CDP CLI: The CDP command line interface version

Release type	Release date	CDW	AKS/EKS	Runtime	Apache Iceberg	CDV	CDP CLI
Hotfix	December 18, 2024	1.9.4-b147	1.30/1.30	2024.0.18.4	1.4.3	7.2.7-b48	0.9.127 Released on: 2024-12-06
GA	December 5, 2024	1.9.3-b166	1.30/1.30	2024.0.18.4	1.4.3	7.2.7-b48	0.9.127 Released on: 2024-12-06
Hotfix	October 4, 2024	1.9.2-b657	1.29/1.29	2024.0.18.2-4	1.4.3	7.2.4-b41	0.9.119 Released on: 2024-08-02

Release type	Release date	CDW	AKS/EKS	Runtime	Apache Iceberg	CDV	CDP CLI
Hotfix	August 15, 2024	1.9.2-b657	1.29/1.29	2024.0.18.1-1	1.4.3	7.2.4-b41	0.9.119 Released on: 2024-08-02
GA	July 26, 2024	1.9.1-b233	1.29/1.29	2024.0.18.0-206	1.4.3	7.2.4-b41	0.9.119 Released on: 2024-08-02
Hotfix	May 2, 2024	1.8.7-b37	1.27/1.28	2024.0.17.2-2	1.3.0	7.2.2-b33	0.9.108 Released on: 2024-03-13

Version mapping for unsupported releases

Review the release date, Cloudera Data Warehouse version, supported AKS and EKS versions, runtime versions (this includes Hive, Impala, Hue, and Iceberg), Cloudera Data Visualization version, and CDP CLI version for the older, unsupported releases.

Release type	Release date	CDW	AKS/EKS	Runtime	Apache Iceberg	CDV	CDP CLI
Hotfix	Mar 26, 2024	1.8.7-b37	1.27/1.28	2024.0.17.1-3	1.3.0	7.1.9-b19	0.9.108 Released on: 2024-03-13
Hotfix	Mar 4, 2024	1.8.6-b264	1.27/1.28	2024.0.17.0-73	1.3.0	7.1.9-b19	0.9.108 Released on: 2024-03-13
GA	Feb 29, 2024	1.8.5-b35	1.27/1.28	2024.0.17.0-73	1.3.0	7.1.9-b19	0.9.108 Released on: 2024-03-13
Hotfix	Jan 18, 2024	1.8.4-b90	1.27/1.26	2023.0.16.3-2	1.3.0	7.1.6.2-3	0.9.99 Released on: 2023-10-11
Hotfix	Jan 10, 2024	1.8.3-b130	1.27/1.26	2023.0.16.3-2	1.3.0	7.1.6.2-3	0.9.99 Released on: 2023-10-11
Hotfix	Dec 19, 2023	1.8.2-b221	1.27/1.26	2023.0.16.2-1 (Dec 19, 2023)	1.3.0	7.1.6.2-3	0.9.99 Released on: 2023-10-11
Hotfix	Nov 20, 2023	1.8.1-b248	1.27/1.26	2023.0.16.1-2 (Dec 1, 2023)	1.3.0	7.1.6.2-3	0.9.99 Released on: 2023-10-11
GA				2023.0.16.0-150 (Nov 20, 2023)	1.3.0		

Release type	Release date	CDW	AKS/EKS	Runtime	Apache Iceberg	CDV	CDP CLI
Hotfix	Oct 16, 2023	1.7.3-b12	1.26/1.26	2023.0.15.1-2	1.3.0	7.1.3-b36	0.9.97 Released on: 2023-09-14
GA	Oct 5, 2023	1.7.2-b321	1.26/1.26	2023.0.15.1-2	1.3.0	7.1.3-b36	0.9.97 Released on: 2023-09-14

Impala shell and Impyla versions

Review the minimum Impala shell and Impyla versions your client users need to connect to a particular version of CDW.

CDW Runtime Version	Impala shell version	Impyla version
2024.0.18.0-206	4.4.0.a1	-
2023.0.15.0-243 - 2023.0.16.0-150	4.3.0a3	0.18.0
2023.0.14.0-155	4.2.0	0.18.0
2021.0.5-b36 - 2023.0.13.0-20	4.1.0a1	0.18a2

December 18, 2024 - Hotfix

Review the fixed issues and changed behaviors in this hotfix release of Cloudera Data Warehouse on Public Cloud.

Fixed issues

Review the fixed issues in this release of the Cloudera Data Warehouse (CDW) service on Public Cloud.

DWX-19898: Databus producer fails with multiple bucket configurations

After a fresh install or DBC upgrade, the databus-producer fails when separate buckets are configured for data and logs. The error occurs because it expects all paths to be in the primary data bucket, but logs are stored in a different bucket.

The issue was resolved by updating the databus-producer version to 1.1.0-b4 in the Cloudera Data Warehouse 1.9.4-b147 release (released December 18, 2024).

December 5, 2024 Release Notes

Review the new feaures, fixes, behavioral changes, and preview features in this release of Cloudera Data Warehouse on Public Cloud.

What's new in Cloudera Data Warehouse on Public Cloud

Review the new features introduced in this release of Cloudera Data Warehouse (CDW) service on CDP Public Cloud.

- CDW features

- CDW on Azure environments
- CDW on AWS environments
- Impala
- Hue

What's new in CDW Public Cloud

Ability to select an instance type for Virtual Warehouses is GA

You can select AWS or Azure compute instance types, such as r6id.4xlarge or Standard_E16_v3, while creating a Virtual Warehouse, both using CDP CLI and UI. See Creating a Virtual Warehouse.

If you select a compute instance type both during environment activation and creating a Virtual Warehouse, then what you select for a Virtual Warehouse takes precedence.

Improvements to Impala Autoscaler Dashboard

The following inprovements were introduced for the Impala Autoscaler Dashboard:

- Ability to select the log-level configuration for the autoscaler and autoscaler metrics containers.
- A new “Understanding The Dashboard” page has been added which explains the metrics displayed on the UI and how they are calculated.
- Empty data points that manifest as gaps in the graphs are skipped. Zero values are accurately displayed.

Ability to view end-of-support information through UI and CDP CLI

Cloudera Data Warehouse releases reach the end of support every six months. The Cloudera Data Warehouse UI displays whether your deployment is nearing its end of support time or is unsupported, enabling you to plan an upgrade. You can also view the upgrade instructions on the UI. The end of support information is also displayed when you run the list-clusters and describe-clusters CDP CLI commands.

Streamlined option for downloading Cloudera Data Warehouse diagnostic bundles

Cloudera Data Warehouse users can now easily download diagnostic bundles with a direct Collect option that reduces the need for prior time interval and log selection adjustments. This update enables faster, more efficient access to relevant diagnostic data.

For more information, see CDW Diagnostic Bundle Documentation, Downloading CDW and Kubernetes diagnostic bundles and Diagnostic bundles for CDW and Kubernetes.

What's new in CDW on Azure environments

Azure AKS 1.30 upgrade

Cloudera supports the Azure Kubernetes Service (AKS) version 1.30. In 1.9.3-b166 (released December 5, 2024), when you activate an Environment, Cloudera Data Warehouse automatically provisions AKS 1.30. To upgrade to AKS 1.30 from an earlier version of Cloudera Data Warehouse, you must backup and restore Cloudera Data Warehouse.

Note: Using the Azure CLI or Azure portal to upgrade the AKS cluster is not supported. Doing so can cause the cluster to become unusable and can cause downtime. For more information about upgrading, see Upgrading an Azure Kubernetes Service (AKS) cluster.

What's new in CDW on AWS environments

AWS EKS 1.30 upgrade

Cloudera supports the AWS Elastic Kubernetes Service (EKS) version 1.30. In 1.9.3-b166 (released December 5, 2024), when you activate an Environment, Cloudera Data Warehouse automatically provisions EKS 1.30. To upgrade to EKS 1.30 from an earlier version of Cloudera Data Warehouse, you must backup and restore Cloudera Data Warehouse. For more information about upgrading, see Upgrading an Amazon Kubernetes Service (EKS) cluster.

What's new in Impala on CDW Public Cloud

Improved log storage location access for S3 and Azure in Cloudera Data Warehouse

Cloudera Data Warehouse now offers a simplified way to locate Impala diagnostic logs in S3 and Azure. With fewer required identifiers and clear directory paths, users can efficiently access the logs they need. Find details on the updated steps in S3 and Azure.

What's new in Hue on CDW Public Cloud

Enhanced AI Integration in Hue SQL AI Assistant

The Hue SQL AI Assistant now supports Cloudera AI Workbench and Cloudera AI Inference service. These integrations enhance the Hue SQL AI Assistant by enabling the use of private models hosted within Cloudera-managed infrastructure. This ensures enhanced security and privacy while leveraging GenAI for the Hue SQL-related tasks.

- Cloudera AI Workbench: This enables you to securely deploy and run your own models within a virtual private cloud. This configuration enhances control and privacy within your environment. For more information, see Configure SQL AI Assistant using Cloudera AI Workbench.
- Cloudera AI Inference service: Helps in a production-grade serving environment for hosting predictive and generative AI models. This service simplifies model deployment and maintenance. For more information, see Configure SQL AI Assistant using Cloudera AI Inference service.

Known issues in Cloudera Data Warehouse on Public Cloud

Review the known issues in this release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud.

- Known issues identified in the December 5, 2024 release on page 10
- Known issues identified before the December 5, 2024 release on page 11
- Technical Service Bulletins on page 23

Known issues identified in the December 5, 2024 release

Impala Virtual Warehouse fails to start due to missing helm release

A failed helm release status prevents helm upgrade operations from executing successfully. This causes the Impala Virtual Warehouse to fail with the following error:

```
time="2024-07-04T13:32:46Z" level=info msg="helm release not fou
nd for input executor group: \"impala-executor-000\" in service
impala-executor"
time="2024-07-04T13:32:46Z" level=error msg="addExecutorGroups
failed with err: \"helm release not found for \\\"impala-executo
r-000\\\"\""
```

Perform one of the following tasks to address this issue:

- Recreate the Virtual Warehouse
- Manually roll back the Helm release using the Helm client.

Look through the release history and identify a release that was in a healthy deployed state. Roll back to the identified release version to bring back the Virtual Warehouse to a healthy state.

Database Catalog goes to an error state after a Data Lake resize or after an upgrade to Azure Flexible Server

After a CDP Data Lake resize or after an upgrade to Azure Flexible Server, the Cloudera Data Warehouse Database Catalog needs to be restarted to consider the change in Data Lake size. Users might restart the Database Catalog using the CDP CLI or by using the Stop and Start functionality from the Cloudera Data Warehouse UI.

Irrespective of how the Database Catalog is restarted, it ends up in an "Error" state.

After you resize the CDP Data Lake or upgrade from Azure Single Server to Flexible Server, start the Database Catalog from the Cloudera Data Warehouse UI and then rebuild it.

Note: The Database Catalog will fail when you start it, however, this step is required and needs to be performed before rebuilding the Database Catalog.

You must also ensure that you restart each Virtual Warehouse as recommended in the Data Lake resizing post-requisites or Post-requisites after upgrading to Azure Flexible Server.

Known issues identified before the December 5, 2024 release

DWX-19451: Data Visualization restore job can fail with ignorable errors

After a successful Data Visualization restoration job, the Data Visualization restore job could be in a failed state with the log displaying ignorable errors.

```
pg_restore: error: could not execute query: ERROR:  sequence "jo
bs_joblog_id_seq" does not exist
Command was: DROP SEQUENCE public.jobs_joblog_id_seq;
pg_restore: error: could not execute query: ERROR:  table "jobs_j
oblog" does not exist
Command was: DROP TABLE public.jobs_joblog;
pg_restore: error: could not execute query: ERROR:  sequence "
jobs_jobcontent_id_seq" does not exist
Command was: DROP SEQUENCE public.jobs_jobcontent_id_seq;
.......
.......
```

This issue occurs because the restore job issues commands to DROP all the objects that will be restored, and if any of these objects do not exist in the destination database, such ignorable errors are reported.

This has no functional impact on the restored Data Visualization application. It is noticed that all the backed up queries, datasets, connections, and dashboards are restored successfully and Data Visualization is available for new queries.

None.

DWX-19003: Unable to set t-shirt size for a Cloudera Data Visualization instance

You cannot set or configure a t-shirt size for a Cloudera Data Visualization instance while creating or editing it from CDW. By default, any Cloudera Data Visualization instance created in CDW release version 1.9.1 uses the “small” compute instance size.

None.

DWX-18950: Hue data restoration fails on Azure CDW clusters having CDW version 1.8.7 or earlier

If you backup your CDW cluster running CDW version 1.8.7 or earlier and if you perform the restore operation without reactivating the latest CDW environment, you might encounter Hue restoration job failure.

None.

DWX-18843: Unable to read Iceberg table from Hive Virtual Warehouse

If you have used Apache Flink to insert data into an Iceberg table that is created from Hive, you cannot read the Iceberg table from the Hive Virtual Warehouse.

Add the engine.hive.enabled table property through the Hive beeline and set the value to "true". You can add this table property either while creating the Iceberg table or use the ALTER TABLE statement to add the table property.

DWX-18489: Hive compaction of Iceberg tables results in a failure

When Cloudera Data Warehouse (CDW) and Data Hub are deployed in the same environment and use the same Hive Metastore (HMS) instance, the Data Hub compaction workers can inadvertently

pick up Iceberg compaction tasks. Since Iceberg compaction is not yet supported in the latest Data Hub version, the compaction tasks will fail when they are processed by the Data Hub compaction workers.

In such a scenario where both CDW and Data Hub share the same HMS instance and there is a requirement to run both Hive ACID and Iceberg compaction jobs, it is recommended that you use the CDW environment for these jobs. If you want to run only Hive ACID compaction tasks, you can choose to use either the CDW or Data Hub environments.

If you want to run the compaction jobs without changing the environment, it is recommended that you use CDW. To avoid interference from Data Hub, change the value of the hive.compactor.worker.threads Hive Server (HS2) property to '0'. This ensures that the compaction jobs are not processed by Data Hub.

1. In Cloudera Manager, click Clusters Hive Configuration to navigate to the configuration page for HMS.
2. Search for hive.compactor.worker.threads and modify the value to '0'.
3. Save the changes and restart the Hive service.

DWX-18854: Compaction cleaner configuration

The compaction cleaner is turned off by default in the CDW database catalog, potentially causing compaction job failures.

To enable the compactor on the Hive Metastore (HMS) instance of the CDW database catalog, set the hive.compactor.cleaner.on property to "true".

DWX-12703: Hue connects to only one Impala coordinator in Active-Active mode

You may not see all Impala queries that have run on the Virtual Warehouse from the **Impala** tab on the Hue **Job Browser**. You encounter this on an Impala Virtual Warehouse that has Impala coordinator configured in an active-active mode. This happens because Hue fetches this information from only one Impala coordinator that is active.

None. You can view the query history from the **Impala Queries** tab on the **Job Browser** page, because the information is fetched from the Hue Query Processor.

Known limitation: CDW does not support S3 Express One Zone buckets

Cloudera does not recommend deploying the Data Lake on S3 Express One Zone buckets. CDW cannot read content present in the S3 Express One Zone buckets. The following limitations apply when using S3 Express buckets:

- You can only use S3 Express buckets with Data Hubs running Runtime 7.2.18 or newer. Data services do not support it, currently.
- S3 Express buckets may not be used for logs and backups.

DWX-15112: Enterprise Data Warehouse database configuration problems after a Helm-related rollback

If a Helm rollback fails due to an incorrect Enterprise Data Warehouse database configuration, the Virtual Warehouse and Database Catalog roll back to a previous configuration. The incorrect Enterprise Data Warehouse configuration persists, and can affect subsequent edit, upgrade, and rebuild operations on the rolled-back Virtual Warehouse or Database Catalog.

None.

DWX-17455: ODBC client using JWT authentication cannot connect to Impala Virtual Warehouse

If you are using the Cloudera ODBC Connector and JWT authentication to connect to a CDW Impala Virtual Warehouse where the Impala coordinators are configured for high availability in an active-active mode, the connection results in a "401 Unauthorized" error. The error is not seen on Impala Virtual Warehouses that have an active-passive high availability or where high availability is disabled.

In the ODBC Data Source Name (DSN), make the following changes:

1. Ensure that SSL is enabled by setting SSL=1.

2. Remove authentication by setting AuthMech=0.
3. Remove the JWTString parameter.
4. Add the parameter http.header.Authorization and set it to the word "Bearer" followed by a space and the JWT string

DSN configuration before the workaround

```
SSL=1
AuthMech=8
JWTString=full_jwt_text
```

DSN configuration after the workaround

```
SSL=1
AuthMech=0
http.header.Authorization=Bearer full_jwt_text
```

HIVE-28055: Merging Iceberg branches requires a target table alias

Hive supports only one level of qualifier when referencing columns. In other words only one dot is accepted. For example, select table.col from ...; is allowed. select db.table.col is not allowed. Using the merge statement to merge Iceberg branches without a target or source table alias causes an exception:

```
org.apache.hadoop.hive.ql.parse.SemanticException: ... Invalid t
able alias or column reference ...
```

Use an alias, for example t, for the target table.

```
merge into mydb.target.branch_branch1 t using mydb.source.branch
_branch1 s on t.id = s.id when matched then update set value = '
matched';
```

DWX-17620: Folder having special characters in its name is not accessible in ABFS

In Cloudera Data Warehouse Public Cloud, from a Virtual Warehouse going to the Azure Blob Filesystem (ABFS), creating a folder, and then performing an action such as Move, causes an error as shown in the following example:

```
Cannot access: abfs://data-files/user/hrt_qa/~@$&()*!+'=;. 404 C
lient Error: The specified path does not exist.
```

None.

Branch FAST FORWARD does not work as expected

The Apache Iceberg spec indicates you can use either one or two arguments to fast forward a branch. The following example shows using two arguments:

```
ALTER TABLE <name> EXECUTE FAST-FORWARD 'x' 'y'
```

However, omitting the second branch name, does not work as documented by Apache Iceberg. The named branch is not fast-forwarded to the current branch. An exception occurs at the Iceberg level.

You must use two arguments to the EXECUTE FAST FORWARD feature to forward a branch.

DWX-17613: Generic error message is displayed when you click on the directory you don't have access to on a RAZ cluster

You see the following error message when you click on an ABFS directory to which you do not have read/write permission on the ABFS File Browser in Hue: There was a problem with your request. This message is generic and does not provide insight into the actual issue.

None.

DWX-17109: ABFS File Browser operations failing intermittently

You may encounter intermittent issues while performing typical operations on files and directories on the ABFS File Broswer, such as moving or renaming files.

None.

CDPD-27918: Hue does not automatically pick up RAZ HA configurations

On a CDP Public Cloud environment in which you have configured RAZ in High Availability mode, Hue in CDW does not pick up all the RAZ host URLs automatically. Therefore, if a RAZ instance to which Hue is connected goes down, Hue becomes unavailable.

You must manually add comma-separated RAZ instances in the Hue Advanced Configuration Snippet.

1. Log in to the CDP Management Console as an Administrator.
2. Go to Environment Data Lake and open Cloudera Manager for your environment.
3. Go to Clusters Ranger RAZ service Instances RAZ server Processes and note the value of the fs.s3a.ext.raz.rest.host.url property from the core-site.xml file. You need this to specify the value of the api_url property in the Hue configuration.

   For Azure environments, note the value of the fs.azure.ext.raz.rest.host.url property.

   For AWS and GCS environments, note the value of the fs.s3a.ext.raz.rest.host.url property.
4. Go to CDW Virtual Warehouse ⋮ Edit CONFIGURATIONS and select the hue-safety-valve from the Configuration files dropdown menu.
5. Add the following lines in the hue-safety-valve field:

```
[desktop]
[[raz]]
is_enabled=true
api_url=https://[***INSTANCE-1***]:6082/,https://[***INSTANCE-
2***]:6082/
```

6. Click Apply Changes.

CDPD-66779: Partitioned Iceberg table not getting loaded with insert select query from Hive

If you create a partitioned table in Iceberg and then try to insert data from another table as shown below, an error occurs.

```
insert into table partition_transform_4 select t, ts from vector
tab10k;
```

Use the CLUSTER BY clause on the partitioned column to insert data. For example:

```
insert into table partition_transform_4 select t, ts from t1 clu
ster by ts;
```

DWX-17703: Non-HA Impala Virtual Warehouse on a private Azure Kubernetes Service (AKS) setup fails

When 'Refresh' and 'Stop' operations run in parallel, Impala might move into an error state. The Refresh operation might think that Impala is in an error state as the coordinator pod is missing.

Rebuild the Impala Virtual Warehouse or restart it using the CLI.

DWX-14923: After JWT authentication, attempting to connect the Impyla client using a user name or password should cause an error

Using a JWT token, you can connect to a Virtual Warehouse as the user who generated the token.

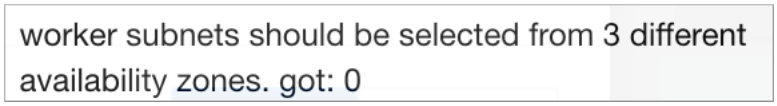
If you connect with the JWT token, and then pass a user name or password from Impyla to the Virtual Warehouse, the connection arguments are silently ignored. Such an action should indicate that it is illegal to specify user or password when using JWT authentication.

None.

DWX-15145: Environment validation popup error after activating an environment

Activating an environment having a public load balancer can cause an environment validation popup error.

To reproduce this problem 1) Create a data lake. 2) Activate an environment having a public load balancer deployment type and subnets in three different availability zones. A environment validation popup can occur even through subnets are in different availability zones. Several different popups can occur, including the following one:



None.

DWX-15144: Virtual Warehouse naming restrictions

You cannot create a Virtual Warehouse having the same name as another Virtual Warehouse even if the like-named Virtual Warehouses are in different environments. You can create a Database Catalog having the same name as another Database Catalog if the Database Catalogs are in different environments.

None.

DWX-13103: Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Restart the prometheus pod so that it gets the istio-proxy container.

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the CreateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersions line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Incorrect diagnostic bundle location

The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

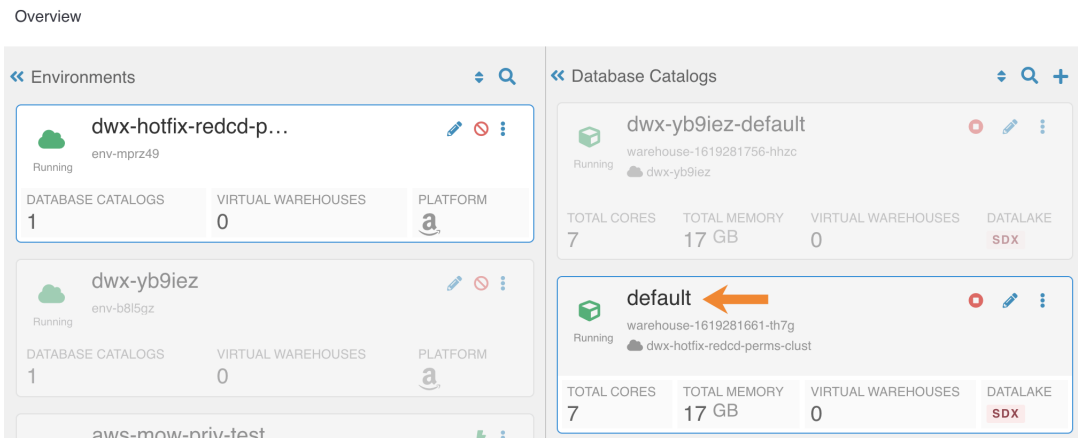
To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

None.

DWX-15064: Hive Virtual Warehouse stops but appears healthy

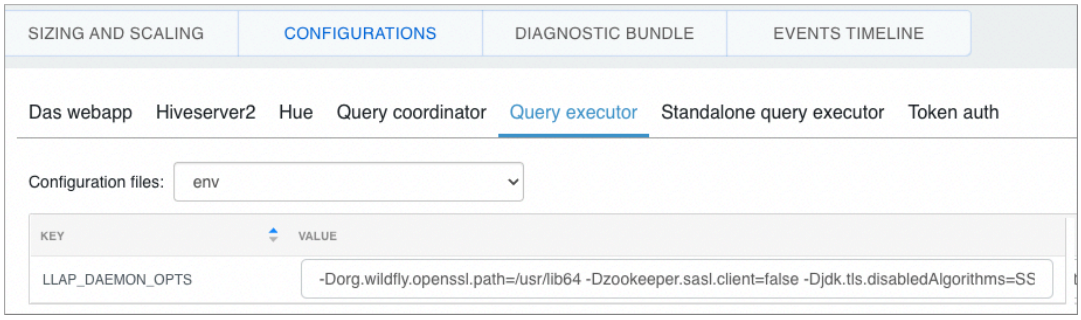
Due to an istio-proxy problem, the query coordinator can unexpectedly enter a not ready state instead of the expected error-state. Subsequently, the Hive Virtual Warehouse stops when reaching the autosuspend timeout without indicating a problem.

None.

DWX-14452: Parquet table query might fail

Querying a table stored in Parquet from Hive might fail with the following exception message: java.lang.RuntimeException: java.lang.StackOverflowError. This problem can occur when the IN predicate has 243 values or more, and a small stack (-Xss = 256k) is configured for Hive in CDW.

Change the value of the Xss in the JVM args to use the default value (1Mb) as follows: Select Edit from the Options menu of your Virtual Warehouse. Click CONFIGURATIONS Query executor and select the env configuration file.



In the third line shown below, change the value of LLAP_DAEMON_OPTS from -Xss256k to -Xss1M, and then click Apply Changes:

FROM:

-Dorg.wildfly.openssl.path=/usr/lib64 -Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Dhttp.maxConnections=12 -Xms24G -Xmx48G -Xss256k ...

TO:

... -Xss1M ...

CDPD-40730: Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

None.

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.

7. Click Apply in the upper right corner of the page to save the configuration.

8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name ‘ec2-18-219-
32-183.us-east-2.compute.amazonaws.com’ does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

DWX-16899: Error while viewing Impala job status on the mini Job Browser

When you click on the application ID after submitting an Impala query on Hue that is running on the environment level, you may notice the following error: 401 Client Error: Unauthorized for url: http://coordinator.impala-xyz.svc.cluster.local:25000/queries?json=true Must authenticate with Basic authentication. This does not happen when you do the same from Hue deployed at a Virtual Warehouse level.

None.

DWX-16895: Incorrect status of Hue pods when you edit the Hue instance properties

When you update a configuration of a Hue instance that is deployed at the environment level, such as increasing or decreasing the size of the Hue instance, you see a success message on the CDW UI. After some time, the status of the Hue instance also changes from “Updating” to “Running”. However, when you list the Hue pods using kubectl, you see that not all backend pods are in the running state–a few of them are still in the init state.

None. The pods come up successfully eventually after a sufficient time has passed.

DWX-16863: The upgrade button is present on the CDW UI, but Hue upgrades are not supported

You see the Upgrade button on the **Query Editor** page in the CDW UI when Hue is deployed at the environment level. However, on CDW version 1.8.1, upgrading the Hue instance that is deployed at the environment level is not supported.

None.

DWX-16893: A user can see all the queries in Job browser

In a Hue instance deployed at the environment level, by design, the Hue instances must not share the saved queries and query history with other Hue instances even for the same user. However, a logged in user is able to view all the queries executed by that user on all the Virtual Warehouses on a particular Database Catalog.

None.

Delay in listing queries in Impala Queries in the Job browser

Listing an Impala query in the Job browser can take an inordinate amount of time.

None.

DWX-14927: Hue fails to list Iceberg snapshots

Hue does not recognize the Iceberg history queries from Hive to list table snapshots. For example, Hue indicates an error at the . before history when you run the following query.

```
select * from <db_name>.<table_name>.history
```

None.

DWX-14968: Connection termination error in Impala queries tab after Hue inactivity

To reproduce the problem: 1) In the Impala job browser, navigate to Impala queries. 2) Wait for a few minutes.

After a few minutes of inactivity, the follow error is displayed:

Refresh the page, or alternatively start a new session.

DWX-15115: Error displayed after clicking on hyperlink below Hue table browser

In Hue, below the table browser, clicking the hyperlink to a location causes an HTTP 500 error because the file browser is not enabled for environments that are not Ranger authorized (RAZ).

None.

DWX-15090: CSRF error intermittently seen in the Hue Job Browser

You may intermittently see the “403 - CSRF” error on the Hue web interface as well as in the Hue logs after running Hive queries from Hue.

Reload the browser or start a new Hue session.

IMPALA-11447: Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

DWX-17210, DWX-13733: Timeout issue querying Iceberg tables from Hive

When querying Iceberg tables from Hive, the queries can faile due to a timeout issue.

1. Add the following configurations to hadoop-core-site for the Database Catalog and the Virtual Warehouse.
   - fs.s3.maxConnections=1000
   - fs.s3a.connection.maximum=1000
2. Restart the Database Catalog and Virtual Warehouse.

DWX-15014: Loading airports table in demo data fails

This issue occurs only when using a non-default Database Catalog. A invalid path error message occurs when using the load command to load the demo data airports table in Iceberg, ORC, or Parquet format.

None.

DWX-14163: Limitations reading Iceberg tables in Avro file format from Impala

The Avro, Impala, and Iceberg specifications describe some limitations related to Avro, and those limitations exist in CDP. In addition to these, the DECIMAL type is not supported in this release.

None.

DEX-7946: Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata if you drop the table during migration from Hive to Iceberg.

Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13062: Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276: Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

IMPALA-11045: Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

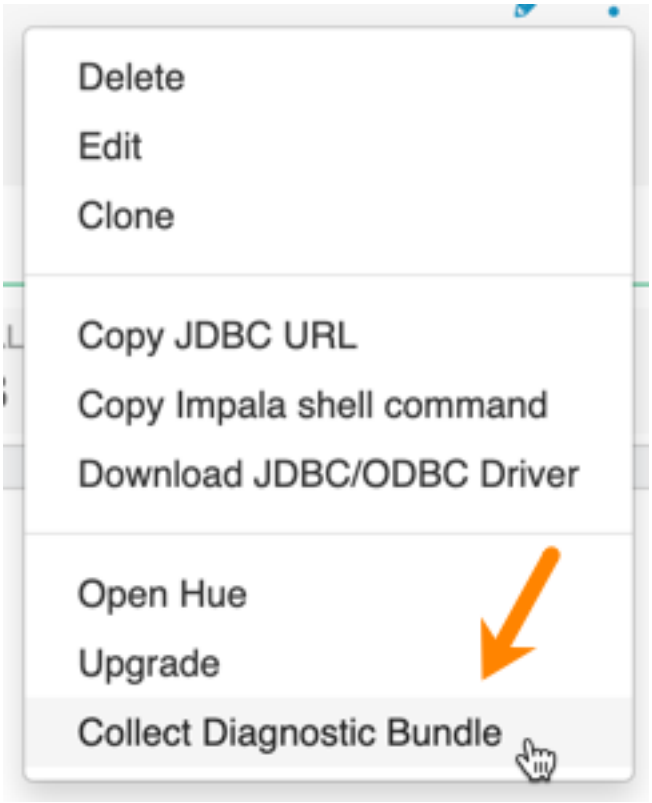
DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



None.

Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

None.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

None.

Technical Service Bulletins

TSB 2023-719: Cloudera Data Warehouse Backup/Restore of CDP Data Visualization incomplete

Cloudera Data Warehouse (CDW) customers using the CDW Automated Backup/Restore feature will encounter an issue with the restoration versions of Cloudera Data Visualization (CDV) older than CDV 7.1.6.2-3 due to a schema change in this release. If the Backup was taken from an older CDW environment that contained a version of CDV older than CDV 7.1.6.2-3, the Restore procedure will succeed. Though once the user opens the CDV Queries tab, the user could encounter the error message: “column jobs_jobschedule.owner_id does not exist…”

None. For the latest update on this issue see the corresponding Knowledge article: TSB 2023-719: Cloudera Data Warehouse Backup/Restore of Cloudera Data Visualization incomplete.

TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

The Cloudera Data Warehouse (CDW) Public Cloud 2023.0.14.0 (DWX-1.6.3) version incorporated a feature for performing automatic metadata synchronization across multiple Apache Impala (Impala) Virtual Warehouses. The feature is enabled by default, and relies on the Hive MetaStore events. When a certain sequence of Data Definition Language (DDL) SQL commands are executed as described below, users may encounter a java.lang.NullPointerException (NPE). The exception causes the event processor to stop processing other metadata operations.

If a CREATE TABLE command (not CREATE TABLE AS SELECT) is followed immediately (approximately within 1 second interval) by INVALIDATE METADATA or REFRESH TABLE command on the same table (either on the same Virtual Warehouse or on a different one), there is a possibility that the second command will not find the table in the catalog cache of a peer Virtual Warehouse and generate an NPE.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

Fixed issues in Cloudera Data Warehouse on Public Cloud

Review the issues fixed in this release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud.

CDPD-74640: Improved query consistency and data writing for Beeline and Hive queries

In concurrent workflows using Beeline, queries occasionally returned incorrect results due to non-thread-safe file handling, especially when hive.query.result.cached.enabled was disabled. Additionally, INSERT OVERWRITE DIRECTORY operations failed to write data correctly to specified directories when query result caching was enabled.

The issue was addressed by implementing thread-safe file handling in HiveSequenceFileInputFormat and adjusting cache handling for directory overwrite queries, ensuring reliable query results and consistent data writes in concurrent workflows.

Apache Jira: HIVE-28530, HIVE-21386, HIVE-25907

CDPD-72985: Compatibility issue in HMS thrift struct for Hive column stats

Hive4 introduced a new field "engine" to differentiate the stats generated by different engines and this was a required field. This breaks the compatibility with clients that are already using Hive 3 or engines that are using customized thrift api like `TrinoDB`.

Made the "engine" field optional in the HMS API, restoring compatibility with Hive 3 clients for column stats retrieval.

Apache Jira: HIVE-27984

CDPD-71484: Improve LLAP performance by reusing FileSystem objects across tasks

Frequent closure of FileSystem objects disabled Hadoop's FileSystem cache, reducing LLAP efficiency.

Adjusted FileSystem handling to close objects once per query and daemon rather than per task, enhancing reuse and maintaining cache functionality.

Apache Jira: HIVE-27884

CDPD-74017: Schema resolution does not work for migrated partitioned Iceberg tables with complex data types

Schema resolution does not work correctly for migrated partitioned Iceberg tables that have complex data types. This fix addresses the field ID generation by taking the number of partitions into account. If none of the partition columns are included in the data file (common scenario), file-level field IDs are adjusted accordingly. You could also come across a scenario where all the partition columns are included in the data files.

However, if some partition columns are included in the data file while other partition columns are not, an error is generated.

Apache Jira: IMPALA-13364

DWX-18975: Configure statestored's memory based on executor group size for Impala Virtual Warehouse with workload-aware autoscaling

For smaller executor group sizes statestored's memory is configured to 1GB while for larger executor group sizes it is configured to 2GB to handle updates from more executors.

DWX-18701: Increasing Catalogd's JVM's Xmx does not increase container's memory limit

Increasing the JVM's Xmx for the Catalogd did not lead to a corresponding increase in the container's memory request and limit, leading to a mismatch between the configuration and actual memory usage. Adjusted the configuration process to ensure that when JVM's Xmx memory is increased, the container's memory request and limits are updated accordingly.

CDPD-73442: Resolution of potential deadlock

This fix addresses a deadlock issue in long-running sessions with an active idle_query_timeout, which caused new queries to hang and prevented existing queries from expiring.

Apache Jira: IMPALA-13313

CDPD-73187: Impala Ranger audit plugin fails to create audit logs

The fix ensures that the Ranger plug-in in Hive and Impala send audit events to the Solr service after upgrading Data Lake to a version that requires SSL for Ranger's audit events.

DWX-18050: Ranger audit logs show origin client's IP address for Impala Virtual Warehouse

Ranger audit logs now show the origin client's IP address for Impala Virtual Warehouse when Impala coordinator's flagfile config has use_xff_address_as_origin=true. This applies to all Impala clients such as impala-shell, impyla, jdbc and odbc clients.

Workaround for Hue: When using Hue client, enable_xff_for_hive_impala=true in hue-safety-valve configuration is additionally also required to get origin client's IP in ranger audit logs.

Note: This change does not fix the issue where Cloudera Data Visualization connections log an internal IP address and not the origin client's IP address in ranger audit.

DWX-19110: Executor deployment issue on workload-aware autoscaling VW creation

Workload-aware autoscaling Impala Virtual Warehouses deployed only one executor for the small group, ignoring the configured group size. The StatefulSet for this group also has a replica count of 1, which does not reflect the expected configuration.

The deployment now respects the configured small group size, ensuring the correct number of executors is deployed. This issue is now fixed.

DWX-19309: Prometheus pods fail due to unencrypted EBS volumes

Prometheus pods switched to EBS volumes failed to start because the default ebs-storageclass did not create encrypted volumes. This caused authorization errors, preventing the pods from running and impacting Grafana, Impala autoscaling, and Hue functionality.

The issue was resolved by ensuring EBS volumes used by Prometheus are encrypted, enabling successful pod deployment and restoring all affected services.

DWX-18448: Impala Virtual Warehouse size changes during updates

When using the update-vw command in the CDP CLI, the --template flag is required to specify the Virtual Warehouse size.

This issue is now fixed.

DWX-16875: Improved memory management for data publishing to Observability dashboard

Data from Virtual Warehouses was intermittently not updating on the Observability dashboard. Restarting the databus-producer deployment temporarily resolved the issue. The pod was being

abruptly killed due to memory limits being exceeded, and the JVM settings were not optimized for dynamic resource allocation.

This issue is now fixed.

DWX-18932: Incorrect High Availability mode displayed for Impala Virtual Warehouses

When creating an Impala Virtual Warehouse (VW) in ACTIVE_PASSIVE High Availability (HA) mode with coordinator auto-shutdown enabled, the CLI incorrectly displayed the HA mode as ACTIVE_ACTIVE.

The logic for determining HA mode has been corrected. This ensures that ACTIVE_PASSIVE and ACTIVE_ACTIVE modes are correctly distinguished.

DWX-19172: Upgrade Cluster Autoscaler to version 1.30.2

The Kubernetes Cluster Autoscaler version is upgraded to 1.30.2 with the corresponding 9.37.0 chart version to ensure support for Amazon Elastic Kubernetes Service (EKS) 1.30.

Deprecation notices in Cloudera Data Warehouse on Public Cloud

Review the features and functionalities that have been or will be removed or deprecated in this release of Cloudera Data Warehouse Public Cloud.

The ability to select a compute instance type during environment activation is deprecated

The ability to select a compute instance type during environment activation is deprecated and will be removed in a future release. In 1.9.3, you can select a compute instance type only when you use CDP CLI to activate an environment in Cloudera Data Warehouse. This option is no longer available on the UI.

The “CDW_CLUSTER_OVERPROVISIONER” entitlement is deprecated

The CDW_CLUSTER_OVERPROVISIONER entitlement allows you to specify the number of compute nodes available on standby in the underlying Kubernetes cluster in the Cloudera Data Warehouse web interface so that autoscaling and new Virtual Warehouse creation happen faster. This entitlement is deprecated and no longer functional.

Behavior changes

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following behavior changes:

Summary: Optimized ephemeral storage for Impala's Catalogd

Ephemeral storage for Catalogd is now optimized for shared nodes, automatically set to the lesser of 24GB or twice the Java heap (Xmx) plus a 1GB buffer.

Before this release: The storage limit of 512MB often caused container evictions due to insufficient space, especially during JVM heap dumps.

After this release: The ephemeral storage limit for Catalogd is now set dynamically based on the JVM heap size (Xmx), calculated as

```
2 * Xmx + Buffer space
```

This approach ensures adequate storage for heap dumps and minimizes the risk of container eviction, resulting in improved stability for Catalogd and other services sharing node resources.

Summary: Optimized memory management for databus-producer deployment

Memory management for the databus-producer deployment has been optimized to improve stability and ensure consistent data publishing to the Observability dashboard.

Before this release:

- Data updates to the Observability dashboard stopped intermittently.
- The pod could be abruptly terminated if memory usage exceeded the requested limit.
- JVM heap memory (-Xmx) was set to a hardcoded value, requiring manual adjustments during updates.

After this release:

- Memory requests and limits are now aligned to prevent abrupt pod termination.
- The JVM heap memory (-Xmx) is dynamically set to 70% of the memory resource limit.
- Enhanced JVM options allow automatic heap dumps in the event of out-of-memory errors, simplifying debugging.

Summary: Deterministic IDs for environments, Database Catalog, and Virtual Warehouses

Before this release: Every time you reactivated an Environment, the Environment ID changed and this led to a change in the path of the logs. The log folders and diagnostic bundle paths were written to the root folder of the storage instead of the configured paths.

After this release: Only the configured paths that are added during the Data Lake creation and prefixed by /tmp are used by the service. Also, the logs are prefixed with the Environment and Data Base Catalog name (wherever applicable) instead of the IDs to simplify manual searching through the logs. The base path of the log folder is also changed and users do not have to go to the external/sys.db/hive folders.

The various log files are still prefixed with the entity IDs in storage and will continue to remain in use to simplify the debugging process but navigating to the entity logs is now made easier.

Summary: Changes to the supported Azure instance types

Before this release: You could only select an instance while activating a Data Warehouse Environment, and the Data Warehouse Control Plane created the node pool from this specific instance. There was no option to select an instance type for a Virtual Warehouse.

After this release: You can select an instance during Environment activation (only through CDP CLI) and Virtual Warehouse creation (through UI and CDP CLI). The Data Warehouse Control Plane requires the following instance types to be allowed, if they are available, in your Azure region to support instance type selection while creating a Virtual Warehouse:

- Standard_E16_v3
- Standard_E16ds_v4
- Standard_E16ads_v5
- Standard_E16ds_v5
- Standard_E16pds_v5

The Data Warehouse Control Plane creates node pools using all these instance types. If you get an error such as the following while activating a Data Warehouse Environment, then increase the quota of the problematic Azure instance type:

```
The VM size of Standard_E16ads_v5 is not allowed in your subscription in loc
ation 'eastus'.
```

You may request that your infrastructure team allow the Azure instance type by asking them to increase the resource quota. A minimal increase in the resource quota is sufficient for the instance to be allowed.

Note: You must understand the distinction between "available" instance types and "allowed" instance types. Cloudera Data Warehouse filters for regional instance type availability (available instance types) but cannot detect quota problems (allowed instance types).

Summary: Simplified CDW Diagnostic Bundle Download Process

The diagnostic bundle download process in CDW has been simplified for an improved user experience.

Before this release: Users had to select specific information types within time intervals or choose a custom time interval. Additionally, they needed to manually adjust options in "Collect For" to include or exclude types of logs for the bundle.

After this release: Users now directly access a simplified “Collect” option, eliminating the need for additional time interval and log selection adjustments.

October 4, 2024 - Hotfix

Review the fixed issues and changed behaviors in this hotfix release of Cloudera Data Warehouse on Public Cloud and learn how to apply this hotfix.

Attention: To apply this hotfix and to upgrade the Runtime version to 2024.0.18.2-4, navigate to the Virtual Warehouses page in the Data Warehouse UI, identify the Virtual Warehouse that you want to upgrade, and click Upgrade for each Virtual Warehouse. It is not necessary to reactivate the Data Warehouse Environment to apply this hotfix.

Fixed issues

This hotfix release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces a Runtime fix.

DWX-18477: (Addendum) Merge query failures with reserved keywords as column names

If a reserved keyword is used as a column name in a query, you must escape the keyword by enclosing it within backticks (`). However, If you have merge queries involving Iceberg target tables that use reserved keywords as column names, the backticks are not retained when the merge queries are rewritten to a join query, and the query fails with a "SemanticException" error.

This issue is now fixed as part of an additional fix that was provided in HIVE-28282.

Known issues

Review the known issues in this release of the Cloudera Data Warehouse (CDW) service on Public Cloud.

DWX-19451: Data Visualization restore job can fail with ignorable errors

After a successful Data Visualization restoration job, the Data Visualization restore job could be in a failed state with the log displaying ignorable errors.

```
pg_restore: error: could not execute query: ERROR:  sequence "jo
bs_joblog_id_seq" does not exist
Command was: DROP SEQUENCE public.jobs_joblog_id_seq;
pg_restore: error: could not execute query: ERROR:  table "jobs_j
oblog" does not exist
Command was: DROP TABLE public.jobs_joblog;
pg_restore: error: could not execute query: ERROR:  sequence "
jobs_jobcontent_id_seq" does not exist
Command was: DROP SEQUENCE public.jobs_jobcontent_id_seq;
.......
.......
```

This issue occurs because the restore job issues commands to DROP all the objects that will be restored, and if any of these objects do not exist in the destination database, such ignorable errors are reported.

This has no functional impact on the restored Data Visualization application. It is noticed that all the backed up queries, datasets, connections, and dashboards are restored successfully and Data Visualization is available for new queries.

None.

DWX-19454: Default Database Catalog does not start in an Azure transparent proxy, private link, and private AKS setup

If you are running a transparent proxy with a private link and private AKS setup in your Azure environment, the Data Warehouse server is unable to connect to the private AKS end point, and the default Database Catalog fails to start.

None.

August 15, 2024 - Hotfix

Review the fixed issues and changed behaviors in this hotfix release of Cloudera Data Warehouse on Public Cloud.

Fixed issues

Review the issues fixed in this hotfix release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud.

DWX-19003: Unable to set t-shirt size for a Cloudera Data Visualization instance

You could not set or configure a t-shirt size for a Cloudera Data Visualization instance while creating or editing it from CDW. By default, any Cloudera Data Visualization instance created in CDW release version 1.9.1 used the small t-shirt size. This issue has been resolved.

DWX-19034: Unable to create an Impala Virtual Warehouse with HA enabled on older runtime versions

Enabling HA on Impala Virtual Warehouses on runtime versions lower than 2024.0.18.0-206 caused errors such as the following on the catalogd pods:

```
I0805 13:33:15.674875     1 thrift-util.cc:198] TSocket::open()
 getaddrinfo() <Host: statestored-0 Port: 24000>Name or service
 not known
I0805 13:33:15.674933     1 thrift-client.cc:82] Couldn't open
 transport for statestored-0:24000 (Could not resolve host for
 client socket.)
```

This was because enabling HA on an Impala Virtual Warehouse enabled Statestore HA, too. Statestore HA is available starting with runtime 2024.0.18.0-206. This issue has been resolved.

DWX-19035: Service discovery does not work with Impala HA

If you created an Impala Virtual Warehouse with HA enabled in CDW 1.9.1-b233, and then tried to link to it from Cloudera Data Visualization, you could not view the required Virtual Warehouse on the **Create New Data Connection** modal. This issue has been resolved.

IMPALA-13270: Addressing IllegalStateException in Complex Views post upgrade

When executing queries that generate runtime filters where the same column identifier appears repeatedly after upgrading to the 2024.0.18.0-206 runtime, you encountered the following error: IllegalStateException: null. This issue has been fixed.

IMPALA-13272: Stability Improvement for analytic functions on collections

The sorting process incorrectly included unnecessary elements, causing errors during array operations and leading to frequent query failures. This issue has been fixed to ensure only complete data entries are used in the sorting process. This prevents crashes and maintains stable execution of analytic functions on collections.

Known issues

Review the known issues in this release of the Cloudera Data Warehouse (CDW) service on Public Cloud.

DWX-19138: The option to enable Impala query logging is unavailable on the CDW UI

You do not see the Log Impala queries option while creating or editing a Virtual Warehouse on the CDW UI.

Use CDP CLI to create or edit a Virtual Warehouse and specify the --impala-query-log option to enable logging Impala queries.

DWX-19451: Data Visualization restore job can fail with ignorable errors

After a successful Data Visualization restoration job, the Data Visualization restore job could be in a failed state with the log displaying ignorable errors.

```
pg_restore: error: could not execute query: ERROR:  sequence "jo
bs_joblog_id_seq" does not exist
Command was: DROP SEQUENCE public.jobs_joblog_id_seq;
pg_restore: error: could not execute query: ERROR:  table "jobs_j
oblog" does not exist
Command was: DROP TABLE public.jobs_joblog;
pg_restore: error: could not execute query: ERROR:  sequence "
jobs_jobcontent_id_seq" does not exist
Command was: DROP SEQUENCE public.jobs_jobcontent_id_seq;
.......
.......
```

This issue occurs because the restore job issues commands to DROP all the objects that will be restored, and if any of these objects do not exist in the destination database, such ignorable errors are reported.

This has no functional impact on the restored Data Visualization application. It is noticed that all the backed up queries, datasets, connections, and dashboards are restored successfully and Data Visualization is available for new queries.

None.

Behavior changes

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following behavior changes:

Summary: Selecting a resource template replaces configuring Cloudera Data Visualization T-shirt sizes

Before this release: You could configure the following sizes for your Cloudera Data Visualization instance:

- Small (default), 8Gb
- Medium, 16Gb
- Large, 24Gb

After this release: You can configure the following CDW resource template for your Cloudera Data Visualization instance:

- Default resources
- Medium resources
- Large resources

July 26, 2024 Release Notes

Review the new feaures, fixes, behavioral changes, and preview features in this release of Cloudera Data Warehouse on Public Cloud.

What's new in Cloudera Data Warehouse on Public Cloud

Review the new features introduced in this release of Cloudera Data Warehouse (CDW) service on CDP Public Cloud.

- CDW features
- CDW on Azure environments
- CDW on AWS environments
- Iceberg
- Hue
- Technical Preview features
- Behavior changes

What's new in CDW Public Cloud

General availability of Virtual Warehouse and Database Catalog workload version selections

The CDW UI now provides a list of workload versions that match your cluster from which you can select one during cluster installation. The Database Catalog list contains versions compatible with your Kubernetes version and your cluster environment (DWX version). The Virtual Warehouse list contains versions compatible with your Kubernetes version, your cluster environment (DWX version), and your Database Catalog version.

General availability of Impala workload-aware autoscaling

Workload-aware autoscaling allocates Impala Virtual Warehouse resources based on the workload that is running. You choose multiple executor group sets size based on your workload requirements, instead of the fixed executor group size of the previous auto-scaling implementation. This feature is now generally available. See Workload Aware Auto-Scaling in Impala.

Improved Impala Autoscaling Dashboard

You can now use the new Impala Autoscaling Dashboard to monitor Impala autoscaling in a warehouse that uses workload-aware autoscaling or the regular autoscaling. You can access the Impala Autoscaling Dashboard by going to the **Virtual Warehouse Details** page's Web UI tab, and clicking the Impala Autoscaler Web UI option. See About the Impala Autoscaling Dashboard.

Ability to forward Prometheus metrics from CDW to an external endpoint

In this release, you can configure Prometheus in CDW to push its metrics to an external endpoint, such as Prometheus, Grafana, Thanos, or some other endpoint. See Forwarding Prometheus metrics from CDW to an endpoint.

Automatically backing up and restoring CDW

This release adds more automation to back up and restore procedures for AWS and Azure environments and clarifies the documentation of the automatic, semi-automatic, and manual procedures.

To get the supported Kubernetes version for this release, you back up your old AWS or Azure environment and start up a new environment using the restoration process. The backup/restore feature saves your environment parameters, making it possible to recreate your environment with the same settings, URL, and connection strings you used in your previous environment.

Ability to configure Impala Statestore high availability

You can now configure high availability for Impala Statestore pods in a Virtual Warehouse, with active and passive modes ensuring continuity and reliability during failovers. See Configuring Impala Statestore high availability.

Downloading the UDF development package from CDW UI

Introducing the ability to download the Impala UDF development package directly from the CDW UI for enhanced convenience and integration, see Building and deploying UDFs.

PostgreSQL replaces SQLite database for Grafana in Cloudera Data Warehouse on Public Cloud

The file-based SQLite database for Grafana has been replaced with PostgreSQL database, providing a more robust experience. You must deactivate and reactivate your environment in CDW to use this feature.

What's new in CDW on Azure environments

Azure AKS 1.29 upgrade

Cloudera supports the Azure Kubernetes Service (AKS) version 1.29. In 1.9.1-b233 (released July 26, 2024), when you activate an environment, CDW automatically provisions AKS 1.29. To upgrade to AKS 1.29 from an earlier version of CDW, you must backup and restore CDW. To avoid compatibility issues between CDW and AKS, upgrade to version 1.29.

Note: Using the Azure CLI or Azure portal to upgrade the AKS cluster is not supported. Doing so can cause the cluster to become unusable and can cause downtime. For more information about upgrading, see Upgrading an Azure Kubernetes Service cluster for CDW.

Addition of new Azure instance types

This release offers the selection of the Standard_E16pds_v5 Azure Virtual Machine, an AKS Ampere® Altra® Arm-based instance type for an Impala Virtual Warehouse. For more information about using the instance type, see Activating an Azure environment from CDW.

CDW provisions Azure Database for PostgreSQL - Flexible Server

Starting with this release, CDW provisions Azure Database for PostgreSQL - Flexible Server instead of Azure Database for PostgreSQL - Single Server. See Enabling a private CDW environment.

What's new in CDW on AWS environments

Amazon EKS 1.29 upgrade

Cloudera supports the Amazon Elastic Kubernetes Service (EKS) version 1.29. In 1.9.1-b233 (released July 26, 2024), when you activate an environment, CDW automatically provisions EKS 1.29. To upgrade to EKS 1.29 from an earlier version of CDW, you must backup and restore CDW. To avoid compatibility issues between CDW and EKS, upgrade to version 1.29. See Upgrading Amazon Kubernetes Service (EKS).

Note about the impact of AWS RDS root certificate rotation in 2024

A CDW Cluster RDS does not use certificate verification for connections to the CDW. Therefore you are not directly impacted by certificate expiration for your CDW Cluster RDS. You can either choose to clear the warnings or rotate the certificate.

To rotate the certificate for the CDW Cluster RDS, follow the step outlined by AWS in Rotating your SSL/TLS certificate to update the certificate. There should be no impact on CDW because the CDW Cluster RDS should not be restarted, Postgres RDS has SupportsCertificateRotationWithoutRestart=true.

For the Datalake RDS, follow instructions shared by the Datalake account team to update the certificate. There maybe some impact to CDW while restarting the Datalake, such as query failures or delays. This could happen because services such as Ranger, Knox, and FreeIPA might be unavailable during this period.

Addition of new AWS instance types

This release offers the selection of the r6gd.4xlarge and r7gd.4xlarge Arm-based instance types for an Impala Virtual Warehouse. For more information about using the instance type, see Activating an AWS environment from CDW.

Ability to use envelope encryption for EKS secrets

Envelope encryption is now added for EKS Secrets through CDW KMS Key by default. See Encrypt Kubernetes secrets with AWS KMS on existing clusters.

What's new in Iceberg on CDW Public Cloud

CDP support for Iceberg version 1.4.3

The Apache Iceberg component has been upgraded from 1.3.0 to 1.4.3.

Support for Iceberg data compaction

You can compact Iceberg tables and optimize them for read operations from Hive and Impala. Compaction is an essential table maintenance activity that creates a new snapshot, which contains the table content in a compact form. See Iceberg data compaction.

SQL support for querying Iceberg metadata tables

Apache Iceberg stores extensive metadata for its tables. From Hive and Impala, you can query the metadata tables as you would query a regular table. For example, you can use projections, joins, filters, and so on. See Query metadata tables feature.

What's new in Hue on CDW Public Cloud

General availability (GA) of the SQL AI Assistant

Hue leverages the power of Large Language Models (LLM) to help you generate SQL queries from natural language prompts and also provides options to optimize, explain, and fix queries, promoting efficient and accurate practices for accessing and manipulating data. You can use several AI services and models such as OpenAI's GPT service, Amazon Bedrock, and Azure's OpenAI service to run the Hue SQL AI assistant.

- To learn more about the supported models and services, limitations, and what data is shared with the LLMs, see About the SQL AI Assistant in CDW.
- To set up and enable the SQL AI Assistant, see About setting up the SQL AI Assistant in CDW.
- To see how to generate, edit, explain, optimize, and fix queries, see Starting the SQL AI Assistant in Hue.

Introduction of task server in Hue and significant improvement in the file upload functionality

A new **Task Server** page has been added to the Hue web interface. The Hue task server enables the following functionalities:

- It improves the file-upload experience, allowing you to upload multiple files up to 5 GB each in parallel.
- It helps you to schedule tasks to clean up Hue documents and the /tmp directory, improving cluster maintenance experience and performance.

See About the Hue task server in CDW.

Related Information

Preview features in Cloudera Data Warehouse on Public Cloud

Preview features in Cloudera Data Warehouse on Public Cloud

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces this technical preview.

Note: Technical previews are considered under development. Do not use these features in production environments.

Enabling the Hive Virtual Warehouse to spill to an EBS volume (Preview)

To prevent failures when query data exceeds memory capacity, you spill data to an EBS volume. The data spills to the Amazon gp3 Elastic Block Store (EBS) volumes. You select the Additional LLAP Spill Disk (EBS) option when you create a Hive Virtual Warehouse. CDP automatically provisions the gp3 volume type for spilling Hive queries when you create or reactivate a Hive Virtual Warehouse on the latest CDW environment. For more information about EBS volumes, see Amazon documentation. Using the EBS volume incurs cost.

Note: You cannot enable the option to spill data to EBS volume after creating a Virtual Warehouse.

Improvements to the shared Hue service (Preview)

- Name change from Query Editor to Shared Hue Service in the left navigation pane in the CDW UI.
- Shared Hue service supports upgrade and rebuild operations similar to other CDW components.
- Added a one-time option to copy saved queries and query history while creating a shared Hue service instance.

For more information, see Deploying shared Hue service in Data Warehouse Public Cloud (Preview).

Ability to log and manage Impala workloads (Preview)

CDW provides you the option to enable logging Impala queries on an existing Virtual Warehouse or while creating a new Impala Virtual Warehouse. By logging the Impala queries in Cloudera Data Warehouse (CDW), you gain increased observability of the workloads running on Impala, which you can use to improve the performance of your Impala Virtual Warehouses.

This feature represents a significant enhancement to query profiling capabilities. You can have Impala archive crucial data from each query's profile into dedicated database tables known as the query history table and live query table. These tables are part of the sys database and are designed to store valuable information that can later be queried using any Impala client, providing a consolidated view of reports from previously executed queries.

For more information, see Impala workload management in Data Warehouse Public Cloud (Preview).

Introducing AI-enhanced UDF development package in Impala (Preview)

- A built-in AI function, ai_generate_text, enabling direct access to Large Language Models (LLMs) from SQL queries by inputting a prompt and retrieving the response.
- This integration into existing workflows simplifies the process, reducing complexity and enhancing the user experience, allowing for quicker setup and deployment of UDFs in Impala.

For more information, see Advantages and use cases of Impala AI functions (Preview).

Support for Impala external JDBC data sources (Preview)

Apache Impala now supports reading from external JDBC data sources. An external JDBC table represents a table or a view in a remote RDBMS database or another Impala cluster. Using external JDBC tables, you can connect Impala to a database, such as MySQL, PostgreSQL, or another Impala cluster and read the data in the remote tables.

For more information, see Using Impala to query external JDBC data sources (Preview).

AWS environment permissions support for the EKS start/stop feature (Preview)

AWS permissions have been expanded to support the Elastic Kubernetes Service (EKS) start/stop feature:

- rds:StartDBInstance
- rds:StopDBInstance
- rds:DescribeDBInstances
- autoscaling:DescribeAutoScalingGroups

Ability to select an instance type for Virtual Warehouses (Preview)

You can now specify AWS or Azure instance types, such as r6id.4xlarge or Standard_E16_v3, that you want to use for your Virtual Warehouse while creating a Virtual Warehouse. You are no longer confined to use the instance types that were specified while activating the environment in CDW. See Creating a Virtual Warehouse with ARM compute instance types using CDP CLI.

Note:

- In CDW 1.9.1-b233 (released July 26, 2024), instance type selection while creating a Virtual Warehouse is only supported using BETA CDP CLI. See CDP CLI (BETA) documentation. You must install and use BETA CDP CLI version 0.9.119.
- You can continue to select an instance type while activating an environment in CDW using both the UI and CDP CLI. The ability to do this will be deprecated in a future release.

Impala support for reading Iceberg equality deletes for NiFi (Preview)

Cloudera supports row-level deletes, and starting with this release you can read equality deletes from Impala with suport added for Apache NiFi. See the Delete data feature.

Related Information

What's new in Cloudera Data Warehouse on Public Cloud

Cloudera preview features documentation

Known issues in Cloudera Data Warehouse on Public Cloud

Review the known issues in this release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud.

- Known issues identified in the July 26, 2024 release on page 35
- Known issues identified before the July 26, 2024 release on page 37
- Technical Service Bulletins on page 48

Known issues identified in the July 26, 2024 release

DWX-19451: Data Visualization restore job can fail with ignorable errors

After a successful Data Visualization restoration job, the Data Visualization restore job could be in a failed state with the log displaying ignorable errors.

```
pg_restore: error: could not execute query: ERROR:  sequence "jo
bs_joblog_id_seq" does not exist
Command was: DROP SEQUENCE public.jobs_joblog_id_seq;
pg_restore: error: could not execute query: ERROR:  table "jobs_j
oblog" does not exist
Command was: DROP TABLE public.jobs_joblog;
pg_restore: error: could not execute query: ERROR:  sequence "
jobs_jobcontent_id_seq" does not exist
Command was: DROP SEQUENCE public.jobs_jobcontent_id_seq;
.......
.......
```

This issue occurs because the restore job issues commands to DROP all the objects that will be restored, and if any of these objects do not exist in the destination database, such ignorable errors are reported.

This has no functional impact on the restored Data Visualization application. It is noticed that all the backed up queries, datasets, connections, and dashboards are restored successfully and Data Visualization is available for new queries.

None.

DWX-19003: Unable to set t-shirt size for a Cloudera Data Visualization instance

You cannot set or configure a t-shirt size for a Cloudera Data Visualization instance while creating or editing it from CDW. By default, any Cloudera Data Visualization instance created in CDW release version 1.9.1 uses the “small” t-shirt size.

None.

DWX-18950: Hue data restoration fails on Azure CDW clusters having CDW version 1.8.7 or earlier

If you backup your CDW cluster running CDW version 1.8.7 or earlier and if you perform the restore operation without reactivating the latest CDW environment, you might encounter Hue restoration job failure.

None.

DWX-18843: Unable to read Iceberg table from Hive Virtual Warehouse

If you have used Apache Flink to insert data into an Iceberg table that is created from Hive, you cannot read the Iceberg table from the Hive Virtual Warehouse.

Add the engine.hive.enabled table property through the Hive beeline and set the value to "true". You can add this table property either while creating the Iceberg table or use the ALTER TABLE statement to add the table property.

DWX-18489: Hive compaction of Iceberg tables results in a failure

When Cloudera Data Warehouse (CDW) and Data Hub are deployed in the same environment and use the same Hive Metastore (HMS) instance, the Data Hub compaction workers can inadvertently pick up Iceberg compaction tasks. Since Iceberg compaction is not yet supported in the latest Data Hub version, the compaction tasks will fail when they are processed by the Data Hub compaction workers.

In such a scenario where both CDW and Data Hub share the same HMS instance and there is a requirement to run both Hive ACID and Iceberg compaction jobs, it is recommended that you use the CDW environment for these jobs. If you want to run only Hive ACID compaction tasks, you can choose to use either the CDW or Data Hub environments.

If you want to run the compaction jobs without changing the environment, it is recommended that you use CDW. To avoid interference from Data Hub, change the value of the hive.compactor.worker.threads Hive Server (HS2) property to '0'. This ensures that the compaction jobs are not processed by Data Hub.

1. In Cloudera Manager, click Clusters Hive Configuration to navigate to the configuration page for HMS.
2. Search for hive.compactor.worker.threads and modify the value to '0'.
3. Save the changes and restart the Hive service.

DWX-18854: Compaction cleaner configuration

The compaction cleaner is turned off by default in the CDW database catalog, potentially causing compaction job failures.

To enable the compactor on the Hive Metastore (HMS) instance of the CDW database catalog, set the hive.compactor.cleaner.on property to "true".

DWX-12703: Hue connects to only one Impala coordinator in Active-Active mode

You may not see all Impala queries that have run on the Virtual Warehouse from the **Impala** tab on the Hue **Job Browser**. You encounter this on an Impala Virtual Warehouse that has Impala coordinator configured in an active-active mode. This happens because Hue fetches this information from only one Impala coordinator that is active.

None. You can view the query history from the **Impala Queries** tab on the **Job Browser** page, because the information is fetched from the Hue Query Processor.

Known limitation: CDW does not support S3 Express One Zone buckets

Cloudera does not recommend deploying the Data Lake on S3 Express One Zone buckets. CDW cannot read content present in the S3 Express One Zone buckets. The following limitations apply when using S3 Express buckets:

- You can only use S3 Express buckets with Data Hubs running Runtime 7.2.18 or newer. Data services do not support it, currently.
- S3 Express buckets may not be used for logs and backups.

Known issues identified before the July 26, 2024 release

DWX-15112: Enterprise Data Warehouse database configuration problems after a Helm-related rollback

If a Helm rollback fails due to an incorrect Enterprise Data Warehouse database configuration, the Virtual Warehouse and Database Catalog roll back to a previous configuration. The incorrect Enterprise Data Warehouse configuration persists, and can affect subsequent edit, upgrade, and rebuild operations on the rolled-back Virtual Warehouse or Database Catalog.

None.

DWX-17455: ODBC client using JWT authentication cannot connect to Impala Virtual Warehouse

If you are using the Cloudera ODBC Connector and JWT authentication to connect to a CDW Impala Virtual Warehouse where the Impala coordinators are configured for high availability in an active-active mode, the connection results in a "401 Unauthorized" error. The error is not seen on Impala Virtual Warehouses that have an active-passive high availability or where high availability is disabled.

In the ODBC Data Source Name (DSN), make the following changes:

1. Ensure that SSL is enabled by setting SSL=1.
2. Remove authentication by setting AuthMech=0.
3. Remove the JWTString parameter.
4. Add the parameter http.header.Authorization and set it to the word "Bearer" followed by a space and the JWT string

DSN configuration before the workaround

```
SSL=1
AuthMech=8
JWTString=full_jwt_text
```

DSN configuration after the workaround

```
SSL=1
AuthMech=0
http.header.Authorization=Bearer full_jwt_text
```

HIVE-28055: Merging Iceberg branches requires a target table alias

Hive supports only one level of qualifier when referencing columns. In other words only one dot is accepted. For example, select table.col from ...; is allowed. select db.table.col is not allowed. Using the merge statement to merge Iceberg branches without a target or source table alias causes an exception:

```
org.apache.hadoop.hive.ql.parse.SemanticException: ... Invalid t
able alias or column reference ...
```

Use an alias, for example t, for the target table.

```
merge into mydb.target.branch_branch1 t using mydb.source.branch
_branch1 s on t.id = s.id when matched then update set value = '
matched';
```

DWX-17620: Folder having special characters in its name is not accessible in ABFS

In Cloudera Data Warehouse Public Cloud, from a Virtual Warehouse going to the Azure Blob Filesystem (ABFS), creating a folder, and then performing an action such as Move, causes an error as shown in the following example:

```
Cannot access: abfs://data-files/user/hrt_qa/~@$&()*!+'=;. 404 C
lient Error: The specified path does not exist.
```

None.

Branch FAST FORWARD does not work as expected

The Apache Iceberg spec indicates you can use either one or two arguments to fast forward a branch. The following example shows using two arguments:

```
ALTER TABLE <name> EXECUTE FAST-FORWARD 'x' 'y'
```

However, omitting the second branch name, does not work as documented by Apache Iceberg. The named branch is not fast-forwarded to the current branch. An exception occurs at the Iceberg level.

You must use two arguments to the EXECUTE FAST FORWARD feature to forward a branch.

DWX-17613: Generic error message is displayed when you click on the directory you don't have access to on a RAZ cluster

You see the following error message when you click on an ABFS directory to which you do not have read/write permission on the ABFS File Browser in Hue: There was a problem with your request. This message is generic and does not provide insight into the actual issue.

None.

DWX-17109: ABFS File Browser operations failing intermittently

You may encounter intermittent issues while performing typical operations on files and directories on the ABFS File Broswer, such as moving or renaming files.

None.

CDPD-27918: Hue does not automatically pick up RAZ HA configurations

On a CDP Public Cloud environment in which you have configured RAZ in High Availability mode, Hue in CDW does not pick up all the RAZ host URLs automatically. Therefore, if a RAZ instance to which Hue is connected goes down, Hue becomes unavailable.

You must manually add comma-separated RAZ instances in the Hue Advanced Configuration Snippet.

1. Log in to the CDP Management Console as an Administrator.
2. Go to Environment Data Lake and open Cloudera Manager for your environment.
3. Go to Clusters Ranger RAZ service Instances RAZ server Processes and note the value of the fs.s3a.ext.raz.rest.host.url property from the core-site.xml file. You need this to specify the value of the api_url property in the Hue configuration.

   For Azure environments, note the value of the fs.azure.ext.raz.rest.host.url property.

   For AWS and GCS environments, note the value of the fs.s3a.ext.raz.rest.host.url property.
4. Go to CDW Virtual Warehouse ⋮ Edit CONFIGURATIONS and select the hue-safety-valve from the Configuration files dropdown menu.
5. Add the following lines in the hue-safety-valve field:

   ```
   [desktop]
   [[raz]]
   is_enabled=true
   ```

```
api_url=https://[***INSTANCE-1***]:6082/,https://[***INSTANCE-
2***]:6082/
```

6. Click Apply Changes.

CDPD-66779: Partitioned Iceberg table not getting loaded with insert select query from Hive

If you create a partitioned table in Iceberg and then try to insert data from another table as shown below, an error occurs.

```
insert into table partition_transform_4 select t, ts from vector
tab10k;
```

Use the CLUSTER BY clause on the partitioned column to insert data. For example:

```
insert into table partition_transform_4 select t, ts from t1 clu
ster by ts;
```

DWX-17703: Non-HA Impala Virtual Warehouse on a private Azure Kubernetes Service (AKS) setup fails

When 'Refresh' and 'Stop' operations run in parallel, Impala might move into an error state. The Refresh operation might think that Impala is in an error state as the coordinator pod is missing.

Rebuild the Impala Virtual Warehouse or restart it using the CLI.

DWX-14923: After JWT authentication, attempting to connect the Impyla client using a user name or password should cause an error

Using a JWT token, you can connect to a Virtual Warehouse as the user who generated the token.

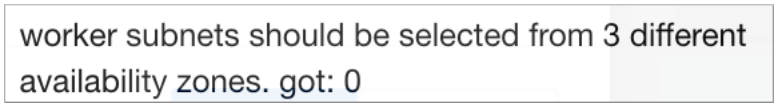
If you connect with the JWT token, and then pass a user name or password from Impyla to the Virtual Warehouse, the connection arguments are silently ignored. Such an action should indicate that it is illegal to specify user or password when using JWT authentication.

None.

DWX-15145: Environment validation popup error after activating an environment

Activating an environment having a public load balancer can cause an environment validation popup error.

To reproduce this problem 1) Create a data lake. 2) Activate an environment having a public load balancer deployment type and subnets in three different availability zones. A environment validation popup can occur even through subnets are in different availability zones. Several different popups can occur, including the following one:



None.

DWX-15144: Virtual Warehouse naming restrictions

You cannot create a Virtual Warehouse having the same name as another Virtual Warehouse even if the like-named Virtual Warehouses are in different environments. You can create a Database Catalog having the same name as another Database Catalog if the Database Catalogs are in different environments.

None.

DWX-13103: Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data

Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Restart the prometheus pod so that it gets the istio-proxy container.

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the CreateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersions line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Incorrect diagnostic bundle location

The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

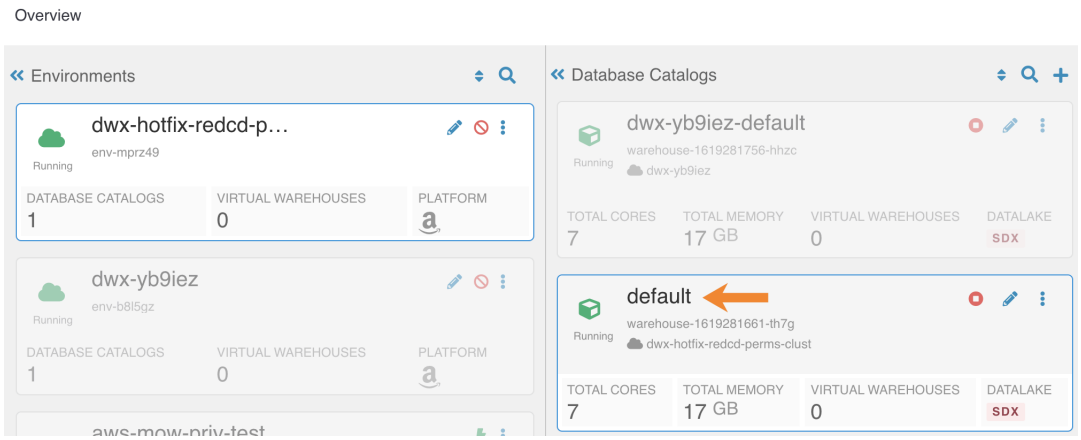
To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

None.

DWX-15064: Hive Virtual Warehouse stops but appears healthy

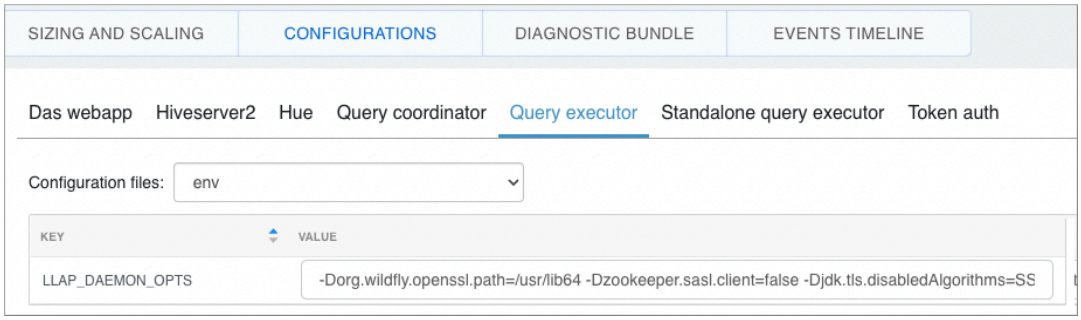
Due to an istio-proxy problem, the query coordinator can unexpectedly enter a not ready state instead of the expected error-state. Subsequently, the Hive Virtual Warehouse stops when reaching the autosuspend timeout without indicating a problem.

None.

DWX-14452: Parquet table query might fail

Querying a table stored in Parquet from Hive might fail with the following exception message: java.lang.RuntimeException: java.lang.StackOverflowError. This problem can occur when the IN predicate has 243 values or more, and a small stack (-Xss = 256k) is configured for Hive in CDW.

Change the value of the Xss in the JVM args to use the default value (1Mb) as follows: Select Edit from the Options menu of your Virtual Warehouse. Click CONFIGURATIONS Query executor and select the env configuration file.



In the third line shown below, change the value of LLAP_DAEMON_OPTS from -Xss256k to -Xss1M, and then click Apply Changes:

FROM:

-Dorg.wildfly.openssl.path=/usr/lib64 -Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Dhttp.maxConnections=12 -Xms24G -Xmx48G -Xss256k ...

TO:

... -Xss1M ...

CDPD-40730: Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

None.

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.

7. Click Apply in the upper right corner of the page to save the configuration.
8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name ‘ec2-18-219-
32-183.us-east-2.compute.amazonaws.com’ does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

DWX-16899: Error while viewing Impala job status on the mini Job Browser

When you click on the application ID after submitting an Impala query on Hue that is running on the environment level, you may notice the following error: 401 Client Error: Unauthorized for url: http://coordinator.impala-xyz.svc.cluster.local:25000/queries?json=true Must authenticate with Basic authentication. This does not happen when you do the same from Hue deployed at a Virtual Warehouse level.

None.

DWX-16895: Incorrect status of Hue pods when you edit the Hue instance properties

When you update a configuration of a Hue instance that is deployed at the environment level, such as increasing or decreasing the size of the Hue instance, you see a success message on the CDW UI. After some time, the status of the Hue instance also changes from “Updating” to “Running”. However, when you list the Hue pods using kubectl, you see that not all backend pods are in the running state–a few of them are still in the init state.

None. The pods come up successfully eventually after a sufficient time has passed.

DWX-16863: The upgrade button is present on the CDW UI, but Hue upgrades are not supported

You see the Upgrade button on the **Query Editor** page in the CDW UI when Hue is deployed at the environment level. However, on CDW version 1.8.1, upgrading the Hue instance that is deployed at the environment level is not supported.

None.

DWX-16893: A user can see all the queries in Job browser

In a Hue instance deployed at the environment level, by design, the Hue instances must not share the saved queries and query history with other Hue instances even for the same user. However, a logged in user is able to view all the queries executed by that user on all the Virtual Warehouses on a particular Database Catalog.

None.

Delay in listing queries in Impala Queries in the Job browser

Listing an Impala query in the Job browser can take an inordinate amount of time.

None.

DWX-14927: Hue fails to list Iceberg snapshots

Hue does not recognize the Iceberg history queries from Hive to list table snapshots. For example, Hue indicates an error at the . before history when you run the following query.

```
select * from <db_name>.<table_name>.history
```

None.

DWX-14968: Connection termination error in Impala queries tab after Hue inactivity

To reproduce the problem: 1) In the Impala job browser, navigate to Impala queries. 2) Wait for a few minutes.

After a few minutes of inactivity, the follow error is displayed:

Refresh the page, or alternatively start a new session.

DWX-15115: Error displayed after clicking on hyperlink below Hue table browser

In Hue, below the table browser, clicking the hyperlink to a location causes an HTTP 500 error because the file browser is not enabled for environments that are not Ranger authorized (RAZ).

None.

DWX-15090: CSRF error intermittently seen in the Hue Job Browser

You may intermittently see the “403 - CSRF” error on the Hue web interface as well as in the Hue logs after running Hive queries from Hue.

Reload the browser or start a new Hue session.

IMPALA-11447: Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

DWX-17210, DWX-13733: Timeout issue querying Iceberg tables from Hive

When querying Iceberg tables from Hive, the queries can faile due to a timeout issue.

1. Add the following configurations to hadoop-core-site for the Database Catalog and the Virtual Warehouse.
   - fs.s3.maxConnections=1000
   - fs.s3a.connection.maximum=1000
2. Restart the Database Catalog and Virtual Warehouse.

DWX-15014: Loading airports table in demo data fails

This issue occurs only when using a non-default Database Catalog. A invalid path error message occurs when using the load command to load the demo data airports table in Iceberg, ORC, or Parquet format.

None.

DWX-14163: Limitations reading Iceberg tables in Avro file format from Impala

The Avro, Impala, and Iceberg specifications describe some limitations related to Avro, and those limitations exist in CDP. In addition to these, the DECIMAL type is not supported in this release.

None.

DEX-7946: Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata if you drop the table during migration from Hive to Iceberg.

Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13062: Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276: Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

IMPALA-11045: Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

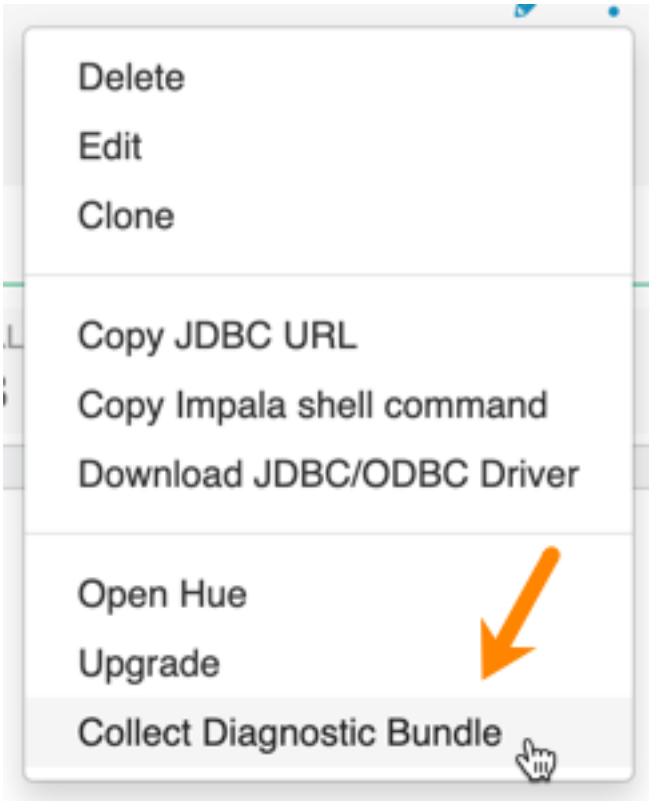
DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



None.

Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

None.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

None.

Technical Service Bulletins

TSB 2023-719: Cloudera Data Warehouse Backup/Restore of CDP Data Visualization incomplete

Cloudera Data Warehouse (CDW) customers using the CDW Automated Backup/Restore feature will encounter an issue with the restoration versions of Cloudera Data Visualization (CDV) older than CDV 7.1.6.2-3 due to a schema change in this release. If the Backup was taken from an older CDW environment that contained a version of CDV older than CDV 7.1.6.2-3, the Restore procedure will succeed. Though once the user opens the CDV Queries tab, the user could encounter the error message: “column jobs_jobschedule.owner_id does not exist…”

None. For the latest update on this issue see the corresponding Knowledge article: TSB 2023-719: Cloudera Data Warehouse Backup/Restore of Cloudera Data Visualization incomplete.

TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

The Cloudera Data Warehouse (CDW) Public Cloud 2023.0.14.0 (DWX-1.6.3) version incorporated a feature for performing automatic metadata synchronization across multiple Apache Impala (Impala) Virtual Warehouses. The feature is enabled by default, and relies on the Hive MetaStore events. When a certain sequence of Data Definition Language (DDL) SQL commands are executed as described below, users may encounter a java.lang.NullPointerException (NPE). The exception causes the event processor to stop processing other metadata operations.

If a CREATE TABLE command (not CREATE TABLE AS SELECT) is followed immediately (approximately within 1 second interval) by INVALIDATE METADATA or REFRESH TABLE command on the same table (either on the same Virtual Warehouse or on a different one), there is a possibility that the second command will not find the table in the catalog cache of a peer Virtual Warehouse and generate an NPE.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

Fixed issues in Cloudera Data Warehouse on Public Cloud

Review the issues fixed in this release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud.

DWX-17794: Custom tags specified on a Virtual Warehouse do not work

Earlier, when you added tags while creating a Virtual Warehouse, the tags did not show on the AWS console. This issue has been fixed.

DWX-15859: Issue backing up and restoring Workload Aware Auto-Scaling (WAAS)

You could not use the CDW backup and restore process to restore the WAAS configurations and you had to manually recreate your Virtual Warehouse by setting the right configurations. This issue has been fixed.

DWX-17620: The folder having special characters in its name is not accessible in ABFS

Earlier, creating a folder and then performing an action such as Move from a Virtual Warehouse going to the Azure Blob Filesystem (ABFS) caused an “404 Client Error: The specified path does not exist” error. This issue has been resolved.

HIVE-28006: Incremental rebuild accuracy for materialized views

Optimized the accuracy of representing incremental rebuild availability for materialized views containing aggregate functions in their definition query.

HIVE-27775: Fix for partition metadata fetch issue during DST shift

Fixes fetching partition metadata from Hive Metastore using Java Data Objects (JDO) by timestamp now does not provide incorrect results in daylight saving time (DST) shift in partition pruning.

HIVE-27924: Materialized view rebuild issue with delete operations

Addressed data correctness issues when rebuilding a materialized view incrementally in scenarios involving delete operations in one of its source tables. This is evident when the materialized view definition query includes joins and aggregates.

IMPALA-12809: Metadata query problem from Impala

When querying metadata from Impala on a cluster with more than a few nodes, the query might fail with a null pointer exception.

The fix addresses issues noticed while querying metadata from Impala on a cluster with more than a few nodes. The change forces Iceberg metadata scanner fragments to be scheduled on the coordinator. A new flag is introduced in the TPlanFragment thrift struct and if the value is 'true', then the fragment is always scheduled on the coordinator.

IMPALA-12742: DELETE/UPDATE Iceberg table partitioned by DATE fails

Fixed data correctness issue by ensuring that Impala CatalogD is able to corretcly parse identity-partitioned DATE values.

IMPALA-13138: Never smallify existing StringValue objects, only new ones during DeepCopy

This issue has been fixed.

Technical Service Bulletins

TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion.

Behavior changes in Cloudera Data Warehouse on Public Cloud

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following behavior changes:

Summary: Changes to the Azure Kubernetes Service (AKS) managed identities

Before this release: Hive inherited the Storage Blob Data Owner role from the Data Lake.

After this release: You must explicitly grant the Storage Blob Data Owner role to the AKS identity on your Data Lake storage account. This permission is required for activating an environment in CDW.

Summary: Overlay Networking replaces kubenet in Azure environments

Before this release: To avoid IP address exhaustion, you enable the kubenet networking feature when you activate an Azure environment to use with CDW.

After this release: To avoid IP address exhaustion, by default, CDW uses CNI overlay networking and replaces kubenet networking. You use the CDP CLI to set CIDR (Classless Inter-Domain Routing) ranges. For more information, see Overlay networking.

Summary: Selecting a resource template replaces configuring Database Catalog T-shirt sizes

Before this release: You could configure the following sizes of the Java Heap for your Database Catalog workload:

- Small (default), 8Gb
- Medium, 16Gb
- Large, 24Gb

After this release: You can configure a CDW resource template, which fulfills all requirements for a Database Catalog for the type of Virtual Warehouse you have, for example a Hive or Impala Virtual Warehouse:

- Default resources
- Medium resources
- Large resources

For more information, see Creating a Database Catalog.

Summary: Impala Autoscaling Dashboard replaces the older Impala Autoscaler Web UI

Before this release: The Impala Autoscaler Web UI provided the tools to monitor how Impala Virtual Warehouses autoscaled.

After this release: You can monitor Impala autoscaling in a warehouse that uses workload-aware autoscaling or the regular autoscaling using the new Impala Autoscaling Dashboard.

Summary: Apache Iceberg component version

Before this release: The supported Iceberg version was 1.3.0.

After this release: The supported Iceberg version is 1.4.3.

Summary: CDW_VERSIONED_DEPLOY entitlement

Before this release: To upgrade Azure Kubernetes Service for CDW to certain workload versions, you needed the CDW_VERSIONED_DEPLOY entitlement.

After this release: The status of Virtual Warehouse and Database Catalog workload version selections is now General availability (GA). The entitlement is no longer needed. For more information, see What's New in this release.

Summary: Uploading files to object store using Hue

Before this release: The Hue server performed the file upload operations. The maximum file size you could upload to S3 buckets or ADLS Gen2 containers was 2 GB. When multiple users performed major upload operations, it impacted the performance of Hue since all the backend calls were synchronous. The Hue server became slow as the upload blocked Hue services. You may have worked around it by disabling the upload from the main Hue server and having another setup with upload enabled and redirecting the end users wanting to upload files to the second instance.

After this release: The files are uploaded using an asynchronous task queue or job queue in Hue. This improves performance and allows you to upload files as large as 5 GB.

Summary: Change in the option name to enable Hue at the environment level

Before this release: The option to enable the shared Hue service (Hue at environment level) from the **Environment Details** page was called Enable Query Editor.

After this release: The option has been renamed to Enable Shared Hue Service.

Summary: Upgrading and rebuilding the shared Hue service

Before this release: In CDW version 1.8.1, you could not upgrade the Hue instances that are deployed at the environment level.

After this release: You can upgrade and rebuild the shared Hue service.

May 2, 2024 - Hotfix

Review the fixed issues and changed behaviors in this hotfix release of Cloudera Data Warehouse on Public Cloud.

Fixed issues

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Upgrade your Virtual Warehouse to get these hotfixes:

- HIVE-28051 A new housekeeping task is added to clean up local folders on Hive Virtual Warehouse startup and periodically after startup, resolving the issue with disk overflow.

  When a Hive LLAP daemon crashes, it can leave behind unnecessary files in the LLAP local directories. These files have to be cleaned periodically without which Hive queries can fail indicating an invalid disk error exception.

  The following properties are introduced to periodically delete the unnecessary files:

  hive.llap.local.dir.cleaner.cleanup.interval

  Specifies the time interval based on which the LocalDirCleaner service in LLAP daemon checks for stale or old files.

  Default value: 2 hours

  hive.llap.local.dir.cleaner.file.modify.time.threshold

  Specifies the threshold time for the LocalDirCLeaner service. If a file is older than the threshold time, the file is deleted.

  Default value: 24 hours
- IMPALA-12827 Fixes event processing errors when write IDs of an AbortTxnEvent are cleaned up by the HMS cleaner housekeeping threads.
- IMPALA-12831 Fix race condition when a table is being invalidated and updated concurrently.
- IMPALA-12832 Implicitly invalidates a table instead of resulting in an ERROR state during event processing.
- IMPALA-12835 Fix event processor, which is not synching file metadata for non-partitioned ACID tables when incremental refresh on transactional tables is turned off.

  Note:

  - This issue only occurs when hms_event_incremental_refresh_transactional_table is set to 'false'.
  - This issue occurs on non-partitioned tables. Partitioned tables are not affected.
- IMPALA-12851 Fix issue of txnId not being added to tableWriteIds mapping in Catalog.
- IMPALA-12855 Fixes the possibility of encountering a NullPointerException when refreshing a partition that has just been dropped.
- IMPALA-12356 Fix for incorrect identification on self events.
- IMPALA-12969 Fix conditional JVM heap leak in array allocation on deserialization failures.

Behavior changes

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following behavior changes:

Summary:

Change in value for the write.delete.mode property

Before this release:

The value for the write.delete.mode property was set to 'merge-on-read'.

After this release:

The default value for the write.delete.mode property is changed to 'copy-on-write'.

This change might result in issues while deleting Iceberg table records.

If you want to continue to use the 'merge-on-read' mode for new Iceberg tables, perform the following steps:

1. Log in to the CDP web interface and navigate to the Data Warehouse service.
2. From the Data Warehouse service, click Database Catalogs, locate your Database Catalog and then click Edit .
3. In the **Database Catalogs detail** page, click Configurations Metastore and select the hive-site configuration file.
4. Click + and add the following configuration key and key value:

   ```
   write.delete.mode=merge-on-read
   ```
5. Click Apply Changes.
6. Click Virtual Warehouses, locate your virtual warehouse and then click Edit .
7. In the **Virtual Warehouse Details** page, click Configurations Hiveserver2 and select the hive-site configuration file.
8. Click + and add the following configuration key and key value:

   ```
   write.delete.mode=merge-on-read
   ```
9. Click Apply Changes.

If you want to set 'merge-on-read' for older Iceberg tables that were created before upgrading to this CDW version, perform the following steps:

1. Run the following queries to modify the table properties of the old Iceberg tables:

   ```
   ALTER TABLE [***OLD TABLE NAME***] SET TBLPROPERTIES('write.
   on.delete'='dummy');
   ALTER TABLE [***OLD TABLE NAME***] SET TBLPROPERTIES('write
   .on.delete'='merge-on-read');
   ```

Unsupported releases

Overview of new features, enhancements, known issues, fixed issues, and changed behavior introduced in earlier, and now unsupported releases of Cloudera Data Warehouse.

March 26, 2024 - Hotfix

Review the known issues and fixed issues in this hotfix release of Cloudera Data Warehouse on Public Cloud.

Fixed issues

This hotfix release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud fixes these issues.

Note: Re-activate your Virtural Warehouse to get these hotfixes.

IMPALA-12787: Concurrently running UPDATE and DELETE operations can no longer revive deleted rows

This issue is fixed.

Technical Service Bulletins

TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables

TSB 2024-750: CDW activation resulting in inaccessible workload endpoints on AWS

CDW activation no longer creates two cloudformation stacks instead of one. This might result in all the workloads being inaccessible depending on which cloudformation stack finishes first. For the latest update on this issue, see the corresponding Knowledge article: Cloudera Customer Advisory 2024-750: CDW activation through UI can result in inaccessible workload endpoints (AWS only)

TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables.

Known issues

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

See Data Visualization release notes for known issues in Cloudera Data Visualization 7.1.9.

Carried over from the previous release: Upgrade-related

TSB 2023-719: Cloudera Data Warehouse Backup/Restore of CDP Data Visualization incomplete

Cloudera Data Warehouse (CDW) customers using the CDW Automated Backup/Restore feature will encounter an issue with the restoration versions of Cloudera Data Visualization (CDV) older than CDV 7.1.6.2-3 due to a schema change in this release. If the Backup was taken from an older CDW environment that contained a version of CDV older than CDV 7.1.6.2-3, the Restore procedure will succeed. Though once the user opens the CDV Queries tab, the user could encounter the error message: “column jobs_jobschedule.owner_id does not exist…”

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-719: Cloudera Data Warehouse Backup/Restore of Cloudera Data Visualization incomplete

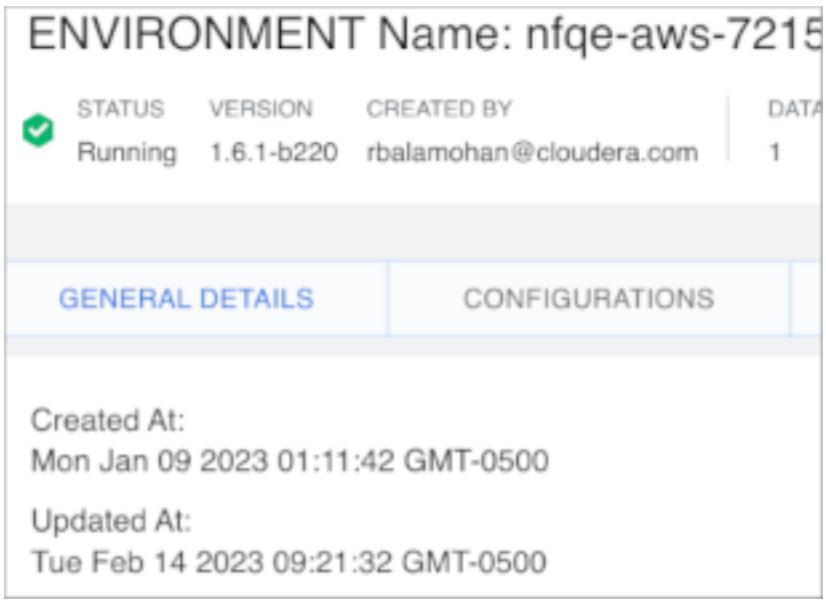
Upgrading to EKS 1.24 could result in Impala coordinators shutting down.

This issue is not seen on the Impala Virtual Warehouse running Runtime 2023.0.15.0-x or later.

Workaround: Manually start the Impala Virtual Warehouse from the UI or cli. Alternatively, replace your runtime with 1.7.1-b755 (released August 30, 2023) or later.

Certain CDW versions cannot upgrade to EKS 1.22

AWS environments activated in version 1.4.1-b86 (released June, 22, 2022) or earlier are not supported for upgrade to EKS 1.22 due to incompatibility of some components with EKS 1.22. To determine the activation version your pre-existing environment, in the Data Warehouse service, expand Environments. In Environments, search for and locate the environment that you want to view. Click Edit. In Environment Details, you see the CDW version.



DWX-15112 Enterprise Data Warehouse database configuration problems after a Helm-related rollback

If a Helm rollback fails due to an incorrect Enterprise Data Warehouse database configuration, the Virtual Warehouse and Database Catalog roll back to a previous configuration. The incorrect Enterprise Data Warehouse configuration persists, and can affect subsequent edit, upgrade, and rebuild operations on the rolled-back Virtual Warehouse or Database Catalog.

Carried over from the previous release: General

DWX-14923 After JWT authentication, attempting to connect the Impyla client using a user name or password should cause an error

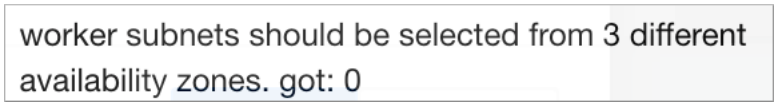
Using a JWT token, you can connect to a Virtual Warehouse as the user who generated the token.

If you connect with the JWT token, and then pass a user name or password from Impyla to the Virtual Warehouse, the connection arguments are silently ignored. Such an action should indicate that it is illegal to specify user or password when using JWT authentication.

DWX-15145 Environment validation popup error after activating an environment

Activating an environment having a public load balancer can cause an environment validation popup error.

To reproduce this problem 1) Create a data lake. 2) Activate an environment having a public load balancer deployment type and subnets in three different availability zones. A environment validation popup can occur even through subnets are in different availability zones. Several different popups can occur, including the following one:



DWX-15144 Virtual Warehouse naming restrictions

You cannot create a Virtual Warehouse having the same name as another Virtual Warehouse even if the like-named Virtual Warehouses are in different environments. You can create a Database Catalog having the same name as another Database Catalog if the Database Catalogs are in different environments.

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data

Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the CreateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersio ns line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

DWX-17703: Non-HA Impala Virtual Warehouse on a private Azure Kubernetes Service (AKS) setup fails

When 'Refresh' and 'Stop' operations run in parallel, Impala might move into an error state. The Refresh operation might think that Impala is in an error state as the coordinator pod is missing.

Rebuild the Impala Virtual Warehouse or restart it using the CLI.

DWX-17109: ABFS File Browser operations failing intermittently

You may encounter intermittent issues while performing typical operations on files and directories on the ABFS File Broswer, such as moving or renaming files.

None.

DWX-17613: Generic error message is displayed when you click on the directory you don't have access to on a RAZ cluster

You see the following error message when you click on an ABFS directory to which you do not have read/write permission on the ABFS File Browser in Hue: There was a problem with your request. This message is generic and does not provide insight into the actual issue.

None.

DWX-17620: Folder having special characters in its name is not accessible in ABFS

In Cloudera Data Warehouse Public Cloud, from a Virtual Warehouse going to the Azure Blob Filesystem (ABFS), creating a folder, and then performing an action such as Move, causes an error as shown in the following example:

```
Cannot access: abfs://data-files/user/hrt_qa/~@$&()*!+'=;. 404 C
lient Error: The specified path does not exist.
```

DWX-15214 DWX-15176 Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

The Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure

(Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed:

```
kubelet Error: failed to create containerd task: failed to creat
e shim task: OCI runtime create failed: runc create failed: unab
le to start container process: exec: “run_httpd.sh”: cannot run
executable found relative to current directory: unknown
```

This issue affects all CDW releases before 1.6.3 (released May 5, 2023)running on CDP Public Cloud on Azure.

The following users are affected: CDW Azure users using Hue in Impala and Hive Virtual Warehouses in environments earlier than version 1.6.3 or later if they upgrade the node image

Workaround: Upgrade the Virtual Warehouses to 1.6.3 and choose Hue version 2023.0.14.0.

Do not choose Hue versions older than 2023.0.14.0 when creating a Virtual Warehouse in environments of version 1.6.3.

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Database Catalog

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

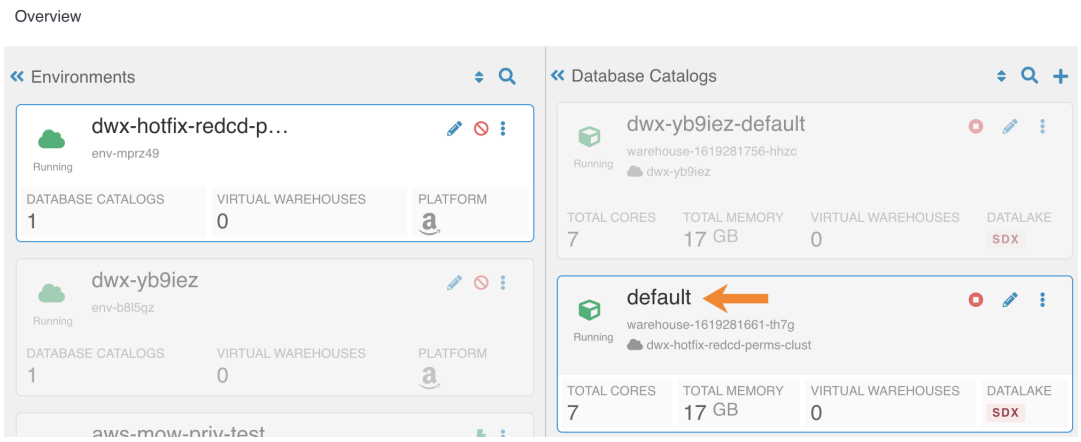
- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

Limited Hive image versions

Hive Virtual Warehouses you create in 1.8.1-b248 (released Nov 20, 2023) and later will run Istio 1.19.0. The new Istio version supports only new versions of Hive helm charts. If you have the CDW_VERSIONED_DEPLOY, only new Hive image versions appear in UI when you create a new Hive Virtual Warehouse.

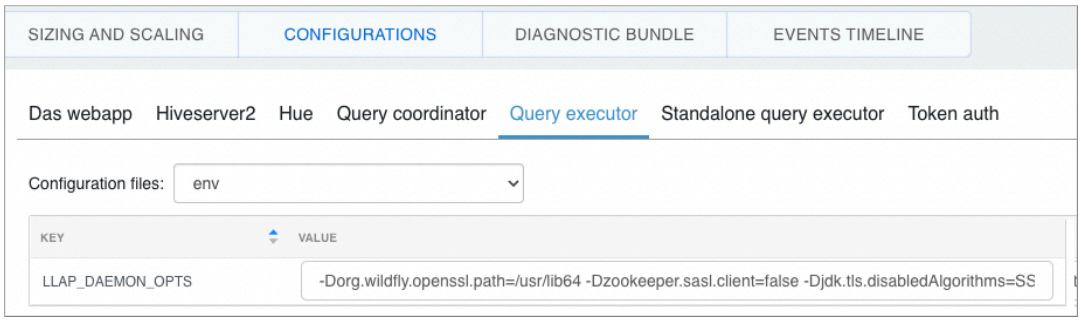
DWX-15064 Hive Virtual Warehouse stops but appears healthy

Due to an istio-proxy problem, the query coordinator can unexpectedly enter a not ready state instead of the expected error-state. Subsequently, the Hive Virtual Warehouse stops when reaching the autosuspend timeout without indicating a problem.

DWX-14452 Parquet table query might fail

Querying a table stored in Parquet from Hive might fail with the following exception message: java.lang.RuntimeException: java.lang.StackOverflowError. This problem can occur when the IN predicate has 243 values or more, and a small stack (-Xss = 256k) is configured for Hive in CDW.

Workaround: Change the value of the Xss in the JVM args to use the default value (1Mb) as follows: Select Edit from the Options menu of your Virtual Warehouse. Click Configurations > Query Executor > and select env.



In the third line shown below, change the value of LLAP_DAEMON_OPTS from -Xss256k to -Xss1M, and then click Apply Changes:

FROM:

-Dorg.wildfly.openssl.path=/usr/lib64 -Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Dhttp.maxConnections=12 -Xms24G -Xmx48G -Xss256k ...

TO:

... -Xss1M ...

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
       kind: ClusterRole
       metadata:
       name: diagnostic-data-generator-role-monitor
       rules:
       - apiGroups:
       - ""
       resources:
       - configmaps
       - events
       - pods
       - persistentvolumeclaims
       - nodes
       verbs:
       - get
       - list
       - apiGroups:
```

```
        - apps
        resources:
        - deployments
        - statefulsets
        verbs:
        - get
        - list
        - apiGroups:
        - "edws.cloudera.com"
        resources:
        - computes
        verbs:
        - get
        - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.

7. Click Apply in the upper right corner of the page to save the configuration.

8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name 'ec2-18-219-
32-183.us-east-2.compute.amazonaws.com' does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Hue Query Editor

CDPD-27918: Hue does not automatically pick up RAZ HA configurations

On a CDP Public Cloud environment in which you have configured RAZ in High Availability mode, Hue in CDW does not pick up all the RAZ host URLs automatically. Therefore, if a RAZ instance to which Hue is connected goes down, Hue becomes unavailable.

You must manually add comma-separated RAZ instances in the Hue Advanced Configuration Snippet.

1. Log in to the CDP Management Console as an Administrator.
2. Go to Environment Data Lake and open Cloudera Manager for your environment.
3. Go to Clusters Ranger RAZ service Instances RAZ server Processes and note the value of the fs.s3a.ext.raz.rest.host.url property from the core-site.xml file. You need this to specify the value of the api_url property in the Hue configuration.

   For Azure environments, note the value of the fs.azure.ext.raz.rest.host.url property.

   For AWS and GCS environments, note the value of the fs.s3a.ext.raz.rest.host.url property.
4. Go to CDW Virtual Warehouse ⋮ Edit CONFIGURATIONS and select the hue-safety-valve from the Configuration files dropdown menu.
5. Add the following lines in the hue-safety-valve field:

```
[desktop]
[[raz]]
is_enabled=true
api_url=https://[***INSTANCE-1***]:6082/,https://[***INSTANCE-
2***]:6082/
```

6. Click Apply Changes.

DWX-16899: Error while viewing Impala job status on the mini Job Browser

When you click on the application ID after submitting an Impala query on Hue that is running on the environment level, you may notice the following error: 401 Client Error: Unauthorized for url: http://coordinator.impala-xyz.svc.cluster.local:25000/queries?json=true Must authenticate with Basic authentication. This does not happen when you do the same from Hue deployed at a Virtual Warehouse level.

None.

DWX-16913: "Results have expired" message while running a CTAS query

You may see the following message on the Hue UI when you submit a CREATE TABLE AS SELECT (CTAS) query from a Hue instance that is deployed on the environment level: "Results have expired, rerun the query if needed". This does not happen when you do the same from Hue deployed at a Virtual Warehouse level.

None.

DWX-16917: Failed to validate proxy privileges error while running queries from Hue

You may see the following error intermittently in Hue's pod logs while running queries from a Hue instance that is deployed at the environment level: "Failed to validate proxy privileges of <username>".

None.

DWX-16895: Incorrect status of Hue pods when you edit the Hue instance properties

When you update a configuration of a Hue instance that is deployed at the environment level, such as increasing or decreasing the size of the Hue instance, you see a success message on the CDW UI. After some time, the status of the Hue instance also changes from "Updating" to "Running". However, when you list the Hue pods using kubectl, you see that not all backend pods are in the running state–a few of them are still in the init state.

None. The pods come up successfully eventually after a sufficient time has passed.

DWX-16863: The upgrade button is present on the CDW UI, but Hue upgrades are not supported

You see the Upgrade button on the **Query Editor** page in the CDW UI when Hue is deployed at the environment level. However, on CDW version 1.8.1, upgrading the Hue instance that is deployed at the environment level is not supported.

None.

DWX-16893: A user can see all the queries in Job browser

In a Hue instance deployed at the environment level, by design, the Hue instances must not share the saved queries and query history with other Hue instances even for the same user. However, a logged in user is able to view all the queries executed by that user on all the Virtual Warehouses on a particular Database Catalog.

None.

Delay in listing queries in Impala Queries in the Job browser

Listing an Impala query in the Job browser can take an inordinate amount of time.

DWX-14927 Hue fails to list Iceberg snapshots

Hue does not recognize the Iceberg history queries from Hive to list table snapshots. For example, Hue indicates an error at the . before history when you run the following query.

```
select * from <db_name>.<table_name>.history
```

DWX-14968 Connection termination error in Impala queries tab after Hue inactivity

To reproduce the problem: 1) In the Impala job browser, navigate to Impala queries. 2) Wait for a few minutes.

After a few minutes of inactivity, the follow error is displayed:

Workaround: Refresh the page, or alternatively start a new session.

DWX-15115 Error displayed after clicking on hyperlink below Hue table browser

In Hue, below the table browser, clicking the hyperlink to a location causes an HTTP 500 error because the file browser is not enabled for environments that are not Ranger authorized (RAZ).

DWX-15090: CSRF error intermittently seen in the Hue Job Browser

You may intermittently see the “403 - CSRF” error on the Hue web interface as well as in the Hue logs after running Hive queries from Hue.

Reload the browser or start a new Hue session.

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Carried over from the previous release: Iceberg

IMPALA-12742 DELETE/UPDATE Iceberg table partitioned by DATE fails

You can partition Iceberg tables using identities, such as int and date. The file path contains the partition value you can read. When you partition by DATE, subsequently running queries to update or delete data from partitioned data creates an incorrect conversion of DATE. The Catalog cannot parse the data and throws an error.

CDPD-66779: Partitioned Iceberg table not getting loaded with insert select query from Hive

If you create a partitioned table in Iceberg and then try to insert data from another table as shown below, an error occurs.

```
insert into table partition_transform_4 select t, ts from vector
tab10k;
```

Use the CLUSTER BY clause on the partitioned column to insert data. For example:

```
insert into table partition_transform_4 select t, ts from t1 clu
ster by ts;
```

Branch FAST FORWARD does not work as expected

The Apache Iceberg spec indicates you can use either one or two arguments to fast forward a branch. The following example shows using two arguments:

```
ALTER TABLE <name> EXECUTE FAST-FORWARD 'x' 'y'
```

However, omitting the second branch name, does not work as documented by Apache Iceberg. The named branch is not fast-forwarded to the current branch. An exception occurs at the Iceberg level.

You must use two arguments to the EXECUTE FAST FORWARD feature to forward a branch.

HIVE-28055 Merging Iceberg branches requires a target table alias

Hive supports only one level of qualifier when referencing columns. In other words only one dot is accepted. For example, select table.col from ...; is allowed. select db.table.col is not allowed. Using the merge statement to merge Iceberg branches without a target or source table alias causes an exception:

```
org.apache.hadoop.hive.ql.parse.SemanticException: ... Invalid t
able alias or column reference ...
```

Use an alias, for example t, for the target table.

```
merge into mydb.target.branch_branch1 t using mydb.source.branch
_branch1 s on t.id = s.id when matched then update set value = '
matched';
```

DWX-17210: Timeout issue querying Iceberg tables from Hive

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

- fs.s3.maxConnections=1000
- fs.s3a.connection.maximum=1000

Restart the Database Catalog and Virtual Warehouse.

DWX-15014 Loading airports table in demo data fails

This issue occurs only when using a non-default Database Catalog. A invalid path error message occurs when using the load command to load the demo data airports table in Iceberg, ORC, or Parquet format.

DWX-14163 Limitations reading Iceberg tables in Avro file format from Impala

The Avro, Impala, and Iceberg specifications describe some limitations related to Avro, and those limitations exist in CDP. In addition to these, the DECIMAL type is not supported in this release.

DEX-7946 Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata if you drop the table during migration from Hive to Iceberg.

Workarounds: Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13733 Timeout issue querying Iceberg tables from Hive

A timeout issue can cause the query to fail.

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

```
fs.s3.maxConnections=1000
fs.s3a.connection.maximum=1000
```

Restart the Database Catalog and Virtual Warehouse.

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: Impala Virtual Warehouse

DWX-17455 ODBC client using JWT authentication cannot connect to Impala Virtual Warehouse

If you are using the Cloudera ODBC Connector and JWT authentication to connect to a CDW Impala Virtual Warehouse where the Impala coordinators are configured for high availability in an active-active mode, the connection results in a "401 Unauthorized" error. The error is not seen on Impala Virtual Warehouses that have an active-passive high availability or where high availability is disabled.

In the ODBC Data Source Name (DSN), make the following changes:

1. Ensure that SSL is enabled by setting SSL=1.
2. Remove authentication by setting AuthMech=0.
3. Remove the JWTString parameter.
4. Add the parameter http.header.Authorization and set it to the word "Bearer" followed by a space and the JWT string

DSN configuration before the workaround

```
SSL=1
AuthMech=8
JWTString=full_jwt_text
```

DSN configuration after the workaround

```
SSL=1
AuthMech=0
http.header.Authorization=Bearer full_jwt_text
```

IMPALA-12809 Metadata query problem from Impala

When querying metadata from Impala on a cluster with more than a few nodes, the query might fail with a null pointer exception. Reported problems occurred when running join queries on metadata, but other metadata queries might fail.

Limit the number of nodes that process the query by setting the NUM_NODES query option to 1:

```
SET NUM_NODES=1
```

TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

The Cloudera Data Warehouse (CDW) Public Cloud 2023.0.14.0 (DWX-1.6.3) version incorporated a feature for performing automatic metadata synchronization across multiple Apache Impala (Impala) Virtual Warehouses. The feature is enabled by default, and relies on the Hive MetaStore events. When a certain sequence of Data Definition Language (DDL) SQL commands are executed as described below, users may encounter a java.lang.NullPointerException (NPE). The exception causes the event processor to stop processing other metadata operations.

If a CREATE TABLE command (not CREATE TABLE AS SELECT) is followed immediately (approximately within 1 second interval) by INVALIDATE METADATA or REFRESH TABLE command on the same table (either on the same Virtual Warehouse or on a different one), there is a possibility that the second command will not find the table in the catalog cache of a peer Virtual Warehouse and generate an NPE.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

DWX-17175: Impala Virtual Warehouse max executor limit issue

The Impala Virtual Warehouse has a limit of max 200 executors. The Executors slider has a max limit of 200 for creating or editing a Virtual Warehouse. If you select a bigger t-shirt size than the default, or set a custom t-shirt size, the default max limit set by the UI could be more than 200 in some cases. Exceeding the max limit causes an error during a future Edit/Upgrade/Rebuild operation.

Workaround: Manually set the max executors limit to a preferred value less than 200, but a multiple of t-shirt size, when creating or editing Impala Virtual Warehouse

DWX-15016 Impala query failure when using Catalog high availability (HA)

Impala queries could fail with InconsistentMetadataFetchException when using HA feature in CDW. This happens when leader election failover for Impala Catalog Service happens due to a transient network error.

Workaround: Disable Catalog HA. In CDW, edit your Virtual Warehouse, and turn off Enable Impala Catalog Server HA.

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"
```

```
        # Save and quit the editor, and then run the following
command to check your changes.

        kubectl get svc nginx-service -n <cluster-name> -o yaml

        # Make sure that the annotation you added is present.
```

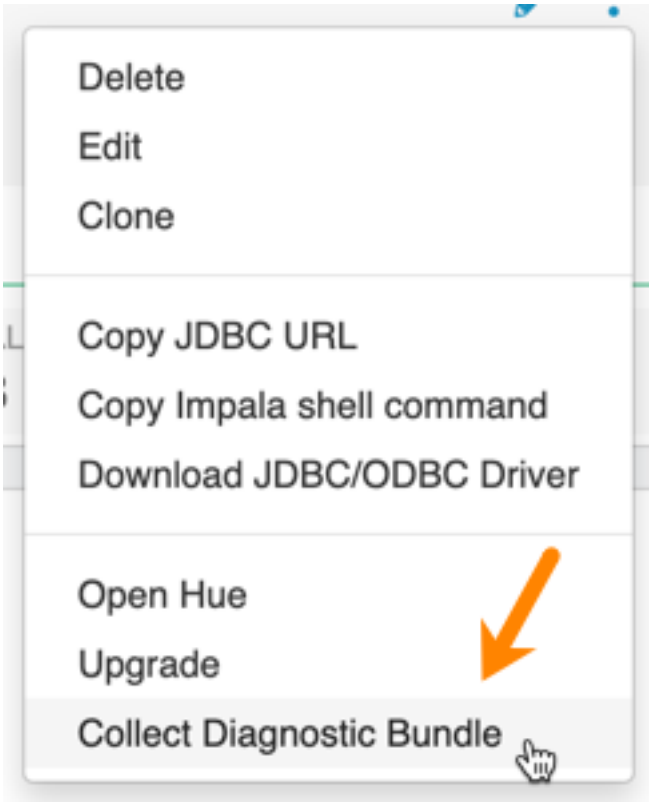
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

UI Issues

DWX-14610 Upgrade icon indicates the Database Catalog is the latest version, but might be incorrect for a few seconds

The cluster fetches data regularly in particular time intervals, or when an action is triggered. Inbetween this time interval, which is very brief, the cluster might not be updated, but will be refreshed in a few seconds.

Technical Service Bulletins

TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion

When working with Apache Hive (Hive) and Apache Iceberg (Iceberg) V2 tables, using the TRUNCATE statement may lead to unintended data deletion. This issue arises when the truncate command is applied to a branch of an Iceberg table. Instead of truncating the branch itself, the command affects the original (main) table, which results in unintended loss of data.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion

March 4, 2024 - Hotfix

Review the known issues, fixed issues, and changed behaviors in this hotfix release of Cloudera Data Warehouse on Public Cloud.

Fixed issues

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud fixes these issues.

Technical Service Bulletins

TSB 2024-743: Restoration of the Database Catalog can be incomplete due to an incorrect Hue Query Processor configuration in CDW

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-743: Restoration of the Database Catalog can be incomplete due to an incorrect Hue Query Processor configuration in CDW.

Known issues

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

See Data Visualization release notes for known issues in Cloudera Data Visualization 7.1.9.

Carried over from the previous release: Upgrade-related

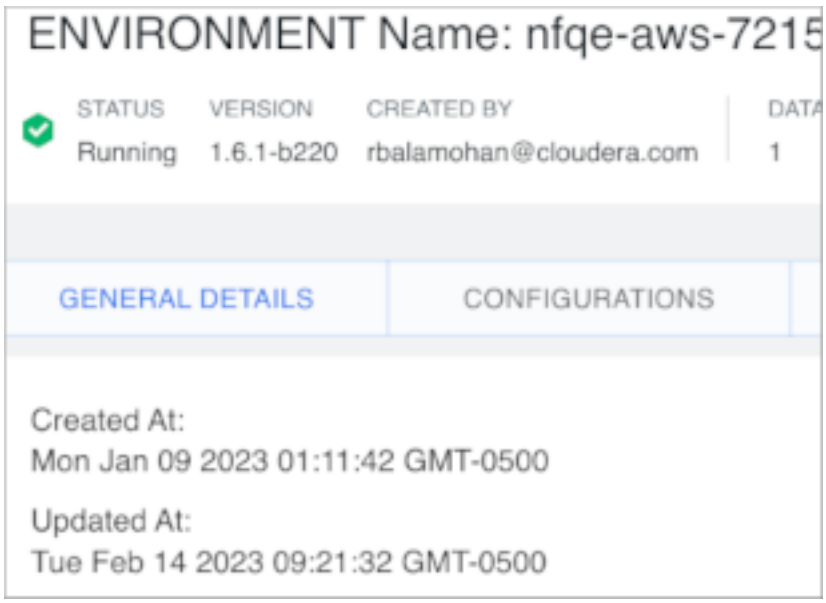
Upgrading to EKS 1.24 could result in Impala coordinators shutting down.

This issue is not seen on the Impala Virtual Warehouse running Runtime 2023.0.15.0-x or later.

Workaround: Manually start the Impala Virtual Warehouse from the UI or cli. Alternatively, replace your runtime with 1.7.1-b755 (released August 30, 2023) or later.

Certain CDW versions cannot upgrade to EKS 1.22

AWS environments activated in version 1.4.1-b86 (released June, 22, 2022) or earlier are not supported for upgrade to EKS 1.22 due to incompatibility of some components with EKS 1.22. To determine the activation version your pre-existing environment, in the Data Warehouse service, expand Environments. In Environments, search for and locate the environment that you want to view. Click Edit. In Environment Details, you see the CDW version.



DWX-15112 Enterprise Data Warehouse database configuration problems after a Helm-related rollback

If a Helm rollback fails due to an incorrect Enterprise Data Warehouse database configuration, the Virtual Warehouse and Database Catalog roll back to a previous configuration. The incorrect Enterprise Data Warehouse configuration persists, and can affect subsequent edit, upgrade, and rebuild operations on the rolled-back Virtual Warehouse or Database Catalog.

Carried over from the previous release: General

DWX-14923 After JWT authentication, attempting to connect the Impyla client using a user name or password should cause an error

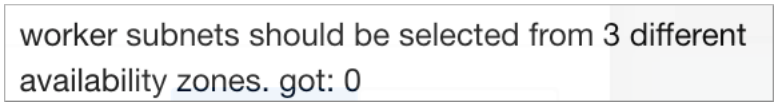
Using a JWT token, you can connect to a Virtual Warehouse as the user who generated the token.

If you connect with the JWT token, and then pass a user name or password from Impyla to the Virtual Warehouse, the connection arguments are silently ignored. Such an action should indicate that it is illegal to specify user or password when using JWT authentication.

DWX-15145 Environment validation popup error after activating an environment

Activating an environment having a public load balancer can cause an environment validation popup error.

To reproduce this problem 1) Create a data lake. 2) Activate an environment having a public load balancer deployment type and subnets in three different availability zones. A environment validation popup can occur even through subnets are in different availability zones. Several different popups can occur, including the following one:



DWX-15144 Virtual Warehouse naming restrictions

You cannot create a Virtual Warehouse having the same name as another Virtual Warehouse even if the like-named Virtual Warehouses are in different environments. You can create a Database Catalog having the same name as another Database Catalog if the Database Catalogs are in different environments.

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data

Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the CreateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersio ns line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

DWX-17703: Non-HA Impala Virtual Warehouse on a private Azure Kubernetes Service (AKS) setup fails

When 'Refresh' and 'Stop' operations run in parallel, Impala might move into an error state. The Refresh operation might think that Impala is in an error state as the coordinator pod is missing.

Rebuild the Impala Virtual Warehouse or restart it using the CLI.

DWX-17109: ABFS File Browser operations failing intermittently

You may encounter intermittent issues while performing typical operations on files and directories on the ABFS File Broswer, such as moving or renaming files.

None.

DWX-17613: Generic error message is displayed when you click on the directory you don't have access to on a RAZ cluster

You see the following error message when you click on an ABFS directory to which you do not have read/write permission on the ABFS File Browser in Hue: There was a problem with your request. This message is generic and does not provide insight into the actual issue.

None.

DWX-17620: Folder having special characters in its name is not accessible in ABFS

In Cloudera Data Warehouse Public Cloud, from a Virtual Warehouse going to the Azure Blob Filesystem (ABFS), creating a folder, and then performing an action such as Move, causes an error as shown in the following example:

```
Cannot access: abfs://data-files/user/hrt_qa/~@$&()*!+'=;. 404 C
lient Error: The specified path does not exist.
```

DWX-15214 DWX-15176 Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

The Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure

(Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed:

```
kubelet Error: failed to create containerd task: failed to creat
e shim task: OCI runtime create failed: runc create failed: unab
le to start container process: exec: “run_httpd.sh”: cannot run
executable found relative to current directory: unknown
```

This issue affects all CDW releases before 1.6.3 (released May 5, 2023)running on CDP Public Cloud on Azure.

The following users are affected: CDW Azure users using Hue in Impala and Hive Virtual Warehouses in environments earlier than version 1.6.3 or later if they upgrade the node image

Workaround: Upgrade the Virtual Warehouses to 1.6.3 and choose Hue version 2023.0.14.0.

Do not choose Hue versions older than 2023.0.14.0 when creating a Virtual Warehouse in environments of version 1.6.3.

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Database Catalog

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

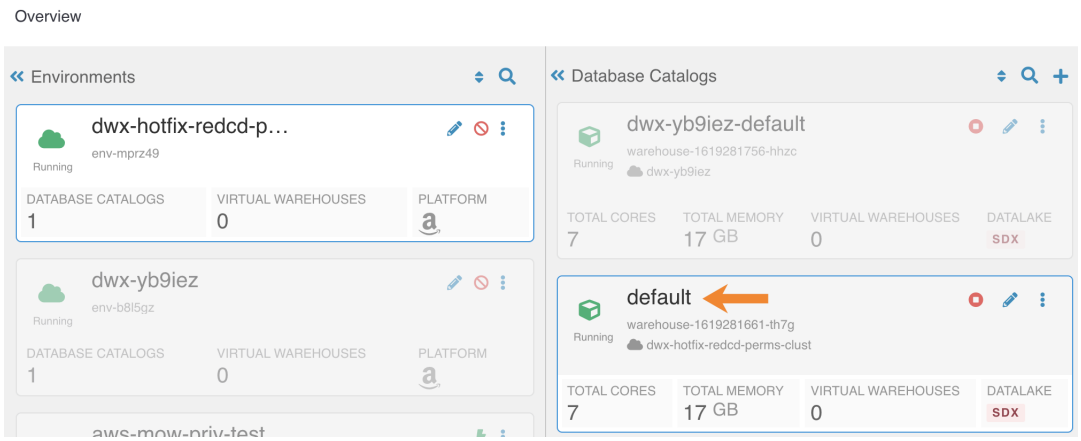
- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

Limited Hive image versions

Hive Virtual Warehouses you create in 1.8.1-b248 (released Nov 20, 2023) and later will run Istio 1.19.0. The new Istio version supports only new versions of Hive helm charts. If you have the CDW_VERSIONED_DEPLOY, only new Hive image versions appear in UI when you create a new Hive Virtual Warehouse.

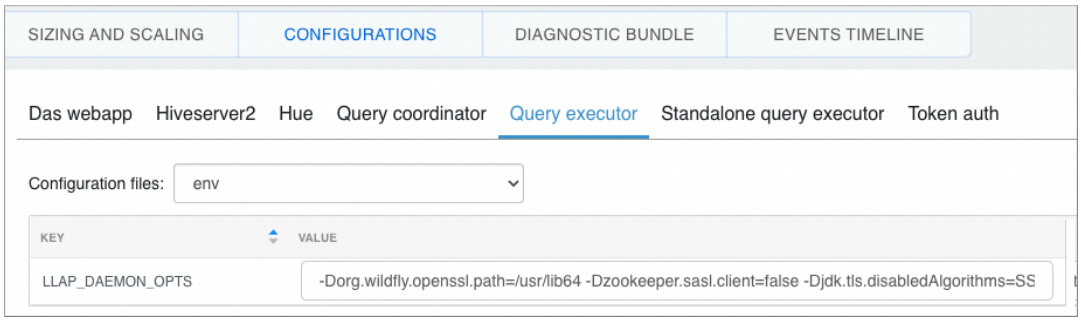
DWX-15064 Hive Virtual Warehouse stops but appears healthy

Due to an istio-proxy problem, the query coordinator can unexpectedly enter a not ready state instead of the expected error-state. Subsequently, the Hive Virtual Warehouse stops when reaching the autosuspend timeout without indicating a problem.

DWX-14452 Parquet table query might fail

Querying a table stored in Parquet from Hive might fail with the following exception message: java.lang.RuntimeException: java.lang.StackOverflowError. This problem can occur when the IN predicate has 243 values or more, and a small stack (-Xss = 256k) is configured for Hive in CDW.

Workaround: Change the value of the Xss in the JVM args to use the default value (1Mb) as follows: Select Edit from the Options menu of your Virtual Warehouse. Click Configurations > Query Executor > and select env.



In the third line shown below, change the value of LLAP_DAEMON_OPTS from -Xss256k to -Xss1M, and then click Apply Changes:

FROM:

-Dorg.wildfly.openssl.path=/usr/lib64 -Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Dhttp.maxConnections=12 -Xms24G -Xmx48G -Xss256k ...

TO:

... -Xss1M ...

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
       kind: ClusterRole
       metadata:
       name: diagnostic-data-generator-role-monitor
       rules:
       - apiGroups:
       - ""
       resources:
       - configmaps
       - events
       - pods
       - persistentvolumeclaims
       - nodes
       verbs:
       - get
       - list
       - apiGroups:
```

```
        - apps
        resources:
        - deployments
        - statefulsets
        verbs:
        - get
        - list
        - apiGroups:
        - "edws.cloudera.com"
        resources:
        - computes
        verbs:
        - get
        - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.

7. Click Apply in the upper right corner of the page to save the configuration.

8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name 'ec2-18-219-
32-183.us-east-2.compute.amazonaws.com' does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Hue Query Editor

CDPD-27918: Hue does not automatically pick up RAZ HA configurations

On a CDP Public Cloud environment in which you have configured RAZ in High Availability mode, Hue in CDW does not pick up all the RAZ host URLs automatically. Therefore, if a RAZ instance to which Hue is connected goes down, Hue becomes unavailable.

You must manually add comma-separated RAZ instances in the Hue Advanced Configuration Snippet.

1. Log in to the CDP Management Console as an Administrator.
2. Go to Environment Data Lake and open Cloudera Manager for your environment.
3. Go to Clusters Ranger RAZ service Instances RAZ server Processes and note the value of the fs.s3a.ext.raz.rest.host.url property from the core-site.xml file. You need this to specify the value of the api_url property in the Hue configuration.

   For Azure environments, note the value of the fs.azure.ext.raz.rest.host.url property.

   For AWS and GCS environments, note the value of the fs.s3a.ext.raz.rest.host.url property.
4. Go to CDW Virtual Warehouse ⋮ Edit CONFIGURATIONS and select the hue-safety-valve from the Configuration files dropdown menu.
5. Add the following lines in the hue-safety-valve field:

```
[desktop]
[[raz]]
is_enabled=true
api_url=https://[***INSTANCE-1***]:6082/,https://[***INSTANCE-
2***]:6082/
```

6. Click Apply Changes.

DWX-16899: Error while viewing Impala job status on the mini Job Browser

When you click on the application ID after submitting an Impala query on Hue that is running on the environment level, you may notice the following error: 401 Client Error: Unauthorized for url: http://coordinator.impala-xyz.svc.cluster.local:25000/queries?json=true Must authenticate with Basic authentication. This does not happen when you do the same from Hue deployed at a Virtual Warehouse level.

None.

DWX-16913: "Results have expired" message while running a CTAS query

You may see the following message on the Hue UI when you submit a CREATE TABLE AS SELECT (CTAS) query from a Hue instance that is deployed on the environment level: "Results have expired, rerun the query if needed". This does not happen when you do the same from Hue deployed at a Virtual Warehouse level.

None.

DWX-16917: Failed to validate proxy privileges error while running queries from Hue

You may see the following error intermittently in Hue's pod logs while running queries from a Hue instance that is deployed at the environment level: "Failed to validate proxy privileges of <username>".

None.

DWX-16895: Incorrect status of Hue pods when you edit the Hue instance properties

When you update a configuration of a Hue instance that is deployed at the environment level, such as increasing or decreasing the size of the Hue instance, you see a success message on the CDW UI. After some time, the status of the Hue instance also changes from "Updating" to "Running". However, when you list the Hue pods using kubectl, you see that not all backend pods are in the running state–a few of them are still in the init state.

None. The pods come up successfully eventually after a sufficient time has passed.

DWX-16863: The upgrade button is present on the CDW UI, but Hue upgrades are not supported

You see the Upgrade button on the **Query Editor** page in the CDW UI when Hue is deployed at the environment level. However, on CDW version 1.8.1, upgrading the Hue instance that is deployed at the environment level is not supported.

None.

DWX-16893: A user can see all the queries in Job browser

In a Hue instance deployed at the environment level, by design, the Hue instances must not share the saved queries and query history with other Hue instances even for the same user. However, a logged in user is able to view all the queries executed by that user on all the Virtual Warehouses on a particular Database Catalog.

None.

Delay in listing queries in Impala Queries in the Job browser

Listing an Impala query in the Job browser can take an inordinate amount of time.

DWX-14927 Hue fails to list Iceberg snapshots

Hue does not recognize the Iceberg history queries from Hive to list table snapshots. For example, Hue indicates an error at the . before history when you run the following query.

```
select * from <db_name>.<table_name>.history
```

DWX-14968 Connection termination error in Impala queries tab after Hue inactivity

To reproduce the problem: 1) In the Impala job browser, navigate to Impala queries. 2) Wait for a few minutes.

After a few minutes of inactivity, the follow error is displayed:

Workaround: Refresh the page, or alternatively start a new session.

DWX-15115 Error displayed after clicking on hyperlink below Hue table browser

In Hue, below the table browser, clicking the hyperlink to a location causes an HTTP 500 error because the file browser is not enabled for environments that are not Ranger authorized (RAZ).

DWX-15090: CSRF error intermittently seen in the Hue Job Browser

You may intermittently see the “403 - CSRF” error on the Hue web interface as well as in the Hue logs after running Hive queries from Hue.

Reload the browser or start a new Hue session.

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Carried over from the previous release: Iceberg

IMPALA-12742 DELETE/UPDATE Iceberg table partitioned by DATE fails

You can partition Iceberg tables using identities, such as int and date. The file path contains the partition value you can read. When you partition by DATE, subsequently running queries to update or delete data from partitioned data creates an incorrect conversion of DATE. The Catalog cannot parse the data and throws an error.

CDPD-66779: Partitioned Iceberg table not getting loaded with insert select query from Hive

If you create a partitioned table in Iceberg and then try to insert data from another table as shown below, an error occurs.

```
insert into table partition_transform_4 select t, ts from vector
tab10k;
```

Use the CLUSTER BY clause on the partitioned column to insert data. For example:

```
insert into table partition_transform_4 select t, ts from t1 clu
ster by ts;
```

Branch FAST FORWARD does not work as expected

The Apache Iceberg spec indicates you can use either one or two arguments to fast forward a branch. The following example shows using two arguments:

```
ALTER TABLE <name> EXECUTE FAST-FORWARD 'x' 'y'
```

However, omitting the second branch name, does not work as documented by Apache Iceberg. The named branch is not fast-forwarded to the current branch. An exception occurs at the Iceberg level.

You must use two arguments to the EXECUTE FAST FORWARD feature to forward a branch.

HIVE-28055 Merging Iceberg branches requires a target table alias

Hive supports only one level of qualifier when referencing columns. In other words only one dot is accepted. For example, select table.col from ...; is allowed. select db.table.col is not allowed. Using the merge statement to merge Iceberg branches without a target or source table alias causes an exception:

```
org.apache.hadoop.hive.ql.parse.SemanticException: ... Invalid t
able alias or column reference ...
```

Use an alias, for example t, for the target table.

```
merge into mydb.target.branch_branch1 t using mydb.source.branch
_branch1 s on t.id = s.id when matched then update set value = '
matched';
```

DWX-17210: Timeout issue querying Iceberg tables from Hive

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

- fs.s3.maxConnections=1000
- fs.s3a.connection.maximum=1000

Restart the Database Catalog and Virtual Warehouse.

DWX-15014 Loading airports table in demo data fails

This issue occurs only when using a non-default Database Catalog. A invalid path error message occurs when using the load command to load the demo data airports table in Iceberg, ORC, or Parquet format.

DWX-14163 Limitations reading Iceberg tables in Avro file format from Impala

The Avro, Impala, and Iceberg specifications describe some limitations related to Avro, and those limitations exist in CDP. In addition to these, the DECIMAL type is not supported in this release.

DEX-7946 Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata if you drop the table during migration from Hive to Iceberg.

Workarounds: Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13733 Timeout issue querying Iceberg tables from Hive

A timeout issue can cause the query to fail.

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

```
fs.s3.maxConnections=1000
fs.s3a.connection.maximum=1000
```

Restart the Database Catalog and Virtual Warehouse.

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: Impala Virtual Warehouse

DWX-17455 ODBC client using JWT authentication cannot connect to Impala Virtual Warehouse

If you are using the Cloudera ODBC Connector and JWT authentication to connect to a CDW Impala Virtual Warehouse where the Impala coordinators are configured for high availability in an active-active mode, the connection results in a "401 Unauthorized" error. The error is not seen on Impala Virtual Warehouses that have an active-passive high availability or where high availability is disabled.

In the ODBC Data Source Name (DSN), make the following changes:

1. Ensure that SSL is enabled by setting SSL=1.
2. Remove authentication by setting AuthMech=0.
3. Remove the JWTString parameter.
4. Add the parameter http.header.Authorization and set it to the word "Bearer" followed by a space and the JWT string

DSN configuration before the workaround

```
SSL=1
AuthMech=8
JWTString=full_jwt_text
```

DSN configuration after the workaround

```
SSL=1
AuthMech=0
http.header.Authorization=Bearer full_jwt_text
```

IMPALA-12809 Metadata query problem from Impala

When querying metadata from Impala on a cluster with more than a few nodes, the query might fail with a null pointer exception. Reported problems occurred when running join queries on metadata, but other metadata queries might fail.

Limit the number of nodes that process the query by setting the NUM_NODES query option to 1:

```
SET NUM_NODES=1
```

TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

The Cloudera Data Warehouse (CDW) Public Cloud 2023.0.14.0 (DWX-1.6.3) version incorporated a feature for performing automatic metadata synchronization across multiple Apache Impala (Impala) Virtual Warehouses. The feature is enabled by default, and relies on the Hive MetaStore events. When a certain sequence of Data Definition Language (DDL) SQL commands are executed as described below, users may encounter a java.lang.NullPointerException (NPE). The exception causes the event processor to stop processing other metadata operations.

If a CREATE TABLE command (not CREATE TABLE AS SELECT) is followed immediately (approximately within 1 second interval) by INVALIDATE METADATA or REFRESH TABLE command on the same table (either on the same Virtual Warehouse or on a different one), there is a possibility that the second command will not find the table in the catalog cache of a peer Virtual Warehouse and generate an NPE.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

DWX-17175: Impala Virtual Warehouse max executor limit issue

The Impala Virtual Warehouse has a limit of max 200 executors. The Executors slider has a max limit of 200 for creating or editing a Virtual Warehouse. If you select a bigger t-shirt size than the default, or set a custom t-shirt size, the default max limit set by the UI could be more than 200 in some cases. Exceeding the max limit causes an error during a future Edit/Upgrade/Rebuild operation.

Workaround: Manually set the max executors limit to a preferred value less than 200, but a multiple of t-shirt size, when creating or editing Impala Virtual Warehouse

DWX-15016 Impala query failure when using Catalog high availability (HA)

Impala queries could fail with InconsistentMetadataFetchException when using HA feature in CDW. This happens when leader election failover for Impala Catalog Service happens due to a transient network error.

Workaround: Disable Catalog HA. In CDW, edit your Virtual Warehouse, and turn off Enable Impala Catalog Server HA.

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"
```

```
        # Save and quit the editor, and then run the following
command to check your changes.

        kubectl get svc nginx-service -n <cluster-name> -o yaml

        # Make sure that the annotation you added is present.
```

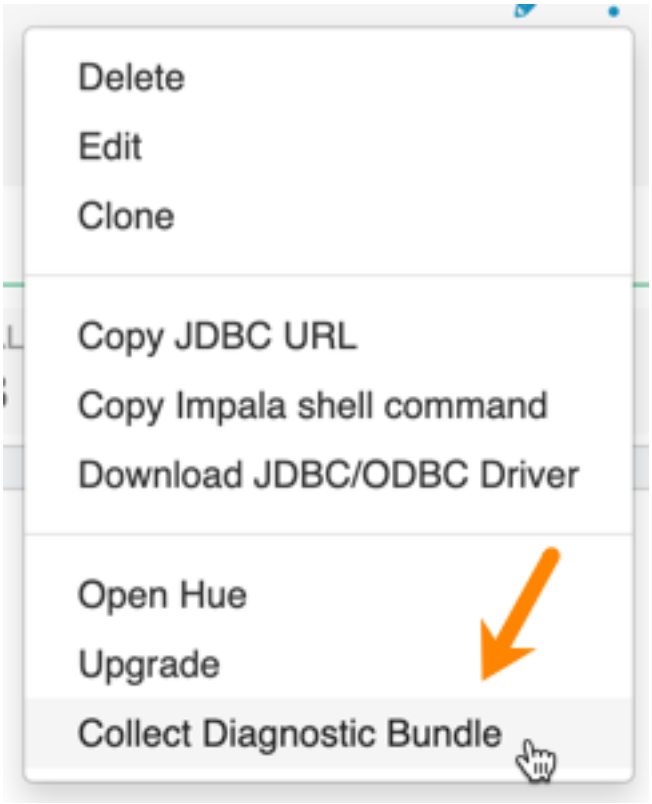
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

UI Issues

DWX-14610 Upgrade icon indicates the Database Catalog is the latest version, but might be incorrect for a few seconds

The cluster fetches data regularly in particular time intervals, or when an action is triggered. Inbetween this time interval, which is very brief, the cluster might not be updated, but will be refreshed in a few seconds.

Technical Service Bulletins

TSB 2023-719: Cloudera Data Warehouse Backup/Restore of CDP Data Visualization incomplete

Cloudera Data Warehouse (CDW) customers using the CDW Automated Backup/Restore feature will encounter an issue with the restoration versions of Cloudera Data Visualization (CDV) older than CDV 7.1.6.2-3 due to a schema change in this release. If the Backup was taken from an older CDW environment that contained a version of CDV older than CDV 7.1.6.2-3, the Restore procedure will succeed. Though once the user opens the CDV Queries tab, the user could encounter the error message: “column jobs_jobschedule.owner_id does not exist…”

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-719: Cloudera Data Warehouse Backup/Restore of Cloudera Data Visualization incomplete.

TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables

Apache Hive (Hive) and Apache Impala (Impala) modify statements (DELETE/UPDATE/MERGE) on Apache Iceberg (Iceberg) V2 tables can corrupt the tables if there is a concurrent table compaction from Apache Spark. The issue happens when the compaction and modify statement run in parallel, and when the compaction job commits before the modify statement. In this case the position delete files of the modify statement still point to the old files. This means the following in case of

- DELETE statements
    - Deleting records pointing to old files have no effect
- UPDATE / MERGE statements
    - Deleting records pointing to old files have no effect
    - The table will also have the newly added data records
    - Rewritten records will still be active

This issue does not affect Apache NiFi (NiFi) and Apache Flink (Flink) as these components write equality delete files.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables.

TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

The Spark Iceberg library includes two procedures - rewrite_data_files and rewrite_position_delete_files. The current implementation of rewrite_data_files has a limitation that the position delete files are not deleted and still tracked by the table metadata, even if they no longer refer to an active data file. This is called the dangling delete problem. To solve this, the rewrite_position_delete_files procedure is implemented in the Spark Iceberg library to remove these old “dangling” position delete files.

Due to the dangling delete limitation, when an Iceberg table with dangling deletes is queried in Impala, Impala tries to optimize select count(*) from iceberg_table query to return the results using stats. This optimization returns incorrect results.

The following conditions must be met for this issue to occur:

- All delete files in the Iceberg table are “dangling”
    - This would occur immediately after running Spark rewrite_data_files AND
        - Before any further delete operations are performed on the table OR
        - Before Spark rewrite_position_delete_files is run on the table

- Only stats optimized plain select count(*) from iceberg_table queries are affected. For example, the query should not have:
  - Any WHERE clause
  - Any GROUP BY clause
  - Any HAVING clause

  Remove dangling deletes: After rewrite_data_files, position delete records pointing to the rewritten data files are not always marked for removal, and can remain tracked by the live snapshot metadata of the table. This is known as the dangling delete problem.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables.

TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion

When working with Apache Hive (Hive) and Apache Iceberg (Iceberg) V2 tables, using the TRUNCATE statement may lead to unintended data deletion. This issue arises when the truncate command is applied to a branch of an Iceberg table. Instead of truncating the branch itself, the command affects the original (main) table, which results in unintended loss of data.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion.

Behavior changes

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

This release is a hot fix for a condition related to DWX server code and the Amazon Relational Database (RDS) that causes an inordinate number of events to be saved. Insufficient database connections can occur and Cloudera Data Warehouse can shut down.

February 29, 2024 Release Notes

Review the new feaures, fixes, behavioral changes, and preview features in this release of Cloudera Data Warehouse on Public Cloud.

What's new in Cloudera Data Warehouse on Public Cloud

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Cloudera Data Warehouse Public Cloud 1.8.5-b35 changes, described in more detail below:

- AWS Elastic Kubernetes Service 1.28 support on page 87
- Accessing S3 buckets in an AWS environment on page 87
- Database Catalog size configuration on page 87
- GA availability of a private CDW environment in Azure Kubernetes Service on page 87
- IAM policies moved from documentation to a public Github repository on page 87
- Metrics for monitoring and troubleshooting Impala Virtual Warehouses on page 88
- CDP CLI commands for creating and updating an AWS and Azure cluster on page 88
- CDP CLI commands for configuring Workload Aware Auto-Scaling on page 88

Cloudera Data Warehouse Runtime 2024.0.17.0-73 changed, described in more detail below:

Iceberg

- Iceberg drop partition feature on page 88
- Iceberg branching and tagging GA on page 88

- Enhanced security of Iceberg metadata on page 89
- Impala support for changing the Iceberg table metadata location on page 89
- Support for copy-on-write (COW) on page 89
- Query metadata tables feature on page 89
- Directed distribution mode on page 89

Impala

- SHOW VIEWS statement on page 89
- Planner changes to improve cardinality estimation on page 89
- Distribute runtime filter aggregation on page 90
- Improvement in catalog observability on page 90
- Caching codegen functions on page 90

AWS Elastic Kubernetes Service 1.28 support

New CDW clusters using AWS environments you activate in this release 1.8.4-b35 (released February, 2024) of CDW will use Amazon Kubernetes (EKS) version 1.28. For more information, see Upgrading Amazon Kubernetes Service.

Accessing S3 buckets in an AWS environment

You can use the CDW UI for configuring access to S3 buckets under certain conditions. You can also configure your own custom encryption key for read/write access from CDW Public Cloud on AWS to the external S3 bucket.

For more information, see Accessing S3 buckets.

Database Catalog size configuration

Using the following CDP CLI commands, you can configure the Java heap for your Database Catalog to small 8Gb (default), medium 16Gb, or large 24 Gb:

Create a default Database Catalog that configures, for example, a large Java Heap size.

```
dw create-dbc --cluster-id=env-sb42vs --name=SSlarge --memory-tshirt-size=la
rge
```

Create a Database Catalog that accepts the default Java Heap size, which is small, by not configuring the size.

```
dw create-dbc --cluster-id=env-sb42vs --name=SSdefault
```

List Database Catalog configuration information:

```
dw list-dbcs --cluster-id=env-sb42vs
```

Using the CDW UI as described in Creating a Database Catalog, you can configure the Database Catalog size.

To avoid unnecessary cloud expenses, do not increase the size unless you experience Java heap issues.

GA availability of a private CDW environment in Azure Kubernetes Service

You can now enable a private CDW environment in Azure Kubernetes Service (AKS). You use the Private CDW option in the CLI, and CDW deploys an Azure Kubernetes Cluster with only private endpoints enabled. The cluster can be accessed only from your Azure network.

IAM policies moved from documentation to a public Github repository

To meet customer requests for tracking changes to IAM policies for CDW, the policies now reside in Github. CDW documentation, such as "Attaching the policy to your cross-account role" provides links to the policies.

Metrics for monitoring and troubleshooting Impala Virtual Warehouses

Additional global metrics and table level event metrics are available for debugging an Impala Virtual Warehouse.

CDP CLI commands for creating and updating an AWS and Azure cluster

You can create and update a CDW cluster in Amazon and Azure environments using the following commands:

- create-aws-cluster --environment-crn <value> [options]
- create-azure-cluster --environment-crn <value> --user-assigned-managed-identity <value> [options]
- update-aws-cluster [options]
- update-azure-cluster [options]

For more information about using the commands, including options, see the CDP CLI reference.

CDP CLI commands for configuring Workload Aware Auto-Scaling

You can perform the following Workload Aware Auto-Scaling (WAAS) configurations as part of a Virtual Warehouse create request:

- Create a Virtual Warehouse with executor group sets
- Choose how many executor group sets to configure
- Configure each executor group set.

  Modify, add, or delete an executor group set as part of a Virtual Warehouse update request. To delete a group set, set the deleteGroupSet option to true for the group set.
- Update executor group sets of a Virtual Warehouse

For example:

```
cdp --profile ${CDP_PROFILE} \
  dw create-vw \
   --cluster-id ${CLUSTER_ID} \
   --dbc-id ${DBC_ID} \
   --vw-type impala \
   --name "impala-$(openssl rand -hex 6)" \
   --template xsmall \
   --impala-ha-settings highAvailabilityMode=ACTIVE_ACTIVE \
   --autoscaling "impalaExecutorGroupSets={small={execGroupSize=1,minExecu
torGroups=1,maxExecutorGroups=1,autoSuspendTimeoutSeconds=301,disableAutoSus
pend=true,triggerScaleUpDelay=21,triggerScaleDownDelay=21},custom1={execGrou
pSize=2,minExecutorGroups=0,maxExecutorGroups=1,autoSuspendTimeoutSeconds=30
2,disableAutoSuspend=true,triggerScaleUpDelay=22,triggerScaleDownDelay=22},c
ustom2={execGroupSize=3,minExecutorGroups=0,maxExecutorGroups=1,autoSuspendT
imeoutSeconds=303,disableAutoSuspend=true,triggerScaleUpDelay=23,triggerScal
eDownDelay=23},custom3={execGroupSize=4,minExecutorGroups=0,maxExecutorGroup
s=1,autoSuspendTimeoutSeconds=304,disableAutoSuspend=true,triggerScaleUpDela
y=24,triggerScaleDownDelay=24},large={execGroupSize=5,minExecutorGroups=0,ma
xExecutorGroups=1,autoSuspendTimeoutSeconds=305,disableAutoSuspend=true,trig
gerScaleUpDelay=25,triggerScaleDownDelay=25}}"
```

Iceberg drop partition feature

You can easily remove a partition from an Iceberg partition using an alter table statement from Impala. Removing a partition does not affect the table schema. The column is not removed from the schema. This feature is offered on a general availability (GA) basis.

Iceberg branching and tagging GA

In this release, branching and tagging is offered on a general availability (GA) basis. From Hive, you can manage the lifecycle of snapshots using the Iceberg branching and Iceberg tagging features. Branches are references to snapshots

that have a lifecycle of their own. Tags identify snapshots you need for auditing and conforming to GDPR. Cloudera recommends that you use this feature in test and development environments. It is not recommended for production deployments.

Enhanced security of Iceberg metadata

In this release, you can prevent overrides of the Iceberg file metadata location by unauthorized users. You accept the default (true), in the Virtual Warehouse to use this feature:

- Hive

  Configure the property hive.iceberg.allow.datafiles.in.table.location.only.
- Impala

  Configure the catalogd property iceberg_restrict_data_file_location.

When set to true, all the data files being read must be within the table location; otherwise, an error occurs. When set to false, an unauthorized party who knows the underlying schema and file location outside the table location can rewrite the manifest files within one table location to point to the data files in another table location to read your data. For more information, see "Changing the table metadata location".

Impala support for changing the Iceberg table metadata location

In this release, you can change the Iceberg table metadata location from Impala as well as Hive.

Support for copy-on-write (COW)

Hive supports the copy-on-write (COW) as well as merge-on-read (MOR) for handling Iceberg row-level updates and deletes. You configure COW or MOR based on your use case and rate of data change.

Query metadata tables feature

From Hive and Impala, you can query Iceberg metadata tables as you would query a Hive table. For example, you can use projections, joins, filters, and so on.

Directed distribution mode

This release implements directed distribution mode. The scheduler collects information about which Iceberg data file is scheduled on which host. Since, the scan node for the data files are on the same host as the Iceberg join node, delete files are sent directly to that specific host. This mode can improve V2 table read performance.

SHOW VIEWS statement

This release introduces the SHOW VIEWS statement, which simplifies the task of listing all views within a specified schema or database. Using this command, you can quickly identify and review views, thereby enhancing performance by reducing metadata scan operations.

Planner changes to improve cardinality estimation

Significant changes have been made to the query planner to improve cardinality estimation, a critical component of workload-aware autoscaling.

In previous versions, Impala would generate a plan first and then search for runtime filters based on the entire plan. In this release, selective runtime filters have been integrated. These filters aim to reduce the cardinality estimates of scan nodes and specific join nodes located above them. This refinement occurs after the generation of runtime filters and before the computation of resource requirements.

Distribute runtime filter aggregation

Aggregating runtime filters during runtime can impose significant memory overhead on the coordinator. To address this issue, we initially introduced local aggregation of runtime filters within a single executor node, aiming to alleviate strain on the coordinator by transmitting filter updates only after local aggregation. However, as Impala clusters scale up, the limitations of local filter aggregation become evident, especially in scenarios involving numerous nodes. This places significant memory stress on the coordinator node. To mitigate this challenge, we have implemented a solution that distributes the runtime filter aggregation across specific Impala backends.

Improvement in catalog observability

This release introduces significant enhancements to the Impala Catalog Web UI, focusing on addressing performance issues associated with delays in processing Hive Metastore (HMS) events. These improvements aim to mitigate the risk of queries using outdated metadata.

Caching codegen functions

In Impala, "codegen" involves generating specialized machine code for each query based on query-specific information. When executing a standard query, the query optimizer generates an optimized query plan, which is then passed to the executor for processing. The codegen capability converts query-specific information into machine code, enhancing query performance through faster execution.

Support ORDER BY for collections of variable length types in SELECT list

This release introduces support for collections of variable length types in the sorting tuple. While it's now possible to include these collection columns in the SELECT list alongside other columns used for sorting, direct sorting by these collection columns is not supported. Additionally, collections of variable-length types can now serve as non-passthrough children of UNION ALL nodes.

It is important to note that structs containing collections, whether of variable or fixed length, are still not supported in the select list for ORDER BY queries.

Here are examples of supported queries:

```
select id, arr_string_1d from collection_tbl order by id;
select id, map_1d from collection_tbl order by id;
```

However, queries such as the following are not supported:

```
select id, struct_contains_map from collection_struct_mix order by id;
```

Attempting to execute such queries will result in the error message "AnalysisException: Sorting is not supported if the select list contains collection(s) nested in struct(s)."

Preview features in Cloudera Data Warehouse on Public Cloud

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces this technical preview.

Note: Technical previews are considered under development. Do not use these features in production environments.

Impala Virtual Warehouse EBS volume type provisioning (Preview)

Impala Virtual Warehouse when configured to spill to EBS volumes can now also spill to the Amazon gp3 Elastic Block Store volumes in addition to gp2 and st1 volumes. CDP automatically provisions the gp3 volume type when conditions exist to reduce cost. The gp3 volume has slightly reduced baseline cost compared to gp2 and other advantages. Provisioning might occur when you specify scratch space limits for spilling Impala queries. Also, when you create an Impala Virtual Warehouse on new CDW environment, CDP uses a gp3 volume. For more information, see Amazon documentation.

Impala and Hive support for Iceberg equality deletes (Preview)

Cloudera supports row-level deletes, and in this release you can read equality deletes from Impala as well as Hive. For more information, see the Delete data feature.

Known issues in Cloudera Data Warehouse on Public Cloud

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

See Data Visualization release notes for known issues in Cloudera Data Visualization 7.1.9.

New Known Issues in this release

DWX-17455 ODBC client using JWT authentication cannot connect to Impala Virtual Warehouse

If you are using the Cloudera ODBC Connector and JWT authentication to connect to a CDW Impala Virtual Warehouse where the Impala coordinators are configured for high availability in an active-active mode, the connection results in a "401 Unauthorized" error. The error is not seen on Impala Virtual Warehouses that have an active-passive high availability or where high availability is disabled.

In the ODBC Data Source Name (DSN), make the following changes:

1. Ensure that SSL is enabled by setting SSL=1.
2. Remove authentication by setting AuthMech=0.
3. Remove the JWTString parameter.
4. Add the parameter http.header.Authorization and set it to the word "Bearer" followed by a space and the JWT string

DSN configuration before the workaround

```
SSL=1
AuthMech=8
JWTString=full_jwt_text
```

DSN configuration after the workaround

```
SSL=1
AuthMech=0
http.header.Authorization=Bearer full_jwt_text
```

IMPALA-12809 Metadata query problem from Impala

When querying metadata from Impala on a cluster with more than a few nodes, the query might fail with a null pointer exception. Reported problems occurred when running join queries on metadata, but other metadata queries might fail.

Limit the number of nodes that process the query by setting the NUM_NODES query option to 1:

```
SET NUM_NODES=1
```

HIVE-28055 Merging Iceberg branches requires a target table alias

Hive supports only one level of qualifier when referencing columns. In other words only one dot is accepted. For example, select table.col from ...; is allowed. select db.table.col is not allowed. Using the merge statement to merge Iceberg branches without a target or source table alias causes an exception:

```
org.apache.hadoop.hive.ql.parse.SemanticException: ... Invalid t
able alias or column reference ...
```

Use an alias, for example t, for the target table.

```
merge into mydb.target.branch_branch1 t using mydb.source.branch
_branch1 s on t.id = s.id when matched then update set value = '
matched';
```

DWX-17620: Folder having special characters in its name is not accessible in ABFS

In Cloudera Data Warehouse Public Cloud, from a Virtual Warehouse going to the Azure Blob Filesystem (ABFS), creating a folder, and then performing an action such as Move, causes an error as shown in the following example:

```
Cannot access: abfs://data-files/user/hrt_qa/~@$&()*!+'=;. 404 C
lient Error: The specified path does not exist.
```

Branch FAST FORWARD does not work as expected

The Apache Iceberg spec indicates you can use either one or two arguments to fast forward a branch. The following example shows using two arguments:

```
ALTER TABLE <name> EXECUTE FAST-FORWARD 'x' 'y'
```

However, omitting the second branch name, does not work as documented by Apache Iceberg. The named branch is not fast-forwarded to the current branch. An exception occurs at the Iceberg level.

You must use two arguments to the EXECUTE FAST FORWARD feature to forward a branch.

DWX-17613: Generic error message is displayed when you click on the directory you don't have access to on a RAZ cluster

You see the following error message when you click on an ABFS directory to which you do not have read/write permission on the ABFS File Browser in Hue: There was a problem with your request. This message is generic and does not provide insight into the actual issue.

None.

DWX-17109: ABFS File Browser operations failing intermittently

You may encounter intermittent issues while performing typical operations on files and directories on the ABFS File Broswer, such as moving or renaming files.

None.

CDPD-27918: Hue does not automatically pick up RAZ HA configurations

On a CDP Public Cloud environment in which you have configured RAZ in High Availability mode, Hue in CDW does not pick up all the RAZ host URLs automatically. Therefore, if a RAZ instance to which Hue is connected goes down, Hue becomes unavailable.

You must manually add comma-separated RAZ instances in the Hue Advanced Configuration Snippet.

1. Log in to the CDP Management Console as an Administrator.
2. Go to Environment Data Lake and open Cloudera Manager for your environment.
3. Go to Clusters Ranger RAZ service Instances RAZ server Processes and note the value of the fs.s3a.ext.raz.rest.host.url property from the core-site.xml file. You need this to specify the value of the api_url property in the Hue configuration.

   For Azure environments, note the value of the fs.azure.ext.raz.rest.host.url property.

   For AWS and GCS environments, note the value of the fs.s3a.ext.raz.rest.host.url property.
4. Go to CDW Virtual Warehouse ⋮ Edit CONFIGURATIONS and select the hue-safety-valve from the Configuration files dropdown menu.

5. Add the following lines in the hue-safety-valve field:

```
[desktop]
[[raz]]
is_enabled=true
api_url=https://[***INSTANCE-1***]:6082/,https://[***INSTANCE-
2***]:6082/
```

6. Click Apply Changes.

CDPD-66779: Partitioned Iceberg table not getting loaded with insert select query from Hive

If you create a partitioned table in Iceberg and then try to insert data from another table as shown below, an error occurs.

```
insert into table partition_transform_4 select t, ts from vector
tab10k;
```

Use the CLUSTER BY clause on the partitioned column to insert data. For example:

```
insert into table partition_transform_4 select t, ts from t1 clu
ster by ts;
```

DWX-17703: Non-HA Impala Virtual Warehouse on a private Azure Kubernetes Service (AKS) setup fails

When 'Refresh' and 'Stop' operations run in parallel, Impala might move into an error state. The Refresh operation might think that Impala is in an error state as the coordinator pod is missing.

Rebuild the Impala Virtual Warehouse or restart it using the CLI.

IMPALA-12742 DELETE/UPDATE Iceberg table partitioned by DATE fails

You can partition Iceberg tables using identities, such as int and date. The file path contains the partition value you can read. When you partition by DATE, subsequently running queries to update or delete data from partitioned data creates an incorrect conversion of DATE. The Catalog cannot parse the data and throws an error.

Carried over from the previous release: Upgrade-related

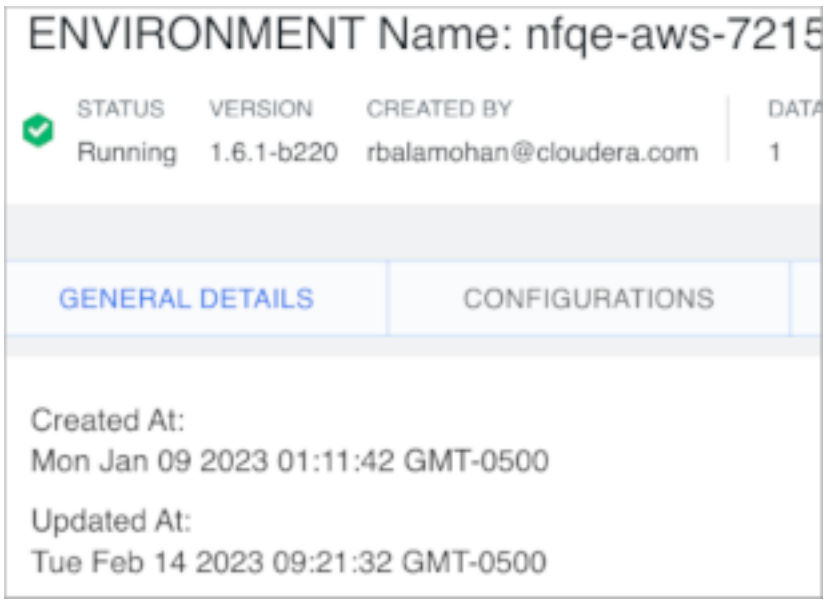
Upgrading to EKS 1.24 could result in Impala coordinators shutting down.

This issue is not seen on the Impala Virtual Warehouse running Runtime 2023.0.15.0-x or later.

Workaround: Manually start the Impala Virtual Warehouse from the UI or cli. Alternatively, replace your runtime with 1.7.1-b755 (released August 30, 2023) or later.

Certain CDW versions cannot upgrade to EKS 1.22

AWS environments activated in version 1.4.1-b86 (released June, 22, 2022) or earlier are not supported for upgrade to EKS 1.22 due to incompatibility of some components with EKS 1.22. To determine the activation version your pre-existing environment, in the Data Warehouse service, expand Environments. In Environments, search for and locate the environment that you want to view. Click Edit. In Environment Details, you see the CDW version.



DWX-15112 Enterprise Data Warehouse database configuration problems after a Helm-related rollback

If a Helm rollback fails due to an incorrect Enterprise Data Warehouse database configuration, the Virtual Warehouse and Database Catalog roll back to a previous configuration. The incorrect Enterprise Data Warehouse configuration persists, and can affect subsequent edit, upgrade, and rebuild operations on the rolled-back Virtual Warehouse or Database Catalog.

Carried over from the previous release: General

DWX-14923 After JWT authentication, attempting to connect the Impyla client using a user name or password should cause an error

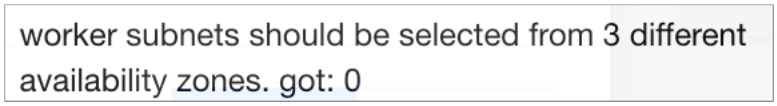
Using a JWT token, you can connect to a Virtual Warehouse as the user who generated the token.

If you connect with the JWT token, and then pass a user name or password from Impyla to the Virtual Warehouse, the connection arguments are silently ignored. Such an action should indicate that it is illegal to specify user or password when using JWT authentication.

DWX-15145 Environment validation popup error after activating an environment

Activating an environment having a public load balancer can cause an environment validation popup error.

To reproduce this problem 1) Create a data lake. 2) Activate an environment having a public load balancer deployment type and subnets in three different availability zones. A environment validation popup can occur even through subnets are in different availability zones. Several different popups can occur, including the following one:



DWX-15144 Virtual Warehouse naming restrictions

You cannot create a Virtual Warehouse having the same name as another Virtual Warehouse even if the like-named Virtual Warehouses are in different environments. You can create a Database Catalog having the same name as another Database Catalog if the Database Catalogs are in different environments.

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data

Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the Cr eateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersions line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

DWX-15214 DWX-15176 Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

The Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed:

```
kubelet Error: failed to create containerd task: failed to creat
e shim task: OCI runtime create failed: runc create failed: unab
le to start container process: exec: “run_httpd.sh”: cannot run
executable found relative to current directory: unknown
```

This issue affects all CDW releases before 1.6.3 (released May 5, 2023)running on CDP Public Cloud on Azure.

The following users are affected: CDW Azure users using Hue in Impala and Hive Virtual Warehouses in environments earlier than version 1.6.3 or later if they upgrade the node image

Workaround: Upgrade the Virtual Warehouses to 1.6.3 and choose Hue version 2023.0.14.0.

Do not choose Hue versions older than 2023.0.14.0 when creating a Virtual Warehouse in environments of version 1.6.3.

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Database Catalog

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

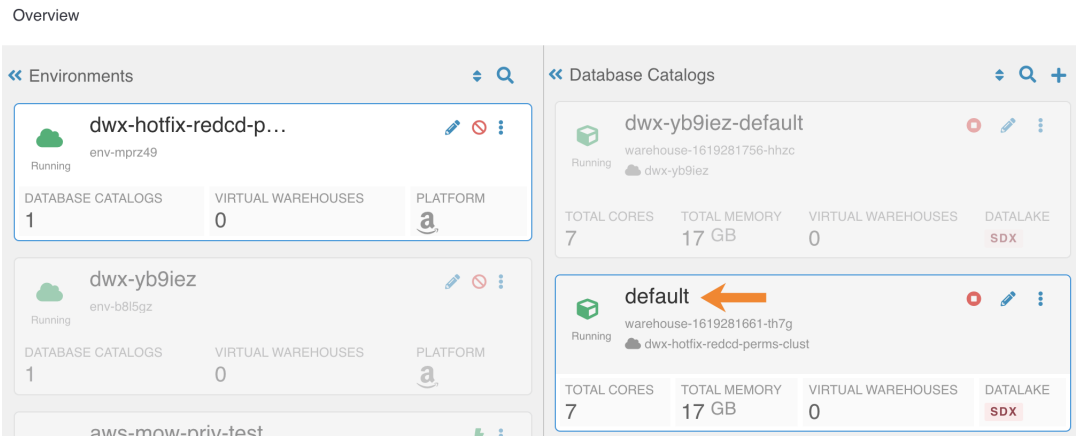
- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

Limited Hive image versions

Hive Virtual Warehouses you create in 1.8.1-b248 (released Nov 20, 2023) and later will run Istio 1.19.0. The new Istio version supports only new versions of Hive helm charts. If you have the CDW_VERSIONED_DEPLOY, only new Hive image versions appear in UI when you create a new Hive Virtual Warehouse.

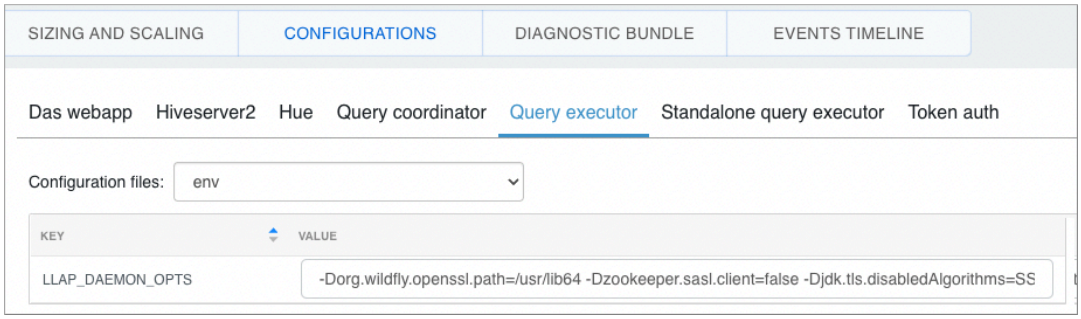
DWX-15064 Hive Virtual Warehouse stops but appears healthy

Due to an istio-proxy problem, the query coordinator can unexpectedly enter a not ready state instead of the expected error-state. Subsequently, the Hive Virtual Warehouse stops when reaching the autosuspend timeout without indicating a problem.

DWX-14452 Parquet table query might fail

Querying a table stored in Parquet from Hive might fail with the following exception message: java.lang.RuntimeException: java.lang.StackOverflowError. This problem can occur when the IN predicate has 243 values or more, and a small stack (-Xss = 256k) is configured for Hive in CDW.

Workaround: Change the value of the Xss in the JVM args to use the default value (1Mb) as follows: Select Edit from the Options menu of your Virtual Warehouse. Click Configurations > Query Executor > and select env.



In the third line shown below, change the value of LLAP_DAEMON_OPTS from -Xss256k to -Xss1M, and then click Apply Changes:

FROM:

-Dorg.wildfly.openssl.path=/usr/lib64 -Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Dhttp.maxConnections=12 -Xms24G -Xmx48G -Xss256k ...

TO:

... -Xss1M ...

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
       kind: ClusterRole
       metadata:
       name: diagnostic-data-generator-role-monitor
       rules:
       - apiGroups:
       - ""
       resources:
       - configmaps
       - events
       - pods
       - persistentvolumeclaims
       - nodes
       verbs:
       - get
       - list
       - apiGroups:
       - apps
       resources:
       - deployments
       - statefulsets
       verbs:
       - get
       - list
       - apiGroups:
       - "edws.cloudera.com"
       resources:
       - computes
       verbs:
```

```
        - get
        - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.

7. Click Apply in the upper right corner of the page to save the configuration.
8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name ‘ec2-18-219-
32-183.us-east-2.compute.amazonaws.com’ does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Hue Query Editor

DWX-16899: Error while viewing Impala job status on the mini Job Browser

When you click on the application ID after submitting an Impala query on Hue that is running on the environment level, you may notice the following error: 401 Client Error: Unauthorized for url: http://coordinator.impala-xyz.svc.cluster.local:25000/queries?json=true Must authenticate with Basic authentication. This does not happen when you do the same from Hue deployed at a Virtual Warehouse level.

None.

DWX-16913: “Results have expired” message while running a CTAS query

You may see the following message on the Hue UI when you submit a CREATE TABLE AS SELECT (CTAS) query from a Hue instance that is deployed on the environment level: “Results have expired, rerun the query if needed”. This does not happen when you do the same from Hue deployed at a Virtual Warehouse level.

None.

DWX-16917: Failed to validate proxy privileges error while running queries from Hue

You may see the following error intermittently in Hue’s pod logs while running queries from a Hue instance that is deployed at the environment level: “Failed to validate proxy privileges of <username>”.

None.

DWX-16895: Incorrect status of Hue pods when you edit the Hue instance properties

When you update a configuration of a Hue instance that is deployed at the environment level, such as increasing or decreasing the size of the Hue instance, you see a success message on the CDW UI. After some time, the status of the Hue instance also changes from “Updating” to “Running”. However, when you list the Hue pods using kubectl, you see that not all backend pods are in the running state–a few of them are still in the init state.

None. The pods come up successfully eventually after a sufficient time has passed.

DWX-16863: The upgrade button is present on the CDW UI, but Hue upgrades are not supported

You see the Upgrade button on the **Query Editor** page in the CDW UI when Hue is deployed at the environment level. However, on CDW version 1.8.1, upgrading the Hue instance that is deployed at the environment level is not supported.

None.

DWX-16893: A user can see all the queries in Job browser

In a Hue instance deployed at the environment level, by design, the Hue instances must not share the saved queries and query history with other Hue instances even for the same user. However, a logged in user is able to view all the queries executed by that user on all the Virtual Warehouses on a particular Database Catalog.

None.

Delay in listing queries in Impala Queries in the Job browser

Listing an Impala query in the Job browser can take an inordinate amount of time.

DWX-14927 Hue fails to list Iceberg snapshots

Hue does not recognize the Iceberg history queries from Hive to list table snapshots. For example, Hue indicates an error at the . before history when you run the following query.

```
select * from <db_name>.<table_name>.history
```

DWX-14968 Connection termination error in Impala queries tab after Hue inactivity

To reproduce the problem: 1) In the Impala job browser, navigate to Impala queries. 2) Wait for a few minutes.

After a few minutes of inactivity, the follow error is displayed:

Workaround: Refresh the page, or alternatively start a new session.

DWX-15115 Error displayed after clicking on hyperlink below Hue table browser

In Hue, below the table browser, clicking the hyperlink to a location causes an HTTP 500 error because the file browser is not enabled for environments that are not Ranger authorized (RAZ).

DWX-15090: CSRF error intermittently seen in the Hue Job Browser

You may intermittently see the “403 - CSRF” error on the Hue web interface as well as in the Hue logs after running Hive queries from Hue.

Reload the browser or start a new Hue session.

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Carried over from the previous release: Iceberg

Concurrent compactions and modify statements can corrupt Iceberg tables

Hive or Impala DELETE/UPDATE/MERGE operations on Iceberg V2 tables can corrupt the tables if there is a concurrent table compaction from Spark. The issue happens if the compaction and modify statement runs in parallel, and if the compaction job commits before the modify statement. In that case the modify statement's position delete files still point to the old files. The results in the case of DELETE and in the case of UPDATE / MERGE are as follows:

- DELETE

  Delete records pointing to old files have no effect.
- UPDATE / MERGE

  Delete records pointing to old files have no effect. The table will also have the newly added data records, which means rewritten records will still be active.

Use one of the following workarounds:

- Do not run compactions and DELETE/UPDATE/MERGE statements in parallel.
- Do not compact the table via Iceberg's RewriteFiles operation. For example do not use Spark's rewriteDataFiles.

DWX-17210: Timeout issue querying Iceberg tables from Hive

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

- fs.s3.maxConnections=1000
- fs.s3a.connection.maximum=1000

Restart the Database Catalog and Virtual Warehouse.

DWX-15014 Loading airports table in demo data fails

This issue occurs only when using a non-default Database Catalog. A invalid path error message occurs when using the load command to load the demo data airports table in Iceberg, ORC, or Parquet format.

DWX-14163 Limitations reading Iceberg tables in Avro file format from Impala

The Avro, Impala, and Iceberg specifications describe some limitations related to Avro, and those limitations exist in CDP. In addition to these, the DECIMAL type is not supported in this release.

DEX-7946 Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata if you drop the table during migration from Hive to Iceberg.

Workarounds: Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13733 Timeout issue querying Iceberg tables from Hive

A timeout issue can cause the query to fail.

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

```
fs.s3.maxConnections=1000
fs.s3a.connection.maximum=1000
```

Restart the Database Catalog and Virtual Warehouse.

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: Impala Virtual Warehouse

TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

The Cloudera Data Warehouse (CDW) Public Cloud 2023.0.14.0 (DWX-1.6.3) version incorporated a feature for performing automatic metadata synchronization across multiple Apache Impala (Impala) Virtual Warehouses. The feature is enabled by default, and relies on the Hive MetaStore events. When a certain sequence of Data Definition Language (DDL) SQL commands are executed as described below, users may encounter a java.lang.NullPointerException (NPE). The exception causes the event processor to stop processing other metadata operations.

If a CREATE TABLE command (not CREATE TABLE AS SELECT) is followed immediately (approximately within 1 second interval) by INVALIDATE METADATA or REFRESH TABLE command on the same table (either on the same Virtual Warehouse or on a different one), there is a possibility that the second command will not find the table in the catalog cache of a peer Virtual Warehouse and generate an NPE.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

DWX-17175: Impala Virtual Warehouse max executor limit issue

The Impala Virtual Warehouse has a limit of max 200 executors. The Executors slider has a max limit of 200 for creating or editing a Virtual Warehouse. If you select a bigger t-shirt size than the default, or set a custom t-shirt size, the default max limit set by the UI could be more than 200 in some cases. Exceeding the max limit causes an error during a future Edit/Upgrade/Rebuild operation.

Workaround: Manually set the max executors limit to a preferred value less than 200, but a multiple of t-shirt size, when creating or editing Impala Virtual Warehouse

DWX-15016 Impala query failure when using Catalog high availability (HA)

Impala queries could fail with InconsistentMetadataFetchException when using HA feature in CDW. This happens when leader election failover for Impala Catalog Service happens due to a transient network error.

Workaround: Disable Catalog HA. In CDW, edit your Virtual Warehouse, and turn off Enable Impala Catalog Server HA.

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.

2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

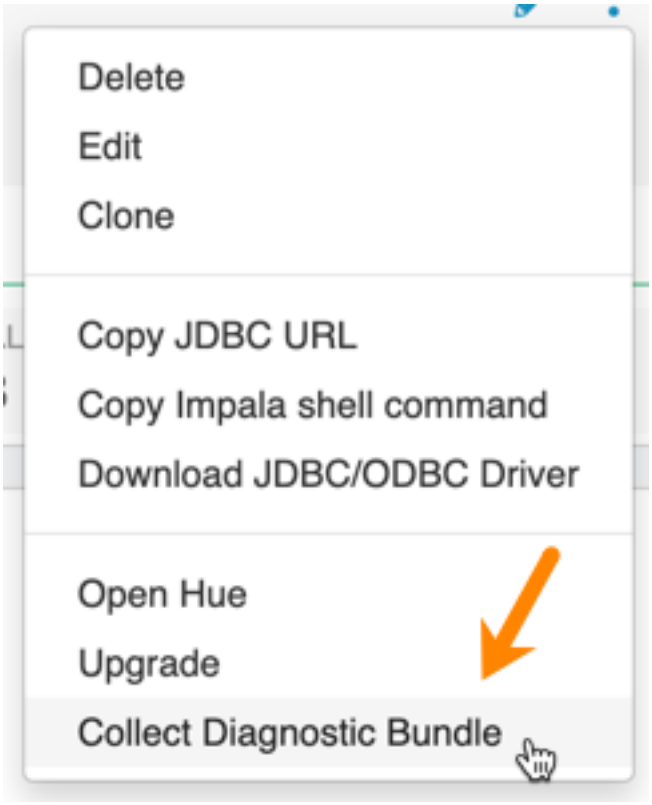
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

UI Issues

DWX-14610 Upgrade icon indicates the Database Catalog is the latest version, but might be incorrect for a few seconds

The cluster fetches data regularly in particular time intervals, or when an action is triggered. Inbetween this time interval, which is very brief, the cluster might not be updated, but will be refreshed in a few seconds.

Technical Service Bulletins

TSB 2023-719: Cloudera Data Warehouse Backup/Restore of CDP Data Visualization incomplete

Cloudera Data Warehouse (CDW) customers using the CDW Automated Backup/Restore feature will encounter an issue with the restoration versions of Cloudera Data Visualization (CDV) older than CDV 7.1.6.2-3 due to a schema change in this release. If the Backup was taken from an older CDW environment that contained a version of CDV older than CDV 7.1.6.2-3, the Restore procedure will succeed. Though once the user opens the CDV Queries tab, the user could encounter the error message: “column jobs_jobschedule.owner_id does not exist…”

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-719: Cloudera Data Warehouse Backup/Restore of Cloudera Data Visualization incomplete.

TSB 2024-743: Restoration of the Database Catalog can be incomplete due to an incorrect Hue Query Processor configuration in CDW

Cloudera Data Warehouse (CDW) customers using the CDW Automated Backup/Restore feature might encounter an issue with the restoration of the Database Catalog, if the hue-query-processor.json configuration file of the Database Catalog has been edited. Even a minor edit to the hue-query-processor.json configuration file can result in this failure.

During the restoration process the Database Catalog will be created, but it will fail to start after a short period of time, and the Database Catalog will be in a Bad Health state on the CDW User Interface.

Inside the Kubernetes Cluster (Azure Kubernetes Service on Azure / Elastic Kubernetes Service on AWS) the StatefulSet of Hue Query Processor (hue-query-processor) is in CrashLoop. This is indicated with the following log in the Hue Query Processor StatefulSet Pod:

```
SQL State  : 3D000
Error Code : 0
Message    : FATAL: database "warehouse-1707832123-abcd_hueqpdb"
 does not exist



              at org.flywaydb.core.internal.jdbc.JdbcUtils.openC
onnection(JdbcUtils.java:65)
```

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-743: Restoration of the Database Catalog can be incomplete due to an incorrect Hue Query Processor configuration in CDW.

TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables

Apache Hive (Hive) and Apache Impala (Impala) modify statements (DELETE/UPDATE/MERGE) on Apache Iceberg (Iceberg) V2 tables can corrupt the tables if there is a concurrent table compaction from Apache Spark. The issue happens when the compaction and modify statement run in parallel, and when the compaction job commits before the modify statement. In this case the position delete files of the modify statement still point to the old files. This means the following in case of

- DELETE statements
  - Deleting records pointing to old files have no effect
- UPDATE / MERGE statements
  - Deleting records pointing to old files have no effect
  - The table will also have the newly added data records
  - Rewritten records will still be active

This issue does not affect Apache NiFi (NiFi) and Apache Flink (Flink) as these components write equality delete files.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables

TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

The Spark Iceberg library includes two procedures - rewrite_data_files and rewrite_position_delete_files. The current implementation of rewrite_data_files has a limitation that the position delete files are not deleted and still tracked by the table metadata, even if they no longer refer to an active data file. This is called the dangling delete problem. To solve this, the rewrite_position_delete_files procedure is implemented in the Spark Iceberg library to remove these old “dangling” position delete files.

Due to the dangling delete limitation, when an Iceberg table with dangling deletes is queried in Impala, Impala tries to optimize select count(*) from iceberg_table query to return the results using stats. This optimization returns incorrect results.

The following conditions must be met for this issue to occur:

- All delete files in the Iceberg table are “dangling”
  - This would occur immediately after running Spark rewrite_data_files AND
    - Before any further delete operations are performed on the table OR
    - Before Spark rewrite_position_delete_files is run on the table
- Only stats optimized plain select count(*) from iceberg_table queries are affected. For example, the query should not have:
  - Any WHERE clause
  - Any GROUP BY clause
  - Any HAVING clause

Remove dangling deletes: After rewrite_data_files, position delete records pointing to the rewritten data files are not always marked for removal, and can remain tracked by the live snapshot metadata of the table. This is known as the dangling delete problem.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables.

TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion

When working with Apache Hive (Hive) and Apache Iceberg (Iceberg) V2 tables, using the TRUNCATE statement may lead to unintended data deletion. This issue arises when the truncate command is applied to a branch of an Iceberg table. Instead of truncating the branch itself, the command affects the original (main) table, which results in unintended loss of data.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion.

Fixed issues in Cloudera Data Warehouse on Public Cloud

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud fixes these issues.

Hive-27643 CPDP-60943 Data Loss during compaction

Data loss no longer occurs during compaction when Apache Ranger policies for masking or row filtering are enabled and compaction users are included in the policies. Compaction queries are now automatically excluded from ALL ranger policies.

Enabling a private CDW environment in Azure Kubernetes Service is unstable

The networking issue at Azure Software Definition Layer (SDN) has been fixed by Microsoft. The issue affected CDW private environments while using the Azure Kubernetes Service (AKS) in 1.7.1-b755 (released Aug 30, 2023) - 1.8.4-b90 (released Jan 18, 2024). Enabling a private CDW environment in AKS is now available in 1.8.5-b35 (released Feb 29, 2024) on a General Availability (GA) basis. Cloudera now supports deploying a private CDW environment in Azure using AKS.

IMPALA-12681: Some local file descriptors not released when using remote spilling

The issue that occurred during remote spilling when writing spilled data to local buffers has been fixed. The disk space occupied by the file can now be reclaimed.

DWX-17368: Missing Root CA certificates in LdapConfig for ldaps should not result in error

If Root CA certificates are not passed as LdapConfigs, the certificates in the default path are used. In public cloud, DWX needs these certs for tls communication. This issue has been fixed by changing the error condition into informational log when CA certificates are not provided as part of LdapConfigs.

DWX-17157: Enable/Disable SSO doesn't set proper Impala configs

The CDW update logic now updates the SSO config related check to compare the current value of `EnableSSO` with previous value of `EnableSSO`.

DWX-17053: Catalog heap to pod memory configuration

The memory issue causing catalogd problems has been fixed by reserving 33% more memory for the catalogd pod than the maximum heap size for JVM.

DWX-17175: Impala Virtual Warehouse max executor limit issue

Earlier, you had to manually set the max executors limit to a preferred value less than 200, but a multiple of t-shirt size, when creating or editing Impala Virtual Warehouse. This issue has been resolved.

Technical Service Bulletins

TSB 2024-745: Impala returns incorrect results for Iceberg V2 tables when optimized operator is being used in CDW

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-745: Impala returns incorrect results for Iceberg V2 tables when optimized operator is being used in CDW.

Behavior changes in Cloudera Data Warehouse on Public Cloud

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following behavior changes:

Summary:

Change in value for the hive.map.groupby.sorted property

Before this release:

The value of the hive.map.groupby.sorted property was set to 'true'.

After this release:

The value of the hive.map.groupby.sorted property is changed to 'false' to disable optimization. The change was introduced to address data correctness issues noticed in query results on tables with CLUSTER BY and SORT BY.

For more information, see HIVE-27876.

Summary:

Change in the upgrade and rebuild functionality

Before this release:

CDW provided the following three options to upgrade and rebuild Database Catalogs and Virtual Warehouses: Rebuild with same image version, rebuild with new image version, and upgrade.

After this release:

The KeepImageVersion option that was displayed when you rebuild the Virtual Warehouse has been removed. The rebuild operation preserves the image version and only redeploys the pods. The upgrade operation updates the CDW entities with the latest image version.

What's new (unsupported releases)

This section lists major features and updates for the Cloudera Data Warehouse (CDW) service on Public Cloud.

January 18, 2024

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

This release is a hot fix for an environment activation issue in CDW and other CDP data services. The latest EKS cluster deployment from AWS changes behaviors of CoreDNS. One of the changes breaks the functions of cdp-coredns-updater, causing a malfunction in CoreDNS. CDW 1.8.4 fixes the corefile issue and is compatible with the latest CoreDNS. For more information, see Known Issues, January 10, 2024.

January 10, 2024

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

AWS announced the Amazon Relational Database (RDS) Postgres 11 end of life is Feb 29, 2024. In Cloudera Data Platform, in an AWS environment, any new CDW created in this release 1.8.3-b130 Jan 10, 2024 or later will support Postgres 13 on RDS.

Any existing CDW in an AWS environments that uses Postgres 11 requires backup and restoration. You backup CDW in an AWS environment that supports Postgres 11 on RDS, and then restore CDW on an environment that supports Postgres 13 on RDS as described in "Backing up and restoring CDW".

December 19, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Upgrade your Virtual Warehouse to get the new CDW Runtime 2023.0.16.2-1 (Dec 19, 2023), which makes the following changes:

- CDPD-48059

  Unable to view and access S3 buckets from Hue configured with RAZ in CDW Public Cloud. For more information, including how to get this fix, see Fixed Issues for December 19, 2023.

- DWX-16923

  Fixes upgrade problem starting Hive Virtual Warehouse after upgrade.
- DWX-17088

  Addresses memory issues by disabling collection of KUBE-EVENTS for diagnostic bundle.
- DWX-17097

  Corrects metering problem causing overcharges.

  DWX-17110

  Populates several Grafana panels that were empty.
- DWX-17111 IMPALA-12486

  Missing database and table count metrics now appear in Grafana. For more information, including how to get this fix, see Fixed Issues for December 19, 2023.
- IMPALA-12580

  Fixes Iceberg exception when queries from Impala filter predicates.

December 1, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Upgrade your Virtual Warehouse to get the new CDW Runtime 2023.0.16.1-2 (released Dec 1, 2023), which fixed the following issues:

- CDPD-62960
- DWX-16690

See fixed issues for more information about the issues.

November 20, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Cloudera Data Warehouse Public Cloud 1.8.1-b248 changes, described in more detail below:

- Azure AKS 1.27 upgrade on page 112
- New AWS instance type on page 112
- Upgraded AWS and Azure environment security on page 112
- Diagnostic bundles for troubleshooting Data Visualization problems on page 112
- Resizing a Data Visualization instance on page 112
- Reduced permissions mode enhancement and tag change on page 113
- dbt adapters for using dbt with Hive, Impala and CDP on page 113

Cloudera Data Warehouse Public Cloud Runtime 2023.0.16.0-150 changes:

Hue

- Hue supports natural language query processing (Preview) on page 114
- Ability to deploy Hue at the environment level (Preview) on page 115

Iceberg

- Enhancement of expiring Iceberg snapshots on page 113
- Truncate partition support for Iceberg on page 113
- Insert into partition and insert overwrite partition support for Iceberg on page 113
- Iceberg branching and tagging technical preview on page 113

Impala

- Impala performance optimization for VALUES() expressions on page 113
- Impala skips scheduling bloom filter from full-build scan on page 114

- Impala events marked as 'Skip' occur prior to a manual REFRESH on page 114
- Allow setting separate mem_limit for coordinators on page 114

Azure AKS 1.27 upgrade

Cloudera supports the kubernetes version 1.27. In 1.8.1-b248 (released November 20, 2023), when you activate an environment, CDW automatically provisions Azure Kubernetes Service (AKS) 1.27. To upgrade to AKS 1.27 from 1.7.3 or earlier, you must backup and restore CDW. To avoid compatibility issues between CDW and AKS, upgrade to version 1.27.

Using the Azure CLI or Azure portal to upgrade the AKS cluster is not supported. Doing so can cause the cluster to become unusable and can cause downtime. For more information about upgrading, see Upgrading an Azure Kubernetes Service cluster for CDW.

New AWS instance type

This release supports the r6id.4xlarge AWS compute instance types. You select this instance type, or another supported one, when you activate your environment in CDW.

Private EKS API Server technical preview

In 1.8.1-b248 (released November 20, 2023), you can establish private connectivity between CDP services running on you cluster and AWS to prevent exposing data on the internet. To set up the Amazon Elastic Kubernetes Service (EKS) cluster in private mode and to enable the private EKS, run the following Beta CDP CLI command:

```
cdp dw create-cluster -\-aws-options "enablePrivateEKS=true" -\-environment-
crn "xyz"
```

The set up configures awsOptions. In this mode, a private endpoint uses the cluster-proxy (ccmv2) networking for control-plane to cluster communication. The freeipa security group is authorized in the eks-cluster's ingress rule, which opens the channel of communication from the control plane side through the freeipa pod.

This feature is a technical preview and not recommended for production deployments. Cloudera recommends that you use this feature in test and development environments only.

Upgraded AWS and Azure environment security

This release upgrades Istio security required for future Kubernetes compatibility. Hive Virtual Warehouses you create in 1.8.1-b248 (released Nov 20, 2023) will run Istio 1.19.0. Because the new Istio version supports only new versions of Hive helm charts, the following limitation exists: If you have the CDW_VERSIONED_DEPLOY, only new Hive image versions appear in UI when you create a new Hive Virtual Warehouse. For more information, see the known issue about Limited Hive image versions.

Diagnostic bundles for troubleshooting Data Visualization problems

You can collect a diagnostic bundle for troubleshooting Data Visualization, as well as Virtual Warehouse, Database Catalog, and environment/cluster. The diagnostic bundle is available for downloading and troubleshooting from the UI. For more information, see Diagnostic bundles for CDW and Kubernetes.

Resizing a Data Visualization instance

The size of the your Data Visualization instance is critical for achieving cost and performance goals. In CDW, after creating a Data Visualization instance you can change its size. You open the Data Visualization instance for editing, and in Size, select the size you want. When you click Apply Changes, the new size takes effect.

Reduced permissions mode enhancement and tag change

The dependent name of the tag key envID has been changed to Cloudera-Resource-Name, which increases strictness. For more information, see Reduced permissions mode template and Minimum set of IAM permissions required for reduced permissions.

dbt adapters for using dbt with Hive, Impala and CDP

You can access the dbt adapters for Hive and Impala from the Cloudera Data Warehouse service, which enable you to use the dbt data management workflow with Cloudera Data Platform. For more information, see Using dbt with Hive, Impala and CDP.

Support for Hive external data sources using data connectors

You can use Hive data connectors to map databases present in external data sources to a local Hive Metastore (HMS). The external data sources can be of different types, such as MySQL, PostgreSQL, Oracle, Redshift, Derby, or other HMS instances. You can create external tables to represent the data, and then query the tables. For more information, see Using Hive data connectors to support external data sources.

Enhancement of expiring Iceberg snapshots

In this release, you have more flexibility to expire snapshots. In addition to expiring snapshots older than a timestamp, you can now expire snapshots based on the following conditions:

- A snapshot having a given ID
- Snapshots having IDs matching a given list of IDs
- Snapshots within the range of two timestamps

You can keep snapshots you are likely to need, for example recent snapshots, and expire old snapshots. For example, you can keep daily snapshots for the last 30 days, then weekly snapshots for the past year, then monthly snapshots for the last 10 years. You can remove specific snapshots to meet the GDPR right to be forgotten requirements.

Truncate partition support for Iceberg

This release introduces the capability to truncate an Iceberg table. Truncation removes all rows from the table. A new snapshot is created. Truncation works for partitioned and unpartitioned tables.

Insert into partition and insert overwrite partition support for Iceberg

From Hive you can insert into, or overwrite data in, Iceberg tables that are statically or dynamically partitioned. For syntax and limitations, see "Insert into/overwrite partition support".

Iceberg branching and tagging technical preview

From Hive, you can manage the lifecycle of snapshots using the Iceberg branching and Iceberg tagging features. Branches are references to snapshots that have a lifecycle of their own. Tags identify snapshots you need for auditing and conforming to GDPR. Branching and tagging is available as a technical preview. Cloudera recommends that you use this feature in test and development environments. It is not recommended for production deployments.

Impala performance optimization for VALUES() expressions

Rewriting expressions in the Impala VALUES() clause can affect performance. If Impala evaluates an expression only once, especially a constant expression, the overhead might outweigh the potential benefits of a rewrite. Consequently, in this release, any attempts to rewrite expressions within the VALUES() clause has no impact. Impala skips expression rewrites entirely for VALUES during the analysis phase except rewrites of expressions separated by a compound vertical bar (||). These expressions are ultimately evaluated and materialized in the backend instead of the analysis phase.

The drop in performance caused by rewriting expressions might not follow a straightforward linear pattern. This drop becomes more pronounced as the number of columns increases. Using code generation for constant expressions in

this context does not provide significant value. As part of this optimization, code generation is turned off for constant expressions within a UNION node if the UNION node is not within a subplan. This applies to all UNION nodes with constant expressions, not just those associated with a VALUES clause.

Here are some examples of queries of disabled expression rewrites and code generation for the UNION operator and VALUES clause.

```
select 1+2+3+4+5,2*1-1,3*3 union all select 1+2+3+4,5,6 union all select 7+1
-2,8+1+1,9-1-1-1;
```

```
insert into test_values_codegen values
  (1+1, '2015-04-09 14:07:46.580465000', base64encode('hello world')),
  (CAST(1*2+2-5 as INT), CAST(1428421382 as timestamp),
   regexp_extract('abcdef123ghi456jkl','.*?(\\d+)',0));
```

Impala skips scheduling bloom filter from full-build scan

PK-FK joins between a dimension table and a fact table are common occurrences in a query. Such joins often do not involve any predicate filters in the dimension table. As a result, a bloom filter generated from this kind of dimension table scan (PK) will most likely contain all values from the fact table column (FK). It becomes ineffective to generate this filter because it is unlikely to reject any rows, especially if the bloom filter size is large and has a high false positive probability (FPP) estimate.

As part of this optimization, Impala skips scheduling bloom filter from join node that has certain characteristics.

Impala events marked as 'Skip' occur prior to a manual REFRESH

If a table has been manually refreshed, the event processor skips any events occurring prior to the manual refresh. This optimization helps catalogd when it lags behind in processing events. Now, event processing has been optimized to determine if any manual refresh was executed after its eventTime. This, in turn, assists CatalogD in swiftly catching up with the HMS events. To activate this optimization, you must set enable_skipping_older_events to true.

Allow setting separate mem_limit for coordinators

The current mem_limit query option applies to all Impala coordinators and executors. This means that the same amount of memory gets reserved, but coordinators typically only handle the task of coordinating the query and thus may not necessarily require all the estimated memory. When we block the estimated memory on coordinators, it hinders the memory available for use by other queries.

The new MEM_LIMIT_COORDINATORS query option functions similarly to the MEM_LIMIT option but sets the query memory limit only on coordinators. This new option addresses the issue related to MEM_LIMIT and is recommended in scenarios where the query needs higher or lower memory on coordinators compared to the planner estimates.

Note: The MEM_LIMIT_COORDINATORS query option does not work in conjunction with MEM_LIMIT. If you set both, only MEM_LIMIT will be applied.

Hue supports natural language query processing (Preview)

Hue leverages the power of Large Language Models (LLM) to help you generate SQL queries from natural language prompts and also provides options to optimize, explain, and fix queries, ensuring efficiency and accuracy in data retrieval and manipulation. You can use several AI services and models such as OpenAI's GPT service, Amazon Bedrock, and Azure's OpenAI service to run the Hue SQL AI assistant. See SQL AI Assistant in Data Warehouse Public Cloud.

Ability to deploy Hue at the environment level (Preview)

Previously, you could run Hue only at a Virtual Warehouse-level. As a result, you would lose all query history when you delete or shut down a Virtual Warehouse. CDW now allows you to deploy Hue at the environment level and allows you to select a Virtual Warehouse from the Hue web interface. Query history and saved queries are retained as long as the environment is active. See Deploying Hue at Environment Level in Data Warehouse Public Cloud.

October 12, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

This release makes a HotFix available for the following issues:

- DWX-16387 : Hue UI unreachable after huebackend restart.
- DWX-16690: Unable to edit any mutable config present in flag file of impala-catalogd.

October 5, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Automatically backing up and restoring CDW

In earlier releases, the manual backup and restore feature replaced the in-place upgrade of AWS environments. This release adds automation to back up and restore procedures for AWS and Azure environments. You can still perform Hive and Impala in-place upgrades.

To get the supported Kubernetes version, you back up your old AWS or Azure environment and start up a new environment using the restoration process. The backup/restore feature saves your environment parameters, making it possible to recreate your environment with the same settings, URL, and connection strings you used in your previous environment.

CDP CLI Start/Stop Azure Kubernetes Service (AKS) Technical Preview

Using the Beta version of the CDP CLI, you can now start and stop a Cloudera Data Warehouse (CDW) cluster on Azure using the following commands:

- resumeCluster <cluster ID>

  Starts the cluster.
- suspendCluster <cluster ID>

  Stops the cluster.

The following prerequisites are required to use resumeCluster and suspendCluster :

- You obtained a cluster ID parameter to pass with the command.
- The cluster you want to stop is running.
- You stopped the Virtual Warehouses and Database Catalogs in the cluster you want to stop
- You stopped the cluster before attempting to start it.
- The cluster you want to start or stop is not in an error state.

The Microsoft documentation describes how resumeCluster and suspendCluster starts and stops the AKS cluster, including every node pool and AKS control plane, on the cloud provider side. The stopped AKS incurs zero cost. The Postgres database, which belongs to the environment, is not stopped.

Note: Heed the Microsoft documentation, which says "Don't repeatedly stop and start your clusters. This can result in errors. Once your cluster is stopped, you should wait at least 15-30 minutes before starting it again."

Limitations

- The capability to start and stop an AKS cluster is not available from the UI.
- You cannot start or stop an AKS cluster from the UI.

- You cannot manually trigger starting or stopping an AKS cluster from the cloud provider either by using the Azure CLI or the Azure portal.

  Manually triggering start and stop on the cloud provider does not synchronize the state of the AKS instance with Cloudera control plane. The AKS continues running and the CDW UI indicates a running state.

Impala Virtual Warehouse spills temporary data to S3 automatically

When you create an Impala Virtual Warehouse, Impala is now automatically configured to write temporary data to the S3 Data Lake bucket. No configuration is required to enable the spill to S3. However, you must configure an Impala policy for permission to the scratch location on the Data Lake bucket.

Kubernetes dashboard adds Azure support and has reached GA status

From AWS or Azure environments, you can use the K8S dashboard to view the state of resources, such as CPU and memory usage, see the status of pods, and download logs. The dashboard can provide insights into the performance and health of a CDW cluster.

August 30, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Cloudera Data Warehouse Public Cloud 1.7.1-b755 changes, described in more detail below:

- Amazon Machine Image updates on page 117
- AWS Elastic Kubernetes Service 1.26 support on page 117
- Amazon EKS 1.23 and from EKS 1.23 to 1.24 upgrades on page 117
- AWS restricted policy updates on page 117
- Automatic backup and restoration of Hue on page 117
- Correcting the Virtual Warehouse Size on page 117
- Deprecation of DAS on page 117
- Instance types available for compute nodes on page 117
- Java tool options now configurable with limitations on page 117
- Managing high partition loads on page 117
- Private Cluster control removed from UI on page 118
- Query results cache support on page 118

Cloudera Data Warehouse Public Cloud Runtime 2023.0.15.0-243 changes:

Hive

- Histogram statistics on page 118

Iceberg

- Iceberg position delete feature support on page 118
- Hive to Iceberg table migration from Impala on page 118

Impala

- Impala support for complex types on page 118
- Support ORDER BY for collections of fixed length types in SELECT list on page 118
- Support collections of fixed length types as non-passthrough children of unions on page 118
- Allow implicit casts between numeric and string types when inserting into table on page 119
- Improved memory estimation for aggregates and pre aggregates on page 119
- Improved CPU costing on page 119
- Added processing cost to control scan parallelism on page 119
- Impala WebUI improvements on page 119
- Skip reloading file metadata for some ALTER_TABLE events on page 120
- JWT auth for Impala on page 120

- Native High Availability (HA) for Impala Catalog Service on page 120
- Codegen for STRUCT type on page 120

Amazon Machine Image updates

This release supports dynamically updating the Amazon Machine Image (AMI) to prevent potential problems running workloads on an old AMI. You can update the AMI of the Cloudformation stack while keeping the current Elastic Kubernetes Service (EKS) version.

AWS Elastic Kubernetes Service 1.26 support

New CDW clusters using AWS environments you activate in this release 1.7.1-b755 (released August 30, 2023) of Cloudera Data Warehouse will use Amazon Kubernetes (EKS) version 1.26. For more information, see Upgrading Amazon Kubernetes Service.

Amazon EKS 1.23 and from EKS 1.23 to 1.24 upgrades

The release includes reduced permissions mode support for the following Amazon Elastic Kubernetes Service upgrades:

- Upgrading to EKS 1.23
- Upgrading from EKS 1.23 to 1.24

AWS restricted policy updates

The AWS restricted policy has been updated to conform to AWS file size requirements. The update divides the policy into two parts as described in the documentation.

Automatic backup and restoration of Hue

The CDW backup and restore processes for Hue have been automated. Manually backing up and restoring Hue is still available, but optional.

Correcting the Virtual Warehouse Size

After creating an Impala Virtual Warehouse, you can tune, or correct the T-shirt size of executor groups that drive the Impala Virtual Warehouse. The size of the executor groups is critical for achieving cost and performance goals.

Deprecation of DAS

Hue now replaces Data Analytics Studio (DAS). DAS has been deprecated and is no longer available in CDW Public Cloud. DAS features to support Hive and Tez such as running queries, defining HPL/SQL, the Job Browser, query explorer, query compare, and more, have been migrated to Hue, and the Hue Query Processor. After you upgrade to this release, you will not see the option to launch DAS from your Virtual Warehouse. Cloudera recommends you use Hue for all use cases where you might have previously used DAS.

Instance types available for compute nodes

In this release, when you activate an AWS environment, you can select the compute instance type you want to use. This release adds additional instance types you can select when you activate an Azure environment.

Java tool options now configurable with limitations

After creating an Impala Virtual Warehouse, you can change the XMX Java tool option.

Managing high partition loads

In this release, you can identify an error related to high partition workloads and tune your Hive Virtual Warehouse to run successfully.

Private Cluster control removed from UI

Enable Private Cluster has been removed from the environment activation dialog. Use CDP CLI for advanced configurations.

Query results cache support

Unified Analytics now supports the caching of Hive/Impala query results. Caching results of repetitive queries can reduce the load.

Histogram statistics

In this release, when you generate column statistics in a Hive Virtual Warehouse in Unified Analytics mode, you can create histogram statistics on columns. By default, the histogram stats are not created. You enable generation of histogram statistics by setting a Hive property: set hive.stats.kll.enable = true;

You can then run the ANALYZE command as usual:

```
ANALYZE TABLE [table_name] COMPUTE STATISTICS for COLUMNS [comma_separated_c
olumn_list];
```

Histogram statistics are supported for numeric data types, date, timestamp and boolean types but not for string/varchar/char columns. Histograms are used to estimate selectivity of range predicates (predicates involving <, <=, >, >= and BETWEEN). The better selectivity estimate allows the optimizer to generate more optimal query plans and improve performance for such queries.

Hive to Iceberg table migration from Impala

In this release, you can use Impala, as well as Hive, to migrate a Hive table to Iceberg tables. You use the ALTER TABLE statement. Syntax is described in Migrate Hive table to Iceberg feature and a step-by-step procedure is covered in Migrating a Hive table to Iceberg.

Iceberg position delete feature support

In this release, Impala, in addition to Hive, can delete Iceberg V2 tables using position delete files, a format defined by the Iceberg Spec. A position delete query evaluates rows from one table against a WHERE clause, and delete all the rows that match WHERE conditions.

Impala support for complex types

Complex types are now supported in the SELECT list. Although collections and structs were previously supported, nesting and mixing of complex types was not. For more information, including limitations, see "Allowing embedding complex types into other complex types" in Complex types.

Support ORDER BY for collections of fixed length types in SELECT list

This release supports collections of fixed length types in the sorting tuple. However, you cannot sort by these collection columns, but they can be in the SELECT list along with other column(s) by which you sort.

Support collections of fixed length types as non-passthrough children of unions

This release adds support for collections of fixed length types as non-passthrough children of unions. Plain UNIONs are not supported yet for any collections, but UNION ALL operations are supported.

Example:

```
select id, int_array from complextypestbl
union all select cast(id as tinyint), int_array from complextypestbl
```

Allow implicit casts between numeric and string types when inserting into table

The current implementation requires explicit casts for numeric and string-based literals. However, this release relaxes the implicit casting rules for these cases. This is controlled through a query option allow_unsafe_casts and turned off by default. This query option allows implicit casting between some numeric types and string types.

Improved memory estimation for aggregates and pre aggregates

This release introduces new query options to improve memory estimation for aggregation nodes. Also introduces better cardinality estimates to help in capping memory limits early on during query planning.

Improved CPU costing

This release introduces some changes to the query planner to improve parallel sizing and resource estimation. These changes are done for workload-aware autoscaling and will be available as query options. These additional query options are added for tuning purposes. This new functionality will allow more customers to enable multi-threaded queries globally for improved performance.

Added processing cost to control scan parallelism

Before this release, when a user executed a query with COMPUTE_PROCESSING_COST=1, Impala relied on the MT_DOP option to decide the degree of parallelism of the scan fragment. This release introduces the scan node's processing cost as another factor to consider raising scan parallelism beyond MT_DOP.

Scan node cost now includes the number of effective scan ranges. Each scan range is given a weight of (0.5% * min_processing_per_thread), which roughly means that one scan node instance can handle at most 200 scan ranges. This release also introduces a new query option MAX_FRAGMENT_INSTANCES_PER_NODE to cap the maximum number of fragment instances per node. This newly introduced query option works in conjunction with PROCESSING_COST_MIN_THREADS.

Impala WebUI improvements

This release enhanced the Impala daemon’s Web UI to display the following additional details:

- Backends start time and version: In a large cluster, you can now use the Impala daemon’s Web UI to view the start time and version for all the backends.
- Query performance characteristics: For a detailed report on how a query was executed and to understand the detailed performance characteristics of a query, you can use the built-in web server’s UI and look at the timeline shown in the Gantt chart. This chart is an alternative to the PROFILE command and is a graphical display in the WebUI that renders timing information and dependencies.
- Export query plan and timeline: To understand the detailed performance characteristics for a query, you issue the PROFILE command in impala-shell immediately after executing a query. As an alternative to the profile download page, this release added support for exporting the graphical query plan and also for downloading the timeline in SVG/HTML format. Once you export the query plan or the timeline, memory resources consumed from the ObjectURLs get cleared.
- Historical/in-flight query performance: You can now use the query list and query details page to analyze historical or in-flight query performance by viewing the memory consumed, the amount of data read, and other information about the query.
- Aggregate CPU node utilization: You can now see the recent aggregate CPU node utilization samples for the different nodes.
- Scaling of timeticks and fragment timing diagram for better accessibility: You can now use the query timeline display to scroll horizontally through the fragment timing diagram and utilization chart. You can also zoom by horizontally scaling through mouse wheel events in addition to increasing/decreasing the precision of timetick values.

Skip reloading file metadata for some ALTER_TABLE events

Before this release, EventProcessor ignored trivial ALTER_TABLE events that only modify tblproperties like "transient_lastDdlTime," "totalSize," "numFilesErasureCoded," and "numFiles". For other non-rename ALTER_TABLE events, it triggered a full refresh on the table, which becomes expensive for tables with a large number of partitions or files.

From this release, to be more cost-efficient, the event processor skips reloading file metadata for some ALTER_TABLE events.

The following list contains the events that skip reloading file metadata:

- changing table comment
- adding/dropping columns
- changing column definition (name/type/comment)
- changing ownership
- setting customized tblproperties

For interoperability purposes, this release introduces a new start-up flag 'file_metadata_reload_properties' to list the table properties that need the file metadata reloaded when the properties are changed.

Note: To disable this optimization (in case of any unexpected issues), set file_metadata_reload_properties to an empty string.

JWT auth for Impala

Impala clients, such as Impala shell, can now authenticate to Impala using a JWT instead of a username/password. To connect to Impala using JWT authentication, specify JWT command-line options to the impala-shell command interpreter and enter the password when prompted.

Native High Availability (HA) for Impala Catalog Service

The High Availability (HA) mode of catalog service in CDW reduces the outage duration of the Impala cluster when the primary catalog service fails. Before this release, catalog HA was supported using the K8s leader election mechanism, and now it is natively supported in Impala.

Codegen for STRUCT type

Codegen uses query-specific information to generate specialized machine code for each query. As an Impala user, when you run a standard query, the query optimizer generates an optimized query plan and passes it to the executor for processing. With the codegen capability for STRUCT type in the SELECT list, the query specific information is converted to machine code for faster execution.

Before this release, having structs in the select list was only supported with codegen turned off. This release lifts this restriction, adding full codegen support for structs in the select list.

Example:

```
select small_struct from complextypes_structs
```

June 29, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

This hotfix release fixes problems in backing up and restoring CDW, primarily the issue restoring a Virtual Warehouse executor count. The workaround to the problem described in TSB-2023-683 is no longer necessary.

May 30, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Azure AKS 1.26 upgrade

The new kubernetes version 1.26 has breaking changes; thus, starting earlier workload (Impala, Hive, Hue) versions is not supported, and this option is now disabled. In this release 1.6.4-b161 (released May 30, 2023), when you activate an environment, CDW automatically provisions Azure Kubernetes Service (AKS) 1.26. To upgrade to AKS 1.26 from 1.6.3 or earlier, you must backup and restore CDW. To avoid compatibility issues between CDW and AKS, upgrade to version 1.26.

Using the Azure CLI or Azure portal to upgrade the AKS cluster is not supported. Doing so can cause the cluster to become unusable and can cause downtime. For more information about upgrading, see Upgrading an Azure Kubernetes Service cluster for CDW.

Azure AKS managed identity now mandatory

You must now specify a user-assigned, managed identity for the Azure Kubernetes Service (AKS) cluster when you activate the Azure environment from the CDP CLI in 1.6.4-b161 (released May 30, 2023). In 1.6.3 (released May 5, 2023), specifying the managed identity is also mandatory when you activate the Azure environment from the UI. A managed identity is required because AKS does not support certificate authentication for service principals. The cluster creation operation will fail if the userAssignedManagedIdentity is not set to your managed identity.

New Backup/Restore CDP CLI commands

The following CDP CLI 0.9.87, including Beta DW CLI commands are available and support backing up and restoring CDW:

- dw backup-cluster
- dw restore-cluster

May 5, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Cloudera Data Warehouse Public Cloud 1.6.3-b319 changes, described in more detail below:

- AWS Kubernetes version 1.24 support on page 122
- Selective support for AWS Kubernetes version 1.22 upgrade on page 122
- Changes to the standard required IAM permissions and AWS restricted policy on page 122
- Synchronized metadata across Impala Virtual Warehouses on page 123
- Upgrade and rebuild functionality on page 123
- New requirements in Azure to provide managed identity on page 123
- Ranger authorization (RAZ) in CDW on page 123
- Accessing buckets in a RAZ environment on page 123
- Configuring token-based authentication in CDW on page 123
- Impala coordinator shutdown in Unified Analytics on page 123
- Anti join support in Unified Analytics on page 124
- Debugging enhancement: HiveServer OOM dump on page 124
- Kubernetes dashboard (Preview) on page 124
- Workload Aware Auto-Scaling (Preview) on page 124

Cloudera Data Warehouse Public Cloud Runtime 2023.0.14.0-155 changes:

- Iceberg manifest caching on page 124
- Iceberg Execute Rollback feature from Impala on page 124
- Incremental rebuild of Iceberg materialized views on page 124
- Query hint for Impala table cardinality on page 124
- Improvement in compaction and Impala metadata refresh on page 125
- Optimize Refresh/Invalidate event processing on page 125
- Support for custom hash schema for Kudu range tables on page 125

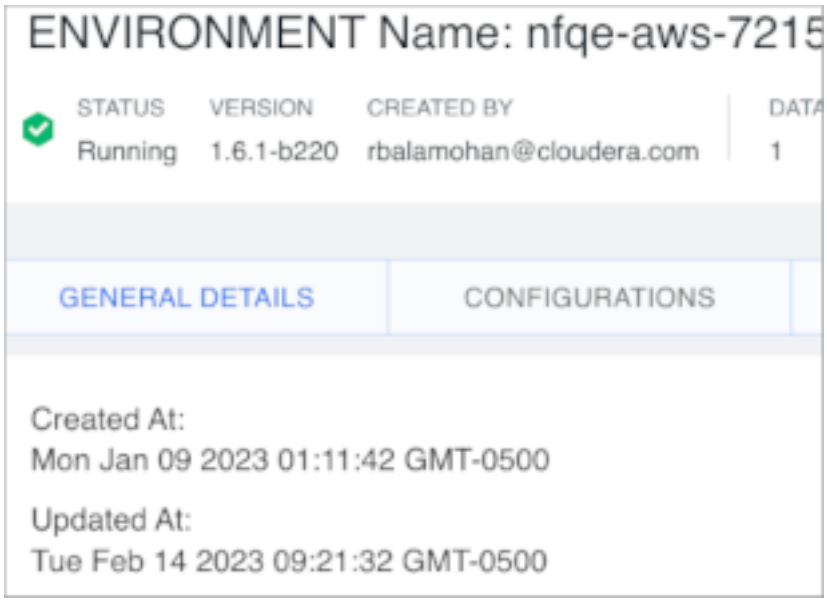
- New shell option "hs2_fp_format" on page 125
- Support for complex types on page 125
- Ability to view Impala query details in Hue
- Ability to access S3 buckets from Hue with RAZ is GA on page 125
- Ability to access ADLS Gen2 containers from Hue with RAZ is GA on page 126
- The “enable_queries_list” configuration has been removed from Hue jobbrowser safety valve section on page 126

AWS Kubernetes version 1.24 support

New CDW clusters using AWS environments you activate in this release 1.6.3-b319 (released May 5, 2023) of Cloudera Data Warehouse will use Amazon Kubernetes (EKS) version 1.24.

Selective support for AWS Kubernetes version 1.22 upgrade

If you activated your pre-existing AWS environment in CDW version 1.4.2-b118 (released Aug 4, 2022) or later, you can upgrade to EKS 1.22 in this release 1.6.3-b319 (released May 5, 2023). AWS environments activated in earlier versions are not supported for upgrade to EKS 1.22 due to incompatibility of some components with EKS 1.22. To determine the activation version of your pre-existing environment, in the Data Warehouse service, expand Environments. In Environments, search for and locate the environment that you want to view. Click Edit. In Environment Details, you see the CDW version.



Changes to the standard required IAM permissions and AWS restricted policy

In this release, as an AWS environment user, you must make the following IAM policy updates:

- AWS restricted policy

  In "Sid": "gocode", add "iam:ListAttachedRolePolicies", only if you have a Ranger Authorization (RAZ) environment.
- Minimum set of IAM permissions required for reduced permissions mode

  Add "autoscaling:UpdateAutoScalingGroup".

  Add "iam:DeleteRolePolicy".

  Add "iam:ListAttachedRolePolicies".

Synchronized metadata across Impala Virtual Warehouses

Using Impala Virtual Warehouses that share a Database Catalog is easier in this CDW release. In past releases, after making changes to data and then refreshing tables or invalidating metadata from your Virtual Warehouse, only the catalog metadata and coordinator metadata for that particular Virtual Warehouse were affected. You had to rerun the commands from each Virtual Warehouse to synchronize metadata across multiple Virtual Warehouses that share a Database Catalog.

This release introduces an enhancement that raises events in the Hive metastore. Catalog daemons process events synchronously across all Virtual Warehouses that share metadata. Metadata is refreshed/invalidated in parallel across all your Virtual Warehouses. You need to run the commands only once in any one of your Virtual Warehouse.

To get this feature, you must upgrade your Virtual Warehouse to this release 1.6.3-b319 that has runtime 2023.0.14.0-155 (released May 5, 2023). You must also set the Impala catalogd flagfile key enable_reload_events to true. Newly created Virtual Warehouses use Impala version 2023.0.14.0-155, which has this feature enabled by default. For more information, including how to disable this feature, see Disabling metadata synchronization.

This enhancement does not synchronize metadata when you refresh tables or invalidate metadata from a Data Hub cluster.

Upgrade and rebuild functionality

The generic Database Catalog and Virtual Warehouse version upgrade process has changed. When you click the Upgrade button to upgrade your Virtual Warehouse or Database Catalog version, the upgrade that occurs includes the rebuild functionality, which cleans up and redeploys your workload. For more information, Rebuilding a Database Catalog and Rebuilding a Virtual Warehouse.

New requirements in Azure to provide managed identity

During the Azure environment activation and private AKS activation, you must provide the resource ID of a user-assigned managed identity to create the Azure Kubernetes cluster resource.

Ranger authorization (RAZ) in CDW

In this release, if your Data Lake enables Ranger Authorization (RAZ), CDW will be RAZ-enabled to provide authentication in Hue, mainly. If you have an AWS environment that predates the option to enable RAZ for Cloudera Data Warehouse, you might want to manually enable RAZ in CDW.

Accessing buckets in a RAZ environment

In an environment that enables Ranger Authorization (RAZ), you must configure permissions to access an S3 bucket. "Accessing buckets in a RAZ environment" describes how to configure the permissions depending on the AWS account that owns the bucket.

Configuring token-based authentication in CDW

Using a JSON web token (JWT), your Virtual Warehouse client user can sign on to your Virtual Warehouse for a period of time instead of entering single-sign on (SSO) credentials every time your user wants to run a query. You use Impyla to access an Impala Virtual Warehouse and a JDBC connection to access a Hive Virtual Warehouse. For more information, see "Configuring token-based authentication".

Impala coordinator shutdown in Unified Analytics

Cloudera Data Warehouse has supported Impala coordinator shutdown for some time, but not in Unified Analytics until this release. When you create an Impala Virtual Warehouse, and enable Unified Analytics, you can configure Impala coordinator shutdown in Unified Analytics.

Anti join support in Unified Analytics

In this release, Unified Analytics enables an optimization for converting a join having a null filter to perform a SQL anti join.

Debugging enhancement: HiveServer OOM dump

In this release, when a hive-server pod out-of-memory (OOM) event occurs, a corresponding dump will be available in the HiveServer dumps directory:

```
<external-bucket>/clusters/<env-id>/<dbc-id>/warehouse/debug-artifacts/hive/
<compute-id>/hive-dumps/<node-name>/heapdump/dump.hprof
```

For example, the full path to the dump file looks something like this:

```
sharecdwdev-bucket-7cv9-dwx-external/clusters/env-9d7cv9/warehouse-167591873
2-q9m2/warehouse/debug-artifacts/hive/compute-1675918927-q8hs/hive-dumps/ip-
192-168-223-136.us-west-2.compute.internal/heapdump/hiveserver2-oom-dump.hpr
of
```

In the example above, ip-192-168-223-136.us-west-2.compute.internal is the name of the HiveServer node on which the pod was running.

Kubernetes dashboard (Preview)

A technical preview of the K8S dashboard is available in this release. You can view the state of resources, such as CPU and memory usage, see the status of pods, and download logs. The dashboard can provide insights into the performance and health of a CDW cluster.

Workload Aware Auto-Scaling (Preview)

Workload Aware Auto-Scaling (WAAS) is available as a technical preview. WAAS allocates Impala Virtual Warehouse resources based on the workload that is running. You choose an executor group set, analogous to a range of nodes instead of the fixed size of the previous auto-scaling implementation. Configuring WAAS requires an entitlement. Contact your account team if you are interested in previewing the feature.

Iceberg manifest caching

Apache Iceberg provides a mechanism to cache the contents of Iceberg manifest files in memory. The manifest caching feature helps to reduce repeated reads of small Iceberg manifest files by Impala Coordinators and Catalogd. You configure manifest caching in CDW. You can enable or disable caching and set a few other caching properties.

Iceberg Execute Rollback feature from Impala

In the event of a problem with your table, you can reset a table to a good state as long as the snapshot of the good table is available. You can roll back the table data based on a snapshot id or a timestamp. The Describe History feature output includes information that helps you use snapshot-related features.

Incremental rebuild of Iceberg materialized views

In this release, the materialized view is automatically updated under certain conditions. After inserting data into a source table of a materialized view, query rewriting is automatically enabled. An incremental rebuild, which updates only the changed part of the view, occurs. Incremental rebuilds tend to update the view much faster than full rebuilds.

Query hint for Impala table cardinality

This release supports a new TABLE_NUM_ROWS query hint to specify a table cardinality for cases where statistics are missing or invalid.

Improvement in compaction and Impala metadata refresh

In this release, handling of the COMMIT_COMPACTION_EVENT has been improved. During compaction, HMS raises events for ACID tables. Impala metadata is refreshed automatically.

Optimize Refresh/Invalidate event processing

Before this release, some metadata consistency issues led to query failures. These failures happen because the metadata updates from multiple coordinators cannot differentiate between self-generated events and those that are generated by a different coordinator. This issue is resolved in this release by adding a coordinator flag to each event, and when processing these events, the coordinator flag is checked to decide whether to ignore the event or consider it.

Support for custom hash schema for Kudu range tables

Impala now includes CREATE TABLE and ALTER TABLE syntax to allow a Kudu custom hash schema. HASH syntax within a partition is similar to the table-level syntax except that HASH clauses must follow the PARTITION clause, and commas are not allowed within a partition. You can use the SHOW HASH SCHEMA statement to view the hash schema information for each partition.

```
CREATE TABLE t1 (id int, c2 int, PRIMARY KEY(id, c2))
       PARTITION BY HASH(id) PARTITIONS 3 HASH(c2) PARTITIONS 4
       RANGE (c2)
       (
       PARTITION 0 <= VALUES < 10
       PARTITION 10 <= VALUES < 20
       HASH(id) PARTITIONS 2 HASH(c2) PARTITIONS 3
       PARTITION 20 <= VALUES < 30
       )
       STORED AS KUDU;
       ALTER TABLE t1 ADD RANGE PARTITION 30 <= VALUES < 40
       HASH(id) PARTITIONS 3 HASH(c2) PARTITIONS 4;
```

New shell option "hs2_fp_format"

This new option sets the printing format specification for floating point values when using HS2 protocol. The default behavior makes the values handled by Python's str() built-in method. Use '16G' to match Beeswax protocol's floating-point output format.

Support for complex types

A SELECT * statement did not expand to complex types to be compatible with earlier versions of Impala that did not support complex types in the result set. In the older versions of Impala, queries using SELECT * skip complex types by default and only expanded to primitive types even when the table contained complex-typed columns. This release adds a new query option EXPAND_COMPLEX_TYPES to include complex types in the SELECT * list.

Ability to view Impala query details in Hue

You can now view Impala query details, query plan, execution summary, and query metrics on the new **Impala Queries** tab on the **Job Browser** page using Hue in CDW, and use this information to tune and optimize your queries. You can also compare two queries and view all query details side-by-side. For more information, see Viewing Impala query details.

Ability to access S3 buckets from Hue with RAZ is GA

Granting fine-grained access to the per-user home directories in Amazon S3 and accessing them from the S3 File Browser in Hue using RAZ is GA. See Accessing S3 bucket from Hue in CDW with RAZ.

Ability to access ADLS Gen2 containers from Hue with RAZ is GA

Granting fine-grained access to the per-user home directories on ADLS Gen2 containers and accessing them from the ABFS File Browser in Hue using RAZ is GA. See Accessing ADLS Gen2 containers from Hue in CDW with RAZ.

The “enable_queries_list” configuration has been removed from Hue jobbrowser safety valve section

The enable_queries_list configuration in the CDW Hue safety valve was used to display or hide the **Queries** tab on the **Job Browser** page. This configuration has been removed. The **Queries** tab is displayed by default based on the type of Virtual Warehouse, whether it is Hive or Impala. You can override the query_store configuration and hide the **Queries** tab. For more information, see Hue configurations in Cloudera Data Warehouse.

February 14, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

AWS Kubernetes version upgrade

With this hot fix, the broken Upgrade button for upgrading an existing cluster to EKS 1.2.1 is now fixed. For more information about upgrading, see What's New, February 7, 2023.

February 7, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

AWS Elastic Kubernetes Service (EKS) version upgrade

The CDW application uses Kubernetes (K8S) clusters to deploy and manage Hive and Impala in the cloud. Kubernetes versions are updated every 3 months on average. When the version is updated, minor versions are deprecated.

To avoid compatibility issues between CDW and AWS resources, you must upgrade the version of Kubernetes that supports your existing CDW clusters to version 1.21.

AWS environments you activate using this release of Cloudera Data Warehouse will use version 1.22.

Important: Check to make sure your AWS environment has been migrated from Helm 2 to Helm 3 before you begin upgrading the Kubernetes version.

If a MIGRATE icon appears in the upper right corner of the environment tile, your AWS environment has not been migrated from Helm 2 to Helm 3. Update the Helm package manager for your environment before you attempt to upgrade its Kubernetes version.

For information about upgrading the AWS EKS Kubernetes version, see Upgrade CDW for EKS upgrade.

Support for Amazon Elastic Kubernetes Service 1.22 cluster

This release 1.6.1-b258 (released Feb 7, 2023) automatically uses and provisions Amazon Elastic Kubernetes Service (EKS) 1.22 when you activate an environment from CDW. In this release, upgrading a cluster to EKS 1.22 is not supported.

Changes to the managed policy ARN, standard IAM permissions, and restricted permissions policy

In this release, as an AWS environment user, you must update the managed policy ARN to handle Kubernetes CSI drivers for EBS and EFS. You must also update your restricted permissions policy if you use it.

Make the following changes:

- Restricted policy changes for updating a managed policy ARN

  Add "arn:aws:iam::<AWS_ACCOUNT>:policy/<noderole-inline-policy>" as shown in "Attaching a managed policy ARN".

- Restricted permissions policy

  Add "elasticfilesystem:PutFileSystemPolicy", to the ResourceTag object and move the "elasticfilesystem:CreateFileSystem", from the CloudFormation object to the RequestTag object, as shown in AWS restricted permissions policy.

Apache Iceberg GA in CDW

This release introduces the general availability of ACID transactions with Iceberg v2 tables from Hive in CDW Runtime 2023.0.13.0-20 (released 2023-2-7)). CDW Runtime 2022.0.12.0-90 (released 2022-12-13), introduced the general availability of ACID transactions with Iceberg V2 tables from Impala. You can run Apache Iceberg ACID transactions within some of the key data services in the Cloudera Data Platform (CDP) public cloud (AWS and Azure), including Cloudera Data Warehouse. From Hive or Impala, you use Apache Iceberg features in CDW, which include time travel, create table as select, and schema and partition evolution.

To access these features, create a new Virtual Warehouse or upgrade an existing one.

Support for Iceberg tables in Avro (Preview)

In this release, you can read Iceberg tables in Avro from Impala. There is a related known issue with using the DECIMAL data type in Avro this release.

Reading Iceberg tables in Avro format from Impala is available as a technical preview. Cloudera recommends that you use this feature in test and development environments. It is not recommended for production deployments.

Enhanced Iceberg support for materialized views

In this release, you can create a materialized view of an Iceberg V1 or V2 table based on an existing Hive table or an Iceberg table. Automatic rewriting of the materialized view occurs under certain conditions.

Iceberg load data inpath feature

From Impala, you can now load data into an Iceberg table using the load data inpath feature.

Querying Data Hub Kudu tables from an Impala Virtual Warehouse using Kudu

After configuring an Impala Virtual Warehouse to connect to Kudu, you can create Kudu tables using Impala clients, such as Hue, the Impala shell, or JDBC/ODBC. You can also ingest data using Spark/NiFi and query using Impala.

Flexible allow lists for Kubernetes cluster and load balancer

In previous releases, there was a single allow list for IP Classless Inter-Domain Routing (CIDRs) from which the Kubernetes cluster or the load balancer should accept incoming connections. In this release, you configure two separate allow lists for accepting incoming connections:

- Enable IP CIDR for Kubernetes cluster
- Enable IP CIDRs for the load balancer

Impala enhancements

This release of Cloudera Data Warehouse includes the following new Impala features:

- Binary support: Impala now supports BINARY columns for all table formats except Kudu. See the BINARY support topic for more information on using this arbitrary-length byte array data type in CREATE TABLE and SELECT statements.
- ALTER VIEW support: Before this release, altering only the VIEW definition, VIEW name, and owner was supported. Impala now supports altering the table properties of a VIEW by using the set tblproperties clause.

- Push down date literals to Kudu scanner: Impala now allows creating and pushing down Kudu predicates from the DATE type.

  Example:

```
select * from functional_kudu.date_tbl where date_col = DATE "1970-01-01";

            PLAN-ROOT SINK
            |
            00:SCAN KUDU [functional_kudu.date_tbl]
            kudu predicates: date_col = DATE '1970-01-01'
            row-size=12B cardinality=1
            ---- DISTRIBUTEDPLAN
            PLAN-ROOT SINK
            |
            01:EXCHANGE [UNPARTITIONED]
            |
            00:SCAN KUDU [functional_kudu.date_tbl]
            kudu predicates: date_col = DATE '1970-01-01'
            row-size=12B cardinality=1
```

- Fix untracked memory in KRPC: Improved memory estimation of queries by accounting for untracked memory in KrpcDataStreamSender.
- Redhat UBI8 images: To address multiple CVEs, Impala images are built using UBI8 base image.

December 13, 2022

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Apache Iceberg V2

This release introduces the general availability of ACID transactions with Iceberg v2 tables from Impala in Cloudera Data Platform (CDP).

Apache Iceberg ACID transactions run within some of the key data services in the Cloudera Data Platform (CDP) public cloud (AWS and Azure), including Cloudera Data Warehouse (CDW).

The following features are included in this release:

- Impala reads of Iceberg V2 tables (row-level deletes or updates)
- Materialized views of Iceberg tables

Iceberg V2 documentation.

November 21, 2022

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

AWS EKS Kubernetes version upgrade

The CDW application uses Kubernetes (K8S) clusters to deploy and manage Hive and Impala in the cloud. Kubernetes versions are updated every 3 months on average. When the version is updated, minor versions are deprecated.

To avoid compatibility issues between CDW and AWS resources, you must upgrade the version of Kubernetes that supports your existing CDW clusters to version 1.21.

AWS environments you activate using this release of Cloudera Data Warehouse will use version 1.22.

Important: Check to make sure your AWS environment has been migrated from Helm 2 to Helm 3 before you begin upgrading the Kubernetes version.

If a MIGRATE icon appears in the upper right corner of the environment tile, your AWS environment has not been migrated from Helm 2 to Helm 3. Update the Helm package manager for your environment before you attempt to upgrade its Kubernetes version.

For information about upgrading the AWS EKS Kubernetes version, see Upgrade CDW for EKS upgrade.

Support for Amazon Elastic Kubernetes Service 1.22 cluster

This release 1.5.1-b110 (released November 22, 2022) automatically uses and provisions Amazon Elastic Kubernetes Service (EKS) 1.22 when you activate an environment from CDW. In this release, upgrading a cluster to EKS 1.22 is not supported.

Forward logs to your observability system

In this release, you can forward logs from environments activated in Cloudera Data Warehouse (CDW) to observability and monitoring systems such as Datadog, New Relic, or Splunk. You configure a CDW environment for these systems using the UI to provide a fluentd configuration.

Rebuild a Database Catalog or Virtual Warehouse

Rebuilding a Database Catalog and rebuilding a Virtual Warehouse, you clean up resources and redeploy your Database Catalog and Virtual Warehouse using your existing DBC or VW image, or the latest available image. You might consider rebuilding to get a feature in a later DBC or VW image, to perform housekeeping, or to troubleshoot a problem.

AWS compute instance type selection is available using the CDP CLI

This release of CDW supports selection of a single AWS compute instance type and additional fallback instance types. From the CDP CLI, you can select the following options:

- computeInstanceTypes

  An array of AWS compute instance types from which you can select a single instance type that the environment of your Virtual Warehouse will use.
- additionalInstanceTypes

  An array of additional AWS instance types for fallback in order of priority to fail over in the event the compute instance type is unavailable. You can specify one of the additional instance types as the default if you do not configure a compute instance type.

When selecting an instance type, consider your computing, memory, networking, and storage needs. To fetch a string of available instance types, use the following CLI command:

```
cdp dw describe-allowed-instance-types
```

Example output:

```
{
     "aws": {
     "default": [
     "r5d.4xlarge"
     ],
     "allowed": [
     "r5d.4xlarge",
     ...
```

Specify the instance types in aws-options of the create-cluster command. For example:

```
cdp \
     --profile ${PROFILE} \
```

```
      dw \
      create-cluster \
      --environment-crn ${AWS_CB_ENV_CRN} \
      --use-private-load-balancer \
      --aws-options computeInstanceTypes=r5dn.4xlarge,additionalInstanceTypes
=r5d.4xlarge,r5dn.4xlarge
```

Use list-clusters and describe-cluster to see cluster settings. CLI documentation link.

Enable CloudWatch Logs

If you use Amazon CloudWatch, you can enable CloudWatch logs when you activate an environment from CDW or edit environment details. Add required permissions to your IAM policy, and then access CloudWatch logs from AWS.

Nodes dashboards added to Grafana

New nodes dashboards show metrics for the AWS/Azure Virtual Machines that host the Kubernetes cluster.

Cloudera Data Warehouse audit events

You can retrieve the following audit events using the CDP CLI that occur in CDW:

CreateEnvironment	CloneDbCatalog	UpgradeHiveVirtualWarehouse
DeleteEnvironment	UpgradeDbCatalog	CreateImpalaVirtualWarehouse
UpdateEnvironment	CreateHiveVirtualWarehouse	DeleteImpalaVirtualWarehouse
UpgradeEnvironment	DeleteHiveVirtualWarehouse	StartImpalaVirtualWarehouse
CreateDbCatalog	StartHiveVirtualWarehouse	StopImpalaVirtualWarehouse
DeleteDbCatalog	StopHiveVirtualWarehouse	UpdateImpalaVirtualWarehouse
StartDbCatalog	UpdateHiveVirtualWarehouse	CloneImpalaVirtualWarehouse
StopDbCatalog	CloneHiveVirtualWarehouse	UpgradeImpalaVirtualWarehouse
UpdateDbCatalog		

For more information about experimental CLI commands for Cloudera Data Warehouse, go to Version Mapping. Click the CDP CLI Reference link for your CDW version. Scroll to Available Commands, and click dw.

Data Analytics Studio (DAS) is deprecated

DAS has been deprecated and is disabled by default. It will be removed in future releases. Cloudera encourages you to use Hue to run Hive LLAP workloads. If you need to use DAS, then you can enable it by setting the das.event-pipeline.enabled property to true in the Database Catalog configurations. For more information, see Enabling Data Analytics Studio in CDW Public Cloud.

Hue Query Processor scan frequency decreased to 5 minutes

The Hue Query Processor scans the event processor pipeline to retrieve the Hive query history and query details and display them on the **Job Browser** page. The scan frequency has been decreased from 2 milliseconds to 5 minutes to optimize resource utilization. As a result, you may notice a delay in viewing the query history and query details on the **Job Browser** page for queries that finish executing in less than 5 minutes. However, you can still view the query history from the Query history tab below the query editor.

September 15, 2022

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Support for Amazon Elastic Kubernetes Service 1.21 cluster

This release 1.4.3-b225 (released September 15, 2022) automatically uses and provisions Amazon Elastic Kubernetes Service (EKS) 1.21 when you activate an environment from CDW. In this release, upgrading a cluster from EKS 1.20 to 1.21 is not supported.

Alert notification configuration

You can now configure alert notifications that appear in user notification system. The notifications supplement alert information in charts on the Manage/Hive/Compaction Observability dashboard in Grafana.

Impala Virtual Warehouse catalog high availability

In earlier releases, the single instance limitation of the Impala Virtual Warehouse catalog made the catalog a single point of failure. In this release, you can optionally configure replication of the Impala catalog for high availability. One instance operates in active mode, the other in passive mode to provide failover. The passive instance serves as a backup and takes over if the active instance goes down.

General availability of Managed Storage Access

CDW capabilities for storing data for multiple tenants is now generally available (GA). You set up a managed storage warehouse for AWS or for Azure.

General availability of SSO to Virtual Warehouses

In this release, the capability to enable SSO (single sign-on) to your Virtual Warehouse from JDBC/ODBC clients is generally available (GA).

Hive to Iceberg table migration disallowed under certain conditions

If you attempt to migrate a Hive table having VARCHAR or CHAR columns to Iceberg, an exception will occur in this release. Migration under these conditions is not allowed in this release because incorrect query results can occur. To work around this limitation, change VARCHAR and CHAR columns to string columns, and then migrate the table to Iceberg.

Impala Autoscaler Web UI

This release includes an addition to the Impala Web UI for debugging the Impala Virtual Warehouse. The new UI provides insight into Autoscaler operations, accessing log messages, and resetting the log level.

Additions to IAM policy, standard and restricted, permissions

In this release, if you want to use Amazon CloudWatch, add logging permissions: to your IAM policy: "logs:CreateLogGroup","logs:CreateLogStream","logs:DescribeLogStreams", and "logs:PutLogEvents". The arn:aws:iam::aws:policy/CloudWatchAgentAdminPolicy must also be added to the restricted permissions.

Using CloudWatch is not required. Add logging permissions for CloudWatch only if you use this Amazon service. For more information about adding the CloudWatch logging permissions, see the following documentation:

- Minimum set of IAM permissions required for reduced permissions mode
- AWS restricted permissions policy

CREATE TABLE LIKE PARQUET FILE feature in Unified Analytics

You can base a new table on a schema in a Parquet file using the CREATE TABLE LIKE PARQUET FILE statement.

August 4, 2022

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Azure AKS 1.22 provisioning

In this release 1.4.2-b118 (released August 4, 2022), when you activate an environment, CDW automatically provisions AKS 1.22.

Do not upgrade your AKS cluster to 1.22 in this release using the Azure CLI. Doing so can cause the cluster to become unusable and can cause downtime. For more information about upgrading, see Upgrading an Azure Kubernetes Service cluster for CDW.

Azure AKS managed identity support

You can now specify a user-assigned, managed identity for the AKS cluster when you activate the Azure environment.

Support for fine-grained access using minimum permissions in Azure environments

You can configure minimum permissions to govern access control between CDW, Azure resources, and the Azure storage account.

Azure Kubernetes Service deployment configuration options

Changes in configuration properties simplify configuring a DNS zone for AKS.

Cluster subdomain definition improvement

In release version 1.4.2-b118 (released August 4, 2022), if you have the CDW_CUSTOM_CLUSTER_ID entitlement, you can define a cluster subdomain to retain your JDBC URLs. When you activate an environment, define your old subdomain using the following format:

env-xxx.dw

Defining the old subdomain in this way retains your old Virtual Warehouse names in the cluster. During environment activation, your old URLS continue to work, including JDBC URLS, Hive and Impala Virtual Warehouse URLS, Grafana URLS, and other URLS.

The new way of defining your old subdomain replaces the behavior in effect for Versions 2021.0.1.-b10 (released August 27, 2021) through 1.4.1-b86 (released June, 22, 2022). For information about the old way of defining a cluster domain, and the JDBC URL changes you had to made, see the August 27, 2021 release notes.

Simplification of Azure activation UI from CDW

Azure documentation for Data Warehouse has been updated to reflect improvements in the UI:

- Creating and registering the Azure environment
- Configuring a custom private DNS Zone
- Enabling kubenet networking
- Activating private AKS
- Custom tags

Apache Iceberg GA in CDW

The following new features are generally available (GA) in Iceberg V1 in CDW:

- Iceberg table rollback from Hive
- From Hive, create an Iceberg table with a metadata location
- Expire snapshots from Hive

Impala BI features were released on a GA basis already.

Impala enhancements

This release of Cloudera Data Warehouse includes the following new Impala features:

- Printing query results in vertical format

  Impala-shell now includes a new command option '-E' or '--vertical' to support printing of query results in vertical format.
- Resolving ORC columns by names

  Before this release, Impala resolved ORC columns by index. In this release, a query option ORC_SCHEMA_RESOLUTION is added to support resolving ORC columns by names.
- Retrieving the data file name

  Impala now supports including a virtual column in a standard SELECT statement select INPUT__FILE__NAME from <tablename> to retrieve the name of the data file that stores the actual row in a table.
- Consolidating the ranger audit logs for the same table

  Impala now consolidates the Ranger audit log entries of column accesses granted by the same policy for columns in the same table, after all the requests for accessing an object are processed.
- BYTES function support

  Impala now supports the BYTES() function. This function returns the number of bytes contained in a byte string.
- Configurable socket timeout for http transport

  Impala supports configuring socket timeout using the new impala-shell config option --http_socket_timeout_s. When you set a reasonable timeout, an impala-shell client can retry in case of connection issues.
- UTF-8 mode support

  Some Impala STRING types now support UTF-8 aware behavior to ensure consistent results for non-ASCII characters in the string in both Hive and Impala.

June 22, 2022

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Apache Iceberg GA in CDW

Apache Iceberg is now generally available (GA) in key data services on CDP, including Cloudera Data Warehouse. Iceberg integration with Cloudera Data Platform (CDP) enhances the Lakehouse architecture by extending multifunction analytics to petabyte scale for multi-cloud and hybrid use-cases. From Impala, you use Apache Iceberg features in CDW, which include time travel, create table as select, and schema and partition evolution. To access these features, create a new Virtual Warehouse or upgrade an existing one.

Cross account role permissions policy

For security reasons, if you do not want to provide PutRolePolicy permission in your cross account role, which would be used later to add an inline policy to the Node instance role, you can attach a managed policy ARN to a node role. This managed policy provides cross account role permissions. To access this feature, create a new NodeInstanceRole manually, and provide the ARN during activation of the environment from CDW.

Zipping unnest on arrays from Views

As part of this release, you can use zipping unnest functionality on arrays from Views. Before this release, this zipping functionality worked for arrays only in Tables but did not support Views as a source. For more information about using this zipping unnest functionality, see Zipping unnest on arrays from Views.

Updates to the Hive and Impala language reference manuals in Hue

Hive and Impala language reference manuals that are built into the Hue web interface have been updated to include the latest syntax and user-defined functions (UDF). Help on Iceberg concepts and syntax is available for Impala.

Improvements have also been made to the query autocomplete predictions and syntax checker to support these changes.

May 10, 2022

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Compaction txn heartbeating after Worker timeout

This hot fix resolved problems with Database Catalogs timeouts.

Inhibit creation of JCEKS secret during updates and upgrades.

This hot fix resolved problems with Ranger policies after upgrading to 2022.0.7.0-80 (released April 27, 2022).

Recognition of regional cloud provider

This hot fix resolved issues with the DWX server recognizing the regional cloud provider and using environment-specific chartmap buckets.

April 27, 2022

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Metering fix

In this release, an important fix related to Cloudera Data Warehouse (CDW) metering has been added. This fix accurately communicates the status of your Virtual Warehouses to the Cloudera control plane, necessary for proper functioning of CDW. To include this fix, upgrade your Virtual Warehouses. Cloudera recommends performing this upgrade as soon as possible.

Unified Analytics

This is the GA release of Unified Analytics that brings SQL equivalency without syntax changes to CDW SQL engines. Unified Analytics brings significant optimization equivalency to these engines, unifying common techniques such as subquery processing, join ordering, and materialized views. Unified Analytics is available as a CDW Kubernetes service in AWS. Unified Analytics documentation includes features, limitations, and how to use Unified Analytics. To take advantage of this feature, create a new Impala Virtual Warehouse and enable the Unified Analytics option, or simply create a new Hive Virtual Warehouse. For more information, see this community blog.

Ability to enable a private CDW environment in Azure Kubernetes Service (AKS)

You can now activate CDW clusters in private mode which enables you to create all cloud resources without the need of public IP addresses. For more information, see Enabling a private CDW environment in Azure Kubernetes Service.

Downloadable link for Beeline CLI tarball

You can now download the Beeline CLI tarball from CDW and use the Beeline client to connect to a Hive Virtual Warehouse and run queries. The archive file contains all the dependent JARs and libraries that are required to run the Beeline script. Upgrade your existing Virtual Warehouse to use this feature.

Support for uploading non-UDF JARs

As part of this release, CDW enables administrators to upload additional non-UDF JARs to the Hive classpath that might be required to support dependency JARs, third-party Serde, or any Hive extensions.

Enhanced complex types support for Impala

This release allows you to use:

- Array in SELECT list
- UNNEST function for arrays in SELECT list

See Querying arrays for more information.

Ability to run Hive LLAP workloads using Hue without the “CDW_HUE_LLAP” entitlement

You can now use Hue to run Hive workloads in CDW without enabling the “CDW_HUE_LLAP” entitlement. Hue packs the combined features of Data Analytics Studio (DAS) and Hue in this new version. If you do not have the “CDW_HUE_LLAP” entitlement enabled in your environment, then you can get Hue by creating a new Hive Virtual Warehouse.

Ability to select the Hue image version

CDW allows you to select a Hue Runtime version from the Hue Image Version drop-down menu while creating a Virtual Warehouse. The compatible Hue image version is automatically selected based on the Hive or Impala image version that you select. If there is no corresponding Hue version for the selected Hive or Impala version, then the latest Hue version is automatically selected. For example, if you select Hive or Impala version 2022.0.7.0-70, then 2022.0.7.0-70 is selected as the Hue image version.

You can also check the Hue image version from the **Virtual Warehouse Details** page.

Hue Query Processor runs at the Database Catalog level

The Hue Query Processor indexes and retrieves Hive query history by reading query event data from a Hadoop Compatible File System (HCFS), such as S3, and writing them into a relational database, such as PostgreSQL. Previously, the Hue Query Processor process ran at the Virtual Warehouse level. Therefore, there was one Query Processor pod running for each Virtual Warehouse associated with a Database Catalog. Now, the Query Processor pod runs at the Database Catalog level. This helps in reducing the read cost in the cloud.

A new tab called “Hue query processor” has been added under the Database Catalog Edit CONFIGURATIONS section. The “hue-query-processor.json” and “hue-event-processor.json” files are no longer available under the Virtual Warehouse Edit CONFIGURATIONS Hue Configuration files drop-down menu.

If you are on an older version of Virtual Warehouse, then you must create a new Database Catalog and Virtual Warehouse to activate this feature.

To tune the data refresh rate, see Configuring the Hue Query Processor scan frequency.

Ability to grant fine-grained access to S3 buckets from the S3 File Browser in Hue (Preview)

For better security and ease of use for users, you can create per-user home directories within your S3 bucket and grant fine-grained access to these user directories from the S3 File Browser in Hue. To enable fine-grained access from the Hue S3 File Browser, you must enable Ranger Authorization Service (RAZ) when you register your AWS environment with CDP, and add the path to your S3 bucket in the hue-safety-valve field in CDW.

Note: You need to contact Cloudera to have this feature enabled.

Ability to grant fine-grained access to ADLS Gen2 containers from the ABFS File Browser in Hue (Preview)

For better security and ease of use for users, you can create per-user home directories within your ADLS Gen2 container and grant fine-grained access to these user directories from the ABFS File Browser in Hue. To enable fine-grained access from the Hue ABFS File Browser, you must enable RAZ when you register your Azure environment with CDP, and add the path to your ABFS container in the hue-safety-valve field in CDW.

Note: You need to contact Cloudera to have this feature enabled.

Performance and security enhancements in Hue

Python 2 has reached the end of life and is no longer supported. Hue in CDW Public Cloud now uses Python 3 which makes use of critical bug fixes and Common Vulnerabilities and Exposures (CVE) fixes for many third-party software dependencies.

Following changes have been made in the Hue codebase in this release of CDW Public Cloud:

- The Hue container now runs in the UBI8 image, which is upgraded from UBI7.
- Python libraries have been upgraded from Python 2.7 to Python 3.8.
- The Django server has been upgraded from version 1.11.29 to 3.2.12.
- Python modules such as django-auth-ldap, django-axes, djangorestframework-simplejwt, Mako, Markdown, python-ldap, django-babel, django-mako, django-cors-headers, djangorestframework, eventlet, sqlparse, and so on have also been upgraded.
- Hue now uses Gunicorn as a front-end server. Previously, Hue used the CherryPy server.

These upgrades bring in significant performance improvement and stability in query execution, uploading, and importing files to S3 or ABFS. Operating System, Python version, and Python module upgrades have resulted in a stable environment and fixed more than 800 security vulnerabilities.

Also, the UBI8 image uses the latest Apache HTTP Server (httpd). This no longer requires you to limit the fully qualified domain name of your Virtual Warehouse to 64 characters.

Hue images now run as non-root users.

Selecting an availability zone when creating a Virtual Warehouse

In AWS environments, you can now accept the default availability zone, or select your preferred zone when creating a Virtual Warehouse. To use this feature, create a new Virtual Warehouse.

AWS restricted permissions policy

As Administrator, to restrict access to your Cloudera Data Warehouse service, your IAM role must include the AWS restricted policy. You must add the policy to your IAM role before you activate the environment in Cloudera Data Warehouse to make the service available to users.

March 8, 2022

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Apache Iceberg technical preview in Cloudera Data Warehouse

Cloudera Data Warehouse (CDW) supports Apache Iceberg as a technical preview. Iceberg is a cloud-native, open table format for organizing petabyte-scale analytic datasets on a file system or object store. In CDW, you can read and write Iceberg tables using Hive or Impala SQL engines. Create a new Virtual Warehouse, or upgrade your existing one, to use Iceberg in CDW.

Automating metadata invalidation after compaction

After compaction of ACID tables, metadata in Impala coordinator local caches might be stale. Invalidation refreshes the metadata and prevents possible query failure. This release introduces automatic metadata invalidation, which you can configure in the Impala Virtual Warehouse. You must upgrade your Database Catalog to get this feature.

Cloudera Data Warehouse availability in Asia Pacific region

Cloudera Data Warehouse (CDW) is now supported in the ap-1 (Australia) regional Control Plane. To use CDW in this regional Control Planes, your CDP administrator must create a new environment. For the list of all supported services for all supported Control Plane regions, see CDP Control Plane regions.

Data Visualization integration in Cloudera Data Warehouse

The status of Data Visualization integration in CDW is GA in this release. CDW integrates Data Visualization for building graphic representations of data, dashboards, and visual applications based on CDW data, or other data sources you connect to. Authorized users can explore data using graphics, such as pie charts and histograms and collaborate using dashboards. BI analysts who can access your environment can use these features. In a future release, the Enable Data Visualization option in the New Virtual Warehouse dialog will be deprecated.

Impala coordinator shutdown

To save cloud costs, you can configure Impala coordinators to automatically shutdown during idle periods. In this release, the coordinator shutdown feature status has changed to GA. The feature is available when you create an Impala Virtual Warehouse.

Impala Virtual Warehouse High Availability

A single Impala coordinator might not handle the number of concurrent queries you want to run or provide the memory your queries require. You can configure multiple active-active coordinators to resolve or mitigate these problems. To use this feature, create a new Virtual Warehouse.

Impala late materialization of columns

The Data Warehouse Runtime release of Impala introduces late materialization, which optimizes certain queries on Parquet tables by limiting table scanning. Only relevant data is materialized to improve query response.

Istio version update

In this release, new environments you create support Istio 1.11.2. In new environments, you must select the latest Cloudera Data Warehouse image version 2021.0.6.96 when you create a Database Catalog and Virtual Warehouse, as described in Known issues.

Log4j Update

Cloudera has released a new version of CDW that upgrades the embedded Log4j version to 2.17.1.

January 19, 2022

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Accessing Impala using the new Impala shell client

You need to provide commands to your client users for installing and launching the Impala shell to connect to your particular Impala Virtual Warehouse version. You can use new options in the UI for accessing Impala from the Impala shell client. You can obtain the commands to download the new Impala shell client from the CDW UI under Impala Virtual Warehouse options. Create a new Virtual Warehouse to get this feature.

Support for regional control planes for eu-1 (Germany)

The Admin must create a new environment to support these regional control planes.

December 22, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

CDW upgrade of embedded Log4j version

Cloudera has released a new version of CDW which upgrades the embedded Log4j version to 2.16. This provides a permanent fix for CVE-2021-44228.

To use this version you must upgrade your Database Catalog(s) and Virtual Warehouse(s) to the latest version, which is 2021.0.4.1-3. Additionally, you can create new Virtual Warehouses using the latest version, and those will also have the fix for this vulnerability.

If you had previously applied the configurations shared in CVE-2021-44228 remediation for CDW, then those are no longer required. Please reach out to Cloudera Support if you have questions.

November 20, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Supported AWS regions now include EU (Milan)

Previously in a Technical Preview state, support for the EU (Milan) region in AWS is in a GA (general availability) state. For more information, see Supported AWS Regions. To use this feature, create a new environment and activate it in CDW.

SSM and CBO federation

This release adds support for interoperation between the cost-based optimizer (CBO) in Cloudera Data Warehouse and the AWS Systems Manager (SSM) service.

Ability to set idle session timeout for Hue

CDW allows you to configure the idle session timeout value in Hue to automatically log out users from the Hue web interface after a period of inactivity. For more information, see Configuring idle session timeout for Hue.

CDW Impala/Hive cross-region connection

This release adds the capability to create external tables on data in S3 buckets that are hosted in a different region from the one where CDP is running. Requests to these buckets can use a proxy. For example, GDPR dictates that you cannot move data from your EMEA region, but your customers want to query the data from CDP in us-east-1. In compliance with GDPR, you can create an external table based on data in a bucket in an EMEA region when you are running CDP in the us-east-1 region.

Hue usability improvements

As Admin in a Ranger Authorization Service (RAZ) environment, you can set the AWS S3 or ADLS path to a user directory in Hue. As the Hue user, you always land in your directory and do not have to repeatedly change URLs.

Hive backports

This release fixes issues with custom compaction queue settings and vectorized built-in functions having compound expressions in PARTITION BY or ORDER BY clause.

October 29, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Hive compaction observability

An alert notification about compaction status, issues, and suggested actions appear in the overview of your Database Catalog, which leads to an informative message about the issue. The following list describes a few of more than 25 notifications:

- Oldest initiated compaction passed threshold
- Large number of compaction failures
- More than one host is initiating compaction

Create an environment in this release to pick up this feature. For more information about compaction observability, see Compaction Observability.

Choosing a CDW version for a Database Catalog or Virtual Warehouse

When creating a Database Catalog or Virtual Warehouse, you can choose the latest version of CDW or an earlier version. Version mapping lists the CDW versions from 2021.0.0-b33 released on August 9, 2021 to the latest release. You can ensure backward compatibility for your scripts, for example, by running your jobs on the same version of CDW. The ability to choose a version of Database Catalog or Virtual Warehouse is available under entitlement CDW_VERSIONED_DEPLOY.

AWS EKS Kubernetes version upgrade

The CDW application uses Kubernetes (K8S) clusters to deploy and manage Hive and Impala in the cloud. Kubernetes versions are updated every 3 months on average. When the version is updated, minor versions are deprecated. Amazon Elastic Kubernetes Service (EKS) is updating to Kubernetes version 1.20 and is ending support for version 1.17.

AWS environments you activate using this release of Cloudera Data Warehouse, and later, will use version 1.20. To avoid compatibility issues between CDW and AWS resources, you must update the version of Kubernetes that supports your existing CDW clusters to version 1.20.

Important: Check to make sure your AWS environment has been migrated from Helm 2 to Helm 3 before you begin upgrading the Kubernetes version.

If a MIGRATE icon appears in the upper right corner of the environment tile, your AWS environment has not been migrated from Helm 2 to Helm 3. Update the Helm package manager for your environment before you attempt to upgrade its Kubernetes version.

For information about upgrading the AWS EKS Kubernetes version, see Upgrade CDW for EKS upgrade.

Hue supports Hive Hybrid Procedural SQL in CDW

You can run Hive Hybrid Procedural SQL (HPL/SQL) using the Hue query editor in CDW. For more information, see How to run a stored procedure from Hue in Cloudera Data Warehouse.

October 22, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Prevention of ineffective scaling in Impala Warehouses

CDW takes advantage of IMPALA-10244 to detect the case of queries being queued because of a resource bottleneck on the Impala Coordinator. In this case, adding more executor groups does not result in the ability to run more queries, and so, the warehouse does not scale up.

September 28, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Retaining PostgreSQL backups in the cluster

When you create a Cloudera Data Warehouse cluster using the CDP CLI create-cluster command, any PostgreSQL backup retention period you set on your Cloud provider side, is observed by CDP. For example, if you configure backupRetentionDays in Azure or BackupRetentionPeriod in AWS, and then create a cluster using the CDP CLI, the cluster will retain the PostgreSQL backups according to your configuration.

August 27, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces no new changes.

CDW Public Cloud endpoints change

The format of Hive, Impala, Hue, and Data Analytics Studio (DAS) endpoints have changed in the Cloudera Data Warehouse (CDW) Public Cloud release dated August 27, 2021:

- FROM:

```
`*.<env-id>.dw.<tenant-id>.<workload-domain-name>`
```

- TO:

```
`*.dw-<env-name>.<tenant-id>.<workload-domain-name>`
```

In the August 27, 2021 release, the format was changed from using the environment ID (env-id) to using the DW environment name (env-name). Consequently, the UI was simplified. If you had the CDW_CUSTOM_CLUSTER_ID entitlement in earlier releases, you might notice the Custom Cluster ID field in Activation Settings has been removed. This field is no longer necessary.

The new endpoint formatting affects the Hive and Impala JDBC URLs, the Impala shell URL, Hue URLs for Hive and Impala, and DAS URLs, as well as the Impala JDBC with coordinator shutdown URL.

Examples of endpoint changes

In these examples, the following values are used for the variables in the endpoint format above:

Variable: env-name

Example value: env1

Description: The name you provided when registering the environment in Management Console.

Variable: env-id

Example value: env-hfwbgd

Description: The random string for the environment ID that gets generated at the time of activation from CDW.

Variable: tenant-id

Example value: 'xcu2-9999'

Description: Your tenant ID.

Variable: workload-domain-name

Example value: dev.abcd.work

Description: The domain name.

Hive Virtual Warehouse name

Example value: hive1

Impala Virtual Warehouse name

Example value: impl1

Table 1:

Endpoint type	New URL (CDW August 27, 2021)	Old URL (CDW August 9, 2021 and earlier)
Hive JDBC	jdbc:hive2://hs2-hive1.dw-env1.xcu2-9999.dev.abcd.work/default;transportMode=http;httpPath=cliservice;socketTimeout=6	jdbc:hive2://hs2-hive1.env-hfwbgd.dw.xcu2-9999.dev.abcd.work/default;transportMode=http;ssl=true;httpPath=cliservice;socketTimeout=6
Impala JDBC	jdbc:impala://coordinator-impl1.dw-env1.xcu2-9999.dev.abcd.work:443/default;AuthMech=3;transportMode=http;httpPath=cliservice;ssl=	jdbc:impala://coordinator-impl1.env-hfwbgd.dw.xcu2-9999.dev.abcd.work:443/default;AuthMech=3;transportMode=http;httpPath=cliservice;ssl=

Endpoint type	New URL (CDW August 27, 2021)	Old URL (CDW August 9, 2021 and earlier)
Impala shell	impala-shell --protocol='hs2-http' --ssl -i 'coordinator-impl1.dw-env1.xcu2-9999.dev.abcd.work:443' -u <userid> –l	impala-shell --protocol='hs2-http' --ssl -i 'coordinator-impl1.env-hfwbgd.dw.xcu2-9999.dev.abcd.work:443' -u <userid> –l
Hue Hive	https://hue-hive1.dw-env1.xcu2-9999.dev.abcd.work/hue	https://hue-hive1.env-hfwbgd.dw.xcu2-9999.dev.abcd.work/hue
Hue Impala	https://hue-impl1.dw-env1.xcu2-9999.dev.abcd.work/hue	https://hue-impl1.env-hfwbgd.dw.xcu2-9999.dev.abcd.work/hue
Impala JDBC with coordinator shutdown enabled	jdbc:impala://impala-proxy-impl1.dw-env1.xcu2-9999.dev.abcd.work:443/default;AuthMech=3;transportMode=http;httpPath=cliservice;ssl=1;UID=<workload username>;PWD=<PASSWORD>	jdbc:impala://impala-proxy-impl1.env-hfwbgd.dw.xcu2-9999.dev.abcd.work:443/default;AuthMech=3;transportMode=http;httpPath=cliservice;ssl=1;UID=<workload username>;PWD=<PASSWORD>
Data Analytics Studio (DAS)	https://hive1.dw-env1.xcu2-9999.dev.abcd.work/	https://hive1.env-hfwbgd.dw.xcu2-9999.dev.abcd.work/

August 9, 2021

There are no new features in this release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud. This release mainly changed the CDW version numbering scheme.

July 29, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Improvements in the CDP CLI

The CDP CLI commands for creating and deleting an environment, and for creating and deleting a Virtual Warehouse, include all options available in the UI.

Auto-shutdown of Impala coordinators for cost saving

When CDP does not need resources to handle Impala queries, coordinators shut down without user intervention.

July 13, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Impala Debug Web UIs are now available in CDW

You can now use the Impala debug web UIs in CDW, which map to equivalent debug web UIs in Cloudera Manager as follows:

CDW Debug Web UI	Cloudera Manager Equivalent
Impala Catalog Web UI	Catalog Server Web UI
Impala Coordinator Web UI	Impala Daemon Web UI
Impala StateStore Web UI	StateStore Web UI

For more information about this feature, see Use the Web UI to debug Impala Virtual Warehouses.

Auto-scaling demos have been disabled

Due to critical security issues discovered in the auto-scaling demo components, the demo has been disabled.

CDW UI improvements

Error messages and UI labels have been enhanced for readability and usefulness.

June 14, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

S3 buckets created by CDW on AWS environments are now encrypted by default with SSE

When you use the latest cross-account role from Management Console to register an environment, S3 buckets created by CDW during AWS environment activation, are now encrypted by default with SSE-S3. This provides server-side object encryption (SSE), which protects data at rest. Before this enhancement, S3 buckets were created with only object-level encryption for every object in the bucket during environment activation.

Important: This only applies to S3 buckets created for managed tables and for external tables, which are actually created by CDW during environment activation. You must ensure that S3 buckets you add to CDW after environment activation are encrypted to meet your organization's security requirements.

Impala Virtual Warehouses can now be recovered after going into an error state

Impala resources are not deleted when the Virtual Warehouse goes into an error state. Consequently, if an Impala Virtual Warehouse goes into an error state, you can now use the Edit option to recover the Virtual Warehouse with its original configurations.

Impala Virtual Warehouses now use node memory more efficiently

Impala coordinators now can use ~13GiB more of available memory on a node to better support concurrent query execution.

June 1, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

S3 Guard disabled for custom Database Catalogs and default Database Catalogs on AWS environments

During new AWS environment activation in CDW:

- If you have not specified a DynamoDB table name during environment registration in Management Console, the S3 Guard feature is disabled for all Database Catalogs and Virtual Warehouses.
- If you have specified a DynamoDB table name during environment registration in Management Console, the S3 Guard feature is enabled for default Database Catalogs and Virtual Warehouses. However, it is disabled for non-default (custom) Database Catalogs.

For all environments that were activated before the June 1, 2021 release, S3 Guard is enabled.

May 21, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

Improved startup times for Impala Virtual Warehouses

The auto-scaler for Impala Virtual Warehouse has been updated to be more efficient thereby improving the startup times for Impala Virtual Warehouses.

CDW UI left navigation menu improvement

The left navigation menu for the CDW UI has been upgraded to be consistent with other CDP experiences.

AWS environments: Helm 2 to Helm 3 migration fixes

Various fixes applied to the Helm 2 to Helm 3 migration process for AWS environments.

May 3, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces these changes.

CDE Airflow integration to schedule ETL jobs for Hive Virtual Warehouses

In this release, CDE Airflow, based on Apache Airflow, which enables you to configure, schedule, and monitor workflows, can now be used to schedule ETL jobs that target Hive Virtual Warehouses running in CDW. For more information, see CDE Airflow ETL workloads in Hive Virtual Warehouses.

AWS environments updates

Helm 2 to Helm 3 migration

Helm is a package manager for the Kubernetes platform. CDW, which runs on Kubernetes, has migrated to Helm version 3 to take advantage of the improved security and upgrading capabilities offered by Helm 3. Any AWS environments that were created prior to this platform migration still run on Helm version 2. Although you can continue to use the Database Catalogs and Virtual Warehouses running on AWS environments that use Helm 2, there are significant limitations. Cloudera recommends that you migrate your AWS environments to Helm 3 as soon as you can.

AWS EKS Kubernetes version upgrade

The CDW application uses Kubernetes (K8S) clusters to deploy and manage Hive, Impala, and Real-time Event Store Analytics (Druid) in the cloud. Kubernetes versions are updated every 3 months on average. When the version is updated, minor versions are deprecated. Amazon Elastic Kubernetes Service (EKS) is updating to Kubernetes version 1.16 and is ending support for version 1.15 on May 2, 2021. To avoid compatibility issues between CDW and AWS resources, the version of Kubernetes that supports your CDW clusters must be updated.

Important: Check to make sure your AWS environment has been migrated from Helm 2 to Helm 3 before you begin upgrading the Kubernetes version.

For more information about upgrading AWS environments for AWS EKS Kubernetes cluster updates, see Upgrade CDW for EKS Kubernetes version upgrade.

Azure environments

Ability to set scratch space limit for spilling Impala queries on Azure environments

Memory-intensive SQL operations need temporary disk space when Impala is close to exceeding its memory limit on a particular host. You can configure the scratch space between 300 GiB and 16 TiB per node while creating an Impala Virtual Warehouse in the Cloudera Data Warehouse (CDW) service. For details, see Setting scratch space limit for spilling Impala queries in Azure environments.

April 9, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces the improvements that are described in this topic.

Virtual Warehouse improvement

Ability to launch and use the Hue app while the Virtual Warehouse is starting up

When you restart a Virtual Warehouse, it takes a few minutes to start and change to the "Running' state. In the older CDW version, you had to wait to launch the Hue app until the Virtual Warehouse is up and running. Now you can open Hue and run queries even when it is in the "Starting" state.

March 31, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces the improvements that are described in this topic.

Ability to use the internal load balancer with Azure Kubernetes Service (AKS)

You can enable an internal (private) load balancer to distribute traffic within your virtual network (VNET). An internal load balancer makes a Kubernetes service accessible only to applications running in the same or peered Azure VNET as the Kubernetes cluster, or applications running in on-premises networks that are connected using Azure ExpressRoute circuits. For more information, see Enabling internal load balancer for AKS clusters.

Behavioral changes

CDPD-24214: Virtual Warehouse compaction issues resolved

Description: The compaction processes, which are performed by Hive Virtual Warehouses, were failing to clean up directories, which caused query performance issues.

Resolution: A fix has been added from Apache HIVE-24900, which resolves this issue.

March 23, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces the new features and improvements that are described in this topic.

Non-transparent proxy support for AWS environments

CDW Public Cloud now supports AWS environments that use non-transparent proxies. Using non-transparent proxies permits you to pass connection or security information along with the connection request that is sent by clients. Some organizations' security policies require the use of non-transparent proxies and now CDW can support that requirement. For more information about this feature, see Use a non-transparent proxy with CDW.

March 12, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces the new features and improvements that are described in this topic.

Support for custom tags on AWS and Azure environments

You can specify tenant-level custom tags to identify and monitor resources within the CDW service for AWS and Azure environments at the time of environment creation. See Custom tags in Azure environments for the list of resources that you can tag in Azure environments. If you are using AWS, you can also tag Virtual Warehouses using custom Virtual Warehouse-level tags. For a complete list of resources that you can tag in AWS environments, see Custom tags in AWS environments.

For information on applying custom tags, see Defining custom tags in the Management Console documentation.

February 23, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces the following new features and behavioral changes.

Reduced permissions mode for AWS environments

This feature enables you to activate an AWS environment for CDW with fewer than half of the standard required IAM permissions on your AWS cross-account IAM role. This is useful if your InfoSec policies do not permit you to use the standard set of required permissions.

When activating an AWS environment in CDW, if the system detects that your account does not have the standard set of required IAM permissions, it invokes this feature automatically. Then you must manually create the cloud resources in your AWS account and tag them appropriately instead of having CDW do this for you. To assist you, CDW pre-populates the required CloudFormation template for you within the AWS Console, but you must perform the manual steps to create the stack. When you are finished using the stack, you must manually delete it in the AWS Console. For more information about this feature, see Reduced permissions mode for AWS environments.

February 10, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces the following new features and behavioral changes.

Virtual Warehouse upgrade fixes

Multiple fixes were made to improve the upgrade process for Virtual Warehouses.

January 29, 2021

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces the new features and improvements that are described in this topic.

Provisioning performance improvements for Virtual Warehouses

The provisioning time for Hive and Impala Virtual Warehouses has been significantly reduced for many cases. Generally, provisioning time has been reduced from 1.5 to 4 times over the previous time required for Virtual Warehouse provisioning.

December 21, 2020

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces the new features and improvements that are described in this topic.

Use Data Visualization with Virtual Warehouses

You can now use CDP Data Visualization, a tool that enables you to explore data by using visual objects, with CDW Virtual Warehouses.

Start and stop Impala Virtual Warehouses

In this release, you can now explicitly start and stop Impala Virtual Warehouses by clicking the start or stop icon in the upper right corner of the Virtual Warehouse tile. In the Virtual Warehouse details page, you can use this feature by clicking the Suspend or Start options in the Actions menu. Use this feature to better control costs and resource usage in your cloud environment.

Note: This feature is only available on newly created Impala Virtual Warehouses.

Azure environments: Use DAS and Hue to create new custom functions

HiveServer2 localizes and packages the User-Defined Functions (UDF) jar files into Azure Blob storage. You can use DAS and Hue to create new custom functions using the external jar file. To create user-defined functions, see Using functions in the Hive documentation.

November 3, 2020

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces the new features and improvements that are described in this topic.

Impala Virtual Warehouse improvements

The following improvements to Impala Virtual Warehouses have been added in this release:

- High availability improvements, including fault-tolerance and multi-threading improvements.
- Full support for reading tables based on the ORC file format. This improves collaboration across the data lifecycle. Now Hive and Impala Virtual Warehouses can use the same data format and the SQL engine can be selected to optimize the workload without restrictions tied to the file format.

- Ability to stop, start, and suspend Impala Virtual Warehouses:

Figure 1: Starting an Impala Virtual Warehouse

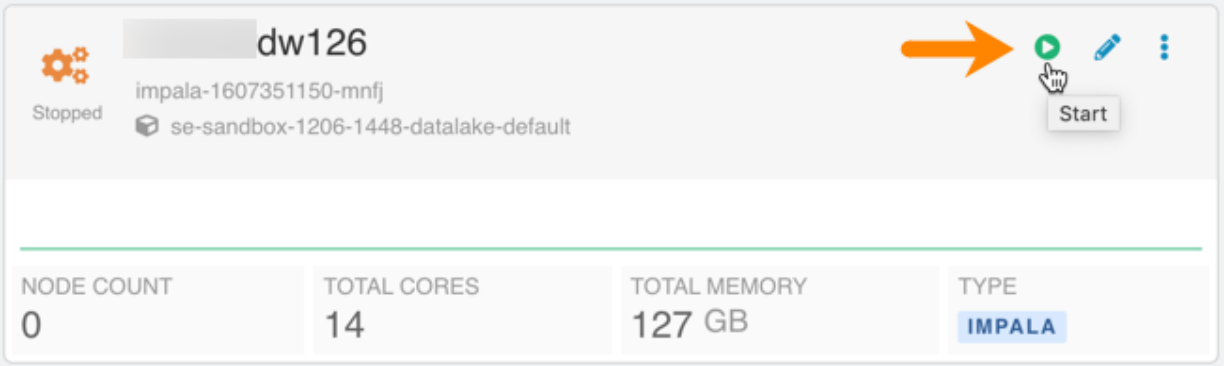


Figure 2: Stopping an Impala Virtual Warehouse

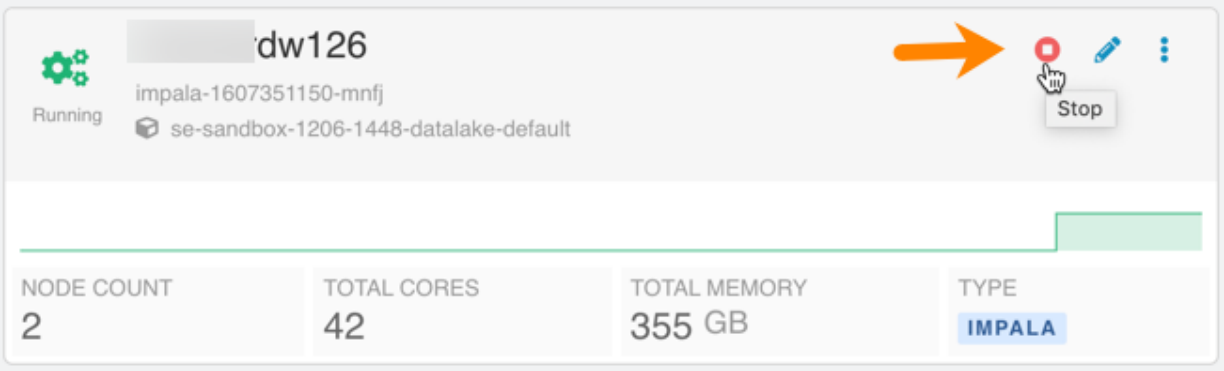
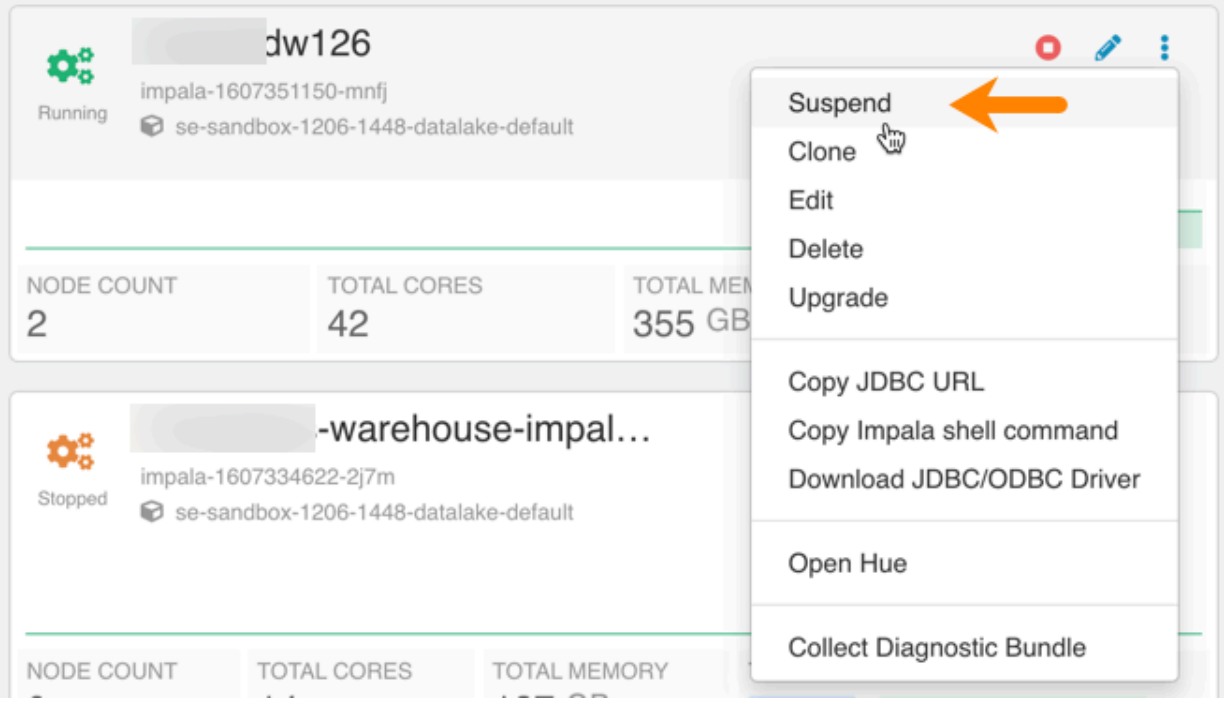


Figure 3: Suspending an Impala Virtual Warehouse



Hive Virtual Warehouse improvements

- Increased stability for Hive ACID and Compaction.
- Performance improvements:
  - For short-running ad-hoc queries by decreasing task startup latency.
  - For late projections by delay joins.
  - For increased optimization of 'shared work.'

Improved security with machine user implementation

When you activate a new environment in Cloudera Data Warehouse (CDW), a machine user is created in the CDP User Management System (UMS) and is synchronized with the FreeIPA Management System (FMS). The FreeIPA machine user is used by most of the services such as Hue, Impala, VizApp, etc. instead of the FreeIPA LDAP admin user to communicate with the services within the Data Lake (for example, Ranger) that is associated with the CDP environment. As a security measure, Cloudera recommends upgrading the CDW environment, Database Catalogs, and the Virtual Warehouses to the latest CDW release.

To upgrade the CDW environment, contact Cloudera Support.

To upgrade the Database Catalog and the Virtual Warehouse, see Upgrading Database Catalogs and Virtual Warehouses in Cloudera Data Warehouse Public Cloud.

Note: Before you upgrade all the Database Catalogs and Virtual Warehouses at the same time, upgrade any one Database Catalog or Virtual Warehouse and let the upgrade process complete. For example, if you have 2 Database Catalogs and 3 Virtual Warehouses in your CDW environment, then complete upgrading any one of the five components. After one of the components has successfully upgraded, you can upgrade the other four in parallel.

When you delete a CDW environment, the machine user is deleted from the UMS and the FMS.

Azure environments

- Executor nodes for Virtual Warehouses can now be scaled down to zero nodes, thereby using zero Azure resources for idle workloads, saving costs.

September 10, 2020

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces the new features and improvements that are described in this topic.

Storage account consolidation for Azure environments

When using an Azure environment, the CDW service no longer creates storage separate from Data Lake storage. This change provides the following benefits:

- Reduced permissions required for the Azure app that is used to register the environment in Management Console:
  - The Storage Blob Contributor role is no longer needed for the storage account.
  - The app used to register the environment with Management Console, can now use the Contributor role instead of the Owner role.
- Reduced management overhead because there are fewer places to look for data and logs that are stored by the CDP platform.
- Simplified custom storage configurations because there no longer is a separate CDW service storage account in addition to the Data Lake storage account.
- Simplified security configuration if additional security controls are needed on storage accounts.
- Log file data and external table data persist in the storage account even after the Virtual Warehouse is deactivated.

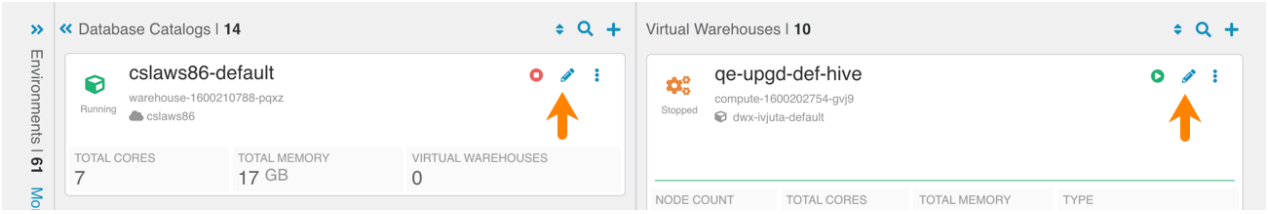
For more information, see Azure environments for CDW.

CDW service endpoint domain names standardized

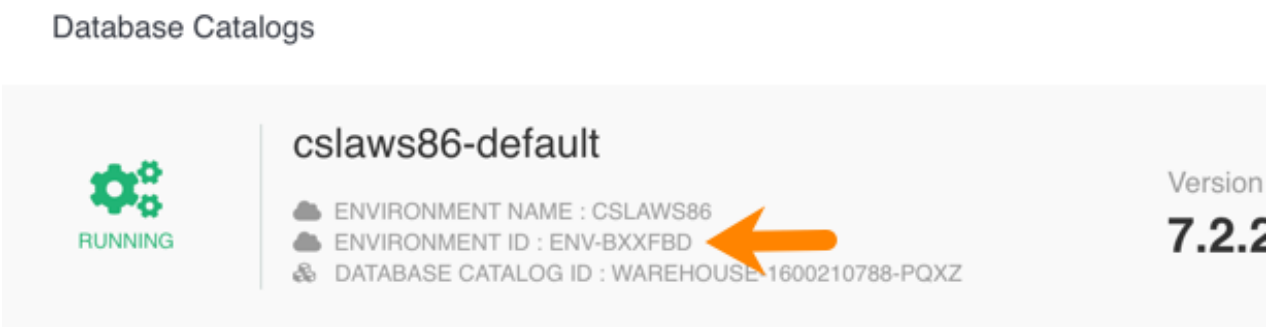
The domain names of CDW service endpoints are now standardized with other CDP service endpoint domain names. The URL syntax for all CDW service endpoints is now:

```
<cdw-environment-ID>.dw.<short-account-ID>.cloudera.site
```

You can get the *<cdw-environment-ID>* from the CDW service UI by clicking the edit icon on either the Data Catalog or the Virtual Warehouse tiles on the Overview page:



Then you can find the *<cdw-environment-ID>* listed at the top of the Database Catalog or Virtual Warehouse details page:



Your *<short-account-ID>* is assigned to you during the account on-boarding process.

August 27, 2020

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces the new features and improvements that are described in this topic.

Overlay networks now available for Azure environments

An overlay network is a software-defined layer of network abstraction that is used to run multiple separate, discrete virtualized network layers over the Azure virtual network. You use overlay networks to increase the number of IP addresses available to support nodes in your CDW cluster on Azure. For more information about using this feature in Azure environments and how to enable it, see Overlay networks for Azure environments.

August 14, 2020

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces the new features and improvements that are described in this topic.

Stability and performance improvements

This release includes numerous stability and performance improvements, improvements to health checks, plus multiple bug fixes.

Ability to configure custom Hue properties

You can configure custom Hue properties that are not directly exposed through the CDW web UI by specifying them in a Hue configuration called hue-safety-valve for a Virtual Warehouse. For more details, see Configuring custom properties using safety valves.

Health check for the Hue load balancer

Unlike earlier, if the Hue server starts and the Hue load balancer has not started yet, Hue does not display an error message when you try to access the Hue endpoint (or the web UI). This is because Hue now checks the health of the load balancer as well as the Hue server before loading the Hue user interface.

July 30, 2020

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud introduces the new features and improvements that are described in this topic.

Support for Microsoft Azure environments now generally available

Support to run CDW service Virtual Warehouses on the Microsoft Azure cloud platform is now generally available. For more details about CDW support for Azure, see Azure support and Azure environments requirements checklist.

June 26, 2020

This release of the Cloudera Data Warehouse service on CDP Public Cloud introduces the following new features and improvements:

Query isolation for scan-heavy, data-intensive queries in Hive LLAP Virtual Warehouses

Hive Virtual Warehouses base auto-scaling on the total scan size of the query. HiveServer, which receives all incoming queries, has a query planner component. When the HiveServer query planner receives queries, it examines the total scan size of each query. That is, it looks at the number of bytes read from the file system required to run the query. If the Query Isolation feature has been enabled for a Virtual Warehouse and a query scans more data than the threshold set in the hive.query.isolation.scan.size.threshold parameter, the planner runs the query in isolation. This means that an isolated standalone executor group is spawned to run the data-intensive query. For more details, see Hive query isolation for scan-heavy, data-intensive queries.

Overlay network support for AWS environments

An overlay network is a software-defined layer of network abstraction that is used to run multiple separate, discrete virtualized network layers over the AWS VPC network. In the case of the CDW service, a custom CNI (Container Network Interface) plugin is used to enable the overlay network. It creates two network spaces:

- A node network space, which derives per-node IP addresses from the VPC.
- A Kubernetes pod network space, which derives per-pod IP addresses from the CNI plugin's own network space.

The overlay network is bridged into the node network. As a result, one IP address is required per node instead of one IP address needed per pod. Consequently, there are more available IP addresses and you can use the CDW service efficiently, auto-scaling Virtual Warehouses as needed to meet the demands of your workloads. For more information, see Use overlay networks for AWS environments in Cloudera Data Warehouse service and Enable overlay networks in AWS environments.

May 14, 2020

This release of the Cloudera Data Warehouse service on CDP Public Cloud introduces the following new features and improvements:

Simplified private networking deployments on AWS

Private networking deployments on AWS for Data Warehouse service have been simplified. When using the Private Load Balancer, Private Worker Nodes deployment mode, both public and private subnets in your AWS VPC are no

longer needed. Now this deployment mode only requires 3 private subnets. For more information, see Supported deployment modes for private networking in AWS.

Impala Virtual Warehouses offer faster SQL analytics

Impala Virtual Warehouses have been improved with the following enhancements:

- More granular control and performance optimizations on multi-threaded query execution.
- Improved read performance for ORC tables with nested type columns.
- Improved automatic updates of metadata.

Generate and download Impala diagnostic bundles

You can now generate and download diagnostic bundles of log files for Impala Virtual Warehouses on AWS environments. For more information, see Downloading Impala diagnostic bundles from AWS.

Launch Grafana dashboards from environment tiles in Cloudera Data Warehouse service

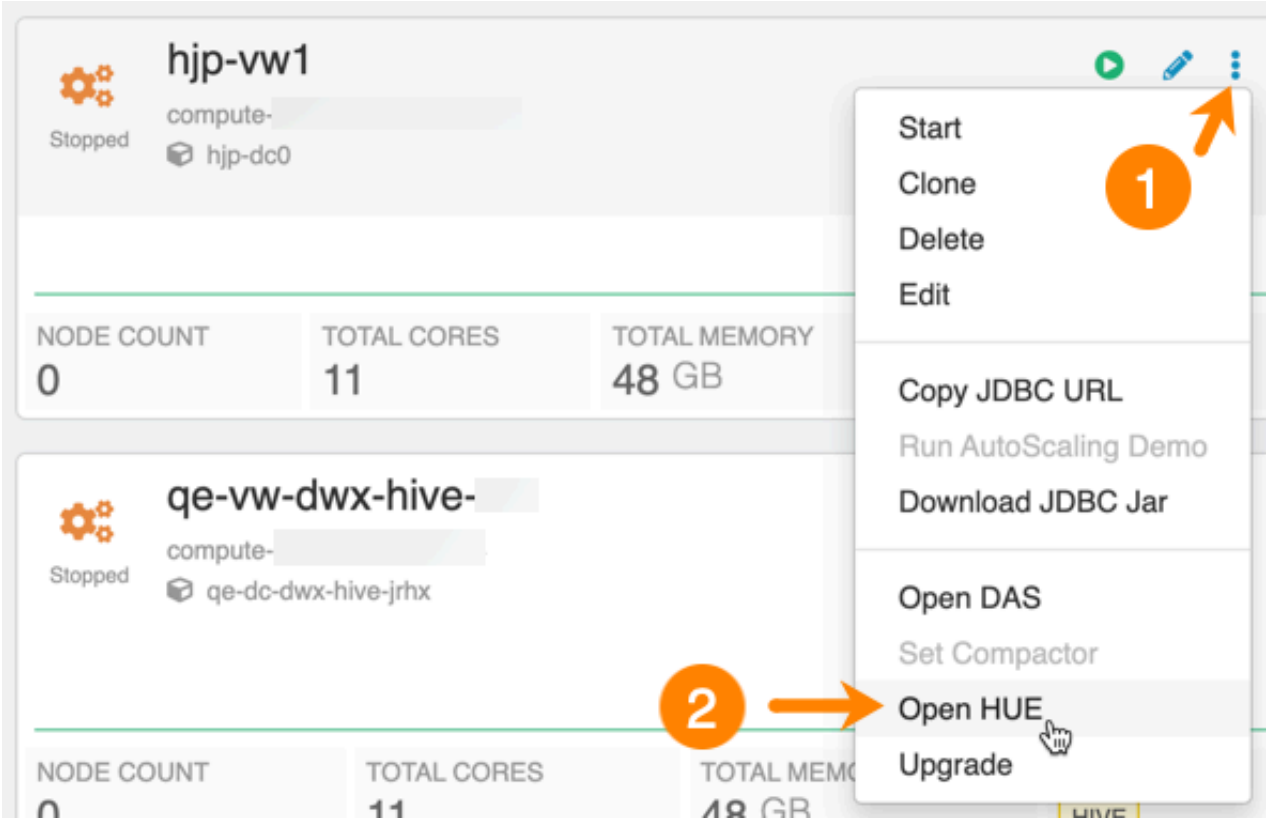
You can now launch Grafana dashboards from environment tiles in the Cloudera Data Warehouse service to monitor the following components:

- Hive
- Impala
- Istio mesh
- Kubernetes clusters

For more information, see Connecting to Grafana dashboards.

Single Sign-On (SSO) support for Hue with Hive Virtual Warehouses

You can now access Hue to query Hive Virtual Warehouses from the Virtual Warehouse tile:



DAS updates

In this release, DAS has been updated as follows:

- You can now specify administrative groups for the DAS web app.
- Multi-DAG support for a single Hive query.

April 7, 2020

This release of the Cloudera Data Warehouse service on CDP Public Cloud introduces the following new features and improvements:

Hue single-sign-on (SSO) support using SAML

In this release, Hue now supports SSO by using SAML. For more information, see Authenticating users with SAML in the Hue documentation.

Private networking feature now available for AWS environments when only private subnets are available

When you are unable to use public subnets in your AWS VPC, CDP Data Warehouse service now supports private networking for your Data Warehouse in this situation. To configure private networking for AWS environments where only private subnets are available to use, register a CDP environment in AWS with three private subnets. Then network administrators must make sure there is outbound internet access from private subnets by way of the transit gateway or by another means. For information about configuring private networking with these AWS environments, see Step 4 in Activating an AWS environment with private subnet support.

Azure support improvements

- Additional regions are now supported, including:
  - Central US
  - East US
  - East US 2
  - West US
  - West US 2
  - Australia East
- Improved identity management.

Improved resiliency for Impala Virtual Warehouses

Resiliency for Impala Virtual Warehouses has been improved by reducing the amount of ephemeral storage used on compute nodes.

March 13, 2020

This release of the Cloudera Data Warehouse service on CDP Public Cloud introduces the following new features and improvements:

Azure support for Data Warehouse service

This release of Data Warehouse service now supports Microsoft Azure as a public cloud provider. For more details about this added support, see Azure support for Data Warehouse service.

Impala Virtual Warehouse improvements

Impala Virtual Warehouses now include the following improvements:

- Improved resiliency for client connections.

- Added more health checks to more accurately indicate the start of a Virtual Warehouse. This helps with query timeouts due to provisioning lags.

Known issues (unsupported releases)

Known issues in unsupported releases.

January 18, 2024

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

See Data Visualization release notes for known issues in Cloudera Data Visualization 7.1.6.

Carried over from the previous release: Upgrade-related

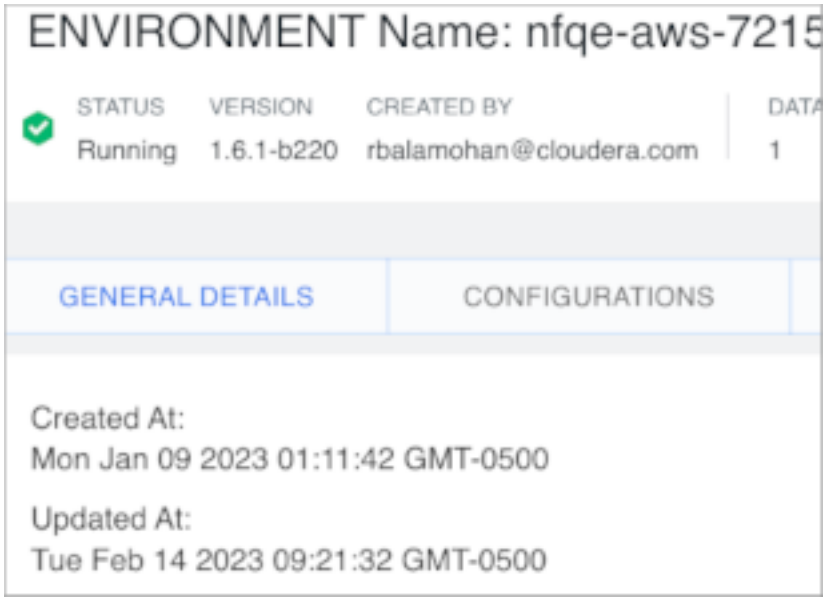
Upgrading to EKS 1.24 could result in Impala coordinators shutting down.

This issue is not seen on the Impala Virtual Warehouse running Runtime 2023.0.15.0-x or later.

Workaround: Manually start the Impala Virtual Warehouse from the UI or cli. Alternatively, replace your runtime with 1.7.1-b755 (released August 30, 2023) or later.

Certain CDW versions cannot upgrade to EKS 1.22

AWS environments activated in version 1.4.1-b86 (released June, 22, 2022) or earlier are not supported for upgrade to EKS 1.22 due to incompatibility of some components with EKS 1.22. To determine the activation version your pre-existing environment, in the Data Warehouse service, expand Environments. In Environments, search for and locate the environment that you want to view. Click Edit. In Environment Details, you see the CDW version.



DWX-15112 Enterprise Data Warehouse database configuration problems after a Helm-related rollback

If a Helm rollback fails due to an incorrect Enterprise Data Warehouse database configuration, the Virtual Warehouse and Database Catalog roll back to a previous configuration. The incorrect Enterprise Data Warehouse configuration persists, and can affect subsequent edit, upgrade, and rebuild operations on the rolled-back Virtual Warehouse or Database Catalog.

Carried over from the previous release: General

Compaction causes data loss under certain conditions

Data loss occurs during compaction when Apache Ranger policies for masking or row filtering are enabled and compaction users are included in the policies.

Workaround: Exclude compaction users from the policies as described in Compaction Prerequisites.

DWX-14923 After JWT authentication, attempting to connect the Impyla client using a user name or password should cause an error

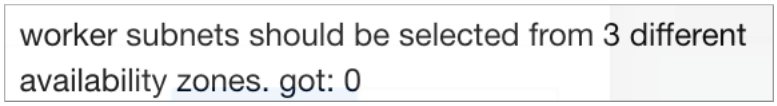
Using a JWT token, you can connect to a Virtual Warehouse as the user who generated the token.

If you connect with the JWT token, and then pass a user name or password from Impyla to the Virtual Warehouse, the connection arguments are silently ignored. Such an action should indicate that it is illegal to specify user or password when using JWT authentication.

DWX-15145 Environment validation popup error after activating an environment

Activating an environment having a public load balancer can cause an environment validatation popup error.

To reproduce this problem 1) Create a data lake. 2) Activate an enviroment having a public load balancer deployment type and subnets in three different availability zones. A environment validation popup can occur even through subnets are in different availability zones. Several differnt popups can occur, including the following one:



DWX-15144 Virtual Warehouse naming restrictions

You cannot create a Virtual Warehouse having the same name as another Virtual Warehouse even if the like-named Virtual Warehouses are in different environments. You can create a Database Catalog having the same name as another Database Catalog if the Database Catalogs are in different environments.

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
```

```
        operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the CreateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersions line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

Enabling a private CDW environment in Azure Kubernetes Service is unstable

Cloudera rolled back the General Availability (GA) status and support for deploying a private CDW environment in Azure using AKS until a solution for the issue is available.

When working together with Microsoft Azure (Azure), a networking issue has been identified at Azure Software Definition Layer (SDN) that impacts the use of CDW private environments while using the Azure Kubernetes Service (AKS). This issue is highly unpredictable and may cause timeouts during Cloudera Data Warehouse (CDW) environment activation, Virtual Warehouse creation, Virtual Warehouse modification, and/or during start and stop operations in CDW.

The following users are affected: CDW users on Azure who are activating a private CDW environment by selecting the Private CDW option on the User Interface (UI) or using CDP Command Line Interface (CLI) enablePrivateAks option for the dw sub-command create-cluster operation within the --aws-options group.

Activation can fail with a timeout issue, and other operation related actions may be interrupted.

DWX-15214 DWX-15176 Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

The Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed:

```
kubelet Error: failed to create containerd task: failed to creat
e shim task: OCI runtime create failed: runc create failed: unab
le to start container process: exec: “run_httpd.sh”: cannot run
executable found relative to current directory: unknown
```

This issue affects all CDW releases before 1.6.3 (released May 5, 2023)running on CDP Public Cloud on Azure.

The following users are affected: CDW Azure users using Hue in Impala and Hive Virtual Warehouses in environments earlier than version 1.6.3 or later if they upgrade the node image

Workaround: Upgrade the Virtual Warehouses to 1.6.3 and choose Hue version 2023.0.14.0.

Do not choose Hue versions older than 2023.0.14.0 when creating a Virtual Warehouse in environments of version 1.6.3.

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Database Catalog

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

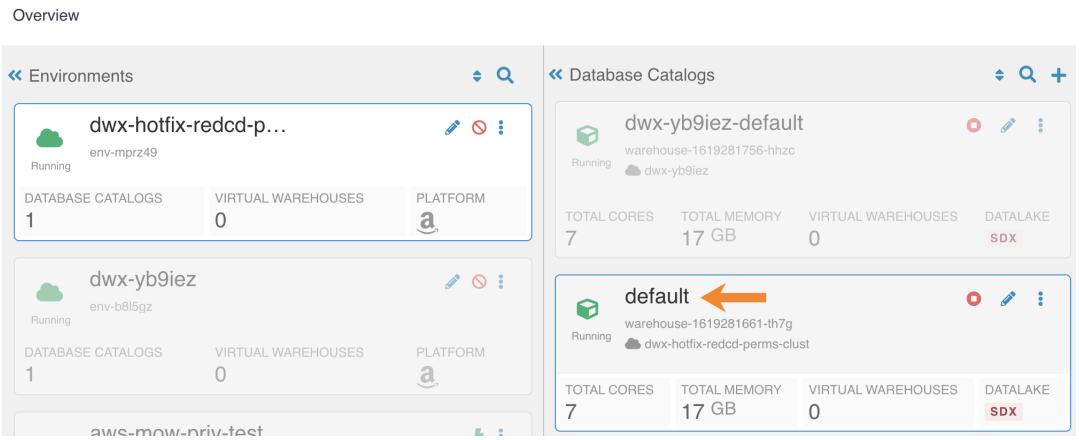
- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

Limited Hive image versions

Hive Virtual Warehouses you create in 1.8.1-b248 (released Nov 20, 2023) and later will run Istio 1.19.0. The new Istio version supports only new versions of Hive helm charts. If you have the CDW_VERSIONED_DEPLOY, only new Hive image versions appear in UI when you create a new Hive Virtual Warehouse.

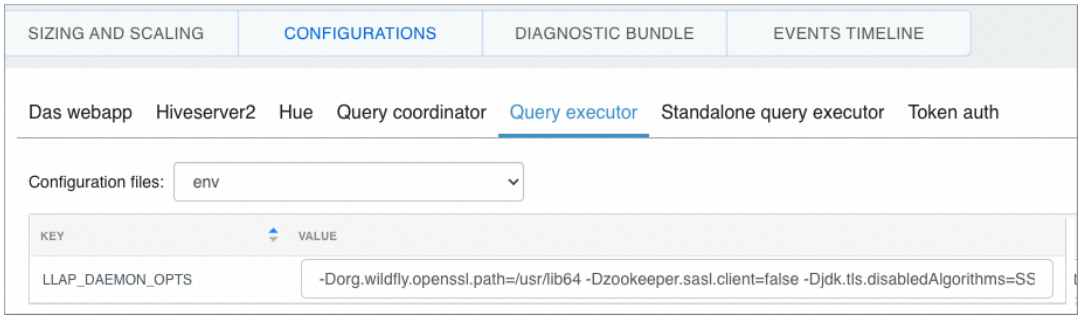
DWX-15064 Hive Virtual Warehouse stops but appears healthy

Due to an istio-proxy problem, the query coordinator can unexpectedly enter a not ready state instead of the expected error-state. Subsequently, the Hive Virtual Warehouse stops when reaching the autosuspend timeout without indicating a problem.

DWX-14452 Parquet table query might fail

Querying a table stored in Parquet from Hive might fail with the following exception message: java.lang.RuntimeException: java.lang.StackOverflowError. This problem can occur when the IN predicate has 243 values or more, and a small stack (-Xss = 256k) is configured for Hive in CDW.

Workaround: Change the value of the Xss in the JVM args to use the default value (1Mb) as follows: Select Edit from the Options menu of your Virtual Warehouse. Click Configurations > Query Executor > and select env.



In the third line shown below, change the value of LLAP_DAEMON_OPTS from -Xss256k to -Xss1M, and then click Apply Changes:

FROM:

-Dorg.wildfly.openssl.path=/usr/lib64 -Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Dhttp.maxConnections=12 -Xms24G -Xmx48G -Xss256k ...

TO:

... -Xss1M ...

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
       kind: ClusterRole
       metadata:
       name: diagnostic-data-generator-role-monitor
       rules:
       - apiGroups:
       - ""
       resources:
       - configmaps
       - events
       - pods
       - persistentvolumeclaims
       - nodes
       verbs:
       - get
       - list
       - apiGroups:
       - apps
       resources:
       - deployments
       - statefulsets
       verbs:
       - get
       - list
       - apiGroups:
       - "edws.cloudera.com"
       resources:
       - computes
       verbs:
```

```
        - get
        - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.

7. Click Apply in the upper right corner of the page to save the configuration.
8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name ‘ec2-18-219-
32-183.us-east-2.compute.amazonaws.com’ does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Hue Query Editor

DWX-16899: Error while viewing Impala job status on the mini Job Browser

When you click on the application ID after submitting an Impala query on Hue that is running on the environment level, you may notice the following error: 401 Client Error: Unauthorized for url: http://coordinator.impala-xyz.svc.cluster.local:25000/queries?json=true Must authenticate with Basic authentication. This does not happen when you do the same from Hue deployed at a Virtual Warehouse level.

None.

DWX-16913: “Results have expired” message while running a CTAS query

You may see the following message on the Hue UI when you submit a CREATE TABLE AS SELECT (CTAS) query from a Hue instance that is deployed on the environment level: “Results have expired, rerun the query if needed”. This does not happen when you do the same from Hue deployed at a Virtual Warehouse level.

None.

DWX-16917: Failed to validate proxy privileges error while running queries from Hue

You may see the following error intermittently in Hue’s pod logs while running queries from a Hue instance that is deployed at the environment level: “Failed to validate proxy privileges of <username>”.

None.

DWX-16895: Incorrect status of Hue pods when you edit the Hue instance properties

When you update a configuration of a Hue instance that is deployed at the environment level, such as increasing or decreasing the size of the Hue instance, you see a success message on the CDW UI. After some time, the status of the Hue instance also changes from “Updating” to “Running”. However, when you list the Hue pods using kubectl, you see that not all backend pods are in the running state–a few of them are still in the init state.

None. The pods come up successfully eventually after a sufficient time has passed.

DWX-16863: The upgrade button is present on the CDW UI, but Hue upgrades are not supported

You see the Upgrade button on the **Query Editor** page in the CDW UI when Hue is deployed at the environment level. However, on CDW version 1.8.1, upgrading the Hue instance that is deployed at the environment level is not supported.

None.

DWX-16893: A user can see all the queries in Job browser

In a Hue instance deployed at the environment level, by design, the Hue instances must not share the saved queries and query history with other Hue instances even for the same user. However, a logged in user is able to view all the queries executed by that user on all the Virtual Warehouses on a particular Database Catalog.

None.

Delay in listing queries in Impala Queries in the Job browser

Listing an Impala query in the Job browser can take an inordinate amount of time.

DWX-14927 Hue fails to list Iceberg snapshots

Hue does not recognize the Iceberg history queries from Hive to list table snapshots. For example, Hue indicates an error at the . before history when you run the following query.

```
select * from <db_name>.<table_name>.history
```

DWX-14968 Connection termination error in Impala queries tab after Hue inactivity

To reproduce the problem: 1) In the Impala job browser, navigate to Impala queries. 2) Wait for a few minutes.

After a few minutes of inactivity, the follow error is displayed:

Workaround: Refresh the page, or alternatively start a new session.

DWX-15115 Error displayed after clicking on hyperlink below Hue table browser

In Hue, below the table browser, clicking the hyperlink to a location causes an HTTP 500 error because the file browser is not enabled for environments that are not Ranger authorized (RAZ).

DWX-15090: CSRF error intermittently seen in the Hue Job Browser

You may intermittently see the “403 - CSRF” error on the Hue web interface as well as in the Hue logs after running Hive queries from Hue.

Reload the browser or start a new Hue session.

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Iceberg

DWX-17210: Timeout issue querying Iceberg tables from Hive

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

- fs.s3.maxConnections=1000
- fs.s3a.connection.maximum=1000

Restart the Database Catalog and Virtual Warehouse.

DWX-15014 Loading airports table in demo data fails

This issue occurs only when using a non-default Database Catalog. A invalid path error message occurs when using the load command to load the demo data airports table in Iceberg, ORC, or Parquet format.

DWX-14163 Limitations reading Iceberg tables in Avro file format from Impala

The Avro, Impala, and Iceberg specifications describe some limitations related to Avro, and those limitations exist in CDP. In addition to these, the DECIMAL type is not supported in this release.

DEX-7946 Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata if you drop the table during migration from Hive to Iceberg.

Workarounds: Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13733 Timeout issue querying Iceberg tables from Hive

A timeout issue can cause the query to fail.

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

```
fs.s3.maxConnections=1000
fs.s3a.connection.maximum=1000
```

Restart the Database Catalog and Virtual Warehouse.

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: Impala Virtual Warehouse

TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

The Cloudera Data Warehouse (CDW) Public Cloud 2023.0.14.0 (DWX-1.6.3) version incorporated a feature for performing automatic metadata synchronization across multiple Apache Impala (Impala) Virtual Warehouses. The feature is enabled by default, and relies on the Hive MetaStore events. When a certain sequence of Data Definition Language (DDL) SQL commands are executed as described below, users may encounter a java.lang.NullPointerException (NPE). The exception causes the event processor to stop processing other metadata operations.

If a CREATE TABLE command (not CREATE TABLE AS SELECT) is followed immediately (approximately within 1 second interval) by INVALIDATE METADATA or REFRESH TABLE command on the same table (either on the same Virtual Warehouse or on a different one), there is a possibility that the second command will not find the table in the catalog cache of a peer Virtual Warehouse and generate an NPE.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

DWX-17175: Impala Virtual Warehouse max executor limit issue

The Impala Virtual Warehouse has a limit of max 200 executors. The Executors slider has a max limit of 200 for creating or editing a Virtual Warehouse. If you select a bigger t-shirt size than the default, or set a custom t-shirt size, the default max limit set by the UI could be more than 200 in some cases. Exceeding the max limit causes an error during a future Edit/Upgrade/Rebuild operation.

Workaround: Manually set the max executors limit to a preferred value less than 200, but a multiple of t-shirt size, when creating or editing Impala Virtual Warehouse

DWX-15016 Impala query failure when using Catalog high availability (HA)

Impala queries could fail with InconsistentMetadataFetchException when using HA feature in CDW. This happens when leader election failover for Impala Catalog Service happens due to a transient network error.

Workaround: Disable Catalog HA. In CDW, edit your Virtual Warehouse, and turn off Enable Impala Catalog Server HA.

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml
```

```
        # Make sure that the annotation you added is present.
```

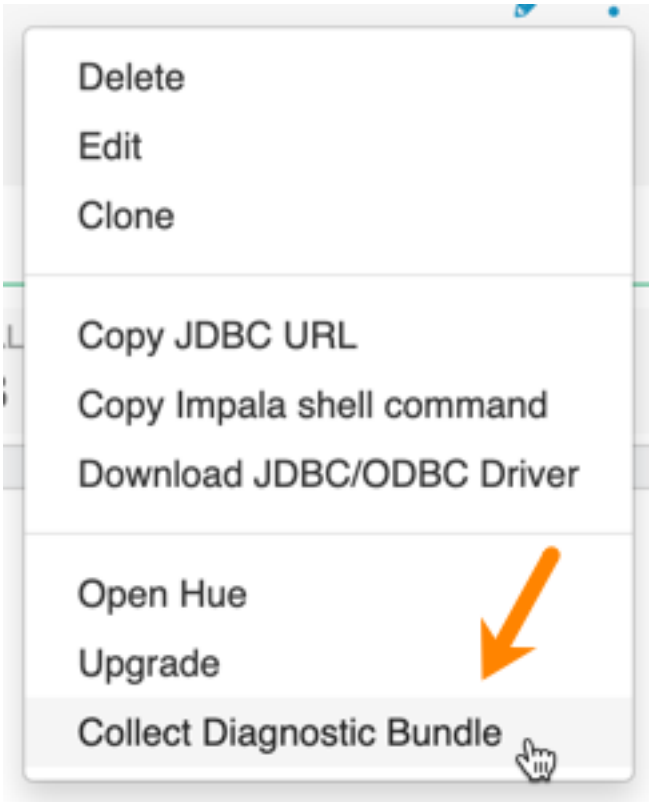
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

UI Issues

DWX-14610 Upgrade icon indicates the Database Catalog is the latest version, but might be incorrect for a few seconds

The cluster fetches data regularly in particular time intervals, or when an action is triggered. Inbetween this time interval, which is very brief, the cluster might not be updated, but will be refreshed in a few seconds.

Technical Service Bulletins

TSB 2023-719: Cloudera Data Warehouse Backup/Restore of CDP Data Visualization incomplete

Cloudera Data Warehouse (CDW) customers using the CDW Automated Backup/Restore feature will encounter an issue with the restoration versions of Cloudera Data Visualization (CDV) older

than CDV 7.1.6.2-3 due to a schema change in this release. If the Backup was taken from an older CDW environment that contained a version of CDV older than CDV 7.1.6.2-3, the Restore procedure will succeed. Though once the user opens the CDV Queries tab, the user could encounter the error message: “column jobs_jobschedule.owner_id does not exist…”

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-719: Cloudera Data Warehouse Backup/Restore of Cloudera Data Visualization incomplete.

TSB 2024-743: Restoration of the Database Catalog can be incomplete due to an incorrect Hue Query Processor configuration in CDW

Cloudera Data Warehouse (CDW) customers using the CDW Automated Backup/Restore feature might encounter an issue with the restoration of the Database Catalog, if the hue-query-processor.json configuration file of the Database Catalog has been edited. Even a minor edit to the hue-query-processor.json configuration file can result in this failure.

During the restoration process the Database Catalog will be created, but it will fail to start after a short period of time, and the Database Catalog will be in a Bad Health state on the CDW User Interface.

Inside the Kubernetes Cluster (Azure Kubernetes Service on Azure / Elastic Kubernetes Service on AWS) the StatefulSet of Hue Query Processor (hue-query-processor) is in CrashLoop. This is indicated with the following log in the Hue Query Processor StatefulSet Pod:

```
SQL State  : 3D000
Error Code : 0
Message    : FATAL: database "warehouse-1707832123-abcd_hueqpdb"
 does not exist



              at org.flywaydb.core.internal.jdbc.JdbcUtils.openC
onnection(JdbcUtils.java:65)
```

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-743: Restoration of the Database Catalog can be incomplete due to an incorrect Hue Query Processor configuration in CDW.

TSB 2024-745: Impala returns incorrect results for Iceberg V2 tables when optimized operator is being used in CDW

Cloudera Data Warehouse (CDW) customers using Apache Impala (Impala) to read Apache Iceberg (Iceberg) V2 tables can encounter an issue of Impala returning incorrect results when the optimized V2 operator is used. The optimized V2 operator is enabled by default in the affected versions below. The issue only affects Iceberg V2 tables that have position delete files.

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-745: Impala returns incorrect results for Iceberg V2 tables when optimized operator is being used in CDW.

TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables

Apache Hive (Hive) and Apache Impala (Impala) modify statements (DELETE/UPDATE/MERGE) on Apache Iceberg (Iceberg) V2 tables can corrupt the tables if there is a concurrent table compaction from Apache Spark. The issue happens when the compaction and modify statement run in parallel, and when the compaction job commits before the modify statement. In this case the position delete files of the modify statement still point to the old files. This means the following in case of

- DELETE statements
  - Deleting records pointing to old files have no effect
- UPDATE / MERGE statements
  - Deleting records pointing to old files have no effect
  - The table will also have the newly added data records
  - Rewritten records will still be active

This issue does not affect Apache NiFi (NiFi) and Apache Flink (Flink) as these components write equality delete files.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables.

TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

The Spark Iceberg library includes two procedures - rewrite_data_files and rewrite_position_delete_files. The current implementation of rewrite_data_files has a limitation that the position delete files are not deleted and still tracked by the table metadata, even if they no longer refer to an active data file. This is called the dangling delete problem. To solve this, the rewrite_position_delete_files procedure is implemented in the Spark Iceberg library to remove these old “dangling” position delete files.

Due to the dangling delete limitation, when an Iceberg table with dangling deletes is queried in Impala, Impala tries to optimize select count(*) from iceberg_table query to return the results using stats. This optimization returns incorrect results.

The following conditions must be met for this issue to occur:

- All delete files in the Iceberg table are “dangling”
  - This would occur immediately after running Spark rewrite_data_files AND
    - Before any further delete operations are performed on the table OR
    - Before Spark rewrite_position_delete_files is run on the table
- Only stats optimized plain select count(*) from iceberg_table queries are affected. For example, the query should not have:
  - Any WHERE clause
  - Any GROUP BY clause
  - Any HAVING clause

For more information, see the Apache Iceberg documentation.

Remove dangling deletes: After rewrite_data_files, position delete records pointing to the rewritten data files are not always marked for removal, and can remain tracked by the live snapshot metadata of the table. This is known as the dangling delete problem.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables.

TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion

When working with Apache Hive (Hive) and Apache Iceberg (Iceberg) V2 tables, using the TRUNCATE statement may lead to unintended data deletion. This issue arises when the truncate command is applied to a branch of an Iceberg table. Instead of truncating the branch itself, the command affects the original (main) table, which results in unintended loss of data.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion

January 10, 2024

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

See Data Visualization release notes for known issues in Cloudera Data Visualization 7.1.6.

TSB 2024-733: Changes in CoreDNS of AWS EKS affect operations in CDW, CDE and CML clusters on Public Cloud

Affected users: All Cloudera Data Platform (CDP) Public Cloud AWS users who have:

- CDW environments with a version lower than CDW 1.8.4.
- CDE on EKS-1.25 and lower version
- CML on EKS-1.25 and lower versions

The latest Amazon Web Services (AWS) Elastic Kubernetes Service (EKS) cluster deployment has changed the behavior of CoreDNS. The changes to the CoreDNS Deployment sets the readinessProbe to use the /ready endpoint, which breaks the current functions of cdp-coredns-updater. This causes a malfunction in CoreDNS that affects the Cloudera Data Warehouse (CDW), Cloudera Data Engineering (CDE) and Cloudera Machine Learning (CML) cluster operations. For the AWS CoreDNS related changes, see important CoreDNS upgrade considerations.

For the latest update on this issue see the corresponding Knowledge article: Cloudera Customer Advisory 2023-733: Don't Activate Or Backup/Restore Clusters in AWS until CDW 1.8.4 released

Carried over from the previous release: Upgrade-related

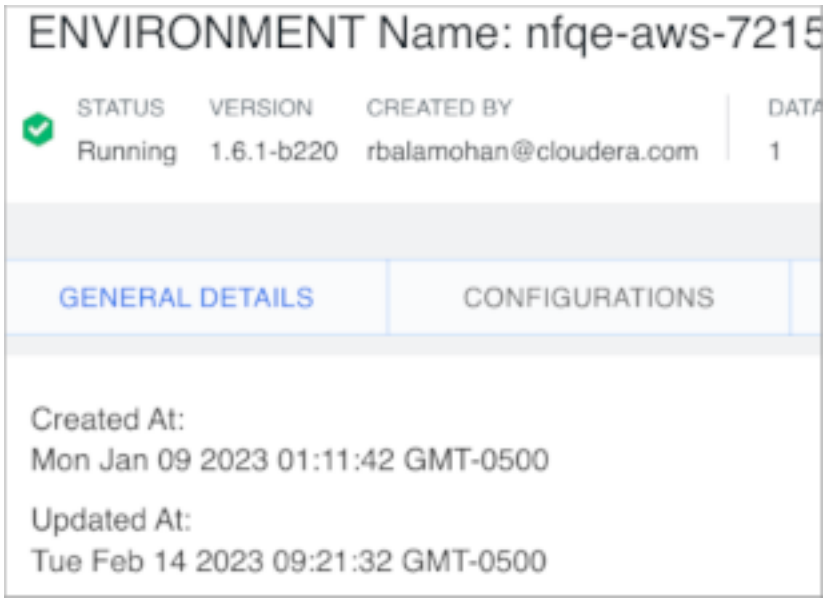
Upgrading to EKS 1.24 could result in Impala coordinators shutting down.

This issue is not seen on the Impala Virtual Warehouse running Runtime 2023.0.15.0-x or later.

Workaround: Manually start the Impala Virtual Warehouse from the UI or cli. Alternatively, replace your runtime with 1.7.1-b755 (released August 30, 2023) or later.

Certain CDW versions cannot upgrade to EKS 1.22

AWS environments activated in version 1.4.1-b86 (released June, 22, 2022) or earlier are not supported for upgrade to EKS 1.22 due to incompatibility of some components with EKS 1.22. To determine the activation version your pre-existing environment, in the Data Warehouse service, expand Environments. In Environments, search for and locate the environment that you want to view. Click Edit. In Environment Details, you see the CDW version.



DWX-15112 Enterprise Data Warehouse database configuration problems after a Helm-related rollback

If a Helm rollback fails due to an incorrect Enterprise Data Warehouse database configuration, the Virtual Warehouse and Database Catalog roll back to a previous configuration. The incorrect Enterprise Data Warehouse configuration persists, and can affect subsequent edit, upgrade, and rebuild operations on the rolled-back Virtual Warehouse or Database Catalog.

Carried over from the previous release: General

Compaction causes data loss under certain conditions

Data loss occurs during compaction when Apache Ranger policies for masking or row filtering are enabled and compaction users are included in the policies.

Workaround: Exclude compaction users from the policies as described in Compaction Prerequisites.

DWX-14923 After JWT authentication, attempting to connect the Impyla client using a user name or password should cause an error

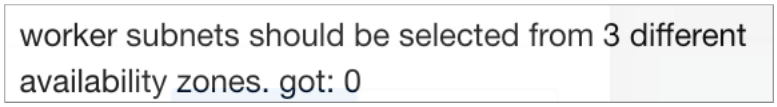
Using a JWT token, you can connect to a Virtual Warehouse as the user who generated the token.

If you connect with the JWT token, and then pass a user name or password from Impyla to the Virtual Warehouse, the connection arguments are silently ignored. Such an action should indicate that it is illegal to specify user or password when using JWT authentication.

DWX-15145 Environment validation popup error after activating an environment

Activating an environment having a public load balancer can cause an environment validatation popup error.

To reproduce this problem 1) Create a data lake. 2) Activate an enviroment having a public load balancer deployment type and subnets in three different availability zones. A environment validation popup can occur even through subnets are in different availability zones. Several differnt popups can occur, including the following one:



DWX-15144 Virtual Warehouse naming restrictions

You cannot create a Virtual Warehouse having the same name as another Virtual Warehouse even if the like-named Virtual Warehouses are in different environments. You can create a Database Catalog having the same name as another Database Catalog if the Database Catalogs are in different environments.

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the Cr eateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersio ns line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
```

```
            "autoscaling:DescribeAutoScalingGroups",
            "autoscaling:DescribeAutoScalingInstances",
            "autoscaling:DescribeTags",
            "autoscaling:DescribeLaunchConfigurations",
            "autoscaling:SetDesiredCapacity",
            "autoscaling:TerminateInstanceInAutoScalingGroup",
            "ec2:DescribeLaunchTemplateVersions",
            "ec2:CreateTags"
            ],
```

7. Save changes.

Carried over from the previous release: Azure

Enabling a private CDW environment in Azure Kubernetes Service is unstable

Cloudera rolled back the General Availability (GA) status and support for deploying a private CDW environment in Azure using AKS until a solution for the issue is available.

When working together with Microsoft Azure (Azure), a networking issue has been identified at Azure Software Definition Layer (SDN) that impacts the use of CDW private environments while using the Azure Kubernetes Service (AKS). This issue is highly unpredictable and may cause timeouts during Cloudera Data Warehouse (CDW) environment activation, Virtual Warehouse creation, Virtual Warehouse modification, and/or during start and stop operations in CDW.

The following users are affected: CDW users on Azure who are activating a private CDW environment by selecting the Private CDW option on the User Interface (UI) or using CDP Command Line Interface (CLI) enablePrivateAks option for the dw sub-command create-cluster operation within the --aws-options group.

Activation can fail with a timeout issue, and other operation related actions may be interrupted.

DWX-15214 DWX-15176 Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

The Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed:

```
kubelet Error: failed to create containerd task: failed to creat
e shim task: OCI runtime create failed: runc create failed: unab
le to start container process: exec: “run_httpd.sh”: cannot run
executable found relative to current directory: unknown
```

This issue affects all CDW releases before 1.6.3 (released May 5, 2023)running on CDP Public Cloud on Azure.

The following users are affected: CDW Azure users using Hue in Impala and Hive Virtual Warehouses in environments earlier than version 1.6.3 or later if they upgrade the node image

Workaround: Upgrade the Virtual Warehouses to 1.6.3 and choose Hue version 2023.0.14.0.

Do not choose Hue versions older than 2023.0.14.0 when creating a Virtual Warehouse in environments of version 1.6.3.

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Database Catalog

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

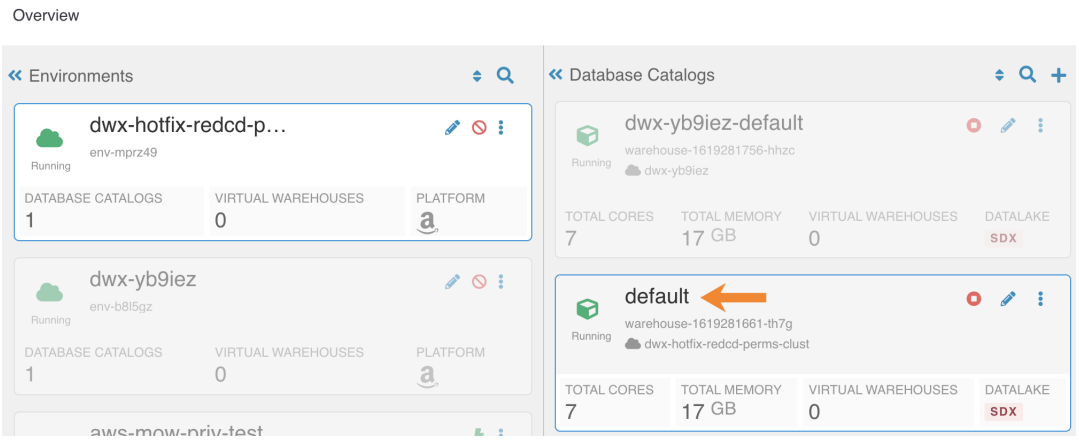
- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

Limited Hive image versions

Hive Virtual Warehouses you create in 1.8.1-b248 (released Nov 20, 2023) and later will run Istio 1.19.0. The new Istio version supports only new versions of Hive helm charts. If you have the CDW_VERSIONED_DEPLOY, only new Hive image versions appear in UI when you create a new Hive Virtual Warehouse.

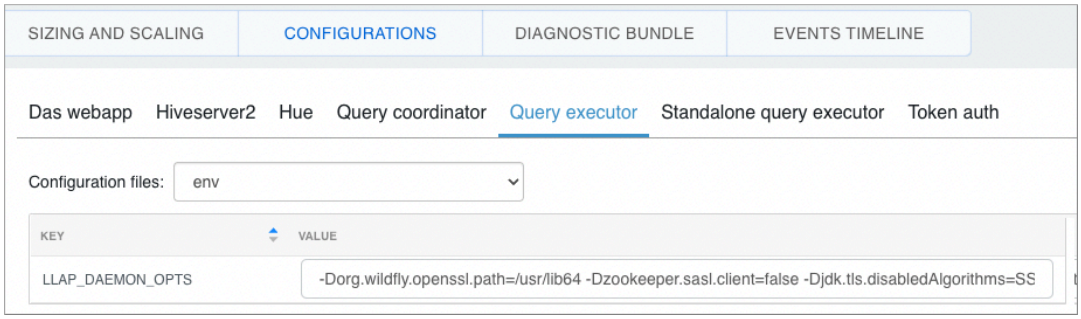
DWX-15064 Hive Virtual Warehouse stops but appears healthy

Due to an istio-proxy problem, the query coordinator can unexpectedly enter a not ready state instead of the expected error-state. Subsequently, the Hive Virtual Warehouse stops when reaching the autosuspend timeout without indicating a problem.

DWX-14452 Parquet table query might fail

Querying a table stored in Parquet from Hive might fail with the following exception message: java.lang.RuntimeException: java.lang.StackOverflowError. This problem can occur when the IN predicate has 243 values or more, and a small stack (-Xss = 256k) is configured for Hive in CDW.

Workaround: Change the value of the Xss in the JVM args to use the default value (1Mb) as follows: Select Edit from the Options menu of your Virtual Warehouse. Click Configurations > Query Executor > and select env.



In the third line shown below, change the value of LLAP_DAEMON_OPTS from -Xss256k to -Xss1M, and then click Apply Changes:

FROM:

-Dorg.wildfly.openssl.path=/usr/lib64 -Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Dhttp.maxConnections=12 -Xms24G -Xmx48G -Xss256k ...

TO:

... -Xss1M ...

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
       kind: ClusterRole
       metadata:
       name: diagnostic-data-generator-role-monitor
       rules:
       - apiGroups:
       - ""
       resources:
       - configmaps
       - events
       - pods
       - persistentvolumeclaims
       - nodes
       verbs:
       - get
       - list
       - apiGroups:
       - apps
       resources:
       - deployments
       - statefulsets
       verbs:
       - get
       - list
       - apiGroups:
       - "edws.cloudera.com"
       resources:
       - computes
       verbs:
```

```
        - get
        - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.

7. Click Apply in the upper right corner of the page to save the configuration.
8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name 'ec2-18-219-
32-183.us-east-2.compute.amazonaws.com' does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Hue Query Editor

DWX-16899: Error while viewing Impala job status on the mini Job Browser

When you click on the application ID after submitting an Impala query on Hue that is running on the environment level, you may notice the following error: 401 Client Error: Unauthorized for url: http://coordinator.impala-xyz.svc.cluster.local:25000/queries?json=true Must authenticate with Basic authentication. This does not happen when you do the same from Hue deployed at a Virtual Warehouse level.

None.

DWX-16913: "Results have expired" message while running a CTAS query

You may see the following message on the Hue UI when you submit a CREATE TABLE AS SELECT (CTAS) query from a Hue instance that is deployed on the environment level: "Results have expired, rerun the query if needed". This does not happen when you do the same from Hue deployed at a Virtual Warehouse level.

None.

DWX-16917: Failed to validate proxy privileges error while running queries from Hue

You may see the following error intermittently in Hue's pod logs while running queries from a Hue instance that is deployed at the environment level: "Failed to validate proxy privileges of <username>".

None.

DWX-16895: Incorrect status of Hue pods when you edit the Hue instance properties

When you update a configuration of a Hue instance that is deployed at the environment level, such as increasing or decreasing the size of the Hue instance, you see a success message on the CDW UI. After some time, the status of the Hue instance also changes from "Updating" to "Running". However, when you list the Hue pods using kubectl, you see that not all backend pods are in the running state–a few of them are still in the init state.

None. The pods come up successfully eventually after a sufficient time has passed.

DWX-16863: The upgrade button is present on the CDW UI, but Hue upgrades are not supported

You see the Upgrade button on the **Query Editor** page in the CDW UI when Hue is deployed at the environment level. However, on CDW version 1.8.1, upgrading the Hue instance that is deployed at the environment level is not supported.

None.

DWX-16893: A user can see all the queries in Job browser

In a Hue instance deployed at the environment level, by design, the Hue instances must not share the saved queries and query history with other Hue instances even for the same user. However, a logged in user is able to view all the queries executed by that user on all the Virtual Warehouses on a particular Database Catalog.

None.

Delay in listing queries in Impala Queries in the Job browser

Listing an Impala query in the Job browser can take an inordinate amount of time.

DWX-14927 Hue fails to list Iceberg snapshots

Hue does not recognize the Iceberg history queries from Hive to list table snapshots. For example, Hue indicates an error at the . before history when you run the following query.

```
select * from <db_name>.<table_name>.history
```

DWX-14968 Connection termination error in Impala queries tab after Hue inactivity

To reproduce the problem: 1) In the Impala job browser, navigate to Impala queries. 2) Wait for a few minutes.

After a few minutes of inactivity, the follow error is displayed:

Workaround: Refresh the page, or alternatively start a new session.

DWX-15115 Error displayed after clicking on hyperlink below Hue table browser

In Hue, below the table browser, clicking the hyperlink to a location causes an HTTP 500 error because the file browser is not enabled for environments that are not Ranger authorized (RAZ).

DWX-15090: CSRF error intermittently seen in the Hue Job Browser

You may intermittently see the “403 - CSRF” error on the Hue web interface as well as in the Hue logs after running Hive queries from Hue.

Reload the browser or start a new Hue session.

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Iceberg

DWX-17210: Timeout issue querying Iceberg tables from Hive

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

- fs.s3.maxConnections=1000
- fs.s3a.connection.maximum=1000

Restart the Database Catalog and Virtual Warehouse.

DWX-15014 Loading airports table in demo data fails

This issue occurs only when using a non-default Database Catalog. A invalid path error message occurs when using the load command to load the demo data airports table in Iceberg, ORC, or Parquet format.

DWX-14163 Limitations reading Iceberg tables in Avro file format from Impala

The Avro, Impala, and Iceberg specifications describe some limitations related to Avro, and those limitations exist in CDP. In addition to these, the DECIMAL type is not supported in this release.

DEX-7946 Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata if you drop the table during migration from Hive to Iceberg.

Workarounds: Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13733 Timeout issue querying Iceberg tables from Hive

A timeout issue can cause the query to fail.

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

```
fs.s3.maxConnections=1000
fs.s3a.connection.maximum=1000
```

Restart the Database Catalog and Virtual Warehouse.

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: Impala Virtual Warehouse

TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

The Cloudera Data Warehouse (CDW) Public Cloud 2023.0.14.0 (DWX-1.6.3) version incorporated a feature for performing automatic metadata synchronization across multiple Apache Impala (Impala) Virtual Warehouses. The feature is enabled by default, and relies on the Hive MetaStore events. When a certain sequence of Data Definition Language (DDL) SQL commands are executed as described below, users may encounter a java.lang.NullPointerException (NPE). The exception causes the event processor to stop processing other metadata operations.

If a CREATE TABLE command (not CREATE TABLE AS SELECT) is followed immediately (approximately within 1 second interval) by INVALIDATE METADATA or REFRESH TABLE command on the same table (either on the same Virtual Warehouse or on a different one), there is a possibility that the second command will not find the table in the catalog cache of a peer Virtual Warehouse and generate an NPE.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

DWX-17175: Impala Virtual Warehouse max executor limit issue

The Impala Virtual Warehouse has a limit of max 200 executors. The Executors slider has a max limit of 200 for creating or editing a Virtual Warehouse. If you select a bigger t-shirt size than the default, or set a custom t-shirt size, the default max limit set by the UI could be more than 200 in some cases. Exceeding the max limit causes an error during a future Edit/Upgrade/Rebuild operation.

Workaround: Manually set the max executors limit to a preferred value less than 200, but a multiple of t-shirt size, when creating or editing Impala Virtual Warehouse

DWX-15016 Impala query failure when using Catalog high availability (HA)

Impala queries could fail with InconsistentMetadataFetchException when using HA feature in CDW. This happens when leader election failover for Impala Catalog Service happens due to a transient network error.

Workaround: Disable Catalog HA. In CDW, edit your Virtual Warehouse, and turn off Enable Impala Catalog Server HA.

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml
```

```
        # Make sure that the annotation you added is present.
```

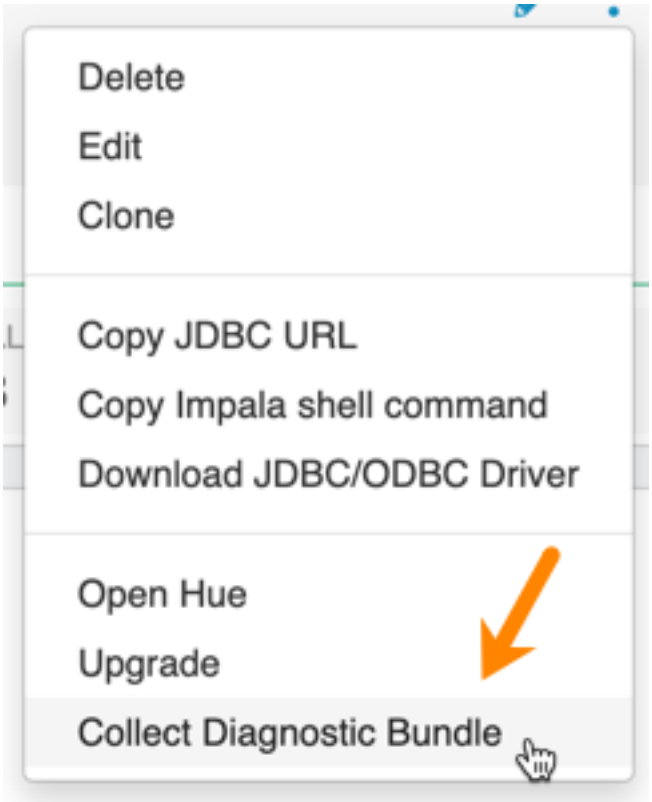
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

UI Issues

DWX-14610 Upgrade icon indicates the Database Catalog is the latest version, but might be incorrect for a few seconds

The cluster fetches data regularly in particular time intervals, or when an action is triggered. Inbetween this time interval, which is very brief, the cluster might not be updated, but will be refreshed in a few seconds.

Technical Service Bulletins

TSB 2023-719: Cloudera Data Warehouse Backup/Restore of CDP Data Visualization incomplete

Cloudera Data Warehouse (CDW) customers using the CDW Automated Backup/Restore feature will encounter an issue with the restoration versions of Cloudera Data Visualization (CDV) older

than CDV 7.1.6.2-3 due to a schema change in this release. If the Backup was taken from an older CDW environment that contained a version of CDV older than CDV 7.1.6.2-3, the Restore procedure will succeed. Though once the user opens the CDV Queries tab, the user could encounter the error message: “column jobs_jobschedule.owner_id does not exist…”

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-719: Cloudera Data Warehouse Backup/Restore of Cloudera Data Visualization incomplete.

TSB 2024-743: Restoration of the Database Catalog can be incomplete due to an incorrect Hue Query Processor configuration in CDW

Cloudera Data Warehouse (CDW) customers using the CDW Automated Backup/Restore feature might encounter an issue with the restoration of the Database Catalog, if the hue-query-processor.json configuration file of the Database Catalog has been edited. Even a minor edit to the hue-query-processor.json configuration file can result in this failure.

During the restoration process the Database Catalog will be created, but it will fail to start after a short period of time, and the Database Catalog will be in a Bad Health state on the CDW User Interface.

Inside the Kubernetes Cluster (Azure Kubernetes Service on Azure / Elastic Kubernetes Service on AWS) the StatefulSet of Hue Query Processor (hue-query-processor) is in CrashLoop. This is indicated with the following log in the Hue Query Processor StatefulSet Pod:

```
SQL State  : 3D000
Error Code : 0
Message    : FATAL: database "warehouse-1707832123-abcd_hueqpdb"
 does not exist



              at org.flywaydb.core.internal.jdbc.JdbcUtils.openC
onnection(JdbcUtils.java:65)
```

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-743: Restoration of the Database Catalog can be incomplete due to an incorrect Hue Query Processor configuration in CDW.

TSB 2024-745: Impala returns incorrect results for Iceberg V2 tables when optimized operator is being used in CDW

Cloudera Data Warehouse (CDW) customers using Apache Impala (Impala) to read Apache Iceberg (Iceberg) V2 tables can encounter an issue of Impala returning incorrect results when the optimized V2 operator is used. The optimized V2 operator is enabled by default in the affected versions below. The issue only affects Iceberg V2 tables that have position delete files.

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-745: Impala returns incorrect results for Iceberg V2 tables when optimized operator is being used in CDW.

TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables

Apache Hive (Hive) and Apache Impala (Impala) modify statements (DELETE/UPDATE/MERGE) on Apache Iceberg (Iceberg) V2 tables can corrupt the tables if there is a concurrent table compaction from Apache Spark. The issue happens when the compaction and modify statement run in parallel, and when the compaction job commits before the modify statement. In this case the position delete files of the modify statement still point to the old files. This means the following in case of

- DELETE statements
  - Deleting records pointing to old files have no effect
- UPDATE / MERGE statements
  - Deleting records pointing to old files have no effect
  - The table will also have the newly added data records
  - Rewritten records will still be active

This issue does not affect Apache NiFi (NiFi) and Apache Flink (Flink) as these components write equality delete files.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables.

TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

The Spark Iceberg library includes two procedures - rewrite_data_files and rewrite_position_delete_files. The current implementation of rewrite_data_files has a limitation that the position delete files are not deleted and still tracked by the table metadata, even if they no longer refer to an active data file. This is called the dangling delete problem. To solve this, the rewrite_position_delete_files procedure is implemented in the Spark Iceberg library to remove these old “dangling” position delete files.

Due to the dangling delete limitation, when an Iceberg table with dangling deletes is queried in Impala, Impala tries to optimize select count(*) from iceberg_table query to return the results using stats. This optimization returns incorrect results.

The following conditions must be met for this issue to occur:

- All delete files in the Iceberg table are “dangling”
  - This would occur immediately after running Spark rewrite_data_files AND
    - Before any further delete operations are performed on the table OR
    - Before Spark rewrite_position_delete_files is run on the table
- Only stats optimized plain select count(*) from iceberg_table queries are affected. For example, the query should not have:
  - Any WHERE clause
  - Any GROUP BY clause
  - Any HAVING clause

For more information, see the Apache Iceberg documentation.

Remove dangling deletes: After rewrite_data_files, position delete records pointing to the rewritten data files are not always marked for removal, and can remain tracked by the live snapshot metadata of the table. This is known as the dangling delete problem.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables.

TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion

When working with Apache Hive (Hive) and Apache Iceberg (Iceberg) V2 tables, using the TRUNCATE statement may lead to unintended data deletion. This issue arises when the truncate command is applied to a branch of an Iceberg table. Instead of truncating the branch itself, the command affects the original (main) table, which results in unintended loss of data.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion.

December 1, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

New Known Issues in this release

See Data Visualization release notes for known issues in Cloudera Data Visualization 7.1.6.

Carried over from the previous release: Upgrade-related

TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

The Spark Iceberg library includes two procedures - rewrite_data_files and rewrite_position_delete_files. The current implementation of rewrite_data_files has a limitation that the position delete files are not deleted and still tracked by the table metadata, even if they no longer refer to an active data file. This is called the dangling delete problem. To solve this, the rewrite_position_delete_files procedure is implemented in the Spark Iceberg library to remove these old “dangling” position delete files.

Due to the dangling delete limitation, when an Iceberg table with dangling deletes is queried in Impala, Impala tries to optimize select count(*) from iceberg_table query to return the results using stats. This optimization returns incorrect results.

The following conditions must be met for this issue to occur:

- All delete files in the Iceberg table are “dangling”
    - This would occur immediately after running Spark rewrite_data_files AND
        - Before any further delete operations are performed on the table OR
        - Before Spark rewrite_position_delete_files is run on the table
- Only stats optimized plain select count(*) from iceberg_table queries are affected. For example, the query should not have:
    - Any WHERE clause
    - Any GROUP BY clause
    - Any HAVING clause

For more information, see the Apache Iceberg documentation.

Remove dangling deletes: After rewrite_data_files, position delete records pointing to the rewritten data files are not always marked for removal, and can remain tracked by the live snapshot metadata of the table. This is known as the dangling delete problem.

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

TSB 2023-719: Cloudera Data Warehouse Backup/Restore of CDP Data Visualization incomplete

Cloudera Data Warehouse (CDW) customers using the CDW Automated Backup/Restore feature will encounter an issue with the restoration versions of Cloudera Data Visualization (CDV) older than CDV 7.1.6.2-3 due to a schema change in this release. If the Backup was taken from an older CDW environment that contained a version of CDV older than CDV 7.1.6.2-3, the Restore procedure will succeed. Though once the user opens the CDV Queries tab, the user could encounter the error message: “column jobs_jobschedule.owner_id does not exist…”

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-719: Cloudera Data Warehouse Backup/Restore of Cloudera Data Visualization incomplete

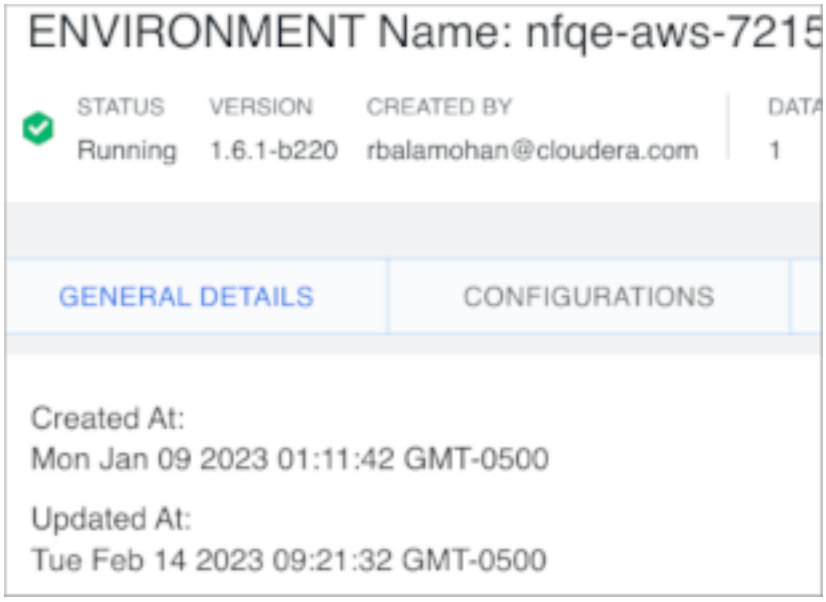
Upgrading to EKS 1.24 could result in Impala coordinators shutting down.

This issue is not seen on the Impala Virtual Warehouse running Runtime 2023.0.15.0-x or later.

Workaround: Manually start the Impala Virtual Warehouse from the UI or cli. Alternatively, replace your runtime with 1.7.1-b755 (released August 30, 2023) or later.

Certain CDW versions cannot upgrade to EKS 1.22

AWS environments activated in version 1.4.1-b86 (released June, 22, 2022) or earlier are not supported for upgrade to EKS 1.22 due to incompatibility of some components with EKS 1.22. To determine the activation version your pre-existing environment, in the Data Warehouse service, expand Environments. In Environments, search for and locate the environment that you want to view. Click Edit. In Environment Details, you see the CDW version.



DWX-15112 Enterprise Data Warehouse database configuration problems after a Helm-related rollback

If a Helm rollback fails due to an incorrect Enterprise Data Warehouse database configuration, the Virtual Warehouse and Database Catalog roll back to a previous configuration. The incorrect Enterprise Data Warehouse configuration persists, and can affect subsequent edit, upgrade, and rebuild operations on the rolled-back Virtual Warehouse or Database Catalog.

Carried over from the previous release: General

Compaction causes data loss under certain conditions

Data loss occurs during compaction when Apache Ranger policies for masking or row filtering are enabled and compaction users are included in the policies.

Workaround: Exclude compaction users from the policies as described in Compaction Prerequisites.

DWX-14923 After JWT authentication, attempting to connect the Impyla client using a user name or password should cause an error

Using a JWT token, you can connect to a Virtual Warehouse as the user who generated the token.

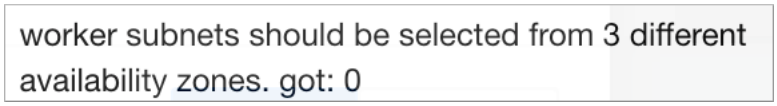
If you connect with the JWT token, and then pass a user name or password from Impyla to the Virtual Warehouse, the connection arguments are silently ignored. Such an action should indicate that it is illegal to specify user or password when using JWT authentication.

DWX-15145 Environment validation popup error after activating an environment

Activating an environment having a public load balancer can cause an environment validatation popup error.

To reproduce this problem 1) Create a data lake. 2) Activate an enviroment having a public load balancer deployment type and subnets in three different availability zones. A environment

validation popup can occur even through subnets are in different availability zones. Several differnt popups can occur, including the following one:



DWX-15144 Virtual Warehouse naming restrictions

You cannot create a Virtual Warehouse having the same name as another Virtual Warehouse even if the like-named Virtual Warehouses are in different environments. You can create a Database Catalog having the same name as another Database Catalog if the Database Catalogs are in different environments.

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the Cr eateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersio ns line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

Enabling a private CDW environment in Azure Kubernetes Service is unstable

Cloudera rolled back the General Availability (GA) status and support for deploying a private CDW environment in Azure using AKS until a solution for the issue is available.

When working together with Microsoft Azure (Azure), a networking issue has been identified at Azure Software Definition Layer (SDN) that impacts the use of CDW private environments while using the Azure Kubernetes Service (AKS). This issue is highly unpredictable and may cause timeouts during Cloudera Data Warehouse (CDW) environment activation, Virtual Warehouse creation, Virtual Warehouse modification, and/or during start and stop operations in CDW.

The following users are affected: CDW users on Azure who are activating a private CDW environment by selecting the Private CDW option on the User Interface (UI) or using CDP Command Line Interface (CLI) enablePrivateAks option for the dw sub-command create-cluster operation within the --aws-options group.

Activation can fail with a timeout issue, and other operation related actions may be interrupted.

DWX-15214 DWX-15176 Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

The Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed:

```
kubelet Error: failed to create containerd task: failed to creat
e shim task: OCI runtime create failed: runc create failed: unab
le to start container process: exec: “run_httpd.sh”: cannot run
executable found relative to current directory: unknown
```

This issue affects all CDW releases before 1.6.3 (released May 5, 2023)running on CDP Public Cloud on Azure.

The following users are affected: CDW Azure users using Hue in Impala and Hive Virtual Warehouses in environments earlier than version 1.6.3 or later if they upgrade the node image

Workaround: Upgrade the Virtual Warehouses to 1.6.3 and choose Hue version 2023.0.14.0.

Do not choose Hue versions older than 2023.0.14.0 when creating a Virtual Warehouse in environments of version 1.6.3.

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Database Catalog

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

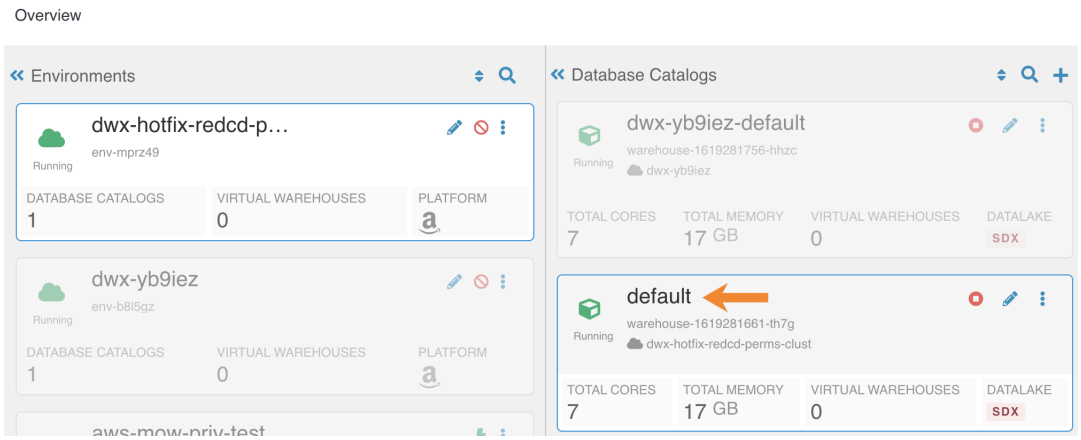
- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

Limited Hive image versions

Hive Virtual Warehouses you create in 1.8.1-b248 (released Nov 20, 2023) and later will run Istio 1.19.0. The new Istio version supports only new versions of Hive helm charts. If you have the CDW_VERSIONED_DEPLOY, only new Hive image versions appear in UI when you create a new Hive Virtual Warehouse.

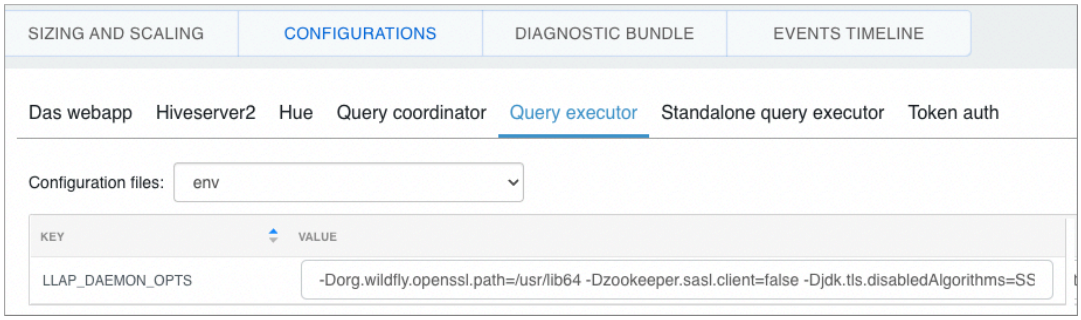
DWX-15064 Hive Virtual Warehouse stops but appears healthy

Due to an istio-proxy problem, the query coordinator can unexpectedly enter a not ready state instead of the expected error-state. Subsequently, the Hive Virtual Warehouse stops when reaching the autosuspend timeout without indicating a problem.

DWX-14452 Parquet table query might fail

Querying a table stored in Parquet from Hive might fail with the following exception message: java.lang.RuntimeException: java.lang.StackOverflowError. This problem can occur when the IN predicate has 243 values or more, and a small stack (-Xss = 256k) is configured for Hive in CDW.

Workaround: Change the value of the Xss in the JVM args to use the default value (1Mb) as follows: Select Edit from the Options menu of your Virtual Warehouse. Click Configurations > Query Executor > and select env.



In the third line shown below, change the value of LLAP_DAEMON_OPTS from -Xss256k to -Xss1M, and then click Apply Changes:

FROM:

-Dorg.wildfly.openssl.path=/usr/lib64 -Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Dhttp.maxConnections=12 -Xms24G -Xmx48G -Xss256k ...

TO:

... -Xss1M ...

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
       kind: ClusterRole
       metadata:
       name: diagnostic-data-generator-role-monitor
       rules:
       - apiGroups:
```

```
      - ""
      resources:
      - configmaps
      - events
      - pods
      - persistentvolumeclaims
      - nodes
      verbs:
      - get
      - list
      - apiGroups:
      - apps
      resources:
      - deployments
      - statefulsets
      verbs:
      - get
      - list
      - apiGroups:
      - "edws.cloudera.com"
      resources:
      - computes
      verbs:
      - get
      - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.

7. Click Apply in the upper right corner of the page to save the configuration.

8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name 'ec2-18-219-
32-183.us-east-2.compute.amazonaws.com' does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Hue Query Editor

DWX-16899: Error while viewing Impala job status on the mini Job Browser

When you click on the application ID after submitting an Impala query on Hue that is running on the environment level, you may notice the following error: 401 Client Error: Unauthorized for url: http://coordinator.impala-xyz.svc.cluster.local:25000/queries?json=true Must authenticate with Basic authentication. This does not happen when you do the same from Hue deployed at a Virtual Warehouse level.

None.

DWX-16913: "Results have expired" message while running a CTAS query

You may see the following message on the Hue UI when you submit a CREATE TABLE AS SELECT (CTAS) query from a Hue instance that is deployed on the environment level: "Results have expired, rerun the query if needed". This does not happen when you do the same from Hue deployed at a Virtual Warehouse level.

None.

DWX-16917: Failed to validate proxy privileges error while running queries from Hue

You may see the following error intermittently in Hue's pod logs while running queries from a Hue instance that is deployed at the environment level: "Failed to validate proxy privileges of <username>".

None.

DWX-16895: Incorrect status of Hue pods when you edit the Hue instance properties

When you update a configuration of a Hue instance that is deployed at the environment level, such as increasing or decreasing the size of the Hue instance, you see a success message on the CDW UI. After some time, the status of the Hue instance also changes from "Updating" to "Running".

However, when you list the Hue pods using kubectl, you see that not all backend pods are in the running state–a few of them are still in the init state.

None. The pods come up successfully eventually after a sufficient time has passed.

DWX-16863: The upgrade button is present on the CDW UI, but Hue upgrades are not supported

You see the Upgrade button on the **Query Editor** page in the CDW UI when Hue is deployed at the environment level. However, on CDW version 1.8.1, upgrading the Hue instance that is deployed at the environment level is not supported.

None.

DWX-16893: A user can see all the queries in Job browser

In a Hue instance deployed at the environment level, by design, the Hue instances must not share the saved queries and query history with other Hue instances even for the same user. However, a logged in user is able to view all the queries executed by that user on all the Virtual Warehouses on a particular Database Catalog.

None.

Delay in listing queries in Impala Queries in the Job browser

Listing an Impala query in the Job browser can take an inordinate amount of time.

DWX-14927 Hue fails to list Iceberg snapshots

Hue does not recognize the Iceberg history queries from Hive to list table snapshots. For example, Hue indicates an error at the . before history when you run the following query.

```
select * from <db_name>.<table_name>.history
```

DWX-14968 Connection termination error in Impala queries tab after Hue inactivity

To reproduce the problem: 1) In the Impala job browser, navigate to Impala queries. 2) Wait for a few minutes.

After a few minutes of inactivity, the follow error is displayed:

Workaround: Refresh the page, or alternatively start a new session.

DWX-15115 Error displayed after clicking on hyperlink below Hue table browser

In Hue, below the table browser, clicking the hyperlink to a location causes an HTTP 500 error because the file browser is not enabled for environments that are not Ranger authorized (RAZ).

DWX-15090: CSRF error intermittently seen in the Hue Job Browser

You may intermittently see the “403 - CSRF” error on the Hue web interface as well as in the Hue logs after running Hive queries from Hue.

Reload the browser or start a new Hue session.

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Iceberg

Concurrent compactions and modify statements can corrupt Iceberg tables

Hive or Impala DELETE/UPDATE/MERGE operations on Iceberg V2 tables can corrupt the tables if there is a concurrent table compaction from Spark. The issue happens if the compaction and modify statement runs in parallel, and if the compaction job commits before the modify statement. In that case the modify statement's position delete files still point to the old files. The results in the case of DELETE and in the case of UPDATE / MERGE are as follows:

- DELETE

  Delete records pointing to old files have no effect.
- UPDATE / MERGE

  Delete records pointing to old files have no effect. The table will also have the newly added data records, which means rewritten records will still be active.

Use one of the following workarounds:

- Do not run compactions and DELETE/UPDATE/MERGE statements in parallel.

- Do not compact the table via Iceberg's RewriteFiles operation. For example do not use Spark's rewriteDataFiles.

DWX-17210: Timeout issue querying Iceberg tables from Hive

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

- fs.s3.maxConnections=1000
- fs.s3a.connection.maximum=1000

Restart the Database Catalog and Virtual Warehouse.

DWX-15014 Loading airports table in demo data fails

This issue occurs only when using a non-default Database Catalog. A invalid path error message occurs when using the load command to load the demo data airports table in Iceberg, ORC, or Parquet format.

DWX-14163 Limitations reading Iceberg tables in Avro file format from Impala

The Avro, Impala, and Iceberg specifications describe some limitations related to Avro, and those limitations exist in CDP. In addition to these, the DECIMAL type is not supported in this release.

DEX-7946 Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata if you drop the table during migration from Hive to Iceberg.

Workarounds: Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13733 Timeout issue querying Iceberg tables from Hive

A timeout issue can cause the query to fail.

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

```
fs.s3.maxConnections=1000
fs.s3a.connection.maximum=1000
```

Restart the Database Catalog and Virtual Warehouse.

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.

- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: Impala Virtual Warehouse

TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

The Cloudera Data Warehouse (CDW) Public Cloud 2023.0.14.0 (DWX-1.6.3) version incorporated a feature for performing automatic metadata synchronization across multiple Apache Impala (Impala) Virtual Warehouses. The feature is enabled by default, and relies on the Hive MetaStore events. When a certain sequence of Data Definition Language (DDL) SQL commands are executed as described below, users may encounter a java.lang.NullPointerException (NPE). The exception causes the event processor to stop processing other metadata operations.

If a CREATE TABLE command (not CREATE TABLE AS SELECT) is followed immediately (approximately within 1 second interval) by INVALIDATE METADATA or REFRESH TABLE command on the same table (either on the same Virtual Warehouse or on a different one), there is a possibility that the second command will not find the table in the catalog cache of a peer Virtual Warehouse and generate an NPE.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

DWX-17175: Impala Virtual Warehouse max executor limit issue

The Impala Virtual Warehouse has a limit of max 200 executors. The Executors slider has a max limit of 200 for creating or editing a Virtual Warehouse. If you select a bigger t-shirt size than the default, or set a custom t-shirt size, the default max limit set by the UI could be more than 200 in some cases. Exceeding the max limit causes an error during a future Edit/Upgrade/Rebuild operation.

Workaround: Manually set the max executors limit to a preferred value less than 200, but a multiple of t-shirt size, when creating or editing Impala Virtual Warehouse

DWX-15016 Impala query failure when using Catalog high availability (HA)

Impala queries could fail with InconsistentMetadataFetchException when using HA feature in CDW. This happens when leader election failover for Impala Catalog Service happens due to a transient network error.

Workaround: Disable Catalog HA. In CDW, edit your Virtual Warehouse, and turn off Enable Impala Catalog Server HA.

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

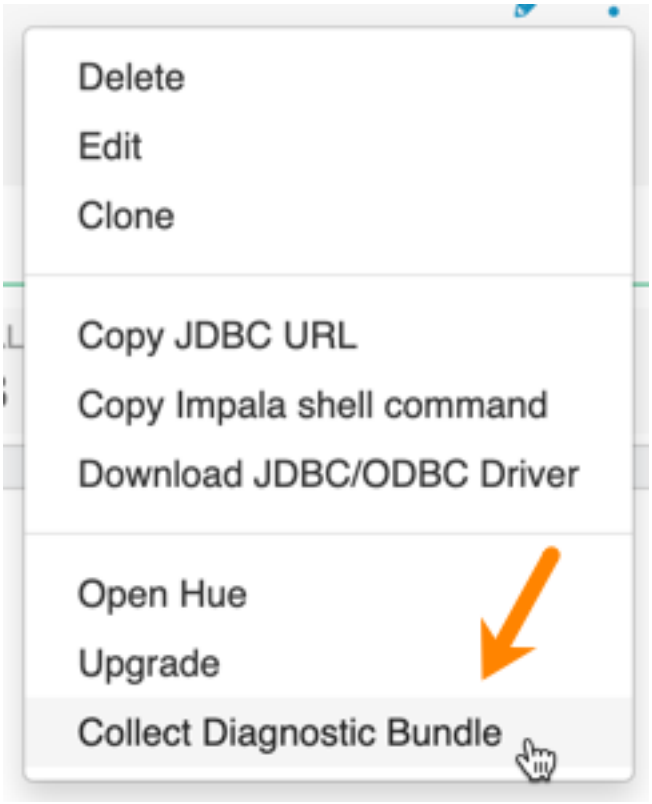
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

UI Issues

DWX-14610 Upgrade icon indicates the Database Catalog is the latest version, but might be incorrect for a few seconds

The cluster fetches data regularly in particular time intervals, or when an action is triggered. Inbetween this time interval, which is very brief, the cluster might not be updated, but will be refreshed in a few seconds.

Technical Service Bulletins

TSB 2024-723: Hue RAZ is using logger role to Read and Upload/Delete (write) files

When using Cloudera Data Hub for Public Cloud (Data Hub) on Amazon Web Services (AWS), users can use the Hue File Browser feature to access the filesystem, and if permitted, read and write directly to the related S3 buckets. As AWS does not provide fine-grained access control, Cloudera Data Platform administrators can use the Ranger Authorization Service (RAZ) capability to take the S3 filesystem, and overlay it with user and group specific permissions, making it easier to allow certain users to have limited permissions, without having to grant those users permissions to the entire S3 bucket.

This bulletin describes an issue when using RAZ with Data Hub, and attempting to use fine-grained access control to allow certain users write permissions.

Through RAZ, an administrator may, for a particular user, specify permissions more limited than what AWS provides for an S3 bucket, allowing the user to have read/write (or other similar

fine grained access) permissions on only a subset of the files and directories within that bucket. However, under specific conditions, it is possible for such user to be able to read and write to the entire S3 bucket through Hue, due to Hue using the logger role (which will have full read/write to the S3 bucket) when using Data Hub with a RAZ enabled cluster. This problem also can affect the Hue service itself, by affecting proper access to home directories causing the service role to not start.

The root cause of this issue is, when accessing Amazon cloud resources, Hue uses the AWS Boto SDK library. This AWS Boto library has a bug that restricts permissions in certain AWS regions in such a way that it provides access to users who should not have it, regardless of RAZ settings. This issue only affects users in specific AWS regions, listed below, and it does not affect all AWS customers.

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-723: Hue Raz is using logger role to Read and Upload/Delete (write) files.

December 19, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

New Known Issues in this release

See Data Visualization release notes for known issues in Cloudera Data Visualization 7.1.6.

DWX-17175: Impala Virtual Warehouse max executor limit issue

The Impala Virtual Warehouse has a limit of max 200 executors. The Executors slider has a max limit of 200 for creating or editing a Virtual Warehouse. If you select a bigger t-shirt size than the default, or set a custom t-shirt size, the default max limit set by the UI could be more than 200 in some cases. Exceeding the max limit causes an error during a future Edit/Upgrade/Rebuild operation.

Workaround: Manually set the max executors limit to a preferred value less than 200, but a multiple of t-shirt size, when creating or editing Impala Virtual Warehouse

DWX-17210: Timeout issue querying Iceberg tables from Hive

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

- fs.s3.maxConnections=1000
- fs.s3a.connection.maximum=1000

Restart the Database Catalog and Virtual Warehouse.

Carried over from the previous release: Upgrade-related

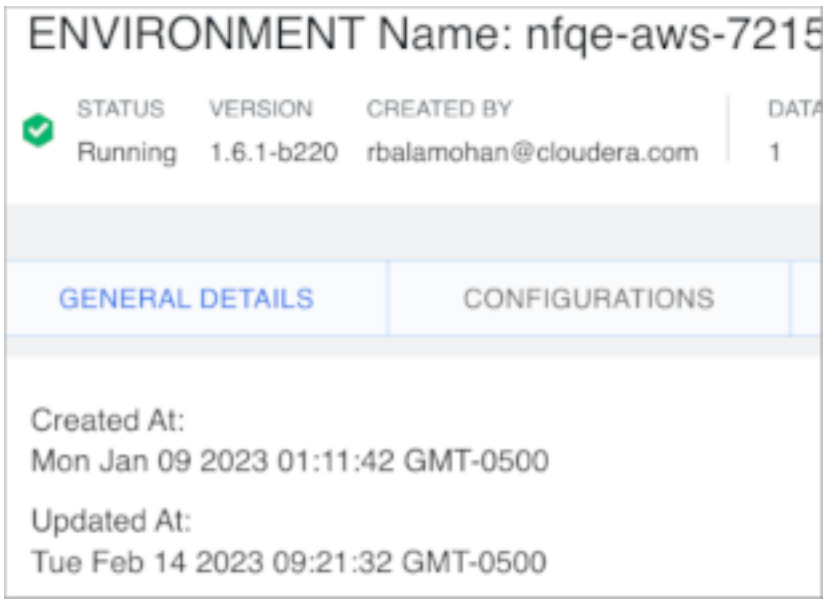
Upgrading to EKS 1.24 could result in Impala coordinators shutting down.

This issue is not seen on the Impala Virtual Warehouse running Runtime 2023.0.15.0-x or later.

Workaround: Manually start the Impala Virtual Warehouse from the UI or cli. Alternatively, replace your runtime with 1.7.1-b755 (released August 30, 2023) or later.

Certain CDW versions cannot upgrade to EKS 1.22

AWS environments activated in version 1.4.1-b86 (released June, 22, 2022) or earlier are not supported for upgrade to EKS 1.22 due to incompatibility of some components with EKS 1.22. To determine the activation version your pre-existing environment, in the Data Warehouse service, expand Environments. In Environments, search for and locate the environment that you want to view. Click Edit. In Environment Details, you see the CDW version.



DWX-15112 Enterprise Data Warehouse database configuration problems after a Helm-related rollback

If a Helm rollback fails due to an incorrect Enterprise Data Warehouse database configuration, the Virtual Warehouse and Database Catalog roll back to a previous configuration. The incorrect Enterprise Data Warehouse configuration persists, and can affect subsequent edit, upgrade, and rebuild operations on the rolled-back Virtual Warehouse or Database Catalog.

Carried over from the previous release: General

Compaction causes data loss under certain conditions

Data loss occurs during compaction when Apache Ranger policies for masking or row filtering are enabled and compaction users are included in the policies.

Workaround: Exclude compaction users from the policies as described in Compaction Prerequisites.

DWX-14923 After JWT authentication, attempting to connect the Impyla client using a user name or password should cause an error

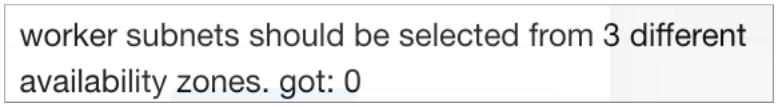
Using a JWT token, you can connect to a Virtual Warehouse as the user who generated the token.

If you connect with the JWT token, and then pass a user name or password from Impyla to the Virtual Warehouse, the connection arguments are silently ignored. Such an action should indicate that it is illegal to specify user or password when using JWT authentication.

DWX-15145 Environment validation popup error after activating an environment

Activating an environment having a public load balancer can cause an environment validatation popup error.

To reproduce this problem 1) Create a data lake. 2) Activate an enviroment having a public load balancer deployment type and subnets in three different availability zones. A environment validation popup can occur even through subnets are in different availability zones. Several differnt popups can occur, including the following one:



DWX-15144 Virtual Warehouse naming restrictions

You cannot create a Virtual Warehouse having the same name as another Virtual Warehouse even if the like-named Virtual Warehouses are in different environments. You can create a Database Catalog having the same name as another Database Catalog if the Database Catalogs are in different environments.

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the CreateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersions line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

Enabling a private CDW environment in Azure Kubernetes Service is unstable

Cloudera rolled back the General Availability (GA) status and support for deploying a private CDW environment in Azure using AKS until a solution for the issue is available.

When working together with Microsoft Azure (Azure), a networking issue has been identified at Azure Software Definition Layer (SDN) that impacts the use of CDW private environments while using the Azure Kubernetes Service (AKS). This issue is highly unpredictable and may cause timeouts during Cloudera Data Warehouse (CDW) environment activation, Virtual Warehouse creation, Virtual Warehouse modification, and/or during start and stop operations in CDW.

The following users are affected: CDW users on Azure who are activating a private CDW environment by selecting the Private CDW option on the User Interface (UI) or using CDP Command Line Interface (CLI) enablePrivateAks option for the dw sub-command create-cluster operation within the --aws-options group.

Activation can fail with a timeout issue, and other operation related actions may be interrupted.

DWX-15214 DWX-15176 Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

The Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed:

```
kubelet Error: failed to create containerd task: failed to creat
e shim task: OCI runtime create failed: runc create failed: unab
```

```
le to start container process: exec: “run_httpd.sh”: cannot run
executable found relative to current directory: unknown
```

This issue affects all CDW releases before 1.6.3 (released May 5, 2023)running on CDP Public Cloud on Azure.

The following users are affected: CDW Azure users using Hue in Impala and Hive Virtual Warehouses in environments earlier than version 1.6.3 or later if they upgrade the node image

Workaround: Upgrade the Virtual Warehouses to 1.6.3 and choose Hue version 2023.0.14.0.

Do not choose Hue versions older than 2023.0.14.0 when creating a Virtual Warehouse in environments of version 1.6.3.

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Database Catalog

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).

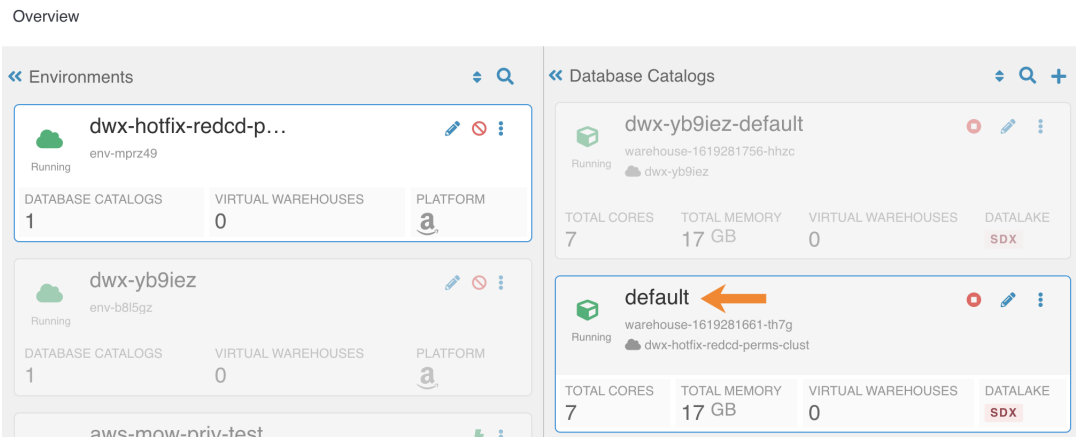
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

Limited Hive image versions

Hive Virtual Warehouses you create in 1.8.1-b248 (released Nov 20, 2023) and later will run Istio 1.19.0. The new Istio version supports only new versions of Hive helm charts. If you have the CDW_VERSIONED_DEPLOY, only new Hive image versions appear in UI when you create a new Hive Virtual Warehouse.

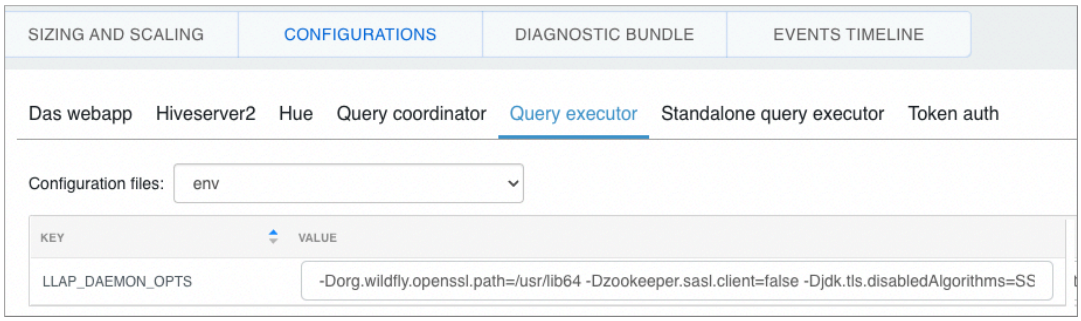
DWX-15064 Hive Virtual Warehouse stops but appears healthy

Due to an istio-proxy problem, the query coordinator can unexpectedly enter a not ready state instead of the expected error-state. Subsequently, the Hive Virtual Warehouse stops when reaching the autosuspend timeout without indicating a problem.

DWX-14452 Parquet table query might fail

Querying a table stored in Parquet from Hive might fail with the following exception message: java.lang.RuntimeException: java.lang.StackOverflowError. This problem can occur when the IN predicate has 243 values or more, and a small stack (-Xss = 256k) is configured for Hive in CDW.

Workaround: Change the value of the Xss in the JVM args to use the default value (1Mb) as follows: Select Edit from the Options menu of your Virtual Warehouse. Click Configurations > Query Executor > and select env.



In the third line shown below, change the value of LLAP_DAEMON_OPTS from -Xss256k to -Xss1M, and then click Apply Changes:

FROM:

-Dorg.wildfly.openssl.path=/usr/lib64 -Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Dhttp.maxConnections=12 -Xms24G -Xmx48G -Xss256k ...

TO:

... -Xss1M ...

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
       kind: ClusterRole
       metadata:
       name: diagnostic-data-generator-role-monitor
       rules:
       - apiGroups:
       - ""
       resources:
       - configmaps
       - events
       - pods
       - persistentvolumeclaims
       - nodes
       verbs:
       - get
       - list
       - apiGroups:
```

```
        - apps
        resources:
        - deployments
        - statefulsets
        verbs:
        - get
        - list
        - apiGroups:
        - "edws.cloudera.com"
        resources:
        - computes
        verbs:
        - get
        - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.

7. Click Apply in the upper right corner of the page to save the configuration.

8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name 'ec2-18-219-
32-183.us-east-2.compute.amazonaws.com' does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Hue Query Editor

DWX-16899: Error while viewing Impala job status on the mini Job Browser

When you click on the application ID after submitting an Impala query on Hue that is running on the environment level, you may notice the following error: 401 Client Error: Unauthorized for url: http://coordinator.impala-xyz.svc.cluster.local:25000/queries?json=true Must authenticate with Basic authentication. This does not happen when you do the same from Hue deployed at a Virtual Warehouse level.

None.

DWX-16913: "Results have expired" message while running a CTAS query

You may see the following message on the Hue UI when you submit a CREATE TABLE AS SELECT (CTAS) query from a Hue instance that is deployed on the environment level: "Results have expired, rerun the query if needed". This does not happen when you do the same from Hue deployed at a Virtual Warehouse level.

None.

DWX-16917: Failed to validate proxy privileges error while running queries from Hue

You may see the following error intermittently in Hue's pod logs while running queries from a Hue instance that is deployed at the environment level: "Failed to validate proxy privileges of <username>".

None.

DWX-16895: Incorrect status of Hue pods when you edit the Hue instance properties

When you update a configuration of a Hue instance that is deployed at the environment level, such as increasing or decreasing the size of the Hue instance, you see a success message on the CDW UI. After some time, the status of the Hue instance also changes from "Updating" to "Running".

However, when you list the Hue pods using kubectl, you see that not all backend pods are in the running state–a few of them are still in the init state.

None. The pods come up successfully eventually after a sufficient time has passed.

DWX-16863: The upgrade button is present on the CDW UI, but Hue upgrades are not supported

You see the Upgrade button on the **Query Editor** page in the CDW UI when Hue is deployed at the environment level. However, on CDW version 1.8.1, upgrading the Hue instance that is deployed at the environment level is not supported.

None.

DWX-16893: A user can see all the queries in Job browser

In a Hue instance deployed at the environment level, by design, the Hue instances must not share the saved queries and query history with other Hue instances even for the same user. However, a logged in user is able to view all the queries executed by that user on all the Virtual Warehouses on a particular Database Catalog.

None.

Delay in listing queries in Impala Queries in the Job browser

Listing an Impala query in the Job browser can take an inordinate amount of time.

DWX-14927 Hue fails to list Iceberg snapshots

Hue does not recognize the Iceberg history queries from Hive to list table snapshots. For example, Hue indicates an error at the . before history when you run the following query.

```
select * from <db_name>.<table_name>.history
```

DWX-14968 Connection termination error in Impala queries tab after Hue inactivity

To reproduce the problem: 1) In the Impala job browser, navigate to Impala queries. 2) Wait for a few minutes.

After a few minutes of inactivity, the follow error is displayed:

Workaround: Refresh the page, or alternatively start a new session.

DWX-15115 Error displayed after clicking on hyperlink below Hue table browser

In Hue, below the table browser, clicking the hyperlink to a location causes an HTTP 500 error because the file browser is not enabled for environments that are not Ranger authorized (RAZ).

DWX-15090: CSRF error intermittently seen in the Hue Job Browser

You may intermittently see the “403 - CSRF” error on the Hue web interface as well as in the Hue logs after running Hive queries from Hue.

Reload the browser or start a new Hue session.

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Iceberg

DWX-15014 Loading airports table in demo data fails

This issue occurs only when using a non-default Database Catalog. A invalid path error message occurs when using the load command to load the demo data airports table in Iceberg, ORC, or Parquet format.

DWX-14163 Limitations reading Iceberg tables in Avro file format from Impala

The Avro, Impala, and Iceberg specifications describe some limitations related to Avro, and those limitations exist in CDP. In addition to these, the DECIMAL type is not supported in this release.

DEX-7946 Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata if you drop the table during migration from Hive to Iceberg.

Workarounds: Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13733 Timeout issue querying Iceberg tables from Hive

A timeout issue can cause the query to fail.

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

```
fs.s3.maxConnections=1000
fs.s3a.connection.maximum=1000
```

Restart the Database Catalog and Virtual Warehouse.

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: Impala Virtual Warehouse

TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

The Cloudera Data Warehouse (CDW) Public Cloud 2023.0.14.0 (DWX-1.6.3) version incorporated a feature for performing automatic metadata synchronization across multiple Apache Impala (Impala) Virtual Warehouses. The feature is enabled by default, and relies on the Hive MetaStore events. When a certain sequence of Data Definition Language (DDL) SQL commands are executed as described below, users may encounter a java.lang.NullPointerException (NPE). The exception causes the event processor to stop processing other metadata operations.

If a CREATE TABLE command (not CREATE TABLE AS SELECT) is followed immediately (approximately within 1 second interval) by INVALIDATE METADATA or REFRESH TABLE command on the same table (either on the same Virtual Warehouse or on a different one), there is a possibility that the second command will not find the table in the catalog cache of a peer Virtual Warehouse and generate an NPE.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

DWX-15016 Impala query failure when using Catalog high availability (HA)

Impala queries could fail with InconsistentMetadataFetchException when using HA feature in CDW. This happens when leader election failover for Impala Catalog Service happens due to a transient network error.

Workaround: Disable Catalog HA. In CDW, edit your Virtual Warehouse, and turn off Enable Impala Catalog Server HA.

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.
```

```
kubectl get svc nginx-service -n <cluster-name> -o yaml

# Make sure that the annotation you added is present.
```

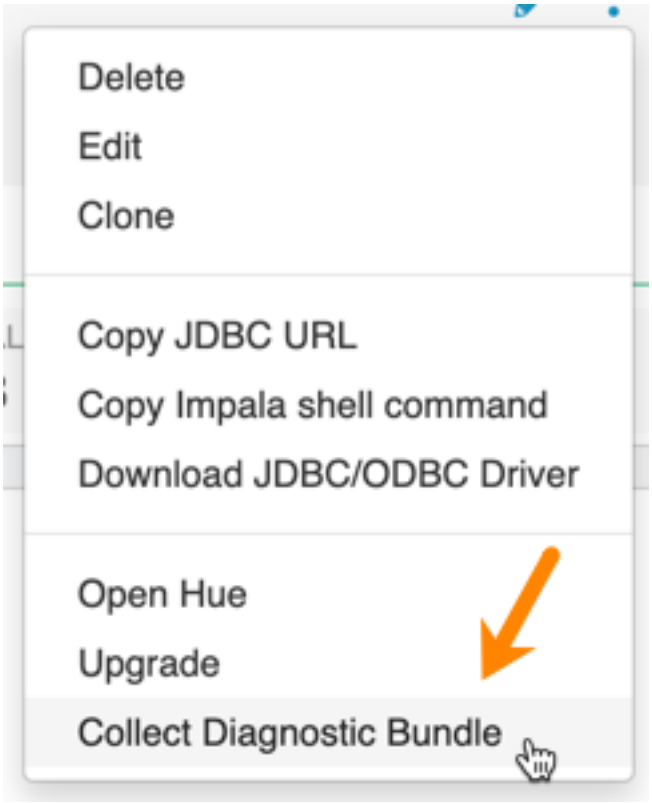
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

UI Issues

DWX-14610 Upgrade icon indicates the Database Catalog is the latest version, but might be incorrect for a few seconds

The cluster fetches data regularly in particular time intervals, or when an action is triggered. Inbetween this time interval, which is very brief, the cluster might not be updated, but will be refreshed in a few seconds.

Technical Service Bulletins

TSB 2023-719: Cloudera Data Warehouse Backup/Restore of CDP Data Visualization incomplete

Cloudera Data Warehouse (CDW) customers using the CDW Automated Backup/Restore feature will encounter an issue with the restoration versions of Cloudera Data Visualization (CDV) older than CDV 7.1.6.2-3 due to a schema change in this release. If the Backup was taken from an older CDW environment that contained a version of CDV older than CDV 7.1.6.2-3, the Restore procedure will succeed. Though once the user opens the CDV Queries tab, the user could encounter the error message: “column jobs_jobschedule.owner_id does not exist…”

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-719: Cloudera Data Warehouse Backup/Restore of Cloudera Data Visualization incomplete.

TSB 2024-743: Restoration of the Database Catalog can be incomplete due to an incorrect Hue Query Processor configuration in CDW

Cloudera Data Warehouse (CDW) customers using the CDW Automated Backup/Restore feature might encounter an issue with the restoration of the Database Catalog, if the hue-query-processor.json configuration file of the Database Catalog has been edited. Even a minor edit to the hue-query-processor.json configuration file can result in this failure.

During the restoration process the Database Catalog will be created, but it will fail to start after a short period of time, and the Database Catalog will be in a Bad Health state on the CDW User Interface.

Inside the Kubernetes Cluster (Azure Kubernetes Service on Azure / Elastic Kubernetes Service on AWS) the StatefulSet of Hue Query Processor (hue-query-processor) is in CrashLoop. This is indicated with the following log in the Hue Query Processor StatefulSet Pod:

```
SQL State  : 3D000
Error Code : 0
Message    : FATAL: database "warehouse-1707832123-abcd_hueqpdb"
 does not exist



              at org.flywaydb.core.internal.jdbc.JdbcUtils.openC
onnection(JdbcUtils.java:65)
```

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-743: Restoration of the Database Catalog can be incomplete due to an incorrect Hue Query Processor configuration in CDW.

TSB 2024-745: Impala returns incorrect results for Iceberg V2 tables when optimized operator is being used in CDW

Cloudera Data Warehouse (CDW) customers using Apache Impala (Impala) to read Apache Iceberg (Iceberg) V2 tables can encounter an issue of Impala returning incorrect results when the optimized V2 operator is used. The optimized V2 operator is enabled by default in the affected versions below. The issue only affects Iceberg V2 tables that have position delete files.

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-745: Impala returns incorrect results for Iceberg V2 tables when optimized operator is being used in CDW.

TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables

Apache Hive (Hive) and Apache Impala (Impala) modify statements (DELETE/UPDATE/MERGE) on Apache Iceberg (Iceberg) V2 tables can corrupt the tables if there is a concurrent table compaction from Apache Spark. The issue happens when the compaction and modify statement run in parallel, and when the compaction job commits before the modify statement. In this case the

position delete files of the modify statement still point to the old files. This means the following in case of

- DELETE statements
  - Deleting records pointing to old files have no effect
- UPDATE / MERGE statements
  - Deleting records pointing to old files have no effect
  - The table will also have the newly added data records
  - Rewritten records will still be active

This issue does not affect Apache NiFi (NiFi) and Apache Flink (Flink) as these components write equality delete files.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables.

TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

The Spark Iceberg library includes two procedures - rewrite_data_files and rewrite_position_delete_files. The current implementation of rewrite_data_files has a limitation that the position delete files are not deleted and still tracked by the table metadata, even if they no longer refer to an active data file. This is called the dangling delete problem. To solve this, the rewrite_position_delete_files procedure is implemented in the Spark Iceberg library to remove these old “dangling” position delete files.

Due to the dangling delete limitation, when an Iceberg table with dangling deletes is queried in Impala, Impala tries to optimize select count(*) from iceberg_table query to return the results using stats. This optimization returns incorrect results.

The following conditions must be met for this issue to occur:

- All delete files in the Iceberg table are “dangling”
  - This would occur immediately after running Spark rewrite_data_files AND
    - Before any further delete operations are performed on the table OR
    - Before Spark rewrite_position_delete_files is run on the table
- Only stats optimized plain select count(*) from iceberg_table queries are affected. For example, the query should not have:
  - Any WHERE clause
  - Any GROUP BY clause
  - Any HAVING clause

For more information, see the Apache Iceberg documentation.

Remove dangling deletes: After rewrite_data_files, position delete records pointing to the rewritten data files are not always marked for removal, and can remain tracked by the live snapshot metadata of the table. This is known as the dangling delete problem.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables.

TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion

When working with Apache Hive (Hive) and Apache Iceberg (Iceberg) V2 tables, using the TRUNCATE statement may lead to unintended data deletion. This issue arises when the truncate

command is applied to a branch of an Iceberg table. Instead of truncating the branch itself, the command affects the original (main) table, which results in unintended loss of data.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion

November 20, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

New Known Issues in this release

See Data Visualization release notes for known issues in Cloudera Data Visualization 7.1.6.

Compaction causes data loss under certain conditions

Data loss occurs during compaction when Apache Ranger policies for masking or row filtering are enabled and compaction users are included in the policies.

Workaround: Exclude compaction users from the policies as described in Compaction Prerequisites.

Limited Hive image versions

Hive Virtual Warehouses you create in 1.8.1-b248 (released Nov 20, 2023) will run Istio 1.19.0. The new Istio version supports only new versions of Hive helm charts. If you have the CDW_VERSIONED_DEPLOY, only new Hive image versions appear in UI when you create a new Hive Virtual Warehouse.

DWX-16899: Error while viewing Impala job status on the mini Job Browser

When you click on the application ID after submitting an Impala query on Hue that is running on the environment level, you may notice the following error: 401 Client Error: Unauthorized for url: http://coordinator.impala-xyz.svc.cluster.local:25000/queries?json=true Must authenticate with Basic authentication. This does not happen when you do the same from Hue deployed at a Virtual Warehouse level.

None.

DWX-16913: “Results have expired” message while running a CTAS query

You may see the following message on the Hue UI when you submit a CREATE TABLE AS SELECT (CTAS) query from a Hue instance that is deployed on the environment level: “Results have expired, rerun the query if needed”. This does not happen when you do the same from Hue deployed at a Virtual Warehouse level.

None.

DWX-16917: Failed to validate proxy privileges error while running queries from Hue

You may see the following error intermittently in Hue’s pod logs while running queries from a Hue instance that is deployed at the environment level: “Failed to validate proxy privileges of <username>”.

None.

DWX-16895: Incorrect status of Hue pods when you edit the Hue instance properties

When you update a configuration of a Hue instance that is deployed at the environment level, such as increasing or decreasing the size of the Hue instance, you see a success message on the CDW UI. After some time, the status of the Hue instance also changes from “Updating” to “Running”. However, when you list the Hue pods using kubectl, you see that not all backend pods are in the running state–a few of them are still in the init state.

None. The pods come up successfully eventually after a sufficient time has passed.

DWX-16863: The upgrade button is present on the CDW UI, but Hue upgrades are not supported

You see the Upgrade button on the **Query Editor** page in the CDW UI when Hue is deployed at the environment level. However, on CDW version 1.8.1, upgrading the Hue instance that is deployed at the environment level is not supported.

None.

DWX-16893: A user can see all the queries in Job browser

In a Hue instance deployed at the environment level, by design, the Hue instances must not share the saved queries and query history with other Hue instances even for the same user. However, a logged in user is able to view all the queries executed by that user on all the Virtual Warehouses on a particular Database Catalog.

None.

CDW backup and restore Data Visualization issue

TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

The Spark Iceberg library includes two procedures - rewrite_data_files and rewrite_position_delete_files. The current implementation of rewrite_data_files has a limitation that the position delete files are not deleted and still tracked by the table metadata, even if they no longer refer to an active data file. This is called the dangling delete problem. To solve this, the rewrite_position_delete_files procedure is implemented in the Spark Iceberg library to remove these old “dangling” position delete files.

Due to the dangling delete limitation, when an Iceberg table with dangling deletes is queried in Impala, Impala tries to optimize select count(*) from iceberg_table query to return the results using stats. This optimization returns incorrect results.

The following conditions must be met for this issue to occur:

- All delete files in the Iceberg table are “dangling”
    - This would occur immediately after running Spark rewrite_data_files AND
        - Before any further delete operations are performed on the table OR
        - Before Spark rewrite_position_delete_files is run on the table
- Only stats optimized plain select count(*) from iceberg_table queries are affected. For example, the query should not have:
    - Any WHERE clause
    - Any GROUP BY clause
    - Any HAVING clause

For more information, see the Apache Iceberg documentation.

Remove dangling deletes: After rewrite_data_files, position delete records pointing to the rewritten data files are not always marked for removal, and can remain tracked by the live snapshot metadata of the table. This is known as the dangling delete problem.

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

TSB 2023-719: Cloudera Data Warehouse Backup/Restore of CDP Data Visualization incomplete

Cloudera Data Warehouse (CDW) customers using the CDW Automated Backup/Restore feature will encounter an issue with the restoration versions of Cloudera Data Visualization (CDV) older than CDV 7.1.6.2-3 due to a schema change in this release. If the Backup was taken from an older CDW environment that contained a version of CDV older than CDV 7.1.6.2-3, the Restore

procedure will succeed. Though once the user opens the CDV Queries tab, the user could encounter the error message: “column jobs_jobschedule.owner_id does not exist…”

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-719: Cloudera Data Warehouse Backup/Restore of Cloudera Data Visualization incomplete

Carried over from the previous release: Upgrade-related

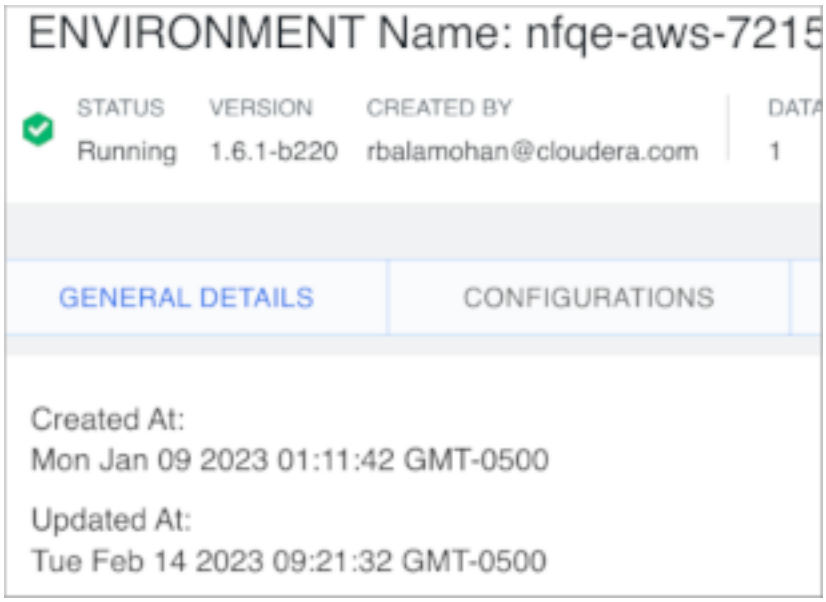
Upgrading to EKS 1.24 could result in Impala coordinators shutting down.

This issue is not seen on the Impala Virtual Warehouse running Runtime 2023.0.15.0-x or later.

Workaround: Manually start the Impala Virtual Warehouse from the UI or cli. Alternatively, replace your runtime with 1.7.1-b755 (released August 30, 2023) or later.

Certain CDW versions cannot upgrade to EKS 1.22

AWS environments activated in version 1.4.1-b86 (released June, 22, 2022) or earlier are not supported for upgrade to EKS 1.22 due to incompatibility of some components with EKS 1.22. To determine the activation version your pre-existing environment, in the Data Warehouse service, expand Environments. In Environments, search for and locate the environment that you want to view. Click Edit. In Environment Details, you see the CDW version.



DWX-15112 Enterprise Data Warehouse database configuration problems after a Helm-related rollback

If a Helm rollback fails due to an incorrect Enterprise Data Warehouse database configuration, the Virtual Warehouse and Database Catalog roll back to a previous configuration. The incorrect Enterprise Data Warehouse configuration persists, and can affect subsequent edit, upgrade, and rebuild operations on the rolled-back Virtual Warehouse or Database Catalog.

Carried over from the previous release: General

DWX-14923 After JWT authentication, attempting to connect the Impyla client using a user name or password should cause an error

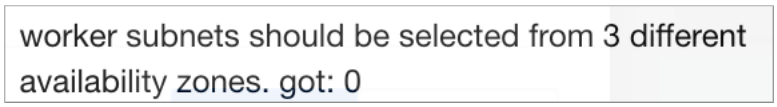
Using a JWT token, you can connect to a Virtual Warehouse as the user who generated the token.

If you connect with the JWT token, and then pass a user name or password from Impyla to the Virtual Warehouse, the connection arguments are silently ignored. Such an action should indicate that it is illegal to specify user or password when using JWT authentication.

DWX-15145 Environment validation popup error after activating an environment

Activating an environment having a public load balancer can cause an environment validatation popup error.

To reproduce this problem 1) Create a data lake. 2) Activate an enviroment having a public load balancer deployment type and subnets in three different availability zones. A environment validation popup can occur even through subnets are in different availability zones. Several differnt popups can occur, including the following one:



DWX-15144 Virtual Warehouse naming restrictions

You cannot create a Virtual Warehouse having the same name as another Virtual Warehouse even if the like-named Virtual Warehouses are in different environments. You can create a Database Catalog having the same name as another Database Catalog if the Database Catalogs are in different environments.

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the CreateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersions line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

Enabling a private CDW environment in Azure Kubernetes Service is unstable

Cloudera rolled back the General Availability (GA) status and support for deploying a private CDW environment in Azure using AKS until a solution for the issue is available.

When working together with Microsoft Azure (Azure), a networking issue has been identified at Azure Software Definition Layer (SDN) that impacts the use of CDW private environments while using the Azure Kubernetes Service (AKS). This issue is highly unpredictable and may cause timeouts during Cloudera Data Warehouse (CDW) environment activation, Virtual Warehouse creation, Virtual Warehouse modification, and/or during start and stop operations in CDW.

The following users are affected: CDW users on Azure who are activating a private CDW environment by selecting the Private CDW option on the User Interface (UI) or using CDP Command Line Interface (CLI) enablePrivateAks option for the dw sub-command create-cluster operation within the --aws-options group.

Activation can fail with a timeout issue, and other operation related actions may be interrupted.

DWX-15214 DWX-15176 Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

The Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed:

```
kubelet Error: failed to create containerd task: failed to creat
e shim task: OCI runtime create failed: runc create failed: unab
le to start container process: exec: “run_httpd.sh”: cannot run
executable found relative to current directory: unknown
```

This issue affects all CDW releases before 1.6.3 (released May 5, 2023)running on CDP Public Cloud on Azure.

The following users are affected: CDW Azure users using Hue in Impala and Hive Virtual Warehouses in environments earlier than version 1.6.3 or later if they upgrade the node image

Workaround: Upgrade the Virtual Warehouses to 1.6.3 and choose Hue version 2023.0.14.0.

Do not choose Hue versions older than 2023.0.14.0 when creating a Virtual Warehouse in environments of version 1.6.3.

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Database Catalog

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

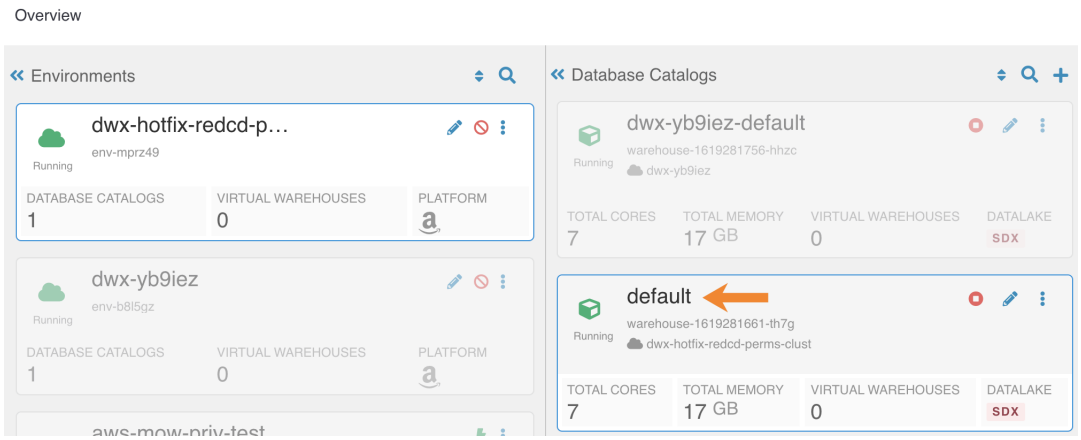
- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

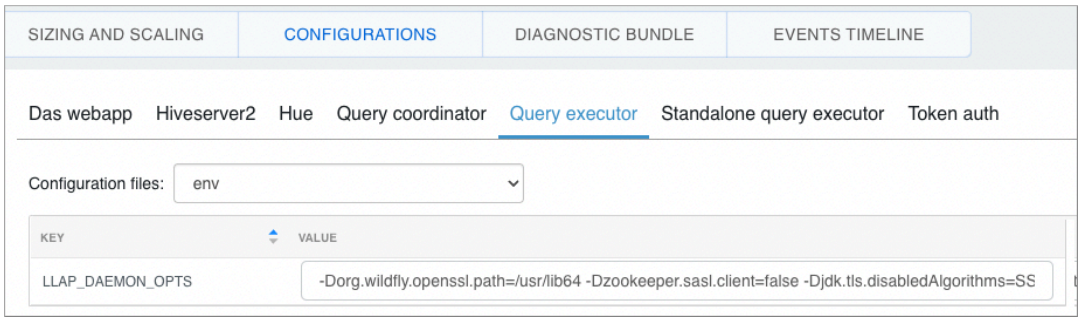
DWX-15064 Hive Virtual Warehouse stops but appears healthy

Due to an istio-proxy problem, the query coordinator can unexpectedly enter a not ready state instead of the expected error-state. Subsequently, the Hive Virtual Warehouse stops when reaching the autosuspend timeout without indicating a problem.

DWX-14452 Parquet table query might fail

Querying a table stored in Parquet from Hive might fail with the following exception message: java.lang.RuntimeException: java.lang.StackOverflowError. This problem can occur when the IN predicate has 243 values or more, and a small stack (-Xss = 256k) is configured for Hive in CDW.

Workaround: Change the value of the Xss in the JVM args to use the default value (1Mb) as follows: Select Edit from the Options menu of your Virtual Warehouse. Click Configurations > Query Executor > and select env.



In the third line shown below, change the value of LLAP_DAEMON_OPTS from -Xss256k to -Xss1M, and then click Apply Changes:

FROM:

-Dorg.wildfly.openssl.path=/usr/lib64 -Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Dhttp.maxConnections=12 -Xms24G -Xmx48G -Xss256k ...

TO:

... -Xss1M ...

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
       kind: ClusterRole
       metadata:
       name: diagnostic-data-generator-role-monitor
       rules:
       - apiGroups:
       - ""
       resources:
       - configmaps
       - events
       - pods
       - persistentvolumeclaims
       - nodes
       verbs:
       - get
       - list
       - apiGroups:
```

```
        - apps
        resources:
        - deployments
        - statefulsets
        verbs:
        - get
        - list
        - apiGroups:
        - "edws.cloudera.com"
        resources:
        - computes
        verbs:
        - get
        - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.

7. Click Apply in the upper right corner of the page to save the configuration.

8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name ‘ec2-18-219-
32-183.us-east-2.compute.amazonaws.com’ does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Hue Query Editor

Delay in listing queries in Impala Queries in the Job browser

Listing an Impala query in the Job browser can take an inordinate amount of time.

DWX-14927 Hue fails to list Iceberg snapshots

Hue does not recognize the Iceberg history queries from Hive to list table snapshots. For example, Hue indicates an error at the . before history when you run the following query.

```
select * from <db_name>.<table_name>.history
```

DWX-14968 Connection termination error in Impala queries tab after Hue inactivity

To reproduce the problem: 1) In the Impala job browser, navigate to Impala queries. 2) Wait for a few minutes.

After a few minutes of inactivity, the follow error is displayed:

Workaround: Refresh the page, or alternatively start a new session.

DWX-15115 Error displayed after clicking on hyperlink below Hue table browser

In Hue, below the table browser, clicking the hyperlink to a location causes an HTTP 500 error because the file browser is not enabled for environments that are not Ranger authorized (RAZ).

DWX-15090: CSRF error intermittently seen in the Hue Job Browser

You may intermittently see the “403 - CSRF” error on the Hue web interface as well as in the Hue logs after running Hive queries from Hue.

Reload the browser or start a new Hue session.

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Iceberg

DWX-15014 Loading airports table in demo data fails

This issue occurs only when using a non-default Database Catalog. A invalid path error message occurs when using the load command to load the demo data airports table in Iceberg, ORC, or Parquet format.

DWX-14163 Limitations reading Iceberg tables in Avro file format from Impala

The Avro, Impala, and Iceberg specifications describe some limitations related to Avro, and those limitations exist in CDP. In addition to these, the DECIMAL type is not supported in this release.

DEX-7946 Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata if you drop the table during migration from Hive to Iceberg.

Workarounds: Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13733 Timeout issue querying Iceberg tables from Hive

A timeout issue can cause the query to fail.

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

```
fs.s3.maxConnections=1000
fs.s3a.connection.maximum=1000
```

Restart the Database Catalog and Virtual Warehouse.

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: Impala Virtual Warehouse

TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

The Cloudera Data Warehouse (CDW) Public Cloud 2023.0.14.0 (DWX-1.6.3) version incorporated a feature for performing automatic metadata synchronization across multiple Apache Impala (Impala) Virtual Warehouses. The feature is enabled by default, and relies on the Hive MetaStore events. When a certain sequence of Data Definition Language (DDL) SQL commands are executed as described below, users may encounter a java.lang.NullPointerException (NPE). The exception causes the event processor to stop processing other metadata operations.

If a CREATE TABLE command (not CREATE TABLE AS SELECT) is followed immediately (approximately within 1 second interval) by INVALIDATE METADATA or REFRESH TABLE command on the same table (either on the same Virtual Warehouse or on a different one), there is

a possibility that the second command will not find the table in the catalog cache of a peer Virtual Warehouse and generate an NPE.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

DWX-15016 Impala query failure when using Catalog high availability (HA)

Impala queries could fail with InconsistentMetadataFetchException when using HA feature in CDW. This happens when leader election failover for Impala Catalog Service happens due to a transient network error.

Workaround: Disable Catalog HA. In CDW, edit your Virtual Warehouse, and turn off Enable Impala Catalog Server HA.

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

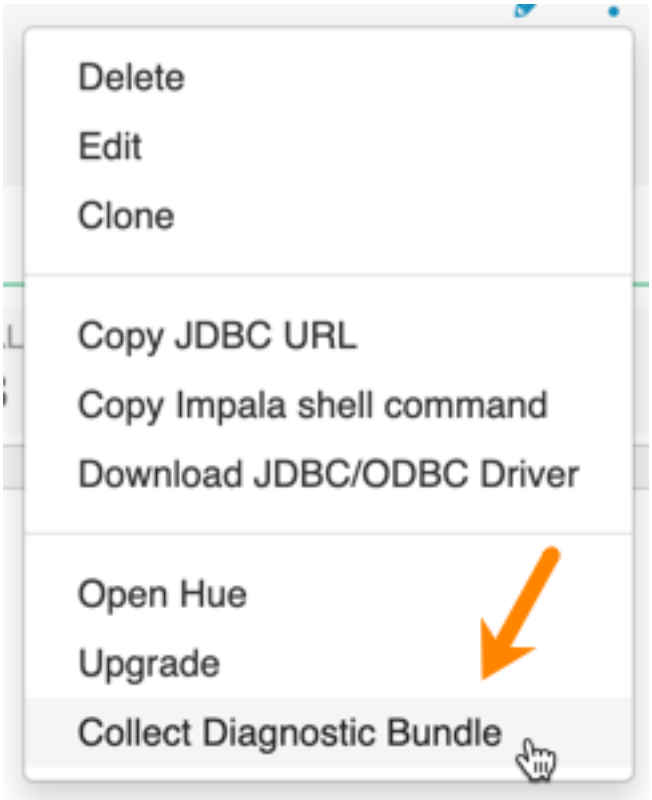
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

UI Issues

DWX-14610 Upgrade icon indicates the Database Catalog is the latest version, but might be incorrect for a few seconds

The cluster fetches data regularly in particular time intervals, or when an action is triggered. Inbetween this time interval, which is very brief, the cluster might not be updated, but will be refreshed in a few seconds.

Technical Service Bulletins

TSB 2024-723: Hue RAZ is using logger role to Read and Upload/Delete (write) files

When using Cloudera Data Hub for Public Cloud (Data Hub) on Amazon Web Services (AWS), users can use the Hue File Browser feature to access the filesystem, and if permitted, read and write directly to the related S3 buckets. As AWS does not provide fine-grained access control, Cloudera Data Platform administrators can use the Ranger Authorization Service (RAZ) capability to take the S3 filesystem, and overlay it with user and group specific permissions, making it easier to allow certain users to have limited permissions, without having to grant those users permissions to the entire S3 bucket.

This bulletin describes an issue when using RAZ with Data Hub, and attempting to use fine-grained access control to allow certain users write permissions.

Through RAZ, an administrator may, for a particular user, specify permissions more limited than what AWS provides for an S3 bucket, allowing the user to have read/write (or other similar fine grained access) permissions on only a subset of the files and directories within that bucket. However, under specific conditions, it is possible for such user to be able to read and write to the entire S3 bucket through Hue, due to Hue using the logger role (which will have full read/write to the S3 bucket) when using Data Hub with a RAZ enabled cluster. This problem also can affect the Hue service itself, by affecting proper access to home directories causing the service role to not start.

The root cause of this issue is, when accessing Amazon cloud resources, Hue uses the AWS Boto SDK library. This AWS Boto library has a bug that restricts permissions in certain AWS regions in such a way that it provides access to users who should not have it, regardless of RAZ settings. This issue only affects users in specific AWS regions, listed below, and it does not affect all AWS customers.

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-723: Hue Raz is using logger role to Read and Upload/Delete (write) files.

TSB 2024-743: Restoration of the Database Catalog can be incomplete due to an incorrect Hue Query Processor configuration in CDW

Cloudera Data Warehouse (CDW) customers using the CDW Automated Backup/Restore feature might encounter an issue with the restoration of the Database Catalog, if the hue-query-processor.json configuration file of the Database Catalog has been edited. Even a minor edit to the hue-query-processor.json configuration file can result in this failure.

During the restoration process the Database Catalog will be created, but it will fail to start after a short period of time, and the Database Catalog will be in a Bad Health state on the CDW User Interface.

Inside the Kubernetes Cluster (Azure Kubernetes Service on Azure / Elastic Kubernetes Service on AWS) the StatefulSet of Hue Query Processor (hue-query-processor) is in CrashLoop. This is indicated with the following log in the Hue Query Processor StatefulSet Pod:

```
SQL State  : 3D000
Error Code : 0
Message    : FATAL: database "warehouse-1707832123-abcd_hueqpdb"
 does not exist



              at org.flywaydb.core.internal.jdbc.JdbcUtils.openC
onnection(JdbcUtils.java:65)
```

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-743: Restoration of the Database Catalog can be incomplete due to an incorrect Hue Query Processor configuration in CDW.

TSB 2024-745: Impala returns incorrect results for Iceberg V2 tables when optimized operator is being used in CDW

Cloudera Data Warehouse (CDW) customers using Apache Impala (Impala) to read Apache Iceberg (Iceberg) V2 tables can encounter an issue of Impala returning incorrect results when the optimized V2 operator is used. The optimized V2 operator is enabled by default in the affected versions below. The issue only affects Iceberg V2 tables that have position delete files.

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-745: Impala returns incorrect results for Iceberg V2 tables when optimized operator is being used in CDW.

TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables

Apache Hive (Hive) and Apache Impala (Impala) modify statements (DELETE/UPDATE/MERGE) on Apache Iceberg (Iceberg) V2 tables can corrupt the tables if there is a concurrent table compaction from Apache Spark. The issue happens when the compaction and modify statement run in parallel, and when the compaction job commits before the modify statement. In this case the position delete files of the modify statement still point to the old files. This means the following in case of

- DELETE statements
  - Deleting records pointing to old files have no effect
- UPDATE / MERGE statements
  - Deleting records pointing to old files have no effect
  - The table will also have the newly added data records
  - Rewritten records will still be active

This issue does not affect Apache NiFi (NiFi) and Apache Flink (Flink) as these components write equality delete files.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables.

TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

The Spark Iceberg library includes two procedures - rewrite_data_files and rewrite_position_delete_files. The current implementation of rewrite_data_files has a limitation that the position delete files are not deleted and still tracked by the table metadata, even if they no longer refer to an active data file. This is called the dangling delete problem. To solve this, the rewrite_position_delete_files procedure is implemented in the Spark Iceberg library to remove these old “dangling” position delete files.

Due to the dangling delete limitation, when an Iceberg table with dangling deletes is queried in Impala, Impala tries to optimize select count(*) from iceberg_table query to return the results using stats. This optimization returns incorrect results.

The following conditions must be met for this issue to occur:

- All delete files in the Iceberg table are “dangling”
  - This would occur immediately after running Spark rewrite_data_files AND
    - Before any further delete operations are performed on the table OR
    - Before Spark rewrite_position_delete_files is run on the table
- Only stats optimized plain select count(*) from iceberg_table queries are affected. For example, the query should not have:
  - Any WHERE clause
  - Any GROUP BY clause
  - Any HAVING clause

For more information, see the Apache Iceberg documentation.

Remove dangling deletes: After rewrite_data_files, position delete records pointing to the rewritten data files are not always marked for removal, and can remain tracked by the live snapshot metadata of the table. This is known as the dangling delete problem.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables.

TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion

When working with Apache Hive (Hive) and Apache Iceberg (Iceberg) V2 tables, using the TRUNCATE statement may lead to unintended data deletion. This issue arises when the truncate command is applied to a branch of an Iceberg table. Instead of truncating the branch itself, the command affects the original (main) table, which results in unintended loss of data.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion

October 16, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

New Known Issues in this release

See Data Visualization release notes for known issues in Cloudera Data Visualization 7.1.3.

Carried over from the previous release: Upgrade-related

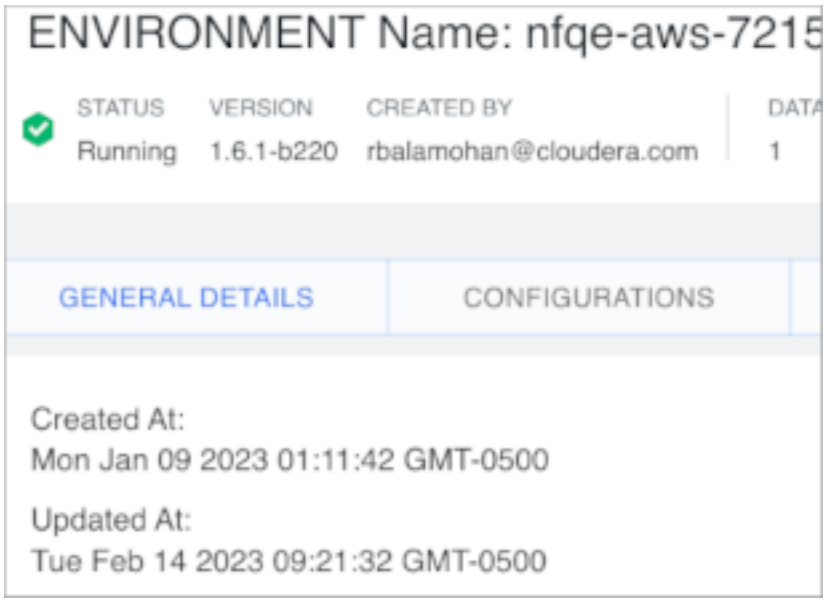
Upgrading to EKS 1.24 could result in Impala coordinators shutting down.

This issue is not seen on the Impala Virtual Warehouse running Runtime 2023.0.15.0-x or later.

Workaround: Manually start the Impala Virtual Warehouse from the UI or cli. Alternatively, replace your runtime with 1.7.1-b755 (released August 30, 2023) or later.

Certain CDW versions cannot upgrade to EKS 1.22

AWS environments activated in version 1.4.1-b86 (released June, 22, 2022) or earlier are not supported for upgrade to EKS 1.22 due to incompatibility of some components with EKS 1.22. To determine the activation version your pre-existing environment, in the Data Warehouse service, expand Environments. In Environments, search for and locate the environment that you want to view. Click Edit. In Environment Details, you see the CDW version.



DWX-15112 Enterprise Data Warehouse database configuration problems after a Helm-related rollback

If a Helm rollback fails due to an incorrect Enterprise Data Warehouse database configuration, the Virtual Warehouse and Database Catalog roll back to a previous configuration. The incorrect Enterprise Data Warehouse configuration persists, and can affect subsequent edit, upgrade, and rebuild operations on the rolled-back Virtual Warehouse or Database Catalog.

Carried over from the previous release: General

Compaction causes data loss under certain conditions

Data loss occurs during compaction when Apache Ranger policies for masking or row filtering are enabled and compaction users are included in the policies.

Workaround: Exclude compaction users from the policies as described in Compaction Prerequisites.

DWX-14923 After JWT authentication, attempting to connect the Impyla client using a user name or password should cause an error

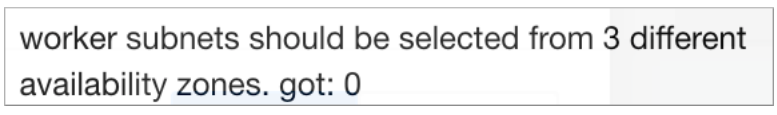
Using a JWT token, you can connect to a Virtual Warehouse as the user who generated the token.

If you connect with the JWT token, and then pass a user name or password from Impyla to the Virtual Warehouse, the connection arguments are silently ignored. Such an action should indicate that it is illegal to specify user or password when using JWT authentication.

DWX-15145 Environment validation popup error after activating an environment

Activating an environment having a public load balancer can cause an environment validatation popup error.

To reproduce this problem 1) Create a data lake. 2) Activate an enviroment having a public load balancer deployment type and subnets in three different availability zones. A environment validation popup can occur even through subnets are in different availability zones. Several differnt popups can occur, including the following one:



DWX-15144 Virtual Warehouse naming restrictions

You cannot create a Virtual Warehouse having the same name as another Virtual Warehouse even if the like-named Virtual Warehouses are in different environments. You can create a Database Catalog having the same name as another Database Catalog if the Database Catalogs are in different environments.

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the Cr eateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersio ns line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

Enabling a private CDW environment in Azure Kubernetes Service is unstable

Cloudera rolled back the General Availability (GA) status and support for deploying a private CDW environment in Azure using AKS until a solution for the issue is available.

When working together with Microsoft Azure (Azure), a networking issue has been identified at Azure Software Definition Layer (SDN) that impacts the use of CDW private environments while using the Azure Kubernetes Service (AKS). This issue is highly unpredictable and may cause timeouts during Cloudera Data Warehouse (CDW) environment activation, Virtual Warehouse creation, Virtual Warehouse modification, and/or during start and stop operations in CDW.

The following users are affected: CDW users on Azure who are activating a private CDW environment by selecting the Private CDW option on the User Interface (UI) or using CDP Command Line Interface (CLI) enablePrivateAks option for the dw sub-command create-cluster operation within the --aws-options group.

Activation can fail with a timeout issue, and other operation related actions may be interrupted.

DWX-15214 DWX-15176 Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

The Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed:

```
kubelet Error: failed to create containerd task: failed to creat
e shim task: OCI runtime create failed: runc create failed: unab
le to start container process: exec: “run_httpd.sh”: cannot run
executable found relative to current directory: unknown
```

This issue affects all CDW releases before 1.6.3 (released May 5, 2023)running on CDP Public Cloud on Azure.

The following users are affected: CDW Azure users using Hue in Impala and Hive Virtual Warehouses in environments earlier than version 1.6.3 or later if they upgrade the node image

Workaround: Upgrade the Virtual Warehouses to 1.6.3 and choose Hue version 2023.0.14.0.

Do not choose Hue versions older than 2023.0.14.0 when creating a Virtual Warehouse in environments of version 1.6.3.

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Database Catalog

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

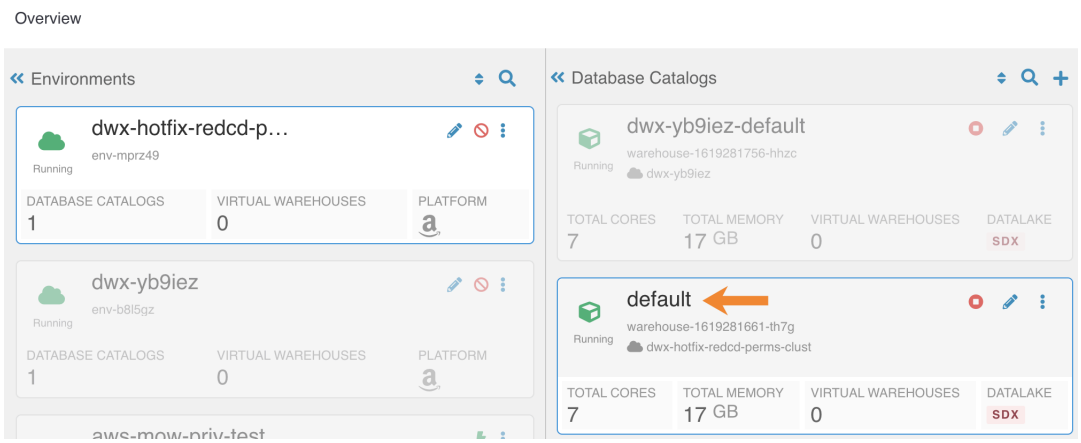
- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

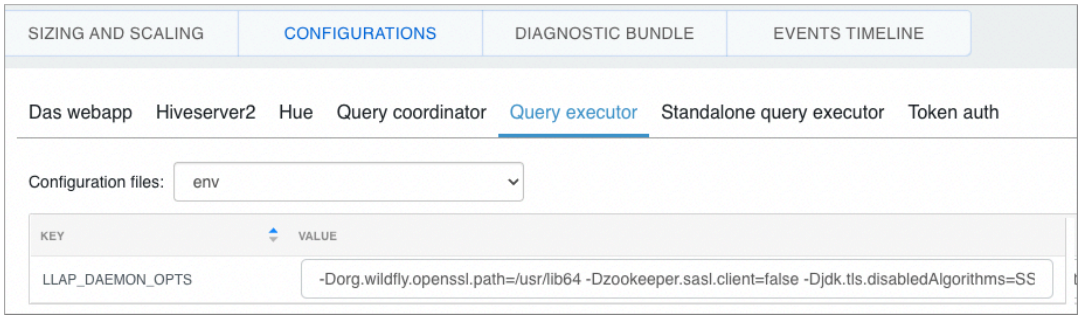
DWX-15064 Hive Virtual Warehouse stops but appears healthy

Due to an istio-proxy problem, the query coordinator can unexpectedly enter a not ready state instead of the expected error-state. Subsequently, the Hive Virtual Warehouse stops when reaching the autosuspend timeout without indicating a problem.

DWX-14452 Parquet table query might fail

Querying a table stored in Parquet from Hive might fail with the following exception message: java.lang.RuntimeException: java.lang.StackOverflowError. This problem can occur when the IN predicate has 243 values or more, and a small stack (-Xss = 256k) is configured for Hive in CDW.

Workaround: Change the value of the Xss in the JVM args to use the default value (1Mb) as follows: Select Edit from the Options menu of your Virtual Warehouse. Click Configurations > Query Executor > and select env.



In the third line shown below, change the value of LLAP_DAEMON_OPTS from -Xss256k to -Xss1M, and then click Apply Changes:

FROM:

-Dorg.wildfly.openssl.path=/usr/lib64 -Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Dhttp.maxConnections=12 -Xms24G -Xmx48G -Xss256k ...

TO:

... -Xss1M ...

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
       kind: ClusterRole
       metadata:
       name: diagnostic-data-generator-role-monitor
       rules:
       - apiGroups:
       - ""
       resources:
       - configmaps
       - events
       - pods
       - persistentvolumeclaims
       - nodes
       verbs:
       - get
       - list
       - apiGroups:
       - apps
       resources:
       - deployments
       - statefulsets
       verbs:
       - get
       - list
       - apiGroups:
       - "edws.cloudera.com"
```

```
      resources:
      - computes
      verbs:
      - get
      - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.

7. Click Apply in the upper right corner of the page to save the configuration.
8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name ‘ec2-18-219-
32-183.us-east-2.compute.amazonaws.com’ does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Hue Query Editor

Delay in listing queries in Impala Queries in the Job browser

Listing an Impala query in the Job browser can take an inordinate amount of time.

DWX-14927 Hue fails to list Iceberg snapshots

Hue does not recognize the Iceberg history queries from Hive to list table snapshots. For example, Hue indicates an error at the . before history when you run the following query.

```
select * from <db_name>.<table_name>.history
```

DWX-14968 Connection termination error in Impala queries tab after Hue inactivity

To reproduce the problem: 1) In the Impala job browser, navigate to Impala queries. 2) Wait for a few minutes.

After a few minutes of inactivity, the follow error is displayed:

Workaround: Refresh the page, or alternatively start a new session.

DWX-15115 Error displayed after clicking on hyperlink below Hue table browser

In Hue, below the table browser, clicking the hyperlink to a location causes an HTTP 500 error because the file browser is not enabled for environments that are not Ranger authorized (RAZ).

DWX-15090: CSRF error intermittently seen in the Hue Job Browser

You may intermittently see the “403 - CSRF” error on the Hue web interface as well as in the Hue logs after running Hive queries from Hue.

Reload the browser or start a new Hue session.

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.

2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Iceberg

DWX-15014 Loading airports table in demo data fails

This issue occurs only when using a non-default Database Catalog. A invalid path error message occurs when using the load command to load the demo data airports table in Iceberg, ORC, or Parquet format.

DWX-14163 Limitations reading Iceberg tables in Avro file format from Impala

The Avro, Impala, and Iceberg specifications describe some limitations related to Avro, and those limitations exist in CDP. In addition to these, the DECIMAL type is not supported in this release.

DEX-7946 Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata if you drop the table during migration from Hive to Iceberg.

Workarounds: Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13733 Timeout issue querying Iceberg tables from Hive

A timeout issue can cause the query to fail.

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

```
fs.s3.maxConnections=1000
```

```
fs.s3a.connection.maximum=1000
```

Restart the Database Catalog and Virtual Warehouse.

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: Impala Virtual Warehouse

TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

The Cloudera Data Warehouse (CDW) Public Cloud 2023.0.14.0 (DWX-1.6.3) version incorporated a feature for performing automatic metadata synchronization across multiple Apache Impala (Impala) Virtual Warehouses. The feature is enabled by default, and relies on the Hive MetaStore events. When a certain sequence of Data Definition Language (DDL) SQL commands are executed as described below, users may encounter a java.lang.NullPointerException (NPE). The exception causes the event processor to stop processing other metadata operations.

If a CREATE TABLE command (not CREATE TABLE AS SELECT) is followed immediately (approximately within 1 second interval) by INVALIDATE METADATA or REFRESH TABLE command on the same table (either on the same Virtual Warehouse or on a different one), there is a possibility that the second command will not find the table in the catalog cache of a peer Virtual Warehouse and generate an NPE.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

MT_DOP related issue

Problem:Impala query with JOIN and LIMIT can hang when MT_DOP is greater than 0.

Workaround: Set MT_DOP to zero.

DWX-15016 Impala query failure when using Catalog high availability (HA)

Impala queries could fail with InconsistentMetadataFetchException when using HA feature in CDW. This happens when leader election failover for Impala Catalog Service happens due to a transient network error.

Workaround: Disable Catalog HA. In CDW, edit your Virtual Warehouse, and turn off Enable Impala Catalog Server HA.

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml
```

```
        # Make sure that the annotation you added is present.
```

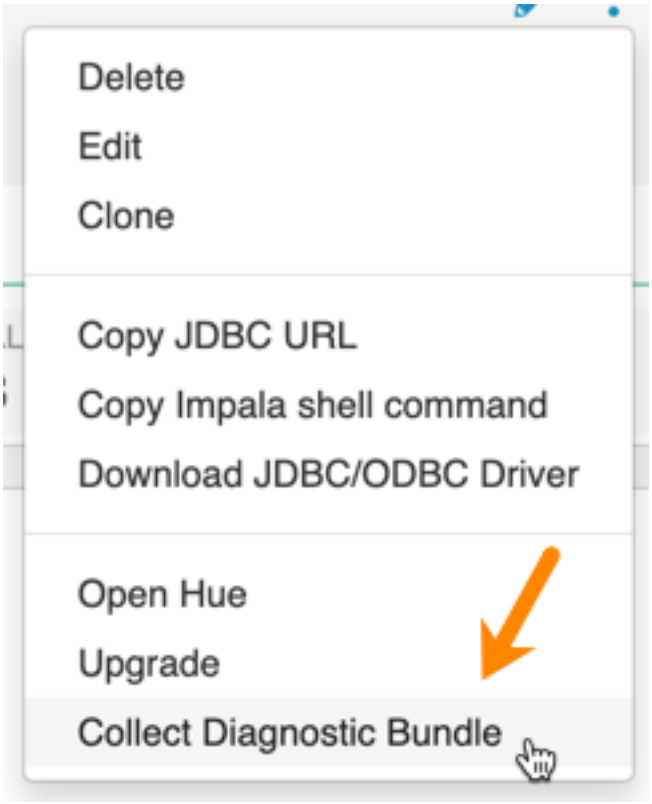
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

UI Issues

DWX-14610 Upgrade icon indicates the Database Catalog is the latest version, but might be incorrect for a few seconds

The cluster fetches data regularly in particular time intervals, or when an action is triggered. Inbetween this time interval, which is very brief, the cluster might not be updated, but will be refreshed in a few seconds.

Technical Service Bulletins

TSB 2024-723: Hue RAZ is using logger role to Read and Upload/Delete (write) files

When using Cloudera Data Hub for Public Cloud (Data Hub) on Amazon Web Services (AWS), users can use the Hue File Browser feature to access the filesystem, and if permitted, read and write

directly to the related S3 buckets. As AWS does not provide fine-grained access control, Cloudera Data Platform administrators can use the Ranger Authorization Service (RAZ) capability to take the S3 filesystem, and overlay it with user and group specific permissions, making it easier to allow certain users to have limited permissions, without having to grant those users permissions to the entire S3 bucket.

This bulletin describes an issue when using RAZ with Data Hub, and attempting to use fine-grained access control to allow certain users write permissions.

Through RAZ, an administrator may, for a particular user, specify permissions more limited than what AWS provides for an S3 bucket, allowing the user to have read/write (or other similar fine grained access) permissions on only a subset of the files and directories within that bucket. However, under specific conditions, it is possible for such user to be able to read and write to the entire S3 bucket through Hue, due to Hue using the logger role (which will have full read/write to the S3 bucket) when using Data Hub with a RAZ enabled cluster. This problem also can affect the Hue service itself, by affecting proper access to home directories causing the service role to not start.

The root cause of this issue is, when accessing Amazon cloud resources, Hue uses the AWS Boto SDK library. This AWS Boto library has a bug that restricts permissions in certain AWS regions in such a way that it provides access to users who should not have it, regardless of RAZ settings. This issue only affects users in specific AWS regions, listed below, and it does not affect all AWS customers.

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-723: Hue Raz is using logger role to Read and Upload/Delete (write) files.

TSB 2024-743: Restoration of the Database Catalog can be incomplete due to an incorrect Hue Query Processor configuration in CDW

Cloudera Data Warehouse (CDW) customers using the CDW Automated Backup/Restore feature might encounter an issue with the restoration of the Database Catalog, if the hue-query-processor.json configuration file of the Database Catalog has been edited. Even a minor edit to the hue-query-processor.json configuration file can result in this failure.

During the restoration process the Database Catalog will be created, but it will fail to start after a short period of time, and the Database Catalog will be in a Bad Health state on the CDW User Interface.

Inside the Kubernetes Cluster (Azure Kubernetes Service on Azure / Elastic Kubernetes Service on AWS) the StatefulSet of Hue Query Processor (hue-query-processor) is in CrashLoop. This is indicated with the following log in the Hue Query Processor StatefulSet Pod:

```
SQL State  : 3D000
Error Code : 0
Message    : FATAL: database "warehouse-1707832123-abcd_hueqpdb"
 does not exist



             at org.flywaydb.core.internal.jdbc.JdbcUtils.openC
onnection(JdbcUtils.java:65)
```

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-743: Restoration of the Database Catalog can be incomplete due to an incorrect Hue Query Processor configuration in CDW.

TSB 2024-745: Impala returns incorrect results for Iceberg V2 tables when optimized operator is being used in CDW

Cloudera Data Warehouse (CDW) customers using Apache Impala (Impala) to read Apache Iceberg (Iceberg) V2 tables can encounter an issue of Impala returning incorrect results when the optimized V2 operator is used. The optimized V2 operator is enabled by default in the affected versions below. The issue only affects Iceberg V2 tables that have position delete files.

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-745: Impala returns incorrect results for Iceberg V2 tables when optimized operator is being used in CDW.

TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables

Apache Hive (Hive) and Apache Impala (Impala) modify statements (DELETE/UPDATE/MERGE) on Apache Iceberg (Iceberg) V2 tables can corrupt the tables if there is a concurrent table compaction from Apache Spark. The issue happens when the compaction and modify statement run in parallel, and when the compaction job commits before the modify statement. In this case the position delete files of the modify statement still point to the old files. This means the following in case of

- DELETE statements
    - Deleting records pointing to old files have no effect
- UPDATE / MERGE statements
    - Deleting records pointing to old files have no effect
    - The table will also have the newly added data records
    - Rewritten records will still be active

This issue does not affect Apache NiFi (NiFi) and Apache Flink (Flink) as these components write equality delete files.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables.

TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

The Spark Iceberg library includes two procedures - rewrite_data_files and rewrite_position_delete_files. The current implementation of rewrite_data_files has a limitation that the position delete files are not deleted and still tracked by the table metadata, even if they no longer refer to an active data file. This is called the dangling delete problem. To solve this, the rewrite_position_delete_files procedure is implemented in the Spark Iceberg library to remove these old “dangling” position delete files.

Due to the dangling delete limitation, when an Iceberg table with dangling deletes is queried in Impala, Impala tries to optimize select count(*) from iceberg_table query to return the results using stats. This optimization returns incorrect results.

The following conditions must be met for this issue to occur:

- All delete files in the Iceberg table are “dangling”
    - This would occur immediately after running Spark rewrite_data_files AND
        - Before any further delete operations are performed on the table OR
        - Before Spark rewrite_position_delete_files is run on the table

- Only stats optimized plain select count(*) from iceberg_table queries are affected. For example, the query should not have:
  - Any WHERE clause
  - Any GROUP BY clause
  - Any HAVING clause

For more information, see the Apache Iceberg documentation.

Remove dangling deletes: After rewrite_data_files, position delete records pointing to the rewritten data files are not always marked for removal, and can remain tracked by the live snapshot metadata of the table. This is known as the dangling delete problem.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables.

October 12, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

New Known Issues in this release

See Data Visualization release notes for known issues in Cloudera Data Visualization 7.1.3.

Carried over from the previous release: Upgrade-related

TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

The Spark Iceberg library includes two procedures - rewrite_data_files and rewrite_position_delete_files. The current implementation of rewrite_data_files has a limitation that the position delete files are not deleted and still tracked by the table metadata, even if they no longer refer to an active data file. This is called the dangling delete problem. To solve this, the rewrite_position_delete_files procedure is implemented in the Spark Iceberg library to remove these old “dangling” position delete files.

Due to the dangling delete limitation, when an Iceberg table with dangling deletes is queried in Impala, Impala tries to optimize select count(*) from iceberg_table query to return the results using stats. This optimization returns incorrect results.

The following conditions must be met for this issue to occur:

- All delete files in the Iceberg table are “dangling”
  - This would occur immediately after running Spark rewrite_data_files AND
    - Before any further delete operations are performed on the table OR
    - Before Spark rewrite_position_delete_files is run on the table
- Only stats optimized plain select count(*) from iceberg_table queries are affected. For example, the query should not have:
  - Any WHERE clause
  - Any GROUP BY clause
  - Any HAVING clause

For more information, see the Apache Iceberg documentation.

Remove dangling deletes: After rewrite_data_files, position delete records pointing to the rewritten data files are not always marked for removal, and can remain tracked by the live snapshot metadata of the table. This is known as the dangling delete problem.

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

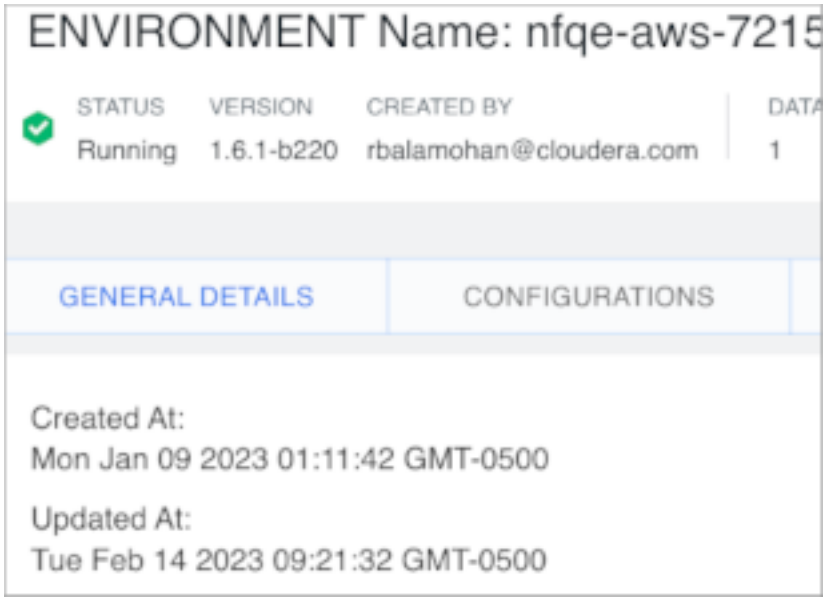
Upgrading to EKS 1.24 could result in Impala coordinators shutting down.

This issue is not seen on the Impala Virtual Warehouse running Runtime 2023.0.15.0-x or later.

Workaround: Manually start the Impala Virtual Warehouse from the UI or cli. Alternatively, replace your runtime with 1.7.1-b755 (released August 30, 2023) or later.

Certain CDW versions cannot upgrade to EKS 1.22

AWS environments activated in version 1.4.1-b86 (released June, 22, 2022) or earlier are not supported for upgrade to EKS 1.22 due to incompatibility of some components with EKS 1.22. To determine the activation version your pre-existing environment, in the Data Warehouse service, expand Environments. In Environments, search for and locate the environment that you want to view. Click Edit. In Environment Details, you see the CDW version.



DWX-15112 Enterprise Data Warehouse database configuration problems after a Helm-related rollback

If a Helm rollback fails due to an incorrect Enterprise Data Warehouse database configuration, the Virtual Warehouse and Database Catalog roll back to a previous configuration. The incorrect Enterprise Data Warehouse configuration persists, and can affect subsequent edit, upgrade, and rebuild operations on the rolled-back Virtual Warehouse or Database Catalog.

Carried over from the previous release: General

Compaction causes data loss under certain conditions

Data loss occurs during compaction when Apache Ranger policies for masking or row filtering are enabled and compaction users are included in the policies.

Workaround: Exclude compaction users from the policies as described in Compaction Prerequisites.

DWX-14923 After JWT authentication, attempting to connect the Impyla client using a user name or password should cause an error

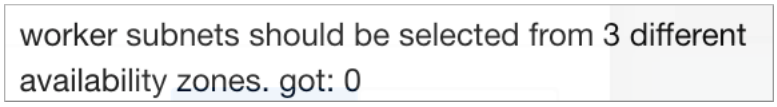
Using a JWT token, you can connect to a Virtual Warehouse as the user who generated the token.

If you connect with the JWT token, and then pass a user name or password from Impyla to the Virtual Warehouse, the connection arguments are silently ignored. Such an action should indicate that it is illegal to specify user or password when using JWT authentication.

DWX-15145 Environment validation popup error after activating an environment

Activating an environment having a public load balancer can cause an environment validatation popup error.

To reproduce this problem 1) Create a data lake. 2) Activate an enviroment having a public load balancer deployment type and subnets in three different availability zones. A environment validation popup can occur even through subnets are in different availability zones. Several differnt popups can occur, including the following one:



DWX-15144 Virtual Warehouse naming restrictions

You cannot create a Virtual Warehouse having the same name as another Virtual Warehouse even if the like-named Virtual Warehouses are in different environments. You can create a Database Catalog having the same name as another Database Catalog if the Database Catalogs are in different environments.

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
```

```
        Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the Cr eateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersio ns line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

Enabling a private CDW environment in Azure Kubernetes Service is unstable

Cloudera rolled back the General Availability (GA) status and support for deploying a private CDW environment in Azure using AKS until a solution for the issue is available.

When working together with Microsoft Azure (Azure), a networking issue has been identified at Azure Software Definition Layer (SDN) that impacts the use of CDW private environments while using the Azure Kubernetes Service (AKS). This issue is highly unpredictable and may cause timeouts during Cloudera Data Warehouse (CDW) environment activation, Virtual Warehouse creation, Virtual Warehouse modification, and/or during start and stop operations in CDW.

The following users are affected: CDW users on Azure who are activating a private CDW environment by selecting the Private CDW option on the User Interface (UI) or using CDP Command Line Interface (CLI) enablePrivateAks option for the dw sub-command create-cluster operation within the --aws-options group.

Activation can fail with a timeout issue, and other operation related actions may be interrupted.

DWX-15214 DWX-15176 Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

The Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed:

```
kubelet Error: failed to create containerd task: failed to creat
e shim task: OCI runtime create failed: runc create failed: unab
le to start container process: exec: “run_httpd.sh”: cannot run
executable found relative to current directory: unknown
```

This issue affects all CDW releases before 1.6.3 (released May 5, 2023)running on CDP Public Cloud on Azure.

The following users are affected: CDW Azure users using Hue in Impala and Hive Virtual Warehouses in environments earlier than version 1.6.3 or later if they upgrade the node image

Workaround: Upgrade the Virtual Warehouses to 1.6.3 and choose Hue version 2023.0.14.0.

Do not choose Hue versions older than 2023.0.14.0 when creating a Virtual Warehouse in environments of version 1.6.3.

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses

old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Database Catalog

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

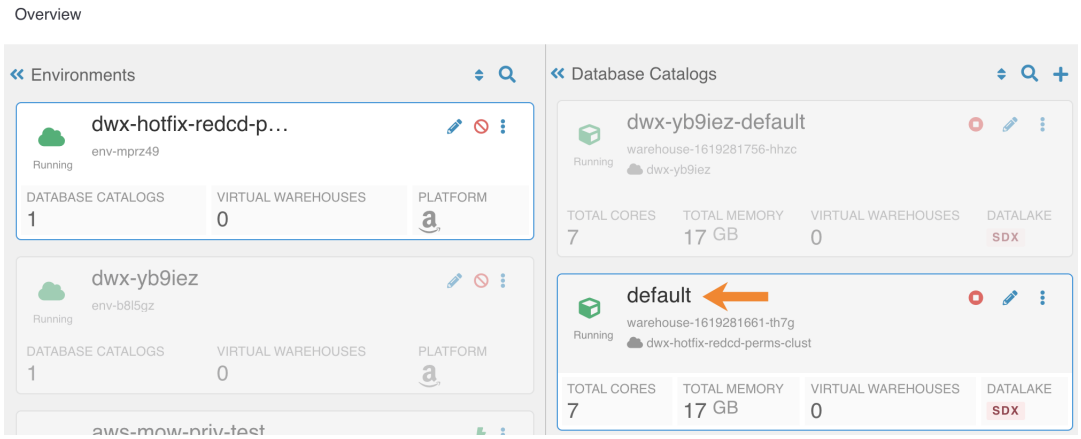
- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

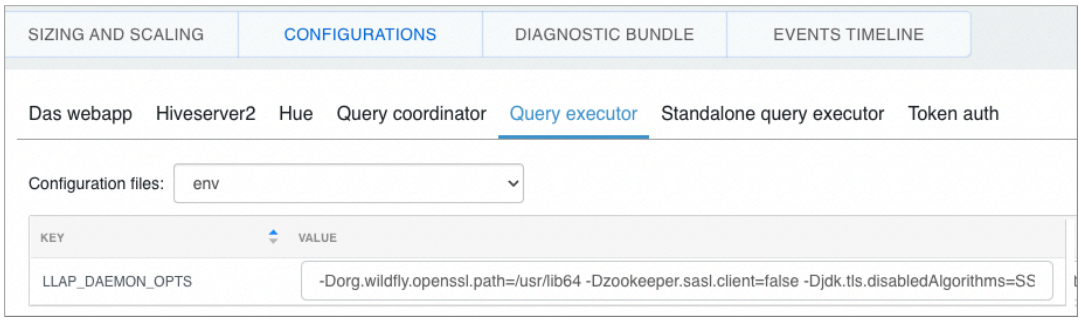
DWX-15064 Hive Virtual Warehouse stops but appears healthy

Due to an istio-proxy problem, the query coordinator can unexpectedly enter a not ready state instead of the expected error-state. Subsequently, the Hive Virtual Warehouse stops when reaching the autosuspend timeout without indicating a problem.

DWX-14452 Parquet table query might fail

Querying a table stored in Parquet from Hive might fail with the following exception message: java.lang.RuntimeException: java.lang.StackOverflowError. This problem can occur when the IN predicate has 243 values or more, and a small stack (-Xss = 256k) is configured for Hive in CDW.

Workaround: Change the value of the Xss in the JVM args to use the default value (1Mb) as follows: Select Edit from the Options menu of your Virtual Warehouse. Click Configurations > Query Executor > and select env.



In the third line shown below, change the value of LLAP_DAEMON_OPTS from -Xss256k to -Xss1M, and then click Apply Changes:

FROM:

-Dorg.wildfly.openssl.path=/usr/lib64 -Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Dhttp.maxConnections=12 -Xms24G -Xmx48G -Xss256k ...

TO:

... -Xss1M ...

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
       kind: ClusterRole
       metadata:
       name: diagnostic-data-generator-role-monitor
       rules:
       - apiGroups:
       - ""
       resources:
       - configmaps
       - events
       - pods
       - persistentvolumeclaims
       - nodes
       verbs:
```

```
        - get
        - list
        - apiGroups:
        - apps
        resources:
        - deployments
        - statefulsets
        verbs:
        - get
        - list
        - apiGroups:
        - "edws.cloudera.com"
        resources:
        - computes
        verbs:
        - get
        - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.

7. Click Apply in the upper right corner of the page to save the configuration.

8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name ‘ec2-18-219-
32-183.us-east-2.compute.amazonaws.com’ does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Hue Query Editor

Delay in listing queries in Impala Queries in the Job browser

Listing an Impala query in the Job browser can take an inordinate amount of time.

DWX-14927 Hue fails to list Iceberg snapshots

Hue does not recognize the Iceberg history queries from Hive to list table snapshots. For example, Hue indicates an error at the . before history when you run the following query.

```
select * from <db_name>.<table_name>.history
```

DWX-14968 Connection termination error in Impala queries tab after Hue inactivity

To reproduce the problem: 1) In the Impala job browser, navigate to Impala queries. 2) Wait for a few minutes.

After a few minutes of inactivity, the follow error is displayed:

Workaround: Refresh the page, or alternatively start a new session.

DWX-15115 Error displayed after clicking on hyperlink below Hue table browser

In Hue, below the table browser, clicking the hyperlink to a location causes an HTTP 500 error because the file browser is not enabled for environments that are not Ranger authorized (RAZ).

DWX-15090: CSRF error intermittently seen in the Hue Job Browser

You may intermittently see the “403 - CSRF” error on the Hue web interface as well as in the Hue logs after running Hive queries from Hue.

Reload the browser or start a new Hue session.

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Iceberg

DWX-15014 Loading airports table in demo data fails

This issue occurs only when using a non-default Database Catalog. A invalid path error message occurs when using the load command to load the demo data airports table in Iceberg, ORC, or Parquet format.

DWX-14163 Limitations reading Iceberg tables in Avro file format from Impala

The Avro, Impala, and Iceberg specifications describe some limitations related to Avro, and those limitations exist in CDP. In addition to these, the DECIMAL type is not supported in this release.

DEX-7946 Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata if you drop the table during migration from Hive to Iceberg.

Workarounds: Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13733 Timeout issue querying Iceberg tables from Hive

A timeout issue can cause the query to fail.

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

```
fs.s3.maxConnections=1000
fs.s3a.connection.maximum=1000
```

Restart the Database Catalog and Virtual Warehouse.

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: Impala Virtual Warehouse

TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

The Cloudera Data Warehouse (CDW) Public Cloud 2023.0.14.0 (DWX-1.6.3) version incorporated a feature for performing automatic metadata synchronization across multiple Apache Impala (Impala) Virtual Warehouses. The feature is enabled by default, and relies on the Hive MetaStore events. When a certain sequence of Data Definition Language (DDL) SQL commands are executed as described below, users may encounter a java.lang.NullPointerException (NPE). The exception causes the event processor to stop processing other metadata operations.

If a CREATE TABLE command (not CREATE TABLE AS SELECT) is followed immediately (approximately within 1 second interval) by INVALIDATE METADATA or REFRESH TABLE command on the same table (either on the same Virtual Warehouse or on a different one), there is

a possibility that the second command will not find the table in the catalog cache of a peer Virtual Warehouse and generate an NPE.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

MT_DOP related issue

Problem:Impala query with JOIN and LIMIT can hang when MT_DOP is greater than 0.

Workaround: Set MT_DOP to zero.

DWX-15016 Impala query failure when using Catalog high availability (HA)

Impala queries could fail with InconsistentMetadataFetchException when using HA feature in CDW. This happens when leader election failover for Impala Catalog Service happens due to a transient network error.

Workaround: Disable Catalog HA. In CDW, edit your Virtual Warehouse, and turn off Enable Impala Catalog Server HA.

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

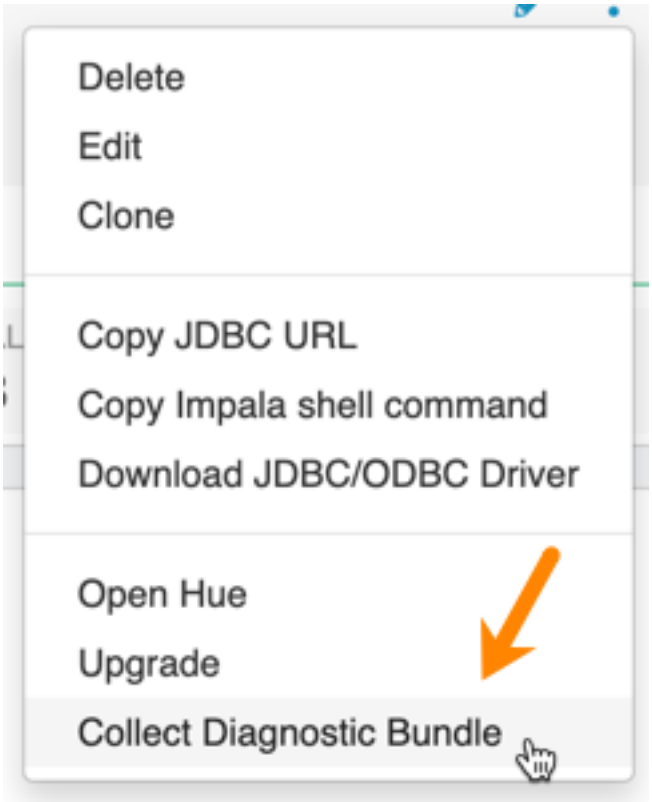
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

UI Issues

DWX-14610 Upgrade icon indicates the Database Catalog is the latest version, but might be incorrect for a few seconds

The cluster fetches data regularly in particular time intervals, or when an action is triggered. Inbetween this time interval, which is very brief, the cluster might not be updated, but will be refreshed in a few seconds.

Technical Service Bulletins

TSB 2024-723: Hue RAZ is using logger role to Read and Upload/Delete (write) files

When using Cloudera Data Hub for Public Cloud (Data Hub) on Amazon Web Services (AWS), users can use the Hue File Browser feature to access the filesystem, and if permitted, read and write directly to the related S3 buckets. As AWS does not provide fine-grained access control, Cloudera Data Platform administrators can use the Ranger Authorization Service (RAZ) capability to take the S3 filesystem, and overlay it with user and group specific permissions, making it easier to allow certain users to have limited permissions, without having to grant those users permissions to the entire S3 bucket.

This bulletin describes an issue when using RAZ with Data Hub, and attempting to use fine-grained access control to allow certain users write permissions.

Through RAZ, an administrator may, for a particular user, specify permissions more limited than what AWS provides for an S3 bucket, allowing the user to have read/write (or other similar fine grained access) permissions on only a subset of the files and directories within that bucket. However, under specific conditions, it is possible for such user to be able to read and write to the entire S3 bucket through Hue, due to Hue using the logger role (which will have full read/write to the S3 bucket) when using Data Hub with a RAZ enabled cluster. This problem also can affect the Hue service itself, by affecting proper access to home directories causing the service role to not start.

The root cause of this issue is, when accessing Amazon cloud resources, Hue uses the AWS Boto SDK library. This AWS Boto library has a bug that restricts permissions in certain AWS regions in such a way that it provides access to users who should not have it, regardless of RAZ settings. This issue only affects users in specific AWS regions, listed below, and it does not affect all AWS customers.

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-723: Hue Raz is using logger role to Read and Upload/Delete (write) files.

TSB 2024-743: Restoration of the Database Catalog can be incomplete due to an incorrect Hue Query Processor configuration in CDW

Cloudera Data Warehouse (CDW) customers using the CDW Automated Backup/Restore feature might encounter an issue with the restoration of the Database Catalog, if the hue-query-processor.json configuration file of the Database Catalog has been edited. Even a minor edit to the hue-query-processor.json configuration file can result in this failure.

During the restoration process the Database Catalog will be created, but it will fail to start after a short period of time, and the Database Catalog will be in a Bad Health state on the CDW User Interface.

Inside the Kubernetes Cluster (Azure Kubernetes Service on Azure / Elastic Kubernetes Service on AWS) the StatefulSet of Hue Query Processor (hue-query-processor) is in CrashLoop. This is indicated with the following log in the Hue Query Processor StatefulSet Pod:

```
SQL State  : 3D000
Error Code : 0
Message    : FATAL: database "warehouse-1707832123-abcd_hueqpdb"
 does not exist



              at org.flywaydb.core.internal.jdbc.JdbcUtils.openC
onnection(JdbcUtils.java:65)
```

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-743: Restoration of the Database Catalog can be incomplete due to an incorrect Hue Query Processor configuration in CDW.

TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables

Apache Hive (Hive) and Apache Impala (Impala) modify statements (DELETE/UPDATE/MERGE) on Apache Iceberg (Iceberg) V2 tables can corrupt the tables if there is a concurrent table compaction from Apache Spark. The issue happens when the compaction and modify statement run in parallel, and when the compaction job commits before the modify statement. In this case the position delete files of the modify statement still point to the old files. This means the following in case of

- DELETE statements
  - Deleting records pointing to old files have no effect
- UPDATE / MERGE statements
  - Deleting records pointing to old files have no effect
  - The table will also have the newly added data records
  - Rewritten records will still be active

This issue does not affect Apache NiFi (NiFi) and Apache Flink (Flink) as these components write equality delete files.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables.

TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion

When working with Apache Hive (Hive) and Apache Iceberg (Iceberg) V2 tables, using the TRUNCATE statement may lead to unintended data deletion. This issue arises when the truncate command is applied to a branch of an Iceberg table. Instead of truncating the branch itself, the command affects the original (main) table, which results in unintended loss of data.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion

October 5, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

New Known Issues in this release

There are no new known issues in this release for CDW. See Data Visualization release notes for known issues in Cloudera Data Visualization 7.1.3.

Carried over from the previous release: Upgrade-related

Upgrading to EKS 1.24 could result in Impala coordinators shutting down.

This issue is not seen on the Impala Virtual Warehouse running Runtime 2023.0.15.0-x or later.

Workaround: Manually start the Impala Virtual Warehouse from the UI or cli. Alternatively, replace your runtime with 1.7.1-b755 (released August 30, 2023) or later.

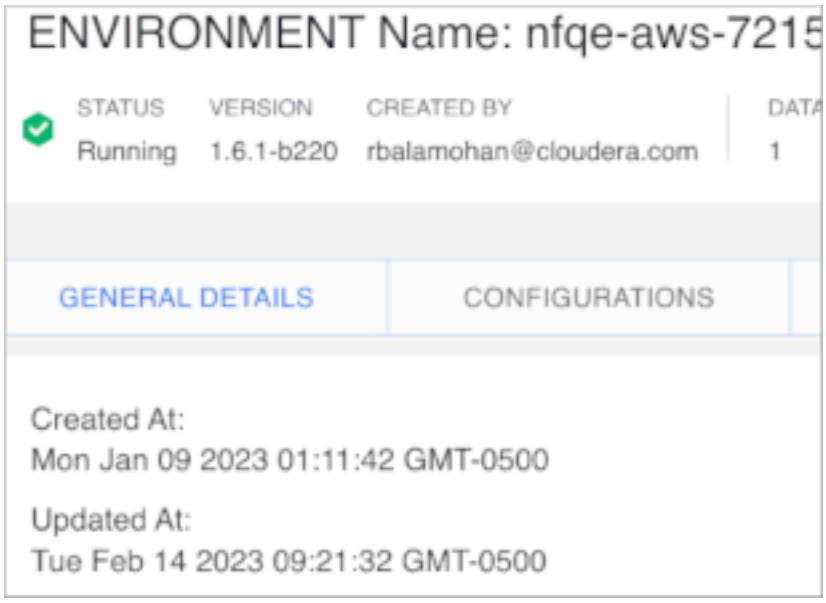
Issue backing up and restoring Workload Aware Auto-Scaling (WAAS)

You cannot use the CDW backup and restore process to restore the WAAS configurations.

Workaround: Recreate your Virtual Warehouse manually setting the right configurations.

Certain CDW versions cannot upgrade to EKS 1.22

AWS environments activated in version 1.4.1-b86 (released June, 22, 2022) or earlier are not supported for upgrade to EKS 1.22 due to incompatibility of some components with EKS 1.22. To determine the activation version your pre-existing environment, in the Data Warehouse service, expand Environments. In Environments, search for and locate the environment that you want to view. Click Edit. In Environment Details, you see the CDW version.



DWX-15112 Enterprise Data Warehouse database configuration problems after a Helm-related rollback

If a Helm rollback fails due to an incorrect Enterprise Data Warehouse database configuration, the Virtual Warehouse and Database Catalog roll back to a previous configuration. The incorrect Enterprise Data Warehouse configuration persists, and can affect subsequent edit, upgrade, and rebuild operations on the rolled-back Virtual Warehouse or Database Catalog.

Carried over from the previous release: General

Compaction causes data loss under certain conditions

Data loss occurs during compaction when Apache Ranger policies for masking or row filtering are enabled and compaction users are included in the policies.

Workaround: Exclude compaction users from the policies as described in Compaction Prerequisites.

DWX-14923 After JWT authentication, attempting to connect the Impyla client using a user name or password should cause an error

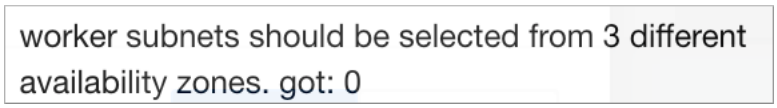
Using a JWT token, you can connect to a Virtual Warehouse as the user who generated the token.

If you connect with the JWT token, and then pass a user name or password from Impyla to the Virtual Warehouse, the connection arguments are silently ignored. Such an action should indicate that it is illegal to specify user or password when using JWT authentication.

DWX-15145 Environment validation popup error after activating an environment

Activating an environment having a public load balancer can cause an environment validatation popup error.

To reproduce this problem 1) Create a data lake. 2) Activate an enviroment having a public load balancer deployment type and subnets in three different availability zones. A environment validation popup can occur even through subnets are in different availability zones. Several differnt popups can occur, including the following one:



DWX-15144 Virtual Warehouse naming restrictions

You cannot create a Virtual Warehouse having the same name as another Virtual Warehouse even if the like-named Virtual Warehouses are in different environments. You can create a Database Catalog having the same name as another Database Catalog if the Database Catalogs are in different environments.

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
      operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the CreateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersions line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

Enabling a private CDW environment in Azure Kubernetes Service is unstable

Cloudera rolled back the General Availability (GA) status and support for deploying a private CDW environment in Azure using AKS until a solution for the issue is available.

When working together with Microsoft Azure (Azure), a networking issue has been identified at Azure Software Definition Layer (SDN) that impacts the use of CDW private environments while using the Azure Kubernetes Service (AKS). This issue is highly unpredictable and may cause timeouts during Cloudera Data Warehouse (CDW) environment activation, Virtual Warehouse creation, Virtual Warehouse modification, and/or during start and stop operations in CDW.

The following users are affected: CDW users on Azure who are activating a private CDW environment by selecting the Private CDW option on the User Interface (UI) or using CDP Command Line Interface (CLI) enablePrivateAks option for the dw sub-command create-cluster operation within the --aws-options group.

Activation can fail with a timeout issue, and other operation related actions may be interrupted.

DWX-15214 DWX-15176 Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

The Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed:

```
kubelet Error: failed to create containerd task: failed to creat
e shim task: OCI runtime create failed: runc create failed: unab
le to start container process: exec: “run_httpd.sh”: cannot run
executable found relative to current directory: unknown
```

This issue affects all CDW releases before 1.6.3 (released May 5, 2023)running on CDP Public Cloud on Azure.

The following users are affected: CDW Azure users using Hue in Impala and Hive Virtual Warehouses in environments earlier than version 1.6.3 or later if they upgrade the node image

Workaround: Upgrade the Virtual Warehouses to 1.6.3 and choose Hue version 2023.0.14.0.

Do not choose Hue versions older than 2023.0.14.0 when creating a Virtual Warehouse in environments of version 1.6.3.

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Database Catalog

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

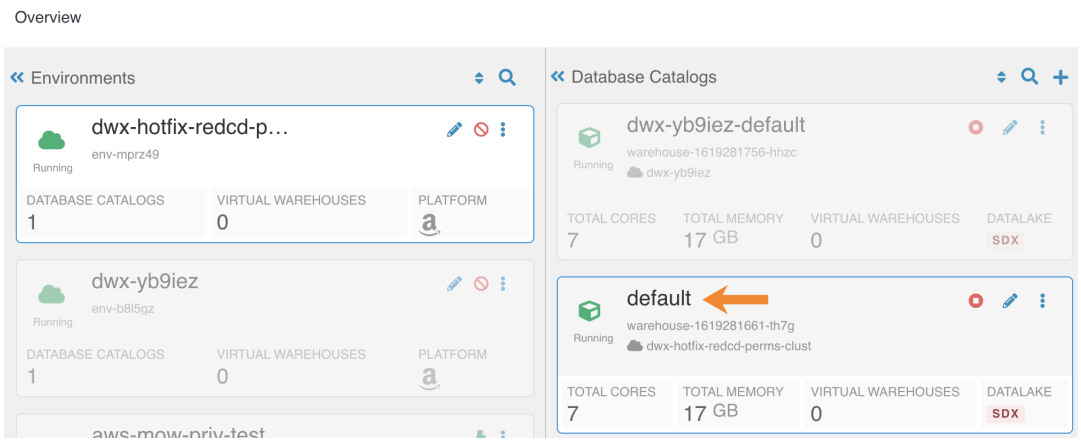
- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

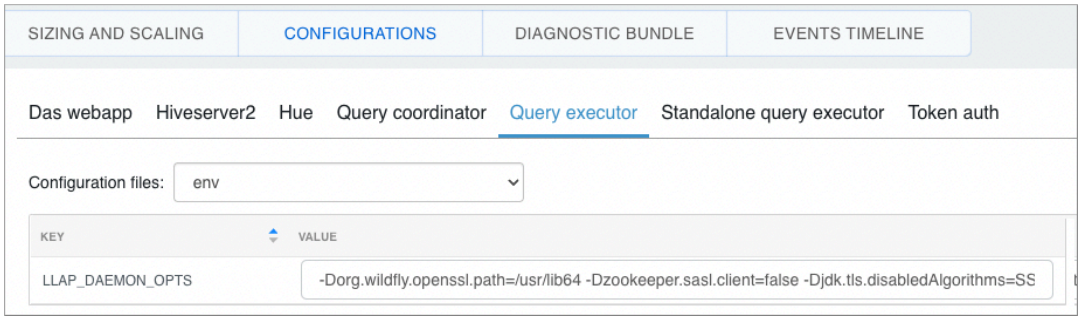
DWX-15064 Hive Virtual Warehouse stops but appears healthy

Due to an istio-proxy problem, the query coordinator can unexpectedly enter a not ready state instead of the expected error-state. Subsequently, the Hive Virtual Warehouse stops when reaching the autosuspend timeout without indicating a problem.

DWX-14452 Parquet table query might fail

Querying a table stored in Parquet from Hive might fail with the following exception message: java.lang.RuntimeException: java.lang.StackOverflowError. This problem can occur when the IN predicate has 243 values or more, and a small stack (-Xss = 256k) is configured for Hive in CDW.

Workaround: Change the value of the Xss in the JVM args to use the default value (1Mb) as follows: Select Edit from the Options menu of your Virtual Warehouse. Click Configurations > Query Executor > and select env.



In the third line shown below, change the value of LLAP_DAEMON_OPTS from -Xss256k to -Xss1M, and then click Apply Changes:

FROM:

-Dorg.wildfly.openssl.path=/usr/lib64 -Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Dhttp.maxConnections=12 -Xms24G -Xmx48G -Xss256k ...

TO:

... -Xss1M ...

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
       kind: ClusterRole
       metadata:
       name: diagnostic-data-generator-role-monitor
       rules:
       - apiGroups:
       - ""
       resources:
       - configmaps
       - events
```

```
      - pods
      - persistentvolumeclaims
      - nodes
      verbs:
      - get
      - list
      - apiGroups:
      - apps
      resources:
      - deployments
      - statefulsets
      verbs:
      - get
      - list
      - apiGroups:
      - "edws.cloudera.com"
      resources:
      - computes
      verbs:
      - get
      - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.

7. Click Apply in the upper right corner of the page to save the configuration.

8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name ‘ec2-18-219-
32-183.us-east-2.compute.amazonaws.com’ does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Hue Query Editor

Delay in listing queries in Impala Queries in the Job browser

Listing an Impala query in the Job browser can take an inordinate amount of time.

DWX-14927 Hue fails to list Iceberg snapshots

Hue does not recognize the Iceberg history queries from Hive to list table snapshots. For example, Hue indicates an error at the . before history when you run the following query.

```
select * from <db_name>.<table_name>.history
```

DWX-14968 Connection termination error in Impala queries tab after Hue inactivity

To reproduce the problem: 1) In the Impala job browser, navigate to Impala queries. 2) Wait for a few minutes.

After a few minutes of inactivity, the follow error is displayed:

Workaround: Refresh the page, or alternatively start a new session.

DWX-15115 Error displayed after clicking on hyperlink below Hue table browser

In Hue, below the table browser, clicking the hyperlink to a location causes an HTTP 500 error because the file browser is not enabled for environments that are not Ranger authorized (RAZ).

DWX-15090: CSRF error intermittently seen in the Hue Job Browser

You may intermittently see the “403 - CSRF” error on the Hue web interface as well as in the Hue logs after running Hive queries from Hue.

Reload the browser or start a new Hue session.

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Iceberg

DWX-15014 Loading airports table in demo data fails

This issue occurs only when using a non-default Database Catalog. A invalid path error message occurs when using the load command to load the demo data airports table in Iceberg, ORC, or Parquet format.

DWX-14163 Limitations reading Iceberg tables in Avro file format from Impala

The Avro, Impala, and Iceberg specifications describe some limitations related to Avro, and those limitations exist in CDP. In addition to these, the DECIMAL type is not supported in this release.

DEX-7946 Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata if you drop the table during migration from Hive to Iceberg.

Workarounds: Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13733 Timeout issue querying Iceberg tables from Hive

A timeout issue can cause the query to fail.

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

```
fs.s3.maxConnections=1000
fs.s3a.connection.maximum=1000
```

Restart the Database Catalog and Virtual Warehouse.

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: Impala Virtual Warehouse

TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

The Cloudera Data Warehouse (CDW) Public Cloud 2023.0.14.0 (DWX-1.6.3) version incorporated a feature for performing automatic metadata synchronization across multiple Apache Impala (Impala) Virtual Warehouses. The feature is enabled by default, and relies on the Hive MetaStore events. When a certain sequence of Data Definition Language (DDL) SQL commands are executed as described below, users may encounter a java.lang.NullPointerException (NPE). The exception causes the event processor to stop processing other metadata operations.

If a CREATE TABLE command (not CREATE TABLE AS SELECT) is followed immediately (approximately within 1 second interval) by INVALIDATE METADATA or REFRESH TABLE command on the same table (either on the same Virtual Warehouse or on a different one), there is

a possibility that the second command will not find the table in the catalog cache of a peer Virtual Warehouse and generate an NPE.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

MT_DOP related issue

Problem:Impala query with JOIN and LIMIT can hang when MT_DOP is greater than 0.

Workaround: Set MT_DOP to zero.

DWX-15016 Impala query failure when using Catalog high availability (HA)

Impala queries could fail with InconsistentMetadataFetchException when using HA feature in CDW. This happens when leader election failover for Impala Catalog Service happens due to a transient network error.

Workaround: Disable Catalog HA. In CDW, edit your Virtual Warehouse, and turn off Enable Impala Catalog Server HA.

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

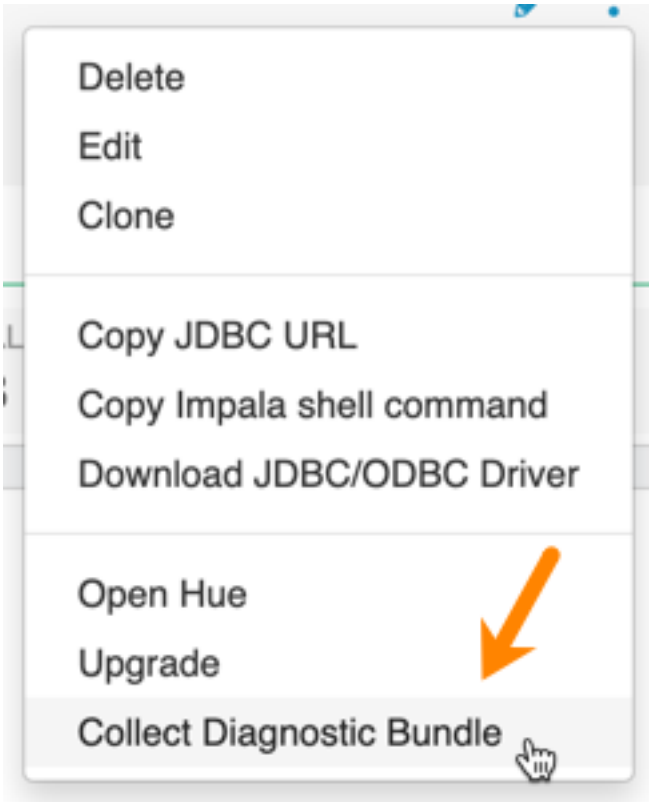
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

UI Issues

DWX-14610 Upgrade icon indicates the Database Catalog is the latest version, but might be incorrect for a few seconds

The cluster fetches data regularly in particular time intervals, or when an action is triggered. Inbetween this time interval, which is very brief, the cluster might not be updated, but will be refreshed in a few seconds.

Technical Service Bulletins

TSB 2024-723: Hue RAZ is using logger role to Read and Upload/Delete (write) files

When using Cloudera Data Hub for Public Cloud (Data Hub) on Amazon Web Services (AWS), users can use the Hue File Browser feature to access the filesystem, and if permitted, read and write directly to the related S3 buckets. As AWS does not provide fine-grained access control, Cloudera Data Platform administrators can use the Ranger Authorization Service (RAZ) capability to take the S3 filesystem, and overlay it with user and group specific permissions, making it easier to allow certain users to have limited permissions, without having to grant those users permissions to the entire S3 bucket.

This bulletin describes an issue when using RAZ with Data Hub, and attempting to use fine-grained access control to allow certain users write permissions.

Through RAZ, an administrator may, for a particular user, specify permissions more limited than what AWS provides for an S3 bucket, allowing the user to have read/write (or other similar fine grained access) permissions on only a subset of the files and directories within that bucket. However, under specific conditions, it is possible for such user to be able to read and write to the entire S3 bucket through Hue, due to Hue using the logger role (which will have full read/write to the S3 bucket) when using Data Hub with a RAZ enabled cluster. This problem also can affect the Hue service itself, by affecting proper access to home directories causing the service role to not start.

The root cause of this issue is, when accessing Amazon cloud resources, Hue uses the AWS Boto SDK library. This AWS Boto library has a bug that restricts permissions in certain AWS regions in such a way that it provides access to users who should not have it, regardless of RAZ settings. This issue only affects users in specific AWS regions, listed below, and it does not affect all AWS customers.

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-723: Hue Raz is using logger role to Read and Upload/Delete (write) files.

TSB 2024-743: Restoration of the Database Catalog can be incomplete due to an incorrect Hue Query Processor configuration in CDW

Cloudera Data Warehouse (CDW) customers using the CDW Automated Backup/Restore feature might encounter an issue with the restoration of the Database Catalog, if the hue-query-processor.json configuration file of the Database Catalog has been edited. Even a minor edit to the hue-query-processor.json configuration file can result in this failure.

During the restoration process the Database Catalog will be created, but it will fail to start after a short period of time, and the Database Catalog will be in a Bad Health state on the CDW User Interface.

Inside the Kubernetes Cluster (Azure Kubernetes Service on Azure / Elastic Kubernetes Service on AWS) the StatefulSet of Hue Query Processor (hue-query-processor) is in CrashLoop. This is indicated with the following log in the Hue Query Processor StatefulSet Pod:

```
SQL State  : 3D000
Error Code : 0
Message    : FATAL: database "warehouse-1707832123-abcd_hueqpdb"
 does not exist



              at org.flywaydb.core.internal.jdbc.JdbcUtils.openC
onnection(JdbcUtils.java:65)
```

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-743: Restoration of the Database Catalog can be incomplete due to an incorrect Hue Query Processor configuration in CDW.

TSB 2024-745: Impala returns incorrect results for Iceberg V2 tables when optimized operator is being used in CDW

Cloudera Data Warehouse (CDW) customers using Apache Impala (Impala) to read Apache Iceberg (Iceberg) V2 tables can encounter an issue of Impala returning incorrect results when the optimized V2 operator is used. The optimized V2 operator is enabled by default in the affected versions below. The issue only affects Iceberg V2 tables that have position delete files.

Knowledge article

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-745: Impala returns incorrect results for Iceberg V2 tables when optimized operator is being used in CDW.

TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables

Apache Hive (Hive) and Apache Impala (Impala) modify statements (DELETE/UPDATE/MERGE) on Apache Iceberg (Iceberg) V2 tables can corrupt the tables if there is a concurrent table compaction from Apache Spark. The issue happens when the compaction and modify statement run in parallel, and when the compaction job commits before the modify statement. In this case the position delete files of the modify statement still point to the old files. This means the following in case of

- DELETE statements
  - Deleting records pointing to old files have no effect
- UPDATE / MERGE statements
  - Deleting records pointing to old files have no effect
  - The table will also have the newly added data records
  - Rewritten records will still be active

This issue does not affect Apache NiFi (NiFi) and Apache Flink (Flink) as these components write equality delete files.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-746: Concurrent compactions and modify statements can corrupt Iceberg tables.

TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

The Spark Iceberg library includes two procedures - rewrite_data_files and rewrite_position_delete_files. The current implementation of rewrite_data_files has a limitation that the position delete files are not deleted and still tracked by the table metadata, even if they no longer refer to an active data file. This is called the dangling delete problem. To solve this, the rewrite_position_delete_files procedure is implemented in the Spark Iceberg library to remove these old “dangling” position delete files.

Due to the dangling delete limitation, when an Iceberg table with dangling deletes is queried in Impala, Impala tries to optimize select count(*) from iceberg_table query to return the results using stats. This optimization returns incorrect results.

The following conditions must be met for this issue to occur:

- All delete files in the Iceberg table are “dangling”
  - This would occur immediately after running Spark rewrite_data_files AND
    - Before any further delete operations are performed on the table OR
    - Before Spark rewrite_position_delete_files is run on the table
- Only stats optimized plain select count(*) from iceberg_table queries are affected. For example, the query should not have:
  - Any WHERE clause
  - Any GROUP BY clause
  - Any HAVING clause

For more information, see the Apache Iceberg documentation.

Remove dangling deletes: After rewrite_data_files, position delete records pointing to the rewritten data files are not always marked for removal, and can remain tracked by the live snapshot metadata of the table. This is known as the dangling delete problem.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables.

TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion

When working with Apache Hive (Hive) and Apache Iceberg (Iceberg) V2 tables, using the TRUNCATE statement may lead to unintended data deletion. This issue arises when the truncate command is applied to a branch of an Iceberg table. Instead of truncating the branch itself, the command affects the original (main) table, which results in unintended loss of data.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion

August 30, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

New Known Issues in this release

Compaction causes data loss under certain conditions

Data loss occurs during compaction when Apache Ranger policies for masking or row filtering are enabled and compaction users are included in the policies.

Workaround: Exclude compaction users from the policies as described in Compaction Prerequisites.

MT_DOP related issue

Problem:Impala query with JOIN and LIMIT can hang when MT_DOP is greater than 0.

Workaround: Set MT_DOP to zero.

Impala Web UI incompatibility

In releases 1.6.5-b10 (released June 29, 2023) and earlier, when catalogd runs in HA mode there is single url for the Web UI. In this release 1.7.1-b755 (released August 30, 2023), when catalogd runs in HA mode there are two urls. A -0 appears in one URL, and -1 appears in the other. For example:

```
catalogd-web-0-impala-nonha.dw-dwx-9abs9a.svbr-nqvp.int.cldr.work
catalogd-web-1-impala-nonha.dw-dwx-9abs9a.svbr-nqvp.int.cldr.work
```

Workaround: Remove the -0 or the -1 from the URL.

Upgrading to EKS 1.24 could result in Impala coordinators shutting down.

This issue is not seen on the Impala Virtual Warehouse running Runtime 2023.0.15.0-x or later.

Workaround: Manually start the Impala Virtual Warehouuse from the UI or cli. Alternatively, replace your runtime with 1.7.1-b755 (released August 30, 2023).

Carried over from the previous release: Upgrade-related

Enabling a private CDW environment in Azure Kubernetes Service is unstable

Cloudera rolled back the General Availability (GA) status and support for deploying a private CDW environment in Azure using AKS until a solution for the issue is available.

When working together with Microsoft Azure (Azure), a networking issue has been identified at Azure Software Definition Layer (SDN) that impacts the use of CDW private environments while using the Azure Kubernetes Service (AKS). This issue is highly unpredictable and may cause timeouts during Cloudera Data Warehouse (CDW) environment activation, Virtual Warehouse creation, Virtual Warehouse modification, and/or during start and stop operations in CDW.

The following users are affected: CDW users on Azure who are activating a private CDW environment by selecting the Private CDW option on the User Interface (UI) or using CDP Command Line Interface (CLI) enablePrivateAks option for the dw sub-command create-cluster operation within the --aws-options group.

Activation can fail with a timeout issue, and other operation related actions may be interrupted.

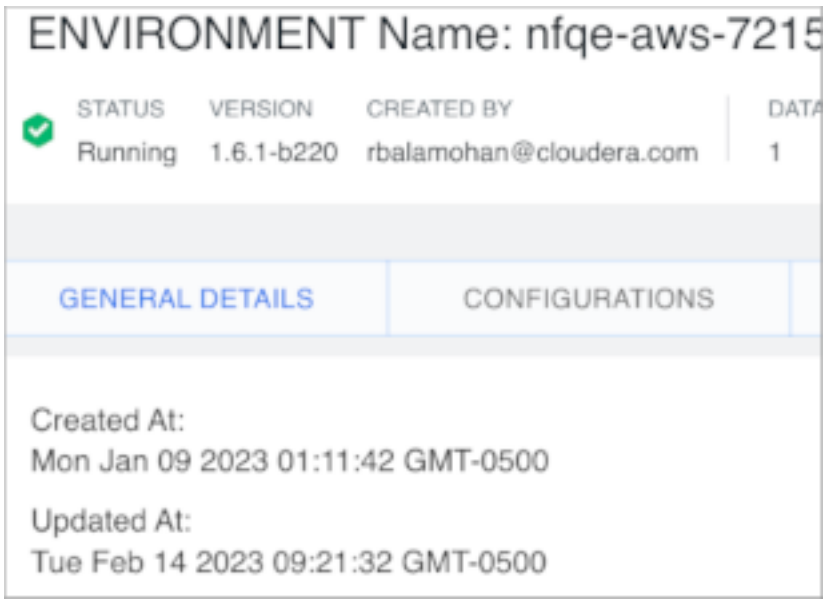
Issue backing up and restoring Workload Aware Auto-Scaling (WAAS)

You cannot use the CDW backup and restore process to restore the WAAS configurations.

Workaround: Recreate your Virtual Warehouse manually setting the right configurations.

Certain CDW versions cannot upgrade to EKS 1.22

AWS environments activated in version 1.4.1-b86 (released June, 22, 2022) or earlier are not supported for upgrade to EKS 1.22 due to incompatibility of some components with EKS 1.22. To determine the activation version your pre-existing environment, in the Data Warehouse service, expand Environments. In Environments, search for and locate the environment that you want to view. Click Edit. In Environment Details, you see the CDW version.



DWX-15112 Enterprise Data Warehouse database configuration problems after a Helm-related rollback

If a Helm rollback fails due to an incorrect Enterprise Data Warehouse database configuration, the Virtual Warehouse and Database Catalog roll back to a previous configuration. The incorrect Enterprise Data Warehouse configuration persists, and can affect subsequent edit, upgrade, and rebuild operations on the rolled-back Virtual Warehouse or Database Catalog.

Carried over from the previous release: General

DWX-14923 After JWT authentication, attempting to connect the Impyla client using a user name or password should cause an error

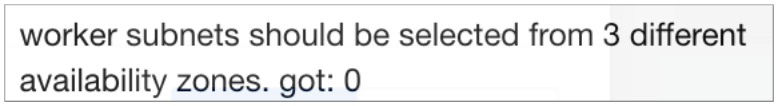
Using a JWT token, you can connect to a Virtual Warehouse as the user who generated the token.

If you connect with the JWT token, and then pass a user name or password from Impyla to the Virtual Warehouse, the connection arguments are silently ignored. Such an action should indicate that it is illegal to specify user or password when using JWT authentication.

DWX-15145 Environment validation popup error after activating an environment

Activating an environment having a public load balancer can cause an environment validatation popup error.

To reproduce this problem 1) Create a data lake. 2) Activate an enviroment having a public load balancer deployment type and subnets in three different availability zones. A environment validation popup can occur even through subnets are in different availability zones. Several differnt popups can occur, including the following one:



DWX-15144 Virtual Warehouse naming restrictions

You cannot create a Virtual Warehouse having the same name as another Virtual Warehouse even if the like-named Virtual Warehouses are in different environments. You can create a Database Catalog having the same name as another Database Catalog if the Database Catalogs are in different environments.

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data

Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the Cr eateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersions line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

DWX-15214 DWX-15176 Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

The Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed:

```
kubelet Error: failed to create containerd task: failed to creat
e shim task: OCI runtime create failed: runc create failed: unab
le to start container process: exec: “run_httpd.sh”: cannot run
executable found relative to current directory: unknown
```

This issue affects all CDW releases before 1.6.3 (released May 5, 2023)running on CDP Public Cloud on Azure.

The following users are affected: CDW Azure users using Hue in Impala and Hive Virtual Warehouses in environments earlier than version 1.6.3 or later if they upgrade the node image

Workaround: Upgrade the Virtual Warehouses to 1.6.3 and choose Hue version 2023.0.14.0.

Do not choose Hue versions older than 2023.0.14.0 when creating a Virtual Warehouse in environments of version 1.6.3.

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Database Catalog

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

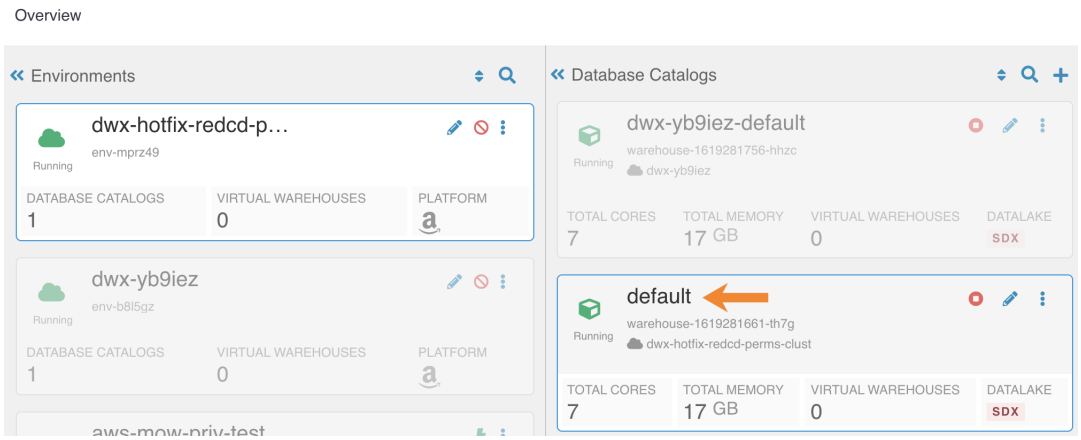
- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

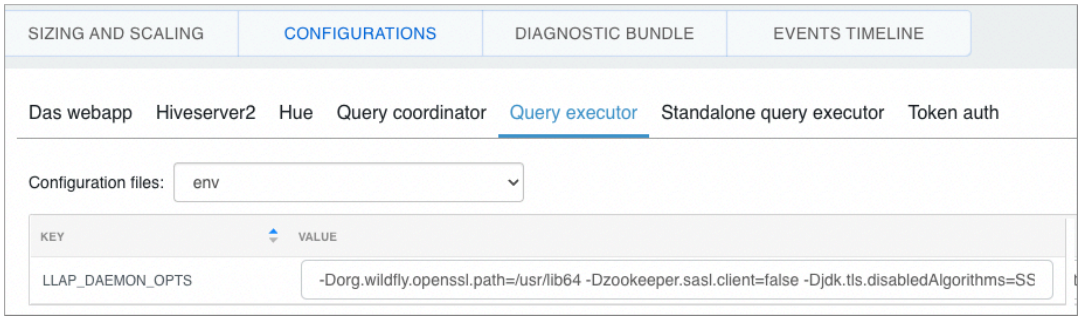
DWX-15064 Hive Virtual Warehouse stops but appears healthy

Due to an istio-proxy problem, the query coordinator can unexpectedly enter a not ready state instead of the expected error-state. Subsequently, the Hive Virtual Warehouse stops when reaching the autosuspend timeout without indicating a problem.

DWX-14452 Parquet table query might fail

Querying a table stored in Parquet from Hive might fail with the following exception message: java.lang.RuntimeException: java.lang.StackOverflowError. This problem can occur when the IN predicate has 243 values or more, and a small stack (-Xss = 256k) is configured for Hive in CDW.

Workaround: Change the value of the Xss in the JVM args to use the default value (1Mb) as follows: Select Edit from the Options menu of your Virtual Warehouse. Click Configurations > Query Executor > and select env.



In the third line shown below, change the value of LLAP_DAEMON_OPTS from -Xss256k to -Xss1M, and then click Apply Changes:

FROM:

-Dorg.wildfly.openssl.path=/usr/lib64 -Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Dhttp.maxConnections=12 -Xms24G -Xmx48G -Xss256k ...

TO:

... -Xss1M ...

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
       kind: ClusterRole
       metadata:
       name: diagnostic-data-generator-role-monitor
       rules:
       - apiGroups:
       - ""
       resources:
       - configmaps
       - events
       - pods
       - persistentvolumeclaims
       - nodes
       verbs:
       - get
       - list
       - apiGroups:
       - apps
       resources:
       - deployments
       - statefulsets
       verbs:
       - get
       - list
       - apiGroups:
       - "edws.cloudera.com"
       resources:
       - computes
       verbs:
       - get
       - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.
7. Click Apply in the upper right corner of the page to save the configuration.
8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name ‘ec2-18-219-
32-183.us-east-2.compute.amazonaws.com’ does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Hue Query Editor

Delay in listing queries in Impala Queries in the Job browser

Listing an Impala query in the Job browser can take an inordinate amount of time.

DWX-14927 Hue fails to list Iceberg snapshots

Hue does not recognize the Iceberg history queries from Hive to list table snapshots. For example, Hue indicates an error at the . before history when you run the following query.

```
select * from <db_name>.<table_name>.history
```

DWX-14968 Connection termination error in Impala queries tab after Hue inactivity

To reproduce the problem: 1) In the Impala job browser, navigate to Impala queries. 2) Wait for a few minutes.

After a few minutes of inactivity, the follow error is displayed:

Workaround: Refresh the page, or alternatively start a new session.

DWX-15115 Error displayed after clicking on hyperlink below Hue table browser

In Hue, below the table browser, clicking the hyperlink to a location causes an HTTP 500 error because the file browser is not enabled for environments that are not Ranger authorized (RAZ).

DWX-15090: CSRF error intermittently seen in the Hue Job Browser

You may intermittently see the “403 - CSRF” error on the Hue web interface as well as in the Hue logs after running Hive queries from Hue.

Reload the browser or start a new Hue session.

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Iceberg

Concurrent compactions and modify statements can corrupt Iceberg tables

Hive or Impala DELETE/UPDATE/MERGE operations on Iceberg V2 tables can corrupt the tables if there is a concurrent table compaction from Spark. The issue happens if the compaction and modify statement runs in parallel, and if the compaction job commits before the modify statement. In that case the modify statement's position delete files still point to the old files. The results in the case of DELETE and in the case of UPDATE / MERGE are as follows:

- DELETE

  Delete records pointing to old files have no effect.
- UPDATE / MERGE

  Delete records pointing to old files have no effect. The table will also have the newly added data records, which means rewritten records will still be active.

Use one of the following workarounds:

- Do not run compactions and DELETE/UPDATE/MERGE statements in parallel.
- Do not compact the table via Iceberg's RewriteFiles operation. For example do not use Spark's rewriteDataFiles.

DWX-15014 Loading airports table in demo data fails

This issue occurs only when using a non-default Database Catalog. A invalid path error message occurs when using the load command to load the demo data airports table in Iceberg, ORC, or Parquet format.

DWX-14163 Limitations reading Iceberg tables in Avro file format from Impala

The Avro, Impala, and Iceberg specifications describe some limitations related to Avro, and those limitations exist in CDP. In addition to these, the DECIMAL type is not supported in this release.

DEX-7946 Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata if you drop the table during migration from Hive to Iceberg.

Workarounds: Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13733 Timeout issue querying Iceberg tables from Hive

A timeout issue can cause the query to fail.

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

```
fs.s3.maxConnections=1000
fs.s3a.connection.maximum=1000
```

Restart the Database Catalog and Virtual Warehouse.

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: Impala Virtual Warehouse

TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

The Cloudera Data Warehouse (CDW) Public Cloud 2023.0.14.0 (DWX-1.6.3) version incorporated a feature for performing automatic metadata synchronization across multiple Apache Impala (Impala) Virtual Warehouses. The feature is enabled by default, and relies on the Hive MetaStore events. When a certain sequence of Data Definition Language (DDL) SQL commands are executed as described below, users may encounter a java.lang.NullPointerException (NPE). The exception causes the event processor to stop processing other metadata operations.

If a CREATE TABLE command (not CREATE TABLE AS SELECT) is followed immediately (approximately within 1 second interval) by INVALIDATE METADATA or REFRESH TABLE command on the same table (either on the same Virtual Warehouse or on a different one), there is a possibility that the second command will not find the table in the catalog cache of a peer Virtual Warehouse and generate an NPE.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

DWX-15016 Impala query failure when using Catalog high availability (HA)

Impala queries could fail with InconsistentMetadataFetchException when using HA feature in CDW. This happens when leader election failover for Impala Catalog Service happens due to a transient network error.

Workaround: Disable Catalog HA. In CDW, edit your Virtual Warehouse, and turn off Enable Impala Catalog Server HA.

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml
```

```
        # Make sure that the annotation you added is present.
```

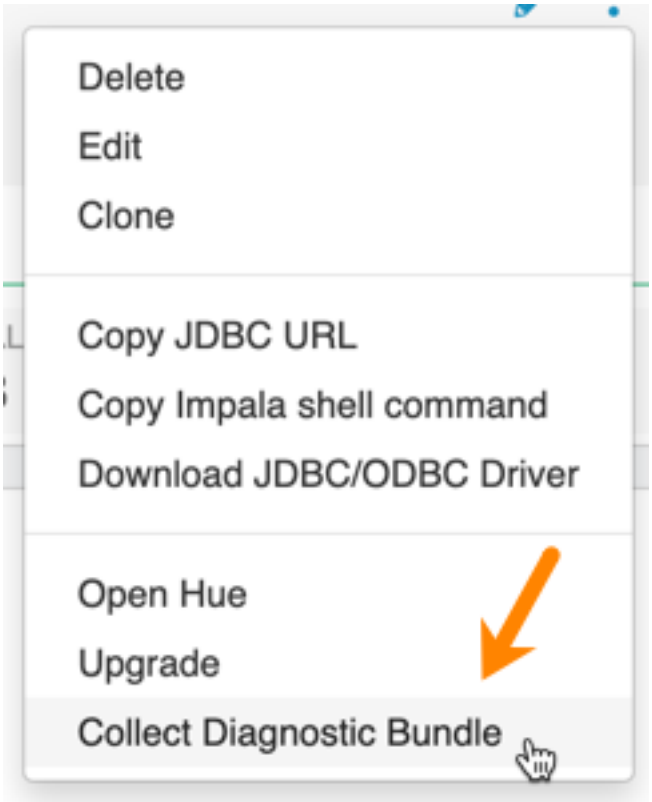
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

UI Issues

DWX-14610 Upgrade icon indicates the Database Catalog is the latest version, but might be incorrect for a few seconds

The cluster fetches data regularly in particular time intervals, or when an action is triggered. Inbetween this time interval, which is very brief, the cluster might not be updated, but will be refreshed in a few seconds.

Technical Service Bulletins

TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

The Spark Iceberg library includes two procedures - rewrite_data_files and rewrite_position_delete_files. The current implementation of rewrite_data_files has a limitation that the position delete files are not deleted and still tracked by the table metadata, even if they no longer refer to an active data file. This is called the dangling delete problem. To solve this, the rewrite_position_delete_files procedure is implemented in the Spark Iceberg library to remove these old “dangling” position delete files.

Due to the dangling delete limitation, when an Iceberg table with dangling deletes is queried in Impala, Impala tries to optimize select count(*) from iceberg_table query to return the results using stats. This optimization returns incorrect results.

The following conditions must be met for this issue to occur:

- All delete files in the Iceberg table are “dangling”
  - This would occur immediately after running Spark rewrite_data_files AND
    - Before any further delete operations are performed on the table OR
    - Before Spark rewrite_position_delete_files is run on the table
- Only stats optimized plain select count(*) from iceberg_table queries are affected. For example, the query should not have:
  - Any WHERE clause
  - Any GROUP BY clause
  - Any HAVING clause

For more information, see the Apache Iceberg documentation.

Remove dangling deletes: After rewrite_data_files, position delete records pointing to the rewritten data files are not always marked for removal, and can remain tracked by the live snapshot metadata of the table. This is known as the dangling delete problem.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables.

TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion

When working with Apache Hive (Hive) and Apache Iceberg (Iceberg) V2 tables, using the TRUNCATE statement may lead to unintended data deletion. This issue arises when the truncate command is applied to a branch of an Iceberg table. Instead of truncating the branch itself, the command affects the original (main) table, which results in unintended loss of data.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-758: Truncate command on Iceberg V2 branches cause unintentional data deletion

June 29, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

New Known Issues in this release

Issue backing up and restoring Workload Aware Auto-Scaling (WAAS)

You cannot use the CDW backup and restore process to restore the WAAS configurations.

Workaround: Recreate your Virtual Warehouse manually setting the right configurations.

Carried over from the previous release: Upgrade-related

Enabling a private CDW environment in Azure Kubernetes Service is unstable

Cloudera rolled back the General Availability (GA) status and support for deploying a private CDW environment in Azure using AKS until a solution for the issue is available.

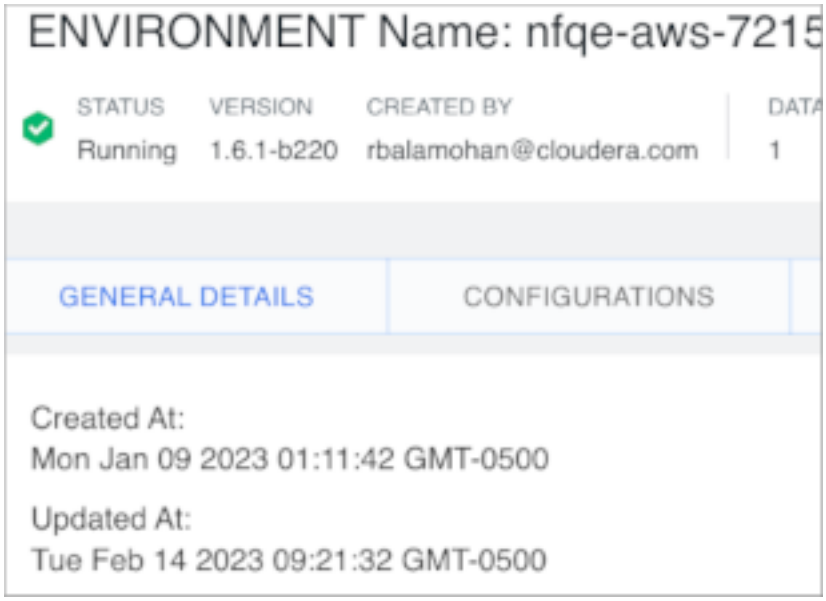
When working together with Microsoft Azure (Azure), a networking issue has been identified at Azure Software Definition Layer (SDN) that impacts the use of CDW private environments while using the Azure Kubernetes Service (AKS). This issue is highly unpredictable and may cause timeouts during Cloudera Data Warehouse (CDW) environment activation, Virtual Warehouse creation, Virtual Warehouse modification, and/or during start and stop operations in CDW.

The following users are affected: CDW users on Azure who are activating a private CDW environment by selecting the Private CDW option on the User Interface (UI) or using CDP Command Line Interface (CLI) enablePrivateAks option for the dw sub-command create-cluster operation within the --aws-options group.

Activation can fail with a timeout issue, and other operation related actions may be interrupted.

Certain CDW versions cannot upgrade to EKS 1.22

AWS environments activated in version 1.4.1-b86 (released June, 22, 2022) or earlier are not supported for upgrade to EKS 1.22 due to incompatibility of some components with EKS 1.22. To determine the activation version your pre-existing environment, in the Data Warehouse service, expand Environments. In Environments, search for and locate the environment that you want to view. Click Edit. In Environment Details, you see the CDW version.



DWX-15114 Incorrect Database Catalog status indication incorrect after an upgrade

After a Database Catalog upgrade, if the Databus producer crashes, the default Database Catalog does not indicate the error. The Database Catalog status instead indicates "running". After a Database Catalog upgrade, if the Databus producer crashes, a non-default (custom) Database Catalog, indicates the correct error status.

DWX-15112 Enterprise Data Warehouse database configuration problems after a Helm-related rollback

If a Helm rollback fails due to an incorrect Enterprise Data Warehouse database configuration, the Virtual Warehouse and Database Catalog roll back to a previous configuration. The incorrect Enterprise Data Warehouse configuration persists, and can affect subsequent edit, upgrade, and rebuild operations on the rolled-back Virtual Warehouse or Database Catalog.

Carried over from the previous release: General

DWX-14923 After JWT authentication, attempting to connect the Impyla client using a user name or password should cause an error

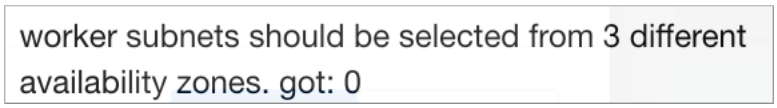
Using a JWT token, you can connect to a Virtual Warehouse as the user who generated the token.

If you connect with the JWT token, and then pass a user name or password from Impyla to the Virtual Warehouse, the connection arguments are silently ignored. Such an action should indicate that it is illegal to specify user or password when using JWT authentication.

DWX-15145 Environment validation popup error after activating an environment

Activating an environment having a public load balancer can cause an environment validatation popup error.

To reproduce this problem 1) Create a data lake. 2) Activate an enviroment having a public load balancer deployment type and subnets in three different availability zones. A environment validation popup can occur even through subnets are in different availability zones. Several differnt popups can occur, including the following one:



DWX-15144 Virtual Warehouse naming restrictions

You cannot create a Virtual Warehouse having the same name as another Virtual Warehouse even if the like-named Virtual Warehouses are in different environments. You can create a Database Catalog having the same name as another Database Catalog if the Database Catalogs are in different environments.

DWX-15151 Runtime error after activating an environment

Activating an environment having a public load balancer can cause a 2083 runtime error

To reproduce this problem 1) Create a data lake. 2) Activate an enviroment having a public load balancer deployment type and subnets in three different availability zones. The following error can occur:

```
RuntimeErr with ErrCode=2083 (cause: 0 'coredns' replica(s) are
ready) Error Code : undefined
```

DWX-15171 Error after disabling SSO.

Disabling SSO for Impala Virtual Warehouse can cause an error.

To reproduce this issue 1) Create an Impala Virtual Warehouse having SSO enabled. 2) Edit the Virtual Warehouse to disable SSO.

The coordinator crashes.

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-6619 Browser auto-close not working on some browsers after token-Based authentication for accessing CDW

The Firefox and Edge browser window does not close automatically after successful authentication.

DWX-9774 Database Catalog or Virtual Warehouse image version problem

Background: In Cloudera Data Warehouse 2021.0.3-b27 - 2021.0.5-b36, you can choose any supported image version when you create a Database Catalog or Virtual Warehouse, assuming you have the CDW_VERSIONED_DEPLOY entitlement.

Problem: In Cloudera Data Warehouse 2021.0.6-b96, you can choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse only if you use an existing environment. An existing environment is one you created in 2021.0.3-b27 - 2021.0.5-b36.

Workaround: In 2021.0.6-b96, use an environment you created before this release to choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse.

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

DWX-15246 Missing entries in subnet selection during CDW environment activation

The subnet selection during environment activation might not display any, or all, of the subnets registered in the environment.

Workaround: Carefully review the list of subnets you can select during activation. If subnets you want are not visible, enter the subnet-ids manually into the list, and then activate the environment.

DWX-14409 Starting a Virtual Warehouse in an AWS environments in the Asia-Pacific Southeast region fails

When you upgrade the Virtual Warehouse in this release and start it, a timing issue causes failure.

Workaround: Continue using the previous release 1.5.1-b110 (released Nov 22, 2022). Do not upgrade to 1.6.1-b282 (released Feb 7, 2023). Delay upgrading your Virtual Warehouse until the next release of CDW.

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the Cr eateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersio ns line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

Enabling a private CDW environment in Azure Kubernetes Service is unstable

Cloudera rolled back the General Availability (GA) status and support for deploying a private CDW environment in Azure using AKS until a solution for the issue is available.

When working together with Microsoft Azure (Azure), a networking issue has been identified at Azure Software Definition Layer (SDN) that impacts the use of CDW private environments while using the Azure Kubernetes Service (AKS). This issue is highly unpredictable and may cause timeouts during Cloudera Data Warehouse (CDW) environment activation, Virtual Warehouse creation, Virtual Warehouse modification, and/or during start and stop operations in CDW.

The following users are affected: CDW users on Azure who are activating a private CDW environment by selecting the Private CDW option on the User Interface (UI) or using CDP Command Line Interface (CLI) enablePrivateAks option for the dw sub-command create-cluster operation within the --aws-options group.

Activation can fail with a timeout issue, and other operation related actions may be interrupted.

DWX-15214 DWX-15176 Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

The Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed:

```
kubelet Error: failed to create containerd task: failed to creat
e shim task: OCI runtime create failed: runc create failed: unab
le to start container process: exec: “run_httpd.sh”: cannot run
executable found relative to current directory: unknown
```

This issue affects all CDW releases before 1.6.3 (released May 5, 2023)running on CDP Public Cloud on Azure.

The following users are affected:

- CDW Azure users using Hue in Impala and Hive Virtual Warehouses in environments earlier than version 1.6.3 if they upgrade the node image
- CDW Azure users creating a new CDW environment with a Hue version earlier than 2023.0.14.0 (released May 5, 2023) during Hive or Impala Virtual Warehouse creation

Workaround: Upgrade the Virtual Warehouses to 1.6.3 or later and choose Hue version 2023.0.14.0.

Do not choose Hue versions older than 2023.0.14.0 when creating a Virtual Warehouse in environments of version 1.6.3.

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Database Catalog

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

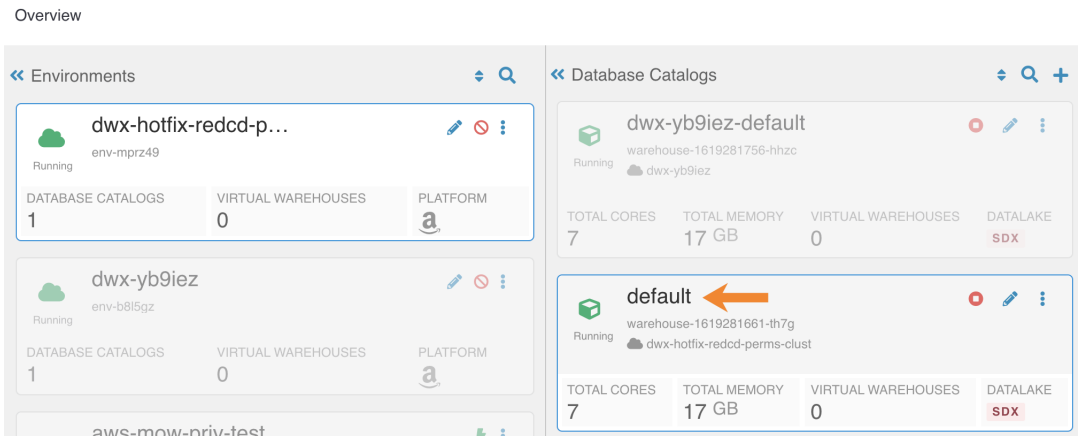
- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

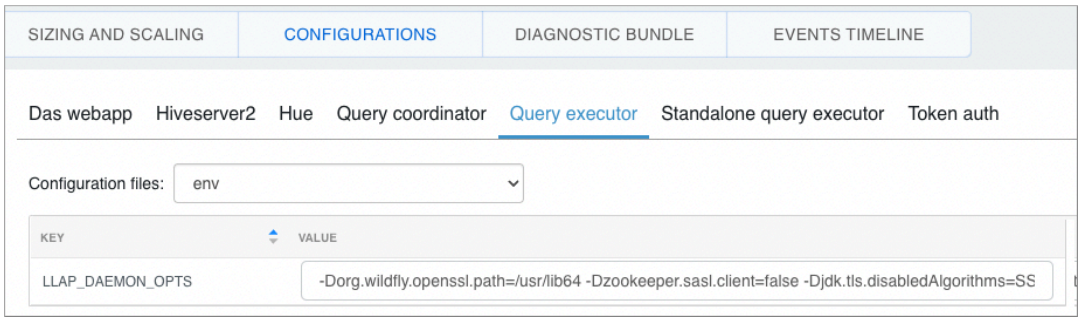
DWX-15064 Hive Virtual Warehouse stops but appears healthy

Due to an istio-proxy problem, the query coordinator can unexpectedly enter a not ready state instead of the expected error-state. Subsequently, the Hive Virtual Warehouse stops when reaching the autosuspend timeout without indicating a problem.

DWX-14452 Parquet table query might fail

Querying a table stored in Parquet from Hive might fail with the following exception message: java.lang.RuntimeException: java.lang.StackOverflowError. This problem can occur when the IN predicate has 243 values or more, and a small stack (-Xss = 256k) is configured for Hive in CDW.

Workaround: Change the value of the Xss in the JVM args to use the default value (1Mb) as follows: Select Edit from the Options menu of your Virtual Warehouse. Click Configurations > Query Executor > and select env.



In the third line shown below, change the value of LLAP_DAEMON_OPTS from -Xss256k to -Xss1M, and then click Apply Changes:

FROM:

-Dorg.wildfly.openssl.path=/usr/lib64 -Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Dhttp.maxConnections=12 -Xms24G -Xmx48G -Xss256k ...

TO:

... -Xss1M ...

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
       kind: ClusterRole
       metadata:
       name: diagnostic-data-generator-role-monitor
       rules:
       - apiGroups:
       - ""
       resources:
       - configmaps
       - events
       - pods
       - persistentvolumeclaims
       - nodes
       verbs:
       - get
       - list
       - apiGroups:
```

```
        - apps
        resources:
        - deployments
        - statefulsets
        verbs:
        - get
        - list
        - apiGroups:
        - "edws.cloudera.com"
        resources:
        - computes
        verbs:
        - get
        - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.

7. Click Apply in the upper right corner of the page to save the configuration.

8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name ‘ec2-18-219-
32-183.us-east-2.compute.amazonaws.com’ does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Hue Query Editor

Delay in listing queries in Impala Queries in the Job browser

Listing an Impala query in the Job browser can take an inordinate amount of time.

DWX-14927 Hue fails to list Iceberg snapshots

Hue does not recognize the Iceberg history queries from Hive to list table snapshots. For example, Hue indicates an error at the . before history when you run the following query.

```
select * from <db_name>.<table_name>.history
```

DWX-14968 Connection termination error in Impala queries tab after Hue inactivity

To reproduce the problem: 1) In the Impala job browser, navigate to Impala queries. 2) Wait for a few minutes.

After a few minutes of inactivity, the follow error is displayed:

Workaround: Refresh the page, or alternatively start a new session.

DWX-15115 Error displayed after clicking on hyperlink below Hue table browser

In Hue, below the table browser, clicking the hyperlink to a location causes an HTTP 500 error because the file browser is not enabled for environments that are not Ranger authorized (RAZ).

DWX-15090: CSRF error intermittently seen in the Hue Job Browser

You may intermittently see the “403 - CSRF” error on the Hue web interface as well as in the Hue logs after running Hive queries from Hue.

Reload the browser or start a new Hue session.

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Iceberg

DWX-15014 Loading airports table in demo data fails

This issue occurs only when using a non-default Database Catalog. A invalid path error message occurs when using the load command to load the demo data airports table in Iceberg, ORC, or Parquet format.

DWX-14163 Limitations reading Iceberg tables in Avro file format from Impala

The Avro, Impala, and Iceberg specifications describe some limitations related to Avro, and those limitations exist in CDP. In addition to these, the DECIMAL type is not supported in this release.

DWX-14286 Loading data to Iceberg error

Using LOAD DATA INPATH from an Impala Virtual Warehouse to load data to an Iceberg table on S3/ADLS when hidden files (files that are prefixed with . or _) are in the path can lead to unexpected query failures. For example:

```
AnalysisException: INPATH contains unsupported LOAD format, file
 '
            s3a://dwx-testdata/impala/sql_test/tests/load_data_in
path/runtime_data/0690a6fa9bfb11ed920c164053429bec/load_data_tes
t/A/impala_data/impala_alltypessmall_data/alltypessmall_parquet_
iceberg/year=2009/month=1/.hiddenfileforloaddatatest
            ' has 'This' magic string.
```

DEX-7946 Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata if you drop the table during migration from Hive to Iceberg.

Workarounds: Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13733 Timeout issue querying Iceberg tables from Hive

A timeout issue can cause the query to fail.

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

```
fs.s3.maxConnections=1000
fs.s3a.connection.maximum=1000
```

Restart the Database Catalog and Virtual Warehouse.

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: Impala Virtual Warehouse

TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

The Cloudera Data Warehouse (CDW) Public Cloud 2023.0.14.0 (DWX-1.6.3) version incorporated a feature for performing automatic metadata synchronization across multiple Apache Impala (Impala) Virtual Warehouses. The feature is enabled by default, and relies on the Hive MetaStore events. When a certain sequence of Data Definition Language (DDL) SQL commands are executed as described below, users may encounter a java.lang.NullPointerException (NPE). The exception causes the event processor to stop processing other metadata operations.

If a CREATE TABLE command (not CREATE TABLE AS SELECT) is followed immediately (approximately within 1 second interval) by INVALIDATE METADATA or REFRESH TABLE command on the same table (either on the same Virtual Warehouse or on a different one), there is a possibility that the second command will not find the table in the catalog cache of a peer Virtual Warehouse and generate an NPE.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

DWX-15016 Impala query failure when using Catalog high availability (HA)

Impala queries could fail with InconsistentMetadataFetchException when using HA feature in CDW. This happens when leader election failover for Impala Catalog Service happens due to a transient network error.

Workaround: Disable Catalog HA. In CDW, edit your Virtual Warehouse, and turn off Enable Impala Catalog Server HA.

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

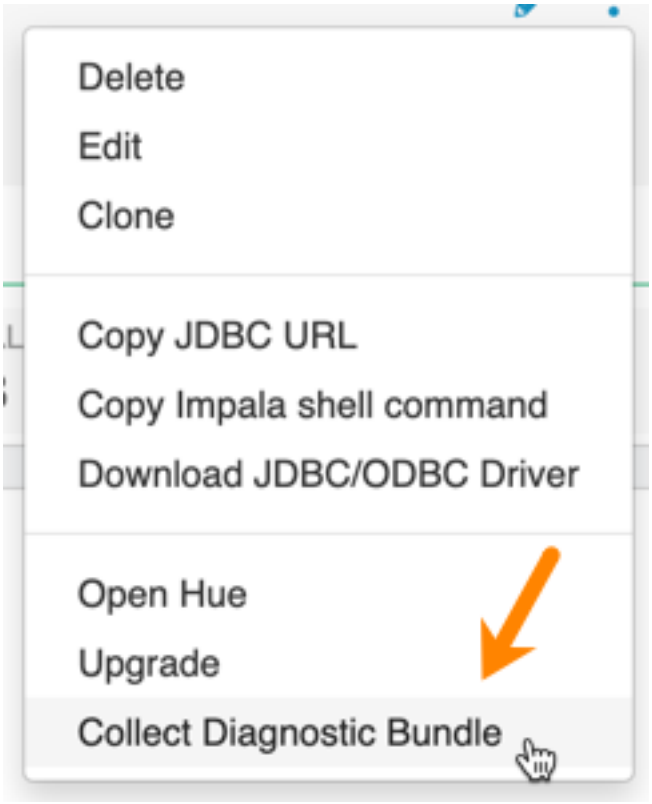
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

UI Issues

DWX-15110 Misaligned dialog in the UI

The Create Virtual Warehouse dialog is not aligned properly when no Virtual Warehouses are running.

DWX-14992 Cosmetic UI problem with Enable DAS link

In Azure environments, no gap appears between the previous text box and Enable DAS.

DWX-14970 Cosmetic UI problem with Virtual Warehouse start label

The start labels of the Impala Virtual Warehouse and Hive Virtual Warehouse are inconsistent with regard to upper- and lower-case.

DWX-14874 No indication that there are no Virtual Warehouses running

When no Data Visualization instance is running, you see "Currently no environments connected to Data Visualization cluster". The same indicator should be displayed when no Virtual Warehouses are running.

DWX-14610 Upgrade icon indicates the Database Catalog is the latest version, but might be incorrect for a few seconds

The cluster fetches data regularly in particular time intervals, or when an action is triggered. Inbetween this time interval, which is very brief, the cluster might not be updated, but will be refreshed in a few seconds.

Technical Service Bulletins

TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

The Spark Iceberg library includes two procedures - rewrite_data_files and rewrite_position_delete_files. The current implementation of rewrite_data_files has a limitation that the position delete files are not deleted and still tracked by the table metadata, even if they no longer refer to an active data file. This is called the dangling delete problem. To solve this, the rewrite_position_delete_files procedure is implemented in the Spark Iceberg library to remove these old “dangling” position delete files.

Due to the dangling delete limitation, when an Iceberg table with dangling deletes is queried in Impala, Impala tries to optimize select count(*) from iceberg_table query to return the results using stats. This optimization returns incorrect results.

The following conditions must be met for this issue to occur:

- All delete files in the Iceberg table are “dangling”
  - This would occur immediately after running Spark rewrite_data_files AND
    - Before any further delete operations are performed on the table OR
    - Before Spark rewrite_position_delete_files is run on the table
- Only stats optimized plain select count(*) from iceberg_table queries are affected. For example, the query should not have:
  - Any WHERE clause
  - Any GROUP BY clause
  - Any HAVING clause

For more information, see the Apache Iceberg documentation.

Remove dangling deletes: After rewrite_data_files, position delete records pointing to the rewritten data files are not always marked for removal, and can remain tracked by the live snapshot metadata of the table. This is known as the dangling delete problem.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables.

May 30, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

New Known Issues in this release

DWX-15701 After a CDW upgrade, a Solr auditing problem exists

After a CDW upgrade, Solr cannot load ranger_audits collection and hence fails to push audit events to Ranger.

DWX-15395 Hive compaction fails in Azure

When RAZ is enabled in Cloudera Data Warehouse Azure environment, Hive compaction fails and returns an error message something like this:

```
ErrorMessage = FAILED: HiveAccessControlException Permission den
ied: user [ed64e526-a445-4945-9ed6-2eaf9560ff08] does not have [
READ] privilege on [[abfs:
```

Workaround:

1) In the CDW UI, in Overview, select a Hive Virtual Warehouse, and click Edit.

2) In Details, click CONFIGURATIONS HiveServer2 . 3) Select hive-site from the drop-down list, and search for metastore.compactor.run.as.user.

4) If the metastore.compactor.run.as.user key does not appear, click + and add it.

5) In Value, enter your Microsoft managed identity, and save the change.

Restart the Virtual Warehouse.

DWX-15473 Reactivating the AWS environment in CDW fails

When activating the AWS environment in CDW, the usage monitor can fail due to a problem connecting to prometheus. The connection problem can be caused by a restrictive AWS account policy. The follow error message occurs:

```
User: arn:aws:sts::OBFUSCATED:assumed-role/env-OBFUSCATED-dwx-st
ack-NodeInstanceRole-OBFUSCATED/i-OBFUSCATED is not authorized to
 perform: elasticfilesystem:TagResource on the specified resource
```

Workaround: After reactivating the CDW Environment, navigate to associated Cloudformation template for CDW, click Resources. Select NodeInstanceRole, and then select **efs inline policy**. Replace the JSON policy with the policy on the kubernetes-sigs github site.

Carried over from the previous release: Upgrade-related

Enabling a private CDW environment in Azure Kubernetes Service is unstable

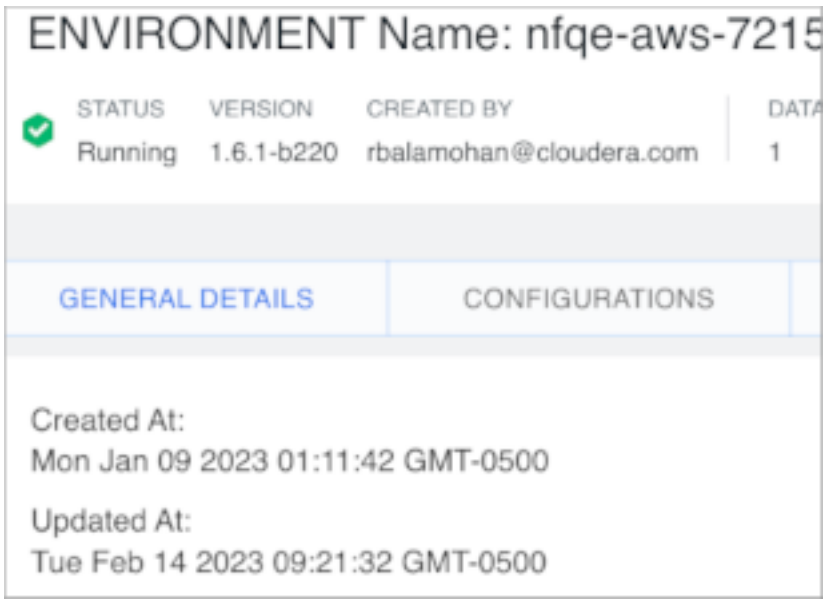
Cloudera rolled back the General Availability (GA) status and support for deploying a private CDW environment in Azure using AKS until a solution for the issue is available. When working together with Microsoft Azure (Azure), a networking issue has been identified at Azure Software Definition Layer (SDN) that impacts the use of CDW private environments while using the Azure Kubernetes Service (AKS). This issue is highly unpredictable and may cause timeouts during Cloudera Data Warehouse (CDW) environment activation, Virtual Warehouse creation, Virtual Warehouse modification, and/or during start and stop operations in CDW.

The following users are affected: CDW users on Azure who are activating a private CDW environment by selecting the Private CDW option on the User Interface (UI) or using CDP Command Line Interface (CLI) enablePrivateAks option for the dw sub-command create-cluster operation within the --aws-options group.

Activation can fail with a timeout issue, and other operation related actions may be interrupted.

Certain CDW versions cannot upgrade to EKS 1.22

AWS environments activated in version 1.4.1-b86 (released June, 22, 2022) or earlier are not supported for upgrade to EKS 1.22 due to incompatibility of some components with EKS 1.22. To determine the activation version your pre-existing environment, in the Data Warehouse service, expand Environments. In Environments, search for and locate the environment that you want to view. Click Edit. In Environment Details, you see the CDW version.



DWX-15114 Incorrect Database Catalog status indication incorrect after an upgrade

After a Database Catalog upgrade, if the Databus producer crashes, the default Database Catalog does not indicate the error. The Database Catalog status instead indicates "running". After a Database Catalog upgrade, if the Databus producer crashes, a non-default (custom) Database Catalog, indicates the correct error status.

DWX-15112 Enterprise Data Warehouse database configuration problems after a Helm-related rollback

If a Helm rollback fails due to an incorrect Enterprise Data Warehouse database configuration, the Virtual Warehouse and Database Catalog roll back to a previous configuration. The incorrect Enterprise Data Warehouse configuration persists, and can affect subsequent edit, upgrade, and rebuild operations on the rolled-back Virtual Warehouse or Database Catalog.

Carried over from the previous release: General

DWX-14923 After JWT authentication, attempting to connect the Impyla client using a user name or password should cause an error

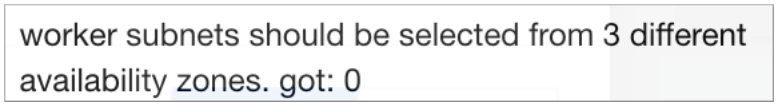
Using a JWT token, you can connect to a Virtual Warehouse as the user who generated the token.

If you connect with the JWT token, and then pass a user name or password from Impyla to the Virtual Warehouse, the connection arguments are silently ignored. Such an action should indicate that it is illegal to specify user or password when using JWT authentication.

DWX-15145 Environment validation popup error after activating an environment

Activating an environment having a public load balancer can cause an environment validatation popup error.

To reproduce this problem 1) Create a data lake. 2) Activate an enviroment having a public load balancer deployment type and subnets in three different availability zones. A environment validation popup can occur even through subnets are in different availability zones. Several differnt popups can occur, including the following one:



DWX-15144 Virtual Warehouse naming restrictions

You cannot create a Virtual Warehouse having the same name as another Virtual Warehouse even if the like-named Virtual Warehouses are in different environments. You can create a Database Catalog having the same name as another Database Catalog if the Database Catalogs are in different environments.

DWX-15151 Runtime error after activating an environment

Activating an environment having a public load balancer can cause a 2083 runtime error

To reproduce this problem 1) Create a data lake. 2) Activate an enviroment having a public load balancer deployment type and subnets in three different availability zones. The following error can occur:

```
RuntimeErr with ErrCode=2083 (cause: 0 'coredns' replica(s) are
ready) Error Code : undefined
```

DWX-15171 Error after disabling SSO.

Disabling SSO for Impala Virtual Warehouse can cause an error.

To reproduce this issue 1) Create an Impala Virtual Warehouse having SSO enabled. 2) Edit the Virtual Warehouse to disable SSO.

The coordinator crashes.

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-6619 Browser auto-close not working on some browsers after token-Based authentication for accessing CDW

The Firefox and Edge browser window does not close automatically after successful authentication.

DWX-9774 Database Catalog or Virtual Warehouse image version problem

Background: In Cloudera Data Warehouse 2021.0.3-b27 - 2021.0.5-b36, you can choose any supported image version when you create a Database Catalog or Virtual Warehouse, assuming you have the CDW_VERSIONED_DEPLOY entitlement.

Problem: In Cloudera Data Warehouse 2021.0.6-b96, you can choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse only if you use an existing environment. An existing environment is one you created in 2021.0.3-b27 - 2021.0.5-b36.

Workaround: In 2021.0.6-b96, use an environment you created before this release to choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse.

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

DWX-15246 Missing entries in subnet selection during CDW environment activation

The subnet selection during environment activation might not display any, or all, of the subnets registered in the environment.

Workaround: Carefully review the list of subnets you can select during activation. If subnets you want are not visible, enter the subnet-ids manually into the list, and then activate the environment.

DWX-14409 Starting a Virtual Warehouse in an AWS environments in the Asia-Pacific Southeast region fails

When you upgrade the Virtual Warehouse in this release and start it, a timing issue causes failure.

Workaround: Continue using the previous release 1.5.1-b110 (released Nov 22, 2022). Do not upgrade to 1.6.1-b282 (released Feb 7, 2023). Delay upgrading your Virtual Warehouse until the next release of CDW.

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the Cr eateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.

6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersions line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

DWX-15214 DWX-15176 Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

The Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed:

```
kubelet Error: failed to create containerd task: failed to creat
e shim task: OCI runtime create failed: runc create failed: unab
le to start container process: exec: “run_httpd.sh”: cannot run
executable found relative to current directory: unknown
```

This issue affects all CDW releases before 1.6.3 (released May 5, 2023)running on CDP Public Cloud on Azure.

The following users are affected:

- CDW Azure users using Hue in Impala and Hive Virtual Warehouses in environments earlier than version 1.6.3 if they upgrade the node image
- CDW Azure users creating a new CDW environment with a Hue version earlier than 2023.0.14.0 (released May 5, 2023) during Hive or Impala Virtual Warehouse creation

Workaround: Upgrade the Virtual Warehouses to 1.6.3 or later and choose Hue version 2023.0.14.0.

Do not choose Hue versions older than 2023.0.14.0 when creating a Virtual Warehouse in environments of version 1.6.3.

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Database Catalog

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

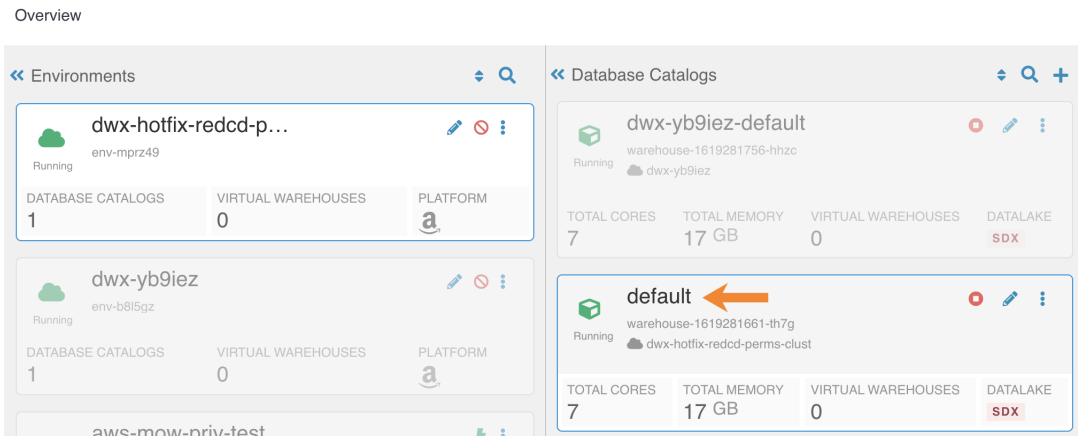
- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

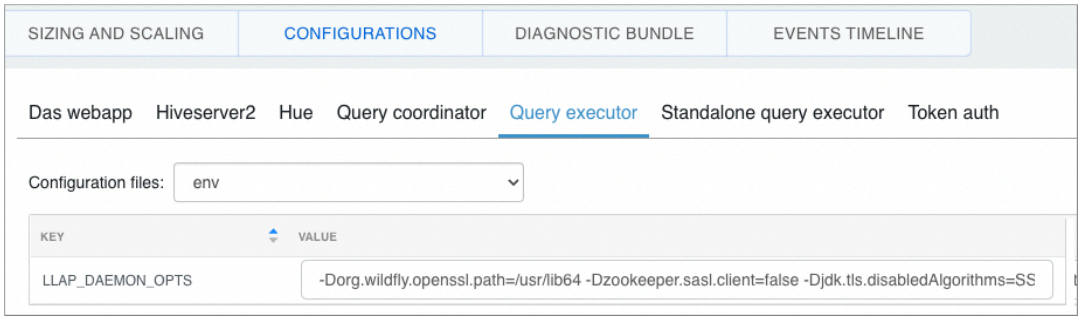
DWX-15064 Hive Virtual Warehouse stops but appears healthy

Due to an istio-proxy problem, the query coordinator can unexpectedly enter a not ready state instead of the expected error-state. Subsequently, the Hive Virtual Warehouse stops when reaching the autosuspend timeout without indicating a problem.

DWX-14452 Parquet table query might fail

Querying a table stored in Parquet from Hive might fail with the following exception message: java.lang.RuntimeException: java.lang.StackOverflowError. This problem can occur when the IN predicate has 243 values or more, and a small stack (-Xss = 256k) is configured for Hive in CDW.

Workaround: Change the value of the Xss in the JVM args to use the default value (1Mb) as follows: Select Edit from the Options menu of your Virtual Warehouse. Click Configurations > Query Executor > and select env.



In the third line shown below, change the value of LLAP_DAEMON_OPTS from -Xss256k to -Xss1M, and then click Apply Changes:

FROM:

-Dorg.wildfly.openssl.path=/usr/lib64 -Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Dhttp.maxConnections=12 -Xms24G -Xmx48G -Xss256k ...

TO:

... -Xss1M ...

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
       kind: ClusterRole
       metadata:
       name: diagnostic-data-generator-role-monitor
       rules:
       - apiGroups:
       - ""
       resources:
       - configmaps
       - events
       - pods
       - persistentvolumeclaims
       - nodes
       verbs:
       - get
       - list
       - apiGroups:
       - apps
       resources:
       - deployments
       - statefulsets
       verbs:
       - get
       - list
       - apiGroups:
       - "edws.cloudera.com"
       resources:
       - computes
       verbs:
       - get
       - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.
7. Click Apply in the upper right corner of the page to save the configuration.
8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name ‘ec2-18-219-
32-183.us-east-2.compute.amazonaws.com’ does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Hue Query Editor

Delay in listing queries in Impala Queries in the Job browser

Listing an Impala query in the Job browser can take an inordinate amount of time.

DWX-14927 Hue fails to list Iceberg snapshots

Hue does not recognize the Iceberg history queries from Hive to list table snapshots. For example, Hue indicates an error at the . before history when you run the following query.

```
select * from <db_name>.<table_name>.history
```

DWX-14968 Connection termination error in Impala queries tab after Hue inactivity

To reproduce the problem: 1) In the Impala job browser, navigate to Impala queries. 2) Wait for a few minutes.

After a few minutes of inactivity, the follow error is displayed:

Workaround: Refresh the page, or alternatively start a new session.

DWX-15115 Error displayed after clicking on hyperlink below Hue table browser

In Hue, below the table browser, clicking the hyperlink to a location causes an HTTP 500 error because the file browser is not enabled for environments that are not Ranger authorized (RAZ).

DWX-15090: CSRF error intermittently seen in the Hue Job Browser

You may intermittently see the “403 - CSRF” error on the Hue web interface as well as in the Hue logs after running Hive queries from Hue.

Reload the browser or start a new Hue session.

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Iceberg

DWX-15014 Loading airports table in demo data fails

This issue occurs only when using a non-default Database Catalog. A invalid path error message occurs when using the load command to load the demo data airports table in Iceberg, ORC, or Parquet format.

DWX-14163 Limitations reading Iceberg tables in Avro file format from Impala

The Avro, Impala, and Iceberg specifications describe some limitations related to Avro, and those limitations exist in CDP. In addition to these, the DECIMAL type is not supported in this release.

DWX-14286 Loading data to Iceberg error

Using LOAD DATA INPATH from an Impala Virtual Warehouse to load data to an Iceberg table on S3/ADLS when hidden files (files that are prefixed with . or _) are in the path can lead to unexpected query failures. For example:

```
AnalysisException: INPATH contains unsupported LOAD format, file
 '
            s3a://dwx-testdata/impala/sql_test/tests/load_data_in
path/runtime_data/0690a6fa9bfb11ed920c164053429bec/load_data_tes
t/A/impala_data/impala_alltypessmall_data/alltypessmall_parquet_
iceberg/year=2009/month=1/.hiddenfileforloaddatatest
            ' has 'This' magic string.
```

DEX-7946 Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata if you drop the table during migration from Hive to Iceberg.

Workarounds: Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13733 Timeout issue querying Iceberg tables from Hive

A timeout issue can cause the query to fail.

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

```
fs.s3.maxConnections=1000
fs.s3a.connection.maximum=1000
```

Restart the Database Catalog and Virtual Warehouse.

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: Impala Virtual Warehouse

TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

The Cloudera Data Warehouse (CDW) Public Cloud 2023.0.14.0 (DWX-1.6.3) version incorporated a feature for performing automatic metadata synchronization across multiple Apache Impala (Impala) Virtual Warehouses. The feature is enabled by default, and relies on the Hive MetaStore events. When a certain sequence of Data Definition Language (DDL) SQL commands are executed as described below, users may encounter a java.lang.NullPointerException (NPE). The exception causes the event processor to stop processing other metadata operations.

If a CREATE TABLE command (not CREATE TABLE AS SELECT) is followed immediately (approximately within 1 second interval) by INVALIDATE METADATA or REFRESH TABLE command on the same table (either on the same Virtual Warehouse or on a different one), there is a possibility that the second command will not find the table in the catalog cache of a peer Virtual Warehouse and generate an NPE.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

DWX-15016 Impala query failure when using Catalog high availability (HA)

Impala queries could fail with InconsistentMetadataFetchException when using HA feature in CDW. This happens when leader election failover for Impala Catalog Service happens due to a transient network error.

Workaround: Disable Catalog HA. In CDW, edit your Virtual Warehouse, and turn off Enable Impala Catalog Server HA.

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

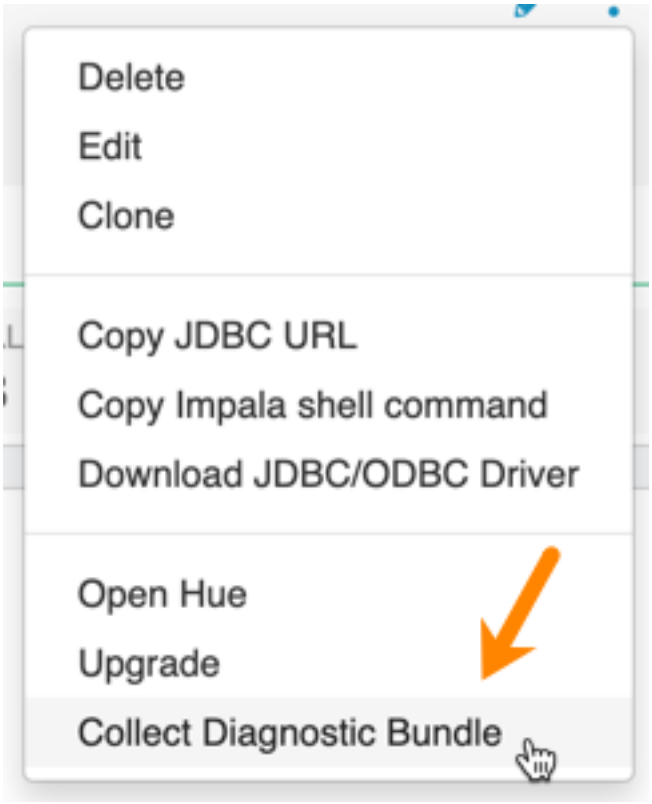
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

UI Issues

DWX-15110 Misaligned dialog in the UI

The Create Virtual Warehouse dialog is not aligned properly when no Virtual Warehouses are running.

DWX-14992 Cosmetic UI problem with Enable DAS link

In Azure environments, no gap appears between the previous text box and Enable DAS.

DWX-14970 Cosmetic UI problem with Virtual Warehouse start label

The start labels of the Impala Virtual Warehouse and Hive Virtual Warehouse are inconsistent with regard to upper- and lower-case.

DWX-14874 No indication that there are no Virtual Warehouses running

When no Data Visualization instance is running, you see "Currently no environments connected to Data Visualization cluster". The same indicator should be displayed when no Virtual Warehouses are running.

DWX-14610 Upgrade icon indicates the Database Catalog is the latest version, but might be incorrect for a few seconds

The cluster fetches data regularly in particular time intervals, or when an action is triggered. Inbetween this time interval, which is very brief, the cluster might not be updated, but will be refreshed in a few seconds.

Technical Service Bulletins

TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

The Spark Iceberg library includes two procedures - rewrite_data_files and rewrite_position_delete_files. The current implementation of rewrite_data_files has a limitation that the position delete files are not deleted and still tracked by the table metadata, even if they no longer refer to an active data file. This is called the dangling delete problem. To solve this, the rewrite_position_delete_files procedure is implemented in the Spark Iceberg library to remove these old “dangling” position delete files.

Due to the dangling delete limitation, when an Iceberg table with dangling deletes is queried in Impala, Impala tries to optimize select count(*) from iceberg_table query to return the results using stats. This optimization returns incorrect results.

The following conditions must be met for this issue to occur:

- All delete files in the Iceberg table are “dangling”
  - This would occur immediately after running Spark rewrite_data_files AND
    - Before any further delete operations are performed on the table OR
    - Before Spark rewrite_position_delete_files is run on the table
- Only stats optimized plain select count(*) from iceberg_table queries are affected. For example, the query should not have:
  - Any WHERE clause
  - Any GROUP BY clause
  - Any HAVING clause

For more information, see the Apache Iceberg documentation.

Remove dangling deletes: After rewrite_data_files, position delete records pointing to the rewritten data files are not always marked for removal, and can remain tracked by the live snapshot metadata of the table. This is known as the dangling delete problem.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables.

May 5, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

New known Issues in this release

TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

The Cloudera Data Warehouse (CDW) Public Cloud 2023.0.14.0 (DWX-1.6.3) version incorporated a feature for performing automatic metadata synchronization across multiple Apache Impala (Impala) Virtual Warehouses. The feature is enabled by default, and relies on the Hive MetaStore events. When a certain sequence of Data Definition Language (DDL) SQL commands are executed as described below, users may encounter a java.lang.NullPointerException (NPE). The exception causes the event processor to stop processing other metadata operations.

If a CREATE TABLE command (not CREATE TABLE AS SELECT) is followed immediately (approximately within 1 second interval) by INVALIDATE METADATA or REFRESH TABLE command on the same table (either on the same Virtual Warehouse or on a different one), there is a possibility that the second command will not find the table in the catalog cache of a peer Virtual Warehouse and generate an NPE.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-684: Automatic metadata synchronization across multiple Impala Virtual Warehouses (in CDW Public Cloud) may encounter an exception

Enabling a private CDW environment in Azure Kubernetes Service is unstable

Cloudera rolled back the General Availability (GA) status and support for deploying a private CDW environment in Azure using AKS until a solution for the issue is available.

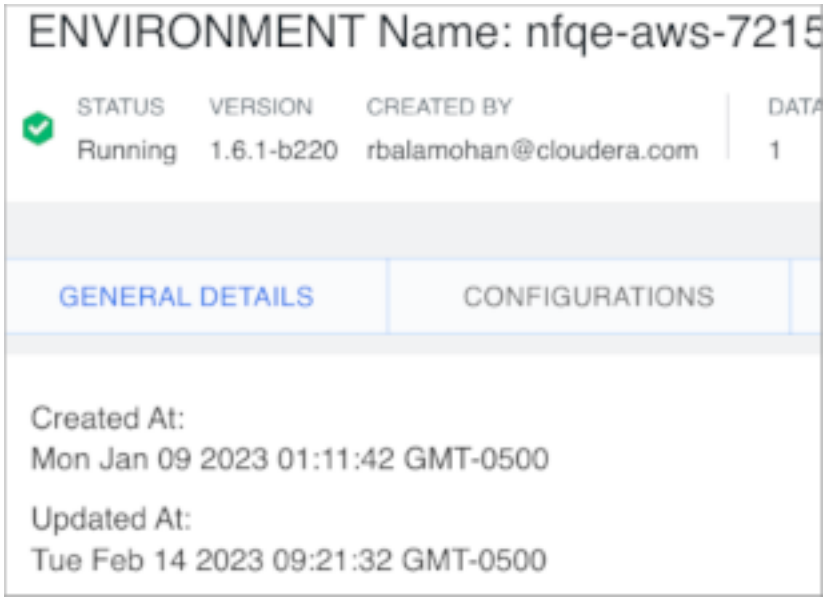
When working together with Microsoft Azure (Azure), a networking issue has been identified at Azure Software Definition Layer (SDN) that impacts the use of CDW private environments while using the Azure Kubernetes Service (AKS). This issue is highly unpredictable and may cause timeouts during Cloudera Data Warehouse (CDW) environment activation, Virtual Warehouse creation, Virtual Warehouse modification, and/or during start and stop operations in CDW.

The following users are affected: CDW users on Azure who are activating a private CDW environment by selecting the Private CDW option on the User Interface (UI) or using CDP Command Line Interface (CLI) enablePrivateAks option for the dw sub-command create-cluster operation within the --aws-options group.

Activation can fail with a timeout issue, and other operation related actions may be interrupted.

Certain CDW versions cannot upgrade to EKS 1.22

AWS environments activated in version 1.4.1-b86 (released June, 22, 2022) or earlier are not supported for upgrade to EKS 1.22 due to incompatibility of some components with EKS 1.22. To determine the activation version your pre-existing environment, in the Data Warehouse service, expand Environments. In Environments, search for and locate the environment that you want to view. Click Edit. In Environment Details, you see the CDW version.



DWX-15473 Reactivating the AWS environment in CDW fails

When activating the AWS environment in CDW, the usage monitor can fail due to a problem connecting to prometheus. The connection problem can be caused by a restrictive AWS account policy. The follow error message occurs:

```
User: arn:aws:sts::OBFUSCATED:assumed-role/env-OBFUSCATED-dwx-st
ack-NodeInstanceRole-OBFUSCATED/i-OBFUSCATED is not authorized to
 perform: elasticfilesystem:TagResource on the specified resource
```

Workaround: After reactivating the CDW Environment, navigate to associated Cloudformation template for CDW, click Resources. Select NodeInstanceRole, and then select **efs inline policy**. Replace the JSON policy with the policy on the kubernetes-sigs github site.

DWX-15395 Hive compaction fails in Azure

When RAZ is enabled in Cloudera Data Warehouse Azure environment, Hive compaction fails and returns an error message something like this:

```
ErrorMessage = FAILED: HiveAccessControlException Permission den
ied: user [ed64e526-a445-4945-9ed6-2eaf9560ff08] does not have [
READ] privilege on [[abfs:
```

Workaround:

1) In the CDW UI, in Overview, select a Hive Virtual Warehouse, and click Edit.

2) In Details, click CONFIGURATIONS HiveServer2 . 3) Select hive-site from the drop-down list, and search for metastore.compactor.run.as.user.

4) If the metastore.compactor.run.as.user key does not appear, click + and add it.

5) In Value, enter your Microsoft managed identity, and save the change.

Restart the Virtual Warehouse.

DWX-15246 Missing entries in subnet selection during CDW environment activation

The subnet selection during environment activation might not display any, or all, of the subnets registered in the environment.

Workaround: Carefully review the list of subnets you can select during activation. If subnets you want are not visible, enter the subnet-ids manually into the list, and then activate the environment.

DWX-15214 DWX-15176 Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

The Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed:

```
kubelet Error: failed to create containerd task: failed to creat
e shim task: OCI runtime create failed: runc create failed: unab
le to start container process: exec: “run_httpd.sh”: cannot run
executable found relative to current directory: unknown
```

This issue affects all CDW releases before 1.6.3 (released May 5, 2023)running on CDP Public Cloud on Azure.

The following users are affected:

- CDW Azure users using Hue in Impala and Hive Virtual Warehouses in environments earlier than version 1.6.3 if they upgrade the node image
- CDW Azure users creating a new CDW environment with a Hue version earlier than 2023.0.14.0 (released May 5, 2023) during Hive or Impala Virtual Warehouse creation

Workaround: Upgrade the Virtual Warehouses to 1.6.3 and choose Hue version 2023.0.14.0.

Do not choose Hue versions older than 2023.0.14.0 when creating a Virtual Warehouse in environments of version 1.6.3.

DWX-15016 Impala query failure when using Catalog high availability (HA)

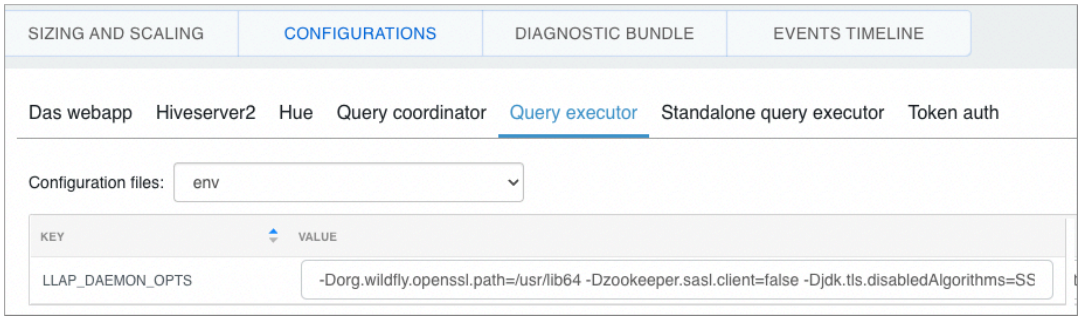
Impala queries could fail with InconsistentMetadataFetchException when using HA feature in CDW. This happens when leader election failover for Impala Catalog Service happens due to a transient network error.

Workaround: Disable Catalog HA. In CDW, edit your Virtual Warehouse, and turn off Enable Impala Catalog Server HA.

DWX-14452 Parquet table query might fail

Querying a table stored in Parquet from Hive might fail with the following exception message: java.lang.RuntimeException: java.lang.StackOverflowError. This problem can occur when the IN predicate has 243 values or more, and a small stack (-Xss = 256k) is configured for Hive in CDW.

Workaround: Change the value of the Xss in the JVM args to use the default value (1Mb) as follows: Select Edit from the Options menu of your Virtual Warehouse. Click Configurations > Query Executor > and select env.



In the third line shown below, change the value of LLAP_DAEMON_OPTS from -Xss256k to -Xss1M, and then click Apply Changes:

FROM:

-Dorg.wildfly.openssl.path=/usr/lib64 -Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Dhttp.maxConnections=12 -Xms24G -Xmx48G -Xss256k ...

TO:

... -Xss1M ...

DWX-15090: CSRF error intermittently seen in the Hue Job Browser

You may intermittently see the “403 - CSRF” error on the Hue web interface as well as in the Hue logs after running Hive queries from Hue.

Reload the browser or start a new Hue session.

DWX-15171 Error after disabling SSO.

Disabling SSO for Impala Virtual Warehouse can cause an error.

To reproduce this issue 1) Create an Impala Virtual Warehouse having SSO enabled. 2) Edit the Virtual Warehouse to disable SSO.

The coordinator crashes.

DWX-15151 Runtime error after activating an environment

Activating an environment having a public load balancer can cause a 2083 runtime error

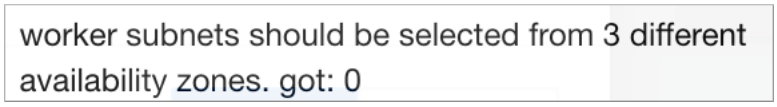
To reproduce this problem 1) Create a data lake. 2) Activate an enviroment having a public load balancer deployment type and subnets in three different availability zones. The following error can occur:

```
RuntimeErr with ErrCode=2083 (cause: 0 'coredns' replica(s) are
ready) Error Code : undefined
```

DWX-15145 Environment validation popup error after activating an environment

Activating an environment having a public load balancer can cause an environment validatation popup error.

To reproduce this problem 1) Create a data lake. 2) Activate an enviroment having a public load balancer deployment type and subnets in three different availability zones. A environment validation popup can occur even through subnets are in different availability zones. Several differnt popups can occur, including the following one:



DWX-15144 Virtual Warehouse naming restrictions

You cannot create a Virtual Warehouse having the same name as another Virtual Warehouse even if the like-named Virtual Warehouses are in different environments. You can create a Database Catalog having the same name as another Database Catalog if the Database Catalogs are in different environments.

DWX-15115 Error displayed after clicking on hyperlink below Hue table browser

In Hue, below the table browser, clicking the hyperlink to a location causes an HTTP 500 error because the file browser is not enabled for environments that are not Ranger authorized (RAZ).

DWX-15114 Incorrect Database Catalog status indication incorrect after an upgrade

After a Database Catalog upgrade, if the Databus producer crashes, the default Database Catalog does not indicate the error. The Database Catalog status instead indicates "running". After a Database Catalog upgrade, if the Databus producer crashes, a non-default (custom) Database Catalog, indicates the correct error status.

DWX-15112 Enterprise Data Warehouse database configuration problems after a Helm-related rollback

If a Helm rollback fails due to an incorrect Enterprise Data Warehouse database configuration, the Virtual Warehouse and Database Catalog roll back to a previous configuration. The incorrect Enterprise Data Warehouse configuration persists, and can affect subsequent edit, upgrade, and rebuild operations on the rolled-back Virtual Warehouse or Database Catalog.

DWX-15110 Misaligned dialog in the UI

The Create Virtual Warehouse dialog is not aligned properly when no Virtual Warehouses are running.

DWX-15064 Hive Virtual Warehouse stops but appears healthy

Due to an istio-proxy problem, the query coordinator can unexpectedly enter a not ready state instead of the expected error-state. Subsequently, the Hive Virtual Warehouse stops when reaching the autosuspend timeout without indicating a problem.

DWX-15014 Loading airports table in demo data fails

This issue occurs only when using a non-default Database Catalog. A invalid path error message occurs when using the load command to load the demo data airports table in Iceberg, ORC, or Parquet format.

DWX-14992 Cosmetic UI problem with Enable DAS link

In Azure environments, no gap appears between the previous text box and Enable DAS.

DWX-14970 Cosmetic UI problem with Virtual Warehouse start label

The start labels of the Impala Virtual Warehouse and Hive Virtual Warehouse are inconsistent with regard to upper- and lower-case.

DWX-14968 Connection termination error in Impala queries tab after Hue inactivity

To reproduce the problem: 1) In the Impala job browser, navigate to Impala queries. 2) Wait for a few minutes.

After a few minutes of inactivity, the follow error is displayed:

Workaround: Refresh the page, or alternatively start a new session.

DWX-14927 Hue fails to list Iceberg snapshots

Hue does not recognize the Iceberg history queries from Hive to list table snapshots. For example, Hue indicates an error at the . before history when you run the following query.

```
select * from <db_name>.<table_name>.history
```

Delay in listing queries in Impala Queries in the Job browser

Listing an Impala query in the Job browser can take an inordinate amount of time.

DWX-14923 After JWT authentication, ttempting to connect the Impyla client using a user name or password should cause an error

Using a JWT token, you can connect to a Virtual Warehouse as the user who generated the token.

If you connect with the JWT token, and then pass a user name or password from Impyla to the Virtual Warehouse, the connection arguments are silently ignored. Such an action should indicate that it is illegal to specify user or password when using JWT authentication.

DWX-14874 No indication that there are no Virtual Warehouses running

When no Data Visualization instance is running, you see "Currently no environments connected to Data Visualization cluster". The same indicator should be displayed when no Virtual Warehouses are running.

DWX-14610 Upgrade icon indicates the Database Catalog is the latest version, but might be incorrect for a few seconds

The cluster fetches data regularly in particular time intervals, or when an action is triggered. Inbetween this time interval, which is very brief, the cluster might not be updated, but will be refreshed in a few seconds.

CDPD-54376: Clicking the home button on the File Browser page redirects to HDFS user directory

When you are previewing a file on any supported filesystem, such as S3 or ABFS, and you click on the Home button, you are redirected to the HDFS user home directory instead of the user home directory on the said filesystem.

None.

Carried over from the previous release: General

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-6619 Browser auto-close not working on some browsers after token-Based authentication for accessing CDW

The Firefox and Edge browser window does not close automatically after successful authentication.

DWX-9774 Database Catalog or Virtual Warehouse image version problem

Background: In Cloudera Data Warehouse 2021.0.3-b27 - 2021.0.5-b36, you can choose any supported image version when you create a Database Catalog or Virtual Warehouse, assuming you have the CDW_VERSIONED_DEPLOY entitlement.

Problem: In Cloudera Data Warehouse 2021.0.6-b96, you can choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse only if you use an existing environment. An existing environment is one you created in 2021.0.3-b27 - 2021.0.5-b36.

Workaround: In 2021.0.6-b96, use an environment you created before this release to choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse.

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

DWX-14409 Starting a Virtual Warehouse in an AWS environments in the Asia-Pacific Southeast region fails

When you upgrade the Virtual Warehouse in this release and start it, a timing issue causes failure.

Workaround: Continue using the previous release 1.5.1-b110 (released Nov 22, 2022). Do not upgrade to 1.6.1-b282 (released Feb 7, 2023). Delay upgrading your Virtual Warehouse until the next release of CDW.

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the CreateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersions line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle. Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Database Catalog

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

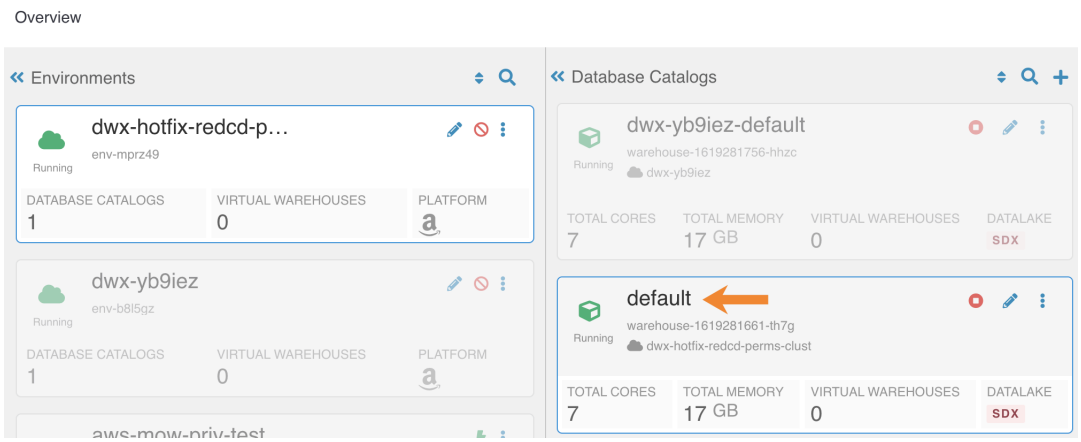
- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
       kind: ClusterRole
       metadata:
       name: diagnostic-data-generator-role-monitor
       rules:
       - apiGroups:
       - ""
       resources:
       - configmaps
       - events
       - pods
       - persistentvolumeclaims
       - nodes
       verbs:
       - get
       - list
       - apiGroups:
       - apps
       resources:
       - deployments
       - statefulsets
       verbs:
       - get
       - list
       - apiGroups:
       - "edws.cloudera.com"
       resources:
       - computes
       verbs:
       - get
       - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.
7. Click Apply in the upper right corner of the page to save the configuration.
8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name ‘ec2-18-219-
32-183.us-east-2.compute.amazonaws.com’ does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Hue Query Editor

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Iceberg

DWX-14163 Limitations reading Iceberg tables in Avro file format from Impala

The Avro, Impala, and Iceberg specifications describe some limitations related to Avro, and those limitations exist in CDP. In addition to these, the DECIMAL type is not supported in this release.

DWX-14286 Loading data to Iceberg error

Using LOAD DATA INPATH from an Impala Virtual Warehouse to load data to an Iceberg table on S3/ADLS when hidden files (files that are prefixed with . or _) are in the path can lead to unexpected query failures. For example:

```
AnalysisException: INPATH contains unsupported LOAD format, file
 '
            s3a://dwx-testdata/impala/sql_test/tests/load_data_in
path/runtime_data/0690a6fa9bfb11ed920c164053429bec/load_data_tes
t/A/impala_data/impala_alltypessmall_data/alltypessmall_parquet_
iceberg/year=2009/month=1/.hiddenfileforloaddatatest
            ' has 'This' magic string.
```

DEX-7946 Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata if you drop the table during migration from Hive to Iceberg.

Workarounds: Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13733 Timeout issue querying Iceberg tables from Hive

A timeout issue can cause the query to fail.

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

```
fs.s3.maxConnections=1000
fs.s3a.connection.maximum=1000
```

Restart the Database Catalog and Virtual Warehouse.

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: Impala Virtual Warehouse

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

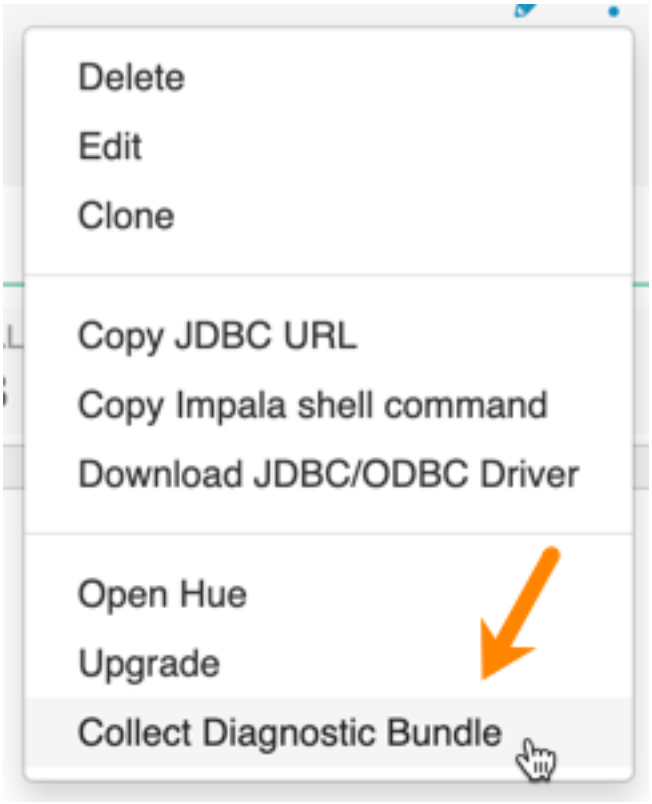
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes so the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

Technical Service Bulletins

TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables

The Spark Iceberg library includes two procedures - rewrite_data_files and rewrite_position_delete_files. The current implementation of rewrite_data_files has a limitation that the position delete files are not deleted and still tracked by the table metadata, even if they no longer refer to an active data file. This is called the dangling delete problem. To solve this, the rewrite_position_delete_files procedure is implemented in the Spark Iceberg library to remove these old “dangling” position delete files.

Due to the dangling delete limitation, when an Iceberg table with dangling deletes is queried in Impala, Impala tries to optimize select count(*) from iceberg_table query to return the results using stats. This optimization returns incorrect results.

The following conditions must be met for this issue to occur:

- All delete files in the Iceberg table are “dangling”
  - This would occur immediately after running Spark rewrite_data_files AND
    - Before any further delete operations are performed on the table OR
    - Before Spark rewrite_position_delete_files is run on the table

- Only stats optimized plain select count(*) from iceberg_table queries are affected. For example, the query should not have:
  - Any WHERE clause
  - Any GROUP BY clause
  - Any HAVING clause

For more information, see the Apache Iceberg documentation.

Remove dangling deletes: After rewrite_data_files, position delete records pointing to the rewritten data files are not always marked for removal, and can remain tracked by the live snapshot metadata of the table. This is known as the dangling delete problem.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-752: Dangling delete issue in Spark rewrite_data_files procedure causes incorrect results for Iceberg V2 tables.

February 7, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

Technical Service Bulletins

TSB-732 2024: Incorrect results are generated by Hive JOIN when bloom filter is activated

The bloom filter implemented in HIVE-23880 was designed to enhance performance for queries with JOIN statements, where one small table and another significantly larger table is joined on partition keys. However, the bloom filter introduced an issue in Apache Hive (Hive), when dynamic semijoin redaction is involved that generates incorrect query results. This issue is corrected in HIVE-26655.

Upstream JIRA

Hive-23880(cause)HIVE-26655(fix)

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-732: Incorrect results are generated by Hive JOIN when bloom filter is activated

New known Issues in this release

Do not upgrade this release to EKS 1.22

The latest version of Cloudera Data Warehouse (CDW) 1.6.1-b258 (released February 7, 2023) provisions AWS Elastic Kubernetes Service (EKS) 1.22 by default in AWS environments; however, upgrading your activated environments to EKS 1.22 or later is not supported. Do not upgrade your EKS clusters to 1.22.

DWX-14163 Limitations reading Iceberg tables in Avro file format from Impala

The Avro, Impala, and Iceberg specifications describe some limitations related to Avro, and those limitations exist in CDP. In addition to these, the DECIMAL type is not supported in this release.

DWX-14286 Loading data to Iceberg error

Using LOAD DATA INPATH from an Impala Virtual Warehouse to load data to an Iceberg table on S3/ADLS when hidden files (files that are prefixed with . or _) are in the path can lead to unexpected query failures. For example:

```
AnalysisException: INPATH contains unsupported LOAD format, file
 '
s3a://dwx-testdata/impala/sql_test/tests/load_data_inpath/runtime
_data/0690a6fa9bfb11ed920c164053429bec/load_data_test/A/impala_d
ata/impala_alltypessmall_data/alltypessmall_parquet_iceberg/year
=2009/month=1/.hiddenfileforloaddatatest
```

```
' has 'This' magic string.
```

Hive can crash due to the default setting of a caching property

Background: The relatively new hive.fetch.task.caching property was designed to prevent files from being cleaned up prematurely during the fetch phase. By default, hive.fetch.task.caching is enabled for queries that have been converted by the hive.fetch.task.conversion optimization.

Problem: The resulting caching activity can use too much heap while running queries that execute fetch tasks with a cache size of 1GB. Frequent garbage collection and high CPU usage can occur, causing Hue and other clients to timeout when returning results.

Workaround: Try setting hive.fetch.task.caching to false to disable the optimization. This action might cause queries to fail due to timing if the table is modified concurrent with the fetching. If problems continue, or cannot tolerate the low risk of a query failure, set hive.fetch.task.conversion to none. In this release, hive.fetch.task.conversion=none might add a little latency to some SELECT queries.

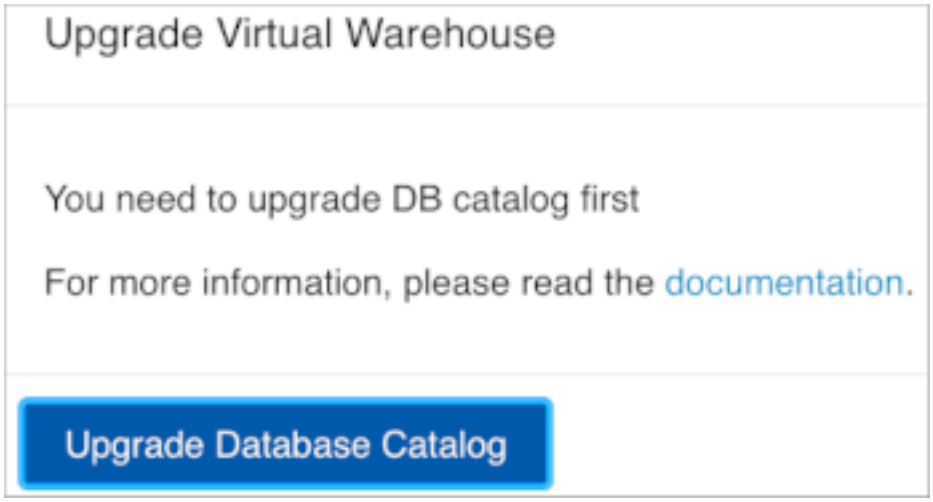
DWX-14409 Starting a Virtual Warehouse in an AWS environments in the Asia-Pacific Southeast region fails

When you upgrade the Virtual Warehouse in this release and start it, a timing issue causes failure.

Workaround: Continue using the previous release 1.5.1-b110 (released Nov 22, 2022). Do not upgrade to 1.6.1-b282 (released Feb 7, 2023). Delay upgrading your Virtual Warehouse until the next release of CDW.

DWX-14376 Upgrade Database Catalog button issue

A UI issue exists when you to upgrade the latest Virtual Warehouse 1.6.1-b258 (released Feb 7, 2023) and you are already running the latest release. The Upgrade Database Catalog button is not working when you attempt this futile Virtual Warehouse upgrade.



Carried over from the previous release: General

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-6619 Browser auto-close not working on some browsers after token-Based authentication for accessing CDW

The Firefox and Edge browser window does not close automatically after successful authentication.

DWX-9774 Database Catalog or Virtual Warehouse image version problem

Background: In Cloudera Data Warehouse 2021.0.3-b27 - 2021.0.5-b36, you can choose any supported image version when you create a Database Catalog or Virtual Warehouse, assuming you have the CDW_VERSIONED_DEPLOY entitlement.

Problem: In Cloudera Data Warehouse 2021.0.6-b96, you can choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse only if you use an existing environment. An existing environment is one you created in 2021.0.3-b27 - 2021.0.5-b36.

Workaround: In 2021.0.6-b96, use an environment you created before this release to choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

Workloads from CDW versions 1.4.2 and earlier cannot be deployed on new AWS environments

You must upgrade the version of Kubernetes that supports your existing CDW clusters to EKS version 1.21. AWS environments you activate using this release of Cloudera Data Warehouse, and later, will use version 1.22.

Because new environments are provisioned automatically to use AWS Elastic Kubernetes Service (EKS) 1.22, deprecated APIs used in workloads of versions 1.4.2-b118 (released August 4, 2022) and earlier are not supported in this version 1.5.1-b110 (released November 22, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads or use workloads from the previous release CDW 1.4.3--b225 (released September 15, 2022). Workloads from CDW 1.4.2-b118 (released August 4, 2022) and

earlier in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with EKS 1.22.

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the Cr eateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.

6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersions line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle. Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Database Catalog

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

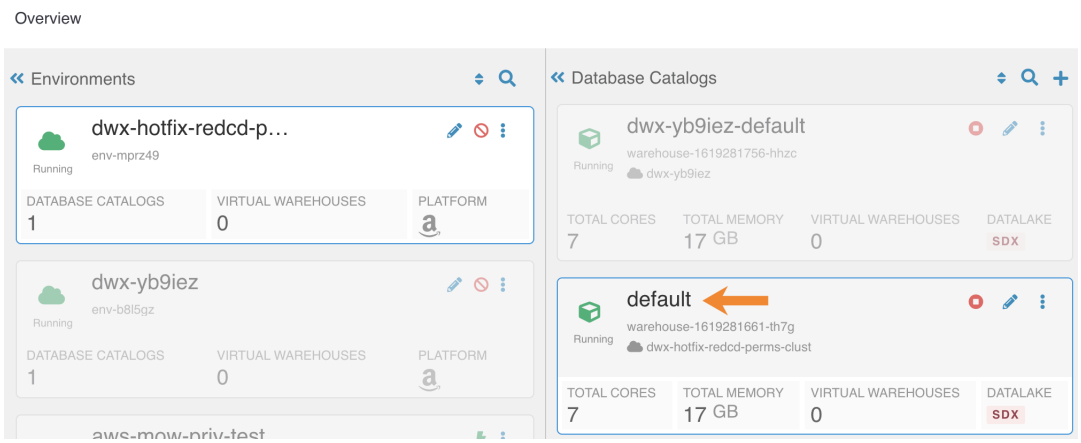
- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
       kind: ClusterRole
       metadata:
       name: diagnostic-data-generator-role-monitor
       rules:
       - apiGroups:
       - ""
       resources:
       - configmaps
       - events
       - pods
       - persistentvolumeclaims
       - nodes
       verbs:
       - get
       - list
       - apiGroups:
       - apps
       resources:
       - deployments
       - statefulsets
       verbs:
       - get
       - list
       - apiGroups:
       - "edws.cloudera.com"
       resources:
       - computes
       verbs:
       - get
       - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.par quet.write.int64.timestamp=true.

DWX-10271: Missing log section in Hive query results

In a Hive Virtual Warehouse, when you run a query in Hue, the query results do not contain a logs section.

DWX-8118: INSERT INTO command fails under certain circumstances

This problem affects users who have a PostgreSQL database as the backend Hive database. If you create a table A and create a table B as select (CTAS) from an empty table A, inserting values into table B fails as follows:

```
Error while compiling statement: FAILED: Execution Error, return
 code 1 from
       org.apache.hadoop.hive.ql.exec.StatsTask.org.apache.thr
ift.transport.TTransportException
```

Workaround: Disable auto-stats gathering: Go to Cloudera Manager Data Warehouse Virtual Warehouses , and click the Hive VW name in the list. In Configuration HiveServer2 , set hive.cbo .enable to false.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.

7. Click Apply in the upper right corner of the page to save the configuration.

8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name 'ec2-18-219-
32-183.us-east-2.compute.amazonaws.com' does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Hue Query Editor

TSB 2023-670: Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed: kubelet Error: failed to create containerd task: failed to create shim task: OCI runtime create failed: runc create failed: unable to start container process: exec: “run_httpd.sh”: cannot run executable found relative to current directory: unknown

The root cause of the issue is a recent node image upgrade in Azure made by Microsoft, where the version of the container runtime (containerd) was upgraded to 1.6.18. This version upgraded Go SDK to 1.19.6, where the behavior of how program execution works has changed. Due to security concerns, the newer version of Go SDK does not resolve a program using an implicit or explicit path entry relative to the current directory (for more information, see the Go documentation), and containerd indirectly uses this Go exec API. The Docker command of the Hue frontend relies on the former behavior. Therefore the command fails on recent node images, because it cannot find the expected executable in the current working directory of the containers anymore. Cloudera first noticed the aforementioned containerd version in the Azure Kubernetes Service (AKS) released on 2023-03-05. Every node image which is newer than the version released on 2023-03-05 is affected by this issue. Unfortunately, Cloudera does not have any insight to when Microsoft rolls out new node images in a given region. Therefore, it is possible that in some regions, the older node images are still in use, where the issue does not arise until a given regional update is applied to the node images.

CDW environments, which are already activated and running, are not affected as long as customers do not trigger a node image upgrade to the latest available version either using Azure Command Line Interface (CLI) or on the Azure portal. If the node image is upgraded to the latest version,

then the Virtual Warehouses also need to be upgraded to the latest CDWH-2023.0.14.0 version. Newly created CDW environments are not affected, but customers are advised to not choose a Hue version lower than CDWH-2023.0.14.0 during any Virtual Warehouse creation because such a configuration is affected by this issue.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-670: TSB Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Iceberg

DEX-7946 Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata if you drop the table during migration from Hive to Iceberg.

Workarounds: Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13733 Timeout issue querying Iceberg tables from Hive

A timeout issue can cause the query to fail.

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

```
fs.s3.maxConnections=1000
fs.s3a.connection.maximum=1000
```

Restart the Database Catalog and Virtual Warehouse.

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: Impala Virtual Warehouse

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

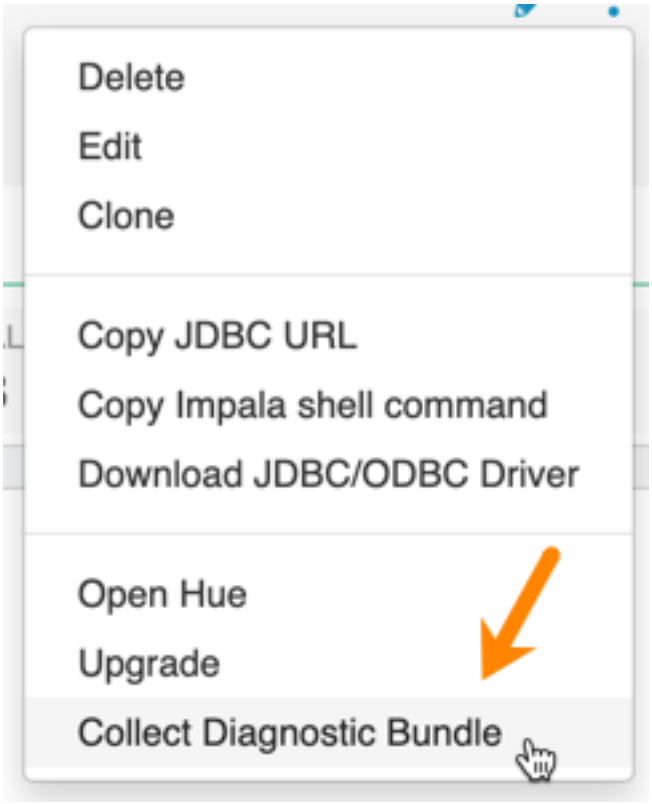
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case so the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

February 14, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

Technical Service Bulletins

TSB-732 2024: Incorrect results are generated by Hive JOIN when bloom filter is activated

The bloom filter implemented in HIVE-23880 was designed to enhance performance for queries with JOIN statements, where one small table and another significantly larger table is joined on partition keys. However, the bloom filter introduced an issue in Apache Hive (Hive), when dynamic semijoin redaction is involved that generates incorrect query results. This issue is corrected in HIVE-26655.

Upstream JIRA

Hive-23880(cause)HIVE-26655(fix)

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-732: Incorrect results are generated by Hive JOIN when bloom filter is activated

December 13, 2022

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

Technical Service Bulletins

CDP CLI CDW command issue when activating an Azure DWX environment

When clients try to activate a Microsoft Azure (Azure) environment on Cloudera Data Platform (CDP) Cloudera Data Warehouse (CDW) for Public Cloud using the CDP Command Line Interface (CLI) dw command's create-cluster sub-command with the --use-private-load-balancer switch, the created load balancer will be public instead of private.

Example command:

cdp \

--profile ${PROFILE} \

dw create-cluster \

--environment-crn ${ENV} \

--use-private-load-balancer

For the latest update on this issue see the corresponding Knowledge article:

TSB 2022-642: CDP CLI CDW command issue when activating an Azure DWX environment

TSB-732 2024: Incorrect results are generated by Hive JOIN when bloom filter is activated

The bloom filter implemented in HIVE-23880 was designed to enhance performance for queries with JOIN statements, where one small table and another significantly larger table is joined on partition keys. However, the bloom filter introduced an issue in Apache Hive (Hive), when dynamic semijoin redaction is involved that generates incorrect query results. This issue is corrected in HIVE-26655.

Upstream JIRA

Hive-23880(cause)HIVE-26655(fix)

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-732: Incorrect results are generated by Hive JOIN when bloom filter is activated

Carried over from the previous release: General

Do not upgrade this release to EKS 1.22

The latest version of Cloudera Data Warehouse (CDW) 1.5.1-b110 (released November 22, 2022) provisions AWS Elastic Kubernetes Service (EKS) 1.22 by default in Azure environments; however, upgrading your activated environments to EKS 1.22 or later is not supported. Do not upgrade your EKS clusters to 1.22.

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-6619 Browser auto-close not working on some browsers after token-Based authentication for accessing CDW

The Firefox and Edge browser window does not close automatically after successful authentication.

DWX-9774 Database Catalog or Virtual Warehouse image version problem

Background: In Cloudera Data Warehouse 2021.0.3-b27 - 2021.0.5-b36, you can choose any supported image version when you create a Database Catalog or Virtual Warehouse, assuming you have the CDW_VERSIONED_DEPLOY entitlement.

Problem: In Cloudera Data Warehouse 2021.0.6-b96, you can choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse only if you use an existing environment. An existing environment is one you created in 2021.0.3-b27 - 2021.0.5-b36.

Workaround: In 2021.0.6-b96, use an environment you created before this release to choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

Workloads from CDW versions 1.4.2 and earlier cannot be deployed on new AWS environments

You must upgrade the version of Kubernetes that supports your existing CDW clusters to EKS version 1.21. AWS environments you activate using this release of Cloudera Data Warehouse, and later, will use version 1.22.

Because new environments are provisioned automatically to use AWS Elastic Kubernetes Service (EKS) 1.22, deprecated APIs used in workloads of versions 1.4.2-b118 (released August 4, 2022) and earlier are not supported in this version 1.5.1-b110 (released November 22, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads or use workloads from the previous release CDW 1.4.3--b225 (released September 15, 2022). Workloads from CDW 1.4.2-b118 (released August 4, 2022) and earlier in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with EKS 1.22.

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the Cr eateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersio ns line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
```

```
            "autoscaling:DescribeAutoScalingGroups",
            "autoscaling:DescribeAutoScalingInstances",
            "autoscaling:DescribeTags",
            "autoscaling:DescribeLaunchConfigurations",
            "autoscaling:SetDesiredCapacity",
            "autoscaling:TerminateInstanceInAutoScalingGroup",
            "ec2:DescribeLaunchTemplateVersions",
            "ec2:CreateTags"
            ],
```

7. Save changes.

Carried over from the previous release: Azure

DWX-13911 Problem activating Azure clusters from CDW using the CDP CLI

Activating an Azure environment from the CDP CLI 0.9.71 using the use-private-load-balancer switch creates a public instead of a private load-balancer. The problem occurs only when using version 0.9.71.

Affected command: cdp dw create-cluster

Example command:

```
cdp \
  --profile ${PROFILE} \
  dw create-cluster \
  --environment-crn ${ENV} \
  --use-private-load-balancer
```

Workaround: Add the following line to the end of the command to create the private load balancer:

```
--aws-options privateSubnetIds=dummy
```

For example, activate the environment using the following command:

```
cdp \
  --profile ${PROFILE} \
  dw create-cluster \
  --environment-crn ${ENV} \
  --use-private-load-balancer \
  --aws-options privateSubnetIds=dummy
```

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Database Catalog

DWX-8980 Data consistency problem occurs when sharing a Database Catalog

When you refresh, or invalidate, metadata in the Database Catalog from one Virtual Warehouse, the coordinators of any other Virtual Warehouses that share the Database Catalog might not receive the updates fast enough. Users can experience inconsistency in query results.

Workaround: Set the timeout for invalidating the tables to a very low value by following the steps below. Impala is forced to read files from the object store more frequently than every 10 minutes (the default timeout). Metadata is refreshed more frequently than normal, which can impact performance. The impact is proportional to the number of small files Impala must read to load metadata.

1) Log in to the CDP web interface and navigate to the Data Warehouse service.

2) In the Data Warehouse service, click Overview, select the first Virtual Warehouse, and click Edit.

3) In Configurations, in Impala catalogd, click +.

4) In Add Custom Configurations, set the invalidate_tables_timeout_s to a small value, and click Add.

The catalogd daemon caches metadata for tables and must be refreshed/invalidated quickly to prevent inconsistency.

5) In Impala coordinator, click +.

6) In Add Custom Configurations, set the invalidate_tables_timeout_s to a small value, and click Add.

The coordinator must also refresh/invalidate data quickly to prevent inconsistency.

7) Repeat these steps for each Virtual Warehouse that shares the same Database Catalog.

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).

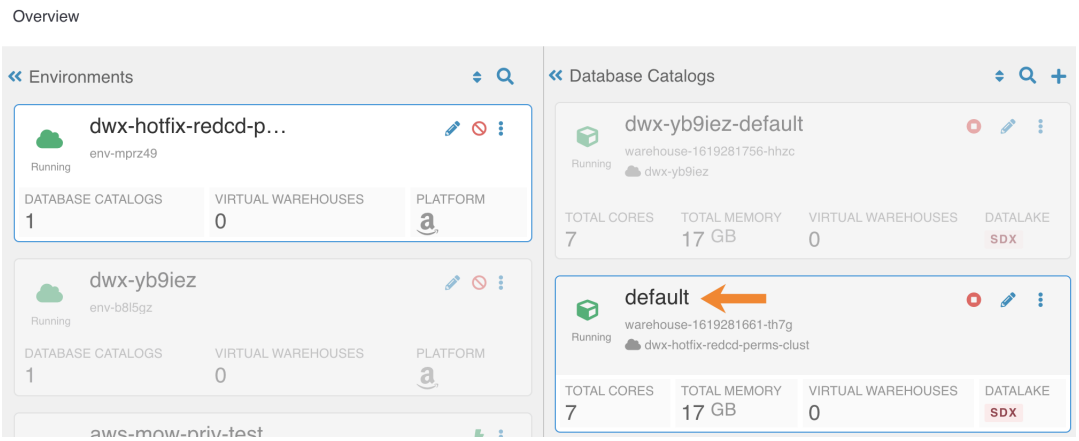
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
       kind: ClusterRole
       metadata:
       name: diagnostic-data-generator-role-monitor
```

```
        rules:
        - apiGroups:
        - ""
        resources:
        - configmaps
        - events
        - pods
        - persistentvolumeclaims
        - nodes
        verbs:
        - get
        - list
        - apiGroups:
        - apps
        resources:
        - deployments
        - statefulsets
        verbs:
        - get
        - list
        - apiGroups:
        - "edws.cloudera.com"
        resources:
        - computes
        verbs:
        - get
        - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.par quet.write.int64.timestamp=true.

DWX-10271: Missing log section in Hive query results

In a Hive Virtual Warehouse, when you run a query in Hue, the query results do not contain a logs section.

DWX-8118: INSERT INTO command fails under certain circumstances

This problem affects users who have a PostgreSQL database as the backend Hive database. If you create a table A and create a table B as select (CTAS) from an empty table A, inserting values into table B fails as follows:

```
Error while compiling statement: FAILED: Execution Error, return
 code 1 from
       org.apache.hadoop.hive.ql.exec.StatsTask.org.apache.thr
ift.transport.TTransportException
```

Workaround: Disable auto-stats gathering: Go to Cloudera Manager Data Warehouse Virtual Warehouses , and click the Hive VW name in the list. In Configuration HiveServer2 , set hive.cbo .enable to false.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses

must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.
7. Click Apply in the upper right corner of the page to save the configuration.

8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name ‘ec2-18-219-
32-183.us-east-2.compute.amazonaws.com’ does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Hue Query Editor

TSB 2023-670: Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed: kubelet Error: failed to create containerd task: failed to create shim task: OCI runtime create failed: runc create failed: unable to start container process: exec: “run_httpd.sh”: cannot run executable found relative to current directory: unknown

The root cause of the issue is a recent node image upgrade in Azure made by Microsoft, where the version of the container runtime (containerd) was upgraded to 1.6.18. This version upgraded Go SDK to 1.19.6, where the behavior of how program execution works has changed. Due to security concerns, the newer version of Go SDK does not resolve a program using an implicit or explicit path entry relative to the current directory (for more information, see the Go documentation), and containerd indirectly uses this Go exec API. The Docker command of the Hue frontend relies on the former behavior. Therefore the command fails on recent node images, because it cannot find the expected executable in the current working directory of the containers anymore. Cloudera first noticed the aforementioned containerd version in the Azure Kubernetes Service (AKS) released on 2023-03-05. Every node image which is newer than the version released on 2023-03-05 is affected by this issue. Unfortunately, Cloudera does not have any insight to when Microsoft rolls out new node images in a given region. Therefore, it is possible that in some regions, the older node images are still in use, where the issue does not arise until a given regional update is applied to the node images.

CDW environments, which are already activated and running, are not affected as long as customers do not trigger a node image upgrade to the latest available version either using Azure Command Line Interface (CLI) or on the Azure portal. If the node image is upgraded to the latest version,

then the Virtual Warehouses also need to be upgraded to the latest CDWH-2023.0.14.0 version. Newly created CDW environments are not affected, but customers are advised to not choose a Hue version lower than CDWH-2023.0.14.0 during any Virtual Warehouse creation because such a configuration is affected by this issue.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-670: TSB Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Iceberg

DEX-7946 Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata is you drop the table during migration from Hive to Iceberg.

Workarounds: Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13733 Timeout issue querying Iceberg tables from Hive

A timeout issue can cause the query to fail.

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

```
fs.s3.maxConnections=1000
fs.s3a.connection.maximum=1000
```

Restart the Database Catalog and Virtual Warehouse.

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: Impala Virtual Warehouse

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

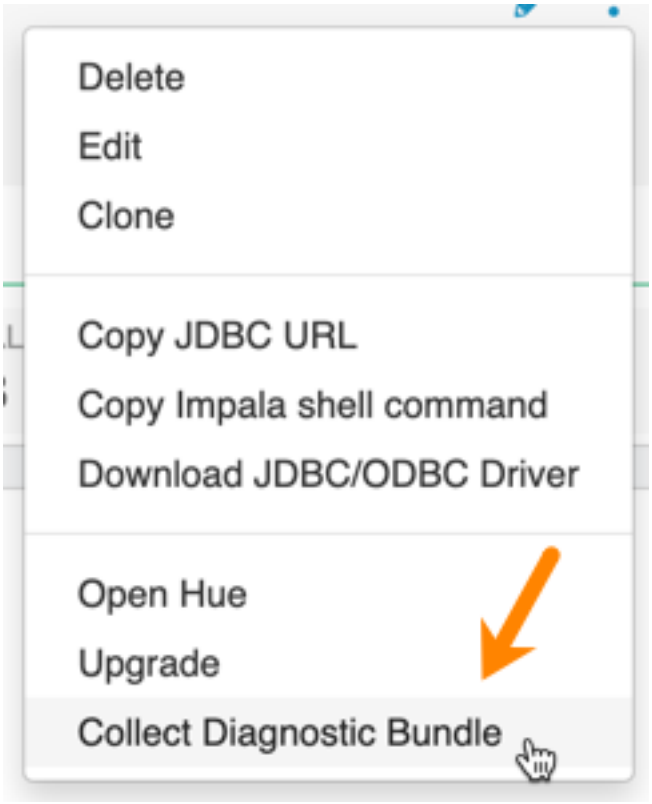
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case so the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

November 21, 2022

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

Technical Service Bulletins

TSB-732 2024: Incorrect results are generated by Hive JOIN when bloom filter is activated

The bloom filter implemented in HIVE-23880 was designed to enhance performance for queries with JOIN statements, where one small table and another significantly larger table is joined on partition keys. However, the bloom filter introduced an issue in Apache Hive (Hive), when dynamic semijoin redaction is involved that generates incorrect query results. This issue is corrected in HIVE-26655.

Upstream JIRA

Hive-23880(cause)HIVE-26655(fix)

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-732: Incorrect results are generated by Hive JOIN when bloom filter is activated

New known Issues in this release

Do not upgrade this release to EKS 1.22

The latest version of Cloudera Data Warehouse (CDW) 1.5.1-b110 (released November 22, 2022) provisions AWS Elastic Kubernetes Service (EKS) 1.22 by default in AWS environments; however,

upgrading your activated environments to EKS 1.22 or later is not supported. Do not upgrade your EKS clusters to 1.22.

Workloads from CDW versions 1.4.2 and earlier cannot be deployed on new AWS environments

You must upgrade the version of Kubernetes that supports your existing CDW clusters to EKS version 1.21. AWS environments you activate using this release of Cloudera Data Warehouse, and later, will use version 1.22.

Because new environments are provisioned automatically to use AWS Elastic Kubernetes Service (EKS) 1.22, deprecated APIs used in workloads of versions 1.4.2-b118 (released August 4, 2022) and earlier are not supported in this version 1.5.1-b110 (released November 22, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads or use workloads from the previous release CDW 1.4.3--b225 (released September 15, 2022). Workloads from CDW 1.4.2-b118 (released August 4, 2022) and earlier in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with EKS 1.22.

DWX-13911 Problem activating Azure clusters from CDW using the CDP CLI

Activating an Azure environment from the CDP CLI 0.9.71 using the use-private-load-balancer switch creates a public instead of a private load-balancer. The problem occurs only when using version 0.9.71.

Affected command: cdp dw create-cluster

Example command:

```
cdp \
  --profile ${PROFILE} \
  dw create-cluster \
  --environment-crn ${ENV} \
  --use-private-load-balancer
```

Workaround: Add the following line to the end of the command to create the private load balancer:

```
--aws-options privateSubnetIds=dummy
```

For example, activate the environment using the following command:

```
cdp \
  --profile ${PROFILE} \
  dw create-cluster \
  --environment-crn ${ENV} \
  --use-private-load-balancer \
  --aws-options privateSubnetIds=dummy
```

DEX-7946 Data loss during migration of a Hive table to Iceberg

In this release, by default the table property 'external.table.purge' is set to true, which deletes the table data and metadata is you drop the table during migration from Hive to Iceberg.

Workarounds: Either one of the following workarounds prevents data loss during table migration:

- Set the table property 'external.table.purge'='FALSE'.
- Do not drop a table during migration from Hive to Iceberg.

DWX-13733 Timeout issue querying Iceberg tables from Hive

A timeout issue can cause the query to fail.

Workaround: Add the following configuration to hadoop-core-site for the Database Catalog and the Virtual Warehouse.

```
fs.s3.maxConnections=1000
```

```
fs.s3a.connection.maximum=1000
```

Restart the Database Catalog and Virtual Warehouse.

Carried over from the previous release: General

TSB 2023-656: TSB Cloudera Data Warehouse does not clean up old Helm releases

Cloudera Data Warehouse (CDW) uses Helm release manager to deploy component releases into the Kubernetes cluster of the customer. However, CDW's current behavior in CDW keeps the Helm release history without any boundary. Since the storage for Helm in CDW uses secrets, this behavior can lead to a very high number of secrets (Impala/Hive scale-up/down creates new releases;clusters were seen with 6000+ secrets). Therefore, parsing and decoding a high number of Helm releases can require a significant amount of memory. As an example, for a cluster with 6000+ secrets, it could be 6-7 GB or more of memory utilized. This may result in performance degradation, instability and/or higher cloud cost because several CDW cluster components rely on secrets, which would cause the memory footprint of these other components to increase significantly.

This issue is addressed in the CDW version 1.5.1-b110 (workload image version 2022.0.11.0-122). Cloudera highly recommends upgrading Apache Impala (Impala) and Apache Hive (Hive) in CDW to 2022.0.11.0-122 or higher versions to ensure performance, stability and/or cloud cost are not impacted by this behavior.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-656: https://my.cloudera.com/knowledge/TSB-2023-656-Cloudera-Data-Warehouse-does-not-clean-up-old?id=368733

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-6619 Browser auto-close not working on some browsers after token-Based authentication for accessing CDW

The Firefox and Edge browser window does not close automatically after successful authentication.

DWX-9774 Database Catalog or Virtual Warehouse image version problem

Background: In Cloudera Data Warehouse 2021.0.3-b27 - 2021.0.5-b36, you can choose any supported image version when you create a Database Catalog or Virtual Warehouse, assuming you have the CDW_VERSIONED_DEPLOY entitlement.

Problem: In Cloudera Data Warehouse 2021.0.6-b96, you can choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse only if you use an existing environment. An existing environment is one you created in 2021.0.3-b27 - 2021.0.5-b36.

Workaround: In 2021.0.6-b96, use an environment you created before this release to choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the CreateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersions line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle. Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Database Catalog

DWX-8980 Data consistency problem occurs when sharing a Database Catalog

When you refresh, or invalidate, metadata in the Database Catalog from one Virtual Warehouse, the coordinators of any other Virtual Warehouses that share the Database Catalog might not receive the updates fast enough. Users can experience inconsistency in query results.

Workaround: Set the timeout for invalidating the tables to a very low value by following the steps below. Impala is forced to read files from the object store more frequently than every 10 minutes (the default timeout). Metadata is refreshed more frequently than normal, which can impact performance. The impact is proportional to the number of small files Impala must read to load metadata.

1) Log in to the CDP web interface and navigate to the Data Warehouse service.

2) In the Data Warehouse service, click Overview, select the first Virtual Warehouse, and click Edit.

3) In Configurations, in Impala catalogd, click +.

4) In Add Custom Configurations, set the invalidate_tables_timeout_s to a small value, and click Add.

The catalogd daemon caches metadata for tables and must be refreshed/invalidated quickly to prevent inconsistency.

5) In Impala coordinator, click +.

6) In Add Custom Configurations, set the invalidate_tables_timeout_s to a small value, and click Add.

The coordinator must also refresh/invalidate data quickly to prevent inconsistency.

7) Repeat these steps for each Virtual Warehouse that shares the same Database Catalog.

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).

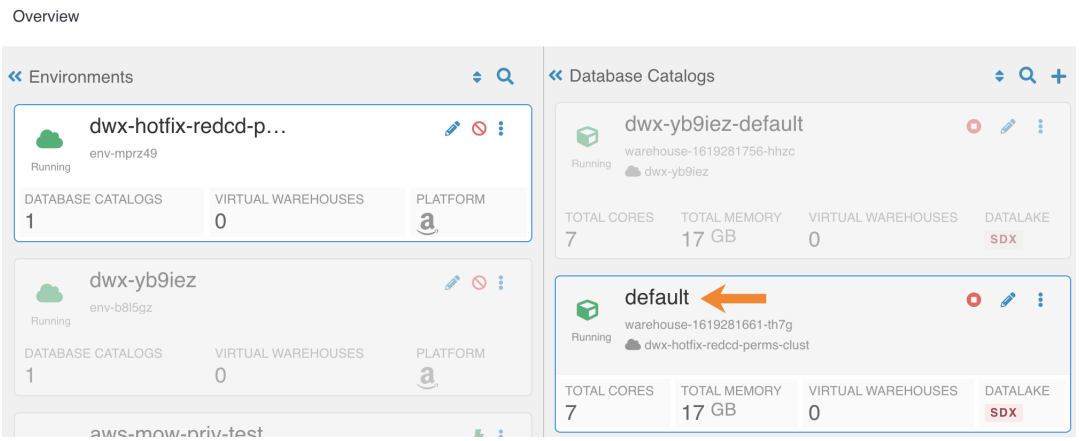
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
       kind: ClusterRole
       metadata:
       name: diagnostic-data-generator-role-monitor
```

```
      rules:
      - apiGroups:
      - ""
      resources:
      - configmaps
      - events
      - pods
      - persistentvolumeclaims
      - nodes
      verbs:
      - get
      - list
      - apiGroups:
      - apps
      resources:
      - deployments
      - statefulsets
      verbs:
      - get
      - list
      - apiGroups:
      - "edws.cloudera.com"
      resources:
      - computes
      verbs:
      - get
      - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.par quet.write.int64.timestamp=true.

DWX-10271: Missing log section in Hive query results

In a Hive Virtual Warehouse, when you run a query in Hue, the query results do not contain a logs section.

DWX-8118: INSERT INTO command fails under certain circumstances

This problem affects users who have a PostgreSQL database as the backend Hive database. If you create a table A and create a table B as select (CTAS) from an empty table A, inserting values into table B fails as follows:

```
Error while compiling statement: FAILED: Execution Error, return
 code 1 from
       org.apache.hadoop.hive.ql.exec.StatsTask.org.apache.thr
ift.transport.TTransportException
```

Workaround: Disable auto-stats gathering: Go to Cloudera Manager Data Warehouse Virtual Warehouses , and click the Hive VW name in the list. In Configuration HiveServer2 , set hive.cbo .enable to false.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses

must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.

7. Click Apply in the upper right corner of the page to save the configuration.

8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name ‘ec2-18-219-
32-183.us-east-2.compute.amazonaws.com’ does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Hue Query Editor

TSB 2023-670: Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed: kubelet Error: failed to create containerd task: failed to create shim task: OCI runtime create failed: runc create failed: unable to start container process: exec: “run_httpd.sh”: cannot run executable found relative to current directory: unknown

The root cause of the issue is a recent node image upgrade in Azure made by Microsoft, where the version of the container runtime (containerd) was upgraded to 1.6.18. This version upgraded Go SDK to 1.19.6, where the behavior of how program execution works has changed. Due to security concerns, the newer version of Go SDK does not resolve a program using an implicit or explicit path entry relative to the current directory (for more information, see the Go documentation), and containerd indirectly uses this Go exec API. The Docker command of the Hue frontend relies on the former behavior. Therefore the command fails on recent node images, because it cannot find the expected executable in the current working directory of the containers anymore. Cloudera first noticed the aforementioned containerd version in the Azure Kubernetes Service (AKS) released on 2023-03-05. Every node image which is newer than the version released on 2023-03-05 is affected by this issue. Unfortunately, Cloudera does not have any insight to when Microsoft rolls out new node images in a given region. Therefore, it is possible that in some regions, the older node images are still in use, where the issue does not arise until a given regional update is applied to the node images.

CDW environments, which are already activated and running, are not affected as long as customers do not trigger a node image upgrade to the latest available version either using Azure Command Line Interface (CLI) or on the Azure portal. If the node image is upgraded to the latest version,

then the Virtual Warehouses also need to be upgraded to the latest CDWH-2023.0.14.0 version. Newly created CDW environments are not affected, but customers are advised to not choose a Hue version lower than CDWH-2023.0.14.0 during any Virtual Warehouse creation because such a configuration is affected by this issue.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-670: TSB Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Iceberg

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: Impala Virtual Warehouse

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses

must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

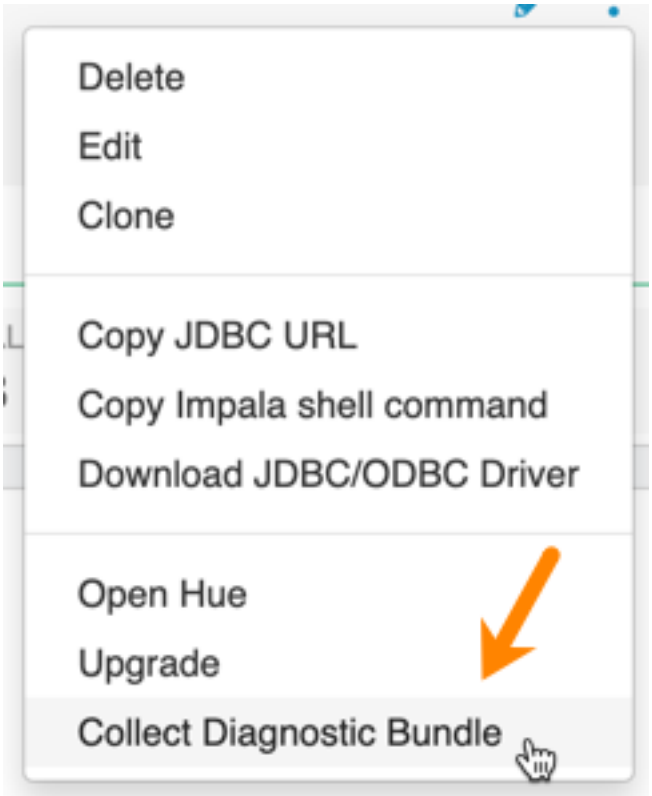
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case so the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

September 15, 2022

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

Technical Service Bulletins

TSB-732 2024: Incorrect results are generated by Hive JOIN when bloom filter is activated

The bloom filter implemented in HIVE-23880 was designed to enhance performance for queries with JOIN statements, where one small table and another significantly larger table is joined on partition keys. However, the bloom filter introduced an issue in Apache Hive (Hive), when dynamic semijoin redaction is involved that generates incorrect query results. This issue is corrected in HIVE-26655.

Upstream JIRA

Hive-23880(cause)HIVE-26655(fix)

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2024-732: Incorrect results are generated by Hive JOIN when bloom filter is activated

New known Issues in this release

DWX-8980 Data consistency problem occurs when sharing a Database Catalog

When you refresh, or invalidate, metadata in the Database Catalog from one Virtual Warehouse, the coordinators of any other Virtual Warehouses that share the Database Catalog might not receive the updates fast enough. Users can experience inconsistency in query results.

Workaround: Set the timeout for invalidating the tables to a very low value by following the steps below. Impala is forced to read files from the object store more frequently than every 10 minutes (the default timeout). Metadata is refreshed more frequently than normal, which can impact performance. The impact is proportional to the number of small files Impala must read to load metadata.

1) Log in to the CDP web interface and navigate to the Data Warehouse service.

2) In the Data Warehouse service, click Overview, select the first Virtual Warehouse, and click Edit.

3) In Configurations, in Impala catalogd, click +.

4) In Add Custom Configurations, set the invalidate_tables_timeout_s to a small value, and click Add.

The catalogd daemon caches metadata for tables and must be refreshed/invalidated quickly to prevent inconsistency.

5) In Impala coordinator, click +.

6) In Add Custom Configurations, set the invalidate_tables_timeout_s to a small value, and click Add.

The coordinator must also refresh/invalidate data quickly to prevent inconsistency.

7) Repeat these steps for each Virtual Warehouse that shares the same Database Catalog.

Diagnostic bundle download fails

After upgrading to version 1.4.3 (released September 15, 2022), downloading the diagnostic bundle for the Hive Virtual Warehouse results in an error. New environments do not have this problem. The problem is caused by missing permissions to access Kubernetes resources. Version 1.4.3 adds a requirement for the role-based access control (rbac) permissions to download the diagnostic bundle.

Workaround: Add the missing rbac permissions as follows:

1) Create the following file:

```
apiVersion: rbac.authorization.k8s.io/v1
kind: ClusterRole
metadata:
  name: diagnostic-data-generator-role-monitor
rules:
- apiGroups:
  - ""
  resources:
  - configmaps
  - events
  - pods
  - persistentvolumeclaims
  - nodes
  verbs:
  - get
  - list
- apiGroups:
  - apps
  resources:
  - deployments
  - statefulsets
  verbs:
  - get
  - list
- apiGroups:
  - "edws.cloudera.com"
  resources:
  - computes
  verbs:
  - get
  - list
```

2) Run the following command to apply the permissions:

```
kubectl apply -f <filename>
```

Workloads from earlier CDW versions cannot be deployed on new AWS environments

Because new environments are provisioned automatically to use AWS Elastic Kubernetes Service (EKS) 1.21, deprecated APIs used in workloads of earlier versions are not supported in this version 1.4.3-b225 (released September 15, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with EKS 1.21.

DWX-13062 Hive-26507 Converting a Hive table having CHAR or VARCHAR columns to Iceberg causes an exception

CHAR and VARCHAR data can be shorter than the length specified by the data type. Remaining characters are padded with spaces. Data is converted to a string in Iceberg. This process can yield incorrect results when you query the converted Iceberg table.

Workaround: Change columns from CHAR or VARCHAR to string types before converting the Hive table to Iceberg.

DWX-13276 Multiple inserts into tables having different formats can cause a deadlock.

Under the following conditions, a deadlock can occur:

- You run a query to insert data into multiple tables comprised of at least one Iceberg table and at least one non-Iceberg table.
- The STAT task locking feature is turned on (default = on).

Workarounds: Perform either one of the following workarounds:

- Run separate queries to insert data into only one table at a time.
- Turn off STAT task locking as follows:

```
set iceberg.hive.request-lock-on-stats-task=false;
```

Carried over from the previous release: General

TSB 2023-656: TSB Cloudera Data Warehouse does not clean up old Helm releases

Cloudera Data Warehouse (CDW) uses Helm release manager to deploy component releases into the Kubernetes cluster of the customer. However, CDW's current behavior in CDW keeps the Helm release history without any boundary. Since the storage for Helm in CDW uses secrets, this behavior can lead to a very high number of secrets (Impala/Hive scale-up/down creates new releases;clusters were seen with 6000+ secrets). Therefore, parsing and decoding a high number of Helm releases can require a significant amount of memory. As an example, for a cluster with 6000+ secrets, it could be 6-7 GB or more of memory utilized. This may result in performance degradation, instability and/or higher cloud cost because several CDW cluster components rely on secrets, which would cause the memory footprint of these other components to increase significantly.

This issue is addressed in the CDW version 1.5.1-b110 (workload image version 2022.0.11.0-122). Cloudera highly recommends upgrading Apache Impala (Impala) and Apache Hive (Hive) in CDW to 2022.0.11.0-122 or higher versions to ensure performance, stability and/or cloud cost are not impacted by this behavior.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-656: https://my.cloudera.com/knowledge/TSB-2023-656-Cloudera-Data-Warehouse-does-not-clean-up-old?id=368733

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures to/from prometheus to any other pods in the Kubernetes cluster. Data Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-6619 Browser auto-close not working on some browsers after token-Based authentication for accessing CDW

The Firefox and Edge browser window does not close automatically after successful authentication.

DWX-9774 Database Catalog or Virtual Warehouse image version problem

Background: In Cloudera Data Warehouse 2021.0.3-b27 - 2021.0.5-b36, you can choose any supported image version when you create a Database Catalog or Virtual Warehouse, assuming you have the CDW_VERSIONED_DEPLOY entitlement.

Problem: In Cloudera Data Warehouse 2021.0.6-b96, you can choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse only if you use an existing environment. An existing environment is one you created in 2021.0.3-b27 - 2021.0.5-b36.

Workaround: In 2021.0.6-b96, use an environment you created before this release to choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-8573: EKS upgrade from DWX UI to K8s v1.20 fails in reduced permissions mode

In reduced permissions mode, attempting to update Amazon Elastic Kubernetes Service (EKS) to Kubernetes version 1.20 fails with an AccessDenied error.

Workaround:In AWS, add the following permissions to your reduced permissions mode JSON IAM permissions policy right after "iam:PutRolePolicy":

```
{
       ...
       "cloudformation:GetTemplate",
```

```
        "cloudformation:GetTemplateSummary",
        "iam:GetRole",
        "eks:ListUpdates",
        "ec2:CreateLaunchTemplateVersion",
        "ec2:DescribeLaunchTemplateVersions",
        "ec2:DescribeLaunchTemplates",
        "autoscaling:UpdateAutoScalingGroup",
        "autoscaling:DeleteAutoScalingGroup",
        "ec2:RunInstances",
        "autoscaling:DescribeScalingActivities",
        "autoscaling:TerminateInstanceInAutoScalingGroup",
        "autoscaling:DescribeScheduledActions",
        "autoscaling:SetDesiredCapacity",
        "iam:PassRole",
        "rds:DescribeDBInstances",
        "ec2:DescribeInstances"
        ...
```

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the Create Tags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.

3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersions line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 2022.0.9.0-120 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle. Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Hue Query Editor

TSB 2023-670: Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed: kubelet Error: failed to create containerd task: failed to create shim task: OCI runtime create failed: runc create failed: unable to start container process: exec: “run_httpd.sh”: cannot run executable found relative to current directory: unknown

The root cause of the issue is a recent node image upgrade in Azure made by Microsoft, where the version of the container runtime (containerd) was upgraded to 1.6.18. This version upgraded Go SDK to 1.19.6, where the behavior of how program execution works has changed. Due to security concerns, the newer version of Go SDK does not resolve a program using an implicit or explicit path entry relative to the current directory (for more information, see the Go documentation), and containerd indirectly uses this Go exec API. The Docker command of the Hue frontend relies on the former behavior. Therefore the command fails on recent node images, because it cannot find the expected executable in the current working directory of the containers anymore. Cloudera first noticed the aforementioned containerd version in the Azure Kubernetes Service (AKS) released on 2023-03-05. Every node image which is newer than the version released on 2023-03-05 is affected by this issue. Unfortunately, Cloudera does not have any insight to when Microsoft rolls out new node images in a given region. Therefore, it is possible that in some regions, the older node images are still in use, where the issue does not arise until a given regional update is applied to the node images.

CDW environments, which are already activated and running, are not affected as long as customers do not trigger a node image upgrade to the latest available version either using Azure Command Line Interface (CLI) or on the Azure portal. If the node image is upgraded to the latest version, then the Virtual Warehouses also need to be upgraded to the latest CDWH-2023.0.14.0 version. Newly created CDW environments are not affected, but customers are advised to not choose a Hue version lower than CDWH-2023.0.14.0 during any Virtual Warehouse creation because such a configuration is affected by this issue.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-670: TSB Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

Carried over from the previous release: Database Catalog

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

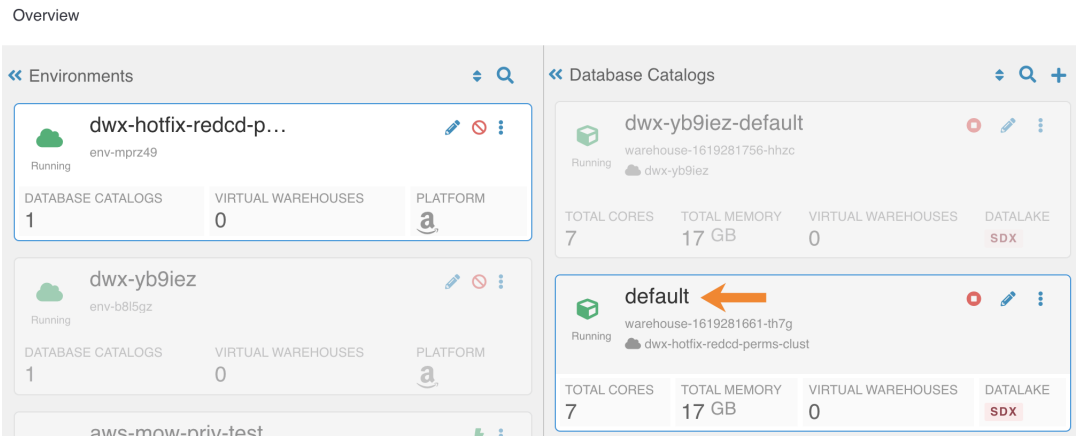
- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 2022.0.9.0-120 (released August 4, 2022).
- You added Database Catalogs (not the automatically generated default Database Catalog) using either one of these CDW versions:
  - 2022.0.8.0-89 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 2022.0.9.0-120 (released August 4, 2022).

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

DWX-10271: Missing log section in Hive query results

In a Hive Virtual Warehouse, when you run a query in Hue, the query results do not contain a logs section.

DWX-8118: INSERT INTO command fails under certain circumstances

This problem affects users who have a PostgreSQL database as the backend Hive database. If you create a table A and create a table B as select (CTAS) from an empty table A, inserting values into table B fails as follows:

```
Error while compiling statement: FAILED: Execution Error, return
 code 1 from
       org.apache.hadoop.hive.ql.exec.StatsTask.org.apache.thr
ift.transport.TTransportException
```

Workaround: Disable auto-stats gathering: Go to Cloudera Manager Data Warehouse Virtual Warehouses , and click the Hive VW name in the list. In Configuration HiveServer2 , set hive.cbo.enable to false.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses

must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.
7. Click Apply in the upper right corner of the page to save the configuration.

8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name ‘ec2-18-219-
32-183.us-east-2.compute.amazonaws.com’ does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Impala Virtual Warehouse

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

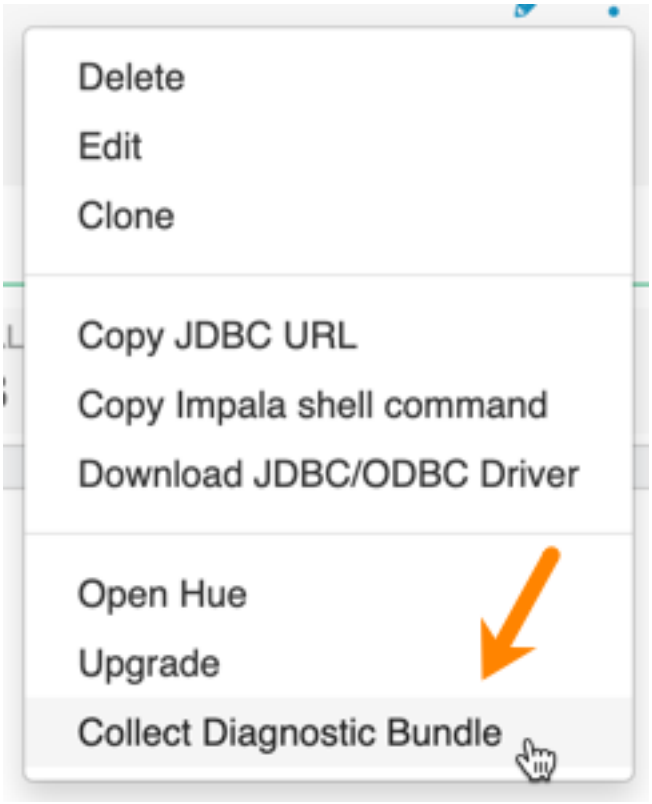
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case so the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

August 4, 2022

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

New known Issues in this release

Do not upgrade this release to AKS 1.22

The latest version of Cloudera Data Warehouse (CDW) 1.4.2-b118 (released August 4, 2022) provisions AKS 1.22 by default in Azure environments; however. upgrading your activated environments to AKS 1.22 or later is not supported. Do not upgrade your AKS clusters to AKS 1.22.

DWX-13103 Cloudera Data Warehouse environment activation problem

When CDW environments are activated, a race condition can occur between the prometheus pod and istiod pod. The prometheus pod can be set up without an istio-proxy container, causing communication failures between prometheus and other pods in the Kubernetes cluster. Data Warehouse prometheus-related functionalities, such as autoscaling, stop working. Grafana dashboards, which get metrics from prometheus, are not populated.

Workaround: Restart the prometheus pod so that it gets the istio-proxy container.

DWX-13094 Diagnostic bundle job fails with missing service account

Due to a change to the name of the service accounts, after upgrading to DWX-1.4.2, the diagnostic bundle job can fail if you used the bundle in earlier versions. Due to a recent. Jobs from the latest

version are trying to run with the new name of the service account, but these are not reinstalled after the upgrade.

Workaround: 1) Create the new service account the job is looking for (diagnostic-data-generator-service-account) as follows:

```
apiVersion: v1
kind: ServiceAccount
metadata:
  name: diagnostic-data-generator-service-account
  namespace: monitor
---
apiVersion: rbac.authorization.k8s.io/v1
kind: ClusterRoleBinding
metadata:
  name: ddg-role-binding-custom
roleRef:
  apiGroup: rbac.authorization.k8s.io
  kind: ClusterRole
  name: diagnostic-data-generator-role-monitor
subjects:
- kind: ServiceAccount
  name: diagnostic-data-generator-service-account
  namespace: monitor
```

2) Bind the service account to the cluster role that grants the necessary permissions to the job using the following command:

```
kubectl create -f <filename>
```

Non-default Database Catalogs created with several earlier CDW versions fails

This issue affects you only if you meet the following conditions:

- You have a versioned CDW deployment and the multi default DBC entitlement.
- You are using this release of CDW version 1.4.2-b118 (released August 4, 2022).
- You added Database Catalogs (not the automatically generated default Database Catalog) using either of these CDW versions:
  - 1.4.1-b86 (released June, 22, 2022)
  - 2022-0.7.1-2 (released May 10, 2022)

Do not attempt to upgrade the Database Catalog you created with these earlier releases. The Database Catalog will fail. Your existing Database Catalog created with the earlier release works fine with CDW Runtime. Recreating your existing Database Catalog updates CDW Runtime for 1.4.2-b118 (released August 4, 2022).

Workloads from earlier CDW versions cannot be deployed on new Azure environments

Because new environments are provisioned automatically to use Azure Kubernetes Service (AKS) 1.22, deprecated APIs used in workloads of earlier versions are not supported in this version 1.4.2-b118 (released August 4, 2022). This issue affects you only if you have the CDW_VERSIONED_DEPLOY entitlement.

Workaround: Create new workloads. Workloads in Database Catalogs, Hive, Impala, Hue, Data Visualization, and are incompatible with AKS 1.22.

CDPD-40730 Parquet change can cause incompatibility

Parquet files written by the parquet-mr library in this version of CDW, where the schema contains a timestamp with no UTC conversion will not be compatible with older versions of Parquet readers. The effect is that the older versions will still consider these timestamps as they would require UTC conversions and will thus end up with a wrong result. You can encounter this problem only when

you write Parquet-based tables using Hive, and tables have the non-default configuration hive.parquet.write.int64.timestamp=true.

Carried over from the previous release: General

TSB 2023-656: TSB Cloudera Data Warehouse does not clean up old Helm releases

Cloudera Data Warehouse (CDW) uses Helm release manager to deploy component releases into the Kubernetes cluster of the customer. However, CDW's current behavior in CDW keeps the Helm release history without any boundary. Since the storage for Helm in CDW uses secrets, this behavior can lead to a very high number of secrets (Impala/Hive scale-up/down creates new releases;clusters were seen with 6000+ secrets). Therefore, parsing and decoding a high number of Helm releases can require a significant amount of memory. As an example, for a cluster with 6000+ secrets, it could be 6-7 GB or more of memory utilized. This may result in performance degradation, instability and/or higher cloud cost because several CDW cluster components rely on secrets, which would cause the memory footprint of these other components to increase significantly.

This issue is addressed in the CDW version 1.5.1-b110 (workload image version 2022.0.11.0-122). Cloudera highly recommends upgrading Apache Impala (Impala) and Apache Hive (Hive) in CDW to 2022.0.11.0-122 or higher versions to ensure performance, stability and/or cloud cost are not impacted by this behavior.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-656: https://my.cloudera.com/knowledge/TSB-2023-656-Cloudera-Data-Warehouse-does-not-clean-up-old?id=368733

DWX-6619 Browser auto-close not working on some browsers after token-Based authentication for accessing CDW

The Firefox and Edge browser window does not close automatically after successful authentication.

DWX-9774 Database Catalog or Virtual Warehouse image version problem

Background: In Cloudera Data Warehouse 2021.0.3-b27 - 2021.0.5-b36, you can choose any supported image version when you create a Database Catalog or Virtual Warehouse, assuming you have the CDW_VERSIONED_DEPLOY entitlement.

Problem: In Cloudera Data Warehouse 2021.0.6-b96, you can choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse only if you use an existing environment. An existing environment is one you created in 2021.0.3-b27 - 2021.0.5-b36.

Workaround: In 2021.0.6-b96, use an environment you created before this release to choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"
```

```
        # Save and quit the editor, and then run the following
command to check your changes.

        kubectl get svc nginx-service -n <cluster-name> -o yaml

        # Make sure that the annotation you added is present.
```

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-8573: EKS upgrade from DWX UI to K8s v1.20 fails in reduced permissions mode

In reduced permissions mode, attempting to update Amazon Elastic Kubernetes Service (EKS) to Kubernetes version 1.20 fails with an AccessDenied error.

Workaround:In AWS, add the following permissions to your reduced permissions mode JSON IAM permissions policy right after "iam:PutRolePolicy":

```
{
        ...
        "cloudformation:GetTemplate",
        "cloudformation:GetTemplateSummary",
        "iam:GetRole",
        "eks:ListUpdates",
        "ec2:CreateLaunchTemplateVersion",
        "ec2:DescribeLaunchTemplateVersions",
        "ec2:DescribeLaunchTemplates",
        "autoscaling:UpdateAutoScalingGroup",
        "autoscaling:DeleteAutoScalingGroup",
        "ec2:RunInstances",
        "autoscaling:DescribeScalingActivities",
        "autoscaling:TerminateInstanceInAutoScalingGroup",
        "autoscaling:DescribeScheduledActions",
        "autoscaling:SetDesiredCapacity",
        "iam:PassRole",
        "rds:DescribeDBInstances",
        "ec2:DescribeInstances"
        ...
```

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
        operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the Create Tags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersions line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Hue Query Editor

TSB 2023-670: Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed: kubelet Error: failed to create containerd task: failed to create shim task: OCI runtime create failed: runc create failed: unable to start container process: exec: “run_httpd.sh”: cannot run executable found relative to current directory: unknown

The root cause of the issue is a recent node image upgrade in Azure made by Microsoft, where the version of the container runtime (containerd) was upgraded to 1.6.18. This version upgraded Go SDK to 1.19.6, where the behavior of how program execution works has changed. Due to security concerns, the newer version of Go SDK does not resolve a program using an implicit or explicit path entry relative to the current directory (for more information, see the Go documentation), and containerd indirectly uses this Go exec API. The Docker command of the Hue frontend relies on the former behavior. Therefore the command fails on recent node images, because it cannot find the expected executable in the current working directory of the containers anymore. Cloudera first noticed the aforementioned containerd version in the Azure Kubernetes Service (AKS) released on 2023-03-05. Every node image which is newer than the version released on 2023-03-05 is affected by this issue. Unfortunately, Cloudera does not have any insight to when Microsoft rolls out new node images in a given region. Therefore, it is possible that in some regions, the older node images are still in use, where the issue does not arise until a given regional update is applied to the node images.

CDW environments, which are already activated and running, are not affected as long as customers do not trigger a node image upgrade to the latest available version either using Azure Command Line Interface (CLI) or on the Azure portal. If the node image is upgraded to the latest version, then the Virtual Warehouses also need to be upgraded to the latest CDWH-2023.0.14.0 version. Newly created CDW environments are not affected, but customers are advised to not choose a Hue

version lower than CDWH-2023.0.14.0 during any Virtual Warehouse creation because such a configuration is affected by this issue.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-670: TSB Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

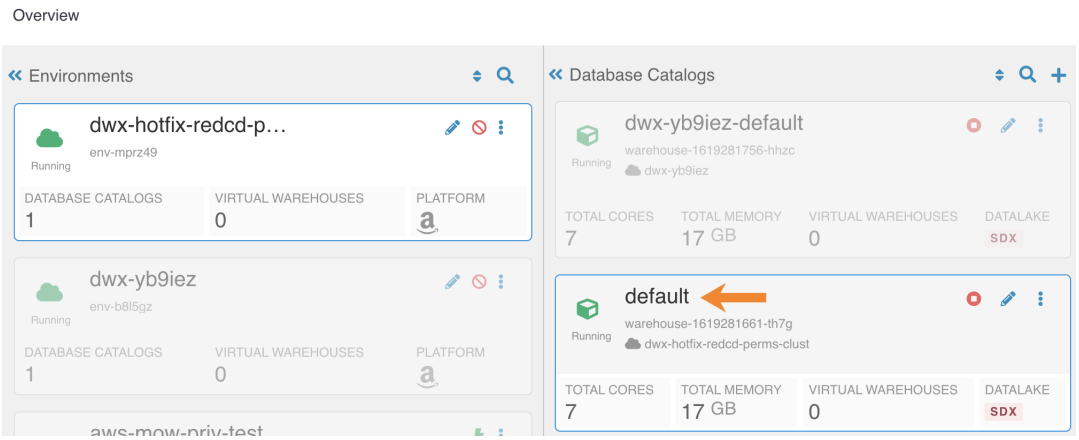
None.

Carried over from the previous release: Database Catalog

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

DWX-10271: Missing log section in Hive query results

In a Hive Virtual Warehouse, when you run a query in Hue, the query results do not contain a logs section.

DWX-8118: INSERT INTO command fails under certain circumstances

This problem affects users who have a PostgreSQL database as the backend Hive database. If you create a table A and create a table B as select (CTAS) from an empty table A, inserting values into table B fails as follows:

```
Error while compiling statement: FAILED: Execution Error, return
 code 1 from
       org.apache.hadoop.hive.ql.exec.StatsTask.org.apache.thr
ift.transport.TTransportException
```

Workaround: Disable auto-stats gathering: Go to Cloudera Manager Data Warehouse Virtual Warehouses , and click the Hive VW name in the list. In Configuration HiveServer2 , set hive.cbo .enable to false.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.

7. Click Apply in the upper right corner of the page to save the configuration.
8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name 'ec2-18-219-
32-183.us-east-2.compute.amazonaws.com' does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Impala Virtual Warehouse

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

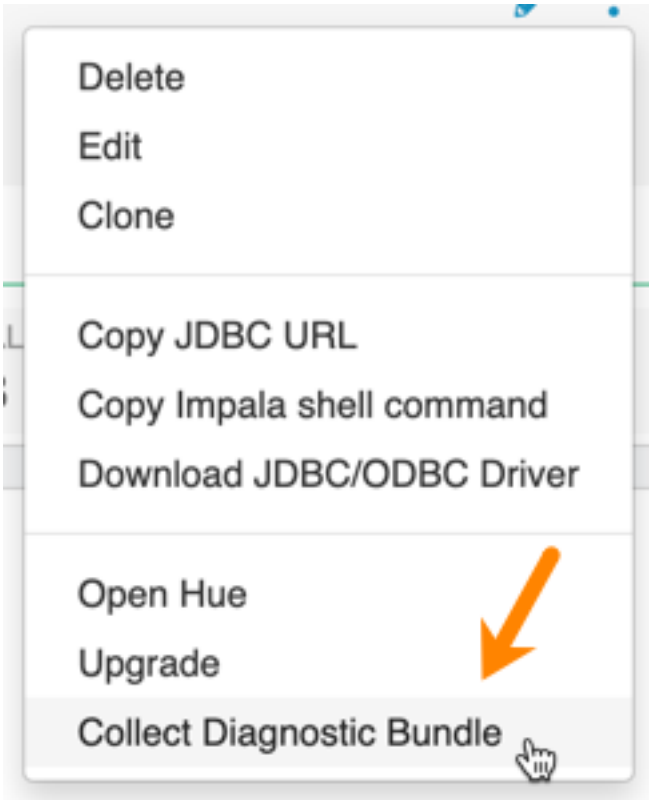
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case so the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

June 22, 2022

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following known issues:

New known Issues in this release

DWX-12758 DBC upgrade fails for some source versions

In this release, if you are using a Database Catalog that was created in CDW 2022.0.7.0-80 (released April 27, 2022) or 2022.0.7.1-2 (released May 10, 2022) a failure can occur when you try to upgrade the Database Catalog.

DWX-12528 Tags are not removed for environment activated in reduced permissions mode

You see a DWX server log error, "... error delete tags" when in reduced permissions mode. You can ignore this error.

DWX-12658 Rebuild materialized view after Database Catalog upgrade

You must rebuild materialized views created in version of the Database Catalog that is older than the version of your Virtual Warehouse. This is expected behavior.

Workaround: Run the following command to rebuild a materialized view after moving to a Virtual Warehouse based on a later version of the Database Catalog:

```
ALTER MATERIALIZED VIEW mv1 REBUILD;
```

DWX-12598 DESCRIBE EXTENDED does not show bucketing information about an Iceberg table.

This issue affects the Hive and Impala Virtual Warehouses. The Hive Virtual Warehouse, which supports DESCRIBE FORMATTED of Iceberg talble, has the same problem

CDPD-40749 IMPALA-11346 Iceberg query failure

When a partitioned legacy table is converted to Iceberg, Impala queries of the table might fail. The problem occurs when the query contains a WHERE clause plus a predicate on the partition columns.

Workaround: Replace Iceberg tables created prior to this GA release with new tables. Create a new table in Iceberg by selecting the old table as shown in Test driving Iceberg from Impala.

Carried over from the previous release: General

TSB 2023-656: TSB Cloudera Data Warehouse does not clean up old Helm releases

Cloudera Data Warehouse (CDW) uses Helm release manager to deploy component releases into the Kubernetes cluster of the customer. However, CDW's current behavior in CDW keeps the Helm release history without any boundary. Since the storage for Helm in CDW uses secrets, this behavior can lead to a very high number of secrets (Impala/Hive scale-up/down creates new releases;clusters were seen with 6000+ secrets). Therefore, parsing and decoding a high number of Helm releases can require a significant amount of memory. As an example, for a cluster with 6000+ secrets, it could be 6-7 GB or more of memory utilized. This may result in performance degradation, instability and/or higher cloud cost because several CDW cluster components rely on secrets, which would cause the memory footprint of these other components to increase significantly.

This issue is addressed in the CDW version 1.5.1-b110 (workload image version 2022.0.11.0-122). Cloudera highly recommends upgrading Apache Impala (Impala) and Apache Hive (Hive) in CDW to 2022.0.11.0-122 or higher versions to ensure performance, stability and/or cloud cost are not impacted by this behavior.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-656: https://my.cloudera.com/knowledge/TSB-2023-656-Cloudera-Data-Warehouse-does-not-clean-up-old?id=368733

DWX-6619 Browser auto-close not working on some browsers after token-Based authentication for accessing CDW

The Firefox and Edge browser window does not close automatically after successful authentication.

DWX-9774 Database Catalog or Virtual Warehouse image version problem

Background: In Cloudera Data Warehouse 2021.0.3-b27 - 2021.0.5-b36, you can choose any supported image version when you create a Database Catalog or Virtual Warehouse, assuming you have the CDW_VERSIONED_DEPLOY entitlement.

Problem: In Cloudera Data Warehouse 2021.0.6-b96, you can choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse only if you use an existing environment. An existing environment is one you created in 2021.0.3-b27 - 2021.0.5-b36.

Workaround: In 2021.0.6-b96, use an environment you created before this release to choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Carried over from the previous release: AWS

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-8573: EKS upgrade from DWX UI to K8s v1.20 fails in reduced permissions mode

In reduced permissions mode, attempting to update Amazon Elastic Kubernetes Service (EKS) to Kubernetes version 1.20 fails with an AccessDenied error.

Workaround:In AWS, add the following permissions to your reduced permissions mode JSON IAM permissions policy right after "iam:PutRolePolicy":

```
{
       ...
       "cloudformation:GetTemplate",
       "cloudformation:GetTemplateSummary",
       "iam:GetRole",
       "eks:ListUpdates",
       "ec2:CreateLaunchTemplateVersion",
       "ec2:DescribeLaunchTemplateVersions",
       "ec2:DescribeLaunchTemplates",
       "autoscaling:UpdateAutoScalingGroup",
       "autoscaling:DeleteAutoScalingGroup",
       "ec2:RunInstances",
       "autoscaling:DescribeScalingActivities",
       "autoscaling:TerminateInstanceInAutoScalingGroup",
       "autoscaling:DescribeScheduledActions",
       "autoscaling:SetDesiredCapacity",
       "iam:PassRole",
       "rds:DescribeDBInstances",
       "ec2:DescribeInstances"
       ...
```

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
       operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
       Description: DynamoDB table name for the SDX S3 file lis
tings, created through S3Guard
       Type: String
       Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the Create Tags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersions line as shown:

```
"Version": "2012-10-17",
          "Statement": [
          {
          "Action": [
          "autoscaling:DescribeAutoScalingGroups",
          "autoscaling:DescribeAutoScalingInstances",
          "autoscaling:DescribeTags",
          "autoscaling:DescribeLaunchConfigurations",
          "autoscaling:SetDesiredCapacity",
          "autoscaling:TerminateInstanceInAutoScalingGroup",
          "ec2:DescribeLaunchTemplateVersions",
          "ec2:CreateTags"
          ],
```

7. Save changes.

Carried over from the previous release: Azure

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle. Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses

old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Carried over from the previous release: Hue Query Editor

TSB 2023-670: Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed: kubelet Error: failed to create containerd task: failed to create shim task: OCI runtime create failed: runc create failed: unable to start container process: exec: “run_httpd.sh”: cannot run executable found relative to current directory: unknown

The root cause of the issue is a recent node image upgrade in Azure made by Microsoft, where the version of the container runtime (containerd) was upgraded to 1.6.18. This version upgraded Go SDK to 1.19.6, where the behavior of how program execution works has changed. Due to security concerns, the newer version of Go SDK does not resolve a program using an implicit or explicit path entry relative to the current directory (for more information, see the Go documentation), and containerd indirectly uses this Go exec API. The Docker command of the Hue frontend relies on the former behavior. Therefore the command fails on recent node images, because it cannot find the expected executable in the current working directory of the containers anymore. Cloudera first noticed the aforementioned containerd version in the Azure Kubernetes Service (AKS) released on 2023-03-05. Every node image which is newer than the version released on 2023-03-05 is affected by this issue. Unfortunately, Cloudera does not have any insight to when Microsoft rolls out new node images in a given region. Therefore, it is possible that in some regions, the older node images are still in use, where the issue does not arise until a given regional update is applied to the node images.

CDW environments, which are already activated and running, are not affected as long as customers do not trigger a node image upgrade to the latest available version either using Azure Command Line Interface (CLI) or on the Azure portal. If the node image is upgraded to the latest version, then the Virtual Warehouses also need to be upgraded to the latest CDWH-2023.0.14.0 version. Newly created CDW environments are not affected, but customers are advised to not choose a Hue version lower than CDWH-2023.0.14.0 during any Virtual Warehouse creation because such a configuration is affected by this issue.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-670: TSB Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.

2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

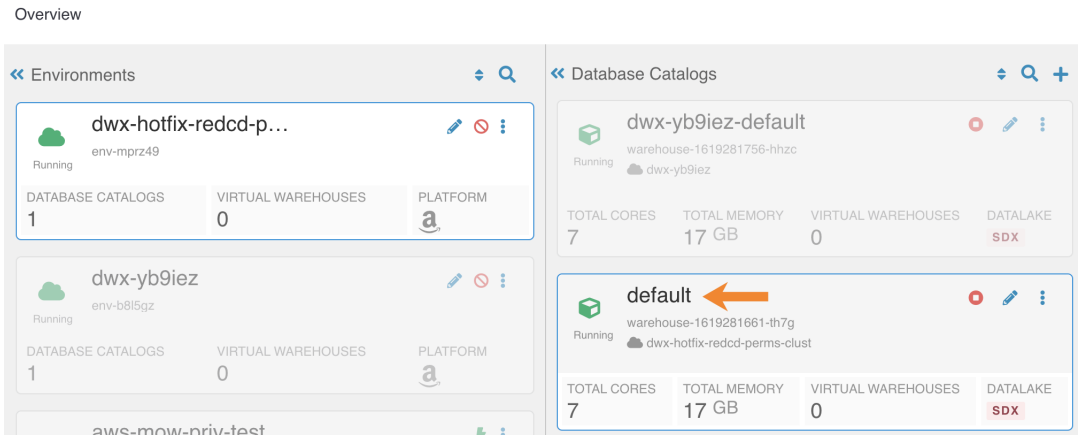
None.

Carried over from the previous release: Database Catalog

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Carried over from the previous release: Hive Virtual Warehouse

DWX-10271: Missing log section in Hive query results

In a Hive Virtual Warehouse, when you run a query in Hue, the query results do not contain a logs section.

DWX-8118: INSERT INTO command fails under certain circumstances

This problem affects users who have a PostgreSQL database as the backend Hive database. If you create a table A and create a table B as select (CTAS) from an empty table A, inserting values into table B fails as follows:

```
Error while compiling statement: FAILED: Execution Error, return
 code 1 from
       org.apache.hadoop.hive.ql.exec.StatsTask.org.apache.thr
ift.transport.TTransportException
```

Workaround: Disable auto-stats gathering: Go to Cloudera Manager Data Warehouse Virtual Warehouses , and click the Hive VW name in the list. In Configuration HiveServer2 , set hive.cbo .enable to false.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.
7. Click Apply in the upper right corner of the page to save the configuration.
8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name ‘ec2-18-219-
32-183.us-east-2.compute.amazonaws.com’ does not
       match the certificate subject provided by the peer (CN=*
.env-c25dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Carried over from the previous release: Impala Virtual Warehouse

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

       # Add the following under the metadata.annotations field
       service.beta.kubernetes.io/aws-load-balancer-ssl-negoti
ation-policy: "ELBSecurityPolicy-TLS-1-2-2017-01"

       # Save and quit the editor, and then run the following
command to check your changes.

       kubectl get svc nginx-service -n <cluster-name> -o yaml

       # Make sure that the annotation you added is present.
```

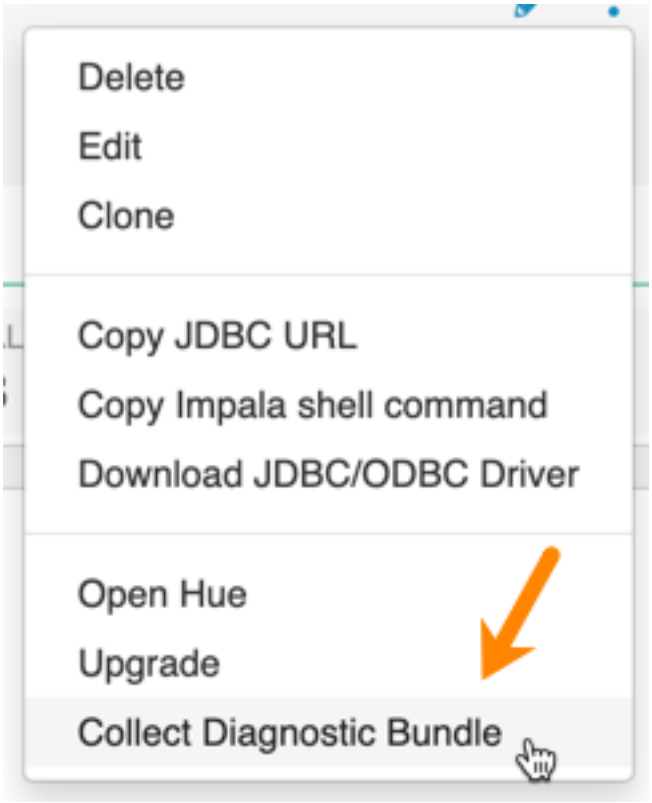
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case so the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

General known issues on public clouds

Learn about the general known issues in Cloudera Data Warehouse (CDW) service on public clouds, the impact or changes to the functionality, and the workaround.

TSB 2023-656: TSB Cloudera Data Warehouse does not clean up old Helm releases

Cloudera Data Warehouse (CDW) uses Helm release manager to deploy component releases into the Kubernetes cluster of the customer. However, CDW's current behavior in CDW keeps the Helm release history without any boundary. Since the storage for Helm in CDW uses secrets, this behavior can lead to a very high number of secrets (Impala/Hive scale-up/down creates new releases;clusters were seen with 6000+ secrets). Therefore, parsing and decoding a high number of Helm releases can require a significant amount of memory. As an example, for a cluster with 6000+ secrets, it could be 6-7 GB or more of memory utilized. This may result in performance degradation, instability and/or higher cloud cost because several CDW cluster components rely on secrets, which would cause the memory footprint of these other components to increase significantly.

This issue is addressed in the CDW version 1.5.1-b110 (workload image version 2022.0.11.0-122). Cloudera highly recommends upgrading Apache Impala (Impala) and Apache Hive (Hive) in CDW to 2022.0.11.0-122 or higher versions to ensure performance, stability and/or cloud cost are not impacted by this behavior.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-656: https://my.cloudera.com/knowledge/TSB-2023-656-Cloudera-Data-Warehouse-does-not-clean-up-old?id=368733

DWX-6619 Browser auto-close not working on some browsers after token-Based authentication for accessing CDW

The Firefox and Edge browser window does not close automatically after successful authentication.

DWX-9774 Database Catalog or Virtual Warehouse image version problem

Background: In Cloudera Data Warehouse 2021.0.3-b27 - 2021.0.5-b36, you can choose any supported image version when you create a Database Catalog or Virtual Warehouse, assuming you have the CDW_VERSIONED_DEPLOY entitlement.

Problem: In Cloudera Data Warehouse 2021.0.6-b96, you can choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse only if you use an existing environment. An existing environment is one you created in 2021.0.3-b27 - 2021.0.5-b36.

Workaround: In 2021.0.6-b96, use an environment you created before this release to choose an image version other than 2021.0.6-b96 when you create a Database Catalog or Virtual Warehouse.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

# Add the following under the metadata.annotations field
service.beta.kubernetes.io/aws-load-balancer-ssl-negotiation-pol
icy: "ELBSecurityPolicy-TLS-1-2-2017-01"

# Save and quit the editor, and then run the following command to
 check your changes.

kubectl get svc nginx-service -n <cluster-name> -o yaml

# Make sure that the annotation you added is present.
```

DWX-5742: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails

Problem: Upgrading multiple Hive and Impala Virtual Warehouses or Database Catalogs at the same time fails.

Workaround: If you need to upgrade or create multiple Hive and Impala Virtual Warehouses or Database Catalogs, perform the upgrade or creation sequentially one at a time.

Known issues in CDW running on AWS environments

Learn about the known issues in Cloudera Data Warehouse service in AWS environments, the impact or changes to the functionality, and workarounds.

AWS availability zone inventory issue

In this release, you can select a preferred availability zone when you create a Virtual Warehouse; however, AWS might not be able to provide enough compute instances of the type that Cloudera Data Warehouse needs.

Workaround: If you experience this AWS issue, try recreating the Virtual Warehouse and choosing a different availability zone.

DWX-8573: EKS upgrade from DWX UI to K8s v1.20 fails in reduced permissions mode

In reduced permissions mode, attempting to update Amazon Elastic Kubernetes Service (EKS) to Kubernetes version 1.20 fails with an AccessDenied error.

Workaround:In AWS, add the following permissions to your reduced permissions mode JSON IAM permissions policy right after "iam:PutRolePolicy":

```
{
 ...
                "cloudformation:GetTemplate",
                "cloudformation:GetTemplateSummary",
                "iam:GetRole",
                "eks:ListUpdates",
                "ec2:CreateLaunchTemplateVersion",
                "ec2:DescribeLaunchTemplateVersions",
                "ec2:DescribeLaunchTemplates",
                "autoscaling:UpdateAutoScalingGroup",
                "autoscaling:DeleteAutoScalingGroup",
                "ec2:RunInstances",
                "autoscaling:DescribeScalingActivities",
                "autoscaling:TerminateInstanceInAutoScalingGrou
p",
                "autoscaling:DescribeScheduledActions",
                "autoscaling:SetDesiredCapacity",
                "iam:PassRole",
                "rds:DescribeDBInstances",
                "ec2:DescribeInstances"
 ...
```

DWX-7613: CloudFormation stack creation using AWS CLI broken for CDW Reduced Permissions Mode

Problem: If you use the AWS CLI to create a CloudFormation stack to activate an AWS environment for use in Reduced Permissions Mode, it fails and returns the following error:

```
An error occurred (ValidationError) when calling the CreateStack
 operation: Parameters: [SdxDDBTableName] must have values
```

The default value of SdxDDBTableName is not being set. If you create the CloudFormation stack using the AWS Console, there is no problem.

Workaround: If you must use the AWS CLI, edit the CloudFormation stack template file as follows:

```
SdxDDBTableName:
    Description: DynamoDB table name for the SDX S3 file listin
gs, created through S3Guard
    Type: String
    Default: " "
```

Then rerun the CloudFormation stack creation command using the AWS CLI.

ENGESC-8271: Helm 2 to Helm 3 migration fails on AWS environments where the overlay network feature is in use and namespaces are stuck in a terminating state

Problem: While using the overlay network feature for AWS environments and after attempting to migrate an AWS environment from Helm 2 to Helm 3, the migration process fails.

Workaround: Run the following kubectl command to determine whether you have any namespaces stuck in a terminating state:

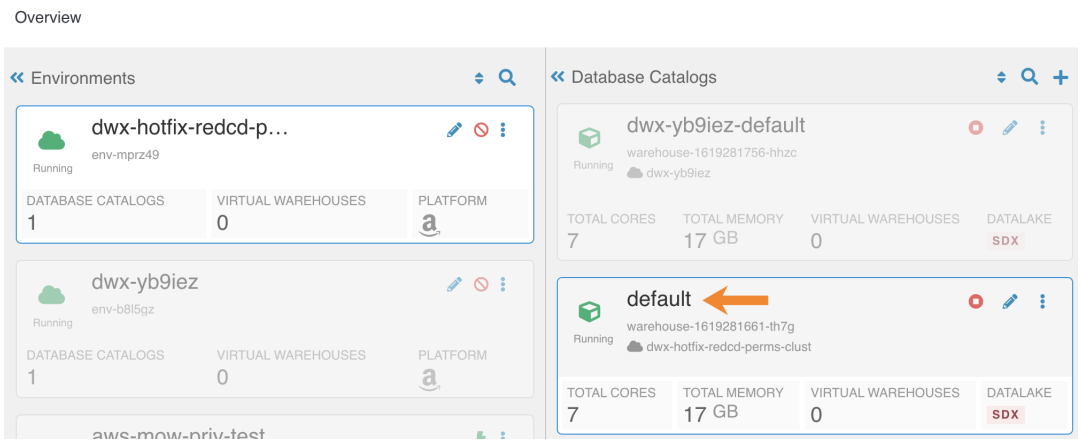
```
kubectl get ns
```

Then contact Cloudera Technical Support to report and get help on this issue.

DWX-7349: In reduced permissions mode, default Database Catalog name does not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6970: Tags do not get applied in existing CDW environments

Problem: You may see the following error while trying to apply tags to Virtual Warehouses in an existing CDW environment: An error occurred (UnauthorizedOperation) when calling the CreateTags operation: You are not authorized to perform this operation and Compute node tagging was unsuccessful. This happens because the ec2:CreateTags privilege is missing from your AWS cluster-autoscaler inline policy for the NodeInstanceRole role.

Workaround: Add the ec2:CreateTags privilege to the cluster-autoscaler inline policy as follows:

1. Log into the AWS IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation panel, choose Roles.
3. Search the list of roles for NodeInstanceRole.
4. Click Permissions.
5. Select cluster-autoscaler and click Edit policy.
6. Add the ec2:CreateTags line in the Actions section after the ec2:DescribeLaunchTemplateVersions line as shown:

```
"Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "autoscaling:DescribeAutoScalingGroups",
                "autoscaling:DescribeAutoScalingInstances",
                "autoscaling:DescribeTags",
                "autoscaling:DescribeLaunchConfigurations",
```

```
                "autoscaling:SetDesiredCapacity",
                "autoscaling:TerminateInstanceInAutoScaling
Group",
                "ec2:DescribeLaunchTemplateVersions",
                "ec2:CreateTags"
            ],
```

7. Save changes.

Known issues in CDW running on Azure environments

Learn about the known issues in Cloudera Data Warehouse service in Azure environments, the impact or changes to the functionality, and the workaround.

Incorrect diagnostic bundle location

Problem:The path you see to the diagnostic bundle is wrong when you create a Virtual Warehouse, collect a diagnostic bundle of log files for troubleshooting, and click Edit Diagnostic Bundle . Your storage account name is missing from the beginning of the path.

Workaround:To find your diagnostic bundle, add your storage account name to the beginning of the path, for example:

```
my-storage-account-name/log-files/clusters/my-env/my-warehouse-x
x-yy/warehouse/debug-artifacts/hive/compute-zz-mvvf/compute-zz-m
vvf-0926210111-0927090111-01234.zip
```

To get your storage account name, in Cloudera Management Console, click Environments, select your environment, and scroll down to Logs Storage and Audits. The first part of the path is your storage account name.

Changed environment credentials not propagated to AKS

Problem: When you change the credentials of a running cloud environment using the Management Console, the changes are not automatically propagated to the corresponding active Cloudera Data Warehouse (CDW) environment. As a result, the Azure Kubernetes Service (AKS) uses old credentials and may not function as expected resulting in inaccessible Hive or Impala Virtual Warehouses.

Workaround: To resolve this issue, you must manually synchronize the changes with the CDW AKS resources. To synchronize the updated credentials, see Update AKS cluster with new service principal credentials in the Azure product documentation.

Known issues in Data Analytics Studio on public clouds

Learn about the known issues related to Data Analytics Studio (DAS) in Cloudera Data Warehouse service on public clouds, the impact or changes to the functionality, and the workaround.

DWX-929: DAS UI displays the internal JDBC URL.

Problem: DAS displays the internal JDBC URL on its About page instead of the correct JDBC URL to use to connect to the data warehouse.

Workaround: To copy the correct JDBC URL to use to connect to the data warehouse, in the Data Warehouse service Overview page, go to Virtual Warehouse , and then click Copy JDBC URL.

DWX-2592: DAS cannot parse certain characters in strings and comments.

Problem: DAS cannot parse semicolons (;) and double hyphens (--) in strings and comments. For example if you have a semicolon in a query such as the following, the query might fail:

```
SELECT * FROM properties WHERE prop_value = "name1;name2";
```

Queries with double hyphens (--) might also fail. For example:

```
SELECT * FROM test WHERE option = '--name';
```

Workaround: If a semicolon is present in a comment, then remove the semicolon before running the query or remove the comment entirely. For example:

```
SELECT * FROM test; -- SELECT * FROM test;
```

Should be changed to:

```
SELECT * FROM test; /* comment; comment */
```

In the same manner, remove any double-hyphens before running queries to avoid failure in DAS.

Older versions of Google Chrome browser might cause issues.

Problem: You might experience problems while using faceted search in older versions of the Google Chrome browser.

Workaround: Use the latest version (71.x or later) of Google Chrome.

BUG-94611: Visual Explain for the same query shows different graphs.

Problem: Visual Explain for the same query shows different graphs on the Compose page and the Query Details page.

Workaround: N/A

Known issues in Hue on public clouds

Learn about the known issues related to the Hue query editor in Cloudera Data Warehouse (CDW) service on public clouds, the impact or changes to the functionality, and the workaround.

TSB 2023-670: Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

Hue frontend for Apache Impala (Impala) and Apache Hive (Hive) Virtual Warehouses created in Cloudera Data Warehouse (CDW) can be stuck in a CrashLoopBackoff state on Microsoft Azure (Azure) platform, making it impossible to reach the Virtual Warehouse through Hue. In this case, the following error message is displayed: kubelet Error: failed to create containerd task: failed to create shim task: OCI runtime create failed: runc create failed: unable to start container process: exec: “run_httpd.sh”: cannot run executable found relative to current directory: unknown

The root cause of the issue is a recent node image upgrade in Azure made by Microsoft, where the version of the container runtime (containerd) was upgraded to 1.6.18. This version upgraded Go SDK to 1.19.6, where the behavior of how program execution works has changed. Due to security concerns, the newer version of Go SDK does not resolve a program using an implicit or explicit path entry relative to the current directory (for more information, see the Go documentation), and containerd indirectly uses this Go exec API. The Docker command of the Hue frontend relies on the former behavior. Therefore the command fails on recent node images, because it cannot find the expected executable in the current working directory of the containers anymore. Cloudera first noticed the aforementioned containerd version in the Azure Kubernetes Service (AKS) released on 2023-03-05. Every node image which is newer than the version released on 2023-03-05 is affected by this issue. Unfortunately, Cloudera does not have any insight to when Microsoft rolls out new node images in a given region. Therefore, it is possible that in some regions, the older node images

are still in use, where the issue does not arise until a given regional update is applied to the node images.

CDW environments, which are already activated and running, are not affected as long as customers do not trigger a node image upgrade to the latest available version either using Azure Command Line Interface (CLI) or on the Azure portal. If the node image is upgraded to the latest version, then the Virtual Warehouses also need to be upgraded to the latest CDWH-2023.0.14.0 version. Newly created CDW environments are not affected, but customers are advised to not choose a Hue version lower than CDWH-2023.0.14.0 during any Virtual Warehouse creation because such a configuration is affected by this issue.

Knowledge article

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-670: TSB Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

IMPALA-11447 Selecting certain complex types in Hue crashes Impala

Queries that have structs/arrays in the select list crash Impala if initiated by Hue.

Workaround: Do not select structs/arrays in Hue.

DWX-8460: Unable to delete, move, or rename directories within the S3 bucket from Hue

Problem: You may not be able to rename, move, or delete directories within your S3 bucket from the Hue web interface. This is because of an underlying issue, which will be fixed in a future release.

Workaround: You can move, rename, or delete a directory using the HDFS commands as follows:

1. SSH into your CDP environment host.
2. To delete a directory within your S3 bucket, run the following command:

```
hdfs dfs -rm -r [***COMPLETE-PATH-TO-S3-BUCKET***]/[***DIREC
TORY-NAME***]
```

3. To rename a folder, create a new directory and run the following command to move files from the source directory to the target directory:

```
hdfs dfs -mkdir [***DIRECTORY-NAME***]
```

```
hdfs dfs -mv [***COMPLETE-PATH-TO-S3-BUCKET***]/[***SOURCE-D
IRECTORY***] [***COMPLETE-PATH-TO-S3-BUCKET***]/[***TARGET-D
IRECTORY***]
```

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5650: Hue only makes the first user a superuser for all Virtual Warehouses within a Data Catalog

Problem: Hue marks the user that logs in to Hue from a Virtual Warehouse for the first time as the Hue superuser. But if multiple Virtual Warehouses are connected to a single Data Catalog, then the first user that logs in to any one of the Virtual Warehouses within that Data Catalog is the Hue superuser.

For example, consider that a Data Catalog DC-1 has two Virtual Warehouses VW-1 and VW-2. If a user named John logs in to Hue from VW-1 first, then he becomes the Hue superuser for all the Virtual Warehouses within DC-1. At this time, if Amy logs in to Hue from VW-2, Hue does not make her a superuser within VW-2.

None.

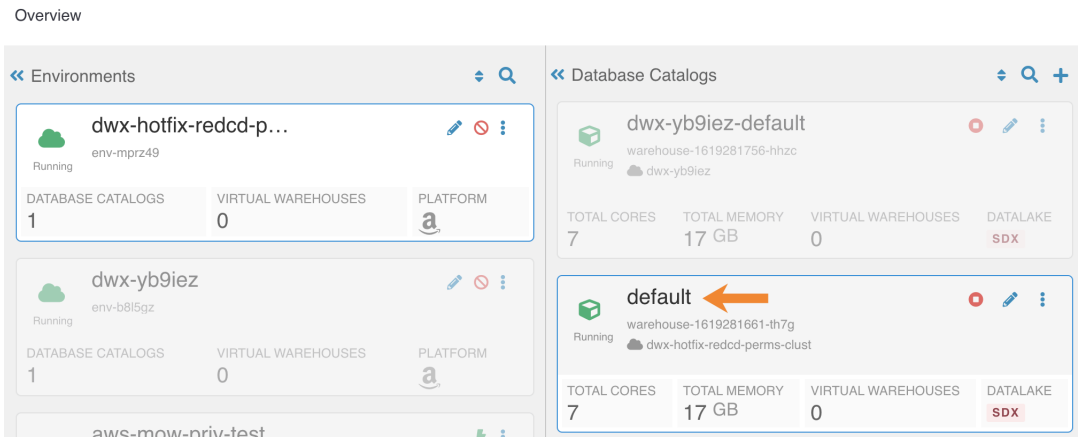
Known issues in Database Catalog on public clouds

Learn about the known issues related to Database Catalog in Cloudera Data Warehouse (CDW) service on public clouds, the impact or changes to the functionality, and the workaround.

DWX-7349: In reduced permissions mode, default Database Catalog name do not include the environment name

Problem:

When you activate an AWS environment in reduced permissions mode, the default Database Catalog name does not include the environment name:



This does not cause collisions because each Database Catalog named "default" is associated with a different environment. For more information about reduced permissions mode, see Reduced permissions mode for AWS environments.

Workaround: None available.

DWX-6167: Maximum connections reached when creating multiple Database Catalogs

Problem:After creating 17 Database Catalogs on one AWS environment, Virtual Warehouses failed to start.

Workaround: Limit the number of Database Catalogs created on one environment to 5. This applies to both AWS and Azure environments.

Known issues in Hive Virtual Warehouses on public clouds

Learn about the known issues related to Hive Virtual Warehouse in Cloudera Data Warehouse (CDW) service on public clouds, the impact or changes to the functionality, and the workaround.

DWX-10271: Missing log section in Hive query results

In a Hive Virtual Warehouse, when you run a query in Hue, the query results do not contain a logs section.

DWX-8118: INSERT INTO command fails under certain circumstances

This problem affects users who have a PostgreSQL database as the backend Hive database. If you create a table A and create a table B as select (CTAS) from an empty table A, inserting values into table B fails as follows:

```
Error while compiling statement: FAILED: Execution Error, return
 code 1 from
          org.apache.hadoop.hive.ql.exec.StatsTask.org.apache.
thrift.transport.TTransportException
```

Workaround: Disable auto-stats gathering: Go to Cloudera Manager Data Warehouse Virtual Warehouses , and click the Hive VW name in the list. In Configuration HiveServer2 , set hive.cbo .enable to false.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

# Add the following under the metadata.annotations field
service.beta.kubernetes.io/aws-load-balancer-ssl-negotiation-pol
icy: "ELBSecurityPolicy-TLS-1-2-2017-01"

# Save and quit the editor, and then run the following command to
 check your changes.

kubectl get svc nginx-service -n <cluster-name> -o yaml

# Make sure that the annotation you added is present.
```

DWX-5926: Cloning an existing Hive Virtual Warehouse fails

Problem: If you have an existing Hive Virtual Warehouse that you clone by selecting Clone from the drop-down menu, the cloning process fails. This does not apply to creating a new Hive Virtual Warehouse.

Workaround: Make the following configuration change to resolve this issue:

1. In the Hive Virtual Warehouse tile, click the edit icon. This launches the Virtual Warehouse details page.
2. In the details page for the Virtual Warehouse, click the Configurations tab.
3. Click the Hiveserver2 sub-tab.
4. Select hive-site from the configuration file drop-down list menu.
5. Search for the configuration property hive.metastore.sasl.enabled.
6. Set the hive.metastore.sasl.enabled configuration property to true.

   Note: If the hive.metastore.sasl.enabled configuration property is already set to true, delete the setting and re-enter it.
7. Click Apply in the upper right corner of the page to save the configuration.

8. Click the Actions menu and select Clone to clone the Hive Virtual Warehouse:

DWX-2690: Older versions of Beeline return SSLPeerUnverifiedException when submitting a query

Problem: When submitting queries to Virtual Warehouses that use Hive, older Beeline clients return an SSLPeerUnverifiedException error:

```
javax.net.ssl.SSLPeerUnverifiedException: Host name 'ec2-18-219-
32-183.us-east-2.compute.amazonaws.com' does not
match the certificate subject provided by the peer (CN=*.env-c2
5dsw.dwx.cloudera.site) (state=08S01,code=0)
```

Workaround: Only use Beeline clients from CDP Runtime version 7.0.1.0 or later.

Known issues in Impala Virtual Warehouses on public clouds

Learn about the known issues related to Impala Virtual Warehouse in Cloudera Data Warehouse service on public clouds, the impact or changes to the functionality, and the workaround.

IMPALA-11039 Impala Virtual Warehouse instability

Problem: Executors can crash during the Parquet scans

Workaround: Disable Parquet late materialization in CDW version 2021.0.6-b96. To the default_query_options startup option, add the following property setting:

```
PARQUET_LATE_MATERIALIZATION_THRESHOLD=-1
```

Upgrade to the next CDW version when available.

CDPD-38204 Top-N query failure

Problem: A Top-N operator crashes in the Impala Virtual Warehouse

Workaround: Set the analytic rank pushdown threshold to zero.To the default_query_options startup option, add the following property:

```
ANALYTIC_RANK_PUSHDOWN_THRESHOLD=0
```

Upgrade to the next CDW version when available.

IMPALA-11045 Impala Virtual Warehouses might produce an error when querying transactional (ACID) table even after you enabled the automatic metadata refresh (version DWX 1.1.2-b2008)

Problem: Impala doesn't open a transaction for select queries, so you might get a FileNotFound error after compaction even though you refreshed the metadata automatically.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Impala Virtual Warehouses might produce an error when querying transactional (ACID) tables (DWX 1.1.2-b1949 or earlier)

Problem: If you are querying transactional (ACID) tables with an Impala Virtual Warehouse and compaction is run on the compacting Hive Virtual Warehouse, the query might fail. The compacting process deletes files and the Impala Virtual Warehouse might not be aware of the deletion. Then when the Impala Virtual Warehouse attempts to read the deleted file, an error can occur. This situation occurs randomly.

Workaround: Run the INVALIDATE METADATA statement on the transactional (ACID) table to refresh the metadata. This fixes the problem until the next compaction occurs. For information about running this statement, see INVALIDATE METADATA statement.

Do not use the start/stop icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier

Problem: If you use the stop/start icons in Impala Virtual Warehouses version 7.2.2.0-106 or earlier, it might render the Virtual Warehouse unusable and make it necessary for you to re-create it.

Workaround: Do not use the stop/start icons in these older Virtual Warehouses. Instead, these older versions automatically suspend and resume the Impala executors depending on the absence or presence of queries, making manual start or stop unnecessary.

DWX-6674: Hue connection fails on cloned Impala Virtual Warehouses after upgrading

Problem: If you clone an Impala Virtual Warehouse from a recently upgraded Impala Virtual Warehouse, and then try to connect to Hue, the connection fails.

Workaround: Create a new Impala Virtual Warehouse and do not clone from a recently upgraded warehouse. Then the connection to Hue from the new Impala Virtual Warehouse succeeds.

DWX-5841: Virtual Warehouse endpoints are now restricted to TLS 1.2

Problem: TLS 1.0 and 1.1 are no longer considered secure, so now Virtual Warehouse endpoints must be secured with TLS 1.2 or later, and then the environment that the Virtual Warehouse uses must be reactivated in CDW. This includes both Hive and Impala Virtual Warehouses. To reactivate the environment in the CDW UI:

1. Deactivate the environment. See Deactivating AWS environments or Deactivating Azure environments.
2. Activate the environment. See Activating AWS environments or Activating Azure environments

Workaround: If environment reactivation is not possible, you can perform manual steps using the kubectl command line tool to pick up the TLS 1.2 endpoint change. Open a terminal window on a system where the kubectl command line tool is installed, log in, and run the following commands:

```
kubectl edit svc nginx-service -n <cluster-name>

# Add the following under the metadata.annotations field
service.beta.kubernetes.io/aws-load-balancer-ssl-negotiation-pol
icy: "ELBSecurityPolicy-TLS-1-2-2017-01"

# Save and quit the editor, and then run the following command to
 check your changes.

kubectl get svc nginx-service -n <cluster-name> -o yaml

# Make sure that the annotation you added is present.
```

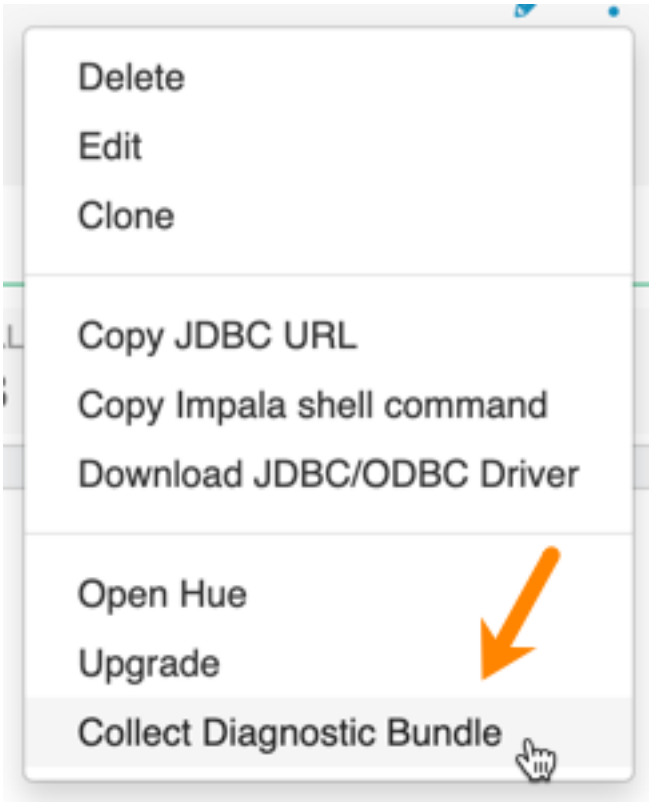
DWX-5276: Upgrading an older version of an Impala Virtual Warehouse can result in error state

Problem: If you upgrade an older version of an Impala Virtual Warehouse (DWX 1.1.1.1-4) to the latest version, the Virtual Warehouse can get into an Updating or Error state.

Workaround: none

DWX-3914: Collect Diagnostic Bundle option does not work on older environments

The Collect Diagnostic Bundle menu option in Impala Virtual Warehouses does not work for older environments:



Data caching:

This feature is limited to 200 GB per executor, multiplied by the total number of executors.

Sessions with Impala continue to run for 15 minutes after the connection is disconnected.

When a connection to Impala is disconnected, the session continues to run for 15 minutes in case so the user or client can reconnect to the same session again by presenting the session_token. After 15 minutes, the client must re-authenticate to Impala to establish a new connection.

Behavior changes (unsupported releases)

Behavior changes in unsupported releases.

November 17, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following behavior changes:

Summary:

Removal of the engine.hive.enabled property

Before this release:

The engine.hive.enabled property was set to "true" to enable the creation of Hive-enabled Iceberg table.

After this release:

The engine.hive.enabled property is removed because Hive is already supported and SQL engines are not required to explicitly specify this property.

Summary:

Change in Backup and Restore

Before this release:

When backing up and restoring CDW, temporarily deploying at least one Virtual Warehouse that runs 2023.0.14.0-15 or later to your environment was not needed to back up Hue.

After this release:

Under certain conditions, you must deploy a Virtual Warehouse that runs 2023.0.14.0-15 or later for a successful Hue backup. For more information, see "Automatically backing up the environment" and "Backing up Hue".

Summary:

Change in supported Instance Metadata Service (IMDS) version

Before this release:

CDP supported AWS EC2 instance metadata from a running instance using only Instance Metadata Service V1 (IMDSv1).

After this release:

To enhance security in CDW clusters, CDP supports Instance Metadata Service V2 (IMDSv2) only in CDW clusters. CDP provisions instances for CDW clusters with IMDSv2. CDP no longer supports IMDSv1. Cloudera has removed the capability to access metadata using IMDSv1 in CDW clusters.

Summary:

Change in the way dates are parsed from string by ignoring trailing invalid characters

Before this release:

HIVE-20007 introduced changes in the way dates were parsed from strings. SQL functions or date operations involving invalid dates returned "null".

After this release:

HIVE-27586 extracts and returns a valid date from a string value if there is a valid date prefix in the string. This fix partially restores the behavior changes introduced as part of HIVE-20007 and also makes the current behavior of handling trailing invalid characters more consistent.

The following table illustrates the behavior changes before and after the fix:

Strong value	Behavior (before HIVE-20007)	Previous behavior (after HIVE-20007)	Current behavior (after HIVE-27586)
2023-08-03_16:02:00	2023-08-03	null	2023-08-03
2023-08-03-16:02:00	2023-08-03	null	2023-08-03
2023-08-0316:02:00	2024-06-11	null	2023-08-03
03-08-2023	0009-02-12	null	0003-08-20
2023-08-03 GARBAGE	2023-08-03	2023-08-03	2023-08-03
2023-08-03TGARBAGE	2023-08-03	2023-08-03	2023-08-03
2023-08-03_GARBAGE	2023-08-03	null	2023-08-03

This change affects various Hive SQL functions and operators that accept dates from string values, such as CAST (V AS DATE), CAST (V AS TIMESTAMP), TO_DATE, DATE_ADD, DATE_DIFF, WEEKOFYEAR, DAYOFWEEK, and TRUNC.

Summary:

Change in the way date and timestamp values are parsed.

Before this release:

Some of the Hive date and timestamp functions, such as unix_timestamp(), from_unixtime(), date _format(), and cast() were enhanced to use the DateTimeFormatter class for printing and parsing date and timestamp objects. Prior to this change, these functions used the SimpleDateFormat class.

After this release:

Starting from this release, a new configurable hive.datetime.formatter property is introduced through HIVE-25576 that enables you to choose between SimpleDateFormat and DateTimeForm atter for the unix_timestamp, from_unixtime, and date_format SQL functions.

Although the DateTimeFormatter class is an improvement over SimpleDateFormat, some users may want to retain the old behavior to ensure compatibility after migration, therefore, making it necessary for introducing this property.

The possible values for the hive.datetime.formatter property are 'DATETIME' and 'SIMPLE' representing DateTimeFormatter and SimpleDateFormat respectively. The default value is set to 'DATETIME'.

Summary:

Change in creation of Database Catalog

Before this release:

Cloudera Data Warehouse (CDW) supported default and non-default Database Catalogs for the latest release. When creating a Database Catalog you selected the type of data lake, depending on how you intend to use CDW.

After this release:

CDW supports default Database Catalogs only, and when you activate the Virtual Warehouse, the Database Catalog is created. A data lake is transparently associated with your Database Catalog based on your activation settings, such as managed storage access, and other factors. You no longer need to select a type of data lake.

August 30, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following behavior changes:

Before this release: There was just a single Web-UI corresponding to the Active Catalog instance.
After this release: Both active and passive Catalog's Web UI are now accessible to the end user.

Prior to this release HA was supported through K8s based leader election. This caused some behavior changes in the Impala Catalog Web UI. Both active and passive Catalog's Web UI are now accessible to the end user. In the past there was just a single Web-UI corresponding to the Active Catalog instance.

See the known issue and workaround related to this change.

February 7, 2023

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud has the following behavior changes:

Summary:

Change in Grafana metrics collection

Before this release:

Many metrics for Grafana monitoring were being stored for 45 days without limits on the disk space consumed. Resource overruns could occur and impact the availability of the entire environment.

After this release:

Storing metrics longer than 15 days or consuming more than 90GB of disk space is not supported. Metrics older than 15 days will be deleted. If the stored metrics consume more than 90GB of disk space, metrics will be deleted regardless of the number of days stored.

See also Grafana Limitations documentation.

Fixed issues (unsupported releases)

Fixed issues in unsupported releases.

January 18, 2024 Fixed issues

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud fixes these issues.

Fixed the problem announced in TSB 2024-733: Changes in CoreDNS of AWS EKS affect operations in CDW, CDE and CML clusters on Public Cloud

This issue that caused a malfunction in CoreDNS that affects the Cloudera Data Warehouse (CDW), Cloudera Data Engineering (CDE) and Cloudera Machine Learning (CML) cluster operations has been fixed.

December 19, 2023 Fixed issues

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud fixes these issues.

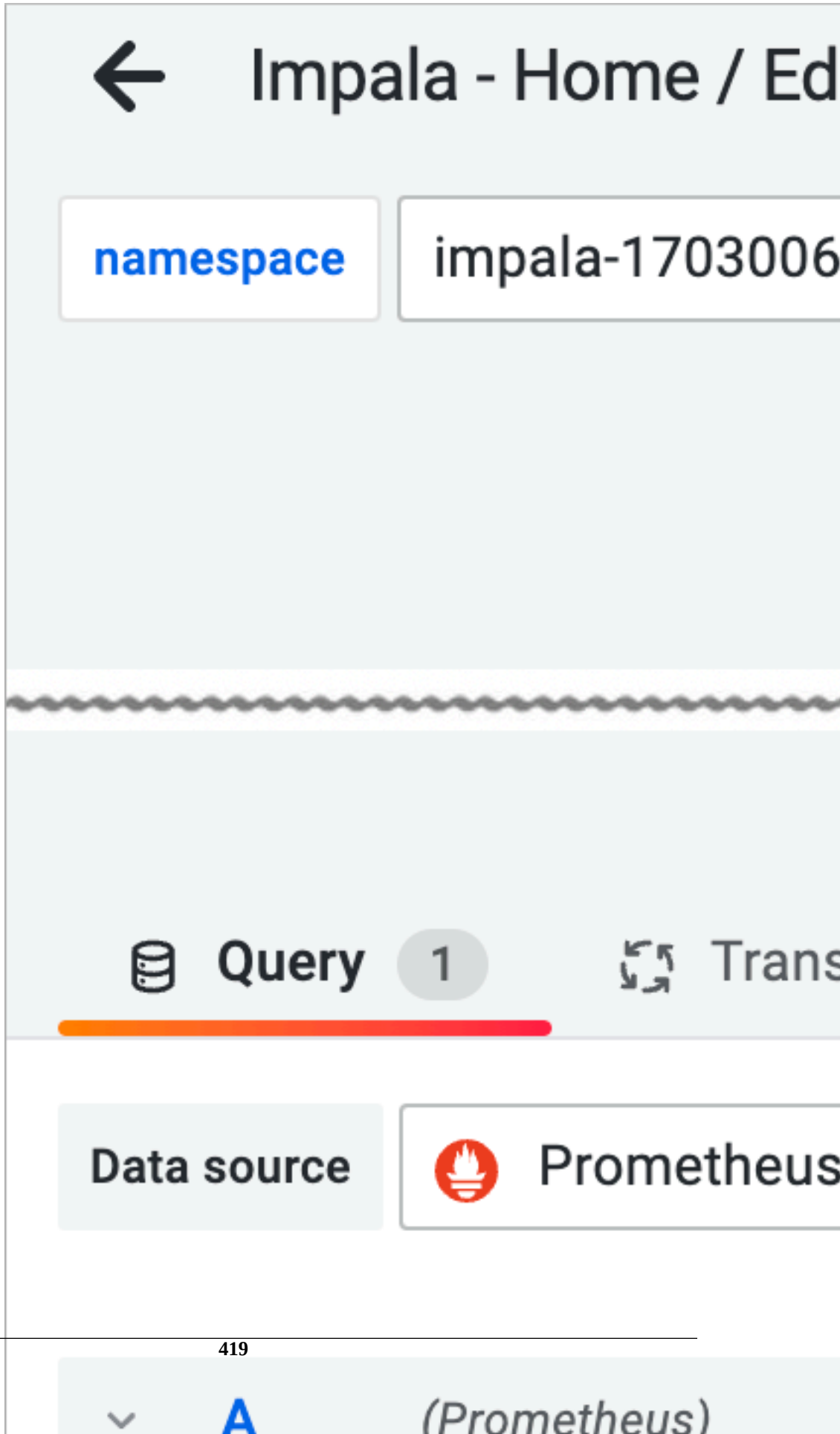
CDPD-48059: Unable to view and access S3 buckets from Hue configured with RAZ in CDW Public Cloud

Earlier, when you configured access to AWS S3 buckets using RAZ in the following regions, you couldn't access S3 buckets from the Hue File Browser: cn-*, eu-central, ap-northeast-2, ap-south-1, us-east-2, ca-central and eu-west. Due to a bug in AWS Boto SDK, accessing S3 buckets from Hue using RAZ was blocked. This issue has been fixed.

DWX-17111 IMPALA-12486 Missing database and table count metrics in Grafana

Catalog metrics for metadata loading have been added to Impala. Getting the fix requires a new environment activation using 1.8.2-b221 Dec 19, 2023 and Impala Virtual Warehouse with 2023.0.16.2-1 (Dec 19, 2023). If activating a new environment is not an option, then you can manually update Grafana charts with database and table counts to make the following change to the Prometheus query:

FROM:


Impala - Home / Ed
namespace
impala-1703006
Query 1
Trans
Data source
Prometheus
A
(Prometheus)

TO: app="coordinator"

Impala Virtual Warehouse version 2023.0.16.2-1 (Dec 19, 2023) is required to make this manual update.

Technical Service Bulletins

TSB 2024-723: Hue RAZ is using logger role to Read and Upload/Delete (write) files

For the latest update on this issue see the corresponding Knowledge Article: TSB 2024-723: Hue Raz is using logger role to Read and Upload/Delete (write) files.

December 1, 2023 Fixed issues

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud fixes these issues.

CDPD-62960: From the left assist panel, the pop-up link to Table Browser displays error 500

Earlier, when you clicked the i button next to one of the tables in the left assist panel and then clicked on Table Browser on the pop-up window, Hue displayed a 500 error and then the Table Browser auto reloaded. This issue has been fixed.

CDPD-63429: Manual configuration needed to support Bedrock using the Boto3 library

Bedrock is now supported with the public version of the Boto3 library. Anthropic Claude is the suggested model to use with Bedrock. Add the following lines in the hue-safety-valve field to get Hue SQL AI Assistant working with Anthropic Claude:

```
[aws]
[[bedrock_account]]
access_key_id = [***ACCESS-KEY***]
secret_access_key = [***SECRET-KEY***]
region = us-east-1
[desktop]
[[ai_interface]]
service='bedrock'
model='claude'
```

November 20, 2023 Fixed issues

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud fixes these issues.

DWX-15859 Issue backing up and restoring Workload Aware Auto-Scaling (WAAS)

You can now use the CDW backup and restore process to restore the WAAS configurations. The issue related to restoring the WAAS configuration is fixed.

IMPALA-12233 MT_DOP issue

Impala query with JOIN and LIMIT no longer hangs when MT_DOP is greater than 0.

DWX-16275 Slow Data Scan on Azure for Impala Virtual Warehouses

The Impala Virtual Warehouse performance issue when using CDW Runtime 2023.0.15.0-243 - 2023.0.15.1-2 is fixed in this release 2023.0.16.2-1 (released December 19, 2023).

DWX-16825 Impala core-site configurations mutability issue

Hadoop core-site and the Hive site properties in the Impala Virtual Warehouse, previously immuatble, are now mutable. In the Impala Virtual Warehouse, click Options ⋮ Edit Configurations . Search or scroll to the property you want to edit and change the value.

VIZ-2269 CDW Virtual Warehouse connection to Data Visualization error

The "not authorized to delegate" error caused by impersonation configurations no longer occurs when creating or editing the Virtual Warehouse connection. For more information, see Data Visualization release notes.

CDPD-61589 Problem reading a file downloaded from Hue on ABFS

Downloading a file from the Hue File Browser, which is configured to use ABFS, no longer causes file corruption. The problem caused by Hue reading 1 byte less for every 1 MB of file content fetched has been fixed. Hue now reads the exact 1 MB chunk size. There is no byte loss.

CDPD-61550 HUE ABFS browser listing limitation

The Hue ABFS browser is no longer limited to listing 5000 files maximum. You can list an unlimited number of files.

October 5, 2023 Fixed issues

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud fixes these issues.

DWX-16424 Impala Web UI incompatibility

The problem that causes two URLs for the Web UI to appear when catalogd ran in HA mode has been fixed. Only a single URL appears now.

August 30, 2023 Fixed issues

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud fixes these issues.

DWX-15181 SAML Auth via ODBC/JDBC does not work with active-active coordinator

The JDBC authentication problem that occurred when the Impala coordinator was configured for high availability (HA) has been fixed.

DWX-15576 Impala autoscaler ConfigMap watcher can return 'unexpected type' error

Autoscaler code was treating every watch event in the same way, which caused errors of the 'unexpected type'. This issue has been fixed to handle different event types appropriately.

DWX-15110 Misaligned dialog in the UI

The Create Virtual Warehouse dialog is now aligned when no Virtual Warehouses are running.

DWX-14970 Cosmetic UI problem with Virtual Warehouse start label

The start labels of the Impala Virtual Warehouse and Hive Virtual Warehouse are now consistent with regard to upper- and lower-case.

DWX-14874 An indicator shows that no Virtual Warehouses running

An indicator is displayed when no Virtual Warehouses are running.

DWX-15171 Disabling SSO for Impala VW lands does not cause an error state

DWX-15749 The Restart dbc cli command has been implemented for Azure

DWX-14334 Impala coordinator, catalog, executors, and statestore CPU graphs in Grafana have been fixed

DWX-15362 The latest version of Impala shell 4.3.0a3 can be downloaded from the Impala shell UI link

DWX-15218 Duplicate application configurations have been removed from Impala Virtual Warehouses in Unified Analytics mode

DWX-15219 Filtering of hadoop-core-site configuration files now works

May 5, 2023 fixed issues

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud fixes these issues.

TSB 2023-670: Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

For the latest update on this issue see the corresponding Knowledge article: TSB 2023-670: Hue frontend may become stuck in CrashLoopBackoff on CDW running on Azure

CDPD-48893: Hue logs get overwritten without a clear root cause

In previous implementations, multiple file handlers would write to a single log file, causing the Hue logs to be overwritten. Hue now uses a socket handler, which solves this problem.

DWX-14630 Issue creating a Virtual Warehouse using the CLI

A problem when creating an Impala Virtual Warehouse using the CLI in 1.6.2 has been fixed. The problem caused the Impala Coordinator auto shutdown to fail when HA was disabled.

DWX-8980 Data consistency problem occurs when sharing a Database Catalog

Events are now raised in the Hive metastore when a refresh/invalidate command is run. Catalog daemons process events synchronously across all Virtual Warehouses that share metadata. The catalog cache across all the Virtual Warehouses that share a Database Catalog is synchronized and consistent.

Feb 7, 2023 fixed issues

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud fixes these issues.

Technical Service Bulletin Nov 2022

CDP CLI CDW command issue when activating an Azure DWX environment

When clients try to activate a Microsoft Azure (Azure) environment on Cloudera Data Platform (CDP) Cloudera Data Warehouse (CDW) for Public Cloud using the CDP Command Line Interface (CLI) dw command's create-cluster sub-command with the --use-private-load-balancer switch, the created load balancer will be public instead of private.

Example command:

cdp \

--profile ${PROFILE} \

dw create-cluster \

--environment-crn ${ENV} \

--use-private-load-balancer

For the latest update on this issue see the corresponding Knowledge article:

TSB 2022-642: CDP CLI CDW command issue when activating an Azure DWX environment

December 13, 2022 fixed issues

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud fixes these issues.

IIMPALA-11721 Impala repeated retries query of updated Iceberg table

In Cloudera Data Warehouse (CDW) Runtime 2022.0.12.0-90 released 2022-12-13, the problem has been fixed that caused Iceberg table loading to fail in local catalog mode if the table gets updated frequently.

November 22, 2022 fixed issues

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud fixes these issues.

TSB 2023-656: Cloudera Data Warehouse does not clean up old Helm releases

For the latest update on this issue see the corresponding Knowledge article: 2023-656: Cloudera Data Warehouse does not clean up old Helm releases

September 15, 2022 fixed issues

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud fixes these issues.

DWX-13094 Diagnostic bundle job fails with missing service account

In Cloudera Data Warehouse (CDW) v1.4.2, the diagnostic bundle job no longer fails if it was used before with any previous versions of CDW. Jobs from the latest version run with the correct name of the service account installed after the upgrade.

August 4, 2022 fixed issues

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud fixes these issues.

DWX-12787 DWX Access Control - Unable to Update the VW with UserGroup

The Virtual Warehouse is now updated with UserGroups.

DWX-12669 HEADROOM property editing disabled for single node Hive Virtual Warehouse and WAIT TIME takes precedence

This issue fixes potential problems caused by setting HEADROOM properties when using a single node Hive VW and by making HEADROOM and WAIT TIME settings mutually exclusive. WAIT TIME settings now take precedence over HEADROOM settings when both are configured.

CDPD-40730 PARQUET-1682 Forward compatibility for TIME/TIMESTAMP

PARQUET-1682 was backported to maintain forward compatibility for TIME/TIMESTAMP. This resolves the incompatibility between Hive's direct Parquet dependency and the transitive Parquet versions brought in by Iceberg.

April 27, 2022 fixed issues

This release of the Cloudera Data Warehouse (CDW) service on CDP Public Cloud fixes these issues.

CDPD-35041 Security Vulnerability with SSO for Hive Virtual Warehouses in CDW Public Cloud

This release fixes the SSO-related problem that occurred only in Hive Virtual Warehouses in Cloudera Data Warehouse version 2021.0.6-b96 released on March 8, 2022. This problem affected tenants who had the CDW_ALLOW_SSO_FOR_LLAP entitlement. In a misconfigured Hive Virtual Warehouse, unauthorized users who knew the JDBC connection string and had a valid userid, could execute arbitrary SQL queries on a Hive Virtual Warehouse. This security vulnerability no longer exists. Hive now properly handles the situation by using LDAP authentication as follows: If SSO authentication is enabled for a Hive Virtual Warehouse and a user bypasses authentication altogether by removing the auth=browser option from the JDBC connection string, Hive uses LDAP authentication.

CVE-2021-44228 remediation for CDW

March 8, 2022 Update

Cloudera has released a new version of CDW that upgrades the embedded Log4j version to 2.17.1.

December 22, 2021 Update

Cloudera has released a new version of CDW which upgrades the embedded Log4j version to 2.16. This provides a permanent fix for CVE-2021-44228.

To use this version you must upgrade your Database Catalog(s) and Virtual Warehouse(s) to the latest version, which is 2021.0.4.1-3. Additionally, you can create new Virtual Warehouses using the latest version, and those will also have the fix for this vulnerability.

If you had previously applied the configurations shared below, then those are no longer required. Please reach out to Cloudera Support if you have additional questions.

Introduction

As mentioned in Cloudera Technical Service Bulletin 2021-545 (Critical vulnerability in log4j2 CVE-2021-44228), the Cloudera Data Warehouse (CDW) data service is impacted by the recent Apache Log4j2 vulnerability. As per that bulletin:

The Apache Security team has released a security advisory for CVE-2021-44228 which affects Apache Log4j2. A malicious user could exploit this vulnerability to run arbitrary code as the user or service account running the affected software. Software products using log4j versions 2.0 through 2.14.1 are affected and log4j 1.x is not affected. Cloudera is making short-term workarounds available for affected software and is in the process of creating new releases containing fixes for this CVE.

Short Term Resolution - CDW

If you are using the CDW data service in CDP Public Cloud, please follow the steps below to implement an immediate manual workaround. This will change the configurations within your running CDW environment so that this vulnerability is mostly, but not completely mitigated. There will still remain some potential attack vectors, as per the official Apache guidance.

Note that this will not upgrade the specific Log4j libraries to a version that does not have this vulnerability. Instead, it protects against most attack vectors via configuration changes. See the “Long Term Resolution - CDW” section below for information on what Cloudera is doing for the permanent fix to the CDW data service.

These manual workaround steps need to be applied to already-running kubernetes pods. Some of the steps involve running kubectl commands to edit the configurations of the pods. Once these kubectl commands are run, if you edit the corresponding Virtual Warehouse (VW) or Database Catalog (DBC) - via the CDW UI or CLI - then the changes you made via kubectl will be overwritten. So either refrain from making further edits to the VW/DBC, or re-apply the kubectl commands.

If, after running the commands on existing pods, new pods are provisioned, then the commands will have to be run on those new pods as well. This would be required in scenarios such as

- Create new VW
- Existing VW that was stopped is now started (manually or via autostart)
- Already-running VW is autoscaled up
- Need to edit some other VW config (via the CDW UI or CLI), in which case the VW will be restarted with new pods that have the original configurations as they existed prior to having run the kubectl commands
- Upgrade of an environment

We understand that it puts additional operational burden to manage this vulnerability while the release with the permanent fix is delivered. If you prefer to simplify the operation while still keeping your CDW environment online, then we recommend you disable autosuspend and auto-scale until the permanent fix is released. If autosuspend is disabled, then you would need to manually suspend the VW whenever it is not needed to avoid extra cost; then when it starts back up the kubectl commands need to be run again on the new pods.

Please see Appendix A below for the manual configuration workaround steps.

Long Term Resolution - CDW

See the “December 22, 2021 Update” section above for details on the permanent fix for CVE-2021-44228. Please note that Cloudera will be releasing a subsequent version containing Log4j 2.17 in January. Moving forward, Cloudera will continue to monitor for other CVEs that are found, and will release updates appropriately to address them.

Cloudera JDBC and ODBC Drivers

The ODBC drivers for Impala and Hive are not affected by this vulnerability.

The Apache Hive JDBC driver uses the SLF4J API, but does not bundle Log4j as a logging provider. If an end user or a client application makes Log4j available as a provider when using the driver, then the user/client must make sure that they are using a Log4j version that is not affected by this vulnerability.

The following JDBC drivers for Impala and Hive are affected. Of those published on the Cloudera Downloads page, only the following versions are impacted. Each of these use Log4j version 2.12.1 or 2.13.3.

Hortonworks JDBC Driver for Apache Hive

- v2.6.12

Hive JDBC Driver

- v2.6.12
- v2.6.13
- v2.6.14

- v2.6.15

Impala JDBC Driver

- v2.6.18
- v2.6.19
- v2.6.20
- v2.6.21
- v2.6.22
- v2.6.23
- v2.6.24

Note that earlier versions of these JDBC drivers use Log4j versions that are not affected by this vulnerability, and can be ignored.

Short Term Resolution - Drivers

Remove the JndiLookup class from the driver jar file. The following command shows an example of how to do this, using the Impala JDBC 2.6.24 version of the driver. This is one command, with all parts to be included on a single line.

```
zip -q -d ~/JDBC-Drivers/ClouderaImpalaJDBC-2.6.24.1029/ClouderaImpalaJDBC42
-2.6.24.1029/ImpalaJDBC42.jar
com/cloudera/impala/jdbc42/internal/apache/logging/log4j/core/lookup/Jndi
Lookup.class
```

More information on this workaround is provided by the Apache Software Foundation here: https://logging.apache.org/log4j/2.x/security.html

Long Term Resolution - Drivers

New versions of the drivers will be released shortly that are not affected by this vulnerability. As these become available Cloudera will notify customers and make the new drivers available on the Cloudera Downloads page.

If there are further questions related to configuration of these drivers, please open a support case with Cloudera.

Appendix A - Manual Configuration Workaround

In several of the steps below it is required to use kubectl. In order to do this you will need the kubeconfig for the kubernetes cluster. You can see how to do this on AWS here, and on Azure here. Also, the kubectl commands require an ID for the virtual warehouse that is being updated. This ID is of the form “compute-##########-####” or “impala-##########-####”, and it can be found underneath the VW name on the CDW Overview page.

Please note that the configuration changes made via the web console (for Data Analytics Studio, Hive Metastore, and Hue) must be completed before the remaining changes, done via kubectl, are applied. Once the web console changes are made, be sure to wait for the DBC/VW to restart completely before continuing on with the kubectl commands.

Data Analytics Studio

1. Go to the DBC
2. Edit
3. Configuration # Das Event Processor # env # Add key

   LOG4J_FORMAT_MSG_NO_LOOKUPS=true
4. Apply

Hive Metastore

1. Go to the DBC
2. Edit

3. Configuration # Metastore # env # SERVICE_OPTS, append

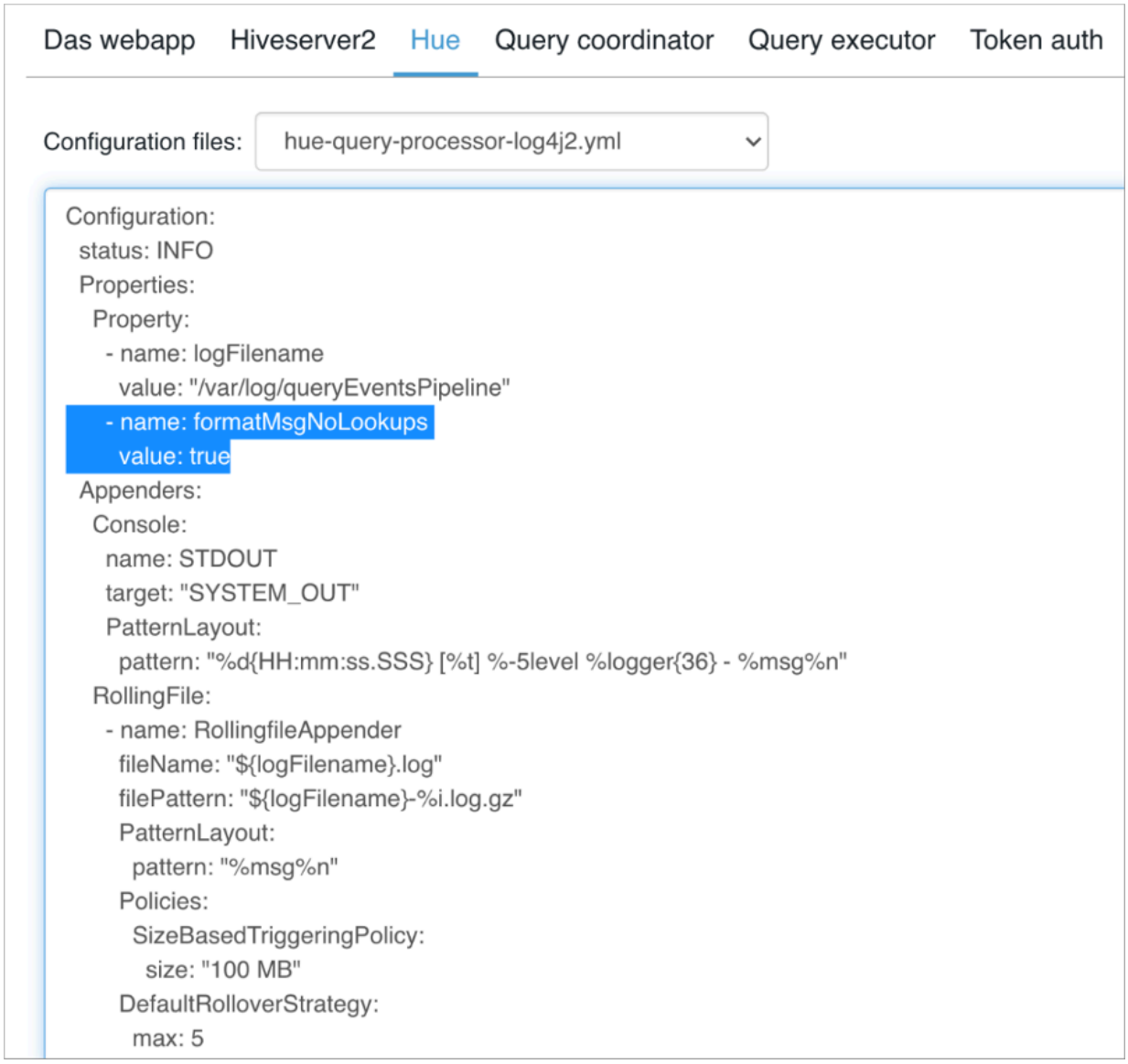
   " -Dlog4j2.formatMsgNoLookups=true" (with a space in front of -D)

   Note: if the env doesn't show SERVICE_OPTS, select any other configuration file and then select "env" again.
4. Apply

Hue

1. Go to the Hive VW
2. Edit
3. Configuration # Hue # env #hue-query-processor-log4j2.yaml # add the following under Property

```
-name: formatMsgNoLookups
 value: true
```



4. Apply

Hive Virtual Warehouse: Hive Server2

Note: These same steps are needed if you have an Impala Virtual Warehouse running in Unified Analytics mode (currently in Tech Preview). In that mode there is a Hive Server2 running within the Impala Virtual Warehouse, and it needs to be updated.

1. Fetch the statefulsets for hiveserver2 pod using “kubectl”

```
$ kubectl -n compute-1639516105-l76g get sts
NAME                          READY   AGE
das-webapp                    1/1     4h20m
hiveserver2                   1/1     4h20m
huebackend                    1/1     4h20m
hueep                         1/1     4h20m
hueqp                         1/1     4h20m
query-coordinator-0           2/2     4h20m
query-executor-0              2/2     4h20m
standalone-compute-operator   1/1     4h20m
```

2. Edit the statefulset and append -Dlog4j2.formatMsgNoLookups=true to the HADOOP_CLIENT_OPTS for container with EDWS_SERVICE_NAME=hiveserver2

```
kubectl -n compute-1639516105-l76g edit sts hiveserver2
```

In the editor, change the following section:

```
containers:
      - env:
        - name: EDWS_SERVICE_NAME
          value: hiveserver2
        - name: HADOOP_CLIENT_OPTS
          value: -DGroup_ID=compute-1639516105-l76g
-Dorg.wildfly.openssl.path=/usr/lib64
            -Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSL
v3,GCM
-Xms8192M   -Xmx11468M -Xss512k -Xloggc:/var/log/hive/hiveserver2-gc-%t.
log
-XX:+UseG1GC
            -XX:+PrintGCDetails -XX:+PrintGCTimeStamps -XX:+PrintGCCause
-XX:+UseGCLogFileRotation
            -XX:NumberOfGCLogFiles=10 -XX:GCLogFileSize=10M
-Dlog4j2.formatMsgNoLookups=true
```

3. This will restart the containers automatically. Wait until all the containers have restarted
4. To confirm that the change has taken effect, ssh to the container and execute something like “ps auxww”. This shows the hiveserver2 process and the command line should include the '-Dlog4j2.formatMsgNoLookups=true' property, which was added above.

```
hive         1  4.6  3.7 14128040 1228016 ?    Ssl  01:31   0:30
/etc/alternatives/jre/bin/java -Dproc_jar -Dproc_hiveserver2
-DGroup_ID=compute-1639516105-l76g -Dorg.wildfly.openssl.path=/usr/lib64
-Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM
-Xms8192M -Xmx11468M -Xss512k -Xloggc:/var/log/hive/hiveserver2-gc-%t.log
-XX:+UseG1GC -XX:+PrintGCDetails -XX:+PrintGCTimeStamps -XX:+PrintGCCause
-XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=10 -XX:GCLogFileSize=10M
-Dlog4j2.formatMsgNoLookups=true
-javaagent:/jmx-exporter/jmx_prometheus_javaagent-0.11.0.jar=35000:/jmx-ex
porter/config.yml -Dlog4j.configurationFile=hive-edws-log4j2.properties
-Djava.util.logging.config.file=/usr/lib/hive/bin/../conf/parquet-loggi
ng.properties -Dyarn.log.dir=/usr/lib/hadoop/logs -Dyarn.log.file=hadoop
.log
-Dyarn.home.dir=/usr/lib/hadoop-yarn -Dyarn.root.logger=INFO,console
-Djava.library.path=/usr/lib/hadoop/lib/native
```

```
-Dhadoop.log.dir=/usr/lib/hadoop/logs -Dhadoop.log.file=hadoop.log -Dhadoo
p.home.dir=/usr/lib/hadoop -Dhadoop.id.str=hive
-Dhadoop.root.logger=INFO,console -Dhadoop.policy.file=hadoop-policy.xml
-Dhadoop.security.logger=INFO,NullAppender org.apache.hadoop.util.RunJar
/usr/lib/hive/lib/hive-service-3.1.3000.2021.0.5-b17.jar
org.apache.hive.service.server.HiveServer2
```

Hive Virtual Warehouse: Query Co-ordinator

1. Fetch the statefulsets for hiveserver2 pod using “kubectl”

```
$ kubectl -n compute-1639516105-l76g get sts
NAME                          READY   AGE
das-webapp                    1/1     4h20m
hiveserver2                   1/1     4h20m
huebackend                    1/1     4h20m
hueep                         1/1     4h20m
hueqp                         1/1     4h20m
query-coordinator-0           2/2     4h20m
query-executor-0              2/2     4h20m
standalone-compute-operator   1/1     4h20m
```

2. Edit the statefulset and append -Dlog4j2.formatMsgNoLookups=true to the JAVA_OPTS for container with EDWS_SERVICE_NAME=query-coordinator

```
kubectl -n compute-1639516105-l76g edit sts query-coordinator-0
containers:
      - env:
        - name: EDWS_SERVICE_NAME
          value: query-coordinator
        - name: JAVA_OPTS
          value: -Dorg.wildfly.openssl.path=/usr/lib64 -Dzookeeper.sasl
.client=false
            -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Xms2048M -Xmx2867M -se
rver
-Djava.net.preferIPv4Stack=true
            -XX:NewRatio=8 -XX:+UseNUMA -XX:+ResizeTLAB
-Xloggc:/var/log/hive/query-coordinator-gc-%t.log
            -XX:+UseG1GC -XX:+PrintGCDetails -XX:+PrintGCTimeStamps -ve
rbose:gc
-XX:+PrintGCCause
            -XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=10
-XX:GCLogFileSize=10M
            -Dlog4j2.formatMsgNoLookups=true
```

3. This will restart the containers automatically. Wait until all the containers have restarted
4. To confirm that the change has taken effect, ssh to the container and execute something like “ps auxww”. This shows the query coordinator process and the command line should include the '-Dlog4j2.formatMsgNoLookups=true' property, which was added above.

```
hive         1  5.9  2.8 8917460 918892 ?      Ssl  01:47   0:24
/etc/alternatives/jre/bin/java -Dproc_querycoordinator -classpath
/etc/hive/conf:/custom-jars/*:/custom-jars/lib/*:/usr/lib/hive/lib/*:/etc
/tez/conf:/usr/lib/tez/*:/usr/lib/tez/lib/*:/etc/hadoop/conf:/usr/lib/ha
doop/lib/*:/usr/lib/hadoop/.//*:/usr/lib/hadoop-hdfs/lib/*:/usr/lib/hado
op-hdfs/.//*:/usr/lib/hadoop-mapreduce/.//*:/usr/lib/hadoop-yarn/lib/*:/
usr/lib/hadoop-yarn/.//*:/efs/udf-jars/conf:/efs/udf-jars/lib/udfs/* -se
rver -Djava.net.preferIPv4Stack=true
-Dlog4j.configurationFile=tez-edws-log4j2.properties
-DisThreadContextMapInheritable=true
-DLog4jContextSelector=org.apache.logging.log4j.core.async.AsyncLoggerCon
textSelector -Dlog4j2.asyncLoggerRingBufferSize=1000000 -Dorg.wildfly.op
enssl.path=/usr/lib64
```

```
-Dzookeeper.sasl.client=false -Djdk.tls.disabledAlgorithms=SSLv3,GCM -Xms2
048M
-Xmx2867M -server -Djava.net.preferIPv4Stack=true -XX:NewRatio=8 -XX:+UseN
UMA
-XX:+ResizeTLAB -Xloggc:/var/log/hive/query-coordinator-gc-%t.log -XX:+U
seG1GC
-XX:+PrintGCDetails -XX:+PrintGCTimeStamps -verbose:gc -XX:+PrintGCCause
-XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=10 -XX:GCLogFileSize=10M
-Dlog4j2.formatMsgNoLookups=true
-javaagent:/jmx-exporter/jmx_prometheus_javaagent-0.11.0.jar=35001:/jmx-ex
porter/config.yml org.apache.tez.dag.app.DAGAppMaster --session
```

Impala Virtual Warehouse

Note: If you have an Impala Virtual Warehouse running in Unified Analytics mode (currently in Tech Preview), then in addition to the steps listed below, please also follow the steps listed above in the “Hive Virtual Warehouse: Hive Server2” section. In Unified Analytics mode there is a Hive Server2 running within the Impala Virtual Warehouse, and it needs to be updated.

- Coordinator
  - Update coordinator statefulset and set -Dlog4j2.formatMsgNoLookups=true in JAVA_TOOL_OPTIONS environment variable

```
$kubectl --kubeconfig ~/kubeconfig edit sts coordinator -n impala-163953
0043-8fnb
...
        - /opt/impala/bin/impalad
        - --flagfile=/opt/impala/conf/flagfile
        command:
        - /opt/impala/bin/daemon_entrypoint.sh
        env:
        - name: JAVA_TOOL_OPTIONS
          value: -Xms2G -Xmx4G -XX:+HeapDumpOnOutOfMemoryError
-XX:HeapDumpPath=/opt/impala/dump.hprof
-XX:+UseConcMarkSweepGC -Dlog4j2.formatMsgNoLookups=true
        - name: POD_NAME
          valueFrom:
            fieldRef:
              apiVersion: v1
              fieldPath: metadata.name
        image:
container-dev.repo.cloudera.com/cloudera/impalad_coordinator:2021.0.5-b1
7
        imagePullPolicy: IfNotPresent
        name: impalad-coordinator
...
```

- Executor
  - Update impala-executor-<xxx> statefulset and set -Dlog4j2.formatMsgNoLookups=true in JAVA_TOOL_OPTIONS environment variable

```
$kubectl --kubeconfig ~/kubeconfig edit sts impala-executor-000 -n
impala-1639530043-8fnb
...
  env:
  - name: JAVA_TOOL_OPTIONS
    value: -Xms2G -Xmx4G -XX:+HeapDumpOnOutOfMemoryError
           -XX:HeapDumpPath=/opt/impala/dump.hprof
           -XX:+UseConcMarkSweepGC -Dhttp.maxConnections=16
           -Dlog4j2.formatMsgNoLookups=true
        image:
```

```
container-dev.repo.cloudera.com/cloudera/impalad_executor:2021.0.5-b17
        imagePullPolicy: IfNotPresent
        lifecycle:
          preStop:
            exec:
              command:
              - /bin/bash
              - /opt/impala/bin/graceful_shutdown_backends.sh
              - "3600"
        name: impalad-executor
 ...
```

- If there are multiple impala-executor-<xxx> statefulsets then the above steps should be repeated for all of them.

- Catalogd
  - Update catalogd deployment and set -Dlog4j2.formatMsgNoLookups=true in JAVA_TOOL_OPTIONS environment variable

```
$kubectl --kubeconfig ~/kubeconfig edit deployment catalogd -n impala-16
39530043-8fnb
...
        env:
        - name: JAVA_TOOL_OPTIONS
          value: -Xms2G -Xmx6G -XX:+HeapDumpOnOutOfMemoryError
-XX:HeapDumpPath=/opt/impala/dump.hprof
            -XX:+UseConcMarkSweepGC
             -Dlog4j2.formatMsgNoLookups=true
        image:
container-dev.repo.cloudera.com/cloudera/catalogd:2021.0.5-b17
        imagePullPolicy: IfNotPresent
        livenessProbe:
          exec:
            command:
            - sh
            - -c
            - curl -I --silent http://localhost:25021/metrics | grep -q
 "HTTP/1.1
              200 OK"
          failureThreshold: 20
          initialDelaySeconds: 1
          periodSeconds: 20
          successThreshold: 1
          timeoutSeconds: 2
        name: catalogd
. . .
```