

Cloudera Data Warehouse Private Cloud 1.5.3

Data Warehouse Release Notes

Date published: 2020-08-17

Date modified: 2024-03-27

CLOUDERA

<https://docs.cloudera.com/>

Legal Notice

© Cloudera Inc. 2024. 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

What's new.....	4
Known issues.....	4
General known issues.....	5
Upgrade-related known issues.....	6
OpenShift cluster environments.....	7
ECS cluster known issues.....	7
Database Catalog known issues.....	7
Hive Virtual Warehouses known issues.....	8
Impala Virtual Warehouses known issues.....	10
Hue known issues.....	10
Known issues in Unified Analytics.....	12
Iceberg-related known issues.....	12
Fixed issues in 1.5.3.....	16
Fixed issues in 1.5.3-CHF1.....	16
CDW component versions.....	17

What's new in Cloudera Data Warehouse Private Cloud

Learn about the new features in Cloudera Data Warehouse (CDW) service on CDP Private Cloud Data Services 1.5.3.

Impala supports updating Iceberg V2 tables

Using the UPDATE statement, you can now update the Iceberg V2 tables from Impala. See, [Update data feature](#).

Added support for downloading root certificates for Impala-shell and Beeline

If your organization is using a custom Certificate Authority (CA) to sign security certificates for Embedded Container Service (ECS) or OpenShift Container Platform (OCP) environments, then as a non-DWAdmin user, you can download the root certificates (TLS) for connecting to Impala or Hive Virtual Warehouses from BI clients such as Impala-shell, Impyla, and Beeline from the CDW user interface. You no longer need to depend on your Administrators to provide you with the certificates. See [Downloading root certificates for Kubernetes environments](#).

CDW validates whether Impala Virtual Warehouses has the complete CA chain

Impala Virtual Warehouses must have the LDAP certificate that contains the complete chain including the root CA certificate, without which it cannot function as expected. For example, an incomplete CA chain may cause an unauthorized error when connecting to Impala with an unclear message in the logs about not being able to connect to the LDAP server. To prevent such hard-to-debug issues later, CDW now validates whether Impala Virtual Warehouse has the complete CA chain during environment activation.



Note: Root CA certificate validation is one of the many validations performed before activating an environment in CDW. These include the port validation, delegation user validation, and the Kerberos keytab configuration validation. You can turn off cluster validation by selecting the Skip cluster validation during environment activation option from Advanced Configurations. However, this is not recommended.

CDW performs port validation during environment activation

The CDW Runtime components such as Hue, Hive, and Impala require access to other services on the CDP Base cluster through specific ports as listed in [Ports Used by Cloudera Runtime Components](#). Now when you activate an environment in CDW, CDW validates whether the following addresses for these Runtime components are present in the configuration, are reachable, and the corresponding ports are open:

- HttpFS WebHDFS address found in the Hue configuration (hadoop.hdfs_clusters.default.webhdfs_url config), if present
- WebHDFS address of the configured HDFS NameNodes (dfs.namenode.https-address config)
- The address of the Ranger service (ranger.plugin.hive.policy.rest.url config)

If the required ports are not open, then the CDW environment activation fails.



Note: Validating ports is one of the many validations performed before activating an environment in CDW. These include the root CA certificate validation, delegation user validation, and the Kerberos keytab configuration validation. You can turn off cluster validation by selecting the Skip cluster validation during environment activation option from Advanced Configurations. However, this is not recommended.

Known issues and limitations in Cloudera Data Warehouse Private Cloud

This section lists known issues and limitations that you might run into while using the Cloudera Data Warehouse (CDW) service in CDP Private Cloud Data Services.

General known issues

This topic describes the general service-wide known issues for Cloudera Data Warehouse (CDW) Private Cloud.

Known issues identified in 1.5.3

DWX-17880: Hive Virtual Warehouse does not start if the bind user contains special characters

The Hive virtual warehouse may fail to start up if you have specified the following special characters in the LDAP bind credential password: < > & ' ". This happens because the HiveServer2 (HS2) pod gets stuck in the CrashLoopBackOff state with the following error in its logs: error parsing conf file:/etc/hive/conf/hive-site.xml com.ctc.wstx.exc.WstxUnexpectedCharException: Unexpected character '&' (code 38) in content after '<' (malformed start element?). at [row,col,system-id]: [388,13,"file:/etc/hive/conf/hive-site.xml"].

1. Change the LDAP bind credentials in the Management Console. ensure that they do not contain the following unsupported special characters: < > & ' " .
2. Reactivate the environment in CDW.

Known issues identified in 1.5.2

VIZ-2269: Issue with impersonation in CDW Impala connections

You may see the following error while creating or editing the data connection to a CDW Impala or Hive Virtual Warehouse and using the connection details auto-populated from the CDW Warehouse drop-down: User <username> is not authorized to delegate to <username>. This happens because the Impersonation and Trusted Impersonation options are both enabled. This affects CDV 7.1.6.

If the message appears when creating a new data connection, try refreshing the page. This will reset the Impersonation and Trusted Impersonation options. If the message appears on the **Edit Data Connection** modal, then copy the hostname, username, HTTP path, and other values present on the **Advanced** tab, and manually edit the existing CDW connections. A manual edit will typically not trigger this bug.

Known issues identified in 1.5.1

DWX-15142 Character restriction on environment name when using FreeIPA server version 4.9.8 and higher

FreeIPA is supported as an authentication mechanism starting with the 1.5.1 release. If you are using FreeIPA version 4.9.8 and higher, then note that the host names are limited to 64 characters. Because the environment name is part of the host name, the environment name must not exceed 17 characters.

None.

Known issues identified in 1.5.0

DWX-18903: Service "postgres-service-default-warehouse" is invalid: spec.externalName error

You see the following error during the Database Catalog creation stage after activating the environment in CDW:

```
Service "postgres-service-default-warehouse" is invalid:
  spec.externalName
  a lowercase RFC 1123 subdomain must consist of lower case
  alphanumeric characters, '-' or '.', and must start and end with
  an alphanumeric character (e.g. 'example.com', regex used for
  validation is '[a-z0-9]([-a-z0-9]*[a-z0-9])?(\.[a-z0-9]([-a-
  z0-9]*[a-z0-9])?)*')
```

. This could happen because if the value of the Hive Metastore Database Host (hive_metastore_database_host) property on the base cluster is not specified in lowercase.

Go to Cloudera Manager Clusters Hive service Configuration and change the value specified in the Hive Metastore Database Host field to be in lowercase.

Known issues identified before 1.4.1

DWX-10403: Executor pods get stuck in pending state with a warning

In rare circumstances, when Impala or Hive executors start up either due to autoscaling or by manually restarting the executors, the pods may get stuck in a pending state with a warning such as "volume node affinity conflict". This happens due to a race condition in the storage class that provides local volumes.

Restart the pods so that they can be rescheduled on new nodes with enough resources.

DWX-8502: HMS health check does not check port 9083

The HMS health check script does not check the health of its service port 9083 and may provide incorrect health status.

None.

Upgrade-related known issues

This topic describes the upgrade-related known issues for Cloudera Data Warehouse (CDW) Private Cloud.

Known issues identified in 1.5.3

No new known issues identified in 1.5.3.

Known issues identified before 1.5.2

DWX-15571: tez.history.logging.proto-base-dir and hive.hook.proto.base-directory locations change after the runtime upgrade

The values of following Hue Query Processor properties change after you upgrade the Database Catalog and Virtual Warehouses to the latest version:



- tez.history.logging.proto-base-dir
- hive.hook.proto.base-directory

CDW changes these locations to match the base cluster configuration for hive.metastore.warehouse.external.dir. You may also see the FileNotFoundException: File does not exist error in the Hue Query Processor logs.



Note: This known issue does not cause any change in the functionality.

To use the old location and to avoid seeing the "File does not exist" error in the logs, you must update the Hue Query Processor configurations as follows:

1. Log in to the Data Warehouse service as an DWAdmin.
2. Go to Database Catalog  Edit CONFIGURATIONS Hue query processor and select hue-query-processor.json from the Configuration files drop-down menu.
3. Note the paths of the tez.history.logging.proto-base-dir and hive.hook.proto.base-directory properties.
4. After you upgrade the Database Catalog, go to Database Catalog  Edit CONFIGURATIONS Hue query processor and select hue-query-processor.json from the Configuration files drop-down menu.
5. Specify the desired path under the dasConf section as shown in the following example:

```
"dasConf" : {
```

```
"hive.hook.proto.base-directory": "/warehouse/tablespace/external/hive/sys.db/query_data",
"tez.history.logging.proto-base-dir": "/warehouse/tablespace/external/hive/sys.db"
},
```

6. Click Apply Changes.

Known issues on OpenShift cluster environments

This topic describes the OpenShift cluster environment known issues for Cloudera Data Warehouse Private Cloud.

Known issues identified in 1.5.3

No new known issues identified in 1.5.3.

Known issues identified before 1.4.1

DWX-4723: HTTP 504 Gateway timeout error returned when you delete an environment

When a CDW Private Cloud environment is deleted in the CDW Private Cloud UI, an HTTP 504 Gateway timeout error can be returned.

Using the OpenShift CLI, increase the route timeout to 60s on the OpenShift cluster:

```
oc annotate route [***ROUTE-NAME***] --overwrite haproxy.router.openshift.io/timeout=60s
```

For more information about setting this timeout, see the [OpenShift documentation](#).

ECS cluster environments

This topic describes the Embedded Container Service (ECS) cluster environment known issues for Cloudera Data Warehouse Private Cloud.

Known issues identified in 1.5.3

No new known issues identified in 1.5.3.

Known issues identified before 1.4.1

BLESC-6074: Impala executor pods not running and queries from Hue fail

You may see the following error after submitting a query from Hue: Latest admission queue reason: Waiting for executors to start. Only DDL queries and queries scheduled only on the coordinator (either NUM_NODES set to 1 or when small query optimization is triggered) can currently run. You may also notice that the Impala executors are not running.

Make sure that the `/etc/resolv.conf` file on the ECS hosts contain a maximum of 2 search domains.

Known issues in Database Catalogs

This topic describes the Database Catalog known issues for Cloudera Data Warehouse Private Cloud.

Known issues identified in 1.5.3

No new known issues identified in 1.5.3.

Known issues identified in 1.5.1

DWX-15302: Upgrade button stays visible even after the upgrade completes

After you upgrade the Database Catalog, the Upgrade button remains visible on the CDW web interface instead of disappearing or getting disabled.

Refresh the page on your browser.

Known issues in Hive Virtual Warehouses

This topic describes the Hive Virtual Warehouse known issues for Cloudera Data Warehouse (CDW) Private Cloud.

Known issues identified in 1.5.3

No new known issues identified in 1.5.3.

Known issues identified in 1.5.2

DWX-16989: Hive query running on Iceberg table fails randomly

Suppose you have disabled the auto-suspend option for a Hive Virtual Warehouse or if the Virtual Warehouse is under continuous load and hence it cannot be stopped by the auto-suspend option. In this situation, using the Iceberg table format may cause the following exceptions to appear in the query coordinator log along with the submitted queries that have failed:

```
org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.token.SecretManager$InvalidToken): token (token for hive: HDFS_DELEGATION_TOKEN owner=hive/dwx-env-host-1.cdp.local@EXAMPLE.CLOUDERA.COM, renewer=hive, realUser=, issueDate=1709813340891, maxDate=1710418140891, sequenceNumber=19784486, masterKeyId=52) is expired, current time: 2024-03-08 04:09:32,835-0800 expected renewal time: 2024-03-08 04:09:00,891-0800
```

```
org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.token.SecretManager$InvalidToken): token (token for hive: HDFS_DELEGATION_TOKEN owner=hive/dwx-env-host-1.cdp.local@EXAMPLE.CLOUDERA.COM, renewer=hive, realUser=, issueDate=1699855596578, maxDate=1700460396578, sequenceNumber=16863242, masterKeyId=39) can't be found in cache
```


```
org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.token.SecretManager$InvalidToken): token (OzoneToken owner=hive/dwx-env-ewxf6g-env.cdp.local@ROOT.EXAMPLE.SITE, renewer=hive, realUser=, issueDate=2024-03-19T21:49:31.033Z, maxDate=2024-03-19T21:50:31.033Z, sequenceNumber=72, masterKeyId=1, strToSign=null, signature=null, awsAccessKeyId=null, omServiceId=ozone1710521984, omCertSerialId=11) is expired, current time: 2024-03-19 21:51:34,293+0000 expected renewal time: 2024-03-19 21:51:31,033+0000
```

```
org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.token.SecretManager$InvalidToken): token (OzoneToken owner=hive/dwx-env-aztlgg-env.cdp.local@ROOT.EXAMPLE.SITE, renewer=hive, realUser=, issueDate=2024-04-09T16:04:12.889Z, maxDate=2024-04-09T17:04:12.889Z, sequenceNumber=29, masterKeyId=1, strToSign=null, signature=null, awsAccessKeyId=null, omServiceId=ozone1711550158, omCertSerialId=2597525731772327) can't be found in cache
```


This happens because the HDFS delegation tokens are not renewed when using the Iceberg table format. After the existing HDFS delegation tokens expire, Hive query coordinator (TEZ App Master) cannot access the tables on the file system during the query planning phase. The problem is independent of the file system--Ozone FS or Hadoop FS. The error only occurs after the HDFS delegation tokens have expired. By default, the delegation tokens expire in one day. However, you can modify the expiration time on the CDP Base cluster.

The problem does not occur if the query coordinator pods in the Hive Virtual Warehouse are stopped manually or by using the auto-suspend functionality within the token expiration period.

Apply this workaround only if you cannot suspend the Hive Virtual Warehouse.

1. Log in to the Data Warehouse service as DWAdmin.
2. Go to the Virtual Warehouses tab and click  Edit Configurations Query Coordinator .
3. Select env from the Configuration files drop-down menu.
4. Add the following value against the JVM_OPTS property:


```
-Diceberg.scan.plan-in-worker-pool=false
```

5. Click Apply Changes.

Known issues identified in 1.5.1

DWX-15480: Hive queries fail with FILE_NOT_FOUND error

ACID directory cache may become outdated in Tez AMs in case of ACID tables that change often, possibly leading to different errors with the same root cause: "split generation works from cache pointing to non-existing files". And you may see the following error in the diagnostic bundles and query logs: FILE_NOT_FOUND: Unable to get file status.

Disable the cache by setting the value of the hive.txn.acid.dir.cache.duration property to -1 by going to Virtual Warehouse  Edit CONFIGURATIONS Hue Configuration files hive-site from the CDW web interface.

DWX-15287: Drop database query for Hive fails with Invalid ACL Exception

You may see the following error in a Hue or beeline session when running DROP DATABASE, DROP TABLE, or ALTER TABLE DROP PARTITION operations on a Hive Virtual Warehouse that is in Stopped state: "org.apache.zookeeper KeeperException\$InvalidACLException: KeeperErrorCode = InvalidACL for /llap-sasl/user-hive".


The exception is caused because the Hive Virtual Warehouse tries to evict the cache in the LLAP executors, but the compute pods of the stopped warehouse are no longer running.



Note: The database or table is deleted despite the exception, only the LLAP executors do not flush their database or table related buffers, because these executors are not running.

Start the Virtual Warehouse before you run the DROP DATABASE, DROP TABLE, or ALTER TABLE DROP PARTITION operations.

Alternatively, you can add the hive.llap.io.proactive.eviction.enabled=false setting in the hive-site.xml file. This method may result in some performance degradation, because LLAP no longer discards the dropped database/table or temp table related buffers.

1. Log in to CDW as DWAdmin.
2. Click  Edit CONFIGURATIONS Hiveserver2 on the Virtual Warehouse tile and select hive-site from the Configuration files drop-down menu.

3. Click **+** and add the following line:

```
hive.llap.io.proactive.eviction.enabled=false
```

4. Click Apply Changes.

Wait for the Virtual Warehouse to refresh and return to Running or Stopped state.

Known issues in Impala Virtual Warehouses

This topic describes the Impala Virtual Warehouse known issues for Cloudera Data Warehouse Private Cloud.

Known issues identified in 1.5.3

No new known issues identified in 1.5.3.

Known issues identified in 1.5.1

DWX-14292: Impala executors and coordinator pods get stuck in pending state

If you have enabled low resource mode while activating an environment in CDW and also enabled the option to customize pod sizes for Impala from the **Advanced Settings** page, then you may notice that the executors and coordinator pods do not request the set amount of memory and CPU. This happens because the low resource mode overrides the default pod configuration.

Low resource mode and custom pod configurations for Impala cannot be used at the same time. To resolve this issue, do not select the Low resource mode option while activating an environment. Instead, create custom pod configurations with lower resources for Impala. You can continue to use lesser resources as defined in the custom pod configuration for Hive.

VIZ-2046: Trusted impersonation does not work in CDW

When you try to create a connection to the Impala Virtual Warehouse from Cloudera Data Visualization (CDV), CDV is unable to use Trusted Impersonation as a mode of authentication. Therefore, you need to enter both the username and password for authentication.

None.

Known issues in Hue

This topic describes the Hue known issues for Cloudera Data Warehouse (CDW) Private Cloud.

Known issues identified in 1.5.3


No new known issues identified in 1.5.3.

Known issues identified in 1.5.2

DWX-15100: Impala doesn't release the query profiles until the session is closed by the client

You may notice that the queries are stuck in the Waiting to be closed state and that the last query submitted from Hue is not displayed on the **Job Browse** page in Hue. Ideally, Impala must finish running the and the client should be able to fetch the results, and Hue should close the query, so that Impala can write the query profile to disk in thrift. However, this does not happen.

You can view the queries in one of the following ways:

- You can click on the query ID that is displayed under the query editor when a query starts to running and view the query profile.
- You can go to the **Job Browser** page's **Impala** tab to view the Impala query profile.
- You can close the Hue session to trigger query closure by clicking  Session Close .

Known issues identified in 1.5.1**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, ABFS, or Ozone 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.

DWX-15090: Intermittently see the CSRF error 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 page or start a new Hue session.

DWX-14911: Export operation failing in a partitioned Hive table

You may see the following error in the Hive server logs when running the EXPORT TABLE query in Hive: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.ExportTask. Distcp operation failed. This issue occurs when the number of files and the size of files exceeds the limits specified in the hive.exec.copyfile.maxsize and hive.exec.copyfile.maxnumfiles properties respectively, and the EXPORT TABLE query launches the distributed copy (distcp) job to copy files of HDFS.

Increase the limit of following properties in HiveServer2:

- hive.exec.copyfile.maxsize to maximum size of files in a directory in bytes
- hive.exec.copyfile.maxnumfiles to maximum number of files in a directory

OPSAPS-66903: Ozone HTTPFS address configuration contains placeholders instead of the real values

The Hue File Browser does not automatically support browsing the Ozone filesystem because the HTTPFS role configuration for the Ozone service contains placeholder text instead of real values. You can view this configuration by going to Cloudera Manager Ozone service Instances HTTPFS Gateway Processes and clicking on the ozone-conf/httpfs-site.xml configuration file.

You must manually configure Hue to access Ozone using the File Browser by following the instructions listed in [Enabling browsing Ozone from Hue on CDW Private Cloud](#).

Known issues identified in 1.5.0**DWX-12616: Hue limitation in CDW Private Cloud**

Following are the known limitations in Hue in CDW Private Cloud 1.5.0:

- Hue Importer does not support importing files more than 200 KB in size
- Hue File Browser does not work if the HDFS service on the base cluster is configured for high availability

None.


DWX-13865: Hue File Browser does not work with HDFS HA

Hue File Browser is not accessible or displays a 403 error when you click on File Browser from the left assist panel in CDW Private Cloud if HDFS is configured for High Availability on the base cluster. Currently, Hue in CDW cannot obtain the hostname and the port from the HttpFS service on the base cluster. This is a known limitation.

You must manually build and specify the WebHDFS URL for Hue in CDW to connect to the HttpFS service on the base cluster.

1. Log in to Cloudera Manager as an Administrator.
2. Go to Clusters Hive Configuration and note the value present in the Kerberos Principal field.

This is the Hive service account name.


3. Go to Clusters HDFS HttpFS Advanced Configuration Snippet (Safety Valve) for httpfs-site.xml and click  to add the following lines:
Name: httpfs.proxyuser.hive.hosts, Value: *
Name: httpfs.proxyuser.hive.groups, Value: *

Replace hive with the actual Hive service account name.
4. Click Save Changes and restart the HDFS service or the HttpFS role.
5. Go to the Instances tab and note the hostname of the HttpFS role.
6. Go to the Configuration tab and note the port for the hdfs.httpfs.http.port configuration from the RESTPort field.

The default value of the hdfs.httpfs.http.port configuration is 14000.

7. Use the hostname and the port to construct the WebHDFS URL as follows:

```
https://[***HOSTNAME***]:[***PORT***]/webhdfs/v1
```

8. Log in to the Data Warehouse service as a DWAdmin.
9. Go to the Virtual Warehouse from which you want to connect Hue to the base cluster's HDFS service and click  Edit .
10. Go to CONFIGURATIONS Hue , select hue-safety-valve from the Configuration files drop-down list and add the following lines:

```
[hadoop]
[[hdfs_clusters]]
[[[default]]]
webhdfs_url=https://[***HOSTNAME***]:[***PORT***]/webhdfs/v1
```

Specify the WebHDFS URL that you constructed earlier.

11. Click Apply changes.

Update the hue-safety-valve configuration for any Hive or Impala Virtual Warehouses from which you want to connect to the base cluster HDFS.

Known issues identified before 1.4.1

DWX-9373: Unable to log into Hue as a local administrator

If you have logged into the CDP Management Console as a local administrator, then you may not be able to log into Hue and you may see the following error: “User is not authorized”.

To access Hue, you must add your username to your organization's LDAP tree or log into the CDP Management Console using your LDAP credentials and then access Hue.

Known issues in Unified Analytics

This topic describes the Unified Analytics known issues for Cloudera Data Warehouse Private Cloud.

Known issues identified in 1.5.3

No new known issues identified in 1.5.3.

Iceberg-related known issues in CDW Private Cloud

This topic describes the Iceberg-related known issues in Cloudera Data Warehouse (CDW) Private Cloud.

Known issues identified in 1.5.3

No new known issues identified in 1.5.3.

Known issues identified in 1.5.2

DWX-16591: Concurrent merge and update Iceberg queries are failing

You may see that the concurrent merge and update Iceberg queries fail with the following error in the Hive application logs: "Base metadata location hdfs://<Location-A> is not same as the current table metadata location '<Location-B>' for default.merge_insert_target_iceberg \rorg.apache.iceberg.exceptions.CommitFailedException". This happens because the corresponding Query A and Query B have overlapping updates. For example, if Query A commits the data and deletes files first, then Query B will fail with validation failure due to conflicting writes. In this case, Query B should invalidate the commit files that are already generated and re-execute the full query on the latest snapshot.

None.

CDPD-59413: Unable to view Iceberg table metadata in Atlas

You may see the following exception in the Atlas application logs when you create an Iceberg table from the CDW data service associated with a CDP Private Cloud Base 7.1.8 or 7.1.7 SP2 cluster: Type ENTITY with name iceberg_table does not exist. This happens because the Atlas server on CDP Private Cloud Base 7.1.8 and 7.1.7 SP2 does not contain the necessary, compatible functionality to support Iceberg tables. This neither affects creating, querying, or modifying of Iceberg tables using CDW nor does it affect creating of policies in Ranger.

On CDP Private Cloud Base 7.1.9, Iceberg table entities are not created in Atlas. You can ignore the following error appearing in the Atlas application logs: ERROR - [NotificationHookConsumer thread-1:] ~ graph rollback due to exception (GraphTransactionInterceptor:200) org.apache.atlas.exception.AtlasBaseException: invalid relationshipDef: hive_table_storagedesc: end type 1: hive_storagedesc, end type 2: iceberg_table

If you are on CDP Private Cloud Base 7.1.7 SP2 or 7.1.8, then you can manually upload the Iceberg model file z1130-iceberg_table_model.json in to the /opt/cloudera/parcels/CDH/lib/atlas/models/1000-Hadoop directory as follows:

1. SSH into the Atlas server host as an Administrator.
2. Change directory to the following:

```
cd /opt/cloudera/parcels/CDH/lib/atlas/models/1000-Hadoop
```

3. Create a file called 1130-iceberg_table_model.json with the following content:

```
{
  "enumDefs": [],
  "structDefs": [],
  "classificationDefs": [],
  "entityDefs": [
    {
      "name": "iceberg_table",
      "superTypes": [
        "hive_table"
      ],
      "serviceType": "hive",
      "typeVersion": "1.0",
      "attributeDefs": [
        {
          "name": "partitionSpec",
          "typeName": "array<string>",
          "cardinality": "SET",
          "isIndexable": false,
          "isOptional": true,

```

```

        "isUnique": false
      }
    ],
  },
  {
    "name": "iceberg_column",
    "superTypes": [
      "hive_column"
    ],
    "serviceType": "hive",
    "typeVersion": "1.0"
  }
],
"relationshipDefs": [
  {
    "name": "iceberg_table_columns",
    "serviceType": "hive",
    "typeVersion": "1.0",
    "relationshipCategory": "COMPOSITION",
    "relationshipLabel": "__iceberg_table.columns",
    "endDef1": {
      "type": "iceberg_table",
      "name": "columns",
      "isContainer": true,
      "cardinality": "SET",
      "isLegacyAttribute": true
    },
    "endDef2": {
      "type": "iceberg_column",
      "name": "table",
      "isContainer": false,
      "cardinality": "SINGLE",
      "isLegacyAttribute": true
    },
    "propagateTags": "NONE"
  }
]
}

```

4. Save the file and exit.
5. Restart the Atlas service using Cloudera Manager.

Technical Service Bulletins

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

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 Private Cloud

This section lists issues fixed in this release of the Cloudera Data Warehouse (CDW) service.

DWX-16826: Executor pods do not get scheduled on the same node after the scale down

Earlier, due to the PersistentVolumeClaim (PVC) behavior, when the executor pods attempted to get rescheduled back on the node on which they were running before scaling down, but could not because other pods got scheduled on that node and consumed the resources. This issue has been fixed by removing the PVCs whenever the Virtual Warehouse scales down. The executor pods can still be scheduled on the same node if the node is healthy and suitable. But the executors no longer wait to get rescheduled on that particular node. They are automatically scheduled on any suitable node.

DWX-17179: Hue backend and Impala catalog pods are scheduled on the same node in an HA deployment

Earlier, multiple replicas of Hue backend, frontend, and Impala catalog pods got scheduled on the same node in HA mode and you had to manually schedule them on other nodes. This issue has been fixed.

DWX-16433: Logging Improvements

Writing production logs and workload profiles using Fluentd have been improved to write logs more reliably to HDFS under heavy load.

DWX-16891: Hive-Ranger integration issue after a refresh

Earlier, CDW displayed an error while compiling a statement: FAILED: HiveAccessControlException Permission denied. This happened because Hive failed to evaluate Ranger group policies when the Virtual Warehouse was either by upgrading or refreshing it, and you had to rebuild the Hive Virtual Warehouse to fix the Ranger integration issues. This issue has been resolved.

Fixed issues in Cloudera Data Warehouse Private Cloud (1.5.3-CHF1)

This section lists issues fixed in this release of the Cloudera Data Warehouse (CDW) service.

DWX-17404: Hive Virtual Warehouse fails to start up when CDW is unable to parse the LDAP bind password

Earlier, the Hive virtual warehouse failed to start up if you specified the following special characters in the LDAP bind credential password: < > & ' " This happened because the HiveServer2 (HS2) pod remained in the CrashLoopBackOff state with the following error in its logs: error parsing conf file:/etc/hive/conf/hive-site.xml com.ctc.wstx.exc.WstxUnexpectedCharException: Unexpected character '&' (code 38) in content after '<' (malformed start element?). at [row,col,system-id]: [388,13,"file:/etc/hive/conf/hive-site.xml"]. This issue has been fixed.



Note: If you are on an older release, then the fix becomes available only when you activate the environment in CDW after upgrading to the Private Cloud Data Services 1.5.3-CHF1 release.

CDPD-67593: Count function in Impala gives incorrect results for Iceberg V2 tables

According to [TSB 2024-752](#), Spark compaction may leave dangling delete files, causing incorrect count results on Iceberg V2 tables. This issue has been temporarily fixed by turning off the count(*) optimizations for Iceberg V2 tables. A permanent fix will be available in a future release.

CDPD-66786: Impala returns incorrect results when the optimized Iceberg V2 operator is used

If you are using Impala to read Iceberg V2 tables, then you may have faced an issue of Impala returning incorrect results when the optimized V2 operator is used. This issue has been resolved by disabling the operator by setting `DISABLE_OPTIMIZED_ICEBERG_V2_READ = 1` as the default value.

HIVE-28102, IMPALA-12860: Data corruption issues in Iceberg tables due to concurrent compaction and modify statements

Earlier, Hive or Impala DELETE, UPDATE, or MERGE operations on Iceberg V2 tables corrupted the tables if there was a concurrent table compaction from Spark. This happened because when the compaction and modify statements ran in parallel, and the compaction job was committed before the modify statement. In that case, the modify statement's position to delete files still points to the old files. This issue has been 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 2024-752: 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](#).

Version information for Cloudera Data Warehouse Private Cloud components

Cloudera Data Warehouse (CDW) uses Hive, Impala, and Hue as its Runtime components and also provides integration with Cloudera Data Visualization. Review the version information of CDW Private Cloud 1.5.3 components.

Table 1: CDW Private Cloud version information

CDW component	Version
Hive	2023.0.15.4
Impala	2023.0.15.4
Hue	2023.0.15.4
Cloudera Data Visualization	7.1.6
CDW server	1.8.0-h2
CDP CLI	0.9.101 to 0.9.107