

Data Catalog Release Notes

Date published: 2023-12-16

Date modified: 2024-07-31

CLOUDERA

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
Fixed issues.....	4
Cumulative hotfixes: Cloudera Data Catalog Private Cloud.....	5
Cloudera Data Catalog Private Cloud 1.5.4-CHF1.....	5
Fixed issues in Cloudera Data Catalog (1.5.4-CHF1).....	5

What's new

This section lists major features and updates for the Data Catalog service.

This release of Cloudera Data Catalog on CDP Private Cloud 1.5.4 includes the following features and enhancements:

- Bug fixes and other performance improvements

Known issues

Learn about the known issues in Data Catalog, the impact or changes to the functionality, and the workaround.

Known Issues

CDPDSS-2706: Data Steward user role is not reflecting

Data Steward role is not reflecting in Control Plane.

None

CDPDSS-3051: Profiler Jobs are not in the sorted order

Profiler jobs are incorrectly ordered when the list cannot fit into a single page under Profiler Jobs . The latest job is displayed in the last page.

CDPDSS-3047: The profiled data is not shown for Ranger Audit Profiler

The profiled data is not shown for Ranger Audit Profiler if the active NameNode is switched over to standby.

Switch the NameNode in standby to active state. For more information on restarting nodes, see [Starting, Stopping, and Restarting Role Instances](#) or [Administrative commands](#).

CDPDSS-3057: Failed profiler job because of missing columns in log entries.

When the logs to be profiled by the Ranger Audit Profiler have missing columns, the profiling job fails. The ranger audit profiler pod log contains the following message: Ranger Audit Profiler failed: Failed to run Ranger Audit Profiler. Error: "[*column1*, '*column2*', ...] not in index".

column1, *column2* show the missing column entries in the logs.

Each Ranger audit profiler job has 3 sub-jobs (identified by the same job ID prefix), any of which could fail because of the above reason. It is possible that one or even two sub-jobs pass till the above scenario is encountered.

None

Fixed issues

This section lists the issues that have been fixed since the previous version.

CDPDSS-2707: Custom Sensitivity Profiler (CSP) regex logic fails intermittently for complex regex

The issue was fixed by changing the regex validation flow. The previously used complex regex had a different type of syntax, which was directly passed on to the code. The profilers failed when these complex regular expressions were passed directly. Therefore, the complex regex strings are encrypted now and they are now stored in the database. Once the profiler execution starts and fetches this complex regex, it first decodes the string and then uses the regex mechanism.

CDPDSS-2830: Data Catalog running on JDK-17 in Private Cloud causes ClassNotFoundException and further failures

Data Catalog running on JDK-17 in Private Cloud causes a `ClassNotFoundException` and further failures which leads to the service being unusable. The service is restored by upgrading the Google Guice dependency manager to 6.0.0.

CDPDSS-2839: Hadoop version upgraded from 7.2.17.200-29 to 7.2.17.301-3

The Hadoop version is upgraded from 7.2.17.200-29 to 7.2.17.301-3 to remediate CVE-2023-44487 and other CVE fixes are implemented as well.

CDPDSS-2843: While upgrading to Data Services 1.5.3, an unnamed module @0x4e38d975 exception error is thrown

The issue causing exceptions from the unnamed module, which previously necessitated a restart of the Data Catalog pod, has been addressed.

Cumulative hotfixes: Cloudera Data Catalog Private Cloud

Review the cumulative hotfixes that have been shipped for Cloudera Data Catalog Private Cloud 1.5.4.

Cloudera Data Catalog Private Cloud 1.5.4-CHF1

Review the features, fixes, and known issues in the Cloudera Data Catalog 1.5.4 Cumulative hotfix 1 release.

Fixed issues in Cloudera Data Catalog (1.5.4-CHF1)

This section lists issues fixed in this release of the Cloudera Data Catalog service.

CDPDSS-3037: Asset Details page does not highlight link for Datasets

The **Asset Details** now correctly highlights links to Datasets.

Datasets / Details / Asset Details

The screenshot shows the 'item' details page for a Hive Table. The 'Properties' section includes:

- Type: HIVE TABLE
- # of Columns: ??
- Data Lake: [blurred]
- Datasets: 8 (highlighted with a red box)
- Owner: hive
- Created On: Mon Jun 17 2024 10:01:51 GMT+0200 (Cent...)
- Last Access Time: Mon Jun 17 2024 10:01:51 GMT+0200...
- Table Type: EXTERNAL_TABLE
- Database: test_dss_db
- DB Catalog: cm
- Parent: test_dss_db

The right-hand side of the page shows the 'Qualified Name' (test_dss_db.item@cm), 'Comment' (+ Add Comment), and 'Description' (+ Add Description).

CDPDSS-2958: Schema rows are not ordered as in their database

Rows under the **Schema** tab in **Asset Details** are ordered based on their position in their parent database.

Datasets / Details / Asset Details

Overview Schema Metadata Audits Policy Access Audits

Search Column Edit

Chart Type	Name	Type	Unique Values *	Null Values	Max	Min	Mean	Comment	Classifications	Terms
▼	l_item_sk	int	17852	0	18,000	1	9,000.5			
▼	l_item_id	char(16)	8713	0						
▼	l_rec_start_date	date	4	18000						
▼	l_rec_end_date	date	3	18000						
▼	l_item_desc	varchar(200)	12819	0						
▼	l_current_price	decimal(7,2)	2917	45	99.99	0.09	9.52			
▼	l_wholesale_cost	decimal(7,2)	2137	46	87.36	0.02	5.75			
▼	l_brand_id	int	952	44	10,016,017	1,001,001	5,502,225.27			
▼	l_brand	char(50)	734	0						
▼	l_class_id	int	15	35	16	1	5.45			
▼	l_class	char(50)	93	0						
▼	l_category_id	int	10	42	10	1	5.49			
▼	l_category	char(50)	11	0						
▼	l_manufact_id	int	1000	38	1,000	1	334.57			
▼	l_manufact	char(50)	971	0						
▼	l_size	char(20)	8	0						

CDPDSS-2959: Asset Details does not provide a link to the parent database

- Now, the parent database can be accessed via a link in **Asset Details**.
- The Qualified Name is no longer trimmed. Longer names are displayed in multiple rows.

Datasets / Details / Asset Details

item

Properties

Type: **HIVE TABLE**

of Columns: **22**

Data Lake:

Datasets: **8**

Owner: **hive**

Created On: **Mon Jun 17 2024 10:01:51 GMT+0200 (Cent...**

Last Access Time: **Mon Jun 17 2024 10:01:51 GMT+020...**

Table Type: **EXTERNAL_TABLE**

Database: **test_dss_db**

DB Catalog: **cm**

Parent: [test_dss_db](#)

Asset Type: **HIVE DB**

Qualified Name

test_dss_db.item@cm

Comment

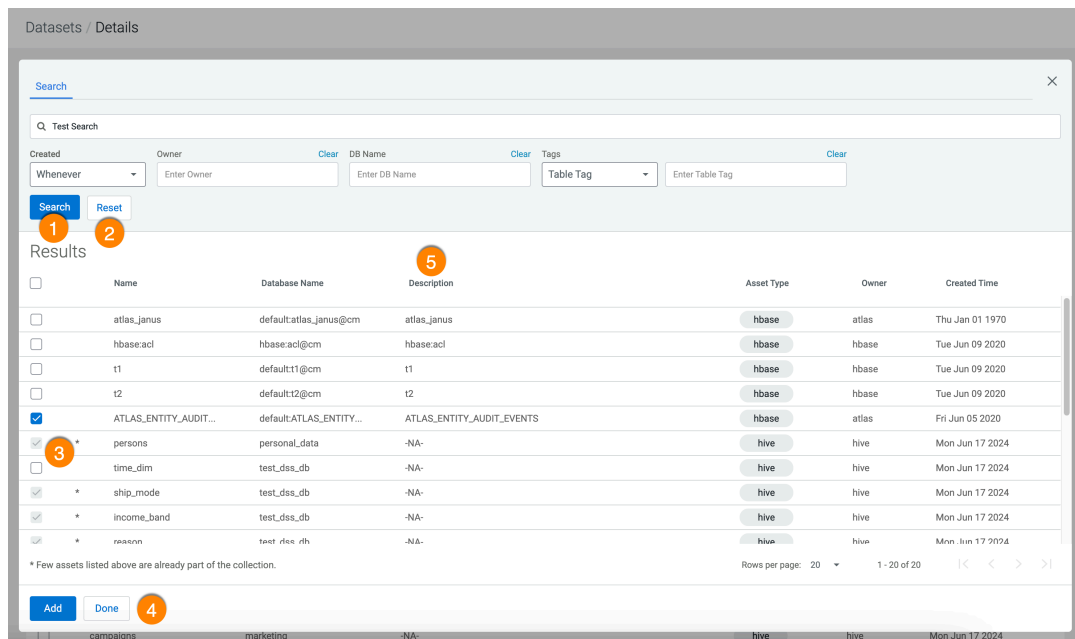
[+ Add Comment](#)

Description

[+ Add Description](#)

CDPDSS-2923: Improvements for the Dataset Detail page when adding assets to a Dataset

1. The Search button is already active after the search terms are entered but before the Enter key is hit.
2. Now, clicking Reset clears all the fields.
3. Already added assets are clearly marked with a check mark as added and non-editable.
4. When adding new assets to the Dataset, the Add and Done buttons are no longer hidden while the page is waiting for the API response.
5. The Description column is added to include the **Datasets Details** menu. The value of the Description field is the same as in Asset Details.



CDPDSS-3042: Edit classification is failing with guid undefined error

Trying to edit the classifications in the **Asset Details** screen no longer results in the "List properties for entity request failed Error from Atlas. Not Found : Received 404. Body {"errorCode": "\ATLAS-404-00-005\", \"errorMessage\": \"Given instance guid undefined is invalid/not found\"}" message.

CDPDSS-3049: Expanded graph is displayed on every page of the Schema tab

Previously, in the **Asset Details** page under the **Schema** tab, the bar graph of profiled assets might stayed visible for a selected asset even when moving to the next page in the list. Now, changing the page correctly closes the previously selected asset, and only displays the assets in the active page.

CDPDSS-3050: Username is not visible in Data Catalog for the Local Administrator user

The username for local administrator users is no longer hidden. It is shown in the bottom left corner of the navigation sidebar.

CDPDSS-3048: Filtering based on Job status Finished is not working

The profiler job filter Finished is working correctly now, displaying the relevant jobs.

CDPDSS-3046: Not able to see the profiled data of Ranger Audit Profiler if the Active NameNode is switched over to standby

The profiled data is shown correctly now after the Ranger Audit Profiler is run.

CDPDSS-3051: Profiler Jobs are not in the sorted order

Profiler jobs are correctly ordered from newest to oldest when the list cannot fit into a single page under Profiler Jobs .