

Data Catalog 1.5.3

Data Catalog Release Notes

Date published: 2023-10-10

Date modified: 2024-03-05

CLOUDERA

<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

What's new.....	4
Known issues.....	4
Fixed issues 1.5.3.....	6
Fixed issues in Cloudera Data Catalog (1.5.3-CHF1).....	7
Known issues in Cloudera Data Catalog (1.5.3-CHF1).....	7

What's new

This release of Cloudera Data Catalog on CDP Private Cloud 1.5.3 includes the following features and enhancements:

- Using the Download CSV option to save queried results.
- 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-2707: Custom Sensitivity Profiler (CSP) regex logic fails intermittently for complex regex

CSP regex logic fails intermittently for complex regex. Due to this intermittently the Cap tags do not get attached to Hive columns on execution of sensitivity profilers.

None

CDPDSS-2706: Data Steward user role is not reflecting

Data Steward role is not reflecting in Control Plane.

None

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

You must restart your Data Catalog Pod.

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

CDPDSS-3077:

Datasets containing deleted assets fail to render the assets list with the following error: pages.dataset.labels.err-undefined-msg(Error Code - undefined)

None

CDPDSS-3353: Profiler Configuration page is giving error, if browser's preferred language is not UK English

The Profilers Configs is may be incorrectly rendered if the web browser's preferred language is not set to en-gb.

CDPDSS-2913: Asset Details view does not provide details of the asset - exception

Entities without lineage information may display the get lineage request failed Error from Atlas Not Found : Received 404..

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

Cloudera Data Catalog running on JDK-17 in PvC causes ClassNotFoundException and further failures which leads to the service not coming up with a Bad Gateways error and being unusable. The following is logged in the Data Catalog pod:

```

----
        INFO from c.g.i.Guice in [main] - - An exception was
        caught and reported. Message:
        java.lang.reflect.InaccessibleObjectException: Unable
        to make protected final
        java.lang.Class
        java.lang.ClassLoader.defineClass(java.lang.String,b
        yte[],int,int,java.security.ProtectionDomain)
        throws java.lang.ClassFormatError accessible: module
        java.base does not "opens java.lang"
        to unnamed module @7a799159 - java.lang.IllegalState
        Exception: Unable to load cache item
        ----
  
```

Use the environment variable in the deployment in the container specification:

```

----
        name: JDK_JAVA_OPTIONS
        value: --add-opens java.base/java.lang=ALL-UNNAMED --
        illegal-access=warn
        ----
  
```

CDPDSS-3346: Profiler pods are stuck in init state

The CDP_ACCESS_KEY_ID and CDP_PRIVATE_KEY keys do not receive values when upgrading from Cloudera Data Catalog 1.5.2 to 1.5.4.

1. Use the following command to restart the DSS App: `kubectl rollout restart deployment/cdp-release-dssapp -n <namespace>`
2. Delete and relaunch the profilers after the upgrade.

CDPDSS-3081: Datasets: Tenant Schema is getting passed as workload username while impersonating a user on ATLAS and Ranger API calls

While making API calls to Apache Atlas and Apache Ranger, there is a potential bug in code where a wrong value for workloadUserName is getting passed which causes the APIs to perform an auth check with a wrong parameter, causing unwanted issues.

None

CDPDSS-3080: Deleted Entities in Atlas cannot be removed from Datasets

Entities deleted in Apache Atlas cannot be removed from Cloudera Data Catalog datasets.

None

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

The username for local administrator users can be hidden instead of showing in the bottom left corner of the navigation sidebar.

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

In the **Asset Details** page under the **Schema** tab, the bar graph of profiled assets might stay visible for a selected asset even when moving to the next page in the list.

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-3042: Edit classification is failing with guid undefined error

Trying to edit the classifications in the **Asset Details** screen can result 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-3037: Asset Details page does not highlight link for Datasets

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

CDPDSS-3524: Profiler pod launch is failing with init error

During the profiler cluster setup, when all the services pods are being deployed, Cloudera Data Catalog attempts to complete the prerequisites to set up a service and request to create a machine user. If the UMS service has not come up properly at that time, the request fails, and does not create the machine user dp_profiler_user required for profiling.

1. Go to Management Console User Management Users .
2. Search the machine user dp_profiler_user. Continue with the following steps, if the machine user is missing.
3. Use the following command to restart Cloudera Data Catalog: `kubectl rollout restart deployment/cdp-release-dssapp -n <***NAMESPACE***>`.
4. Delete and relaunch the profilers.

CDPDSS-3472: Profiler jobs are showing up twice when two data lakes are present

Cloudera Data Catalog does not support more than one data lakes (or base clusters). Connecting two data lakes will result in all profiler jobs showing as duplicated in the **Profilers** page.

CDPDSS-3395: Incorrect results are being displayed after navigation, when "owners" filter is selected

In the **Search** page, when returning to the list of results after opening the **Asset Details** of one the results and repeating the search with the same filter, other entities belonging to owners different than the selected can appear.

CDPDSS-3381: The Ranger Audit Profiler is showing status as FINISHED/SUCCESS even though there is error with the profilers

The Ranger Audit Profiler can shows an incorrect status when the NameNode is stopped, and it is can get stuck in FINISHED.

CDPDSS-2688: Hive Column On-demand Profiler is failing but the Asset Details page is not updated with 'Failed State'

A failed Hive Column On-Demand Profiler job is incorrectly not updated to Failed state in Asset Details Profilers .

Fixed issues 1.5.3

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

CDPDSS-2708: Kerberos kinit expires leading to all Atlas API calls failure

Kerberos kinit expires after 48 hours. Due to this scenario, all Atlas APIs start failing leading to blockage on Data Catalog functionality.

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-2673: Download CSV feature is not supported

CSV files can be successfully downloaded.

Fixed issues in Cloudera Data Catalog (1.5.3-CHF1)

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

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

The issue when Cloudera Data Catalog running on JDK-17 in PvC caused ClassNotFoundException and further failures which led to the service not coming up and unusable was fixed.

Known issues in Cloudera Data Catalog (1.5.3-CHF1)

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

CDPDSS-3346: Profiler pods are stuck in init state

The CDP_ACCESS_KEY_ID and CDP_PRIVATE_KEY keys do not receive values when upgrading from Cloudera Data Catalog 1.5.2 to 1.5.4.

1. Use the following command to restart the DSS App: `kubectl rollout restart deployment/cdp-release-dssapp -n <namespace>`
2. Delete and relaunch the profilers after the upgrade.

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

CDPDSS-3081: Datasets: Tenant Schema is getting passed as workload username while impersonating a user on ATLAS and Ranger API calls

While making API calls to Apache Atlas and Apache Ranger, there is a potential bug in code where a wrong value for workloadUserName is getting passed which causes the APIs to perform an auth check with a wrong parameter, causing unwanted issues.

None

CDPDSS-3080: Deleted Entities in Atlas cannot be removed from Datasets

Entities deleted in Apache Atlas cannot be removed from Cloudera Data Catalog datasets.

None

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-3042: Edit classification is failing with guid undefined error

Trying to edit the classifications in the **Asset Details** screen can result 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-3037: Asset Details page does not highlight link for Datasets

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

CDPDSS-3524: Profiler pod launch is failing with init error

During the profiler cluster setup, when all the services pods are being deployed, Cloudera Data Catalog attempts to complete the prerequisites to set up a service and request to create a machine user. If the UMS service has not come up properly at that time, the request fails, and does not create the machine user dp_profiler_user required for profiling.

1. Go to Management Console User Management Users .
2. Search the machine user dp_profiler_user. Continue with the following steps, if the machine user is missing.
3. Use the following command to restart Cloudera Data Catalog: `kubectl rollout restart deployment/cdp-release-dssapp -n <***NAMESPACE***>`.
4. Delete and relaunch the profilers.

CDPDSS-3472: Profiler jobs are showing up twice when two data lakes are present

Cloudera Data Catalog does not support more than one data lakes (or base clusters). Connecting two data lakes will result in all profiler jobs showing as duplicated in the **Profilers** page.

CDPDSS-3395: Incorrect results are being displayed after navigation, when "owners" filter is selected

In the **Search** page, when returning to the list of results after opening the **Asset Details** of one the results and repeating the search with the same filter, other entities belonging to owners different than the selected can appear.

CDPDSS-3381: The Ranger Audit Profiler is showing status as FINISHED/SUCCESS even though there is error with the profilers

The Ranger Audit Profiler can shows an incorrect status when the NameNode is stopped, and it is can get stuck in FINISHED.

CDPDSS-3353: Profiler Configuration page is giving error, if browser's preferred language is not UK English

The **Profilers Configs** is may be incorrectly rendered if the web browser's preferred language is not set to en-gb.

CDPDSS-2688: Hive Column On-demand Profiler is failing but the Asset Details page is not updated with 'Failed State'

A failed Hive Column On-Demand Profiler job is incorrectly not updated to Failed state in **Asset Details Profilers** .