

Data Catalog 1.5.2

Data Catalog Release Notes

Date published: 2023-11-02

Date modified:

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
Fixed issues.....	5

What's new

This release (1.5.2) marks the General Availability (GA) release of the Cloudera Data Catalog service on CDP Private Cloud Data Services.

- Cloudera Data Catalog service is available when you upgrade to CDP Private Cloud Data Services 1.5.2 version
- Data Catalog enables Data Stewards and Business Analysts across the enterprise to work with data assets in the following ways:
 - Organize and curate data globally
 - Understand where relevant data is located
 - Understand how data access is secured, protected, and audited
 - Understand how data is interpreted for use
 - The Asset 360 degree view of entities within Atlas helps analysts find high quality curated data and leverage it for decision making.
 - Support for Profiling of Hive tables, to examine, analyze and review data, providing organizations with a high level view of data quality.

Known issues

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

Known Issues

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

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

On Cloudera Manager ---> administration ---> security ----> kerberos credentials tab --> search for dpprofiler and then click on regenerate selected. Besides this user needs to refresh the dss-app pod on the namespace using the command: `kubectl rollout restart deployment/cdp-release-dssapp -n <namespace name>`

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-2673: Download CSV feature is not supported

A CSV file is not getting downloaded.

None

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.

None

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-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-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..

Fixed issues

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

ENGESC-24257: Control Plane upgrade failed with "must be no more than 63 characters"

Creation of routes with the long namespace for the Data Catalog service did not succeed The system was stuck in that state. The upgrade process was not complete and the post-upgrade tasks were pending.

There is a maximum of 63 characters allowed in the DNS label. The strings between the dots, for routes built in OpenShift. You can specify a custom hostname to get up to 253 characters in the route as long as none of the DNS labels exceed 63 characters.