

## Cloudera Data Catalog Release Notes

Date published: 2019-11-14

Date modified: 2025-10-17



# 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

<b>What's new.....</b>	<b>5</b>
October 17, 2025.....	5
Apache Parquet CVE-2025-30065.....	5
July 23, 2025.....	6
June 17, 2025.....	6
April 8, 2025.....	6
January 16, 2025.....	8
 <b>Older releases.....</b>	 <b>9</b>
Cloudera Data Catalog releases for 2024.....	9
December 18, 2024.....	9
June 03, 2024.....	9
May 16, 2024.....	9
April 3, 2024.....	10
Cloudera Data Catalog releases for 2023.....	10
December 15, 2023.....	10
March 13, 2023.....	10
January 17, 2023.....	11
Cloudera Data Catalog releases for 2022.....	11
April 1, 2022.....	11
March 2, 2022.....	11
February 16, 2022.....	11
Cloudera Data Catalog releases for 2021.....	12
December 9, 2021.....	12
October 26, 2021.....	12
August 27, 2021.....	12
October 5, 2021.....	12
July 26, 2021.....	12
May 31, 2021.....	13
April 5, 2021.....	13
March 25, 2021.....	13
March 15, 2021.....	13
Cloudera Data Catalog releases for 2020.....	13
December 14, 2020.....	14
October 29, 2020.....	14
August 27, 2020.....	14
August 12, 2020.....	14
July 25, 2020.....	14
June 30, 2020.....	15
June 17, 2020.....	15
June 5, 2020.....	15
May 6, 2020.....	15
April 7, 2020.....	15
January 6, 2020.....	16
Cloudera Data Catalog releases for 2019.....	16
November 14, 2019.....	16
September 23, 2019.....	16
August 22, 2019.....	16

**Known issues..... 17**

**Fixed issues..... 18**

## What's new

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

### October 17, 2025

This release (3.1.3) of the Cloudera Data Catalog service introduces the following new changes:

- The Cluster Sensitivity Profiler and the Statistics Collector Profiler support incremental profiling to reduce required time and compute resources during repeated profiling jobs in Compute Cluster enabled environments.

For more information, see:

- [Configuring the Statistics Collector Profiler](#)
- [Configuring the Data Compliance Profiler](#)
- The new **Asset Filtering Rules** tab in **Job Summary** shows the relevant Allow and Deny list rules for each Data Compliance and Statistics Collector Profiler job.
- Bug fixes and improvements.

#### Related Information

[Known issues](#)

[Fixed issues](#)

### Apache Parquet CVE-2025-30065

Cloudera released Cloudera Data Catalog 3.1.3 for Cloudera Data Services on cloud to address a critical vulnerability in the parquet-avro module of Apache Parquet.

#### Background:

On April 1, 2025, a critical vulnerability in the parquet-avro module of Apache Parquet ([CVE-2025-30065](#), [CVSS score 10.0](#)) was announced.

Cloudera has determined the list of affected products, and is issuing this TSB to provide details of remediation for affected versions.

Upgraded versions are being released for all currently affected [supported releases](#) of Cloudera products. Customers using older versions are advised to upgrade to a [supported release](#) that has the remediation, once it becomes available.

#### Vulnerability Details:

Exploiting this vulnerability is only possible by modifying the accepted schema used for translating Parquet files and subsequently submitting a specifically crafted malicious file.

[CVE-2025-30065](#) | Schema parsing in the parquet-avro module of Apache Parquet 1.15.0 and previous versions allows bad actors to execute arbitrary code.

#### CVE:NVD - CVE-2025-30065

Severity (Critical):

[CVSS:4.0/AV:N/AC:L/AT:N/PR:N/UI:N/VC:H/VI:H/VA:H/SC:H/SI:H/SA:H](#)

#### Impact:

Schema parsing in the parquet-avro module of Apache Parquet 1.15.0 and previous versions allows bad actors to execute arbitrary code. Attackers may be able to modify unexpected objects or data that was assumed to be safe from modification. Deserialized data or code could be modified without using the provided accessor functions, or unexpected functions could be invoked.

Deserialization vulnerabilities most commonly lead to undefined behavior, such as memory modification or remote code execution.

#### Addressed in release:

Upgrade to the following release containing the fix:

- Cloudera Data Services on cloud (formerly Public Cloud Data Services)
  - Release 3.1.3



**Note:** Cloudera Data Catalog with Compute Cluster enabled environments supports all versions of Cloudera on cloud, V1 environments support only CDP Public Cloud 7.2.18 and earlier versions.

For the latest update on this issue, see the corresponding Knowledge article:

- [Cloudera Customer Advisory 2025-847: Cloudera's remediation actions for Apache Parquet CVE-2025-30065](#)

## July 23, 2025

This release (3.1.2) of the Cloudera Data Catalog service introduces the following new change:

You are now able to approve tags recommended by the Data Compliance Profiler before applying them to your assets and syncing them to Apache Atlas. This means that you can review tag suggestions to correct mistakenly applied tags, which would otherwise lead to unexpected changes of tag-based Apache Ranger policies.

#### Related Information

[Approving Data Compliance Profiler tags](#)

## June 17, 2025

This release (3.1.1) of the Cloudera Data Catalog service introduces the following new changes:

- Profilers in Compute Cluster enabled environments now support profiling text based files. Custom delimiters are supported for Hive tables where the Hive Metastore contains them. Cloudera Data Catalog now supports LazySimpleSerDe for Hive tables. For more information, see the [Apache Hive Developer Guide](#).
- This release contains fixes and updates for the changes in the 3.1.0 release. For more information, see the list of fixed and known issues.

#### Related Information

[Known issues](#)

[Fixed issues](#)

## April 8, 2025

This release (3.1.0) of the Cloudera Data Catalog service introduces the following new changes:

#### Improved services for profilers

- Thanks to the improved Cluster Setup API, the configuration of profilers is simplified
  - Executor related settings only specify the maximum number of workers, an internal service manages the autoscaling within this range

#### Redesigned profiler setup

- Settings for instance sizing and autoscaling are introduced

## Improved profiler UI

The improved profilers present a more user friendly UI and several extended capabilities for Compute Cluster enabled environments.

- New names for profilers in Compute Cluster enabled environments:
  - The Cluster Sensitivity Profiler is now called Data Compliance profiler.
  - The Hive Column Profiler is now called Statistics Collector profiler.
  - The Ranger Audit Profiler is now called Activity Profiler.
- Redesigned **Profilers** menu for easier access to jobs, configurations and their history, asset filtering and tag rules:
  - The individual profilers show new metrics
    - Number of profiled assets of the last job
    - Job duration of the last job
    - The profilers menu also shows the next jobs' start time and the number of completions
  - The CRON expression based scheduler is supplemented with a natural language based scheduler
  - **Asset Filtering Rules** is expanded with the list of assets affected by your rule set
  - You can now access the **Configuration History** of a profiler, where you can check your changes in a sequential order
  - The **Job Summary** page is introduced new metrics:
    - Workers details:
      - Worker Memory limit
      - Threads per workers
      - Number of workers
    - Last run check details
  - The **Job Summary** page provides the list of profiled assets.

## Redesigned and expanded Tag Rules for Compute Cluster enabled environments

- Profiling table names is introduced next to column values or column names.
- Atlas classifications (Cloudera Data Catalog tags) can be used in a more granular way thanks to the distinction between parent and child tags.
- Tag rules are data lake specific in Compute Cluster enabled environments compared to being valid for all data lakes in VM-based environments.
- The new **Tag Rules** tab offers filters to allow for faster searching and displays:
  - List of applied parent and child tags
  - Tag rule status (Can be used to filter for tag rules not yet validated by Dry Run)
  - Rule types
  - You can filter for tag rules that apply child tags
- The initial loading time of rules has been decreased.
- You can upload regex patterns in CSV files for easier handling.
- Now you can specify weightage for column value based matching (which was fixed at 85% before). The column weightage and column name weightage add up to 100%.
- When profiling column values, you can upload a sample set of column values instead of defining a regex pattern.
- You can review your configuration before finalizing your tag rule.
- Dry Run: Before deploying your tag rules, you have to test them with actual table data.
- New API calls are available.

## New file formats for Compute Cluster based profilers

Compute Cluster based profilers also support the ORC and Avro file format.

## January 16, 2025

This release (2025-M1) of the Cloudera Data Catalog service introduces the following new changes:

### Containerized architecture for profilers

Cloudera Data Catalog introduces a new containerized architecture for Profilers for Compute Cluster enabled environments, providing a scalable environment:

- Only the required amount of Kubernetes pods are launched based on the size of the database to be profiled. You need to pay only for the used cloud resources only while they are used by the profilers.
- Also, the deployment of the containerized profiler architecture is more streamlined and quicker than the previous VM-based architecture.
- Moreover, the containerized nature of the architecture means that later upgrades can be carried out easier, without the need for multiple dependencies as in the VM-based architecture which used multiple services.
- Profilers now also support the following file formats:
  - VM-based environments: CSV, ORC
  - Compute Cluster enabled environments: CSV and Parquet
  - Hive Column Profilers and Cluster Sensitivity Profilers also support profiling Iceberg Tables, including with On-Demand Profilers.



#### Important:

- Currently, Kubernetes based profilers are only supported in AWS environments.
- In Compute Cluster enabled environments, profilers only support tables which are stored on AWS S3 storage.

For more information, see [Profiler architecture in Compute Cluster enabled environment](#).



#### Note:

The VM-based architecture (using the Cluster) is deprecated from the 3.0.0 release but remains available until 7.2.18 is supported (Sept 2025). Therefore, based profilers will also not be available in versions after 7.2.18. Only Compute Cluster enabled environment will be able to run profilers after version 7.2.18.

For more information, see [Cloudera Support lifecycle policy](#).

### Redesigned Dashboard menu

- A new **Dashboard** is introduced to give a overview of your data lakes and profilers including:
  - Data lake type and status
  - Profiler status
  - Last 10 assets bookmarked by you
  - Last run of profiler
  - Number of assets profiled

### Redesigned Search menu

The **Search** menu is reorganized so information is easier to access. You can expand each entity result to see their qualified name, database, classification and assigned terms. You can use these to check if your query returns the expected results.

### Improved display of comments in Asset Details

Following this release, you can hover over the **Comment** field for individual schema entries in **Asset Details** to preview longer comments without opening them.



### Common time format

**Asset Details** and other menus will use the same time format for a more readable overview: MM/DD/YYYY hh:mm A.

### Removed features

The following features have been removed:

- Tag Rules Luhn check algorithm
- Tag Rules File-based Allow and Deny list
- Tag Rules Lookup files
- Tagging multiple assets in the **Search** menu

## Older releases

Overview of new features, enhancements, and changed behavior introduced in earlier releases of Cloudera Data Catalog.

## Cloudera Data Catalog releases for 2024

Release summaries:

### December 18, 2024

This release (2.0.28) of the Cloudera Data Catalog service introduces the following new changes:

#### Column name based tagging in Cluster Sensitivity Profiler

You can override the sampling to profile data in a column based on the column name matching a preset regular expression pattern instead matching the certain percentage of the columns values. This can be used for assets with skewed proportions where relying on the sampling would not result in correct tagging.

For more information, see [Configuring the Cluster Sensitivity Profiler](#) and [Setting up column name based tagging](#).

### June 03, 2024

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

This release of the Cloudera Data Catalog service introduces the following new changes:

This release only contains fixes and updates to prepare Cloudera Data Catalog for the changes in the upcoming 3.0.0 release.


### May 16, 2024

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

This release of the Cloudera Data Catalog service introduces the following new changes:

Iceberg tables are now supported by the Cloudera Data Catalog service:

- You are able to filter for them in the Search page.
- Iceberg tables can be viewed in the Asset Details page.
- Iceberg tables can be added to a dataset.

- All subcomponents of Cloudera Data Catalog support JDK 17.
-  **Important:** Profilers do not support Iceberg tables in this release.

## April 3, 2024

This release of the Cloudera Data Catalog service provides you with a notable behavior change which you must note and act accordingly.

While upgrading your cluster from Cloudera Runtime version **7.2.17** to **7.2.18**, and specifically during the OS upgrade step, the cluster goes into the failure state. The following message is seen:

\_\_NODE\_FAILURE:

New node(s) could not be added to the cluster. Reason Please find more details on Cloudera Manager UI. Failed command(s): Start(id=1546339088): Failed to start role profc6cf3856-PROFILER\_SCHEDULER\_AGENT-484032cb8f17cacf9e684efe50 of service profiler\_scheduler in cluster cdp-dc-profilers-258395ef\_

### Impact on Cloudera Data Catalog profilers:

If the Cloudera Data Hub is not created, then the Cloudera Data Catalog profilers will not be created in Cloudera Runtime 7.2.18 version.

To overcome this scenario, you must use the following process to bring up the Cloudera Data Catalog profilers in the Cloudera Runtime 7.2.18 version.

First you must delete your existing 7.2.17 clusters. For more information, see [Deleting profiler cluster](#).

Next, after you upgrade to the 7.2.18 Data Lake, then you can launch the Cloudera Data Catalog profilers. For more information, see [Launch profiler cluster](#).



**Note:** There is no data loss expected on the users' side or the Profiler analysis. However, the only loss that could be expected is related to the last runtime value of the profiler and the profiler run history. The Profiler Last Runtime history refers to the records of how many runs of the profiler are displayed on the history page. It includes information on whether the runs were completed successfully or resulted in failures.

## Cloudera Data Catalog releases for 2023

Release summaries:

### December 15, 2023

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following features as generally available:

Cloudera Data Catalog is now integrated with AWS Glue Data Catalog which enables the users to browse and discover data as well as register data into Cloudera Shared Data Experience ( using metadata translation or copy), so that it can be used with Cloudera Data Hub and other relevant experiences.

### March 13, 2023

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following features as generally available:

- Supporting High Availability for profiler services.
- Navigating to tables and databases in Hue using Cloudera Data Catalog.
- Using the Download CSV option to save queried results.

## January 17, 2023

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following additions:

- Cloudera Cloudera Data Catalog is now additionally supported in the EU and APAC regional Control Plane.
- Associating Customer Sensitivity Profiler tags to Hive assets fails on the 7.2.16 Cloudera Runtime cluster. You must update the configurations. For more information, see [Associating tags to Hive assets](#).

## Cloudera Data Catalog releases for 2022

Release summaries:

### April 1, 2022

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following additions:

Profiler cluster launch requires specific instance type of the Master node.

### March 2, 2022

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following additions:

- Support for Cloudera Runtime 7.2.14
- No end-user manual process needed for backing up and restoring profiler data
- Post-upgrade (to Cloudera Runtime 7.2.14 version), deleting profiler cluster does not require saving the modified Custom Sensitivity Profiler rules

### February 16, 2022

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following additions:

- Deleting profiler cluster with multiple scenarios
- Bug fixes

## Cloudera Data Catalog releases for 2021

Release summaries:

### December 9, 2021

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following additions:

- Deleting profiler cluster
- Bug fixes

### October 26, 2021

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following addition:

Available location paths for the Backup and Restore scripts based on the Cloudera Runtime version.

### August 27, 2021

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following addition:

- Bug fixes. Refer to the Fixed Issues section in the Release Notes document.
- Refreshed screenshots for profiler configuration and related pages.
- Profiler cluster creation using the new UI
- The option selected in Depth drop-down and Show Process nodes are now displayed on the upper-left corner of the Lineage module.

### October 5, 2021

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following addition:

- Support for deleting the profiler cluster
- Updated asset details
- Bug fixes and refreshed UI

### July 26, 2021

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following addition:

- Creating classifications on the search page or on the classification widget.
- Bug fixes. Refer to the Fixed Issues section in the Release Notes document.
- Updated in-help documentation for multiple components in the UI.

## May 31, 2021

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following addition:

- Launching Profilers using the Command-line Interface (CLI).
- Bug fixes. Refer to the Fixed Issues section in the Release Notes document.

## April 5, 2021

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following addition:

Support for Business Glossary and usage of "terms" to associate datasets.

## March 25, 2021

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following addition:

Backing up and Restoring Profiler Database.

## March 15, 2021

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following additions:

- Searching for asset types using additional options in Cloudera Data Catalog and Apache Atlas
- Bug fixes

## Cloudera Data Catalog releases for 2020

Release summaries:

## December 14, 2020

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following additions:

- Support for AWS and Azure in Cloudera Runtime 7.2.6 version
- Atlas entity audits in the **Asset Details** page
- Profiler jobs tracking status
- Audit graphs for HBase tables

## October 29, 2020

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following additions:

- Support for searching assets across multiple data lakes.
- A new tab called Contents is introduced to list the contents of the select entity.
- Support for tagging multiple assets.

## August 27, 2020

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following additions:

- Multiple entity type selections to search for assets
- Other UI fixes and improvements

## August 12, 2020

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following additions:

- Multiple filter type selections on the **Search** page
- Other UI fixes and improvements

## July 25, 2020

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following additions:

- Support for improved user experience for the **Search** page

- Atlas tags option is now directly accessible from the menu bar

## June 30, 2020

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following additions:

- Tag management
- On-Demand profiler support for Hive Column Statistics Profiler and Cluster Sensitivity Profiler
- Last Run Check Configuration
- Support for adding Hive asset to one or more datasets from the **Asset Details** screen

## June 17, 2020

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following changes and bug fixes:

- UI improvements for search capabilities
- Ability to search Hive and HBase assets using Column tag

## June 5, 2020

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following changes:

- About accessing Data Lakes
- Support for prepopulating the Asset owners
- Support for selecting more entity types for searching assets

## May 6, 2020

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following changes and bug fixes:

Navigation Support for Hive entity within Lineage.

## April 7, 2020

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following new features:

- Profiler Tag Rules
- Hive Column Statistical Profiler

## January 6, 2020

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments.

Cloudera Data Catalog introduces the following new features:

- Ability to run Sensitivity Profiler jobs
- Ability to collaborate with other users - you can now rate datasets, add comments about asset collections and respond to the comments of other users.

## Cloudera Data Catalog releases for 2019

Release summaries:

### November 14, 2019

This release of the Cloudera Data Catalog service introduces the following new features.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments (on-premises, cloud, and IOT).

Cloudera Data Catalog introduces the following new features:

- Ability to view Profiler jobs
- Ability to edit Profiler configuration
- Ability to toggle a Profiler on and off.

### September 23, 2019

This release of the Cloudera Data Catalog service introduces the following new features:

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments (on-premises, cloud, and IOT).

Cloudera Data Catalog enables data stewards 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 is interpreted for use
- Understand how data is created and modified
- Understand how data access is secured and protected, and audit use

### August 22, 2019

This release of the Cloudera Data Catalog service introduces the following new features:

This is the first release of the Cloudera Data Catalog service.

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across enterprise data lakes. Cloudera Data Catalog helps you understand data across multiple clusters and across multiple environments (on-premises, cloud, and IOT).

Cloudera Data Catalog enables data stewards 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 is interpreted for use
- Understand how data is created and modified
- Understand how data access is secured and protected, and audit use

## Known issues

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

### **CDPDSS-2488: In MGMT service of DH CM, the telemetry command is failing bringing the master node down**

When a Cloudera Data Hub cluster is created, the management service of Cloudera Manager becomes unavailable because the telemetry command fails on Hive. Because of this, the Master node of Cloudera Data Hub goes down causing a Node Failure error on the cluster.

In Cloudera Data Catalog, due to the failure of the master node of Cloudera Data Hub, the **Profiler** and the **On-Demand Profiler** page in the **Asset Details** page does not load.

It is recommended to use profilers on a Compute Cluster enabled environment to avoid this issue.

### **CDPDSS-3403: Altered ICEBERG table is not available in Compute Cluster enable environment**

Iceberg tables modified or renamed with the ALTER operation appear with both old and new "name", however, neither entry can be accessed. Apache Atlas synchronizes the table with the altered name with its full qualified name instead of its name shown in the **Search** results.

### **CDPDSS-3412: Profiling is not working on migrated ICEBERG table**

Hive tables migrated to Iceberg with the ALTER TABLE statement cannot be profiled in Compute Cluster enabled environments. Cluster Sensitivity Profilers and Hive Column Profilers will not be able to apply tags to the values.

### **CDPDSS-3387: Profiling of struct or composite data type are not supported in Cloudera Data Catalog**

Profiling of struct or composite data type are not supported in Cloudera Data Catalog. When profiling tables with such data, the profiling job will be stuck in the Undetermined status.

### **CDPDSS-3561: The number of columns is displayed incorrectly when the DB or table name is too long**

In **Asset Details**, the number of columns (**# of Columns**) is displayed as zero when the database or table name is too long.

### **CDPDSS-3852: Number of Columns count is getting incremented on creating new tables**

In **Search Asset Details**, the **# number of Columns** value may not match the real number of columns in the table.

None

### **CDPDSS-4051: Node Group setup API responds with HTTP code 202 even for quota Issues - (400 Bad Request)**

If the resource quota request fails after a profiler setup attempt, the cloud provider may return a bad request response with a 400 status code. Although the profiler launch seems successful first, the following error message is returned and the profiler is unavailable: The auto-scaling node group is unavailable at the moment but profilers seem to have been launched. This has caused an inconsistency on Cloudera Data Catalog. Please delete all profilers and relaunch them to continue profiling your assets.

None

### **CDPDSS-4042: Even though profiler is launched, audits charts and on demand profilers are missing**

If the distroXCrn is not received in the API response after a profiler is launched, on-demand profilers and audit logs are not visible.

Try to reload the **Profilers** page. If the reload is not working, delete your browser's cache and try using Incognito mode.

**CDPDSS-4058: The basic cron expression returns a Not a Number for leap year dates**

If you set the **Basic Schedule** of a profiler to a leap day in a leap year, the schedule cannot be saved.

Use a day different from the leap day in your schedule. Alternatively, you can use cron expressions with leap days.

**CDPDSS-4133: Column count is incorrect in the pagination footer**

In **Asset Details Schema** can show a larger number of total rows in the pagination counter than available.

**CDPDSS-4134: When user rejects a tag there should be a dialogue box to reconfirm the rejection**

There is no dialog for confirmation in **Asset Details Classifications** when rejecting recommended classifications.

**CDPDSS-4151: Edit Tag Rule screen lacks Save button in intermediate step**

While editing a tag rule for the Data Compliance Profiler, there is no Save button available in the intermediate steps. As a result, users are forced to proceed to the final step just to save their changes.

**CDPDSS-4266: Stuck profiler pods should be deleted**

In Compute Cluster enabled environments, the profilers remain stuck if the Compute Cluster is unable to provision the necessary resources. As a result, the subsequently scheduled jobs (of the same kind) will not be scheduled until the resources are provisioned.

Go to the compute cluster and delete the stuck profiler pod. Update the profiler configs to compensate for the resource crunch, to help the profilers to get scheduled properly next time.

## Fixed issues

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

**CDPDSS-81: Cloudera Data Catalog Issue**

Problem: If an error occurs while fetching results, an error message appears briefly and disappears. Once the message disappears, the results pane is stuck in the Loading state.

**CDPDSS-81: Cloudera Data Catalog Issue**

Problem: Despite having no privileges to add, modify, or delete classifications, users are able to edit the tag flow on the **Asset Details** page. However, users will not be able to save the changes.

**CDPDSS-81: Cloudera Data Catalog Issue**

Problem: On the **Asset Details** page, the Audit tab fails to load if the logged in user does not have Audit read privilege.

**CDPDSS-363: The 403 error message seen in Cloudera Data Catalog > Table / Audit tab**

Problem: Error messaging needs to be improved for users accessing Table and Audit tab without sufficient Ranger role.

**CDPDSS-437: Unauthorized message for tag management flow does not have column name**

Problem: Unauthorized column names and comma visible for tag management flow.

**CDPDSS-444: Duplicate tag based policy on Asset Details Policy tab**

Problem: When a specific tag is present both on the table as well as on any column of the same table, and a couple of requests to Ranger with the same tag and it displays the policy twice on list.

**CDPDSS-527: Cloudera Data Catalog specific user roles not supported.**

Problem: The following user roles are not supported:

- DataCatalogCspRuleManager
- DataCatalogCspRuleViewer

Adding or removing these roles in the User Management will not have any impact on Cloudera Data Catalog user access.

#### **CDPDSS-81: Cloudera Data Catalog Issue**

Problem: When you navigate to the **Cloudera Data Catalog** page from the **Environments** page or **Data Lakes** page, the context is not retained and the **Cloudera Data Catalog** page loads with the last used context.

Workaround: You must change the context manually by selecting the right data lake.

#### **CDPDSS-302: Profiler Manager server requires a manual restart after installation for DP Tags to reflect in Atlas**

Problem: Once the profiler is launched, run the sensitivity profiler, the profiler recognises the sensitive tags but fails to push them to Atlas.

Workaround: You must restart the profiler manager. Note that restarting the profiler manager is a one-time task.

#### **CDPDSS-360: Sensitive and Hive Column Profilers do not support managed Hive tables.**

Problem: Profilers cannot access the storage file of the managed table.

#### **CDPDSS-517: Dataset View: Asset list displays loading icon if user is not authorized on Environment on which the dataset is created.**

Problem: The dataset view page breaks if user does not have access to the Environment on which the dataset is created.

#### **CDPDSS-570: The Filter type list should be refreshed based on the data lake selection.**

Problem: Selecting the filter type from the drop-down does not display the supported entity type and also in the radio button list.

#### **CDPDSS-547: Data lake name missing on the Asset Details page.**

Problem: The data lake name was not visible on the Asset Details page .

#### **CDPDSS-594: Database filter and create dataset button gets hidden**

Problem: When you clear the query used the clear button, the databaser filter gets hidden and create dataset button on the search page gets disabled.

#### **CDPDSS-636: User able to view assets of a dataset which is available in an inaccessible data lake.**

Problem: Users are able to view the assets in a data lake for which they have no permission or access.

#### **CDPDSS-672: On the Asset-Details page, when you make changes and save them, the Scheme page is refreshed, the Schema tab is disabled, and Overview tab becomes operational.**

Problem: When you perform some operations on the **Asset Details** page and save the changes, the **Schema** tab gets inactive.

#### **CDPDSS-727: CSP Custom Rule Validation is getting stuck in "Validation Running" status**

Problem: The rule gets stuck in the Validation Running status.

Workaround: Keep sensitivity profiler enabled when running CSP rule validation.

#### **CDPDSS-955: Cloudera Data Catalog uses wrong attribute to read createTime of entities and misinterprets lastAccessTime attribute.**

Problem: Cloudera Data Catalog does not match the Created On date with Atlas.

#### **CDPDSS-949: Asset registration pertaining to GLUE integration is not possible.**

Problem: Asset registration failing with GLUE integration.

**CDPDSS-921: Table created few days ago and configured recently is getting backlisted.**

Problem: When an asset created date is older than the filter applied for black-listing, the asset still gets blacklisted (not picked during the profiler run).

**CDPDSS-1007: On-Demand Profiler remains in "in-progress" state if the profiler job fails.**

Problem: The On-Demand Profiler remains in the "in-progress" state if the profiler job fails due to some exception (For example: Permission denied exception).

**CDPDSS-144: Filter Results Count Mismatch**

Problem: When you search and filter the results using Created Before, the count displayed in the filters panel is incorrect. But, the number of search results is accurate.

**CDPDSS-144: Search icon within dataset does not work**

Problem: When you try to search for an asset within a dataset and click the search icon, it does not yield any results.

Workaround: Press enter to view the search results after entering a string in the search box.

**CDPDSS-143: Edit button appears after creating a new dataset**

Problem: After creating a new dataset, when you visit the dataset, the Edit button appears instead of Add Assets.

Workaround: Refresh your browser and reopen the dataset. You will be able to see the Add Assets button.

**CDPDSS-81: When you select a table tag, owner filters along with the counts disappear**

Problem: When you select a table tag, owner filters along with the counts disappear.

Workaround: Make sure you select table tags before selecting the owner filters.

**CDPDSS-789: Adding system tag 'dp' prevents further editing of tags on the assets**

Problem: Adding system tag 'dp' prevents further editing of tags on the assets

Workaround: Choose not to add the tag 'dp' on any of the assets.

**CDPDSS-795: Search bar in schema tab goes empty, but previous results are shown**

Problem: Search results displays previous results.

Workaround: Refresh the page.

**CDPDSS-797: Schema search tab displays all results for not existing search string**

Problem: Search by terms which have no hits and returns all assets.

**CDPDSS-1055: Multiple tagging across different types throws 504 Gateway Timeout error**

Problem: 504 Bad Gateway error is displayed.

**CDPDSS-1075: When one specific asset is tagged to multiple tags from the search page, an incomplete alert is displayed**

Problem: Asset name display is inconsistent. Only the first tag is displayed, remaining tag messages do not display the asset name.

**CDPDSS-1084: Confusing terminologies.**

Problem: Different usage terminologies. Not consistent.

**CDPDSS-1373: "Unauthorised" message is displayed by middleware on the Cloudera Data Catalog UI for every API failing 401**

Problem: The Cloudera Data Catalog UI shows a pop-up at the bottom with message: "unauthorised request" and no context.

**CDPDSS-1375: Search enhancements by TERM**

Problem: In Cloudera Data Catalog, if the term is included in the asset search query, the count of the assets next to the Data Lakes displays NA. It is because, Atlas search with Glossary term is returns approximate count as -1.

**CDPDSS-1410: While navigating to another node from the Lineage tags filter, the tool-tip is retained.**

Problem: From the lineage tags filter, the tool-tip is retained.

**CDPDSS-1414: The "Cancel" button should exit the "Classifications" pane.**

Problem: The "Cancel" button does not get away from the the "Classification" pane.

**CDPDSS-1415: The Attribute name field in the "add classification" gets truncated when its very long and it cannot be read completely.**

Problem: The attribute name field cannot be read fully.

**CDPDSS-1416: Intermittent: Clicking add classifications reloads the lineage.**

Problem: Clicking "add classifications" reloads the lineage but takes a very long before completely loading it.

**CDPDSS-1474: [Glossary Terms] Difference in the Slider panel used in Asset Details and Search pages.**

Problem: Indifferent display of slider panels in **Asset Details** and search pages.

**CDPDSS-1475: Glossary Term | Allow selection on clicking any part of the row.**

Problem: While associating a term to column or entity, the user has to click on the + icon for selection and not just tap on any object within the row.

**CDPDSS-1484: Atlas and Ranger services UI link for Cloudera Base on premises clusters is broken.**

Problem: The managerUri attribute in Classic Cluster does not contain the Knox gateway URL.

**CDPDSS-1498: For AWS Glue assets, users cannot add terms.**

Problem: An error is displayed when Glue asset user tries to add terms.

**CDPDSS-1501: On applying the term name in the Glossary filter on the search page, the count for data lakes is rendered as NA**

Problem: The result count for data lakes displays are NA but the result list loads correctly.

**CDPDSS-1664: Inherited attributes are not getting displayed for tags in the Schema tab slider for specific user cases.**

Problem: Missing attributes in the Schema tab slider search bar.

**CDPDSS-1179: Cloudera Data Catalog search via Atlas's /v2/search/basic API doesn't retrieve results when the search text mentioned in the entity filter criteria (like searching by DB / table name) has special characters (like + - & | ! ( ) { } [ ] ^ " ~ \* ? :)**

Workaround: No known workaround via Cloudera Data Catalog.

**CDPDSS-1346: Asset Details view of HiveTable not displaying all columns**

Workaround: N/A

**CDPDSS-1276: Error is thrown when advanced configuration values are changed to 0 or left blank**

Problem: Error is thrown when advanced configuration values are changed to 0 or left blank and saved

**CDPDSS-1506: For specific databases, the terms associated to them are not getting deleted**

Workaround: N/A

**CDPDSS-1595: Cloudera Data Catalog Profilers are not functioning as expected for Ranger Authorization Service (RAZ) enabled Data Lakes**

Problem: When you are using the RAZ enabled clusters, you can launch the Cloudera Data Catalog profilers but the profilers do not function as expected.

Workaround: To enable RAZ with Azure, you must update the existing Ranger policy and also add a couple of new Ranger policies:

- Existing Policy: Default: Hive warehouse locations
- Add dpprofiler user and assign Read + List permission.

Later, add the following policies:

- Policy Name: Profiler locations (Anything that Ranger can accept)

Depending on your storage location, note the following and set them accordingly:

If you have a sample location like:

```
abfs://<Storage account>@<Storage account
container>.dfs.core.windows.net/<sub_directory>/dpprofiler
Storage account:<Storage account>
Storage account container : <Storage account Container>
Relative path: /*/dpprofiler
```

Select All permissions to allow conditions to take effect.

- Policy Name :Profiler Ranger Audit Logs (Anything that Ranger can accept)

Depending on your Storage location, note the following and set them accordingly:

If you have a sample location like:

```
abfs://<Storage
account>@<Storage account container>.dfs.core.windows.net/<sub_d
irectory>/ranger/audit
Storage account:<Storage account>
Storage account container : <Storage account Container>
Relative path: /*/ranger/audit
```

Select Read + List permissions for allow conditions.

Restart the Profiler Manager service in the Cloudera Data Hub Cloudera Manager instance.

#### **CDPDSS-1748: Release specific features call fails with 404 error on the search page**

Problem: Release specific features call is failing with 404 error for data lakes with no profiler cluster.

#### **CDPDSS-1774: In the edit dataset page, on selecting all the assets and deselecting a few does not work**

Problem: In the edit dataset page, on selecting all the assets and deselecting a few of the assets, deselects them in the UI but on hitting the 'delete' (trash icon), all the assets (selected and unselected ones) are deleted.

#### **CDPDSS-1801: Options drop-down for a dataset stays open even though drop-down is opened for another dataset**

Problem: The drop-down for all the datasets stay open.

#### **CDPDSS-1807: Overlapping b/w icon and inline text in text boxes**

Problem: This overlapping of icons and inline text is removed from all the pages.

#### **CDPDSS-1808: UI Issues with new 'add asset' modal to add assets to dataset N/A**

Problem: Directly clicking on 'search' field of 'Table Tag' shows no results found.

Typing "\_" in the search field erases the entire search string

'Clear' text button is not working for Table tags , however it works for column tags

#### **CDPDSS-1897: Show process nodes option is displayed for entity types which do not have process nodes**

Problem: The toggle option to show the "process nodes" is present even for entity types which do not have process nodes.

Workaround: N/A

#### **CDPDSS-1954: User cannot delete profilers when no entities are displayed in search result**

Problem: The "Action" button containing the option to delete profilers is also removed when the search results yield no entity type.

**CDPDSS-1988: Asset names are not clickable for registered assets in Glue Data Lake**

Problem: Once the asset is registered observe that the 'asset name' is in blue but not clickable.

**CDPDSS-2125: Unable to create profiler cluster for Cloudera Runtime version 7.2.11 and below****CDPDSS-2130: Unregistered Glue tables consume more time to load Asset Details page.****CDPDSS-2045: Duplicate API calls in Profiler job listing screen****CDPDSS-2047: Azure | Profilers HA | Issue with DB connection**

The issue can be solved by upgrading the data lake and the profiler cluster to Cloudera Public Cloud 7.2.14 or later.

**CDPDSS-2032: Profiler Manager Service call is failing for Classic Clusters****CDPDSS-2120: Test run on CSP tag creation fails with 404 error****CDPDSS-2140: System tags status is marked as 'suspended' on a medium duty cluster with Profiler Management Services on Cloudera Shared Data Experience**

Problem: The system tags are in a suspended state instead of default state which is 'deployed'.

**CDPDSS-1965: AWS S3 V2 Directory type nodes could not be clicked in the Asset Details page**

The issue was fixed by enabling AWS S3 V2 Directory type nodes in a configuration file.

**CDPDSS-2141: Delete profiler option should be available in Cloudera Data Catalog UI for all the states**

Problem: This functionality is present on CloudBreak UI and hence can be made available from Cloudera Data Catalog App as well.

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

Problem: 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-2925: The getNextScheduledRun API fails if createtime value is in the Allow List.**

The issue was fixed by updating the Cloudera Data Catalog Profiler configuration.

Workaround: Remove asset filter rules with createTime filter and recreate them.

**CDPDSS-2926: Profilers page displayed Start Time and Last Updated time in epoch time format**

The Profilers page was updated to use the UTC time format.

**CDPDSS-2928: Refreshing the Profiler page deletes configuration details**

The issue was fixed by updating the Profiler page.

**CDPDSS-2929: The Asset Details page reports a 404 error for impala\_process\_execution type assets.**

The issue was fixed by removing the unsupported lineage widget for impala\_process\_execution type assets.

**CDPDSS-2934: Classification and Glossary filters are not loaded on search page intermittently**

The issue was fixed by updating the entity search components.

**CDPDSS-2948: 400 Bad Request is generated in the Asset details page for Table Type "VIRTUAL\_VIEW"**

The issue was fixed by adding a tooltip and disabling the ODP run button for the empty location.

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

The Asset Details now correctly highlights links to Datasets.

Datasets / Details / Asset Details

item

Properties

Type: **HIVE TABLE**

# of Columns: 22

Data Lake: **hive**

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+0200...

Table Type: **EXTERNAL\_TABLE**

Database: **test\_dss\_db**

DB Catalog: **cm**

Parent: **test\_dss\_db**

Qualified Name

test\_dss\_db.item@cm

Comment

+ Add Comment

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

OverviewSchemaMetadata AuditsPolicyAccess Audits

Q 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						

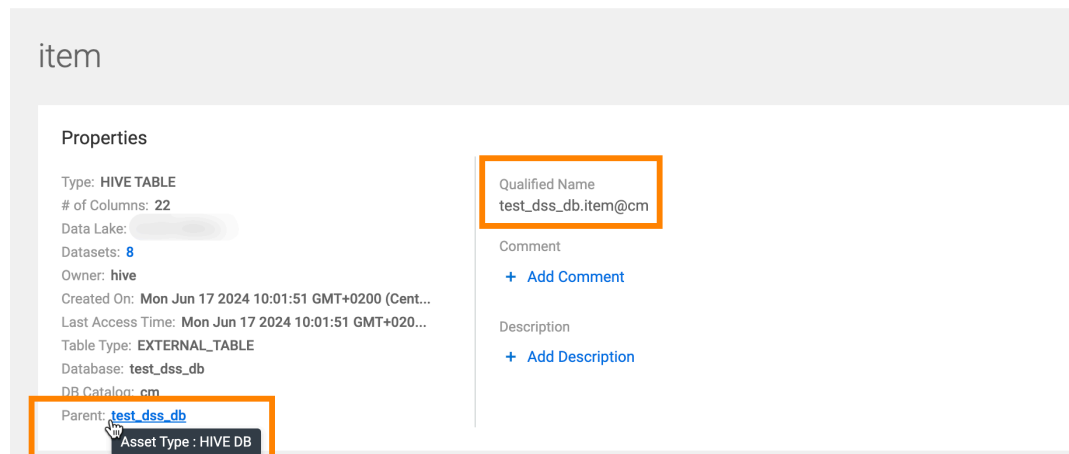
CDPDSS-2959: Asset Details does not provide a link to the parent database

- 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.

24



Datasets / Details / 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 Cloudera 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-3077: The Datasets Details page fails to load the list of assets**

Datasets containing deleted assets no longer fail to render the assets list with the following error: pages.dataset.labels.err-undefined-msg(Error Code - undefined)

**CDPDSS-3366: Hbase Table: The Asset Details page is distorted with very long description under Schema tab**

After the update, the description length no longer disrupts the **Schema** tab in **Asset Details**.

**CDPDSS-3319: No filter options for Access Type for HIVE DB type assets**

After the update, all filter options are correctly displayed for Hive DB type assets in **Asset Details** Access Audits .

**CDPDSS-3318: Access Audits doesn't have the filter for Access Type INSERT**

After the update, the INSERT filter option is correctly displayed for relevant assets in **Asset Details** Access Audits .

**CDPDSS-3064: Cancel and Suspend button is disabled on revert**

In **Profilers Tag Rules** , when trying to revert a suspended tag rule, the Cancel and Suspend buttons are no longer disabled.

**CDPDSS-3371: Asset Details is distorted with very long description**

After the update, the very long strings in the **Description** field are trimmed and no longer distort **Asset Details** by being displayed in full length.

**CDPDSS-3351: Unable to see comments on the Hive columns in Asset Details**

Full comments can be displayed in **Asset Details** while hovering over the **Comment** field compared to the earlier limited view.

**CDPDSS-3169: Asset Details Page - Optimize API to avoid 504 Error**

Before the update, by some tables with thousands of columns and one ALTER statement on each column to add comment, the `/api/atlas/v2/entity/guid/` API call took more than one minute and caused the timeout. Optimizing the API calls related to **Asset Details** led to improved performance and avoiding potential 504 errors when opening assets.

**CDPDSS-3073: API Status code is 200 even if add classification is failed because of wrong datatype for attribute**

When a classification is added to an asset with a mismatching attribute value type, the correct error message is shown, including the correct status code 400.

**CDPDSS-3061: Add assets is not displaying all the assets when searched based on Owner/DBName/Tags**

When adding assets to a dataset in **Dataset Details**, all assets will be correctly displayed when filtering the available assets by DB Name.

**CDPDSS-3031: Number of Jobs count is not matching in the Profilers/Jobs page**

The number of profiler jobs now correctly matches the between the **Filters** panel and list of profiler jobs in **ProfilerJobs** if the Rows per page: 500 setting is used.

**CDPDSS-3030: Rows displayed in the Profilers/Jobs page is not matching with the selected Rows per page value**

The list of profilers in **Profiler Jobs** correctly matches the number set in the **The Rows per page** setting.

**CDPDSS-3417: Incorrect results are being displayed after navigation, when "owners" filter is selected**

In the **Search** page, 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 no longer appear. Only entities belonging to the selected owners are displayed.

**CDPDSS-2571: Events filtered from Atlas that are not assets or objects**

The API call to Apache Atlas creating the default quick search response was streamlined by removing events that are not relevant in the Cloudera Data Catalog context, resulting in less clutter in the default view of the **Search** page. When searching for Hive or Iceberg columns based on Owner, Entity Tag or Glossary Terms, select the asset Type first.

**CDPDSS-3569: [Search Page ] Filters are retained when a filter is selected and then a different data lake is selected**

When switching data lakes on the **Search** page, the previously selected search filters are cleared after the update.

**CDPDSS-3396: Occasionally getting Bad Gateway(Error Code - 502) error in DC UI**

In a Compute Cluster enabled environments, the messages Error Occurred while processing: Bad Gateway is no longer occurring in the **Asset Details Overview** tab failing to display any data.

**CDPDSS-3576: No filter options for 'Access Type' of 'Type HIVE Column'**

The Access Type filter correctly displays all filters in **Asset Details Access Audits**

**CDPDSS-3579: Columns of different tables are shown when we have pagination in Schema tab**

Switching between entities requiring and not requiring pagination to display all their child entities no longer mixes up the listed child entities between the parent entities.

**CDPDSS-3478: The Ranger Audit Profiler is failing with ERROR "DataFrame" with objects without an attribute "limit"**

In Compute Cluster enabled environments, the Activity Profiler no longer fails when there are no logs to process in a dataframe.

**CDPDSS-3304: CRON jobs are not running in the time set in CRON Expression**

CRON jobs are running according to the profiler schedule settings, considering the local timezone.

**CDPDSS-3820: UI: Wrong Data Lake name is getting displayed in Asset Details page**

**Asset Details** displays the correct data lake.

**CDPDSS-3575: No link from iceberg column to iceberg table**

In **Asset Details**, the link from Iceberg column is provided to the parent Iceberg Table.

**CDPDSS-3578: Number of Assets is not matching with Datasets**

Now, **# of Assets** column in the **Add to Dataset** wizard in **Asset Details** is correctly matching with the number of assets in the datasets.

**CDPDSS-3935: UID is allowed to be sent with job start signal**

Profilers jobs are no longer moved to UNKNOWN status if they were started for a profiler pod that was stopped but sent a profiler job start signal in error.

**CDPDSS-3865: Switching from Basic Schedule to Cron Expression results in error**

Switching from the CRON based expression to the natural language based (Basic) schedule and moving back to the CRON based expression no longer results in an error for missing fields. Any changes made to the natural language based (Basic) schedule will get retained.

None

**CDPDSS-3870: Tag Rule Status is getting removed in the Edit Tag Rule wizard**

When editing an existing tag rule, the **Tag Rule Deployment Status** settings are no longer hidden in case the rule has been validated with a dry-run.


None

**CDPDSS-3880: Filter for associated tag does not return any data**

In **Profilers Data Compliance Profiler Tag Rules** the Associated Tag filter now shows the parent tags when selected.

None

**CDPDSS-3879: Removing the assets from Dry Run list takes two clicks**

In **Profilers Data Compliance Profiler Tag Rules Dry Run Test** the selected assets in **Test Connection with Catalog Data** can be removed by a single click from the field instead of clicking the  icon twice as before.

None

**CDPDSS-3862: Date range filter and pagination is not showing any results in Job History**

Applying the Time Range filter when the second page is selected in **Profilers Details Job History** shows results in **Job History** correctly.

None

**CDPDSS-3873: No validation for existing Tag Rules in Create/Edit Tag Rule flow**

Cloudera Data Catalog now checks if the **Tag Rule Name** was already used for a previously created tag rule. You will be prompted to use a unique tag rule name.

**CDPDSS-3877: Unable to search for classifications while creating Apache Atlas tags**

While creating a new Atlas tag for a tag rule for the Data Compliance Profiler in **Create Tag Rule Create New Tag**, the **Classification (optional)** field now offers the available classifications already present in Atlas for selection. The selected classifications will become the parent for the new tag.

None

**CDPDSS-3323: Profiler Jobs are running forever if the primary Kubernetes pod is shut down**

If the primary Kubernetes pod is forced shut down, or shuts down because of an error, the Kubernetes scheduler requests new resources for new pods. A CRON job runs every midnight (UTC) and reconciles such entries by checking for running pods and moving the jobs which would otherwise remain in the RUNNING state to an UNKNOWN state.

**CDPDSS-3945: Unable to create tag rules**

In VM-based environments, creating a tag, deleting a tag, updating a tag, and running a test no longer fails with error message: None.get .

**CDPDSS-3922: Searching assets in datasets with "\" giving error**

The **Dataset Details** page does not get stuck when trying to search terms containing the character \, however, the character \ is not supported in dataset search.

**CDPDSS-3921: Profilers page is not reflecting the status of the profiler**

The **Profilers** page no longer offers setting up new profilers while the previous ones are still being deleted.

**CDPDSS-3900: Setup Profiler is failing for Azure datalake**

Profiler setup is supported in Azure based environments.


**CDPDSS-3896: User should be able to close the "Select Instance" from the dropdown icon**

In **Profilers**, during the **Setup Cluster** process, users are now able to close the list of available instance types with the drop-down icon the instance search field.

**CDPDSS-3906: Search page pagination is not working as expected**

After changing the displayed page count and clearing the search filters, Cloudera Data Catalog shows the correct search results.

**CDPDSS-3926: Clicking on the cross icon also closes the instance list**

In **Setup Cluster**, clicking the  icon no longer closes the instance list, but only clears the instance search field.

**CDPDSS-3905: Parent tags are shown for classification search results**

When searching for tags or classifications in **Profilers Data Compliance Profiler Create Tag Rules Create Atlas Tag Classification (optional)**, tags are shown when trying to inherit attributes for the new child tag.

**CDPDSS-3834: Profiling of text format with custom delimiter is not supported**

Profiling now supports text files with custom delimiters by detecting and applying the correct delimiter during data loading.

**CDPDSS-4002: Pagination support for the list of profiled assets per job**

Pagination support is provided for the list of profiled assets for a particular profiling job in **Profilers Job History Profiled Assets**. The user interface no longer becomes delayed or unresponsive, if the list is too long.

**CDPDSS-3401: Profiler is shown as running in DSS App on deleting default compute cluster**

The profiler deletion process logic was updated for better error handling. Added logic to handle cases where the compute cluster is not in a running state or the node group is unavailable. In a Compute Cluster enabled environment, after deleting the default Compute Cluster, profilers are no longer displayed as still running in the **Dashboard**.

After recreating the Compute Cluster, delete and relaunch the profilers.

**CDPDSS-3364: Not able to delete the profilers from DSS App when underlying compute cluster deleted**

The profiler deletion process logic was updated for better error handling. When deleting the compute cluster in Compute Cluster enabled environments, profilers can be deleted now.

**CDPDSS-4057: All info (?) buttons incorrectly functioning as Save buttons**

In **Profiler Configuration**, the ? icons no longer work as Save buttons.

**CDPDSS-4142: Tag Pasting Not Working in Entity Tag Filter**

The search and event logic is update so that now, the **Entity Tag Filter in Search** correctly triggers a search when a tag is pasted into it.

**CDPDSS-4167: Unable to view anything in Asset360 Page - 400 error CDH - 500.68873829**

After the necessary Apache Atlas update, a Basic Search query sent to the 7.3.1.500 version of Atlas no longer results in a bad request response, resulting in an empty **Asset Details** page.

#### **CDPDSS-4084: [Hydrogen UI] Profiler jobs should not be case sensitive**

The search method for Profiler Details Job History is updated so profiler job names are normalized to lower case letter when searching.

#### **CDPDSS-4064: Cron job runs in UTC time-zone as multi-cron expressions are not supported**

Previously, when a Basic Cron schedule was set up in a way that it could not be described as a single CRON expression, like `* * 1 6 *` in a non-GMT timezone, which means "every minute on the 1st of June," the job will only run according to the UTC timezone. This happens because multiple cron expressions are not supported.

As a solution now, both Both the Basic and Cron Expression scheduler (Unix in Compute Cluster enabled environments cron jobs) will be created according to the UTC timezone instead of the local timezone of the user.

#### **CDPDSS-4063: Specific end of month dates are classified as invalid**

The following cron expressions are now supported in the profiler scheduler:

```
0 10 30 * *
0 10 29 * *
0 10 31 * *
```

Cloudera Data Catalog no longer returns the Please recheck the cron expression. Day of month must be a valid number message for the 29th, 30th and 31s of any month.

#### **CDPDSS-3815: Cron expression scheduled for the next years can lead to incorrect schedule**

A new validation method in Compute Cluster enabled environments ensures that Cron expression must run within the next 12 months, else an error message is produced.

#### **CDPDSS-3163: Datasets missing owners**

The updates fixes the issues where in **Dataset** the **Created by** and the **Author** fields remained empty when there was no data available. Now, Cloudera Data Catalog uses the value in the **user** field, trimming the domain if its an email.

#### **CDPDSS-4095: [Dataset] API calls are getting cancelled for dataset name**

The dataset search logic is improved by adding a timeout period. Searches are no longer stopped too soon.

#### **CDPDSS-4248: Unable to Edit Datasets in Data Catalog**

The UTC timestamp calculation for datasets is updated to enable editing dataset in all time zones.