

Managing Database Catalogs in CDW

Date published: 2020-08-17

Date modified: 2023-06-13



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

Rebuild Database Catalog..... 4
 Back up & restore Hue documents.....5

Add Database Catalog.....6

Rebuilding a Database Catalog

You can clean up resources and redeploy your Database Catalog using your existing image, or the latest image. You might want to rebuild your Database Catalog to get a feature in a later Database Catalog image, to perform housekeeping, or to troubleshoot a problem.

About this task



Note: The ability to rebuild a Database Catalog is in Technical Preview and is not recommended for production deployments. Cloudera recommends that you try this feature in test or development environments.

Upgrading your Database Catalog also rebuilds it. Rebuilding the Database Catalog redeploys resources in the cluster while keeping the configurations and other settings you made.

You can rebuild the Database Catalog in the following ways:

- Using the [Beta version DW CLI](#)
- Using the Rebuild button in the UI
- Using the Upgrade button in the UI

In this task, you rebuild the Database Catalog using the Rebuild button in the UI.

Rebuilding a Database Catalog has the following limitations:


- Rebuilding deprecated versions of the Database Catalog is not supported and will fail.
- Rebuilding does not preserve the changes you made directly, using `kubectl` for example, to cluster resources in the affected namespaces.

Changes you make to the Database Catalog using either the UI or CDP CLI are preserved.

Before you begin

- You must obtain the DWAdmin role.
- To preserve Hue documents, such as saved queries, you need to save a backup of the documents as described in *Backing up and restoring Hue documents*.

Procedure

1. Log in to the CDP web interface, navigate to **Data Warehouse Database Catalogs**, and select your Database Catalog.
2. Click .
3. Select **Rebuild**.

4. Rebuild the Database Catalog.

- If your Database Catalog does not have the latest image, you see the version number of your current image and can select Keep current image version, and then click Rebuild Database Catalog.



- Alternatively, you can accept the default value for Keep current image version, which is off, and then click Rebuild Database Catalog. The Database Catalog will be rebuilt using the latest image.



- If your Database Catalog has the latest image, click Rebuild Database Catalog.

5. Restore Hue documents as described *Backing up and restoring Hue documents*.

Related Information

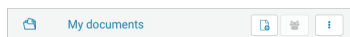
[Backing up and restoring Hue documents](#)

Backing up and restoring Hue documents

You can back up and restore Hue data, such as a Hive or Impala query you saved in a JSON document, in the Hue database to prevent losing the documents after rebuilding the Database Catalog.

About this task

In Hue, you can view the JSON documents you created or imported in My documents.



Procedure

- Backup Hue documents you want to preserve from the command line.

```
# on hue pod
./build/env/bin/hue dumpdata -o /tmp/data.json
# on local machine
kubectl --kubeconfig ~/kconfig cp compute-1675582336-knmf/huebackend-0:/
tmp/data.json -c hue ./data.json
```

- Rebuild the Database Catalog as described in the previous topic.
- Restore the Hue documents in the rebuilt Database Catalog.

```
#on local machine
kubectl --kubeconfig ~/kconfig cp ./data.json compute-1675582336-knmf/huebackend-0:/tmp/data.json -c hue

# on new hue pod
./build/env/bin/hue loaddata --ignorenonexistent /tmp/data.json
```

Adding a new Database Catalog

A Database Catalog is automatically created when you activate an environment in Cloudera Data Warehouse (CDW). You can add additional Database Catalogs if you want a standalone data warehouse that is not shared with other authorized users of the environment.

About this task

When you activate an environment from the Data Warehouse, a default Database Catalog is created and named after your environment. This HMS instance associated with the default Database Catalog is the same HMS as the one used by your CDP environment. You can add additional Database Catalogs if you want standalone data warehouses based on a new HMS instance. When you create a new Database Catalog, you specify which environment to use. If you make a change to the default database catalog, the change is reflected in the environment where the default Database Catalog resides.



Note: Cloudera recommends that you use the default Database Catalog in production environments. While CDW enables you to create custom Database Catalogs, these are not supported by Cloudera.

You can optionally load demo data in Hue when you create a new Database Catalog.

Before you begin

You need to obtain the DWAdmin role.

You must also enable the Create multiple Database Catalogs option on the [Advanced Configurations Advanced Settings](#) page.

Procedure

1. Log in to the CDP web interface, navigate to [Data Warehouse Database Catalogs Add New](#).
2. In Name, specify a Database Catalog name.



Note: CDW uses deterministic namespace and adds a prefix to the Database Catalog name. The length of the namespace ID after CDW applies a prefix to the Database Catalog name, including the hyphen (-), should not exceed 63 characters. You can specify the Database Catalog name 53 characters long.

3. In Environments, select the name of an activated environment.
4. For example purposes, turn on Load Demo Data to use sample airline data in Hue.
5. Click Create to create the new Database Catalog.