

..

CDW Top Tasks

Date published: 2022-08-15

Date modified:

CLOUDBERA

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

Activate environments.....4

Explore Data Lake.....5

Visualize data..... 6

Activating an environment in CDW Private Cloud

An environment is a logical entity that represents the association of your Private Cloud user account with the compute resources. After you register your OpenShift or Embedded Container Service (ECS) environment with CDP, activate the environment in Cloudera Data Warehouse (CDW) to set up a Database Catalog and create Virtual Warehouses. Activating an environment causes CDP to connect to the Kubernetes cluster, which provides the computing resources for the Database Catalog. This is the first step in setting up CDW.

Activating OpenShift environments



Note: If you have more than one OCP clusters managed using different instances of Cloudera Manager, but using the same AD server and using the same environment name, then go to the [Advanced Configuration Advanced Settings](#) page and disable the [Use deterministic namespace names](#) option before activating the environment in CDW.

1. Log in to the Data Warehouse service as an DWAdmin.
2. Expand the Environments column by clicking More... and locate the environment that you want to activate.
3. Click on the activate icon.
4. Specify the storage class name, delegation username and password, and the authentication mode to use for connecting to Hive Virtual Warehouses and click **ACTIVATE**.

For local caching, ensure that the platform administrator uses the Local Storage Operator to create a local file system on an SSD/NVMe for each OpenShift worker node and then mounts it to a known location on the worker node. Make sure that this local caching location allows temporary data to be stored in a way that supports performance.

The delegation user and password can authenticate users through an LDAP service account.

Activating ECS environments



Note: A “default” environment is created by the Control Plane service when you add a Private Cloud cluster. If you have more than one ECS clusters managed using different instances of Cloudera manager, but using the same Active Directory (AD) server and using the same “default” environment, then go to the [Advanced Configuration Advanced Settings](#) page and disable the [Use deterministic namespace names](#) option before activating the environment in CDW.

1. Log in to the Data Warehouse service as an DWAdmin.
2. Expand the Environments column by clicking More... and locate the environment that you want to activate.
3. Click on the activate icon.
4. Specify the delegation username and password to impersonate authorization requests from Hue to the Impala engine, and then click **ACTIVATE**.

The delegation user and password can authenticate users through an LDAP service account.

Next steps

A default database Catalog is automatically created when you activate an environment in CDW. After the Database Catalog is created, you can create Hive, Impala, or Impala Virtual Warehouses with the Unified Analytics mode.

If Analysts in your organization use JDBC or ODBC SQL clients, then you can download the drivers from the CDW UI and connect them to the Virtual Warehouses. Analysts can also use Hue, the built-in SQL editor to query and explore data.

You can also link to Cloudera Data Visualization instance from the CDW UI which line of business users can use to visualize data and gain insights.

Exploring a Data Lake

In Cloudera Data Warehouse (CDW) Private Cloud, a default Database Catalog is linked to the Hive MetaStore (HMS) on the CDP Private Cloud Base cluster. As soon as you activate an environment in CDW, you can explore and query databases and tables that are present on your base cluster. If you are new to CDP, then you can also import data using Hue and run queries.

Before you begin

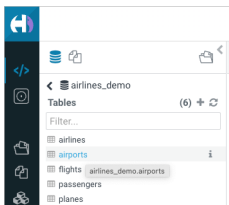
- You must have the DWAdmin role.
- You must have activated an environment in CDW.
- You must have a default Database Catalog running.

About this task

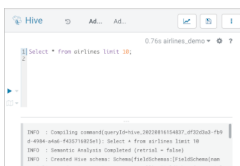
You set up a Virtual Warehouse with the minimum, default configurations for learning how to explore a Data Lake. You do not need to configure auto-scaling and other features to explore a Data Lake. Ensure you delete the Virtual Warehouse after your exploration.

Procedure

1. Log in to the Data Warehouse service as DWAdmin.
2. Click Virtual WarehousesAdd New.
3. Set up the Virtual Warehouse:
 - Specify a Name for the Virtual Warehouse.
 - In Type, click the SQL engine you prefer: Hive or Impala.
 - Select your Database Catalog and User Group if you have been assigned a user group.
 - In Size, select the number of executors, for example xsmall-2Executors.
 - Accept default values for other settings.
4. Click Create.
5. After your Virtual Warehouse starts running, click Hue, and expand Tables to explore available data.



6. Explore data lake contents by running queries. For example, select all data from the airlines table.



Visualizing data using Cloudera Data Visualization

Data Visualization is a dashboarding and application building service that enables you to explore data through the common language of visuals. You create a Cloudera Data Visualization (CDV) instance from the Cloudera Data Warehouse (CDW) UI, link your data sets to CDV, and build visuals.

Before you begin

Create a CDV instance from the CDW UI by going to [Data Visualization Create](#) . Specify a name, select an environment, select user groups and admin groups, and select a size for the CDV instance.

New Data Visualization X

Name * ⓘ

Environments *

User Groups ⓘ

You can select groups present in the CDP user management system. The group(s) must also exist in the external LDAP Identity Provider. Nested groups are not supported. Only users that are direct members of the group are allowed access.

Admin Groups * ⓘ

You can select groups present in the CDP user management system. The group(s) must also exist in the external LDAP Identity Provider. Nested groups are not supported. Only users that are direct members of the group are allowed access.

Size *

[Create](#)

Procedure

1. Create a CDW data connection as described in [Creating a CDW data connection in Data Visualization](#).
2. Create a visual as described in [Creating a visual](#).