

Getting Started with Public Cloud

Date published: 2024-01-01

Date modified: 2024-01-01

CLOUDERA

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

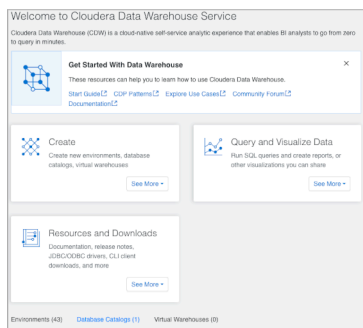
- Getting started in CDW.....4**
 - Prerequisites..... 5
 - Granting CDP users and groups access to CDW 5
 - Creating a Database Catalog..... 6
 - Creating your first Virtual Warehouse..... 7

Getting started in CDW

In Cloudera Data Warehouse (CDW), you can deploy and administer cloud data warehousing. You can move workloads to the cloud securely and manage user access to warehouse data, or just parts of the data, through Apache Ranger fine-grained control. You learn how to get started in CDW.

As CDW administrator, you need privileges to access a CDP environment. As a CDW user, you need privileges to access warehouse data as described in the subtopics, "Prerequisites" and "Granting CDP users and groups access" below.

To navigate to Overview, log into Cloudera Data Platform, and click Data Warehouse:



If you have permission to access an environment, you can create a Virtual Warehouse:

- A Hive Virtual Warehouse for Hive users
- An Impala Virtual Warehouse for Impala users
- Unified Analytics-enabled Virtual Warehouse.

Unified Analytics offers Hive/Impala-equivalent SQL syntax for many SQL queries, plus optimizations and enhancements.

You provide users with the URL of your Virtual Warehouse, and they can access the data you authorize using tools you download from Resources and Downloads and provide to them. From Resources and Downloads, you can install or download the following software or drivers:

- DBT Hive and DBT Impala

Adapters for [using dbt](#) with Hive or Impala.

- CDP CLI

Install a tool to help you manage your Service Manager-managed cluster instances. The CLI can be used for automating cluster creation and termination.

- Hive JDBC Jar

Download the Hive JDBC driver for clients to connect to a Virtual Warehouse.

- Beeline CLI

Download the Beeline CLI tarball for clients to install and connect to a Virtual Warehouse.

- Workload Insights

Download this software for getting recommendations for creating materialized views.


- Impala JDBC/ODBC Driver

Download the Impala JDBC/ODBC driver for clients to connect to a Virtual Warehouse.

- Unified Analytics JDBC Driver

Download a JDBC driver for clients to connect to a Unified Analytics Virtual Warehouse.

Exiting CDW

In the left navigation, click .



Select Management Console to leave CDW to manage environments, users, and CDP services, or select a different menu item.

Prerequisites

To get started in Cloudera Data Warehouse (CDW), your data must conform to supported compression codecs, and you must obtain CDP resource roles to grant users access to an environment. Users can then get started on CDW tasks, such as activating the environment from CDW.

Unsupported compression

CDW does not support LZO compression due to licensing of the LZO library. You cannot query tables having LZO compression in Virtual Warehouses, which use CDW Impala or Hive LLAP engines.

CDP resource roles

You need to have the PowerUser role to change the following CDP resource roles that are associated with the CDW service.

- **DWAdmin:** This role enables users or groups to grant a CDP user or group the ability to activate, terminate, launch, stop, or update services in Virtual Warehouses.
- **DWUser:** This role enables users or groups to view and use CDW clusters (Virtual Warehouses) that are associated with specific environments.

A CDP PowerUser assigns these roles to users who require access to the Virtual Warehouses that are associated with specific environments. A default Database Catalog is associated with your environment. After granting these roles to users and groups, they then have Single Sign-On (SSO) access to the Virtual Warehouses.

Granting CDP users and groups access to CDW

You assign the DWAdmin and DWUser roles to users and groups who must use or manage Virtual Warehouses in Cloudera Data Warehouse (CDW) service on public cloud environments. These roles are not for accessing Kubernetes endpoints in the cluster.

Procedure

1. Log in to the CDP web UI using your corporate credentials or other credentials that your administrator gave to you.
2. In the left navigation menu, click Environments.
3. Search for the environment you want to grant access to, and then click the environment name.
4. In the environment's Clusters page, click the Actions drop-down menu on the right, and select Manage Access.
5. In the Access page, in the Select group or user text box, type the name of the user or group you want to add, and then click that user or group's name.
6. In Update Resource Roles for..., select the DWAdmin or DWUser or both roles, and then click Update Roles. You should receive a success message.
7. Click the environment name at the top of the page to navigate back to the Clusters page.

8. In the Clusters page, click Actions, and select Synchronize Users to FreeIPA.

Depending on how many users have access to the environment, this synchronization process can take a few seconds or a few minutes.

Results

The users and groups you have granted the DWAdmin or DWUser resource roles to can now use CDW service with the environment.

Creating a Database Catalog

When you activate an environment in CDW, a default Database Catalog is created. You can create a new Database Catalog from the Data Warehouse UI.

About this task

You can configure the following sizes of the Java Heap for your Database Catalog workload:

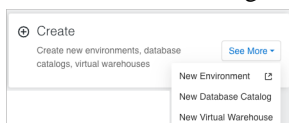
- Small (default), 8Gb
- Medium, 16Gb
- Large, 24Gb

The size you configure for the Database Catalog determines the container size of Hive Metastore (HMS) for storing your workload metadata. To avoid unnecessary cloud expenses, do not increase the size unless you experience Java heap issues.

Given Ranger permissions, you can access any objects or data sets created in the Data Hub or the Data Engineering clusters from CDW Virtual Warehouses and vice versa. The CDW service sets up the Kubernetes cluster, which provides the computing resources for the Database Catalog. The CDW service uses the existing data lake that was set up for the environment, including all data, metadata, and security. The following procedure shows you steps to create a Database Catalog to replace your default Database Catalog.

Procedure

1. Navigate to Cloudera Data Warehouse Overview.
2. Click Database Catalogs, in Create, click See More New Database Catalog .



3. In Name, specify a Database Catalog name.
4. In Environments, select the name of your environment.

If you do not see the environment you want in the drop-down list, you might need to activate the environment.

5. In Select Size, accept the default (Small), or select Medium or Large.

6. In Database Catalog Image Version, select the [Hive version](#) for constructing the Hive Metastore.
7. Optionally turn on Load Demo Data to use sample airline data in Hue.
8. Click Create Database Catalog.

Creating your first Virtual Warehouse

After meeting prerequisites, you learn how to activate the environment from Cloudera Data Warehouse (CDW). You then create a Virtual Warehouse.

Before you begin

Before you can create a Virtual Warehouse in CDW service, you must have completed the following tasks:

- Register an environment with Management Console. This sets up your Data Lake. Depending on which cloud platform you are using, see [Register an AWS environment](#) or [Register an Azure environment](#).
- Grant the DWAdmin role to the user or group that needs to create a Virtual Warehouse. This user or group determines which environment and Data Lake your Virtual Warehouse uses. See [Granting CDP users access to Cloudera Data Warehouse service](#).

Procedure

1. Activate the environment that uses the Data Lake. See [Activating AWS environments](#) or [Activating Azure environments](#).
A default Database Catalog is created.
2. Add a new Virtual Warehouse. See [Creating a Virtual Warehouse](#).

Results

You can start executing workloads in the new Virtual Warehouse.