Deploying shared Hue service in Data Warehouse Public Cloud (Preview)

Date published: 2023-11-20
Date modified: 2024-07-26

Legal Notice

Contents

About deploying the shared Hue service

Cloudera Data Warehouse (CDW) allows you to deploy a shared Hue service at an environment level. Learn about the advantages and limitations of deploying a shared Hue service and some FAQs that can help you understand more about the feature.

Note: This feature is in technical preview and not recommended for production deployments. Cloudera recommends that you try this feature in test and development environments.

Advantages of deploying a shared Hue service

By deploying a shared Hue service, you can manage costs by keeping only those Virtual Warehouses running which your users need at that time. Data Analysts only need to know or bookmark one Hue instance URL and can run queries on any Virtual Warehouses available to them.

Each shared Hue service instance has its own database where queries and query history is saved. You can create isolation for different teams within your organization by deploying multiple shared Hue instances.

Moreover, the shared Hue service remains active as long as the environment is active.

Limitations

- When you use the Importer to create tables from files in Hue, by default, Hue creates a Hive table if Hive is available and uses the first Hive Virtual Warehouse that was created. You cannot select a Virtual Warehouse using which you want to create a table by importing a file.
  To create an Impala table using the Importer, you must first select the editor type as Impala, and then click **+** on the **Table Browser**.

Access control for the shared Hue service

You can specify user groups you created in the CDP Management Console, similar to how you specify while creating the Virtual Warehouses.

When you specify user groups while creating the shared Hue instance or Virtual Warehouses, the subset of users who have access to the Hue instance as well as the Virtual Warehouse can submit queries through that Virtual Warehouse instance.

If you do not specify user groups while creating a shared Hue instance, then all users within your organization can access the Hue UI. If you do not specify user groups for a Virtual Warehouse, all users within your organization can submit queries through that Virtual Warehouse.

As a best practice, specify user groups while creating Hue and Virtual Warehouse instances so that specific users have access to specific compute resources.

Key differences in database management approach for Virtual Warehouse-level Hue and shared Hue service

All Hue instances linked to a Database Catalog through Virtual Warehouses within a CDW environment share a single database. The Hue database is not deleted unless you deactivate the environment. If you delete a Virtual Warehouse and create a new one, the Hue instance linked to that Virtual Warehouse continues to display old query history and saved queries.

Each shared Hue service instance has its Hue database. CDW does not delete the Hue database when you delete the shared Hue service instance. The Hue database exists in the backend until a database administrator manually deletes it. Each Hue database is named after the shared Hue service name. In case you have deleted a shared Hue service instance, you can reuse the Hue database by specifying the name of the Hue instance you deleted. This brings back the query history and saved queries.

CDW provides you a one-time option to copy the Hue database content from the Hue database linked to a Database Catalog to the shared Hue service database while creating a new shared Hue service instance. The data between the two databases is not synchronized after the initial copy event.

Enabling the shared Hue service

By deploying a shared Hue service, you can retain the Hue instance with its query history and saved queries even if the Virtual Warehouses are deleted. You can enable this option by editing the environment configuration.

Steps

1. Log in to the Data Warehouse service as a DWAdmin.
2. Go to the **Environment(s)** tab, locate the environment in which you want to enable this feature, and click the More Options icon > **Edit**.

3. Go to the **CONFIGURATIONS** tab, select the **Enable Shared Hue Service** option, and click **Apply Changes**.
   A new **Shared Hue Service** option is added on the left navigation pane on the CDW UI.

Creating a shared Hue instance

After you enable the option to deploy Hue at the environment level, you can create any number of Hue instances. Each Hue instance has its own database. The Hue instances deployed at the environment level do not share query history or saved queries.

Steps

1. Log in to the Data Warehouse service as a DWAdmin.
2. Click the **Shared Hue Service** option from the left navigation pane of the CDW UI.
3. Create a shared Hue instance by clicking **ADD NEW** on the **Shared Hue Service** page.
   The **Create Shared Hue Service** modal is displayed.
4. Specify a name for your Hue instance and select an environment from the drop-down menu.
   a. Select a size for the shared Hue instance from the **Size** drop-down menu.
      The default value is **1 x Hue Node**. This indicates the number of Hue backend pods you want to create.
   b. Select one of the following options from the **Select the Hue database initialization strategy** drop-down menu to initialize a database for this Hue instance.
      i. **Reuse if Hue data is present**: Select this option to use an existing Hue database for your new shared Hue instance.
      ii. **Copy Virtual Warehouse database**: Select this option to copy the contents of the Hue database within a Database Catalog into the shared Hue service database. This helps you to copy the shared queries and query history.
      Note: This is a one-time copy operation. After the data is copied from the Database Catalog to the shared Hue database, the data between the two databases is not synchronized.
      iii. **Flush the current database content**: Select this option to create a new database for the shared Hue instance.
      Select this option to recreate a shared Hue service instance with the same name, but a fresh database. The existing database linked to the Hue instance is dropped.

   You can optionally specify user groups and tags in the same way you specify while creating a Virtual Warehouse for fine-grained access control.
5. Click **Create**.
   A new shared Hue service instance is created.

After you finish

Click **Editor** to open Hue.
You can select the Virtual Warehouse you want to use from the **Virtual Warehouse** drop-down menu on the Hue web interface.

Rebuilding a shared Hue service

The rebuild operation deletes and recreates the Hue pods while preserving Hue's image version and configurations. By rebuilding the Hue service, you fix pods that are in a bad state, thereby improving performance.

Steps

1. Log in to the Data Warehouse service as DWAdmin.
2. Click **Shared Hue Service** from the left navigation pane.
3. Locate the shared Hue service instance you want to rebuild and click the More icon > **Rebuild**.
4. Review the message on the **Review Shared Hue Service** modal and click **Review Shared Hue Service**.
   The "Rebuild in progress" message is displayed.

Upgrading a shared Hue instance

If you are on an older version of the shared Hue service, you can upgrade the Hue image version by upgrading Hue from the Shared Hue Service page in Cloudera Data Warehouse.

Steps

1. Log in to the Data Warehouse service as DWAdmin.
2. Click **Shared Hue Service** from the left menu.
3. Locate the shared Hue service instance you want to upgrade and click **Upgrade**.
   The **Upgrade Shared Hue Service** modal is displayed.
4. Click **Upgrade**.
   The shared Hue service is upgraded to the latest available image version.

FAQs

Can I still use Hue which is deployed at the Virtual Warehouse level?

Yes, you can continue to access and use Hue from a particular Virtual Warehouse even after deploying the shared Hue service at the environment level.

Can I create more than one shared Hue service instance?

Yes, you can create any number of shared Hue service instances. However, Cloudera recommends deploying a single Hue unless isolating saved queries is a requirement. When you create multiple shared Hue instances, each instance has its own database. The shared Hue service instances do not share query history or saved queries.

Can I view queries submitted from other BI tools?

Hue superusers and administrators can view all queries submitted from all Virtual Warehouses linked to a Database Catalog. Other logged-in users can view only their queries on the **Impala Queries** and **Hive Queries** tabs.

Where can I specify advanced Hue configurations (safety valve) for the shared Hue instance?

On the **Shared Hue Service** page, click the more options icon > **Edit** corresponding to the Hue instance that you want to configure, go to the **CONFIGURATIONS** tab, and select **hue-safety-valve** from the **Configuration files** drop-down menu.