

Getting Started with on cloud

Date published: 2024-01-01

Date modified: 2024-12-05

CLOUDERA

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

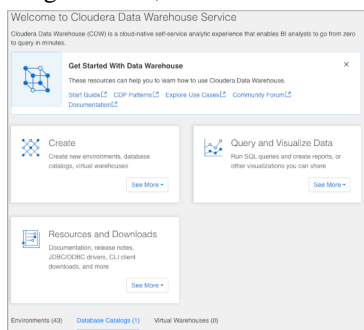
Getting started in Cloudera Data Warehouse.....	4
Prerequisites.....	5
Granting Cloudera users and groups access to Cloudera Data Warehouse.....	5
Creating your first Virtual Warehouse.....	6
Set up SQL AI Assistant.....	6
Prerequisites for configuring Hue SQL AI Assistant.....	7
Configure SQL AI Assistant using Cloudera AI Workbench.....	8
Configure SQL AI Assistant using the Cloudera AI Inference service.....	9
Configure SQL AI Assistant using the Microsoft Azure OpenAI service.....	9
Configure SQL AI Assistant using the Amazon Bedrock Service.....	10
Configure SQL AI Assistant using the OpenAI platform.....	10
Complete list of model-related configurations for setting up the Hue SQL AI Assistant.....	11
Hue SQL AI Assistant FAQ.....	13
About deploying the shared Hue service.....	14
Access control for the shared Hue service.....	15
Creating a shared Hue instance.....	15
Rebuilding a shared Hue service.....	16
Upgrading a shared Hue instance.....	16
FAQ for shared Hue service.....	16

Getting started in Cloudera Data Warehouse

In Cloudera Data Warehouse, 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 Cloudera Data Warehouse.

As Cloudera Data Warehouse administrator, you need privileges to access a Cloudera environment. As a Cloudera Data Warehouse user, you need privileges to access warehouse data as described in the subtopics, *Prerequisites* and *Granting Cloudera users and groups access*.

To get started, click Data Warehouse on the Cloudera home page. The **Overview** page is displayed:



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

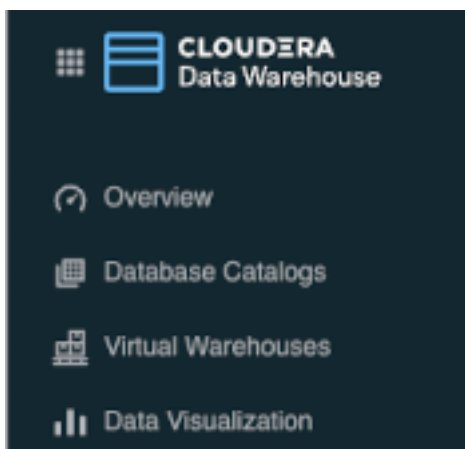
Download the Impala JDBC/ODBC driver for clients to connect to a Virtual Warehouse.

- Unified Analytics JDBC Driver

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

Exiting Cloudera Data Warehouse

In the left navigation, click  as shown in the following image:



Select Management Console to leave Cloudera Data Warehouse to manage environments, users, and Cloudera services, or select a different menu item.

Prerequisites

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

Unsupported compression

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

Cloudera resource roles

You need to have the PowerUser role to change the following Cloudera resource roles that are associated with the Cloudera Data Warehouse service.

- DWAdmin: This role enables users or groups to grant a Cloudera 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 Cloudera Data Warehouse clusters (Virtual Warehouses) that are associated with specific environments.

A Cloudera 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 Cloudera users and groups access to Cloudera Data Warehouse

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

Procedure

1. Log in to the Cloudera web interface.
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 Cloudera Data Warehouse service with the environment.

Creating your first Virtual Warehouse

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

Before you begin

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

- Register an environment with Cloudera 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 Cloudera 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.

About setting up the Hue SQL AI Assistant

Administrators are required to set up and enable the SQL AI Assistant before analysts can use it to generate, edit, optimize, explain, and fix queries using natural language in Hue.

First, you must obtain clearance from your organization's infosec team to ensure it is safe to use the SQL AI Assistant because some of the table metadata and data, as mentioned in the previous section, is shared with the LLM.

Next, select and prepare one of the following AI services of your choice for hosting an LLM, and then configure the SQL AI Assistant in Hue:

- Cloudera AI Workbench
- Cloudera AI Inference service
- Microsoft Azure OpenAI service
- Amazon Bedrock service
- OpenAI platform

Prerequisites for configuring Hue SQL AI Assistant

To configure the SQL AI Assistant in Hue, you must pass the token required for connecting to the LLM service. Learn about the open and secure approaches to pass the tokens, and use the one that fits your organization policy.

(Recommended) Secure approach for passing a token

In this approach, you use Kubernetes' method of distributing secrets. You first encode the credentials and then add the encoded bit as a data item in the HUE_AI_INTERFACE_TOKEN property. The token becomes available in the Hue pod as an environment variable.

About this task



Note: Secrets are lost when you rebuild the Virtual Warehouse. You need to redo this step to continue using encoded credentials.

Procedure

1. Use a base64 encoding tool to convert your token to a base-64 representation by running the following command:

```
echo -n '[***MY-TOKEN***]' | base64
```

Replace `[***MY-TOKEN***]` with the token value you want to encode.

2. Open a terminal session and run the following command to add the encoded secret:

```
kubectl edit secret hue-secret -n [***VIRTUAL-WAREHOUSE-NAMESPACE***]
```

Replace `[***VIRTUAL-WAREHOUSE-NAMESPACE***]` with the actual Virtual Warehouse ID (same as the namespace) in which you want to add the secret.

3. Add the encoded value returned for your token in the HUE_AI_INTERFACE_TOKEN property as follows:

```
...
apiVersion: v1
data:
  HADOOP_CREDSTORE_PASSWORD: [***ENCODED-HADOOP-CREDSTORE-PASSWORD***]
  HUE_AI_INTERFACE_TOKEN: [***ENCODED-TOKEN-VALUE***]
kind: Secret
```

Replace `[***ENCODED-TOKEN-VALUE***]` with the actual encoded value returned for your token.

Open approach for passing a token

In this approach, you specify the token value in the hue-safety-valve field in Cloudera Data Warehouse. The credentials are saved in a configuration file in the plain text format.



Note: Cloudera recommends that you use the open approach to pass tokens in test deployments, for proof of concept use cases. Use the [Secure approach for passing a token](#) in production deployments.

Here's a list of the open token values in the hue-safety-valve field to configure the SQL AI Assistant:

For Open token

Microsoft Azure OpenAI

```
[desktop]
[[ai_interface]]
  service='azure'
  model_name='[***DEPLOYMENT-NAME***]'
  base_url="https://[***RESOURCE***].openai.azure.com/"
  token="[***RESOURCE-KEY***]"
```

AWS

```
[aws]
[[bedrock_account]]
  access_key_id='[***ACCESS-KEY***]'
  secret_access_key='[***SECRET-KEY***]'
  region='us-east-1'
[desktop]
[[ai_interface]]
  service='bedrock'
  model='claude'
```

OpenAI

```
[desktop]
[[ai_interface]]
  service='openai'
  token='[***API-KEY***]'
```


Configure SQL AI Assistant using Cloudera AI Workbench

This topic describes how to deploy and configure the SQL AI Assistant using the Cloudera AI Workbench. With the added support for Cloudera AI Workbench, you can securely deploy and run your own models within a virtual private cloud. This self-contained integration offers enhanced control and privacy within your environment.

Before you begin

To know more about creating and deploying models using Cloudera AI Workbench, see [Create and deploy the model](#).

Procedure

1. Upon successful completion of model deployment, log in to the Cloudera Data Warehouse service.
2. Go to the Virtual Warehouses tab, locate the Virtual Warehouse on which you want to enable this feature, and click  Edit .
3. Go to Configurations Hue , select hue-safety-valve from the Configuration files drop-down menu, and add the following lines :

```
[desktop]
[[ai_interface]]
  service='cml'
  model='llama'
  model_ref='[***Place model access key here***]'
  base_url='https://[***RESOURCE***].cloudera.site/model'
```

4. Click Apply Changes.

Results

You see ✨ Assistant on the Hue SQL editor, where the SQL AI Assistant utilizes the model hosted in the Cloudera AI Workbench.


Configure SQL AI Assistant using the Cloudera AI Inference service

This topic describes configuring the SQL AI Assistant using the Cloudera AI Inference service.

Before you begin

To know more about installing and setting up the Cloudera AI Inference service, see [Prerequisites for setting up the Cloudera AI Inference service](#).

Procedure

1. Upon installing and setting up the Cloudera AI Inference service, log in to the Cloudera Data Warehouse service.
2. Go to the Virtual Warehouses tab, locate the Virtual Warehouse on which you want to enable this feature, and click  Edit .
3. Go to Configurations Hue , select hue-safety-valve from the Configuration files drop-down menu, and add the following lines :

```
[[ai_interface]]
  service='caii'
  model_name='[***Place MODEL name here***] '
  base_url="https://[***RESOURCE***]/v1"
```

4. Click Apply Changes.

Results

You see ✨ Assistant on the Hue SQL editor, where the SQL AI Assistant utilizes the model hosted in Cloudera AI Inference service.


Configure SQL AI Assistant using the Microsoft Azure OpenAI service

Microsoft Azure allows for dedicated deployments of OpenAI GPT models. You can use Azure's OpenAI service instead of the publicly hosted OpenAI APIs, as it enables data processing within your Azure Virtual Network (VNet) network. GPT models can also be integrated with the Hue SQL AI Assistant using Azure's OpenAI service.

Before you begin

Obtain a Microsoft Azure subscription by working with your organization's IT team and registering for access to the Azure OpenAI service. For more information, see [Create and deploy an Azure OpenAI Service resource](#).

Procedure

1. Log in to the Cloudera Data Warehouse service
2. Go to the Virtual Warehouses tab, locate the Virtual Warehouse on which you want to enable this feature, and click  Edit .
3. Go to Configurations Hue , select hue-safety-valve from the Configuration files drop-down menu, and add the following lines :

```
[desktop]
[[ai_interface]]
  service='azure'
```

```
model_name=' [***DEPLOYMENT-NAME***] '
base_url="https://[***RESOURCE***].openai.azure.com/"
```

4. Click Apply Changes.

Results

You see ✨ Assistant on the Hue SQL editor, and the SQL AI Assistant will connect to the specified model on the Microsoft Azure OpenAI service.


Configure SQL AI Assistant using the Amazon Bedrock Service

This topic describes how to configure the SQL AI Assistant using the Amazon Bedrock Service.

Before you begin

You must have an AWS account with Bedrock access. For more information on accessing keys, see [Amazon Bedrock](#).

Procedure

1. Log in to the Cloudera Data Warehouse service.
2. Go to the Virtual Warehouses tab, locate the Virtual Warehouse on which you want to enable this feature, and click  Edit .
3. Go to Configurations Hue , select hue-safety-valve from the Configuration files drop-down menu, and add the following lines :

```
[aws]
[[bedrock_account]]
  access_key_id_script='echo $AWS_BEDROCK_ACCESS_KEY_ID'
  secret_access_key_script='echo $AWS_BEDROCK_SECRET_ACCESS_KEY'
  region='us-east-1'
[desktop]
[[ai_interface]]
  service='bedrock'
  model='claude'
```

AWS_BEDROCK_ACCESS_KEY_ID and AWS_BEDROCK_SECRET_ACCESS_KEY must be added as encoded values under hue-secret. For more information, see [Secure approach for passing a token](#).

4. Click Apply Changes.

Results

You see ✨ Assistant on the Hue SQL editor, and the SQL AI Assistant will connect to the specified model in the Amazon Bedrock service.

Configure SQL AI Assistant using the OpenAI platform


This topic describes how to set up SQL AI Assistant and connect to a model on the OpenAI platform.

Before you begin

You must have created an account with the OpenAI platform.

Procedure

1. Log in to the Cloudera Data Warehouse service

- Go to the Virtual Warehouses tab, locate the Virtual Warehouse on which you want to enable this feature, and click  Edit .
- Go to Configurations Hue , select hue-safety-valve from the Configuration files drop-down menu, and add the following lines :

```
[desktop]
[[ai_interface]]
    service='openai'
```

You can specify the model_name (optional) and define the model. If no model is defined, the default model (gpt-3.5-turbo-16k) will be used.

- Click Apply Changes.

Results

You see  Assistant on the Hue SQL editor, and the SQL AI Assistant will connect to the specified model on the OpenAI platform.

Complete list of model-related configurations for setting up the Hue SQL AI Assistant

Review the list of service, model, and semantic search-related configurations used for custom configuring the AI services and models you want to use with the SQL AI Assistant and how to specify them in the Hue Advanced Configuration Snippet in the Cloudera Data Warehouse web interface.

List of service and model-related configurations

IN Cloudera Data Warehouse, you can configure by going to Virtual Wharehouse Configurations Hue hue-safety-valve . The following template shows the base structure of the configurations and adding the following lines:

```
[desktop]
[[ai_interface]]
    [***CONFIG-KEY1***]='[***VALUE***]'
    [***CONFIG-KEY2***]='[***VALUE***]'
[[semantic_search]]
    [***CONFIG-KEY1***]='[***VALUE***]'
    [***CONFIG-KEY2***]='[***VALUE***]'
```

AI interface-related configurations

Here is the complete list of configurations under [[ai_interface]], which allows you to specify the service and model to be used:

AI interface config key	Description
service	API service to be used for AI tasks. AI is disabled when a service is not configured. For example, Workbench and Cloudera AI Inference service are API services.
service_version	API service version to be used for AI tasks.
trusted_service	Indicates whether the LLM is trusted or not. Turn on to disable the warning. The default value is False.
model	The AI model you want to use for AI tasks. For example, gpt and llama.
model_name	The fully qualified name of the model to be used. For example, gpt-3.5-turbo-16k.
model_ref	The `model_ref` is a placeholder for adding the access key of the specific model you want to use.
base_url	Service API base URL.
add_table_data	When enabled, sample rows from the table are added to the prompt. The default value is True.
table_data_cache_size	Size of the LRU cache used for storing table sample data.

AI interface config key	Description
auto_fetch_table_meta_limit	Number of tables to load from a database, initially.
token	Service API secret token.
token_script	Provides a secure way to get the service API secret token.
enabled_sql_tasks	A comma-separated list of SQL-related AI tasks available in the Editor.

User Input Validation for Hue SQL AI

Following is the complete list of configurations under `[[ai_interface]]`. It helps to specify the input validation to enhance security and optimize performance.

AI interface config key	Description
user_input_max_length	Ensure the configured user input length is not exceeded. The default limit is 1000, but you can configure it to a higher value if needed.
user_input_remove_characters	Remove specific characters from user input, such as newlines (<code>\n</code>), tab spaces (<code>\t</code>), and others, to ensure clean and consistent formatting.
user_input_banned_keyphrases	Block user input if certain configured keyphrases are found.
user_input_banned_regex	Block user input if a configured regex pattern match is found.
user_input_block_html	Escape HTML tags to prevent malicious activities and ensure secure input handling. This config accepts a boolean value: True to escape HTML tags or False to allow raw HTML. The default value is set to False.

The following sample configuration sets the validations for user input:

```
[[ai_interface]]
service='azure'
model_name='[***DEPLOYMENT-NAME***]'
base_url='https://[***RESOURCE***].cloudera.site/model'
token='[***RESOURCE-KEY***]'
user_input_max_length=1000
user_input_remove_characters="&\n\r\t"
user_input_banned_keyphrases=""
user_input_banned_regex=""
user_input_block_html="False"
```

Semantic search-related configurations

Specify the semantic search-related configurations used for RAG under the `[[semantic_search]]` section, as listed in the following table:

Semantic search config key	Description
relevancy	The technology you want to use for semantic search. Acceptable values are <code>vector_search</code> or <code>v</code> .
embedding_model	The model you want to use for data-embedding. This must be compatible with SentenceTransf
top_k	Number of top-ranking items returned by semantic search.

Semantic search config key	Description
cache_size	Size of the LRU cache used for storing embedding.

Hue SQL AI Assistant FAQ

A collection of frequently asked questions about Hue SQL AI Assistant.

- [General Questions](#) on page 13
- [Using SQL AI Assistant](#) on page 13
- [AI Models and Security](#) on page 14
- [Configuration and Setup](#) on page 14

General Questions

What is the SQL AI Assistant in Hue?


The SQL AI Assistant in Hue is an AI-powered tool integrated into the SQL editor that helps users generate, edit, optimize, fix, and summarize SQL queries using natural language. It leverages large language models (LLMs) to assist data analysts in making SQL development faster, easier, and less error-prone.

Which SQL dialects does the SQL AI Assistant support?

Multiple SQL dialects are supported, including Hive, Impala, and Trino.

Using SQL AI Assistant

How do I launch the SQL AI Assistant?

Click the  Assistant to expand the SQL AI toolbar, which provides buttons for generating, editing, explaining, optimizing, and fixing SQL statements.

For more information, see [About setting up the Hue SQL AI Assistant](#).

What happens when I click 'Generate' in the SQL AI Assistant?

Clicking "Generate" allows you to enter a natural language query, which the assistant converts into an SQL query. The generated SQL is presented along with assumptions made by the LLM.

For more information, see [Generating SQL from natural language in Hue](#).

Can I create a query that joins multiple databases when using the Hue SQL AI Assistant?

Yes, the Hue SQL AI Assistant supports multi-database queries. You can select multiple databases in the AI Assistant Settings pop-up, allowing you to create queries that join tables across different databases.

For more information, see [Multi database support for SQL query](#).

How does the 'Edit' function work?

The "Edit" button allows users to modify an active SQL statement. If an NQL comment precedes the statement, it can be reused by pressing Tab. Users can also enter new instructions for modifications.

For more information, see [Editing the query in natural language in Hue](#).

What do 'Optimize' and 'Fix' do?

- "Optimize" improves SQL query structure and performance while maintaining the original results.
- "Fix" automatically corrects syntactic errors and misspellings in the SQL query.

For more information, see [Optimizing a query in Hue](#) and [Fixing a query in Hue](#).

How does the 'Explain' function work?

The "Explain" button provides a natural language summary and explanation of the selected SQL query, which can be inserted as a comment in the editor.

For more information, see [Getting an explanation of a SQL query in natural language in Hue](#).

AI Models and Security

Which AI models does the SQL AI Assistant support?

The Hue SQL AI Assistant supports Cloudera AI Workbench and Cloudera AI Inference service, along with several third-party services. Using the Cloudera integrations enhances the Hue SQL AI Assistant by enabling the use of private models hosted within Cloudera-managed infrastructure. This ensures enhanced security and privacy while leveraging GenAI for the Hue SQL-related tasks. For more information, see [Supported services](#).

How does the SQL AI Assistant handle data privacy?

The SQL AI Assistant shares only the data that the logged-in user is authorized to access. It uses a Retrieval Augmented Generation (RAG)-based architecture to limit the number of tables sent per request. However, there is currently no way to explicitly exclude certain tables from being shared.

Configuration and Setup

What AI services are supported for integration?

Supported services include:

- [Cloudera AI Workbench](#)
- [Cloudera AI Inference service](#)
- [Microsoft Azure OpenAI](#)
- [OpenAI API](#)
- [Amazon Bedrock](#)

Is it necessary to train the Hue SQL AI Assistant on the database schema before using it?

Training is not necessary for the Hue SQL AI Assistant. Once connected, the assistant can begin querying immediately. However, in cases where table names are similar or column names are repeated across multiple tables, it is recommended to ensure that the database metadata is well-maintained. Specifically:

- Table and column comments should be clear and descriptive as they are used by the assistant for context and disambiguation.
- Consider using distinct table and column naming conventions to reduce confusion.
- Ensure that schema and table relationships are properly defined, as this helps the assistant understand context.

Does the Hue SQL AI Assistant only work based on the prompt, or does it also have information about the present databases? For example, can it correct the name of a table or database if I write it incorrectly within the prompt?

Yes, in most cases, the Hue SQL AI Assistant can catch issues with your prompt, such as misspelled table names. This is because the user prompt is packaged alongside information about the SQL dialect in use, the database metadata, and sample rows from relevant tables. However, if another table exists with the misspelled name, the assistant might be misdirected.

About deploying the shared Hue service

Cloudera Data Warehouse 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.

Advantages of deploying a shared Hue service

By deploying a shared Hue service, you can manage costs by keeping only those Virtual Warehouses running that 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 are saved. Moreover, the shared Hue service remains active as long as the environment is active.

Limitations

When you use the Importer in Hue to create tables from files, Hue creates a Hive table by default (if Hive is available) and uses the first available Hive Virtual Warehouse. You cannot select a specific Virtual Warehouse during file import. 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 Cloudera Management Console, similar to how you specified them 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 Cloudera Data Warehouse 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 own Hue database. Cloudera Data Warehouse 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.

Cloudera Data Warehouse provides you with 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.

Creating a shared Hue instance

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.

Procedure

1. Log in to the Cloudera Data Warehouse service as a DWAdmin
2. Click the Shared Hue Service option from the left navigation pane of the Cloudera Data Warehouse 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. 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.
 - Reuse if Hue data is present: Select this option to use an existing Hue database for your new shared Hue instance.
 - 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.

What to do next

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.

Procedure

1. Log in to the Cloudera Data Warehouse service as a DWAdmin.
2. Click the Shared Hue Service option from the left navigation pane of the Cloudera Data Warehouse UI.
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.

Procedure

1. Log in to the Cloudera Data Warehouse service as a DWAdmin.
2. Click the Shared Hue Service option from the left navigation pane of the Cloudera Data Warehouse UI.
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.

FAQ for shared Hue service

A collection of frequently asked questions about deploying a shared Hue service.

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 instance 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.