

Cloudera Runtime 7.3.1

## Hue Overview

Date published: 2020-05-21

Date modified: 2024-12-10

# CLOUdera

<https://docs.cloudera.com/>

# 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

<b>Hue overview.....</b>	<b>4</b>
<b>About Hue Query Processor.....</b>	<b>6</b>
<b>About the Hue SQL AI Assistant.....</b>	<b>6</b>

## Hue overview

Hue is a web-based interactive query editor that enables you to interact with databases and data warehouses. Data architects, SQL developers, and data engineers use Hue to create data models, clean data to prepare it for analysis, and to build and test SQL scripts for applications.

Hue packs the combined abilities of Data Analytics Studio (DAS) such as query optimization, query debugging framework, and rich query editor experience of Hue, making Hue the next generation SQL assistant on CDP. You can search Hive query history, view query details, visual explain plan, and DAG information, compare two queries, and download debug bundles for troubleshooting from the Job Browser page.

Hue offers powerful execution, debugging, and self-service capabilities to the following key Big Data personas:

- Business Analysts
- Data Engineers
- Data Scientists
- Power SQL users
- Database Administrators
- SQL Developers

Business Analysts (BA) are tasked with exploring and cleaning the data to make it more consumable by other stakeholders, such as the data scientists. With Hue, they can import data from various sources and in multiple formats, explore the data using File Browser and Table Browser, query the data using the smart query editor, and create dashboards. They can save the queries, view old queries, schedule long-running queries, and share them with other stakeholders in the organization. They can also use Cloudera Data Visualization (Viz App) to get data insights, generate dashboards, and help make business decisions.

Data Engineers design data sets in the form of tables for wider consumption and for exploring data, as well as scheduling regular workloads. They can use Hue to test various Data Engineering (DE) pipeline steps and help develop DE pipelines.

Data scientists predominantly create models and algorithms to identify trends and patterns. They then analyze and interpret the data to discover solutions and predict opportunities. Hue provides quick access to structured data sets and a seamless interface to compose queries, search databases, tables, and columns, and execute query faster by leveraging Tez and LLAP. They can run ad hoc queries and start the analysis of data as pre-work for designing various machine learning models.

Power SQL users are advanced SQL experts tasked with analyzing and fine-tuning queries to improve query throughput and performance. They often strive to meet the TPC decision support (TPC-DS) benchmark. Hue enables them to run complex queries and provides intelligent recommendations to optimize the query performance. They can further fine-tune the query parameters by comparing two queries, viewing the explain plan, analyzing the Directed Acyclic Graph (DAG) details, and using the query configuration details. They can also create and analyze materialized views.

The Database Administrators (DBA) provide support to the data scientists and the power SQL users by helping them to debug long-running queries. They can download the query logs and diagnostic bundles, export and process log events to a database, or download it in JSON file format for further analysis. They monitor all the queries running in the system and terminate long-running queries to free up resources. Additionally, they may also provision databases, manage users, monitor and maintain the database and schemas for the Hue service, and provide solutions to scale up the capacity or migrate to other databases, such as PostgreSQL, Oracle, MySQL, and so on.

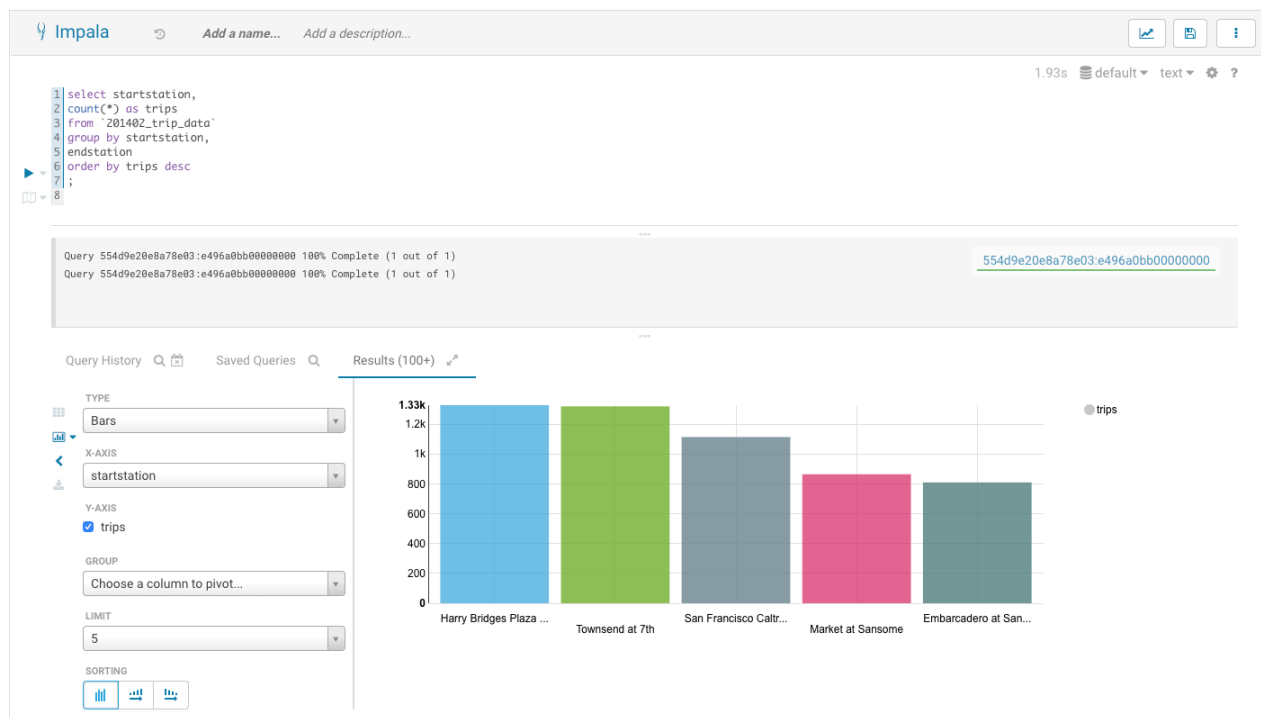
All Hue users can download logs and share them with their DBA or Cloudera Support for debugging and troubleshooting purposes.

SQL Developers can use Hue to create data sets to generate reports and dashboards that are often consumed by other Business Intelligence (BI) tools, such as Cloudera Data Visualization.

Hue can sparsely be used as a Search Dashboard tool, usually to prototype a custom search application for production environments.

For example, the following image shows a graphical representation of Impala SQL query results that you can generate with Hue:

**Figure 1: Impala SQL query results generated with Hue**



You can use Hue to:

- Explore, browse, and import your data through guided navigation in the left panel of the page.

From the left panel, you can:

- Browse your databases
- Drill down to specific tables
- View HDFS directories and cloud storage
- Discover indexes and HBase or Kudu tables
- Find documents

Objects can be tagged for quick retrieval, project association, or to assign a more "human-readable" name, if desired.

- Query your data, create a custom dashboard, or schedule repetitive jobs in the central panel of the page.

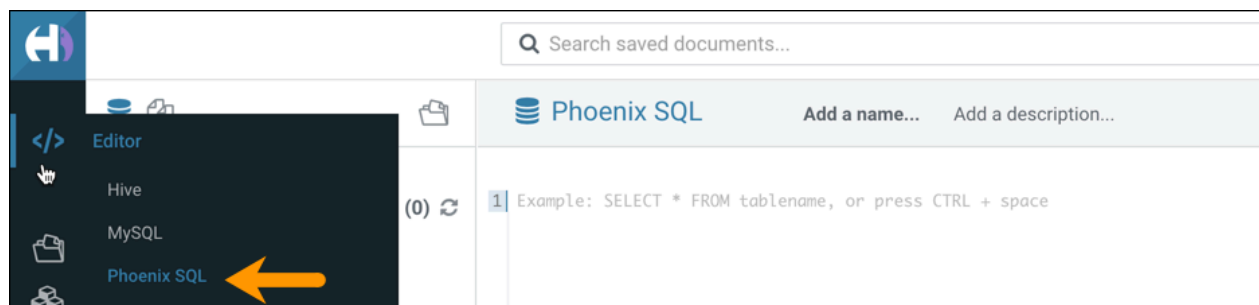
The central panel of the page provides a rich toolset, including:

- Versatile editors that enable you to create a wide variety of scripts.
- Dashboards that you can create "on-the-fly" by dragging and dropping elements into the central panel of the Hue interface. No programming is required. Then you can use your custom dashboard to explore your data.
- Schedulers that you can create by dragging and dropping, just like the dashboards feature. This feature enables you to create custom workflows and to schedule them to run automatically on a regular basis. A monitoring interface shows the progress, logs, and makes it possible to stop or pause jobs.
- Get expert advice on how to complete a task using the assistance panel on the right.

The assistant panel on the right provides expert advice and hints for whatever application is currently being used in the central panel. For example, in the above image, Impala SQL hints are provided to help construct queries in the central panel.

- (Hive-only) View query details such as query information, visual explain, query timeline, query configuration, Directed Acyclic Graph (DAG) information, DAG flow, DAG swimlane, DAG counters, and DAG configurations.
- (Impala-only) View query details such as query type, duration, default database, request pool, CPU time, rows produced, peak memory, HDFS bytes read, coordinator details, execution plan, execution summary, and counters and metrics.
- (Hive-only) Terminate Hive queries.
- (Hive-only) Download debug bundles.
- Compare two queries.
- View query history.

Hue also provides a simple SQL interface to create, access, and query HBase tables using Apache Phoenix in addition to HBase shell and database API.



#### Related Information

[Apache Phoenix and SQL](#)

## About Hue Query Processor

The Hue Query Processor service indexes Hive and Impala query history and provides APIs to access them. Enabling the processor allows you to view only Hive query history and query details on the Queries tab on the Hue Job Browser page.



**Note:** Hue does not display Impala query history and query details on the Queries tab on the **Job Browser** page.



**Note:** Query Processor is set up by default in clusters created using the Data Engineering template. Administrators have the privilege to add or remove services in Data Hub clusters.

You can enable or disable Hue to display the **Queries** tab by selecting or deselecting the Query Processor Service option in the Hue Advanced Configuration Snippet in Cloudera Manager.

#### Related Information

[Adding the Query Processor service to a cluster](#)

[Removing the Query Processor service from a cluster](#)

[Enabling the Queries tab in Hue](#)

## About the Hue SQL AI Assistant

Learn about the AI models and services that Hue uses to run the SQL AI Assistant and its limitations. Review what data is shared with the LLM models before you start using the SQL AI Assistant with Hue.

A SQL AI Assistant has been integrated into Hue with the capability to leverage the power of Large Language Models (LLMs) for various SQL tasks. It helps you to create, edit, optimize, fix, and succinctly summarize queries using natural language and makes SQL development faster, easier, and less error-prone. Both Hive and Impala dialects are supported.

### AI models and services that Hue uses

The SQL AI Assistant supports various LLMs and hosting services. The models run on cloud infrastructure, and the AI Assistant can be configured to use them remotely. Cloudera has tested with GPT running in Open AI, Microsoft Azure, and Amazon Bedrock. The following service-model combinations are supported:

Service Provider	Model	Model Versions
OpenAI	OpenAI GPT	<ul style="list-style-type: none"> <li>gpt-3.5-turbo</li> <li>gpt-3.5-turbo-16k</li> </ul> Current GPT version is 3.5 turbo. You can configure GPT 4 for better results.
Microsoft Azure	OpenAI GPT	<ul style="list-style-type: none"> <li>gpt-3.5-turbo</li> <li>gpt-3.5-turbo-16k</li> </ul>
Amazon Bedrock	Anthropic Claude	<ul style="list-style-type: none"> <li>anthropic.claude-v1</li> <li>anthropic.claude-v2</li> </ul>
Amazon Bedrock	Amazon Titan	<ul style="list-style-type: none"> <li>amazon.titan-text-express-v1</li> </ul>



#### Note:

You must have access to the Hugging Face to download the required sentence transformer model, and ensure your system can connect to the internet, specifically, [huggingface.co](https://huggingface.co). Since these models aren't pre-bundled, they must be downloaded during setup. For more information, see [Hugging Face](#)

For better results, Cloudera recommends you to use the SQL AI assistant with the Azure OpenAI service. This ensures that the models run in your Virtual Private Cloud (VPC) network.

The SQL AI Assistant uses a Retrieval Augmented Generation (RAG)-based architecture for augmenting results. It uses the sentence-transformer library for semantic search, and Hue can be configured with any of the [pre-trained models](#) for better multi-lingual support. By default, “all-MiniLM-L6-v2” models are used.

Embedding Model	Language Support
all-MiniLM-L6-v2	English
distiluse-base-multilingual-cased-v1	Arabic, Chinese, Dutch, English, French, German, Italian, Korean, Polish, Portuguese, Russian, Spanish, and Turkish.

### What data is shared with the LLM models

The following details are shared with the LLMs:

- Everything that a user inputs
- Dialect in use
- Table details such as table name, column names, column data types and related keys, partitions, and constraints that the logged-in user has access to.
- Three sample rows from the tables (as per the best practices specified in [Evaluating the Text-to-SQL Capabilities of Large Language Models](#))

### Limitations

#### Non-deterministic nature

LLMs are non-deterministic, which means you cannot guarantee the same output for the same input every time, and it can lead to different responses to similar queries.

**Ambiguity**

LLMs may struggle to handle ambiguous queries or contexts. SQL queries often rely on specific and unambiguous language, but LLMs can misinterpret or generate ambiguous SQL queries, leading to incorrect results.

**Hallucinations**

In the context of LLMs, hallucination refers to a phenomenon where these models generate text or responses that are incorrect, nonsensical, or fabricated. Occasionally you might see incorrect identifiers or literals in the response.