Cloudera Runtime 7.3.1

Using Hue

Date published: 2020-07-28
Date modified: 2024-12-10

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

About using Hue 5
Accessing and using Hue 5
Viewing Hive query details 11
Viewing Hive query history 11
Viewing Hive query information 11
Viewing explain plan for a Hive query 12
Viewing Hive query timeline 13
Viewing configurations for a Hive query 13
Viewing DAG information for a Hive query 15
Viewing Impala query details 18
Viewing Impala query history 18
Viewing Impala query information 19
Viewing the Impala query execution plan 19
Viewing the Impala query metrics 20
Viewing Impala profiles in Hue 21
Terminating Hive queries 21
Comparing Hive and Impala queries in Hue 22
Start SQL AI Assistant 24
Generate SQL from NQL 24
Edit query in natural language 25
Explain query in natural language 26
Optimize SQL query 28
Fixing a query in Hue 29
Generate comment for a SQL query 30
Enable stored procedures in Hue 31
Run stored procedure from Hue 32
Using SQL to query HBase from Hue 33
Querying existing HBase tables 34
Enabling the SQL editor autocompleter 34
Rerunning a query from the Job Browser page 35
Using governance-based data discovery 35
Searching metadata tags 36
Using Amazon S3 with Hue 36
Enabling S3 browser for Hue configured with IDBroker 37
Enabling S3 browser for Hue configured without IDBroker 38
Enabling S3 File Browser for Hue with RAZ in DataHub 38
Using Azure Data Lake Storage Gen2 with Hue 40
Enabling ABFS file browser for Hue configured with IDBroker 40
Enabling ABFS file browser for Hue configured without IDBroker 41
Enabling ABFS File Browser in Hue with RAZ in DataHub 42
Using Google Cloud Storage with Hue 43
Prerequisites for enabling GS File Browser 44
Enabling GS File Browser with RAZ 44
Disabling the automatic creation of user home directories 45
Granting permission to access S3, ABFS, GS File Browser in Hue 45
Creating tables in Hue by importing files 46
Supported non-ASCII and special characters in Hue 47
Options to rerun Oozie workflows in Hue 49
Unsupported features in Hue 50

Known limitations in Hue..... 50

About using Hue

Hue provides a one-stop querying experience in Cloudera Data Platform (CDP) that leverages Hive and Impala SQL queries.

Accessing and using Hue

Get started using Hue by analyzing and visualizing your data with Impala and Hive SQL query engines.

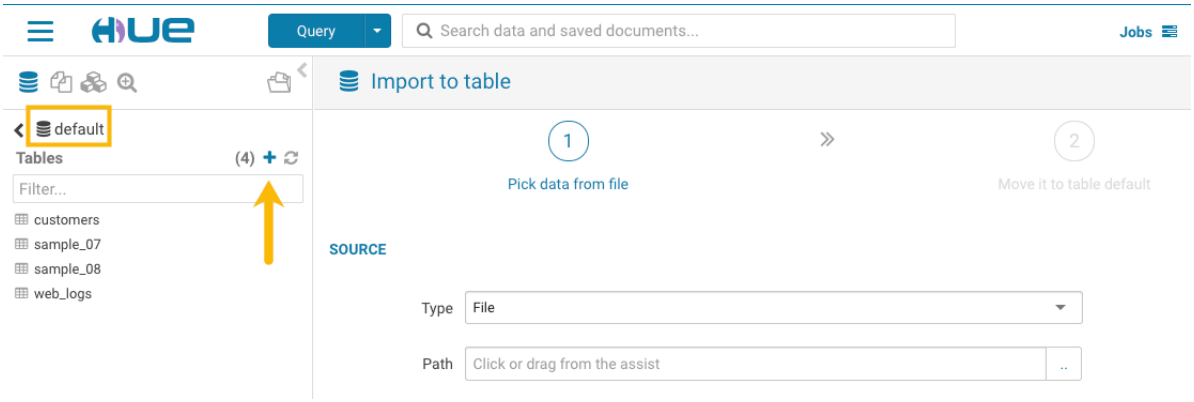
About this task

To try Hue without having an account, try running sample queries on http://demo.gethue.com/.

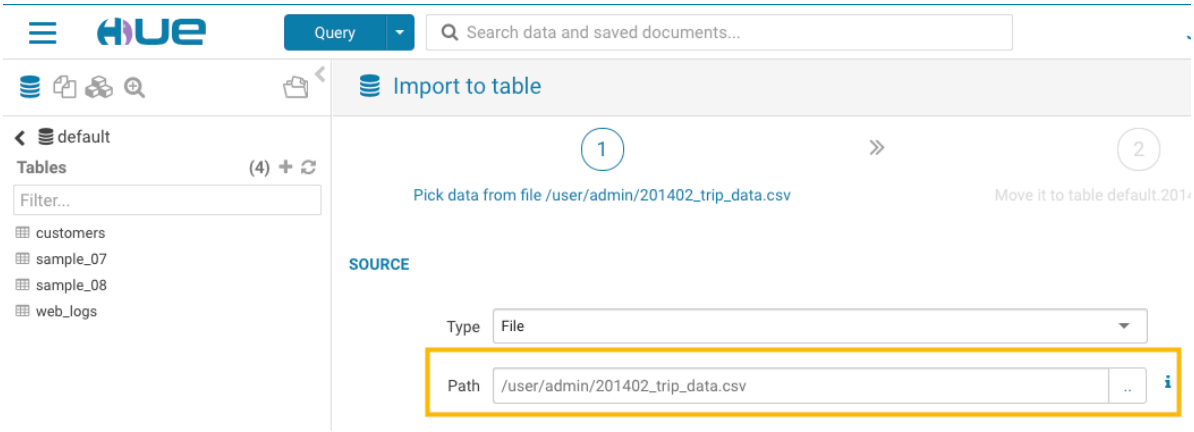
Procedure

1. Download and unzip one year of bike trips from the Bay Area Bike Share program. This file is about 80 MB in size.

2. Create a table from the ~/babs_open_data_year_1/201402_babs_open_data/201402_trip_data.csv file found in the unzipped babs_open_data_year_1.zip file:
   a) In the Cloudera Manager, select Hue WebUI Hue Load Balanced to launch Hue.
   b) In the left navigation panel of Hue, make sure the default database is selected, and click the plus sign to create a table as shown in the following image:



   If the default database is not selected, click the "less than" icon that is next to the database icon in the left panel. This enables you to select the default database.
   c) In the center panel Importer UI, set Type to File.
   d) Drag the 201402_trip_data.csv file to the Path field as shown in the following image:



   e) Set the formats as follows:
      - Field Separator = Comma (,)
      - Record Separator = New line
      - Quote Character = Double Quote

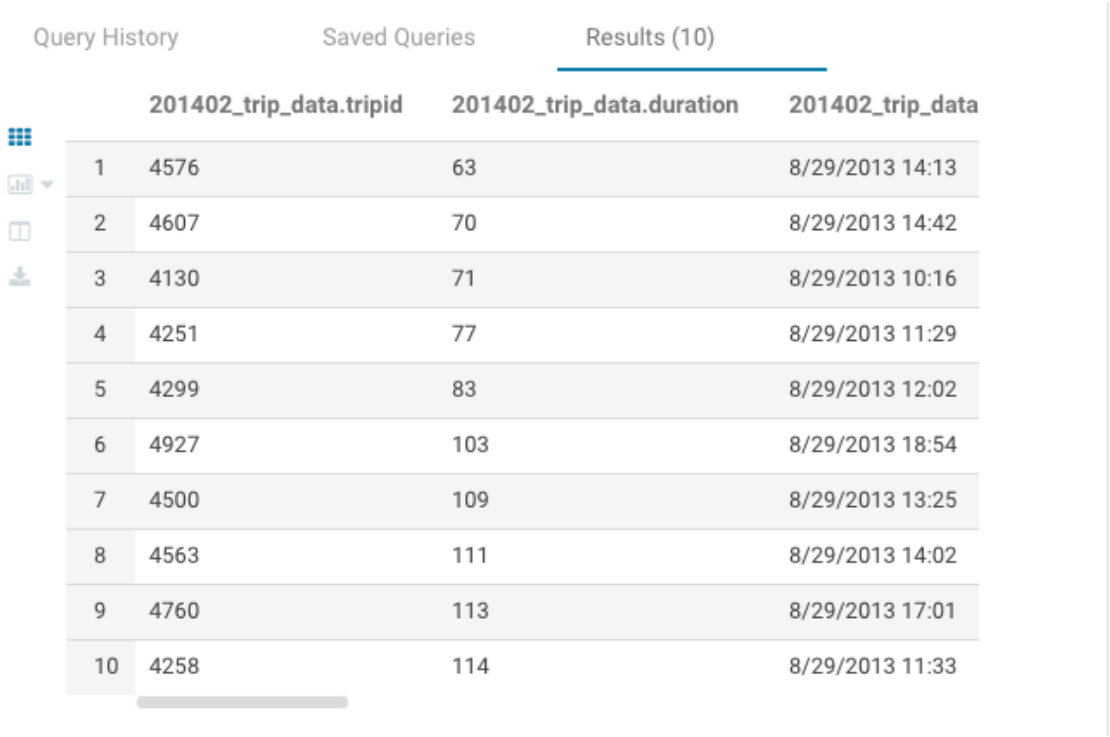
      Then click Next at the bottom of the page.
   f) Set the properties Format = Text.
   g) Edit the FIELDS as follows:
      - Rename Bike # to Bike ID
      - Change the data type of ZipCode to string.
      - Remove all of the spaces in the Name fields.

      Then click Submit at the bottom of the page.

3. Click Query at the top of the page and select Editor Hive to open the Hive editor and then create a query.
   - Enter the following query into the editor window:

     ```
     SELECT * FROM default.201402_trip_data
     LIMIT 10;
     ```

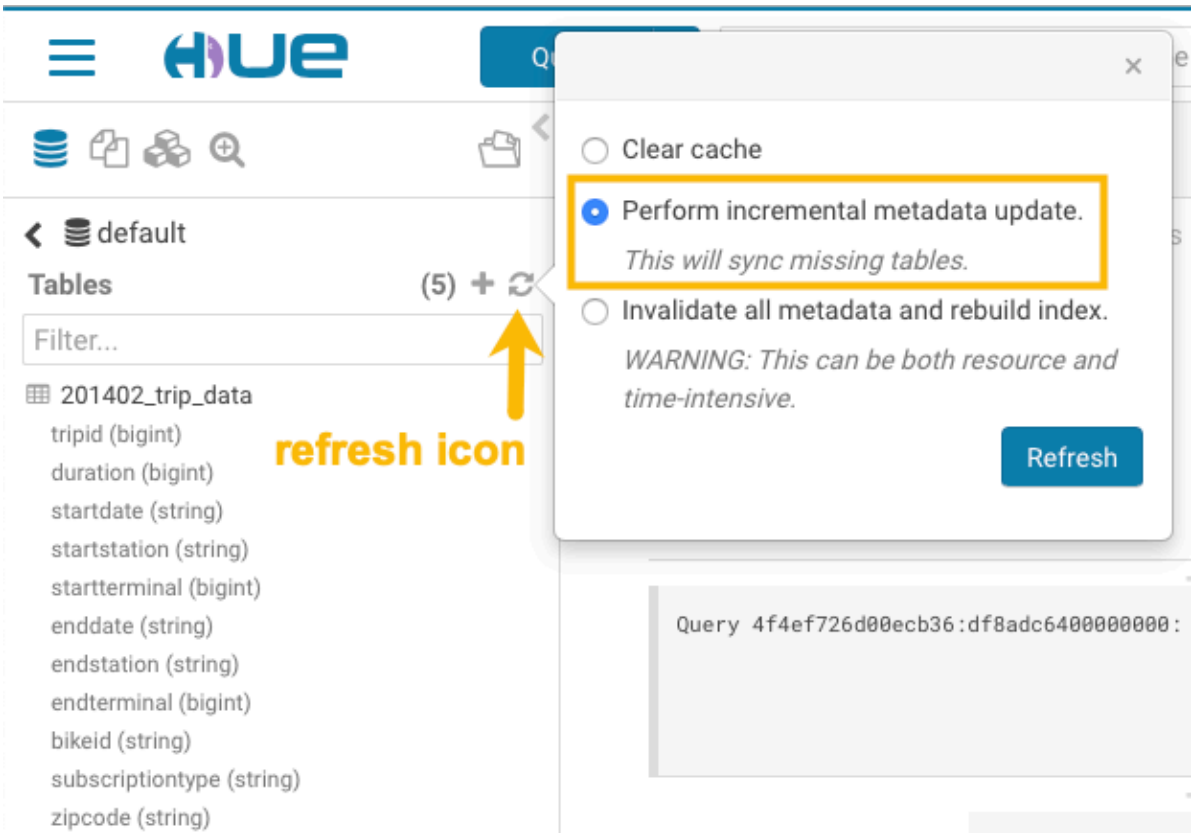
   - Click the execute icon to run the query. The following rows are returned:



Query History | Saved Queries | Results (10)

	201402_trip_data.tripid	201402_trip_data.duration	201402_trip_data
1	4576	63	8/29/2013 14:13
2	4607	70	8/29/2013 14:42
3	4130	71	8/29/2013 10:16
4	4251	77	8/29/2013 11:29
5	4299	83	8/29/2013 12:02
6	4927	103	8/29/2013 18:54
7	4500	109	8/29/2013 13:25
8	4563	111	8/29/2013 14:02
9	4760	113	8/29/2013 17:01
10	4258	114	8/29/2013 11:33

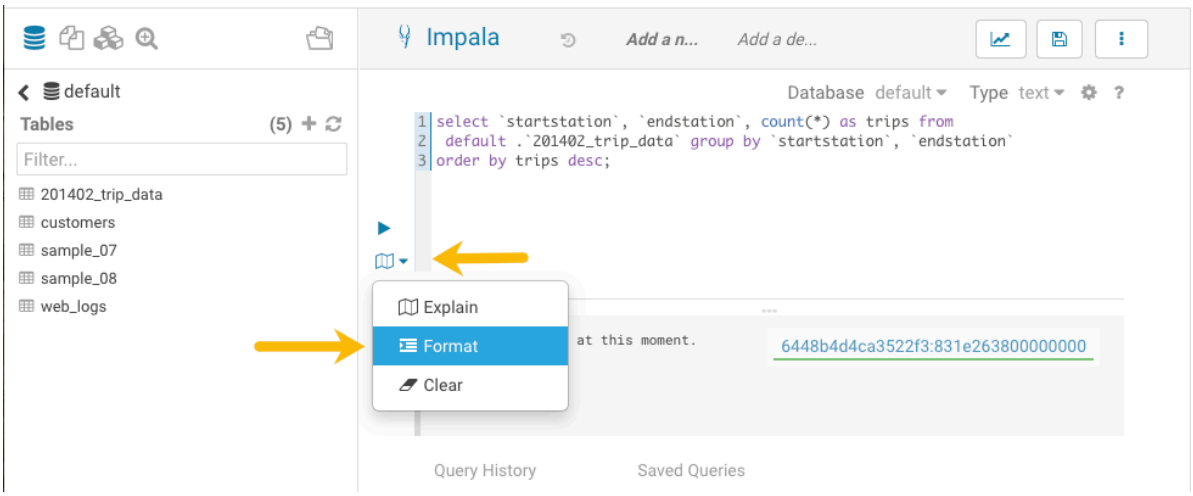
4. Click Query at the top of the page and select Editor Impala to open the Impala SQL editor and then create a query.
   a. In the left panel, click the refresh icon and select Perform incremental metadata update to make the new table visible to Impala:



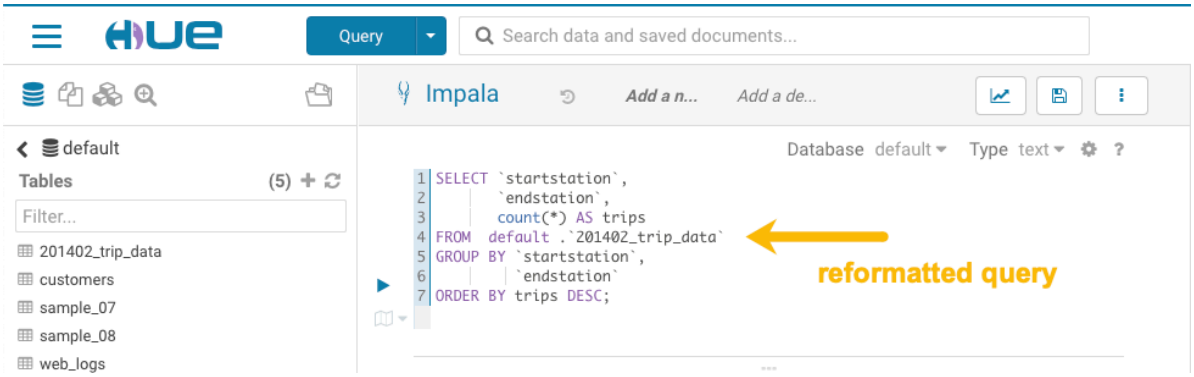
   b. Enter the following query into the editor window:

```
select 'startstation', 'endstation', count(*) as trips from default.'20
1402_trip_data'
group by 'startstation', 'endstation' order by trips desc;
```

c. Click the down arrow just under the execution icon and select Format:



This reformats the query:

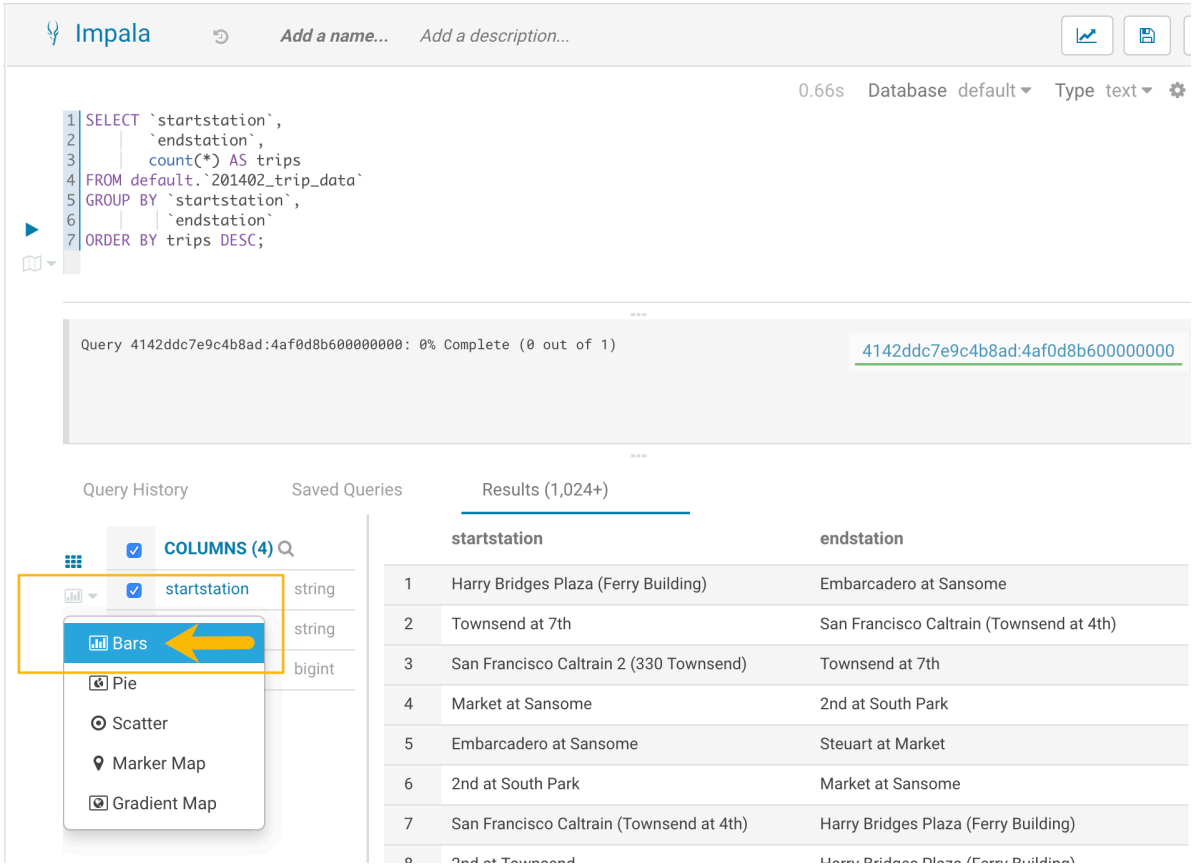


d. Click the save icon , enter a query name, and click Save.

e. Click the execute icon to run the query.

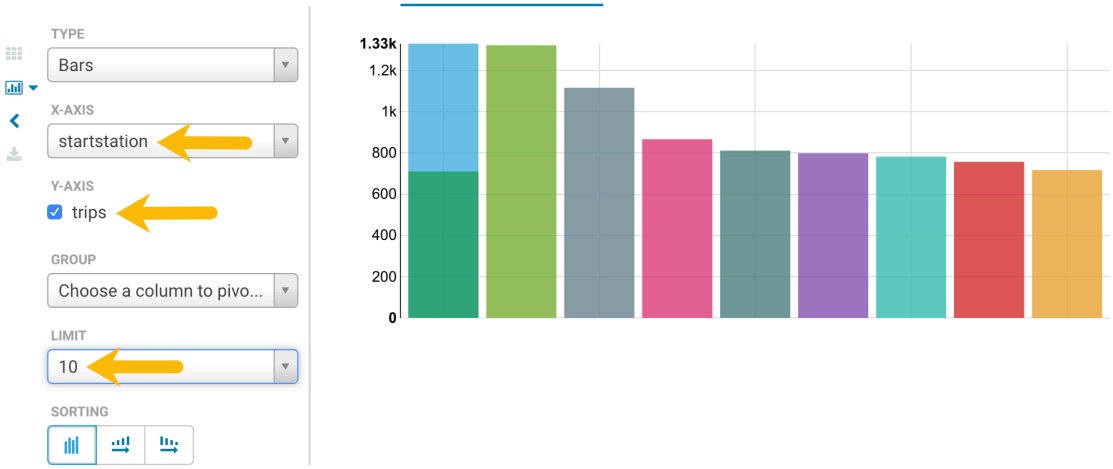
5. Create a bar chart that is based on the query results:

   a. Click the chart icon and then select Bars.



   b. Set the bar chart elements as follows:

   - X-AXIS = startstation
   - Y-AXIS = trips
   - LIMIT = 10



6. Create a pie chart by clicking the chart icon again and then select Pie.
7. Download the query results by clicking and selecting in what format you want to download, copy, or export the results.

Viewing Hive query details

You can search Hive query history, compare two queries, download debug bundles for troubleshooting, and view query details, a graphical representation of the query execution plan, and DAG information on the Job Browser page in Hue.

Viewing Hive query history

The Queries tab on the Job Browser page in Hue displays all the queries that were run from various query interfaces, such as Beeline, Hive Warehouse Connector (HWC), Tableau, Hue, and other JDBC BI clients and tools.

About this task

Only Query Processor Administrators can view historical queries of all users to monitor resource utilization and control costs from the Hue Job Browser. Non-admin users can view only their queries.

Procedure

1. Log in to the Hue web interface.
2. Click Jobs from the left assist panel.

   The **Job Browser** page is displayed.
3. Click Queries.

   The Hive queries that were run for the past seven days are displayed. You can select the time period for which you want to view the historical data.

   You can also filter queries by their status.

Related Information

Adding Query Store Administrator users on Data Hub clusters

Viewing Hive query information

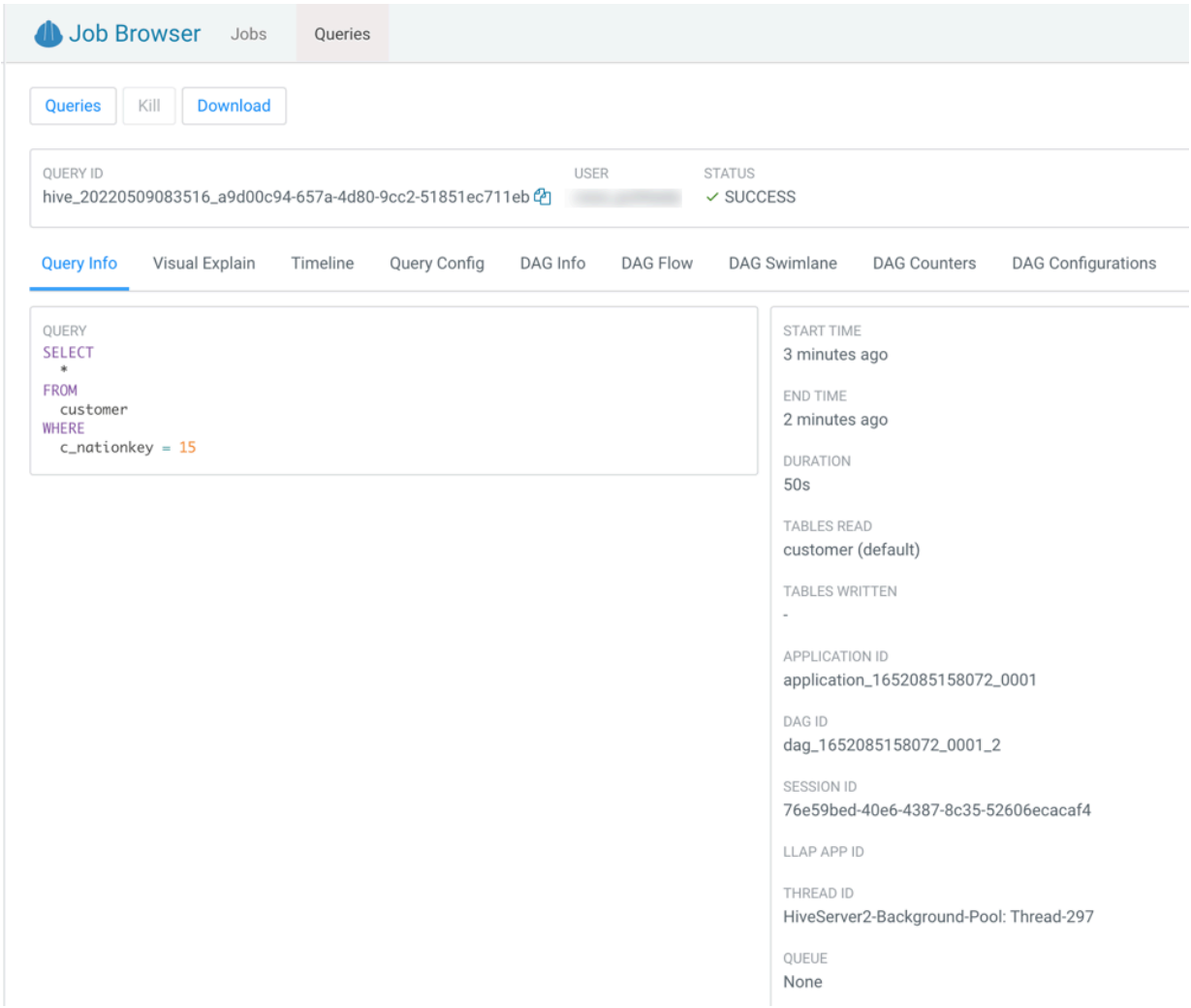
The Query Info tab provides information such as, the Hive query ID, the user who executed the query, the start time, the end time, the total time taken to execute the query, the tables that were read and written, application ID, Directed Acyclic Graph (DAG) IDs, session ID, LLAP app ID, thread ID, and the queue against which the query was run.

Procedure

1. Log in to the Hue web interface.
2. Click Jobs from the left assist panel.

   The **Job Browser** page is displayed.

3. Go to the **Queries** tab and click on the query for which you want to view the query details.
   The following image shows the **Query Info** tab on the Hue web interface:



Viewing explain plan for a Hive query

The Visual Explain feature provides a graphical representation of the query execution plan. The Explain plan is read from right to left. It provides details about every stage of query execution.

Procedure

1. Log in to the Hue web interface.
2. Click Jobs from the left assist panel.
   The **Job Browser** page is displayed.
3. Go to the **Queries** tab and click on the query for which you want to view the query details.

4. Click on Visual Explain.

   The following image shows the **Visual Explain** tab on the Hue web interface:

5. (Optional) Click ⬇ to download the query explain plan in JSON format.

Viewing Hive query timeline

The Timeline tab provides a visual representation of Hive performance logs and shows the time taken by each stage of the query execution.

About this task

Following are the stages in which a query is executed:

- Pre-execution and DAG construction: It is the first phase of query execution and is executed on the Hive engine. It constitutes the time taken to compile, parse, and build the Directed Acyclic Graph (DAG) for the next phase of the query execution.
- DAG submission: It is the second phase in which the DAG that was generated in Hive is submitted to the Tez engine for execution.
- DAG runtime: It shows the time taken by the Tez engine to execute the DAG.
- Post-execution: It is the last phase of query execution in which the files in S3/ABFS are moved or renamed.

Duration data about each phase are distilled into more granular metrics based on query execution logs.

Procedure

1. Log in to the Hue web interface.
2. Click Jobs from the left assist panel.

   The **Job Browser** page is displayed.
3. Go to the **Queries** tab and click on the query for which you want to view the query details.
4. Click on Timeline.

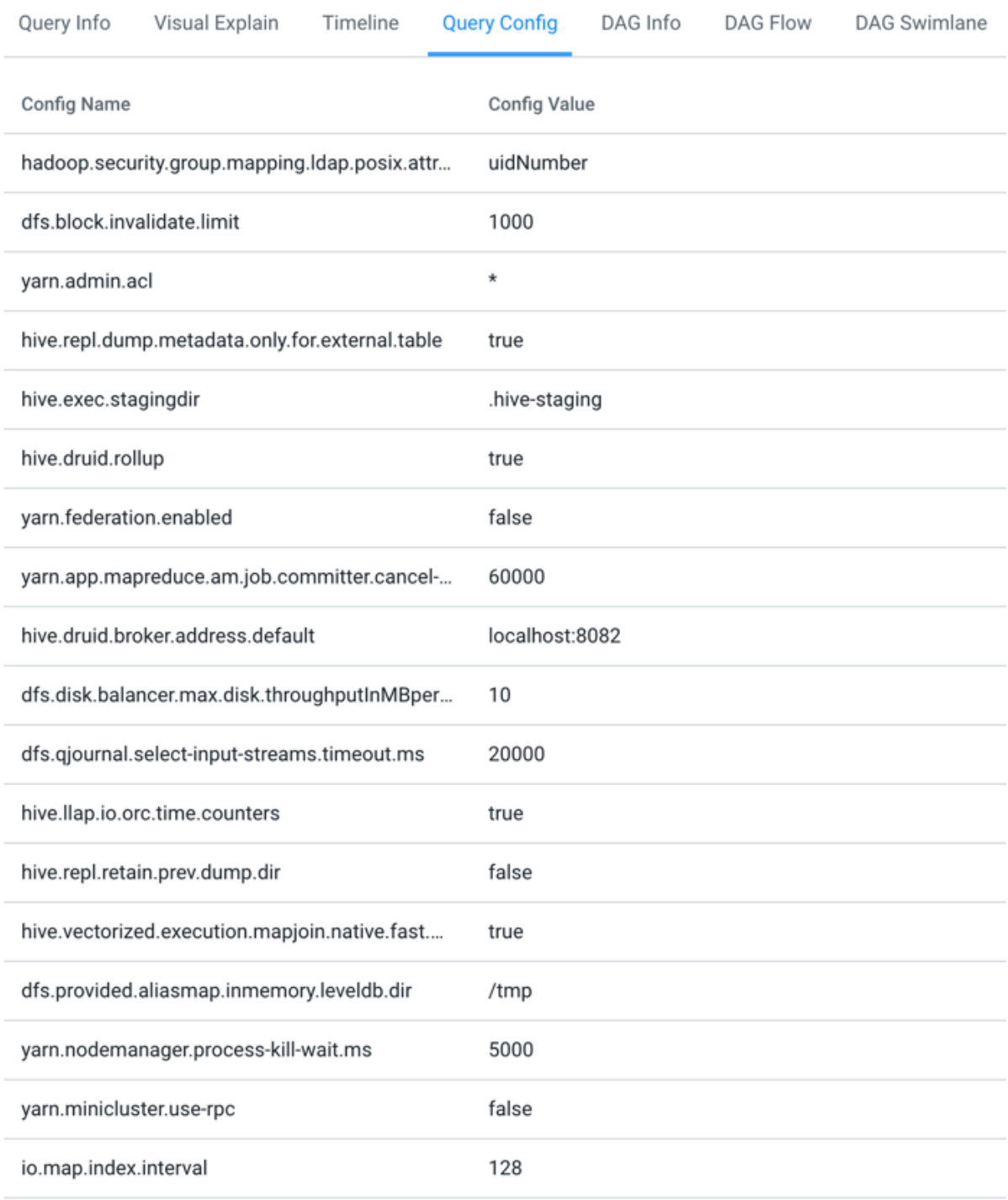
   The following image shows the **Timeline** tab on the Hue web interface:

Viewing configurations for a Hive query

The Query Config tab provides the configuration properties and settings that are used in a Hive query. You can use this tab to verify that configuration property values align with your expectations.

Procedure

1. Log in to the Hue web interface.
2. Click Jobs from the left assist panel.
   The **Job Browser** page is displayed.
3. Go to the **Queries** tab and click on the query for which you want to view the query details.
4. Click on Query Config.
   The following image shows the **Query Config** tab on the Hue web interface:



Viewing DAG information for a Hive query

Directed Acyclic Graph (DAG) is created by the Hive engine every time you query the Hive Virtual Warehouse. The Hive SQL queries are compiled and converted into a Tez execution graph also known as a DAG. DAG is a collection of vertices where each vertex executes a fragment of the query or script. Hue provides a web interface to view detailed information about DAGs.

About this task

Directed connections between vertices determine the order in which they are executed. For example, the vertex to read a table must be run before a filter can be applied to the rows of that table. As another example, consider a vertex that reads a user table that is very large and distributed across multiple computers and multiple racks. Reading the table is achieved by running many tasks in parallel.

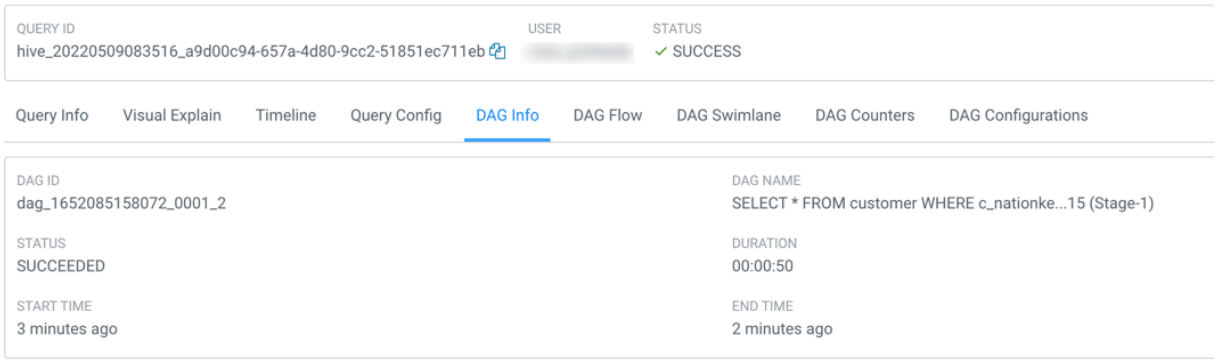
Important: The DAG information tabs (**DAG Info**, **DAG Flow**, **DAG Swimlane**, **DAG Counters**, **DAG Configurations**) are displayed only if the Tez engine is used for query execution. The Tez engine is typically utilized for complex queries.

Procedure

1. Log in to the Hue web interface.
2. Click Jobs from the left assist panel.

   The **Job Browser** page is displayed.
3. Go to the **Queries** tab and click on the query for which you want to view the query details.
4. Click DAG Info to see the DAG ID, DAG name, the status of the query, the time taken to execute the DAG, start time, and end time.

   The following image shows the **DAG Info** tab on the Hue web interface:



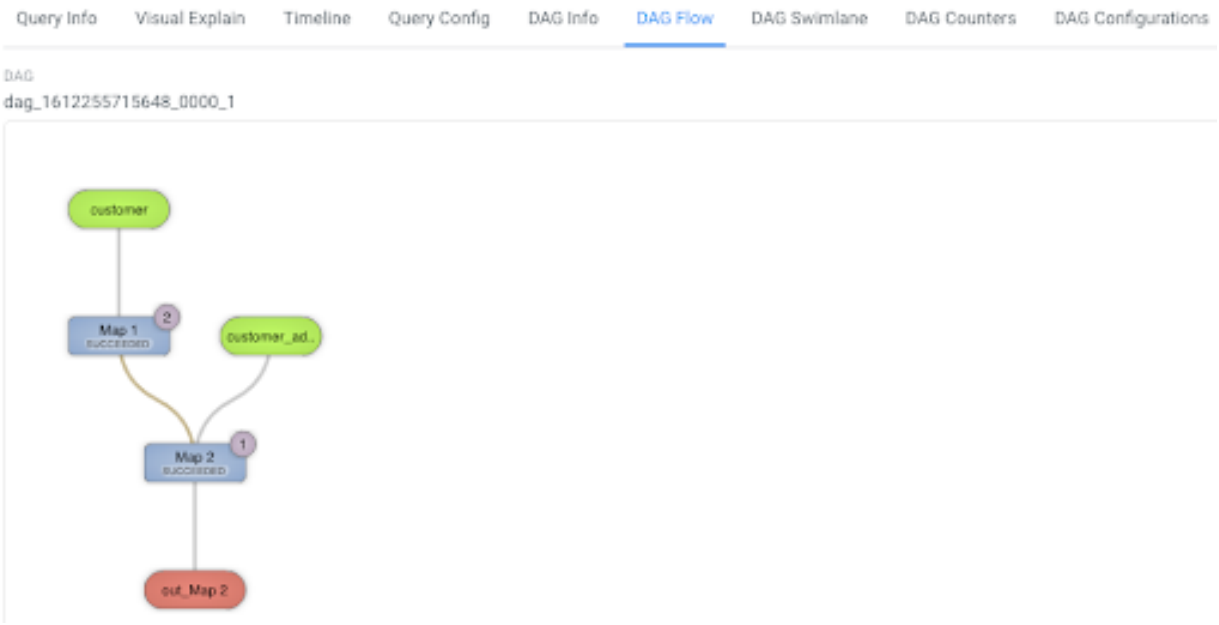
   The following table lists and describes the status of the Tez job:

Status	Description
Submitted	The DAG is submitted to Tez but is not running
Running	The DAG is currently running
Succeeded	The DAG was completed successfully
Failed	The DAG failed to complete successfully
Killed	The DAG was stopped manually
Error	An internal error occurred when executing the DAG

5. Click DAG Flow to see the DAG in the form of a flowchart.

   You can gain insight into the complexity and the progress of executing jobs, and investigate the vertices that have failures or are taking a long time to complete.

   The following image shows the **DAG Flow** tab on the Hue web interface::



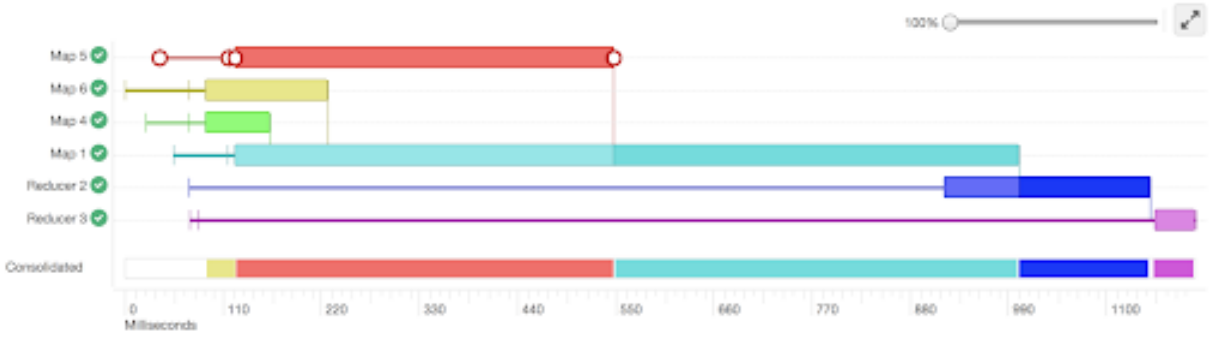
   Here, the input to vertices Map 1 and Map 2 are the tables displayed in green boxes. Next, Map 2 depends on the result set generated by Map 1. Map 2 is the last vertex in the DAG flow and after it completes its execution, the query output is written to a file in a filesystem such as S3 or ABFS.

   There are a few options to change the layout of the DAG flow. You can hide the input and the output nodes to view only the task vertices by clicking the Toggle source/sink visibility button. You can switch between the horizontal and vertical orientation by clicking the Toggle orientation button.

6. Click DAG Swimlane to see the DAG of the vertices against time.

   Each mapping and reducing task is a vertex. Each horizontal bar of the swimlane represents the total time taken by the vertex to complete the execution. The vertical lines indicate the time when the vertex was initialized, the time when the vertex started, the time when the first task started, the time when the last task was completed, and the time when the vertex finished its execution. When you mouse over the vertical line, the bubble displays the stage of the vertex execution and provides a timestamp. The vertical lines connecting two vertices denote the dependency of a vertex on another vertex.

   The following image shows the **DAG Swimlane** tab on the Hue web interface:



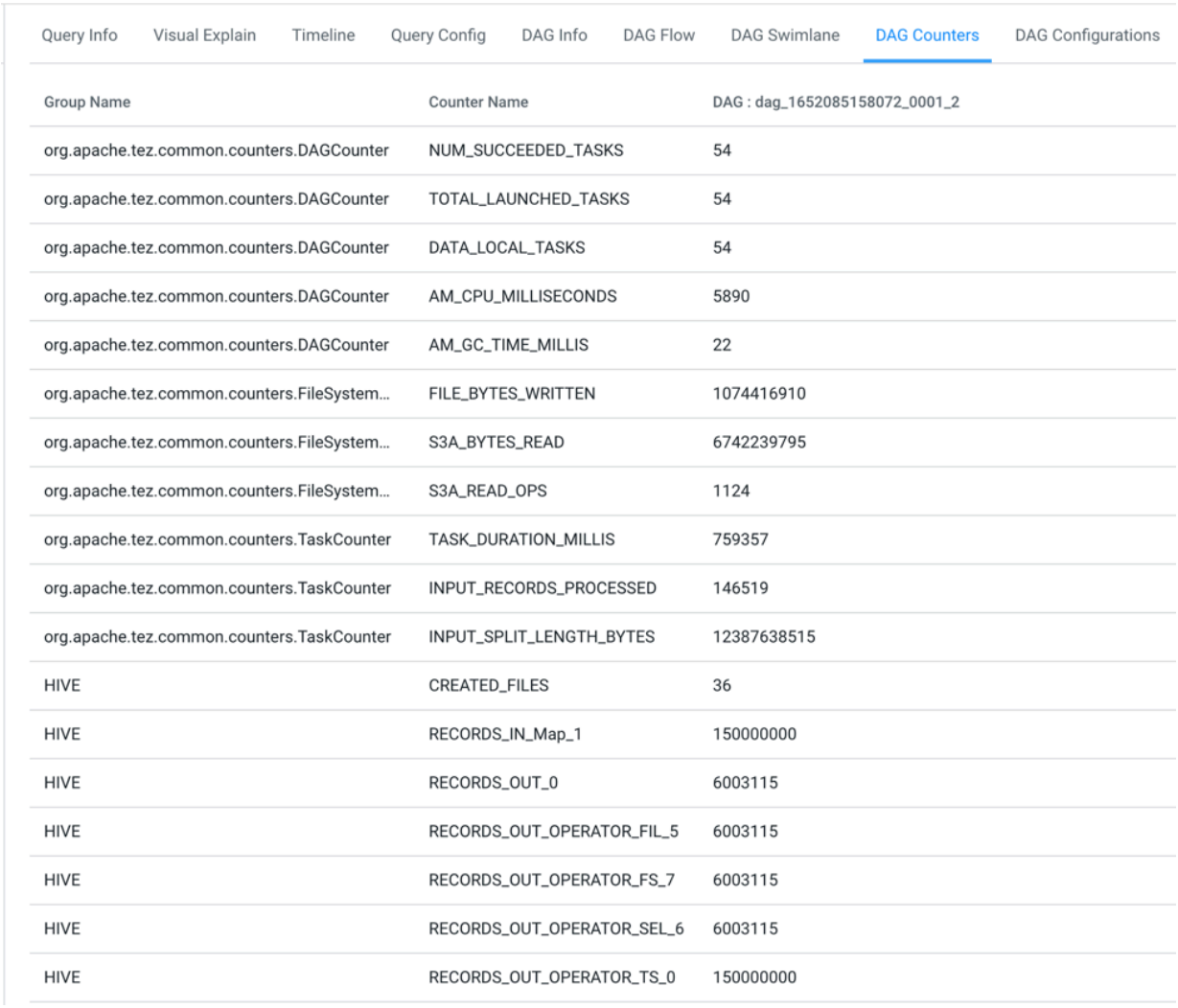
   In this example, Map 1 depends on the results of Map 5. Map 1 will finish its execution only when Map 5 finishes its execution successfully. Similarly, Reducer 2 depends on Map 1 to complete its execution.

   The consolidated timeline shows the percentage of time each vertex took to complete executing.

7. Click DAG Counters to see details such as the number of bytes read and written, number of tasks that initiated and ran successfully, amount of CPU and memory consumed, and so on.

   The **DAG Counters** tab provides a way to measure the progress or the number of operations that occur within a generated DAG. Counters are used to gather statistics for quality control purposes or problem diagnosis.

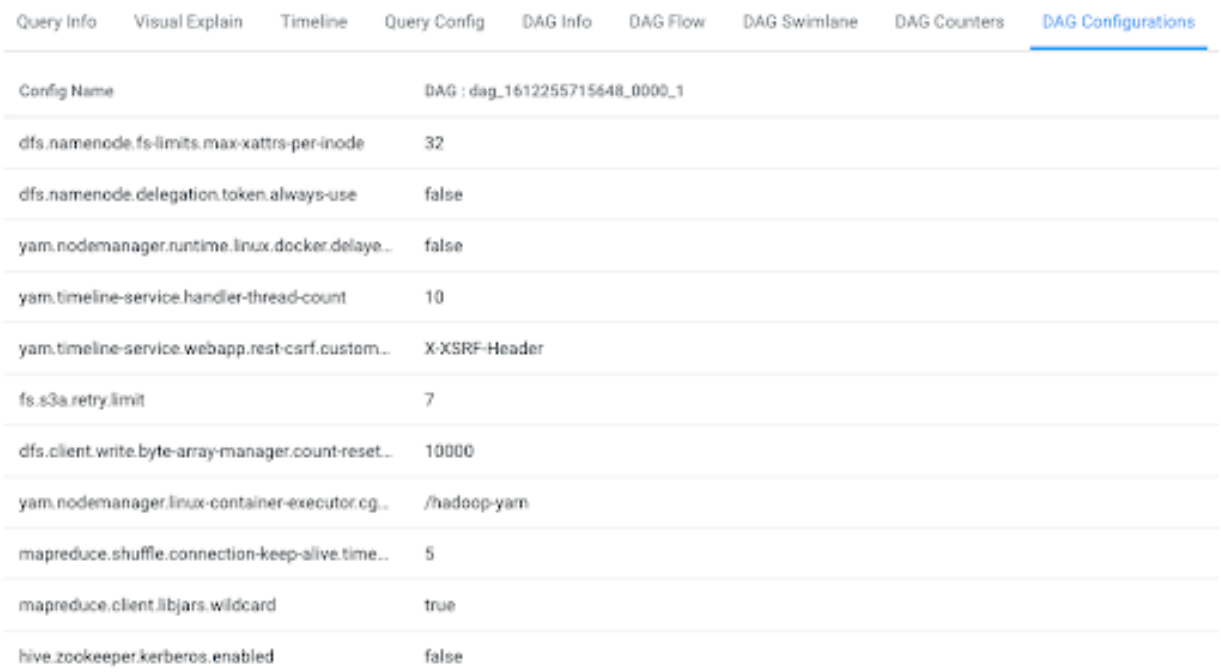
   The following image shows the **DAG Counters** tab on the Hue web interface:



Query Info | Visual Explain | Timeline | Query Config | DAG Info | DAG Flow | DAG Swimlane | DAG Counters | DAG Configurations

Group Name	Counter Name	DAG : dag_1652085158072_0001_2
org.apache.tez.common.counters.DAGCounter	NUM_SUCCEEDED_TASKS	54
org.apache.tez.common.counters.DAGCounter	TOTAL_LAUNCHED_TASKS	54
org.apache.tez.common.counters.DAGCounter	DATA_LOCAL_TASKS	54
org.apache.tez.common.counters.DAGCounter	AM_CPU_MILLISECONDS	5890
org.apache.tez.common.counters.DAGCounter	AM_GC_TIME_MILLIS	22
org.apache.tez.common.counters.FileSystem...	FILE_BYTES_WRITTEN	1074416910
org.apache.tez.common.counters.FileSystem...	S3A_BYTES_READ	6742239795
org.apache.tez.common.counters.FileSystem...	S3A_READ_OPS	1124
org.apache.tez.common.counters.TaskCounter	TASK_DURATION_MILLIS	759357
org.apache.tez.common.counters.TaskCounter	INPUT_RECORDS_PROCESSED	146519
org.apache.tez.common.counters.TaskCounter	INPUT_SPLIT_LENGTH_BYTES	12387638515
HIVE	CREATED_FILES	36
HIVE	RECORDS_IN_Map_1	150000000
HIVE	RECORDS_OUT_0	6003115
HIVE	RECORDS_OUT_OPERATOR_FIL_5	6003115
HIVE	RECORDS_OUT_OPERATOR_FS_7	6003115
HIVE	RECORDS_OUT_OPERATOR_SEL_6	6003115
HIVE	RECORDS_OUT_OPERATOR_TS_0	150000000

8. Click DAG Configurations to see the Tez configuration details for a query that has a DAG associated with it.
   The following image shows the **DAG Configurations** tab on the Hue web interface:



Viewing Impala query details

You can view Impala query details, query plan, execution summary, and query metrics on the new Impala Queries tab on the Job Browser page in Hue, and use this information to tune and optimize your queries. You can also view Impala query profiles on the Impala tab.

Viewing Impala query history

The Impala Queries tab on the Job Browser page in Hue displays all the queries that were run from various query interfaces, such as Impala-shell, Impyla, Hue, and other JDBC BI clients and tools.

About this task

Only Query Processor Administrators can view historical queries of all users to access history of queries run in that Impala Virtual Warehouse from the Hue Job Browser. Non-admin users can only view only their own queries.

Procedure

1. Log in to the Hue web interface.
2. Click Jobs from the left assist panel.
   The **Job Browser** page is displayed.
3. Click Queries.
   The Impala queries that were run for the past seven days are displayed. You can select the time period for which you want to view the historical data.
   You can also search using the query ID, sort queries by various parameters such as duration, peak memory, and so on, and filter queries by their status.

Related Information

Adding Query Store Administrator users on Data Hub clusters

Viewing Impala query information

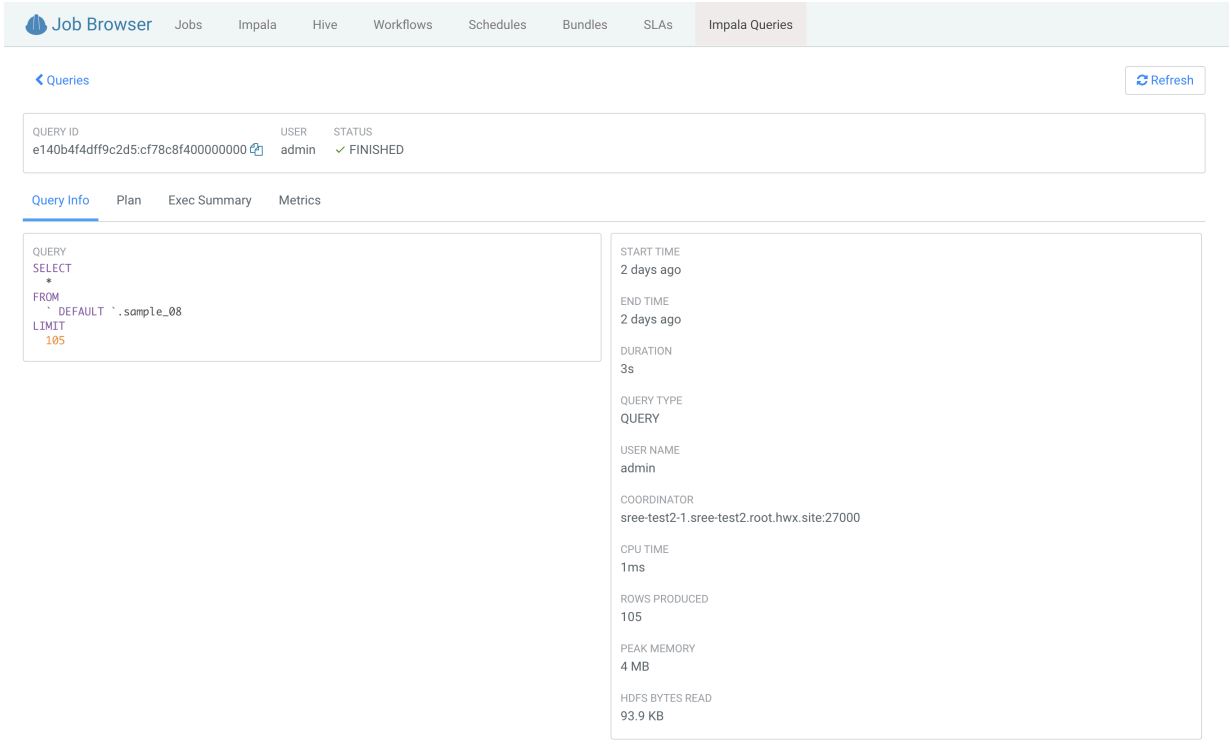
The Query Info tab in Hue provides information such as, the Impala query ID, the user who executed the query, the start time, the end time, the total time taken to execute the query, the coordinator that received the query, CPU time, rows produced, peak memory, and HDFS bytes read.

Procedure

1. Log in to the Hue web interface.
2. Click Jobs from the left assist panel.

   The **Job Browser** page is displayed.
3. Go to the **Impala Queries** tab and click on the query for which you want to view the query details.

   The following image shows the **Query Info** tab on the Hue web interface:



Viewing the Impala query execution plan

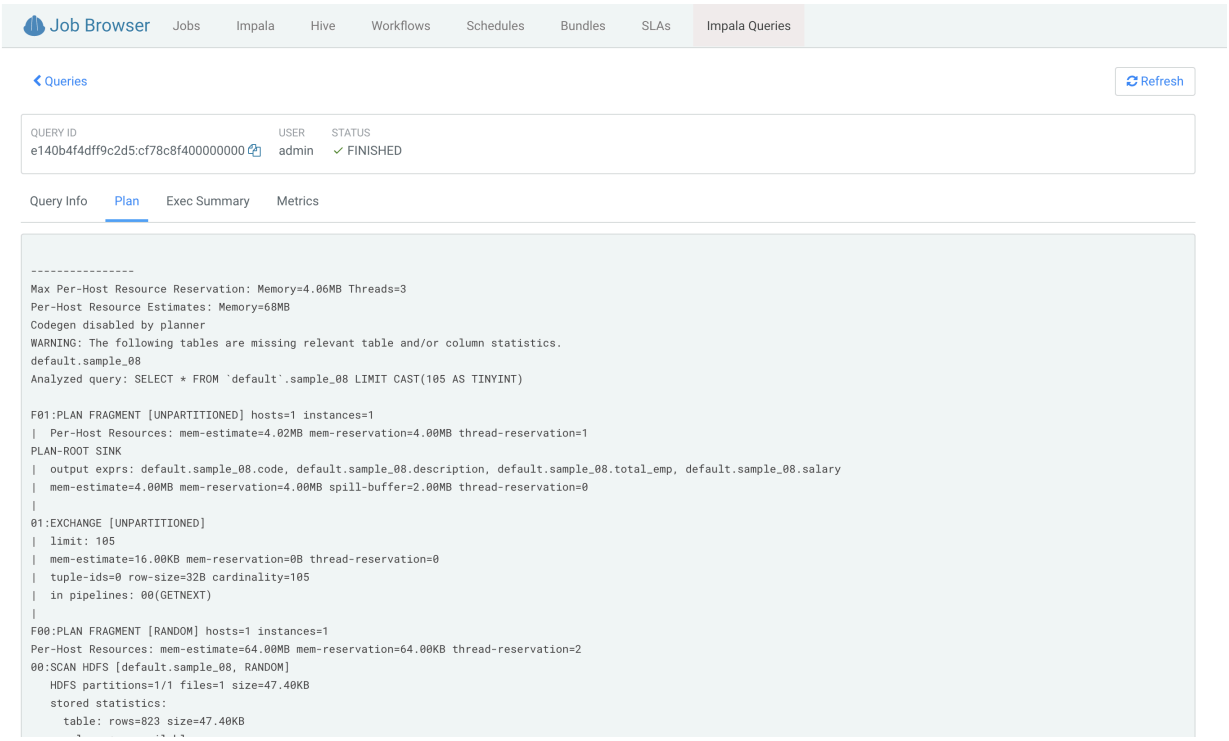
The query execution plan in Hue provides details on how the query will be executed, the operators involved, and other information before the query is submitted to the Impala engine.

Procedure

1. Log in to the Hue web interface.
2. Click Jobs from the left assist panel.

   The **Job Browser** page is displayed.

3. Go to the **Impala Queries** tab and click on the query for which you want to view the execution plan.
   The following image shows the **Plan** tab on the Hue web interface:



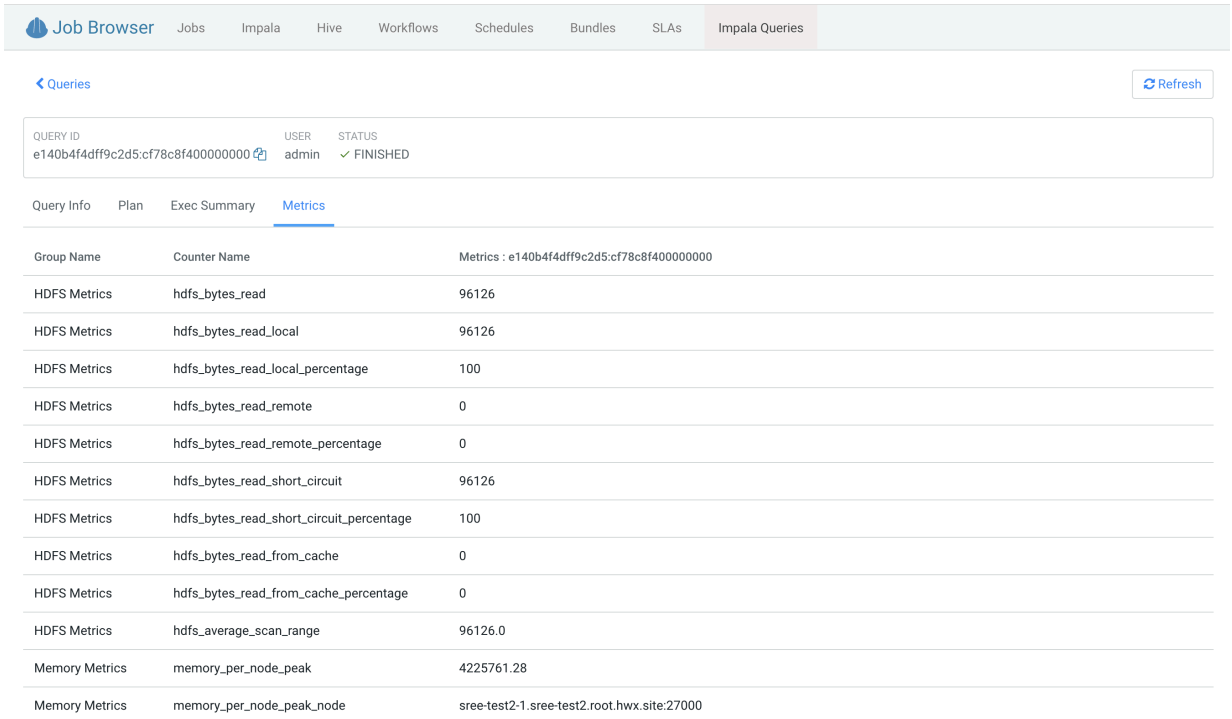
Viewing the Impala query metrics

You can view detailed, aggregated metrics for various counters such as hdfs_bytes_read, memory_per_node_peak, thread_cpu_time, and so on, on the Metrics tab in Hue.

Procedure

1. Log in to the Hue web interface.
2. Click Jobs from the left assist panel.
   The **Job Browser** page is displayed.

3. Go to the **Impala Queries** tab and click on the query for which you want to view the query metrics.
   The following image shows the **Metrics** tab on the Hue web interface:



Group Name	Counter Name	Metrics : e140b4f4dff9c2d5:cf78c8f400000000
HDFS Metrics	hdfs_bytes_read	96126
HDFS Metrics	hdfs_bytes_read_local	96126
HDFS Metrics	hdfs_bytes_read_local_percentage	100
HDFS Metrics	hdfs_bytes_read_remote	0
HDFS Metrics	hdfs_bytes_read_remote_percentage	0
HDFS Metrics	hdfs_bytes_read_short_circuit	96126
HDFS Metrics	hdfs_bytes_read_short_circuit_percentage	100
HDFS Metrics	hdfs_bytes_read_from_cache	0
HDFS Metrics	hdfs_bytes_read_from_cache_percentage	0
HDFS Metrics	hdfs_average_scan_range	96126.0
Memory Metrics	memory_per_node_peak	4225761.28
Memory Metrics	memory_per_node_peak_node	sree-test2-1.sree-test2.root.hwx.site:27000

Viewing Impala profiles in Hue

When Impala executes any query, it captures the runtime details of the execution in a query profile. You can now view the Impala query profile from Hue's Job Browser page.

About this task

Query Processor Administrators can view query details of all users unlike the non-admin users who can view query details only for their queries. This enables the Query Processor Administrators to provide a comprehensive report on past queries and running queries.

Procedure

1. Go to the Cloudera Data Warehouse (CDW) web interface and open Hue from your Impala Virtual Warehouse.
2. Click Jobs on the left-assist panel to go to the **Job Browser** page and then click on the Impala tab.
3. Click on the query for which you want to view the Impala query profile.
   The query execution details are displayed.
4. Click on the Profile tab.

Terminating Hive queries

If a query is running for longer than expected, or you have accidentally triggered it, then you can stop the query to free up the resources. Hue also allows you to stop multiple queries at once.

About this task

Note: This feature is available only for Hive queries. Only admin users or Hue superusers can stop running queries.

Procedure

1. Log in to the Hue web interface.
2. Click Jobs from the left assist panel.

   The **Job Browser** page is displayed.
3. Go to the **Queries** tab.

   A list of queries that were run is displayed.
4. Select the queries that you want to stop and click Kill.

Comparing Hive and Impala queries in Hue

You can compare two queries to know how each query is performing in terms of speed and cost-effectiveness. Hue compares various aspects of the two queries, based on which you can identify what changed between the executions of those two queries, and you can debug performance-related issues between different runs of the same query.

About this task

The query comparison report provides you a detailed side-by-side comparison of your queries.

For Hive queries, it includes recommendations for optimizing each query, metadata about the queries, visual explain for each query, query timeline, query configuration, Directed Acyclic Graph (DAG) information, DAG flows, DAG swimlanes, DAG counters, and DAG configurations.

For Impala queries, the query comparison report includes query details, execution plan details, and the aggregated metrics for both the queries and provides a variance between the two.

Procedure

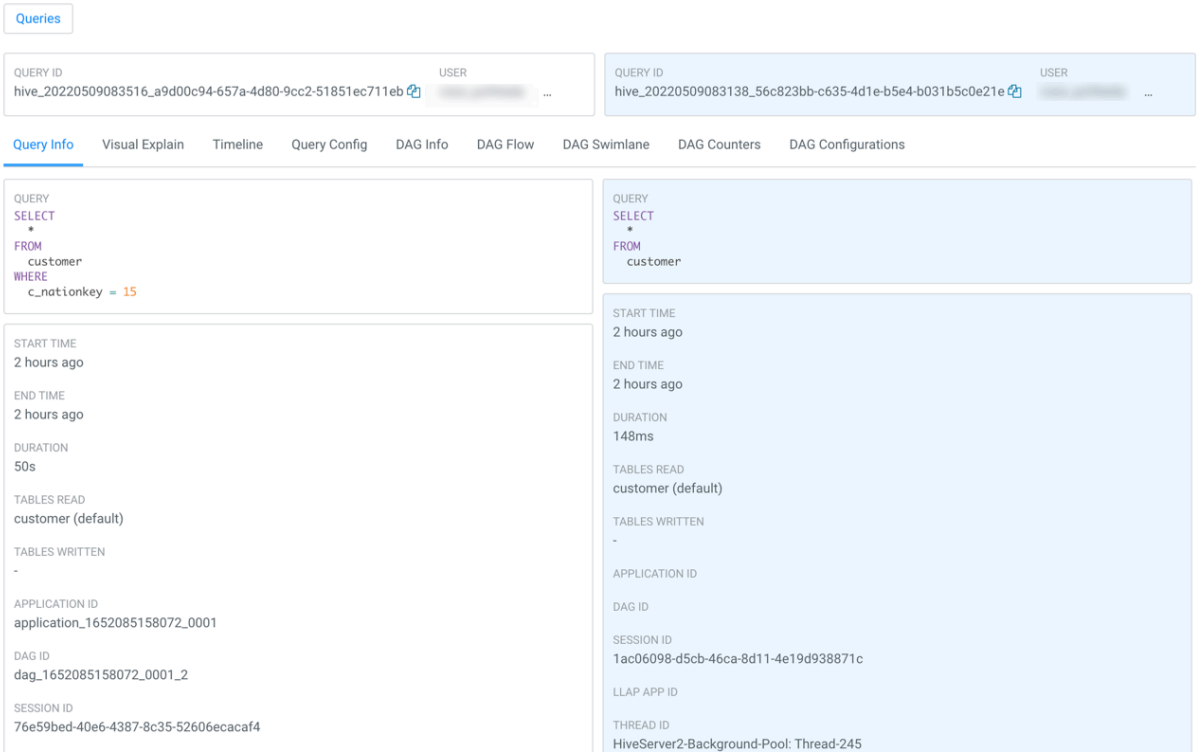
1. Log in to the Hue web interface.
2. Click Jobs from the left assist panel.

   The **Job Browser** page is displayed.
3. Go to the **Queries** tab.

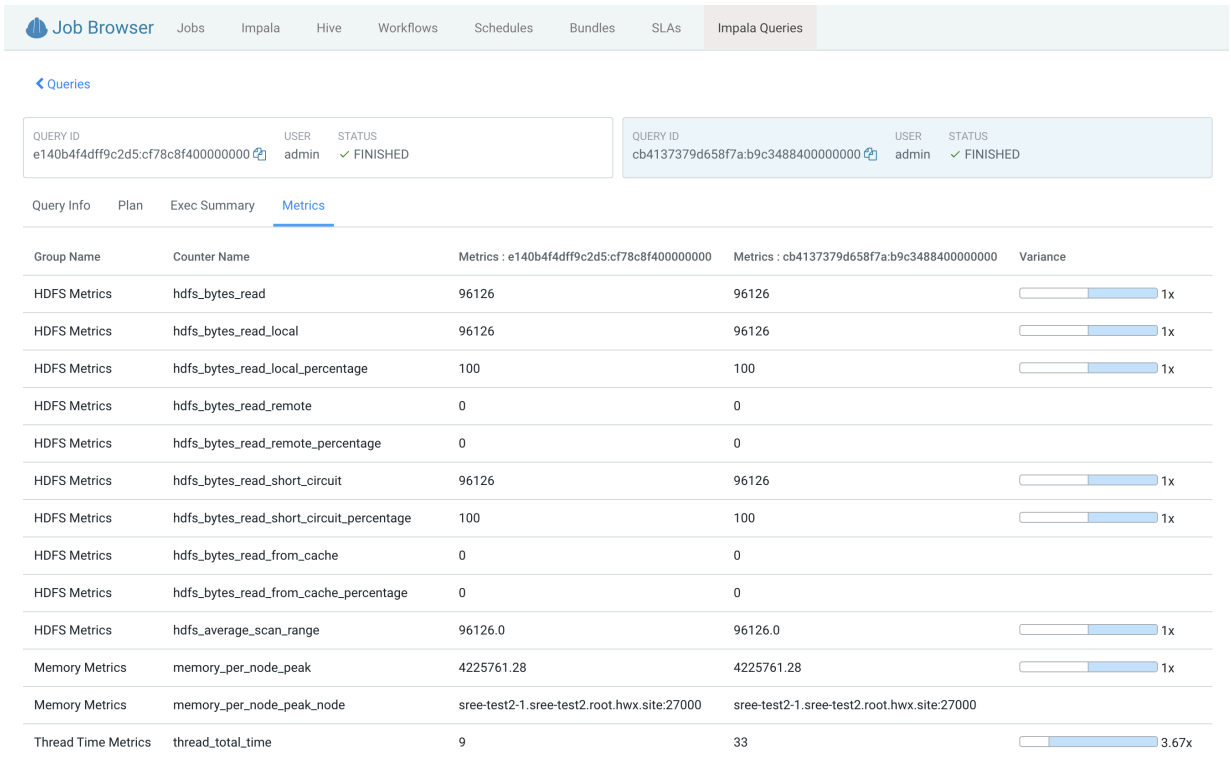
   A list of queries that were run is displayed.

4. Select the two queries you want to compare and click Compare.

   Query comparison report for Hive queries:



   Query comparison report for Impala queries:



Starting the SQL AI Assistant in 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. You can also generate comments and insert them into your queries to improve readability.

About this task

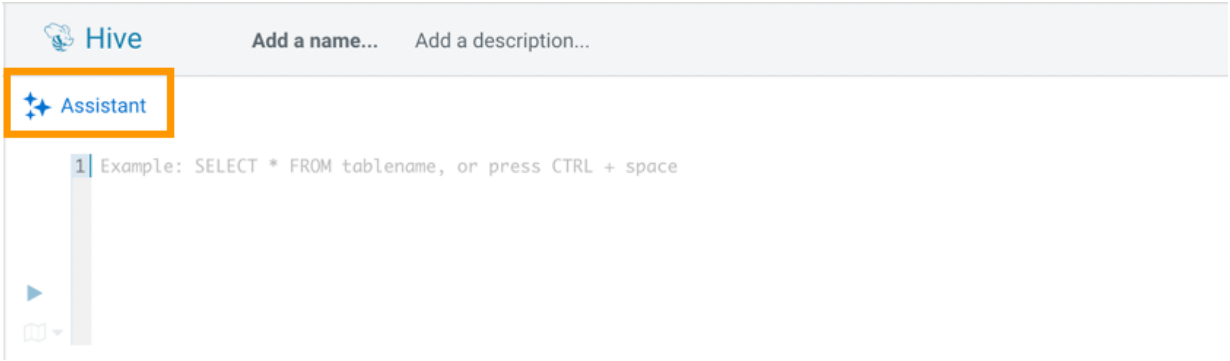
Attention: The SQL AI Assistant operates only on the database that you have selected in the Hue editor, and not necessarily on the one that is displayed on the left-assist bar.

Before you begin

Your administrator must have configured and set up the required infrastructure for you to use the SQL AI Assistant. See About setting up the Hue SQL AI Assistant.

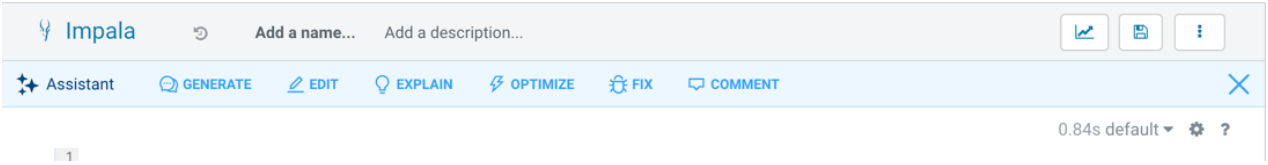
Procedure

1. Log in to the Hue web interface.
2. Click Assistant on the Hue SQL editor:



Results

The following options are displayed:



Generating SQL from natural language in Hue

The SQL AI Assistant in Cloudera Data Warehouse (CDW) helps you to generate SQL queries by entering a prompt in natural language. You can then insert the generated SQL in the Hue SQL editor and run it as usual.

Before you begin

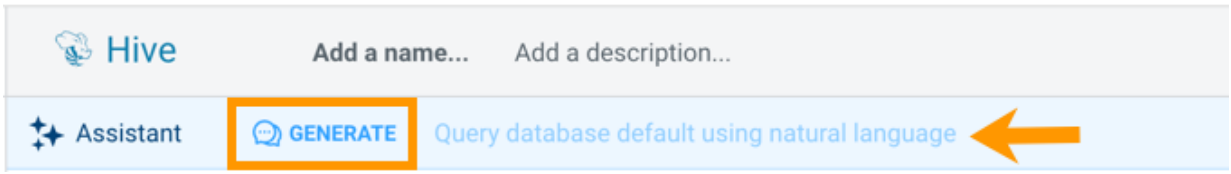
Your administrator must have configured and set up the required infrastructure for you to use the SQL AI Assistant. See About setting up the Hue SQL AI Assistant.

Procedure

1. Log in to the Hue web interface.

2. Click Assistant on the Hue SQL editor:

3. Click GENERATE.



A SQL query is generated based on your input prompt. Click Insert to insert the query into the editor and run it.

Editing the query in natural language in Hue

You can edit a query in natural language and generate it again to fine-tune your results using the Hue SQL AI Assistant. You can then insert the SQL query into the editor and run it as usual.

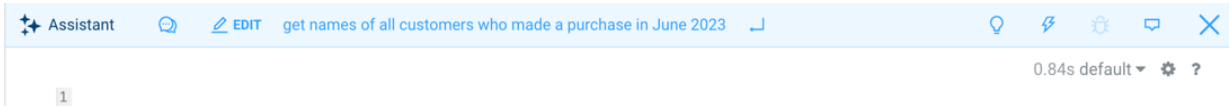
Before you begin

Your administrator must have configured and set up the required infrastructure for you to use the SQL AI Assistant. See About setting up the Hue SQL AI Assistant.

Procedure

1. Log in to the Hue web interface.
2. Click Assistant on the Hue SQL editor:

3. Click EDIT.



Select the query from the list and edit it as needed. Then press enter to regenerate the SQL.

Getting an explanation of a SQL query in natural language in Hue

You can use the SQL AI Assistant to breakdown and understand a complex SQL query in natural language.

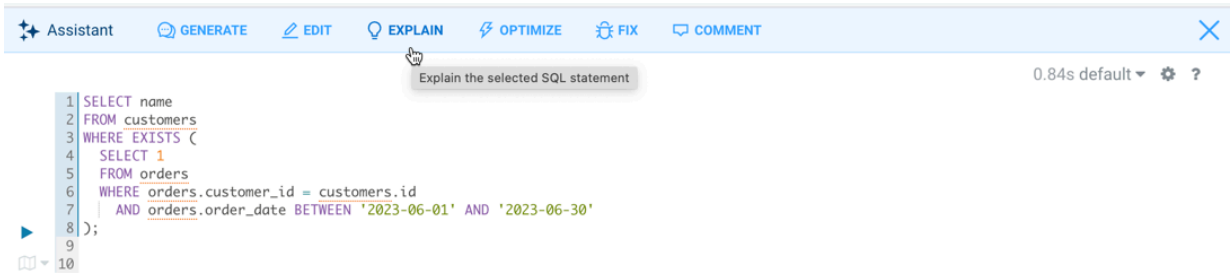
Before you begin

Your administrator must have configured and set up the required infrastructure for you to use the SQL AI Assistant. See About setting up the Hue SQL AI Assistant.

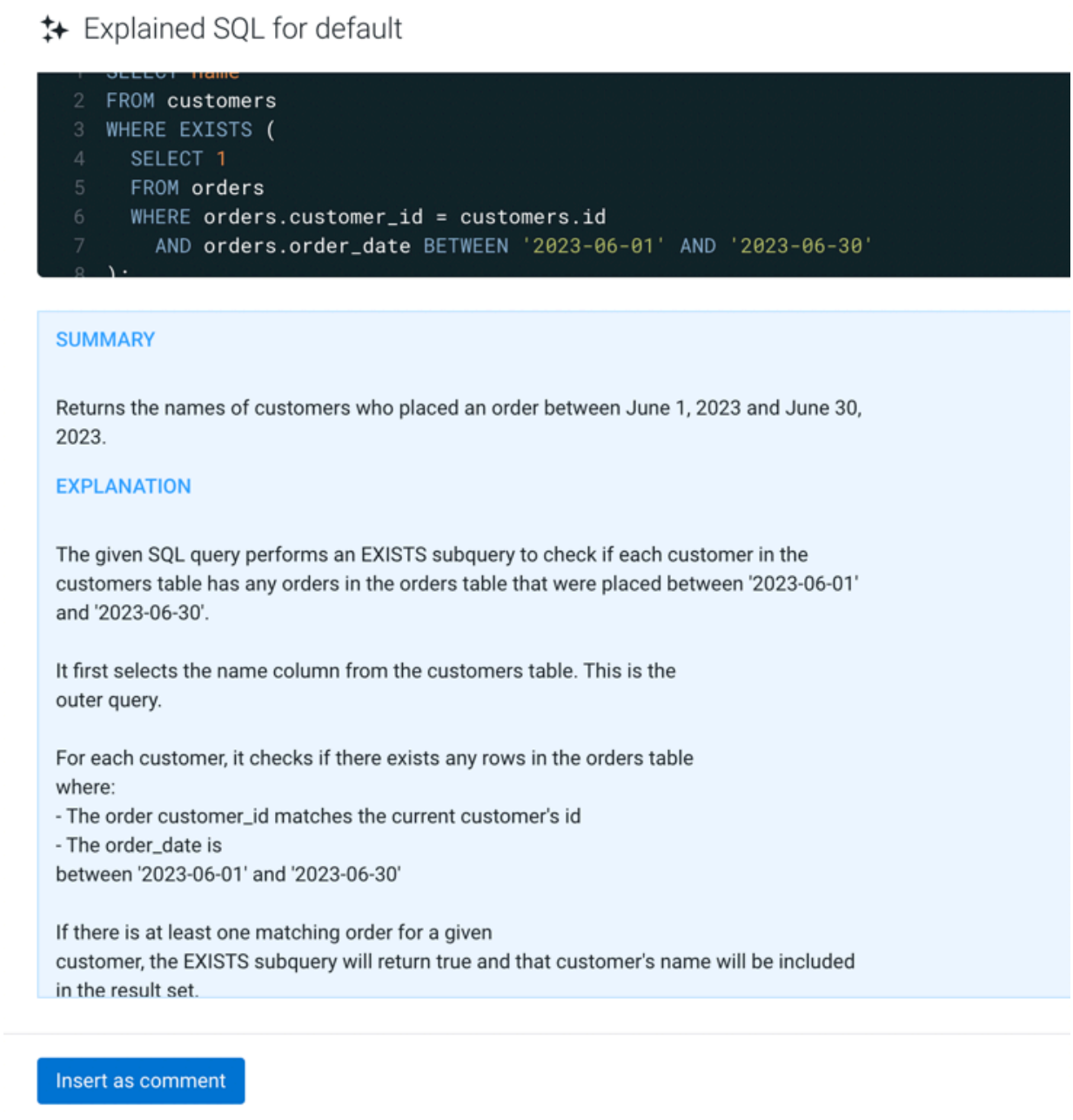
Procedure

1. Log in to the Hue web interface.
2. Insert a SQL query in the Hue editor that you would like to understand in natural language.
3. Click Assistant on the Hue SQL editor:

4. Click EXPLAIN.



The LLM generates the explanation of the SQL query.



Optimizing a query in Hue

You can use the SQL AI Assistant to optimize a SQL query. Hue identifies the issues in the source query, optimizes it, and provides the optimized version of the SQL query. Hue also summarizes the issues and how it optimized the query in natural language.

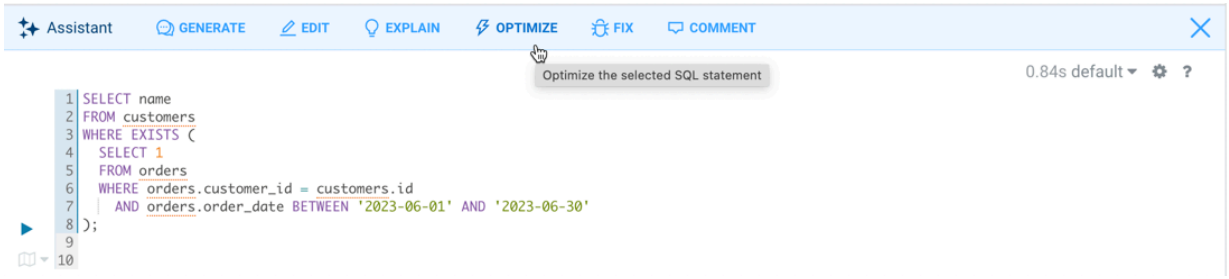
Before you begin

Your administrator must have configured and set up the required infrastructure for you to use the SQL AI Assistant. See About setting up the Hue SQL AI Assistant.

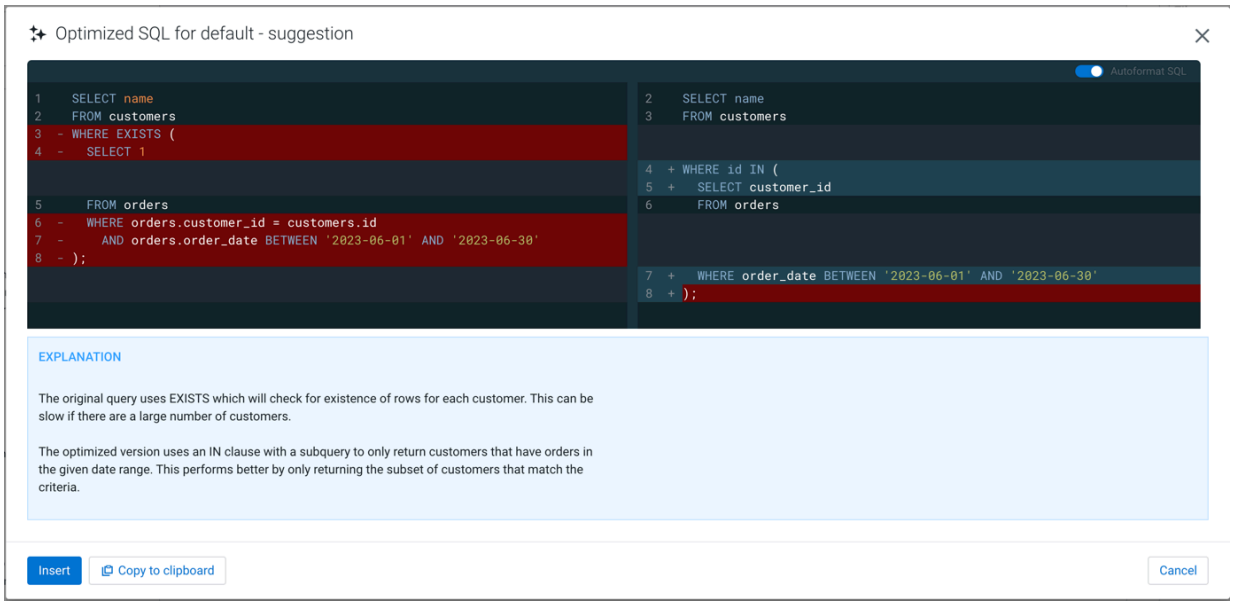
Procedure

1. Log in to the Hue web interface.
2. Insert a SQL query in the Hue editor that you would like to optimize.
3. Click Assistant on the Hue SQL editor:

4. Click OPTIMIZE.



Hue displays the original and the optimized SQL query side-by-side. It also provides an explanation of the issues in the original query and how it was optimized.



Fixing a query in Hue

You can use the SQL AI Assistant to fix a broken SQL query. Hue identifies the issues in SQL syntax and provides the corrected version.

Before you begin

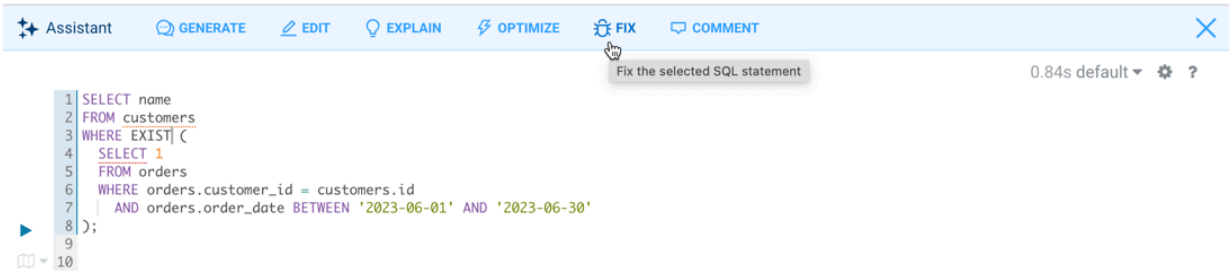
Your administrator must have configured and set up the required infrastructure for you to use the SQL AI Assistant. See About setting up the Hue SQL AI Assistant.

Procedure

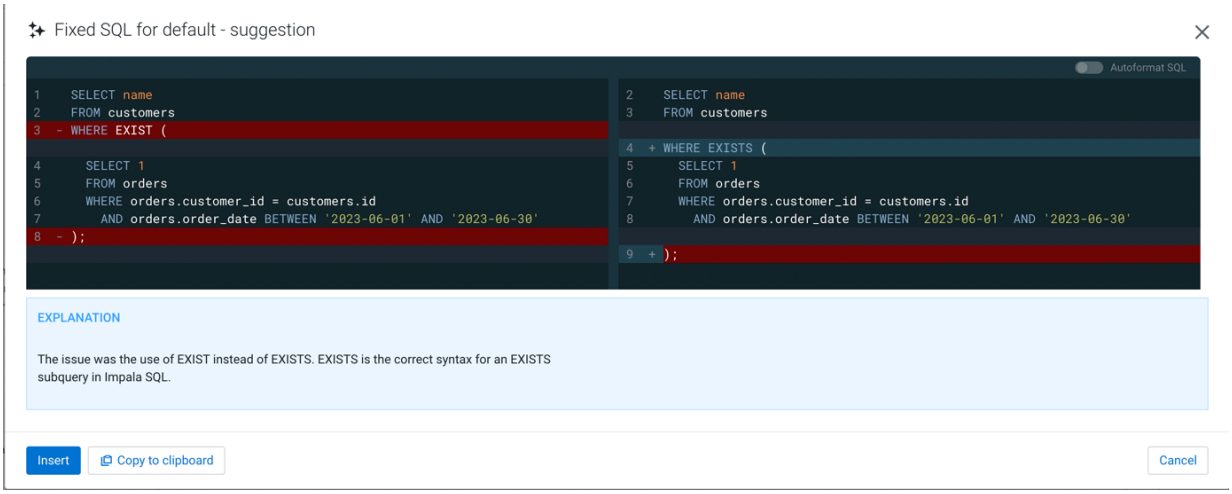
1. Log in to the Hue web interface.
2. Insert a SQL query in the Hue editor that you would like to fix.

3. Click Assistant on the Hue SQL editor:

4. Click FIX.



Hue displays the original and the fixed SQL query in a side-by-side comparison.



Click Insert to insert the fixed query in the Hue editor and run it.

Generating a comment for a query in Hue

The SQL AI Assistant can generate a comment explaining what SQL query does. You can insert it into the query to improve readability.

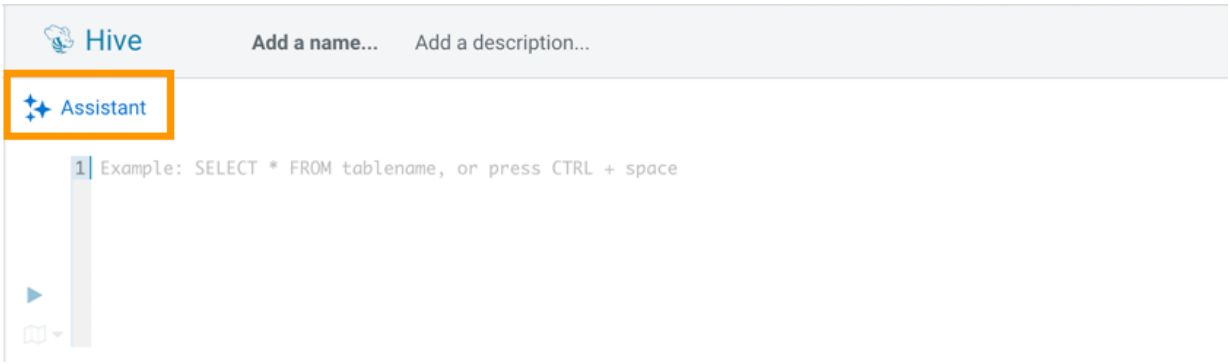
Before you begin

Your administrator must have configured and set up the required infrastructure for you to use the SQL AI Assistant. See About setting up the Hue SQL AI Assistant.

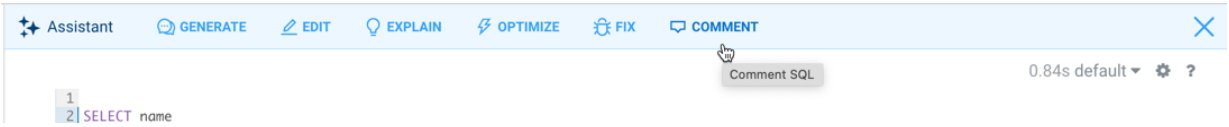
Procedure

1. Log in to the Hue web interface.
2. Insert a SQL query in the Hue editor for which you want to generate a comment.

3. Click Assistant on the Hue SQL editor:



4. Click COMMENT.



The SQL AI Assistant generates a detailed comment for the input SQL query.

Click Insert to insert the comment into the query.

Enabling stored procedures for Hive in DataHub

To create, edit, and drop procedures and functions that are written in Hive Hybrid Procedural SQL (HPL/SQL) using the Hue query editor, you must enable the hplsql option in the Hue Advanced Configuration Snippet.

About this task

Note: Hue enables you to switch between Hive and HPL/SQL interpreters. By default, the regular Hive interpreter is enabled when you add the Hue service to your cluster. To enable the HPL/SQL interpreter, you must update Hue's Advanced Configuration Snippet in Cloudera Manager. However, updating Hue's Advanced Configuration Snippet overrides the default configuration. Therefore, to use both Hive and HPL/SQL interpreters, you must enable both by updating Hue's Advanced Configuration Snippet.

Procedure

1. Log in to the CDP web interface as an Administrator.
2. Go to Management Console Environments , select your environment and go to the Data Lake tab.
3. Click on the CM URL to open Cloudera Manager.
4. Go to Clusters Hue Configuration and add the following lines in the Hue Service Advanced Configuration Snippet (Safety Valve) for hue_safety_valve.ini field:

```
[notebook]
  [[interpreters]]
    [[[hive]]]
      name=Hive
      interface=hiveserver2
    [[[hplsql]]]
      name=Hplsql
      interface=hiveserver2
```

5. Click Save Changes.
6. Restart the Hue service.
7. Go to Clusters Hive on Tez Configuration and add the following property name and its value in the Hive Service Advanced Configuration Snippet (Safety Valve) for hive-site.xml field:

   Name: hive.security.authorization.sqlstd.confwhitelist.append

   Value: QUERY_EXECUTOR|HPLSQL
8. Click Save Changes.
9. Restart the Hive on Tez service.

How to run a stored procedure from Hue in DataHub

HPL/SQL allows you to implement business logic using variables, expressions, flow-of-control statements, and iterations. HPL/SQL makes SQL-on-Hadoop more dynamic. You can leverage your existing procedural SQL skills, and use functions and statements to make your typical ETL development more productive. Hue provides a smart interface to run stored procedures.

Note: This feature is available only for Hive queries.

The following example creates a procedure and returns records by passing a cursor:

```
print 'Hello world';/
CREATE PROCEDURE greet(name STRING)
BEGIN
  PRINT 'Hello ' || name;
END;/
CREATE PROCEDURE even(cur OUT SYS_REFCURSOR)
BEGIN
  OPEN cur FOR
SELECT n FROM NUMBERS
WHERE MOD(n, 2) == 0;
END;/
```

```
CREATE PROCEDURE set_message(IN name STRING, OUT result STRING)
BEGIN
 SET result = 'Hello, ' || name || '!';
END;
-- Call the procedure and print the results
DECLARE str STRING;
CALL set_message('world', str);
PRINT str;
```

Attention: In the hplsql mode, you must terminate the commands using the forward slash (/). The semicolon (;) is used throughout procedure declarations and can no longer be relied upon to terminate a query in the editor.

Note: HPL/SQL does not support all types of Hive statements, such as JOIN or EXPLAIN. Refer to the HPL/SQL Reference for more information.

Related Information

Enabling stored procedures for Hive in DataHub

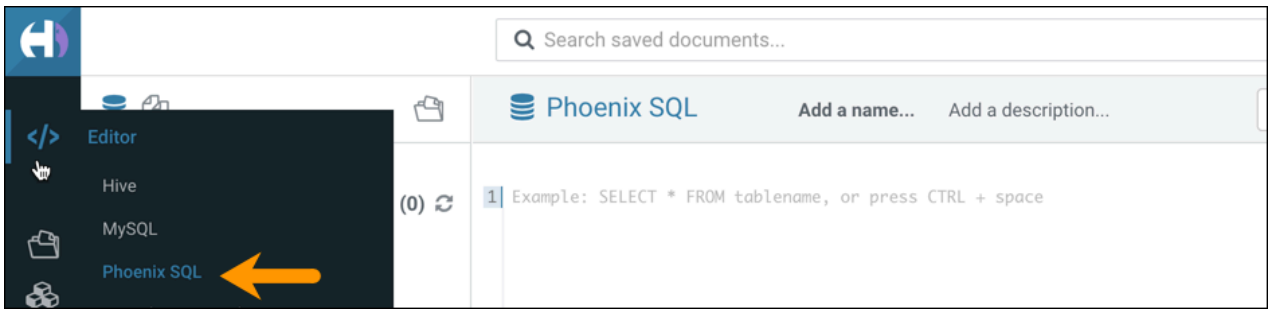
Using SQL to query HBase from Hue

Hue provides a simple SQL interface to create and manipulate SQL tables that are stored in HBase, and define and manipulate views on HBase tables using Apache Phoenix in addition to HBase shell and database API.

Cloudera does not recommend manipulating Phoenix tables from HBase as this can lead to data loss.

The SQL connector is shipped with Hue so that you do not have to download and configure it yourself.

Following are some examples to create and manipulate the Phoenix SQL tables from the Hue editor:



Creating a table

```
CREATE TABLE IF NOT EXISTS Company (company_id INTEGER PRIMARY KEY, name VAR
CHAR(225));
```

Upserting values in the table

```
UPSERT INTO Company VALUES(1, 'Cloudera');
UPSERT INTO Company VALUES(2, 'Apache');
```

Querying the table

```
SELECT * FROM Company;
```

Deleting a record

```
DELETE FROM Company WHERE COMPANY_ID=1;
```

Dropping the table

```
DROP TABLE Company;
```

Related Information

Apache Phoenix SQL command reference

Apache Phoenix Grammar

Querying existing HBase tables

To use SQL for querying data from existing HBase tables, you must create a view in Phoenix pointing to the HBase table.

To map the existing tables to the views, run the following statement using the Phoenix editor on the Hue web interface:

```
CREATE VIEW if not exists "[***HBASE-TABLE-NAME***]" ( pk VARCHAR PRIMARY KE
Y, val VARCHAR );
```

Related Information

Mapping a Phoenix table to an existing HBase table

Enabling the SQL editor autocompleter

Autocompleter provides finely tuned SQL suggestions for Hive and Impala dialects while you enter queries into the editor window. See Brand new Autocompleter for Hive and Impala in the Hue blog.

About this task

Autocompleter is enabled by default. To manually enable or disable it, open the editor configuration panel and edit settings as follows:

Procedure

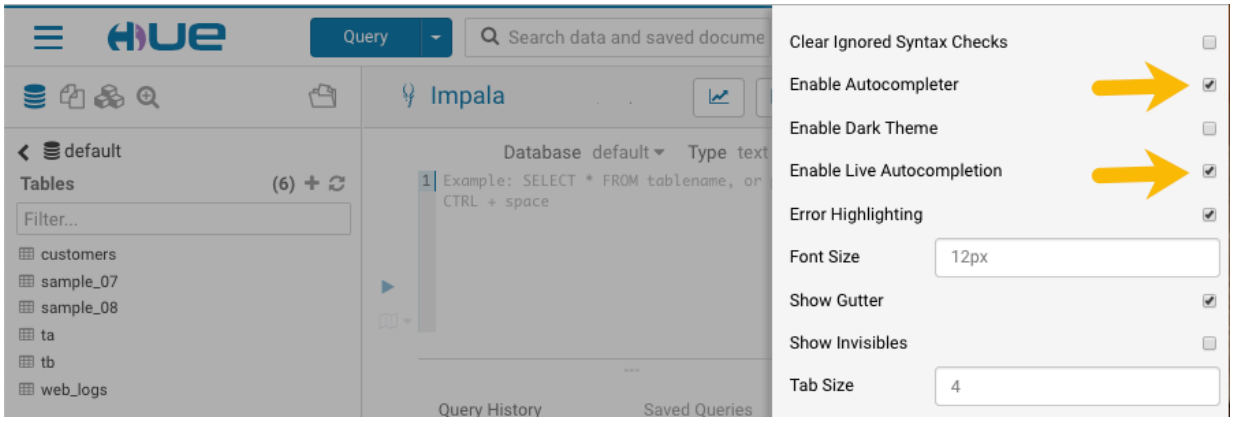
1. Log in to Hue and go to either the Hive or Impala editor.
2. Place your cursor in the editor window and then use one of the following keyboard shortcuts to open the editor configuration panel:
   - On a Mac system, use the Command key followed by a hyphen and then a comma:

     Command-,
   - On a Windows system, use the Ctrl key followed by a hyphen and then a comma:

     Ctrl-,

Tip: Type a question mark (?) anywhere but in the active editor window to open a menu of editor keyboard shortcuts.

3. To enable autocompletion, check the box adjacent to Enable Autocompleter. When you check Enable Autocompleter, Enable Live Autocompletion is automatically enabled as well. Place your cursor in the editor window to close the configuration panel.



4. To disable autocompletion:
   - Uncheck Enable Live Autocompletion but leave Enable Autocompleter checked, and then place your cursor in the editor window to close the configuration panel. This disables live autocompletion, but if you want to use autocompletion while building your queries in the editor, enter the following key stroke sequence to activate autocompletion: Ctrl + Space Key
   - Uncheck both Enable Autocompleter and Enable Live Autocompletion, and then click in the editor to close the configuration panel. This disables all autocompletion functionality.

Rerunning a query from the Job Browser page

You may need to reconfigure the query parameters post a query execution, based on the results. You can modify and rerun your historical queries from the Job Browser page.

About this task

Note: Only Query Processor Administrators can view and rerun historical queries of all users from the **Job Browser** page. Non-admin users can view and rerun only their queries.

Procedure

1. Log in to the Hue web interface.
2. Select Jobs from the left assist panel and click Queries on the **Job Browser** page.
3. Click on the query that you want to rerun.

   The query details are displayed.
4. Click Re Execute.

   You are redirected to the query editor where you can make the required changes to the pre-populated query.
5. Click to submit the query.

Using governance-based data discovery

Hue can use the metadata tagging, indexing, and search features available in Apache Atlas data management. After integrating Hue with Atlas, classifications and indexed entities can be accessed and viewed in Hue. This topic shows you how to use metadata classifications in Hue.

Integration between Hue and Atlas is enabled by default, but if your administrator has disabled it, it must be re-enabled before you can use governance-based data discovery.

You can create tags to classify your data both from Atlas and Hue.

Searching metadata tags

The SQL Editor in Hue provides a search text box where you can search on the metadata tags or classifications that are associated with your databases, tables, and columns.

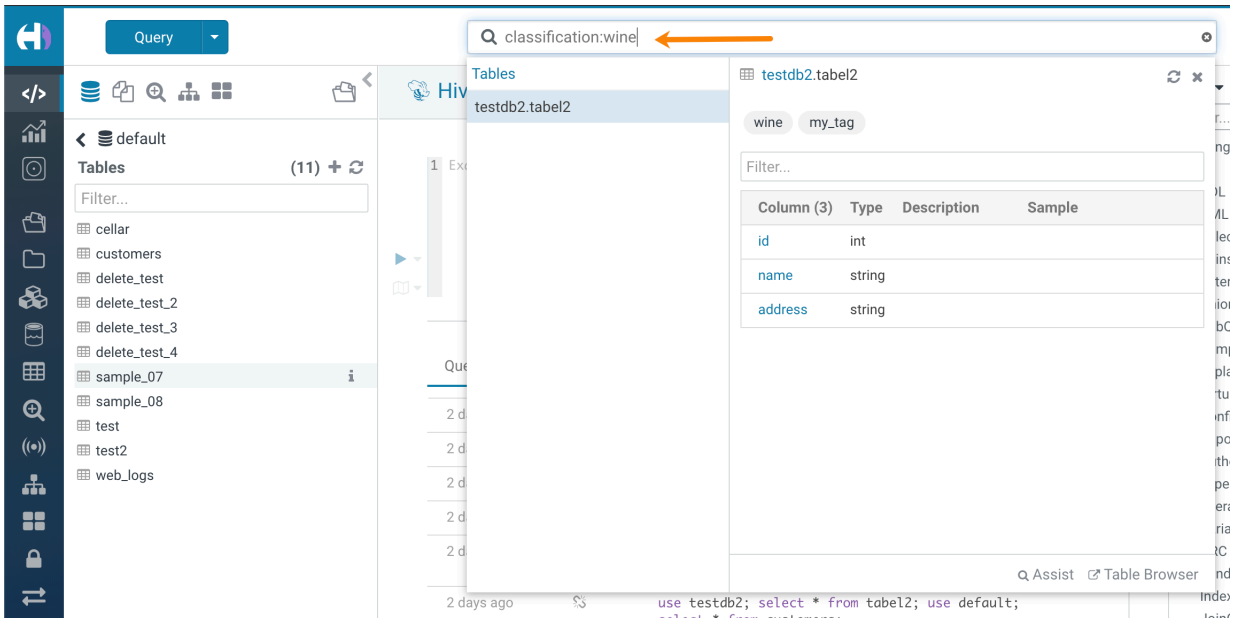
About this task

You can search for tags or classifications in either the Hive or the Impala editors.

Note: On clusters that use Apache Ranger for role-based access control, the Search mechanism does not display counts of popular values. Ranger ensures that Hue users can view only entities to which their user role (as configured and managed by Ranger) has been granted specific permissions.

Procedure

1. Go to Query Editor Impala or Hive.
2. To locate the tags or classifications in Apache Atlas, in the metadata search box located just to the right of the Query drop-down menu, type a tag: or classification: facet followed by its name. For example, type classification: wine as shown in the following image:



After you type the search facet and the tag or classification name in the search box, the *<database>.<table>* where the tag or classification is found is returned. Click the *<database>.<table>* to view the tags and classifications that have been defined for it.

Using Amazon S3 with Hue

Hue can read to and write to an Amazon S3 bucket.

Note:

You can use the S3 browser to store and retrieve data from Amazon S3 buckets into Hue and vice versa. If your CDP cluster is secured using Knox, then the Hue users will be able to access the S3 folders as per the access rules set on the individual folders. The access permissions are governed by IDBroker.

In CDP, you can enable the S3 File Browser in Hue in the following three ways:

- With IDBroker
- Without IDBroker
- With Ranger Authorization Service

Note: If you have enabled RAZ while registering your AWS environment with CDP, then Hue uses RAZ as the default mechanism for enabling the S3 File Browser.

Enabling S3 browser for Hue configured with IDBroker

You can access the S3 buckets from Hue to upload files and tables to S3 and import CSV files as tables directly in Hue by enabling the S3 browser on the Hue user interface.

Before you begin

If you have set up authentication using Knox IDBroker on your cluster, then Hue automatically detects and uses the IDBroker mappings from your cluster's core-site.xml file. Verify that the following property is present in the core-site.xml file:

```
<property>
  <name>fs.s3a.ext.cab.address</name>
  <value>https://<idbrokerurl>:8444/gateway</value>
</property>
```

This property is automatically appended to the core-site.xml file when you enable Knox IDBroker on your cluster.

Procedure

1. Sign in to Cloudera Manager as an Administrator.
2. Go to Clusters Hue service Configuration .
3. Enter the following in the Hue Server Advanced Configuration Snippet (Safety Valve) for hue_safety_valve_server.ini field:

   ```
   [desktop]
   # Remove the file browser from the blacklisted apps.
   # Tweak the app_blacklist property to suit your app configuration.
   app_blacklist=spark,zookeeper,hive,hbase,search,oozie,jobsub,pig,sqoop,sec
   urity
   [aws]
   has_iam_detection=true
   [[aws_accounts]]
   [[[default]]]
   region=[***AWS-REGION***]
   # Set a particular S3 bucket as the default
   [filebrowser]
   remote_storage_home=s3a://[***S3-BUCKET-NAME***]
   ```

   The custom configuration is stored in the hue_safety_valve_server.ini file.
4. Click Save Changes.
5. Restart the Hue service.

   The S3 file browser icon appears on the left Assist panel as well as on the left navigation bar on the Hue web interface.

What to do next

You must manually grant the following application permission to non-admin users and groups for them to be able to view and access S3 File Browser in Hue: filebrowser.s3_access:Access to S3 from filebrowser and filepicker..

You must also add the CDP users and groups to IAM role mappings.

Related Information

Granting permission to access S3 and ABFS File Browser in Hue

Adding CDP user/group to IAM role mappings

Enabling S3 browser for Hue configured without IDBroker

You can access the S3 buckets from Hue to upload files and tables to S3 and import CSV files as tables directly in Hue by enabling the S3 browser. You must specify the AWS access key and secret access key information along with the AWS region in the Hue Server Advanced Configuration Snippet if you have not set up authentication using Knox IDBroker.

Procedure

1. Sign in to Cloudera Manager as an Administrator.
2. Go to Clusters Hue service Configuration .
3. Enter the following in the Hue Server Advanced Configuration Snippet (Safety Valve) for hue_safety_valve_server.ini field:

```
[desktop]
# Remove the file browser from the blacklisted apps.
# Tweak the app_blacklist property to suit your app configuration.
app_blacklist=spark,zookeeper,hive,hbase,search,oozie,jobsub,pig,sqoop,sec
urity
[aws]
[[aws_accounts]]
[[[default]]]
access_key_id=[***AWS-ACCESS-KEY***]
secret_access_key=[***SECRET-ACCESS-KEY***]
region=[***AWS-REGION***]

# Set a particular S3 bucket as the default
[filebrowser]
remote_storage_home=s3a://[***S3-BUCKET-NAME***]
```

   The custom configuration is stored in the hue_safety_valve_server.ini file.
4. Click Save Changes.
5. Restart the Hue service.

   The S3 file browser icon appears on the left Assist panel as well as on the left navigation bar on the Hue web interface.

What to do next

You must manually grant the following application permission to non-admin users and groups for them to be able to view and access S3 File Browser in Hue: filebrowser.s3_access:Access to S3 from filebrowser and filepicker.

Related Information

Granting permission to access S3 and ABFS File Browser in Hue

Enabling S3 File Browser for Hue with RAZ in DataHub

Hue offers you the capability to browse S3 buckets, upload files to S3, and create tables by importing files from S3. With RAZ, you can grant fine-grained access to per-user home directories and other directories within the S3 bucket using Ranger policies.

About this task

Hue administrators can create home directories for users by selecting the Create home directory option on the user's profile in Hue. If fine-grained authorization is enabled to access S3 buckets, then user home directories are automatically created when a user logs into Hue. You can disable automatic creation of user directories.

Before you begin

1. Register an AWS environment with the Enable Ranger authorization for AWS S3 option enabled. You can use the CDP web interface or the CDP CLI to complete this task.
2. Create a Data Hub cluster with Data Engineering or Data Mart cluster template.
3. Create the following Ranger policies:
   - Hadoop SQL policy (all - database, table, column, all - url).

     Note: You must grant permissions to individual users or groups in these Ranger policies. To grant permissions to all users, you can specify {USER} in the Permission section.
   - S3 (cm_S3) policy (Default: User Home).

     You must specify the bucket name in the S3 Bucket field and the directory path in the Path field of the cm_S3 Ranger policy.
4. Grant appropriate permissions to the users in CDP User Management Service (UMS). For example, EnvironmentUser.

Procedure

1. Log in to the CDP web interface as an Administrator.
2. Go to Management Console Environments , select your environment and go to the Data Lake tab.
3. Click on the CM URL to open Cloudera Manager.
4. Go to Clusters Ranger RAZ service Instances RAZ server Processes and note the value of the fs.s3a.ext.raz.rest.host.url property from the core-site.xml file. You need this to specify the value of the api_url property in the Hue configuration.
5. Go to Clusters Hue service Configuration Hue Service Advanced Configuration Snippet (Safety Valve) for hue_safety_valve.ini and add the following lines:

```
[desktop]
app_blacklist=spark,zookeeper,hbase,impala,search,sqoop,security,pig
[[raz]]
is_enabled=true
api_url=[***fs.s3a.ext.raz.rest.host.url***]
[aws]
has_iam_detection=false

[[aws_accounts]]
[[[default]]]
region=[***AWS-REGION***]
host=s3.[***AWS-REGION***].amazonaws.com
allow_environment_credentials=false

[filebrowser]
remote_storage_home=s3a://[***S3-BUCKET-NAME***]/user
```

   (RAZ-HA) You can specify a comma-separated list of URLs in the api_url property. For example:

```
api_url=https://[***INSTANCE-1***]:6082/,https://[***INSTANCE-2***]:6082/
```

6. Click Save Changes.
7. Restart the Hue service.

Results

S3 File Browser icon is added to the left assist pane on the Hue web interface. When you click on the S3 File Browser icon, you are redirected to your user home directory.

Related Information

Introduction to RAZ on AWS environments

Disabling the automatic creation of user home directories

Using Azure Data Lake Storage Gen2 with Hue

Hue can read to and write to an Azure Data Lake Storage (ADLS) Gen2.

Note:

You can use the ABFS file browser to store and retrieve data from ADLS Gen2 into Hue and vice versa. If your CDP cluster is secured using Knox, then the Hue users will be able to access the ADLS as per the access rules set on the individual components. The access permissions are governed by IDBroker.

Only Hue superusers can view and access the ABFS file browser.

In CDP, you can enable the ABFS File Browser in Hue in the following three ways:

- With IDBroker
- Without IDBroker
- With Ranger Authorization Service

Enabling ABFS file browser for Hue configured with IDBroker

You can directly access the Azure file storage from Hue to upload files and tables to Azure file storage and import CSV files as tables directly in Hue by enabling the ABFS file browser on the Hue user interface.

Before you begin

If you have set up authentication using Knox IDBroker on your cluster, then Hue automatically detects and uses the IDBroker mappings from your cluster's core-site.xml file. Verify that the following property is present in the core-site.xml file:

```
<property>
  <name>fs.azure.ext.cab.address</name>
  <value>https://<idbrokerurl>:8444/gateway</value>
</property>
```

This property is automatically appended to the core-site.xml file when you enable Knox IDBroker on your cluster.

You can also check whether there is a mapped role for the group(s) associated with the authenticated user by going to Management Console Environments Actions Manage Access IDBroker Mappings . If no role is mapped, then you may see the following error: Failed to obtain storage credentials from IDBroker with error: 403 Client Error: Forbidden for url: <url> { "error": "There is no mapped role for the group(s) associated with the authenticated user.", "auth_id": "<user-id>" }. To resolve this, add the user and the Data Access role to the current mappings on the **IDBroker Mappings** page.

Procedure

1. Sign in to Cloudera Manager as an Admin user.
2. Go to Clusters Hue service Configuration .
3. Specify the following in the Hue Server Advanced Configuration Snippet (Safety Valve) for hue_safety_valve_server.ini field:

```
[desktop]
# Remove the file browser from the blacklisted apps.
# Tweak the app_blacklist property to suit your app configuration.
app_blacklist=spark,zookeeper,hive,hbase,search,oozie,jobsub,pig,sqoop,sec
urity
[azure]
[[abfs_clusters]]
[[[default]]]
```

```
  fs_defaultfs=abfs://<container_name>@<storage_name>.dfs.core.windows.net
 webhdfs_url=https://<storage_name>.dfs.core.windows.net/
```

The custom configuration is stored in the hue_safety_valve_server.ini file.

4. (Optional) If the Knox server and Hue are not present on the same host, then add the add the Hue server's FQDN to the trusted origins property in the Hue Server Advanced Configuration Snippet (Safety Valve) for hue_safety_valve_server.ini field:

```
[desktop]
[[session]]
trusted_origins=[***HUE-SERVER-FQDN***]
```

5. Click Save Changes.
6. Restart the Hue service.

   The ABFS file browser icon appears on the left Assist panel as well as on the left navigation bar on the Hue web interface.

What to do next

You must manually grant the following application permission to non-admin users and groups for them to be able to view and access ABFS File Browser in Hue: filebrowser.abfs_access:Access to ABFS from filebrowser and file picker.

You must also add the CDP users and groups to IAM role mappings.

Related Information

Granting permission to access S3 and ABFS File Browser in Hue

Adding CDP user/group to IAM role mappings

Enabling ABFS file browser for Hue configured without IDBroker

You can directly access the Azure file storage from Hue to upload files and tables to Azure file storage and import CSV files as tables directly in Hue by enabling the ABFS file browser on the Hue user interface. You must specify the Azure client ID and the client secret ID information along with the tenant ID in the Hue Server Advanced Configuration Snippet if you have not set up authentication using Knox IDBroker.

Procedure

1. Sign in to Cloudera Manager as an Admin user.
2. Go to Clusters Hue service Configuration .
3. Specify the following in the Hue Server Advanced Configuration Snippet (Safety Valve) for hue_safety_valve_server.ini field:

```
[desktop]
# Remove the file browser from the blacklisted apps.
# Tweak the app_blacklist property to suit your app configuration.
app_blacklist=spark,zookeeper,hive,hbase,search,oozie,jobsub,pig,sqoop,sec
urity
[azure]
  [[azure_accounts]]
    [[[default]]]
      client_id=<client_id>
      client_secret=<client_secret_id>
      tenant_id=<tenant_id>

    [[abfs_clusters]]
      [[[default]]]
  fs_defaultfs=abfs://<container_name>@<storage_name>.dfs.core.windows.net
```

```
webhdfs_url=https://<storage_name>.dfs.core.windows.net/
```

The custom configuration is stored in the hue_safety_valve_server.ini file.

4. Click Save Changes.
5. Restart the Hue service.

   The ABFS file browser icon appears on the left Assist panel as well as on the left navigation bar on the Hue web interface.

What to do next

You must manually grant the following application permission to non-admin users and groups for them to be able to view and access S3 File Browser in Hue: filebrowser.abfs_access:Access to ABFS from filebrowser and filepicker.

Related Information

Granting permission to access S3 and ABFS File Browser in Hue

Enabling ABFS File Browser in Hue with RAZ in DataHub

Hue offers you the capability to browse Azure Data Lake Storage (ADLS) Gen2 file system, upload files to ABFS, and create tables by importing files from ABFS containers. With RAZ, you can grant fine-grained access to per-user home directories and other directories within the ADLS containers using Ranger policies.

About this task

Hue administrators can create home directories for users by selecting the Create home directory option on the user's profile in Hue. If fine-grained authorization is enabled to access ADLS containers, then user home directories are automatically created when a user logs into Hue. You can disable automatic creation of user directories.

Note: If you have enabled RAZ while registering your Azure environment with CDP, then Hue uses RAZ as the default mechanism for enabling the ABFS File Browser

Before you begin

You must enable RAZ while registering your environment with CDP. See Introduction to RAZ on Azure environments" to get started with RAZ and complete the following steps:

1. Register an Azure environment with the Enable Ranger authorization for ADLS Gen2 option enabled. You can use the CDP web interface or the CDP CLI to complete this task.
2. Create a Data Hub cluster with Data Engineering or Data Mart cluster template.
3. Create the following Ranger policies:
   - Hadoop SQL policy (all - database, table, column, all - url).

     Note: You must grant permissions to individual users or groups in these Ranger policies. To grant permissions to all users, you can specify {USER} in the Permission section.

   - ABFS (cm_ADLS) policy (Default: User Home)

     You must specify the storage account name in the Storage Account field and the directory path of the container and its sub-directories in the Storage Account Container field of the cm_ADLS Ranger policy.
4. Grant appropriate permissions to the users in CDP User Management Service (UMS). For example, EnvironmentUser.

Procedure

1. Log in to the CDP web interface as an Administrator.
2. Go to Management Console Environments , select your environment and go to the Data Lake tab.
3. Click on the CM URL to open Cloudera Manager.

4. Go to Clusters Ranger RAZ service Instances RAZ server Processes and note the value of the fs.azure.ext.raz.rest.host.url property from the core-site.xml file. You need this to specify the value of the api_url property in the Hue configuration.
5. Go to Clusters Hue service Configuration Hue Service Advanced Configuration Snippet (Safety Valve) for hue_safety_valve.ini and add the following lines:

```
[desktop]
app_blacklist=spark,zookeeper,hbase,impala,search,sqoop,security,pig
[[raz]]
is_enabled=true
api_url=[***fs.azure.ext.raz.rest.host.url***]

[azure]
[[abfs_clusters]]
[[[default]]]
fs_defaultfs=abfs://[***CONTAINER-FOR-DATA-ACCESS***]@[***STORAGE-ACCOUNT
***].dfs.core.windows.net
webhdfs_url=https://[***STORAGE-ACCOUNT***].dfs.core.windows.net

[filebrowser]
remote_storage_home=abfs://[***CONTAINER-FOR-DATA-ACCESS***]/user
```

   For example:

   - fs_defaultfs=abfs://storage-fs@mystorage.dfs.core.windows.net
   - webhdfs_url=https://mystorage.dfs.core.windows.net
   - remote_storage_home=abfs://storage-fs/user

   (RAZ-HA) You can specify a comma-separated list of URLs in the api_url property. For example:

```
api_url=https://[***INSTANCE-1***]:6082/,https://[***INSTANCE-2***]:6082/
```

6. Click Save Changes.
7. Restart the Hue service.

Results

ABFS File Browser icon is added to the left assist pane on the Hue web interface. When you click on the ABFS File Browser icon, you are redirected to your user home directory.

Related Information

Introduction to RAZ on Azure environments

Disabling the automatic creation of user home directories

Using Google Cloud Storage with Hue

Hue can read to and write to Google Cloud Storage (GS) buckets.

Hue offers you the capability to browse GS buckets, upload files to GS, and create tables by importing files from GS. Using Ranger Authorization Service (RAZ), you can grant fine-grained access to per-user home directories and other directories within the GS bucket using Ranger policies.

Hue administrators can create home directories for users by selecting the Create home directory option on the user's profile in Hue. If fine-grained authorization is enabled to access GS buckets, then user home directories are automatically created when a user logs into Hue. You can disable the automatic creation of user directories.

Related Information

Disabling the automatic creation of user home directories

Prerequisites for enabling GS File Browser in Hue with RAZ in DataHub

First, you must register your Google Cloud environment with CDP with the Enable Ranger authorization for Google Cloud Storage option enabled. Next, you must set up Ranger policies for fine-grained access control for your user and Google Cloud Storage (GS) buckets.

Procedure

1. Log in to the CDP Management Console as an Administrator.
2. Register a GCP environment with the Enable Ranger authorization for Google Cloud Storage option enabled. You can complete this task using the CDP web interface or the CDP CLI.
3. Create a Data Hub cluster with a Data Engineering or Data Mart cluster template.
4. Create the following Ranger policies:
   a) Hadoop SQL policy (all - database, table, column, all - url).

      Note: You must grant permissions to individual users or groups in these Ranger policies. To grant permissions to all users, specify {USER} in the **Permission** section.

   b) GS (cm_gs) policy (Default: User Home).

      You must specify the bucket name in the GS Bucket field and the directory path in the Path field of the cm_gs Ranger policy.
5. Grant appropriate permissions to the users in the CDP User Management Service (UMS). For example, EnvironmentUser--grants permission to view Data Hub clusters and set the workload password for the environment.

What to do next

Enable the GS File Browser in Hue by configuring [desktop] properties in the Hue Advanced Configuration Snippet using Cloudera Manager.

Related Information

Introduction to RAZ on GCP environments

Enabling GS File Browser for Hue with RAZ in DataHub

After you have registered your Google Cloud environment with CDP with the Enable Ranger authorization for Google Cloud Storage (GS) option enabled, created a DataHub cluster, created the required Ranger policies, and granted access to your users, you can enable the GS File Browser using Cloudera Manager

Procedure

1. Log in to the CDP web interface as an Administrator.
2. Go to Management Console Environments , select your environment and go to the Data Lake tab.
3. Click on the link under CM URL to open Cloudera Manager.
4. Go to Clusters Ranger RAZ service Instances RAZ server Processes and note the value of the fs.s3a.ext.raz.rest.host.url property from the core-site.xml file. You need this to specify the value of the api_url property in the Hue configuration.
5. Go to Clusters Hue service Configuration Hue Service Advanced Configuration Snippet (Safety Valve) for hue_safety_valve.ini and add the following lines:

```
[desktop]
app_blacklist=spark,zookeeper,hbase,impala,search,sqoop,security,pig
[[raz]]
is_enabled=true
is_raz_gs_enabled=true
api_url=[***fs.s3a.ext.raz.rest.host.url***]

[filebrowser]
```

```
remote_storage_home=gs://[***GS-BUCKET-NAME***]/user
```

(RAZ-HA) You can specify a comma-separated list of URLs in the api_url property. For example:

```
api_url=https://[***INSTANCE-1***]:6082/,https://[***INSTANCE-2***]:6082/
```

6. Click Save Changes.
7. Restart the Hue service.

Results

GS File Browser icon is added to the left assist panel on the Hue web interface. When you click the GS File Browser icon, you are redirected to your user home directory.

Disabling the automatic creation of user home directories

If you have enabled fine-grained authorization to access S3 buckets, ADLS Gen 2 storage, or Google Cloud Storage buckets, then Hue is configured to automatically create user home directories, by default. You can disable the automatic creation of user home directories by setting the autocreate_user_dir flag to false in the Hue Service Advanced Configuration Snippet.

Procedure

1. Log in to the CDP web interface as an Administrator.
2. Go to Data Hub Clusters and select your cluster.
3. Click on the CM URL to open Cloudera Manager.
4. Go to Clusters Hue service Configuration Hue Service Advanced Configuration Snippet (Safety Valve) for hue_safety_valve.ini and add the following lines:

```
[desktop]
[[raz]]
autocreate_user_dir=false
```

5. Click Save Changes.
6. Restart the Hue service.

Granting permission to access S3, ABFS, GS File Browser in Hue

Only admin users can view and access S3, ABFS, and GS File Browser in Hue after enabling it. You must manually grant application permissions to non-admin users and groups for them to be able to view and access S3, ABFS, and GS File Browsers in Hue.

About this task

The following table lists the application permissions for each cloud storage type:

Cloud storage	Hue application permission
S3	filebrowser.s3_access:Access to S3 from filebrowser and filepicker.
ABFS (ADLS Gen2)	filebrowser.abfs_access:Access to ABFS from filebrowser and filepicker.
ADLS (Gen1)	filebrowser.adls_access:Access to ADLS from filebrowser and filepicker.
GS	Access to GS from filebrowser and filepicker.

Before you begin

You can only assign Hue application permissions to groups or users within a group. Add the users to a group to whom you want to grant access to the S3, ABFS, or GS File Browsers.

Important: The "default" group in Hue does not have the permissions required to view S3, ABFS, or GS File Browsers, by default. If your users belong to the default group, then you must manually grant the required permissions.

Procedure

1. Sign in to Hue as an Administrator.
2. Go to admin Manage Users Groups .
3. Click on the group to whom you want to grant the filebrowser application permissions.
4. On the **Edit [***GROUP-NAME***]** page, select the required permission under the permission section and click Update group.

Creating tables in Hue by importing files

Using Hue Importer, you can create Hive, Impala, and Iceberg tables from CVS and XLSX files. After enabling the File Browser for your cloud provider, you can import the file into Hue to create tables.

About this task

You can upload and import a 200 KB file from your local computer, or import a file up to 3 GB from cloud storage (S3, ADLS Gen2, and Google Cloud Storage (GS) buckets) using the Importer.

Note: To enable uploading a small local file from your computer, you must go to Cloudera Manager Clusters Hue Configuration and add the following lines in the Hue Service Advanced Configuration Snippet (Safety Valve) for hue_safety_valve.ini field:

```
[indexer]
  # Flag to turn on the direct upload of a small file.
  enable_direct_upload=true
```

You can create managed and external tables. When you create managed tables, the data is moved to HDFS in the Cloudera warehouse workspace. When you create external tables, your data remains in its original location.

Note: Superusers can browse all directories up to the root. Other users can only browse and access buckets and directories that they are granted access to.

Before you begin

- Enable user access to cloud storage buckets/containers in Ranger
- Enable the File Browser for your cloud provider

Procedure

1. Log in to the Hue web interface.
2. Click Importer from the left-assist panel.

   Alternatively, click + to create tables under the Databases Tables list view. This opens the Hue Importer.
3. Under SOURCE, select Remote File from the Type drop-down menu.
4. Click .. at the end of the Path field.

   The **Choose a file** modal is displayed.

5. (Non-RAZ deployment only) Click on your cloud provider, type the following depending on your cloud provider, and press enter:

   (AWS): s3a://[***BUCKET-NAME***]

   (Azure): abfs://[***CONTAINER-NAME***]

   (GCS): gs://[***BUCKET-NAME***]
6. Browse and select the file you want to use to create a table.

   Hue displays the preview of the table along with the format.

   Hue automatically detects the field separator, record separator, and the quote character from the file. If you want to override a specific setting, then you can change it by selecting a different value from the drop-down menu.
7. Click Next.

   The table destination and properties are displayed.
8. Under Extras, deselect Store in Default location and Transaction table options.

   This is required to create external tables.
9. Select the Copy file option.

   Selecting this option retains the source file in the original location and creates a copy of the file to create a table.

   If you do not select this option, then Hue moves the file from the source location to a new folder, making it unavailable in the original location.
10. Set the table destination, partitions, and change the column data types.
11. Verify the settings and click Submit to create the table.

    The CREATE TABLE query is triggered.

    Hue displays the logs and opens the Table Browser from which you can view the newly created table when the operation completes successfully.

Related Information

Enabling S3 File Browser for Hue with RAZ in DataHub

Enabling ABFS File Browser in Hue with RAZ in DataHub

Enabling GS File Browser for Hue with RAZ in DataHub

Supported non-ASCII and special characters in Hue

Auto-generated files may often introduce non-alphanumeric characters in the file and directory names that Hue does not support. This might cause the files or directories to not appear on the Hue File Browser. Review the list of non-alphanumeric, non-ASCII, diacritics (accents) characters supported in Hue for the following operations: upload, create, list in folder, view, and rename.

Table 1: Supported characters on HDFS

Special character symbol	Description	Filename support	Folder name support
~	Tilde	Fully supported	Fully supported
@	Ampersat	Fully supported	Fully supported
#	Hash	Partially supported. Not supported for rename operations.	Unsupported
$	Dollar sign	Fully supported	Fully supported
&	Ampersand	Fully supported	Fully supported
(	Left paranthesis	Fully supported	Fully supported
)	Right paranthesis	Fully supported	Fully supported
*	Asterisk	Fully supported	Fully supported
!	Exclamation mark	Fully supported	Fully supported

Special character symbol	Description	Filename support	Folder name support
+	Plus	Fully supported	Fully supported
=	Equal	Fully supported	Fully supported
:	Colon	Unsupported	Unsupported
;	Semicolon	Fully supported	Fully supported
,	Comma	Fully supported	Fully supported
.	Period	Fully supported	Fully supported
?	Question mark Not supported with Knox.	Fully supported	Fully supported
/	Forward slash	Unsupported	Unsupported
\	Backslash	Unsupported	Unsupported
'	Apostrophe or single quote	Fully supported	Fully supported
%	Percentage	Fully supported	Fully supported
T######-ääö	Non-standard alphabets with diacritics and accents.	Fully supported	Fully supported

Table 2: Supported characters on S3

Special character symbol	Description	Filename support	Folder name support
~	Tilde	Fully supported	Fully supported
@	Ampersat	Fully supported	Fully supported
#	Hash	Partially supported. Not supported for view and rename operations.	Unsupported
$	Dollar sign	Fully supported	Fully supported
&	Ampersand	Fully supported	Fully supported
(	Left paranthesis	Fully supported	Fully supported
)	Right paranthesis	Fully supported	Fully supported
*	Asterisk	Fully supported	Fully supported
!	Exclamation mark	Fully supported	Fully supported
+	Plus	Fully supported. Not supported for create operations on RAZ-enabled environments.	Fully supported
=	Equal	Fully supported	Fully supported
:	Colon	Unsupported	Unsupported
;	Semicolon	Fully supported	Fully supported
,	Comma	Fully supported	Fully supported
.	Period	Fully supported	Fully supported
?	Question mark Not supported with Knox.	Fully supported	Partially supported. Not supported for list, upload, and view operations.
/	Forward slash	Unsupported	Unsupported
\	Backslash	Partially supported. Not supported for upload operations.	Partially supported. Not supported for upload operations.
'	Apostrophe or single quote	Fully supported	Fully supported

Special character symbol	Description	Filename support	Folder name support
%	Percentage	Fully supported	Fully supported
T######-ääö	Non-standard alphabets with diacritics and accents.	Fully supported	Fully supported

Table 3: Supported characters on ABFS

Special character symbol	Description	Filename support	Folder name support
~	Tilde	Fully supported	Fully supported
@	Ampersat	Fully supported	Fully supported
#	Hash	Partially supported. Not supported for view and rename operations.	Unsupported
$	Dollar sign	Fully supported	Fully supported
&	Ampersand	Fully supported	Fully supported
(	Left paranthesis	Fully supported	Fully supported
)	Right paranthesis	Fully supported	Fully supported
*	Asterisk	Fully supported	Fully supported
!	Exclamation mark	Fully supported	Fully supported
+	Plus	Fully supported	Fully supported
=	Equal	Fully supported	Fully supported
:	Colon	Unsupported	Unsupported
;	Semicolon	Fully supported	Fully supported
,	Comma	Fully supported	Fully supported
.	Period	Fully supported	Fully supported
?	Question mark Not supported with Knox.	Partially supported. Not supported for view and rename operations.	Partially supported. Not supported for list, rename, and view operations.
/	Forward slash	Unsupported	Unsupported
\	Backslash	Unsupported	Unsupported
'	Apostrophe or single quote	Fully supported	Fully supported
%	Percentage	Fully supported	Fully supported
T######-ääö	Non-standard alphabets with diacritics and accents.	Fully supported	Fully supported

Options to rerun Oozie workflows in Hue

Oozie workflows consume time and resources to run. You can optimize the rerun of a failed Oozie workflow by selecting one of the following two options: “All or skip successful” or “Only failed”. These options enable you to select individual actions within your workflow that you want to rerun.

Important: The option to rerun the an Oozie workflow is enabled in the following conditions:

- In case of an external workflow, the rerun button is enabled if the workflow was run by the same user who is logged in, or if the user is a Hue superuser.
- If the workflow is created in Hue, then the rerun button is enabled only for the user who originally ran the workflow, or for the Hue superuser.

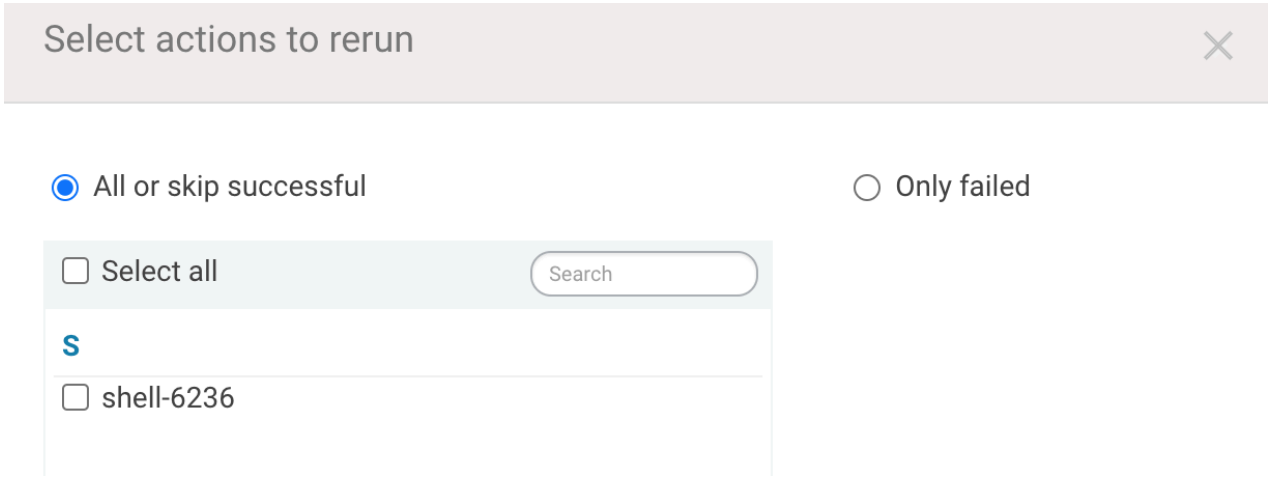
The All or skip successful option enables you to either rerun both the successful and failed actions or skip the actions that ran successfully and run only the failed ones. Using this option, you have more control over selecting the actions that you want to rerun from the list of successfully run actions.

If you select the All or skip successful option but do not select any or all of the successfully run actions, then Hue reruns the whole Oozie workflow.

If you select the All or skip successful option and select some of the successfully run actions, then Hue reruns the selected actions and the failed actions.

Select the Only failed option to only rerun the failed actions within your workflow.

The following image shows the available rerun options on the Hue web interface:



Unsupported features in Hue

Learn about the Hue features that are not supported by Cloudera.

Unsupported options in Hue Importer

The following options are displayed on the Hue **Importer** page under SOURCE Path , but are not supported:

- External Database

  Creating an external database using the Hue Importer is not supported. Cloudera recommends that you create a database using a SQL query.
- Manually

Known limitations in Hue

Review the known limitations in Hue.

Hue has the following limitations:

- Node depth for graphing Oozie workflows because of performance issues. See Improved Oozie Workflow display of large Graphs.
- You must use the Cloudera-provided Apache Load balancer to serve static content, because:
  - It serves static JavaScript, CSS, and Webpack files for client requests and reduces the load from the backend Python web server.
  - The Hue load balancer uses a sticky cookie session to route requests to the same backend as the Python web server, which talks to the same coordinator.
- Hue can only show logs from either Spark1 or Spark2, not both at a time.

- Spark notebook is not supported.
- External RDBMS in the query editor is not supported out of the box by default. Cloudera support will assist on a best-effort basis. Cloudera recommends that you raise issues in the open-source github community.
- Impala queries stay in the “executing” state so that Hue can display results when users are ready
- We need to limit the amount of data available to download from Hive/Impala because massive downloads cause performance degradation. Multiple simultaneous downloads of result sets could also degrade performance.
- Upstream features and connectors may not function properly in CDP. Cloudera recommends that you raise issues in the open-source github community.