Cloudera Runtime 1.0.0

Using Hue

Date published: 2020-07-28
Date modified: 2024-12-05

https://docs.cloudera.com/

Legal Notice

Contents

About using Hue

Hue provides a one-stop querying experience in Cloudera that leverages Hive, Impala, and Unified Analytics SQL engines.

Accessing and using Hue in Cloudera Data Warehouse

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/.

Before you begin

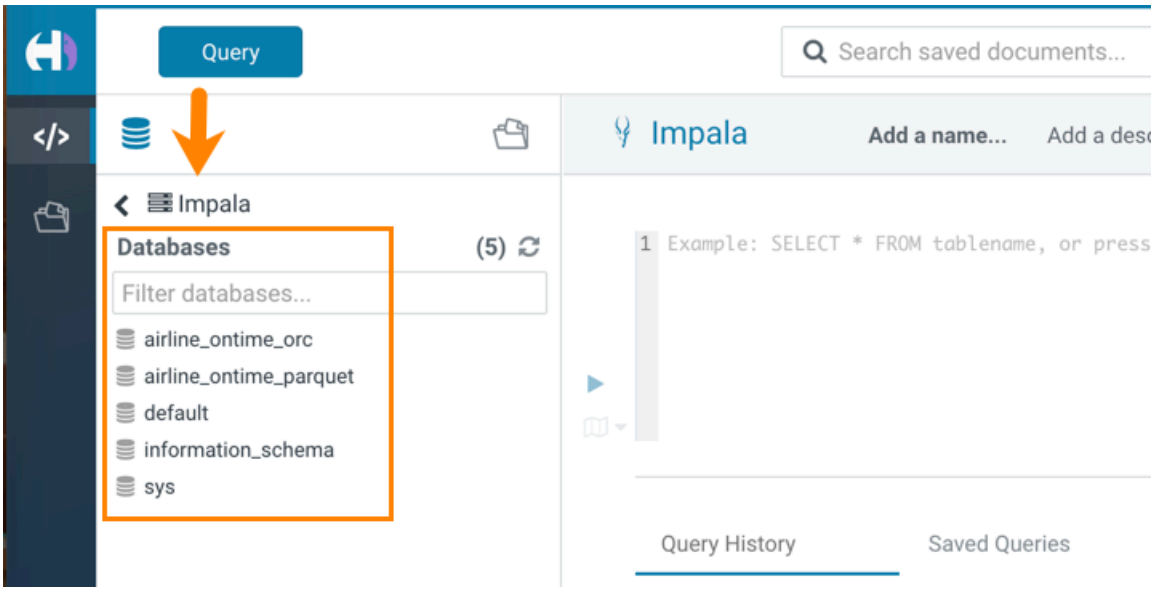
Hue uses your LDAP credentials that you have configured for the Cloudera cluster.

Procedure

1. Log into the Cloudera web interface and navigate to the Cloudera Data Warehouse service.
2. In the Cloudera Data Warehouse service, navigate to the **Overview** page.

Note: You can also launch Hue from the **Virtual Warehouse** page using the same steps.

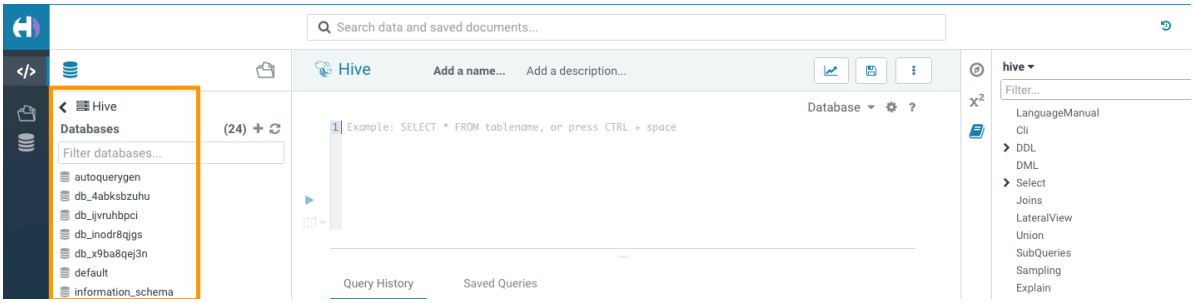
3. To run Impala queries:
   a) On the **Overview** page under Virtual Warehouses, click on the Hue button.
   The query editor is displayed:



   b) Click a database to view the tables it contains.
   When you click a database, it sets it as the target of your query in the main query editor panel.
   c) Type a query in the editor panel and click the run icon ▶ to run the query.

   Note: Use the Impala language reference to get information about syntax in addition to the SQL auto-complete feature that is built in. To view the language reference, click the book icon to the right of the query editor panel.

4. To run Hive queries:
   a) On the **Overview** page under Virtual Warehouses, click on the Hue button.
   The Hive query editor is displayed:



   b) Click a database to view the tables it contains.
   When you click a database, it sets it as the target of your query in the main query editor panel.
   c) Type a query in the editor panel and click the run icon ▶ to run the query.

   Note: Use the Hive language reference to get information about syntax in addition to the SQL auto-complete feature that is built in. To view the language reference, click the book icon to the right of the query editor panel.

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 on all Hive Virtual Warehouses within a Database Catalog 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.

Queries are retained in the backend database for 30 days by default, after which they are cleaned up. You can change the clean-up interval from the Database Catalog configurations.

Procedure

1. Go to the Cloudera Data Warehouse web interface and open Hue from your Virtual Warehouse.
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 in Cloudera Data Warehouse

Viewing Hive query information

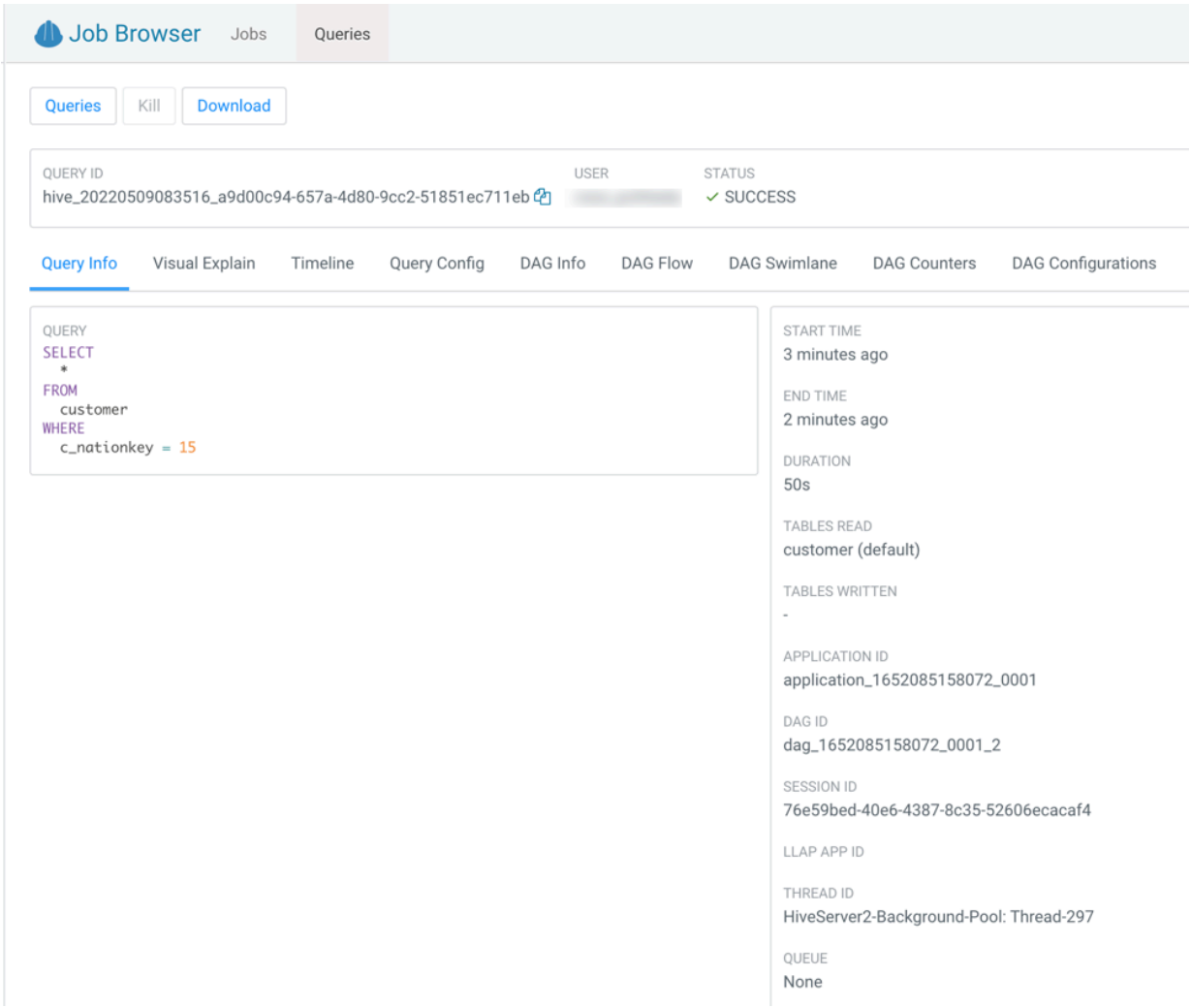
The Query Info tab provides information such as, the Hive query ID, the user who executed the query, the start time, the end time, the total time taken to execute the query, the tables that were read and written, application ID, Directed Acyclic Graph (DAG) IDs, session ID, LLAP app ID, thread ID, and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

Procedure

1. Go to the Cloudera Data Warehouseweb interface and open Hue from your Virtual Warehouse.
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. Go to the Cloudera Data Warehouse web interface and open Hue from your Virtual Warehouse.
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.

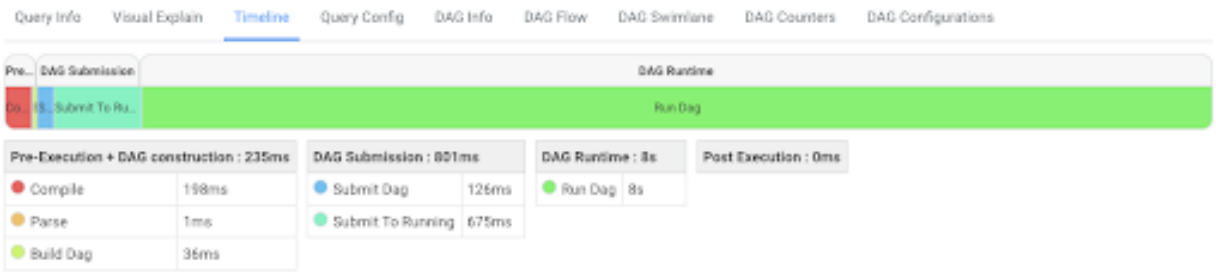
Duration data about each phase are distilled into more granular metrics based on query execution logs.

Procedure

1. Go to the Cloudera Data Warehouse web interface and open Hue from your Virtual Warehouse.
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. Go to the Cloudera Data Warehouse web interface and open Hue from your Virtual Warehouse.
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.

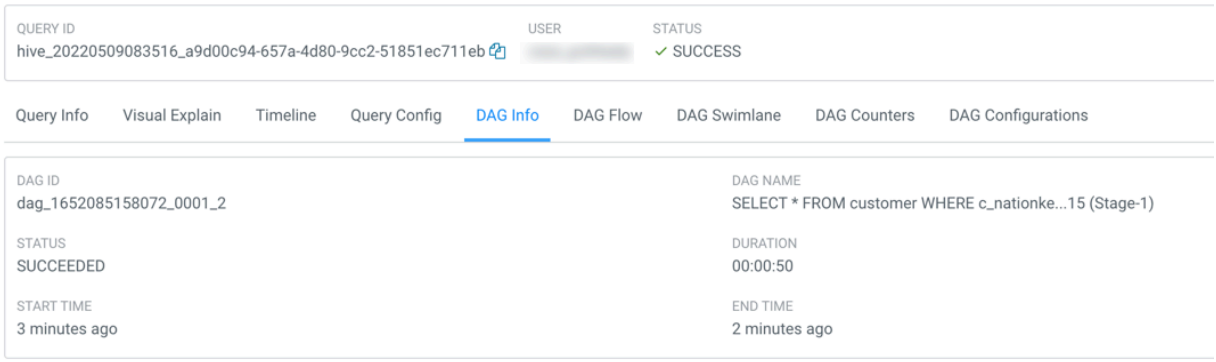
Important: The DAG information tabs (**DAG Info**, **DAG Flow**, **DAG Swimlane**, **DAG Counters**, **DAG Configurations**) are displayed only if the Tez engine is used for query execution. The Tez engine is typically utilized for complex queries.

Procedure

1. Go to the Cloudera Data Warehouse web interface and open Hue from your Virtual Warehouse.
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:



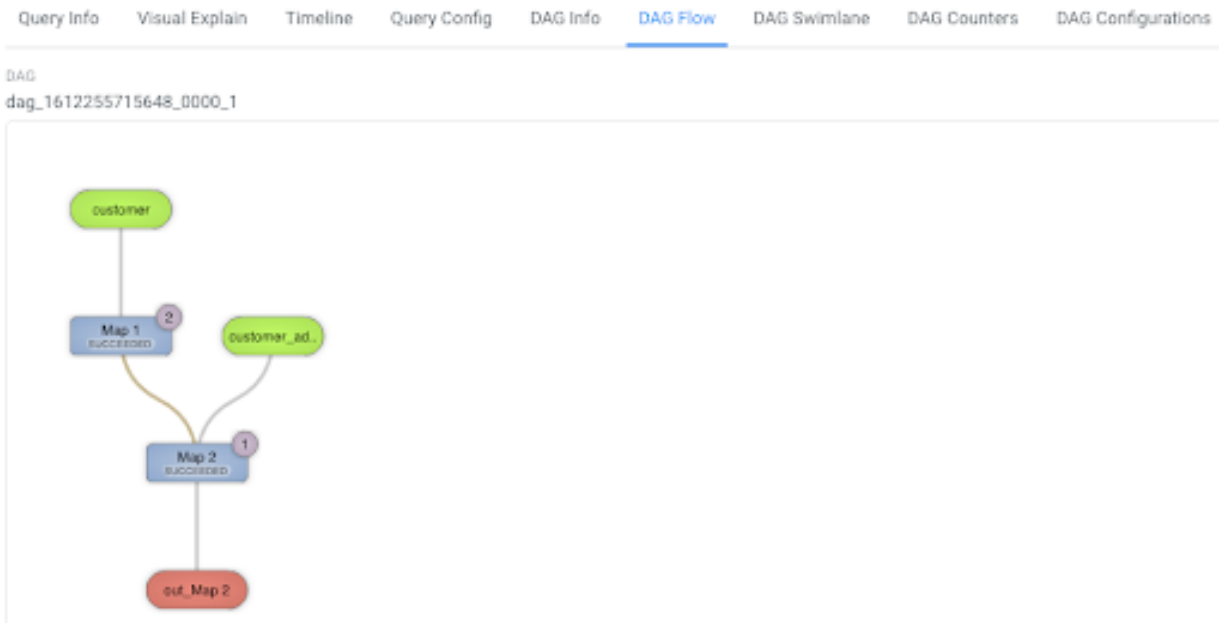
   The following table lists and describes the status of the Tez job:

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:

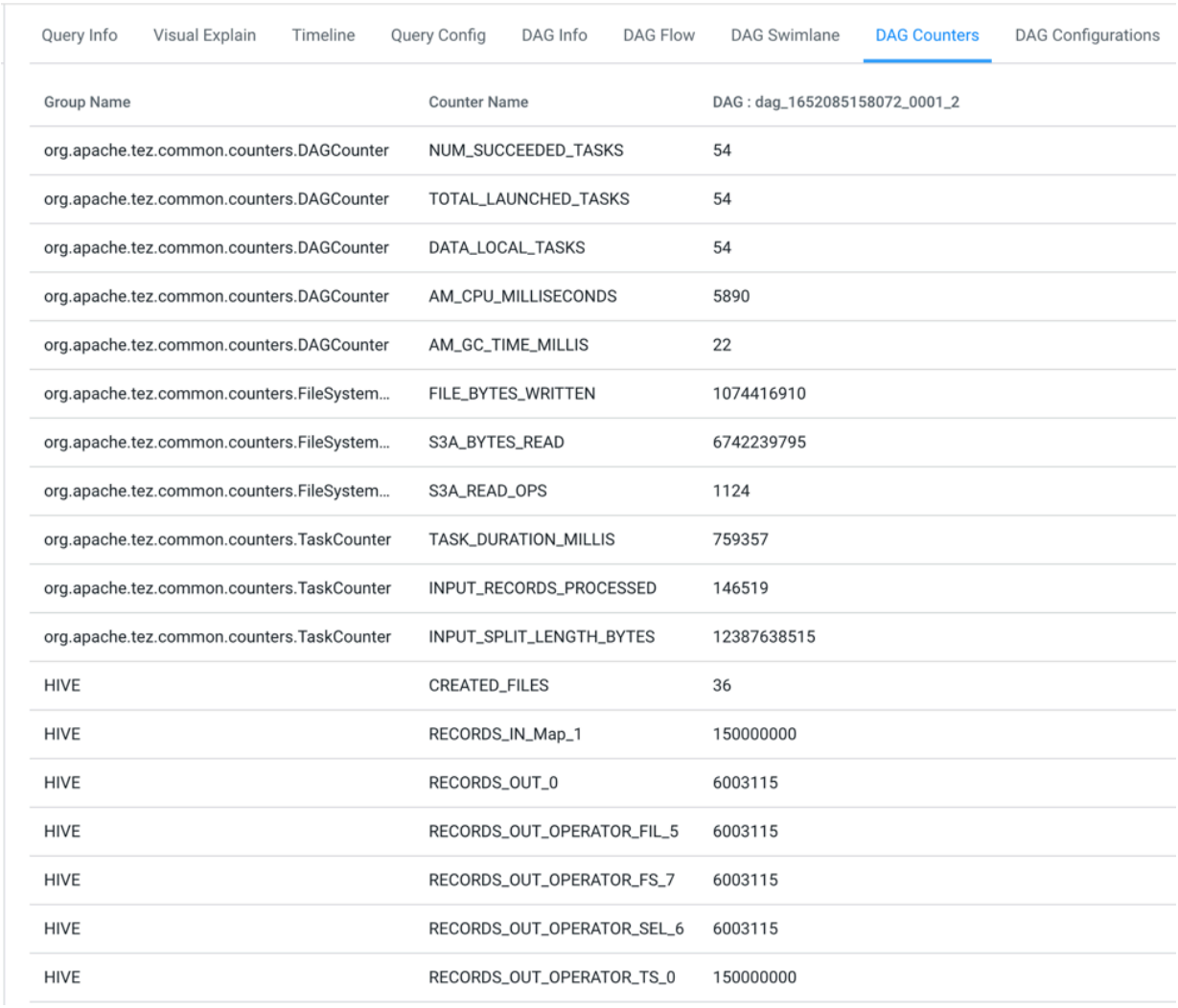
   In this example, Map 1 depends on the results of Map 5. Map 1 will finish its execution only when Map 5 finishes its execution successfully. Similarly, Reducer 2 depends on Map 1 to complete its execution.

   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.

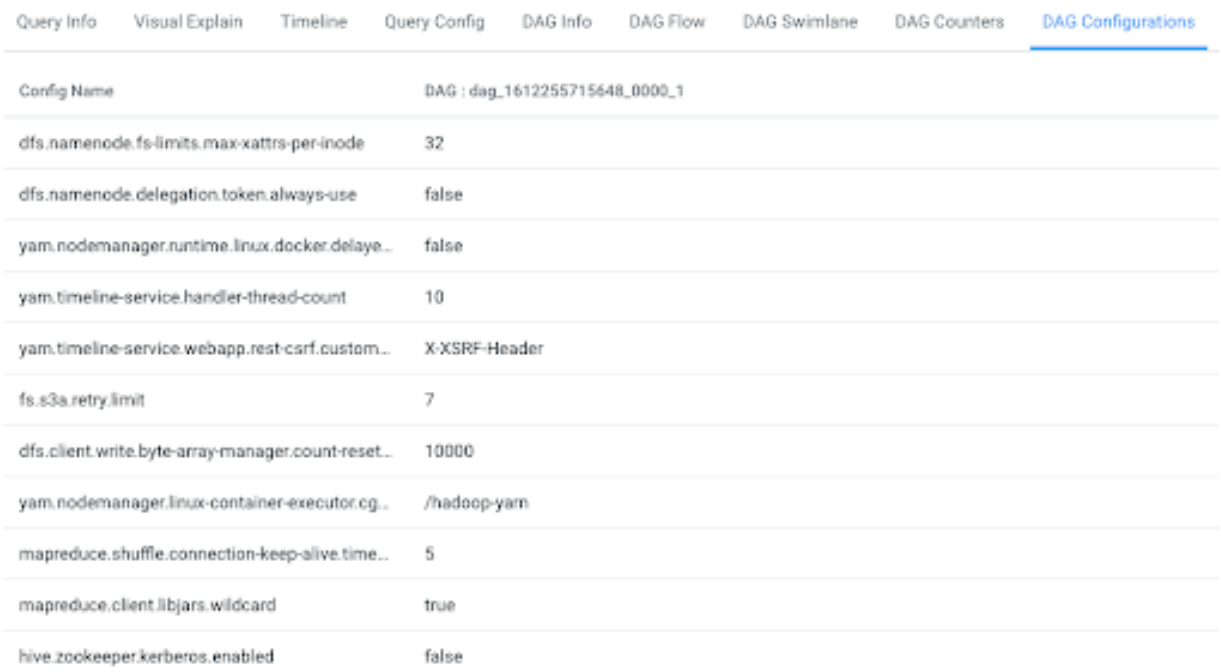
   The following image shows the **DAG Counters** tab on the Hue web interface:



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.

Viewing Impala query history

The Impala Queries tab on the Job Browser page in Hue displays all the queries that were run on all Impala Virtual Warehouses within a Database Catalog 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 monitor resource utilization and control costs from the Hue Job Browser. Non-admin users can view only their queries.

Queries are retained in the backend database for 30 days by default, after which they are cleaned up. You can change the clean-up interval from the Database Catalog configurations.

Note: Impala queries may take up to 25 minutes to appear on the **Impala Queries** tab after they are run. This is a known limitation in Cloudera Data Warehouse.

Procedure

1. Go to the Cloudera Data Warehouse web interface and open Hue from your Virtual Warehouse.
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.

Viewing Impala query information

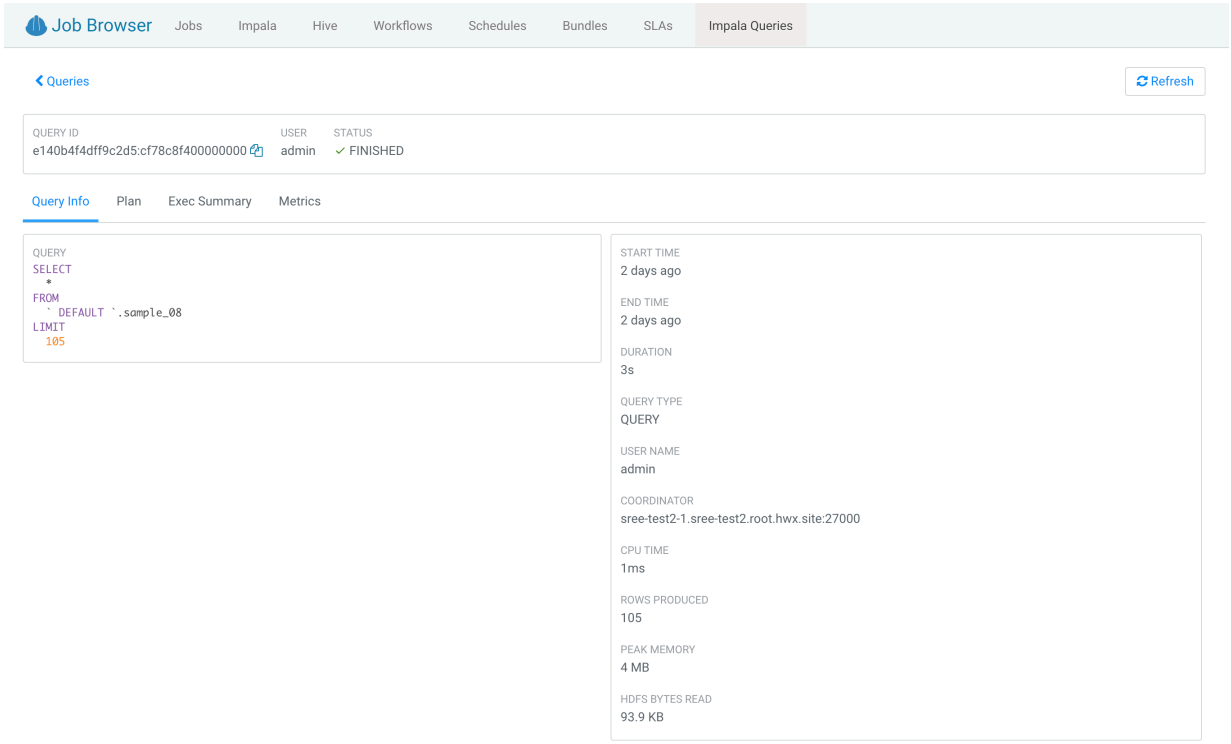
The Query Info tab in Hue provides information such as, the Impala query ID, the user who executed the query, the start time, the end time, the total time taken to execute the query, the coordinator that received the query, CPU time, rows produced, peak memory, and HDFS bytes read.

Procedure

1. Go to the Cloudera Data Warehouse web interface and open Hue from your Virtual Warehouse.
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

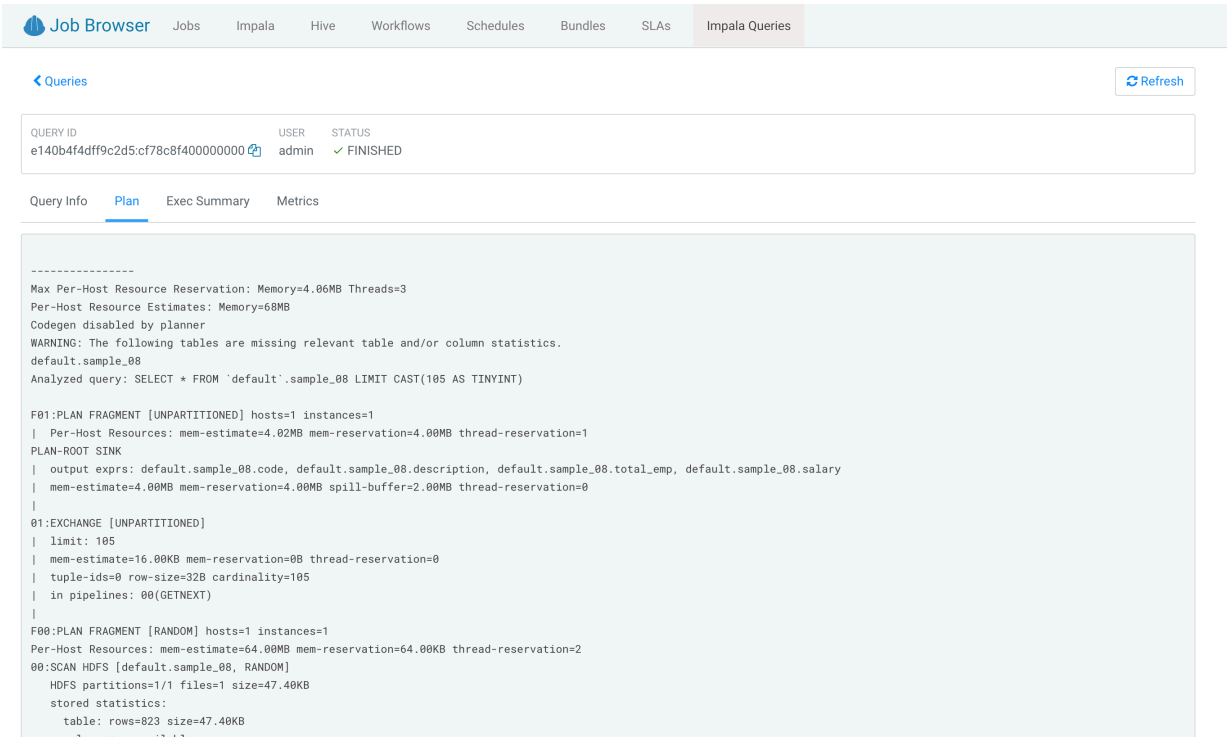
The query execution plan in Hue provides details on how the query will be executed, the operators involved, and other information before the query is submitted to the Impala engine.

Procedure

1. Go to the Cloudera Data Warehouse web interface and open Hue from your Virtual Warehouse.
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:



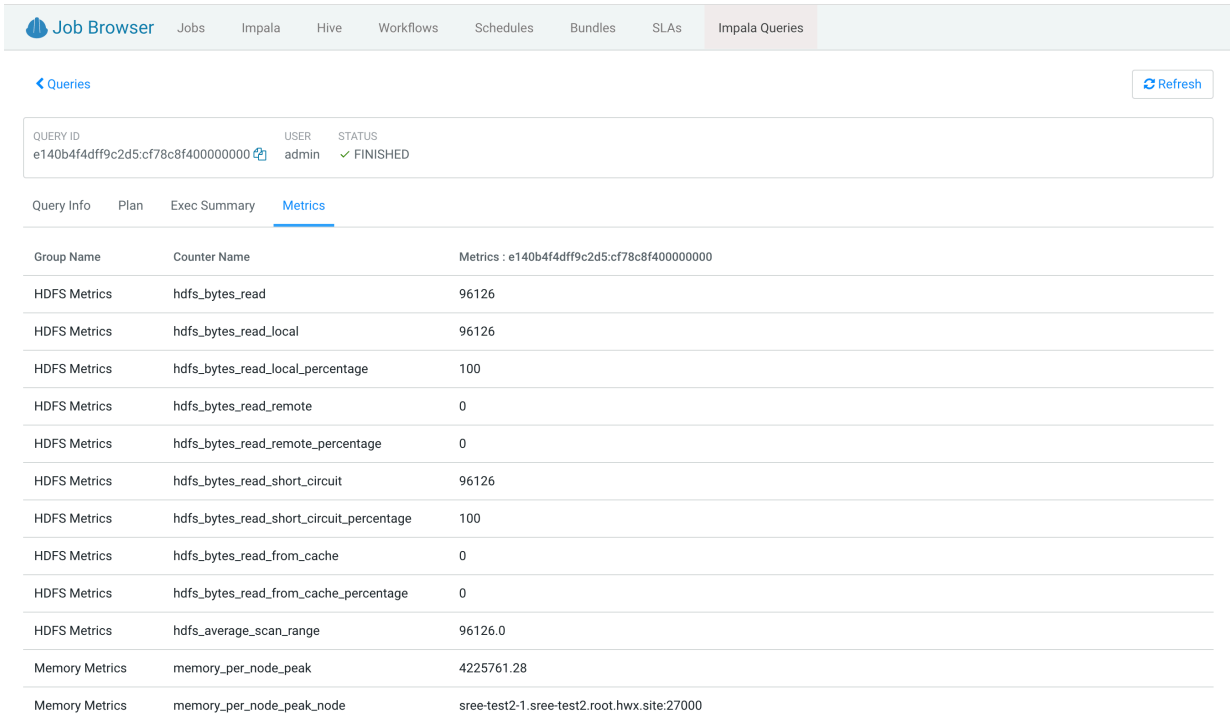
Viewing the Impala query metrics

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. Go to the Cloudera Data Warehouse web interface and open Hue from your Virtual Warehouse.
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 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. Go to the Cloudera Data Warehouse web interface and open Hue from your Virtual Warehouse.
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

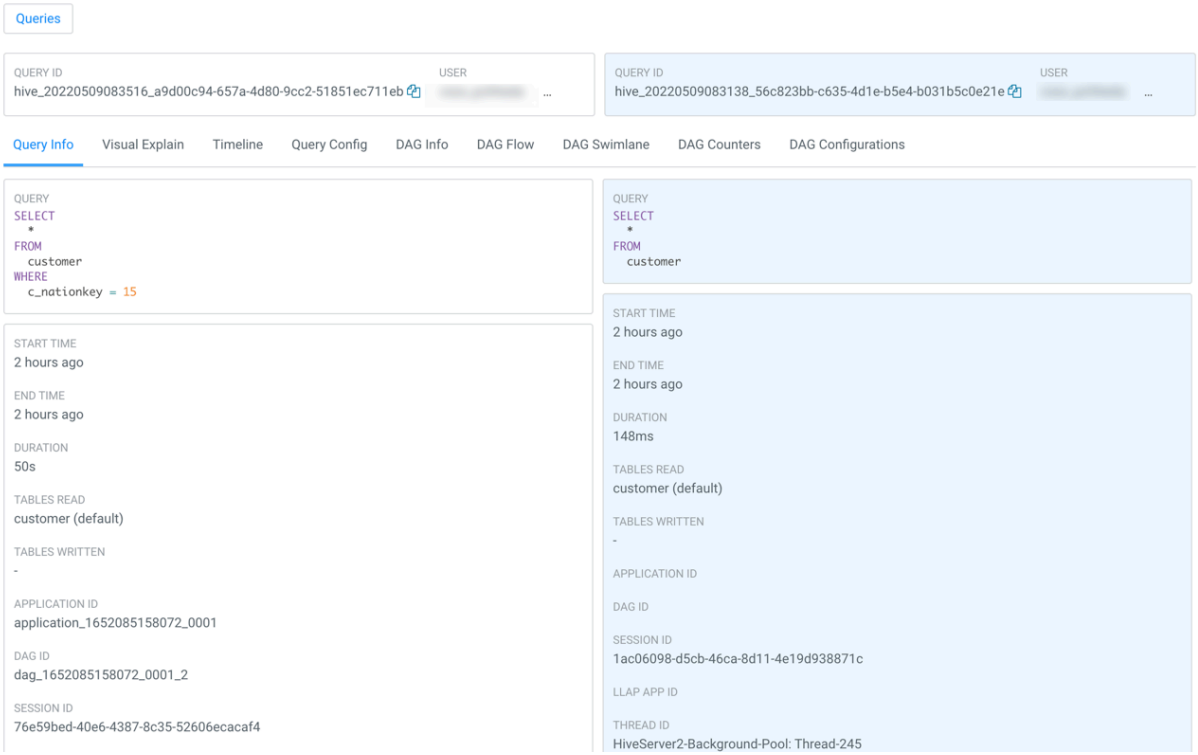
1. Go to the Cloudera Data Warehouse web interface and open Hue from your Virtual Warehouse.
2. Click Jobs from the left assist panel.

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

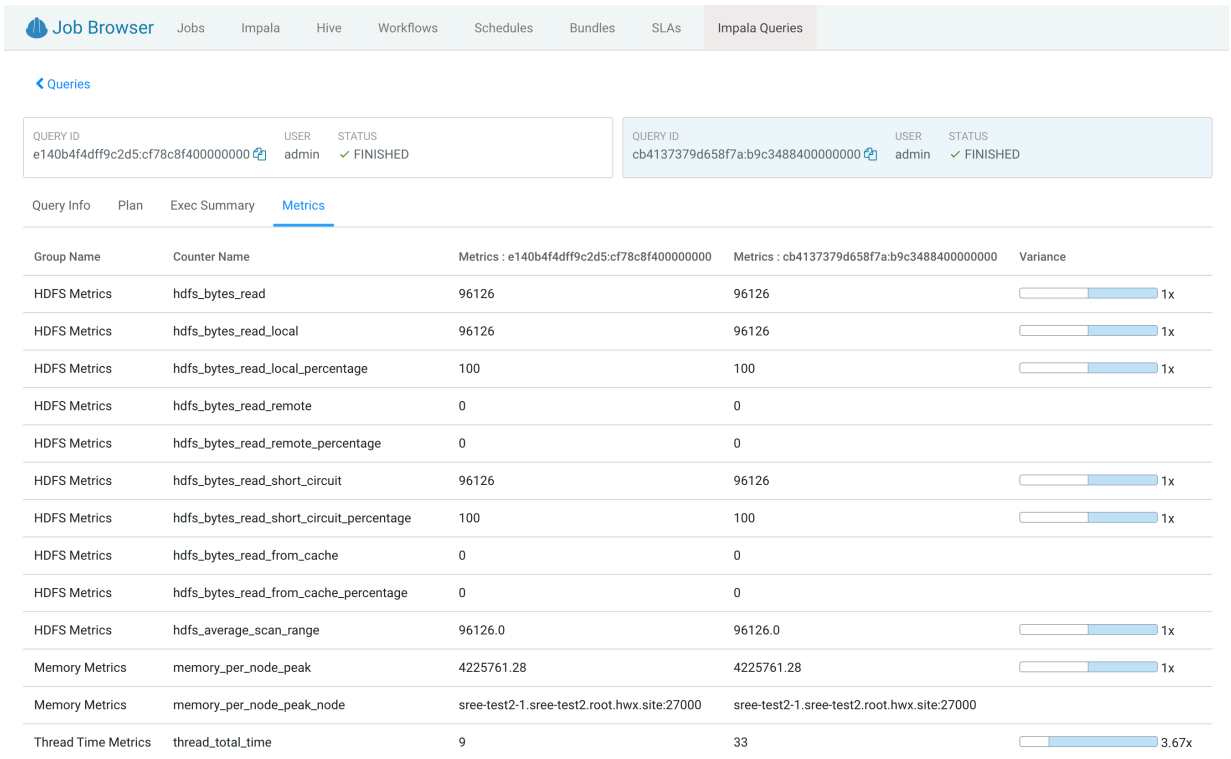
   A list of queries that were run is displayed.

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

   Query comparison report for Hive queries:



   Query comparison report for Impala queries:



Group Name	Counter Name	Metrics : e140b4f4dff9c2d5:cf78c8f400000000	Metrics : cb4137379d658f7a:b9c3488400000000	Variance
HDFS Metrics	hdfs_bytes_read	96126	96126	1x
HDFS Metrics	hdfs_bytes_read_local	96126	96126	1x
HDFS Metrics	hdfs_bytes_read_local_percentage	100	100	1x
HDFS Metrics	hdfs_bytes_read_remote	0	0	
HDFS Metrics	hdfs_bytes_read_remote_percentage	0	0	
HDFS Metrics	hdfs_bytes_read_short_circuit	96126	96126	1x
HDFS Metrics	hdfs_bytes_read_short_circuit_percentage	100	100	1x
HDFS Metrics	hdfs_bytes_read_from_cache	0	0	
HDFS Metrics	hdfs_bytes_read_from_cache_percentage	0	0	
HDFS Metrics	hdfs_average_scan_range	96126.0	96126.0	1x
Memory Metrics	memory_per_node_peak	4225761.28	4225761.28	1x
Memory Metrics	memory_per_node_peak_node	sree-test2-1.sree-test2.root.hwx.site:27000	sree-test2-1.sree-test2.root.hwx.site:27000	
Thread Time Metrics	thread_total_time	9	33	3.67x

How to run a stored procedure from Hue in Cloudera Data WarehouseHow to run a stored procedure from Hue in Cloudera Data Hub

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. In Cloudera Data Warehouse, Hue provides a smart interface to run stored procedures.

Note: This feature is available only for Hive queries.

To run stored procedures from Hue, create a Hive Virtual Warehouse in Cloudera Data Warehouse and enable the hplsql option in the hue-safety-valve field.

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 Cloudera Data Warehouse

Enabling stored procedures for Hive in Cloudera Data Warehouse

To create, edit, and drop procedures and functions that are written in Hive Hybrid Procedural SQL (HPL/SQL) using the Hue query editor in Cloudera Data Warehouse, you must enable the hplsql option in the hue-safety-valve field.

About this task

Important: Hue enables you to switch between Hive and HPL/SQL interpreters. By default, the regular Hive interpreter is enabled when you create a Hive Virtual Warehouse. To enable the HPL/SQL interpreter, you must update the configuration in the hue-safety-valve field in your Hive Virtual Warehouse. However, updating the hue-safety-valve overrides the default configuration. Therefore, to use both Hive and HPL/SQL interpreters, you must enable both by updating the configuration in the hue-safety-valve field.

Procedure

1. Log in to the Cloudera Data Warehouse service as an administrator.
2. Go to Virtual Warehouse Edit CONFIGURATIONS Hue and select hue-safety-valve from the Configuration files drop-down list.
3. Add the following lines in the hue-safety-valve:

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

4. Click APPLY.
5. Restart the Virtual Warehouse.

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

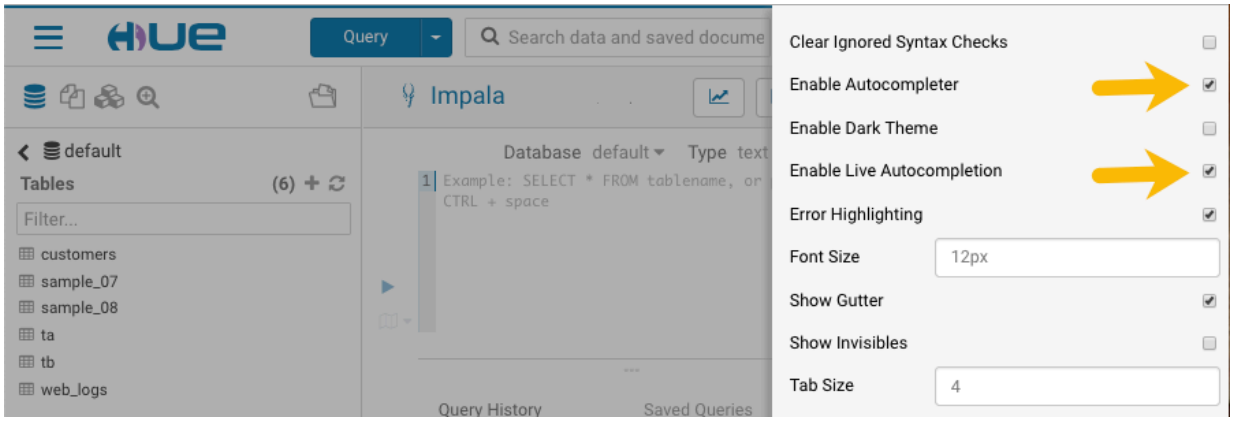
1. Log in to Hue and go to either the Hive or Impala editor.
2. Place your cursor in the editor window and then use one of the following keyboard shortcuts to open the editor configuration panel:
   - On a Mac system, use the Command key followed by a hyphen and then a comma:

     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.

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.

In Cloudera Data Warehouse, you can only view tags that are created in Atlas in Hue. You must create tags in Atlas.

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

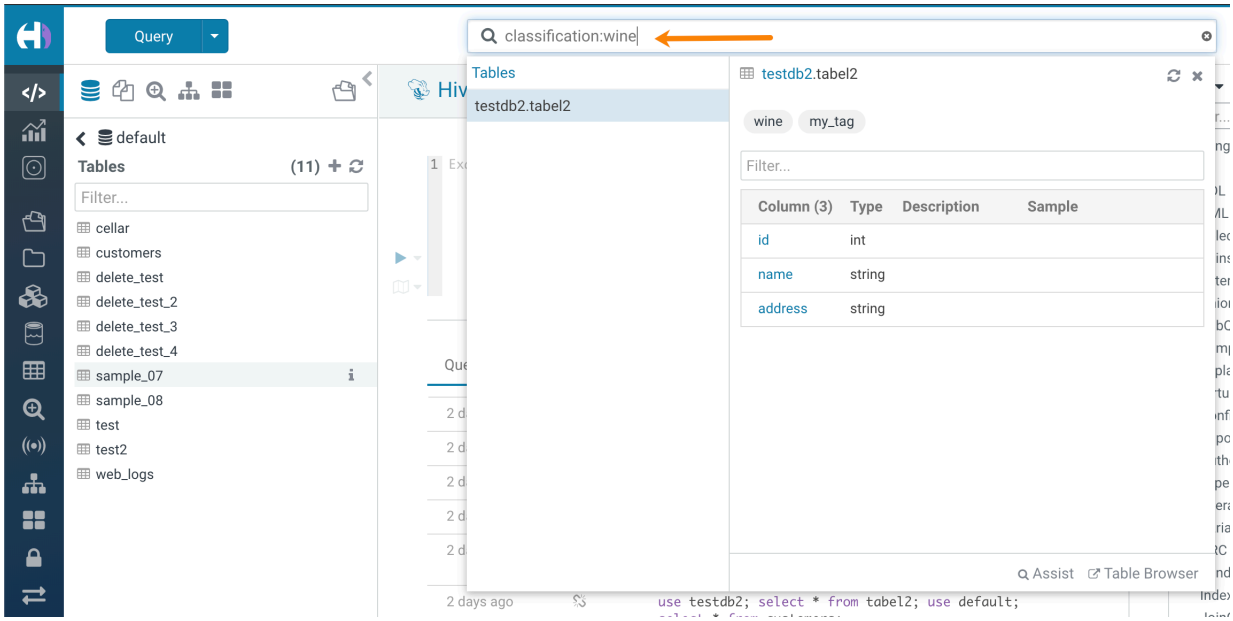
You can search for tags or classifications in either the Hive or the Impala editors.

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:

In Cloudera Data Warehouse, you can browse Amazon S3 buckets from Hue in the following ways:

- With Ranger Authorization Service (RAZ)
- Without Ranger Authorization Service (RAZ)

If you have registered your Cloudera on cloud environment using RAZ, policies attached to the Ranger RAZ Service role at the Data Lake-level can control access to external S3 buckets. If your Virtual Warehouse predates the capability to use RAZ configurations and policies, then you must manually enable RAZ for Cloudera Data Warehouse and then configure Hue to access S3 buckets. Your Data Lake must be RAZ-enabled to manually enable RAZ exclusively for Cloudera Data Warehouse.

Related Information

Registering a RAZ-enabled AWS environment

Enabling RAZ manually in Cloudera Data Warehouse on cloud

Accessing S3 bucket from Hue in Cloudera Data Warehouse with RAZ

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

About this task

If you have enabled RAZ while registering your AWS environment with Cloudera, then Hue uses RAZ as the default mechanism for enabling the S3 File Browser. Before you can enable the S3 File Browser in Hue, you must complete the following prerequisites:

Procedure

1. Follow the instructions listed in Introduction to RAZ on AWS environments to register an AWS environment with the Enable Ranger authorization for AWS S3 option enabled. You can use the Cloudera web interface or the Cloudera CLI to complete this task.

   Note: You must enable RAZ while registering your environment with Cloudera.

2. Log in to the Cloudera Management Console as a DWAdmin or DWUser and go to the Cloudera Data Warehouse service.
3. Click Open Ranger on your Database Catalog.
4. Create the following Ranger policies:
   a) Hadoop SQL policy (all - database, table, column, all - url).

      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.

   b) S3 (cm_S3) policy (Default: User Home)

      You must grant permissions to the following users in the Permissions section for the user home directory (/user/{USER}): {USER}.

      Specify the bucket name in the S3 Bucket field and the directory path in the Path field of the cm_S3 Ranger policy.

   c) S3 (cm_S3) policy (Default: user)

      You must grant permissions to the following users in the Permissions section for the root directory (/user/): hive, impala.
5. You must also grant appropriate permissions to the users in Cloudera User Management Service (UMS). For example, EnvironmentUser.

Enabling the S3 File Browser for Hue in Cloudera Data Warehouse with RAZ

The S3 File Browser in Hue is enabled by default. However, you must set the path to your S3 directory in the hue-safety-valve field to avoid a 403 error when you click on the S3 File Browser.

Procedure

1. Sign in to Cloudera Data Warehouse DWAdmin or DWUser.
2. Go to the Virtual Warehouse from which you want to access the S3 buckets and click .
3. Go to CONFIGURATIONS Hue and select hue-safety-valve from the Configuration files drop-down menu.
4. Add the path to your S3 bucket under the [filebrowser] section as follows:

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

   (Optional) Per-user home directories are created by default. To disable automatic user directory creation, you can add the following lines in the hue-safety-valve as follows:

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

5. Click APPLY.
   You should be able to view the icon for the S3 File Browser on the left assist panel on the Hue web interface.

Accessing S3 bucket from Hue in Cloudera Data Warehouse without RAZ

To enable access to S3 buckets from Hue without RAZ, you must have onboarded to Cloudera on cloud and must meet the requirements listed in this section.

Only Hue superusers can view and access the S3 File Browser.

Creating roles and synchronizing users to FreeIPA

You must be an EnvironmentAdmin or EnvironmentUser to browse S3 buckets and create tables by importing CSV files. You must also synchronize users to FreeIPA. This is required for users accessing the Data Lake.

About this task

To assign roles to users, see Assigning resource roles to users in the *Cloudera Management Console* documentation.

To synchronize users to FreeIPA, see Performing user sync in the *Cloudera Management Console* documentation.

Related Information

Assigning resource roles to users
Performing user sync

Adding an external S3 bucket to your Cloudera Data Warehouse environment

If you try to access an external S3 bucket from the Hue web interface without adding it to the Cloudera Data Warehouse environment, then Impala or Hive may display the “AccessDeniedException 403” exception. Make sure that your Cloudera Data Warehouse environment has access to the S3 buckets that you want to access from Hue.

About this task

When you create a Virtual Warehouse in the Cloudera Data Warehouse service, a cluster is created in your AWS account. This cluster has two buckets. One bucket is used for managed data and the other is used for external data. Access to these two buckets is controlled by AWS instance profiles.

To add read/write access to external S3 buckets that reside in the same AWS account as the Cloudera Data Warehouse service cluster or that are different from the account where the Cloudera Data Warehouse service cluster resides, see the corresponding links in the Related information section.

Procedure

1. Sign in to the Cloudera Management Console as an administrator.
2. Go to Cloudera Data Warehouse Environments and click the More… menu.
3. Search and locate the environment in which you want to add the S3 bucket and click the edit icon.
   The **Environment Details** page is displayed.
4. Specify the name of the S3 bucket you want to configure access to in the Add External S3 Bucket text box.
   If the bucket belongs to another AWS account, then select the Bucket belongs to different AWS account option.
5. Select the access mode.
   Read-only access is sufficient to import data in Hue.
6. Click Add Bucket to save the configuration.
   A success message is displayed.
7. Click APPLY to update the Cloudera Data Warehouse environment.

Tip: If you configure read only access to an external S3 bucket, there is no need to restart Virtual Warehouses. However, if you configure read/write access to an external S3 bucket, you must restart Virtual Warehouses by suspending them and starting them again.

Related Information

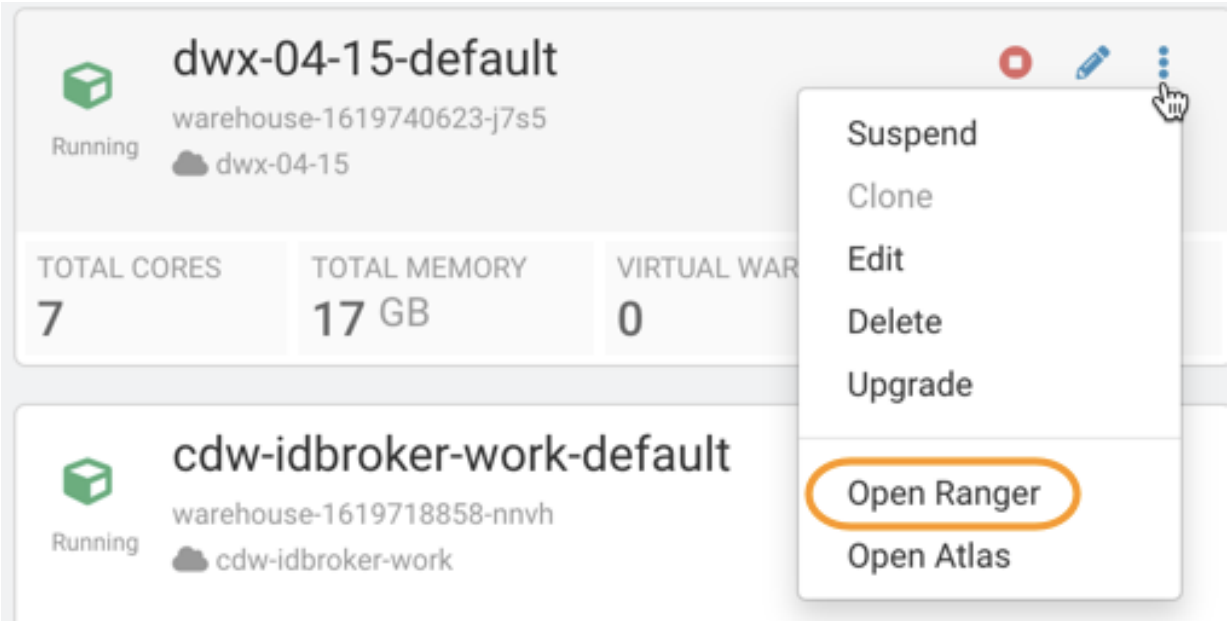
Adding Cloudera Data Warehouse cluster access to external S3 buckets in the same AWS account
Adding Cloudera Data Warehouse cluster access to external S3 buckets in a different AWS account
AWS instance profiles

Adding users to Hadoop SQL Ranger policies

You must grant the Hadoop SQL Ranger permissions to enable your users to access specific tables and secure your data from unauthorized access.

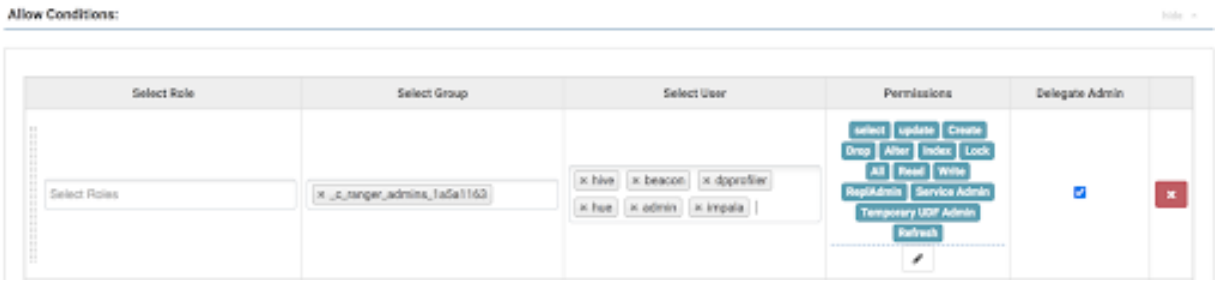
Procedure

1. Sign in to Cloudera Data Warehouse.
2. Click the ⋮ Open Ranger option on your Database Catalog.



3. On the **Ranger Service Manager** page, click Hadoop SQL.
4. Select the all - url policy.

   The **Edit Policy** page is displayed.
5. Under the Add Conditions section, add the users under the Select User column and add permissions such as Create, Alter, Drop, Select, and so on from the Permissions column.



Tip: To grant permissions to all users, you can specify {USER} in the Select User column.

6. Scroll to the bottom of the page and click Save.

Enabling the S3 File Browser for Hue in Cloudera Data Warehouse without RAZ

To enable access to S3 buckets from the Hue web interface in a non-RAZ environment, you must add the AWS environment details in the hue-safety-valve configuration from your Virtual Warehouse. After enabling the S3 File Browser, you can browse the S3 buckets, create folders, and upload files from your computer, and import files to create tables.

Procedure

1. Sign in to Cloudera Data Warehouse.
2. Go to the Virtual Warehouse from which you want to access the S3 buckets and click  .
3. On the **Virtual Warehouses** detail page, click the Hue tab and select hue-safety-valve from the drop-down menu.
4. Add the following configuration for Hive or Impala Virtual Warehouse in the space provided:

   For the Hive Virtual Warehouse:

   ```
   [desktop]
   # Remove the file browser from the blocked list of apps.
   # Tweak the app_blacklist property to suit your app configuration.
   app_blacklist=oozie,search,hbase,security,pig,sqoop,spark,impala

   [aws]
   [[aws_accounts]]
   [[[default]]]
   access_key_id=[***AWS-ACCESS-KEY***]
   secret_access_key=[***SECRET-ACCESS-KEY***]
   region=[***AWS-REGION***]
   [filebrowser]
   # (Optional) To set a specific home directory path:
   remote_storage_home=s3a://[***S3-BUCKET-NAME***]
   ```

   For Impala Virtual Warehouse:

   ```
   [desktop]
   # Remove the file browser from the blocked list of 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***]

   [filebrowser]
   # (Optional) To set a specific home directory path:
   remote_storage_home=s3a://[***S3-BUCKET-NAME***]
   ```

5. Click APPLY.

   The S3 File Browser icon appears on the left Assist panel on the Hue web interface after the Virtual Warehouse restarts.

Creating tables by importing CSV files from AWS S3 in

You can create tables in Hue by importing CSV files stored in S3 buckets. Hue automatically detects the schema and the column types, thus helping you to create tables without using the CREATE TABLE syntax.

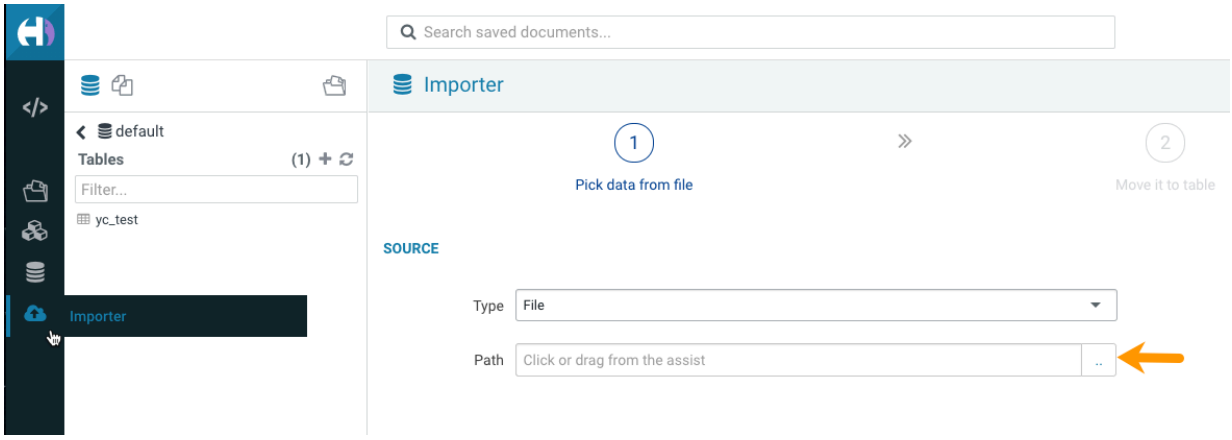
About this task

The maximum file size supported is three gigabytes.

(Non-RAZ deployment) Only Hue Superusers can access S3 buckets and import files to create tables. To create tables by importing files from S3, you must assign and authorize use of a specific bucket on S3 bucket for your environment. The bucket then appears like a home directory on the Hue web interface.

Procedure

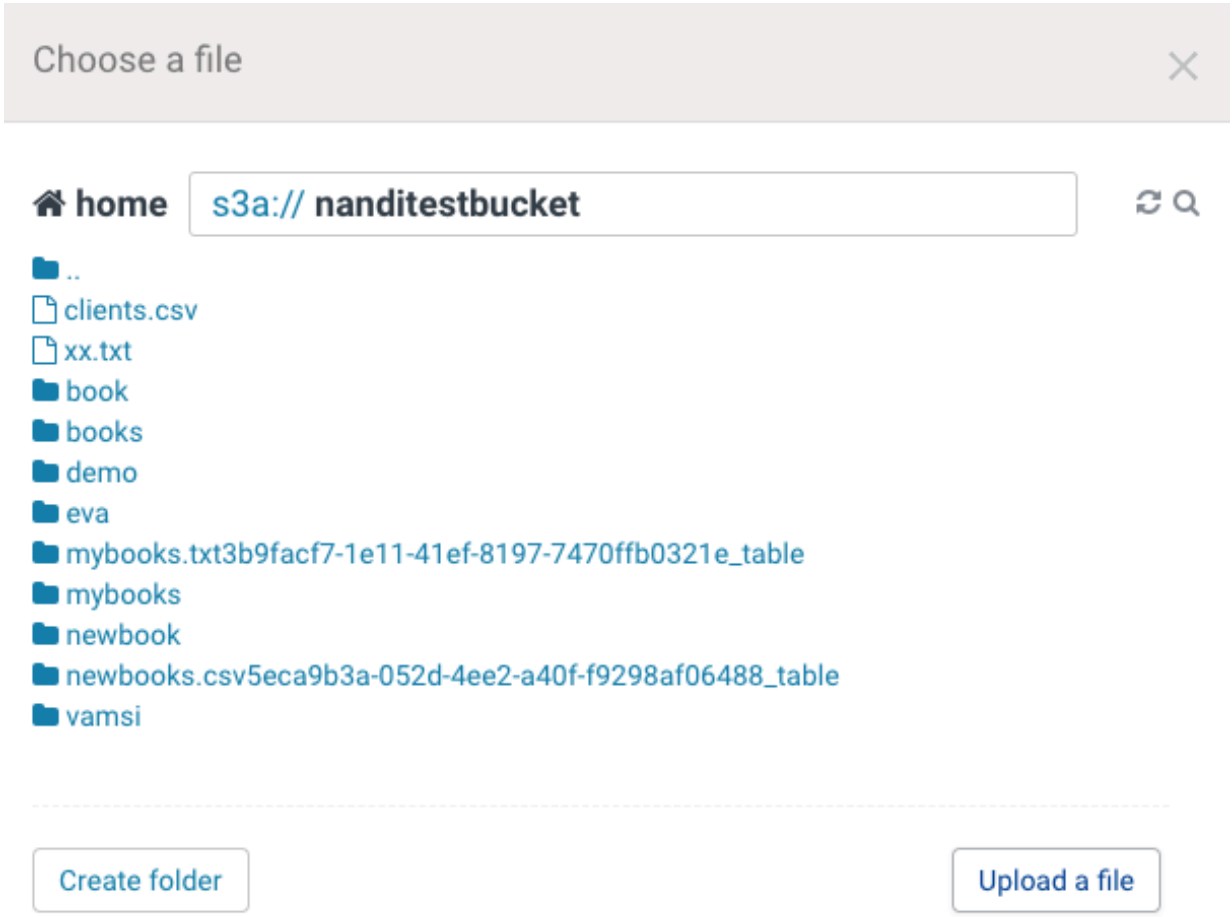
1. Log in to the service.
2. On the **Overview** page, select the Virtual Warehouse in which you want to create the table and click on Hue.
3. From the left assist panel, click on Importer.
4. On the Importer screen, click .. at the end of the Path field:



Choose a file pop-up is displayed.

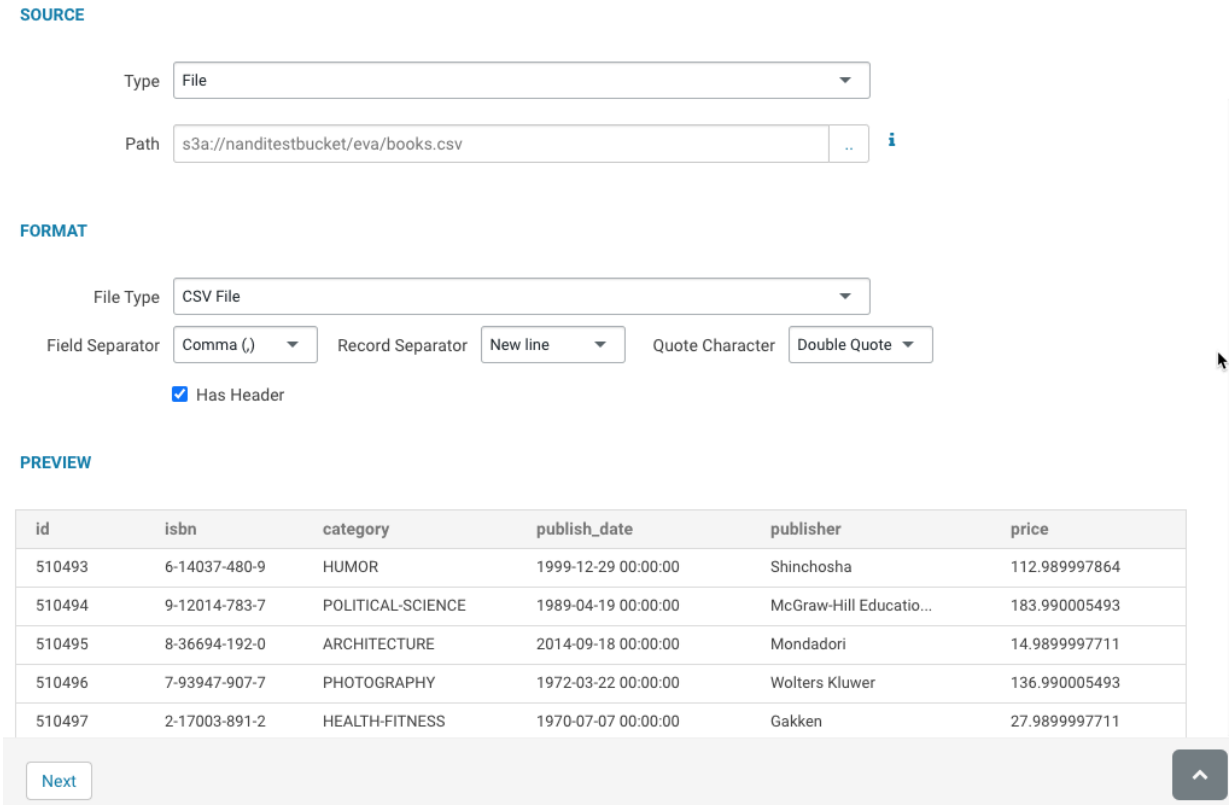
5. (Non-RAZ deployment) Type s3a:// in the address text box and press enter.

   The S3 buckets associated with the environment are displayed. You can narrow down the list of results using the search option.



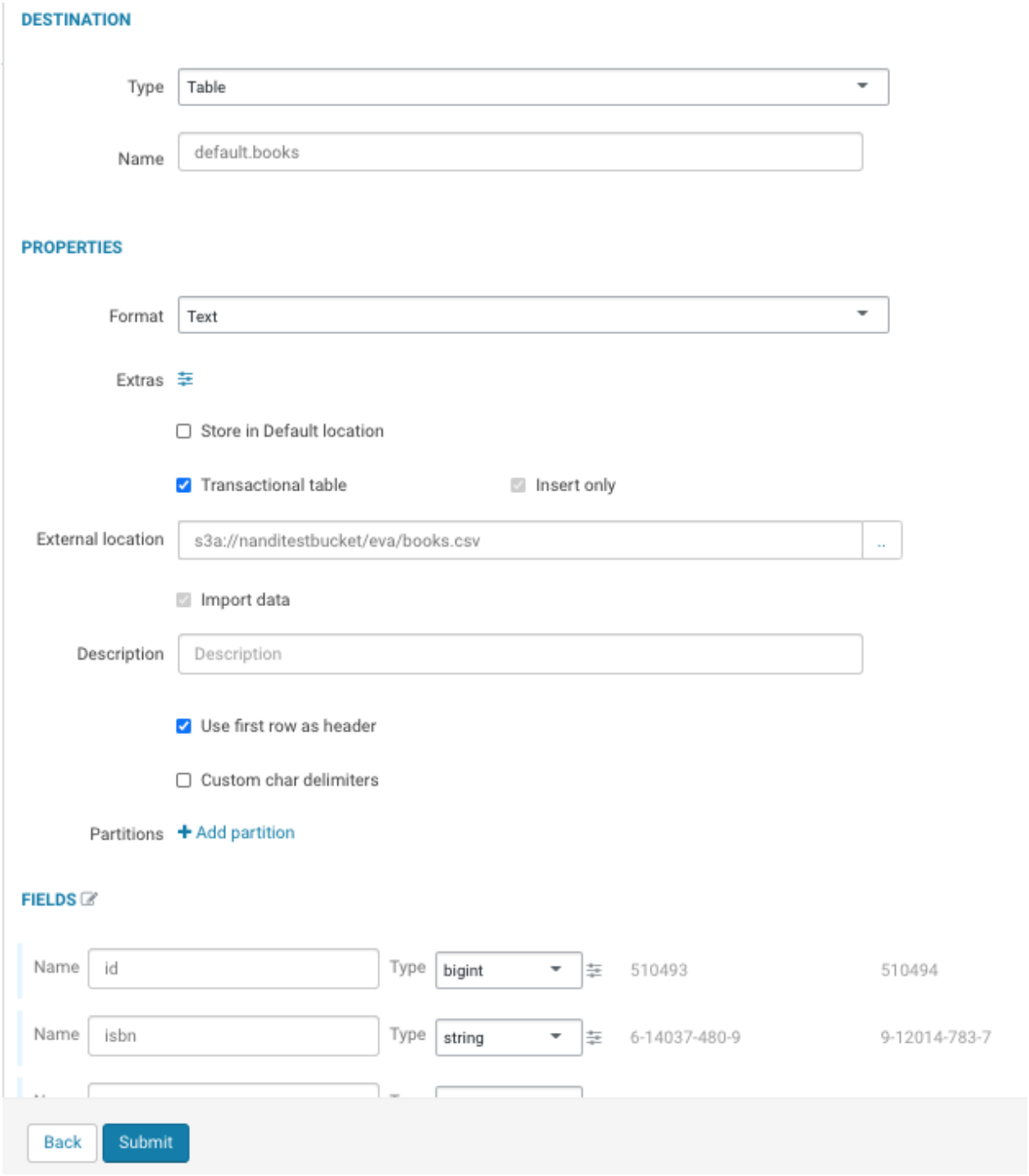
   If the file is present on your computer, then you can upload it to S3 by clicking Upload a file. To do this, you must have enabled read/write access to the S3 bucket from the environment.

6. Select the CSV file that you want to import into Hue.
   Hue displays the preview of the table along with the format:

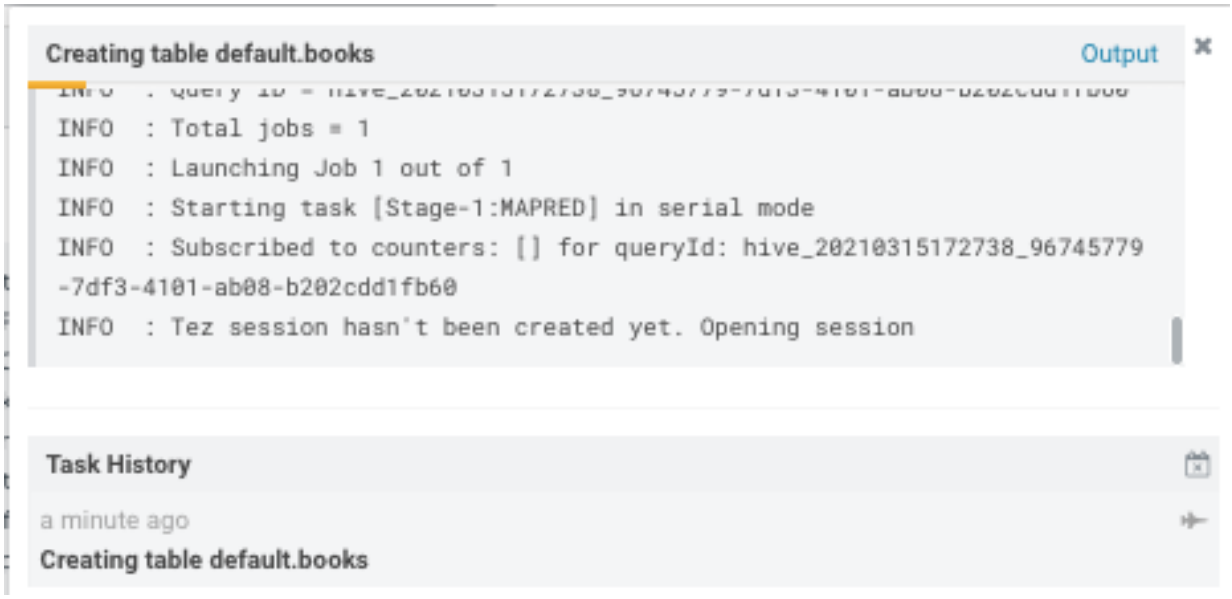


Hue automatically detects the field separator, record separator, and the quote character from the CSV file. If you want to override a specific setting, then you can change it by selecting a different value from the drop-down menu.

7. Click Next.
   On this page, you can set the table destination, partitions, and change the column data types.



8. 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.

Using Azure Data Lake Storage Gen2 with Hue

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

In Cloudera Data Warehouse, you can browse ADLS Gen2 storage from Hue in the following ways:

- With Ranger Authorization Service (RAZ)
- Without Ranger Authorization Service (RAZ)

Related Information

Registering a RAZ-enabled Azure environment

Accessing ADLS Gen2 containers from Hue in Cloudera Data Warehouse with RAZ

Hue offers you the capability to browse ADLS Gen2 containers, upload files to ADLS Gen2 containers, and create tables by importing files from ABFS. With Ranger Authorization (RAZ), you can grant fine-grained access to per-user home directories.

About this task

If you have enabled RAZ while registering your AWS environment with Cloudera, then Hue uses RAZ as the default mechanism for enabling the ABFS File Browser. Before you can enable the ABFS File Browser in Hue, you must complete the following prerequisites:

Procedure

1. Follow the instructions listed in Introduction to RAZ on Azure environments to register an Azure environment with the Enable Ranger authorization for ADLS Gen2 option enabled. You can use the Cloudera web interface or the Cloudera CLI to complete this task.

   Note: You must enable RAZ while registering your environment with Cloudera.

2. Log in to the Cloudera Management Console as a DWAdmin or DWUser and go to the Cloudera Data Warehouse service.
3. Click ⋮ Open Ranger on your Database Catalog.
4. Create the following Ranger policies:
   a) Hadoop SQL policy (all - database, table, column, all - url).
      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.
   b) ABFS (cm_ADLS) policy (Default: User Home)
      You must grant permissions to the following users in the Permissions section for the user home directory: {USER}.
   c) ABFS (cm_ADLS) policy (Default: user)
      You must grant permissions to the following users in the Permissions section for the root directory (/user/): hive, impala.
5. You must also grant appropriate permissions to the users in Cloudera User Management Service (UMS). For example, EnvironmentUser.
6. 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.

Enabling the ABFS File Browser for Hue in Cloudera Data Warehouse with RAZ

The ABFS File Browser in Hue is enabled by default. However, you must set the path to your ADLS Gen2 container in the hue-safety-valve field to avoid a 403 error when you click on the ABFS File Browser.

Procedure

1. Sign in to Cloudera Data Warehouse DWAdmin or DWUser.
2. Go to the Virtual Warehouse from which you want to access the ADLS Gen2 containers and click ✎.
3. Go to CONFIGURATIONS Hue and select hue-safety-valve from the Configuration files drop-down menu.
4. Add the path to your ADLS Gen2 container under the [filebrowser] section as follows:

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

   (Optional) Per-user home directories are created by default. To disable automatic user directory creation, you can add the following lines in the hue-safety-valve as follows:

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

5. Click APPLY.
   You should be able to view the icon for the ABFS File Browser on the left assist panel on the Hue web interface.

Accessing ADLS Gen2 containers from Hue in Cloudera Data Warehouse without RAZ

To enable access to Azure Data Lake Storage (ADLS) Gen2 containers from Hue, you must have onboarded to Cloudera on cloud and must meet the requirements listed in this section.

Only Hue superusers can view and access the ABFS File Browser.

Creating an Azure storage account

You need an Azure storage account to use ABFS with Hue.

Procedure

1. Sign in to the Microsoft Azure portal as an administrator.
2. On the Create storage account Advanced page, enable Data Lake Storage Gen2 so that the objects and files within your account can be organized into a hierarchy of directories and nested subdirectories in the same way that the file system on your computer is organized.

Setting storage location base

You must specify the Storage Location Base to configure a default ADLS Gen2 base storage location for the Cloudera environment when you register your Azure environment with Cloudera.

About this task

While registering an Azure environment Cloudera Management Console, set the Storage Location Base in the Data Access section as follows:

```
abfs://storage-fs@[***AZURE-STORAGE-ACCOUNT-NAME***].dfs.core.windows.net
```

This location is used to read and store data.

Enabling the ABFS File Browser

To enable access to ADLS Gen2 containers from the Hue web interface, you must add the Azure environment details in the hue-safety-valve configuration from your Virtual Warehouse. After enabling the ABFS File Browser, you can browse the ADLS Gen2 containers, create folders, and upload files from your computer, and import files to create tables.

Procedure

1. Sign in to Cloudera Data Warehouse.
2. Go to the Virtual Warehouse from which you want to access the ADLS Gen2 containers and click  .
3. On the **Virtual Warehouses** detail page, click the Hue tab and select hue-safety-valve from the drop-down menu.
4. Add the following configuration for Hive or Impala Virtual Warehouse in the space provided:

   For the Hive Virtual Warehouse:

   ```
   [desktop]
   # Remove the file browser from the blocked list of apps.
   # Tweak the app_blacklist property to suit your app configuration.
   app_blacklist=oozie,search,hbase,security,pig,sqoop,spark,impala
   [azure]
     [[azure_accounts]]
       [[[default]]]
         client_id=[***AZURE-ACCOUNT-CLIENT-ID***]
         client_secret=[***AZURE-ACCOUNT-CLIENT-SECRET***]
         tenant_id=[***AZURE-ACCOUNT-TENANT-ID***]

       [[abfs_clusters]]
         [[[default]]]
     fs_defaultfs=abfs://[***CONTAINER-NAME***]@[***AZURE-STORAGE-ACCOUNT-N
   AME***]>.dfs.core.windows.net
                webhdfs_url=https://[***AZURE-STORAGE-ACCOUNT-NAME***].dfs.c
   ore.windows.net/
   ```

   For Impala Virtual Warehouse:

   ```
   [desktop]
   # Remove the file browser from the blocked list of 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=[***AZURE-ACCOUNT-CLIENT-ID***]
      client_secret=[***AZURE-ACCOUNT-CLIENT-SECRET***]
      tenant_id=[***AZURE-ACCOUNT-TENANT-ID***]

    [[abfs_clusters]]
      [[[default]]]
  fs_defaultfs=abfs://[***CONTAINER-NAME***]@[***AZURE-STORAGE-ACCOUNT-NA
ME***]>.dfs.core.windows.net
              webhdfs_url=https://[***AZURE-STORAGE-ACCOUNT-NAME***].df
s.core.windows.net/
```

Make sure that the container name and the Azure storage account name that you specify under the abfs_clusters section is same as what you specified under Data Access Storage Location Base while activating the Azure environment, so that Hive or Impala has permission to access the uploaded files.

5. Click APPLY.
   The ABFS File Browser icon appears on the left Assist panel on the Hue web interface after the Virtual Warehouse restarts.

Creating tables by importing CSV files from ABFS

You can create tables in Hue by importing CSV files stored in ABFS. Hue automatically detects the schema and the column types, thus helping you to create tables without using the CREATE TABLE syntax.

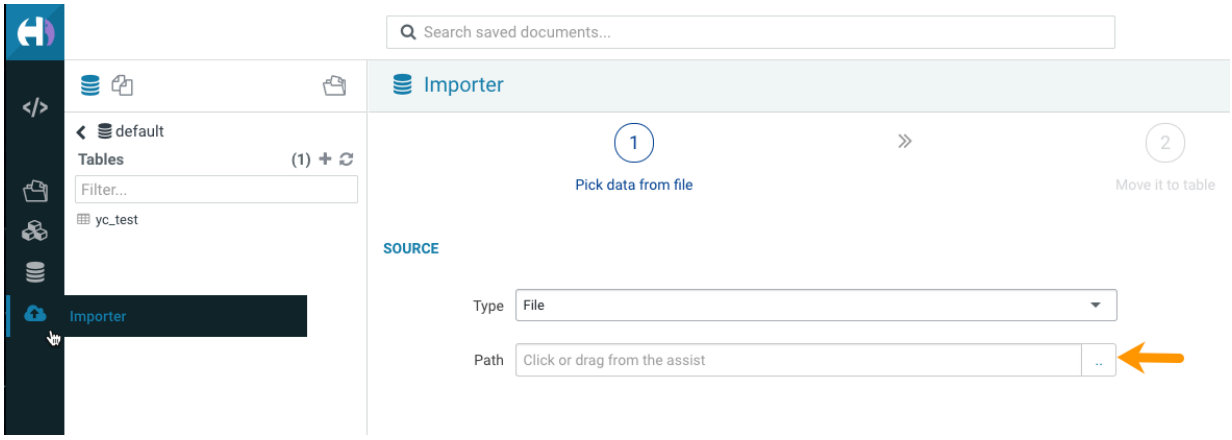
About this task

The maximum file size supported is three gigabytes.

(Non-RAZ deployment) Only Hue Superusers can access ADLS Gen2 containers and import files to create tables. To create tables by importing files from ADLS, you must assign and authorize use of a specific bucket on ADLS Gen2 containers for your environment. The bucket then appears like a home directory on the Hue web interface.

Procedure

1. In the service **Overview** page, select the Virtual Warehouse in which you want to create the table, click the options menu in the upper right corner and click Open Hue.
2. From the left assist panel, click on Importer.
3. On the **Importer** screen, click .. at the end of the Path field:



Choose a file pop-up is displayed.

4. (Non-RAZ deployment) Type abfs://[***CONTAINER-NAME***] in the address text box and press enter.

   The ABFS containers created under the Azure storage account are displayed.

   You can narrow down the list of results using the search option.

   If the file is present on your computer, then you can upload it to ABFS by clicking Upload a file.
5. Select the CSV file that you want to import into Hue.

   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 CSV file. If you want to override a specific setting, then you can change it by selecting a different value from the drop-down menu.
6. Click Next.

   On this page, you can set the table destination, partitions, and change the column data types.
7. 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.

Granting permission to access S3 and ABFS File Browser in Hue

Only admin users can view and access S3 or ABFS 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 and ABFS 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 or ABFS File Browsers.

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

Procedure

1. Open Hue from Cloudera Data Warehouse Virtual Warehouse as an EnvironmentAdmin.
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 Data 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.

Uploading files using Hue with the task server enabled in Cloudera Data Warehouse

The task server is enabled by default. When task server is in the enabled state, the files are uploaded using an asynchronous task queue or job queue. This improves performance and allows you to upload multiple files as large as 5 GB each in parallel.

Before you begin

You must have access to the cloud storage (AWS S3 or Azure ADLS Gen2).

Procedure

1. Log in to the Hue web interface as a normal user.
2. Go to the Hue File Browser.
3. Click Schedule Upload.
4. Click Select files to browse the files from your local system, and then click Upload.

   Alternatively, you can drag and drop the file into the file upload dialogue box.

Note: 5 GB is the maximum supported upload file size per file.

List of supported non-alphanumeric characters for file and directory names in Hue

Auto-generated files may often introduce non-alphanumeric characters in the filenames which are not supported by Hue. This might cause the files or directories to not appear on the Hue File Browser. Review the list of non-alphanumeric characters supported in Hue to avoid running into this issue.

Table 1: Non-alphanumeric characters supported in Hue

Special character symbol	Description
~	Tilde
@	Ampersat
#	Hash
$	Dollar sign
&	Ampersand
(	Left paranthesis
)	Right paranthesis

Special character symbol	Description
*	Asterisk
!	Exclamation mark
+	Plus
=	Equal
:	Colon Not supported with Knox.
;	Semicolon
,	Comma
.	Period
?	Question mark Not supported with Knox.
/	Forward slash Not supported with Knox.
\	Backslash
'	Apostrophe or single quote

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 Cloudera. Cloudera recommends that you raise issues in the open-source github community.