

Cloudera Runtime 7.2.6

Using Hue

Date published: 2020-07-28

Date modified: 2020-12-11

CLOUDERA

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

Legal Notice

© Cloudera Inc. 2024. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

Using Hue.....	4
Enabling the SQL editor autocompleter.....	10
Using governance-based data discovery.....	11
Searching metadata tags.....	11
Using Amazon S3 with Hue.....	12
Enabling S3 browser for Hue configured with IDBroker.....	12
Enabling S3 browser for Hue configured without IDBroker.....	13
Populating an S3 bucket.....	14
Creating a table from an Amazon S3 file.....	14
Exporting query results to Amazon S3.....	15
Using Azure Data Lake Storage Gen2 with Hue.....	17
Enabling ABFS file browser for Hue configured with IDBroker.....	17
Enabling ABFS file browser for Hue configured without IDBroker.....	18
List of supported non-alphanumeric characters for file and directory names in Hue.....	19
Granting permission to access S3 and ABFS File Browser in Hue.....	20

Using Hue

Get started using Hue by analyzing and visualizing your data with Impala, a high-speed, low-latency SQL query engine.

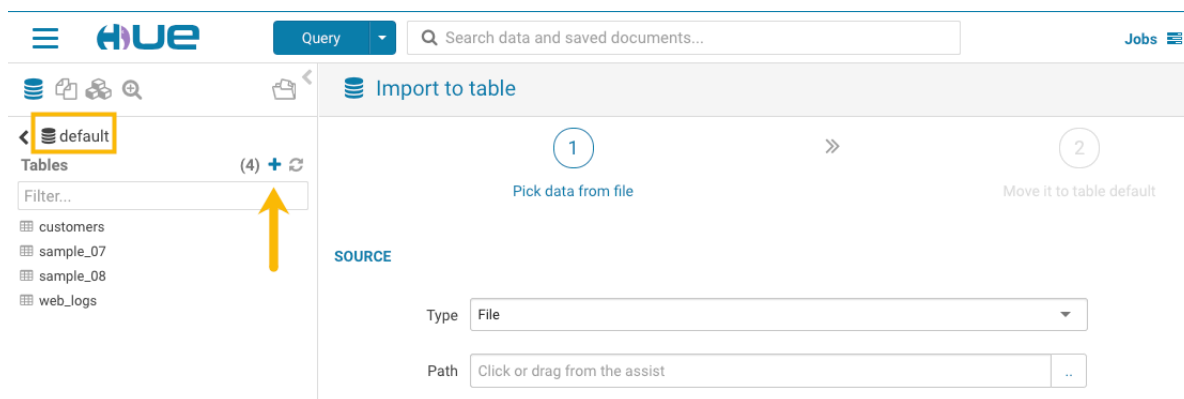
About this task

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.

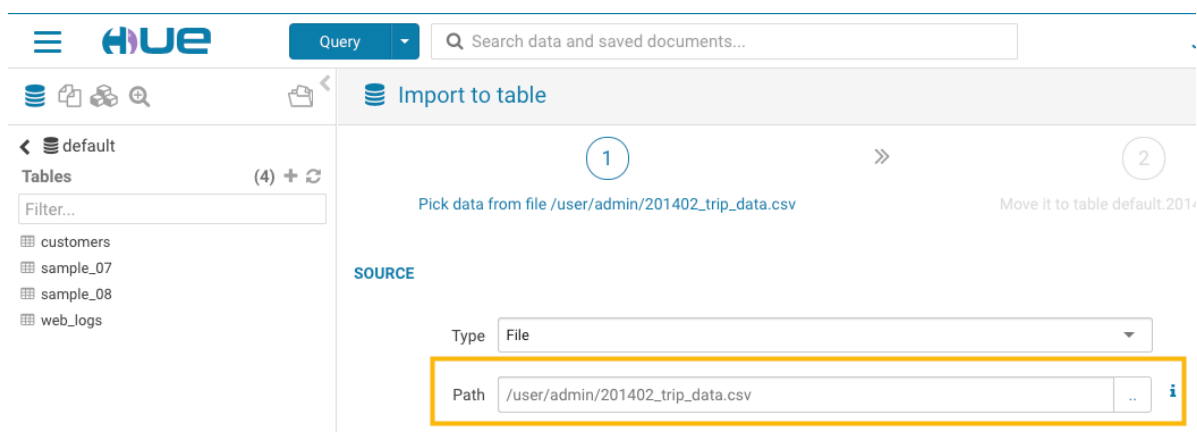
2. Create a table from the `~/babs_open_data_year_1/201402_babs_open_data/201402_trip_data.csv` file found in the unzipped `babs_open_data_year_1.zip` file:
 - a) In the Cloudera Manager Admin Console, select HueWebUIHue Load Balanced to launch Hue.
 - b) In the left navigation panel of Hue, make sure 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.

that is next to the database icon in the

- 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.

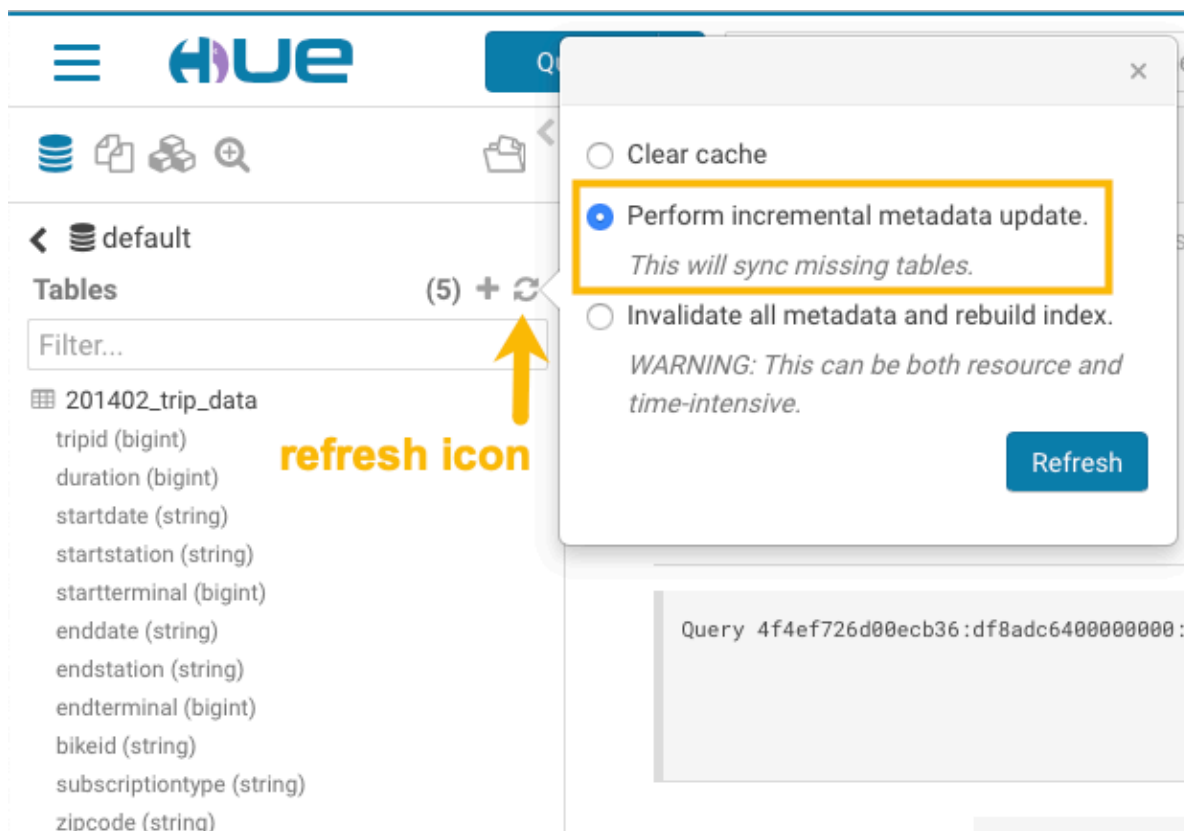
3. Click Query at the top of the page and select EditorHive to open the Hive editor and then create a query.
- Enter the following query into the editor window:

```
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

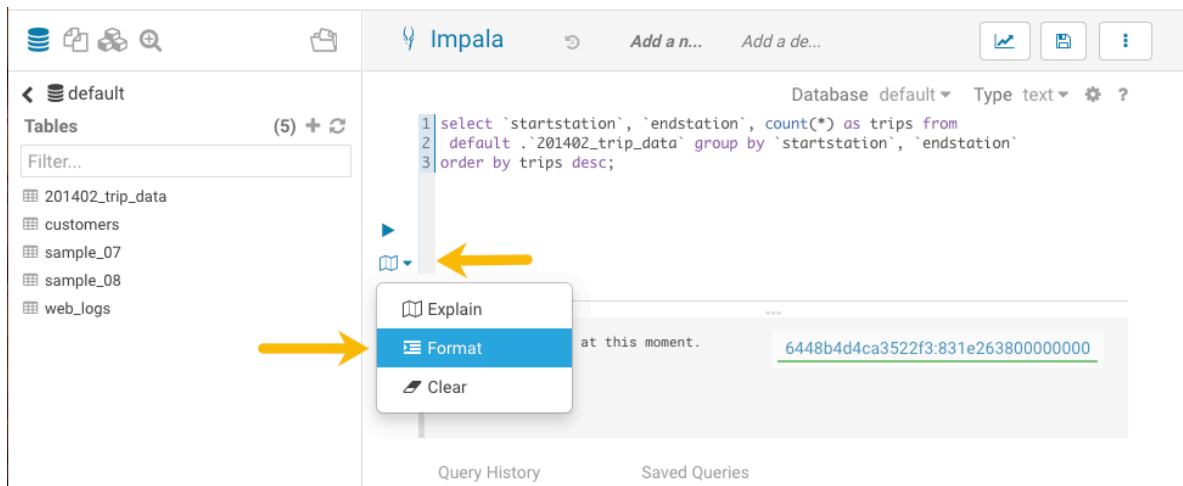
4. Click Query at the top of the page and select EditorImpala to open the Impala SQL editor and then create a query.
 - a. In the left panel, click the refresh icon and select Perform incremental metadata update to make the new table visible to Impala:



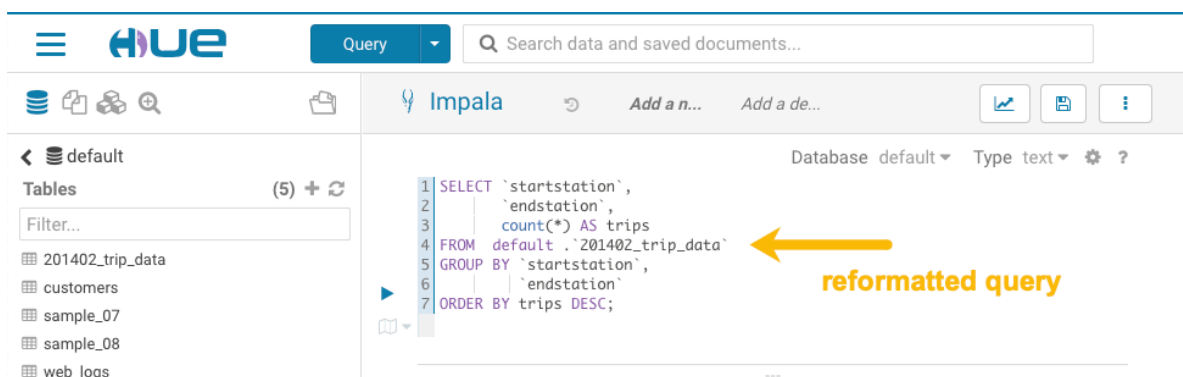
- b. Enter the following query into the editor window:

```
select 'startstation', 'endstation', count(*) as trips from default.'201402_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.

Impala

0.66s Database default Type text

```

1 SELECT `startstation`,
2        `endstation`,
3        count(*) AS trips
4 FROM default.`201402_trip_data`
5 GROUP BY `startstation`,
6          `endstation`
7 ORDER BY trips DESC;

```

Query 4142ddc7e9c4b8ad:4af0d8b600000000: 0% Complete (0 out of 1)

4142ddc7e9c4b8ad:4af0d8b600000000

Query History Saved Queries Results (1,024+)

COLUMNS (4)

- ☒ startstation string
- ☒ endstation string
- ☒ trips bigint

Bar chart options:

- ☒ Bars
- ☐ Pie
- ☐ Scatter
- ☐ Marker Map
- ☐ Gradient Map

	startstation	endstation
1	Harry Bridges Plaza (Ferry Building)	Embarcadero at Sansome
2	Townsend at 7th	San Francisco Caltrain (Townsend at 4th)
3	San Francisco Caltrain 2 (330 Townsend)	Townsend at 7th
4	Market at Sansome	2nd at South Park
5	Embarcadero at Sansome	Steuart at Market
6	2nd at South Park	Market at Sansome
7	San Francisco Caltrain (Townsend at 4th)	Harry Bridges Plaza (Ferry Building)
8	2nd at Townsend	Harry Bridges Plaza (Ferry Building)

b. Set the bar chart elements as follows:

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

TYPE: Bars

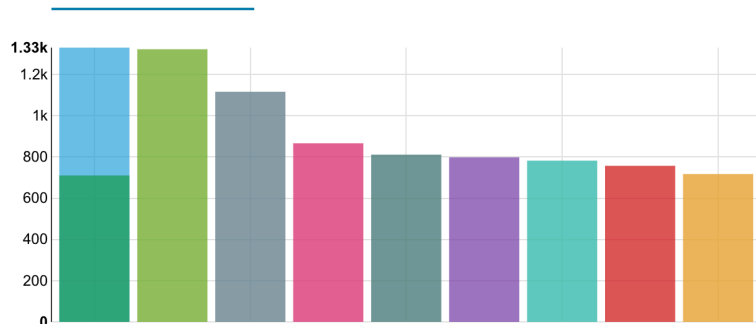
X-AXIS: startstation

Y-AXIS: ☒ trips

GROUP: Choose a column to pivo...

LIMIT: 10

SORTING: ☒ Ascending ☐ Descending ☐ Grouped



6.

Create a pie chart by clicking the chart icon again



and then select Pie.

7.

Download the query results by clicking the download icon



and selecting in what format you want to download, copy, or export the results.

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-,

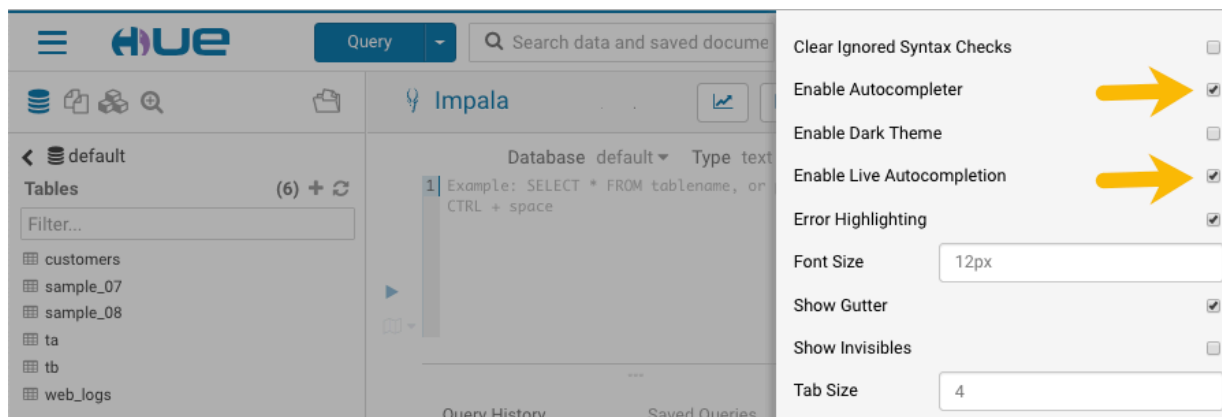
- On a Windows system, use the Ctrl key followed by a hyphen and then a comma:

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.

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.

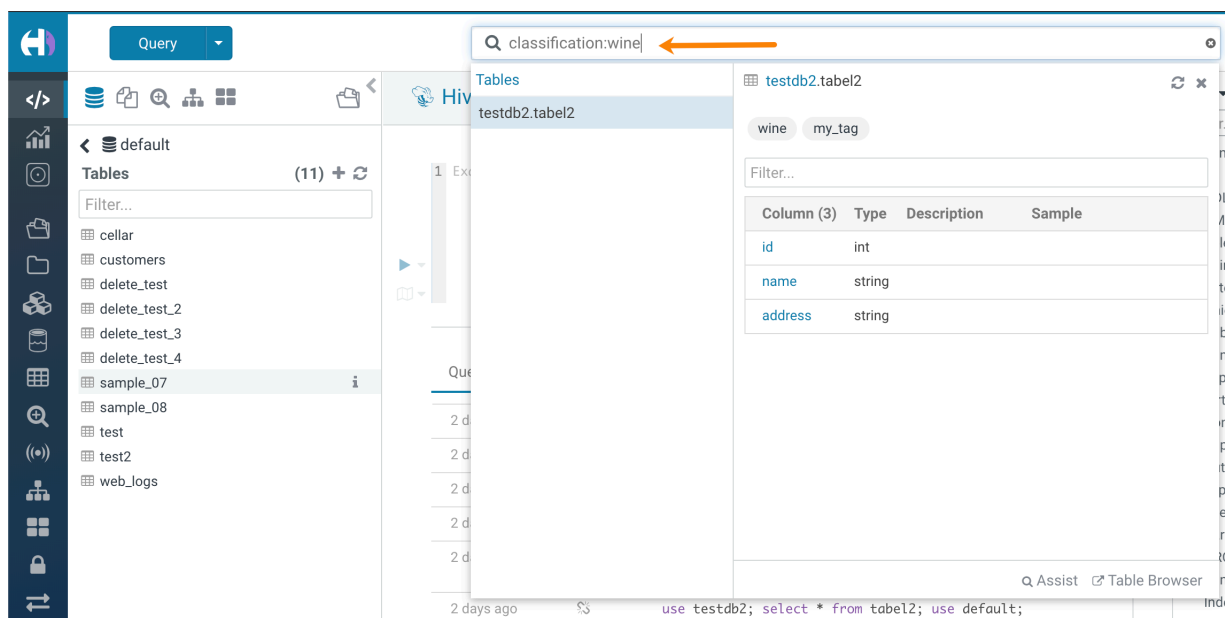


Note: On clusters that use Apache Ranger for role-based access control, the Search mechanism does not display counts of popular values. Ranger ensures that Hue users can view only entities to which their user role (as configured and managed by Ranger) has been granted specific permissions.

Procedure

1. Go to Query Editor Impala or Hive.

- 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.

Only Hue superusers can view and access the S3 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,security
[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,security
```

```
[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.

Populating an S3 bucket

Use the Hue Web UI to populate buckets in Amazon S3.

About this task

Use open data from the U.S. Geological Survey to demonstrate how to populate and S3 bucket with Hue.

Procedure

1. Download [30 days of earthquake data](#) (`all_month.csv`) from the [USGS](#) (~2 MB).
2. Log on to the Hue Web UI from Cloudera Manager.
3. Select File BrowserS3 Browser.
4. Click NewBucket, name it "quakes_<any unique id>" and click Create.



Tip: Unique bucket names are important per S3 [bucket naming conventions](#).

5. Navigate into the bucket by clicking the bucket name
6. Click NewDirectory, name it "input" and click Create.
7. Navigate into the directory by clicking the directory name.
8. Click Upload and select, or drag, `all_month.csv`. The path is `s3a://quakes/input/all_month.csv`.




Important: Do not add anything else to the "input" directory—no extra files, no directories.

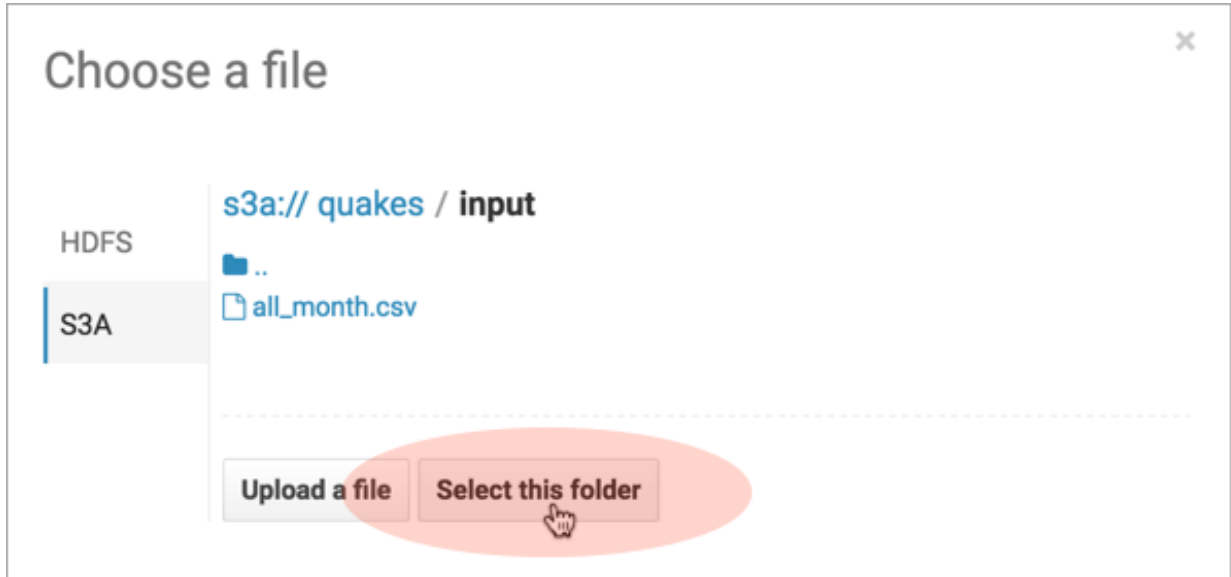
Creating a table from an Amazon S3 file

Using Hue to create a table from an Amazon S3 file streamlines the process.

Procedure

1. Go to the Metastore Manager by clicking Data BrowsersMetastore Tables.
2. Create a new table from a file by clicking .
3. Enter a Table Name such as "earthquakes".

4. Browse for the Input Directory, s3a://quakes/input/, and click Select this folder.



5. Select Create External Table from the Load Data menu and click Next.
6. Delimit by Comma(,) and click Next.
7. Click Create Table.
8. Click the Browse Data icon to automatically generate a SELECT query in the Hive editor:

```
SELECT * FROM `default`.`earthquakes` LIMIT 10000;
```

Exporting query results to Amazon S3

Use Hue to export query results to Amazon S3 as a custom file, a MapReduce file, or as a table.

Procedure

1. Run and Export Results in Hive

a)

Run the query by clicking Execute

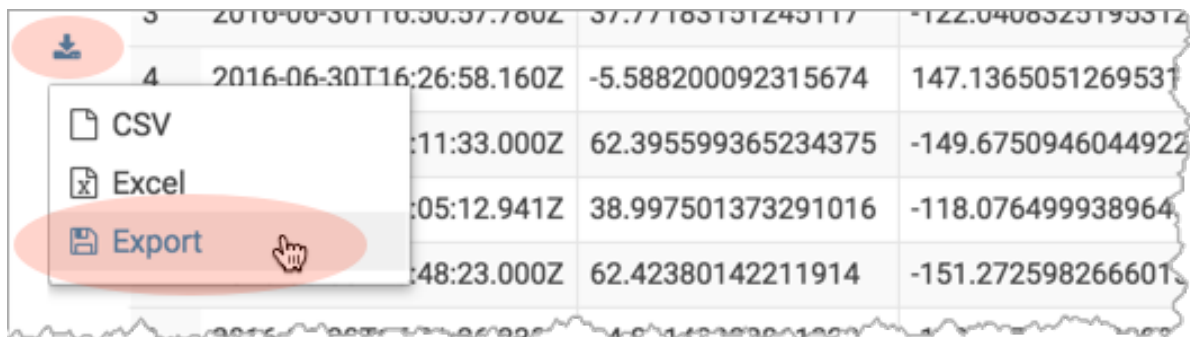


b)



Click Get Results

c) Select Export to open the Save query result dialog.



2. Save Results as Custom File

a) Select In store (max 10000000 cells) and open the Path to CSV file dialog.

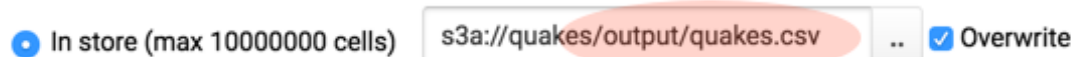
b) Navigate into the bucket, s3a://quakes.

c) Create folder named, "output."

d) Navigate into the output directory and click Select this folder.

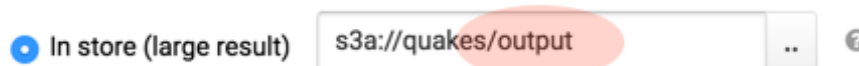
e) Append a file name to the path, such as quakes.csv.

f) Click Save. The results are saved as s3a://quakes/output/quakes.csv.



3. Save Results as MapReduce files

- Select In store (large result) and open the Path to empty directory dialog.
- Navigate into the bucket, s3a://quakes.
- If you have not done so, create a folder named, "output."
- Navigate into the output directory and click Select this folder.
- Click Save. A MapReduce job is run and results are stored in s3a://quakes/output/.

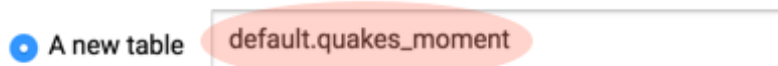
**4. Save Results as Table**

- Run a query for "moment" earthquakes and export:

```
SELECT time, latitude, longitude, mag
FROM `default`.`earthquakes`
WHERE magtype IN ( 'mw', 'mwb', 'mwc', 'mwr', 'mww' );
```

- Select A new table and input <database>.<new table name>.
- Click Save.
-

Click Browse Data  to view the new table.



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.

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,jobsup,pig,sqoop,security
[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 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,security
[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.

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.

The following table lists the supported non-alphanumeric characters in Hue:

Table 1: Non-alphanumeric characters supported in Hue

Special character symbol	Description
~	Tilde
@	Ampersat
#	Hash
\$	Dollar sign
&	Ampersand
(Left parenthesis
)	Right parenthesis
*	Asterisk
!	Exclamation mark

Special character symbol	Description
+	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

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. 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.