

Cloudera Runtime 7.1.3

## Using Hue

Date published: 2020-07-28

Date modified: 2020-08-10

# CLOUDERA

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

# Legal Notice

© Cloudera Inc. 2024. All rights reserved.

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

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

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

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

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

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

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

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

# Contents

<b>Using Hue.....</b>	<b>4</b>
<b>Enabling the SQL editor autoCompleter.....</b>	<b>9</b>
<b>Using governance-based data discovery.....</b>	<b>10</b>
Searching metadata tags.....	10
<b>Using Amazon S3 with Hue.....</b>	<b>11</b>
Populating an S3 bucket.....	11
Creating a table from an Amazon S3 file.....	11
Exporting query results to Amazon S3.....	12

## Using Hue

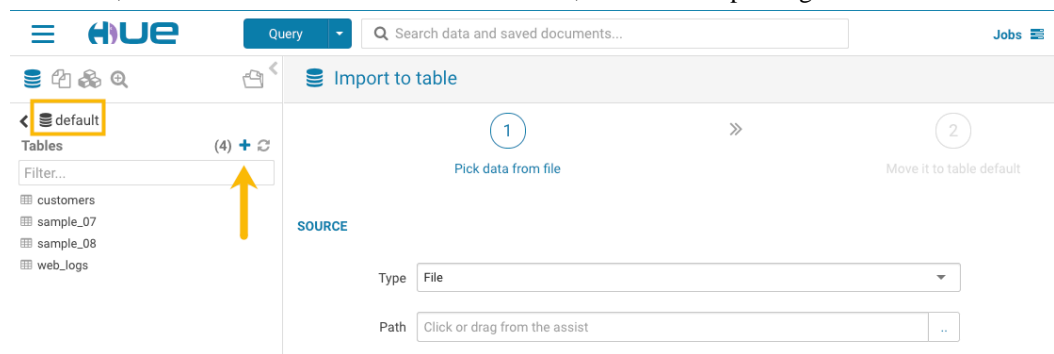
Get started using Hue by analyzing and visualizing your data with Impala, a high-speed, low-latency SQL query engine.

### About this task

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

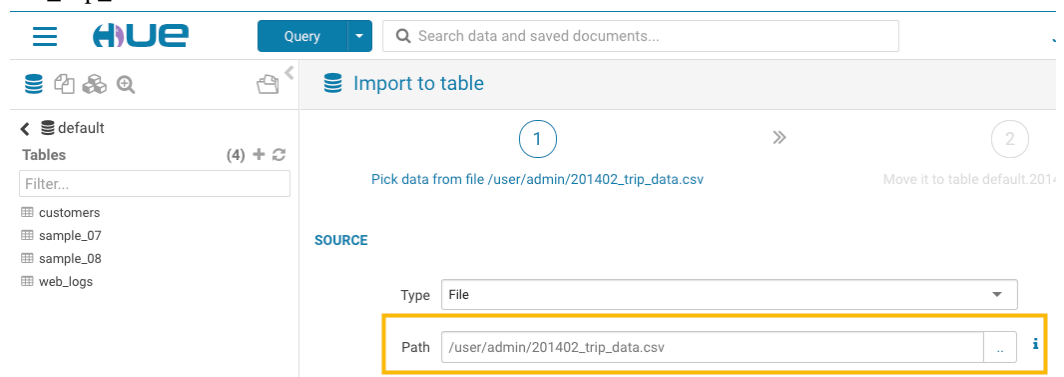
### Procedure

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 panel of Hue, make sure the default database is selected, and click the plus sign to create a table:



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:



- 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:

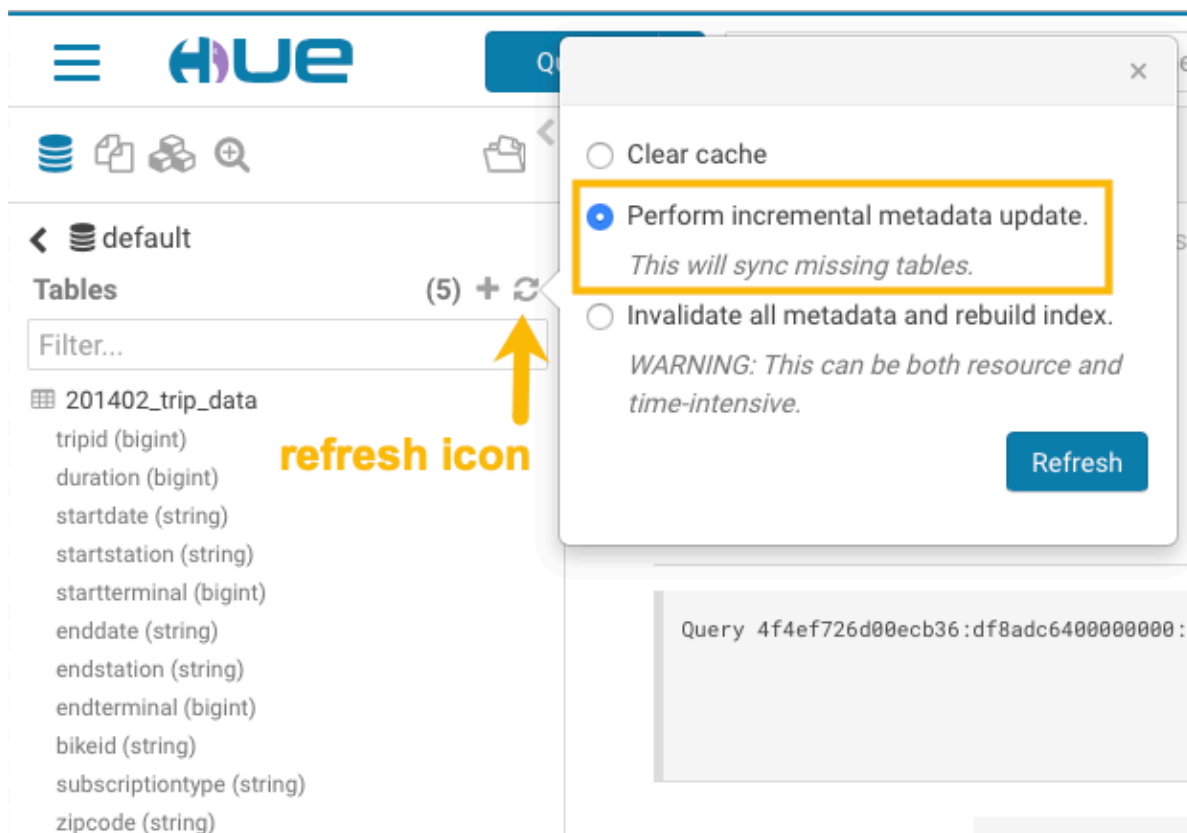
```
SELECT * FROM default.201402_trip_data
LIMIT 10;
```

- 

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

	Query History	Saved Queries	Results (10)																																																						
			<table><thead><tr><th></th><th>201402_trip_data.tripid</th><th>201402_trip_data.duration</th><th>201402_trip_data</th></tr></thead><tbody><tr><td></td><td>1</td><td>4576</td><td>63</td><td>8/29/2013 14:13</td></tr><tr><td></td><td>2</td><td>4607</td><td>70</td><td>8/29/2013 14:42</td></tr><tr><td></td><td>3</td><td>4130</td><td>71</td><td>8/29/2013 10:16</td></tr><tr><td></td><td>4</td><td>4251</td><td>77</td><td>8/29/2013 11:29</td></tr><tr><td></td><td>5</td><td>4299</td><td>83</td><td>8/29/2013 12:02</td></tr><tr><td></td><td>6</td><td>4927</td><td>103</td><td>8/29/2013 18:54</td></tr><tr><td></td><td>7</td><td>4500</td><td>109</td><td>8/29/2013 13:25</td></tr><tr><td></td><td>8</td><td>4563</td><td>111</td><td>8/29/2013 14:02</td></tr><tr><td></td><td>9</td><td>4760</td><td>113</td><td>8/29/2013 17:01</td></tr><tr><td></td><td>10</td><td>4258</td><td>114</td><td>8/29/2013 11:33</td></tr></tbody></table>		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
	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:

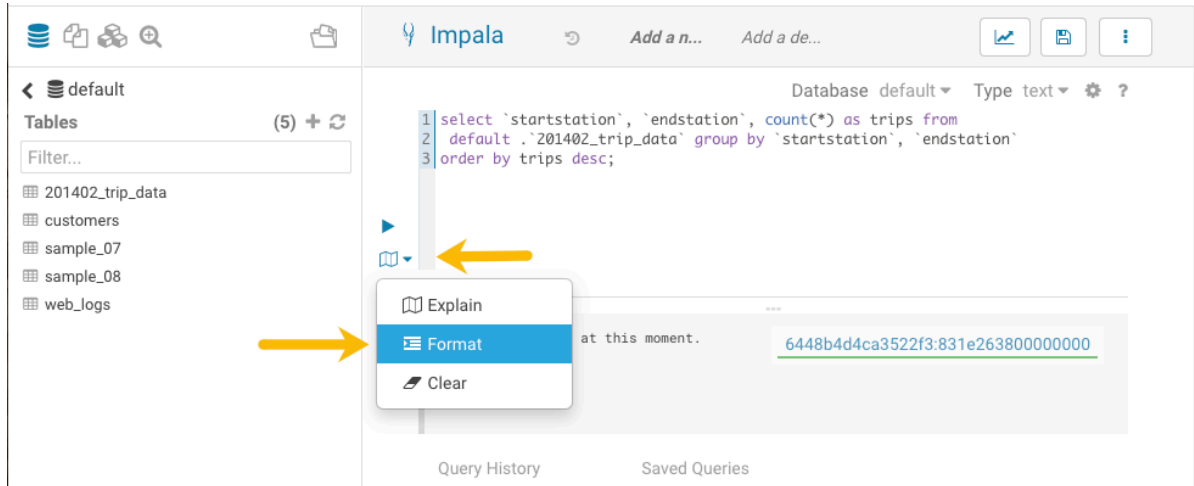


The screenshot shows the Hue interface with a modal dialog box open. The dialog box has three radio button options: 'Clear cache', 'Perform incremental metadata update.', and 'Invalidate all metadata and rebuild index.'. The 'Perform incremental metadata update.' option is selected and highlighted with a yellow box. Below this option is the text 'This will sync missing tables.'. The 'Invalidate all metadata and rebuild index.' option has a warning below it: 'WARNING: This can be both resource and time-intensive.'. A blue 'Refresh' button is at the bottom right of the dialog. In the background, the left panel shows a table list with a refresh icon (a circular arrow) next to '(5) +'. A yellow arrow points to this icon with the text 'refresh icon' written in yellow.

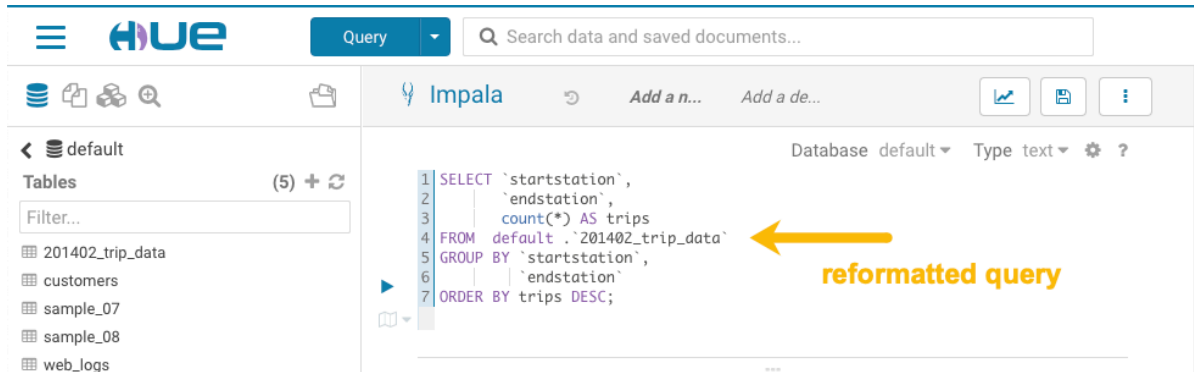
- b. Enter the following query into the editor window:

```
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 interface showing a query and its results. The query is:

```

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;
    
```

The results table shows the top 10 stations by trip count:

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)

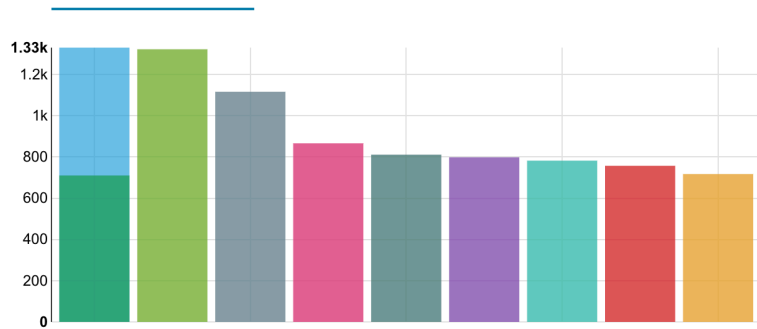
The 'Columns' panel on the left shows 'startstation' selected for the X-axis and 'trips' selected for the Y-axis. The 'Bars' chart type is highlighted in the dropdown menu.

b. Set the bar chart elements as follows:

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

Configuration panel for the bar chart:

- TYPE: Bars
- X-AXIS: startstation
- Y-AXIS: trips
- GROUP: Choose a column to pivo...
- LIMIT: 10
- SORTING: [Icons for sorting options]






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:

- For Macs, use the Command key followed by a hyphen and then a comma:

Command-,

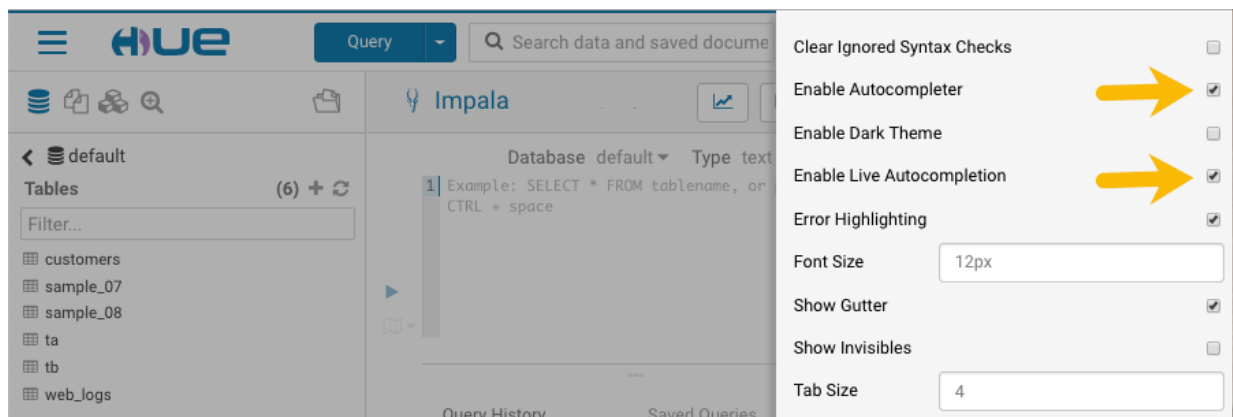
- For Windows, use the Ctrl key followed by a hyphen and then a comma:

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

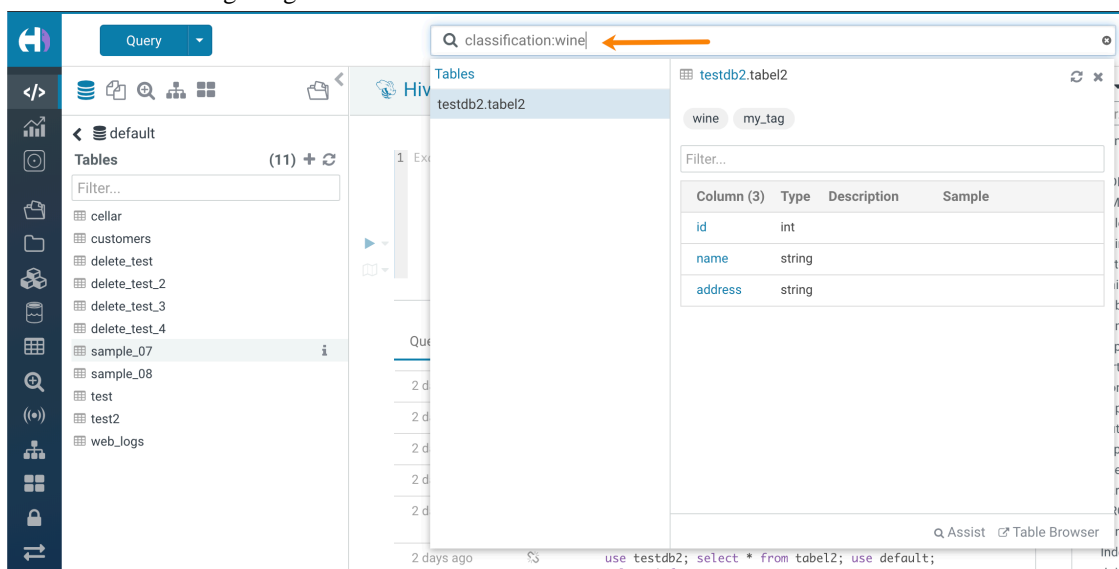
The SQL Editor in Hue provides a search text box where you can search on the metadata tags or classifications that are associated with your databases, tables, and columns.

#### About this task

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

#### 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.

### 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

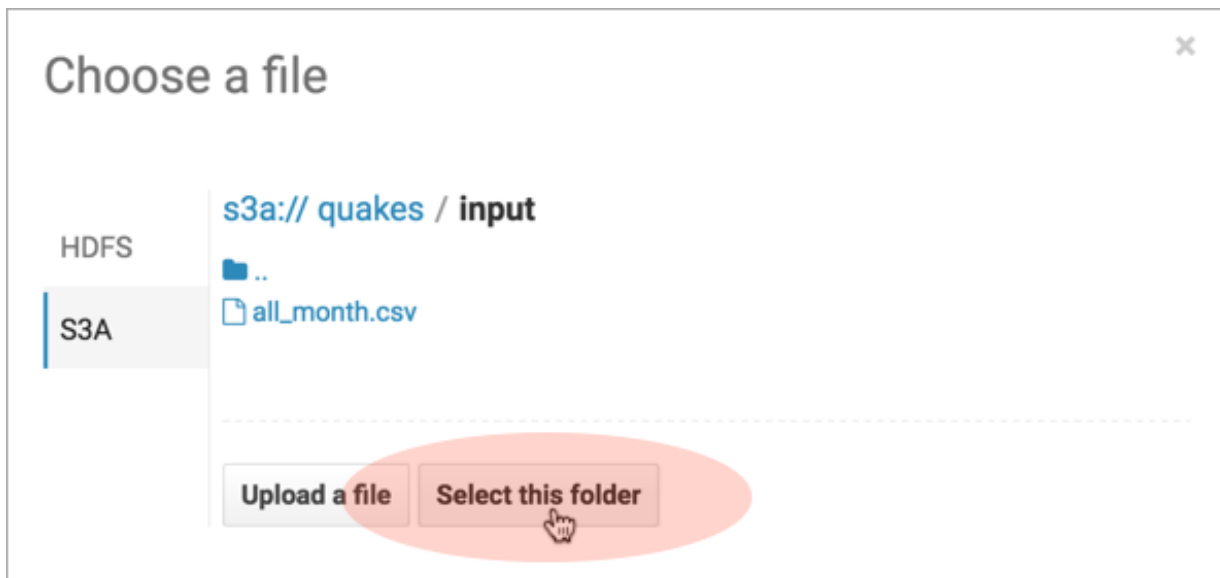
Using Hue to create a table from an Amazon S3 file streamlines the process.

#### About this task

#### 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.

### About this task

## Procedure

### 1. Run and Export Results in Hive

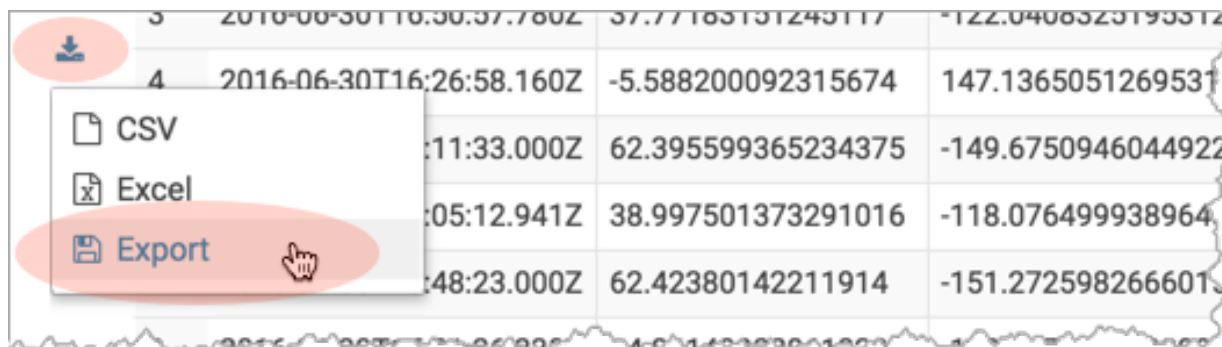
a.

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.

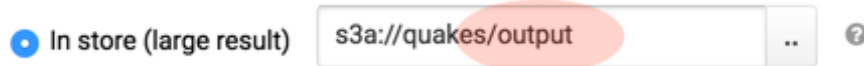
e. Append a file name to the path, such as quakes.csv.

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.

