

Uploading the Bike Share Data into Hive

Slides for the Hive Assignment

Download the data from

<http://www.bayareabikeshare.com/open-data>

Utilize the browser within Quickstart VM and download the files for both 2014 and 2015 into the Quickstart VM into a location of your choice within the VM (Desktop, etc.)

cloudera-quickstart-vm-5.4.2-0-vmware - VMware Workstation 12 Player (Non-commercial use only)

Player | Applications | Places | System

Sat Nov 14, 6:27 AM cloudera

Hue - Metastore Manager - Tables - Mozilla Firefox

about:sessionrestore x Cloudera Live : Welcom... x Hue - Impala Editor - Qu... x Hue - Metastore Man... x

quickstart.cloudera:8888/metastore/tables/

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

HUE Query Editors Data Browsers Workflows Search Security File Browser Job Browser cloudera

Metastore Manager

DATABASE

default

ACTIONS

- Create a new table from a file
- Create a new table manually

Databases > default

Search for table name

View Browse Data Drop

<input type="checkbox"/>	Table Name
<input type="checkbox"/>	categories
<input type="checkbox"/>	orders
<input type="checkbox"/>	products
<input type="checkbox"/>	sample_07
<input type="checkbox"/>	sample_08
<input type="checkbox"/>	tokenized_access_logs

Metastore Tables

Hue - Metastore Mana...

6:27 AM 11/14/2015

cloudera-quickstart-vm-5.4.2-0-vmware - VMware Workstation 12 Player (Non-commercial use only)

Player

Sat Nov 14, 6:30 AM cloudera

Hue - Metastore Manager - Create table from file - Mozilla Firefox

about:sessionrestore x Cloudera Live : Welcom... x Hue - Impala Editor - Qu... x Hue - Metastore Man... x

quickstart.cloudera:8888/beeswax/create/import_wizard/default

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

HUE Query Editors Data Browsers Workflows Search Security File Browser Job Browser cloudera

Metastore Manager

DATABASE

default

ACTIONS

- Create a new table from a file
- Create a table manually

Databases > default > Create a new table from a file

Step 1: Choose File Step 2: Choose Delimiter Step 3: Define Columns

Name Your Table and Choose A File

Table Name
Name of the new table. Table names must be globally unique. Table names tend to correspond to the directory where the data will be stored.

Description
Use a table comment to describe the table. For example, note the data's provenance and any caveats users need to know.

Input File ..
The HDFS path to the file on which to base this new table definition. It can be compressed (gzip) or not.

Import data from file ☒
Check this box to import the data in this file after creating the table definition. Leave it unchecked to define an empty table.

Warning: The selected file is going to be moved during the import.

Next

Hue - Metastore Mana...

6:30 AM 11/14/2015

cloudera-quickstart-vm-5.4.2-0-vmware - VMware Workstation 12 Player (Non-commercial use only)

Player ▾ | Applications Places System

Sat Nov 14, 6:32 AM cloudera

Hue - Metastore Manager - Create table from file - Mozilla Firefox

about:sessionrestore x Cloudera Live : Welcom... x Hue - Impala Editor - Qu... x Hue - Metastore Man... x

quickstart.cloudera:8888/ beeswax/create/import_wizard/default

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

File Browser Job Browser cloudera

Metastore Manager

DATABASE

default ▾

ACTIONS

Create a new table from a file

Create a new table manually

Databases > c

Step 1: Choose File

Name Your T

Table Name

Name of the new table. Table names must be globa

Description

Trip Data

Use a table comment to describe the table. For exa

Input File

/user/user_name/data_dir

The HDFS path to the file on which to base this nev

Import data from file ☒

Check this box to import the data in this file after c

Warning: The selected file is going to be moved

Next

Choose a file

Home / user / cloudera

Upload a file

File Upload

cloudera Desktop 201402_babs_open_data

Name	Size	Modified
201402_station_data.csv	5.1 KB	09/15/2015
201402_status_data.csv	593.6 MB	09/15/2015
201402_trip_data.csv	16.4 MB	09/15/2015
201402_weather_data.csv	78.9 KB	09/15/2015
README.txt	2.7 KB	09/15/2015

Add Remove All Files

Cancel Open

cloudera-quickstart-vm-5.4.2-0-vmware - VMware Workstation 12 Player (Non-commercial use only)

Player ▾ | [Icons] | Sat Nov 14, 6:33 AM | cloudera

Hue - Metastore Manager - Create table from file - Mozilla Firefox

about:sessionrestore x Cloudera Live : Welcom... x Hue - Impala Editor - Qu... x Hue - Metastore Man... x

quickstart.cloudera:8888/beeswax/create/import_wizard/default

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

HUE Query Editors Data Browsers Workflows Search Security File Browser Job Browser cloudera

Metastore Manager

DATABASE

default ▾

ACTIONS

Create a new table from a file

Create a new table manually

Databases > c

Step 1: Choose File

Name Your Table

Table Name

Description Trip Data

Input File /user/user_name/data_dir

Import data from file ☒

Warning: The selected file is going to be moved

Next

Choose a file

Home / user / cloudera

Upload a file

File Upload

cloudera Desktop 201402_babs_open_data

Name	Size	Modified
201402_station_data.csv	5.1 KB	09/15/2015
201402_status_data.csv	593.6 MB	09/15/2015
201402_trip_data.csv	16.4 MB	09/15/2015
201402_weather_data.csv	78.9 KB	09/15/2015
README.txt	2.7 KB	09/15/2015

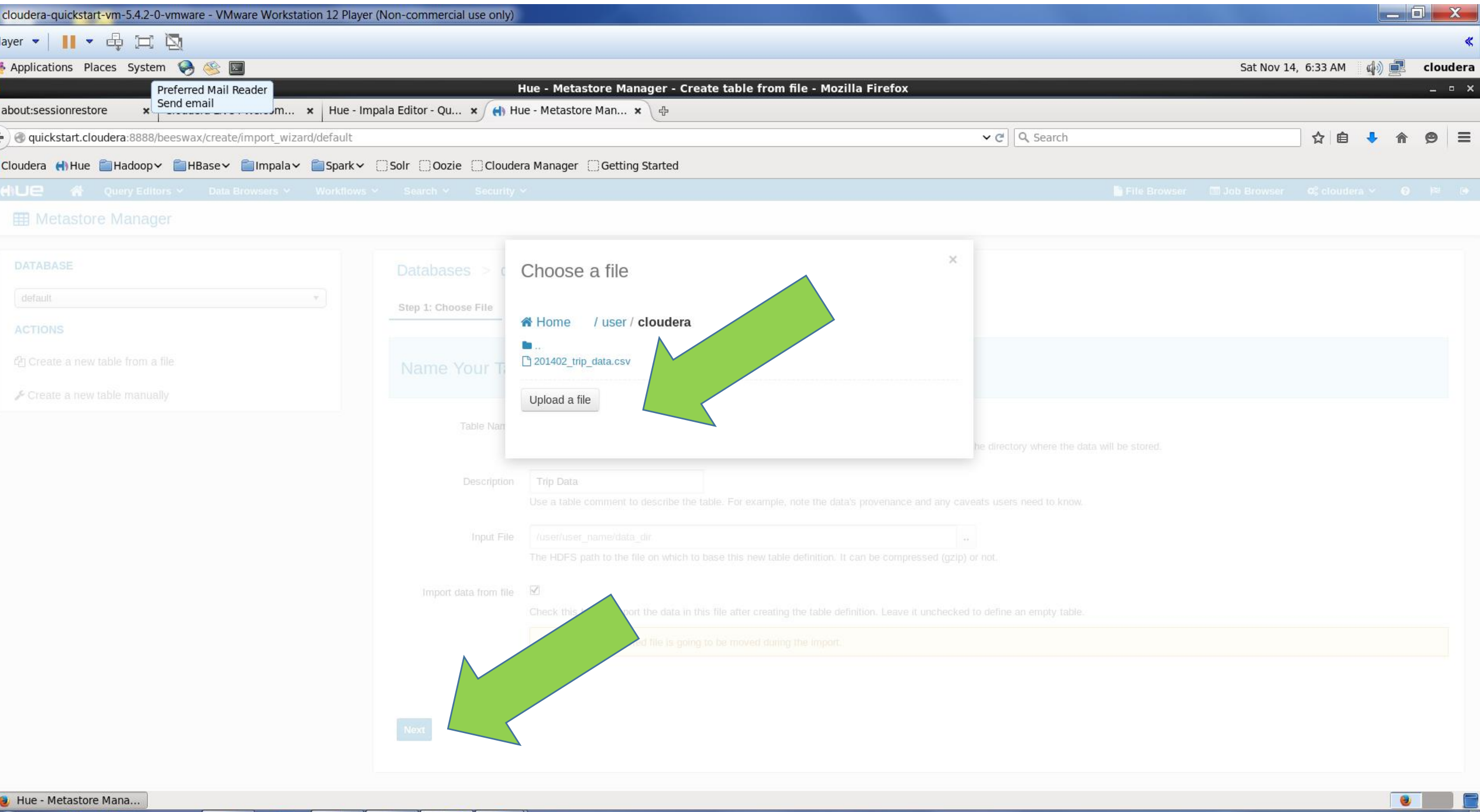
Add Remove

All Files ▾

Cancel Open

Hue - Metastore Mana...

6:33 AM 11/14/2015



cloudera-quickstart-vm-5.4.2-0-vmware - VMware Workstation 12 Player (Non-commercial use only)

Player

Sat Nov 14, 6:40 AM cloudera

Hue - Metastore Manager - Create table from file - Mozilla Firefox

about:sessionrestore x Cloudera Live : Welcom... x Hue - Impala Editor - Qu... x Hue - Metastore Man... x

quickstart.cloudera:8888/beeswax/create/import_wizard/default

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

HUE Query Editors Data Browsers Workflows Search Security File Browser Job Browser cloudera

Metastore Manager

ACTIONS

- Create a new table from a file
- Create a new table manually

Choose a Delimiter

Beeswax has determined that this file is delimited by **commas**.

Delimiter

Enter the column delimiter which must be a single character. Use syntax like "\001" or "\t" for special characters.

Table preview

col_1	col_2	col_3	col_4	col_5	col_6	col_7	col_8	col_9	col_10	col_11
Trip ID	Duration	Start Date	Start Station	Start Terminal	End Date	End Station	End Terminal	Bike #	Subscription Type	Zip Code
4576	63	8/29/2013 14:13	South Van Ness at Market	66	8/29/2013 14:14	South Van Ness at Market	66	520	Subscriber	94127
4607	70	8/29/2013 14:42	San Jose City Hall	10	8/29/2013 14:43	San Jose City Hall	10	661	Subscriber	95138
4130	71	8/29/2013 10:16	Mountain View City Hall	27	8/29/2013 10:17	Mountain View City Hall	27	48	Subscriber	97214
4251	77	8/29/2013 11:29	San Jose City Hall	10	8/29/2013 11:30	San Jose City Hall	10	26	Subscriber	95060
4299	83	8/29/2013 12:02	South Van Ness at Market	66	8/29/2013 12:04	Market at 10th	67	319	Subscriber	94103
4927	103	8/29/2013 18:54	Golden Gate at Polk	59	8/29/2013 18:56	Golden Gate at Polk	59	527	Subscriber	94109
4500	109	8/29/2013 13:25	Santa Clara at Almaden	4	8/29/2013 13:27	Adobe on Almaden	5	679	Subscriber	95112
4563	111	8/29/2013 14:02	San Salvador at 1st	8	8/29/2013 14:04	San Salvador at 1st	8	687	Subscriber	95112
4760	113	8/29/2013 17:01	South Van Ness at Market	66	8/29/2013 17:03	South Van Ness at Market	66	553	Subscriber	94103

We will leave the default delimiter as is – comma separated

cloudera-quickstart-vm-5.4.2-0-vmware - VMware Workstation 12 Player (Non-commercial use only)

Player ▾ | Applications Places System | Sat Nov 14, 6:42 AM | cloudera

Hue - Metastore Manager - Create table from file - Mozilla Firefox

about:sessionrestore x Cloudera Live : Welcom... x Hue - Impala Editor - Qu... x Hue - Metastore Man... x

quickstart.cloudera:8888/ beeswax/create/import_wizard/default

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

HUE | Query Editors | Data Browsers | Workflows | Search | Security | File Browser | Job Browser | cloudera

Metastore Manager

DATABASE

default ▾

ACTIONS

- Create a new table from a file
- Create a new table manually

Databases > default > Create a new table from a file

Step 1: Choose File Step 2: Choose Delimiter **Step 3: Define Columns**

Define your columns

Use first row as column names Bulk edit column names

Column name	Column Type	Sample Row #1	Sample Row #2
TripID	smallint	4607	4607
Duration	tinyint	63	70
StartDate	string	8/29/2013 14:13	8/29/2013 14:42
StartStation	string	South Van Ness at Market	San Jose City Hall
StartTerminal	tinyint	66	10
EndDate	string	8/29/2013 14:14	8/29/2013 14:43
EndStation	string	South Van Ness at Market	San Jose City Hall

Change Tinyint to an Int to preserve large values

Hue - Metastore Mana...

6:42 AM 11/14/2015

Refresh to see the newly uploaded Bike share database



Player

Applications Places System

Hue - Hive Editor - Query - Mozilla Firefox

about:sessionrestore Cloudera Live : Welcom... Hue - Impala Editor - Qu... Hue - Hive Editor - Query

quickstart.cloudera:8888/beeswax/execute/query/67#query/results

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

HUE Query Editors Data Browsers Workflows Search Security

Hive Editor Query Editor My Queries Saved Queries History

Assist Settings

DATABASE default

Table name...

bike_share_trips
tripid (smallint)
duration (int)
startdate (string)
startstation (string)
startterminal (tinyint)
enddate (string)
endstation (string)
endterminal (tinyint)
bike (smallint)
subscriptiontype (...)
zipcode (int)
categories
customers
departments
intermediate_access_logs
order_items
orders
products
sample_07
sample_08
tokenized_access_logs

1 SELECT * from
2 bike_share_trips
3

Execute Save Save as... Explain or create a New query

Recent queries Query Log Columns Results Chart

	bike_share_trips.tripid	bike_share_trips.duration	bike_share_trips.startdate	bike_share_trips.startstation	bike_share_trips.startterminal	bike_share_trips.enddate	bike_share_trips.endstation
0	4576	63	8/29/2013 14:13	South Van Ness at Market	66	8/29/2013 14:14	South Van Ness at Market
1	4607	70	8/29/2013 14:42	San Jose City Hall	10	8/29/2013 14:43	San Jose City Hall
2	4130	71	8/29/2013 10:16	Mountain View City Hall	27	8/29/2013 10:17	Mountain View City Hall
3	4251	77	8/29/2013 11:29	San Jose City Hall	10	8/29/2013 11:30	San Jose City Hall
4	4299	83	8/29/2013 12:02	South Van Ness at Market	66	8/29/2013 12:04	Market at 10th
5	4927	103	8/29/2013 18:54	Golden Gate at Polk	59	8/29/2013 18:56	Golden Gate at Polk
6	4500	109	8/29/2013 13:25	Santa Clara at Almaden	4	8/29/2013 13:27	Adobe on Almaden
7	4563	111	8/29/2013 14:02	San Salvador at 1st	8	8/29/2013 14:04	San Salvador at 1st
8	4760	113	8/29/2013 17:01	South Van Ness at Market	66	8/29/2013 17:03	South Van Ness at Market
9	4258	114	8/29/2013 11:33	San Jose City Hall	10	8/29/2013 11:35	MLK Library
10	4549	125	8/29/2013 13:52	Spear at Folsom	49	8/29/2013 13:55	Embarcadero at Bryant
11	4498	126	8/29/2013 13:23	San Pedro Square	6	8/29/2013 13:25	Santa Clara at Almaden

Hue - Hive Editor - Qu...

9:21 AM 11/14/2015

Start querying the data