

Assisted Practice 6.1: Working with Hive Query Editor Using Metadata

Problem Scenario: In this demonstration, you will work with the Hive Query editor and external table

Objective: In this demonstration, you will use Hive Query editor to create an EXTERNAL table where even after dropping tables, data will remain intact, and metadata will be cleaned.

Dataset Name: "drivers.csv"

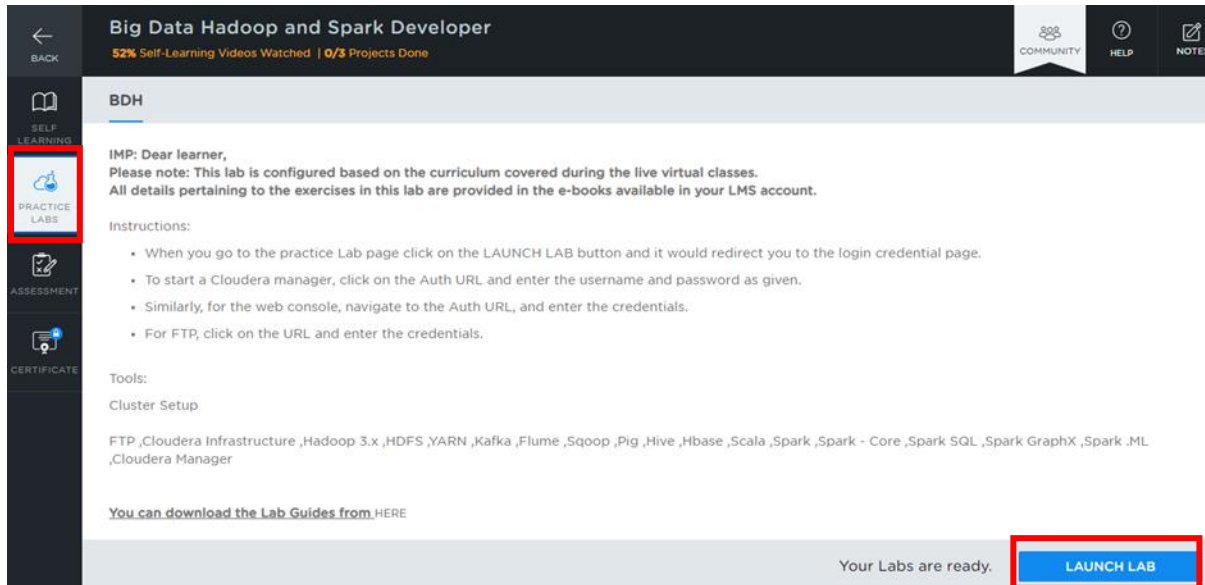
Steps to Perform:

Step 1: Download the dataset named "**drivers.csv**" from the Course Resources section

Step 2: Log in to your LMS account

Step 3: Open the course "**Big Data Hadoop and Spark Developer**"

Step 4: On the left side, click on the "**PRACTICE LABS**" tab and then click on the "**LAUNCH LAB**" button



Big Data Hadoop and Spark Developer
52% Self-Learning Videos Watched | 0/3 Projects Done

BDH

IMP: Dear learner,
Please note: This lab is configured based on the curriculum covered during the live virtual classes.
All details pertaining to the exercises in this lab are provided in the e-books available in your LMS account.

Instructions:

- When you go to the practice Lab page click on the LAUNCH LAB button and it would redirect you to the login credential page.
- To start a Cloudera manager, click on the Auth URL and enter the username and password as given.
- Similarly, for the web console, navigate to the Auth URL, and enter the credentials.
- For FTP, click on the URL and enter the credentials.

Tools:

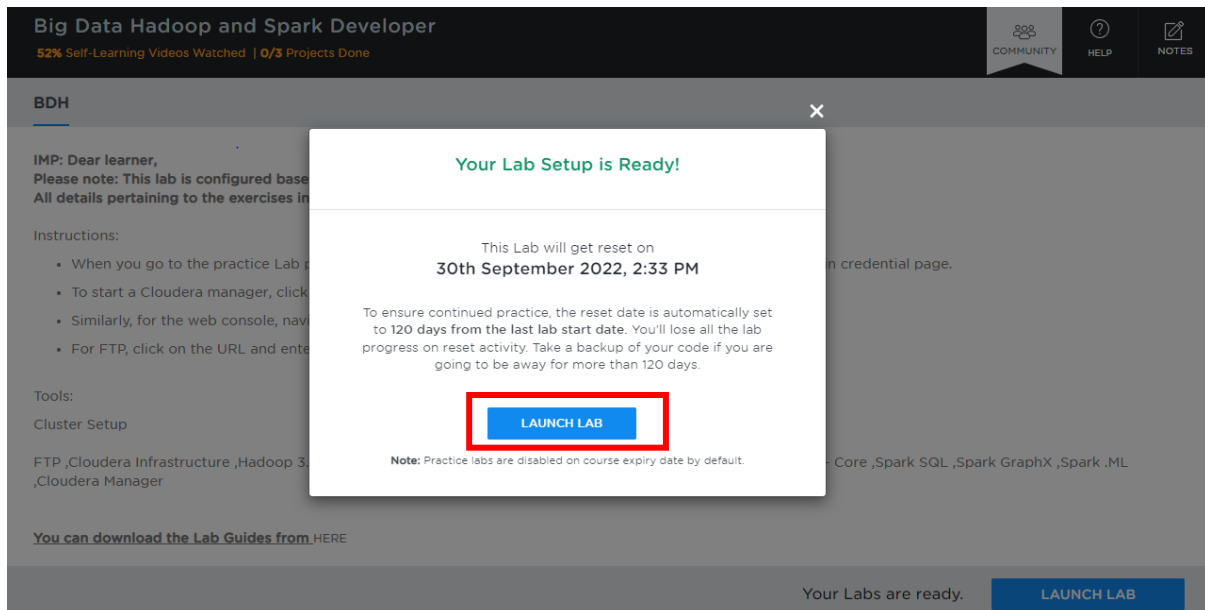
Cluster Setup

FTP ,Cloudera Infrastructure ,Hadoop 3.x ,HDFS ,YARN ,Kafka ,Flume ,Sqoop ,Pig ,Hive ,Hbase ,Scala ,Spark ,Spark - Core ,Spark SQL ,Spark GraphX ,Spark .ML ,Cloudera Manager

You can download the Lab Guides from [HERE](#)

Your Labs are ready. **LAUNCH LAB**

Step 5: Again, click on the **“LAUNCH LAB”** button



Big Data Hadoop and Spark Developer
52% Self-Learning Videos Watched | 0/3 Projects Done

BDH

IMP: Dear learner,
Please note: This lab is configured based on the curriculum covered during the live virtual classes.
All details pertaining to the exercises in this lab are provided in the e-books available in your LMS account.

Instructions:

- When you go to the practice Lab page click on the LAUNCH LAB button and it would redirect you to the login credential page.
- To start a Cloudera manager, click on the Auth URL and enter the username and password as given.
- Similarly, for the web console, navigate to the Auth URL, and enter the credentials.
- For FTP, click on the URL and enter the credentials.

Tools:

Cluster Setup

FTP ,Cloudera Infrastructure ,Hadoop 3.x ,HDFS ,YARN ,Kafka ,Flume ,Sqoop ,Pig ,Hive ,Hbase ,Scala ,Spark ,Spark - Core ,Spark SQL ,Spark GraphX ,Spark .ML ,Cloudera Manager

You can download the Lab Guides from [HERE](#)

Your Labs are ready. **LAUNCH LAB**

Your Lab Setup is Ready!

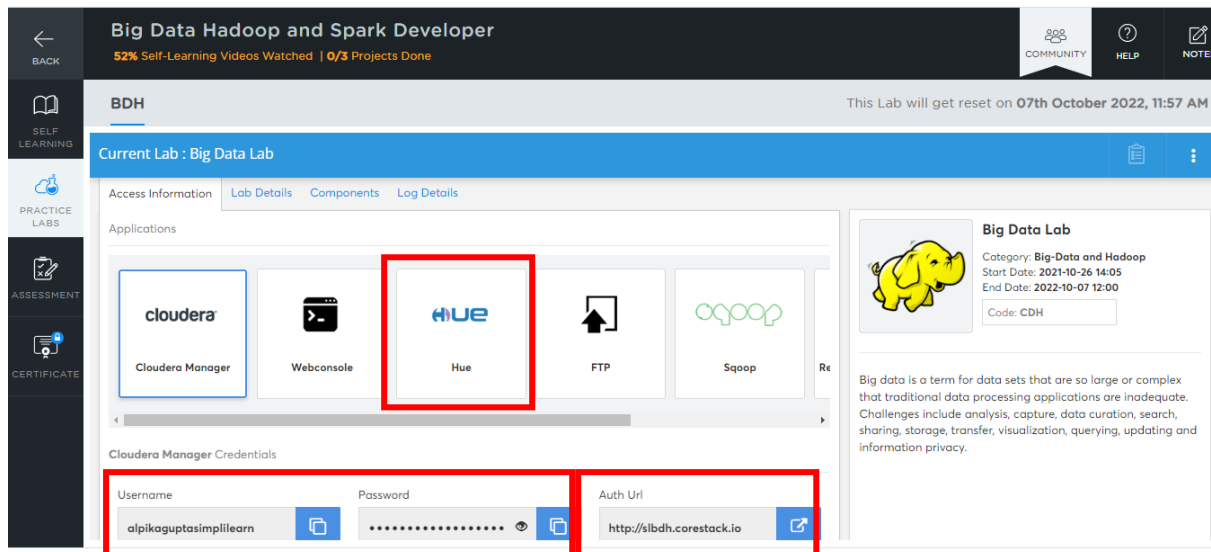
This Lab will get reset on
30th September 2022, 2:33 PM

To ensure continued practice, the reset date is automatically set to 120 days from the last lab start date. You'll lose all the lab progress on reset activity. Take a backup of your code if you are going to be away for more than 120 days.

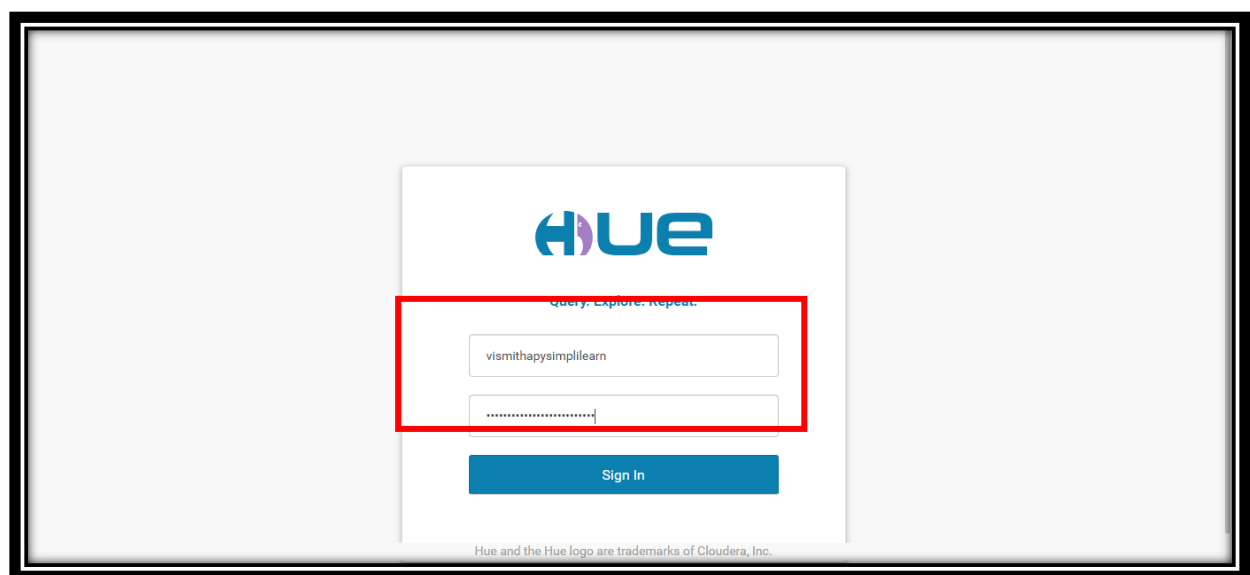
LAUNCH LAB

Note: Practice labs are disabled on course expiry date by default.

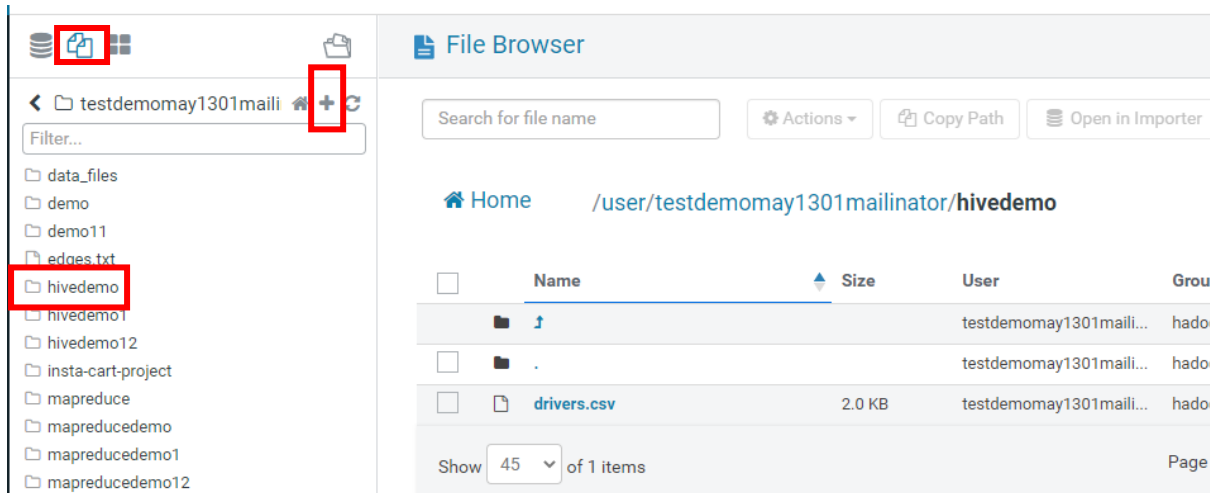
Step 6: Click on **“Hue”** and click on the **“Auth Url”** to upload the dataset and copy the **“Username”** and the **“Password”** provided to log in to the **“Hue”**



Step 7: Paste the **“Username”** and the **“Password”** on the login window and click on **“Sign In”**



Step 8: Create a directory named **“hivedemo”** and click on the **“HDFS”** icon and click on the **“+”** symbol to upload the dataset

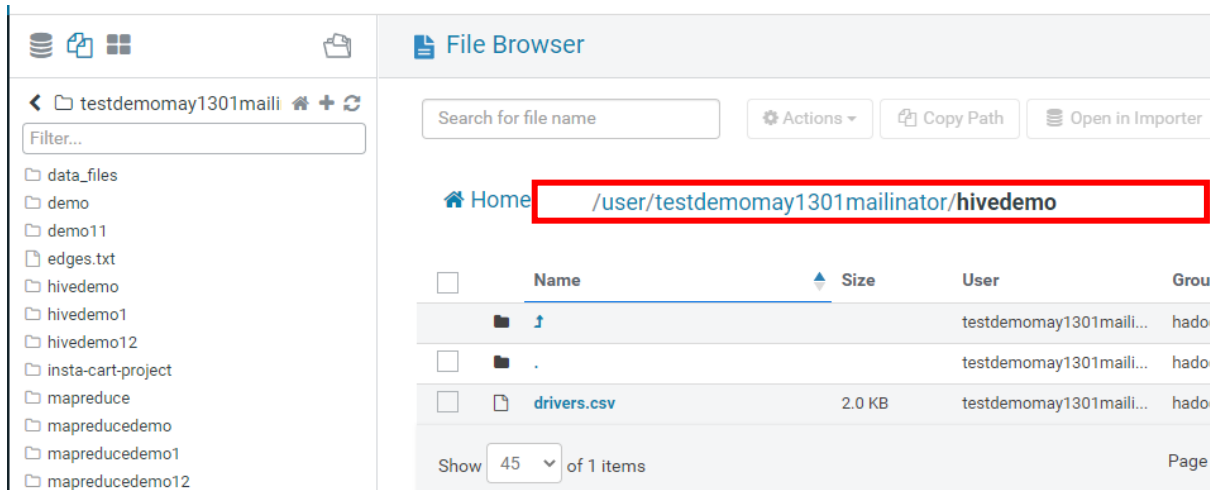


The screenshot shows the 'File Browser' interface. On the left sidebar, the file 'hivedemo' is selected. The main view shows the path '/user/testdemomay1301mailinator/hivedemo' highlighted. Below the path, there is a table listing files and folders.

Name	Size	User	Group
↑		testdemomay1301maili...	hado
.		testdemomay1301maili...	hado
drivers.csv	2.0 KB	testdemomay1301maili...	hado

At the bottom, it says 'Show 45 of 1 items' and 'Page'.

Step 9: Copy the path of the dataset that has been uploaded

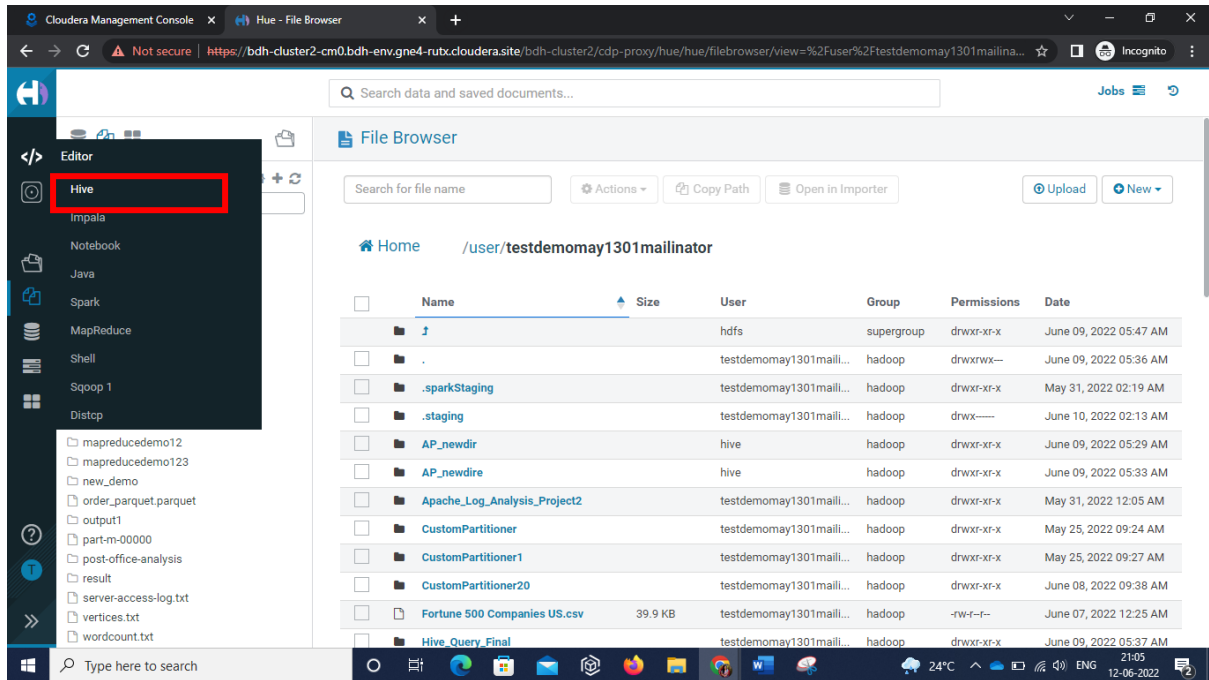


The screenshot shows the 'File Browser' interface. The path '/user/testdemomay1301mailinator/hivedemo' is highlighted in a red box. Below the path, there is a table listing files and folders.

Name	Size	User	Group
↑		testdemomay1301maili...	hado
.		testdemomay1301maili...	hado
drivers.csv	2.0 KB	testdemomay1301maili...	hado

At the bottom, it says 'Show 45 of 1 items' and 'Page'.

Step 10: Click on the "Hive" icon and open the Hive editor



Step 11: Create a database and a table in it to store the data as shown below:

create database Hive_Test;

CREATE EXTERNAL TABLE Hive_Test.drivers

(

driverId INT,

name STRING,

ssn BIGINT,

location STRING,

certified STRING,

wageplan STRING

)

row format delimited fields terminated by ','

LOCATION '/user/testdemomay1301mailinator/new_demo/'

TBLPROPERTIES("skip.header.line.count"="1");

Note: The parameter LOCATION determines where exactly the new hive table will be created.

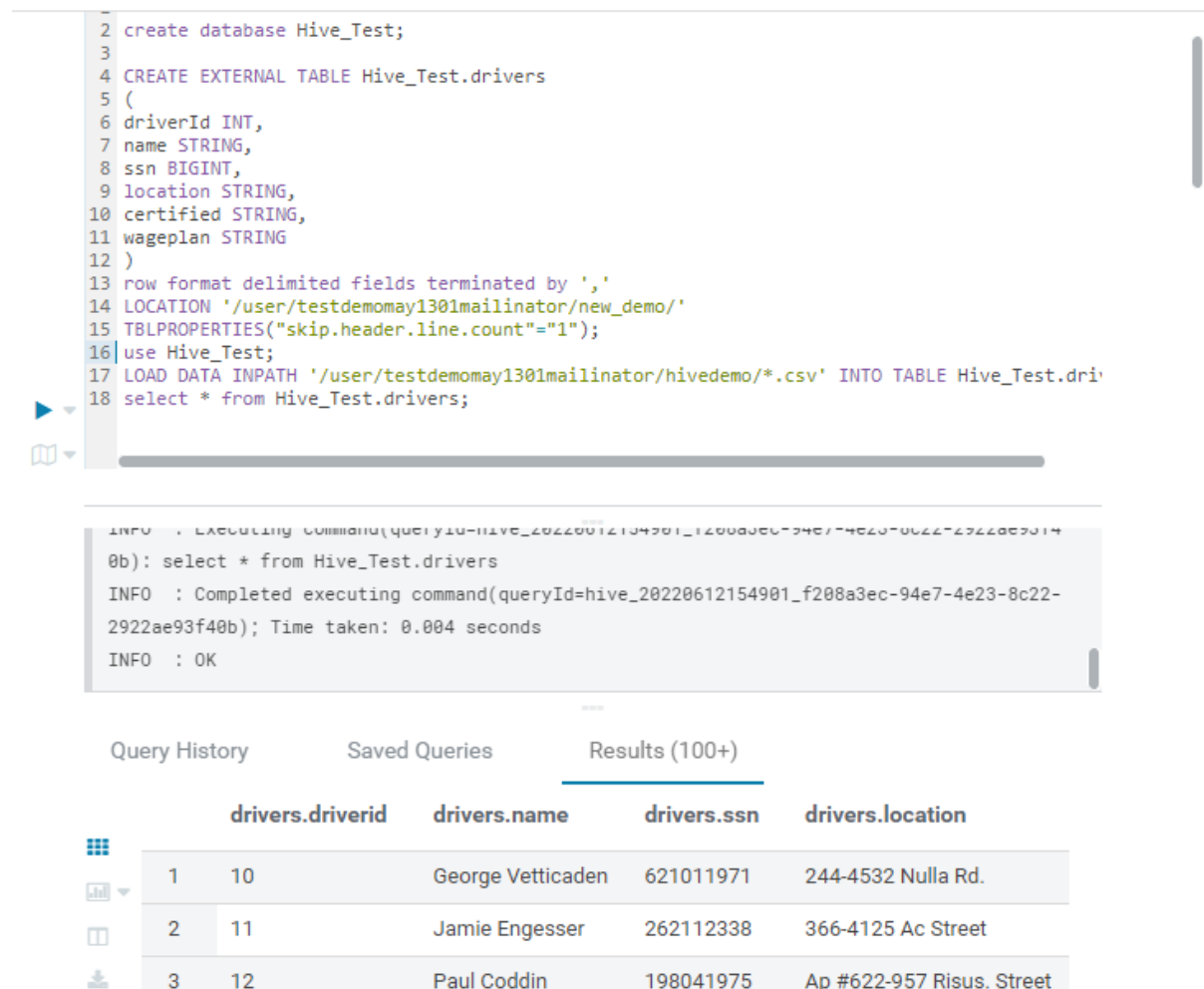
Step 12: Next, upload data by writing the below command:

use Hive_Test;

LOAD DATA INPATH '/user/testdemomay1301mailinator/hivedemo/*.csv' INTO
TABLE Hive_Test.drivers;

Step 13: Verify the data by executing the below command:

select * from Hive_Test.drivers;



```

2 create database Hive_Test;
3
4 CREATE EXTERNAL TABLE Hive_Test.drivers
5 (
6 driverId INT,
7 name STRING,
8 ssn BIGINT,
9 location STRING,
10 certified STRING,
11 wageplan STRING
12 )
13 row format delimited fields terminated by ','
14 LOCATION '/user/testdemomay1301mailinator/new_demo/'
15 TBLPROPERTIES("skip.header.line.count"="1");
16 use Hive_Test;
17 LOAD DATA INPATH '/user/testdemomay1301mailinator/hivedemo/*.csv' INTO TABLE Hive_Test.drivers;
18 select * from Hive_Test.drivers;

```

```

INFO : Executing command(queryId=hive_20220612154901_f208a3ec-94e7-4e23-8c22-2922ae93f40b): select * from Hive_Test.drivers
INFO : Completed executing command(queryId=hive_20220612154901_f208a3ec-94e7-4e23-8c22-2922ae93f40b); Time taken: 0.004 seconds
INFO : OK

```

	drivers.driverid	drivers.name	drivers.ssn	drivers.location
1	10	George Vetticaden	621011971	244-4532 Nulla Rd.
2	11	Jamie Engesser	262112338	366-4125 Ac Street
3	12	Paul Coddin	198041975	Ap #622-957 Risus. Street

Step 14: Now, you will be able to see **"drivers.csv"** which will be created under the new_demo folder

⚙️ Actions ▼
📄 Copy Path
📂 Open in Importer

[🏠 Home](#)
/user/testdemomay1301mailinator/new_demo

<input type="checkbox"/>	Name	Size	User	Group
<input type="checkbox"/>	📁 ↑		testdemomay1301maili...	hadoop
<input type="checkbox"/>	📁 .		testdemomay1301maili...	hadoop
<input type="checkbox"/>	📄 drivers.csv	2.0 KB	testdemomay1301maili...	hadoop

Show 45 ▼ of 1 items
Page

Step 15: So, you can see the data, as well as metadata, is present

[🏠 Home](#)
Page to of 1
⏪ ⏩ ⏴ ⏵

/ user/ testdemomay1301mailinator/ new_demo/ drivers .csv

```

driverId,name,ssn,location,certified,wage-plan
10,George Vetticaden,621011971,244-4532 Nulla Rd.,N,miles
11,Jamie Engesser,262112338,366-4125 Ac Street,N,miles
12,Paul Coddin,198041975,Ap #622-957 Risus. Street,Y,hours
13,Joe Niemiec,139907145,2071 Hendrerit. Ave,Y,hours
14,Adis Cesir,820812209,Ap #810-1228 In St.,Y,hours
15,Rohit Bakshi,239005227,648-5681 Dui- Rd.,Y,hours
16,Tom McCuch,363303105,P.O. Box 313- 962 Parturient Rd.,Y,hours
17,Eric Mizell,123808238,P.O. Box 579- 2191 Gravida. Street,Y,hours
18,Grant Liu,171010151,Ap #928-3159 Vestibulum Av.,Y,hours
19,Ajay Singh,160005158,592-9430 Nonummy Avenue,Y,hours
20,Chris Harris,921812303,883-2691 Proin Avenue,Y,hours
21,Jeff Markham,209408086,Ap #852-7966 Facilisis St.,Y,hours
22,Nadeem Asghar,783204269,154-9147 Aliquam Ave,Y,hours
23,Adam Diaz,928312208,P.O. Box 260- 6127 Vitae Road,Y,hours
24,Don Hilborn,254412152,4361 Ac Road,Y,hours

```

Step 16: Now, drop the table as shown below:

Command:

drop table Hive_Test.drivers

Step 17: After this verify the table using the below command:

Command:

select * from Hive_Test.drivers;

0s hiveap111 ▾ ⚙ ?

```

4 CREATE EXTERNAL TABLE Hive_Test.drivers
5 (
6   driverId INT,
7   name STRING,
8   ssn BIGINT,
9   location STRING,
10  certified STRING,
11  wageplan STRING
12 )
13 row format delimited fields terminated by ','
14 LOCATION '/user/testdemomay1301mailinator/new_demo/'
15 TBLPROPERTIES("skip.header.line.count"="1");
16 use Hive_Test;
17 LOAD DATA INPATH '/user/testdemomay1301mailinator/hivedemo/*.csv' INTO TABLE Hive_Test.drivers;
18
19
20 drop table Hive_Test.drivers;
21 select * from Hive_Test.drivers;

```

Error while compiling statement: FAILED: SemanticException [Error 10001]: line 20:43 Table not found 'drivers'

Step 18: As you can see the error: no table found, go to /user/testdemomay1301mailinator/new_demo/drivers.csv and the data can be still seen there

[Home](#)

Page 1 to 1 of 1



/ user/ testdemomay1301mailinator/ new_demo/ drivers.csv

```

driverId,name,ssn,location,certified,wage-plan
10,George Vetticaden,621011971,244-4532 Nulla Rd.,N,miles
11,Jamie Engesser,262112338,366-4125 Ac Street,N,miles
12,Paul Coddin,198041975,Ap #622-957 Risus. Street,Y,hours
13,Joe Niemiec,139907145,2071 Hendrerit. Ave,Y,hours
14,Adis Cesir,820812209,Ap #810-1228 In St.,Y,hours
15,Rohit Bakshi,239005227,648-5681 Dui- Rd.,Y,hours
16,Tom McCuch,363303105,P.O. Box 313- 962 Parturient Rd.,Y,hours
17,Eric Mizell,123808238,P.O. Box 579- 2191 Gravida. Street,Y,hours
18,Grant Liu,171010151,Ap #928-3159 Vestibulum Av.,Y,hours
19,Ajay Singh,160005158,592-9430 Nonummy Avenue,Y,hours
20,Chris Harris,921812303,883-2691 Proin Avenue,Y,hours
21,Jeff Markham,209408086,Ap #852-7966 Facilisis St.,Y,hours
22,Nadeem Asghar,783204269,154-9147 Aliquam Ave,Y,hours
23,Adam Diaz,928312208,P.O. Box 260- 6127 Vitae Road,Y,hours
24,Don Hilborn,254412152,4361 Ac Road,Y,hours
25,Jean-Philippe Playe,913310051,P.O. Box 812- 6238 Ac Rd.,Y,hours

```