

Assisted Practice 8: Data Upload from HDFS to HBase

Problem Scenario: Create a table using namespace in HBase shell to upload bulk data from HDFS

Objective: In this demonstration, you will create a table in HBase to store bulk data that is stored on HDFS.

Dataset Name: data.csv

Tasks to Perform:

1. Load the data (data.csv) into HDFS
2. Create a table to store data using namespace in the HBase shell
3. Execute the command to fetch data from HDFS into the table
4. Verify whether the data is in the table or not by using the scan command

Steps to Perform:

Step 1: Log in to your LMS account

Step 2: Open the course “**Big Data Hadoop and Spark developer**”

Step 3: On the left side, click on the “**PRACTICE LABS**” tab and click on the “**LAUNCH LAB**” button

SELF

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

Big Data Hadoop and Spark Developer

52% Self-Learning Videos Watched | 0/3 Projects Done

COMMUNITY HELP NOTES

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!

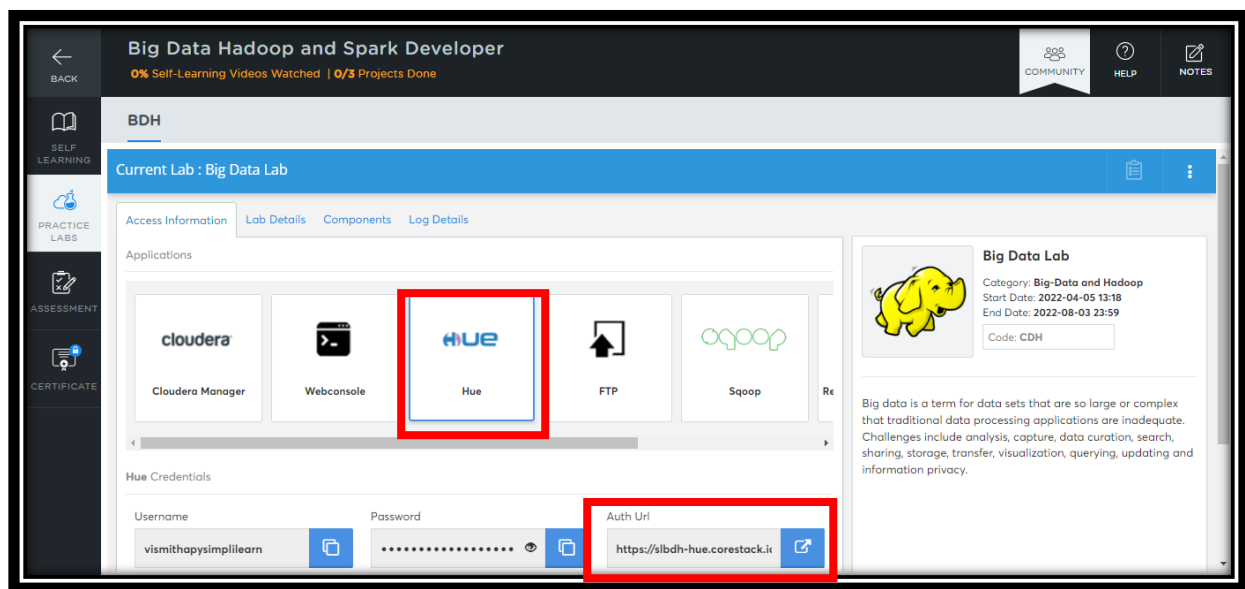
This Lab will get reset on
07th October 2022, 11:57 AM

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.

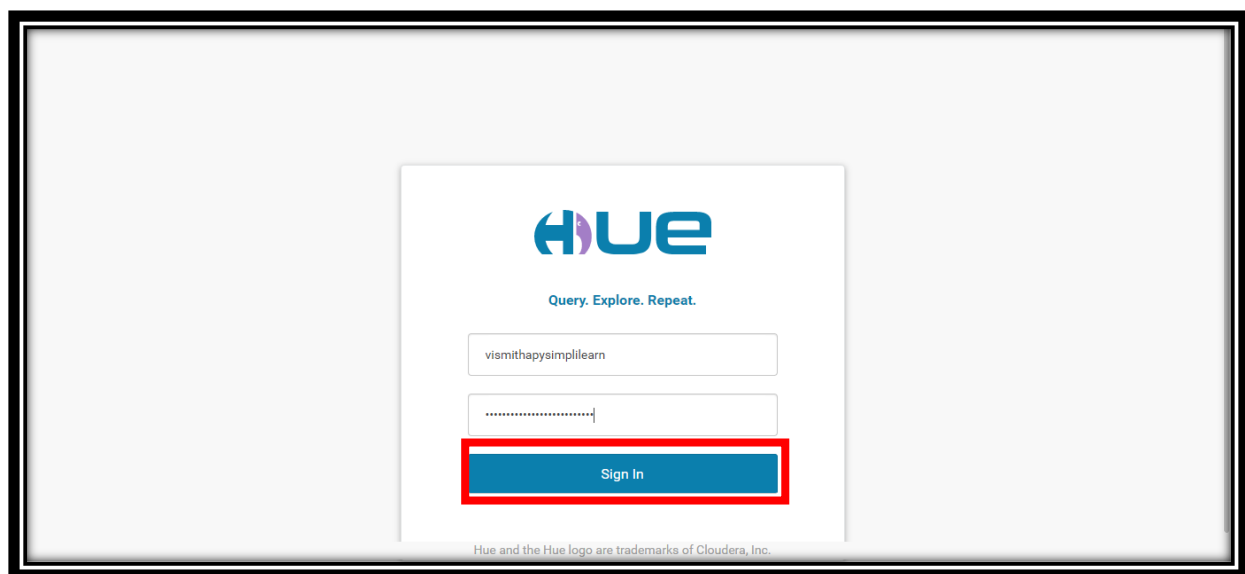
LAUNCH LAB

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

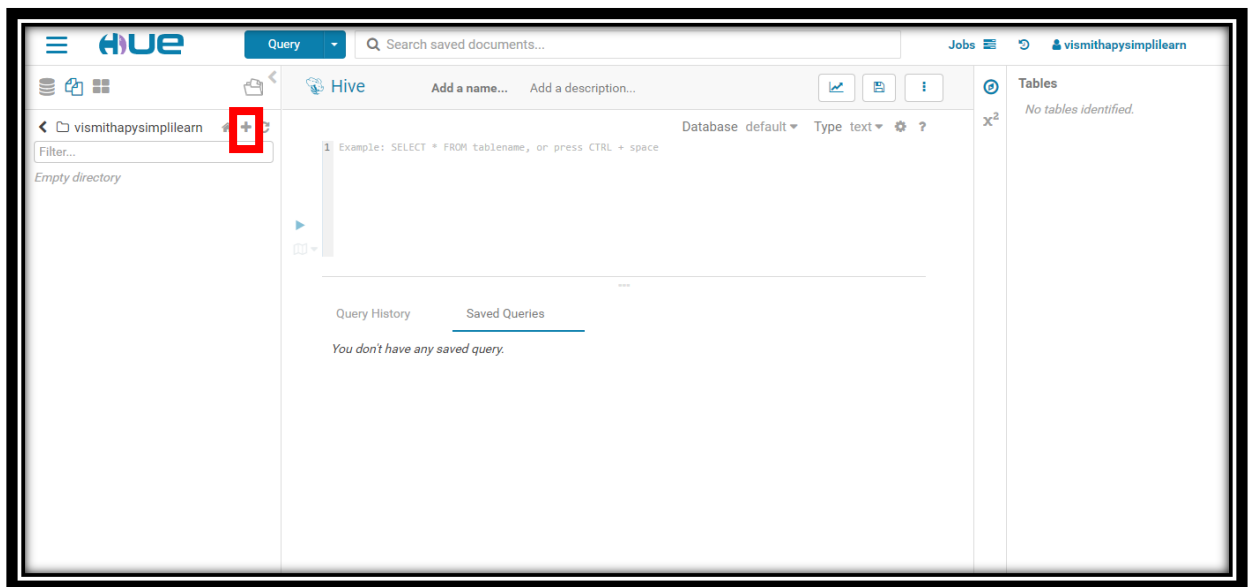
Step 5: Click on “**Hue**” and click on the “**Auth Url**” to upload the dataset and copy the “**Username**” and the “**Password**” provided to log in to the “**Hue**”



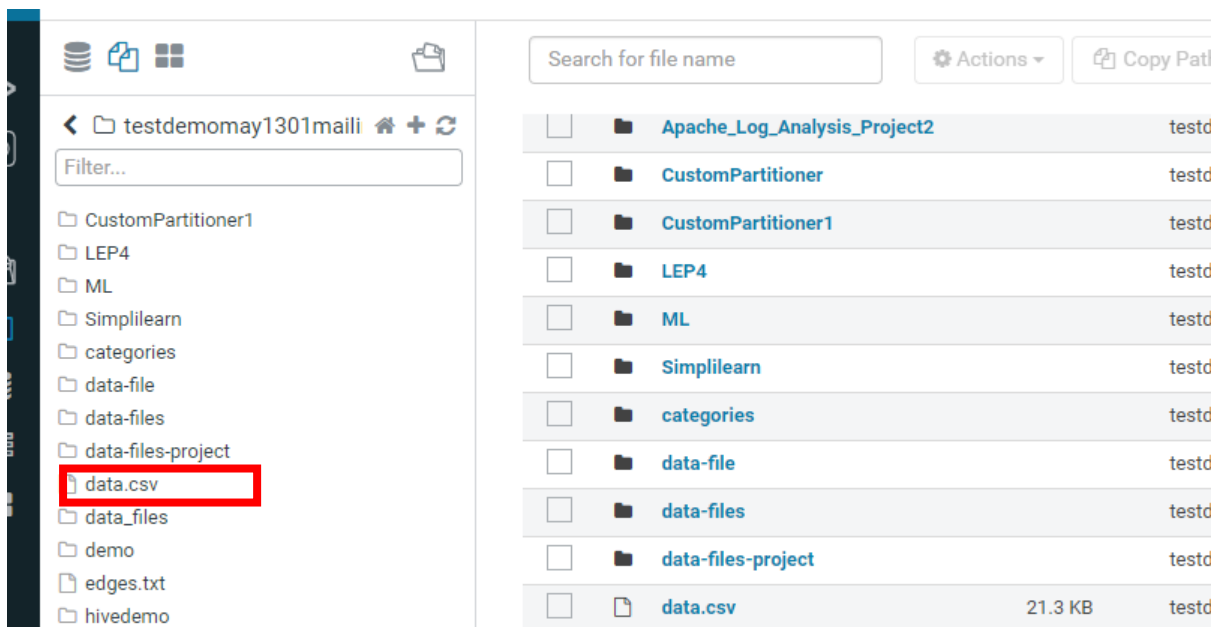
Step 6: Paste the **“Username”** and the **“Password”** on the login window and click on sign in



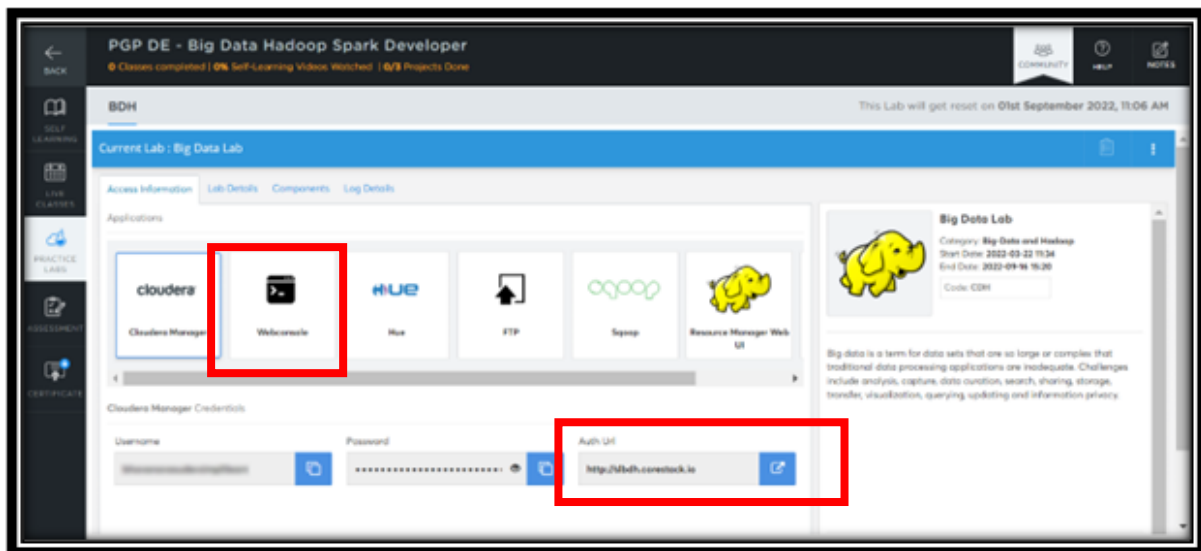
Step 7: Click on the **“HDFS”** icon and click on the **“+”** symbol to upload the dataset



Step 8: Select the downloaded dataset file and upload it to **"HDFS."** In addition, by right-clicking, copy the path from the dataset that has been uploaded



Step 9: Click on **"Webconsole"** and click on the **"Auth Url"**



Step 10: Copy the “**Username**” and the “**Password**” provided to log in to the Webconsole

Step 11: Paste the “**Username**” and the “**Password**” on the console and click on enter

Note: The password will not be visible when pasted on the console



Step 12: Login into the HBase shell using the below command:

Command: hbase shell

```
[testdemomay1301mailinator@bdh-cluster2-edgenode10 ~]$ hbase shell
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
For Reference, please visit: http://hbase.apache.org/2.0/book.html#shell
Version 2.2.6.7.2.12.4-1, r9f8987ebec2d1969a758180f996bb8bf7b46d576, Wed Jan 26 03:29:50 UTC
Took 0.0012 seconds
```

Step 13: Create a namespace **"demo11"**, and then by using the namespace create a table **"employee_demo"** using the below commands:

Command:

- a) `create_namespace 'demo11'`
- b) `create 'demo11:employee_demo', 'personal_data', 'professional_data'`

```
hbase:002:0> create_namespace 'demo11'
Took 0.4295 seconds
hbase:003:0> create 'demo11:employee_demo', 'personal_data', 'professional_data'
Created table demo11:employee_demo
Took 8.3310 seconds
=> Hbase::Table - demo11:employee_demo
hbase:004:0>
```

Step 14: To come out from the HBase shell click on Ctrl+Z, and execute the below command to fetch data from an HDFS into the table:

Command:

```
hbase org.apache.hadoop.hbase.mapreduce.ImportTsv -Dimporttsv.separator=,
-
Dimporttsv.columns="HBASE_ROW_KEY,personal_data:emp_id,personal_data:emp_Name,personal_data:city,personal_data:emailid,personal_data:age,professional_data:designation,professional_data:salary,professional_data:contact_num"
demo11:employee_demo /user/testdemomay1301mailinator/data.csv
```

```
[testdemomay1301mailinator@bdh-cluster2-edgenode10 ~]$ hbase org.apache.hadoop.hbase.mapreduce.ImportTsv -Dimporttsv.separator=,
-Dimporttsv.columns="HBASE_ROW_KEY,personal_data:emp_id,personal_data:emp_Name,personal_data:city,personal_data:emailid,personal_data:age,professional_data:designation,professional_data:salary,professional_data:contact_num" demo11:employee_demo /user/testdemomay1301mailinator/data.csv
```

Step 15: Next, log in to the HBase shell and verify whether the data is in the table or not by using the below command:

Command:

scan 'demo11:employee_demo'

```
hbase:001:0> scan 'demo11:employee_demo'
ROW COLUMN+CELL
10 column=personal_data:age, timestamp=1654157828955, value=44
10 column=personal_data:city, timestamp=1654157828955, value=Pemberton
10 column=personal_data:emailid, timestamp=1654157828955, value=augue@sem.com
10 column=personal_data:emp_Name, timestamp=1654157828955, value=Lynn Tyler
10 column=personal_data:emp_id, timestamp=1654157828955, value=62770
10 column=professional_data:contact_num, timestamp=1654157828955, value=1-607-671-59
10 column=professional_data:designation, timestamp=1654157828955, value=at fringilla
ulputate velit eu sem. Pellentesque ut ipsum ac mi eleifend
10 column=professional_data:salary, timestamp=1654157828955, value=5022
2 column=personal_data:age, timestamp=1654157828955, value=61
2 column=personal_data:city, timestamp=1654157828955, value=Sherani
2 column=personal_data:emailid, timestamp=1654157828955, value=eros.Proin@convallis
2 column=personal_data:emp_Name, timestamp=1654157828955, value=Adrienne Rojas
2 column=personal_data:emp_id, timestamp=1654157828955, value=2051
2 column=professional_data:contact_num, timestamp=1654157828955, value=1-684-429-94
2 column=professional_data:designation, timestamp=1654157828955, value=Cras eget ni
eger id magna et ipsum cursus vestibulum. Mauris magna. Duis dignissim tempor arc
2 column=professional_data:salary, timestamp=1654157828955, value=4831
22 column=personal_data:age, timestamp=1654157828955, value=30
22 column=personal_data:city, timestamp=1654157828955, value=Castiglione del Lago
```