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:

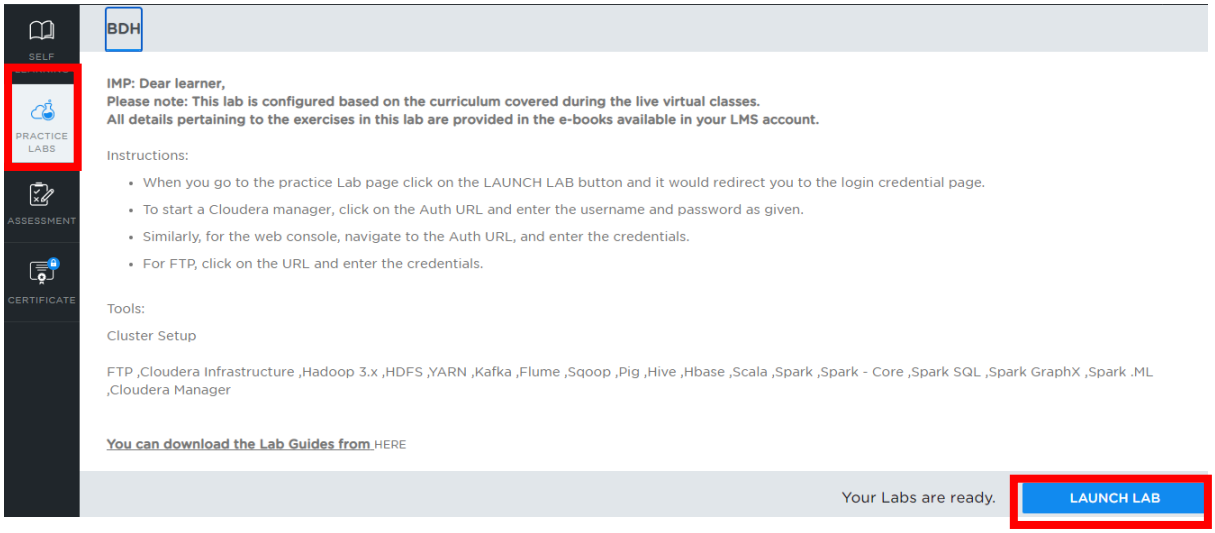
1. Load the data (data.csv) into HDFS
2. Create a table to store data using namespace in the HBase shell
3. Execute the command to fetch data from HDFS into the table
4. Verify whether the data is in the table or not by using the scan command

Steps to Perform:

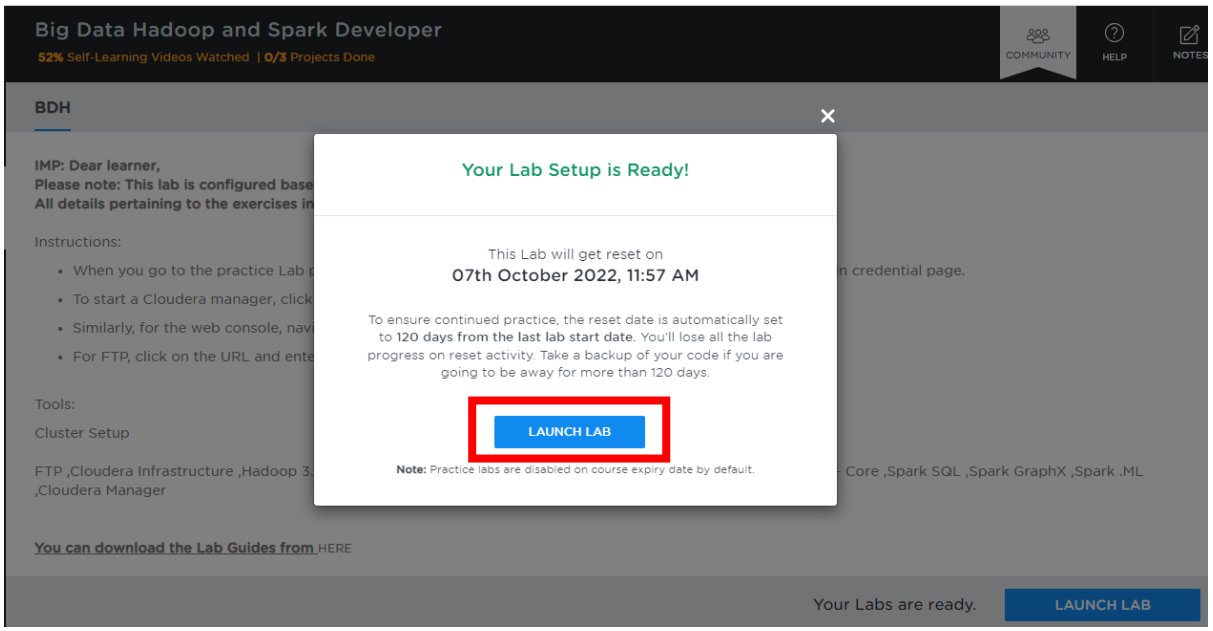
Step 1: Log in to your LMS account

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

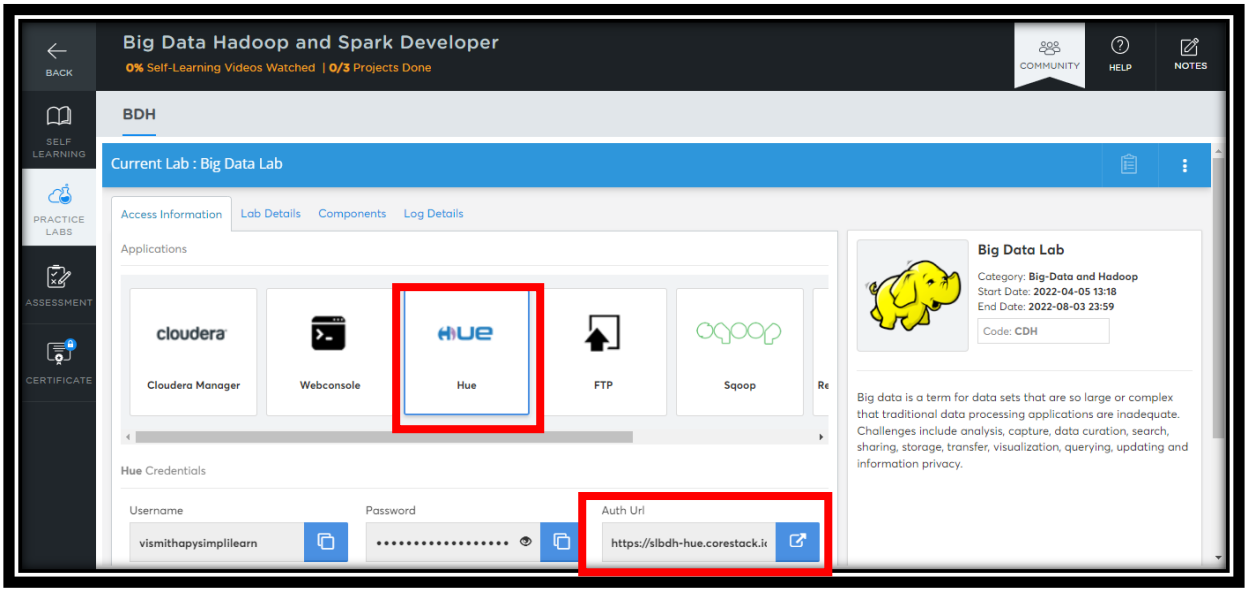
Step 3: On the left side, click on the "**PRACTICE LABS**" tab and click on the "**LAUNCH LAB**" button



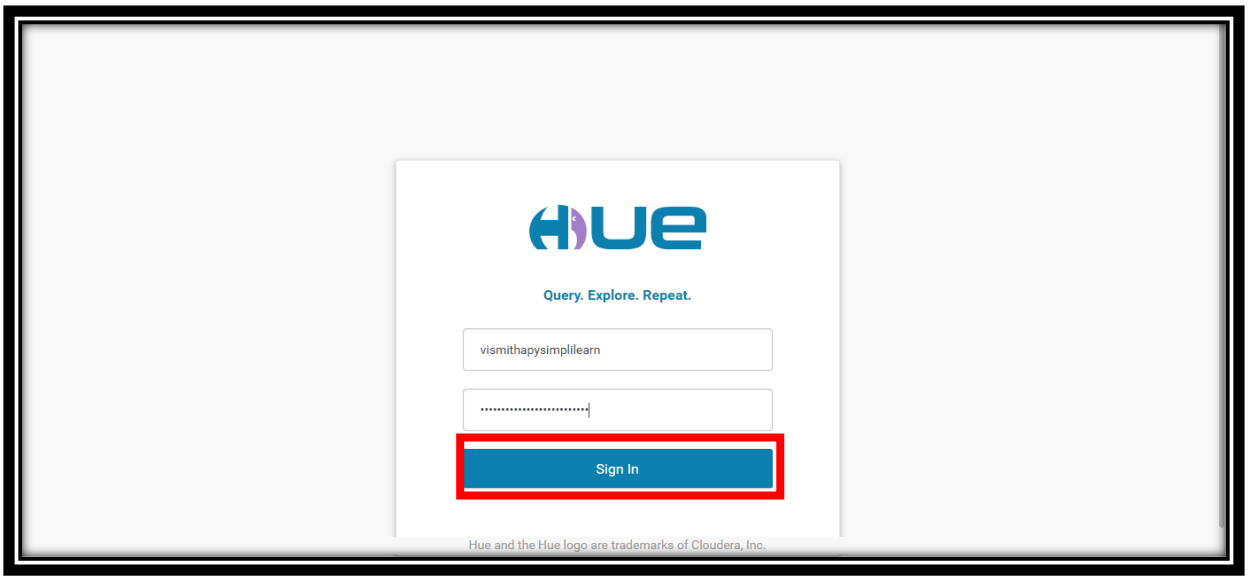
Step 4: Again, click on the "**LAUNCH LAB**" button



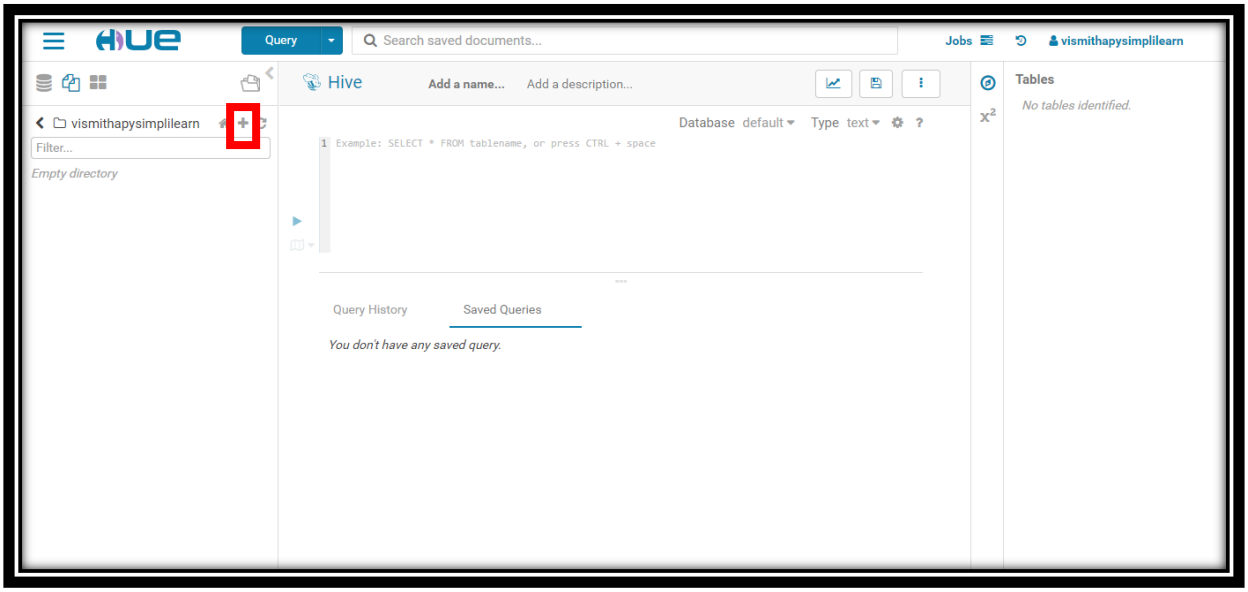
Step 5: Click on **"Hue"** and click on the **"Auth Url"** to upload the dataset and copy the "**Username**" and the "**Password**" provided to log in to the **"Hue"**



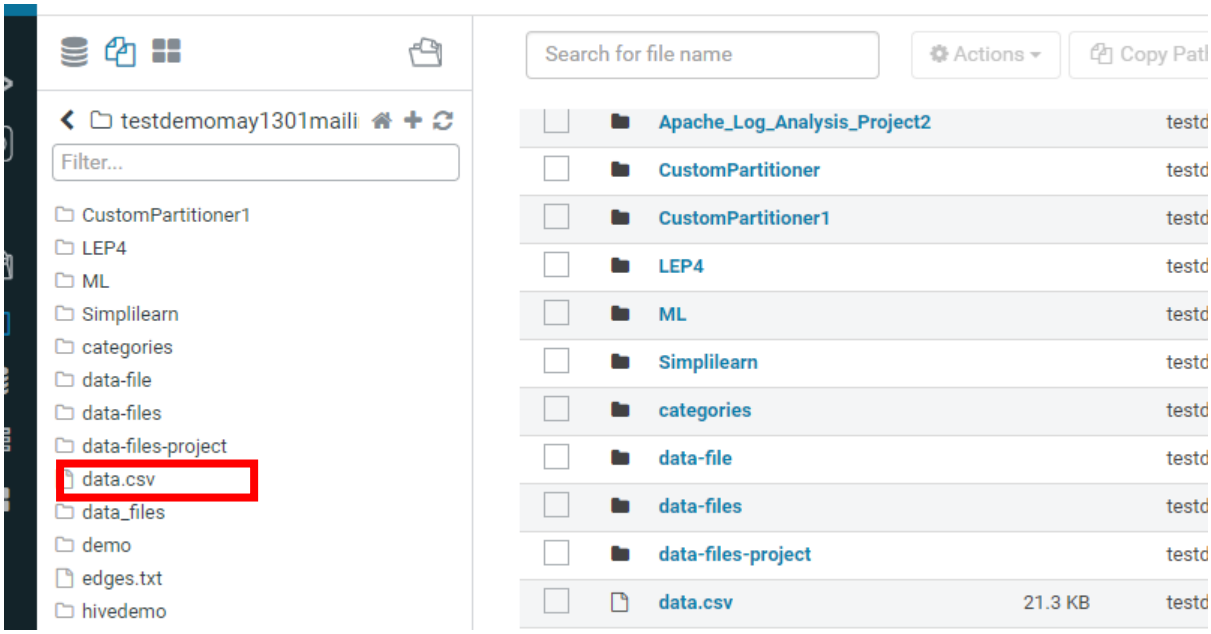
Step 6: Paste the **"Username"** and the **"Password"** on the login window and click on sign in



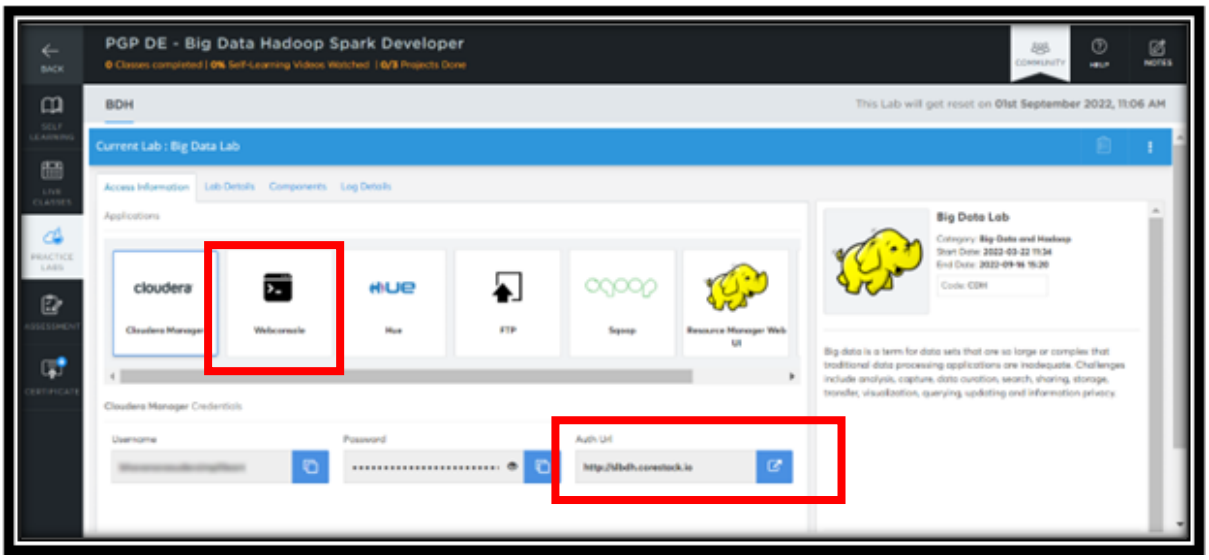
Step 7: Click on the **"HDFS"** icon and click on the **"+"** symbol to upload the dataset



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**"



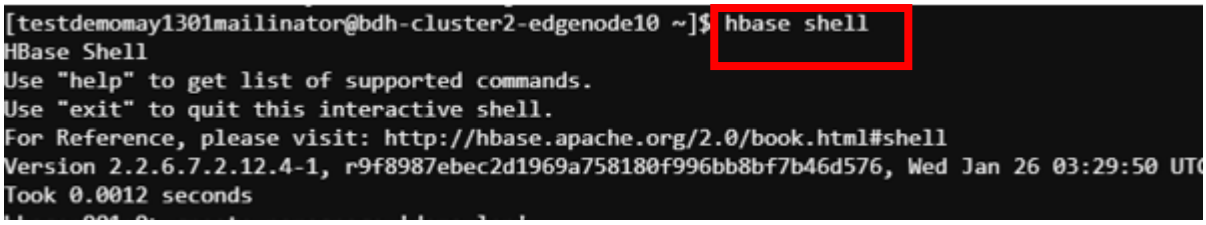
Step 10: Copy the "**Username**" and the "**Password**" provided to log in to the Webconsole

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:

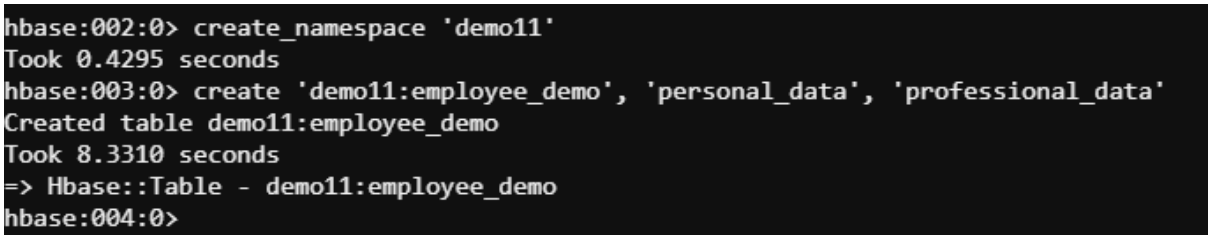
Command: hbase shell



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

Command:

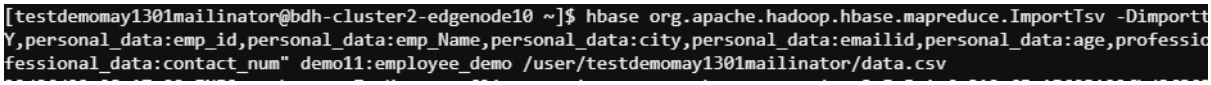
a) create_namespace 'demo11'
b) create 'demo11:employee_demo', 'personal_data', 'professional_data'



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



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
```