Big Data Hadoop and Spark Developer

Lesson-End Project Solution

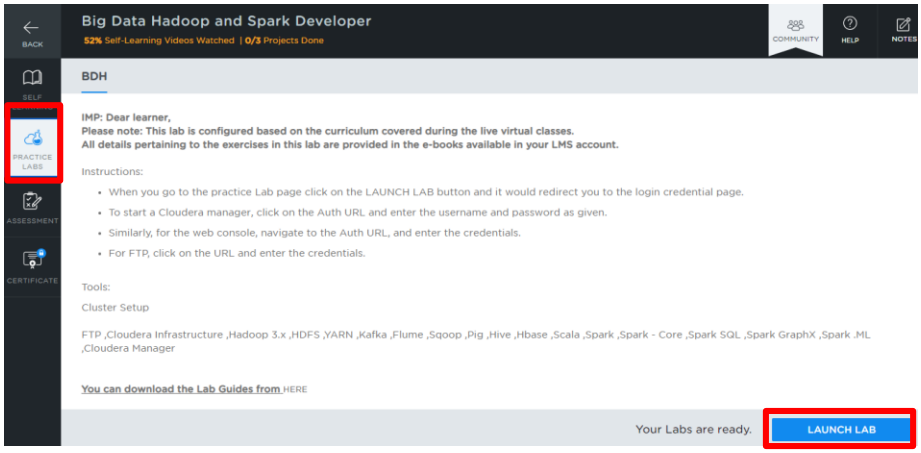
Uploading Data from HDFS to HBase

Steps to Perform:

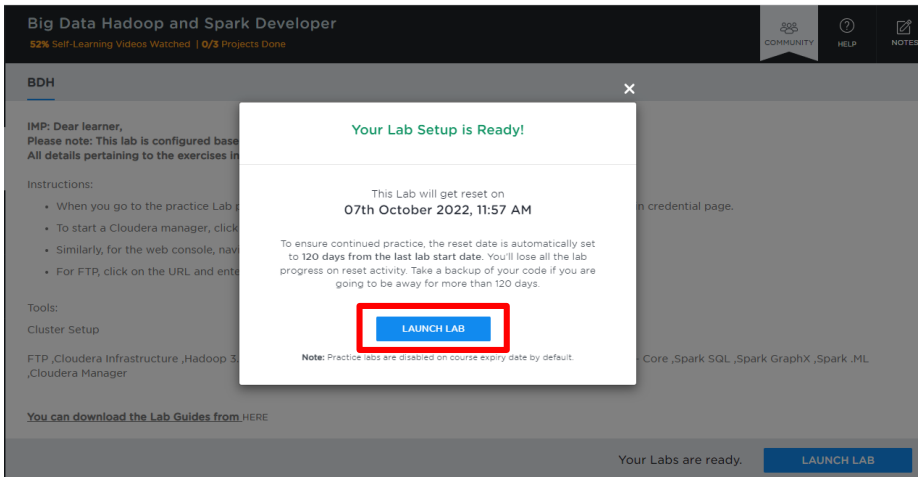
Step 1: Log in to your LMS account

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

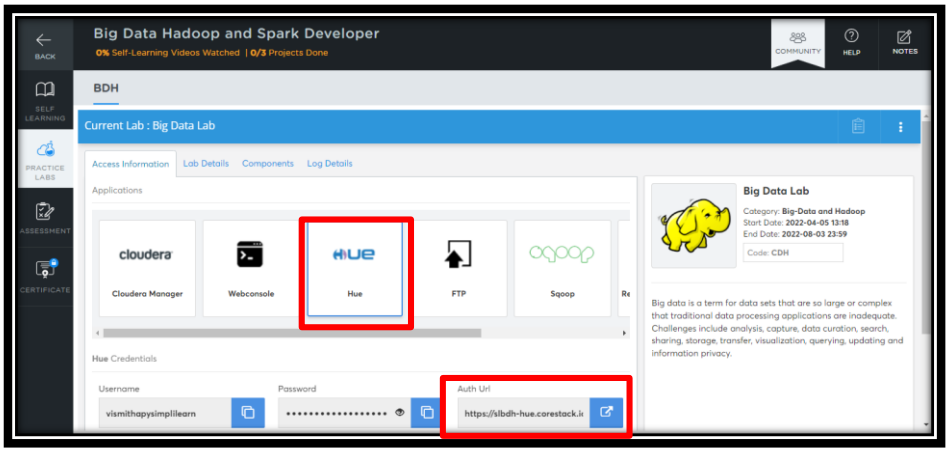
Step 3: Click on the "**PRACTICE LABS**" tab on the left side and select "**LAUNCH LAB**"



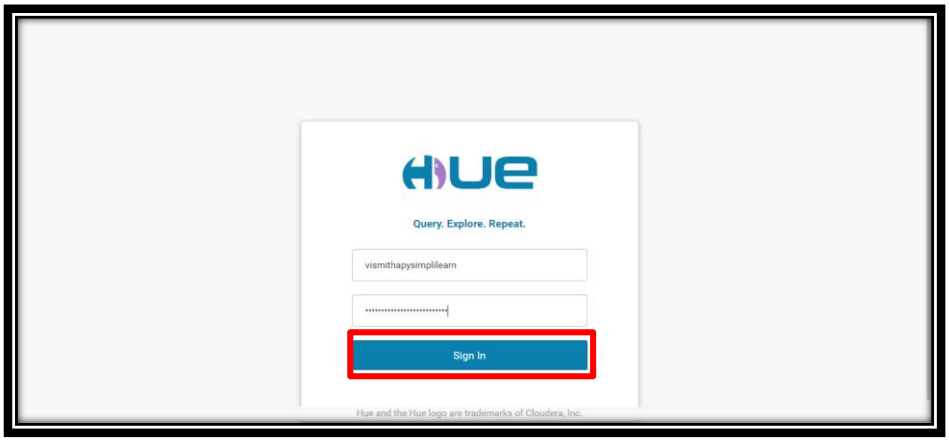
Step 4: Click on the "**LAUNCH LAB**" button



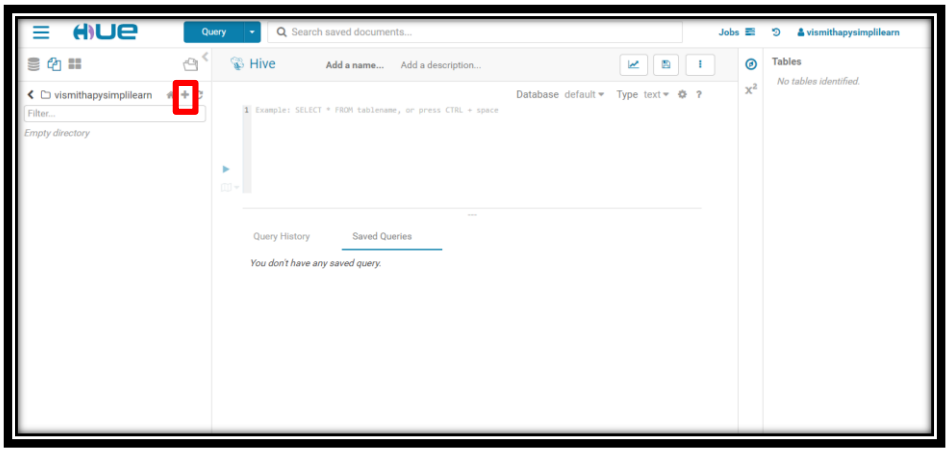
Step 5: Click on **"HUE"** and click on the **"Auth Url"** to upload the dataset. Copy the "**Username**" and "**Password**" provided to log in to **"HUE"**



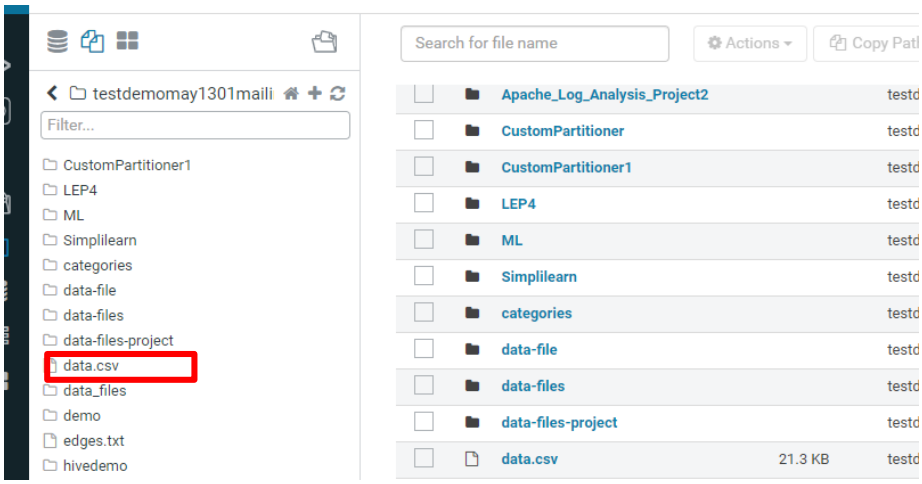
Step 6: Paste the **"Username"** and **"Password"** on the login window and click on sign In



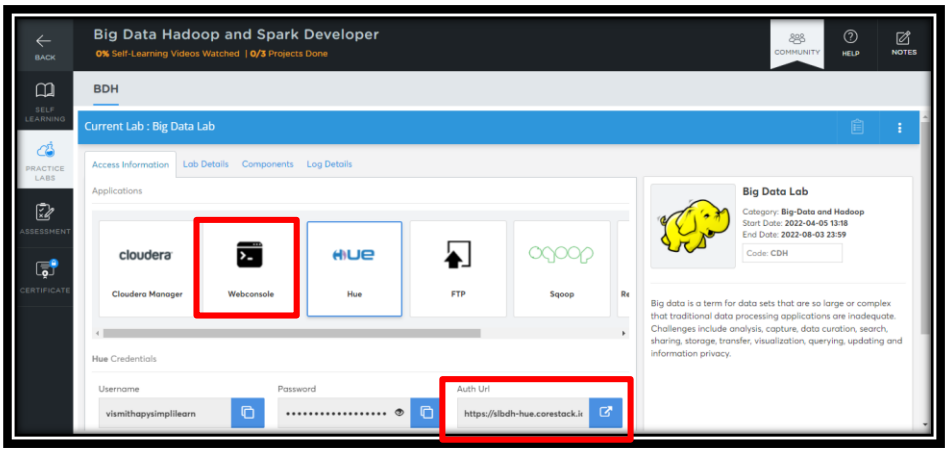
Step 7: Click on the **“HDFS”** icon and then the **“+”** symbol to upload the dataset



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 then on "**Auth Url**"



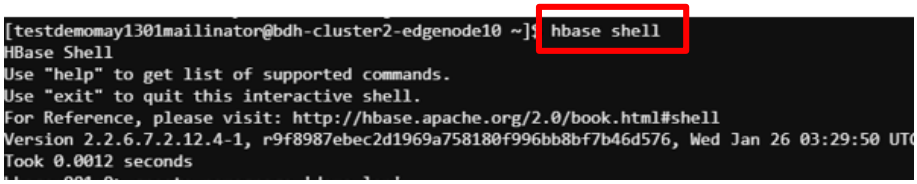
Step 10: Copy the "**Username**" and "**Password**" provided to log in to Webconsole

Step 11: Paste the "**Username**" and "**Password**" on the console and click on Enter

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

Step 12: Log in to the HBase shell using the below command:

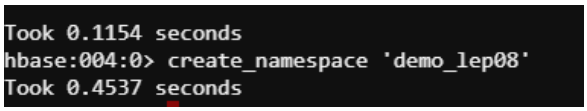
Command: hbase shell



Step 13: Create a namespace **"demo_lep08".** By using the namespace create a table **"employee_demo"** using the below commands**:**

Command:

a) create_namespace 'demo_lep08'
b) create 'demo_lep08:employee_demo', 'personal_data', 'professional_data'



```
hbase:005:0> create 'demo_lep08:employee_demo', 'personal_data', 'professional_data'

Created table demo_lep08:employee_demo
Took 8.3405 seconds
=> Hbase::Table - demo_lep08:employee_demo
hbase:006:0>
hbase:007:0>
```

Step 14: Display the metadata of the table using the command below:

Command:

describe 'demo_lep08':employee_demo'

```
hbase:007:0> describe 'demo_lep08:employee_demo'
Table demo_lep08:employee_demo is ENABLED
demo_lep08:employee_demo
COLUMN FAMILIES DESCRIPTION
{NAME => 'personal_data', VERSIONS => '1', NEW_VERSION_BEHAVIOR => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', MIN_VER
SIONS => '0', REPLICATION_SCOPE => '0', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', COMPRESSION => 'NONE', BLOCKCACHE => 'true', BLOCKSIZE => '65536'}

{NAME => 'professional_data', VERSIONS => '1', NEW_VERSION_BEHAVIOR => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', MIN
_VERSIONS => '0', REPLICATION_SCOPE => '0', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', COMPRESSION => 'NONE', BLOCKCACHE => 'true', BLOCKSIZE => '65536'}

2 row(s)
Quota is disabled
Took 0.3878 seconds
hbase:008:0>
```

Step 15: Insert the data into the table as shown below:

```
hbase:008:0> put 'demo_lep08:employee_demo','1','personal_data:emp_id','1001'
Took 0.1921 seconds
hbase:009:0> put 'demo_lep08:employee_demo','1','personal_data:emp_Name','mateuz'
Took 0.0041 seconds
hbase:010:0> put 'demo_lep08:employee_demo','1','personal_data:city','NY'
Took 0.0041 seconds
hbase:011:0> put 'demo_lep08:employee_demo','1','personal_data:emailid','Chhaya@microsoft.com'
Took 0.0033 seconds
hbase:012:0> put 'demo_lep08:employee_demo','1','personal_data:age','18'
Took 0.0035 seconds
hbase:013:0> put 'demo_lep08:employee_demo','1','personal_data:hobbies','Reading'
Took 0.0034 seconds
hbase:014:0> put 'demo_lep08:employee_demo','1','personal_data:favsports','Badminaton'
Took 0.0035 seconds
hbase:015:0> put 'demo_lep08:employee_demo','1','professional_data:designation','CTO'
Took 0.0035 seconds
hbase:016:0> put 'demo_lep08:employee_demo','1','professional_data:salary',130000
Took 0.0034 seconds
hbase:017:0> put 'demo_lep08:employee_demo','1','professional_data:contact_num',123658715
Took 0.0035 seconds
hbase:018:0> put 'demo_lep08:employee_demo','1','professional_data:hike',11
Took 0.0039 seconds
hbase:019:0> put 'demo_lep08:employee_demo','1','professional_data:reportingmanager',Barabara
Traceback (most recent call last):
NameError (uninitialized constant Barabara)
hbase:020:0> put 'demo_lep08:employee_demo','1','professional_data:reportingmanager','Barabara'
Took 0.0043 seconds
hbase:021:0>
```

Step 16: Read the data to ensure if it is present, as shown below:

```
hbase:021:0> get 'demo_lep08:employee_demo','1'
COLUMN                                  CELL
 personal_data:age                      timestamp=1654810346469, value=18
 personal_data:city                     timestamp=1654810346440, value=NY
 personal_data:emailid                  timestamp=1654810346455, value=Chhaya@microsoft.com
 personal_data:emp_Name                 timestamp=1654810346416, value=mateuz
 personal_data:emp_id                   timestamp=1654810346398, value=1001
 personal_data:favsports                timestamp=1654810346497, value=Badminaton
 personal_data:hobbies                  timestamp=1654810346484, value=Reading
 professional_data:contact_num          timestamp=1654810346548, value=123658715
 professional_data:designation          timestamp=1654810346511, value=CTO
 professional_data:hike                 timestamp=1654810346565, value=11
 professional_data:reportingmanager     timestamp=1654810396422, value=Barabara
 professional_data:salary               timestamp=1654810346526, value=130000
1 row(s)
Took 0.0200 seconds
hbase:022:0> get 'demo_lep08:employee_demo','1','personal_data:emailid'
COLUMN                                  CELL
 personal_data:emailid                  timestamp=1654810346455, value=Chhaya@microsoft.com
1 row(s)
Took 0.0048 seconds
hbase:023:0>
```

Step 17: Delete the data as shown below:

```
hbase:023:0> truncate 'demo_lep08:employee_demo'
Truncating 'demo_lep08:employee_demo' table (it may take a while):
Disabling table...
Truncating table...
Took 22.5074 seconds
hbase:024:0>
```

Step 18: Execute the below command to fetch bulk data from an HDFS into the table:
Command:

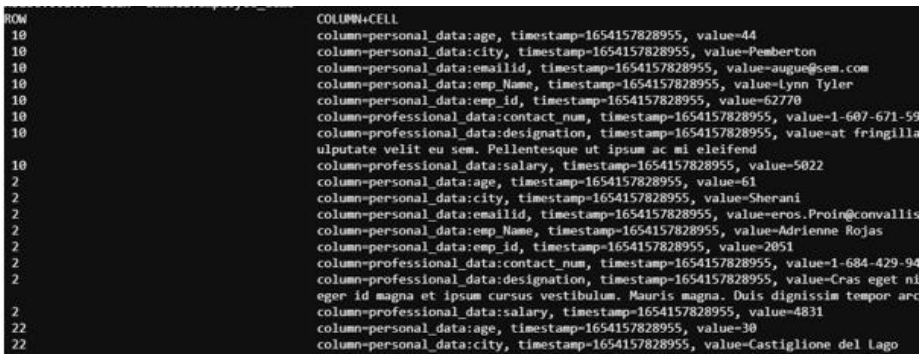
hbase org.apache.hadoop.hbase.mapreduce.ImportTsv -Dimporttsv.separator=, -Dimporttsv.columns="HBASE_ROW_KEY,personal_data:emp_id,personal_data:emp_Name,personal_data:city,personal_data:emailid,personal_data:age,professional_data:designation,professional_data:salary,professional_data:contact_num" demo_lep08:employee_demo /user/testdemomay1301mailinator/data.csv

```
[testdemomay1301mailinator@bdh-cluster2-edgenode10 ~]$ hbase org.apache.hadoop.hbase.mapreduce.ImportTsv -Dimporttsv.separator=,  -Dimport
Y,personal_data:emp_id,personal_data:emp_Name,personal_data:city,personal_data:emailid,personal_data:age,professional_data:designation,pro
fessional_data:contact_num" demo_lep08:employee_demo /user/testdemomay1301mailinator/data.csv
```

Step 19: Verify whether the data is in the table by using the command below:

Command:

scan **‘demo_lep08: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
```