

Assisted Practice: Execution of MapReduce Job

Problem Scenario: Write the commands to perform partitions using the JAR files for the execution of a MapReduce job

Objective: In this demonstration, you will use JAR files and the wordcount.txt file to perform the partitions in the execution of a MapReduce job.

Dataset Name: “wordcount.txt”

Tasks to Perform:

1. Download the **Hadoop-mapreduce-example.jar** file and **wordcount.txt** file
2. Log in to the FTP using the username and password from the lab and upload the file
3. Log in to the Webconsole using the username and password from the lab and create a new directory **demo** in HDFS using the **mkdir** command
4. Push the **wordcount.txt** file into the directory using the put command
5. Execute the command to move the **Hadoop-mapreduce-example.jar** file to the HDFS directory
6. View the files in the **Output** folder with the part files

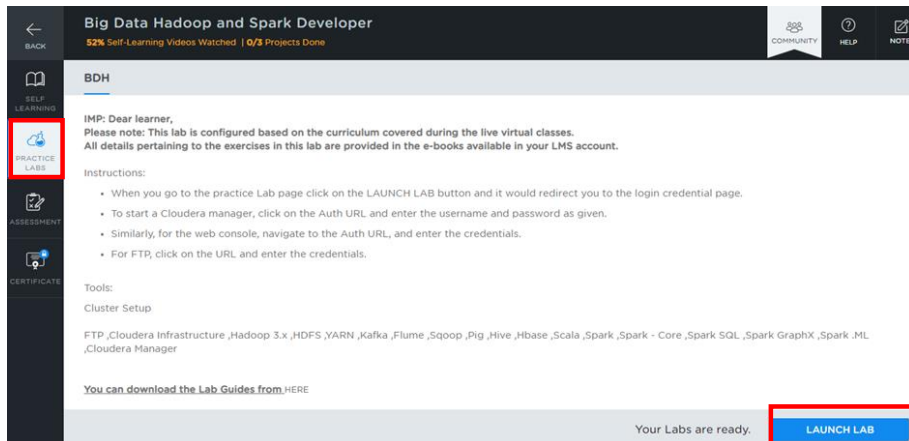
Steps to Perform:

Step 1: Download the dataset named “wordcount.txt” and the JAR file 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 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:

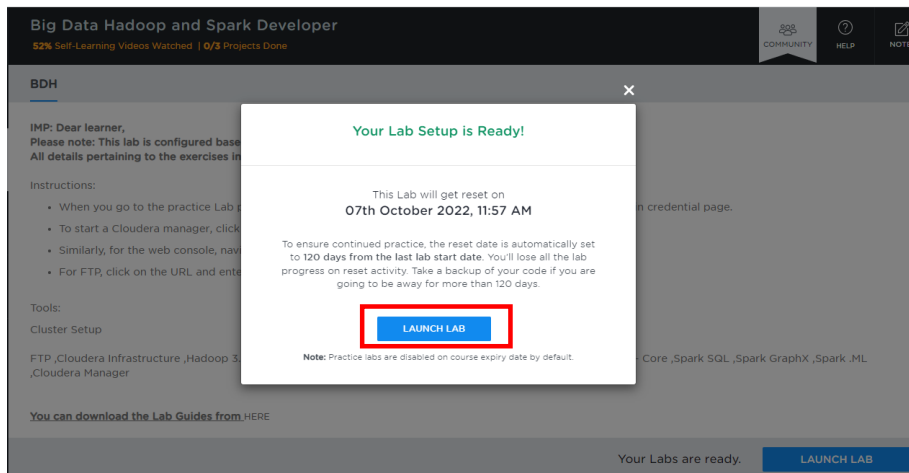
Cluster Setup

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
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.

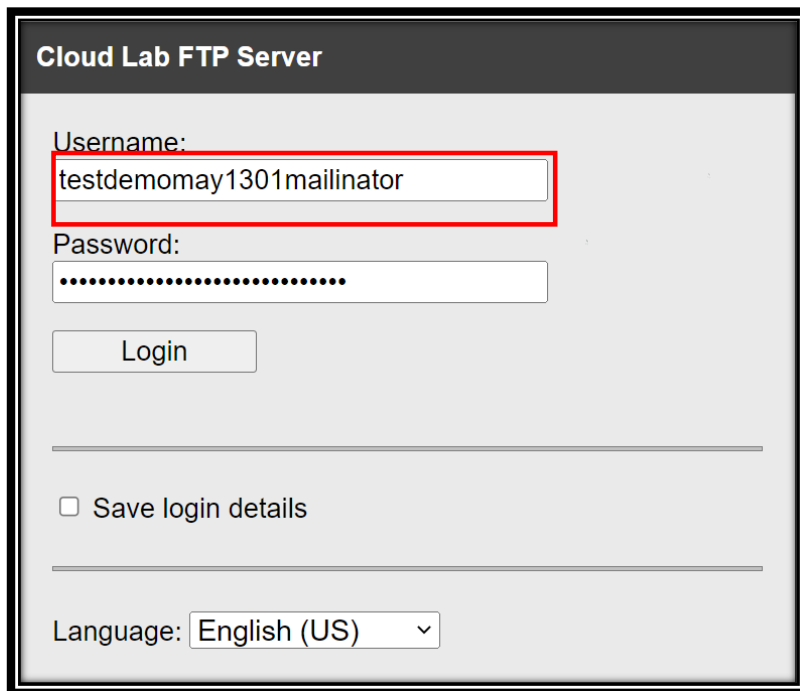
LAUNCH LAB

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

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

The screenshot displays the 'Big Data Hadoop and Spark Developer' interface. The top navigation bar includes 'BACK', '52% Self Learning Videos Watched | 0/3 Projects Done', 'COMMUNITY', 'HELP', and 'NOTES'. The main header shows 'BDH' and a reset notice: 'This Lab will get reset on 07th October 2022, 11:57 AM'. The left sidebar contains 'SELF LEARNING', 'PRACTICE LABS', 'ASSESSMENT', and 'CERTIFICATE'. The main content area is titled 'Current Lab : Big Data Lab' and includes tabs for 'Access Information', 'Lab Details', 'Components', and 'Log Details'. Under 'Applications', there are icons for 'cloudera', 'Webconsole', 'Hue', 'FTP' (highlighted with a red box), and 'Sqoop'. Below this, the 'Cloudera Manager Credentials' section shows 'Username' as 'alpiguptasimplilearn', 'Password' as masked dots, and 'Auth Url' as 'http://vlabh.coresstack.ie' (highlighted with a red box). A right sidebar provides lab details: 'Big Data Lab', 'Category: Big Data and Hadoop', 'Start Date: 2021-10-26 14:05', 'End Date: 2022-10-07 12:00', and 'Code: CDH'. A descriptive paragraph about big data is also present.

Step 7: Paste the **“Username”** and the **“Password”** on the login window and click on **“Login”**.



Cloud Lab FTP Server

Username:

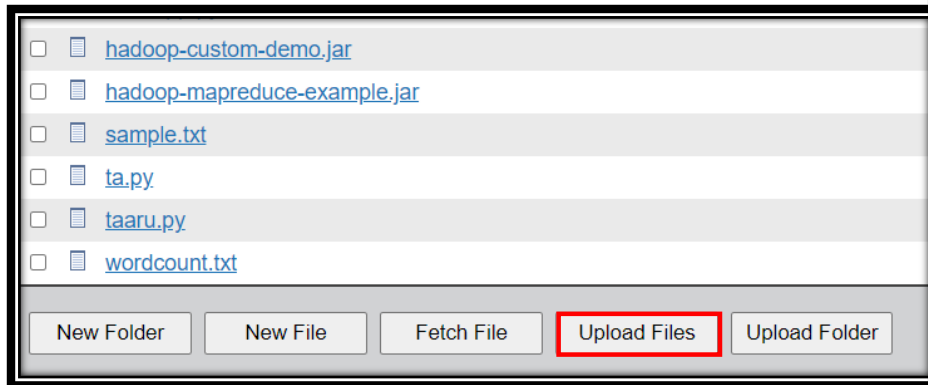
Password:

Login

☐ Save login details

Language:

Step 8: Click on the **“Upload Files”** icon and upload the **“wordcount.txt”** and **“Hadoop-mapreduce-example.jar”** file into the FTP



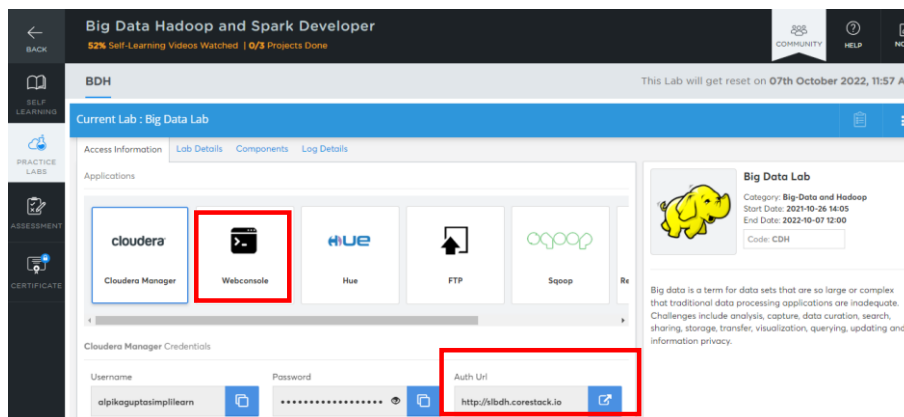
Step 9: Go back to the lab window and click on **“Webconsole”** and click on the **“Auth Url”**

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

Commented [SB1]: Shouldn't this be Step 10? Change all the following steps numberings too accordingly

Commented [AG2R1]: sure

Commented [AG3R1]: done



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: Create a directory using mkdir named “demo.” After this, load the “wordcount.txt” file into HDFS using the below command:

Command:

```
hdfs dfs -mkdir demo
```

```
hdfs dfs -put wordcount.txt demo
```

Note: Make sure you have the “Hadoop-mapreduce-example.jar” file present in FTP using ls command

```
[testdemomay1301mailinator@bdh-cluster2-edgenode10 ~]$ hdfs dfs -mkdir demo
[testdemomay1301mailinator@bdh-cluster2-edgenode10 ~]$ hdfs dfs -put wordcount.txt demo
[testdemomay1301mailinator@bdh-cluster2-edgenode10 ~]$ ls
abc      data_files  hadoop-mapreduce-example.jar  lin_reg.py  mapexample.py  sample.txt  swap.py      wordcount.txt
convert.py dictionary.py kmeans.py      log_reg.py  Sample         Sample.txt  WordCount.java
[testdemomay1301mailinator@bdh-cluster2-edgenode10 ~]$ hdfs dfs -put wordcount.txt demo/
put: 'demo/wordcount.txt': File exists
[testdemomay1301mailinator@bdh-cluster2-edgenode10 ~]$ yarn jar hadoop-mapreduce-example.jar demo/wordcount.txt demo/
```

Commented [SB4]: The sentence is not clear. Please check and rephrase.

Commented [AG5R4]: Ls is a command so when we write ls then we can see the no of files present in the FTP

Commented [AG6R4]: done

Step 13: Now, execute the below command and see if your job gets executed successfully

Note: Change the username of the Hadoop directory to “testdemomay1301mailinator” as assigned in your Lab

Command:







```
yarn jar hadoop-mapreduce-example.jar
org.simplilearn.demo.mapreduce.wordcount.WordCount /user/
testdemomay1301mailinator/demo/ wordcount.txt demo/Output
```

Commented [SB7]: Is the spacing correct here? Please check

Commented [AG8R7]: After wordcount spacing will come. It is correct

```
[testdemomay1301mailinator@bdh-cluster2-edgenode10 ~]$ yarn jar hadoop-mapreduce-example.jar org.simplilearn.demo.mapreduce.wordcount.WordCount /user/testdemomay1301mailinator/demo/wordcount.txt demo/Output
WARNING: YARN_OPTS has been replaced by HADOOP_OPTS. Using value of YARN_OPTS.
22/05/25 16:20:11 INFO hdfs.DFSClient: Created token for testdemomay1301mailinator: HDFS_DELEGATION_TOKEN owner=testdemomay1301mailinator@BDH-ENV-ONE4-RUTX_CLUSTER
```

Step 14: Login into the Hue and you will be able to see the “demo” directory. Open the “demo” directory and you will see one more folder named “Output” where you will be able to see the part files

<input type="checkbox"/>	Name	Size
<input type="checkbox"/>	 ↑	
<input type="checkbox"/>	 .	
<input type="checkbox"/>	 _SUCCESS	0 bytes
<input type="checkbox"/>	 part-r-00000	45 bytes
<input type="checkbox"/>	 part-r-00001	46 bytes
<input type="checkbox"/>	 part-r-00002	47 bytes