

Assisted Practice: Execution of MapReduce Job

Using Custom Partitioner

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

Objective: In this demonstration, you will explore the execution of a MapReduce job in Simplilearn's Lab by creating a custom partitioner.

Dataset Name: "wordcount.txt"

Tasks to Perform:

1. Download the **hadoop-custom-demo.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 **CustomPartitioner20** 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-custom-demo.jar** file to the HDFS directory
6. View the files in the **final** folder with the part files

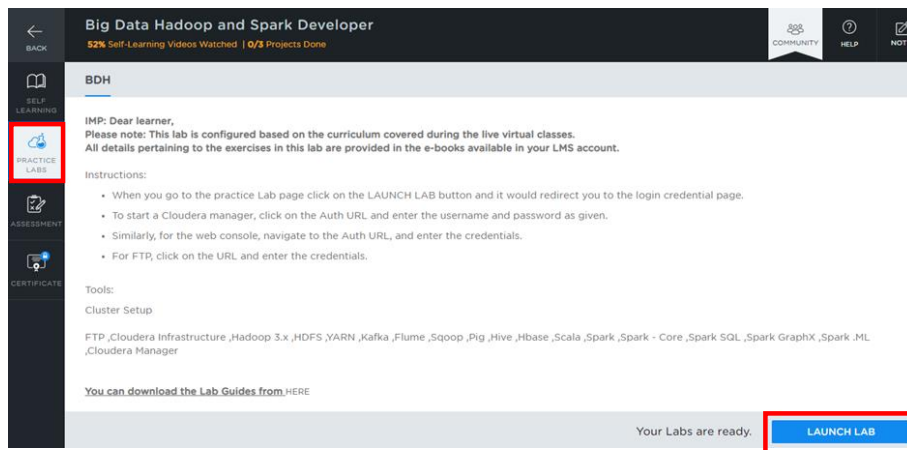
Steps to Perform:

Step 1: Download the dataset named "wordcount.txt" and the JAR file from the course resources section

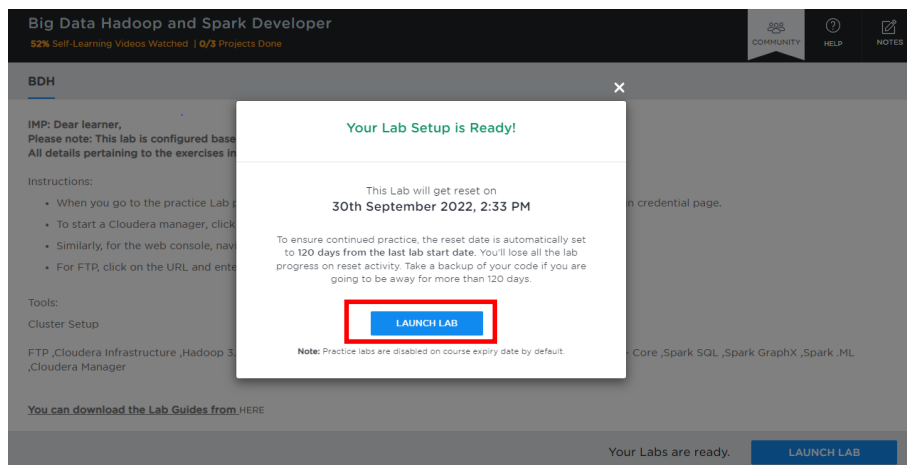
Step 2: Log in to your LMS account

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



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



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' (BDH) lab interface. The main navigation sidebar on the left includes options for BACK, SELF LEARNING, PRACTICE LABS, ASSESSMENT, and CERTIFICATE. The current lab is 'Big Data Lab', which will reset on 07th October 2022, 11:57 AM. The 'Components' tab is active, showing a grid of application tiles: Cloudera Manager, Webconsole, Hue, FTP, and Sqoop. The FTP tile is highlighted with a red box. Below this grid, the 'Cloudera Manager Credentials' section contains fields for Username (alpiguptasimplilearn), Password (masked), and Auth Url (http://kibdh.coreslack.io). The Auth Url field is also highlighted with a red box. On the right, a 'Big Data Lab' information card provides details about the lab category, dates, and code (CDH), along with a brief definition of big data.

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

Cloud Lab FTP Server

Username:

Password:

☐ Save login details

Language:

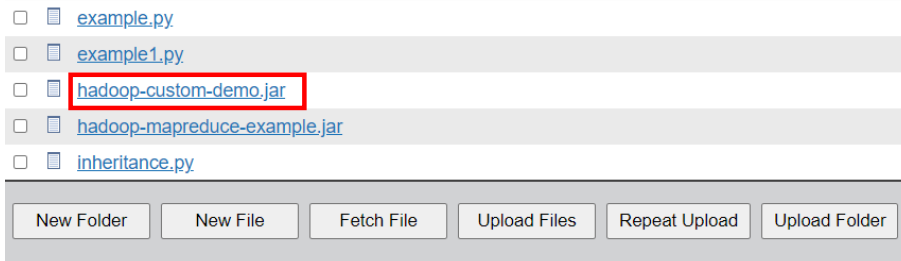
Step 8: Click on the “Upload Files” icon and upload the “wordcount.txt” and “hadoop-custom-demo.jar” files into the FTP

☐ [search.py](#)

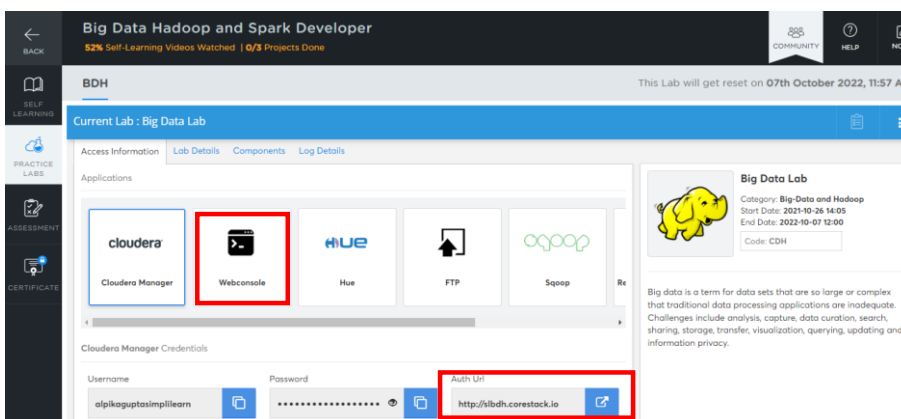
☐ [spark-streaming-example.py](#)

☐ [swap.py](#)

☐ [wordcount.txt](#)



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**”

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

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

Commented [AG2R1]: sure

Commented [AG3R1]: done

```
bdh-cluster2-edgenode10 login: testdemomay1301mailinator
Password:
Last login: Tue Jun  7 10:44:11 on pts/0

Welcome to CentOS 7

=====
 *                               :
=====

Password for testdemomay1301mailinator@BDH-ENV.GNE4-RUTX.CLOUDERA.SITE:
[testdemomay1301mailinator@bdh-cluster2-edgenode10 ~]$
```

Step 12: Create a directory using mkdir named **“CustomPartitioner20.”** After this, load the **“wordcount.txt”** file into HDFS using the below command:

Command:

```
hdfs dfs -mkdir CustomPartitioner20
```

```
hdfs dfs -put wordcount.txt CustomPartitioner20
```

Note: Make sure you have the **“hadoop-custom-demo.jar”** file present in FTP using the ls command

```
[testdemomay1301mailinator@bdh-cluster2-edgenode10 ~]$ hdfs dfs -mkdir CustomPartitioner20
[testdemomay1301mailinator@bdh-cluster2-edgenode10 ~]$ hdfs dfs -put wordcount.txt CustomPartitioner20
[testdemomay1301mailinator@bdh-cluster2-edgenode10 ~]$ ls
13_4AP.py      classDemo.py  demoll       example.py    kmeans.py     mapexample.py
abc            convert.py    derby.log    hadoop-custom-demo.jar  Lesson_13_Dataset.csv  map.py
abstractAP.py  data.csv      dictionary.py hadoop-mapreduce-example.jar  lin_reg.py      metastore_db
abstract.py    data_files    encapsulation.py inheritance.py  LoggerAnalysis.jar  module.py
Apache-log.log data_files_CEP example1.py   insta-cart    log_reg.py     poly.py
[testdemomay1301mailinator@bdh-cluster2-edgenode10 ~]$
```

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-custom-demo.jar
```

```
/user/testdemomay1301mailinator/CustomPartitioner20/wordcount.txt
```





```
/user/testdemomay1301mailinator/CustomPartitioner20/final
```

```
testdemomay1301mailinator@bdh-cluster2-edgenode10 ~]$ yarn jar hadoop-custom-demo.jar /user/testdemomay1301mailinator/CustomPartitioner20/wordcount.txt /user/testdemomay1301mailinator/CustomPartitioner20/final
```

Step 14: Login into the Hue and you will be able to see the






“**CustomPartitioner20**” directory. Open the “**CustomPartitioner20**” directory and you will see one more folder named “**final**” where you will be able to see the part files

[Home](#) /user/testdemomay1301mailinator/CustomPartitioner20

Name	Size	User	Group
 ↑		testdemomay1301maili...	testdemomay1301maili...
 .		testdemomay1301maili...	testdemomay1301maili...
 final		testdemomay1301maili...	testdemomay1301maili...
 wordcount.txt	790 bytes	testdemomay1301maili...	testdemomay1301maili...

w 45 of 2 items Page 1

[Home](#) /user/testdemomay1301mailinator/CustomPartitioner20/final

<input type="checkbox"/>	Name	Size	User	Group
<input type="checkbox"/>	 ↑		testdemomay1301maili...	testdemomay1301maili...
<input type="checkbox"/>	 .		testdemomay1301maili...	testdemomay1301maili...
<input type="checkbox"/>	 _SUCCESS	0 bytes	testdemomay1301maili...	testdemomay1301maili...
<input type="checkbox"/>	 part-r-00000	176 bytes	testdemomay1301maili...	testdemomay1301maili...
<input type="checkbox"/>	 part-r-00001	589 bytes	testdemomay1301maili...	testdemomay1301maili...

Show of 3 items Page