SMIT R PATEL

19162121031

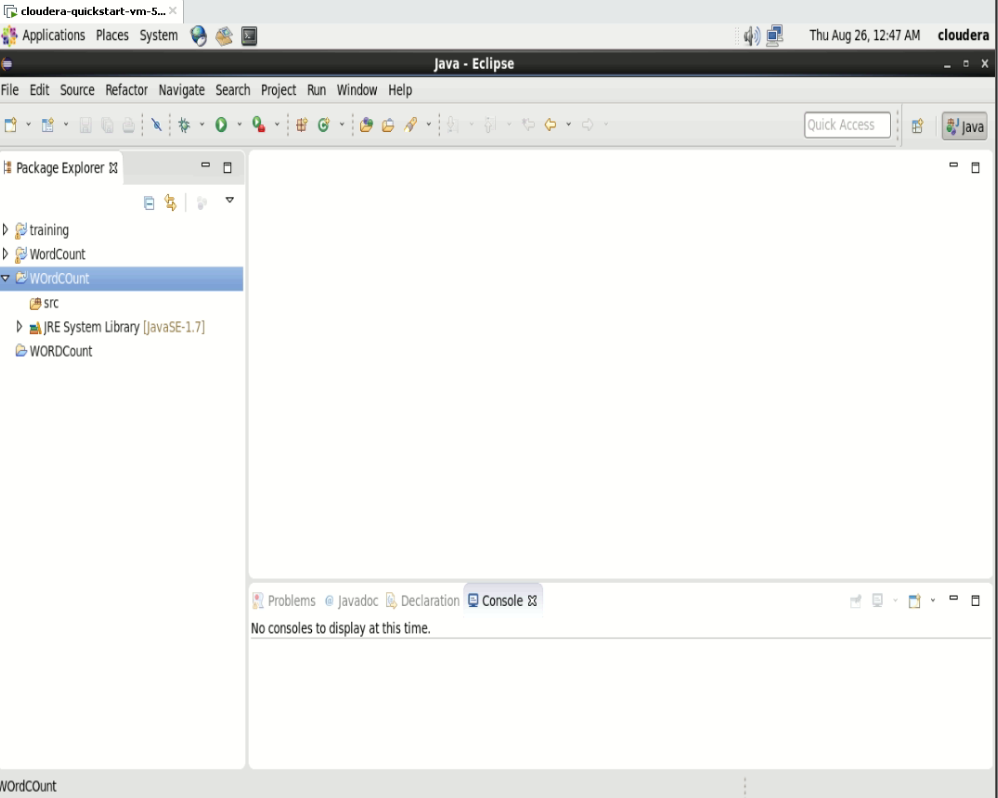
SEM 5

PRACTICAL 4

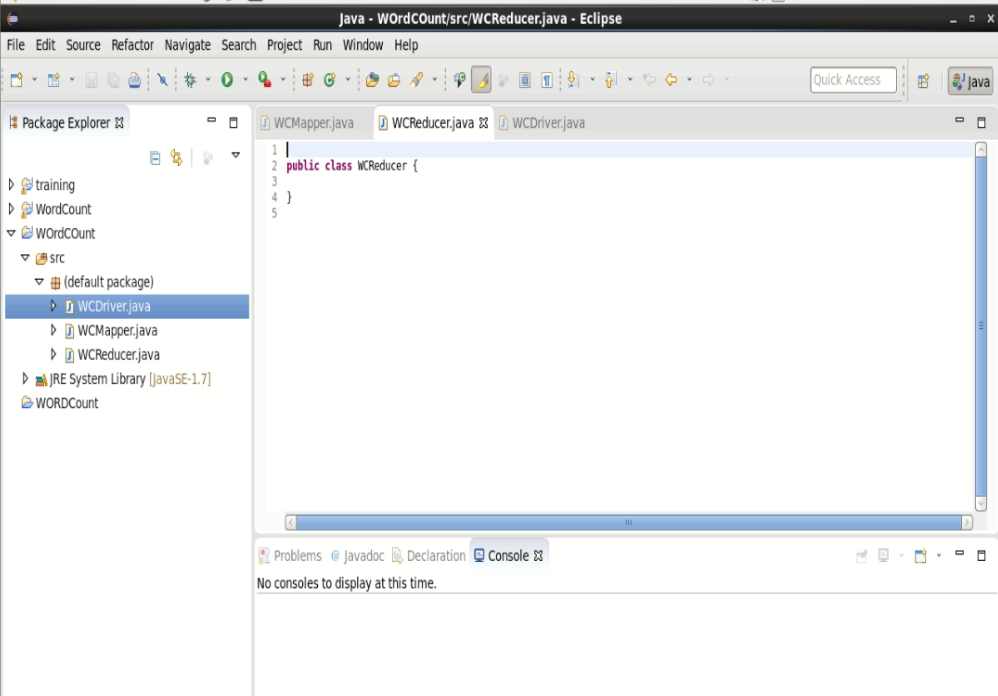
BIG DATA AND ANALYTICS

Aim: Understanding how to Execute WordCount Program in MapReduce using Cloudera Distribution Hadoop (CDH). You are deployed as a trainer in a multi-national company that is planning to adapt to big data practices. The batch of employees you are training- are freshers to Hadoop. You need to teach them to execute a basic Word Count MapReduce program. Counting the number of words in any language is a piece of cake like in C, C++, Python, Java, etc. MapReduce also uses Java but it is very easy if you know the syntax on how to write it. It is the basic of MapReduce. You will first learn how to execute this code similar to “Hello World” program in other languages. So here are the steps which show how to write a MapReduce code for Word Count.

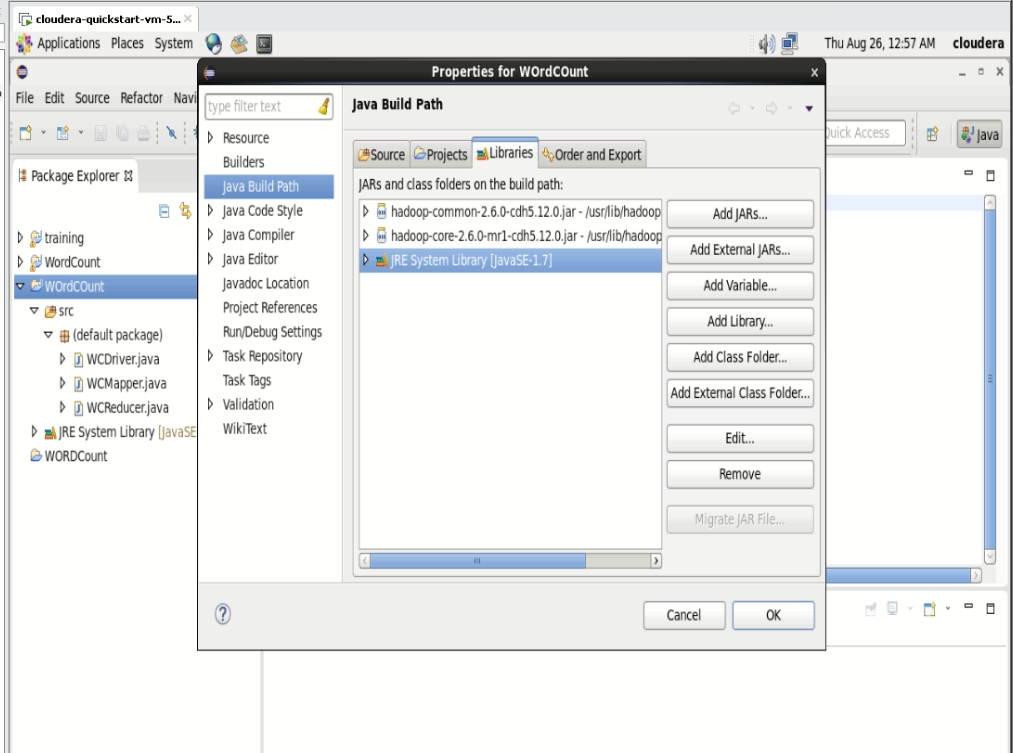
Tasks: 1. First Open Eclipse -> then select File -> New -> Java Project ->Name it WordCount -> then Finish.

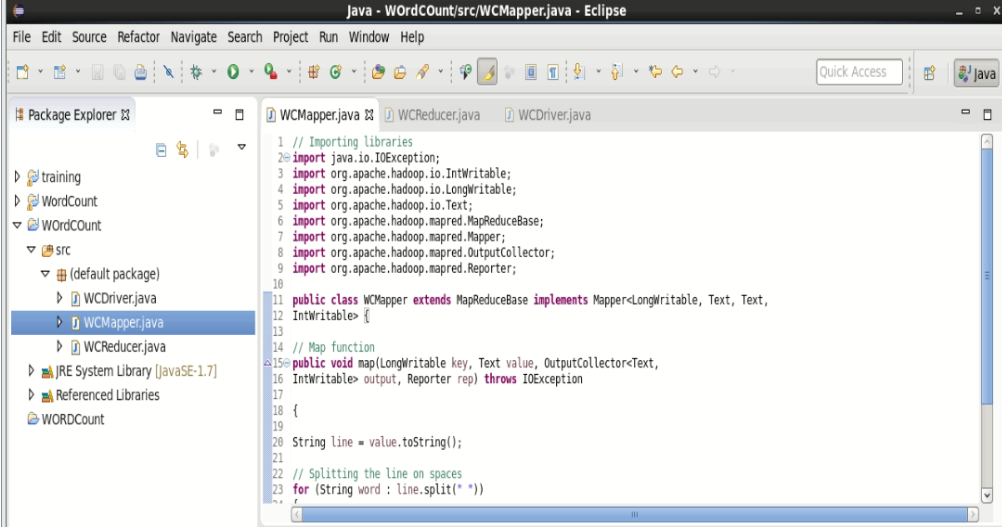


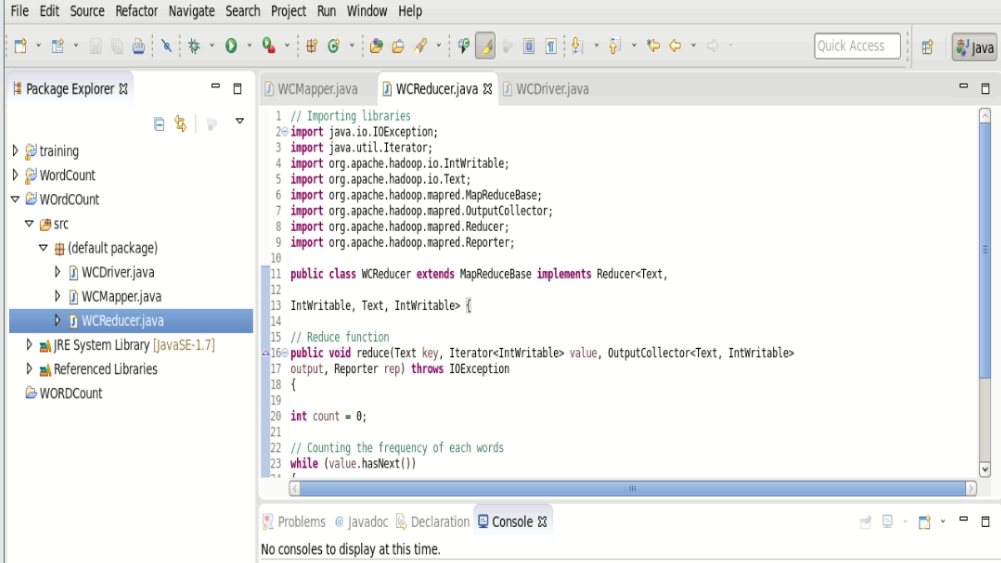
2. Create Three Java Classes into the project. Name them WCDriver(having the main function), WCMapper, WCReducer.

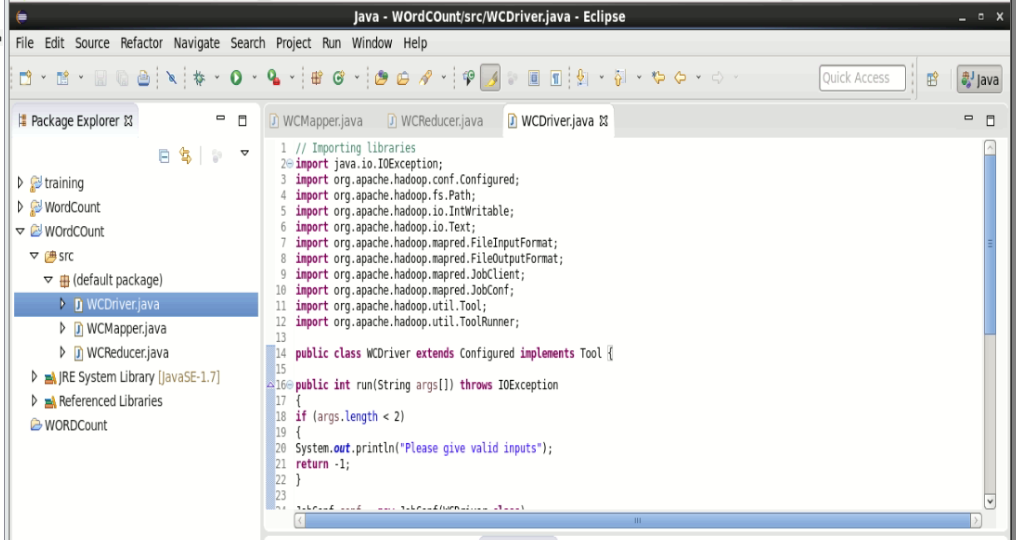


3. You have to include two Reference Libraries for that: Right Click on Project -> then select Build Path-> Click on Configure Build Path. You will see an “Add External JARs” option on the Right-Hand Side. Click on it and add the below mention files. You can find these files in /usr/lib/ 1. /usr/lib/hadoop-0.20-mapreduce/hadoop-core-2.6.0-mr1- cdh5.13.0.jar 2. /usr/lib/hadoop/hadoop-common-2.6.0-cdh5.13.0.jar

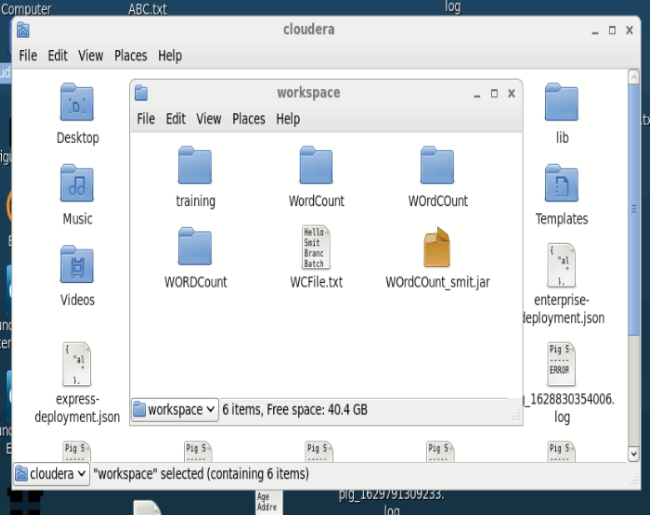


4. Writing Code in WCDriver(having the main function), WCMapper, WCReducer.

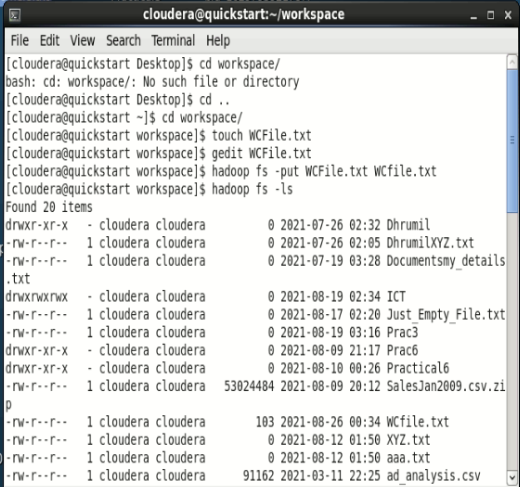


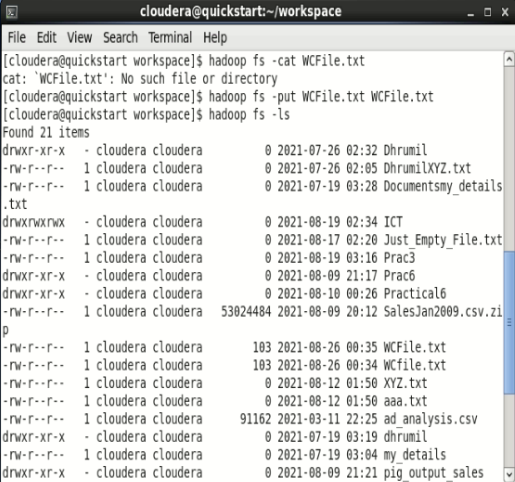


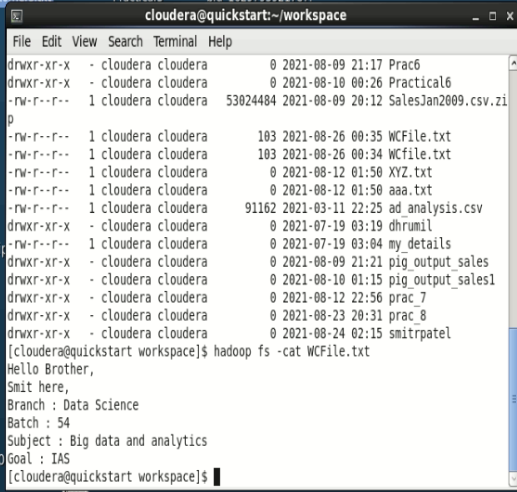
5. After writing the MapReduce code into the classes, make a jar file. Right Click on Project-> Click on Export-> Select export destination as Jar File-> Name the jar File (WordCount.jar) -> Click on next -> at last Click on Finish. Now copy this file into the Workspace directory of Cloudera.



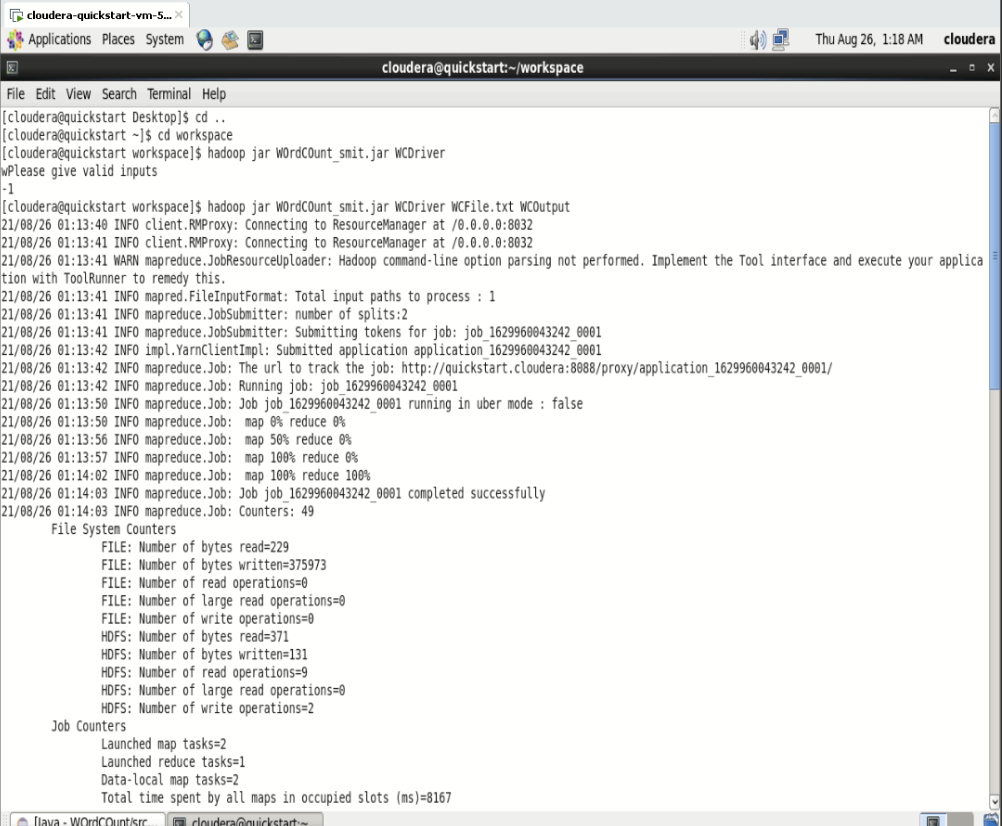
6. Open the terminal on CDH and change the directory to the workspace. You can do this by using “cd workspace/” command. Now, create a text file (WCFile.txt) and move it to HDFS. For that open terminal and write this code (remember you should be in the same directory as jar file you have created just now). Then, run this command to copy the file input file into the HDFS. “hadoop fs -put WCFile.txt WCFile.txt”

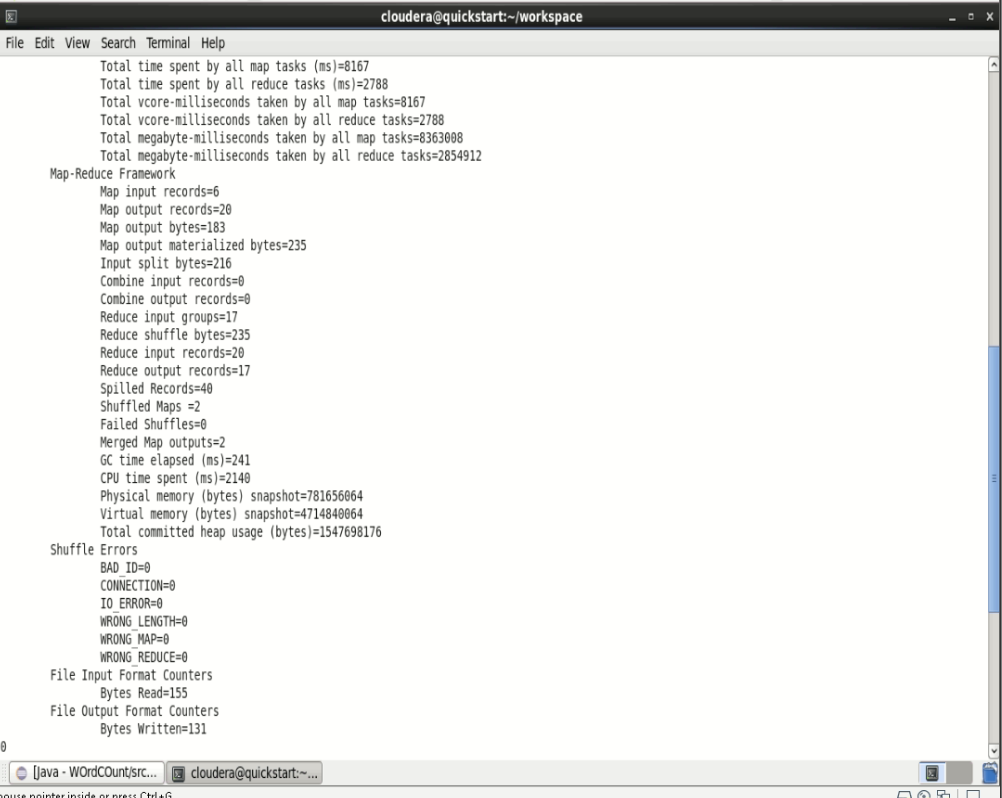






7. Run this command to store the output of map reduce in hadoop’s directory. Command: hadoop jar WordCount.jar WCDriver WCFile.txt WCOutput





8. After Executing the code, you can see the result in WCOutput file or by writing following command on terminal. Command: hadoop fs -cat WCOutput/part-00000

