

**CSE488: Big Data Analytics**

**[SPRING 2023]**

**Assignment**

**LAB-02**

**Submitted by:**

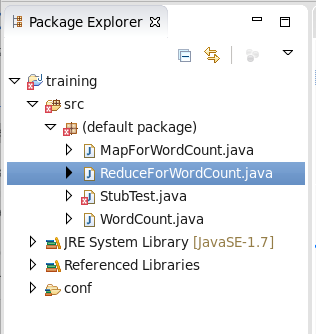
Student ID: 2019-2-60-072

Student Name: Mujahidul Islam

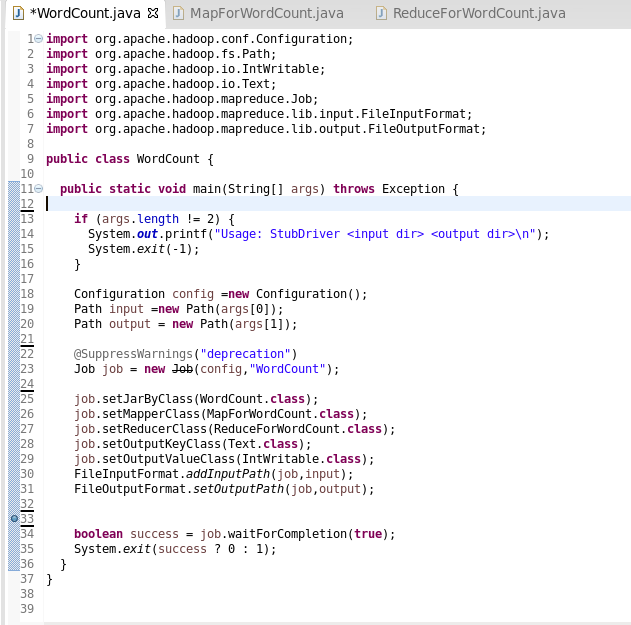
Showing the execution of the WordCount application in Hadoop MapReduce Environment step by step.

Step A: Writing Code

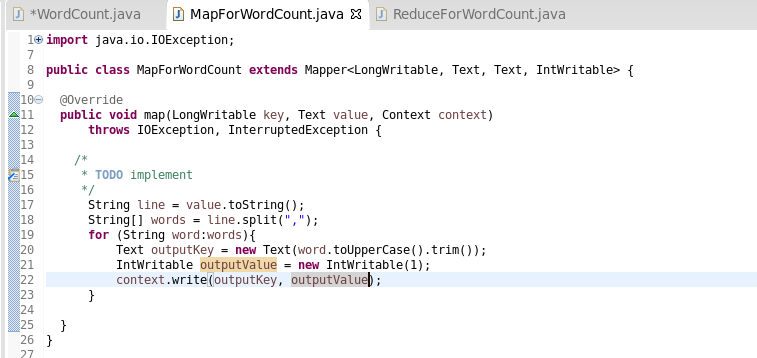
1. Launch Eclipse on Cloudera Environment.
2. Rename the file under /training/src/default package according to the file name given in image. [Remain the subtest as it is]



1. Write the code as it is-



Code for main class



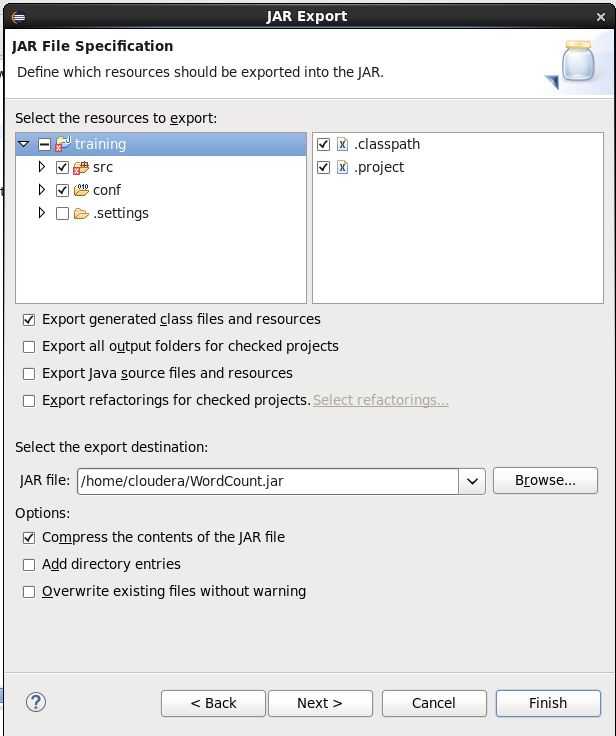
Code for Mapper Class



Code for Reducer Class

Step B: Creating Jar file

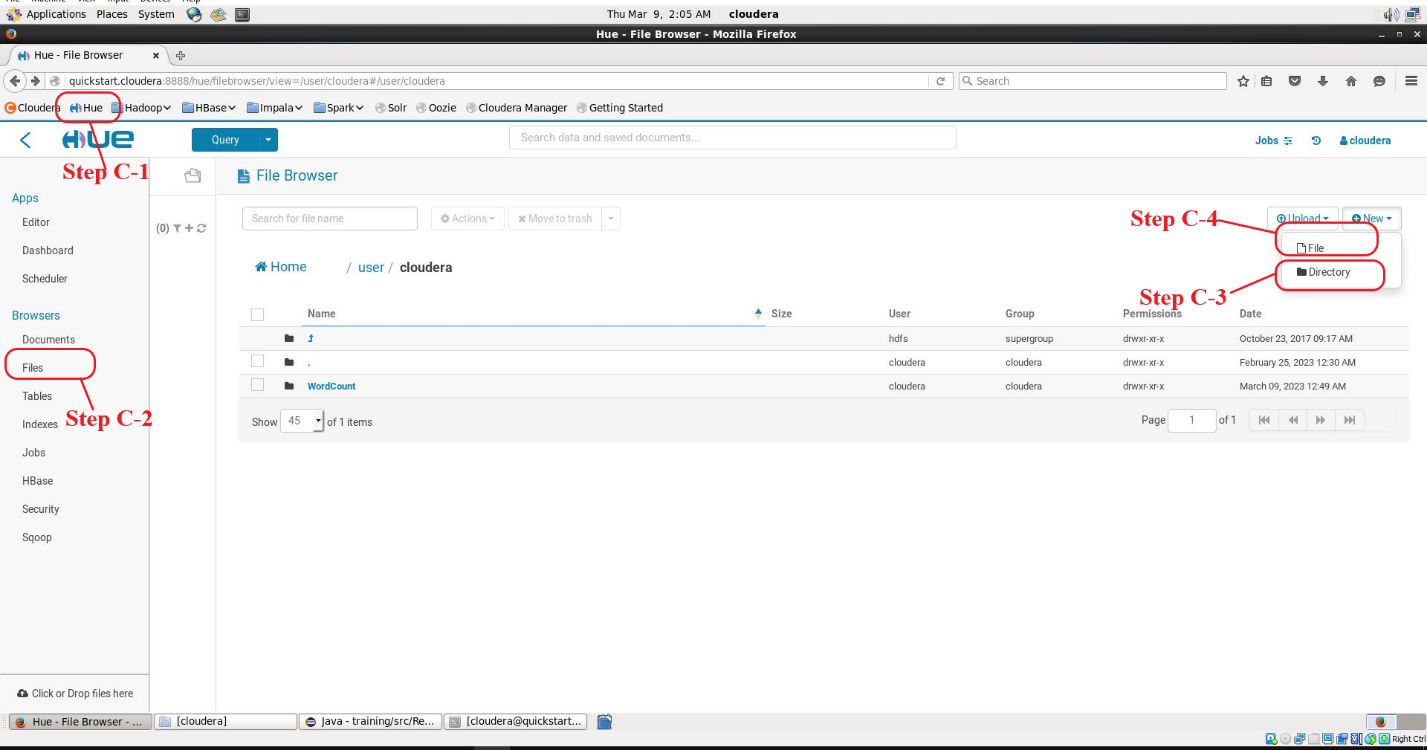
1. To create a jar file navigate through file🡪export. A window will pop-up. Select java🡪jar. Click next.
2. Make sure to check the resources to export where you need the jar file. Then create e destination and give file a name. (For instance WordCount.jar).

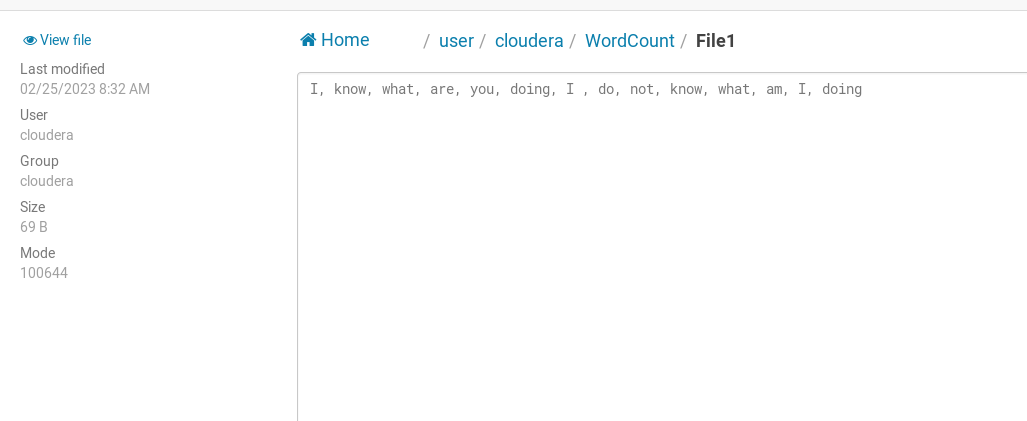


1. Click next🡪next🡪next🡪finish.

Step C: Creating Input file

1. Launch web browser (mozilla) on eclipse. Click on hue tab. [Note: You may ask the username and password. Both are ‘cloudera’]
2. Open navigation and open files tab.
3. Create a new directory. Name this according to you.
4. Under the directory create file
5. Open the file and edit it according to you. Make sure the file is comma separated. [Make sure to create multiple file]
6. Save the files.





Input file

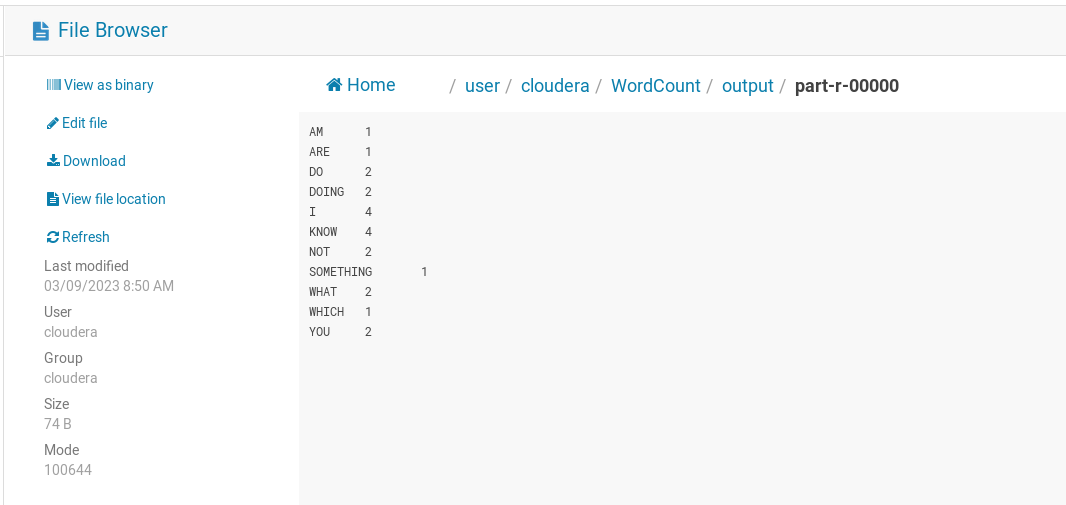
Step D: Executing the program through terminal

1. Open terminal from eclipse.
2. Execute this command line “Hadoop jar jarfilename.jar mainfilename inputfiledirectory outputfiledirectory/outputfilename”

For my programme I used this command “hadoop jar WordCount.jar WordCount /user/cloudera/WordCount /user/cloudera/WordCount/output”

1. Press enter. A output file should be generated in you declared output directory.





Output File