Lab 2

2 a. Code add to WordCount to remove output directory

FileSystem fs = FileSystem.*get*(getConf());

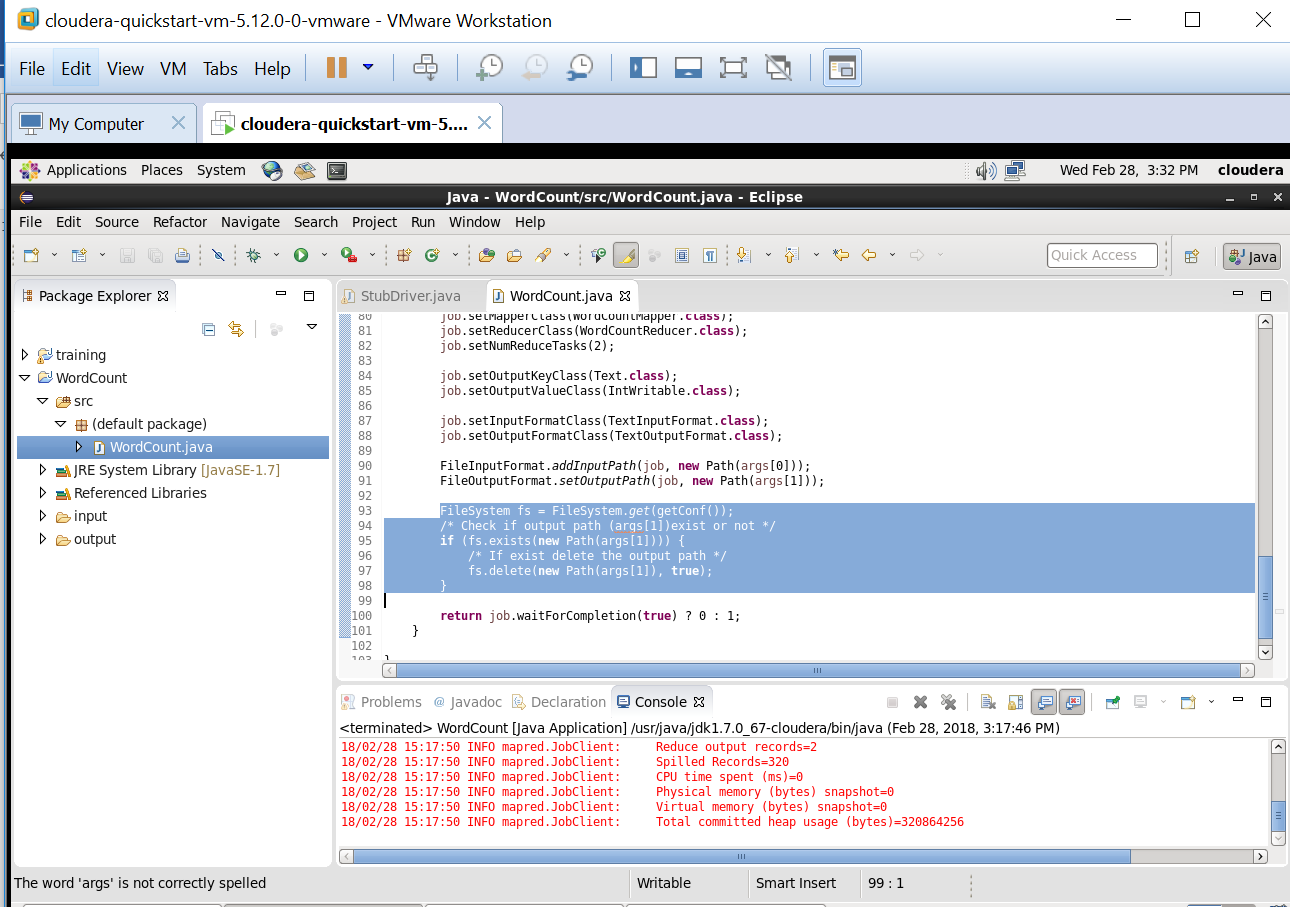
/\* Check if output path (args[1])exist or not \*/

**if** (fs.exists(**new** Path(args[1]))) {

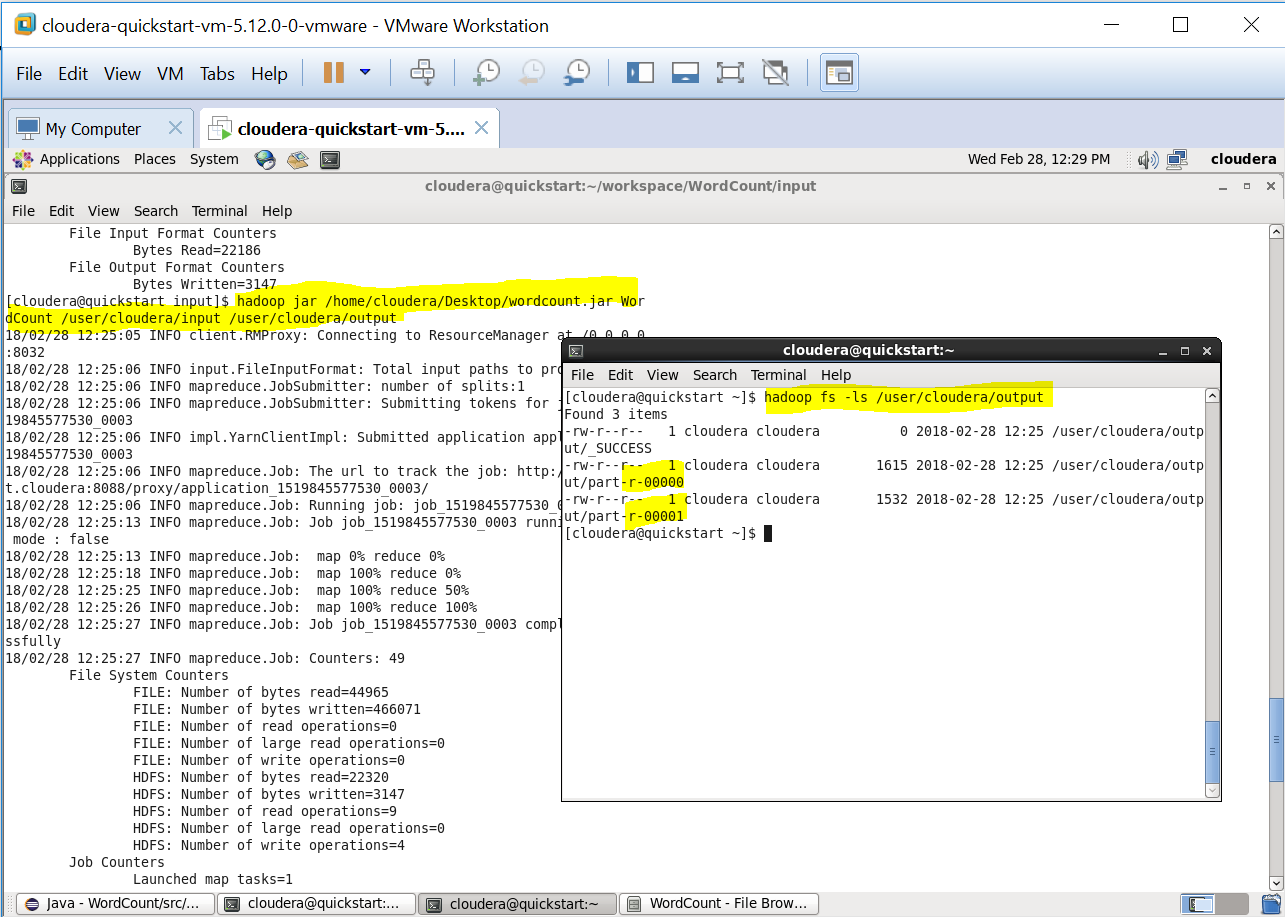
/\* If exist delete the output path \*/

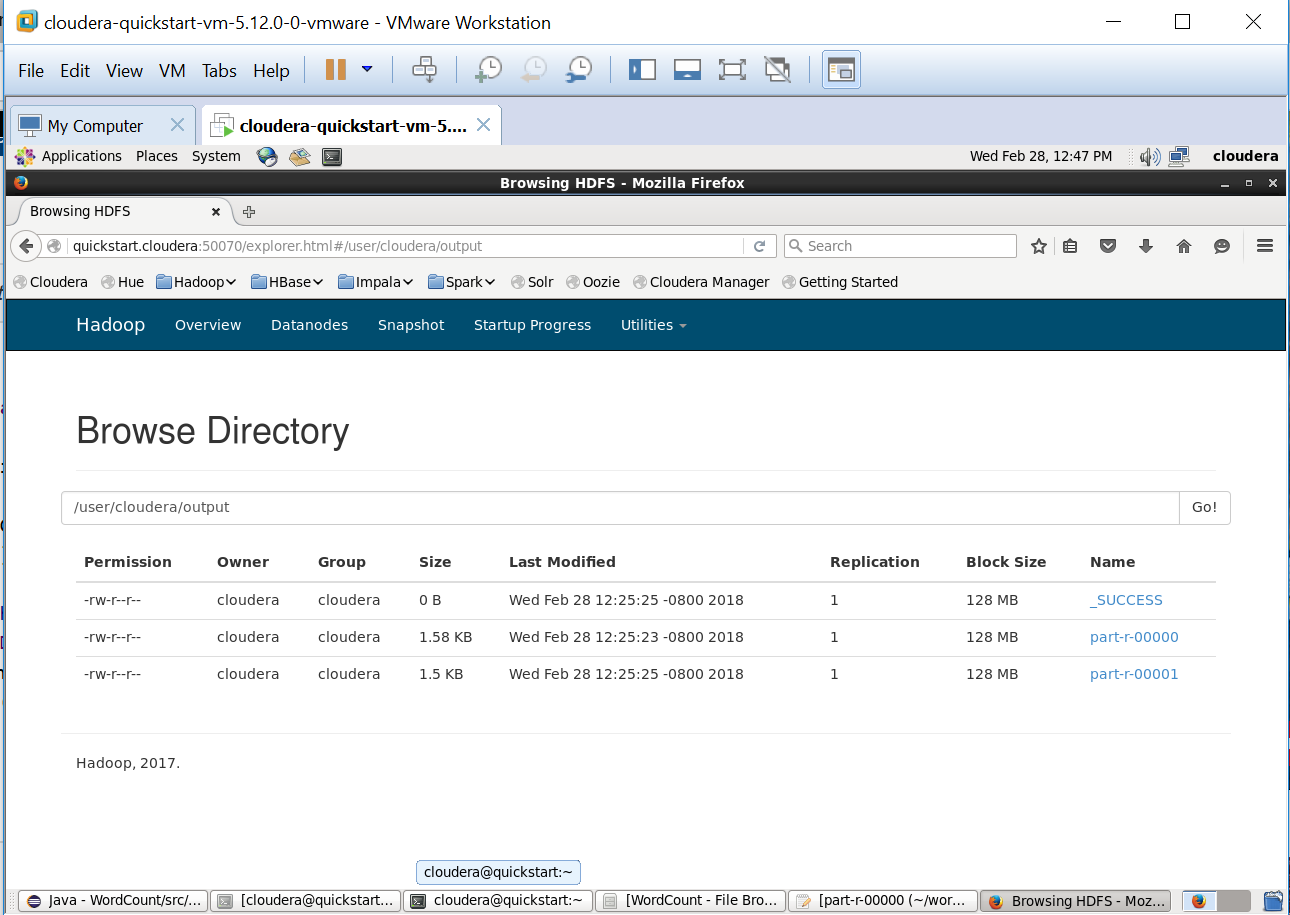
fs.delete(**new** Path(args[1]), **true**);

}

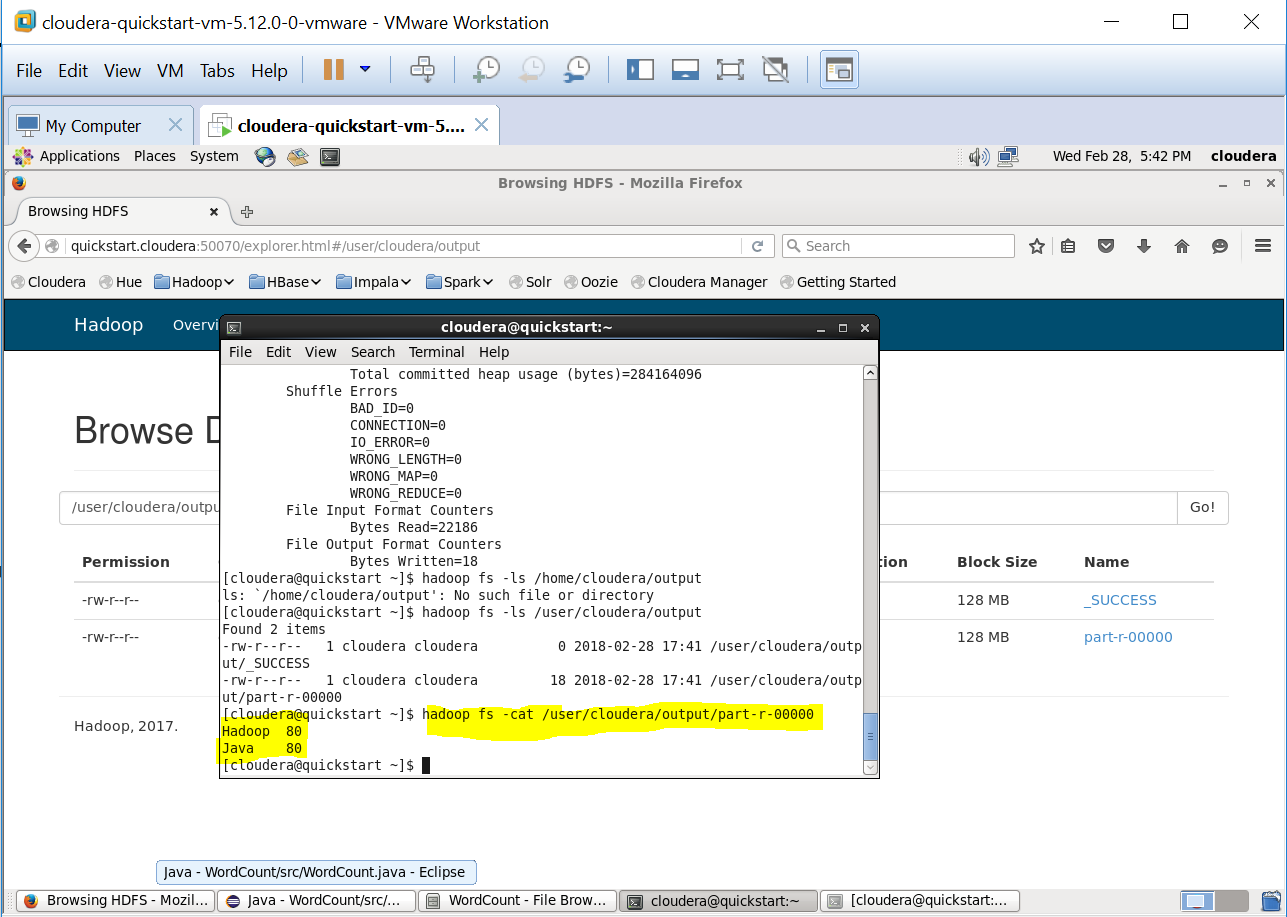


b. Word Count program in pseudo distributed mode with 2 reducers





c. Modify the WordCount program to output the counts of only the words “Hadoop” and “Java”. (Count Hadoop and hadoop as same word!)



d. Modify the WordCount program to output the counts of only those words which appear in the document at least 25 times

