

STUDENT NAME: LEE YU XIAN

STUDENT NUMBER: 7233164

1. Start off by creating a .java file with the necessary imports of java and HADOOP.
2. In the main function, define a new Configuration and Job class.
3. For this task, setup a MapperClass, CombinerClass and a ReducerClass.
4. Add 2 InputPaths and set them as the first and second args.
5. Add an OutputPath and set them as the third args.
6. In the mapper function, store and sort out both of the input's values in ascending order.
7. In the reducer function, begin comparing the values from the 2 lists.
8. If the value of the first list is more than or equals to the value of the second list, store the value.
9. If the value of the first list is less than the value of the second list, ignore it.
10. At the end of the function, set and write out the values which are stored from step 5.
11. Using the terminal, be sure to start HADOOP's namenode, datanode, resourcemanager, nodemanager and historyserver.
12. Use the command jps to make sure they are running.
13. Define an environment variable by using the export HADOOP_CLASSPATH command.
14. Use the echo command to make sure that HADOOP_CLASSPATH is set.
15. Compile the codes using the javac -cp command on the terminal.
16. Create the .jar file using the jar cvf command on the terminal.
17. Using HADOOP's jar command (\$HADOOP_HOME/bin/hadoop jar), run the .jar file created on step 16, with the first input being "measurement.txt", and second input being "car.txt", and third output being any other name.
18. Using HADOOP's fs -cat command (\$HADOOP_HOME/bin/hadoop fs -cat), display the output by locating the output file from step 17 and pathing to the "part-r*" file.