**MAP-REDUCE ON SAMPLE DATA**

1) Create the single node cluster :-

-using previously given document.Browse it using link :- http://localhost:50070/

2)Create the input directory and place the files to be analysed in it.

3) mapper-reduces code :-

$HADOOP\_HOME :-/home/arjun/hadoop-2.2.0

3.1) write a Mapper-Reducer code :-Given in the attachments.

3.2)Compile the code to check for errors

command:-javac -classpath $HADOOP\_HOME/share/hadoop/common/hadoop-common- 2.2.0.jar:$HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-client core2.2.0.jar:$HADOOP\_HOME/share/hadoop/common/lib/commons-cli-1.2.jar -d . wordcount.java

3.3)convert the output files into jar

command:- jar cf wc.jar wordcount\*.class

4)Putting data on datanodes :-

4.1) Make the HDFS directories required to execute MapReduce jobs:

command:- bin/hdfs dfs -mkdir /user

command:- bin/hdfs dfs -mkdir /user/<username>

4.2) copy file on the hdfs running:

command:-bin/hdfs dfs -put /home/arjun/hadoop-2.2.0/input /user/<username>

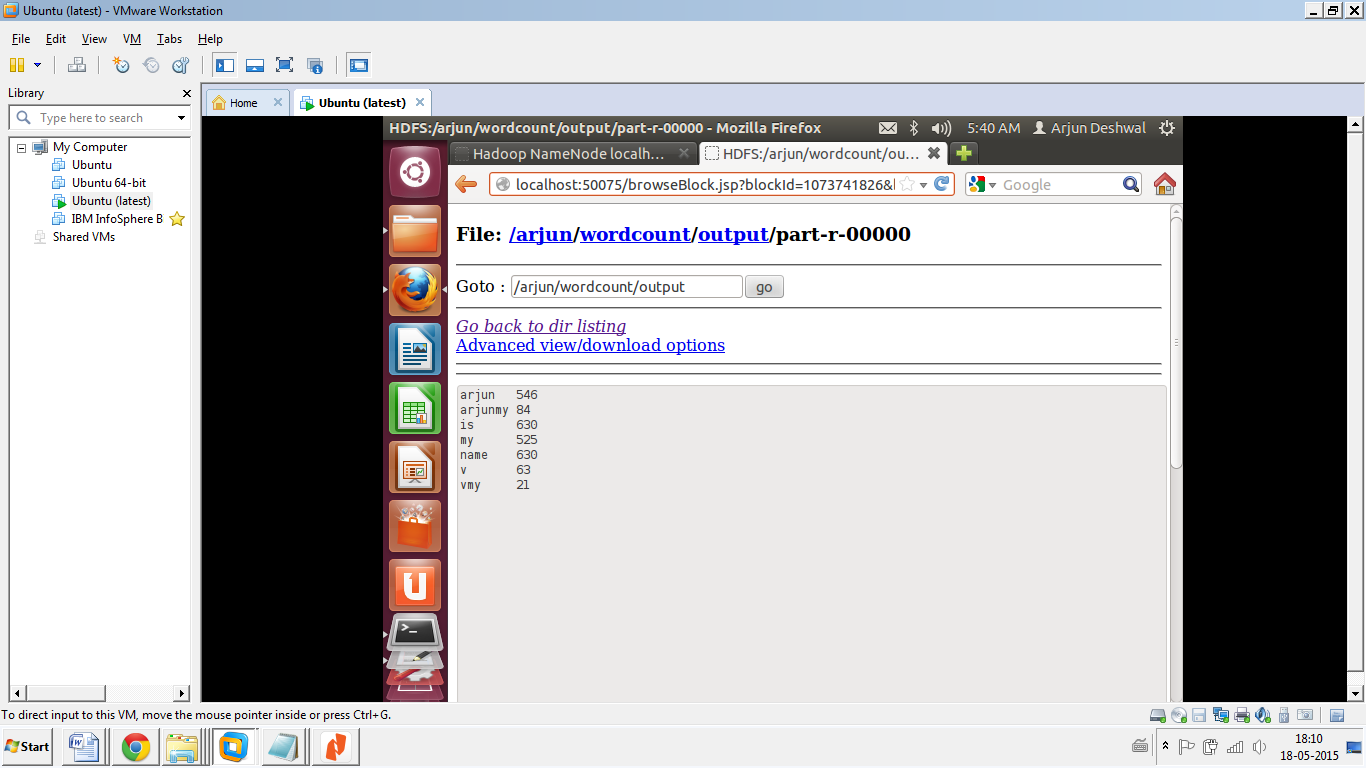
4.3) Run jar file on input

command:- bin/hadoop jar wc.jar wordcount /arjun/wordcount/input /arjun/wordcount/output

5)output

5.1)output can be seen on in the file kept on hdfs folder output

command:- bin/hdfs dfs -cat /arjun/wordcount/output/\*



OUTPUT