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>

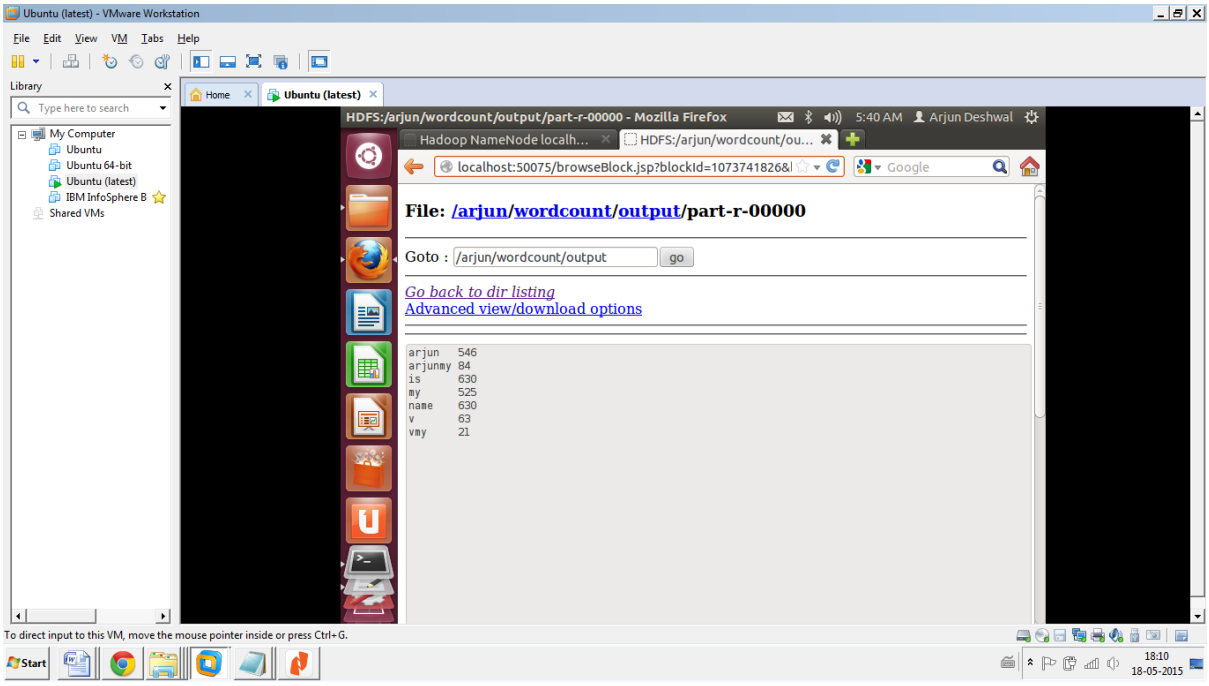
4.3) Run jar file on input

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/*


Ubuntu (latest) - VMware Workstation
File Edit View VM Tabs Help
Library
Type here to search
My Computer
Ubuntu
Ubuntu 64-bit
Ubuntu (latest)
IBM InfoSphere B
Shared VMs
Home
Ubuntu (latest)
HDFS:/arjun/wordcount/output/part-r-00000 - Mozilla Firefox
5:40 AM
Arjun Deshwal
Hadoop NameNode localh...
HDFS:/arjun/wordcount/ou...
localhost:50075/browseBlock.jsp?blockId=1073741826&l
Google
File: /arjun/wordcount/output/part-r-00000
Goto : /arjun/wordcount/output
go
Go back to dir listing
Advanced view/download options
arjun 546
arjunmy 84
is 630
my 525
name 630
v 63
vmy 21
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
Start
18:10
18-05-2015

OUTPUT