COMMANDS TO START HADOOP IN PSEUD0\_DISTRIBUTED MODE and EXECUTE MAP REDUCE PROGRAM

MapReduce(MR) Job is a Java Program that runs on the top of Hadoop Cluster. When executing for the first time, the process of executing a MR Job involves three steps described below. From second time onwards, you can directly execute the MR job using the third step.

1) Formatting the Hadoop File System(HDFS)

2) Starting the Services :

* Hadoop Services(NameNode,Data Node and Secondary Name Node) and
* MR Services(Job Tracker and Task Tracker)

3) Converting the .java file into jar file and executing on cluster.

Execute the following instructions in the CDH installed CentOS terminal.

I) **Formatting the Hadoop File System(HDFS):**

for x in `cd /etc/init.d ; ls hadoop-\*` ; do sudo service $x stop ; done

sudo -u hdfs hdfs namenode -format

sudo su

cd /var/lib/hadoop-hdfs/cache/hdfs/dfs/data

rm -rf \*

exit

**2) Starting the Services :**

for x in `cd /etc/init.d ; ls hadoop-hdfs-\*` ; do sudo service $x start ; done

(please check http://localhost:50070/dfshealth.html#tab-overview on browser and make sure live nodes

as 1)

sudo /usr/lib/hadoop/libexec/init-hdfs.sh

for x in `cd /etc/init.d ; ls hadoop-0.20-mapreduce-\*` ; do sudo service $x start ; done

sudo -u hdfs hadoop fs -mkdir -p /user/$USER

sudo -u hdfs hadoop fs -chown $USER /user/$USER

**3) Converting the .java file into jar file and executing on cluster.**

hadoop fs -mkdir $INPUTDIRECTORY (Replace $INPUTDIRECTORY with a relevant name but use the same

from here on for $INPUTDIRECTORY)

hadoop fs -put $INPUTFILES /user/$USER/$INPUTDIRECTORY (Here $INPUTFILES refers to the files on the local system that are needed to inserted into HDFS)

javac -classpath /usr/lib/hadoop/\*:/usr/lib/hadoop-0.20-mapreduce/\* -d . $FILENAME.java

($FILENAME should be same as your java file name)

jar -cvf $JAR\_FILE\_NAME.jar \*.class

hadoop jar $JAR\_FILE\_NAME.jar $CLASSNAME $INPUTDIRECTORY $OUTPUTDIRECTORYNAME (Input

Directory should exist before(Use mkdir() command as directed before ). Output Directory SHOULDNOT

EXIST BEFORE.)

hdfs dfs -ls -R /user/$USER/$OUTPUTDIRECTORYNAME

hdfs dfs -cat $OUTPUTDIRECTORYNAME/$OUTPUTNAME (output name typically will be in the form of

part-r-00000)