**\*-ASSIGNMENT -1**

data and creates several small chunks of data.

**Reduce stage :** This stage is the combination of the Shuffle stage and the Reduce stage. The Reducer’s job is to process the data that comes from the mapper. After processing, it produces a new set of output, which will be stored in the HDFS.

**Inserting Data into HDFS:**

•The MapReduce framework operates on <key, value> pairs, that is, the framework views the input to the job as a set of <key, value> pairs and produces a set of <key, value> pairs as the output of the job, conceivably of different types.

• The key and the value classes should be in serialized manner by the framework and hence, need to implement the Writable interface. Additionally, the key classes have to implement the Writable-Comparable interface to facilitate sorting by the framework.

•Input and Output types of a MapReduce job: (Input) <k1,v1> -> map -><k2, v2>-> reduce -><k3, v3> (Output).

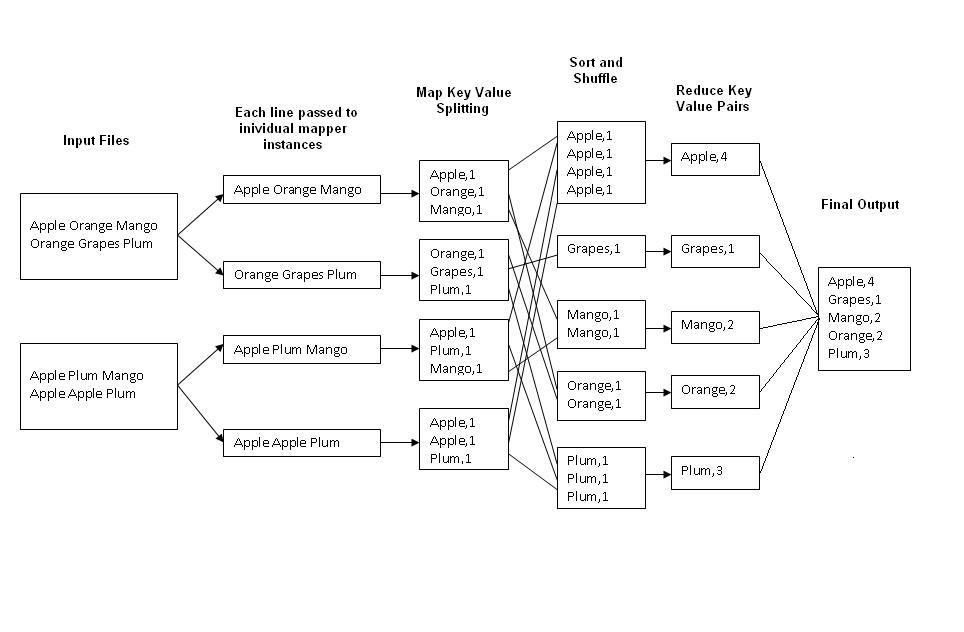


Fig.1 : An Example Program to Understand working of MapReduce Program.

**Steps for Compilation & Execution of Program:**

Su hadoopuser

#sudo mkdir analyzelogs

ls

#sudo chmod -R 777 analyzelogs/

cd

ls cd ..(to move to home directory)

pwd

ls

cd

pwd

#sudo chown -R hadoop1 analyzelogs/

cd

ls

#cd analyzelogs/

ls cd ..

# Copy the Files (Mapper.java,Reduce.java,Driver.java to Analyzelogs Folder)

# **#sudocp /home/Sonal/Desktop/assignment3/\* -/analyzelogs/**

(Convert access\_log\_short.txt into access\_log\_short.csv)

Start HADOOP

#start-dfs.sh

#start-yarn.sh

#jps

cd

cd analyzelogs

ls pwd

ls

#ls -ltr

#ls -al

#sudo chmod +r \*.\*

pwd

#export CLASSPATH="$HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduceclient-core-2.9.0.jar:$HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-clientcommon-2.9.0.jar:$HADOOP\_HOME/share/hadoop/common/hadoop-common-2.9.0.jar**:~/analyzelogs/SalesCountry**/\*:$HADOOP\_HOME/lib/\*"

(This should be PWD)

# Compile Java Files

# javac -d . SalesMapper.java SalesCountryReducer.java SalesCountryDriver.java ls

#cd SalesCountry/

ls cd ..

#sudo gedit Manifest.txt

Main-class:SalesCountry.SalesCountryDriver

(Press enter)

#jar -cfm analyzelogs.jar Manifest.txt SalesCountry/\*.class ls cd

#cd analyzelogs/

# Create Directory on Hadoop

#sudo mkdir ~/input2000

ls

pwd

#sudo cp access\_log\_short.csv ~/input2000/

# $HADOOP\_HOME/bin/hdfs dfs -put ~/input2000 /

# $HADOOP\_HOME/bin/hadoop jar analyzelogs.jar /input2000 /output2000

# $HADOOP\_HOME/bin/hdfs dfs -cat /output2000/part-00000

# stop-all.sh

# jps

For GUI

Go to browser(localhost:50070)

Go to utilities(browse directory)