**9.Aim: Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm.**

Now that we’ve seen how the FileSystem (fs) shell can be used to execute Hadoop commands to interact with HDFS, the same fs shell can be used to launch MapReduce jobs. In this section, we will walk through the steps required to run a MapReduce program. The source code for a MapReduce program is contained in a compiled .jar file. Hadoop will load the JAR into HDFS and distribute it to the data nodes, where the individual tasks of the MapReduce job will be executed. Hadoop ships with some example MapReduce programs to run. One of these is a distributed WordCount program which reads text files and counts how often words occur.

**Running the WordCount program**

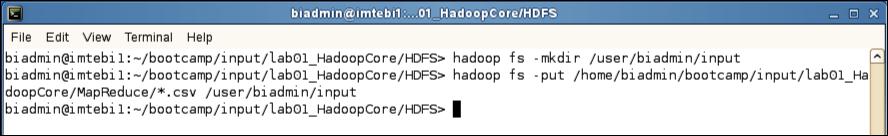
First we need to copy the data files from the local file system to HDFS.

Step 1:Execute the commands below to copy the input files into HDFS.



**hadoop fs -mkdir /user/biadmin/input**

**hadoop fs -put /home/biadmin/bootcamp/input/lab01\_HadoopCore/MapReduce/\*.csv /user/biadmin/input**

****

**Copy input files into HDFS**

Step 2: Review the files have been copied with the following command:



**hadoop fs -ls input**

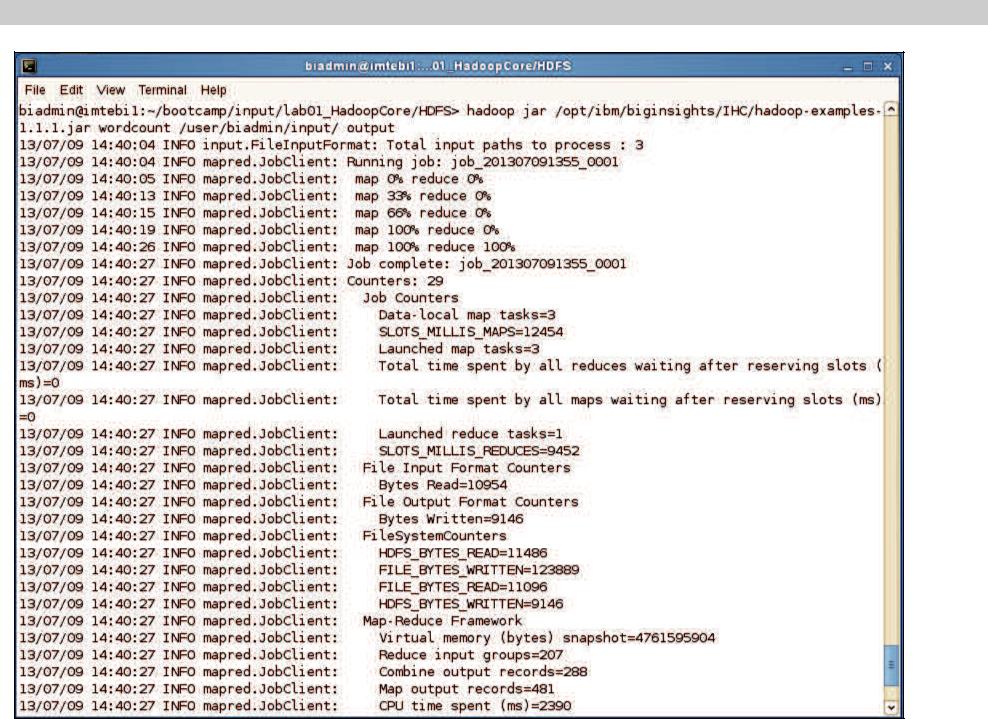
****

**List copied files into HDFS**

Step 3: Now we can run the wordcount job with the command below, where “/user/biadmin/input/” is where the input files are, and “output” is the directory where the output of the job will be stored. The “output” directory will be created automatically when executing the command below.

****

**hadoop jar /opt/ibm/biginsights/IHC/hadoop-examples-1.1.1.jar wordcount /user/biadmin/input/ output**

****

**WordCount MapReduce job running**

Step 4: Now review the output of step 3:

In this case, the output was not split into multiple files.



**hadoop fs -ls output**

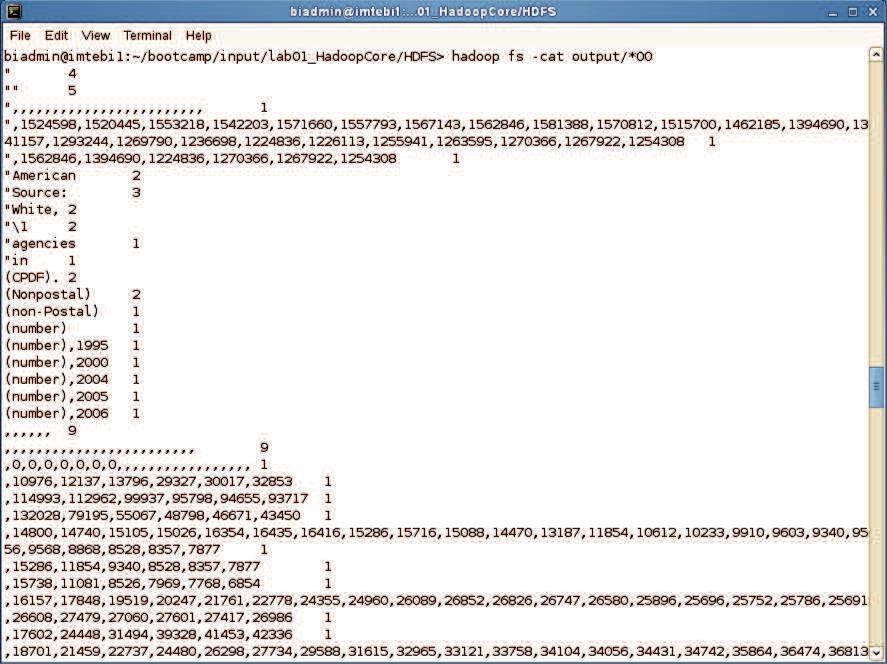
****

**MapReduce result files**

Step 5: To view the contents of the part-r-0000 file issue the command below:



**hadoop fs -cat output/\*00**

****

**MapReduce output**