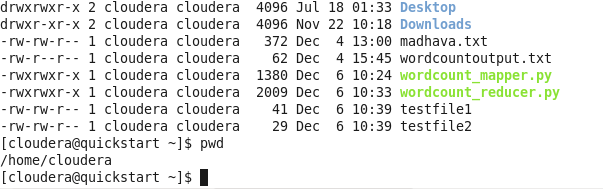
1. Created wordcount\_mapper.py and wordcount\_reducer.py files in local file system using gedit and change permissions as shown below.



1. Created testfile1 and testfile2 in local file system as shown in the above picture.
2. Created the directory called “input” under HDFS File system using the below command.

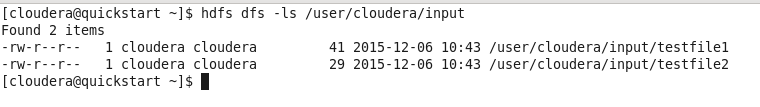
hdfs dfs -mkdir /user/cloudera/input

1. Copied the testfile1 and testfile2 files from local file system to HDFS file system using the below command.

hdfs dfs -put /home/cloudera/testfile1 /user/cloudera/input

hdfs dfs -put /home/cloudera/testfile2 /user/cloudera/input

1. Please see the files copied to HDFS file system as shown below.



1. Ran the below command

hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming.jar \

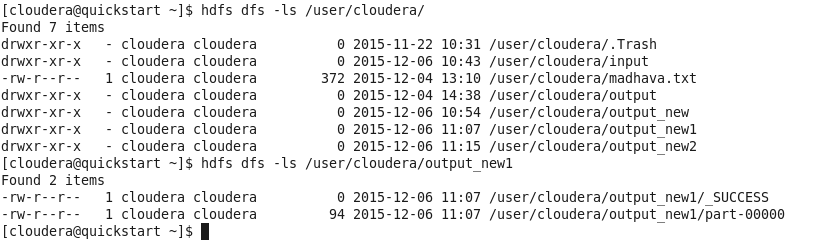
-input /user/cloudera/input \

-output /user/cloudera/output\_new1 \

-mapper /home/cloudera/wordcount\_mapper.py \

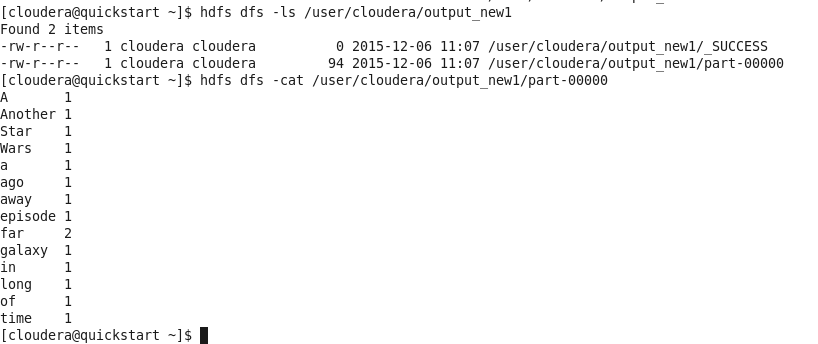
-reducer /home/cloudera/wordcount\_reducer.py

1. After running, the output\_new1 folder as shown below.



1. By using the below command we can see the result

hdfs dfs -cat /user/cloudera/output\_new1/part-r-00000



1. Now, running the mapper and reducer program again by setting number of reduce tasks as shown below.

hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming.jar \

-input /user/cloudera/input \

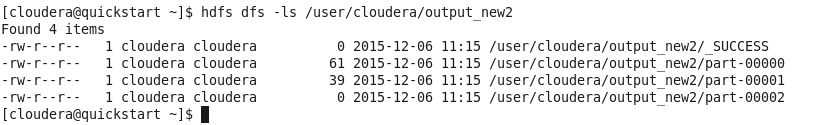
-output /user/cloudera/output\_new2 \

-mapper /home/cloudera/wordcount\_mapper.py \

-reducer /home/cloudera/wordcount\_reducer.py \

-numReduceTasks 0

1. Now, the output will be in three files.



1. Here is content of each output file.



