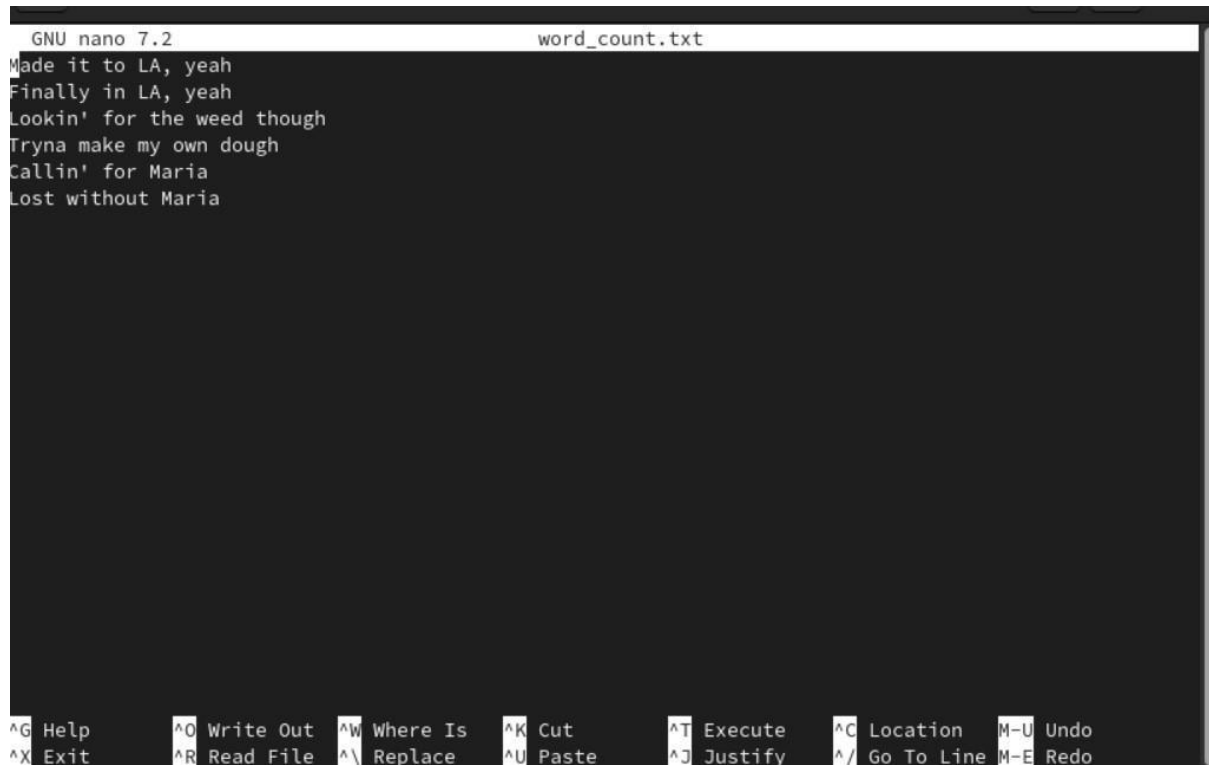


**EXP NO: 2****RUN A BASIC WORD COUNT MAP REDUCE PROGRAM TO  
UNDERSTAND MAP REDUCE PARADIGM****\$nano word\_count.txt**

```
GNU nano 7.2 word_count.txt
Made it to LA, yeah
Finally in LA, yeah
Lookin' for the weed though
Tryna make my own dough
Callin' for Maria
Lost without Maria

^G Help      ^O Write Out  ^W Where Is   ^K Cut        ^T Execute    ^C Location   M-U Undo
^X Exit      ^R Read File  ^_ Replace    ^U Paste      ^J Justify    ^_ Go To Line M-E Redo
```

**\$nano mapper.py****\$nano reducer.py****\$start-all.sh****\$hdfs dfs -mkdir /exp2****\$hdfs dfs -copyFromLocal ~/DA-Lab/exp2/word\_count.txt /exp2****\$chmod 777 mapper.py reducer.py****\$hadoop jar \$HADOOP\_STREAMING -input /exp2/word\_count.txt -output /exp2/output  
mapper ~/DA-Lab/exp2/mapper.py -reducer ~/DA-Lab/exp2/reducer.py**

```

Sep 13 10:53
hadoop@ubuntu: ~
at org.apache.hadoop.mapred.Mapper.run(Mapper.java:466)
at org.apache.hadoop.mapred.Mapper.run(Mapper.java:359)
at org.apache.hadoop.mapred.LocalJobRunner$Job$MapTaskRunnable.run(LocalJobRunner.java:271)
at java.util.concurrent.Executors$RunnableAdapter.call(Executors.java:511)
at java.util.concurrent.FutureTask.run(FutureTask.java:266)
at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149)
at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)
at java.lang.Thread.run(Thread.java:750)
2024-09-13 10:45:21,095 INFO mapreduce.Job: Job job_local1681841077_0001 running in uber mode : false
2024-09-13 10:45:21,096 INFO mapreduce.Job: map 0% reduce 0%
2024-09-13 10:45:21,011 INFO mapreduce.Job: Job job_local1681841077_0001 failed with state FAILED due to: NA
2024-09-13 10:45:21,026 INFO mapreduce.Job: Counters: 0
2024-09-13 10:45:21,026 ERROR streaming.StreamJob: Job not successful!
Streaming Command Failed!
hadoop@ubuntu: ~$ pwd
/home/hadoop
hadoop@ubuntu: ~$ nano mapper.py
hadoop@ubuntu: ~$ nano reducer.py
hadoop@ubuntu: ~$ chmod +x ~/mapper.py
hadoop@ubuntu: ~$ chmod +x ~/reducer.py
hadoop@ubuntu: ~$ hadoop jar /home/hadoop/hadoop/share/hadoop/tools/lib/hadoop-streaming-3.3.6.jar \
-input /word_count_in_python/word_count.txt \
-output /word_count_in_python/new_output \
-mapper ~/mapper.py \
-reducer ~/reducer.py
2024-09-13 10:49:38,633 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-09-13 10:49:38,717 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-09-13 10:49:38,730 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-09-13 10:49:38,730 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2024-09-13 10:49:38,895 ERROR streaming.StreamJob: Error Launching job : Output directory hdfs://localhost:9800/word_count_in_python/new_output already exists
Streaming Command Failed!
hadoop@ubuntu: ~$ hdfs dfs -rm -r /word_count_in_python/new_output
Deleted /word_count_in_python/new_output
hadoop@ubuntu: ~$ hadoop jar /home/hadoop/hadoop/share/hadoop/tools/lib/hadoop-streaming-3.3.6.jar \
-input /word_count_in_python/word_count.txt \
-output /word_count_in_python/new_output \
-mapper ~/mapper.py \
-reducer ~/reducer.py
2024-09-13 10:50:30,227 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-09-13 10:50:31,018 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-09-13 10:50:31,018 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-09-13 10:50:31,032 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2024-09-13 10:50:31,256 INFO mapred.FileInputFormat: Total input files to process : 1
2024-09-13 10:50:31,282 INFO mapreduce.JobSubmitter: number of splits:1
2024-09-13 10:50:31,385 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1258122800_0001
2024-09-13 10:50:31,385 INFO mapreduce.JobSubmitter: Execution with tokens: 1

```

\$hdfs dfs -cat /exp2/output/\*

```

Sep 13 10:53
hadoop@ubuntu: ~
Merged Map outputs=1
GC time elapsed (ms)=8
Total committed heap usage (bytes)=522190848
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=149
File Output Format Counters
Bytes Written=177
2024-09-13 10:50:33,526 INFO streaming.StreamJob: Output directory: /word_count_in_python/new_output
hadoop@ubuntu: ~$ hdfs dfs -ls /word_count_in_python/new_output
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2024-09-13 10:50 /word_count_in_python/new_output/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 177 2024-09-13 10:50 /word_count_in_python/new_output/part-00000
hadoop@ubuntu: ~$ hdfs dfs -cat /word_count_in_python/new_output/part-00000
Callin 1
Finally 1
LA 2
Lookin 1
Lost 1
Made 1
Marta 1
Martia 1
Night 1
dive 1
dough 1
for 2
in 2
it 1
make 1
marino 1
my 1
own 1
the 2
though 1
tryns 1
weed 1
without 1
yeah 2
hadoop@ubuntu: ~$

```