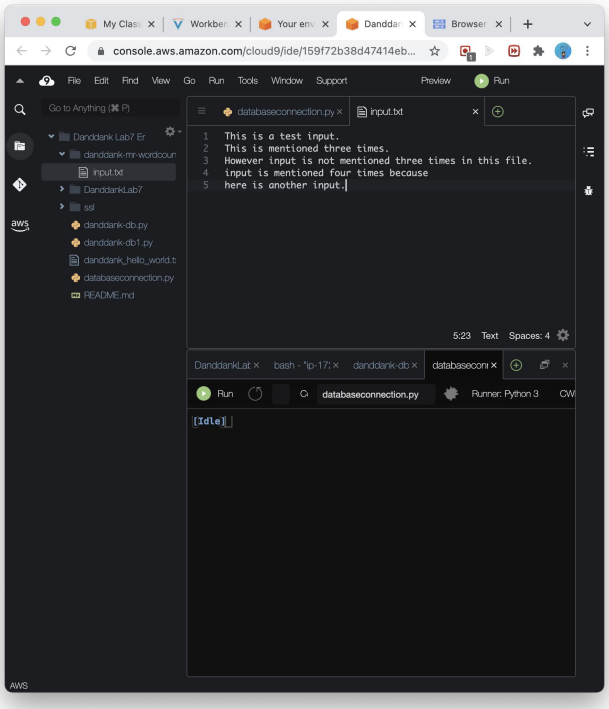
Nalongsone Danddank Student ID : 14958950 StarID: jf3893pd
Email: nalongsone.danddank@my.metrostate.edu\

ICS 432 - 01 — Distributed and Cloud Computing Fall 2021

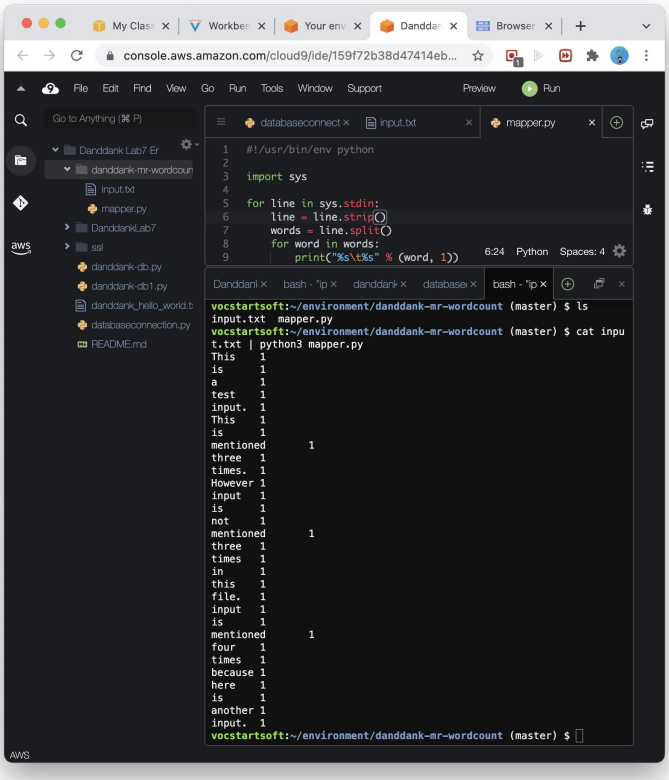
Lab #9 : Running a map-reduce job on a Hadoop cluster

Exercise 1: Running map-reduce jobs on the cloud

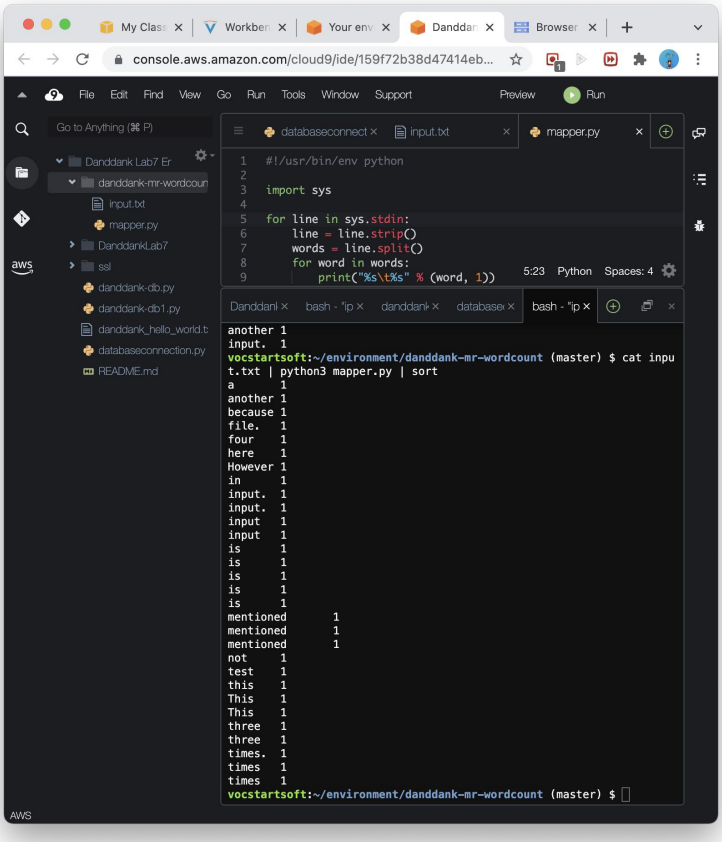
Lab report screen-shot #1:



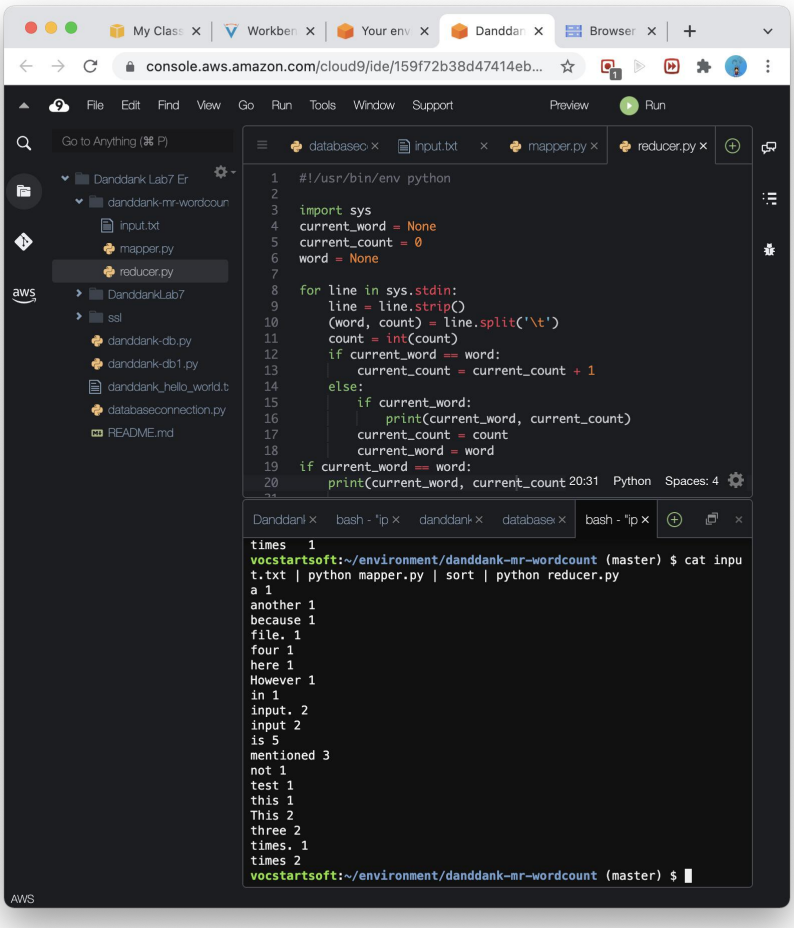
Lab report screen-shot #2:



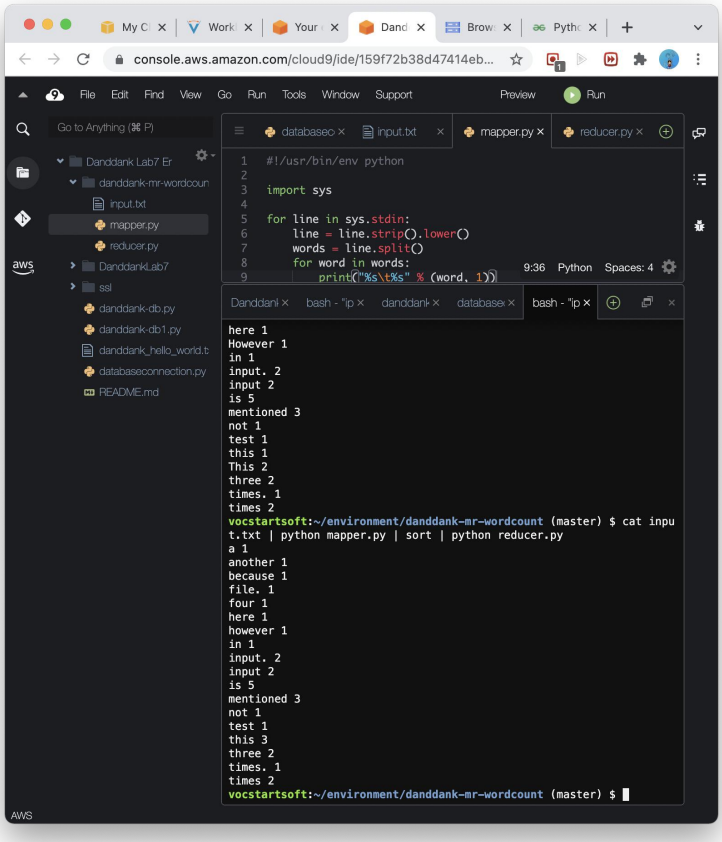
Lab report screen-shot #3:



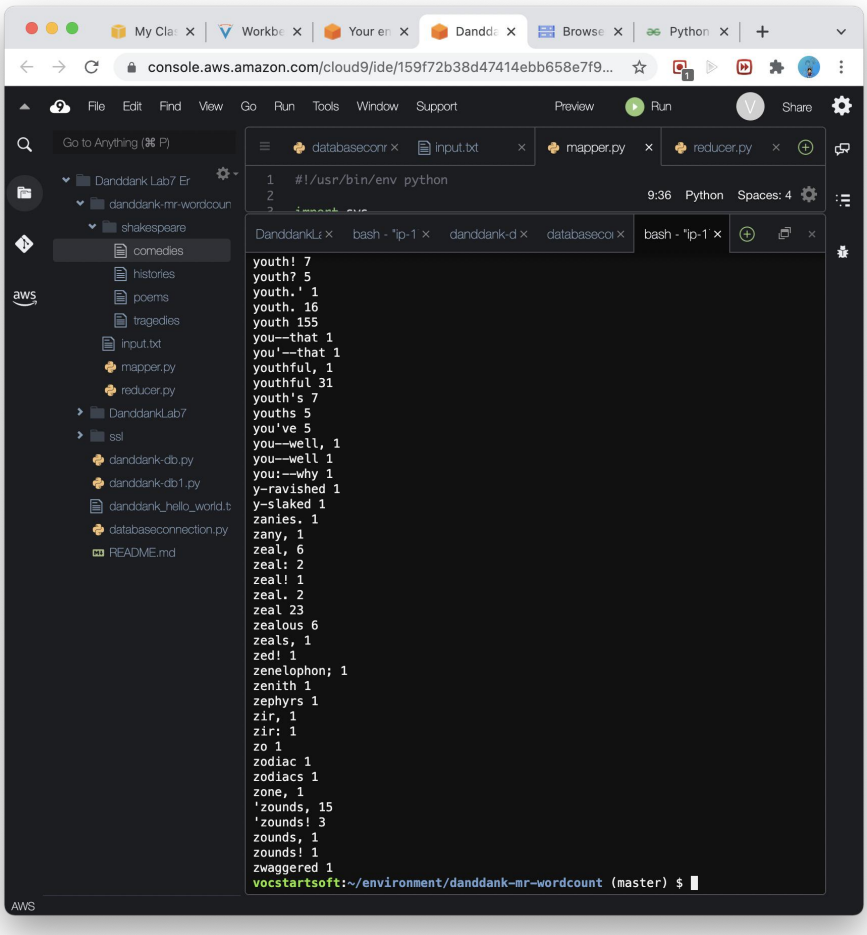
Lab report screen-shot #4:



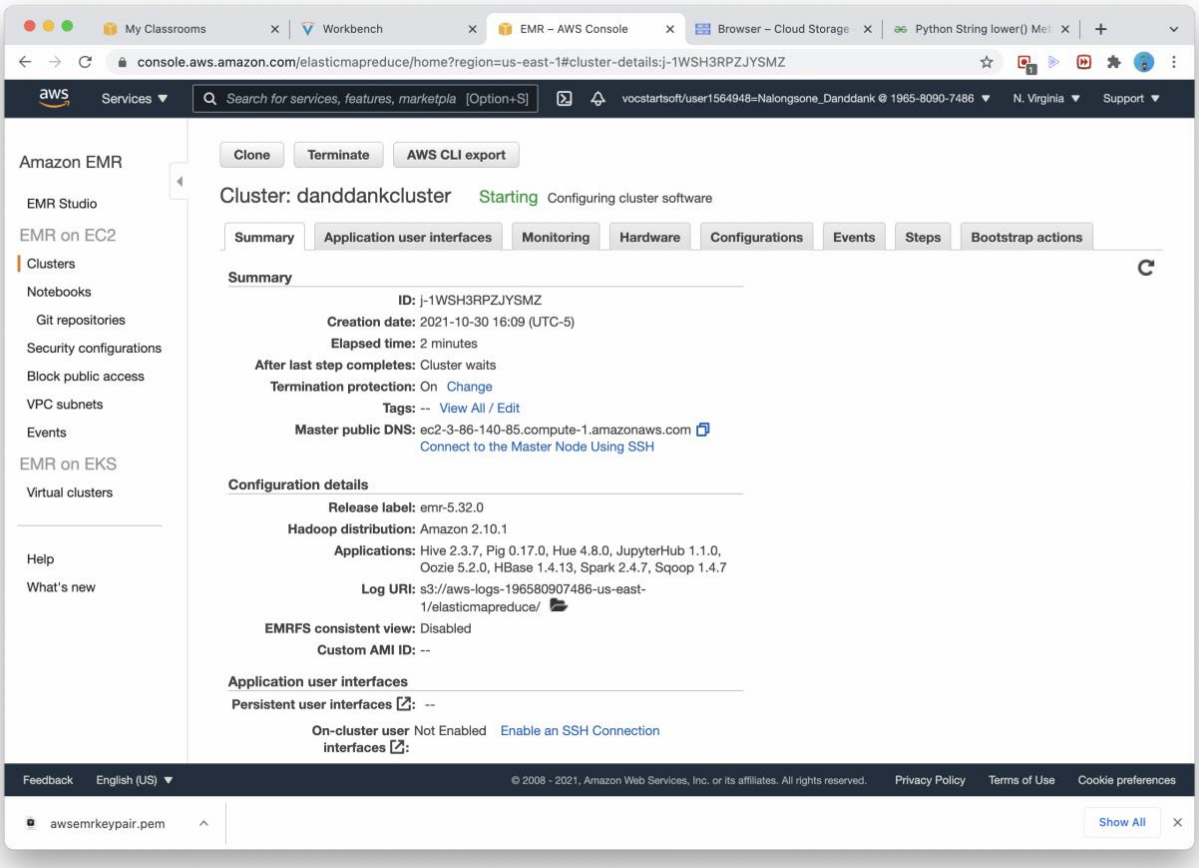
Lab report screen-shot #5:



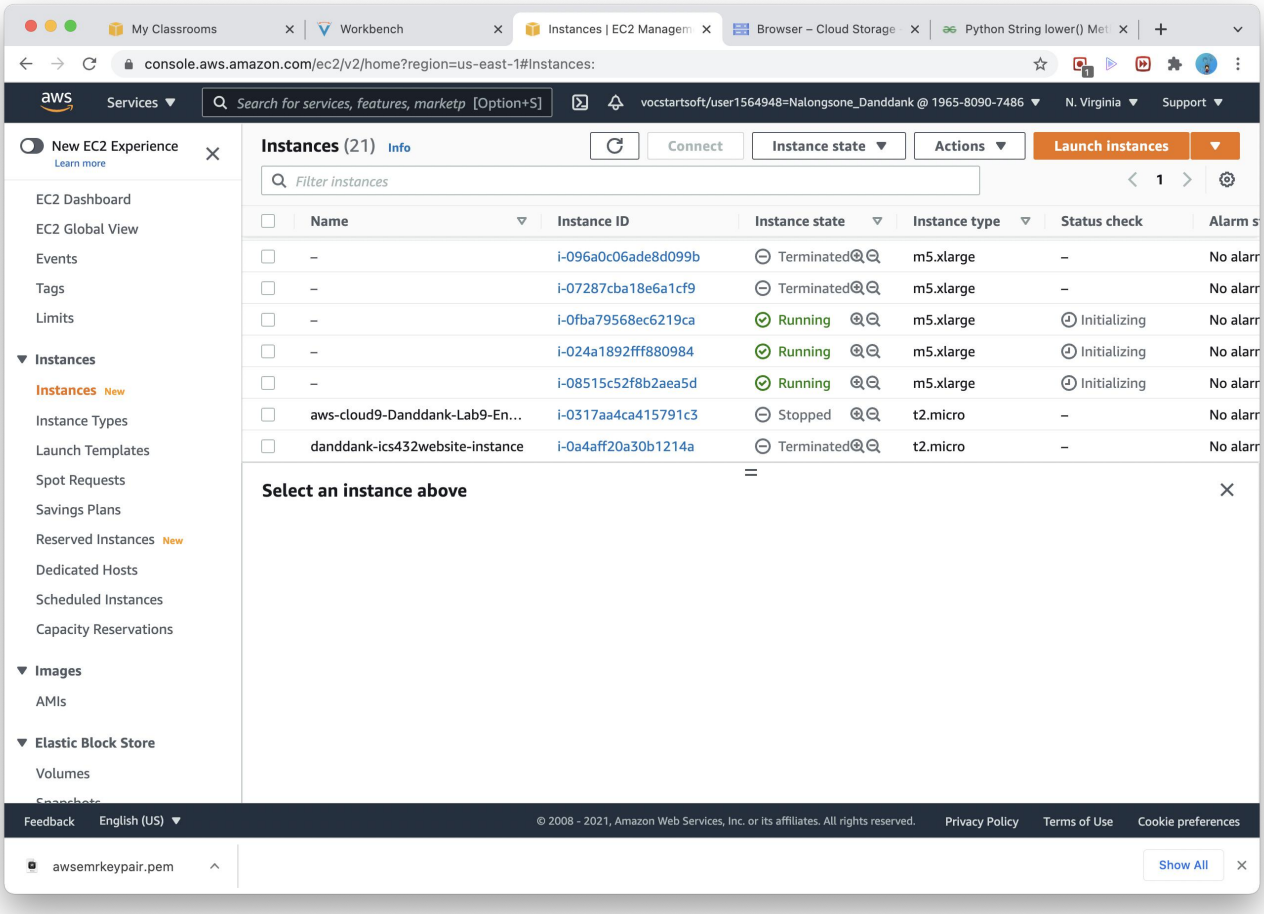
Lab report screen-shot #6:



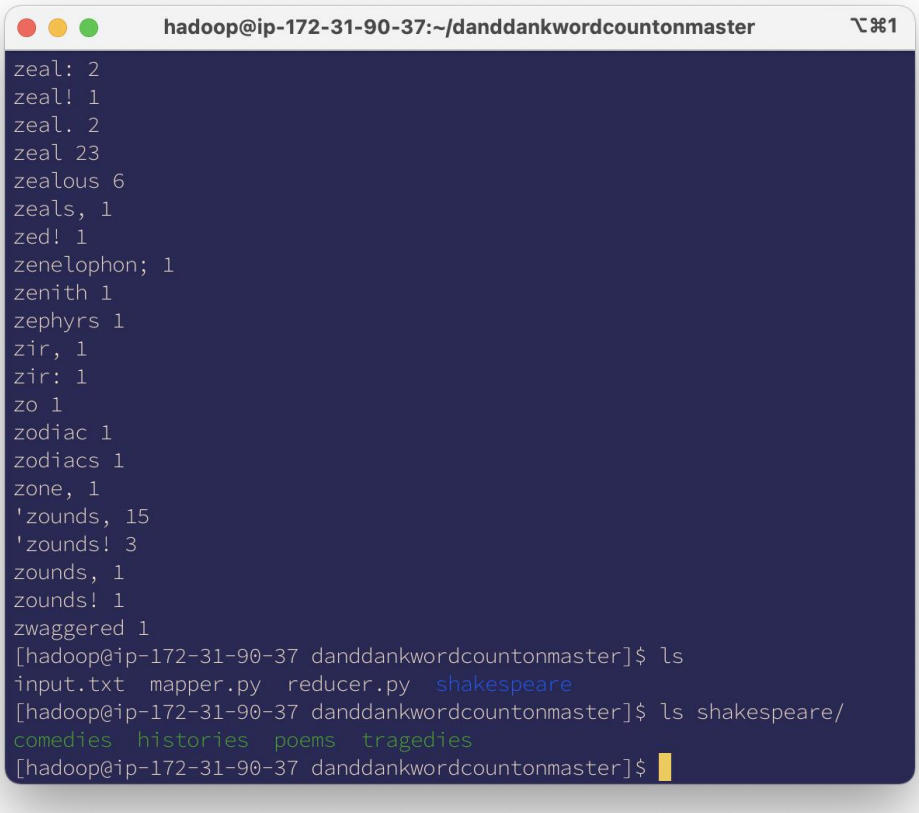
Lab report screen-shot #7:



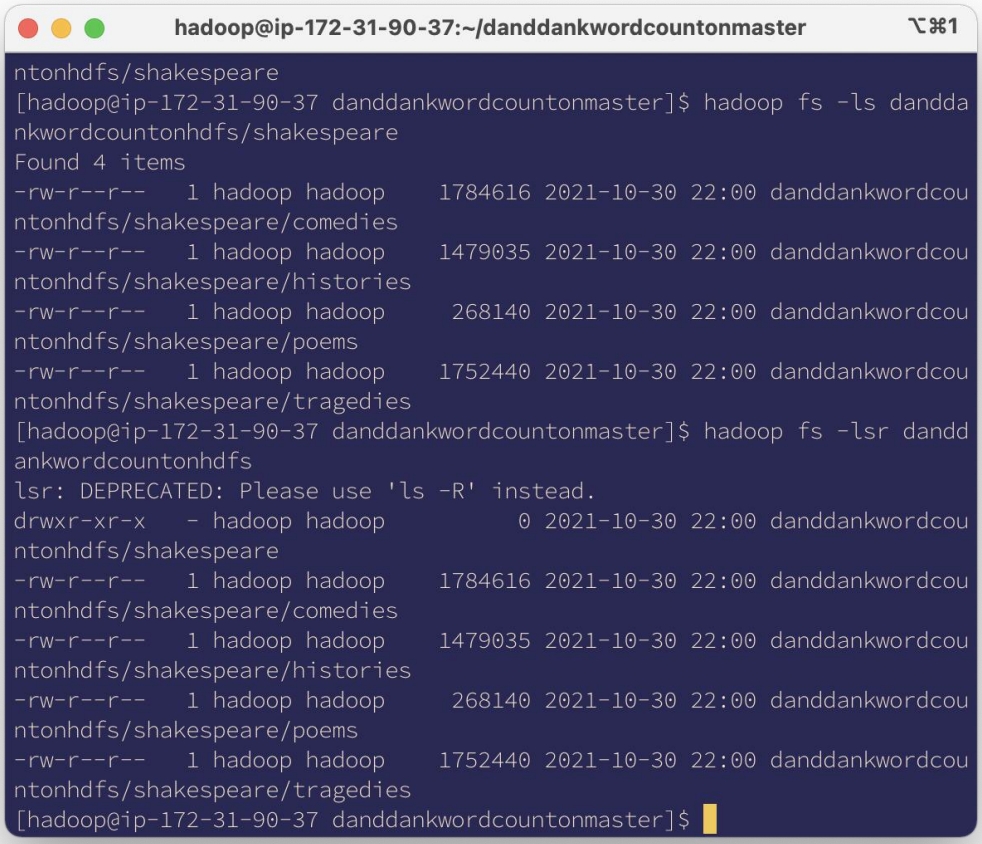
Lab report screen-shot #8:



Lab report screen-shot #9:



Lab report screen-shot #10:



Lab report screen-shot #11:

a. hadoop fs -mkdir danddankFolder

b. hadoop fs -put input.txt danddankFolder/input.txt

c. hadoop fs -ls danddankFolder

d. hadoop fs -cat danddankFolder/input.txt

e. hadoop fs -mv danddankFolder/input.txt danddankFolder/newInput.txt

f. hadoop fs -mkdir nalongsoneFolder

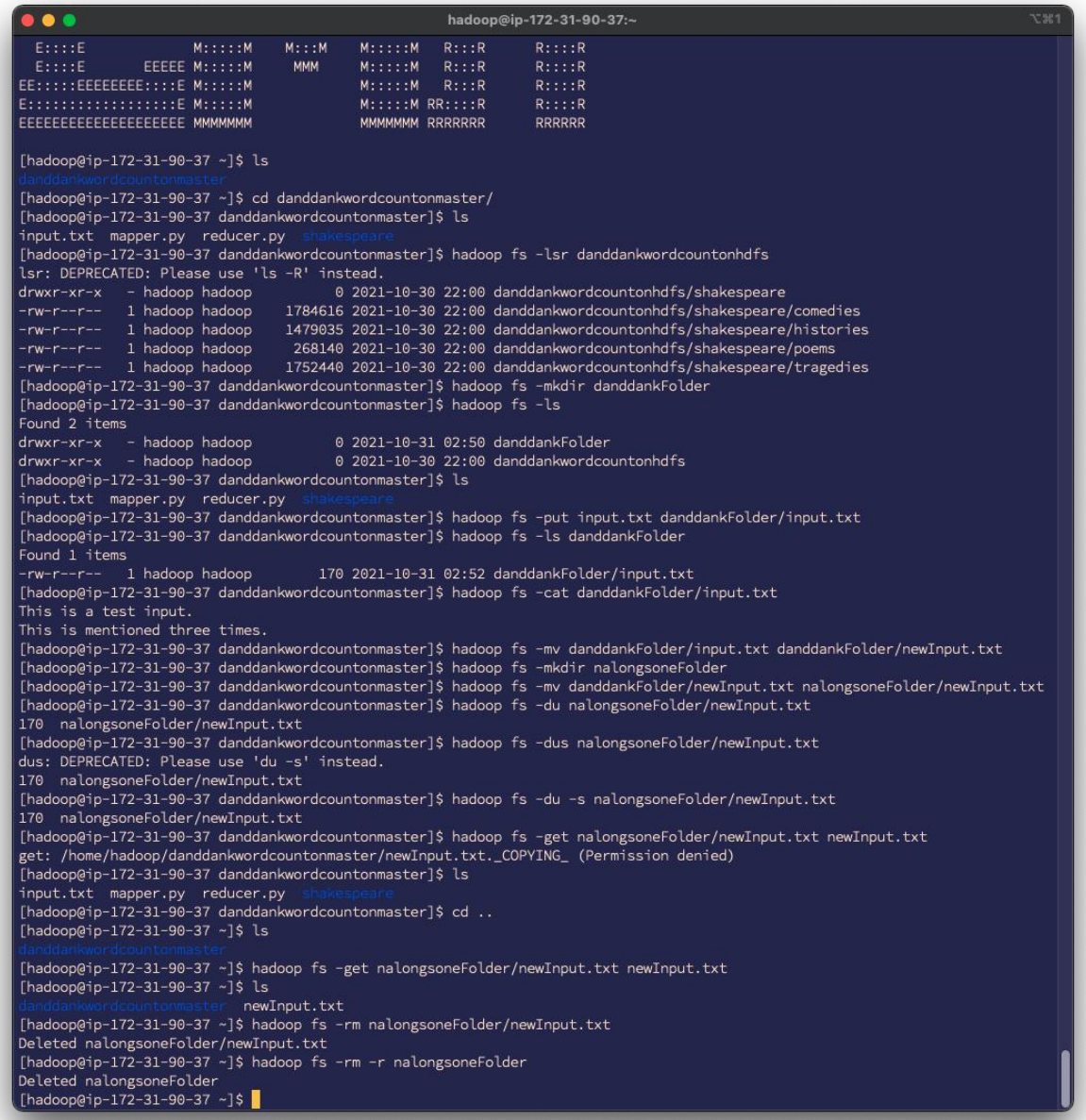
Had oop fs -mv danddankFolder/newInput.txt nalongsoneFolder/newInput.txt

g. hadoop fs -du nalongsoneFolder/newInput.txt

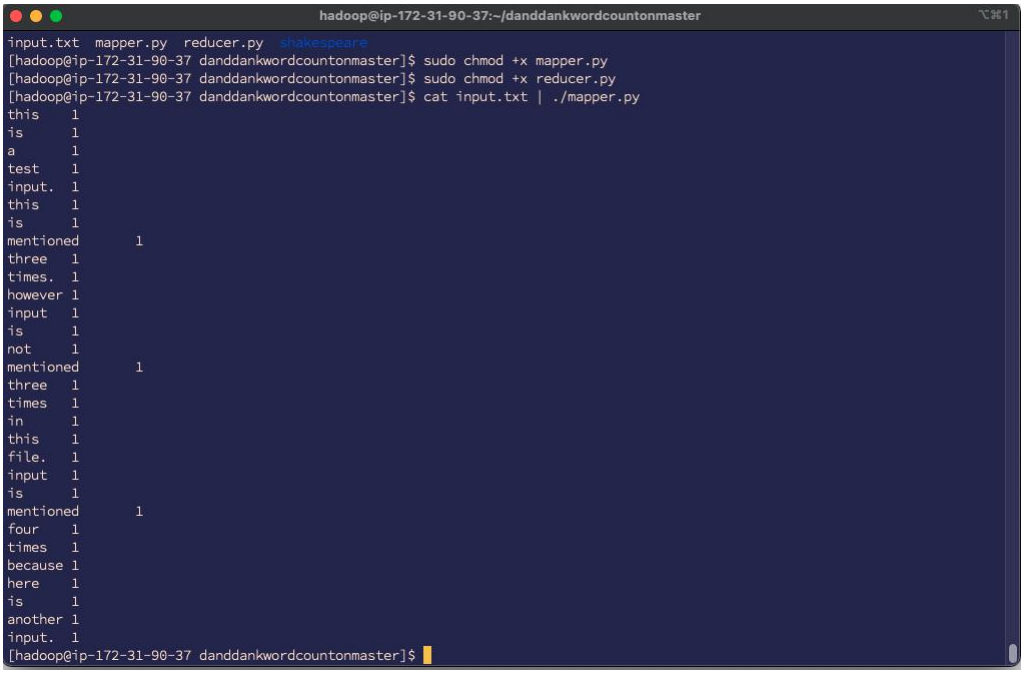
h. hadoop fs -get nalongsoneFolder/newInput.txt newInput.txt

i. hadoop fs -rm nalongsoneFolder/newInput.txt

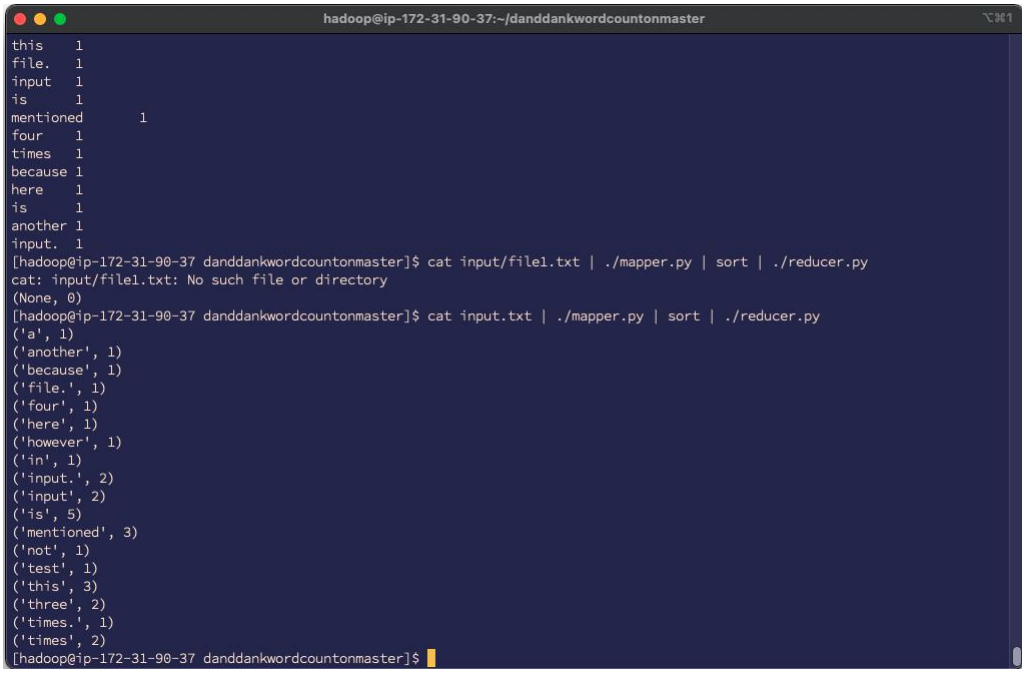
j. hadoop fs -rm -r nalongsoneFolder



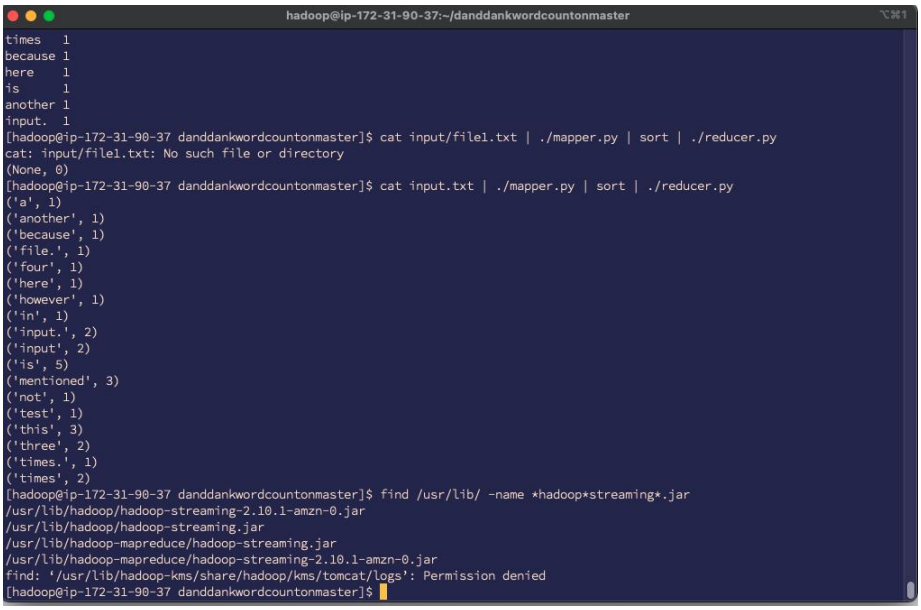
Lab report screen-shot #12:



Lab report screen-shot #13:

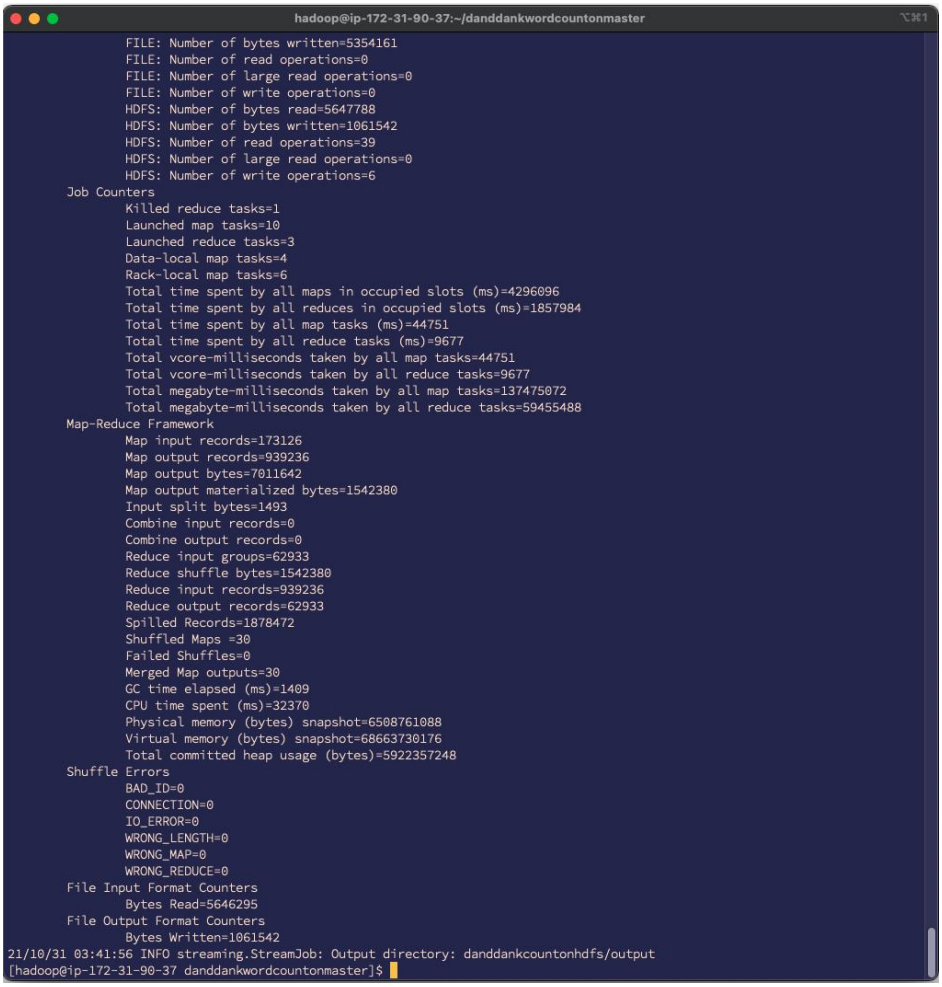


Lab report screen-shot #14:

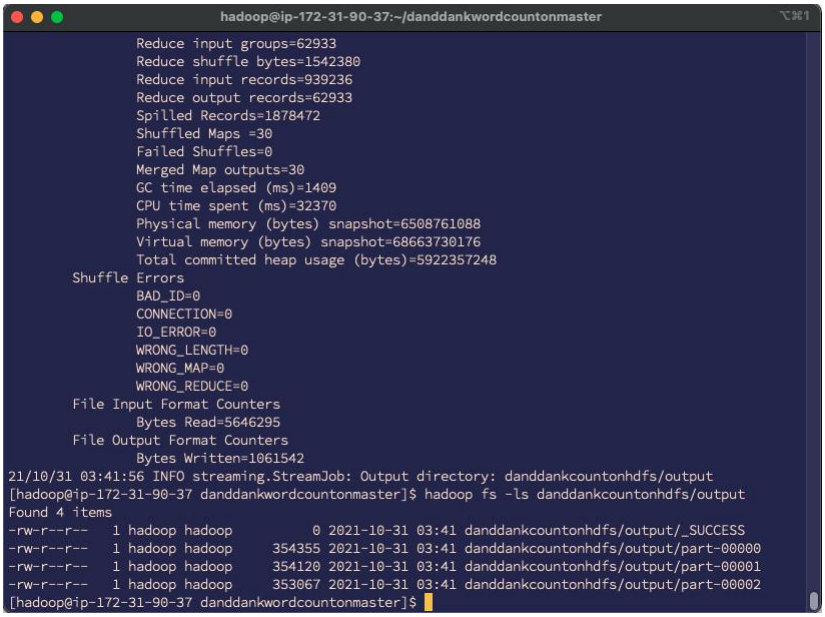


Lab report screen-shot #15:

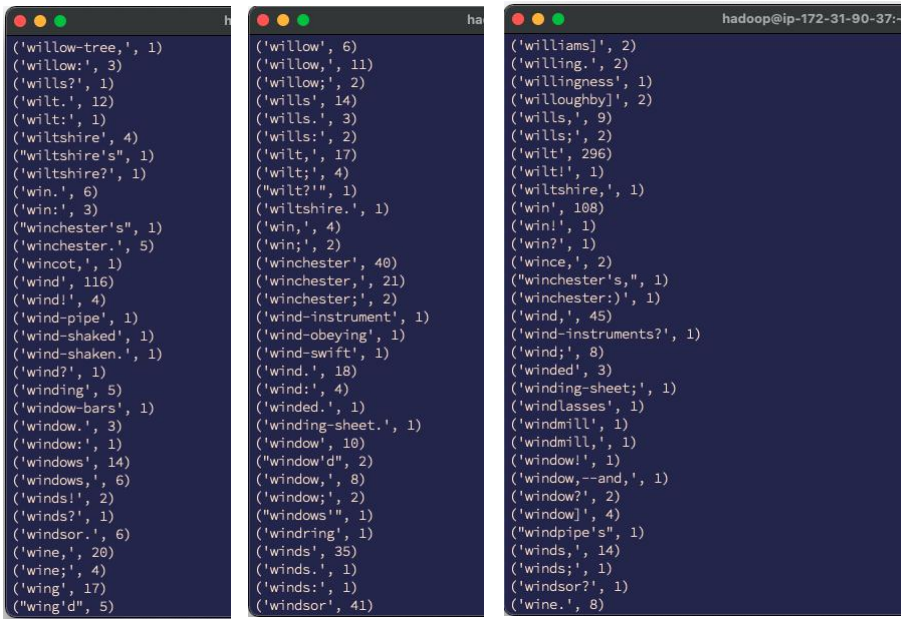
How many data-local map tasks were executed?	**4**
How many rack-local map tasks were executed?	**6**
How many reduce tasks?	**3**
Total time spent by all maps?	**4296096**
Total time spent by all reduce tasks?	**9677**
How many map-input records?	**173126**
How many map output records?	**939236**
How many reduce input groups?	**62933**



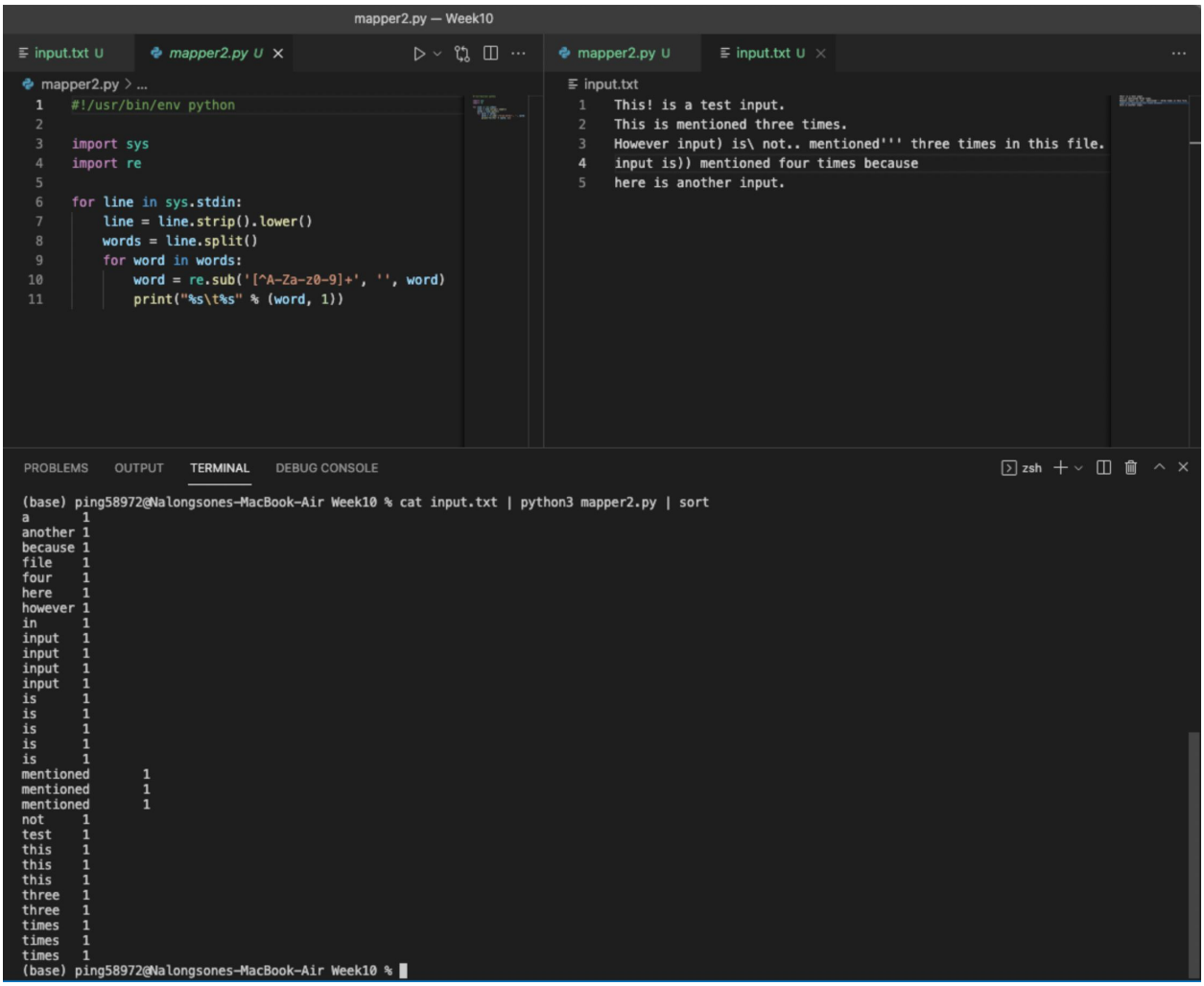
Lab report screen-shot #16:



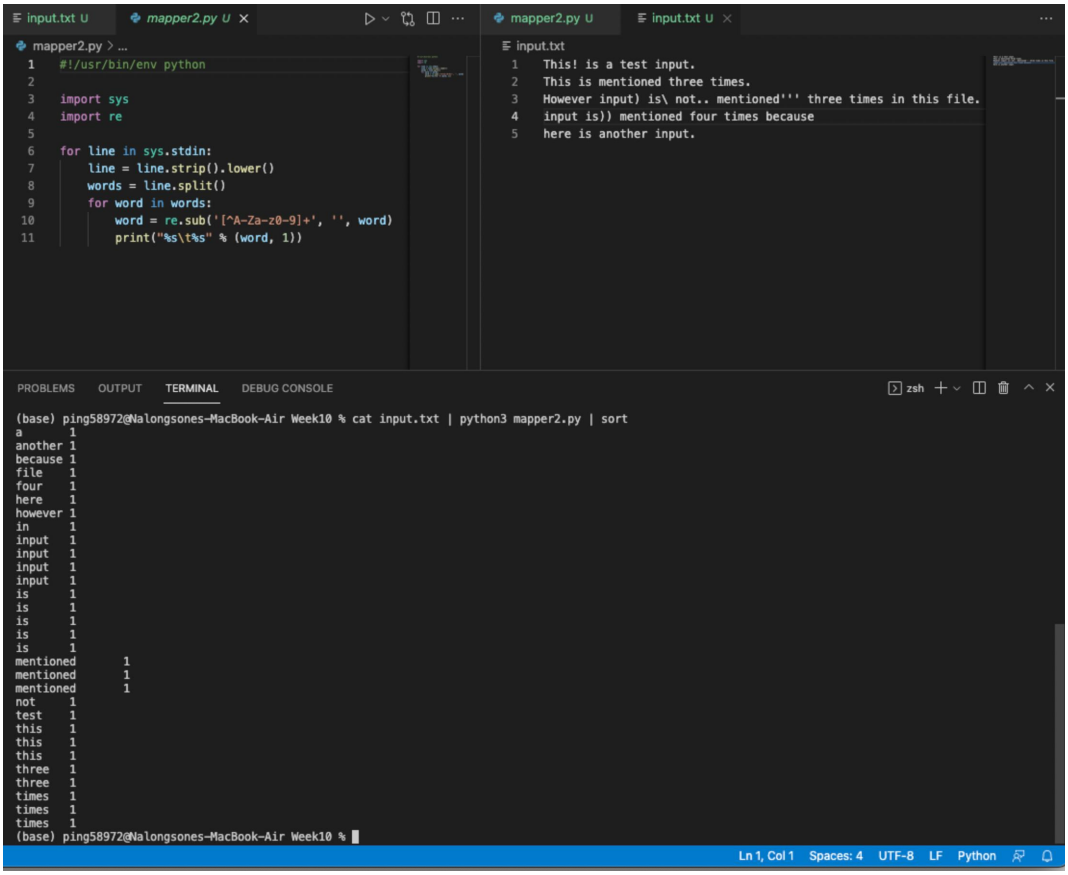
Lab report screen-shot #17:



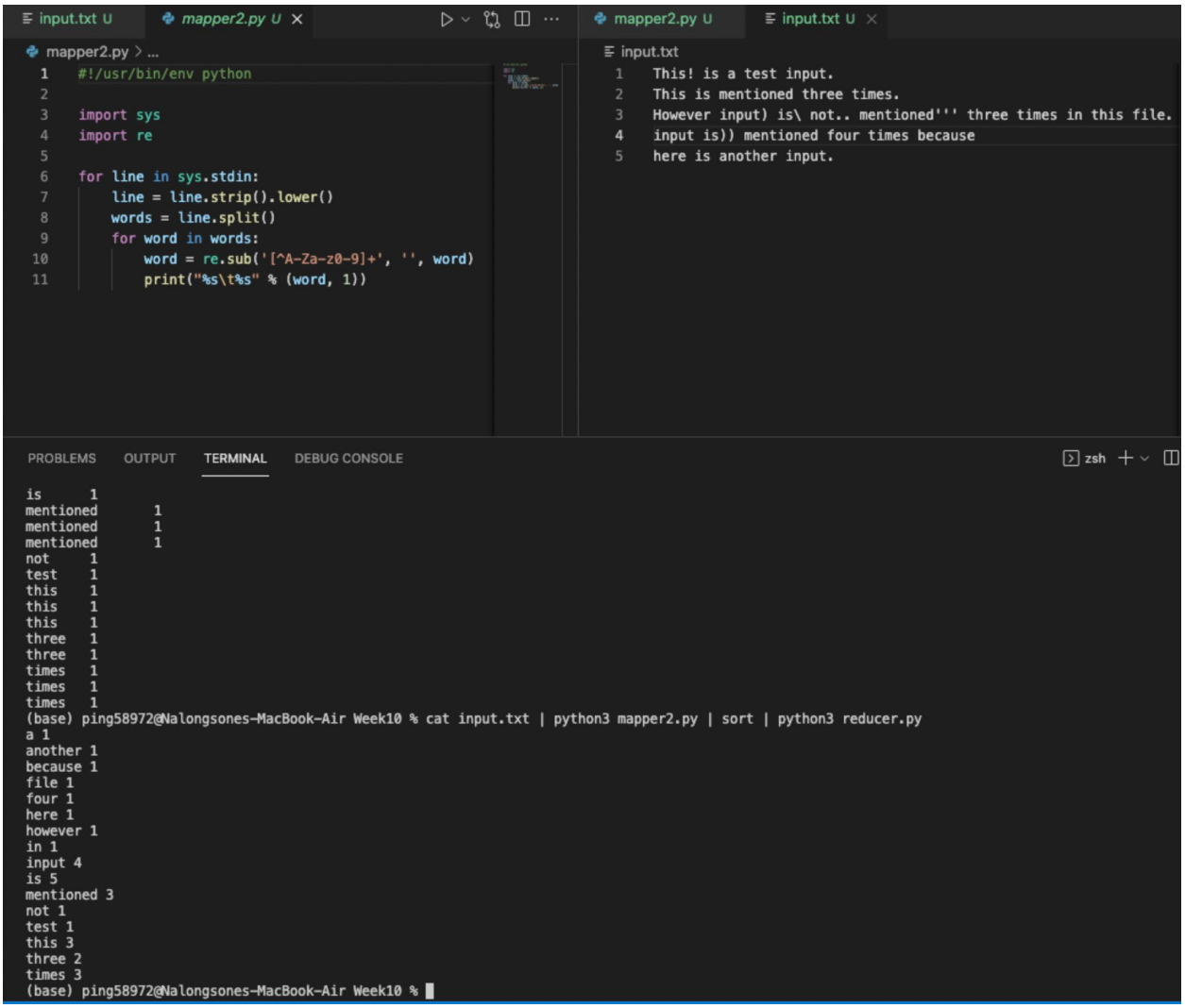
Lab report screen-shot #18:



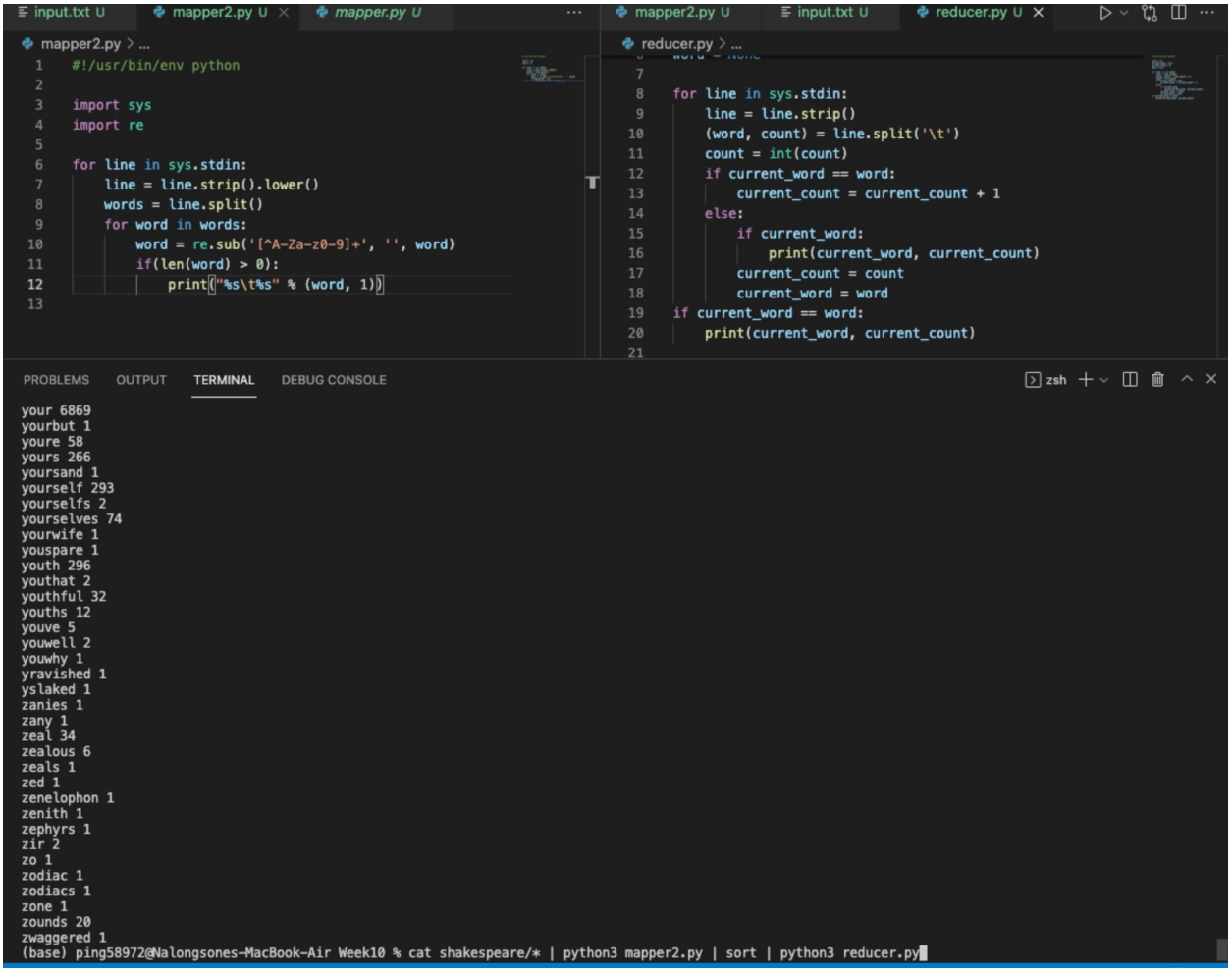
Lab report screen-shot #19:



Lab report screen-shot #20:

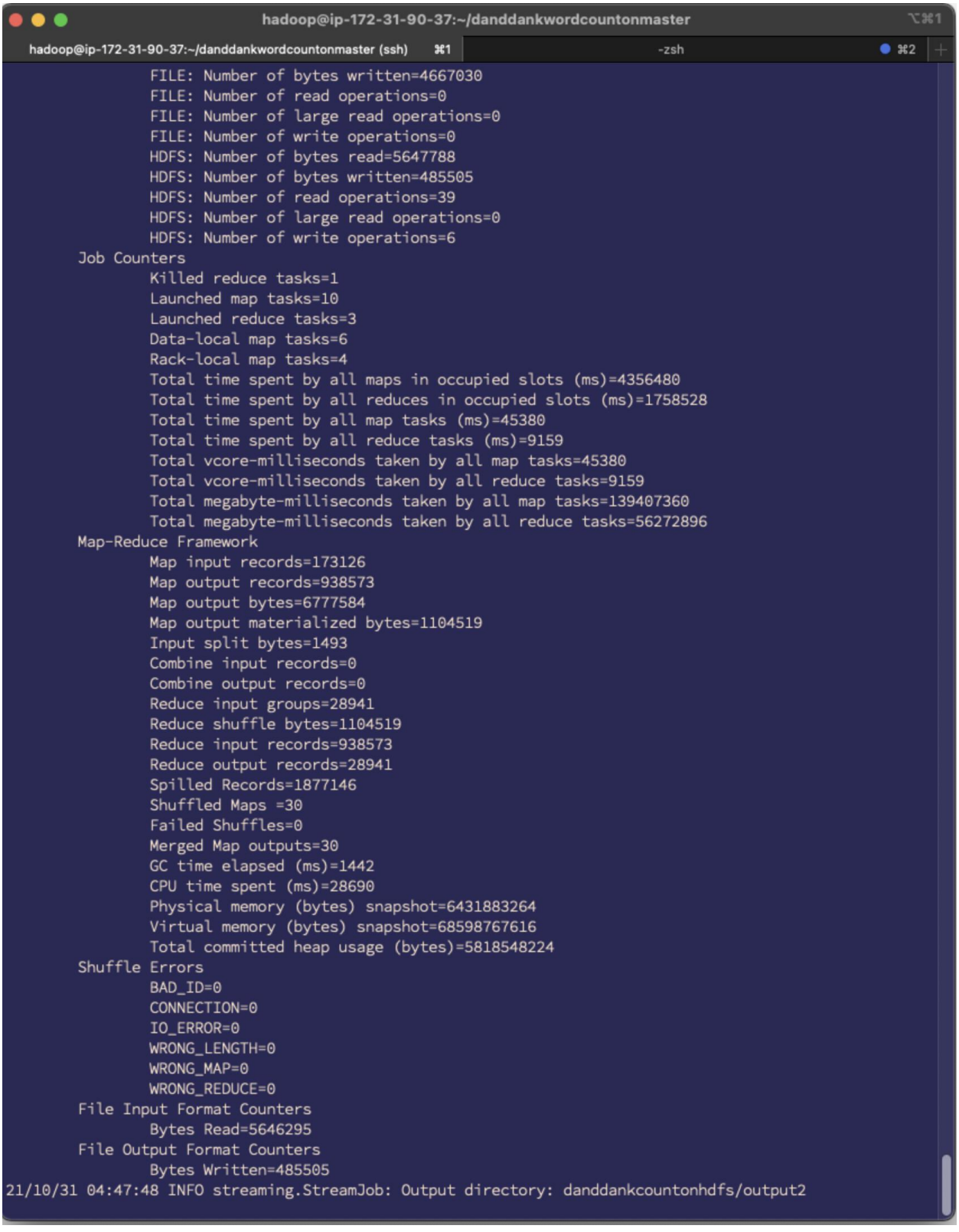


Lab report screen-shot #21:



Lab report screen-shot 22:

	mapper.py	**mapper2.py**
How many data-local map tasks were executed?	**4**	**6**
How many rack-local map tasks were executed?	**6**	**4**
How many reduce tasks?	**3**	**3**
Total time spent by all maps?	**4296096**	**4356480**
Total time spent by all reduce tasks?	**9677**	**9159**
How many map-input records?	**173126**	**173126**
How many map output records?	**939236**	**938573**
How many reduce input groups?	**62933**	**28941**


hadoop@ip-172-31-90-37:~/danddankwordcountonmaster
hadoop@ip-172-31-90-37:~/danddankwordcountonmaster (ssh) ⌘1
-zsh ⌘2
FILE: Number of bytes written=4667030
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=5647788
HDFS: Number of bytes written=485505
HDFS: Number of read operations=39
HDFS: Number of large read operations=0
HDFS: Number of write operations=6
Job Counters
Killed reduce tasks=1
Launched map tasks=10
Launched reduce tasks=3
Data-local map tasks=6
Rack-local map tasks=4
Total time spent by all maps in occupied slots (ms)=4356480
Total time spent by all reduces in occupied slots (ms)=1758528
Total time spent by all map tasks (ms)=45380
Total time spent by all reduce tasks (ms)=9159
Total vcore-milliseconds taken by all map tasks=45380
Total vcore-milliseconds taken by all reduce tasks=9159
Total megabyte-milliseconds taken by all map tasks=139407360
Total megabyte-milliseconds taken by all reduce tasks=56272896
Map-Reduce Framework
Map input records=173126
Map output records=938573
Map output bytes=6777584
Map output materialized bytes=1104519
Input split bytes=1493
Combine input records=0
Combine output records=0
Reduce input groups=28941
Reduce shuffle bytes=1104519
Reduce input records=938573
Reduce output records=28941
Spilled Records=1877146
Shuffled Maps =30
Failed Shuffles=0
Merged Map outputs=30
GC time elapsed (ms)=1442
CPU time spent (ms)=28690
Physical memory (bytes) snapshot=6431883264
Virtual memory (bytes) snapshot=68598767616
Total committed heap usage (bytes)=5818548224
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=5646295
File Output Format Counters
Bytes Written=485505
21/10/31 04:47:48 INFO streaming.StreamJob: Output directory: danddankcountonhdfs/output2

Lab report screen-shot #23: