**TERMWORK4**

**1.** hduser@ubuntu:~$ cd Documents/

2. hduser@ubuntu:~/Documents$ nano Term4data.csv

1, M, 25000, 2, Agree

2, F, 50000, 1, Disagree

3,M,75000,0,Neutral

4,F,80000,2,Agree

5,F,10000,1,DisAgree

6,F,20000,3,Neutral

7,M,17000,0,DisAgree

8,F,15000,0,DisAgree

9,M,60000,1,Agree

10,F,45000,1,Agree

11,F,46000,3,DisAgree

12,F,50000,3,Neutral

3. hduser@ubuntu:~/Documents$ nano mapper.py

#!/usr/bin/env python

import sys

for line in sys.stdin:

line = line.strip()

line = line.split(",")

if len(line) >=2:

pid = line[0]

opinion = line[4]

print '%s\t%s' % (pid, opinion)

4. hduser@ubuntu:~/Documents$ cat Term4data.csv| python mapper.py

5.hd.user@ubuntu:~/Documents$ nano reducer.py

#!/usr/bin/env python

import sys

opiniondic={}

count=0

for line in sys.stdin:

line = line.strip()

pid, opinion = line.split('\t')

if opinion in opiniondic:

opiniondic[opinion].append(count+1)

else:

opiniondic[opinion] = []

opiniondic[opinion].append(count+1)

for op in opiniondic.keys():

count=len(opiniondic[op])

print '%s\t%s'% (op,count)

6. hduser@ubuntu:~/Documents$ cat Term4data.csv| python mapper.py | python reducer.py

7. hduser@ubuntu:~/Documents$,star-all.sh

8. hduser@ubuntu:~/Documents$ jps

9. hduser@ubuntu:~/Documents$ hdfs dfs -mkdir /T4

10 hduser@ubuntu:~/Documents$ hdfs dfs -copyFromLocal /home/hduser/Documents/Term4data.csv /T4

11. hduser@ubuntu:~/Documents$ hdfs dfs -ls /

12. hduser@ubuntu:~/Documents$ chmod 777 mapper.py reducer.py

13. hduser@ubuntu:~/Documents$ ls -l

14. hduser@ubuntu:~/Documents$ hadoop jar /home/hduser/Documents/hadoop-streaming-2.7.3.jar \

> -input /T4/ Term4data.csv \

> -output /T4/output \

> -mapper /home/hduser/Documents/mapper.py \

> -reducer /home/hduser/Documents/reducer.py

15. hduser@ubuntu:~/Documents$ hdfs dfs -cat /T4 /output/part-00000