Assignment #1
CSC 555
Zehong Zhuang

Part 1

a)
1024
1048576
4096
42
17
15
7

b)
(1, -1, 0)
(3, 3, 6)
3.741657
3.741657
(13, 8, 7)
[(8, 4, 13), (5, 5, 7), (4, 1, 7)]
[(33,16,54),(22,15,34),(16,6,27)]

c)
0.0192
0.0192
0.0192
0.0432

d)
i.
```
SELECT unique(Project.PID), Project.Name
FROM Project, Assign
Where Project.PID NOT IN (
SELECT PID FROM Assign);
```
ii.
```
select ID
from employee
where ID = (
SELECT EID
From Assign
Having count (EID)>3
Group by EID);
```
iii.
```
select PID, Name
from Project
where PID in (
```

select PID
from Assign
having count (EID)<=2
group by PID);
e)
5
f)
i.
Mapper reads the input file.
Produce pair values (?, grade) from each block.
All the pair values will be transferred to Reducers by framework
Reducers perform average function to generate output (?, avg(grade))
ii.
Mapper read the input file from each block
Produce pair values (First, Grade) from each block.
All the pair values will be transferred to Reducers by framework
Reducers perform average function to generate output (First, avg(grade))
iii
Mapper read the input file from each block
Produce pair values (First, Last, Grade) from each block
All the pair values will be transferred to Reducers by framework
Reducers perform max function to generate output (First, Last, max(grade))

Part 2

2.

```
[ec2-user@ip-172-31-36-25 ~]$ cp myfile.txt mycopy.txt
[ec2-user@ip-172-31-36-25 ~]$ ls
mycopy.txt  myfile.txt
[ec2-user@ip-172-31-36-25 ~]$
```

3.

```
[ec2-user@ip-172-31-36-25 ~]$ cd CSC555
[ec2-user@ip-172-31-36-25 CSC555]$ ls
mycopy.txt  myfile.txt
[ec2-user@ip-172-31-36-25 CSC555]$
```

6.

```
[ec2-user@ip-172-31-36-25 CSC555]$ zip myzipfile mycopy.txt myfile.txt
  adding: mycopy.txt (stored 0%)
  adding: myfile.txt (stored 0%)
[ec2-user@ip-172-31-36-25 CSC555]$ mv myzipfile.zip /home/ec2-user/
[ec2-user@ip-172-31-36-25 CSC555]$ cd
[ec2-user@ip-172-31-36-25 ~]$ ls
CSC555  myzipfile.zip
[ec2-user@ip-172-31-36-25 ~]$ unzip myzipfile.zip
Archive:  myzipfile.zip
 extracting: mycopy.txt
 extracting: myfile.txt
[ec2-user@ip-172-31-36-25 ~]$
```

7.

88k

```
-rw-rw-r-- 1 ec2-user ec2-user  386 Jan 21 04:59 myzipfile.zip
[ec2-user@ip-172-31-36-25 ~]$ ls -lh
total 88K
-rw-rw-r-- 1 ec2-user ec2-user 73K Aug  9  2000 grail
-rw-rw-r-- 1 ec2-user ec2-user  34 Jan 21 04:56 mycopy.txt
-rw-rw-r-- 1 ec2-user ec2-user  34 Jan 21 04:54 myfile.txt
-rw-rw-r-- 1 ec2-user ec2-user 386 Jan 21 04:59 myzipfile.zip
[ec2-user@ip-172-31-36-25 ~]$
```

11.

```
[ec2-user@ip-172-31-36-25 ~]$ chmod u-r myfile.txt
[ec2-user@ip-172-31-36-25 ~]$ more myfile.txt
more: cannot open myfile.txt: Permission denied
[ec2-user@ip-172-31-36-25 ~]$
```

12.

```
[ec2-user@ip-172-31-36-25 ~]$ python tx.py
Word 1 false
Word 1 false
[ec2-user@ip-172-31-36-25 ~]$ cat tx.py
file=open("mycopy.txt","r")
for line in file:
    array=[]
    word=line.split()
    if word in array:
        print("Word 1 true")
    else:
        array.append(word)
        print("Word 1 false")
```

Part 3

```
[ec2-user@ip-172-31-36-25 ~]$ hadoop fs -put bioproject.xml /data/
[ec2-user@ip-172-31-36-25 ~]$ hadoop fs -ls /data
Found 1 items
-rw-r--r--   1 ec2-user supergroup  231149003 2019-01-21 05:54 /data/bioproject.xml
[ec2-user@ip-172-31-36-25 ~]$
                Bytes Read=231153099
        File Output Format Counters
                Bytes Written=20056175

real    1m11.060s
user    0m3.769s
sys     0m0.247s
[ec2-user@ip-172-31-36-25 ~]$ hadoop fs -du /data/wordcount1/
0         /data/wordcount1/_SUCCESS
20056175  /data/wordcount1/part-r-00000
```