**CSC 555 Assignment 1**

**Yiyang Yang**

**Part 1**

1. 34 / 5 \* 4 = 28 mins
2. 34 / 10 \* 4 = 16 mins
3. 34 / 10 \* 4 = 16 mins
4. 2 + 2 = 4 mins
5. Run time won’t be changed, same as the time in a 28 mins.

**Part 2**

1. 1.

2.

3.

4.

5.

6.

1. PR(Q) = PR(X)\*0 + PR(Y)\*1/2 + PR(Z)\*1/3 = 1/2 \* 0 + 1/2 \* 1/2 + 0 \* 1/3 = 1/4
2. For the first node, since it is self-loop, the page rank for this node is 1.

Add a node to the first one, the page rank is 1 \* 1/2 = 1/2.

Add another node to it, the page rank is 1 \* 0 + 1/2 \* 1 = 1/2.

Add nth node to it , the page rank of the last is 1/2.

So, the page rank for first node is 1, and all the other node is 1/2.

**Part 3**

1. (5 + 15 + 15 + 20 + 12 + 20 + 23 + 25 + 22)/9 = 17.444
2. None-overlapping window:

1pm - 3pm: (5 + 15 + 15)/3 = 11.667

4pm – 6pm: (20 + 12 + 20)/3 = 14

7pm – 9pm: (23 + 25 + 22)/3 = 17.5

1. Forward 2 hours each time:

1pm – 3pm: (5 + 15 + 15)/3 = =11.667

3pm – 5pm: (15 + 20 + 12)/3 = 15.667

5pm – 7pm: (12 + 20 + 23)/3 = 18.333

7pm – 9pm: (23 + 25 + 22)/3 = 17.5

**Part 4**

#!/usr/bin/python

import sys

for line in sys.stdin:

words = line.strip().split(' ')

for word in words: if word != '' and word != '<, >, “, &, . !, ?, / , .': # Ignore blanks

print "%s\t1" % (word)

#!/usr/bin/python

import sys

curr\_id = None

curr\_cnt = 1

id = None # The input comes from standard input (line by line)

for line in sys.stdin:

line = line.strip() # parse the line and split it by '\t'

ln = line.split('\t') # grab the key

id = ln[0]

if curr\_id == id:

curr\_cnt += 1

else:

if curr\_id >= 3: # output the count, single key completed

print '%s\t%d' % (curr\_id, curr\_cnt)

curr\_id = id

curr\_cnt = 1 # output the last key!

if curr\_id == id:

print '%s\t%d' % (curr\_id, curr\_cnt)

**Part 5**

a)

put 'employees', 'ID1', 'public: postion', 'manager'

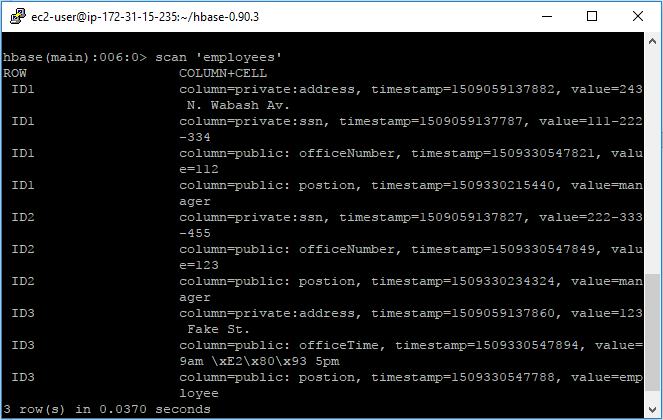
put 'employees', 'ID2', 'public: postion', 'manager'

put 'employees', 'ID3', 'public: postion', 'employee'

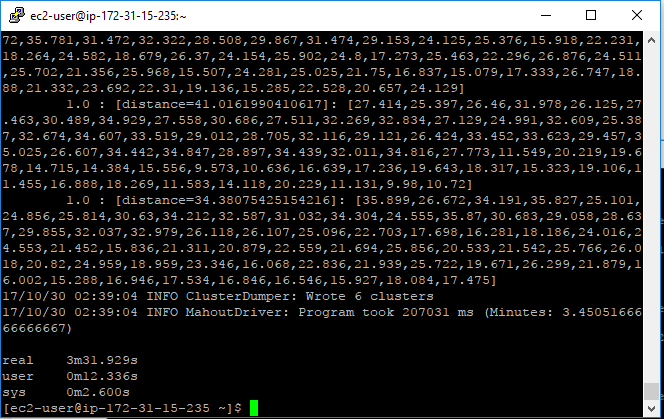
put 'employees', 'ID1', 'public: officeNumber', '112'

put 'employees', 'ID2', 'public: officeNumber', '123'

put 'employees', 'ID3', 'public: officeTime', '9am – 5pm'



b)



First page of clusteranalyze.txt

