

Assignment #4
CSC 555
Zehong Zhuang

1)

- a) $78 \times 3 / 5 = 46.8 \times 1 \text{ min} = 46.8 \text{ mins}$, round up to 47 mins
- b) $78 / 10 = 7.8 \times 1 \text{ min} = 7.8 \text{ mins}$, round up to 8 mins
- c) $78 \times 3 / 10 = 23.4 \times 1 \text{ min} = 23.4 \text{ mins}$, round up to 24 mins
- d) 1 min
- e) $78 \times 3 / 100 = 2.34 \times 1 \text{ min} = 2.34 \text{ mins}$, round up to 3 mins

2)

- a) 19 iterations after it converges

```
[1,] 0.3075256 0.3059082 0.3069458 0.3062744
> c19=c18%%a
> c19
      [,1]      [,2]      [,3]      [,4]
[1,] 0.3075256 0.3059082 0.3069458 0.3062744
> c20=c18%%a
> c20
      [,1]      [,2]      [,3]      [,4]
[1,] 0.3075256 0.3059082 0.3069458 0.3062744
```

- b) Drop dead-end node Q and P

14 iterations after it converges

```
[1,]
> d14=d13%%d
> d14
      [,1]      [,2]      [,3]
[1,] 0.3671875 0.3645833 0.3671875
> d15=d13%%d
> d15
      [,1]      [,2]      [,3]
[1,] 0.3671875 0.3645833 0.3671875
```

$[A, Y, Z] = [0.3671875, 0.3645833, 0.3671875]$

$\text{PageRankQ} = 0 \times 0.3671875 + 1/2 \times 0.3645833 + 1/2 \times 0.3671875$
 $= 0.3658854$

- c) A=1
B=1/2
N=1/2

3)

Mapper

```
#!/usr/bin/python
import sys

for line in sys.stdin:
    line=line.strip()
    split=line.split('|')
    for i in split:
        if i.startswith('800'):
            firstFour=split[0:4]
        elif i.startswith('EMP'):
            firstFour=split[0:4]
```

Reducer

```
#!/usr/bin/python

import sys

currentKey=None
emp=None
cust=None

for line in sys.stdin:
    split=line.strip().split('\t')
    key=split[1:3]
    value='\t'.join(split[0,4])
    if currentKey==key:
        if value.startswith('800'):
            cust.append(value)
        if value.startswith('EMP'):
            emp.append(value)
    else:
        if currentKey:
            lenEmp=len(emp)
            lenCust=len(cust)
            if (lenEmp*lenCust>0):
                print currentKey, '\t', emp[-1],cust[-1]
        emp=[]
        cust=[]
        currentKey=key
        if value.startswith('800'):
            cust.append(value)
        elif value.startswith('EMP'):
            emp.append(value)
            lenEmp=len(emp)
            lenCust=len(cust)
            if (lenEmp*lenCust>0):
                print currentKey, '\t', emp[-1], cust[-1]
```

```
17/03/07 00:20:30 INFO Streaming.StreamJob: Output directory: EmpCust/output_1
[ec2-user@ip-172-31-3-189 ~]$
[ec2-user@ip-172-31-3-189 ~]$ hadoop jar hadoop-streaming-2.6.4.jar -input EmpCu
st -output EmpCust/output_1 -mapper 'nameMap.py' -reducer 'nameReduce.py' -file
nameReduce.py -file nameMap.py
```

4)

a) Using 4-node Clusters

```
hbase(main):025:0> scan 'employees'
ROW COLUMN+CELL
ID1 column=private:LastName, timestamp=1551802646965, value=Fruits
ID1 column=private:address, timestamp=1551802568164, value=243 N. Wabash Av.
ID1 column=private:firstName, timestamp=1551802604788, value=Apple
ID1 column=private:ssn, timestamp=1551802553577, value=111-222-334
ID1 column=public:country, timestamp=1551802697454, value=CHINA
ID2 column=private:LastName, timestamp=1551802657321, value=Fruits
ID2 column=private:firstName, timestamp=1551802625011, value=Banana
ID2 column=private:ssn, timestamp=1551802560055, value=222-333-445
ID2 column=public:country, timestamp=1551802709862, value=USA
ID3 column=private:address, timestamp=1551802564296, value=123 State St.
ID4 column=others:color, timestamp=1551802897625, value=red
ID4 column=others:gender, timestamp=1551802863036, value=male
ID4 column=others:shape, timestamp=1551802931417, value=square
ID5 column=others:color, timestamp=1551802908939, value=green
ID5 column=others:gender, timestamp=1551802880622, value=female
ID5 column=others:shape, timestamp=1551802948382, value=line
5 row(s) in 0.0240 seconds
```

b)

```
[ec2-user@ip-172-31-3-189 ~]$ cat clusteranalyze.txt
{"identifier":"VL-584","r":[3.573,3.49,3.294,3.32,3.468,3.344,3.618,3.699,3.813,3.724,3.506,3.896,3.70
8,3.554,3.7,3.717,3.418,3.984,3.976,3.952,4.377,4.436,4.328,5.494,4.848,4.959,5.124,5.823,5.665,6.024,
6.256,6.398,6.448,6.475,6.122,6.183,6.078,5.736,5.444,4.684,4.65,5.059,4.372,4.743,4.587,4.718,5.198,4
.519,4.59,4.55,4.378,4.544,4.468,4.42,4.553,4.738,4.658,4.587,4.873,5.198],"c":["30.22,30.146,30.391,29
.469,29.741,29.664,29.548,30.032,28.861,29.501,29.111,28.992,29.094,29.128,28.342,28.443,28.495,28.363
,28.196,28.944,28.900,27.737,27.935,27.41,26.737,25.956,26.46,25.866,25.885,25.049,24.479,23.716,22.86
7,22.276,20.392,20.601,18.8,29.109,18.457,18.602,18.42,18.005,17.876,17.485,17.375,17.046,16.752,16.94
9,17.53,17.498,17.165,17.602,16.906,16.069,16.623,16.285,16.635,16.39,15.805,16.611],"n":123}
Weight : [props - optional]: Point:
1.0 : [distance=28.84341517866862]: [24.672,35.261,30.486,34.447,27.469,33.179,32.28,31.612,33
.215,30.145,25.664,26.51,23.344,22.746,23.703,25.613,27.95,30.915,27.055,32.099,28.053,25.602,25.857,2
3.649,23.729,20.707,26.265,24.739,23.297,28.814,28.322,24.125,27.636,19.49,20.211,23.685,17.537,24.913
,23.852,17.429,18.166,26.208,16.25,18.389,19.903,17.949,26.284,16.435,22.171,16.566,14.734,20.814,15.6
15,25.051,17.75,22.335,12.816,20.545,17.145,16.969]
1.0 : [distance=29.44680567725921]: [25.592,29.951,34.188,25.658,26.887,23.573,34.07,32.134,24
.226,32.835,28.736,22.764,27.075,31.695,23.068,28.177,30.347,21.692,23.52,25.869,20.738,26.484,25.945,
26.356,26.61,27.923,22.344,18.341,25.289,17.043,23.898,21.45,21.012,26.453,19.442,19.78,23.152,16.66,2
3.176,24.844,21.37,24.335,22.465,17.06,12.203,11.832,15.639,14.378,17.319,18.499,10.786,17.209,15.585,
17.023,19.042,18.056,17.958,15.153,9.625,17.662]
1.0 : [distance=30.842581038308975]: [30.696,32.667,34.223,33.183,34.835,33.391,33.175,32.804,
24.116,25.19,22.739,25.053,32.679,31.196,32.16,29.381,23.589,31.786,24.265,30.298,21.2,26.239,30.859,2
9.984,21.029,27.869,18.415,19.499,23.458,24.589,25.958,23.921,26.189,27.101,27.984,21.713,20.958,20.11
,16.171,26.001,21.95,22.971,17.464,19.791,14.989,15.0,24.85,20.741,23.414,16.101,15.681,15.673,23.288,
17.766,21.817,16.371,12.439,18.907,17.32,12.94]
1.0 : [distance=23.834311378345646]: [32.702,35.219,30.143,24.275,28.156,26.281,26.887,29.739,
28.588,32.115,28.952,31.654,23.86,24.503,26.14,26.283,25.897,31.719,24.259,31.153,27.673,28.435,27.952
,23.764,20.125,24.848,27.495,24.808,20.754,24.518,18.523,22.455,25.533,19.716,17.452,17.822,18.375,18
.684,17.331,20.561,21.989,20.922,15.342,21.997,23.533,19.106,20.59,15.386,23.64,15.969,16.974,18.554,18
.152,14.431,18.404,12.034,16.727,17.414,10.661,12.707]
1.0 : [distance=24.253642666500138]: [30.343,33.085,28.13,31.294,28.719,30.306,26.441,29.986,2
5.757,26.601,27.699,27.233,29.376,31.373,30.535,24.821,23.137,24.924,30.362,29.024,28.737,19.135,19.31
8,22.184,24.326,21.256,24.222,24.839,24.351,18.481,21.962,20.152,18.972,22.825,22.988,23.799,18.61,17.
205,17.968,22.92,21.987,22.731,18.08,19.168,20.863,19.833,16.372,19.79,16.253,15.409,16.462,19.237,14.
938,12.695,16.116,19.813,17.155,19.612,19.827,13.522]
1.0 : [distance=31.80467747403478]: [35.808,31.081,24.708,33.25,27.58,32.197,33.551,25.853,30.
312,29.219,30.36,25.589,31.237,30.415,26.281,29.581,25.528,24.473,30.356,19.168,25.713,25.939,29.067,2
2.784,24.974,18.374,19.958,16.594,25.982,25.78,21.492,16.749,24.357,19.331,23.435,25.199,22.319,22.317
,16.496,14.443,21.78,24.062,17.402,22.303,12.297,17.604,11.538,12.63,11.392,12.564,18.107,18.251,16.26
4,13.074,19.478,17.523,8.629,16.786,7.773,12.647]
1.0 : [distance=27.84568402556544]: [26.244,25.101,24.183,28.193,28.75,29.573,24.273,27.766,26
.017,23.423,28.825,30.355,30.079,30.958,30.375,24.422,26.932,29.466,20.477,30.426,23.994,24.956,22.684
,27.437,25.091,20.378,26.951,19.092,20.62,22.601,24.356,27.178,26.322,23.844,22.201,23.775,17.645,19.5
97,21.821,23.409,21.782,22.818,19.625,21.251,21.802,16.1,25.75,22.902,19.219,18.175,16.457,20.239,17.7
02,13.948,19.203,14.477,21.579,20.673,17.885,13.123]
1.0 : [distance=34.77347440414934]: [27.758,33.859,29.879,25.174,32.274,32.354,25.113,26.755,2
2.779,27.211,24.251,24.493,28.648,26.506,21.308,23.732,29.583,25.733,27.46,28.056,17.796,21.303,20.008
,19.436,22.987,27.368,25.791,16.325,23.785,24.122,25.072,20.572,24.287,24.497,17.704,18.971,14.808,18.
94,21.819,18.776,23.273,21.283,21.968,18.279,12.465,11.504,9.975,12.097,10.054,10.406,15.096,16.509,10
.029,9.826,14.062,9.928,8.387,17.475,14.684,14.503]
1.0 : [distance=30.104961888202755]: [34.707,24.586,34.949,35.057,31.932,34.388,25.907,28.421,
28.444,25.239,26.056,24.762,33.058,23.791,30.8,22.383,31.129,30.908,28.38,29.889,22.698,23.037,21.353,
21.311,26.264,20.623,27.406,25.404,28.99,28.413,19.282,24.602,26.614,19.501,22.286,18.976,25.599,20.64
4,16.278,26.959,18.411,25.464,17.051,17.173,21.066,21.719,20.119,24.455,15.362,15.673,13.323,23.259,17
.306,14.847,18.058,16.086,17.895,20.662,16.57,18.433]
```