Make sure you have the ‘web-BerkStan.txt’ in your hdfs directory

1.scp the web-BerkStan.txt file to linux machine

scp web-BerkStan.txt username@linux.dc.engr.scu.edu:/home/username/

2. SSH to SCU cluster system

3.hdfs dfs -copyFromLocal web-BerkStan.txt

**Pagerank\_poweriteration.py (Pagerank implementation with power iteration method with partitions)**

1.scp Pagerank\_poweriteration.py to scu linux machine

2.From the folder containing Pagerank\_poweriteration.py file enter pyspark --master='local[20]'

3.In pyspark terminal enter the following commands:

importPagerank\_poweriteration as P

P.pageRank(‘web-BerkStan.txt’,10,20,sc) //10 partitions, 20-iterations (user can change the input values)

**Pagerank\_poweriteration\_noPartition.py (Pagerank implementation with power iteration method without partitions)**

1.scp Pagerank\_poweriteration\_noPartition.py to scu linux machine

2.From the folder containing Pagerank\_poweriteration\_noPartition.py file enter pyspark --master='local[20]'

3.In pyspark terminal enter the following commands:

import Pagerank\_poweriteration\_noPartition as PN

PN.pageRank(‘web-BerkStan.txt’,20,sc) //20-iterations (user can change the input values)

**Pagerank\_naive.py (Pagerank naive implementation with partitions)**

1.scp Pagerank\_naive.py to scu linux machine

2.From the folder containing Pagerank\_naive.py file enter pyspark --master='local[20]'

3.In pyspark enter the following commands:

import Pagerank\_naive as N

N.pageRank(‘web-BerkStan.txt’,10,20,sc) //10 partitions, 20-iterations (user can change the input values)

**Pagerank\_naive\_noPartition.py (Pagerank naive implementation without partitions)**

1.scp Pagerank\_naive\_noPartition.py to scu linux machine

2.From the folder containing Pagerank\_naive\_noPartition.py file enter

pyspark --master='local[20]'

3.In pyspark enter the following commands:

import Pagerank\_naive\_noPartition as NN

NN.pageRank(‘web-BerkStan.txt’,20,sc) //20-iterations(user can change the input values)