Jorge Quezada

Big Data CSC 4760

April 13, 2020

Assignment 4

Assignment 4 – PageRank Spark

1. Explanation of the source code: What are the functions of your lambda functions? Which kind of intermediate results are generated?

lambda line : line.split(" ")

lambda x : (int(x[0]), [int(y) for y in x[1:]])

lambda v : 1 / float(nNumOfNodes)

lambda x\_y\_z : [(i, x\_y\_z[1][1] / len(x\_y\_z[1][0])) for i in x\_y\_z[1][0]]

lambda x, y : x + y

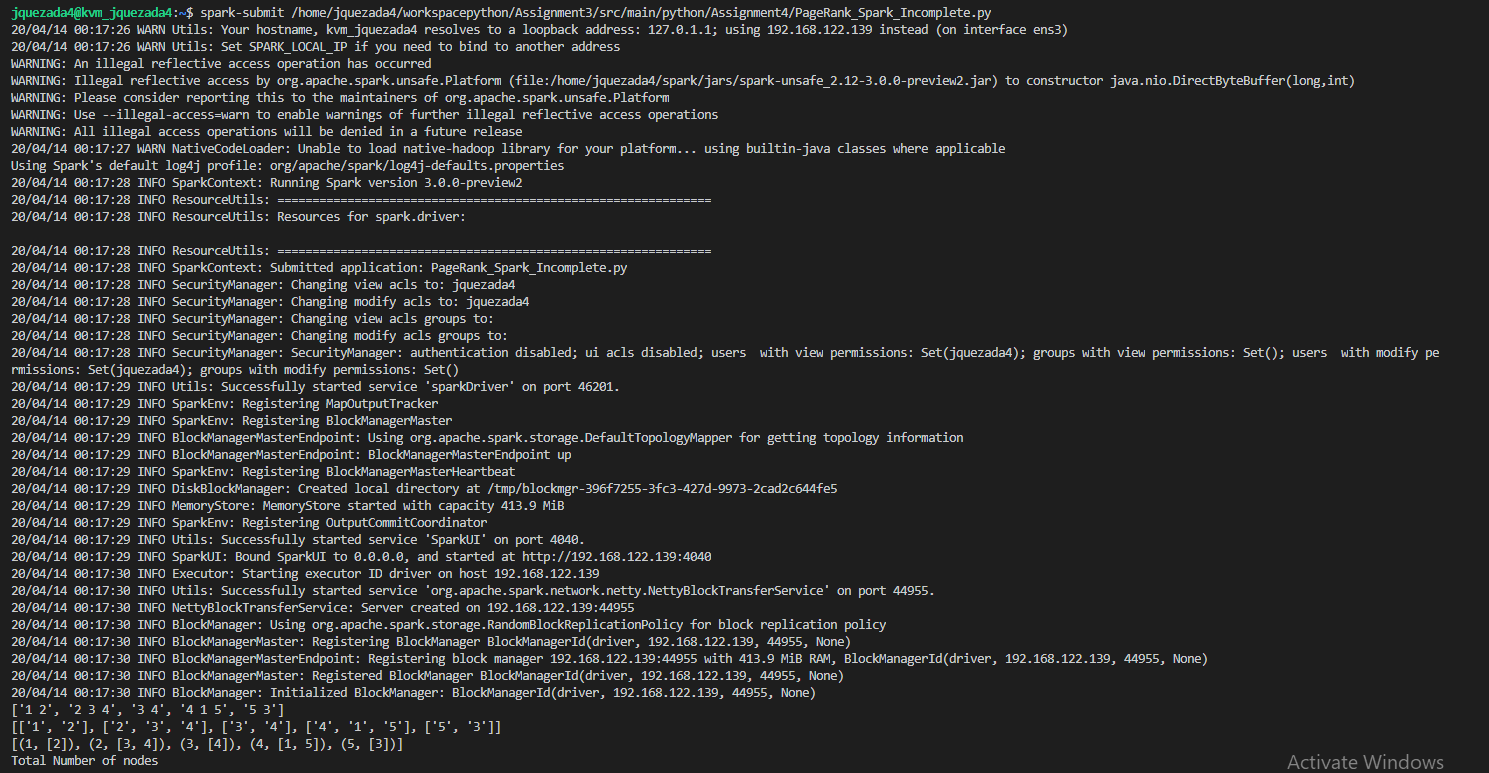
lambda v : 0.85 \* v + 0.15 / float(nNumOfNodes)

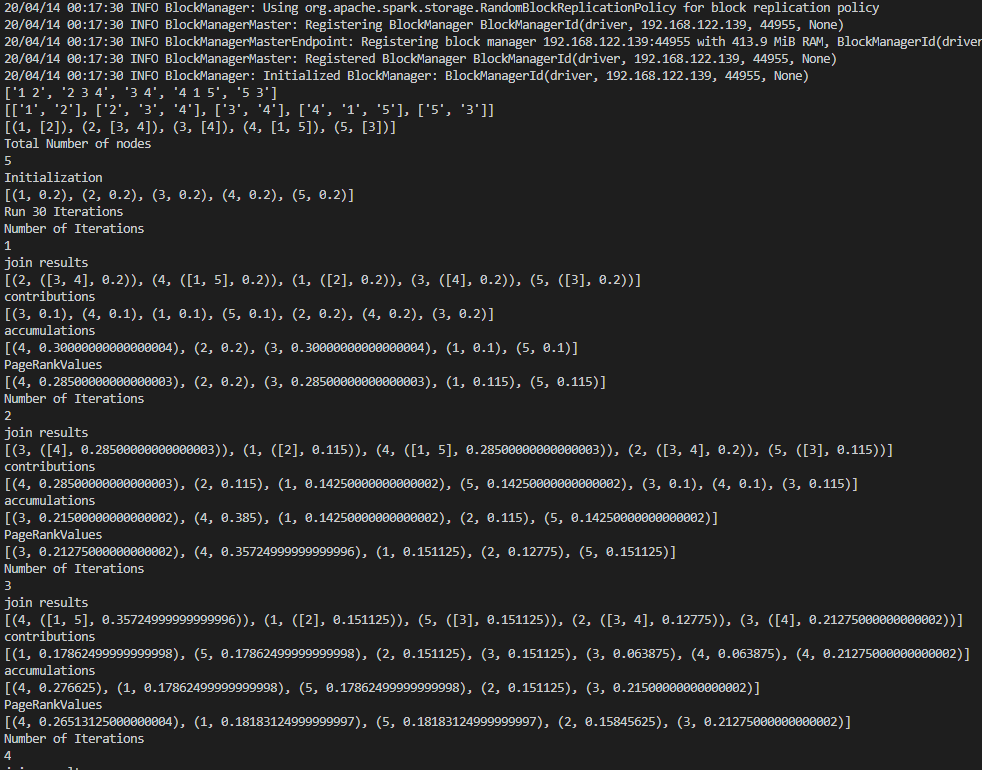
These were the functions for my lambda functions. For the first line, you had to get the values from the adjacency list so we split the lines according to the spaces. The next lambda function allows us to get PageRank values so that then we can collect them. The 3rd line allows us to calculate the PageRank Values again to be mapped. The 4th line allows us to loop through each tuple format of values and correctly map those to our contributions. Then our 5th line lets us add the values together, so we know how much we accumulated. Our 6th line calculates the PageRank values with the rate of decay. Some of the intermediate results generated were the join results, the contributions, the accumulations, and the PageRank Values for that iteration.

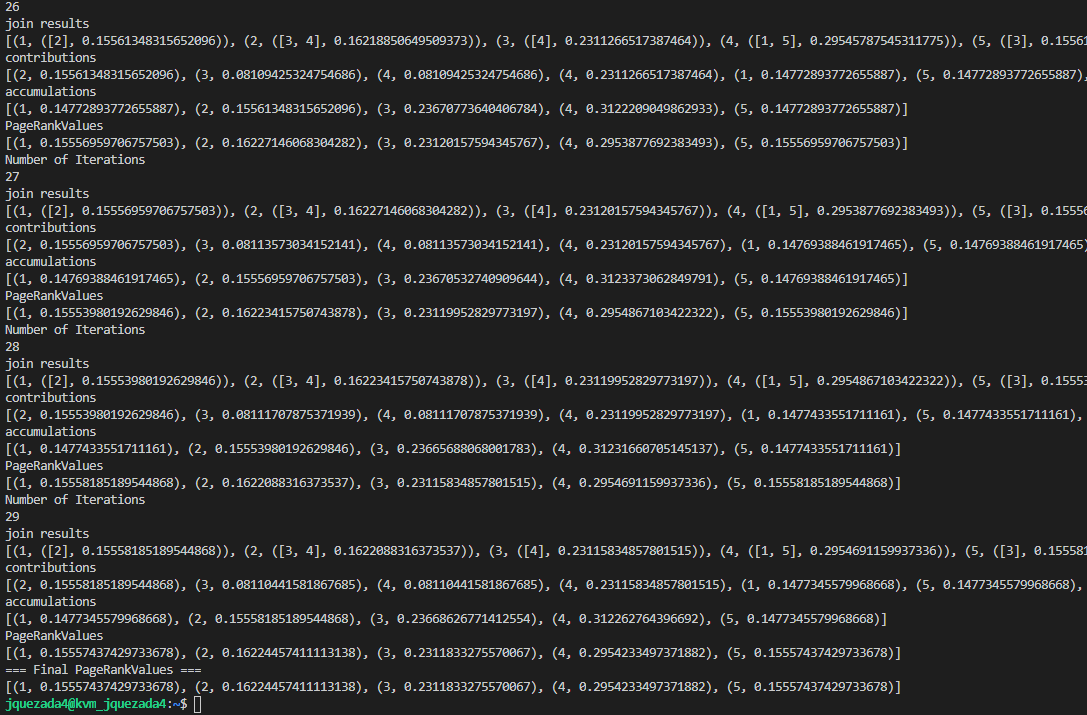
2) Experimental Results

Some errors I was getting while working with this code was when I was working with the tuples. After looking up some ways on how to deal with tuples in Python3 I was able to correctly get an output. I was also having trouble getting an output text file, but this was occurring because I had not correctly fixed my tuples yet.

2.1) Screenshots of the key steps. For example, the screenshot for the outputs in the terminal when you run the command. It will demonstrate that your program has no bug.







2.2) Explain your results. Does your implementation give the exact PageRank values?

My implementation of PageRank gives the exact value up to the max size of a float. These values are for a given decay factor of .85.

