BIG DATA AND ANALYTICS ASSIGNMENT # 2

Due Data: 10 March, 2023

Assignments are to be done in same group as previous. No late assignments will be accepted.

HONOR POLICY

This assignment is a learning opportunity that will be evaluated based on your ability to think, work through a problem in a logical manner. You may however discuss verbally or via email the assignment with your classmates or the course instructor, and use the Internet to do your research, but the written work should be your own. Plagiarized reports or code will get a zero. If in doubt, ask the course instructor.

TASK 1# /70

In this task you will download the following data:

https://snap.stanford.edu/data/soc-LiveJournal1.html

After downloading the data you will explore the data and try to map each id with it's all the friend nodes.

0, [2,3,4,5,6,…,46]
1,[5,20,135,2409,8715, …., 1413125]

Using this prepared data you will tell me how all the nodes will be related to x, y and z number if a direct connection exists well and good. But if not you have too tell me through which persons you can approach the x, y and z person.

X will be Roll no of First Student :-> 19i-2000 := X will be 2000
Y will be Roll no of First Student :-> 19i-2010 := Y will be 2010
Z will be Roll no of First Student :-> 19i-1900 := Z will be 1900

You have to map shortest path from node X, Y, Z to every other nodes. For example if your roll no is 0 then

1, [5]

One can Approach 0 with the help of 5 similarly also 0 can be also approached from 20.

1,[20]

So any Path is acceptable you have to provide me with any of the one. And another possibility will be invalid.

This will be done for all the id **[0 to n ids in Data]** with your group members roll no. You will submit 3 csv having **id** and **path**.

This will be done using regular python.

TASK 2# /80

Now you have to implement the Above Logic using Hadoop. And Produce Hadoop Generated File in CSV Format. You cannot use any Regular Python in this step.

For all these Task you will provide me a report in which you will tell me how you implemented your logic. You will compare time taken by task one and task two. Tell me the problems and challenges you faced. What you learned from the given Assignment.

Best of Luck 😃