

1. GENERAL INSTRUCTIONS

These questions require thought but do not require long answers. Please be as concise as possible. You should submit your answers as a write-up in a PDF format and code via **Moodle**.

2. QUESTION

Write a MapReduce program in Hadoop that implements a simple “People You Might Know” social network friendship recommendation algorithm. The key idea is that if two people have a lot of mutual friends, then the system should recommend that they connect with each other.

3. INPUT

Download the input file from the **Moodle**. The name of the file is *FriendData.txt*. The input file contains the adjacency list and has multiple lines in the following format:

`<User><TAB><Friends>`

Here, `<User>` is a unique integer ID corresponding to a unique user and `<Friends>` is a comma-separated list of unique IDs corresponding to the friends of the user with the unique ID `<User>`. Note that the friendship is mutual (i.e. edges are undirected): if A is friend with B , then B is also friend with A . The data provided is consistent with that rule as there is an explicit entry for each side of each edge.

4. ALGORITHM

Let us use a simple algorithm such that, for each user U , the algorithm recommends $N = 10$ users who are not already friends with U , but have the largest number of mutual friends in common with U .

5. OUPUT

The output should contain one line per user in the following format:

`<User><TAB><Recommendations>`

where `<User>` is a unique ID corresponding to a user and `<Recommendations>` is a comma-separated list of unique IDs corresponding to the algorithms recommendation of people that `<User>` might know, ordered by decreasing number of mutual friends. Even if a user has fewer than 10 second-degree friends, output all of them in decreasing order of the number of mutual friends. If a user has no friends, you can provide an empty list of recommendations. If there are multiple users with the same number of mutual friends, ties are broken by ordering them in a numerically ascending order of their user IDs.

Also, please provide a description of how you are going to use MapReduce jobs to solve this problem. Don't write more than 3 to 4 sentences for this: we only want a very high-level description of your strategy to tackle this problem.

6. WHAT TO SUBMIT

- (1) Submit the source code with the filename as `Your_Roll_Number.*`. Here `*` may `.java` or `.py` or `.scala`.
- (2) Include in your writeup a short paragraph describing your algorithm to tackle this problem. Name this file as `Your_Roll_Number.pdf`
- (3) Include in your writeup the recommendations for the users with following user IDs: 924, 8941, 8942, 9019, 9020, 9021, 9022, 9990, 9992, 9993.