Big Data

Project 1- Warm Up!

Find Friends

As you know, part of this course is to be familiar with some programming languages such as python, java, and scala. Therefore, using one of these languages, which are the main languages for the Big Data, please see the following:

1- You will be given a file named <u>hobbies.txt</u>.
   a. This file contains a group of fictitious Facebook users and their hobbies.
   b. Each line in the file contains a user/username and a list of hobbies of that user.
   c. The data in each line is delimited by commas.
   d. For instance in the line: 2254,reading,coding,swimming,playing soccer,
      i. The user/username is: 2254
      ii. The hobbies are: reading, coding, swimming, and playing soccer
      iii. The number and type of hobbies may differ from one user to another.
2- This file will be your data set that your code has to read to be able to implement a code that does the following:
   a. Finding circles/networks of friends:
      i. In each circle you will report, all the users should share at least **x** number of hobbies
      ii. **x** is a variable that a user can input to the program.
      iii. Circles of friends should be written to a file named <u>circles.txt.</u>
      iv. Each line should have the usernames in the circle/network you found, tab character, and list of shared hobbies.
      v. for example, a line may look like: 2254,552,1258   reading,swimming,hiking
   b. Finding popular users:
      i. Popularity is based on being part of at least **y** circles/networks.
      ii. **y** can be variable that a user can input to the program.
      iii. Popular users should be written to a file named <u>popular.txt.</u> Each user and how many circles/networks the user belongs to, should be in separate line and separated by the tab character.
      iv. For instance: 2254   5
      v. This step should occur after step (**a.).**
      vi. **Hint:** You may want to save the circles you found in part (**a.**) in some data structure so that you can us them in this part.

Notes:

1- The project is to be done in groups of 3 or less. Groups should be from same section. Forming groups, if you want to have a group, is the responsibility of students. Therefore, not finding a group is not an excuse not to do the project; you still can do it on your own.

2- You should be developing this project under the Linux machine (the Cloudera virtual machine) you should have installed at the beginning of this semester, without the need to install any special packages or libraries except the default compilers and libraries.

3- Name the solution file facebook.py, facebook.java, or facebook.scala.

4- Only one code file should be submitted per group. Your code should start with a block of comment.

5- This comment block has:
   a. Students names, ids, and sections
   b. Instructions on how to run the code in each of the steps above, i.e. step a., b.

6- You have to make sure that your code runs error-free, especially compilation errors. **We will not debug or fix any errors.** Very low score is expected in this case.

7- Be careful about the Path names. Always assume current folder/directory.

8- The command to run your code would be similar to: python2.6 facebook.py 5 6
   a. 5 refers to the x in step **a.**, and 6 refers to y in step **b.**

9- Copying and cheating will have serious consequences. So, avoid that.

10- Due date is: 9/14/2016

Good Luck!