

# CptS 415 / Assignment-05

## 1. MapReduce

---

Facebook updates the “common friends” of you and response to hundreds of millions of requests every day.

The friendship information is stored as a pair: `(Person, [List of Friends])` for every user in the social network.

Write a MapReduce program *to return a dictionary of common friends* of the form:

```
user_pair, list_of_common_friends = \
    (user_i, user_j),
    [
        list(
            #Common Friends of user_i and user_j
        )
    ]

# for all pairs of i and j who are friends.
```

Python

The order of i and j you returned should be the same as the lexicographical order of their names.

You need to give the pseudo-code of a main function, and both Map() and Reduce() function. Specify the key/value pair and their semantics (what are they referring to?).

### Solution

This is easy. I'm going to write it in Python.

```
# assumptions:
# - partitions: friends in the each person's list are distributed
#   across one or more partitions

user_i
user_j

# get min of the lists
```

Python

```
iterable_friends = user_i.friends if user_i.friends.count < user_j.friends.  
  
common_friends = list()  
for i in range(iterable_friends):  
    pass
```

## 2. Graph Parallel Models: MR for Graph Processing

---

a.

Consider the common friends problem in Problem 1.a. We study a “2-hop common contact problem”, where a list should be returned for any pair of friends  $i$  and  $j$ , such that the list contains all the users that can reach both  $i$  and  $j$  within 2 hops. Write a MR algorithm to solve the problem and give the pseudo code.

### Solution

b.

We described how to compute distances with mapReduce. Consider a class of  $d$ -bounded reachability queries as follows. Given a graph  $G$ , two nodes  $u$  and  $v$  and an integer  $d$ , it returns a Boolean answer **YES**, if the two nodes can be connected by a path of length no greater than  $d$ . Otherwise, it returns **NO**. Write an MR program to compute the query  $Q(G, u, v, d)$  and give the pseudo code.

Provide necessary correctness and complexity analysis.

### Solution

## 3. Hadoop

---

Hadoop Program:

The attached CSV file contains hourly normal recordings for temperature and dew point temperature at Asheville Regional Airport, NC, USA. *The unit of measurement is in tenths of a degree Fahrenheit*. For example, 344 is 34.4 F.

Write a program using Hadoop to compute and output daily average measurements for temperature and dew point temperature.

The daily average measurements *should include measurements for 24-hour period*.  
For example, from:

```
20100101 00:00 (2010, January 1st, 00:00)
```

to:

```
20100101 23:00 (2010, January 1st, 23:00)
```

Output the result in the format shown below - the columns are date and the combined result (separated by comma) of daily temperature and daily dew point temperature:

```
20100101      377.04, 285.58      Plain Text
20100102      378.67, 286.92
....         .... , ....
```

You may write the application in Java, C/C++ or Python language. Provide both source code and compiled code, if applicable, for your program.

## Solution

- First, I need to look up the API for Hadoop in Python/C++.