

Develop an efficient Map reduce algorithm to solve the following real-world problem.

- Find set of common friends between any two users on Facebook. • Facebook has a list of friends ☐ Note that friends are a bi-directional on Facebook.
☐ If I'm your friend, you're mine.
- FB have lots of disk space and they serve hundreds of millions of requests every day. • They pre-compute calculations when they can to reduce the processing time of requests.
- One common processing request is the 'Common Friends' that is you visit someone's profile, you see a list of friends that you have in common.
- This list doesn't change frequently so it'd be wasteful to recalculate it every time you visited the profile.
- Your task is to use map reduce and calculate everyone's common friends once a day and store those results. Later on it's just a quick lookup.
- Assume the friends are stored as Person->[List of Friends]
A -> B C D
B -> A C D E
C -> A B D E
D -> A B C E
E -> B C D

```
In [248]: %%file q1.py
from mrjob.job import MRJob
from mrjob.step import MRStep

class CommonFriendsMR(MRJob):
    def mapper(self, _, line):
        user, friends = line.strip().split('->')
        friendsList = friends.split(' ')

        for i in range(len(friendsList)):
            for j in range(i + 1, len(friendsList)):
                yield tuple(sorted([friendsList[i], friendsList[j]])), user

    def reducer(self, key, values):
        yield key, list(values)

    def steps(self):
        return [
            MRStep(mapper=self.mapper, reducer=self.reducer)
        ]

if __name__ == '__main__':
    CommonFriendsMR.run()
```

Overwriting q1.py

In [249]: `python q1.py file.txt`

```
["B","C"]      ["A"]
["B","D"]      ["A"]
["B","E"]      ["A"]
["C","D"]      ["A","B"]
["C","E"]      ["A"]
["C","F"]      ["B"]
["D","E"]      ["A"]
["D","F"]      ["B"]
```

No configs found; falling back on auto-configuration

No configs specified for inline runner

Creating temp directory C:\Users\DELL\AppData\Local\Temp\q1.DELL.20240224.114445.318526

Running step 1 of 1...

job output is in C:\Users\DELL\AppData\Local\Temp\q1.DELL.20240224.114445.318526\output

Streaming final output from C:\Users\DELL\AppData\Local\Temp\q1.DELL.20240224.114445.318526\output...

Removing temp directory C:\Users\DELL\AppData\Local\Temp\q1.DELL.20240224.114445.318526...

In []: