

ASSIGNMENT DISTRIBUTED SYSTEM

Problem 1 :: Facebook has a list of friends (note that friends are a bi-directional thing on Facebook. If X is friend of Y then Y will be friend of X). We've decided to pre-compute calculations when we can to reduce the processing time of requests. One common processing request can be "You and X have 20 friends in common" feature. When 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 (sure you could use a decent caching strategy, but then I wouldn't be able to continue writing about map reduce for this problem). So your job is that Given a social network like Facebook which is having tens of millions of users, you have to implement a MapReduce program to identify "common friends" among all pairs of users

- Using MapReduce find common friends of Facebook data.
- Performance of same problem when running on single machine and when running on clusters(draw graph which shows relation b/w nodes and performance).

Links:

- <https://evantamle.wordpress.com/2016/03/14/implement-finding-common-friend-with-map-reduce/>
- <http://stevekrenzel.com/finding-friends-with-mapreduce>