

CS5542 Big Data Apps and Analytics

LAB ASSIGNMENT #4

Lab Submission before 2016/2/17 (W) 11:59PM.

Submit a short lab report including screenshots and source code & data to GitHub. Post your GitHub URL through Lab 4 form at <http://goo.gl/forms/Q5djVdlOzN>

1. Hadoop MapReduce Algorithm
Implement MapReduce algorithm for finding Facebook common friends problem and run the MapReduce job on Apache Hadoop. Write a report including your algorithm and result screenshots.
2. Smartphone/Watch Application
Implement a smartwatch/smartphone application using existing speech services/image services (e.g., IBM Alchemyapi, Face++) related to your project.

NOTE:

Finding Facebook common friends: Facebook has a list of friends (note that friends are a bi-directional thing on Facebook. If I'm your friend, you're mine). They also have lots of disk space and they serve hundreds of millions of requests everyday. They've decided to pre-compute calculations when they can to reduce the processing time of requests. One common processing request is the "You and Joe have 230 friends in common" feature. When you visit someone's profile, you see a list of friends that you have in common. We're going to use MapReduce so that we can calculate everyone's common friends once a day and store those results. Later on it's just a quick lookup. We've got lots of disk, it's cheap.

Example (What is the Key/Value Pair?)

Assume the friends are stored as Person->[List of Friends], our friends list is then:

A -> B C D
B -> A C D E
C -> A B D E
D -> A B C E
E -> B C D

The result after reduction is:

(A B) -> (C D)
(A C) -> (B D)
(A D) -> (B C)
(B C) -> (A D E)
(B D) -> (A C E)
(B E) -> (C D)
(C D) -> (A B E)
(C E) -> (B D)
(D E) -> (B C)

When D visits B's profile, we can quickly look up (B D) and see that they have three friends in common, (A C E).

List of APIs

- **IBM Alchemyapi:**
<https://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/services-catalog.html>
- **Face++:** <http://www.faceplusplus.com/api-overview/>
- **More:**
https://docs.google.com/spreadsheets/d/1exidX76xuwQWifzgdFE_niLF8iEt96dTpvOpxWylJnc/edit#gid=1328142336