Report for Lab Assignment 4

**Question 1** : 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.

**Description**:

used mapreduce so that we can calculate everyone's common friends once a day and store those results.

Given Input as :

A B C D

B A C D E

C A B D E

D A B C E

E B C D

Each line will be an argument to a mapper. For every friend in the list of friends, the mapper will output a key-value pair. The key will be a friend along with the person. The value will be the list of friends. The key will be sorted so that the friends are in order, causing all pairs of friends to go to the same reducer.

For map(A -> B C D) :

(A B) -> B C D

(A C) -> B C D

(A D) -> B C D

For map(B -> A C D E) : (Note that A comes before B in the key)

(A B) -> A C D E

(B C) -> A C D E

(B D) -> A C D E

(B E) -> A C D E

And so on…

Before we send these key-value pairs to the reducers, we group them by their keys and get:

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

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

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

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

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

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

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

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

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

Each line will be passed as an argument to a reducer. The reduce function will simply intersect the lists of values and output the same key with the result of the intersection. For example, reduce((A B) -> (A C D E) (B C D)) will output (A B) : (C D) and means that friends A and B have C and D as common friends.

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)

Now 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).

The ouput looks like:

AB CD

AC BD

AD BC

BC ADE

BD ACE

BE CD

CD ABE

CE BD

DE BC