**Exam-1 submission**

**Task1:**

**Introduction:**  Facebook being one of the most popular used social media applications in todays world and generated terabytes of data every day. The task is to find the mutual friends for all the input pairs same as Facebook using a Big data technology Hadoop Map and reduce algorithm.

**Objective:** To create a Hadoop map reduce algorithm to find the mutual friends for the given input pairs.

**Approach:**

The below flow diagram shows the flow of data between Mapper and reducer Phase.

**A picture containing computer

Description automatically generated**

**1.Mapper class:**

1. initially each line of the input file is taken and splitter up in to two columns as user\_accounts and mutuals based on the symbol ‘->’.
2. Then for each mutual friend in the if account\_user compared with(mutual)<0 then the account\_user,mutual is returned as a pair or mutual,account\_user is returned as a pair.
3. The output of the mapper phase will be with pairs along with one of their mutual friends. The output is written into conf.

A screenshot of text

Description automatically generated

A screenshot of a social media post

Description automatically generated

**Reducer class:**

1. In the reducer Phase each output from the conf is taken.
2. Hash functions are used to eliminate the duplicates. Whereas string builder functions are used to store the output result.
3. Each pair is compared, and the non-duplicate mutual friends is added to the string builder
4. Finally, the output is stored into the result and is written to conf.

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

**Main class**

1. In the main class the all the setup is being done like the configuration and job setup and also checking whether the columns are split into 2 or not.
2. All the class names, setting up mapper and reducer classes.
3. Specifying the input and output file formats.

**A screenshot of a cell phone

Description automatically generated**

**A screenshot of a cell phone

Description automatically generated**

**Workflow:**

1. Initially a java project named FbMutualFriends is created in the eclipse and a class named FbMutualFriends is created.
2. All the external libraries of Hadoop are added
3. A jar file named FbMutualFriends is created from the FbMutualFriends class

1.This is the input file which is saved as input.txt in the local system with the following values.

**A screenshot of a cell phone

Description automatically generated**

2.This is viewing the input file in command prompt.A picture containing knife, table

Description automatically generated

3.Created a directory named exam question1 in Hdfs.

i. Imported this input.txt file into the Hdfs.

ii. Now performing the Fb Mutual friends Hadoop Map reduce algorithm on the input file using the FbMutualFriends jar and storing it in the question1output directory.

A screenshot of a cell phone

Description automatically generated

4.Sucessful execution of Map and reduce phase.

A screenshot of a cell phone

Description automatically generated

5.Output of the Fb mutual friends in the command prompt

A screenshot of a cell phone

Description automatically generated

6.Ouput visualization in the Hue.

A screenshot of a cell phone

Description automatically generated

**Task4**

1. Facebook being one of the most popular used social media applications in today’s world and generated terabytes of data every day. The task is to find the mutual friends for all the input pairs same as Facebook using a Big data technology Hadoop Map and reduce algorithm.
2. As people are extensively using the Facebook newer accounts have been created and people now a days are more willing to socialize. So, with the help of Big data technology Hadoop Map reduce we can find the mutual between two individuals As It is highly scalable and the processing is very fast and efficient even for the high volumes of data and people can find whether they know the person or not with the help of mutual contacts.
3. I have done the Task1 hadoop map and reduce algorithm for finding Face book mutual friends.
4. The challenge I faced was to eliminate the duplicate key value pairs in the reducer class but later came up with the idea of hash function.
5. Everyone in the team have divided the tasks equally and individually working on them but incase of facing challenges we as a team are to figure that out.

**References:**

**1.** <https://umkc.app.box.com/s/xk4jj0do3p7fa3swx6utcuazx3wny8j8>

**2.** <https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html>