Name: Keerti **CS6350 :Big data Management Analytics and Management – Assignment 1**

NetID: KXK190012

Question 1: MutualFriendsQ1.java

Question 2: UserTopTenFriendsQ2.java

Question 3: InMemoryJoinMutualFrdsDoB.java

Question 4: ReduceSideJoinTopTenMFMaxAge.java

**Steps:**

1. Created a directory as /keerti/input in hdfs
   1. Command : hadoop dfs -mkdir /Keerti
   2. hadoop dfs -mkdir /keerti/input
2. Using sftp, transfer the input files soc-LiveJournalAdj1.txt and userdata.txt and the jar files of the above mentioned java files
   1. Command: put C:/soc-LiveJournalAdj1.txt ---same way for all the other files
3. Then move the input files from the server to the hdfs directory
   1. Command: hdfs dfs -put userdata.txt /Keerti/input/

**Now execute each jar files with the input file(s)**

**Question 1: Write a MapReduce program in Hadoop that implements a simple “Mutual/Common friend list of two friends".**

**Command**: hadoop jar MutualFriendsQ1.jar /keerti/input/soc-LiveJournalAdj1.txt /keerti/outputQ1

To see the output :

hdfs dfs -get /keerti/input/outputQ1

cd outputQ1

ls

cat part-r-00000

Output:

(0, 1) : 520

(20, 28193) : 1

(1, 29826)

(6222, 19272) : 19263,19280,19281,19282

(28041, 28056) : 6245,28054,28061

**Question 2: Find friend pairs whose common friend number are within the top-10 in all the pairs.**

**Command**: hadoop jar UserTopTenFriends.jar /keerti/input/soc-LiveJournalAdj1.txt /keerti/outputq2\_01/keerti/outputQ2

To see the output :

hdfs dfs -get /keerti/input/outputQ2

cd outputQ2

ls

cat part-r-00000

Output files are sent separately.



**Question 3: Given any two Users (they are friend) as input, output the list of the names and the date of birth (mm/dd/yyyy) of their mutual friends.**

**Command:** hadoop jar InMemoryJoinMutualFrdsDoB.jar 0 1 /keerti/input/soc-LiveJournalAdj1.txt temp\_output1 /keerti/input/userdata.txt /keerti/outputQ3

To see the output :

hdfs dfs -get /keerti/input/outputQ3

cd outputQ3

ls

cat part-r-00000



**Question 4:**

**Step 1: Calculate the maximum age of the direct friends of each user.**

**Step 2: Sort the users based on the calculated maximum age in descending order as described in step 1.**

**Step 3. Output the top 10 users from step 2 with their address and the calculated maximum age.**

**Command:** hadoop jar ReduceSideJoinTopTenMFMaxAge.jar /keerti/input/soc-LiveJournalAdj1.txt /keerti/input/userdata.txt /keerti/temp\_op1 /keerti/temp\_op2 /keerti/outputQ4

To see the output :

hdfs dfs -get /keerti/input/outputQ4

cd outputQ4

ls

cat part-r-00000

