**CS 6350.001 Big Data Management and Analytics**

**Instructor: Prof. Latifur Khan**

**Homework 1**

Submitted By:

**Shyam Patharla (sxp178231)**

Given below are the steps for running the code for all the questions.

**Running Code for Question 1**

Jar File: MapReduce.jar

Class: Part1

1. Create a directory on HDFS and put the input files:

> hdfs dfs -mkdir /test

> hdfs dfs -put <location of soc-LiveJournal1Adj.txt on PC> /test

> hdfs dfs -put <location of userdata.txt PC> /test

2. Delete the output directory if it already exists:

> hdfs dfs -rm -r /test/out

3. Run the jar file:

> hadoop jar <Location of MapReduce.jar on PC> Part1 /test/soc-LiveJournal1Adj.txt /test/out

4. Read the output file:

> hdfs dfs -cat /test/out/part-r-00000

5. To get output for specific pairs, run:

> hdfs dfs -cat /test/out/part-r-00000 | grep "<userid\_1>,<userid\_2> <press ctrl+v> <press tab>"

**Output:**

0,1 5,20

20,28193 1

1,29826

6222,19272

28041,28056

**Running Code for Question 2**

Input Files:

1. soc-LiveJournal1Adj.txt

Jar File: MapReduce.jar

Class: Part2

1. Delete the output directories if they already exist:

> hdfs dfs -rm -r /test/out1

> hdfs dfs -rm -r /test/out2

2. Run the jar file:

> hadoop jar <Location-of-MapReduce.jar-on-local-PC> Part2 /test/soc-LiveJournal1Adj.txt /test/out1 /test/out2

3. Read the output file:

> hdfs dfs -cat /test/out2/part-r-00000

**Output**

16539,40423 64 1357,1918,3403,12781,14920,16538,20562,32085,32574,34977,35612,36635,40419,40434,40455,40497,40426,40429,40430,40432,40433,40435,40436,40439,40440,40441,40442,40443,40445,40448,40451,40452,40453,40458,40459,40461,40462,40464,40466,40469,40470,40471,40473,40474,40475,40476,40478,40482,40483,40484,40485,40486,40487,40488,40489,40490,40491,40492,40493,40494,40495,40496,40498,40499

3610,3634 61 1725,3611,3612,3613,3614,3616,3617,3619,3621,3624,3625,3626,3627,3628,3629,3630,3631,3632,3633,3635,3636,3637,3638,3639,3640,3641,3642,3643,3644,3645,3646,3648,3649,3650,3651,3652,3653,3655,3657,3658,3659,3660,3661,3662,3663,3664,3665,3666,3667,3668,3669,3670,3671,3672,3673,3674,3675,3677,3679,3681,3683

31482,31545 60 31481,31483,31484,31486,31490,31491,31493,31495,31501,31508,31510,31511,31521,31523,31529,31531,31540,31547,31556,31561,31565,31573,31575,31485,31488,31489,31496,31497,31498,31499,31500,31502,31507,31509,31512,31514,31520,31522,31524,31525,31532,31534,31535,31536,31539,31543,31544,31548,31557,31558,31564,31566,31569,31570,31576,31577,31578,31581,31582,31583

41851,41903 58 10014,10055,10059,10066,13947,19365,30219,36696,36708,36713,36720,36728,36735,36742,36807,36827,36852,36870,36894,36927,37011,37035,37038,37063,37132,37179,37218,37247,37269,37303,37314,37346,37374,37430,37441,37448,37493,37532,37561,37674,37722,37765,39235,39237,41888,18912,36923,41935,44049,44101,44137,44178,44201,45022,43018,44149,44191,44199

2689,29425 55 1688,2708,2667,2668,16713,21555,21562,27286,29428,29430,29431,29432,29433,29434,29435,29436,29437,29438,29442,29444,29445,29447,29449,29450,29451,29452,29454,29459,29461,29464,29465,29466,29468,29469,29470,29473,29474,29476,29477,29478,29479,29482,29485,29487,29492,29494,29495,29496,29499,29504,29505,29506,29508,29509,29513

47034,47047 55 47035,47038,47040,47041,47042,47043,47044,47045,47046,47050,47051,47052,47054,47055,47056,47057,47058,47059,47060,47062,47063,47066,47067,47068,47070,47071,47073,47074,47075,47076,47077,47078,47079,47082,47083,47084,47085,47086,47087,47088,47089,47090,47091,47096,47097,47098,47099,47100,47102,47103,47104,47108,47110,47113,47109

32867,32869 54 24792,32858,32861,32862,32863,32864,32865,32866,32868,32860,32870,32871,32872,32873,32874,32876,32878,32880,32881,32883,32884,32888,32889,32891,32894,32895,32897,32898,32899,32900,32901,32903,32905,32906,32908,32911,32912,32913,32914,32915,32922,32931,32934,32937,32938,32940,32942,32945,32930,32932,32933,32939,32941,32943

4069,16215 52 2084,3004,3525,3958,3961,4041,4048,4070,4281,6279,7957,10287,10715,10898,10987,11005,11577,16192,16196,16197,16198,16200,16204,16205,16208,16211,16212,16214,660,2660,4159,8508,10408,16218,16229,22091,22102,22118,22158,22883,23211,27410,30811,30864,42719,42810,42811,43238,43709,47669,48067,48920

7983,8003 52 4285,4752,5052,4707,6284,11693,12066,13290,18084,21847,21849,21850,21853,21857,21858,21859,21860,21861,21862,21863,21864,21865,21866,21867,21868,21872,21873,21874,21878,21880,21883,21885,21886,21887,21891,21894,21895,21897,21898,21899,21900,21901,21902,21903,21904,21905,21907,21909,21910,21911,21912,21916

4425,4447 52 1384,4396,4398,4420,4426,4431,4433,4441,4442,4451,4455,4471,4482,5249,7014,11756,23744,32352,32356,32360,32386,32394,32408,32417,32418,32423,32425,32440,32728,32729,32730,32737,32738,32740,32741,32754,32761,32766,32774,32779,32782,32784,32801,32806,32812,32814,35008,35182,35186,35261,38000,40136

**Running Code for Question 3**

Input Files:

1. soc-LiveJournal1Adj.txt

2. userdata.txt

Jar File: MapReduce.jar

Class: Part3

1. Delete the output directories if they already exist:

> hdfs dfs -rm -r /test/out1

> hdfs dfs -rm -r /test/out2

2. Run the jar file:

> hadoop jar <Location-of-MapReduce.jar-on-local-PC> Part2 /test/userdata.txt /test/soc-LiveJournal1Adj.txt /test/out1 /test/out2 <userid\_1> <userid\_2>

3. Read the output file:

> hdfs dfs -cat /test/out2/part-r-00000

**Output**

Sample output:

0,1 [Juan Figueroa: Columbia,Beth Watson: Lake Worth]

**Running Code for Question 4**

Input Files:

1. soc-LiveJournal1Adj.txt

2. userdata.txt

Jar File: MapReduce.jar

Class: Part4

1. Delete the output directories if they already exist:

> hdfs dfs -rm -r /test/out1

> hdfs dfs -rm -r /test/out2

2. Run the jar file:

> hadoop jar <Location-of-MapReduce.jar-on-local-PC> Part2 /test/userdata.txt /test/soc-LiveJournal1Adj.txt /test/out1 /test/out2

3. Read the output file:

> hdfs dfs -cat /test/out2/part-r-00000

**Output:**

43227 Ray,4803 Cliffside Drive,Etna,New York,0

43327 Evelyn,4370 Freshour Circle,San Antonio,Texas,0

38556 Richard,3091 Hickory Lane,Washington,Washington DC,0

38314 Jacob,4200 Arthur Avenue,Sterling,Illinois,0

21236 Debra,4944 Church Street,New York,New York,0

41230 Dustin,3451 Brentwood Drive,Austin,Texas,0

41605 Nancy,2754 Briarwood Drive,Camden,New Jersey,0

37782 Ellen,2799 Wright Court,Hamilton,Alabama,0

37764 Everett,4127 Hawks Nest Lane,Stlouis,Missouri,0

37739 Bonnie,3241 Armbrester Drive,Santa Monica,California,0

37656 Amanda,4975 Cedarstone Drive,Toledo,Ohio,0

41932 Brooke,4522 Pride Avenue,Brooklyn,New York,0

7607 Donald,3344 Neville Street,Bloomington,Indiana,0

818 Forrest,3223 Olen Thomas Drive,Wichita Falls,Texas,0

37489 Anthony,4838 Bloomfield Way,Rockland,Maine,0