

Assignment 2

Ishan Desai

Net ID: ijd140030

Q1. Write a MapReduce program in Hadoop that implements a simple "Mutual/Common friend list of two friends".

Answer

File name: Eclipse Files/Assignment2_1.scala

To run: Import the file in your Scala IDE for eclipse. Make sure you add all the external jars required. Save your soc-LiveJournal1Adj.txt in your local hdfs system in the folder “input”. And Run your code.

To view output: Output will be displayed on console.

OR

File name: Spark-Shell Files/Assignment2_2.scala

To run: Save your soc-LiveJournal1Adj.txt in your local hdfs system in the folder “input” . And Run the code line by line in your spark-shell.

To view output: Output will be displayed on console.

Output for given input points

- 0,1 20,5
- 20,28913 1
- 1, 29826
- 6222, 19272 19263,19280,19281,19282
- 28041, 28056 6245,28054,28061

Q2. Given any two Users (they are friend) as input, output the list of the user id of their mutual friends.

Answer

File name: Eclipse Files/Assignment2_2.scala

To run: Import the file in your Scala IDE for eclipse. Make sure you add all the external jars required. Save your soc-LiveJournal1Adj.txt in your local hdfs system in the folder "input". And Run your code.

To view output: Output will be displayed on console.

OR

File name: Spark-Shell Files/Assignment2_2.scala

To run: Save your soc-LiveJournal1Adj.txt in your local hdfs system in the folder "input". And Run the code line by line in your spark-shell.

To view output: Output will be displayed on console.

Output for some example input points

- 6222, 19272 19263,19280,19281,19282
- 28041, 28056 6245,28054,28061

Q3. 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.

Answer

File name: Eclipse Files/Assignment2_3.scala

To run: Import the file in your Scala IDE for eclipse. Make sure you add all the external jars required. Save your soc-LiveJournal1Adj.txt in your local hdfs system in the folder "input". And the file userdata.txt in you local hdfs system in the folder "userdata".And Run your code.

To view output: Output will be displayed on console.

OR

File name: Spark-Shell Files

To run: Save your soc-LiveJournal1Adj.txt in your local hdfs system in the folder "input". And the file userdata.txt in you local hdfs system in the folder "userdata". And Run the code line by line in your spark-shell.

To view output: Output will be displayed on console.

Output for some example input points

- 0,1 [Juan:9/12/1991,Beth:8/27/1970]
- 6222, 19272 [Betty:6/10/1958,Hilton:6/22/1965,Charles:2/2/1986,Willie:5/18/1965]
- 28041, 28056 [Stephanie:9/21/1965,Carl:11/10/1981,Taylor:8/17/1973]

Q4. Reduce-side join and Job Chaining

Answer

File name: Eclipse Files/Assignment2_4.scala

To run: Import the file in your Scala IDE for eclipse. Make sure you add all the external jars required. Save your soc-LiveJournal1Adj.txt in your local hdfs system in the folder “input”. And the file userdata.txt in you local hdfs system in the folder “userdata”.And Run your code.

To view output: Output will be displayed on console.

OR

File name: Spark-Shell Files/Assignment2_4.scala

To run: Save your soc-LiveJournal1Adj.txt in your local hdfs system in the folder “input”. And the file userdata.txt in you local hdfs system in the folder “userdata”. And Run the code line by line in your spark-shell.

To view output: Output will be displayed on console.

Output

Debra,4670 Elmwood Avenue,Scottsdale,Arizona,86
 Freda,3587 Cerullo Road,Louisville,Kentucky,86
 Byron,2221 Ethels Lane,Lakeland,Florida,86
 John,2201 Valley Street,Riverside Burlington,New Jersey,86
 Sue,1208 May Street,Pittsburg,Kentucky,86
 James,3284 Isaacs Creek Road,Chandlerville,Illinois,86
 Loretta,2867 Graystone Lakes,Macon,Georgia,86
 Chad,2049 Christie Way,Andover,Massachusetts,86
 William,1510 Lowndes Hill Park Road,Bakersfield,California,86
 Mary,598 Fairmont Avenue,Kansas City,Missouri,86
 Pamela,575 Bassell Avenue,Hot Springs,Arkansas,86

Q5. Write a scala program using Apache Spark to find Top N bi-gram

Answer

File name: Eclipse Files/Assignment2_5.scala

To run: Import the file in your Scala IDE for eclipse. Make sure you add all the external jars required. And Run your code.

To view output: Output will be displayed on console.

OR

File name: Spark-Shell Files/Assignment2_4.scala

To run: Run the code line by line in your spark-shell.

To view output: Output will be displayed on console.

Input: "Alice is testing spark application. Testing spark is fun"

Output: ("test" , "spark") 2

Note: I have also attached the eclipse project file for your convenience.