**BIG DATA PROGRAMMING**

**PROJECT-2**

**Case1-FacebookMutualFriendsUsingSpark**

**Team Members and collaboration:**

Roshini varada --**Facebook Mutual Friends Using Spark**

Sarika Reddy Kota -- **Spark Data Frames**

Pallavi Arikatla – **Spark Streaming**

Zakari, Abdulmuhaymin –**Spark Graph Frames**

**Idea: (Question4- Part-1)**

To identify the Common Friends of any two people in a social network with the help of map-reduce using spark. In spark the MapReduce algorithm is hundred times faster than the Hadoop MapReduce and is efficient. In spark we use the transformations such as group by, map, flat-map and reduce to perform the operation.

**Usage or the real time scenario: (Question4- Part-2)**

Social networking sites identifies the common friends between people. When a person visits another person profile then they can see the mutual contacts

**Example:**

If P1 and P2 has P3,P4 as their common friends, If P1 visits P2's profile P1 can identify P3,P4 as their common friends and vice versa.

**Approach and solution: (Question4- Part-3)**

**Flow Chart or Pictorial Representation:**

1.The identification of mutual friends for a unit takes place in 5 phases. They are

a.Input

b.GroupBy

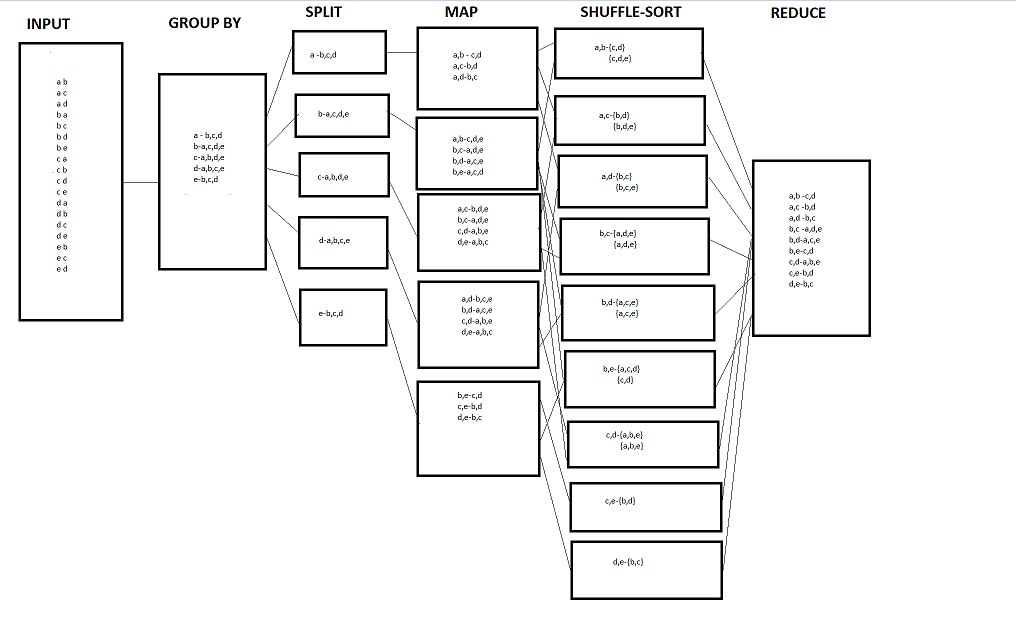
c.Split

d.Map

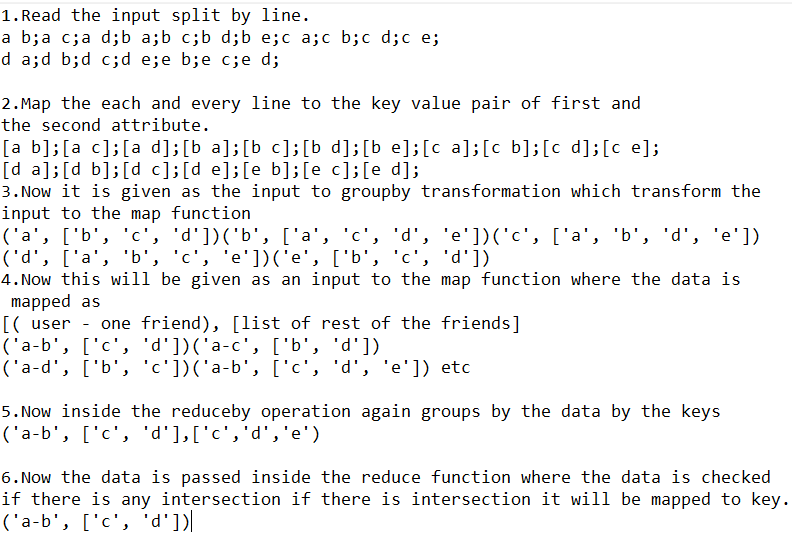
e.Shuffle

f.Reduce

The below diagram represents how the input transforms in each phase.



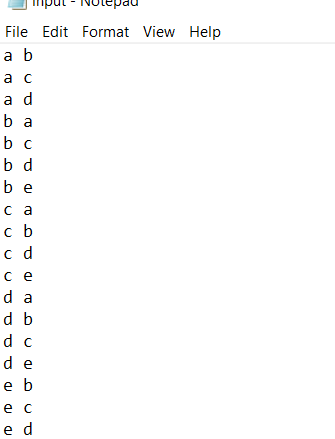
**c)Algorithm For the approach**



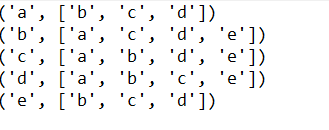
**Implementation**

**a)Implementing mapreduce using spark with a simple example.**

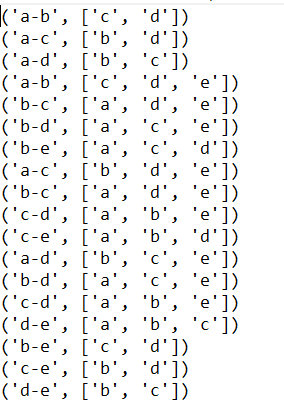
1.Here the input taken is the first attribute indicates the user and the second attribute indicates the friend of the user.



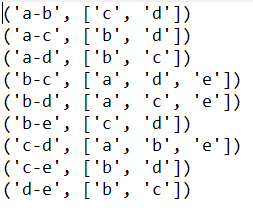
2.After this the users are mapped with their friends using groupby key.



3.After grouping the data is split and the mapping is done for each user. Here each user is paired with the other user and the and the list of the other friends is formed.

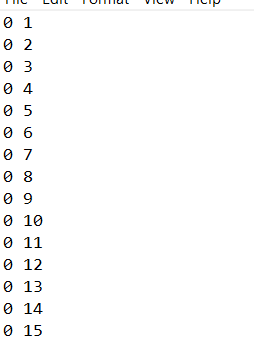


4.Then the data is reduced using the reduce by operation after the shuffle sorting.

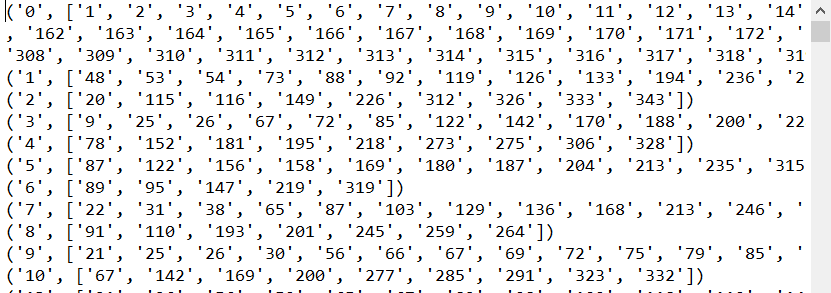


**b)Implementing mapreduce using spark data set given.**

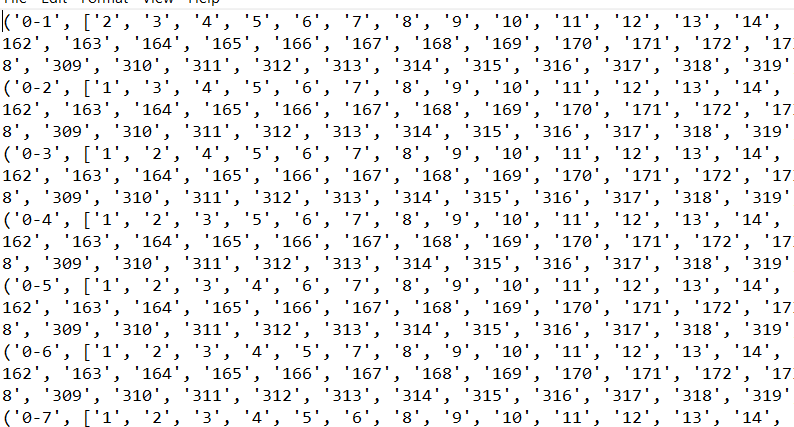
1.The facebook\_combined data is taken as the input.



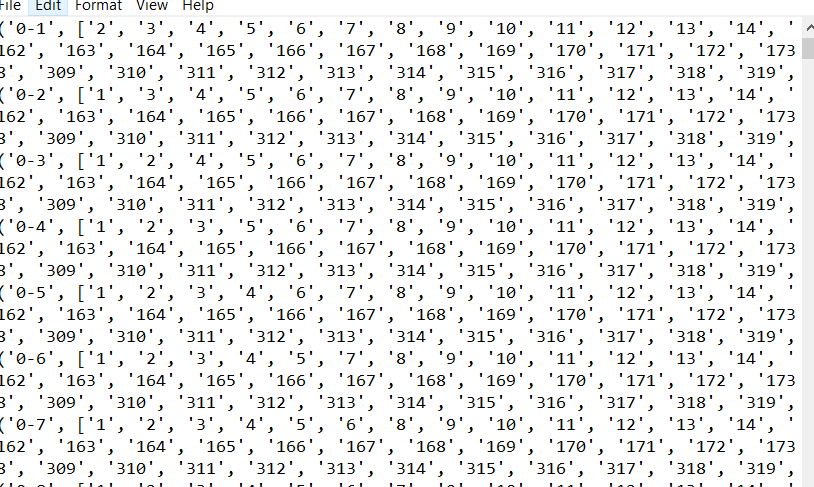
2.Output after grouping.



3.Output after mapping



4.The final output after reduction.



**Challenges faced (Question4- Part-4)**

1.The coding part was smooth without any difficulties.

**Integration and Milestones (Question4- Part-5)**

Since all the cases are independent from each other there are no difficulties faced in the integration part.

**Links**

**Github Link**

<https://github.com/RoshiniVarada/BDP_Project2/tree/master/CASE-1>

**Link for Sourcecode**

<https://github.com/RoshiniVarada/BDP_Project2/tree/master/CASE-1/Sourcecode>

**Link for YouTube Video**

**Team Members Links**

**Use-case1 -FacebookMutualFriendsUsingSpark**

Wiki-link- <https://github.com/RoshiniVarada/BDP_Project2/wiki/CASE1>

**Use-case2-Spark Data Frames**

Wiki-link- <https://github.com/RoshiniVarada/BDP_Project2/wiki/CASE2>

**Use-case3- Spark Streaming**

Wiki-link- <https://github.com/RoshiniVarada/BDP_Project2/wiki/CASE3>

**Use-case4-Spark-Graphx**

Wiki-link- <https://github.com/RoshiniVarada/BDP_Project2/wiki/CASE4>

Submitted by

Roshini Varada(16302628)