Name: Panja, Kumara Satya Gopal

UMKC ID: 16222771

Class ID: 32

Tutorial Lab-2

Objective:

To write a spark program with at least two spark transformations and two spark actions

Technologies Used:

Spark, Intellij

Use Case:

Here I took finding mutual Facebook friends as Use case. I wrote spark program for finding mutual friends.

Input:

A-->BCD

B-->ACDE

C-->ABDE

D-->ABCE

E-->BCD

Output:

AB-->CD

AC-->BD

AD-->BC

BC-->ADE

BD-->ACE

BE-->CD

CD-->ABE

CE-->BD

DE-->BC

Transformations Used:

1. Map: It returns a new distributed dataset formed by passing each element of the source through a function func.

2. Intersection: It returns a new RDD that contains the intersection of elements in the source dataset and the argument.

Actions Used:

1. reduce: It aggregates the elements of the dataset using a function func. The function should be commutative and associative so that it can be computed correctly in parallel.

2. for each: It runs a function func on each element of the dataset.

Description:

Here I wrote a spark program for finding mutual friends.

I used Map, Intersection transformations and reduce, for each actions here.

In input, A-->BCD means A is a person who has three friends (B,C,D).

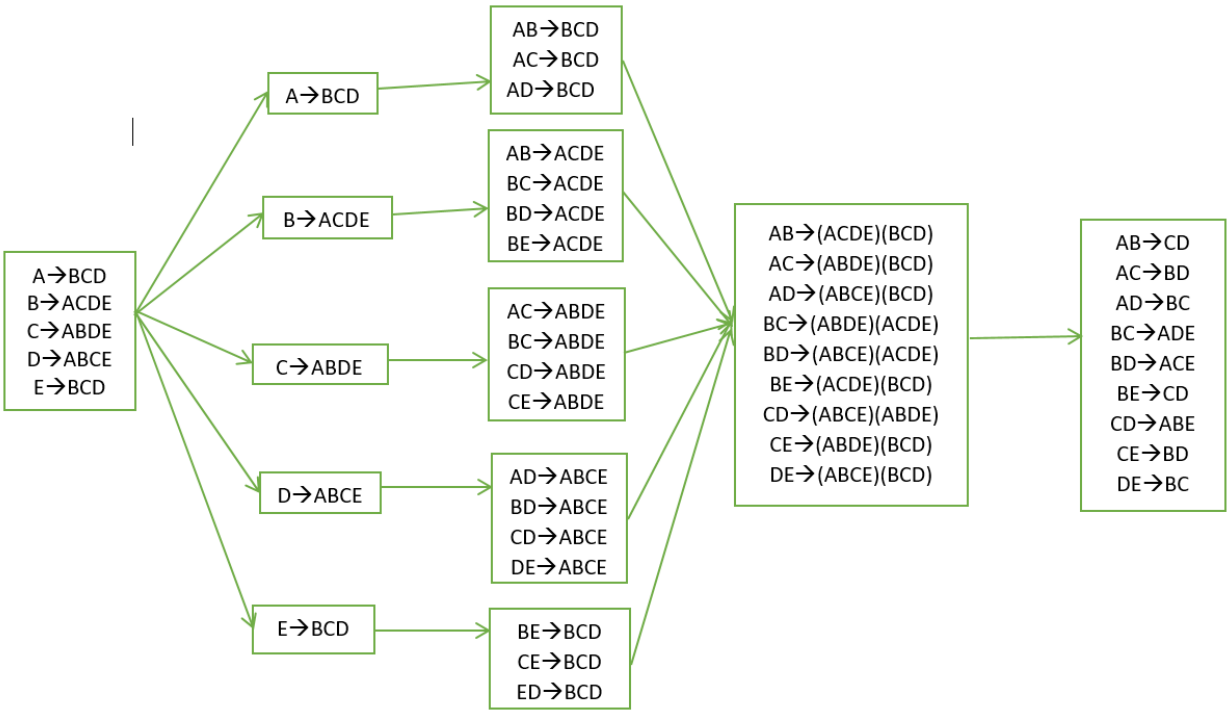
In output, CD-->ABE means C and D have three mutual friends (A,B,E).

It uses map reduce algorithm for finding Mutual friends.

Map Reduce Paradigm:

Input Splitting Mapping Reducing Output

o



Output Screenshots:

```
[cloudera@quickstart ~]$ hadoop fs -ls fbout
Found 2 items
-rw-r--r--   1 cloudera cloudera          0 2016-02-17 18:04 fbout/_SUCCESS
-rw-r--r--   1 cloudera cloudera         66 2016-02-17 18:04 fbout/part-r-00000
[cloudera@quickstart ~]$ hadoop fs -cat fbout/part-r-00000
AB      CD
AC      BD
AD      BC
BC      ADE
BD      ACE
BE      CD
CD      ABE
CE      BD
DE      BC
```

```
3 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
3 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunn
g.
2 INFO input.FileInputFormat: Total input paths to process : 1
2 INFO mapreduce.JobSubmitter: number of splits:1
3 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1455759153571_0003
4 INFO impl.YarnClientImpl: Submitted application application_1455759153571_0003
5 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1455759153571_0003/
5 INFO mapreduce.Job: Running job: job_1455759153571_0003
1 INFO mapreduce.Job: Job job_1455759153571_0003 running in uber mode : false
1 INFO mapreduce.Job:  map 0% reduce 0%
3 INFO mapreduce.Job:  map 100% reduce 0%
1 INFO mapreduce.Job:  map 100% reduce 100%
3 INFO mapreduce.Job: Job job_1455759153571_0003 completed successfully
3 INFO mapreduce.Job: Counters: 49
```

References:

http://spark.apache.org/docs/latest/programming-guide.html#transformations

http://stevekrenzel.com/finding-friends-with-mapreduce