CS 5542 Big Data Applications and Analytics

LAB ASSIGNMENT #4

- Sricharan Pamuru(20)

1. Hadoop MapReduce Algorithm

FileZilla is used to load both the jar file as well as input file.

```
$ hadoop fs -put FB_input.txt FB_input.txt
```

```
$ hadoop fs -put MutualFriends_FB.jar MutualFriends_FB.jar
```

Server is logged into by using Putty

The files are listed and the directory is entered into

```
$ cd SriCharan
$ ls
FB_input.txt   MutualFriends_FB.jar
```

The Hadoop MapReduce program is run.

```
$ hadoop jar MutualFriends_FB.jar MutualFriends FB_input.txt output_FB
16/02/22 00:25:30 INFO client.RMProxy: Connecting to ResourceManager at KC-SCE-C
S5542-1/172.16.2.241:8032
16/02/22 00:25:30 WARN mapreduce.JobResourceUploader: Hadoop command-line option
 parsing not performed. Implement the Tool interface and execute your applicatio
n with ToolRunner to remedy this.
16/02/22 00:25:31 INFO input.FileInputFormat: Total input paths to process : 1
16/02/22 00:25:31 INFO mapreduce.JobSubmitter: number of splits:1
16/02/22 00:25:31 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_14
55690915780_0030
16/02/22 00:25:31 INFO impl.YarnClientImpl: Submitted application application_14
55690915780_0030
16/02/22 00:25:31 INFO mapreduce.Job: The url to track the job: http://KC-SCE-CS
5542-1:8088/proxy/application_1455690915780_0030/
16/02/22 00:25:31 INFO mapreduce.Job: Running job: job_1455690915780_0030
16/02/22 00:25:40 INFO mapreduce.Job: Job job_1455690915780_0030 running in uber
 mode : false
16/02/22 00:25:40 INFO mapreduce.Job:  map 0% reduce 0%
16/02/22 00:25:47 INFO mapreduce.Job:  map 100% reduce 0%
16/02/22 00:25:57 INFO mapreduce.Job:  map 100% reduce 50%
16/02/22 00:26:03 INFO mapreduce.Job:  map 100% reduce 100%
16/02/22 00:26:04 INFO mapreduce.Job: Job job_1455690915780_0030 completed succe
ssfully
```

The output is shown by using the code executed.

```
$ hadoop fs -cat output_FB/part-r-00000
AB        CD
AD        BC
BC        ADE
BE        CD
CD        ABE
DE        BC
$ hadoop fs -cat output_FB/part-r-00001
AC        BD
BD        ACE
CE        BD
```

Therefore, the output of the Hadoop program is used to show mutual friends on Facebook for the given example.