

1. Hadoop MapReduce Algorithm Implement MapReduce algorithm for finding Facebook common friend's problem and run the MapReduce job on Apache Hadoop. Write a report including your algorithm and result screenshots.

Description:

Algorithm used:

Map (Person, Friends List)

{

Step1: pairing the friends, for example D-> A B C E, E-> B C D can be paired as D,E

Step 2: Mapping the common friends of the paired D & E. Here it is DE->B, DE->C.

Step 3: Emit (Paired Friends, Friends of paired friends)

}

Reduce (Paired Friends, Friends of paired friends)

{

Step 1: Reduce the no. of steps .For example DE -> (B,C)

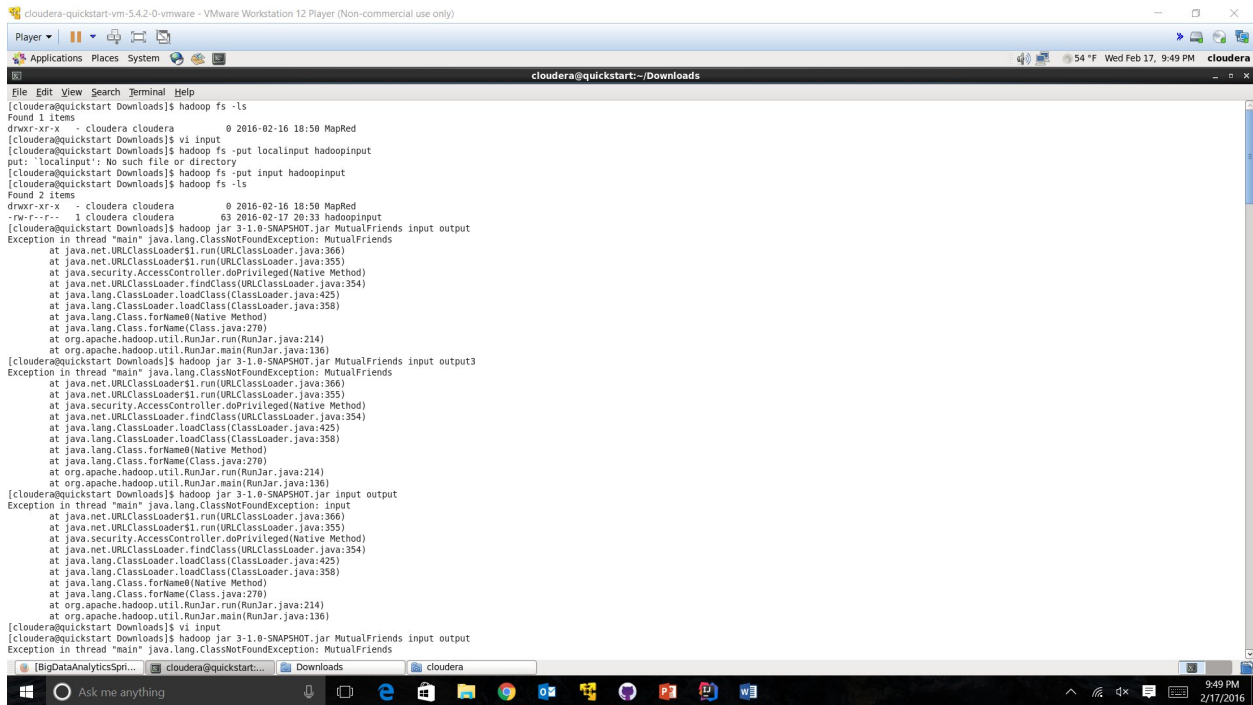
Step 2: Emit (Paired Friends, Common friends)

}

Screenshots:

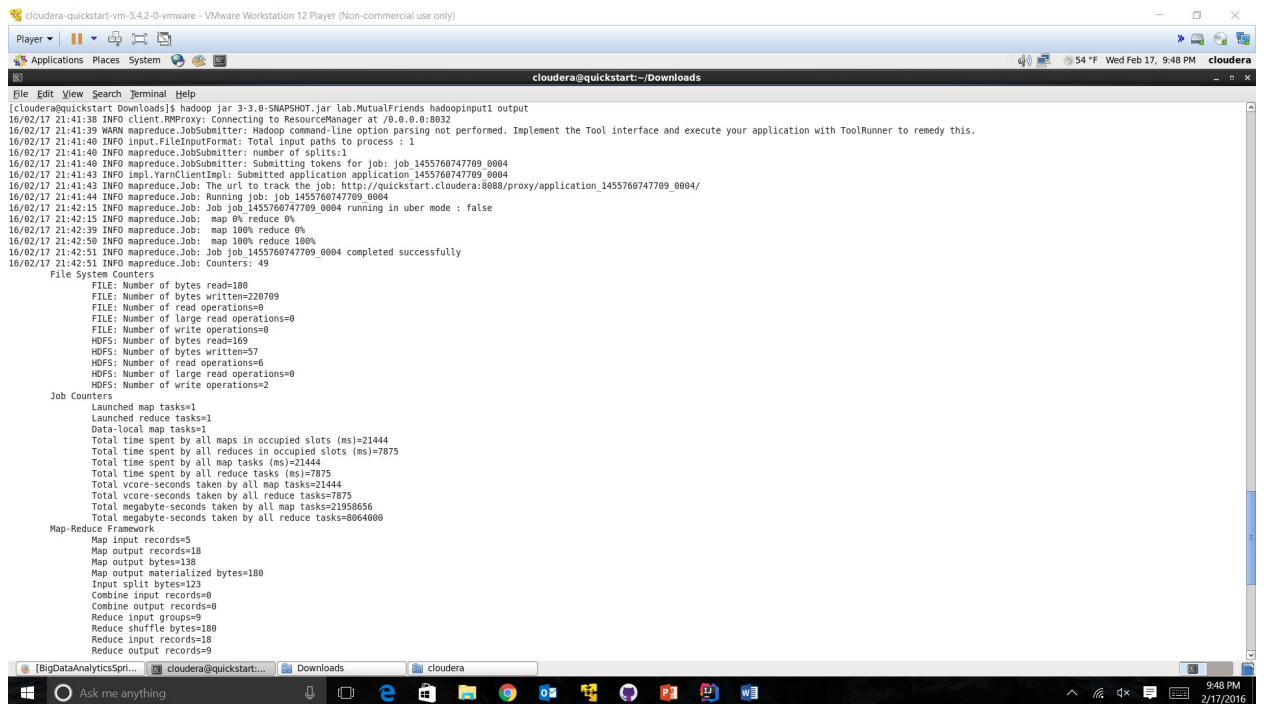
Lab Assignment 4

Varun Chavakula
Class ID: 4



The screenshot shows a VMware Workstation 12 Player window titled "cloudera-quickstart-vm-5.4.2-0-vmware - VMware Workstation 12 Player (Non-commercial use only)". The terminal window is titled "cloudera@quickstart:~/Downloads" and displays the following commands and output:

```
[cloudera@quickstart Downloads]$ hadoop fs -ls
Found 1 items
drwxr-xr-x  - cloudera cloudera      0 2016-02-16 18:50 MapRed
[cloudera@quickstart Downloads]$ vi input
[cloudera@quickstart Downloads]$ hadoop fs -put localinput hadoopinput
put: 'localinput': No such file or directory
[cloudera@quickstart Downloads]$ hadoop fs -put input hadoopinput
[cloudera@quickstart Downloads]$ hadoop fs -ls
Found 2 items
drwxr-xr-x  - cloudera cloudera      0 2016-02-16 18:50 MapRed
-rw-r--r--  1 cloudera cloudera      63 2016-02-17 20:33 hadoopinput
[cloudera@quickstart Downloads]$ hadoop jar 3-1.0-SNAPSHOT.jar MutualFriends input output
Exception in thread "main" java.lang.ClassNotFoundException: MutualFriends
    at java.net.URLClassLoader$1.run(URLClassLoader.java:366)
    at java.net.URLClassLoader$1.run(URLClassLoader.java:355)
    at java.security.AccessController.doPrivileged(Native Method)
    at java.net.URLClassLoader.findClass(URLClassLoader.java:354)
    at java.lang.ClassLoader.loadClass(ClassLoader.java:425)
    at java.lang.ClassLoader.loadClass(ClassLoader.java:358)
    at java.lang.Class.forName0(Native Method)
    at java.lang.Class.forName(Class.java:278)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:214)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
[cloudera@quickstart Downloads]$ hadoop jar 3-1.0-SNAPSHOT.jar MutualFriends input output3
Exception in thread "main" java.lang.ClassNotFoundException: MutualFriends
    at java.net.URLClassLoader$1.run(URLClassLoader.java:366)
    at java.net.URLClassLoader$1.run(URLClassLoader.java:355)
    at java.security.AccessController.doPrivileged(Native Method)
    at java.net.URLClassLoader.findClass(URLClassLoader.java:354)
    at java.lang.ClassLoader.loadClass(ClassLoader.java:425)
    at java.lang.ClassLoader.loadClass(ClassLoader.java:358)
    at java.lang.Class.forName0(Native Method)
    at java.lang.Class.forName(Class.java:278)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:214)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
[cloudera@quickstart Downloads]$ hadoop jar 3-1.0-SNAPSHOT.jar input output
Exception in thread "main" java.lang.ClassNotFoundException: input
    at java.net.URLClassLoader$1.run(URLClassLoader.java:366)
    at java.net.URLClassLoader$1.run(URLClassLoader.java:355)
    at java.security.AccessController.doPrivileged(Native Method)
    at java.net.URLClassLoader.findClass(URLClassLoader.java:354)
    at java.lang.ClassLoader.loadClass(ClassLoader.java:425)
    at java.lang.ClassLoader.loadClass(ClassLoader.java:358)
    at java.lang.Class.forName0(Native Method)
    at java.lang.Class.forName(Class.java:278)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:214)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
[cloudera@quickstart Downloads]$ vi input
[cloudera@quickstart Downloads]$ hadoop jar 3-1.0-SNAPSHOT.jar MutualFriends input output
Exception in thread "main" java.lang.ClassNotFoundException: MutualFriends
```



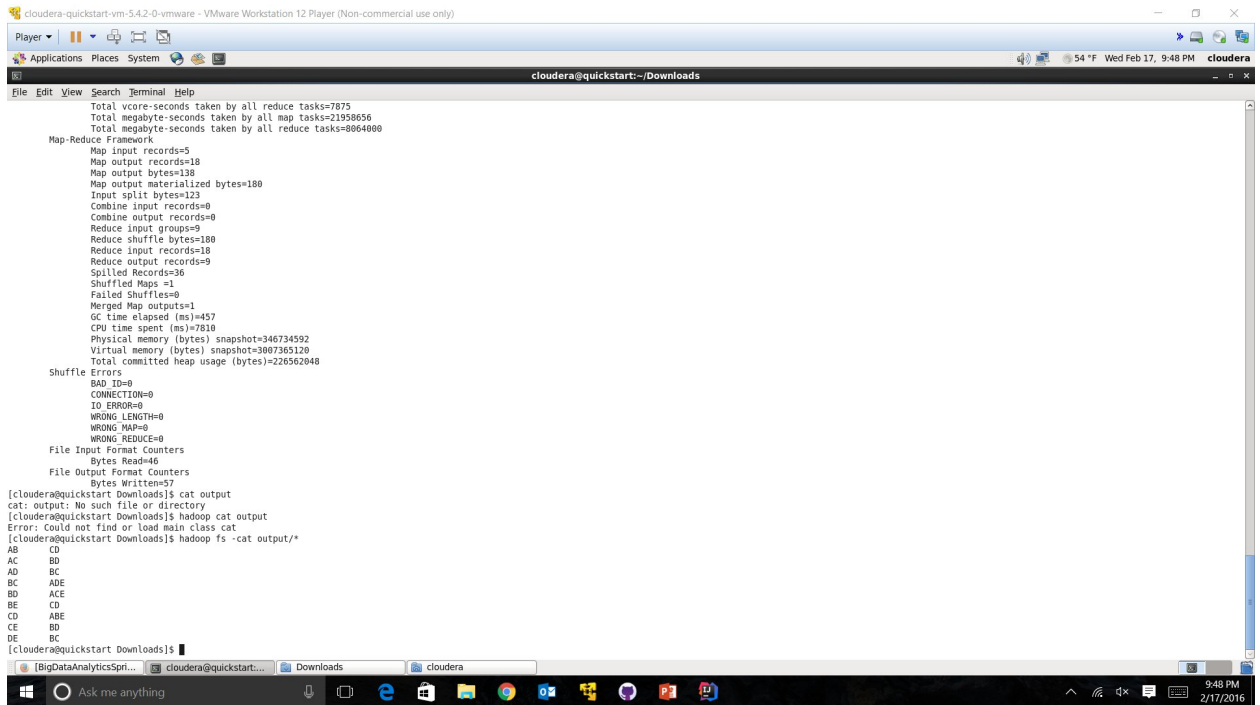
The screenshot shows a VMware Workstation 12 Player window titled "cloudera-quickstart-vm-5.4.2-0-vmware - VMware Workstation 12 Player (Non-commercial use only)". The terminal window is titled "cloudera@quickstart:~/Downloads" and displays the following commands and output:

```
[cloudera@quickstart Downloads]$ hadoop jar 3-3.0-SNAPSHOT.jar lab.MutualFriends hadoopinput output
16/02/17 21:41:30 INFO ClientRMProxy: Connecting to ResourceManager at /0.0.0.0:8032
16/02/17 21:41:39 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
16/02/17 21:41:40 INFO Input.FileInputFormat: Total input paths to process : 1
16/02/17 21:41:40 INFO mapreduce.JobSubmitter: number of splits:1
16/02/17 21:41:40 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1455760747709_0004
16/02/17 21:41:43 INFO impl.YarnClientImpl: Submitted application application_1455760747709_0004
16/02/17 21:41:43 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1455760747709_0004/
16/02/17 21:41:44 INFO mapreduce.Job: Running job: job_1455760747709_0004
16/02/17 21:42:15 INFO mapreduce.Job: Job job_1455760747709_0004 running in uber mode : false
16/02/17 21:42:15 INFO mapreduce.Job: map 0% reduce 0%
16/02/17 21:42:39 INFO mapreduce.Job: map 100% reduce 0%
16/02/17 21:42:50 INFO mapreduce.Job: map 100% reduce 100%
16/02/17 21:42:51 INFO mapreduce.Job: Job job_1455760747709_0004 completed successfully
16/02/17 21:42:51 INFO mapreduce.Job: Counters: 49

File System Counters
  FILE: Number of bytes read=180
  FILE: Number of bytes written=220709
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=169
  HDFS: Number of bytes written=57
  HDFS: Number of read operations=0
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2

Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=21444
  Total time spent by all reduces in occupied slots (ms)=7875
  Total time spent by all map tasks (ms)=21444
  Total time spent by all reduce tasks (ms)=7875
  Total vcore-seconds taken by all map tasks=21444
  Total vcore-seconds taken by all reduce tasks=7875
  Total megabyte-seconds taken by all map tasks=21958656
  Total megabyte-seconds taken by all reduce tasks=8064000

Map-Reduce Framework
  Map input records=5
  Map output records=18
  Map output bytes=138
  Map output materialized bytes=180
  Input split bytes=123
  Combine input records=0
  Combine output records=0
  Reduce input groups=9
  Reduce shuffle bytes=180
  Reduce input records=18
  Reduce output records=9
```



The screenshot shows a VMware Workstation 12 Player window titled "cloudera-quickstart-vm-5.4.2-0-vmware - VMware Workstation 12 Player (Non-commercial use only)". Inside the player, a terminal window titled "cloudera@quickstart: ~/Downloads" displays the output of a Hadoop MapReduce job. The output includes statistics for the Map-Reduce Framework, Shuffle Errors, and File Input/Output. The job completed successfully, with a final output of 57 bytes written to the file "output".

```
cloudera@quickstart:~/Downloads
Total vcore-seconds taken by all reduce tasks=7875
Total megabyte-seconds taken by all map tasks=21958656
Total megabyte-seconds taken by all reduce tasks=8864800
Map-Reduce Framework
  Map input records=5
  Map output records=18
  Map output bytes=138
  Map output materialized bytes=188
  Input split bytes=123
  Combine input records=0
  Combine output records=0
  Reduce input groups=9
  Reduce shuffle bytes=180
  Reduce input records=18
  Reduce output records=9
  Spilled Records=36
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=457
  CPU time spent (ms)=7818
  Physical memory (bytes) snapshot=346734592
  Virtual memory (bytes) snapshot=3007365120
  Total committed heap usage (bytes)=226562848
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=46
File Output Format Counters
  Bytes Written=57
[cloudera@quickstart Downloads]$ cat output
cat: output: No such file or directory
[cloudera@quickstart Downloads]$ hadoop cat output
Error: Could not find or load main class cat
[cloudera@quickstart Downloads]$ hadoop fs -cat output/*
AB
CD
AC
BD
AD
BC
BC
ADE
BD
ACE
BE
CD
CD
ABE
CE
BD
DE
BC
[cloudera@quickstart Downloads]$
```

2. Smartphone/Watch Application Implement a smartwatch/smartphone application using existing speech services/ image services (e.g., IBM Alchemyapi, Face++) related to your project.

Description: I have created an app that uses the camera of our smart phone to capture images. This is done by using the camera API.