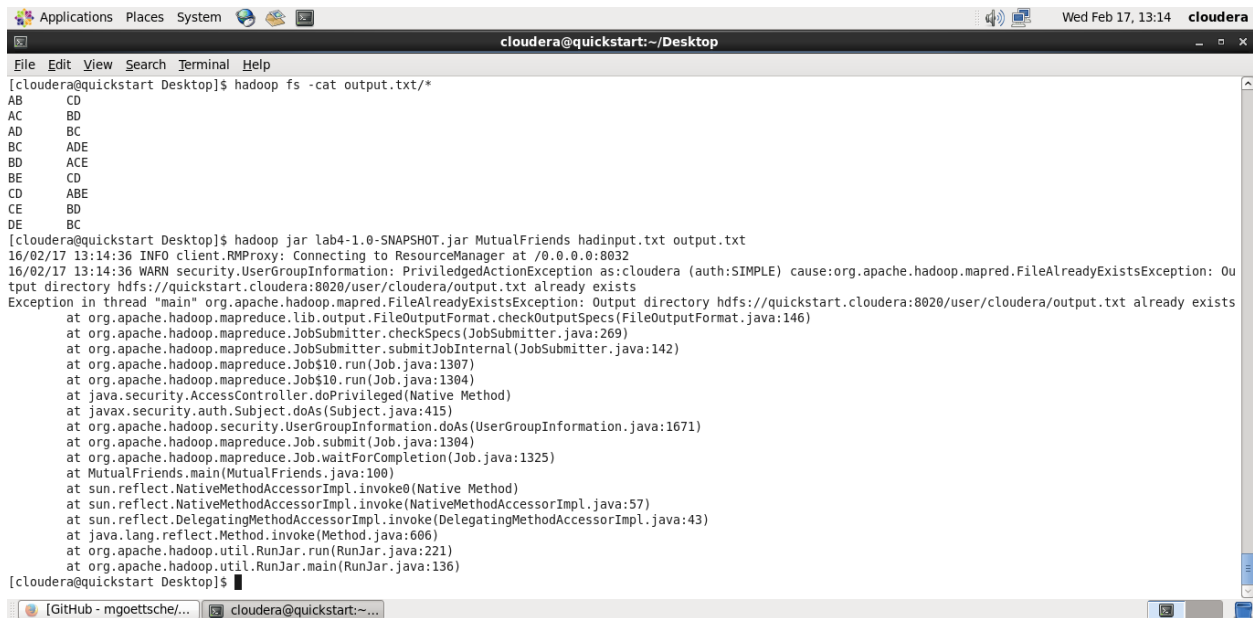
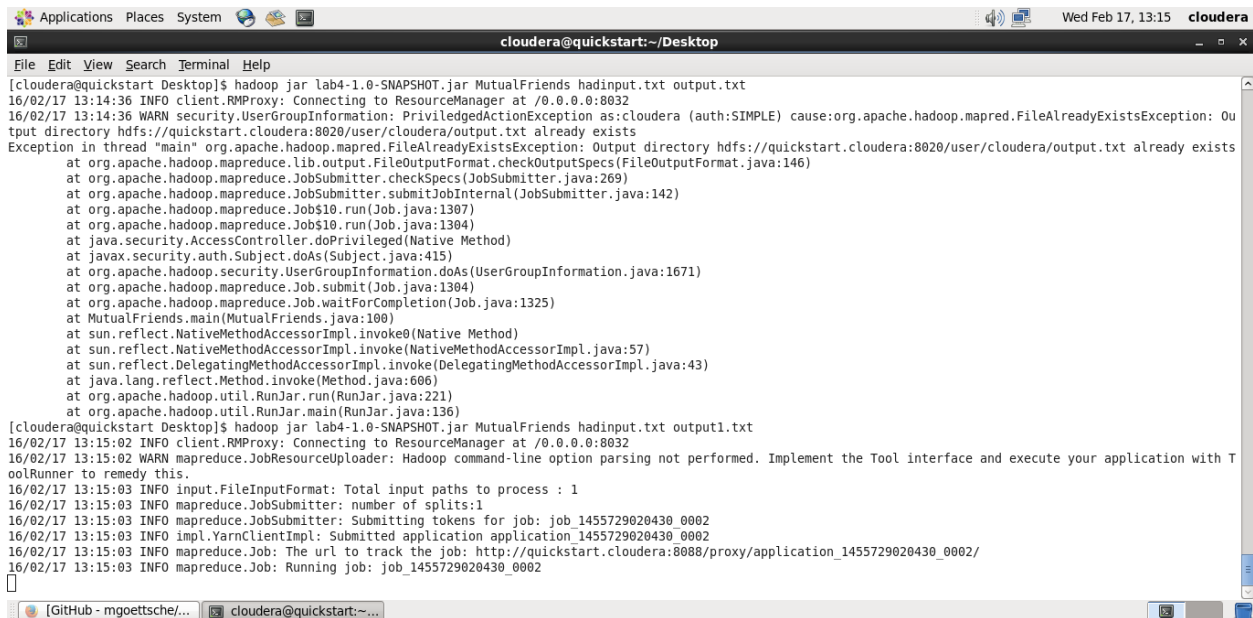


## 1. Mapreduce implementation for Facebook mutual friends concept



```
Applications Places System cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
[cloudera@quickstart Desktop]$ hadoop fs -cat output.txt/*
AB      CD
AC      BD
AD      BC
BC      ADE
BD      ACE
BE      CD
CE      ABE
DE      BD
DE      BC

[cloudera@quickstart Desktop]$ hadoop jar lab4-1.0-SNAPSHOT.jar MutualFriends hadinput.txt output.txt
16/02/17 13:14:36 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
16/02/17 13:14:36 WARN security.UserGroupInformation: PrivilegedActionException as:cloudera (auth:SIMPLE) cause:org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://quickstart.cloudera:8020/user/cloudera/output.txt already exists
Exception in thread "main" org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://quickstart.cloudera:8020/user/cloudera/output.txt already exists
    at org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.checkOutputSpecs(FileOutputFormat.java:146)
    at org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:269)
    at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:142)
    at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1307)
    at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1304)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:415)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1671)
    at org.apache.hadoop.mapreduce.Job.submit(Job.java:1304)
    at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1325)
    at MutualFriends.main(MutualFriends.java:100)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:606)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
[cloudera@quickstart Desktop]$
```



```
Applications Places System cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
[cloudera@quickstart Desktop]$ hadoop jar lab4-1.0-SNAPSHOT.jar MutualFriends hadinput.txt output.txt
16/02/17 13:14:36 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
16/02/17 13:14:36 WARN security.UserGroupInformation: PrivilegedActionException as:cloudera (auth:SIMPLE) cause:org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://quickstart.cloudera:8020/user/cloudera/output.txt already exists
Exception in thread "main" org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://quickstart.cloudera:8020/user/cloudera/output.txt already exists
    at org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.checkOutputSpecs(FileOutputFormat.java:146)
    at org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:269)
    at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:142)
    at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1307)
    at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1304)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:415)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1671)
    at org.apache.hadoop.mapreduce.Job.submit(Job.java:1304)
    at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1325)
    at MutualFriends.main(MutualFriends.java:100)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:606)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
[cloudera@quickstart Desktop]$ hadoop jar lab4-1.0-SNAPSHOT.jar MutualFriends hadinput.txt output1.txt
16/02/17 13:15:02 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
16/02/17 13:15:02 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
16/02/17 13:15:03 INFO input.FileInputFormat: Total input paths to process : 1
16/02/17 13:15:03 INFO mapreduce.JobSubmitter: number of splits:1
16/02/17 13:15:03 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1455729020430_0002
16/02/17 13:15:03 INFO impl.YarnClientImpl: Submitted application application_1455729020430_0002
16/02/17 13:15:03 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8080/proxy/application_1455729020430_0002/
16/02/17 13:15:03 INFO mapreduce.Job: Running job: job_1455729020430_0002
[cloudera@quickstart Desktop]$
```

```
Applications Places System cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
tput directory hdfs://quickstart.cloudera:8020/user/cloudera/output.txt already exists
Exception in thread "main" org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://quickstart.cloudera:8020/user/cloudera/output.txt already exists
    at org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.checkOutputsSpecs(FileOutputFormat.java:146)
    at org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:269)
    at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:142)
    at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1307)
    at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1304)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:415)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1671)
    at org.apache.hadoop.mapreduce.Job.submit(Job.java:1304)
    at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1325)
    at MutualFriends.main(MutualFriends.java:100)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:606)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
[cloudera@quickstart Desktop]$ hadoop jar lab4-1.0-SNAPSHOT.jar MutualFriends hadinput.txt output1.txt
16/02/17 13:15:02 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
16/02/17 13:15:02 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with T
oolRunner to remedy this.
16/02/17 13:15:03 INFO input.FileInputFormat: Total input paths to process : 1
16/02/17 13:15:03 INFO mapreduce.JobSubmitter: number of splits:1
16/02/17 13:15:03 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1455729020430_0002
16/02/17 13:15:03 INFO impl.YarnClientImpl: Submitted application application_1455729020430_0002
16/02/17 13:15:03 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8080/proxy/application_1455729020430_0002/
16/02/17 13:15:03 INFO mapreduce.Job: Running job: job_1455729020430_0002
16/02/17 13:15:14 INFO mapreduce.Job: Job job_1455729020430_0002 running in uber mode : false
16/02/17 13:15:14 INFO mapreduce.Job: map 0% reduce 0%
16/02/17 13:15:23 INFO mapreduce.Job: map 100% reduce 0%
```

```
Applications Places System cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
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
  Spilled Records=36
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=139
  CPU time spent (ms)=1530
  Physical memory (bytes) snapshot=339156992
  Virtual memory (bytes) snapshot=3008630784
  Total committed heap usage (bytes)=226562048
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=49
File Output Format Counters
  Bytes Written=57
[cloudera@quickstart Desktop]$
```

```
Applications Places System cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
16/02/17 13:15:03 INFO mapreduce.Job: Running job: job_1455729020430_0002
16/02/17 13:15:14 INFO mapreduce.Job: Job job_1455729020430_0002 running in uber mode : false
16/02/17 13:15:14 INFO mapreduce.Job: map 0% reduce 0%
16/02/17 13:15:23 INFO mapreduce.Job: map 100% reduce 0%
16/02/17 13:15:32 INFO mapreduce.Job: map 100% reduce 100%
16/02/17 13:15:32 INFO mapreduce.Job: Job job_1455729020430_0002 completed successfully
16/02/17 13:15:32 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=180
    FILE: Number of bytes written=223173
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=172
    HDFS: Number of bytes written=57
    HDFS: Number of read operations=6
    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)=6658
    Total time spent by all reduces in occupied slots (ms)=7055
    Total time spent by all map tasks (ms)=6658
    Total time spent by all reduce tasks (ms)=7055
    Total vcore-seconds taken by all map tasks=6658
    Total vcore-seconds taken by all reduce tasks=7055
    Total megabyte-seconds taken by all map tasks=6817792
    Total megabyte-seconds taken by all reduce tasks=7224320
  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
    Spilled Records=36
    Shuffled Maps=1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=139
    CPU time spent (ms)=1530
    Physical memory (bytes) snapshot=339156992
    Virtual memory (bytes) snapshot=3008630784
    Total committed heap usage (bytes)=226562048
  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=49
  File Output Format Counters
    Bytes Written=57
[cloudera@quickstart Desktop]$
```

```
Applications Places System cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
  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
    Spilled Records=36
    Shuffled Maps=1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=139
    CPU time spent (ms)=1530
    Physical memory (bytes) snapshot=339156992
    Virtual memory (bytes) snapshot=3008630784
    Total committed heap usage (bytes)=226562048
  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=49
  File Output Format Counters
    Bytes Written=57
[cloudera@quickstart Desktop]$
```

```
Applications Places System cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
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)=139
CPU time spent (ms)=1530
Physical memory (bytes) snapshot=339156992
Virtual memory (bytes) snapshot=3008630784
Total committed heap usage (bytes)=226562048
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=49
File Output Format Counters
Bytes Written=57
[cloudera@quickstart Desktop]$ hadoop fs -cat output1.txt/*
AB      CD
AC      BD
AD      BC
BC      ADE
BD      ACE
BE      CD
CD      ABE
CE      BD
DE      BC
[cloudera@quickstart Desktop]$
```

## 2. Android text to speech conversion

