

Homework Assignment 2

Name: Quach, Hoa

NetID: gt4789

1. Work *Example 4-16* (to the last command for *index* on page 139) on Amazon Elastic MapReduce. Type out all the commands in each step of the process and print out a screenshot of the final results in Amazon EMR.

List of Commands

```
javac -cp src/:hadoop-common-2.6.1.jar:hadoop-mapreduce-  
client-core-2.6.1.jar:commons-cli-2.0.jar -d .  
MapFileWriteDemo.java
```

```
jar -cvf MapFileWriteDemo.jar MapFileWriteDemo.class
```

```
hdfs dfs -ls /user/hadoop
```

```
ls -l
```

```
hadoop MapFileWriteDemo numbers.map
```

```
hdfs dfs -ls numbers.map
```

```
hadoop fs -text numbers.map/data | head
```

```
hadoop fs -text numbers.map/index
```

```
hadoop@ip-172-31-55-123:~
e-2.6.1.jar:commons-cli-2.0.jar -d . MapFileWriteDemo.java
Note: MapFileWriteDemo.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
[hadoop@ip-172-31-55-123 ~]$ jar -cvf MapFileWriteDemo.jar MapFileWriteDemo.class
added manifest
adding: MapFileWriteDemo.class(in = 1829) (out= 1007) (deflated 44%)
[hadoop@ip-172-31-55-123 ~]$ hdfs dfs -ls /user/hadoop
Found 1 items
drwxr-xr-x - hadoop hadoop 0 2016-04-25 19:40 /user/hadoop/input
[hadoop@ip-172-31-55-123 ~]$ hadoop MapFileWriteDemo numbers.map
16/04/25 19:42:14 INFO metrics.MetricsSaver: MetricsConfigRecord disabledInCluster: false instanceEngineCycleSec: 60 clusterEngineCycleSec: 60 disableClusterEngine: true maxMemoryMb: 3072 maxInstanceCount: 500 lastModified: 1461613070545
16/04/25 19:42:14 INFO metrics.MetricsSaver: Created MetricsSaver j-E8168UGCKW69:i-0f798ab33b7b93a34:MapFileWriteDemo:13177 period:60 /mnt/var/em/raw/i-0f798ab33b7b93a34_20160425_MapFileWriteDemo_13177_raw.bin
16/04/25 19:42:14 INFO zlib.ZlibFactory: Successfully loaded & initialized native-zlib library
16/04/25 19:42:14 INFO compress.CodecPool: Got brand-new compressor [.deflate]
16/04/25 19:42:14 INFO compress.CodecPool: Got brand-new compressor [.deflate]
[hadoop@ip-172-31-55-123 ~]$ hdfs dfs -ls numbers.map
Found 2 items
-rw-r--r-- 1 hadoop hadoop 47898 2016-04-25 19:42 numbers.map/data
-rw-r--r-- 1 hadoop hadoop 251 2016-04-25 19:42 numbers.map/index
[hadoop@ip-172-31-55-123 ~]$ hadoop fs -text numbers.map/data | head
16/04/25 19:42:24 INFO zlib.ZlibFactory: Successfully loaded & initialized native-zlib library
16/04/25 19:42:24 INFO compress.CodecPool: Got brand-new decompressor [.deflate]
1 One, two, buckle my shoe
2 Three, four, shut the door
3 Five, six, pick up sticks
4 Seven, eight, lay them straight
5 Nine, ten, a big fat hen
6 One, two, buckle my shoe
7 Three, four, shut the door
8 Five, six, pick up sticks
9 Seven, eight, lay them straight
10 Nine, ten, a big fat hen
[hadoop@ip-172-31-55-123 ~]$ hadoop fs -text numbers.map/index
16/04/25 19:42:29 INFO zlib.ZlibFactory: Successfully loaded & initialized native-zlib library
16/04/25 19:42:29 INFO compress.CodecPool: Got brand-new decompressor [.deflate]
16/04/25 19:42:29 INFO compress.CodecPool: Got brand-new decompressor [.deflate]
16/04/25 19:42:29 INFO compress.CodecPool: Got brand-new decompressor [.deflate]
16/04/25 19:42:29 INFO compress.CodecPool: Got brand-new decompressor [.deflate]
1 128
129 6079
257 12054
385 18030
513 24002
641 29976
769 35947
897 41922
[hadoop@ip-172-31-55-123 ~]$
```

2. Work *Example 4-17* (to the end of page 142) on Amazon Elastic MapReduce. Type out all the commands in each step of the process and print out a screenshot of the final results (showing the file directory) in Amazon EMR.

List of Commands

```
javac -cp src/:hadoop-common-2.6.1.jar:hadoop-mapreduce-  
client-core-2.6.1.jar:commons-cli-2.0.jar:hadoop-mapreduce-  
examples-2.2.0.jar -d . MapFileFixer.java
```

```
javac -cp src/:hadoop-common-2.6.1.jar:hadoop-mapreduce-  
client-core-2.6.1.jar:commons-cli-2.0.jar:hadoop-mapreduce-  
examples-2.2.0.jar -d . SequenceFileWriteDemo.java
```

```
hadoop SequenceFileWriteDemo numbers.seq
```

```
jar -cvf MapFileFixer.jar MapFileFixer.class
```

```
hdfs dfs -ls /user/hadoop
```

```
ls -l
```

```
hdfs dfs -mkdir -p /user/hadoop/input2
```

```
hdfs dfs -copyFromLocal /home/hadoop/MapFileFixer.jar /user/hadoop/input2
```

```
hadoop jar hadoop-mapreduce-examples-2.2.0.jar sort -r 1 -  
inFormat  
org.apache.hadoop.mapreduce.lib.input.SequenceFileInputFormat -  
outFormat  
org.apache.hadoop.mapreduce.lib.output.SequenceFileOutputFormat -  
outKey org.apache.hadoop.io.IntWritable -outValue  
org.apache.hadoop.io.Text numbers.seq numbers.map
```

```
hadoop fs -mv numbers.map/part-r-00000 numbers.map/data
```

```
hadoop MapFileFixer numbers.map
```

```
hadoop@ip-172-31-54-97:~  
Total time spent by all reduce tasks (ms)=7221  
Total vcore-milliseconds taken by all map tasks=4946  
Total vcore-milliseconds taken by all reduce tasks=7221  
Total megabyte-milliseconds taken by all map tasks=7122240  
Total megabyte-milliseconds taken by all reduce tasks=20796480  
Map-Reduce Framework  
  Map input records=100  
  Map output records=100  
  Map output bytes=3100  
  Map output materialized bytes=656  
  Input split bytes=129  
  Combine input records=0  
  Combine output records=0  
  Reduce input groups=100  
  Reduce shuffle bytes=656  
  Reduce input records=100  
  Reduce output records=100  
  Spilled Records=200  
  Shuffled Maps =1  
  Failed Shuffles=0  
  Merged Map outputs=1  
  GC time elapsed (ms)=179  
  CPU time spent (ms)=1990  
  Physical memory (bytes) snapshot=642514944  
  Virtual memory (bytes) snapshot=5371191296  
  Total committed heap usage (bytes)=807927808  
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=4788  
File Output Format Counters  
  Bytes Written=4005  
Job ended: Sun May 01 19:49:20 UTC 2016  
The job took 32 seconds.  
[hadoop@ip-172-31-54-97 ~]$ hadoop fs -mv numbers.map/part-r-00000 numbers.map/data  
[hadoop@ip-172-31-54-97 ~]$ hadoop MapFileFixer numbers.map  
16/05/01 19:49:34 INFO metrics.MetricsSaver: MetricsConfigRecord disabledInCluster: false instanceEngineCycleSec: 60 clusterEngineCycleSec: 60 disableClusterEngine: true maxMemoryMb: 3072 maxInstanceCount: 500 lastModified: 1462131920031  
16/05/01 19:49:34 INFO metrics.MetricsSaver: Created MetricsSaver j-1ITKJ50Y3J0S0:i-01ee591be73801c7d:MapFileFixer:11915 period:60 /mnt/var/em/raw/i-01ee591be73801c7d_20160501_MapFileFixer_11915_raw.bin  
16/05/01 19:49:35 INFO zlib.ZlibFactory: Successfully loaded & initialized native-zlib library  
16/05/01 19:49:35 INFO compress.CodecPool: Got brand-new compressor [deflate]  
Created MapFile numbers.map with 100 entries  
[hadoop@ip-172-31-54-97 ~]$
```