

Assignment 2

1. Work *Example 4-16* (to the last command for *index* on page 139) on CSUEB Hadoop or 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 CSUEB Hadoop or in Amazon EMR.

Command to Compile Java files:

```
javac -classpath /home/student7/hadoop-common-2.6.1.jar:/home/student7/hadoop-mapreduce-client-core-2.6.1.jar:/home/student7/commons-cli-2.0.jar -d . MapFileWriteDemo.java
```

Export Hadoop Path

```
export HADOOP_CLASSPATH=/home/student7
```

Command to Run the program to build the Map file

```
hadoop MapFileWriteDemo /home/student7_lu/numbers.map
```

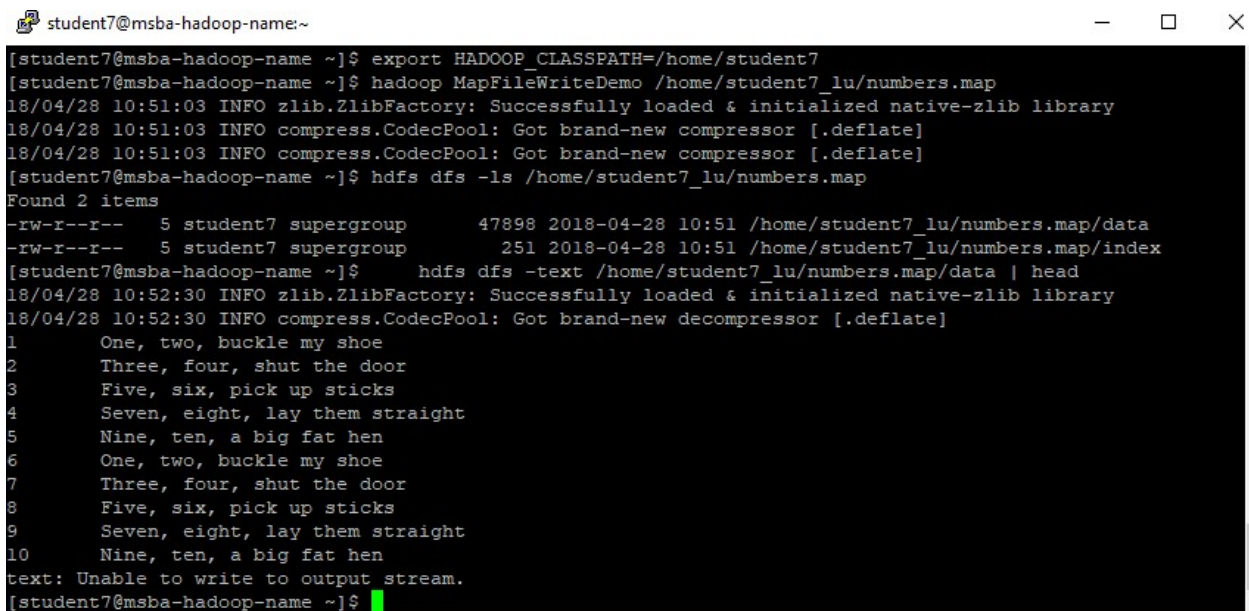
Command to look for the map file (found 2 items: data and index)

```
hdfs dfs -ls /home/student7_lu/numbers.map
```

Using command text to look at data and index file

```
hdfs dfs -text /home/student7_lu/numbers.map/data | head
```

```
hdfs dfs -text /home/student7_lu/numbers.map/index
```



```
student7@msba-hadoop-name:~  
[student7@msba-hadoop-name ~]$ export HADOOP_CLASSPATH=/home/student7  
[student7@msba-hadoop-name ~]$ hadoop MapFileWriteDemo /home/student7_lu/numbers.map  
18/04/28 10:51:03 INFO zlib.ZlibFactory: Successfully loaded & initialized native-zlib library  
18/04/28 10:51:03 INFO compress.CodecPool: Got brand-new compressor [.deflate]  
18/04/28 10:51:03 INFO compress.CodecPool: Got brand-new compressor [.deflate]  
[student7@msba-hadoop-name ~]$ hdfs dfs -ls /home/student7_lu/numbers.map  
Found 2 items  
-rw-r--r--  5 student7 supergroup      47898 2018-04-28 10:51 /home/student7_lu/numbers.map/data  
-rw-r--r--  5 student7 supergroup       251 2018-04-28 10:51 /home/student7_lu/numbers.map/index  
[student7@msba-hadoop-name ~]$ hdfs dfs -text /home/student7_lu/numbers.map/data | head  
18/04/28 10:52:30 INFO zlib.ZlibFactory: Successfully loaded & initialized native-zlib library  
18/04/28 10:52:30 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  
text: Unable to write to output stream.  
[student7@msba-hadoop-name ~]$
```

```
student7@msba-hadoop-name:~  
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  
text: Unable to write to output stream.  
[student7@msba-hadoop-name ~]$ hdfs dfs -text /home/student7_lu/numbers.map/index  
18/04/28 11:00:26 INFO zlib.ZlibFactory: Successfully loaded & initialized native-zlib library  
18/04/28 11:00:26 INFO compress.CodecPool: Got brand-new decompressor [.deflate]  
18/04/28 11:00:26 INFO compress.CodecPool: Got brand-new decompressor [.deflate]  
18/04/28 11:00:26 INFO compress.CodecPool: Got brand-new decompressor [.deflate]  
18/04/28 11:00:26 INFO compress.CodecPool: Got brand-new decompressor [.deflate]  
1      128  
129     6079  
257     12054  
385     18030  
513     24002  
641     29976  
769     35947  
897     41922  
[student7@msba-hadoop-name ~]$
```

2. Work *Example 4-17* (to the end of page 142) on CSUEB Hadoop or 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 CSUEB Hadoop or in Amazon EMR.(Re-creating the index for a MapFile)

Command to Compile Java files:

```
javac -classpath /home/student7/hadoop-common-2.6.1.jar:/home/student7/hadoop-mapreduce-client-core-2.6.1.jar:/home/student7/commons-cli-2.0.jar -d . MapFileFixer.java
```

Command to sort the sequence file numbers.seq into a new directory called number.map

```
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  
/home/student7/numbers.seq /home/student7_lu/numbers.map2
```

Command to look for new directory numbers.map2

```
hadoop fs -ls /home/student7_lu/numbers.map2/
```

Command to rename the MapReduce output to be the data file


```
hadoop fs -mv /home/student7_lu/numbers.map2/part-r-00000  
/home/student7_lu/numbers.map2/data
```

Export Hadoop Path

```
export HADOOP_CLASSPATH=/home/student7
```

Create the index file

```
hadoop MapFileFixer /home/student7_lu/numbers.map2/
```

 student7@msba-hadoop-name:~

```
[student7@msba-hadoop-name ~]$ 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 /home/student7/numbers.seq /home/student7_lu/numbers.map2
18/05/02 21:01:09 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
Running on 1 nodes to sort from hdfs://msba-hadoop-name:9000/home/student7/numbers.seq into hdfs://msba-hadoop-name:9000/home/student7_lu/numbers.map2 with 1 reduces.
Job started: Wed May 02 21:01:09 PDT 2018
18/05/02 21:01:09 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
18/05/02 21:01:10 INFO input.FileInputFormat: Total input files to process : 1
18/05/02 21:01:10 INFO mapreduce.JobSubmitter: number of splits:1
18/05/02 21:01:10 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1513807751472_1249
18/05/02 21:01:10 INFO impl.YarnClientImpl: Submitted application application_1513807751472_1249
18/05/02 21:01:10 INFO mapreduce.Job: The url to track the job: http://msba-hadoop-name:8088/proxy/application_1513807751472_1249/
18/05/02 21:01:10 INFO mapreduce.Job: Running job: job_1513807751472_1249
18/05/02 21:01:16 INFO mapreduce.Job: Job job_1513807751472_1249 running in uber mode : false
18/05/02 21:01:16 INFO mapreduce.Job:  map 0% reduce 0%
18/05/02 21:01:20 INFO mapreduce.Job:  map 100% reduce 0%
18/05/02 21:01:25 INFO mapreduce.Job:  map 100% reduce 100%
18/05/02 21:01:25 INFO mapreduce.Job: Job job_1513807751472_1249 completed successfully
18/05/02 21:01:26 INFO mapreduce.Job: Counters: 49
      File System Counters
        FILE: Number of bytes read=3306
        FILE: Number of bytes written=410611
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
```

```

    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: Wed May 02 21:01:26 PDT 2018
The job took 16 seconds.
[student7@msba-hadoop-name ~]$ hadoop fs -ls /home/student7_lu/numbers.map2/
Found 2 items
-rw-r--r--  5 student7 supergroup          0 2018-05-02 21:01 /home/student7_lu
/numbers.map2/_SUCCESS
-rw-r--r--  5 student7 supergroup      4005 2018-05-02 21:01 /home/student7_lu
/numbers.map2/part-r-00000

```

```

File Output Format Counters
    Bytes Written=4005
Job ended: Wed May 02 21:01:26 PDT 2018
The job took 16 seconds.
[student7@msba-hadoop-name ~]$ hadoop fs -ls /home/student7_lu/numbers.map2/
Found 2 items
-rw-r--r--  5 student7 supergroup          0 2018-05-02 21:01 /home/student7_lu
/numbers.map2/_SUCCESS
-rw-r--r--  5 student7 supergroup      4005 2018-05-02 21:01 /home/student7_lu
/numbers.map2/part-r-00000
[student7@msba-hadoop-name ~]$ hadoop fs -mv /home/student7_lu/numbers.map2/part-r-00000 /home/
[student7@msba-hadoop-name ~]$ hadoop MapFileFixer /home/student7_lu/numbers.map2/
Error: Could not find or load main class MapFileFixer
[student7@msba-hadoop-name ~]$ hadoop fs -mv /home/student7_lu/numbers.map2/part-r-00000 /home/
mv: `/home/student7_lu/numbers.map2/data': File exists
[student7@msba-hadoop-name ~]$ hadoop MapFileFixer /home/student7_lu/numbers.map2/
Error: Could not find or load main class MapFileFixer
[student7@msba-hadoop-name ~]$ export HADOOP_CLASSPATH=/home/student7
[student7@msba-hadoop-name ~]$ hadoop MapFileFixer /home/student7_lu/numbers.map2/
18/05/02 21:32:04 INFO zlib.ZlibFactory: Successfully loaded & initialized native-zlib li
brary
18/05/02 21:32:04 INFO compress.CodecPool: Got brand-new compressor [.deflate]
Created MapFile /home/student7_lu/numbers.map2 with 100 entries
[student7@msba-hadoop-name ~]$

```