Question 1. Work Examples 7-2, 7-3, and 7-4 on CSUEB Hadoop. Type out all the commands in each step of the process and print out a screenshot of the final results in CSUEB Hadoop.

Command 1 - Preliminary step, compile Java classes with classpath

```
javac -classpath
/home/student29/hadoop-common-2.6.1.jar:/home/student29/hadoop-mapreduce-client-core-2.6
.1.jar:/home/student29/commons-cli-2.0.jar -d . WholeFileInputFormat.java
WholeFileRecordReader.java SmallFilesToSequenceFileConverter.java JobBuilder.java
```

Command 2 - creating a hadoop-example.jar JAR file which will be used as the Hadoop MapReduce job

```
jar -cvf hadoop-example.jar WholeFile*.class SmallFilesToSequenceFileConverter*.class
JobBuilder.class
```

Command 3 - Transfer the Jar (optional) and smallfiles directory into the hadoop filesystem to perform MapReduce job

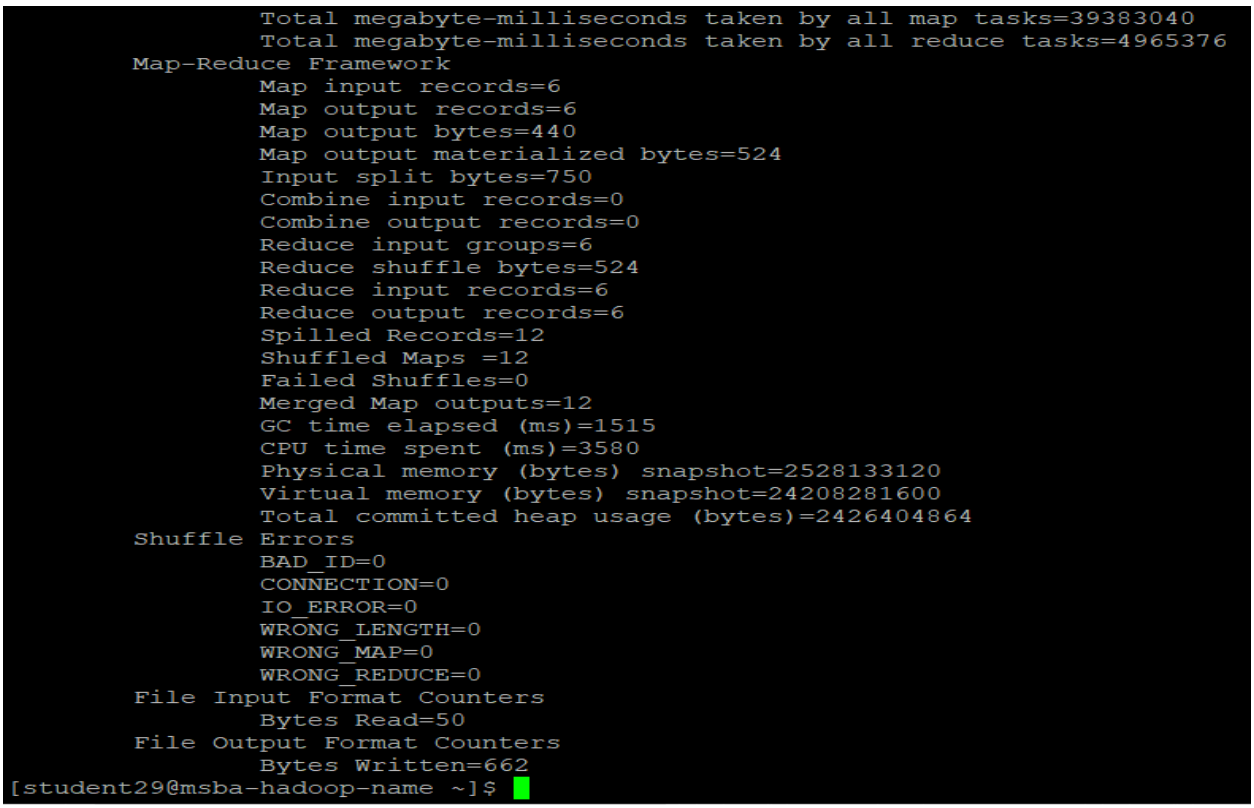
```
hdfs dfs -copyFromLocal hadoop-example.jar /home/student29/     \
hdfs dfs -copyFromLocal smallfiles /home/student29/
```

Command 4 - Run hadoop-example.jar MapReduce job with some configurations.

```
hadoop jar /home/student29/hadoop-example.jar SmallFilesToSequenceFileConverter -D
mapred.reduce.tasks=2 /home/student29/smallfiles /home/student29/output11
```

Command 5 - Display the output from the reducers where each part file contains a portion of the final output data.

```
hadoop fs -text /home/student29/output11/part-r-00000       \
hadoop fs -text /home/student29/output11/part-r-00001
```



Final Output for Question 1 - Displays output from Command 5

```
[student29@msba-hadoop-name ~]$ hdfs dfs -ls /home/student29/output11
Found 3 items
-rw-r--r--   5 student29 supergroup          0 2024-03-22 13:49 /home/student29/output11/_SUCCESS
-rw-r--r--   5 student29 supergroup        326 2024-03-22 13:49 /home/student29/output11/part-r-00000
-rw-r--r--   5 student29 supergroup        336 2024-03-22 13:49 /home/student29/output11/part-r-00001
[student29@msba-hadoop-name ~]$ hadoop fs -text /home/student29/output11/part-r-00000
hdfs://msba-hadoop-name:9000/home/student29/smallfiles/a.txt    61 61 61 61 61 61 61 61 61 61
hdfs://msba-hadoop-name:9000/home/student29/smallfiles/c.txt    63 63 63 63 63 63 63 63 63 63
hdfs://msba-hadoop-name:9000/home/student29/smallfiles/e.txt
[student29@msba-hadoop-name ~]$ hadoop fs -text /home/student29/output11/part-r-00001
hdfs://msba-hadoop-name:9000/home/student29/smallfiles/b.txt    62 62 62 62 62 62 62 62 62 62
hdfs://msba-hadoop-name:9000/home/student29/smallfiles/d.txt    64 64 64 64 64 64 64 64 64 64
hdfs://msba-hadoop-name:9000/home/student29/smallfiles/f.txt    66 66 66 66 66 66 66 66 66 66
```

Question 2. Work Example 8-1 on CSUEB Hadoop. Type out all the commands in each step of the process and print out a screenshot of the final results (the counters) in CSUEB Hadoop.

Command 1 - Preliminary step, compile Java classes with classpath

```
javac -classpath
/home/student29/hadoop-common-2.6.1.jar:/home/student29/hadoop-mapreduce-client-core-2.6
.1.jar:/home/student29/commons-cli-2.0.jar -d . MaxTemperatureWithCounters.java
MaxTemperatureMapper.java MaxTemperatureReducer.java NcdcRecordParser.java
JobBuilder.java
```

Command 2 - creating max-temp-job.jar JAR file which will be used as the Hadoop MapReduce job

```
jar -cvf max-temp-job.jar MaxTemperatureWithCounters*.class MaxTemperatureMapper.class
MaxTemperatureReducer.class NcdcRecordParser.class JobBuilder.class
```

Command 3 - Transfer the max-temp-job.jar Jar (optional) and 1930 directory into the hadoop filesystem to perform MapReduce job

```
hdfs dfs -copyFromLocal 1930/ /home/student29/    \
hdfs dfs -copyFromLocal max-temp.jar /home/student29/
```

Command 4 - Run max-temp-job.jar MapReduce job

```
hadoop jar max-temp-job.jar MaxTemperatureWithCounters /home/student29/1930
/home/student29/output-counters
```

Final Output for Question 2 - Prints out the counters from Command 4

```
                Shuffled Maps =121
                Failed Shuffles=0
                Merged Map outputs=121
                GC time elapsed (ms)=27329
                CPU time spent (ms)=72360
                Physical memory (bytes) snapshot=41788686336
                Virtual memory (bytes) snapshot=368567398400
                Total committed heap usage (bytes)=38525206528
        MaxTemperatureWithCounters$Temperature
                MISSING=3665
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        TemperatureQuality
                1=85580
                2=17
                9=3665
        File Input Format Counters
                Bytes Read=1649606
        File Output Format Counters
                Bytes Written=9
[student29@msba-hadoop-name ~]$
```