ITM_6273_02
Homework Assignment #3

Quach, Hoa
NetID: gt4789

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.**

   Hint 1: need to create a jar file including five classes:
   WholeFileInputFormat.class,
   WholeFileRecordReader.class,
   SmallFilesToSequenceFileConverter.class,
   SmallFilesToSequenceFileConverter$SequenceFileMapper.class,
   JobBuilder.class

   Note. *SmallFilesToSequenceFileConverter.class* is the main class.
   SmallFilesToSequenceFileConverter$SequenceFileMapper.class is a nested/inner class of *SmallFilesToSequenceFileConverter.class*.
   JobBuilder.java can be found in Hadoop-Book-Master/common/src/main/java
   Smallfiles folder can be found in Hadoop-Book-Master/input

   Hint 2: *-conf conf/Hadoop-localhost.xml* is not needed in the hadoop jar command. So, your command will be like this:

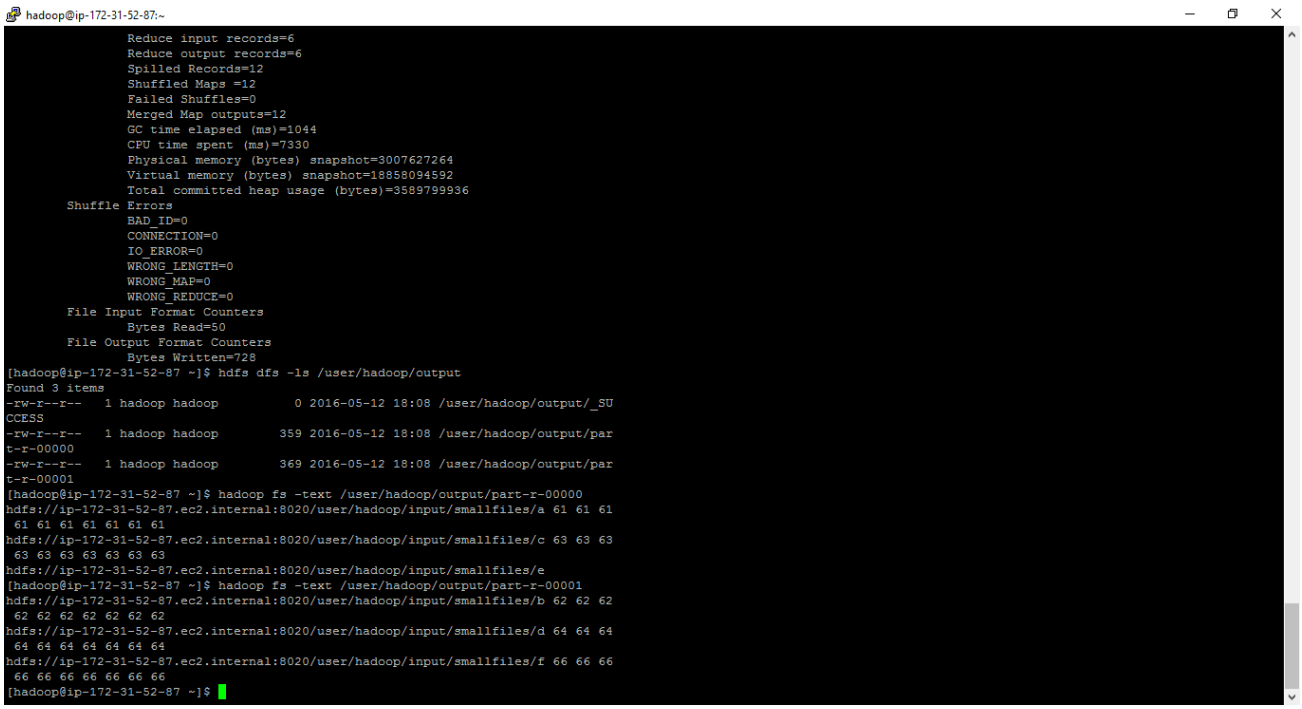
   hadoop jar /home/jwu/hadoop-example.jar SmallFilesToSequenceFileConverter -D mapred.reduce.tasks=2 /home/jwu/smallfiles /home/jwu/output11

   Commands
   ****Note:** All .java files have been placed in the root directory of /home/hadoop/ directory in AWS.

```
mkdir /home/hadoop/Assignment_3_1

javac -cp src/:hadoop-common-2.6.1.jar:hadoop-mapreduce-
client-core-2.6.1.jar:commons-cli-2.0.jar -d
Assignment_3_1/ *.java

jar -cvf
Assignment_3_1/SmallFilesToSequenceFileConverter.jar -C
Assignment_3_1/ .

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

**After manually uploading smallfiles directory from local desktop to AWS Hadoop server via WinSCP:

```
hdfs dfs -copyFromLocal /home/hadoop/smallfiles
/user/hadoop/input

hdfs dfs -ls /user/hadoop/input/smallfiles

hadoop jar
Assignment_3_1/SmallFilesToSequenceFileConverter.jar
SmallFilesToSequenceFileConverter -D mapred.reduce.tasks=2
/user/hadoop/input/smallfiles/ /user/hadoop/output

hdfs dfs -ls /user/hadoop/output

hadoop fs -text /user/hadoop/output/part-r-00000

hadoop fs -text /user/hadoop/output/part-r-00001
```



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.**

   Hint: need to create a jar file including five classes:
   MaxTemperatureWithCounters.class,
   MaxTemperatureMapperWithCounters.class,
   NcdcRecordParser.class,
   JobBuilder.class,
   MaxTemperatureReducer.class.

   Note. 1) *NcdcRecordParser.java* can be found in Hadoop-Book-Master/common/src/main/java, 2) you must use the data of year 1930 as input data to run the program (download data at: ftp://ftp.ncdc.noaa.gov/pub/data/noaa/), and 3) your results shall be different from those run over the complete dataset of 100 years, which are shown on page 265 in the textbook.

   Note:
   **To download all 1930 files from ftp://ftp.ncdc.noaa.gov/pub/data/noaa/1930 to /user/hadoop/ (AWS)

```
mkdir /home/hadoop/1930

cd 1930

wget -r --no-directories
ftp://ftp.ncdc.noaa.gov/pub/data/noaa/1930/

ls -l

cd ../
```

   Commands using AWS

   ****Note:** All .java files have been placed in the root directory of /home/hadoop/ directory in AWS and problem#1 java codes have been removed (except JobBuilder.java).

```
mkdir /home/hadoop/MaxTemperature

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

jar -cvf MaxTemperature/MaxTemperatureWithCounters.jar -C
MaxTemperature/ .
```

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

hdfs dfs -copyFromLocal /home/hadoop/1930/
/user/hadoop/input2

hdfs dfs -ls /user/hadoop/input2/1930/

hadoop jar MaxTemperature/MaxTemperatureWithCounters.jar
MaxTemperatureWithCounters /user/hadoop/input2/1930/
/user/hadoop/output1930

hdfs dfs -ls /user/hadoop/output1930
```