Homework 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 for compiling Java file:

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

Command for setting class path:

export HADOOP\_CLASSPATH=/home/student22

Command to build a MapFile:

hadoop MapFileWriteDemo /home/22student22/numbers.map

Commands to Display data:

hdfs dfs -text /home/22student22/numbers.map/data | head

hdfs dfs -text /home/22student22/numbers.map/data | tail

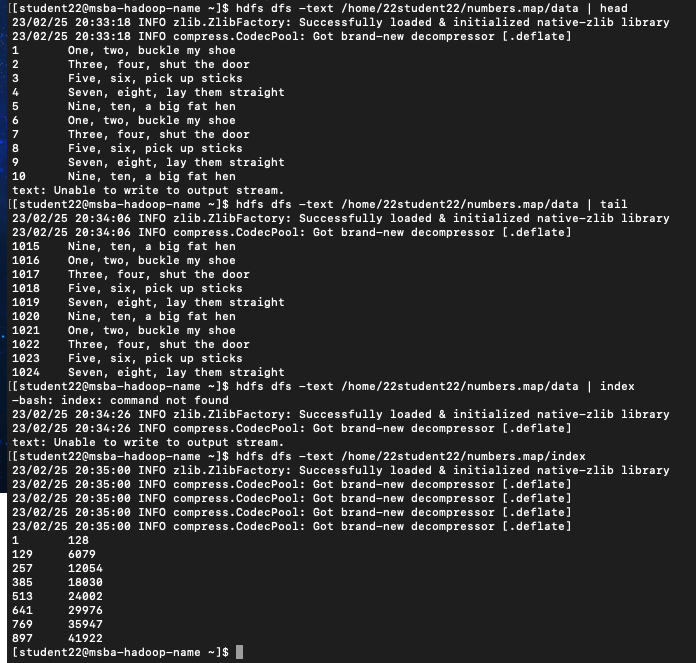
hdfs dfs -text /home/22student22/numbers.map/index

Command to check a MapFile:

hdfs dfs -ls /home/22student22/numbers.map

Text

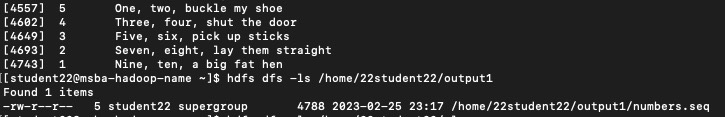
Description automatically generated



1. 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.

Command to check a numbers.seq file:

hdfs dfs -ls /home/22student22/assignment2/numbers.map



Command for compiling Java file:

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

Sort the sequence file numbers.seq (generated in *Example 4-14*) into a new directory called number.map that will become the MapFile.

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/22student22/output1/numbers.seq /home/22student22/assignment2/numbers.map

Check output MapReduce file generated:

hdfs dfs -ls /home/22student22/assignment2/number.map

Command for Rename the MapReduce output to be the data file:

hdfs dfs -mv /home/22student22/assignment2/numbers.map/part-r-00000 /home/22student22/assignment2/numbers.map/data

Command to Rename the MapReduce output to be the data file:

hdfs dfs -mv /home/22student22/assignment2/numbers.map/part-r-00000 /home/22student22/assignment2/ numbers.map/data

Text

Description automatically generated

A picture containing diagram

Description automatically generated