**Hadoop File System Commands List:**

https://hadoop.apache.org/docs/r2.7.1/hadoop-project-dist/hadoop-common/FileSystemShell.html

**CSUEB Hadoop URL**

URL: 134.154.190.204

NameNode: http://msba-hadoop-name:50070/

ResourceManager: http://msba-hadoop-name:8088/

MapReduce JobHistory Server: http://msba-hadoop-name:19888/

Location on local disk: /home/student0

Location on HDFS: /home/user0

Change the directory permission: hdfs dfs -chmod go-rx /home/user0

**Commands for Chapter 2 - Amazon**

How to submit a customer Jar:

<http://docs.aws.amazon.com/ElasticMapReduce/latest/DeveloperGuide/emr-launch-custom-jar-cli.html>

In Amazon console, set up Jar (max-temperature.jar) and then set up arguments:

MaxTemperature

s3://chapter2/input

s3://chapter2/output

Create an Input directory

hdfs dfs -mkdir /user/hadoop/input0/

In Amazon SSH command interface:

Amazon local drive folder is /home/hadoop; Amazon HDFS folder is /user/hadoop

Copy file from S3 to Local Disk:

aws s3 cp s3://chapter2/MaxTemperatureMapper.java .

aws s3 cp s3://textbook-example/URLCat.java .

Copy file from Local Disk to HDFS

hdfs dfs -copyFromLocal file.gz .

hdfs dfs -copyFromLocal sample.txt.gz .

hdfs dfs -copyFromLocal sample.txt .

hdfs dfs -copyFromLocal sample.txt /user/hadoop/input0/

List files in local drive:

ls

List files in HDFS:

hdfs dfs -ls /user/hadoop/

Delete a file from local drive:

rm filename

Delete a file from HDFS:

hadoop fs -rm -r /home/jwu/chapter2/sample.txt

hadoop fs -rm -r /user/hadoop/sample.txt

hdfs dfs -rm -r /user/hadoop/sample.txt

Compile Java files:

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 MaxTemperatureReducer.java MaxTemperatureMapper.java

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

Create a Jar file:

jar -cvf max-temperature.jar MaxTemperature\*.class

jar -cvf old-max-temperature.jar OldMaxTemperature\*.class

*(-c – create new archive; -v – generate verbose output on standard output; -f – specify archive file name)*

Run a Jar file

(Need to delete the classes; otherwise the system is running the main class instead of Jar; this indicates Amazon by default sets “export HADOOP\_CLASSPATH=/home/hadoop”)

export HADOOP\_CLASSPATH=$JAVA\_HOME/lib/tools.jar

hadoop jar max-temperature.jar MaxTemperature /user/hadoop/input0/sample.txt /user/hadoop/output01

hadoop jar max-temperature.jar MaxTemperature s3://itm6273weather/input1 output5

hadoop jar max-temperature.jar MaxTemperature /user/hadoop/input/ /user/hadoop/output1

hadoop jar old-max-temperature.jar OldMaxTemperature /user/hadoop/input/ /user/hadoop/output4

Display the output on screen:

hdfs dfs -cat /user/hadoop/chapter2/output3/part-r-00000

hadoop fs -cat /user/hadoop/output01/part-r-00001

hadoop fs -cat /user/hadoop/output4/part-00000

**Commands for Chapter 2 - CSUEB**

CSUEB Hadoop Host Address:

134.154.190.204

Compile Java files:

javac -classpath /home/jwu/jar4compile/hadoop-common-2.6.1.jar:/home/jwu/jar4compile/hadoop-mapreduce-client-core-2.6.1.jar:/home/jwu/jar4compile/commons-cli-2.0.jar -d . MaxTemperature.java MaxTemperatureMapper.java MaxTemperatureReducer.java

Create a Jar file:

jar -cvf max-temperature.jar ./MaxTemperature\*.class

Make a directory in HDFS:

hdfs dfs -mkdir /home/user0/chapter2/

Delete a directory or a file from HDFS:

hadoop fs -rm -r /home/user0/chapter2/sample.txt

*(The -r option is equivalent to –R; deletes the directory and any content under it recursively.)*

Copy a file from local drive to HDFS:

hdfs dfs -copyFromLocal /home/student0/sample.txt /home/user0/chapter2/sample.txt

hdfs dfs -copyFromLocal /home/jwu/sample.txt.gz /home/jwu/chapter2

Run a Jar file on Hadoop:

hadoop jar /home/student0/max-temperature.jar MaxTemperature /home/user0/chapter2/sample.txt /home/user0/output2/

Display the output on screen:

hdfs dfs -cat /home/jwu/chapter2/output3/part-r-00000

hadoop fs -cat /home/jwu/chapter2/output2/part-r-00000

*(The above two commands are the same.)*

**Commands for Chapter 3 – Amazon**

Commands for HDFS Operations:

hdfs dfs -copyFromLocal quangle.txt .

hdfs dfs -copyToLocal quangle.txt quangle.copy.txt

md5sum quangle.txt quangle.copy.txt

Commands for Example 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 . URLCat.java

hadoop URLCat hdfs:///user/hadoop/quangle.txt

Commands for Example 3-2:

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

hadoop FileSystemCat hdfs:///user/hadoop/quangle.txt

Commands for using Hadoop archives

hadoop fs -ls -R /home/jwu

hadoop archive -archiveName files.har -p /home/jwu input1 input2 /home/jwu

hadoop fs -ls /home/jwu/files.har

hadoop fs -ls -R har:///home/jwu/files.har

**Commands for Chapter 4 – Amazon**

Commands for Example 4-4:

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

jar -cvf max-temperature-compression.jar MaxTemperature\*.class

hdfs dfs -mkdir /user/hadoop/input1/

hdfs dfs -copyFromLocal sample.txt.gz /user/hadoop/input1/

hadoop jar max-temperature-compression.jar MaxTemperatureWithCompression /user/hadoop/input1/ /user/hadoop/output5

hdfs dfs -copyToLocal /user/hadoop/output5/ /home/hadoop/

gunzip -c output5/part-r-00001.gz

Commands for Example 4-14:

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

hadoop SequenceFileWriteDemo numbers.seq

Commands for Example 4-15:

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

hadoop SequenceFileReadDemo numbers.seq

hdfs dfs -text numbers.seq | head

hdfs dfs -text numbers.seq | tail

**Commands for Chapter 5 – CSUEB**

Commands for Example 5-5:

javac -classpath /home/student0/hadoop-common-2.6.1.jar:/home/student0/hadoop-mapreduce-client-core-2.6.1.jar:/home/student0/commons-cli-2.0.jar:/home/student0/mrunit-1.0.0-hadoop2.jar:/home/student0/junit-4.11.jar -d . MaxTemperatureMapperTest.java MaxTemperatureMapper.java

jar -cvf max-mapper-test-v1.jar MaxTemperatureMapper\*.class org

export HADOOP\_CLASSPATH=/home/student0

hadoop jar max-mapper-test-v1.jar org.junit.runner.JUnitCore MaxTemperatureMapperTest

**Commands for Chapter 7 – CSUEB**

Commands for Example 7-5:

javac -classpath /home/student0/hadoop-common-2.6.1.jar:/home/student0/hadoop-mapreduce-client-core-2.6.1.jar:/home/student0/commons-cli-2.0.jar -d . PartitionByStationUsingMultipleOutputs.java NcdcRecordParser.java JobBuilder.java

jar -cvf partition-by-station.jar PartitionByStationUsingMultipleOutputs\*.class NcdcRecordParser.class JobBuilder.class

hdfs dfs -copyFromLocal 1930 /home/user0/

export HADOOP\_CLASSPATH=/home/student0

hadoop jar partition-by-station.jar PartitionByStationUsingMultipleOutputs /home/user0/1930 /home/user0/output6

**Commands for Chapter 11 – CSUEB**

Commands for Example 11-2:

pig -x local

A = LOAD 'piginput/A' AS (fruit:chararray);

DUMP A;

DESCRIBE A;

javac -classpath /home/student0/hadoop-common-2.6.1.jar:/home/student0/hadoop-mapreduce-client-core-2.6.1.jar:/home/student0/commons-cli-2.0.jar:/home/student0/pig-0.11.0.jar -d . Trim.java

jar -cvf pig-trim.jar com/hadoopbook/pig/Trim.class

REGISTER pig-trim.jar;

B = FOREACH A GENERATE com.hadoopbook.pig.Trim(fruit);

BUMP B;

DESCRIBE B;

**Commands for Chapter 12 – CSUEB**

Commands may be needed to activate Hive on CSUEB Hadoop:

# ls -l | grep meta

should show:

drwxr-xr-x. 5 student0 hadoop 4096 May 17 13:40 metastore\_db

# mv metastore\_db metastore\_db.old

# schematool -dbType derby -initSchema

# hive

set hive.metastore.warehouse.dir;

LOAD DATA LOCAL INPATH 'sample.txt'

OVERWRITE INTO TABLE records;

hdfs dfs -ls /user/hive/warehouse/records/

! <command>: Executes a shell command from the Hive shell.

dfs <dfs command>: Executes a dfs command from the Hive shell.

Commands for Example 12-4 on Page 454:

javac -classpath /home/student1/hadoop-common-2.6.1.jar:/home/student1/hadoop-mapreduce-client-core-2.6.1.jar:/home/student1/commons-cli-2.0.jar:/home/student1/hive-exec-0.13.0.jar -d . Strip.java

jar -cvf hive-strip.jar com/hadoopbook/hive/Strip.class

ADD JAR hive-strip.jar;

CREATE TEMPORARY FUNCTION strip AS 'com.hadoopbook.hive.Strip';

DROP TABLE IF EXISTS dummy;

CREATE TABLE dummy (value STRING);

LOAD DATA LOCAL INPATH 'dummy2.txt'

OVERWRITE INTO TABLE dummy;

(original dummy.txt doesn’t work; need to use mine)

SELECT strip(' bee ') FROM dummy;

SELECT strip('banana', 'ab') FROM dummy;

SELECT STRIP(' bee ') FROM dummy;

SELECT STRIP('0123456789') FROM dummy;

SELECT strip('bananaabf') FROM dummy;

SELECT strip('acbanabcnbcaacb', 'abc') FROM dummy;