

HDFS

1. Create txt file and write sample input using this command :
Ex : vi sample_input.txt
2. Put txt file on a specific location using this command
Ex : hdfs dfs -put sample_input.txt /PATH/

Map Reduce

1. Define variable in Linux using “export” command to save path of my libraries or Jars to use it while compiling my code.
Ex : export LIB_PATH=/usr/lib/Hadoop
2. Using “javac” command to compiling Java source code into bytecode.
The javac command is a tool in Java Development Kit (JDK) used for compiling Java source code into bytecode that can be executed by the Java Virtual Machine (JVM). It is the primary Java compiler that is used to compile Java programs on a command line interface.

Ex : javac -classpath \$JARS wordcount_classes/WordCount.java
3. Create a “Jar”
A "jar" refers to a Java archive file that contains compiled Java classes and associated resources, such as images or configuration files. Jars are used to package and distribute MapReduce jobs that can be executed on a Hadoop cluster.

Ex : export JARS=\$LIB_PATH/client-0.20/hadoop-core.jar:\$LIB_PATH/lib/commons-cli-1.2.jar:\$LIB_PATH/client/hadoop-common.jar:\$LIB_PATH/hadoop-annotations-2.6.0-cdh5.4.2.jar

Ex : jar -cvf wordcount.jar -C wordcount_classes/ .
4. After step 3, I have a Jar ready to run a Hadoop command that will execute a MapReduce job and take a text file (sample_input.txt) created in the HDFS phase.
Ex : hadoop jar wordcount.jar WordCount /user/cloudera/input /user/cloudera/output