

Step 1:

Open Virtual Box and start Cloudera

Step 2:

Open Eclipse on Cloudera

Step 3:

Create a Java Project

-> Enter Project Name

-> Import Libraries

-> Libraries->FileSystem->usr->lib->hadoop->all jar files

-> Add external jars

-> FileSystem->usr->lib->hadoop->client-> all jar files

Step 4:

Create a new JavaClass = WordCount

Step 5:

Copy the code from "<https://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html>"

Step 6:

Export the .jar file of project in /home/cloudera/WordCount.jar

Step 7:

Create a file for input

-> Create a txt file in this folder

cat > /home/cloudera/inputFile.txt (add in text)

Step 8:

Moving this input file to the hadoop file system

-> List Directory in hdfs

hdfs dfs -ls /

-> Create Directory in hdfs

hdfs dfs -mkdir /inputnew

-> List Directory in hdfs

hdfs dfs -ls /

-> Move this file to hadoop env

hdfs dfs -put /home/cloudera/inputFile.txt /inputnew/

-> View this file in hadoop

hdfs dfs -cat /inputnew/inputFile.txt

Step 9:

Run this file using the give command

hadoop jar /home/cloudera/wordcount.jar WordCount

/inputnew/inputFile.txt /output\_new