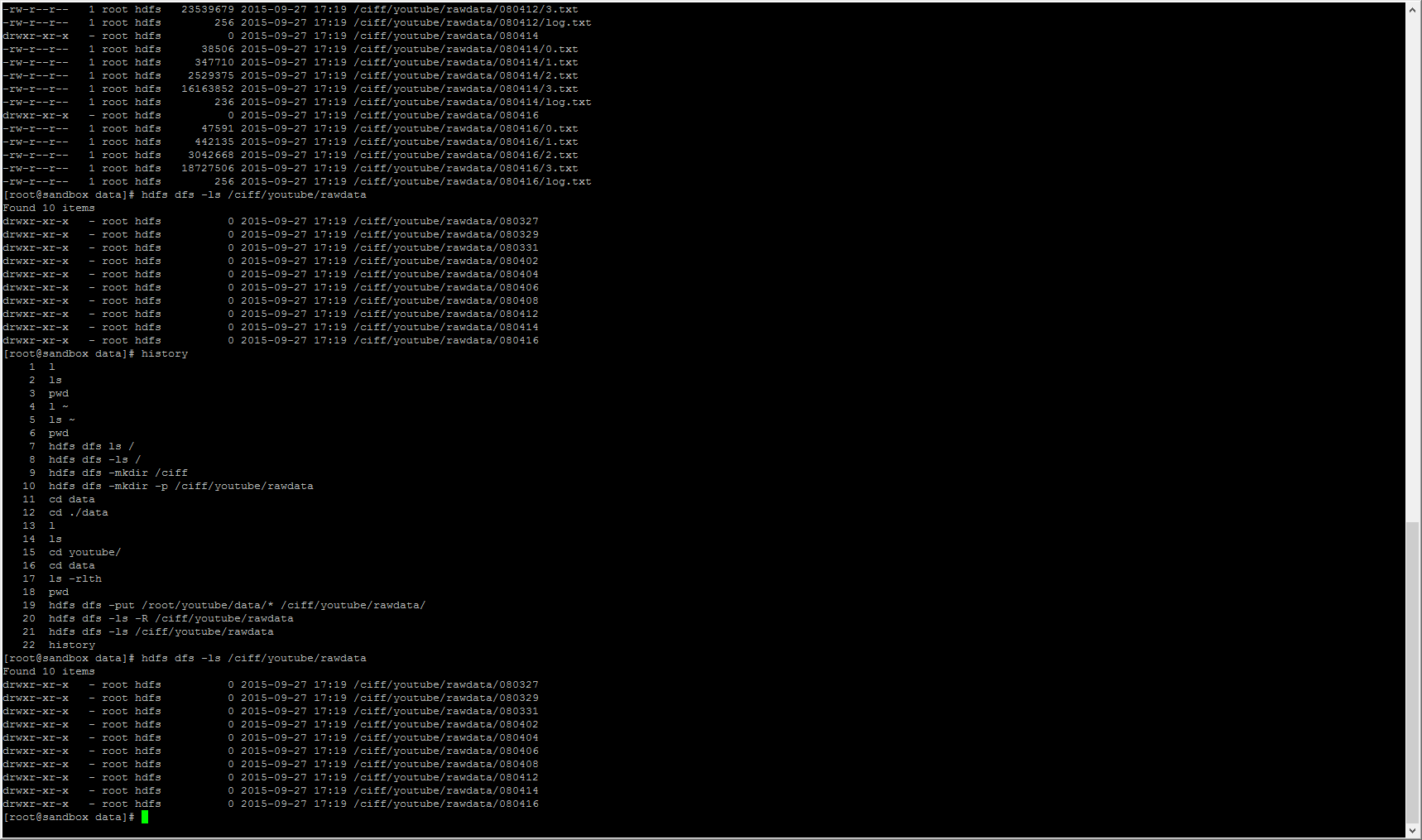
HDFS

1) Ingesta básica en HDFS

> hdfs dfs -put /root/youtube/data/\* /ciff/youtube/rawdata/

> hdfs dfs -ls /ciff/youtube/rawdata



2) Ejercicio 2: Postprocesado básico

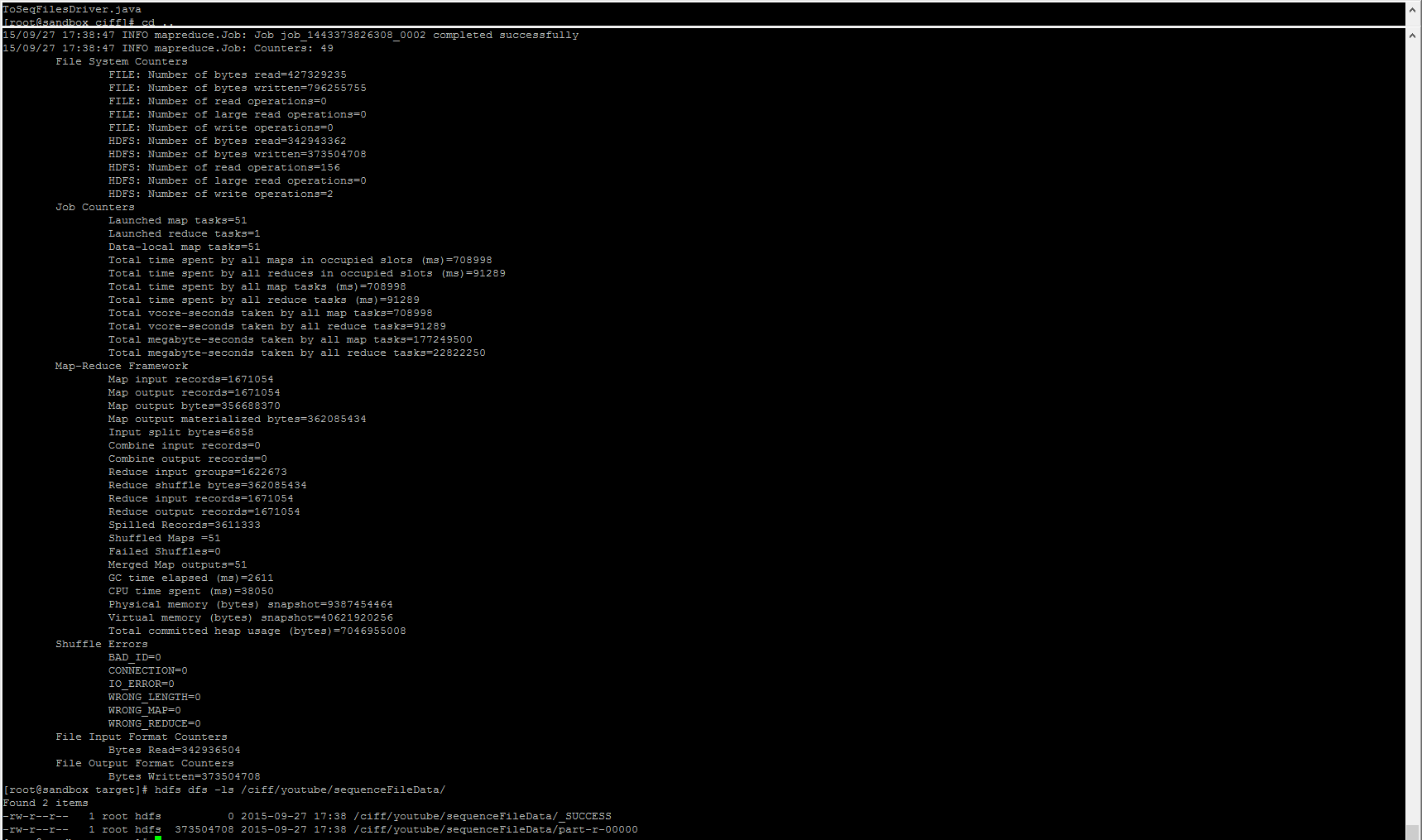
> mvn package

> cd target



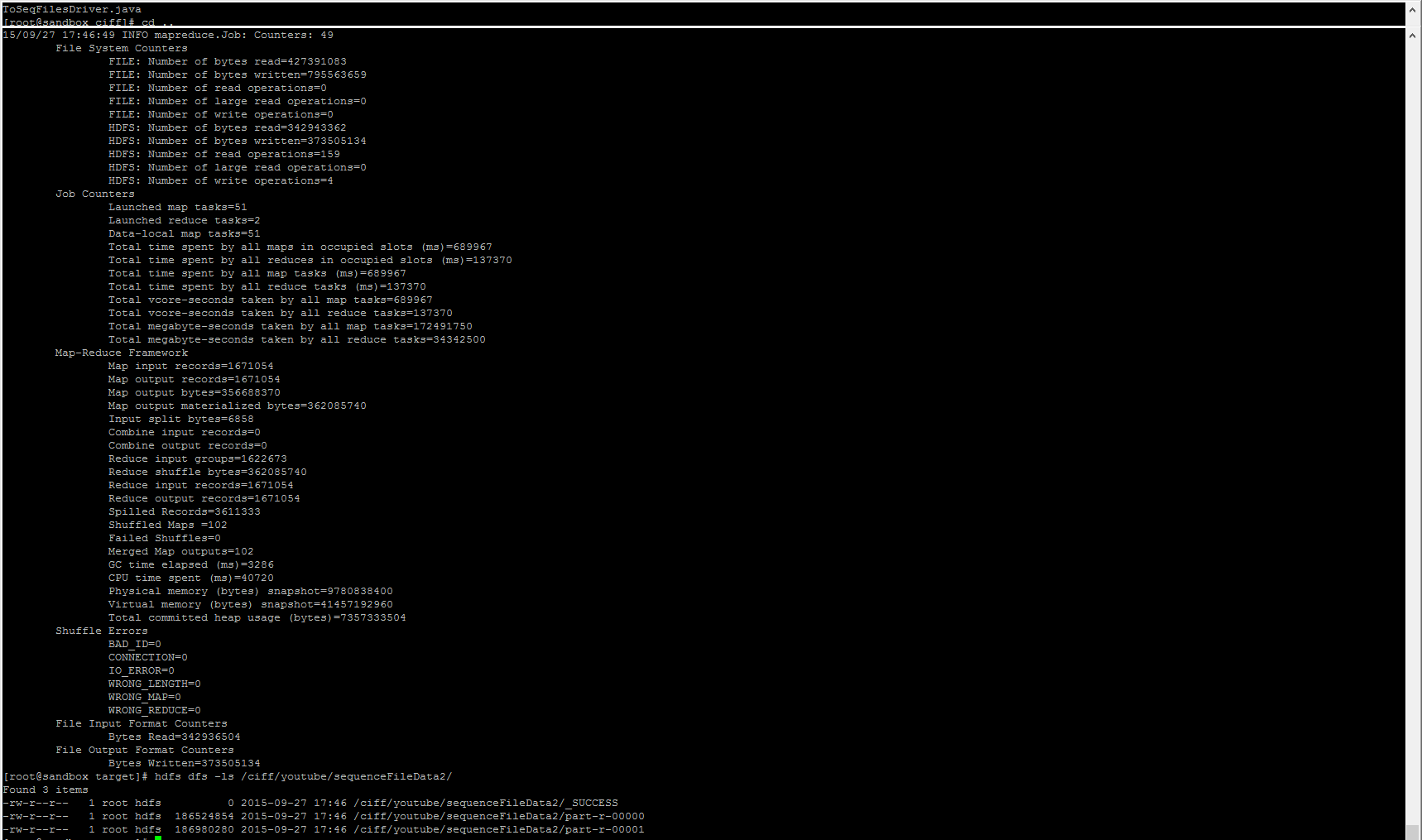
> hadoop jar toseqfiles-1.0-SNAPSHOT.jar ciff.ToSeqFilesDriver /ciff/youtube/rawdata/\*/ /ciff/youtube/sequenceFileData/

> hdfs dfs -ls /ciff/youtube/sequenceFileData/



> hadoop jar toseqfiles-1.0-SNAPSHOT.jar ciff.ToSeqFilesDriver -D mapred.reduce.tasks=2 /ciff/youtube/rawdata/\*/ /ciff/youtube/sequenceFileData2/

> hdfs dfs -ls /ciff/youtube/sequenceFileData2/



3) Ejercicio 3

> hadoop jar /usr/hdp/2.2.4.2-2/hadoop-mapreduce/hadoop-mapreduce-examples.jar sort -r 2 -inFormat org.apache.hadoop.mapreduce.lib.input.KeyValueTextInputFormat -outFormat org.apache.hadoop.mapreduce.lib.output.SequenceFileOutputFormat

-outKey org.apache.hadoop.io.Text -outValue org.apache.hadoop.io.Text /ciff/youtube/rawdata/\*/ /ciff/youtube/sequenceFileData3/

> hdfs dfs -ls /ciff/youtube/sequenceFileData3/

