

# Map reduce

## Iniciar servicios

### Hadoop

Para iniciar los servicios de hadoop. En la virtual (desde la terminal):

- **/home/hadoop/Desktop/curso/bin/start.sh hadoop**
- **/home/hadoop/Desktop/curso/bin/start.sh yarn**

Para verificar, en cualquier terminal al correr el comando **jps** (lista todos los procesos java), deberían ver:

- **NameNode**
- **DataNode**
- **ResourceManager**
- **NodeManager**

Además, deberían tener habilitadas las interfaces gráficas (en firefox):

- <http://localhost:50070> (hadoop - hdfs)
- <http://localhost:8088> (hadoop -yarn)

Usar los scripts de start es lo mismo que desde donde este instalado hadoop (que llamaremos \$HADOOP\_HOME):

- **\$HADOOP\_HOME/sbin/hadoop-daemon.sh start namenode**
- **\$HADOOP\_HOME/sbin/hadoop-daemon.sh start datanode**
- **\$HADOOP\_HOME/sbin/yarn-daemon.sh start resourcemanager**
- **\$HADOOP\_HOME/sbin/yarn-daemon.sh start nodemanager**

### Eclipse

Para iniciar el eclipse. En la virtual (desde la terminal):

- **/home/hadoop/Desktop/curso/bin/start.sh eclipse**

Para importar los proyectos (mapreduce-example):

- Click derecho > import > project
- Maven project

## Proyecto mapreduce-example

Descomprimir:

- Click derecho extraer aqui, o sino...

- 
- **tar xvfz <ruta a mapreduce-example.tar.gz>** (desde la terminal)

## Comandos útiles hadoop

Desde la terminal (asumiendo que \$HADOOP\_HOME es la carpeta donde esta hadoop instalado - /home/hadoop/Desktop/curso/hadoop-2.8.1 en la virtual):

- **\$HADOOP\_HOME/bin/hdfs dfs** (listar los comandos de hadoop hdfs)
- **\$HADOOP\_HOME/bin/hdfs dfs -ls <path-en-hdfs>**  
Ejemplo  
`/home/hadoop/Desktop/curso/hadoop-2.8.1/bin/hdfs dfs -ls /`
- **\$HADOOP\_HOME/bin/hdfs dfs -copyFromLocal <path-local> <path-en-hdfs>**  
Ejemplo  
`/home/hadoop/Desktop/curso/hadoop-2.8.1/bin/hdfs dfs \  
-copyFromLocal \  
/etc/hosts \  
/hosts`
- **\$HADOOP\_HOME/bin/hdfs dfs -copyToLocal <path-en-hdfs> <path-local>**  
Ejemplo  
`/home/hadoop/Desktop/curso/hadoop-2.8.1/bin/hdfs dfs \  
-copyToLocal \  
/README.txt \  
/home/hadoop/`
- **\$HADOOP\_HOME/bin/hdfs dfs -rm -r <path-en-hdfs>**  
Ejemplo  
`/home/hadoop/Desktop/curso/hadoop-2.8.1/bin/hdfs dfs -rm -r /  
README.txt`
- **\$HADOOP\_HOME/bin/hadoop jar <path-al-jar> <clase-del-jar>  
<argumentos>**

## Pasos para correr ejemplos

Los archivos de titanic.csv y ratings.csv los pueden enviar a hdfs con el comando **copyFromLocal** visto arriba. Luego:

- Ir a la carpeta del proyecto (desde la terminal)
- **mvn package** (crea el jar empaquetado)

- 
- `cd target/`
  - `$HADOOP_HOME/bin/hadoop jar mapreduce-example-0.0.1-SNAPSHOT.jar it.dbland.mapreduce.wordCount.WordCountExample /README.txt /mapred-output` (es importante mencionar que fallará si la carpeta de salida ya existe en hdfs. Eliminar con `$HADOOP_HOME/bin/hdfs dfs -rm -r /mapred-output`)