**HADOOP**

**Hdfs dfs**

* Acceso a datos en HDFS a través de la línea de comandos:
  + - Copiar un fichero de disco local a HDFS
      * **$ hadoop fs -put /…/foo.txt /…/foo.txt**
    - Listar el directorio Home del usuario
      * **$ hadoop fs -ls**
    - Listar el directorio root del usuario
      * **$ hadoop fs –ls /**
    - Mostrar el contenido de un fichero, por ej: /user/fred/bar.txt
      * **$ hadoop fs -cat /user/fred/bar.txt**
    - Copiar un fichero de HDFS al local, por ejemplo baz.txt
      * **$ hadoop fs -get /user/fred/bar.txt baz.txt**
    - Crear un directorio en el home de HDFS del usuario llamado input
      * **$ hadoop fs -mkdir input**
    - Borrar un directorio y todo su contenido
      * **$ Hadoop fs -rm -r input\_old**

**MAPREDUCE (workcount)**

La carpeta wordcount, como hemos visto, ya contiene los javas compilados y el jar creado, por lo que solo tenemos que ejecutar el submit del job hadoop usando nuestro fichero JAR para contar las ocurrencias de palabras contenidas en nuestra carpeta “shakespeare”. Nuestro jar contiene las clases java compiladas dentro de un paquete llamado “solutions”, por eso se le llama de este modo.hadoop

**“hadoop jar wordcount/wc.jar solution.WordCount shakespeare /user/cloudera/wordcounts”**

Para ver la lista de Jobs que se están ejecutando

**“mapred job -list”**

Si conocemos la id de un job, lo podemos matar. Recordemos que cerrando un terminal no se mata el job. Para ello, ejecutamos en otra terminal lo siguiente

**“mapred job -kill jobid”**