Exercício 1

([URL IMF Class](https://campusformacionnebrija.imf.com/general/prt/prg/leccgen/newud.php?codcurso=C7274C13&unidad=2&trozo=28&param=contenido))

Partiendo de las letras de las siguientes canciones:

* Nine Inch Nails, "March of the Pigs"
* Pink Floyd, "Pigs (Three Different Ones)"
* Dave Matthews Band, "Pig"

**Se pide:**

1. Crear ficheros de texto en la máquina Ubuntu.
2. Subir los ficheros a HDFS.
3. Ejecutar un job wordcount y resolver las siguientes cuestiones:
   1. Acceder al jobhistory e indicar los valores de "Status" y "Map Total" del job con el nombre "word count".
   2. Palabra que se repite más veces.
   3. Palabra que se repite seis veces.
   4. Cuántas veces aparece la palabra “eyes”.

**Solución:**

**1. Crear ficheros de texto en la máquina Ubuntu**

Crear ficheros de texto con la letra de las canciones:

*mkdir /home/bigdata/hadoop/practica1*

Crear los ficheros con la letra de las canciones:

*sudo nano cancion1.txt && sudo nano cancion2.txt && sudo nano cancion3.txt*

*(colar conteúdo das letras e gravar)*

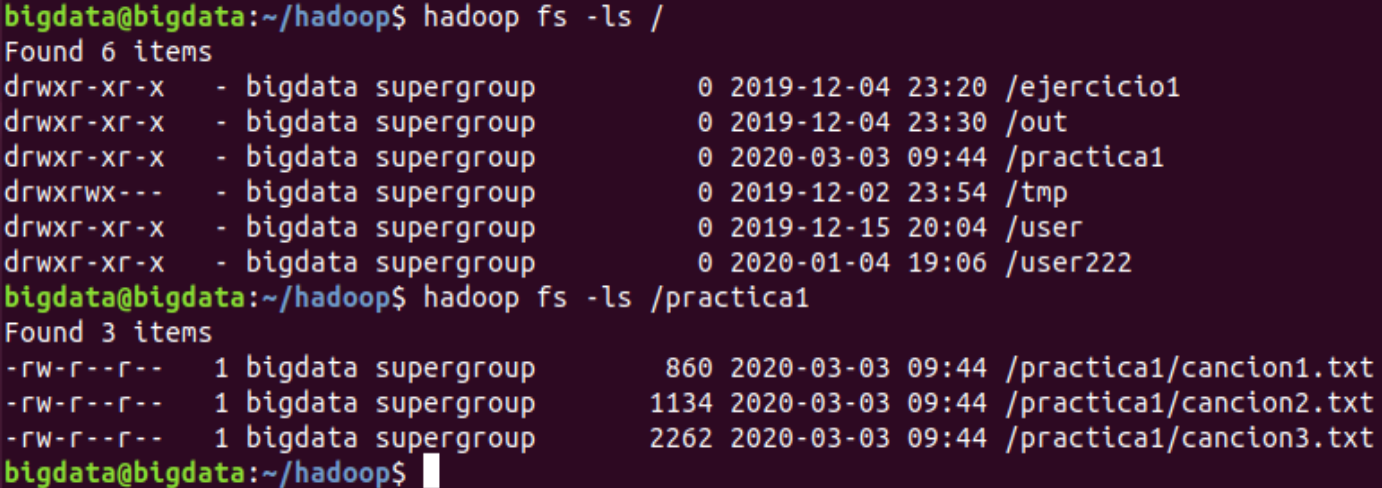
**2. Subir los ficheros a HDFS**

Añadir los ficheros a HDFS:

*bigdata@bigdata:~/hadoop$ hdfs dfs -copyFromLocal practica1 /practica1*

Para verificar:

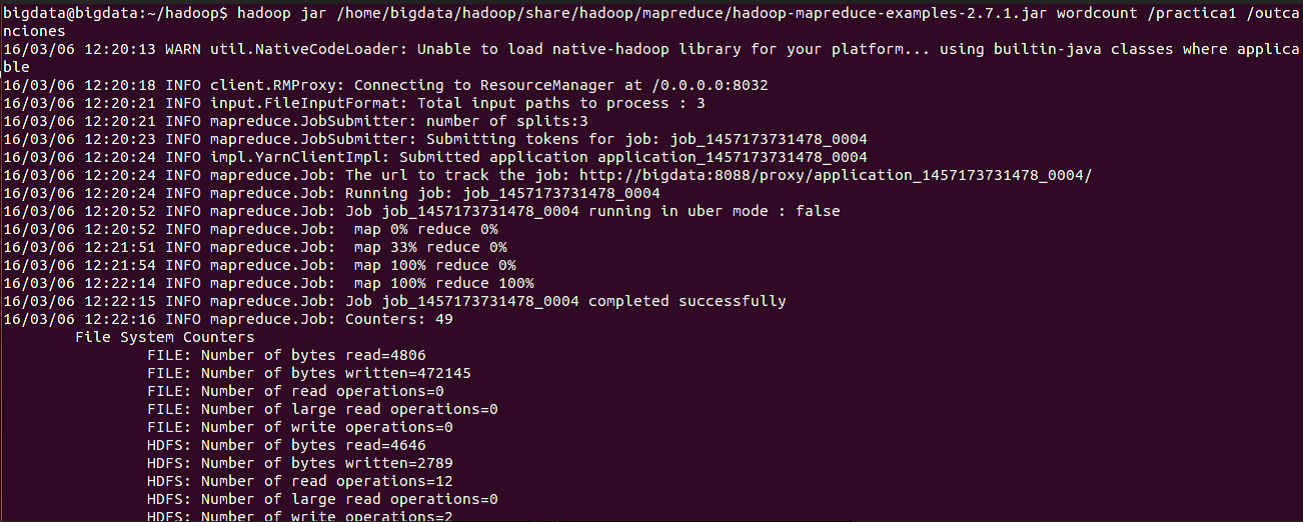
*hadoop fs -ls / && hadoop fs -ls /practica1*

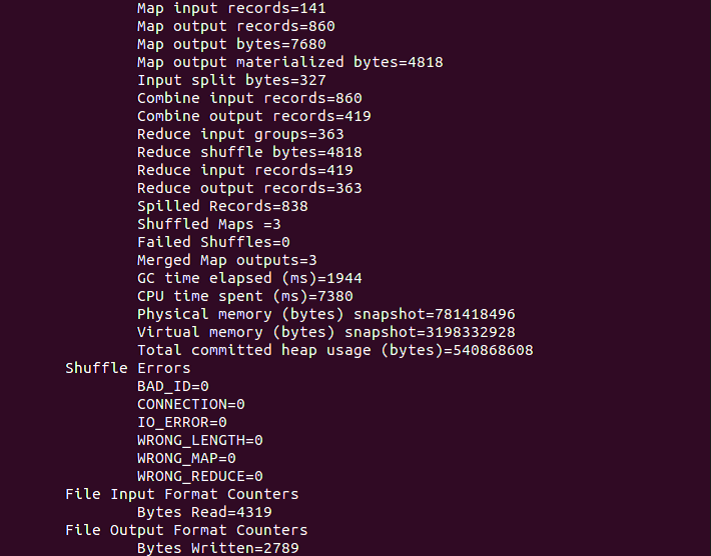


**3. Ejecutar un wordcount y resolver las siguientes cuestiones**

Ejecutar wordcount:

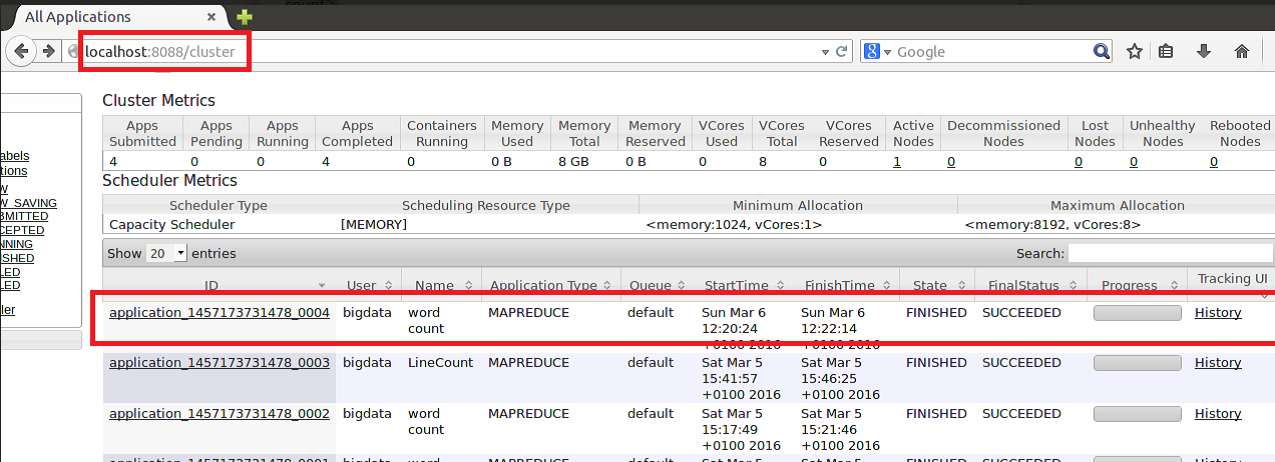
hadoop jar /home/bigdata/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.8.0.jar wordcount /practica1 /outcanciones



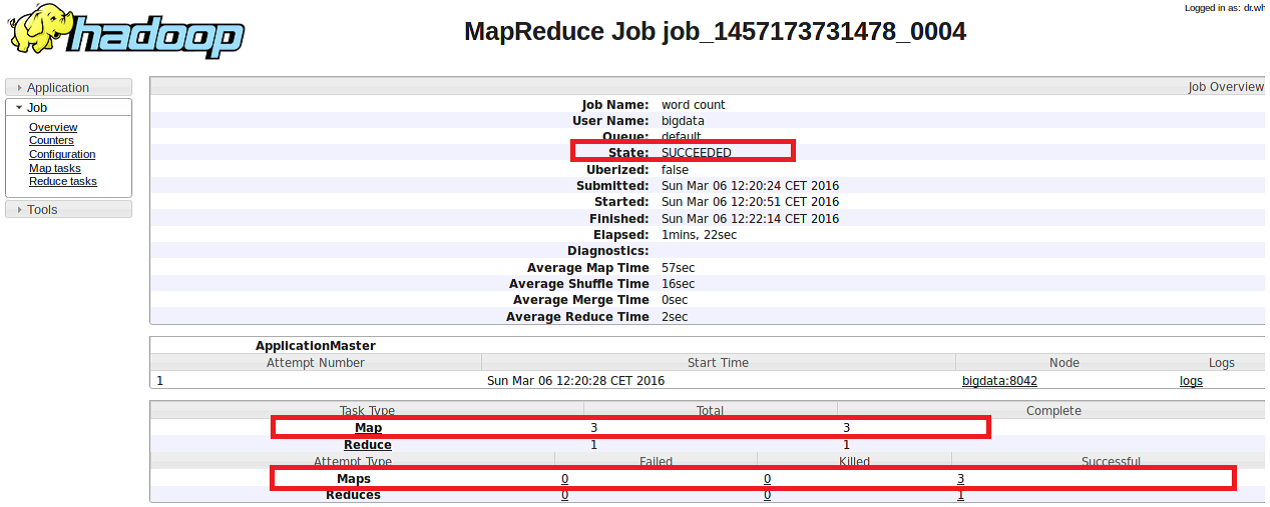


A. Acceder al jobhistory e indicar los valores de "Status" y "Map Total" del job con el nombre "word count":

* Acceder al jobhistoy:



* Comprobar que el Status es SUCCEEDED y el Map Total es 3.



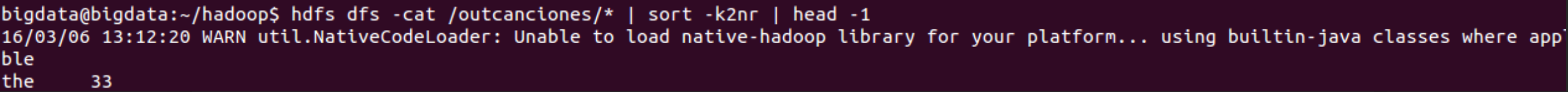
B. Palabra que se repite más veces:

* Revisar la salida:

*bigdata@bigdata:~/hadoop$ bin/hdfs dfs -cat /outcanciones/\**

Para obtener la palabra que más se repite, ordenar la salida por la segunda columna numérica en orden descendente y nos quedamos con la primera:

*hdfs dfs -cat /outcanciones/\* | sort -k2nr | head -1*

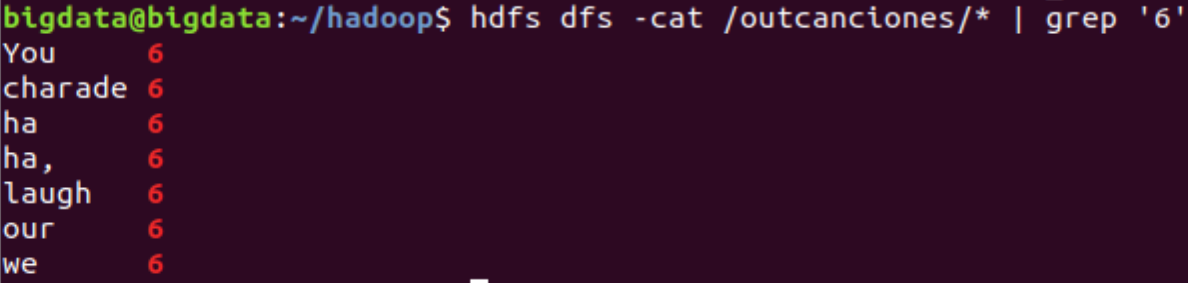


El resultado es la plabra **“the”,**con **33 apariciones**.

C. Palabra que se repite seis veces

* Usamos el comando **grep -** is used to search text. It searches the given file for lines containing a match to the given strings or words.

hdfs dfs -cat /outcanciones/\* | grep '6'

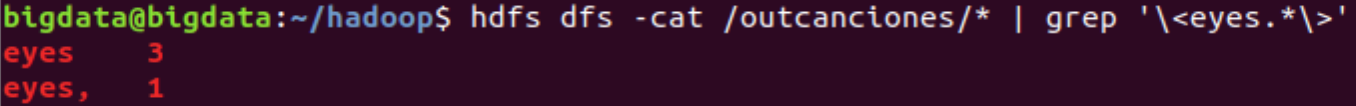


Obtenemos las palabras: “You”, “charade”, “ha”, “laugh”, “our” y “we”.

D. Cuántas veces aparece la palabra “eyes”

* Obtenemos las veces que aparece la palabra “eyes” con el siguiente comando:

*hdfs dfs -cat /outcanciones/\* | grep '\<eyes.\*\>'*



Vemos que aparece **cuatro veces**.