

EJERCICIOS DE HDFS

WORD COUNT

Partiendo de las letras de las siguientes canciones:

NINE INCH NAILS LYRICS "March Of The Pigs"	PINK FLOYD "Pigs (Three Different Ones)"	DAVE MATTHEWS BAND "Pig"
step right up march push crawl right up on your knees please greed feed (no time to hesitate) I want a little bit I want a piece of it I think he's losing it I want to watch it come down don't like the look of it don't like the taste of it don't like the smell of it I want to watch it come down all the pigs are all lined up I give you all that you want take the skin and peel it back now doesn't that make you feel better? shove it up inside surprise! lies stains like the blood on your teeth bite chew suck away the tender parts I want to break it up I want to smash it up I want to fuck it up I want to watch it come down maybe afraid of it let's discredit it let's pick away at it I want to watch it come down now doesn't that make you feel better? the pigs have won tonight now they can all sleep soundly and everything is all right	Big man, pig man, ha ha, charade you are You well heeled big wheel, ha ha, charade you are And when your hand is on your heart You're nearly a good laugh Almost a joker With your head down in the pig bin Saying "keep on digging" Pig stain on your fat chin What do you hope to find? When you're down in the pig mine You're nearly a laugh You're nearly a laugh But you're really a cry. Bus stop rat bag, ha ha, charade you are You fucked up old hag, ha ha, charade you are You radiate cold shafts of broken glass You're nearly a good laugh Almost worth a quick grin You like the feel of steel You're hot stuff with a hat pin And good fun with a hand gun You're nearly a laugh You're nearly a laugh But you're really a cry. Hey you Whitehouse, ha ha, charade you are You house proud town mouse, ha ha, charade you are You're trying to keep our feelings off the street You're nearly a real treat All tight lips and cold feet And do you feel abused?!.....!.....!.....! You gotta stem the evil tide And keep it all on the inside Mary you're nearly a treat Mary you're nearly a treat But you're really a cry.	Isn't it strange How we move our lives for another day Like skipping a beat What if a great wave should wash us all away Just thinking out loud Don't mean to dwell on this dying thing But looking at blood It's alive right now Deep and sweet within Pouring through our veins Intoxicate moving wine to tears Drinking it deep Then an evening spent dancing It's you and me This love will open our world From the dark side we can see a glow of something bright There's much more than we see here Don't burn the day away Is this not enough This blessed sip of life Is it not enough Staring down at the ground Oh then complain and pray more from above Greedy little pig Stop just watch your world trickle away Oh it's your problem now It'll all be dead and gone in a few short years Just love will open our eyes Just love will put the hope in our minds Much more than we could ever know Don't burn the day away Come sister my brother Shake up your bones shake up your feet I'm saying open up And let the rain come pouring in Wash out this tired notion That the best is yet to come But while you're dancing on the ground Don't think of when you're gone

		<p> Love love what more is there We need the light of love in here Don't beat your head Dry your eyes Let the love in there There are bad times But that's ok Just look for love in it Don't burn the day away Look Here are we On this starry night staring into space And I must say I feel as small as dust Lying down here What point could there be troubling Head down wondering what will become of me Why concern we cannot see But no reason to abandon it Time is short but that's all right Maybe I'll go in the middle of the night Take your hands from your eyes, my love Everything must end some time Don't burn the day away Come sister my brother Shake up your bones shake up your feet I'm saying open up And let the rain come flooding in Wash out this tired notion That the best is yet to come But while you're dancing on the ground Don't think of when you're gone Love love what more is there We need the light of love in here Don't beat your head Dry your eyes Let the love in there There are bad times But that's ok Just look for love in it </p>
--	--	--

1. Crear ficheros de texto en la máquina ubuntu

En primer lugar compruebo si están arrancados los demonios:

```
bigdata@bigdata:~/hadoop$ jps
9856 DataNode
10083 SecondaryNameNode
10235 ResourceManager
10366 NodeManager
24203 Jps
8267 JobHistoryServer
9720 NameNode
bigdata@bigdata:~/hadoop$
```

Veo que están arrancados.

Si no estuvieran tendría que arrancarlos con los siguientes comandos:

`./sbin/start-dfs.sh`

`./sbin/start-yarn.sh`

`./sbin/mr-jobhistory-daemon.sh start historyserver`

Me encuentro en la carpeta `home/bigdata/hadoop`. Para acceder más rápido he creado una variable de entorno:

`HADOOP_HOME=/home/bigdata/hadoop`

Con el comando `cd HADOOP_HOME` puedo acceder a la carpeta directamente.

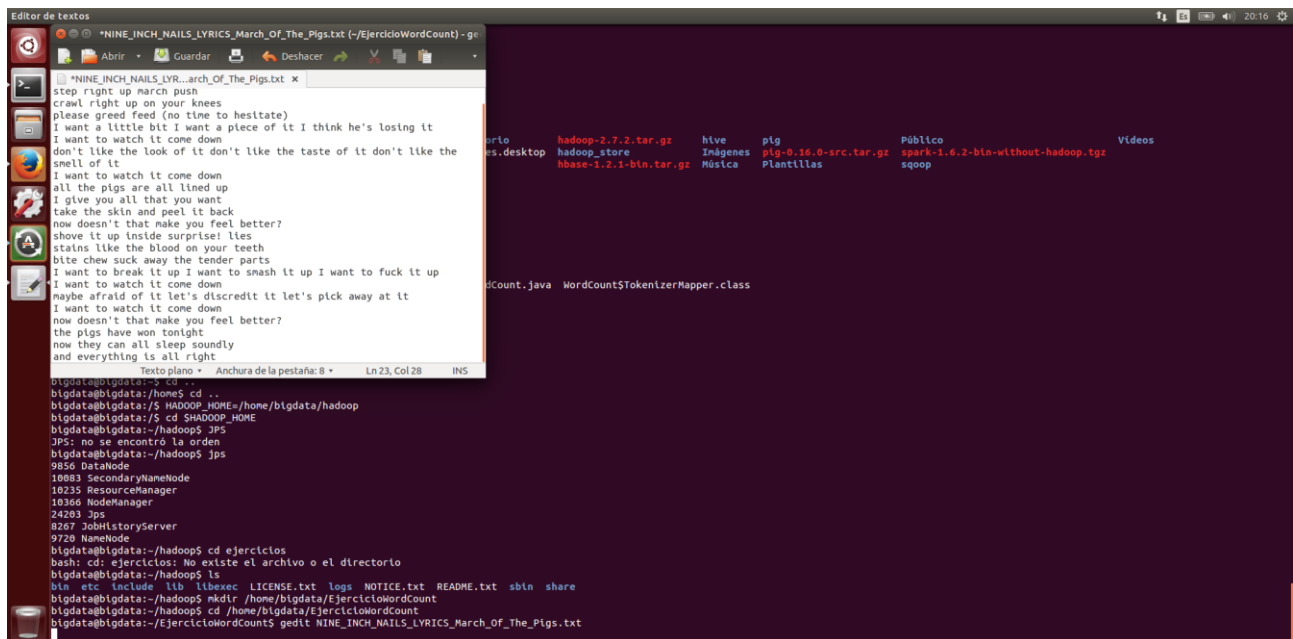
Voy a crear en local una carpeta para el ejercicio WordCount con el comando `mkdir /home/bigdata/EjercicioWordCount`.

A continuación me ubico en la carpeta con el comando `cd`.

```
bigdata@bigdata:~/hadoop$ mkdir /home/bigdata/EjercicioWordCount
bigdata@bigdata:~/hadoop$ cd /home/bigdata/EjercicioWordCount
bigdata@bigdata:~/EjercicioWordCount$
```

Voy a crear 3 ficheros de texto, uno por canción. En cada uno de los ficheros voy a pegar el contenido de la canción, guardar y cerrar.

`gedit NINE_INCH_NAILS_LYRICS_March_Of_The_Pigs.txt`



gedit PINK_FLOYD_Pigs.txt

gedit DAVE_MATTHEWS_BAND_Pig.txt

Compruebo que los ficheros se han creado:

```
bigdata@bigdata:~/EjercicioWordCount$ ls
DAVE_MATTHEWS_BAND_Pig.txt  NINE_INCH_NAILS_LYRICS_March_Of_The_Pigs.txt  PINK_FLOYD_Pigs.txt
bigdata@bigdata:~/EjercicioWordCount$
```

Compruebo el contenido de los ficheros:

```
bigdata@bigdata:~/EjercicioWordCount$ head DAVE_MATTHEWS_BAND_Pig.txt
DAVE MATTHEWS BAND "Pig"
Isn't it strange
How we move our lives for another day
Like skipping a beat
What if a great wave should wash us all away
Just thinking out loud
Don't mean to dwell on this dying thing
But looking at blood
It's alive right now
Deep and sweet within
bigdata@bigdata:~/EjercicioWordCount$ head NINE_INCH_NAILS_LYRICS_March_Of_The_Pigs.txt
NINE INCH NAILS LYRICS "March Of The Pigs"
step right up march push
crawl right up on your knees
please greed feed (no time to hesitate)
I want a little bit I want a piece of it I think he's losing it
I want to watch it come down
don't like the look of it don't like the taste of it don't like the smell of it
I want to watch it come down
all the pigs are all lined up
I give you all that you want
bigdata@bigdata:~/EjercicioWordCount$ head PINK_FLOYD_Pigs.txt
PINK FLOYD "Pigs (Three Different Ones)"
Big man, pig man, ha ha, charade you are
You well heeled big wheel, ha ha, charade you are
And when your hand is on your heart
You're nearly a good laugh
Almost a joker With your head down in the pig bin
Saying "keep on digging" Pig stain on your fat chin
What do you hope to find?
When you're down in the pig mine
You're nearly a laugh You're nearly a laugh
bigdata@bigdata:~/EjercicioWordCount$
```

2. Subir los ficheros a HDFS

Creo en HDFS una carpeta Canciones con `hdfs dfs -mkdir /Canciones` y subo los ficheros de texto a ella:

`hdfs dfs -copyFromLocal EjercicioWordCount / /Canciones`

Compruebo si los ficheros se han subido con: `hdfs dfs -ls -R /Canciones`

```
bigdata@bigdata:~$ hdfs dfs -mkdir /Canciones
16/08/02 20:54:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
bigdata@bigdata:~$ hdfs dfs -copyFromLocal EjercicioWordCount / /Canciones
16/08/02 20:55:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
copyFromLocal: /boot/System.map-3.13.0-32-generic (Permiso denegado)
copyFromLocal: /dev/autofs (Permiso denegado)
-copyFromLocal: Pathname /Canciones/dev/block/11:0 from /Canciones/dev/block/11:0 is not a valid DFS filename.
Usage: hadoop fs [generic options] -copyFromLocal [-f] [-p] [-l] <localsrc> ... <dst>
bigdata@bigdata:~$ hdfs dfs -ls -R /Canciones
16/08/02 20:56:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
drwxr-xr-x  - bigdata supergroup          0 2016-08-02 20:55 /Canciones/EjercicioWordCount
-rw-r--r--  1 bigdata supergroup       2212 2016-08-02 20:55 /Canciones/EjercicioWordCount/DAVE_MATTHEWS_BAND_Pig.txt
-rw-r--r--  1 bigdata supergroup        882 2016-08-02 20:55 /Canciones/EjercicioWordCount/NINE_INCH_NAILS_LYRICS_March_Of_The_Pigs.txt
-rw-r--r--  1 bigdata supergroup       1151 2016-08-02 20:55 /Canciones/EjercicioWordCount/PINK_FLOYD_Pigs.txt
drwxr-xr-x  - bigdata supergroup          0 2016-08-02 20:55 /Canciones/bin
-rw-r--r--  1 bigdata supergroup       1017016 2016-08-02 20:55 /Canciones/bin/bash
```

3. Ejecutar un word count y resolver las preguntas del formulario.

`hadoop jar /home/bigdata/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.2.jar wordcount`

`/Canciones/EjercicioWordCount /ejercicioCanciones_out`

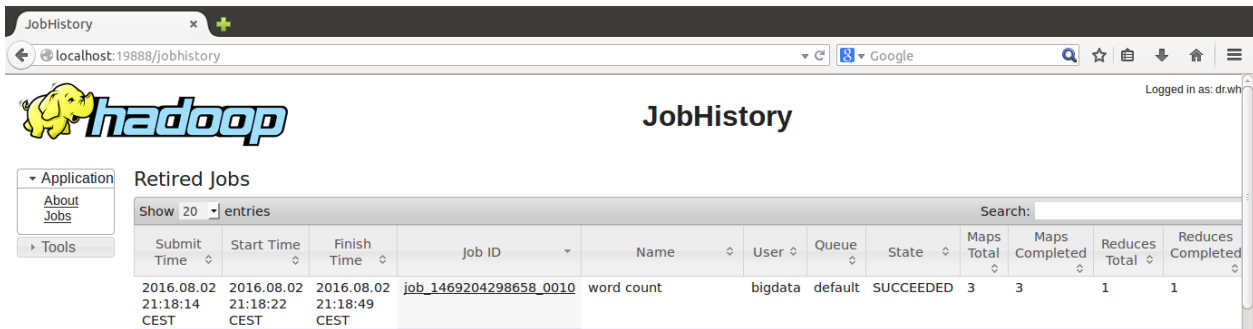
```
bigdata@bigdata:~$
16/08/02 21:18:13 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
16/08/02 21:18:13 INFO Input.FileInputFormat: Total input paths to process : 3
16/08/02 21:18:14 INFO mapreduce.JobSubmitter: number of splits:3
16/08/02 21:18:14 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1469204298658_0010
16/08/02 21:18:14 INFO UnlVarClientImpl: Submitted application application_1469204298658_0010
16/08/02 21:18:15 INFO mapreduce.Job: The url to track the job: http://bigdata:8088/proxy/application_1469204298658_0010/
16/08/02 21:18:15 INFO mapreduce.Job: Running job: job_1469204298658_0010
16/08/02 21:18:24 INFO mapreduce.Job: Job job_1469204298658_0010 running in uber mode : false
16/08/02 21:18:24 INFO mapreduce.Job: map 0% reduce 0%
16/08/02 21:18:42 INFO mapreduce.Job: map 100% reduce 0%
16/08/02 21:18:50 INFO mapreduce.Job: map 100% reduce 100%
16/08/02 21:18:51 INFO mapreduce.Job: Job job_1469204298658_0010 completed successfully
16/08/02 21:18:51 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=4808
  FILE: Number of bytes written=480377
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=4808
  HDFS: Number of bytes written=2791
  HDFS: Number of read operations=12
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
Job Counters
  Launched map tasks=3
  Launched reduce tasks=1
  Data-local map tasks=3
  Total time spent by all maps in occupied slots (ms)=49101
  Total time spent by all reduces in occupied slots (ms)=5601
  Total time spent by all map tasks (ms)=49101
  Total time spent by all reduce tasks (ms)=5601
  Total vcore-milliseonds taken by all map tasks=49101
  Total vcore-milliseonds taken by all reduce tasks=5601
  Total megabyte-milliseonds taken by all map tasks=50279424
  Total megabyte-milliseonds taken by all reduce tasks=5735424
Map-Reduce Framework
  Map input records=134
  Map output records=860
  Map output bytes=7682
  Map output materialized bytes=4820
  Input split bytes=437
  Combine input records=860
  Combine output records=419
  Reduce input groups=363
  Reduce shuffle bytes=4820
  Reduce input records=419
  Reduce output records=363
  Spilled Records=898
  Shuffled Maps=3
  Failed Shuffles=0
File Output Format Counters
  Bytes Written=2791
bigdata@bigdata:~$ hdfs dfs -ls -R /ejercicioCanciones_out
16/08/02 21:22:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
-rw-r--r--  1 bigdata supergroup          0 2016-08-02 21:18 /ejercicioCanciones_out/_SUCCESS
-rw-r--r--  1 bigdata supergroup       2791 2016-08-02 21:18 /ejercicioCanciones_out/part-r-00000
bigdata@bigdata:~$
```

Ejecuto `hdfs dfs -ls -R /ejercicioCanciones_out` y veo que hay un fichero de salida creado.

```
File Output Format Counters
  Bytes Written=2791
bigdata@bigdata:~$ hdfs dfs -ls -R /ejercicioCanciones_out
16/08/02 21:22:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
-rw-r--r--  1 bigdata supergroup          0 2016-08-02 21:18 /ejercicioCanciones_out/_SUCCESS
-rw-r--r--  1 bigdata supergroup       2791 2016-08-02 21:18 /ejercicioCanciones_out/part-r-00000
bigdata@bigdata:~$
```

Ejecuto `hdfs dfs -cat /ejercicioCanciones_out/* > CancionesOut.txt` para llevar el fichero de salida a local y poder consultar el contenido, lo renombro CancionesOut.txt

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



The screenshot shows the Hadoop JobHistory web interface. The browser address bar shows 'localhost:19888/jobhistory'. The page title is 'JobHistory'. On the left, there is a sidebar with 'Application', 'About Jobs', and 'Tools'. The main content area is titled 'Retired Jobs' and contains a table with columns: Submit Time, Start Time, Finish Time, Job ID, Name, User, Queue, State, Maps Total, Maps Completed, Reduces Total, and Reduces Completed. The table shows one job: 'word count' with Job ID 'job_1469204298658_0010', User 'bigdata', Queue 'default', State 'SUCCEEDED', Maps Total '3', Maps Completed '3', Reduces Total '1', and Reduces Completed '1'.

Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed	Reduces Total	Reduces Completed
2016-08-02 21:18:14 CEST	2016-08-02 21:18:22 CEST	2016-08-02 21:18:49 CEST	job_1469204298658_0010	word count	bigdata	default	SUCCEEDED	3	3	1	1

Status es "SUCCEEDED" y Map Total es 3.

B. Palabra que se repite más veces

Con ayuda del comando cat CancionesOut.txt veo cuantas veces se repiten todas las palabras en el fiche CancionesOut.txt
Veo que la palabra "the" se repite más: 33 veces

```
tears 1
teeth 1
tender 1
than 2
that 3
that's 3
the 33
then 1
there 5
they 1
thing 1
```

C. Palabra que se repite 6 veces

Para saber cuál palabra se repite 6 veces ejecuto:
grep 6 CancionesOut.txt

Obtengo lista de palabras que se repiten 6 veces

```
bigdata@bigdata:~$ grep 6 CancionesOut.txt
You 6
charade 6
ha 6
ha, 6
laugh 6
our 6
we 6
```

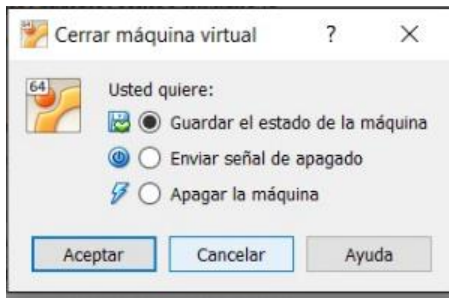
D. Cuántas veces aparece la palabra eyes

Para saber cuantas veces aparece la palabra "eyes" ejecuto:
grep eyes CancionesOut.txt

Obtengo lista de veces que se repite la palabra "eyes".

```
bigdata@bigdata:~$ grep eyes CancionesOut.txt
eyes 3
eyes, 1
bigdata@bigdata:~$
```

Para finalizar cierre y guardo la máquina Archivo/cerrar y guardar el estado de la máquina.



Si no hago eso tendía que parar los demonios:

```
cd $HADOOP_HOME
```

```
./sbin/stop-dfs.sh
```

```
./sbin/stop-yarn.sh
```

```
./sbin/mr-jobhistory-daemon.sh stop historyserver
```