

Solucion tarea 6 parte 1

Configuramos el Classpath para que hadoop encuentre las librerías de Java y lo añadimos a ".bashrc" para que persista.

```
[kali㉿kali)-[~]
$ export HADOOP_CLASSPATH=$JAVA_HOME/lib/tools.jar

[kali㉿kali)-[~]
$ echo 'export HADOOP_CLASSPATH=$JAVA_HOME/lib/tools.jar' >> ~/.bashrc

[kali㉿kali)-[~]
$ source ~/.bashrc
```

Creamos las carpetas para meter el .java

```
[kali㉿kali)-[~]
$ cd /home/kali/Desktop/

[kali㉿kali)-[~/Desktop]
$ mkdir -p ~/Practicas && cd ~/Practicas
```

Comprobamos que el .java está dentro con el quijote.txt y log1.log

```
[kali㉿kali)-[~/Desktop/recursos]
$ ls
contador-palabras.java  log1.log  quijote.txt

[kali㉿kali)-[~/Desktop/recursos]
$
```

Creamos la carpeta en el hdfs llamada prácticas y subimos los ficheros quijote.txt y log1.log

```
\[\e]0;\u0@h: \w\|a\|\[\033[;32m\]\-\(\[\033[1;34m\]\u0@h\[\033[;32m\]\)-[\[\033[0;1m\]\w\[\033[;32m\]]\n\[\033[;
\]\-\[\033[1;34m\]$[\033[0m\] hadoop fs -mkdir -p /practicas

\[\e]0;\u0@h: \w\|a\|\[\033[;32m\]\-\(\[\033[1;34m\]\u0@h\[\033[;32m\]\)-[\[\033[0;1m\]\w\[\033[;32m\]]\n\[\033[;
\]\-\[\033[1;34m\]$[\033[0m\] hadoop fs -put quijote.txt /practicas/
put: `/practicas/quijote.txt': File exists

\[\e]0;\u0@h: \w\|a\|\[\033[;32m\]\-\(\[\033[1;34m\]\u0@h\[\033[;32m\]\)-[\[\033[0;1m\]\w\[\033[;32m\]]\n\[\033[;
\]\-\[\033[1;34m\]$[\033[0m\] hadoop fs -put log1.log /practicas/
put: `/practicas/log1.log': File exists

\[\e]0;\u0@h: \w\|a\|\[\033[;32m\]\-\(\[\033[1;34m\]\u0@h\[\033[;32m\]\)-[\[\033[0;1m\]\w\[\033[;32m\]]\n\[\033[;
\]\-\[\033[1;34m\]$[\033[0m\] hadoop fs -ls /practicas
Found 3 items
-rw-r--r--    1 kali supergroup      0 2025-11-28 05:20 /practicas/log1.log
-rw-r--r--    1 kali supergroup  2161063 2025-11-12 10:15 /practicas/quijote.txt
drwxr-xr-x   - kali supergroup      0 2025-11-12 16:47 /practicas/resultado

\[\e]0;\u0@h: \w\|a\|\[\033[;32m\]\-\(\[\033[1;34m\]\u0@h\[\033[;32m\]\)-[\[\033[0;1m\]\w\[\033[;32m\]]\n\[\033[;
\]\-\[\033[1;34m\]$[\033[0m\]
```

Iniciamos el History Server

```
\[\e]0;\u0@h: \w\@[\u033[;32m\]—(\[\u033[1;34m\]\u@h[\u033[;32m\])-[\\[\u033[0;1m\]\w\[\\033[;32m\]]\n\\[\u033[;32m\]\]—\\[\u033[1;34m\]$\\[\u033[0m\] $HADOOP_HOME/sbin/mr-jobhistory-daemon.sh start historyserver
WARNING: Use of this script to start the MR JobHistory daemon is deprecated.
WARNING: Attempting to execute replacement "mapred --daemon start" instead.

\[\e]0;\u0@h: \w\@[\u033[;32m\]—(\[\u033[1;34m\]\u@h[\u033[;32m\])-[\\[\u033[0;1m\]\w\[\\033[;32m\]]\n\\[\u033[;32m\]\]—\\[\u033[1;34m\]$\\[\u033[0m\] ]
```

Compilamos el primer programa

```
\[\e]0;\u0@h: \w\@[\u033[;32m\]—(\[\u033[1;34m\]\u@h[\u033[;32m\])-[\\[\u033[0;1m\]\w\[\\033[;32m\]]\n\\[\u033[;32m\]\]—\\[\u033[1;34m\]$\\[\u033[0m\] hadoop com.sun.tools.javac.Main ContarPalabras.java
\[\e]0;\u0@h: \w\@[\u033[;32m\]—(\[\u033[1;34m\]\u@h[\u033[;32m\])-[\\[\u033[0;1m\]\w\[\\033[;32m\]]\n\\[\u033[;32m\]\]—\\[\u033[1;34m\]$\\[\u033[0m\] ]
```

Empaquetamos en JAR y verificamos el contenido

```
\[\e]0;\u0@h: \w\@[\u033[;32m\]—(\[\u033[1;34m\]\u@h[\u033[;32m\])-[\\[\u033[0;1m\]\w\[\\033[;32m\]]\n\\[\u033[;32m\]\]—\\[\u033[1;34m\]$\\[\u033[0m\] jar cf contarpalabras.jar Contar*.class
\[\e]0;\u0@h: \w\@[\u033[;32m\]—(\[\u033[1;34m\]\u@h[\u033[;32m\])-[\\[\u033[0;1m\]\w\[\\033[;32m\]]\n\\[\u033[;32m\]\]—\\[\u033[1;34m\]$\\[\u033[0m\] jar tf contarpalabras.jar
META-INF/
META-INF/MANIFEST.MF
ContarPalabras$IntSumReducer.class
ContarPalabras$TokenizerMapper.class
ContarPalabras.class

\[\e]0;\u0@h: \w\@[\u033[;32m\]—(\[\u033[1;34m\]\u@h[\u033[;32m\])-[\\[\u033[0;1m\]\w\[\\033[;32m\]]\n\\[\u033[;32m\]\]—\\[\u033[1;34m\]$\\[\u033[0m\] ]
```

Ejecuta el job indicando JAR, clase principal y rutas HDFS (las dos capturas es el resultado de ejecutar el comando)

```
\[ \e ]0; \u0\h: \w\ a\[ [\ \033[;32m\] —(\[\ \033[1;34m\]\u@ \h\[ \033[;32m\])-[ [\ \033[0;1m\]\w\[ \033[;32m\]]\n\[ \033[;32m\]\]—\[ \033[1;34m\]$\[ \033[0m\] hadoop jar contarpalabras.jar ContarPalabras /practicas/quijote.txt /resultado3
2025-11-28 05:36:35,728 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at localhost/127.0.0.1:8032
2025-11-28 05:36:36,255 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/kali/.staging/job_1764324830287_0001
2025-11-28 05:36:37,329 INFO input.FileInputFormat: Total input files to process : 1
2025-11-28 05:36:37,452 INFO mapreduce.JobSubmitter: number of splits:1
2025-11-28 05:36:37,855 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1764324830287_0001
2025-11-28 05:36:37,856 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-11-28 05:36:38,111 INFO conf.Configuration: resource-types.xml not found
2025-11-28 05:36:38,111 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-11-28 05:36:38,651 INFO impl.YarnClientImpl: Submitted application application_1764324830287_0001
2025-11-28 05:36:38,815 INFO mapreduce.Job: The url to track the job: http://kali:8088/proxy/application_1764324830287_0001/
2025-11-28 05:36:38,816 INFO mapreduce.Job: Running job: job_1764324830287_0001
2025-11-28 05:36:51,444 INFO mapreduce.Job: Job job_1764324830287_0001 running in uber mode : false
2025-11-28 05:36:51,445 INFO mapreduce.Job: map 0% reduce 0%
2025-11-28 05:37:02,260 INFO mapreduce.Job: map 100% reduce 0%
2025-11-28 05:37:11,692 INFO mapreduce.Job: map 100% reduce 100%
2025-11-28 05:37:12,738 INFO mapreduce.Job: Job job_1764324830287_0001 completed successfully
2025-11-28 05:37:12,862 INFO mapreduce.Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=4457125
        FILE: Number of bytes written=9466339
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=2161171
        HDFS: Number of bytes written=448894
        HDFS: Number of read operations=8
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Data-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=8649
        Total time spent by all reduces in occupied slots (ms)=6600
        Total time spent by all map tasks (ms)=8649
        Total time spent by all reduce tasks (ms)=6600
        Total vcore-milliseconds taken by all map tasks=8649
        Total vcore-milliseconds taken by all reduce tasks=6600
        Total megabyte-milliseconds taken by all map tasks=8856576
        Total megabyte-milliseconds taken by all reduce tasks=6758400
```

```
Map-Reduce Framework
  Map input records=37861
  Map output records=384260
  Map output bytes=3688599
  Map output materialized bytes=4457125
  Input split bytes=108
  Combine input records=0
  Combine output records=0
  Reduce input groups=40059
  Reduce shuffle bytes=4457125
  Reduce input records=384260
  Reduce output records=40059
  Spilled Records=768520
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=313
  CPU time spent (ms)=5430
  Physical memory (bytes) snapshot=637251584
  Virtual memory (bytes) snapshot=5454680064
  Total committed heap usage (bytes)=735051776
  Peak Map Physical memory (bytes)=359354368
  Peak Map Virtual memory (bytes)=2727661568
  Peak Reduce Physical memory (bytes)=277897216
  Peak Reduce Virtual memory (bytes)=2727018496
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=2161063
File Output Format Counters
  Bytes Written=448894
```

Comprobamos los resultados

```
\[\e]0;\u001b: \w\|]\[\033[;32m\]—(\[\033[1;34m\]\u001b\h\[\033[;32m\])-[\\033[0;1m\]
\]└\[\033[1;34m\]$\\033[0m\] hadoop fs -ls /resultado3
Found 2 items
-rw-r--r--    1 kali supergroup          0 2025-11-28 05:37 /resultado3/_SUCCESS
-rw-r--r--    1 kali supergroup  448894 2025-11-28 05:37 /resultado3/part-r-00000

\[\e]0;\u001b: \w\|]\[\033[;32m\]—(\[\033[1;34m\]\u001b\h\[\033[;32m\])-[\\033[0;1m\]
\]└\[\033[1;34m\]$\\033[0m\] hadoop fs -cat /resultado3/part-r-00000 | head -n 50
!Mal      1
"Al       1
"Cuando   2
"Cuidados  1
"De       2
"Defects," 1
"Desnudo   1
"Dijo     1
"Dime     1
"Don      1
"Donde    1
"Dulcinea  1
"El       2
"Esta     1
"Harto    1
"Iglesia,  1
"Information 1
"Más      2
"No       5
"Nunca    1
"Plain    2
"Project   5
"Que      1
"Quien    1
"Right    1
"Salta    1
"Sancho   1
"Si       3
"Tened    1
"Toda     1
"Vengan   1
"Vete,    1
"Viose    1
"Vivo     1
"Vuestro   1
"Y        1
"a        3
"alli     1
"bebe    1
"caballeros 1
"cada    1
"castigame 1
"cortés,   1
```

Compilo AnalizarLog.java

```
\[\e]0;\u001b: \w\|]\[\033[;32m\]—(\[\033[1;34m\]\u001b\h\[\033[;32m\])-[\\033[0;1m\]\w\[\033[;32m\]]\n\[\033[;32m\]
\]└\[\033[1;34m\]$\\033[0m\] hadoop com.sun.tools.javac.Main AnalizarLog.java

\[\e]0;\u001b: \w\|]\[\033[;32m\]—(\[\033[1;34m\]\u001b\h\[\033[;32m\])-[\\033[0;1m\]\w\[\033[;32m\]]\n\[\033[;32m\]
\]└\[\033[1;34m\]$\\033[0m\] []
```

Añado el .class al JAR existente

```
\[\e]0;\u001b: \w\|]\[\033[;32m\]—(\[\033[1;34m\]\u001b\h\[\033[;32m\])-[\\033[0;1m\]\w\[\033[;32m\]]\n\[\033[;32m\]
\]└\[\033[1;34m\]$\\033[0m\] jar uf contarpalabras.jar Analizar*.class

\[\e]0;\u001b: \w\|]\[\033[;32m\]—(\[\033[1;34m\]\u001b\h\[\033[;32m\])-[\\033[0;1m\]\w\[\033[;32m\]]\n\[\033[;32m\]
\]└\[\033[1;34m\]$\\033[0m\] []
```

Lanza el job con el fichero de logs en HDFS

```
\[e]0;\u001b: \w\|[\u033[;32m]\|---(\u033[1;34m]\u001b\h[\u033[;32m])-[\u033[0;1m]\w\|\u033[;32m]\n\|\u033[1;\u033[1;34m\$[\u033[0m] hadoop jar contarpalabras.jar AnalizarLog /practicas/log1.log /resultado_log
2025-11-28 06:15:16,878 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at localhost/127.0.0.1:8032
2025-11-28 06:15:17,217 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-11-28 06:15:17,243 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/kali/.staging/job_1764324830287_0002
2025-11-28 06:15:17,581 INFO input.FileInputFormat: Total input files to process : 1
2025-11-28 06:15:18,140 INFO mapreduce.JobSubmitter: number of splits:1
2025-11-28 06:15:18,514 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1764324830287_0002
2025-11-28 06:15:18,514 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-11-28 06:15:18,797 INFO conf.Configuration: resource-types.xml not found
2025-11-28 06:15:18,797 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-11-28 06:15:18,957 INFO impl.YarnClientImpl: Submitted application application_1764324830287_0002
2025-11-28 06:15:19,078 INFO mapreduce.Job: The url to track the job: http://kali:8088/proxy/application_1764324830287_0002/
2025-11-28 06:15:19,085 INFO mapreduce.Job: Running job: job_1764324830287_0002
```

Analizamos la salida

```
\[e]0;\u001b: \w\|[\u033[;32m]\|---(\u033[1;34m]\u001b\h[\u033[;32m])-[\u033[0;1m]\w\|\u033[;32m]\n\|\u033[1;\u033[1;34m\$[\u033[0m] hadoop fs -ls /resultado_log
Found 2 items
-rw-r--r-- 1 kali supergroup          0 2025-11-28 06:15 /resultado_log/_SUCCESS
-rw-r--r-- 1 kali supergroup          0 2025-11-28 06:15 /resultado_log/part-r-00000
```