



CURSO DE ESPECIALIZACIÓN EN INTELIGENCIA ARTIFICIAL Y BIGDATA

BIGDATA APLICADO

TAREA EVALUABLE 1.1

Autor: Carlos Sánchez Recio.
FECHA

Índice

Apartado 1	1
Apartado 2	6
Apartado 3	9

Apartado 1

En el libro de apuntes hemos visto cómo configurar un clúster Hadoop formado por tres nodos, un maestro y dos esclavos. En esta tarea se pide añadir otro nodo esclavo al clúster.

En mi caso, hice uso de un servidor dedicado a virtualizar que poseo para facilitarme la tarea en cuanto adaptadores de red se refiere, ya que utilizo los de la propia red interna. Se trata de un equipo reciclado de 4ta generación de Intel con 24GB de RAM, con pendiente configuración de almacenamiento redundante (actualmente disponibles están unos 600GB), cuyo sistema operativo es un hipervisor de tipo 1, Proxmox. Todo más que suficiente para realizar la tarea.



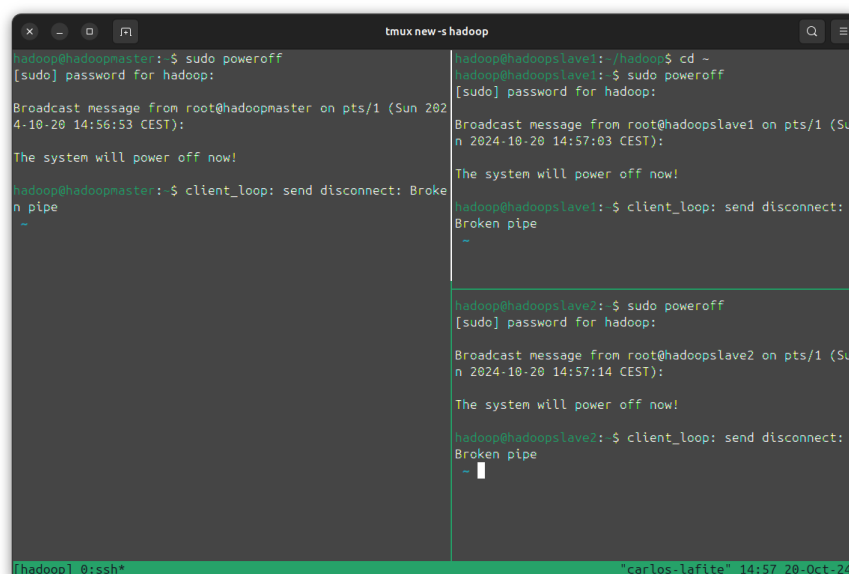
Type	Description	Disk usage...	Memory us...	CPU usage	Uptime	Host CPU ...	Host Mem...	Tags
node	proxmox	1.6 %	28.2 %	0.5% of 8 ...	6 days 17:16...			
qemu	100 (hadoopmaster)	0.0 %	91.4 %	0.9% of 2 ...	03:06:17	0.2% of 8C...	7.8 %	
qemu	101 (hadoopslave1)	0.0 %	74.2 %	0.3% of 2 ...	03:12:56	0.1% of 8C...	6.4 %	
qemu	102 (hadoopslave2)	0.0 %	74.4 %	0.3% of 2 ...	03:08:49	0.1% of 8C...	6.4 %	
sdn	localnetwork (proxmox)							
storage	internet-access (proxmox)	8.2 %						
storage	local (proxmox)	1.6 %						

Figura 1. 1: Panel de control de Proxmox con los 3 nodos ya configurados.

La configuración de IPs para estos equipos es la siguiente:

- hadoopmaster - 192.168.1.120
- hadoopslave1 - 192.168.1.121
- hadoopslave2 - 192.168.1.122
- hadoopslave3 - 192.168.1.123 (para esta tarea)

Para la creación del tercer *DataNode* clonaré uno de los actuales y realizaré la configuración de red tal y como se explican en los apuntes de este bloque. Para ello primero de todo pauso (apagar) todas las máquinas virtuales mediante la ejecución del comando **poweroff** (o **shutdown -h now**) en cada uno de los nodos del clúster (observar los *hostnames*, son 3 sesiones ssh diferentes usando **tmux**).



```

tmux new -s hadoop

hadoop@hadoopmaster:~$ sudo poweroff
[sudo] password for hadoop:
Broadcast message from root@hadoopmaster on pts/1 (Sun 2024-10-20 14:56:53 CEST):

The system will power off now!

hadoop@hadoopmaster:~$ client_loop: send disconnect: Broken pipe
~

hadoop@hadoopslave1:~/hadoop$ cd -
hadoop@hadoopslave1:~$ sudo poweroff
[sudo] password for hadoop:
Broadcast message from root@hadoopslave1 on pts/1 (Sun 2024-10-20 14:57:03 CEST):

The system will power off now!

hadoop@hadoopslave1:~$ client_loop: send disconnect: Broken pipe
~

hadoop@hadoopslave2:~$ sudo poweroff
[sudo] password for hadoop:
Broadcast message from root@hadoopslave2 on pts/1 (Sun 2024-10-20 14:57:14 CEST):

The system will power off now!

hadoop@hadoopslave2:~$ client_loop: send disconnect: Broken pipe
~

```

Figura 1. 2: Apagar todas las máquinas virtuales.

En mi caso para saber de forma gráfica si están apagadas, solo hay que mirar si aparece el icono ► sobre la máquina virtual (comparar con la Figura 1).

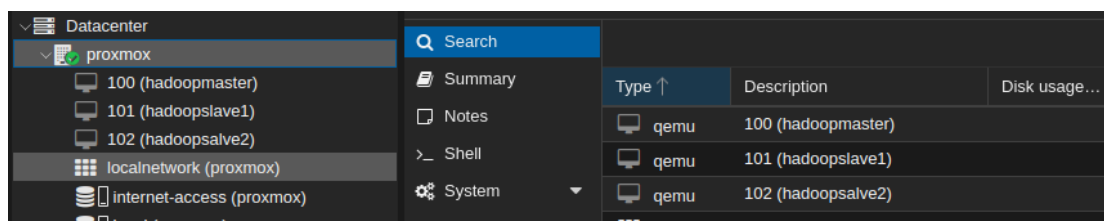


Figura 1. 3: todas las máquinas virtuales apagadas desde la interfaz web de Proxmox.

El siguiente paso es hacer click derecho sobre la máquina virtual a clonar (en mi caso clonaré el segundo esclavo), asignarle un id (automático) y nombre a la nueva máquina virtual, y esperar a que se realice la clonación.

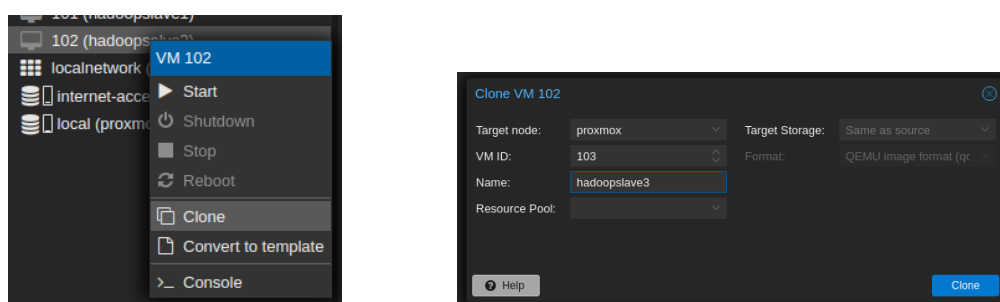
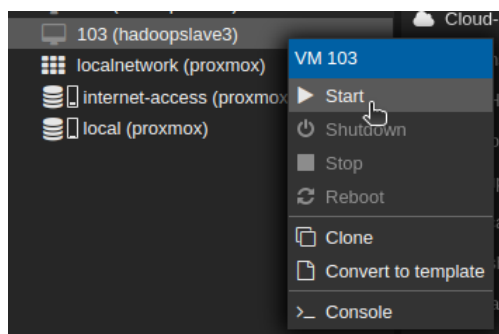
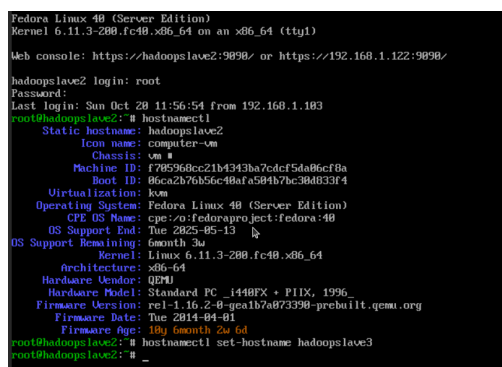


Figura 1. 4: Clonar máquina virtual en Proxmox.

Una vez finalizada la clonación, se inicia la máquina virtual en cuestión y se realiza la configuración mostrada en los apuntes de este bloque. Mostraré únicamente las capturas de pantalla del proceso ya que la explicación de cada paso resultaría redundante a estas alturas.



(a) Iniciar la máquina virtual desde la interfaz gráfica de Proxmox.



(b) Configurar el *hostname* del nuevo nodo.



(c) Configurar la nueva IP mediante el comando *nmtui*.

Figura 1. 5: Configuración del nuevo nodo esclavo o *DataNode* de Hadoop.

```

tmux new-s hadoop

- ssh root@192.168.1.120
root@192.168.1.120's password:
Web console: https://hadoopmaster:9898/ or https://192.168.1.120:9898/
Last login: Sun Oct 20 11:56:17 2024 from 192.168.1.103
root@hadoopmaster:~#

- ssh root@192.168.1.121
root@192.168.1.121's password:
Web console: https://hadoopslave1:9898/ or https://192.168.1.121:9898/
Last login: Sun Oct 20 11:56:37 2024 from 192.168.1.103
root@hadoopslave1:~#

- ssh root@192.168.1.122
root@192.168.1.122's password:
Web console: https://hadoopslave2:9898/ or https://192.168.1.122:9898/
Last login: Sun Oct 20 11:56:54 2024 from 192.168.1.103
root@hadoopslave2:~#

- ssh root@192.168.1.123
The authenticity of host '192.168.1.123 (192.168.1.123)' can't be established.
ED25519 key fingerprint is SHA256:KZINGVC2XBVZ2R2LICWEEZrMBHTrLyMEDUL22A19o.
This host key is known by the following other names/addresses:
- ./ssh/known_hosts:1: [hashed name]
- ./ssh/known_hosts:4: [hashed name]
- ./ssh/known_hosts:5: [hashed name]
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added '192.168.1.123' (ED25519) to the list of known hosts.
root@192.168.1.123's password:
Web console: https://hadoopslave3:9898/ or https://192.168.1.123:9898/
Last login: Sun Oct 20 15:50:41 2024
root@hadoopslave3:~#

```

(d) Crear una sesión SSH para cada uno de los nodos (usando **tmux**).

```

GNU nano 7.2 /etc/hosts
# Loopback entries; do not change.
# For historical reasons, localhost precedes localhost.localdomain:
127.0.0.1 localhost localhost.localdomain localhost4 localhost4.localdomain4
::1 localhost localhost.localdomain localhost6 localhost6.localdomain6
# See hosts(5) for proper format and other examples:
# 192.168.1.10 foo.example.org foo
# 192.168.1.13 bar.example.org bar
192.168.1.120 hadoopmaster
192.168.1.121 hadoopslave1
192.168.1.122 hadoopslave2
192.168.1.123 hadoopslave3

```

(e) Añadir la IP y *hostname* en el archivo `/etc/hosts` de **todos** los nodos del clúster.

```

GNU nano 7.2 /home/hadoop/hadoop/etc/hadoop/hdfs-site.xml Modified
-->
<!-- Put site-specific property overrides in this file. -->

<configuration>
  <property>
    <name>dfs.replication</name>
    <value>3</value>
  </property>
  <property>
    <name>dfs.name.dir</name>
    <value>/home/hadoop/hadoopdata/hdfs/namenode</value>
  </property>
  <property>
    <name>dfs.data.dir</name>
    <value>/home/hadoop/hadoopdata/hdfs/datanode3</value>
  </property>
</configuration>

```

(f) Actualizar el archivo `$HADOOP_HOME/etc/hadoop/hdfs-site.xml` en el tercer *DataNode*.

Figura 1. 5: (*cont.*) Configuración del nuevo nodo esclavo o *DataNode* de Hadoop.

```
GNU nano 7.2 /home/hadoop/hadoop/etc/hadoop/workers Modified
hadoopmaster
hadoopslave1
hadoopslave2
hadoopslave3
```

(g) Añadir el nuevo esclavo en el archivo `$HADOOP_HOME/etc/hadoop/workers` del *NameNode*.

```
GNU nano 7.2 /etc/hosts
# Loopback entries; do not change.
# For historical reasons, localhost precedes localhost.localdomain:
127.0.0.1 localhost localhost.localdomain localhost4 localhost4.localdomain4
::1 localhost localhost.localdomain localhost6 localhost6.localdomain6
# See hosts(5) for proper format and other examples:
# 192.168.1.10 foo.example.org foo
# 192.168.1.13 bar.example.org bar
192.168.1.120 hadoopmaster
192.168.1.121 hadoopslave1
192.168.1.122 hadoopslave2
192.168.1.123 hadoopslave3
```

(h) Añadir la IP y *hostname* en el archivo `/etc/hosts` de **todos** los nodos del clúster.

```
tmux new -s hadoop
hadoop@hadoopmaster:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [hadoopmaster]
Starting datanodes
hadoopslave3: Warning: Permanently added 'hadoopslave3' (ED25519) to the list of known hosts.
Starting secondary namenodes [hadoopmaster]
Starting resourcemanager
Starting nodeManagers
hadoop@hadoopmaster:~$ jps
1328 DataNode
2192 Jps
1286 NameNode
1735 ResourceManager
1517 SecondaryNameNode
1853 NodeManager
hadoop@hadoopmaster:~$

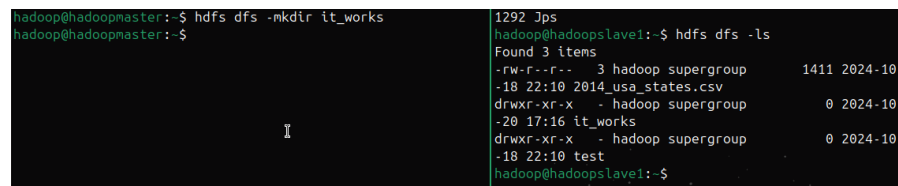
hadoop@hadoopslave1:~$ jps
1176 NodeManager
1852 DataNode
1292 Jps
hadoop@hadoopslave1:~$

hadoop@hadoopslave2:~$ jps
1872 DataNode
1193 NodeManager
1389 Jps
hadoop@hadoopslave2:~$

hadoop@hadoopslave3:~$ jps
1289 DataNode
1323 NodeManager
1439 Jps
hadoop@hadoopslave3:~$
```

(i) Iniciar todos los nodos y comprobar los estados.

Figura 1. 5: (*cont.*) Configuración del nuevo nodo esclavo o *DataNode* de Hadoop.



```
hadoop@hadoopmaster:~$ hdfs dfs -mkdir it_works
hadoop@hadoopmaster:~$
1292 Jps
hadoop@hadoopslave1:~$ hdfs dfs -ls
Found 3 items
-rw-r--r-- 3 hadoop supergroup 1411 2024-10
-18 22:10 2014_usa_states.csv
drwxr-xr-x - hadoop supergroup 0 2024-10
-20 17:16 it_works
drwxr-xr-x - hadoop supergroup 0 2024-10
-18 22:10 test
hadoop@hadoopslave1:~$
```

(j) Comprobar que los nodos se sincronicen.

Figura 1. 5: (*cont.*) Configuración del nuevo nodo esclavo o *DataNode* de Hadoop.

Con esto se puede dar por finalizada la configuración del tercer *DataNode*.

Apartado 2

Tal y como hemos comentado en la actividad de aprendizaje, las Administraciones Públicas, desde el ámbito municipal al internacional, publican gran cantidad de datos mediante sus portales de datos abiertos. Entre otros, tenemos a nuestra disposición el portal de datos abiertos del Gobierno de las Islas Baleares ([Dades Obertes Illes Balears](#)), el del Gobierno de España ([Datos Gobierno Español](#)) y el de la Comisión Europea ([Datos Comisión Europea](#)). Se trata de un sistema jerárquico, por lo que por ejemplo en el portal español podemos encontrar también los datos publicados por las administraciones autonómicas y municipales del país. A lo largo del curso veremos otros repositorios en los que podemos encontrar conjuntos de datos utilizados por la comunidad de ciencia de datos e inteligencia artificial.

En esta tarea vamos a trabajar con datos de calidad del aire de diversas comunidades autónomas. Una vez configurado el clúster Hadoop con 4 nodos, debes descargar desde el nodo maestro los archivos CSV correspondientes a los siguientes conjuntos de datos que puedes encontrar en el portal español de datos abiertos.

- [Estaciones de control de calidad del aire de las Islas Baleares](#). El enlace directo para descargar el archivo CSV es [link](#) (si tiene problemas para acceder a este archivo, también lo puede descargar desde [aquí](#)).
- [Estaciones de control de calidad del aire de la Comunidad de Madrid](#). El enlace directo para descargar el archivo CSV es [link](#).
- [Datos medidos en las estaciones de Cataluña](#). El enlace directo para descargar el archivo CSV es [link](#).

A continuación, debes ejecutar algunas operaciones utilizando los pedidos de FS Shell sobre tu clúster de 4 nodos:

1. Crea un nuevo directorio en HDFS llamado `calidadaire`. Dentro de este directorio crea otras tres: `illes.balears.csv`, `madrid` y `catalunya`.
2. Sube los tres archivos CSV de los datos al directorio `calidadaire`.
3. Cambia el nombre a los archivos, de modo que se llamen respectivamente `illes.balears.csv`, `madrid.csv` y `catalunya.csv`.
4. Copia el archivo `illes.balears.csv` en el directorio `illes.balears.csv` y el archivo `madrid.csv` en el directorio `madrid`.
5. Mueve el archivo `catalunya.csv` al directorio `catalunya`.
6. Borra todos los archivos que hayan quedado en el directorio `calidadaire/`
7. Muestra el contenido del directorio `Cataluña`.
8. Muestra el contenido del archivo de las Islas Baleares.

Antes de mostrar las capturas de pantalla de los comandos ejecutados cabe mencionar que de cara a optimizar los comandos a ejecutar he decidido no seguir al pie de la letra los puntos anteriormente listados y usar una secuencia de comandos más eficiente. A fin de cuentas el resultado final es el mismo, pero sin tanto comando (no ha sido una optimización muy bestia, pero es solo para aclarar que no he seguido los pasos uno por uno).

Para obtener todos los archivos necesarios, teniendo en cuenta que mi entorno gráfico reside en mi máquina anfitrión, he utilizado el comando `wget` seguido del enlace al archivo `.csv` para

todos los enlaces proporcionados. Para cambiar los nombres a los archivos, sencillamente hago uso del comando mv.

```
hadoop@hadoopmaster:~$ ls
2014_usa_states.csv
calidad_aire_estaciones.csv
Estacions_Control_Qualitat_Aire_Illes_Balears.csv
hadoop@hadoopmaster:~$ mv calidad_aire_estaciones.csv madrid.csv
hadoop@hadoopmaster:~$ mv rows.csv catalunya.csv
hadoop@hadoopmaster:~$ mv Estacions_Control_Qualitat_Aire_Illes_Balears.csv baleares.csv
hadoop@hadoopmaster:~$ ls
2014_usa_states.csv
baleares.csv
catalunya.csv
hadoop@hadoopmaster:~$
```

Figura 2. 1: Archivos .csv obtenidos y cambio de nombre.

Mediante los comandos de Hadoop y mkdir, como se muestra en la imagen siguiente, creo los directorios necesarios. Luego muevo dichos archivos al directorio correspondiente.

```
hadoop@hadoopmaster:~$ hdfs dfs -mkdir -p calidad_aire/illes_balears
hadoop@hadoopmaster:~$ hdfs dfs -mkdir -p calidad_aire/madrid
hadoop@hadoopmaster:~$ hdfs dfs -mkdir -p calidad_aire/catalunya
hadoop@hadoopmaster:~$

hadoop@hadoopslave1:~$ hdfs dfs -ls
Found 4 items
-rw-r--r-- 3 hadoop supergroup 1411 2024-10-18 22:10 2014_usa_states.csv
drwxr-xr-x - hadoop supergroup 0 2024-10-20 18:37 calidad_aire
drwxr-xr-x - hadoop supergroup 0 2024-10-20 17:16 it_works
drwxr-xr-x - hadoop supergroup 0 2024-10-18 22:10 test
hadoop@hadoopslave1:~$ hdfs dfs -ls calidad_aire
Found 3 items
drwxr-xr-x - hadoop supergroup 0 2024-10-20 18:37 calidad_aire/catalunya
drwxr-xr-x - hadoop supergroup 0 2024-10-20 18:36 calidad_aire/illes_balears
drwxr-xr-x - hadoop supergroup 0 2024-10-20 18:36 calidad_aire/madrid
hadoop@hadoopslave1:~$
```

(a) Crear los directorios mediante instrucciones de Hadoop.

```
hadoop@hadoopmaster:~$ hdfs dfs -put illes_balears.csv calidad_aire/illes_balears
hadoop@hadoopmaster:~$ hdfs dfs -put madrid.csv calidad_aire/madrid
hadoop@hadoopmaster:~$ hdfs dfs -put catalunya.csv calidad_aire/catalunya
hadoop@hadoopmaster:~$

hadoop@hadoopslave1:~$ hdfs dfs -ls calidad_aire
Found 3 items
drwxr-xr-x - hadoop supergroup 0 2024-10-20 18:40 calidad_aire/catalunya
drwxr-xr-x - hadoop supergroup 0 2024-10-20 18:39 calidad_aire/illes_balears
drwxr-xr-x - hadoop supergroup 0 2024-10-20 18:39 calidad_aire/madrid
hadoop@hadoopslave1:~$ hdfs dfs -ls calidad_aire/madrid
Found 1 items
-rw-r--r-- 3 hadoop supergroup 7593 2024-10-20 18:39 calidad_aire/madrid/madrid.csv
hadoop@hadoopslave1:~$ hdfs dfs -ls calidad_aire/illes_balears
Found 1 items
-rw-r--r-- 3 hadoop supergroup 5765 2024-10-20 18:39 calidad_aire/illes_balears/illes_balears.csv
hadoop@hadoopslave1:~$ hdfs dfs -ls calidad_aire/catalunya
Found 1 items
-rw-r--r-- 3 hadoop supergroup 730269498 2024-10-20 18:40 calidad_aire/catalunya/catalunya.csv
hadoop@hadoopslave1:~$
```

(b) Mover los archivos al directorio correspondiente mediante instrucciones de Hadoop.

Figura 2. 2: Creación de directorios y mover archivos al correspondiente.

Mediante los comandos mostrados, no es necesario borrar archivos que han sido movidos de forma temporal como se dicta en el enunciado.

```
hadoop@hadoopslave1:~$ hdfs dfs -ls calidad_aire/catalunya
Found 1 items
-rw-r--r-- 3 hadoop supergroup 730269498 2024-10-20 18:40 calidad_aire/catalunya/catalunya.csv
hadoop@hadoopslave1:~$
```

Figura 2. 3: Mostrar el contenido del directorio calidad_aire/catalunya/.

```

Codi Local Estació,Nom Estació,Ubicació,Municipi,Altitud,Propietat,Paràmetres mesurats,Tipus d'àrea,Tipus
d'estació,Localització,Enllaç
7040003,Parc de Bellver,Castell de Bellver,Palma,117,Ministeri de Medi Ambient i Conselleria de Medi Ambie
nt,"SO2,NO,NO2,O3,PM10,DD,VV,TMP,HR,RS,PRB,LL",Suburbana,Fons,POINT (2.62169444444 39.5645278),http://www.
caib.es/sites/atmosfera/ca/estacio_de_control_de_la_qualitat_de_laire_de_parc_de_bellver_-_palma_de_mallor
ca_mallorca-3249/
7040006,Sant Joan de Déu,Hospital Sant Joan de Déu,Palma,5,GESA,"SO2,NO,NO2,PM10,O3,DD,VV,TMP,LL",Urbana,I
ndustrial,POINT (2.7 39.5469444),http://www.caib.es/sites/atmosfera/ca/estacio_de_control_de_la_qualitat_d
e_laire_de_lhospital_sant_joan_de_deu_-_palma_de_mallorca_mallorca-35681/
7032002,Pous,"Carretera de Maó a Ciutadella, Km. 1",Maó,55,GESA,"SO2,NO,NO2,PM10,O3,VV,DD,TMP",Urbana,Indu
strial,POINT (4.25297222222 39.8909167),http://www.caib.es/sites/atmosfera/ca/estacio_de_control_de_la_qua
litat_de_laire_de_pous_-_mao_menorca-3261/
7026002,Dalt Vila,"C/ Major, 20",Eivissa,44,GESA,"SO2,NO,NO2,O3,DD,VV,TMP",Urbana,Industrial,POINT (1.4348
8888889 38.9071944),http://www.caib.es/sites/atmosfera/ca/estacio_de_control_de_la_qualitat_de_laire_de_da
lt_vila_eivissa-3258/
7040004,Parc Bit,Edifici Telecomunicacions - Parc Bit,Palma,109,GESA,"SO2,NO,NO2,O3,DD,VV,TMP",Rural,Fons,
POINT (2.63055555556 39.6347222),http://www.caib.es/sites/atmosfera/ca/estacio_de_control_de_la_qualitat_d
e_laire_del_parc_bit_-_palma_de_mallorca_mallorca-3265/
7054001,Torrent,"Subestació de Torrent - Camí Vell de Sant Mateu, s/n",Eivissa,12,GESA,"SO2,NO,NO2,O3,PM10
,DD,VV,TMP",Rural,Fons,POINT (1.43275 38.9364167),http://www.caib.es/sites/atmosfera/ca/estacio_de_control
_de_la_qualitat_de_laire_de_torrent_-_sta_eulària_des_riu_eivissa-3253/
7032003,Port de Maó,c/ Villacarlos s/n,Maó,42,GESA,"SO2,NO,NO2,PM10,O3,DD,VV,TMP",Suburbana,Industrial,POI
NT (4.25705555556 39.8926389),http://www.caib.es/sites/atmosfera/ca/estacio_de_control_de_la_qualitat_de_l
aire_del_port_de_mao_-_mao_menorca/
7015001,Ciutadella,Carrer Carnissers - Polígon industrial,Ciutadella,30,Ministeri de Medi Ambient i Consel
leria de Medi Ambient,"NO,NO2,O3,PM10,TMP,HR,PRB,LL",Suburbana,Fons,POINT (3.85758333333 40.0106111),http:
//www.caib.es/sites/atmosfera/ca/estacio_de_control_de_la_qualitat_de_laire_de_ciutadella_menorca-3255/
70046001,Sant Antoni de Portmany,Carretera Corona - Poliesportiu Can Coix,Sant Antoni de Portmany,65,Minis
teri de Medi Ambient i Conselleria de Medi Ambient,"NO,NO2,O3,PM10,DD,VV,TMP,HR,PRB,LL",Suburbana,Fons,POI
NT (1.31013888889 38.9943889),http://www.caib.es/sites/atmosfera/ca/estacio_de_control_de_la_qualitat_de_l
aire_de_sant_antoni_de_portmany_eivissa-3263/
7019001,Cases de Menut,"Cases de Menut, Serra de Tramuntana",Escorca,584,Conselleria de Medi Ambient,O3,Ru
ral,Fons,POINT (2.90072222222 39.83025),http://www.caib.es/sites/atmosfera/ca/estacio_de_control_de_la_qua
litat_de_laire_de_cases_de_menut_-_ozo_mallorca-80678/
7003001,S'Albufera,"Carretera de Sa Pobla a Alcúdia, Km. 6",Alcúdia,6,GESA,"SO2,NO,NO2,PM10,O3,DD,VV,TMP,P
RB",Rural,Industrial,POINT (3.07525 39.7911667),http://www.caib.es/sites/atmosfera/ca/estacio_de_control_d
e_la_qualitat_de_laire_de_salbufera_-_alcudia_mallorca-3264/
7044001,Sa Pobla,"Carretera de Sa Pobla a Llubí, Km. 2,5 - Finca Sa Canova",Sa Pobla,7,GESA,"SO2,NO,NO2,PM
10,O3,DD,VV,TMP",Rural,Fons,POINT (3.016 39.7484722),http://www.caib.es/sites/atmosfera/ca/estacio_de_cont
rol_de_la_qualitat_de_laire_de_sa_pobla_mallorca-3252/
7032001,Sant Lluís,"Carretera de Maó a Sant Lluís, Km. 2",Maó,65,GESA,"SO2,NO,NO2,PM10,O3,DD,VV,TMP",Urban
a,Industrial,POINT (4.25883333333 39.8748333),http://www.caib.es/sites/atmosfera/ca/estacio_de_control_de_
la_qualitat_de_laire_de_sant_lluis_-_mao_menorca-3248/
7026001,Can Misses,Hospital de Can Misses,Eivissa,25,GESA,"SO2,NO,NO2,PM10,O3,DD,VV,TMP",Urbana,Industrial
,POINT (1.41447222222 38.9140556),http://www.caib.es/sites/atmosfera/ca/estacio_de_control_de_la_qualitat_
de_laire_de_can_misses_eivissa-3260/
7010001,Hospital Joan March,Hospital Joan March,Bunyola,172,TIRME,"SO2,NO,NO2,O3,PM2,5,PM10,DD,VV,TMP,HR,P
RB,LL",Rural,Fons,POINT (2.68777777778 39.6788333),http://www.caib.es/sites/atmosfera/ca/estacio_de_contro
l_de_la_qualitat_de_laire_de_lhospital_joan_march_-_bunyola_mallorca-3259/
Lloseta,Depuradora de Lloseta,Lloseta,,CEMEX,"PM10,PM2,5",Rural,Fons,POINT (2.86491666667 39.7089444),htt
p://www.caib.es/sites/atmosfera/ca/estacio_de_control_de_la_qualitat_de_laire_de_lloseta_mallorca-3267/
7040002,Foners,"Carrer Foners, cantonada amb l'Av. Gabriel Alomar i Vilallonga",Palma,23,Ministeri de Medi
Ambient i Conselleria de Medi Ambient,"SO2,NO,NO2,CO,O3,BEN,TOL,XIL,DD,W,TMP,HR,LL,PM10",Urbana,Trànsit,P
OINT (2.65702777778 39.57125),http://www.caib.es/sites/atmosfera/ca/estacio_de_control_de_la_qualitat_de_l
aire_de_carrer_foners_-_palma_de_mallorca_mallorca-3250/
7027002,Sa Vinyeta d'Inca,Carretera vella d'Inca,Inca,118,GESA,"SO2,NO,NO2,PM10,O3,DD,VV,TMP",Suburbana,Fo
ns,POINT (2.91944444444 39.7155556),http://www.caib.es/sites/atmosfera/ca/estacio_de_control_de_la_qualita
t_de_laire_de_sa_vinyeta_-_inca_mallorca-3251/
7003004,Alcúdia,Antiga central tèrmica d'Alcúdia I - Port d'Alcúdia,Alcúdia,15,GESA,"SO2,NO,NO2,PM10,O3,DD
:

```

Figura 2. 4: Mostrar el contenido del archivo calidad_aire/illes_balears/illes_balears.csv.

Apartado 3

Por último, utilizando la interfaz web de HDFS, muestra cómo está replicado el archivo con los datos de calidad del aire de Cataluña. Tienes que incluir todas las capturas de pantalla para ver cuántos bloques tiene el archivo y en qué nodos se encuentra cada uno de los bloques.

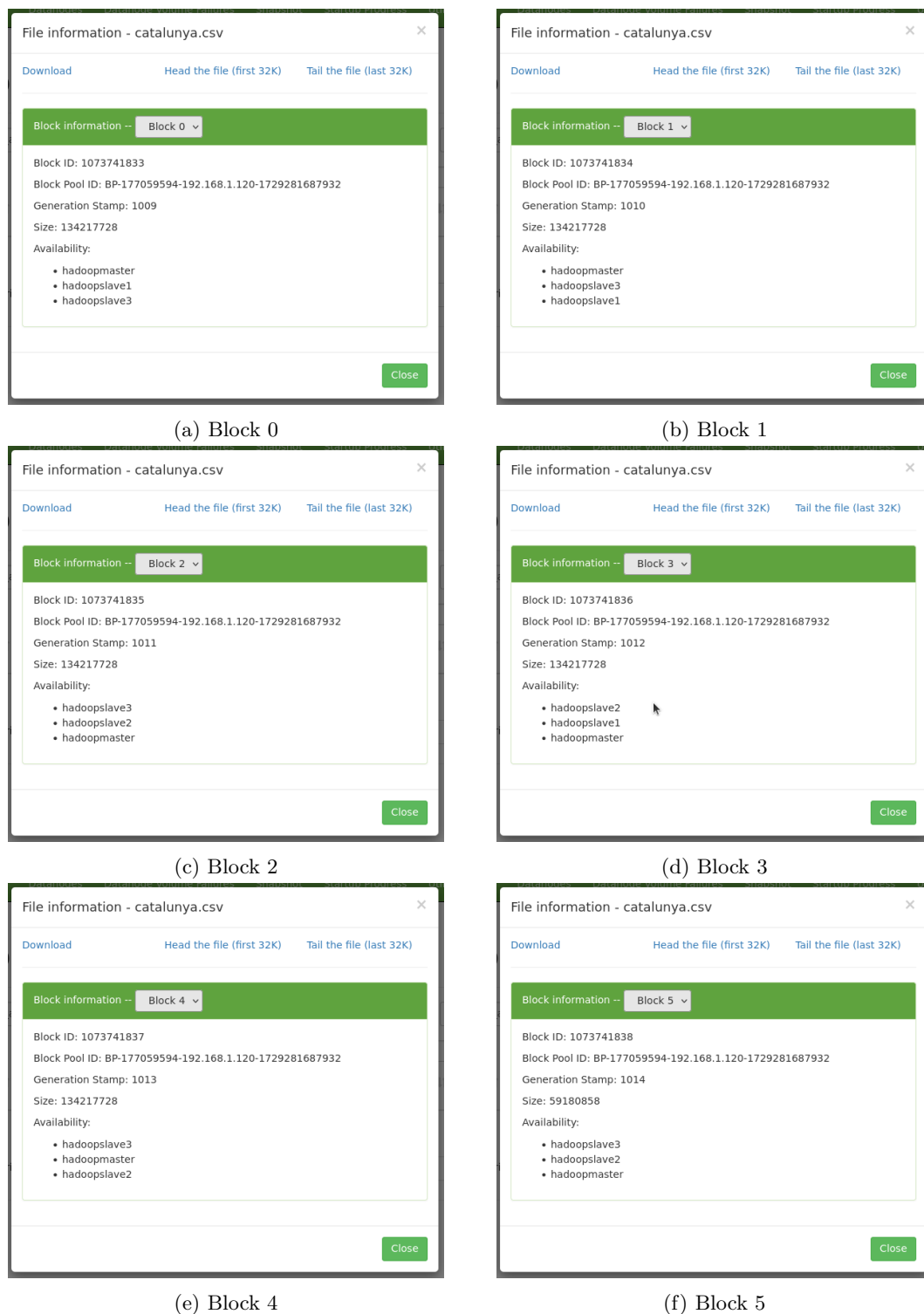


Figura 3. 1: Replicación del archivo `calidad_aire/catalunya/catalunya.csv`.

Como se puede ver, se ha dividido en 6 bloques en total, todos del máximo de la capacidad

permitida por Hadoop (128MB) exceptuando el último que corresponde al restante del archivo. La replicación por los nodos únicamente coincide en que el nodo maestro siempre tiene una copia del bloque. La replicación en los nodos restantes (replicación 3) son repartidos entre los 3 *DataNode* configurados para esta tarea.