

Tarea evaluable

CE_5075 1.1

Big data aplicado

Índice

Comentario sobre mi caso en la práctica:	2
1 - AÑADIR UN NUEVO NODO ESCLAVO AL CLÚSTER	3
2 - OPERACIONES EN FS SHELL	8
1. Crear directorios	8
2: Subir ficheros .csv a los directorios	9
3: Cambio nombres de archivos .csv	9
4: Copiar “illes_balears.csv” y “madrid.csv” en directorios	9
5: Mover archivo “catalunya.csv” al directorio “catalunya”	10
6: Borrar archivos que queden en “qualitat_aire”	10
7: Mostrar el contenido del directorio “catalunya”	11
8: Mostrar el contenido del archivo “illes_balears.csv”	11
3 - INTERFAZ WEB	11

Comentario sobre mi caso en la práctica:

Utilizo un MacBook con chip M3, lo que implica que mi sistema operativo es de Apple y basado en arquitectura ARM. Dado que VirtualBox no es compatible con ARM, tuve que buscar una alternativa, optando finalmente por VMWare. A pesar de algunos inconvenientes iniciales, he logrado aprender a utilizar VMWare y seguir los pasos de configuración indicados en los apuntes. Quería mencionarlo por si las capturas de pantalla generan dudas, ya que no estoy usando VirtualBox.

Un último comentario es que en algunas capturas aparece un Warning, en concreto “WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable”.

No he conseguido solucionar el warning de momento pero no afecta en nada a los ejercicios ya que he podido realizar toda la práctica sin problemas. Por esa razón he decidido destacar con un recuadro rojo los comandos y resultados necesarios para poder hacer una revisión más sencilla y guiada. Trataré de solucionarlo para las próximas entregas. Gracias.

1 - AÑADIR UN NUEVO NODO ESCLAVO AL CLÚSTER

El primer paso es añadir al archivo **/etc/hosts** el que será el nuevo nodo esclavo y su IP correspondiente. Debemos añadirlo en los tres nodos activos que tenemos (hadoopmaster, hadoopslave1 y hadoopslave2)

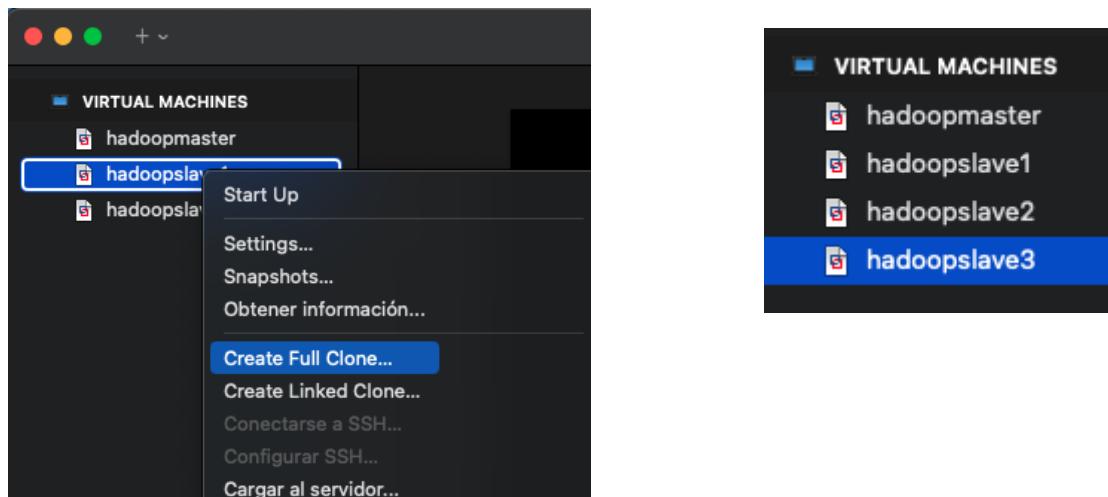
```
root@hadoopmaster:~# nano /etc/hosts
```

```
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

#192.168.10.128 hadoopmaster
10.10.10.1 hadoopmaster
10.10.10.2 hadoopslave1
10.10.10.3 hadoopslave2
10.10.10.4 hadoopslave3
```

Después detenemos las máquinas virtuales y clonamos una de ellas, le cambiamos el nombre para asignarle “*hadoopslave3*”. En mi caso he clonado hadoopslave1 por tema de almacenamiento. Ya que hadoopmaster tiene descargado GNOME para la interfaz gráfica y el navegador Firefox he decidido que era mejor opción clonar una de las máquinas virtuales que contaba con mayor capacidad.

Una vez clonado, accedemos a la configuración de los adaptadores de red y generamos una nueva dirección MAC para cada adaptador.



Mostrar todo hadoopslave3: Network Adapter Agregar dispositivo...

Conectar adaptador de red

Este adaptador de red se ha configurado para utilizar:

Internet Sharing	<input checked="" type="radio"/> Share with my Mac	The virtual machine shares the IP address of the Mac on the external network. The Mac provides Network Address Translation (NAT) for network traffic from the virtual machine.
Bridged Networking	<input checked="" type="radio"/> Autodetect	
	<input type="radio"/> USB 10/100/1000 LAN	
	<input type="radio"/> Wi-Fi	
Custom	<input checked="" type="radio"/> Private to my Mac	

▼ Opciones avanzadas

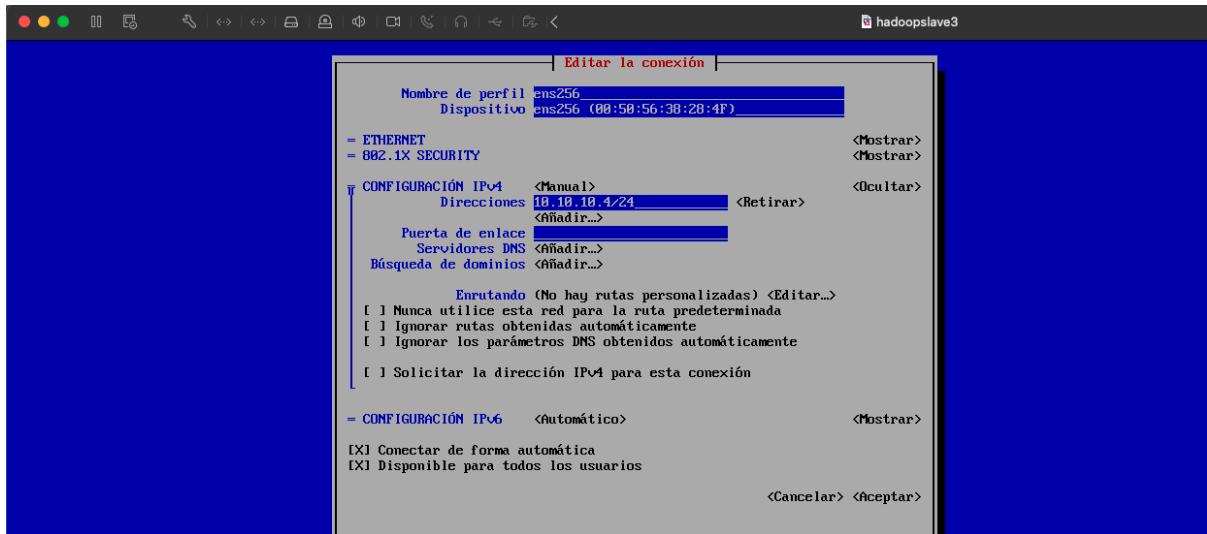
Dirección MAC: 00:50:56:3E:DF:60 **Generar**

Configuración de simulación de red

Transición entrante	Transición saliente
Ancho de banda: Unlimited	Ancho de banda: Unlimited
Kbps: Unlimited	Kbps: Unlimited
Pérdida de paquetes (%): 0.0	Pérdida de paquetes (%): 0.0
Latencia (ms): 0	Latencia (ms): 0

Eliminar adaptador de red ?

Luego pasamos a configurar la IP del nuevo nodo esclavo con la herramienta “nmtui”. Le asignamos la misma ip que definimos anteriormente en los archivos /etc/hosts, en este caso 10.10.10.4



Comprobamos que la opción “Conectar de forma automática” esté activada y guardamos. Lo único que queda ahora antes de probar la conexión entre máquinas es modificar el nombre de host de la máquina virtual

```
root@hadoopslave1:~# hostnamectl set-hostname hadoopslave3
```

Reiniciamos la máquina y podemos confirmar que se ha cambiado el nombre

```
Fedora Linux 40 (Server Edition)
Kernel 6.8.5-301.fc40.aarch64 on an aarch64 (tty1)

Web console: https://hadoopslave3:9090/ or https://192.168.1.10:9090

hadoopslave3 login: root
Password:
Last login: Mon Oct 21 01:58:47 on tty1
root@hadoopslave3:~# _
```

Podemos pasar a hacer ping entre las máquinas para asegurar que la conexión funciona correctamente. Tanto desde los nodos que ya teníamos creados como desde el nuevo nodo esclavo

```
hadoop@hadoopmaster:~$ ping hadoopslave3
PING hadoopslave3 (10.10.10.4) 56(84) bytes of data.
64 bytes from hadoopslave3 (10.10.10.4): icmp_seq=1 ttl=64 time=0.784 ms
64 bytes from hadoopslave3 (10.10.10.4): icmp_seq=2 ttl=64 time=0.665 ms

hadoop@hadoopslave1:~$ ping hadoopslave3
PING hadoopslave3 (10.10.10.4) 56(84) bytes of data.
64 bytes from hadoopslave3 (10.10.10.4): icmp_seq=1 ttl=64 time=0.354 ms
64 bytes from hadoopslave3 (10.10.10.4): icmp_seq=2 ttl=64 time=0.768 ms

hadoop@hadoopslave2:~$ ping hadoopslave3
PING hadoopslave3 (10.10.10.4) 56(84) bytes of data.
64 bytes from hadoopslave3 (10.10.10.4): icmp_seq=1 ttl=64 time=0.713 ms
64 bytes from hadoopslave3 (10.10.10.4): icmp_seq=2 ttl=64 time=0.818 ms
64 bytes from hadoopslave3 (10.10.10.4): icmp_seq=3 ttl=64 time=0.976 ms

hadoop@hadoopslave3:~$ ping hadoopmaster
PING hadoopmaster (10.10.10.1) 56(84) bytes of data.
64 bytes from hadoopmaster (10.10.10.1): icmp_seq=1 ttl=64 time=0.584 ms
64 bytes from hadoopmaster (10.10.10.1): icmp_seq=2 ttl=64 time=1.15 ms
^C
--- hadoopmaster ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 1008ms
rtt min/avg/max/mdev = 0.584/0.866/1.148/0.282 ms
hadoop@hadoopslave3:~$ ping hadoopslave1
PING hadoopslave1 (10.10.10.2) 56(84) bytes of data.
64 bytes from hadoopslave1 (10.10.10.2): icmp_seq=1 ttl=64 time=0.750 ms
64 bytes from hadoopslave1 (10.10.10.2): icmp_seq=2 ttl=64 time=0.589 ms
^C
--- hadoopslave1 ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 1051ms
rtt min/avg/max/mdev = 0.589/0.669/0.750/0.080 ms
hadoop@hadoopslave3:~$ ping hadoopslave2
PING hadoopslave2 (10.10.10.3) 56(84) bytes of data.
64 bytes from hadoopslave2 (10.10.10.3): icmp_seq=1 ttl=64 time=0.612 ms
64 bytes from hadoopslave2 (10.10.10.3): icmp_seq=2 ttl=64 time=0.899 ms

--- hadoopslave2 ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 1055ms
rtt min/avg/max/mdev = 0.612/0.755/0.899/0.143 ms
^C
```

Continuamos cambiando el directorio relacionado con el datanode de hadoopslave3 para que no sea el mismo que hadoopslave1 (ya que lo hemos clonado de ese nodo esclavo).

Simplemente accedemos a **\$HADOOP_HOME/etc/hadoop/hdfs-site.xml** y modificamos la propiedad **dfs.dta.dir** en la que pondremos **/home/hadoop/hadoopdata/hdfs/datanode3**.

```

<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>

```

Tras modificar este archivo sólo quedará dirigirnos a hadoopmaster y añadir el nuevo nodo hadoopslave3 al archivo **\$HADOOP_HOME/etc/hadoop/workers**. Esto servirá para que el nodo maestro sepa cuáles son los nodos del clúster o “workers”.

```

GNU nano 7.2
localhost
hadoopmaster
hadoopslave1
hadoopslave2
hadoopslave3

```

Ya podemos ejecutar el comando “*start-all.sh*” para arrancar el clúster desde el nodo maestro. Usando “*jps*” en cada nodo podemos ver que los procesos están en marcha. Además vemos que desde hadoopslave3 tenemos acceso al directorio “prova3” creado anteriormente.

```

hadoop@hadoopmaster:~/hadoop/etc/hadoop$ jps
1760 NameNode
2422 ResourceManager
2231 SecondaryNameNode
1963 DataNode
2798 NodeManager
3151 Jps

```

```

hadoop@hadoopslave1:~$ jps
1751 Jps
1608 NodeManager
1486 DataNode

```

```

hadoop@hadoopslave2:~$ jps
1763 Jps
1620 NodeManager
1498 DataNode

```

```

hadoop@hadoopslave3:~/hadoop/etc/hadoop$ jps
1411 DataNode
1670 Jps
1532 NodeManager

```

```

hadoop@hadoopslave3:~/hadoop/etc/hadoop$ hdfs dfs -ls prova3
2024-10-21 20:55:39,282 WARN util.NativeCodeLoader: Unable to load native-had
Found 2 items
-rw-r--r-- 3 hadoop supergroup 1411 2024-10-20 13:22 prova3/2014_usa_
-rw-r--r-- 3 hadoop supergroup 359317765 2024-10-20 23:47 prova3/rows.csv

```

2 - OPERACIONES EN FS SHELL

- 1. Crear directorios

Para crear el directorio “qualitat_aire” primero arrancamos el clúster con “start-all.sh” en el nodo maestro “hadoopmaster”. Una vez el clúster está activado creamos el directorio con el comando “**hdfs dfs -mkdir qualitat_aire**”

```
hadoop@hadoopmaster:~$ hdfs dfs -mkdir qualitat_aire
2024-10-22 23:24:41,858 WARN util.NativeCodeLoader: U
sses where applicable
hadoop@hadoopmaster:~$
```

A continuación podemos crear los tres subdirectorios. Con los comandos:

- **hdfs dfs -mkdir qualitat_aire/illes_balears**
- **hdfs dfs -mkdir qualitat_aire/madrid**
- **hdfs dfs -mkdir qualitat_aire/catalunya**

Por último comprobamos los directorios creados usando:

“**hdfs dfs -ls qualitat_aire**”

```
sses where applicable
hadoop@hadoopmaster:~$ hdfs dfs -mkdir qualitat_aire/illes_balears
2024-10-22 23:36:15,445 WARN util.NativeCodeLoader: Unable to load
sses where applicable
hadoop@hadoopmaster:~$ hdfs dfs -mkdir qualitat_aire/madrid
2024-10-22 23:36:28,721 WARN util.NativeCodeLoader: Unable to load
sses where applicable
hadoop@hadoopmaster:~$ hdfs dfs -mkdir qualitat_aire/catalunya
```

```
hadoop@hadoopmaster:~$ hdfs dfs -ls qualitat_aire
2024-10-22 23:37:19,040 WARN util.NativeCodeLoader: Unable to load native-hadoop library
sses where applicable
Found 3 items
drwxr-xr-x  - hadoop supergroup          0 2024-10-22 23:36 qualitat_aire/catalunya
drwxr-xr-x  - hadoop supergroup          0 2024-10-22 23:36 qualitat_aire/illes_balears
drwxr-xr-x  - hadoop supergroup          0 2024-10-22 23:36 qualitat_aire/madrid
```

- 2: Subir ficheros .csv a los directorios

Nos ubicamos en el directorio donde están los archivos y desde ahí ejecutamos:

- **hdfs dfs -put {archivo.csv} qualitat_aire**

```
hadoop@hadoopmaster:~$ hdfs dfs -put Estacions_Control_Qualitat_Aire_Illes_Balears.csv qualitat_aire
2024-10-22 23:50:20,579 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
sses where applicable
hadoop@hadoopmaster:~$ hdfs dfs -put calidad_aire_estaciones.csv qualitat_aire
2024-10-22 23:50:20,579 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
sses where applicable
```

```
hadoop@hadoopmaster:~$ hdfs dfs -put rows.csv qualitat_aire
```

```
hadoop@hadoopmaster:~$ hdfs dfs -ls qualitat_aire
2024-10-23 00:07:53,004 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
sses where applicable
Found 6 items
-rw-r--r-- 3 hadoop supergroup 5765 2024-10-22 23:50 qualitat_aire/Estacions_Control_Qualitat_Aire_Illes_Balears.csv
-rw-r--r-- 3 hadoop supergroup 7593 2024-10-22 23:50 qualitat_aire/calidad_aire_estaciones.csv
drwxr-xr-x - hadoop supergroup 0 2024-10-22 23:36 qualitat_aire/catalunya
drwxr-xr-x - hadoop supergroup 0 2024-10-22 23:36 qualitat_aire/illes_balears
drwxr-xr-x - hadoop supergroup 0 2024-10-22 23:36 qualitat_aire/madrid
-rw-r--r-- 3 hadoop supergroup 730413175 2024-10-23 00:06 qualitat_aire/rows.csv
```

- 3: Cambio nombres de archivos .csv

Con el comando “**hdfs dfs -mv {nombre_antiguo} {nombre_nuevo}**” renombramos los archivos.

```
hadoop@hadoopmaster:~$ hdfs dfs -mv qualitat_aire/rows.csv qualitat_aire/catalunya.csv
```

```
hadoop@hadoopmaster:~$ hdfs dfs -mv qualitat_aire/calidad_aire_estaciones.csv qualitat_aire/madrid.csv
```

```
hadoop@hadoopmaster:~$ hdfs dfs -mv qualitat_aire/Estacions_Control_Qualitat_Aire_Illes_Balears.csv qualitat_aire/il
les_balears.csv
```

- 4: Copiar “illes_balears.csv” y “madrid.csv” en directorios

Para copiar los archivos deseados usamos el comando “**hdfs dfs -cp {ruta_antigua} {ruta_nueva}**”

```
hadoop@hadoopmaster:~$ hdfs dfs -cp qualitat_aire/illes_balears.csv qualitat_aire/illes_balears/
2024-10-23 00:33:05,094 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
hadoop@hadoopmaster:~$ hdfs dfs -cp qualitat_aire/madrid.csv qualitat_aire/madrid/
2024-10-23 00:33:57,681 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
```

```
hadoop@hadoopmaster:~$ hdfs dfs -ls qualitat_aire/illes_balears
2024-10-23 00:35:20,689 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
using builtin-java classes where applicable
Found 1 items
-rw-r--r-- 3 hadoop supergroup 5765 2024-10-23 00:33 qualitat_aire/illes_balears/illes_balears.csv
hadoop@hadoopmaster:~$ hdfs dfs -ls qualitat_aire/madrid
2024-10-23 00:35:28,290 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
using builtin-java classes where applicable
Found 1 items
-rw-r--r-- 3 hadoop supergroup 7593 2024-10-23 00:33 qualitat_aire/madrid/madrid.csv
```

- 5: Mover archivo “catalunya.csv” al directorio “catalunya”

Volvemos a usar el comando “**hdfs dfs -mv {ruta_antigua} {ruta_nueva}**” pero en este caso simplemente para mover el archivo y no para renombrarlo.

```
hadoop@hadoopmaster:~$ hdfs dfs -mv qualitat_aire/catalunya.csv qualitat_aire/catalunya/
2024-10-23 00:39:50,612 WARN util.NativeCodeLoader: Unable to load native-hadoop library
in java classes where applicable
```

```
hadoop@hadoopmaster:~$ hdfs dfs -ls qualitat_aire/catalunya
2024-10-23 00:39:59,649 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your p
latform; falling back to Java classes where applicable
Found 1 items
-rw-r--r-- 3 hadoop supergroup 730413175 2024-10-23 00:06 qualitat_aire/catalunya/catalunya.csv
```

- 6: Borrar archivos que queden en “qualitat_aire”

Primero revisamos los archivos que quedan en el directorio “qualitat_aire”.

Después a través del comando “**hdfs dfs -rm {ruta/nombre_archivo}**”

Luego volvemos a comprobar los archivos del directorio para asegurarnos de que ya no se encuentran.

```
hadoop@hadoopmaster:~$ hdfs dfs -ls qualitat_aire
2024-10-23 00:45:36,083 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
platform; falling back to Java classes where applicable
Found 5 items
drwxr-xr-x - hadoop supergroup 0 2024-10-23 00:39 qualitat_aire/catalunya
drwxr-xr-x - hadoop supergroup 0 2024-10-23 00:33 qualitat_aire/illes_balears
-rw-r--r-- 3 hadoop supergroup 5765 2024-10-22 23:50 qualitat_aire/illes_balears.csv
drwxr-xr-x - hadoop supergroup 0 2024-10-23 00:33 qualitat_aire/madrid
-rw-r--r-- 3 hadoop supergroup 7593 2024-10-22 23:50 qualitat_aire/madrid.csv
```

```
hadoop@hadoopmaster:~$ hdfs dfs -rm qualitat_aire/illes_balears.csv
2024-10-23 00:46:03,559 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
platform; falling back to Java classes where applicable
Deleted qualitat_aire/illes_balears.csv
hadoop@hadoopmaster:~$ hdfs dfs -rm qualitat_aire/madrid.csv
2024-10-23 00:46:12,349 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
platform; falling back to Java classes where applicable
Deleted qualitat_aire/madrid.csv
```

```
hadoop@hadoopmaster:~$ hdfs dfs -ls qualitat_aire
2024-10-23 00:46:16,198 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
platform; falling back to Java classes where applicable
Found 3 items
drwxr-xr-x - hadoop supergroup 0 2024-10-23 00:39 qualitat_aire/catalunya
drwxr-xr-x - hadoop supergroup 0 2024-10-23 00:33 qualitat_aire/illes_balears
drwxr-xr-x - hadoop supergroup 0 2024-10-23 00:33 qualitat_aire/madrid
```

- 7: Mostrar el contenido del directorio “catalunya”

Usamos “**hdfs dfs -ls qualitat_aire/catalunya**” para mostrar su contenido

```
hadoop@hadoopmaster:~$ hdfs dfs -ls qualitat_aire/catalunya
2024-10-23 00:55:11,375 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r--    3 hadoop supergroup  730413175 2024-10-23 00:06 qualitat_aire/catalunya/catalunya.csv
```

- 8: Mostrar el contenido del archivo “illes_balears.csv”

Usamos “**hdfs dfs -cat qualitat_aire/illes_balears/illes_balears.csv**” para mostrar el contenido del archivo.

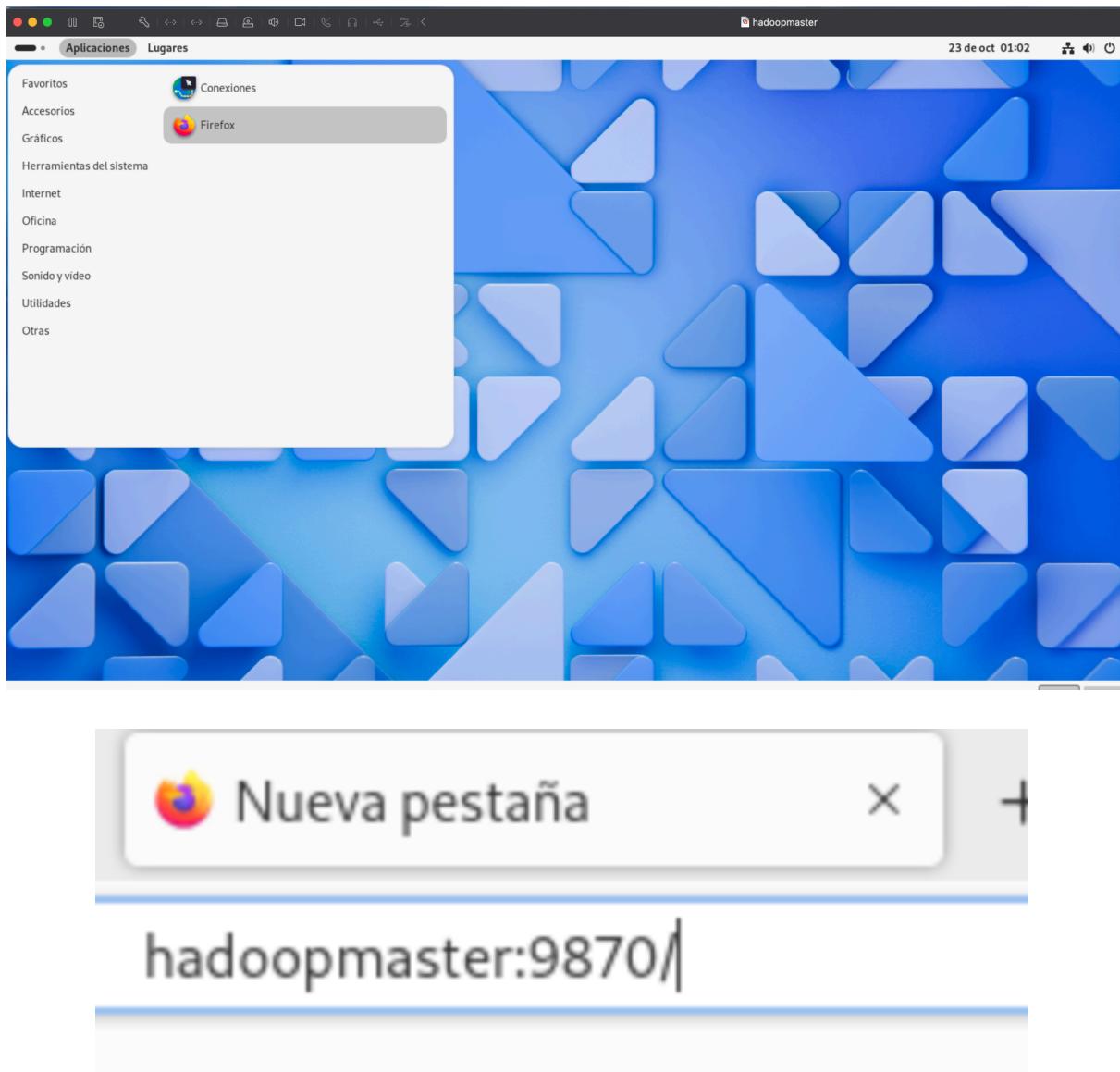
```
hadoop@hadoopmaster:~$ hdfs dfs -cat qualitat_aire/illes_balears/illes_balears.csv
2024-10-23 00:57:39,924 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Codi Local_Estadí,Non_Estadí,Ubicació,Municipi,Altitud,Propietat,Parametres_mesurats,Tipus_d'àrea,Tipus_d'estació,Localització,Entitat
7040003,Parc de Bellver,Castell de Bellver,Palma,117,Ministeri de Medi Ambient i Conselleria de Medi Ambient,"S02,NO,N02,03,PM10,DD,VV,TMP,HR,PRB,LL",Suburbana,Fons,POINT (2.621694
44444 39.5645278),http://www.caib.es/sites/atmosfera/ca/estacio_de_control_de_la_qualitat_de_laire_de_parc_de_bellver_-_palma_de_mallorca_mallorca-3249/
7040006,Sant Joan de Déu,Hospital Sant Joan de Déu,Palma,5,GESE,"S02,NO,N02,PM10,03,DD,VV,TMP,LL",Urbana,Industrial,POINT (2.7 39.5469444),http://www.caib.es/sites/atmosfera/ca/estaci
o_de_control_de_la_qualitat_de_laire_de_hospital_sant_joan_de_deu_-_palma_de_mallorca_mallorca-35681/
7032092,Pous,"Carretera de Maó a Ciutadella, Km. 1",Maó,55,GESE,"S02,NO,N02,PM10,03,DD,VV,TMP",Urbana,Industrial,POINT (4.2529722222 39.8909167),http://www.caib.es/sites/atmosfera/ca
/estacio_de_control_de_la_qualitat_de_laire_de_pous_-_mao_menorca-3261/
7025002,Dalt Vila,"C/ Major, 20",Eivissa,44,GESE,"S02,NO,N02,03,DD,VV,TMP",Urbana,Industrial,POINT (1.43488888889 38.9071944),http://www.caib.es/sites/atmosfera/ca/estacio_de_control_
de_la_qualitat_de_laire_de_dalt_vila_eivissa-3258/
7040004,Parc Bit,Edifici Telecomunicacions - Parc Bit,Palma,109,GESE,"S02,NO,N02,03,DD,VV,TMP",Rural,Fons,POINT (2.63055555556 39.6347222),http://www.caib.es/sites/atmosfera/ca/estaci
o_de_control_de_la_qualitat_de_laire_del_parc_bit_-_palma_de_mallorca_mallorca-3265/
7054001,Torrent,"Subestació de Torrent - Camí Vell de Sant Mateu, s/n",Eivissa,12,GESE,"S02,NO,N02,03,PM10,DD,VV,TMP",Rural,Fons,POINT (1.43275 38.9364167),http://www.caib.es/sites/at
mosfera/ca/estacio_de_control_de_la_qualitat_de_laire_de_torrent_-_sta_eularia_des_riu_eivissa-3253/
7032093,Port de Maó,c/ Villacarlos s/n,Maó,42,GESE,"S02,NO,N02,PM10,03,DD,VV,TMP",Suburbana,Industrial,POINT (4.25705555556 39.8926389),http://www.caib.es/sites/atmosfera/ca/estacia_d
e_control_de_la_qualitat_de_laire_del_port_de_maa_-_maa_menorca/
7015001,Ciutadella,Carrer Carriners - Polígon industrial,Ciutadella,30,Ministeri de Medi Ambient i Conselleria de Medi Ambient,"NO,N02,03,PM10,DD,VV,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 - Pollesportiu Can Coix,Sant Antoni de Portmany,65,Ministeri de Medi Ambient i Conselleria de Medi Ambient,"NO,N02,03,PM10,DD,VV,TMP,
HR,PRB,LL",Suburbana,Fons,POINT (1.31013888889 38.9943889),http://www.caib.es/sites/atmosfera/ca/estacio_de_control_de_la_qualitat_de_laire_de_santantonideportmany_eivissa-3263/
7010001,Cases de Menut, Serra de Tramuntana",Escorba,584,Conselleria de Medi Ambient,03,Rural,Fons,POINT (2.9007222222 39.83025),http://www.caib.es/sites/atmosfera/ca
/estacio_de_control_de_la_qualitat_de_laire_de_cases_de_menut_-_ozo_mallorca-80678/
7003001,S'Albufera,"Carretera de Sa Pobla a Alcúdia, Km. 6",Alcúdia,6,GESE,"S02,NO,N02,PM10,03,DD,VV,TMP,PRB",Rural,Industrial,POINT (3.07525 39.7911667),http://www.caib.es/sites/atmo
sfera/ca/estacio_de_control_de_la_qualitat_de_laire_de_sabufera_-_alcudia_mallorca-3264/
7044001,Sa Pobla,"Carretera de Sa Pobla a Llubi, Km. 2.5 - Finca Sa Canova",Sa Pobla,7,GESE,"S02,NO,N02,PM10,03,DD,VV,TMP",Rural,Fons,POINT (3.016 39.7484722),http://www.caib.es/sites
/atmosfera/ca/estacio_de_control_de_la_qualitat_de_laire_de_sa_pobla_mallorca-3252/
7032001,Sant Lluís,"Carretera de Maó a Sant Lluís, Km. 2",Maó,65,GESE,"S02,NO,N02,PM10,03,DD,VV,TMP",Urbana,Industrial,POINT (4.25883333333 39.8748333),http://www.caib.es/sites/atmos
fera/ca/estacio_de_control_de_la_qualitat_de_laire_de_sant_lluis_-_mao_menorca-3248/
7025001,Can Misses,Hospital de Can Misses,Eivissa,25,GESE,"S02,NO,N02,PM10,03,DD,VV,TMP",Urbana,Industrial,POINT (1.4144722222 38.9140556),http://www.caib.es/sites/atmosfera/ca/estaci
o_de_control_de_la_qualitat_de_laire_de_can_misses_eivissa-3260/
7010001,Hospital Joan March,Hospital Joan March,Bunyola,172,TIRME,"S02,NO,N02,03,PM2.5,PM10,DD,VV,TMP,HR,PRB,LL",Rural,Fons,POINT (2.68777777778 39.6788333),http://www.caib.es/sites/a
tmosfera/ca/estacio_de_control_de_la_qualitat_de_laire_de_lhospital_joan_march_-_bunyola_mallorca-3259/
,Illeseta,Depuradora de Illeseta,Illeseta,,CFEMEX,"PM10,PM2.5",Rural,Fons,POINT (2.86491666667 39.7089444),http://www.caib.es/sites/atmosfera/ca/estacio_de_control_de_la_qualitat_de_laire
_de_illeseta_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,"S02,NO,N02,CO,03,BEN,TOL,XIL,DD,W,TMP
,HR,LL,PM10",Urbana,Trànsit,POINT (2.65702777778 39.57125),http://www.caib.es/sites/atmosfera/ca/estacio_de_control_de_la_qualitat_de_laire_de_carrer_foners_-_palma_de_mallorca_mallor
ca-3250/
7027002,Sa Vinyeta d'Inca,Carretera vella d'Inca,Inca,118,GESE,"S02,NO,N02,PM10,03,DD,VV,TMP",Suburbana,Fons,POINT (2.91944444444 39.7155556),http://www.caib.es/sites/atmosfera/ca/est
acio_de_control_de_la_qualitat_de_laire_de_sa_vinyeta_-_inca_mallorca-3251/
7003004,Alcúdia,Antiga central tèrmica d'Alcúdia I - Port d'Alcúdia,Alcúdia,15,GESE,"S02,NO,N02,PM10,03,DD,VV,TMP,PRB",Urbana,Fons,POINT (3.1454722222 39.8348056),http://www.caib.es/
sites/atmosfera/ca/estacio_de_control_de_la_qualitat_de_laire_dalculdia_i_mallorca-3266/
7042001,Can Llompart,Carretera de Pollença a Campanet,Pollença,25,GESE,"S02,NO,N02,PM10,03,DD,VV,TMP,PRB,LL",Rural,Fons,POINT (3.0422222222 39.8447222),http://www.caib.es/sites/atmos
fera/ca/estacio_de_control_de_la_qualitat_de_laire_de_can_llompert_-_pollenca_mallorca-3254/
hadoop@hadoopmaster: $
```

3 - INTERFAZ WEB

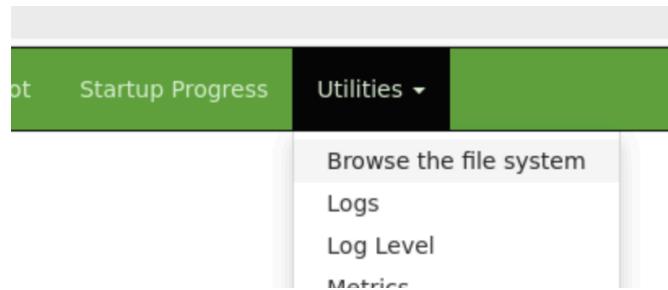
El primer paso sería iniciar la interfaz gráfica usando GNOME que hemos configurado antes. Para arrancarla ejecutamos el comando “**startx**”.

```
hadoop@hadoopmaster:~$ startx
```

Una vez dentro de la interfaz gráfica accedemos al navegador web y escribimos la url a la que queremos acceder, en este caso a la interfaz web de Hadoop en concreto al puerto 9870 de hadoopmaster.



Accedemos a “Browse the files system” para investigar los archivos con los que hemos estado trabajando.



Lo primero que podemos observar es que el archivo “rows.csv” se ha renombrado a “rows.csv.1_COPYING_”.

Antes de profundizar en esto, quiero hacer una aclaración: durante el ejercicio, al notar que debía descargar el archivo “rows.csv” en el apartado 2, asumí que esto podría causar conflictos con el archivo que ya había descargado con ese mismo nombre. Para evitar problemas, decidí eliminar el archivo previamente descargado. Sin embargo, al llegar al apartado 3, me di cuenta de que no era necesario haberlo borrado.

Ahora, en cuanto al archivo “rows.csv.1_COPYING_”, este es un archivo temporal que HDFS genera al ejecutar el comando `-put`. Por lo general, HDFS elimina este archivo temporal una vez que la operación de carga se ha completado. Sin embargo, en mi caso, recibí un error durante el proceso, lo que probablemente se debió a problemas de almacenamiento.

A pesar de esto, decidí continuar con los ejercicios, ya que no estaba seguro del impacto que había tenido ese error, especialmente porque el archivo “rows.csv” se encontraba en el directorio deseado (qualitat_air). Debido a esta interrupción, el archivo temporal no se eliminó, lo que explica la presencia del sufijo “COPYING”.

A screenshot of the HDFS Web UI showing the "Browse Directory" page. The URL is "/user/hadoop". The page includes a search bar, a file list table, and navigation buttons. The table has columns: Permission, Owner, Group, Size, Last Modified, Replication, Block Size, Name, and a delete icon. The table shows four entries: "prova" (0 B, Oct 19 20:18, 0, 0 B, prova), "prova3" (0 B, Oct 20 23:47, 0, 0 B, prova3), "qualitat_air" (0 B, Oct 23 00:46, 0, 0 B, qualitat_air), and "rows.csv.1_COPYING_" (315.52 MB, Oct 22 23:54, 3, 128 MB, rows.csv.1_COPYING_). Navigation buttons at the bottom include "Previous", "1", and "Next".

Dentro de qualitat_aire/catalunya encontramos el archivo "catalunya.csv"

Browse Directory

/user/hadoop/qualitat_aire/catalunya									Go!	File	Upload	Folder	Trash
Show 25 entries		Search:											
	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name					
<input type="checkbox"/>	-rw-r--r--	hadoop	supergroup	696.58 MB	Oct 23 00:06	3	128 MB	catalunya.csv					
Showing 1 to 1 of 1 entries													
Previous 1 Next													

Vemos que está formado por 6 bloques. Y los nodos en los que se encuentra

BLOQUE 0:

Nodes Datanode Volume Failures Snapshot Startup Progress Utilities

File information - catalunya.csv

Download Head the file (first 32K) Tail the file (last 32K)

Block information -- Block 0

Block ID: 107374184
Block Pool ID: BP-1842134217728
Generation Stamp: 1
Size: 134217728
Availability:

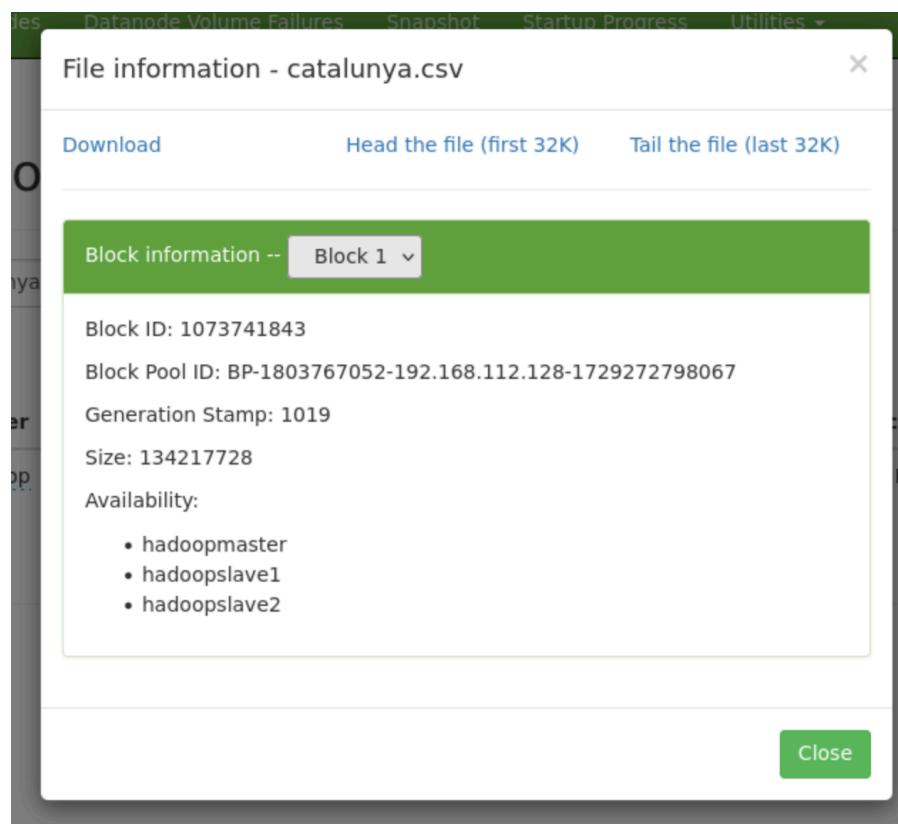
- hadoopmaster
- hadoopslave1
- hadoopslave2

2.168.112.128-1729272798067

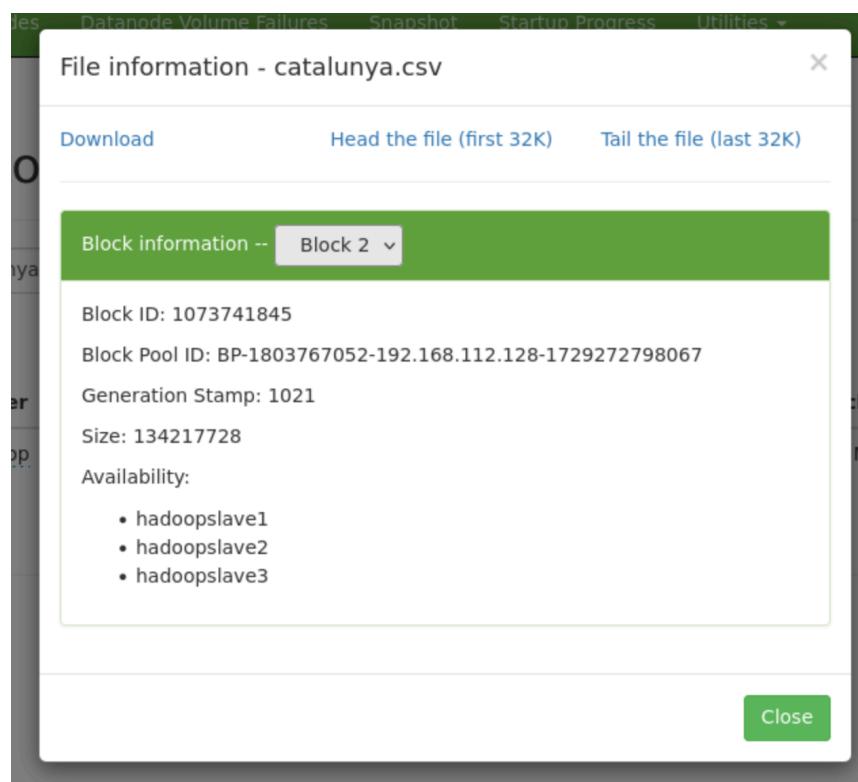
Block 1 Block 2 Block 3 Block 4 Block 5

Close

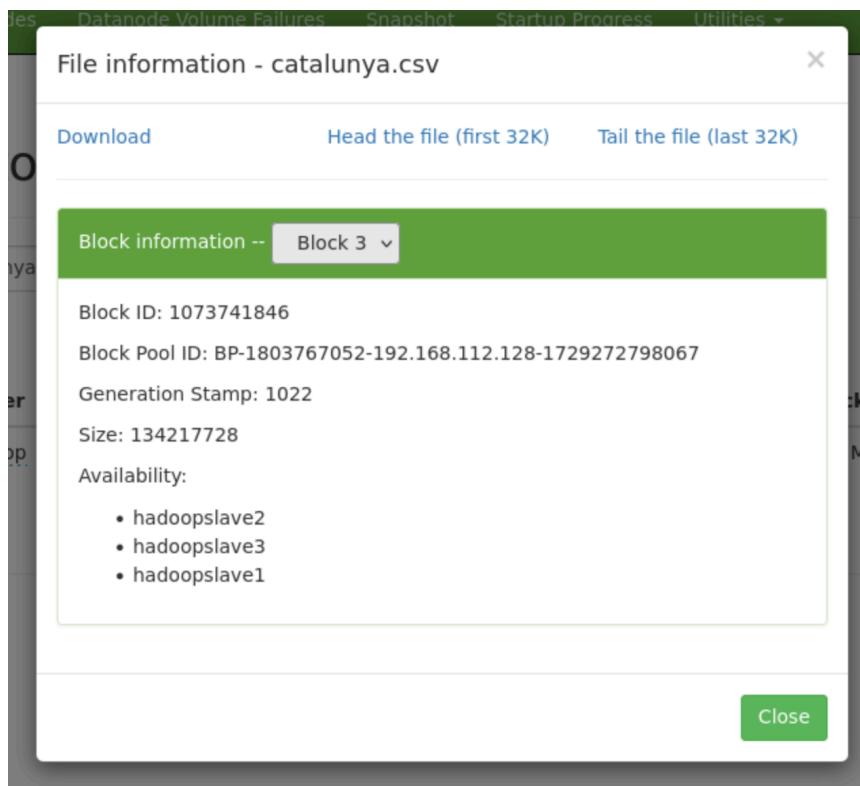
BLOQUE 1:



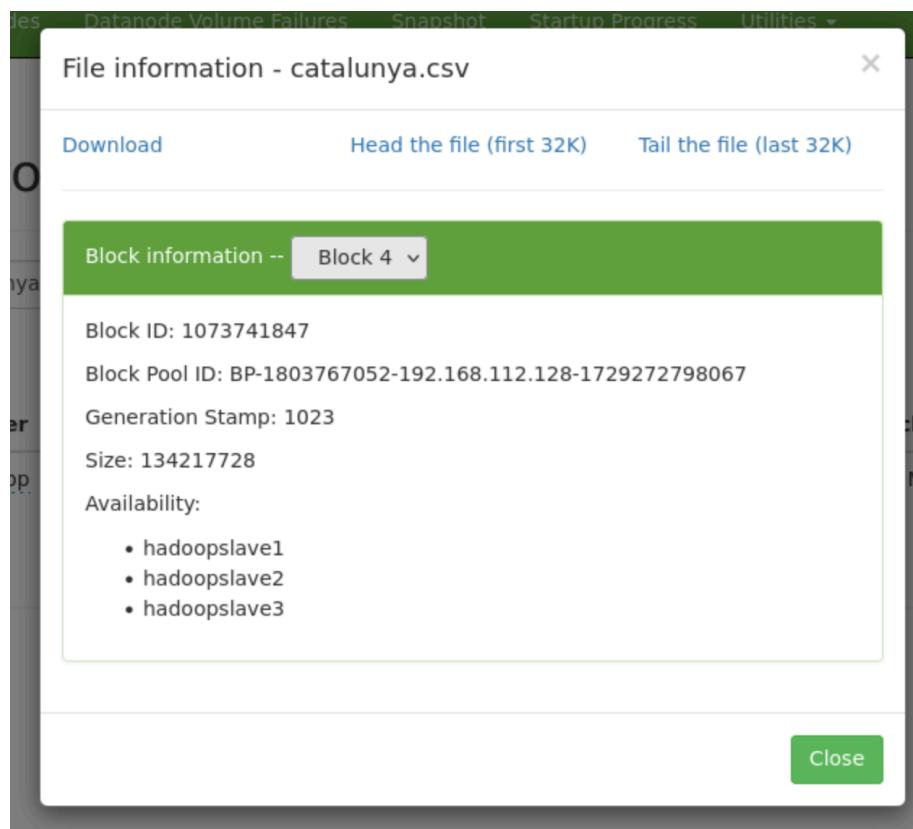
BLOQUE 2:



BLOQUE 3:



BLOQUE 4:



BLOQUE 5:

