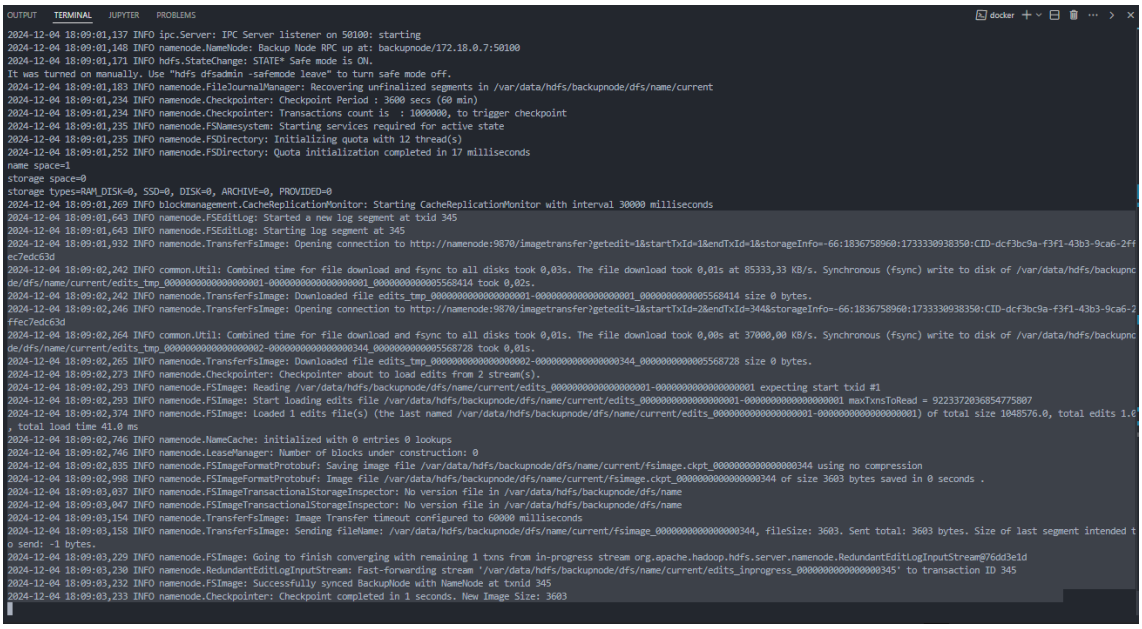
1. **Servidor de Backup.**

1.1.0. Captura de pantalla de los metadatos del namenode antes de iniciar el backupnode.

```
root@namenode:/var/data/hdfs/namenode/current# ls -l
total 2064
-rw-r--r-- 1 hdadmin hadoop     214 Dec  4 17:48 VERSION
-rw-r--r-- 1 hdadmin hadoop 1048576 Dec  4 17:49 edits_0000000000000000001-0000000000000000001
-rw-r--r-- 1 hdadmin hadoop 1048576 Dec  4 18:03 edits_inprogress_0000000000000000002
-rw-r--r-- 1 hdadmin hadoop     402 Dec  4 17:48 fsimage_0000000000000000000
-rw-r--r-- 1 hdadmin hadoop      62 Dec  4 17:48 fsimage_0000000000000000000.md5
-rw-r--r-- 1 hdadmin hadoop       2 Dec  4 17:56 seen_txid
root@namenode:/var/data/hdfs/namenode/current#
```

Nos introducimos dentro de la carpeta /var/data/hdfs/namenode/current y vemos el contenido que hay dentro de esta con el comando “ls -l”.

1.1.1. Captura de pantalla en la que se vean los mensajes que genera el servicio de backup, destacando aquellos en los que se vea como se hace el checkpoint.



Cuando tenemos listo el contenedor de backup, realizamos “hdfs namenode -backup” para ejecutar el servicio de backup. Una vez iniciado, habrá que esperar un tiempo para que el servicio realice el primer backup. Después de unos minutos, realizará los primeros guardados como se ve en las capturas.

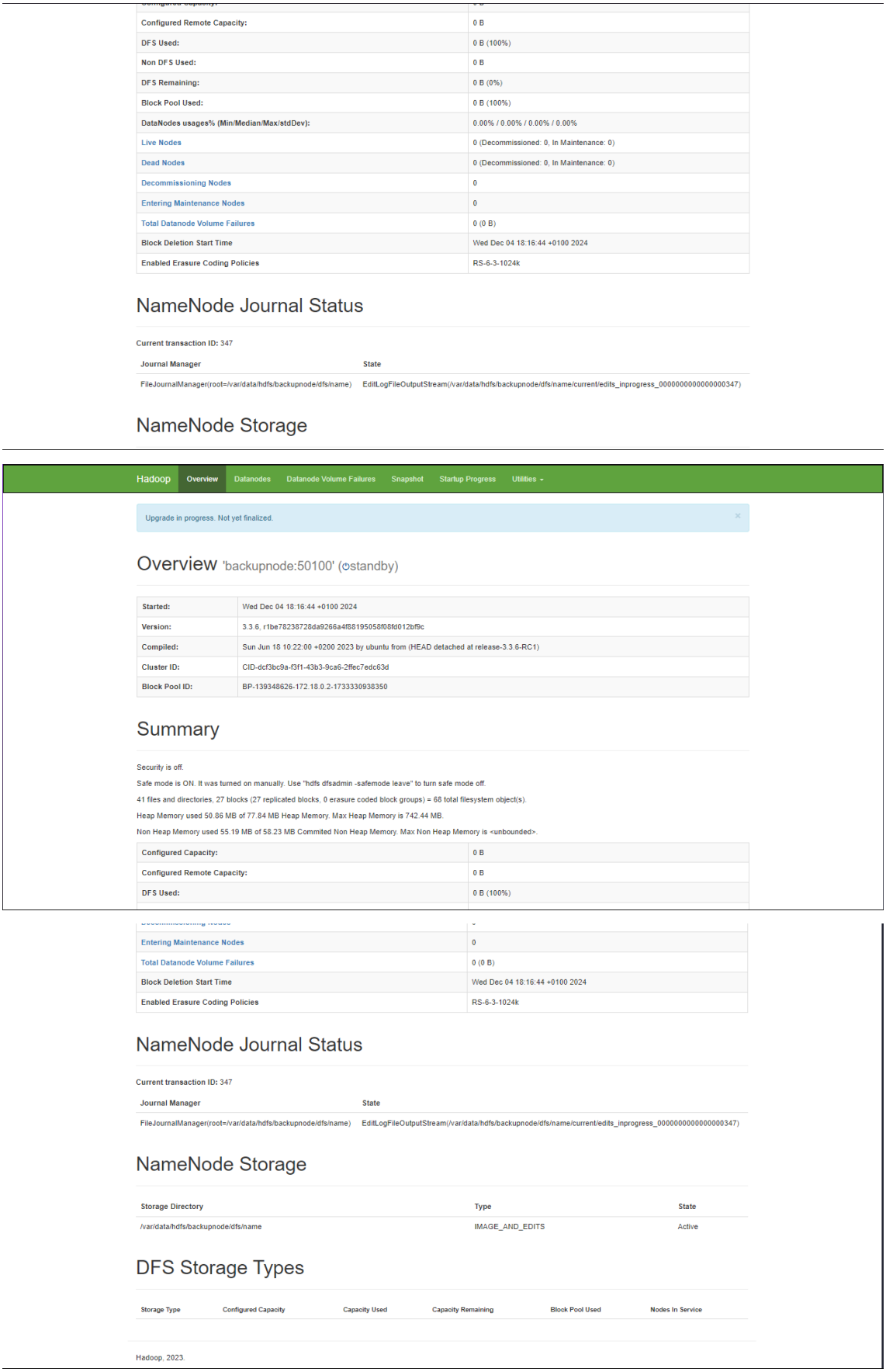
1.1.3 Captura de pantalla en la que se compare el contenido del directorio del backup con el directorio con los metadatos de NameNode, antes y una vez que el servicio de backup se ha completado.

```
root@backupnode:/var/data/hdfs/backupnode/dfs/name/current# ls -l
total 2100
-rw-r--r-- 1 hdadmin hadoop     214 Dec  4 18:09 VERSION
-rw-r--r-- 1 hdadmin hadoop 1048576 Dec  4 18:09 edits_0000000000000000001-0000000000000000001
-rw-r--r-- 1 hdadmin hadoop   38190 Dec  4 18:09 edits_0000000000000000002-0000000000000000344
-rw-r--r-- 1 hdadmin hadoop 1048576 Dec  4 18:09 edits_inprogress_0000000000000000345
-rw-r--r-- 1 hdadmin hadoop    3603 Dec  4 18:09 fsimage_0000000000000000344
-rw-r--r-- 1 hdadmin hadoop      62 Dec  4 18:09 fsimage_0000000000000000344.md5
root@backupnode:/var/data/hdfs/backupnode/dfs/name/current#
```

```
root@namenode:/var/data/hdfs/namenode/current# ls -l
total 2104
-rw-r--r-- 1 hdadmin hadoop     214 Dec  4 17:48 VERSION
-rw-r--r-- 1 hdadmin hadoop 1048576 Dec  4 17:49 edits_0000000000000000001-0000000000000000001
-rw-r--r-- 1 hdadmin hadoop   38190 Dec  4 18:09 edits_0000000000000000002-0000000000000000344
-rw-r--r-- 1 hdadmin hadoop 1048576 Dec  4 18:09 edits_inprogress_0000000000000000345
-rw-r--r-- 1 hdadmin hadoop     402 Dec  4 17:48 fsimage_0000000000000000000
-rw-r--r-- 1 hdadmin hadoop      62 Dec  4 17:48 fsimage_0000000000000000000.md5
-rw-r--r-- 1 hdadmin hadoop       4 Dec  4 18:09 seen_txid
root@namenode:/var/data/hdfs/namenode/current#
```

Se puede ver en las dos capturas como el contenido es prácticamente el mismo en los dos.

1.1.3. Captura de pantalla del interfaz web del nodo de backup.



Configured Remote Capacity:	0 B
DFS Used:	0 B (100%)
Non DFS Used:	0 B
DFS Remaining:	0 B (0%)
Block Pool Used:	0 B (100%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Block Deletion Start Time	Wed Dec 04 18:16:44 +0100 2024
Enabled Erasure Coding Policies	RS-6-3-1024k

NameNode Journal Status

Current transaction ID: 347

Journal Manager	State
FileJournalManager(root=/var/data/hdfs/backupnode/dfs/name)	EditLogFileOutputStream(/var/data/hdfs/backupnode/dfs/name/current/edits_inprogress_0000000000000000347)

NameNode Storage

Hadoop | Overview | Datanodes | Datanode Volume Failures | Snapshot | Startup Progress | Utilities

Upgrade in progress. Not yet finalized.

Overview 'backupnode:50100' (standby)

Started:	Wed Dec 04 18:16:44 +0100 2024
Version:	3.3.6, r1be78238728da9266a4f88195058f08fd012bf9c
Compiled:	Sun Jun 18 10:22:00 +0200 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1)
Cluster ID:	CID-dcf3bc9a-f3f1-43b3-9ca6-2ffec7edc63d
Block Pool ID:	BP-139348626-172.18.0.2-1733330938350

Summary

Security is off.

Safe mode is ON. It was turned on manually. Use "hdfs dfsadmin -safemode leave" to turn safe mode off.

41 files and directories, 27 blocks (27 replicated blocks, 0 erasure coded block groups) = 68 total filesystem object(s).

Heap Memory used 50.86 MB of 77.84 MB Heap Memory. Max Heap Memory is 742.44 MB.

Non Heap Memory used 55.19 MB of 58.23 MB Commited Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	0 B
Configured Remote Capacity:	0 B
DFS Used:	0 B (100%)
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Block Deletion Start Time	Wed Dec 04 18:16:44 +0100 2024
Enabled Erasure Coding Policies	RS-6-3-1024k

NameNode Journal Status

Current transaction ID: 347

Journal Manager	State
FileJournalManager(root=/var/data/hdfs/backupnode/dfs/name)	EditLogFileOutputStream(/var/data/hdfs/backupnode/dfs/name/current/edits_inprogress_0000000000000000347)

NameNode Storage

Storage Directory	Type	State
/var/data/hdfs/backupnode/dfs/name	IMAGE_AND_EDITS	Active

DFS Storage Types

Storage Type	Configured Capacity	Capacity Used	Capacity Remaining	Block Pool Used	Nodes In Service

Hadoop, 2023.

2. Añadir un nuevo DataNode/NodeManager.

2.1. Las líneas de los ficheros de log del namenode y del resourcemanager que muestran que se han incluido los nodos indicados en los ficheros include.

```
hdadmin@namenode:~/hadoop/logs$ grep -i "include" hadoop-hdadmin-resourcemanager-namenode.log
2024-12-04 17:49:26,088 INFO org.apache.hadoop.util.HostsFileReader: Refreshing hosts (include/exclude) list
2024-12-04 17:56:47,374 INFO org.apache.hadoop.util.HostsFileReader: Refreshing hosts (include/exclude) list
2024-12-04 18:28:39,960 INFO org.apache.hadoop.util.HostsFileReader: Refreshing hosts (include/exclude) list
2024-12-13 16:14:28,757 INFO org.apache.hadoop.util.HostsFileReader: Refreshing hosts (include/exclude) list
2024-12-19 19:55:05,080 INFO org.apache.hadoop.util.HostsFileReader: Refreshing hosts (include/exclude) list
2024-12-19 20:19:59,074 INFO org.apache.hadoop.util.HostsFileReader: Refreshing hosts (include/exclude) list
2024-12-31 11:01:58,647 INFO org.apache.hadoop.util.HostsFileReader: Refreshing hosts (include/exclude) list
2024-12-31 11:14:50,302 INFO org.apache.hadoop.util.HostsFileReader: Refreshing hosts (include/exclude) list
2024-12-31 11:14:50,302 INFO org.apache.hadoop.util.HostsFileReader: Adding a node "datanode1" to the list of included hosts from /opt/bd/hadoop/etc/hadoop/yarn.include
2024-12-31 11:14:50,362 INFO org.apache.hadoop.util.HostsFileReader: Adding a node "datanode2" to the list of included hosts from /opt/bd/hadoop/etc/hadoop/yarn.include
2024-12-31 11:14:50,362 INFO org.apache.hadoop.util.HostsFileReader: Adding a node "datanode3" to the list of included hosts from /opt/bd/hadoop/etc/hadoop/yarn.include
2024-12-31 11:14:50,362 INFO org.apache.hadoop.util.HostsFileReader: Adding a node "datanode4" to the list of included hosts from /opt/bd/hadoop/etc/hadoop/yarn.include
```

```
hdadmin@namenode:~/hadoop/logs$ grep -i "include" hadoop-hdadmin-namenode-namenode.log
2024-12-31 11:14:45,656 INFO org.apache.hadoop.util.HostsFileReader: Adding a node "datanode1" to the list of included hosts from /opt/bd/hadoop/etc/hadoop/dfs.include
2024-12-31 11:14:45,668 INFO org.apache.hadoop.util.HostsFileReader: Adding a node "datanode2" to the list of included hosts from /opt/bd/hadoop/etc/hadoop/dfs.include
2024-12-31 11:14:45,669 INFO org.apache.hadoop.util.HostsFileReader: Adding a node "datanode3" to the list of included hosts from /opt/bd/hadoop/etc/hadoop/dfs.include
2024-12-31 11:14:45,669 INFO org.apache.hadoop.util.HostsFileReader: Adding a node "datanode4" to the list of included hosts from /opt/bd/hadoop/etc/hadoop/dfs.include
```

Se han buscado las líneas correspondientes con el comando “grep”. Se ve comom se han incluido correctamente los datanodes.

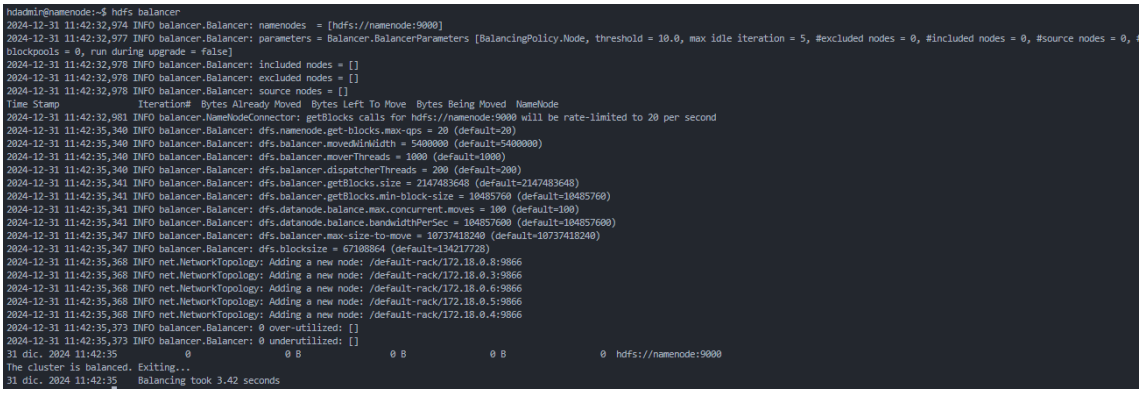
2.2. Los pasos indicados para añadir un nuevo datanode/nodemanager, con las salidas de los comandos `hdfs dfsadmin -report` y `yarn node -list`.

```
hdadmin@datanode5:~$ yarn node -list
2024-12-31 11:36:10,795 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at resourcemanager/172.18.0.2:8032
Total Nodes:5
         Node-Id             Node-State Node-Http-Address       Number-of-Running-Containers
 datanode1:36779                RUNNING    datanode1:8042                                  0
 datanode3:38601                RUNNING    datanode3:8042                                  0
 datanode5:36311                RUNNING    datanode5:8042                                  0
 datanode4:33683                RUNNING    datanode4:8042                                  0
 datanode2:45865                RUNNING    datanode2:8042                                  0
hdadmin@datanode5:~$ hdfs dfsadmin -report | grep -i "datanode"
Live datanodes (4):
Name: 172.18.0.3:9866 (datanode1.hadoop-cluster)
Hostname: datanode1
Name: 172.18.0.4:9866 (datanode2.hadoop-cluster)
Hostname: datanode2
Name: 172.18.0.5:9866 (datanode3.hadoop-cluster)
Hostname: datanode3
Name: 172.18.0.6:9866 (datanode4.hadoop-cluster)
Hostname: datanode4
```

```
hdadmin@datanode5:~$ yarn node -list
2024-12-31 11:40:27,189 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at resourcemanager/172.18.0.2:8032
Total Nodes:5
         Node-Id             Node-State Node-Http-Address       Number-of-Running-Containers
 datanode1:36779                RUNNING    datanode1:8042                                  0
 datanode3:38601                RUNNING    datanode3:8042                                  0
 datanode5:36311                RUNNING    datanode5:8042                                  0
 datanode4:33683                RUNNING    datanode4:8042                                  0
 datanode2:45865                RUNNING    datanode2:8042                                  0
hdadmin@datanode5:~$ hdfs dfsadmin -report | grep -i "datanode"
Live datanodes (5):
Name: 172.18.0.3:9866 (datanode1.hadoop-cluster)
Hostname: datanode1
Name: 172.18.0.4:9866 (datanode2.hadoop-cluster)
Hostname: datanode2
Name: 172.18.0.5:9866 (datanode3.hadoop-cluster)
Hostname: datanode3
Name: 172.18.0.6:9866 (datanode4.hadoop-cluster)
Hostname: datanode4
Name: 172.18.0.8:9866 (datanode5)
Hostname: datanode5
```

Vemos cómo se ha añadido un nuevo datanode5 al hdfs pero no al yarn.

2.3. Salida de la ejecución del balanceador de carga. Indica también cuántos datos se han movido y cuántos bloques tiene el datanode5.



Se realiza el balanceo pero no mueve ningún bloque, por lo que datanode5 tiene 0 bloques.

3. Retirar un DataNode/NodeManager

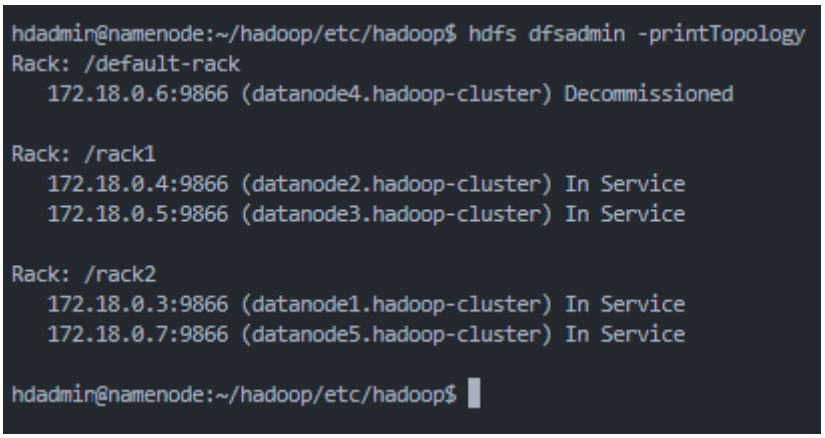
Configured Capacity:	3.93 TB
Configured Remote Capacity:	0 B
DFS Used:	2.37 GB (0.06%)
Non DFS Used:	45.63 GB
DFS Remaining:	3.69 TB (93.72%)
Block Pool Used:	2.37 GB (0.06%)
DataNodes usages% (Min/Median/Max/stdDev):	0.04% / 0.07% / 0.07% / 0.01%
Live Nodes	5 (Decommissioned: 1, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion (including replicas)	0
Block Deletion Start Time	Tue Dec 31 11:14:44 +0100 2024
Last Checkpoint Time	Tue Dec 31 11:02:06 +0100 2024
Enabled Erasure Coding Policies	RS-6-3-1024k

Show 20 entries

Node Labels	Rack	Node State	Node Address	Node HTTP Address	Last health-update	Health-report	Co
	/default-rack	RUNNING	datanode1:36779	datanode1:8042	mar. dic. 31 12:00:47 +0100 2024		0
	/default-rack	RUNNING	datanode3:38601	datanode3:8042	mar. dic. 31 12:00:49 +0100 2024		0
	/default-rack	RUNNING	datanode5:36311	datanode5:8042	mar. dic. 31 11:59:59 +0100 2024		0
	/default-rack	RUNNING	datanode2:45865	datanode2:8042	mar. dic. 31 12:00:48 +0100 2024		0

Vemos en las capturas anteriores como hay 5 datanodes pero 1 decomisionado. Solo tenemos del datanode1, datanode3, datanode5 y datanode2. El datanode4 no está activo.

4. **Rack awareness**



Vemos que se han divido los contenedores en dos racks.