

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.

```
2024-12-04 18:09:01,137 INFO ipc.Server: IPC Server listener on 50100: starting
2024-12-04 18:09:01,148 INFO namenode.NameNode: Backup Node RPC up at: backupnode/172.18.0.7:50100
2024-12-04 18:09:01,171 INFO hdfs.StateChange: STATE* Safe mode is ON.
2024-12-04 18:09:01,183 INFO namenode.FileJournalManager: Recovering unfinalized segments in /var/data/hdfs/backupnode/dfs/name/current
2024-12-04 18:09:01,234 INFO namenode.Checkpointer: Checkpoint Period : 3600 secs (60 min)
2024-12-04 18:09:01,234 INFO namenode.Checkpointer: Transactions count is : 1000000, to trigger checkpoint
2024-12-04 18:09:01,235 INFO namenode.FSNamesystem: Starting services required for active state
2024-12-04 18:09:01,235 INFO namenode.FSJournalManager: Initializing quota with 12 thread(s)
2024-12-04 18:09:01,252 INFO namenode.FSJournalManager: Quota initialization completed in 17 milliseconds
name space=1
storage space=0
storage types=RAM_DISK=0, SSD=0, DISK=0, ARCHIVE=0, PROVIDED=0
2024-12-04 18:09:01,269 INFO blockmanagement.CacheReplicationMonitor: Starting CacheReplicationMonitor with interval 30000 milliseconds
2024-12-04 18:09:01,643 INFO namenode.FSEditLog: Started a new log segment at txid 345
2024-12-04 18:09:01,643 INFO namenode.FSEditLog: Starting log segment at 345
2024-12-04 18:09:01,932 INFO namenode.TransferFsImage: Opening connection to http://namenode:9870/ImageTransfer?getedit=1&startTxId=1&endTxId=1&storageInfo=-66:1836758960:1733330938350:CID-dcf3bc9a-f3f1-43b3-9ca6-2ffec7edc63d
2024-12-04 18:09:02,242 INFO common.Util: Combined time for file download and fsync to all disks took 0,03s. The file download took 0,01s at 85333,33 KB/s. Synchronous (fsync) write to disk of /var/data/hdfs/backupnode/dfs/name/current/edits_tmp_0000000000000000001-0000000000000000001 took 0,01s.
2024-12-04 18:09:02,242 INFO namenode.TransferFsImage: Downloaded file edits_tmp_0000000000000000001-0000000000000000001_0000000000000000001 size 0 bytes.
2024-12-04 18:09:02,246 INFO namenode.TransferFsImage: Opening connection to http://namenode:9870/ImageTransfer?getedit=1&startTxId=2&endTxId=344&storageInfo=-66:1836758960:1733330938350:CID-dcf3bc9a-f3f1-43b3-9ca6-2ffec7edc63d
2024-12-04 18:09:02,264 INFO common.Util: Combined time for file download and fsync to all disks took 0,01s. The file download took 0,00s at 37000,00 KB/s. Synchronous (fsync) write to disk of /var/data/hdfs/backupnode/dfs/name/current/edits_tmp_0000000000000000002-0000000000000000002 took 0,01s.
2024-12-04 18:09:02,265 INFO namenode.TransferFsImage: Downloaded file edits_tmp_0000000000000000002-0000000000000000002_0000000000000000002 size 0 bytes.
2024-12-04 18:09:02,273 INFO namenode.Checkpointer: Checkpointer about to load edits from 2 stream(s).
2024-12-04 18:09:02,293 INFO namenode.FSImage: Reading /var/data/hdfs/backupnode/dfs/name/current/edits_0000000000000000001-0000000000000000001 expecting start txid #1
2024-12-04 18:09:02,293 INFO namenode.FSImage: Start loading edits file /var/data/hdfs/backupnode/dfs/name/current/edits_0000000000000000001-0000000000000000001 maxTxnsToRead = 9223372036854775807
2024-12-04 18:09:02,374 INFO namenode.FSImage: Loaded 1 edits file(s) (the last named /var/data/hdfs/backupnode/dfs/name/current/edits_0000000000000000001-0000000000000000001) of total size 1048576, total edits 1, total load time 41,0 ms
2024-12-04 18:09:02,746 INFO namenode.NameCache: Initialized with 0 entries 0 lookups
2024-12-04 18:09:02,746 INFO namenode.LeaseManager: Number of blocks under construction: 0
2024-12-04 18:09:02,835 INFO namenode.FSImageFormatProtobuf: Saving image file /var/data/hdfs/backupnode/dfs/name/current/fsimage.chkpt_000000000000000000344 using no compression
2024-12-04 18:09:02,998 INFO namenode.FSImageFormatProtobuf: Image file /var/data/hdfs/backupnode/dfs/name/current/fsimage.chkpt_000000000000000000344 of size 3603 bytes saved in 0 seconds .
2024-12-04 18:09:03,037 INFO namenode.FSImageTransactionalStorageInspector: No version file in /var/data/hdfs/backupnode/dfs/name
2024-12-04 18:09:03,047 INFO namenode.FSImageTransactionalStorageInspector: No version file in /var/data/hdfs/backupnode/dfs/name
2024-12-04 18:09:03,154 INFO namenode.TransferFsImage: Image Transfer timeout configured to 60000 milliseconds
2024-12-04 18:09:03,158 INFO namenode.TransferFsImage: Sending file name: /var/data/hdfs/backupnode/dfs/name/current/fsimage_000000000000000000344, fileSize: 3603. Sent total: 3603 bytes. Size of last segment intended to send: -1 bytes.
2024-12-04 18:09:03,228 INFO namenode.FSImage: Going to finish converging with remaining 1 tons from in-progress stream org.apache.hadoop.hdfs.server.namenode.RedundantEditLogInputStream@7f6d43e1d
2024-12-04 18:09:03,230 INFO namenode.RedundantEditLogInputStream: Fast-forwarding stream /var/data/hdfs/backupnode/dfs/name/current/edits_inprogress_000000000000000000345 to transaction ID 345
2024-12-04 18:09:03,232 INFO namenode.FSImage: Successfully synced BackupNode with NameNode at txid 345
2024-12-04 18:09:03,233 INFO namenode.Checkpointer: Checkpoint completed in 1 seconds. New Image Size: 3603
```


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) EditLogOutputStream(/var/data/hdfs/backupnode/dfs/name/current/edits_inprogress_000000000000000347)	

NameNode Storage

Overview

Upgrade in progress. Not yet finalized.

Overview 'backupnode:50100' (standby)

Started:	Wed Dec 04 18:16:44 +0100 2024
Version:	3.3.6, r1be78238728da9266a4f88195058f08f012bf9c
Compiled:	Sun Jun 18 10:22:00 +0200 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1)
Cluster ID:	CID-dcf3bc9a-fd11-43b3-9ca6-2fec7edc63d
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 Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	0 B
Configured Remote Capacity:	0 B
DFS Used:	0 B (100%)

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) EditLogOutputStream(/var/data/hdfs/backupnode/dfs/name/current/edits_inprogress_000000000000000347)	

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
--------------	---------------------	---------------	--------------------	-----------------	------------------

2. Añadir un nuevo DataNode/NodeManager.

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

```
hadoop@namenode:~/hadoop/logs$ grep -i "include" hadoop-hadmin-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 11:14:50,362 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

hadoop@namenode:~/hadoop/logs$ grep -i "include" hadoop-hadmin-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 como 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`.

```
hadoop@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
hadoop@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

hadoop@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
hadoop@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.

```
hadoop@namenode-5$ hdfs balancer
2024-12-31 11:42:32,974 INFO balancer.Balancer: namenodes = [hdfs://namenode:9000]
2024-12-31 11:42:32,977 INFO balancer.Balancer: parameters = Balancer.BalancerParameters [BalancingPolicy.Node, threshold = 10.0, max idle iteration = 5, #excluded nodes = 0, #included nodes = 0, #source nodes = 0, #blockpools = 0, run during upgrade = false]
2024-12-31 11:42:32,978 INFO balancer.Balancer: included nodes = []
2024-12-31 11:42:32,978 INFO balancer.Balancer: excluded nodes = []
2024-12-31 11:42:32,978 INFO balancer.Balancer: source nodes = []
Time Stamp      Iteration#      Bytes Already Moved      Bytes Left To Move      Bytes Being Moved      NameNode
2024-12-31 11:42:32,981 INFO balancer.NameNodeConnector: getBlocks calls for hdfs://namenode:9000 will be rate-limited to 20 per second
2024-12-31 11:42:35,340 INFO balancer.Balancer: dfs.namenode.get-blocks.max-ops = 20 (default=20)
2024-12-31 11:42:35,340 INFO balancer.Balancer: dfs.balancer.movedata.kwidth = 5400000 (default=5400000)
2024-12-31 11:42:35,340 INFO balancer.Balancer: dfs.balancer.moveThreads = 1000 (default=1000)
2024-12-31 11:42:35,340 INFO balancer.Balancer: dfs.balancer.dispatcherThreads = 200 (default=200)
2024-12-31 11:42:35,341 INFO balancer.Balancer: dfs.balancer.getBlocks.size = 2147483648 (default=2147483648)
2024-12-31 11:42:35,341 INFO balancer.Balancer: dfs.balancer.getBlocks.min-block-size = 10485760 (default=10485760)
2024-12-31 11:42:35,341 INFO balancer.Balancer: dfs.datanode.balance.max.concurrent.moves = 100 (default=100)
2024-12-31 11:42:35,341 INFO balancer.Balancer: dfs.datanode.balance.bandwidthPerSec = 104857600 (default=104857600)
2024-12-31 11:42:35,347 INFO balancer.Balancer: dfs.balancer.max-size-to-move = 10737418240 (default=10737418240)
2024-12-31 11:42:35,347 INFO balancer.Balancer: dfs.blocksize = 67108864 (default=134217728)
2024-12-31 11:42:35,368 INFO net.NetworkTopology: Adding a new node: /default-rack/172.18.0.8:9866
2024-12-31 11:42:35,368 INFO net.NetworkTopology: Adding a new node: /default-rack/172.18.0.3:9866
2024-12-31 11:42:35,368 INFO net.NetworkTopology: Adding a new node: /default-rack/172.18.0.6:9866
2024-12-31 11:42:35,368 INFO net.NetworkTopology: Adding a new node: /default-rack/172.18.0.5:9866
2024-12-31 11:42:35,373 INFO balancer.Balancer: 0 over-utilized: []
2024-12-31 11:42:35,373 INFO balancer.Balancer: 0 underutilized: []
31 dic. 2024 11:42:35      0      0 B      0 B      0 B      0 hdfs://namenode:9000
The cluster is balanced. Exiting...
31 dic. 2024 11:42:35      Balancing took 3.42 seconds
```

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

Show20entries

Node Labels	Rack	Node State	Node Address	Node HTTP Address	Last health-update	Health-report	Cc
	/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

```
hdadmin@namenode:~/hadoop/etc/hadoop$ hdfs dfsadmin -printTopology
Rack: /default-rack
  172.18.0.6:9866 (datanode4.hadoop-cluster) Decommissioned

Rack: /rack1
  172.18.0.4:9866 (datanode2.hadoop-cluster) In Service
  172.18.0.5:9866 (datanode3.hadoop-cluster) In Service

Rack: /rack2
  172.18.0.3:9866 (datanode1.hadoop-cluster) In Service
  172.18.0.7:9866 (datanode5.hadoop-cluster) In Service

hdadmin@namenode:~/hadoop/etc/hadoop$
```

Vemos que se han dividido los contenedores en dos racks.