

HDFS CLUSTER

Parte de preparacion del nodo

IPs y nombres de los hosts

```
victorgonzbaj@namenode:~$
```

```
victorgonzbaj@datanode1:~$
```

```
victorgonzbaj@datanode2:~$
```

Versión Java

```
javac 1.8.0_432
```

Fichero authorizes-keys de todos los hosts

```
hadoop@namenode:~$ cat .ssh/authorized_keys
ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQACQD11641Se60JFHKS1e8OhdsuNaBRHw+BQ126uNQV
d+hbJl+PKCVHSEKERB3JUCSpd0LVL/5nuPj+6qZQMeqZfysFMLbs918h3e6b4qeoUXCZGBLP7oeZ
856R+wCONPKysdLoZgrogISVLSjJ1INWwuc81/MU3p1c3X8DzSQW8013rDs9yU+n+484sc25e15E
S36UZEulaLXDBMKg3LOp8ozJ12L089J9CS+Kk63n6zVVRUjUbTmeF1cXkNwgb1x0JMaq1faVMSw04
EoJnTDn3f3YFEqz83dcdeFnaThoSUGQhsVhK2P1T560B/Z1785o1NuI340h2vgQx5EeQqQhnpZ
U9vAYMR2r19m130wMoobJVCntvsvY8yVquntuccCYLSVj+7z2pnlWHC1eW7gRgn16e6h250Qa
AgGras5B52DZXVfocMK1lWbxxKTfV8E8SPNBjLRAqZKZwSr1WYQSElK6BzUwNpH+4lWun+NC
FJ1924472Kduqs+SSPloFq+ZnwKTjdiOn2qCwO99eAKj76ZLDFMaoyDAG2eIITf3J6e6H5K1Dz
CT85KjDQUN4zyBy/oKruMfYgmsh+LYZc5yl/TXNkIntawNt6CIMI1tzh9m/XBdLx3sR/Zbo76N05rE
Ews= hadoop@namenode

hadoop@datanode1:~$ cat .ssh/authorized_keys
ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQACQD11641Se60JFHKS1e8OhdsuNaBRHw+BQ126uNQV
d+hbJl+PKCVHSEKERB3JUCSpd0LVL/5nuPj+6qZQMeqZfysFMLbs918h3e6b4qeoUXCZGBLP7oeZ
856R+wCONPKysdLoZgrogISVLSjJ1INWwuc81/MU3p1c3X8DzSQW8013rDs9yU+n+484sc25e15E
S36UZEulaLXDBMKg3LOp8ozJ12L089J9CS+Kk63n6zVVRUjUbTmeF1cXkNwgb1x0JMaq1faVMSw04
EoJnTDn3f3YFEqz83dcdeFnaThoSUGQhsVhK2P1T560B/Z1785o1NuI340h2vgQx5EeQqQhnpZ
U9vAYMR2r19m130wMoobJVCntvsvY8yVquntuccCYLSVj+7z2pnlWHC1eW7gRgn16e6h250Qa
AgGras5B52DZXVfocMK1lWbxxKTfV8E8SPNBjLRAqZKZwSr1WYQSElK6BzUwNpH+4lWun+NC
FJ1924472Kduqs+SSPloFq+ZnwKTjdiOn2qCwO99eAKj76ZLDFMaoyDAG2eIITf3J6e6H5K1Dz
CT85KjDQUN4zyBy/oKruMfYgmsh+LYZc5yl/TXNkIntawNt6CIMI1tzh9m/XBdLx3sR/Zbo76N05rE
Ews= hadoop@namenode

hadoop@datanode2:~$ cd -
hadoop@datanode2:~$ cat .ssh/authorized_keys
ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQACQD11641Se60JFHKS1e8OhdsuNaBRHw+BQ126uNQV
d+hbJl+PKCVHSEKERB3JUCSpd0LVL/5nuPj+6qZQMeqZfysFMLbs918h3e6b4qeoUXCZGBLP7oeZ
856R+wCONPKysdLoZgrogISVLSjJ1INWwuc81/MU3p1c3X8DzSQW8013rDs9yU+n+484sc25e15E
S36UZEulaLXDBMKg3LOp8ozJ12L089J9CS+Kk63n6zVVRUjUbTmeF1cXkNwgb1x0JMaq1faVMSw04
EoJnTDn3f3YFEqz83dcdeFnaThoSUGQhsVhK2P1T560B/Z1785o1NuI340h2vgQx5EeQqQhnpZ
U9vAYMR2r19m130wMoobJVCntvsvY8yVquntuccCYLSVj+7z2pnlWHC1eW7gRgn16e6h250Qa
AgGras5B52DZXVfocMK1lWbxxKTfV8E8SPNBjLRAqZKZwSr1WYQSElK6BzUwNpH+4lWun+NC
FJ1924472Kduqs+SSPloFq+ZnwKTjdiOn2qCwO99eAKj76ZLDFMaoyDAG2eIITf3J6e6H5K1Dz
CT85KjDQUN4zyBy/oKruMfYgmsh+LYZc5yl/TXNkIntawNt6CIMI1tzh9m/XBdLx3sR/Zbo76N05rE
Ews= hadoop@namenode
```

lsblk de los nodos

```
victorgonzbaj@namenode:~$ lsblk
NAME MAJ:MIN RM SIZE RO TYPE MOUNTPOINTS
loop0 7:0 0 4K 1 loop /snap/bare/s
loop1 7:1 0 73,9M 1 loop /snap/core22/1722
loop2 7:2 0 73,9M 1 loop /snap/core22/1748
loop3 7:3 0 269,8M 1 loop /snap/firefox/4793
loop4 7:4 0 595,1M 1 loop /snap/gnome-42-2284/176
loop5 7:5 0 11,1M 1 loop /snap/firmware-updater/147
loop6 7:6 0 10,7M 1 loop /snap/firmware-updater/127
loop7 7:7 0 275,3M 1 loop /snap/firefox/5647
loop8 7:8 0 516M 1 loop /snap/gnome-42-2284/282
loop9 7:9 0 91,7M 1 loop /snap/gtk-common-themes/1535
loop10 7:10 0 10,5M 1 loop /snap/snap-store/1173
loop11 7:11 0 10,8M 1 loop /snap/snap-store/1248
loop12 7:12 0 38,8M 1 loop /snap/snapd/21759
loop13 7:13 0 500K 1 loop /snap/snapd-desktop-integration/178
loop14 7:14 0 568K 1 loop /snap/snapd-desktop-integration/253
loop15 7:15 0 44,4M 1 loop /snap/snapd/23545
sda 8:0 0 64G 0 disk
└─sda1 8:1 0 1M 0 part
└─sda2 8:2 0 64G 0 part /
└─sdb 8:16 0 100G 0 disk
└─sdb1 8:17 0 100G 0 part /mnt/data
sr0 11:0 1 1024M 0 rom

victorgonzbaj@datanode1:~$ lsblk
NAME MAJ:MIN RM SIZE RO TYPE MOUNTPOINTS
loop0 7:0 0 4K 1 loop /snap/bare/s
loop1 7:1 0 73,9M 1 loop /snap/core22/1722
loop2 7:2 0 269,8M 1 loop /snap/firefox/4793
loop3 7:3 0 275,3M 1 loop /snap/firefox/5647
loop4 7:4 0 10,7M 1 loop /snap/firmware-updater/127
loop5 7:5 0 73,9M 1 loop /snap/core22/1748
loop6 7:6 0 595,1M 1 loop /snap/gnome-42-2284/176
loop7 7:7 0 26 1 loop /snap/kf6-core22/40
loop8 7:8 0 11,1M 1 loop /snap/firmware-updater/147
loop9 7:9 0 91,7M 1 loop /snap/gtk-common-themes/1535
loop10 7:10 0 4,5M 1 loop /snap/kpat/92
loop11 7:11 0 10,8M 1 loop /snap/snap-store/1248
loop12 7:12 0 10,5M 1 loop /snap/snap-store/1173
loop13 7:13 0 44,4M 1 loop /snap/snapd/23545
loop14 7:14 0 38,8M 1 loop /snap/snapd/21759
loop15 7:15 0 500K 1 loop /snap/snapd-desktop-integration/178
loop16 7:16 0 568K 1 loop /snap/snapd-desktop-integration/253
sda 8:0 0 64G 0 disk
└─sda1 8:1 0 1M 0 part
└─sda2 8:2 0 64G 0 part /
└─sdb 8:16 0 100G 0 disk
└─sdb1 8:17 0 100G 0 part /mnt/data
sr0 11:0 1 1024M 0 rom

victorgonzbaj@datanode2:~$ lsblk
NAME MAJ:MIN RM SIZE RO TYPE MOUNTPOINTS
loop0 7:0 0 4K 1 loop /snap/bare/s
loop1 7:1 0 73,9M 1 loop /snap/core22/1722
loop2 7:2 0 73,9M 1 loop /snap/core22/1748
loop3 7:3 0 269,8M 1 loop /snap/firefox/4793
loop4 7:4 0 595,1M 1 loop /snap/gnome-42-2284/176
loop5 7:5 0 11,1M 1 loop /snap/firmware-updater/147
loop6 7:6 0 10,7M 1 loop /snap/firmware-updater/127
loop7 7:7 0 10,5M 1 loop /snap/snap-store/1173
loop8 7:8 0 91,7M 1 loop /snap/gtk-common-themes/1535
loop9 7:9 0 10,8M 1 loop /snap/snap-store/1248
loop10 7:10 0 44,4M 1 loop /snap/snapd/23545
loop11 7:11 0 500K 1 loop /snap/snapd-desktop-integration/178
loop12 7:12 0 38,8M 1 loop /snap/snapd/21759
loop13 7:13 0 568K 1 loop /snap/snapd-desktop-integration/253
sda 8:0 0 64G 0 disk
└─sda1 8:1 0 1M 0 part
└─sda2 8:2 0 64G 0 part /
└─sdb 8:16 0 100G 0 disk
└─sdb1 8:17 0 100G 0 part /mnt/data
sr0 11:0 1 1024M 0 rom
```

Configuración de Hadoop

Ver variables HADOOP_HOME y PATH del namenode

```
hadoop@namenode:~$ echo $HADOOP_HOME
/home/hadoop/hadoop-3.4.0
hadoop@namenode:~$ echo $PATH
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/snap/bin:/home/hadoop/hadoop-3.4.0/bin:/home/hadoop/hadoop-3.4.0/sbin:/home/hadoop/hadoop-3.4.0/bin
hadoop@namenode:~$
```

Configuración del archivo core-site.xml en cada nodo.

<pre><configuration> <property> <name>hadoop.tmp.dir</name> <value>/mnt/data/tmpdata</value> </property> <property> <name>fs.default.name</name> <value>hdfs://namenode:9000</value> </property> </configuration> hadoop@namenode:~\$</pre>	<pre><configuration> <property> <name>hadoop.tmp.dir</name> <value>/mnt/data/tmpdata</value> </property> <property> <name>fs.default.name</name> <value>hdfs://namenode:9000</value> </property> </configuration> hadoop@datanode1:~\$</pre>	<pre><configuration> <property> <name>hadoop.tmp.dir</name> <value>/mnt/data/tmpdata</value> </property> <property> <name>fs.default.name</name> <value>hdfs://namenode:9000</value> </property> </configuration> hadoop@datanode2:~\$</pre>
---	--	--

Namenode sin ser datanode

Configuración del archivo hdfs-site.xml en cada nodo.

<pre><configuration> <property> <name>dfs.namenode.name.dir</name> <value>file:/mnt/data/hdfs/namenode</value> </property> <!-- <property> <name>dfs.data.dir</name> <value>file:/mnt/data/hdfs/datanode</value> </property--> <property> <name>dfs.replication</name> <value>2</value> </property> <!-- <property> <name>dfs.hosts.exclude</name> <value>hadoop-3.4.0/etc/hadoop/noworkers</value> </property--> </configuration> hadoop@namenode:~\$</pre>	<pre> You may obtain a copy of the License at http://www.apache.org/licenses/LICENSE-2.0 Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License. See accompanying LICENSE file. --> <!-- Put site-specific property overrides in this file. --> <configuration> <property> <name>dfs.data.dir</name> <value>file:/mnt/data/hdfs/datanode</value> </property> </configuration> Firefox vtanode1:~\$</pre>	<pre> You may obtain a copy of the License at http://www.apache.org/licenses/LICENSE-2.0 Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License. See accompanying LICENSE file. --> <!-- Put site-specific property overrides in this file. --> <configuration> <property> <name>dfs.data.dir</name> <value>file:/mnt/data/hdfs/datanode</value> </property> </configuration> hadoop@datanode2:~\$</pre>
--	---	---

Información del sistema hdfs, espacio disponible

Overview 'namenode:9000' (✔active)

Started:	Tue Feb 04 20:57:37 +0100 2025
Version:	3.4.0, rbd8b77f398f626bb7791783192ee7a5dfaec760
Compiled:	Mon Mar 04 07:35:00 +0100 2024 by root from (HEAD detached at release-3.4.0-RC3)
Cluster ID:	CID-0a84a87e-39b0-449a-b650-3e5b9af34002
Block Pool ID:	BP-1317947451-192.168.1.1-1738613828265

Summary

Security is off.
Safemode is off.
1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).
Heap Memory used 202.15 MB of 262 MB Heap Memory. Max Heap Memory is 871.5 MB.
Non Heap Memory used 53.6 MB of 55.28 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	195.74 GB
Configured Remote Capacity:	0 B
DFS Used:	60 KB (0%)
Non DFS Used:	96 KB
DFS Remaining:	185.71 GB (94.88%)
Block Pool Used:	60 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	2 (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)
Number of Under-Replicated Blocks	0

Crear algunas carpetas y subir algún fichero grande a alguna de ellas

```
hadoop@namenode:~$ hdfs dfs -mkdir /prueba
hadoop@namenode:~$ hdfs dfs -put text.txt /
hadoop@namenode:~$ hdfs dfs -ls /
Found 2 items
drwxr-xr-x   - hadoop supergroup          0 2025-02-04 21:05 /prueba
-rw-r--r--   2 hadoop supergroup    2141520 2025-02-04 21:05 /text.txt
hadoop@namenode:~$
```

Mostrar la estructura y algún fichero con bloques replicados y en que nodos están



Namenode siendo datanode

Mostrar archivo hdfs-site.xml de ese hosts. Mostrar archivo Workers.

```
<!-- Put site-specific property overrides in this file. -->

<configuration>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:/mnt/data/hdfs/namenode</value>
  </property>
  <property>
    <name>dfs.data.dir</name>
    <value>file:/mnt/data/hdfs/datanode</value>
  </property>
  <property>
    <name>dfs.replication</name>
    <value>3</value>
  </property>
  <!--<property>
    <name>dfs.hosts.exclude</name>
    <value>hadoop-3.4.0/etc/hadoop/noworkers</value>
  </property-->
</configuration>
hadoop@namenode:~$
```

```
hadoop@namenode:~$ cat hadoop-3.4.0/etc/hadoop/workers
namenode
datanode1
datanode2
```

Eliminar el fichero de antes y volverlo a subir

```
hadoop@namenode:~$ hdfs dfs -rm /text.txt
Deleted /text.txt
hadoop@namenode:~$ hdfs dfs -put text.txt /
hadoop@namenode:~$
```

Mostrar de nuevo en un navegador los nodos activos y los bloques que ocupa el fichero

Overview 'namenode:9000' (✔active)

Started:	Tue Feb 04 21:11:48 +0100 2025
Version:	3.4.0, rbd8b77f398f626bb7791783192ee7a5dfaeec760
Compiled:	Mon Mar 04 07:35:00 +0100 2024 by root from (HEAD detached at release-3.4.0-RC3)
Cluster ID:	CID-0a84a87e-39b0-449a-b650-3e5b9af34002
Block Pool ID:	BP-1317947451-192.168.1.1-1738613828265

Summary

Security is off.

Safemode is off.

3 files and directories, 1 blocks (1 replicated blocks, 0 erasure coded block groups) = 4 total filesystem object(s).

Heap Memory used 168.76 MB of 233 MB Heap Memory. Max Heap Memory is 871.5 MB.

Non Heap Memory used 58.42 MB of 59.97 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	293.61 GB
Configured Remote Capacity:	0 B
DFS Used:	6.27 MB (0%)
Non DFS Used:	5.22 MB
DFS Remaining:	278.55 GB (94.87%)
Block Pool Used:	6.27 MB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	3 (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)

File information - text.txt

Download

Head the file (first 32K)

Tail the file (last 32K)

Block information -- Block 0 ▾

Block ID: 1073741827

Block Pool ID: BP-1317947451-192.168.1.1-1738613828265

Generation Stamp: 1003

Size: 2141520

Availability:

- namenode
- datanode1
- datanode2

Close

Modificar hdfs-site.xml para disminuir el tiempo de latencia de comprobación de nodos activos y mostrar esa configuración

```
<property>
  <name>dfs.namenode.name.dir</name>
  <value>file:/mnt/data/hdfs/namenode</value>
</property>
<property>
  <name>dfs.data.dir</name>
  <value>file:/mnt/data/hdfs/datanode</value>
</property>
<property>
  <name>dfs.replication</name>
  <value>3</value>
</property>
<property>
  <name>dfs.namenode.heartbeat.recheck-interval</name>
  <value>5000</value>  <!--ms-->
<property>
  <name>dfs.heartbeat.interval</name>
  <value>3</value>  <!--s-->
</property>
<!--<property>
  <name>dfs.hosts.exclude</name>
  <value>hadoop-3.4.0/etc/hadoop/noworkers</value>
</property>-->
</configuration>
hadoop@namenode:~$
```

Decomisar un host que no sea el namenode

```
</property>
<property>
  <name>dfs.replication</name>
  <value>3</value>
</property>
<property>
  <name>dfs.namenode.heartbeat.recheck-interval</name>
  <value>5000</value>  <!--ms-->
<property>
  <name>dfs.heartbeat.interval</name>
  <value>3</value>  <!--s-->
</property>
<property>
  <name>dfs.hosts.exclude</name>
  <value>hadoop-3.4.0/etc/hadoop/noworkers</value>
</property>
</configuration>
hadoop@namenode:~$ cat hadoop-3.4.0/etc/hadoop/noworkers
namenode
hadoop@namenode:~$
```

Mostrar en el navegador la nueva situación y como ha replicado los bloques de nuevos

Decommissioning

Node	Under replicated blocks	Blocks with no live replicas	Under Replicated Blocks In files under construction	Decommission duration
/default-rack/namenode:9866 (192.168.1.1:9866)	1	0	0	42587ms

File information - texto2.txt

Download

Head the file (first 32K)

Tail the file (last 32K)

Block information --

Block 0

Block ID: 1073741828

Block Pool ID: BP-1317947451-192.168.1.1-1738613828265

Generation Stamp: 1004

Size: 2141520

Availability:

- datanode2
- datanode1

Close