

# BDA02-02

## Practica Creación de un clúster HDFS

DNI: 39465322J

Alumno : Bieito Sousa Barreiro

Url Entrega : [Creación de un clúster HDFS](#)

### Objetivo

- Demostrar capacidad para crear una red de equipos interconectados.

# Enunciado

1.- Realiza una conexión ssh de tu nodo1 a tu nodo4 usando el nombre del nodo y sin necesidad de introducir contraseña.

Captura pantalla donde se vea que usas el nombre del nodo en lugar de su IP y que no has tenido que introducir contraseña.

```
hadoop@bie1:~$ ssh hadoop@bie4
Welcome to Ubuntu 22.04.3 LTS (GNU/Linux 5.15.0-94-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

System information as of sáb 17 feb 2024 17:25:15 UTC

System load:  0.0               Processes:            119
Usage of /:   29.9% of 23.45GB   Users logged in:     1
Memory usage: 11%               IPv4 address for enp0s3: 192.168.0.104
Swap usage:   0%

 * Strictly confined Kubernetes makes edge and IoT secure. Learn how MicroK8s
just raised the bar for easy, resilient and secure K8s cluster deployment
.

https://ubuntu.com/engage/secure-kubernetes-at-the-edge

El mantenimiento de seguridad expandido para Applications está desactivado

Se pueden aplicar 55 actualizaciones de forma inmediata.
3 de estas son actualizaciones de seguridad estándares.
Para ver estas actualizaciones adicionales, ejecute: apt list --upgradable

Active ESM Apps para recibir futuras actualizaciones de seguridad adicionales.
Vea https://ubuntu.com/esm o ejecute «sudo pro status»

Last login: Sat Feb 17 16:45:48 2024 from 192.168.0.21
hadoop@bie4:~$
```

2.- Comprueba en la web <https://hadoop.apache.org/releases> la ruta del paquete hadoop-3.3.6 para tu plataforma, descárgalo con wget y descomprímelo de manera que su contenido quede en la ruta /home/hadoop/hadoop .

Repite esta operación en todos los nodos. Realiza una captura de pantalla del comando `ls -la /home/hadoop/hadoop` .

```
hadoop@bie1:~$ ls -la /home/hadoop/hadoop
total 120
drwxr-xr-x 10 hadoop hadoop 4096 jun 18 2023 .
drwxr-x-- 6 hadoop hadoop 4096 feb 16 01:46 ..
drwxr-xr-x 2 hadoop hadoop 4096 jun 18 2023 bin
drwxr-xr-x 3 hadoop hadoop 4096 jun 18 2023 etc
drwxr-xr-x 2 hadoop hadoop 4096 jun 18 2023 include
drwxr-xr-x 3 hadoop hadoop 4096 jun 18 2023 lib
drwxr-xr-x 4 hadoop hadoop 4096 jun 18 2023 libexec
-rw-rw-r-- 1 hadoop hadoop 24276 jun 14 2023 LICENSE-binary
drwxr-xr-x 2 hadoop hadoop 4096 jun 18 2023 licenses-binary
-rw-rw-r-- 1 hadoop hadoop 15217 jun 9 2023 LICENSE.txt
-rw-rw-r-- 1 hadoop hadoop 29473 jun 9 2023 NOTICE-binary
-rw-rw-r-- 1 hadoop hadoop 1541 jun 9 2023 NOTICE.txt
-rw-rw-r-- 1 hadoop hadoop 175 jun 9 2023 README.txt
drwxr-xr-x 3 hadoop hadoop 4096 jun 18 2023 sbin
drwxr-xr-x 4 hadoop hadoop 4096 jun 18 2023 share
hadoop@bie1:~$
```

```
hadoop@bie2:~$ ls -la /home/hadoop/hadoop
total 120
drwxr-xr-x 10 hadoop hadoop 4096 feb 16 01:49 .
drwxr-x-- 6 hadoop hadoop 4096 feb 16 01:48 ..
drwxr-xr-x 2 hadoop hadoop 4096 feb 16 01:48 bin
drwxr-xr-x 3 hadoop hadoop 4096 feb 16 01:49 etc
drwxr-xr-x 2 hadoop hadoop 4096 feb 16 01:49 include
drwxr-xr-x 3 hadoop hadoop 4096 feb 16 01:48 lib
drwxr-xr-x 4 hadoop hadoop 4096 feb 16 01:48 libexec
-rw-rw-r-- 1 hadoop hadoop 24276 feb 16 01:48 LICENSE-binary
drwxr-xr-x 2 hadoop hadoop 4096 feb 16 01:48 licenses-binary
-rw-rw-r-- 1 hadoop hadoop 15217 feb 16 01:48 LICENSE.txt
-rw-rw-r-- 1 hadoop hadoop 29473 feb 16 01:49 NOTICE-binary
-rw-rw-r-- 1 hadoop hadoop 1541 feb 16 01:48 NOTICE.txt
-rw-rw-r-- 1 hadoop hadoop 175 feb 16 01:48 README.txt
drwxr-xr-x 3 hadoop hadoop 4096 feb 16 01:48 sbin
drwxr-xr-x 4 hadoop hadoop 4096 feb 16 01:49 share
hadoop@bie2:~$
```

```
hadoop@bie3:~$ ls -la /home/hadoop/hadoop
total 120
drwxr-xr-x 10 hadoop hadoop 4096 feb 16 01:50 .
drwxr-x-- 6 hadoop hadoop 4096 feb 16 01:49 ..
drwxr-xr-x 2 hadoop hadoop 4096 feb 16 01:49 bin
drwxr-xr-x 3 hadoop hadoop 4096 feb 16 01:50 etc
drwxr-xr-x 2 hadoop hadoop 4096 feb 16 01:50 include
drwxr-xr-x 3 hadoop hadoop 4096 feb 16 01:49 lib
drwxr-xr-x 4 hadoop hadoop 4096 feb 16 01:49 libexec
-rw-rw-r-- 1 hadoop hadoop 24276 feb 16 01:49 LICENSE-binary
drwxr-xr-x 2 hadoop hadoop 4096 feb 16 01:49 licenses-binary
-rw-rw-r-- 1 hadoop hadoop 15217 feb 16 01:49 LICENSE.txt
-rw-rw-r-- 1 hadoop hadoop 29473 feb 16 01:50 NOTICE-binary
-rw-rw-r-- 1 hadoop hadoop 1541 feb 16 01:49 NOTICE.txt
-rw-rw-r-- 1 hadoop hadoop 175 feb 16 01:49 README.txt
drwxr-xr-x 3 hadoop hadoop 4096 feb 16 01:49 sbin
drwxr-xr-x 4 hadoop hadoop 4096 feb 16 01:50 share
hadoop@bie3:~$
```

```
hadoop@bie4:~$ ls -la /home/hadoop/hadoop
total 120
drwxr-xr-x 10 hadoop hadoop 4096 feb 16 01:57 .
drwxr-x-- 6 hadoop hadoop 4096 feb 16 01:56 ..
drwxr-xr-x 2 hadoop hadoop 4096 feb 16 01:56 bin
drwxr-xr-x 3 hadoop hadoop 4096 feb 16 01:57 etc
drwxr-xr-x 2 hadoop hadoop 4096 feb 16 01:57 include
drwxr-xr-x 3 hadoop hadoop 4096 feb 16 01:56 lib
drwxr-xr-x 4 hadoop hadoop 4096 feb 16 01:56 libexec
-rw-rw-r-- 1 hadoop hadoop 24276 feb 16 01:56 LICENSE-binary
drwxr-xr-x 2 hadoop hadoop 4096 feb 16 01:56 licenses-binary
-rw-rw-r-- 1 hadoop hadoop 15217 feb 16 01:56 LICENSE.txt
-rw-rw-r-- 1 hadoop hadoop 29473 feb 16 01:57 NOTICE-binary
-rw-rw-r-- 1 hadoop hadoop 1541 feb 16 01:56 NOTICE.txt
-rw-rw-r-- 1 hadoop hadoop 175 feb 16 01:56 README.txt
drwxr-xr-x 3 hadoop hadoop 4096 feb 16 01:56 sbin
drwxr-xr-x 4 hadoop hadoop 4096 feb 16 01:57 share
hadoop@bie4:~$
```

3.- Edita el archivo “/etc/environment” para añadir lo siguiente:

- Una línea con:  
`JAVA_HOME="/usr/lib/jvm/java-8-openjdk-amd64"`
- Una línea con: `HADOOP_HOME="/home/hadoop/hadoop"`
- Añadir al final del `PATH` las rutas de `bin` y `sbin` de hadoop (ojo con los dos puntos y las comillas): `/home/hadoop/hadoop/bin:/home/hadoop/hadoop/sbin`

Realiza una captura del comando `cat /etc/environment`. Explica con tus palabras para que sirve lo que acabas de hacer en este archivo.

El archivo `/etc/environment`

1. archivo de configuración
2. se utiliza para establecer **variables de entorno globales** para **todos los usuarios del sistema**.
3. **variables de entorno** se aplican a **todos los procesos y programas** que se ejecutan en el sistema.

Necesitamos definir un entorno en el cual Hadoop pueda ejecutar acciones en cualquiera de los nodos o en varios de los mismos de manera simultanea, sin preocuparnos de la configuración individual de cada nodo.

Para ello tenemos que definir un entorno de manera global y satisfacer en el las necesidades de hadoop en este caso seria el acceso a la ejecución de binarios y archivos de configuración tanto de java como del propio hadoop para ello definimos en `/etc/environment` dos variables

- `HADOOP_HOME` con la carpeta raíz de hadoop la cual contiene archivos de configuración y ejecutables de hadoop
- `JAVA_HOME` con la carpeta raíz de hadoop la cual contiene archivos de configuración y ejecutables de java.

a su vez es necesario definir en `$PATH` la ruta de los binarios tanto de java como de hadoop

```
hadoop@bie1:~$ cat /etc/environment
PATH="/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/snap/bin:/home/hadoop/hadoop/bin:/home/hadoop/hadoop/sbin"
JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
HADOOP_HOME="/home/hadoop/hadoop"
hadoop@bie1:~$

hadoop@bie2:~$ cat /etc/environment
PATH="/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/snap/bin:/home/hadoop/hadoop/bin:/home/hadoop/hadoop/sbin"
JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
HADOOP_HOME="/home/hadoop/hadoop"
hadoop@bie2:~$

hadoop@bie3:~$ cat /etc/environment
PATH="/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/snap/bin:/home/hadoop/hadoop/bin:/home/hadoop/hadoop/sbin"
JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
HADOOP_HOME="/home/hadoop/hadoop"
hadoop@bie3:~$

hadoop@bie4:~$ cat /etc/environment
PATH="/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/snap/bin:/home/hadoop/hadoop/bin:/home/hadoop/hadoop/sbin"
JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
HADOOP_HOME="/home/hadoop/hadoop"
hadoop@bie4:~$
```

4.- Modifica la configuración del archivo `home/hadoop/hadoop/etc/hadoop/core-site.xml` para indicar que la propiedad `fs.defaultFS` tiene un valor de `hdfs://nodo1:9000` . Donde `nodo1` será uno de tus nodos que has elegido para la función de NameNode. Haz lo mismo en todos los nodos. Realiza la captura de pantalla del comando `cat /home/hadoop/hadoop/etc/core-site.xml` en cualquiera de los nodos.

PROBLEMAS

SALIDA

CONSOLA DE DEPURACIÓN

TERMINAL

PUERTOS

```
hadoop@bie1:~$ cat /home/hadoop/hadoop/etc/hadoop/core-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  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>fs.defaultFS</name>
    <value>hdfs://bie1:9000</value>
  </property>

</configuration>
hadoop@bie1:~$
```

bash bda02-02-task

nodo1 - bie1

nodo2 - bie2

nodo3 - bie3

nodo4 - bie4

¿Qué estamos indicando en esa configuración?

Este archivo `core-site.xml` se utiliza para especificar la configuración centralizada para el sistema de archivos y otros aspectos fundamentales del clúster de Hadoop

type	Concepto	descripción
Propiedad	<code>fs.defaultFS</code>	Propiedad que define el sistema de archivos por defecto que utilizará Hadoop
value	<code>hdfs://bie1:9000</code>	<code>hdfs</code> Sistema de archivos por defecto utilizado por Hadoop será HDFS (Hadoop Distributed File System) <code>bie1</code> la dirección del namenode <code>:9000</code> puerto 9000.

¿Por qué todos los nodos comparten la la misma configuración en el archivo core-site.xml?

En este caso definimos que el nodo bie1 actuara como namenode esta configuración se aplica en todos los nodos, puesto que es una configuración global y todos ellos tienen que saber la dirección del nodo que actuara como namenode, es decir donde se guardaran la tabla direcciones.

```
<configuration>
<!--
`fs.defaultFS`
  propiedad que define el sistema de archivos por defecto que utilizará Hadoop , in
  hdfs://bie1:9000=
`hdfs`    sistema de archivos por defecto utilizado por Hadoop será HDFS (Hadoop Dis
`bie1` la dirección del namenode
`:9000`  puerto 9000.
-->
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://bie1:9000</value>
</property>

</configuration>
```

## 5.- En el nodo que realizará las funciones de NameNode modifica el archivo

/home/hadoop/hadoop/etc/hadoop/hdfs-site.xml para incluir las siguientes propiedades:

- **Nombre:** dfs.namenode.name.dir **con valor** /home/hadoop/discogrande/namenode
- **Nombre:** dfs.replication **con valor** 2

Realiza una captura de pantalla del comando

```
cat /home/hadoop/hadoop/etc/hadoop/hdfs-site.xml .
```

¿Qué indicamos con la primera propiedad?

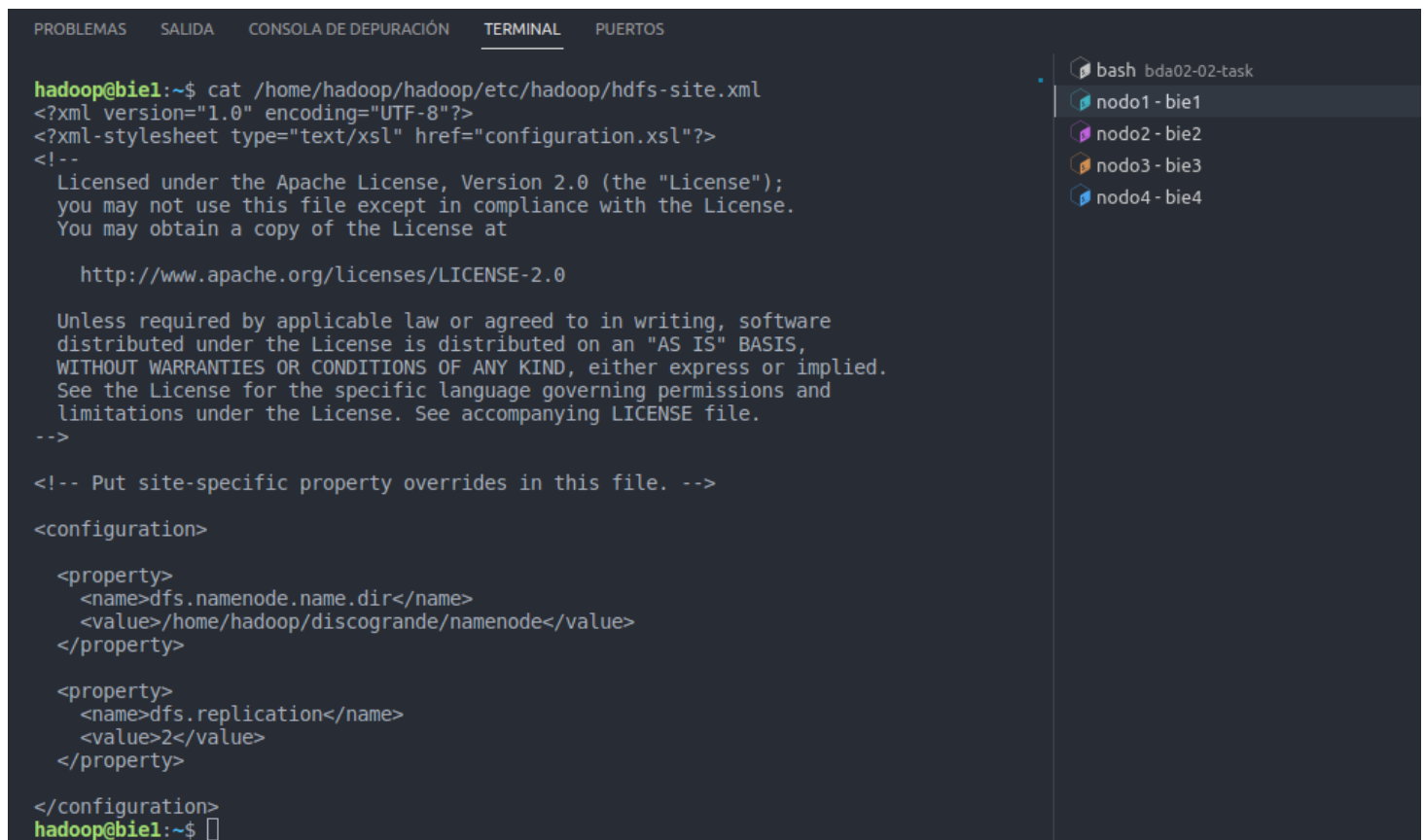
dfs.namenode.name.dir Ruta donde esta guarda la tabla de direcciones de namenode

¿Qué indicamos con la segunda propiedad?

dfs.replication : Cuando guardas un archivo en Hadoop, el archivo se divide en fragmentos de datos llamados "bloques", pore defecto (128 MB). estos bloques se guardan en n nodos siendo n el valor de dfs.replication, en este caso como el valor es 2 se guarda cada "bloque" en 2 nodos.

¿Por qué no replicamos el contenido de este archivo en el resto de los nodos?

en este archivo el nodo1 actúa como namenode, por lo cual unicamente definimos en el la configuración relativa al proceso de namenode, mientras que nodo2,3,4 van a actuará como datanode y definimos en ello de manera idéntica la configuración relativa a datanode



```
PROBLEMAS  SALIDA  CONSOLA DE DEPURACIÓN  TERMINAL  PUERTOS

hadoop@biel:~$ cat /home/hadoop/hadoop/etc/hadoop/hdfs-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  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.namenode.name.dir</name>
    <value>/home/hadoop/discogrande/namenode</value>
  </property>

  <property>
    <name>dfs.replication</name>
    <value>2</value>
  </property>

</configuration>
hadoop@biel:~$
```

 /home/hadoop/hadoop/etc/hadoop/hdfs-site.xml **en** namenode

```
<!--
more configuration property
https://hadoop.apache.org/docs/r2.7.0/hadoop-project-dist/hadoop-hdfs/hdfs-default.

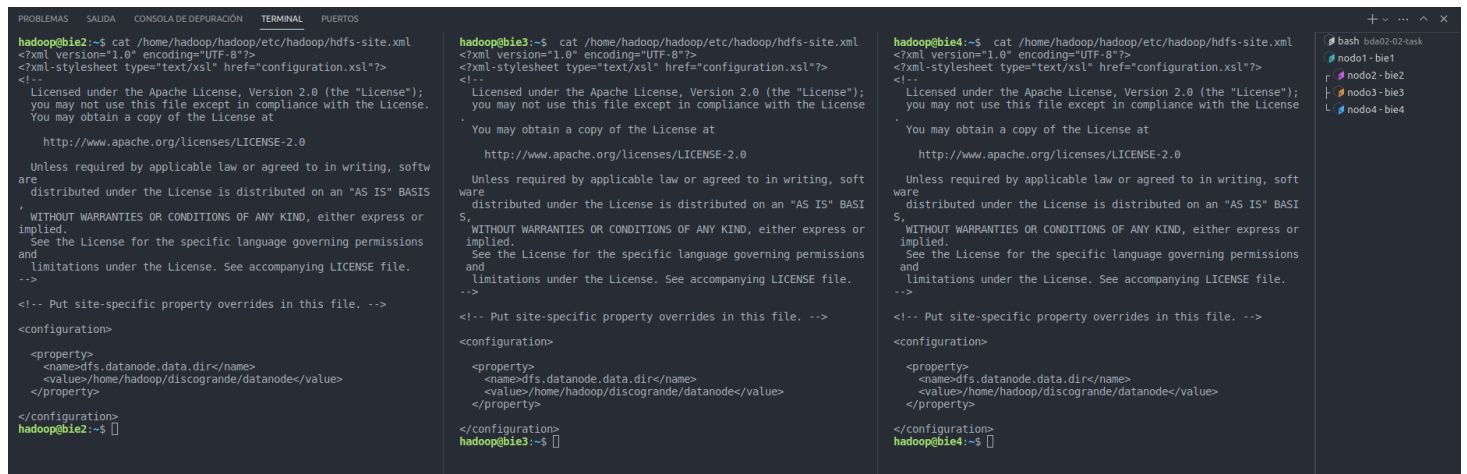
nota:
al modificar este archivo se puede configurar en caliente
hdfs dfsadmin -reconfig namenode biel:9000 start
-->
<configuration>
<!--
dfs.namenode.name.dir
file://${hadoop.tmp.dir}/dfs/name
Determines where on the local filesystem the DFS name node should store the name ta
- - - - -
Ruta donde esta guarda la table de direcciones de namenode
-->
<property>
  <name>dfs.namenode.name.dir</name>
  <value>/home/hadoop/discogrande/namenode</value>
</property>
<!--
dfs.replication
default 3
Default block replication. The actual number of replications can be specified when
- - - - -
Cuando guardas un archivo en Hadoop, el archivo se divide en fragmentos de datos 11
-->
<property>
  <name>dfs.replication</name>
  <value>2</value>
</property>

</configuration>
```



6.- En cada uno de los nodos que realizarán las funciones de DataNode modifica el archivo `/home/hadoop/hadoop/etc/hadoop/hdfs-site.xml` para que incluya la propiedad de `dfs.datanode.data.dir` con el valor `/home/hadoop/discogrande/datanode`.

Captura la pantalla del comando `cat /home/hadoop/hadoop/etc/hadoop/hdfs-site`



```
hadoop@bie2:~$ cat /home/hadoop/hadoop/etc/hadoop/hdfs-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
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.datanode.data.dir</name>
    <value>/home/hadoop/discogrande/datanode</value>
  </property>
</configuration>
hadoop@bie2:~$
```

```
hadoop@bie3:~$ cat /home/hadoop/hadoop/etc/hadoop/hdfs-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License
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.datanode.data.dir</name>
    <value>/home/hadoop/discogrande/datanode</value>
  </property>
</configuration>
hadoop@bie3:~$
```

```
hadoop@bie4:~$ cat /home/hadoop/hadoop/etc/hadoop/hdfs-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License
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.datanode.data.dir</name>
    <value>/home/hadoop/discogrande/datanode</value>
  </property>
</configuration>
hadoop@bie4:~$
```

Explica con tus palabras la posible razón por la que el mismo archivo “hdfs-site.xml” es distinto en NameNodes y en DataNodes.

En esta practica por optimización decidimos definir que en nodo1 se establece la configuración de la tabla de namenode, por lo tanto se encapsula la configuración de todo lo relacionado con namenode en nodo1, asimismo decidimos que nodo2,3,4 actuaran como datanode.

Justifica con tus palabras si consideras que sería bueno (o no) fusionar ambas versiones de “hdfs-site.xml” en una sola en la que aparezcan todas las propiedades.

si tenemos los recursos disponibles para poder aislar un nodo para que su función sea exclusiva de namenode seria lo optimo, teniendo en cuenta que en este nodo se define la tabla de direcciones que determina en que nodo y en que espacio de memoria esta guardaos los datos, y que hadoop para ejecutar sus acciones con relación a los datos el único nodo imprescindible es el que contenga namenode, si me parece una buena practica incluso tener un nodo cuya función exclusiva sea actuar de back-up del nodo namenode por si este cae no estaría mal.

## 7.- Formatea el sistema de ficheros HDFS en el NameNode.

Captura la pantalla del comando que has usado y su salida. Indica qué comando has usado. Indica desde qué nodo has ejecutado el comando.

```
hdfs namenode -form
```

```
hadoop@biel:~$ hdfs namenode -form
2024-02-19 04:47:22,357 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = biel/192.168.0.101
STARTUP_MSG: args = [-form]
STARTUP_MSG: version = 3.3.6
STARTUP_MSG: classpath = /home/hadoop/hadoop/etc/hadoop:/home/hadoop/hadoop/share/hadoop/common/lib/kerb-identity-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/commons-logging-1.1.3.jar:/home/hadoop/hadoop/share/hadoop/common/lib/commons-compress-1.21.jar:/home/hadoop/hadoop/share/hadoop/common/lib/curator-recipes-5.2.0.jar:/home/hadoop/hadoop/share/hadoop/common/lib/commons-beanutils-1.9.4.jar:/home/hadoop/hadoop/share/hadoop/common/lib/jackson-core-2.12.7.jar:/home/hadoop/hadoop/share/hadoop/common/lib/netty-transport-rxtx-4.1.89.Final.jar:/home/hadoop/hadoop/share/hadoop/common/lib/jetty-util-ajax-9.4.51.v20230217.jar:/home/hadoop/hadoop/share/hadoop/common/lib/netty-codec-stomp-4.1.89.Final.jar:/home/hadoop/hadoop/share/hadoop/common/lib/netty-resolver-dns-native-macos-4.1.89.Final-osx-aarch64.jar:/home/hadoop/hadoop/share/hadoop/common/lib/protobuf-java-2.5.0.jar:/home/hadoop/hadoop/share/hadoop/common/lib/kerb-simplekdc-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/netty-resolver-dns-native-macos-4.1.89.Final-osx-x86_64.jar:/home/hadoop/hadoop/share/hadoop/common/lib/jackson-core-asl-1.9.13.jar:/home/hadoop/hadoop/share/hadoop/common/lib/paranamer-2.3.jar:/home/hadoop/hadoop/share/hadoop/common/lib/reload4j-1.2.22.jar:/home/hadoop/hadoop/share/hadoop/common/lib/netty-codec-socks-4.1.89.Final.jar:/home/hadoop/hadoop/share/hadoop/common/lib/re2j-1.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/hadoop-auth-3.3.6.jar:/home/hadoop/hadoop/share/hadoop/yarn/hadoop-yarn-server-tests-3.3.6.jar:/home/hadoop/hadoop/share/hadoop/yarn/hadoop-yarn-api-3.3.6.jar:/home/hadoop/hadoop/share/hadoop/yarn/hadoop-yarn-client-3.3.6.jar:/home/hadoop/hadoop/share/hadoop/yarn/hadoop-yarn-services-core-3.3.6.jar:/home/hadoop/hadoop/share/hadoop/yarn/hadoop-yarn-server-sharedcachemanager-3.3.6.jar:/home/hadoop/hadoop/share/hadoop/yarn/hadoop-yarn-server-applicationhistoryservice-3.3.6.jar:/home/hadoop/hadoop/share/hadoop/yarn/hadoop-yarn-services-api-3.3.6.jar:/home/hadoop/hadoop/share/hadoop/yarn/hadoop-yarn-registry-3.3.6.jar:/home/hadoop/hadoop/share/hadoop/yarn/hadoop-yarn-common-3.3.6.jar:/home/hadoop/hadoop/share/hadoop/yarn/hadoop-yarn-server-web-proxy-3.3.6.jar:/home/hadoop/hadoop/share/hadoop/yarn/hadoop-yarn-server-router-3.3.6.jar:/home/hadoop/hadoop/share/hadoop/yarn/hadoop-yarn-applications-unmanaged-am-launcher-3.3.6.jar:/home/hadoop/hadoop/share/hadoop/yarn/hadoop-yarn-server-common-3.3.6.jar:/home/hadoop/hadoop/share/hadoop/yarn/hadoop-yarn-applications-mr-mo-core-3.3.6.jar
STARTUP_MSG: build = https://github.com/apache/hadoop.git -r 1be78238728da9266a4f88195058f08fd012bf9c; compiled by 'ubuntu' on 2023-06-18T08:22Z
STARTUP_MSG: java = 1.8.0_392
/*****
2024-02-19 04:47:22,375 INFO namenode.NameNode: registered UNIX signal handlers for [TERM, HUP, INT]
2024-02-19 04:47:22,529 INFO namenode.NameNode: createNameNode [-form]
Usage: hdfs namenode [-backup] |
    [-checkpoint] |
    [-format [-clusterid cid] [-force] [-nonInteractive]] |
    [-upgrade [-clusterid cid] [-renameReserved<k-v pairs>]] |
    [-upgradeOnly [-clusterid cid] [-renameReserved<k-v pairs>]] |
    [-rollback] |
    [-rollingUpgrade <rollback|started>] |
    [-importCheckpoint] |
    [-initializeSharedEdits] |
    [-bootstrapStandby [-force] [-nonInteractive] [-skipSharedEditsCheck]] |
    [-recover [-force]] |
    [-metadataVersion]
2024-02-19 04:47:22,582 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at biel/192.168.0.101
/*****
hadoop@biel:~$
```

En este momento en el que aún no hay ningún namenode ni datanode encendido, ¿qué dirías que está haciendo este formateo?

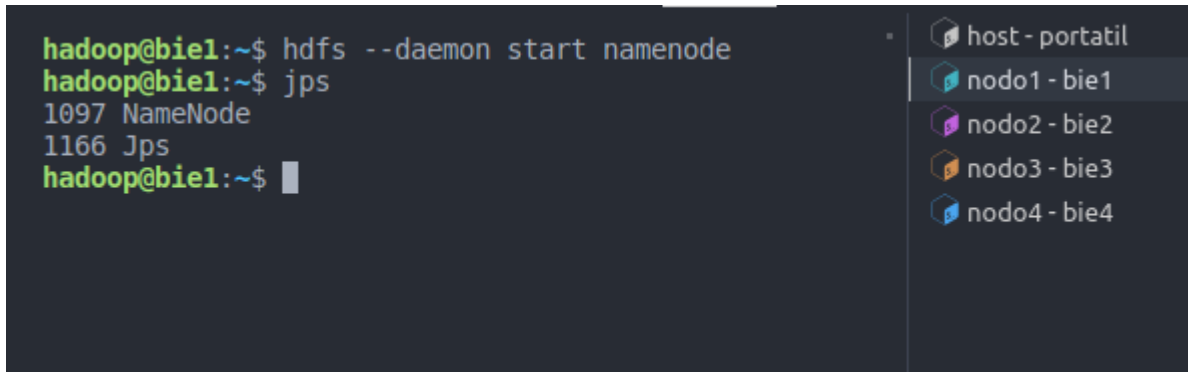
El formateo prepara el directorio con un sistema de archivos tipo HDFS que esta optimizado con una estructura para el uso de hadoop, con lo cual es necesario realizarlo antes de levantar el namenode dado que tiene que leer y escribir en el mismo.

8.- Arranca únicamente el namenode desde su nodo. Captura pantalla en la que se vea el comando que usas y su salida.

nota la primera vez que lo ejecute :

- error : no se había formateado correctamente namenode
- warning : se creo el archivo de logs

```
hdfs --daemon start namenode
```



```
hadoop@bie1:~$ hdfs --daemon start namenode
hadoop@bie1:~$ jps
1097 NameNode
1166 Jps
hadoop@bie1:~$
```

host - portatil  
nodo1 - bie1  
nodo2 - bie2  
nodo3 - bie3  
nodo4 - bie4

9.- Desde el namenode realiza una captura de pantalla de la salida del comando

```
hdfs dfsadmin -report .
```

Interpreta y explica lo que puedas de la salida del comando.

```
hdfs dfsadmin -report

hadoop@biel:~$ hdfs dfsadmin -report
Configured Capacity: 0 (0 B)
Present Capacity: 0 (0 B)
DFS Remaining: 0 (0 B)
DFS Used: 0 (0 B)
DFS Used%: 0.00%
Replicated Blocks:
    Under replicated blocks: 0
    Blocks with corrupt replicas: 0
    Missing blocks: 0
    Missing blocks (with replication factor 1): 0
    Low redundancy blocks with highest priority to recover: 0
    Pending deletion blocks: 0
Erasure Coded Block Groups:
    Low redundancy block groups: 0
    Block groups with corrupt internal blocks: 0
    Missing block groups: 0
    Low redundancy blocks with highest priority to recover: 0
    Pending deletion blocks: 0

-----
```

En este caso la salida tenemos todos los valores a 0 puesto que al no tener levantado ningún datanode no disponemos de espacio ni capacidad

concepto	descripción
Configured Capacity	Capacidad total configurada del sistema de archivos en el clúster HDFS.
Present Capacity	Capacidad actual del sistema de archivos en el clúster HDFS.
DFS Remaining	Espacio restante disponible en el sistema de archivos distribuido HDFS.
DFS Used	Espacio utilizado en el sistema de archivos distribuido HDFS.

concepto	descripción
DFS Used%	Porcentaje del espacio utilizado en el sistema de archivos distribuido HDFS.
Replicated Blocks	Información sobre los bloques replicados en el sistema de archivos distribuido HDFS, incluyendo la cantidad de bloques que están bajo replicados, los bloques con réplicas corruptas, los bloques que faltan, etc.
Erasure Coded Block Groups	Información sobre grupos de bloques codificados por borrado en el sistema de archivos distribuido HDFS, incluyendo la cantidad de grupos de bloques con baja redundancia, los grupos de bloques con bloques internos corruptos, los grupos de bloques que faltan, etc.

```

hadoop@biel1:~$ hdfs dfsadmin -report
Configured Capacity: 0 (0 B)
Present Capacity: 0 (0 B)
DFS Remaining: 0 (0 B)
DFS Used: 0 (0 B)
DFS Used%: 0.00%
Replicated Blocks:
  Under replicated blocks: 0
  Blocks with corrupt replicas: 0
  Missing blocks: 0
  Missing blocks (with replication factor 1): 0
  Low redundancy blocks with highest priority to recover: 0
  Pending deletion blocks: 0
Erasure Coded Block Groups:
  Low redundancy block groups: 0
  Block groups with corrupt internal blocks: 0
  Missing block groups: 0
  Low redundancy blocks with highest priority to recover: 0
  Pending deletion blocks: 0

```

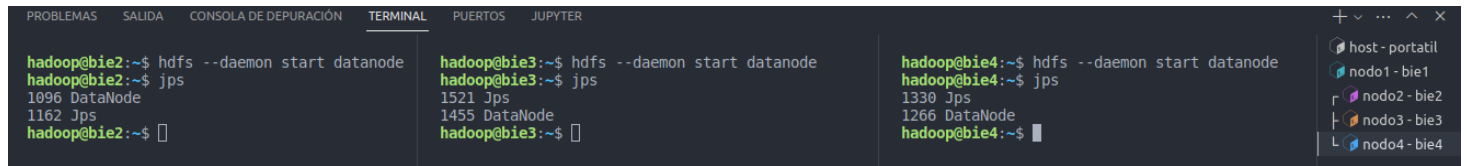
```

-----
hadoop@biel1:~$ █

```

10.- Arranca los datanodes de uno en uno desde cada nodo.

Captura pantalla en la que se vea el comando que usas y su salida en alguno de los nodos.



```
hadoop@bie2:~$ hdfs --daemon start datanode
hadoop@bie2:~$ jps
1096 DataNode
1162 Jps
hadoop@bie2:~$

hadoop@bie3:~$ hdfs --daemon start datanode
hadoop@bie3:~$ jps
1521 Jps
1455 DataNode
hadoop@bie3:~$

hadoop@bie4:~$ hdfs --daemon start datanode
hadoop@bie4:~$ jps
1330 Jps
1266 DataNode
hadoop@bie4:~$
```

11.- Desde el namenode realiza una captura de pantalla de la salida del comando

```
hdfs dfsadmin -report .
```

Si tienes una interfaz gráfica también puedes capturar a web en la ip del namenode:9870 , apartado datanodes .

Interpreta y explica lo que puedas de la salida del comando.

```
hdfs --daemon start datanode  
hdfs dfsadmin -report
```

Nos sale primero la información del nameNode :

Capacidad total configurada del sistema de archivos (293.61 GB)

Capacidad actual del sistema de archivos (278.57 GB)

Espacio restante disponible en el sistema de archivos (278.57 GB)

Espacio utilizado 73728 (72 KB)

Porcentaje del espacio utilizado en el sistema de archivos (0.00%)

(293.61 GB) es la suma de la capacidad de los 3 datanode (3\*97.87 GB)

(278.57 GB) es la suma del espacio restante disponible de los 3 datanode (3\*92.86 GB)

nos sale que tenemos activos 3 datanode

de cada datanode , ip el nombre y el puerto :9866 en cada nodo , numero de bloques

Configured Capacity: 315264577536 (293.61 GB)

Present Capacity: 299108155392 (278.57 GB)

DFS Remaining: 299108081664 (278.57 GB)

DFS Used: 73728 (72 KB)

DFS Used%: 0.00%

Replicated Blocks:

Under replicated blocks: 0

Blocks with corrupt replicas: 0

Missing blocks: 0

Missing blocks (with replication factor 1): 0

Low redundancy blocks with highest priority to recover: 0

Pending deletion blocks: 0

Erasure Coded Block Groups:

Low redundancy block groups: 0

Block groups with corrupt internal blocks: 0

Missing block groups: 0

Low redundancy blocks with highest priority to recover: 0

Pending deletion blocks: 0

-----  
Live datanodes (3):

Name: 192.168.0.102:9866 (bie2)

Hostname: bie2

Decommission Status : Normal

Configured Capacity: 105088192512 (97.87 GB)

DFS Used: 24576 (24 KB)

Non DFS Used: 45056 (44 KB)

DFS Remaining: 99702693888 (92.86 GB)

DFS Used%: 0.00%

DFS Remaining%: 94.88%

Configured Cache Capacity: 0 (0 B)

Cache Used: 0 (0 B)

Cache Remaining: 0 (0 B)

Cache Used%: 100.00%

Cache Remaining%: 0.00%

Xceivers: 0

Last contact: Sat Feb 17 20:22:26 UTC 2024

Last Block Report: Sat Feb 17 20:19:59 UTC 2024

Num of Blocks: 0

Name: 192.168.0.103:9866 (bie3)



Hostname: bie3  
Decommission Status : Normal  
Configured Capacity: 105088192512 (97.87 GB)  
DFS Used: 24576 (24 KB)  
Non DFS Used: 45056 (44 KB)  
DFS Remaining: 99702693888 (92.86 GB)  
DFS Used%: 0.00%  
DFS Remaining%: 94.88%  
Configured Cache Capacity: 0 (0 B)  
Cache Used: 0 (0 B)  
Cache Remaining: 0 (0 B)  
Cache Used%: 100.00%  
Cache Remaining%: 0.00%  
Xceivers: 0  
Last contact: Sat Feb 17 20:22:25 UTC 2024  
Last Block Report: Sat Feb 17 20:19:34 UTC 2024  
Num of Blocks: 0

Name: 192.168.0.104:9866 (bie4)  
Hostname: bie4  
Decommission Status : Normal  
Configured Capacity: 105088192512 (97.87 GB)  
DFS Used: 24576 (24 KB)  
Non DFS Used: 45056 (44 KB)  
DFS Remaining: 99702693888 (92.86 GB)  
DFS Used%: 0.00%  
DFS Remaining%: 94.88%  
Configured Cache Capacity: 0 (0 B)  
Cache Used: 0 (0 B)  
Cache Remaining: 0 (0 B)  
Cache Used%: 100.00%  
Cache Remaining%: 0.00%  
Xceivers: 0  
Last contact: Sat Feb 17 20:22:25 UTC 2024  
Last Block Report: Sat Feb 17 20:20:10 UTC 2024  
Num of Blocks: 0

```
hadoop@bie1:~$ hdfs dfsadmin -report
Configured Capacity: 315264577536 (293.61 GB)
Present Capacity: 299108167680 (278.57 GB)
DFS Remaining: 299108069376 (278.57 GB)
DFS Used: 98304 (96 KB)
DFS Used%: 0.00%
Replicated Blocks:
    Under replicated blocks: 0
    Blocks with corrupt replicas: 0
    Missing blocks: 0
    Missing blocks (with replication factor 1): 0
    Low redundancy blocks with highest priority to recover: 0
    Pending deletion blocks: 0
Erasure Coded Block Groups:
    Low redundancy block groups: 0
    Block groups with corrupt internal blocks: 0
    Missing block groups: 0
    Low redundancy blocks with highest priority to recover: 0
    Pending deletion blocks: 0
```

-----  
Live datanodes (3):

```
Name: 192.168.0.102:9866 (bie2)
Hostname: bie2
Decommission Status : Normal
Configured Capacity: 105088192512 (97.87 GB)
DFS Used: 32768 (32 KB)
Non DFS Used: 40960 (40 KB)
DFS Remaining: 99702689792 (92.86 GB)
DFS Used%: 0.00%
DFS Remaining%: 94.88%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 0
Last contact: Mon Feb 19 05:15:44 UTC 2024
Last Block Report: Mon Feb 19 05:05:08 UTC 2024
Num of Blocks: 0
```

```
Name: 192.168.0.103:9866 (bie3)
Hostname: bie3
Decommission Status : Normal
Configured Capacity: 105088192512 (97.87 GB)
DFS Used: 32768 (32 KB)
Non DFS Used: 40960 (40 KB)
DFS Remaining: 99702689792 (92.86 GB)
DFS Used%: 0.00%
DFS Remaining%: 94.88%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 0
Last contact: Mon Feb 19 05:15:42 UTC 2024
Last Block Report: Mon Feb 19 05:05:13 UTC 2024
Num of Blocks: 0
```

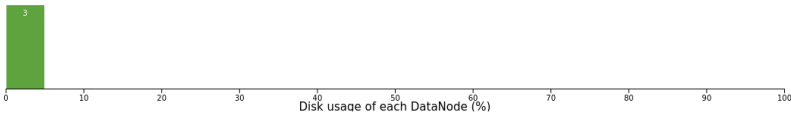
```
Name: 192.168.0.104:9866 (bie4)
Hostname: bie4
Decommission Status : Normal
Configured Capacity: 105088192512 (97.87 GB)
DFS Used: 32768 (32 KB)
Non DFS Used: 40960 (40 KB)
DFS Remaining: 99702689792 (92.86 GB)
DFS Used%: 0.00%
DFS Remaining%: 94.88%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 0
Last contact: Mon Feb 19 05:15:44 UTC 2024
Last Block Report: Mon Feb 19 05:06:04 UTC 2024
Num of Blocks: 0
```

hadoop@bie1:~\$

Datanode Information

In service Down Decommissioning Decommissioned Decommissioned & dead  
Entering Maintenance In Maintenance In Maintenance & dead

Datanode usage histogram



In operation

Datanode State

All

Show

25

entries

Search:

Node	Http Address	Last contact	Last Block Report	Used	Non DFS Used	Capacity	Blocks	Block pool used	Version
In service /default-rack/bie3:9866 (192.168.0.103:9866)	http://bie3:9864	2s	16m	32 KB	40 KB	97.87 GB	0	32 KB (0%)	3.3.6
In service /default-rack/bie2:9866 (192.168.0.102:9866)	http://bie2:9864	1s	16m	32 KB	40 KB	97.87 GB	0	32 KB (0%)	3.3.6
In service /default-rack/bie4:9866 (192.168.0.104:9866)	http://bie4:9864	1s	15m	32 KB	40 KB	97.87 GB	0	32 KB (0%)	3.3.6

Showing 1 to 3 of 3 entries

Previous1Next

Entering Maintenance

No nodes are entering maintenance.

12.- Realiza una captura de la salida del comando `hdfs dfs -df -h` ejecutado en cualquiera de los nodos.

Interpreta con tus palabras el resultado.

```
hdfs dfs -df -h
```

```
hadoop@biel:~$ hdfs dfs -df -h
```

Filesystem	Size	Used	Available	Use%
hdfs://biel:9000	293.6 G	96 K	278.6 G	0%

Filesystem : hdfs://biel:9000 que es la dirección del namenode

Size: 293.6 G tamaño total

Used: 96 K espacio utilizado

Available: 278.6 G espacio disponible

Use%: 0% porcentaje de espacio utilizado

PROBLEMAS   SALIDA   CONSOLA DE DEPURACIÓN   TERMINAL   PUERTOS   JUPYTER

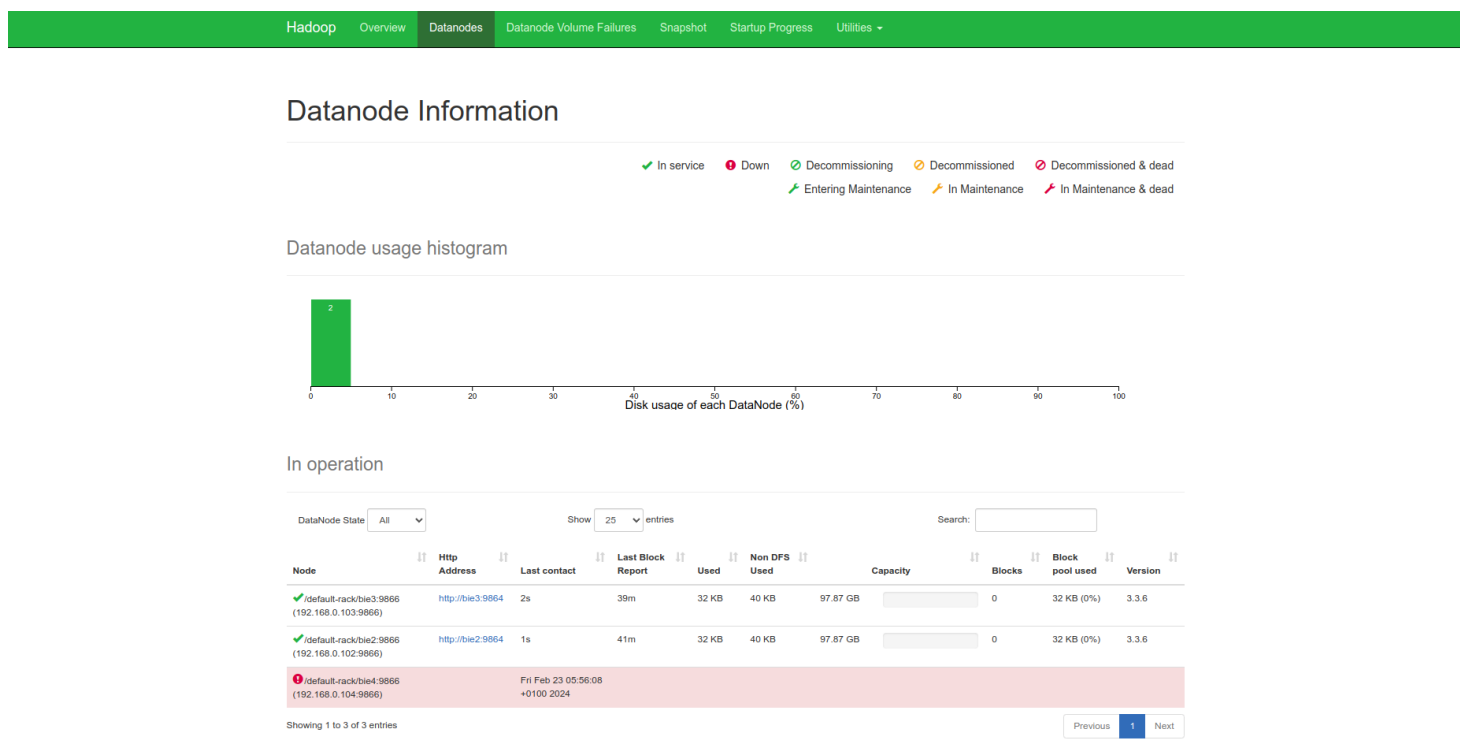
```
hadoop@biel:~$ hdfs dfs -df -h
Filesystem      Size  Used Available Use%
hdfs://biel:9000 293.6 G 96 K   278.6 G   0%
hadoop@biel:~$
```

13.- Apaga un datanode y confirma que se muestra como caído.  
Realiza una captura de pantalla de la salida del comando donde se pueda ver esta información.

tiramos el nodo4

```
hadoop@bie4:~$ hdfs --daemon stop datanode
```

tras 10 min aparece - Dead datanodes (1) - Live datanodes (2)



```
hadoop@bie4:~$ hdfs --daemon stop datanode
hadoop@bie4:~$ hdfs dfsadmin -report
Configured Capacity: 210176385024 (195.74 GB)
Present Capacity: 199405445120 (185.71 GB)
DFS Remaining: 199405379584 (185.71 GB)
DFS Used: 65536 (64 KB)
DFS Used%: 0.00%
Replicated Blocks:
    Under replicated blocks: 0
    Blocks with corrupt replicas: 0
    Missing blocks: 0
    Missing blocks (with replication factor 1): 0
    Low redundancy blocks with highest priority to recover: 0
    Pending deletion blocks: 0
Erasure Coded Block Groups:
    Low redundancy block groups: 0
    Block groups with corrupt internal blocks: 0
    Missing block groups: 0
    Low redundancy blocks with highest priority to recover: 0
    Pending deletion blocks: 0
```

-----  
Live datanodes (2):

[...]

Dead datanodes (1):

```
Name: 192.168.0.104:9866 (bie4)
Hostname: bie4
Decommission Status : Normal
Configured Capacity: 105088192512 (97.87 GB)
DFS Used: 32768 (32 KB)
Non DFS Used: 40960 (40 KB)
DFS Remaining: 99702689792 (92.86 GB)
DFS Used%: 0.00%
DFS Remaining%: 94.88%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 0
Last contact: Fri Feb 23 04:56:07 UTC 2024
```

Last Block Report: Fri Feb 23 04:26:35 UTC 2024

Num of Blocks: 0

Con los parámetros por defecto, Hadoop dará por caído un nodo cuando no tenga conexión durante los últimos 10 minutos y medio.

Justifica con tus palabras qué valor de tiempo especificarías en segundos como tope para dar un nodo por caído en un clúster de 100 máquinas dentro de un mismo CPD.

hay muchos nodos y están en el mismo cpd es mucho mas probable que la mayoría de las veces las caídas se deban a errores del nodo puesto que es poco probable los problemas de conexión, la replicación de los bloques, del nodo caído al ser 100 maquinas conectadas en el mismo cpd no supondrían una carga maxima de recursos, por lo tanto decremental el tiempo por defecto 10 min 600segundos en un /2 con lo cual 5 min 300 segundos.

Justifica con tus palabras qué valor de tiempo especificarías en segundos como tope para dar un nodo por caído en un clúster de miles de máquinas repartidas En CPDs de distintos países.

están repartidos por distintos países y son miles de maquinas con lo cual es mucho mas probable que la mayoría de las veces las caídas se deban a problemas de conectividad y realmente el nodo no tener fallos sino problemas de conexión, por lo tanto incrementaría el tiempo por defecto 10 min 600segundos en un x2 con lo cual 20 min 1200 segundos .

14.- Cambia en el archivo `hdfs-site.xml` las siguientes propiedades:

- Nombre: `dfs.heartbeat.interval` , Valor: 1
- Nombre: `dfs.namenode.heartbeat.recheck-interval` , Valor: 500

Al reiniciar los nodos, el tiempo de detección de nodos caídos será de 11 segundos.

 `/home/hadoop/hadoop/etc/hadoop/workers`

Cuando se ejecutan tareas de procesamiento de datos en un clúster de Hadoop, el framework Hadoop consulta este archivo para determinar qué nodos están disponibles para realizar el trabajo.

```
# default localhost
bie2
bie3
bie4
```





/home/hadoop/hadoop/etc/hadoop/hdfs-site.xml **en** namenode

```

<!--
more configuration property
https://hadoop.apache.org/docs/r2.7.0/hadoop-project-dist/hadoop-hdfs/hdfs-default.

nota:
al modificar este archivo se puede configurar en caliente
hdfs dfsadmin -reconfig namenode biel:9000 start
-->
<configuration>
<!--
dfs.namenode.name.dir
file://${hadoop.tmp.dir}/dfs/name
Determines where on the local filesystem the DFS name node should store the name ta
- - - - -
Ruta donde esta guarda la table de direcciones de namenode
-->
<property>
  <name>dfs.namenode.name.dir</name>
  <value>/home/hadoop/discogrande/namenode</value>
</property>
<!--
dfs.replication
default 3
Default block replication. The actual number of replications can be specified when
- - - - -
Cuando guardas un archivo en Hadoop, el archivo se divide en fragmentos de datos 11
-->
<property>
  <name>dfs.replication</name>
  <value>2</value>
</property>
<!--
dfs.heartbeat.interval
default 3
Determines datanode heartbeat interval in seconds.
- - - - -
tiempo que tarde en mili-segundos en dar un latido
-->
<property>
  <name>dfs.heartbeat.interval</name>
  <value>1</value>
</property>
<!--

```

```

dfs.namenode.heartbeat.recheck-interval
default          300000
300000           This time decides the interval to check for expired datanodes. With t
- - - - -

En el proceso de declarar un nodo como muerto
tras no responder un nodo tras 10 latidos
hadoop establece w periodos de gracia
tiempo en mili-segundos de cada periodo de gracia :
-->
<property>
  <name>dfs.namenode.heartbeat.recheck-interval</name>
  <value>500</value>
</property>
<!--
dfs.hosts
Names a file that contains a list of hosts that are permitted to connect to the nam
- - - - -

En el proceso de declarar un nodo como muerto
tras no responder un nodo tras 10 latidos
hadoop establece w periodos de gracia
tiempo en mili-segundos de cada periodo de gracia :
-->
<property>
  <name>dfs.hosts</name>
  <value>/home/hadoop/hadoop/etc/hadoop/workers</value>
</property>

</configuration>

```

## Actualizar cambios

```

# modificar en caliente hadoop/workers
hdfs dfsadmin -refreshNodes
# volver a leer la configuración de en caliente
hdfs dfsadmin -reconfig namenode biel:9000 start

```

Realiza una captura de pantalla donde se vea un nodo caído con pocos segundos de retraso.

```
# nodo4
echo "Hora actual" && date "+%Y-%m-%d %H:%M:%S" && echo "paro datanode4" && hdfs --da
# nodo1
echo "Hora actual" && date "+%Y-%m-%d %H:%M:%S" && echo "espero 12 segundos" && sleep
```

nodo4

```
hadoop@bie4:~$ echo "Hora actual"
&& date "+%Y-%m-%d %H:%M:%S"
&& echo "paro datanode4"
&& hdfs --daemon stop datanode
Hora actual
# date "+%Y-%m-%d %H:%M:%S"
2024-02-23 06:25:36
# hdfs --daemon stop datanode
paro datanode4
```

nodo1

```
echo "Hora actual"
&& date "+%Y-%m-%d %H:%M:%S"
&& echo "espero 12 segundos"
&& sleep 12
&& hdfs dfsadmin -report
&& tail -n3 hadoop/logs/hadoop-hadoop-namenode-bie1.log

Hora actual
# date "+%Y-%m-%d %H:%M:%S"
2024-02-23 06:25:37
# echo "espero 12 segundos"
espero 12 segundos
#hdfs dfsadmin -report
[...]
Live datanodes (2):
[...]
Dead datanodes (1):

Name: 192.168.0.104:9866 (bie4)
Hostname: bie4
Decommission Status : Normal
[...]
Last contact: Fri Feb 23 06:25:35 UTC 2024
Last Block Report: Fri Feb 23 06:24:29 UTC 2024
Num of Blocks: 0
# tail -n3 hadoop/logs/hadoop-hadoop-namenode-bie1.log
2024-02-23 06:24:29,190 INFO BlockStateChange: BLOCK* processReport 0x9a4329e6ab89b2d
2024-02-23 06:24:29,190 INFO BlockStateChange: BLOCK* processReport 0x9a4329e6ab89b2d
2024-02-23 06:27:29,484 INFO org.apache.hadoop.net.NetworkTopology: Removing a node:
```

In operation

DataNode State

All

Show

25

entries

Search:

Node	Http Address	Last contact	Last Block Report	Used	Non DFS Used	Capacity	Blocks	Block pool used	Version
<div><div></div><div>/default-rack/bie3:9866</div><div>(192.168.0.103:9866)</div></div>	<a href="http://bie3:9864">http://bie3:9864</a>	2s	125m	32 KB	40 KB	97.87 GB	0	32 KB (0%)	3.3.6
<div><div></div><div>/default-rack/bie2:9866</div><div>(192.168.0.102:9866)</div></div>	<a href="http://bie2:9864">http://bie2:9864</a>	1s	10m	32 KB	40 KB	97.87 GB	0	32 KB (0%)	3.3.6
<div><div></div><div>/default-rack/bie4:9866</div><div>(192.168.0.104:9866)</div></div>		Fri Feb 23 07:25:36 +0100 2024							

Showing 1 to 3 of 3 entries

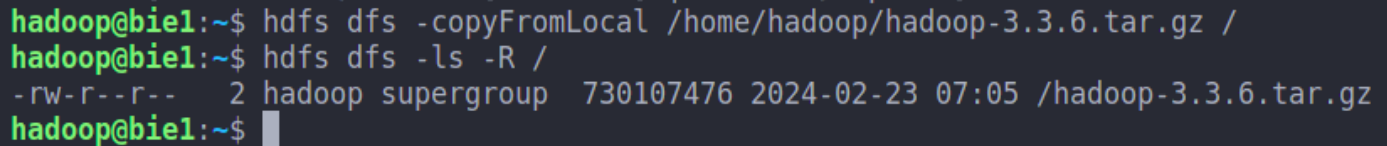
Previous

1

Next

15.- Copia el archivo que has descargado previamente (hadoop-3.3.6.tar.gz) a HDFS con el siguiente comando: `hdfs dfs -copyFromLocal /home/hadoop/hadoop-3.3.6.tar.gz /`.  
Captura la pantalla con el comando `ls` aplicado a `hdfs` para listar los documentos que están en la raíz de HDFS.

```
hdfs dfs -copyFromLocal /home/hadoop/hadoop-3.3.6.tar.gz /  
hdfs dfs -ls -R / # verificamos  
# -rw-r--r--    2 hadoop supergroup  730107476 2024-02-23 07:05 /hadoop-3.3.6.tar.gz
```

A terminal window with a dark background and green text. It shows the execution of two HDFS commands. The first command copies a local file to the root of HDFS. The second command lists the contents of the root directory, showing a single file named 'hadoop-3.3.6.tar.gz' with permissions '-rw-r--r--', owned by 'hadoop' and 'supergroup', with a size of 730107476 bytes, and a timestamp of 2024-02-23 07:05.

```
hadoop@biel:~$ hdfs dfs -copyFromLocal /home/hadoop/hadoop-3.3.6.tar.gz /  
hadoop@biel:~$ hdfs dfs -ls -R /  
-rw-r--r--    2 hadoop supergroup  730107476 2024-02-23 07:05 /hadoop-3.3.6.tar.gz  
hadoop@biel:~$
```

16.- Desde la interfaz web del namenode en el puerto 9870 averigua en qué nodos está replicado el archivo que has subido.

Realiza una captura donde se pueda ver esta información.

información común a todos los bloques

Block ID: 1073741836

Block Pool ID: BP-762343052-192.168.0.101-1708201041376

información por bloque

block	nodo	Stamp	Size
block 0	bie2 bie4	1010	134217728
block 1	bie3 bie2	1010	134217728
block 2	bie2 bie4	1009	134217728
block 3	bie2 bie3	1010	134217728
block 4	bie4 bie2	1011	134217728
block 5	bie3 bie4	1012	59018836

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

Browse Directory

/

Show 25 entries

Permission

Owner

-rwxr-xr-x

hadoop

Showing 1 to 1 of 1 entries

Hadoop, 2023.

File information - hadoop-3.3.6.tar.gz

Download

Head the file (first 32K)

Tail the file (last 32K)

Block information

Block 0

Block ID: 1073741831

Block Pool ID: BP-762343052-192.168.0.101-1708201041376

Generation Stamp: 1007

Size: 134217728

Availability:

- bie2
- bie4

Close

Search:

Name

hadoop-3.3.6.tar.gz

Previous

1

Next

Indica con tus palabras qué debería pasar si uno de los nodos que mantiene una réplica cae

Si uno de los nodos se cae.

esperamos 10 latidos por el valor de `dfs.heartbeat.interval` segundos

esperamos dos periodos de gracia de `dfs.namenode.heartbeat.recheck-interval` milisegundos

si en ese tiempo el datanode no da señal entra en modo dead.

cuando entra en este modo , los bloques de este nodo se replican en otro nodo de manera

temporal , para mantener siempre el numero de replicas de bloques `dfs.replication` una vez

que se recupera el datanode, para mantener el numero de `dfs.replication` se eliminan las copias efectuadas en los otros bloques de manera temporal.