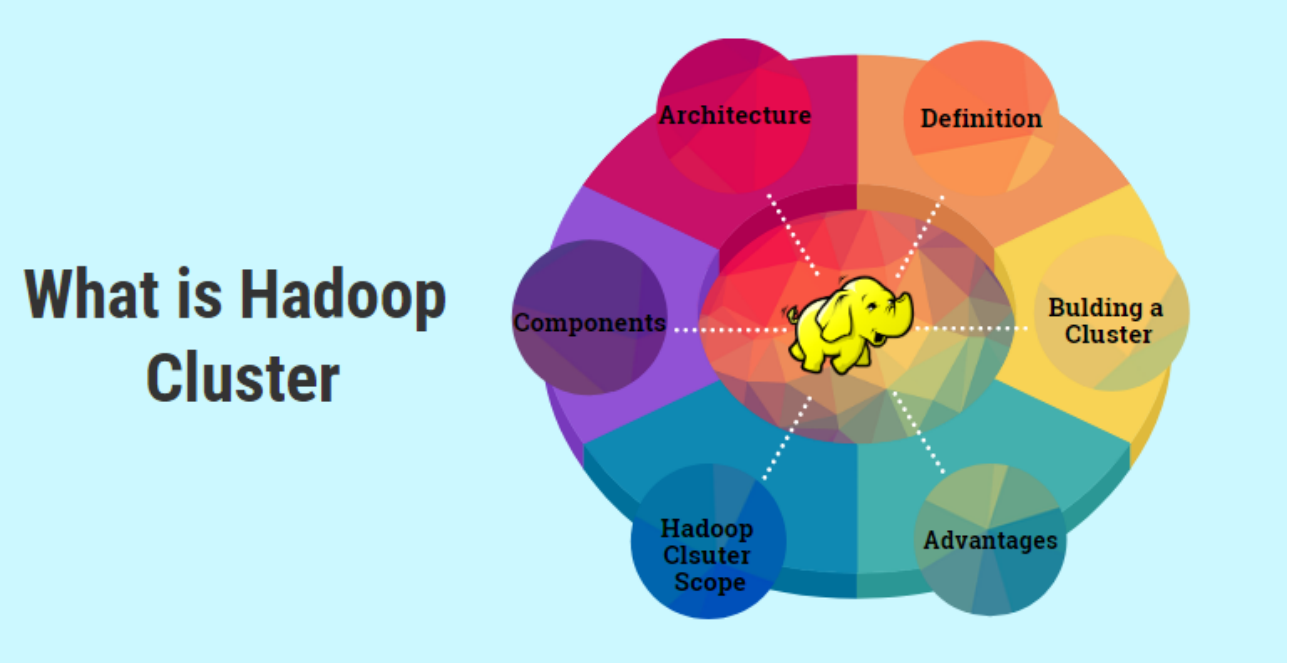
Práctica 2.2

Clúster Hadoop



Alumno:Alvaro Lucio-Villegas de Cea

Indice

Creacion de Clusters

Partimos de la máquina configurada en pseudo cluster de la práctica anterior. Esta primera máquina virtual será el **Master Node**. Cambia el nombre de la máquina a **hadoop-master**

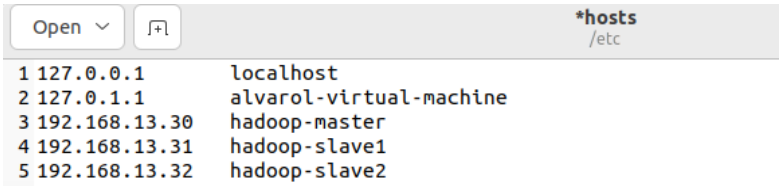
```
hostnamectl set-hostname hadoop-master
```

```
alvarol@alvarol-virtual-machine:~$ hostnamectl set-hostname hadoop-master
==== AUTHENTICATING FOR org.freedesktop.hostname1.set-static-hostname ===
Authentication is required to set the statically configured local hostname, as well as the pretty hostname.
Authenticating as: alvarol,,, (alvarol)
Password:
==== AUTHENTICATION COMPLETE ===
alvarol@alvarol-virtual-machine:~$
```

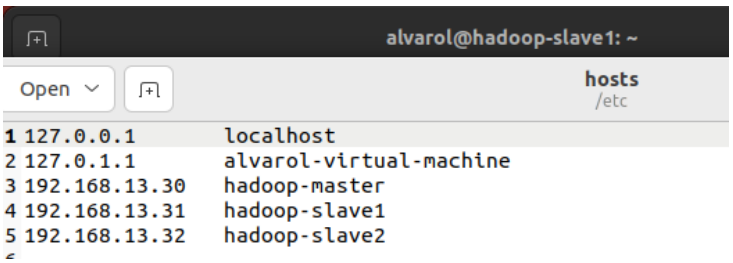
- Clona la máquina virtual tantas veces como nodos vayas a utilizar en el cluster. Al realizar la clonación, recuerda generar nuevas direcciones MAC para evitar conflictos a nivel de red.
- Comprueba las direcciones IP de cada una de las máquinas y edita el fichero **/etc/hosts** de cada máquina para que contenga los nombres de todos los nodos del cluster:

```
192.168.13.30 hadoop-master
192.168.13.31 hadoop-slave1
192.168.13.32 hadoop-slave2
```

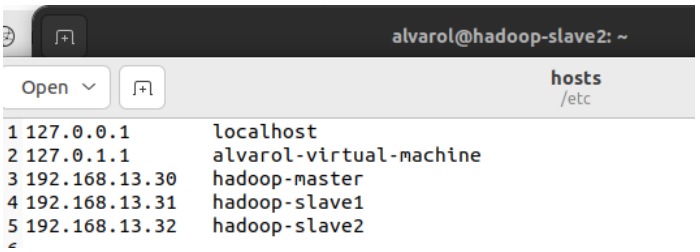
Maestro



Esclavo1



Esclavo2



- En cada uno de los nodos esclavo cambia el nombre de la máquina a **hadoop-slaveX**

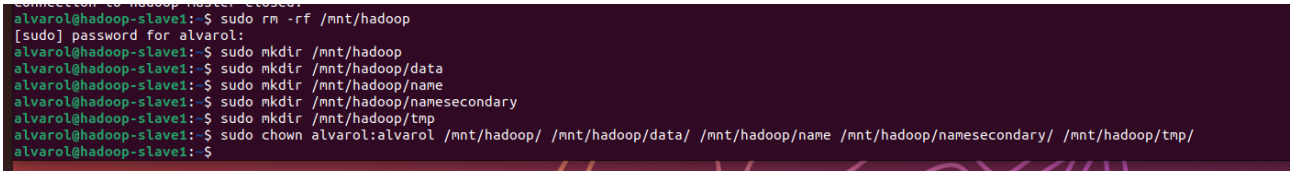
```
hostnamectl set-hostname hadoop-slave1

hostnamectl set-hostname hadoop-slave2
```

- En cada Slave Node, genera una nueva clave SSH, y copiala al Master Node usando **ssh-copy-id**. De esta forma el Master Node podrá ejecutar comandos en los Slave Nodes:

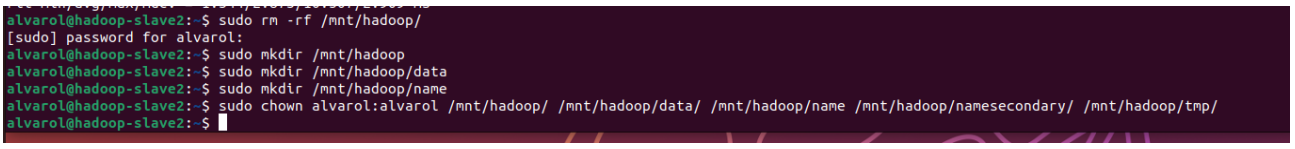
```
ssh-keygen -t rsa

ssh-copy-id ivan@hadoop-master -f
```

•Ahora vamos a borrar y volver a crear las distintas carpetas en cada Slave Node: (+info: https://stackoverflow.com/questions/34967491/datanode-error-namesystem-getdatanode)

Esclavo1



Esclavo2



- En el Master Node, añade la siguiente propiedad al fichero /usr/share/hadoop/etc/hadoop/core-site.xml

```
<property>
        <name>fs.defaultFS</name>
        <value>hdfs://hadoop-master:9000</value>
</property>
```

```
<configuration>
        <property>
                <name>hadoop.tmp.dir</name>
                <value>/mnt/hadoop/tmp</value>
        </property>
        <property>
                <name>fs.defaultFS</name>
                <value>hdfs://hadoop-master:9000</value>
        </property>

</configuration>
```

- Añade la siguiente propiedad a core-site.xml para que Hadoop almacene los archivos temporales en el nuevo directorio.

```
<property>
        <name>hadoop.tmp.dir</name>
        <value>/mnt/hadoop/tmp</value>
 </property>
```

```
<configuration>
        <property>
                <name>hadoop.tmp.dir</name>
                <value>/mnt/hadoop/tmp</value>
        </property>
        <property>
                <name>fs.defaultFS</name>
                <value>hdfs://hadoop-master:9000</value>
        </property>
        <property>
                <name>hadoop.tmp.dir</name>
                <value>/mnt/hadoop/tmp</value>
        </property>

</configuration>
```

- En el Master Node, añade los siguientes parámetros a **/usr/share/hadoop/etc/hadoop/hdfs-site.xml**

```
<property>
        <name>dfs.replication</name>
        <value>2</value>
</property>
<property>
        <name>dfs.datanode.data.dir</name>
        <value>/mnt/hadoop/data</value>
</property>
<property>
        <name>dfs.datanode.name.dir</name>
        <value>/mnt/hadoop/name</value>
</property>
<property>
        <name>dfs.datanode.checkpoint.dir</name>
        <value>/mnt/hadoop/namesecondary</value>
</property>
<property>
      <name>dfs.permissions</name>
      <value>false</value>
</property>
```

Open

hdfs-site.xml
/usr/share/hadoop/etc/hadoop

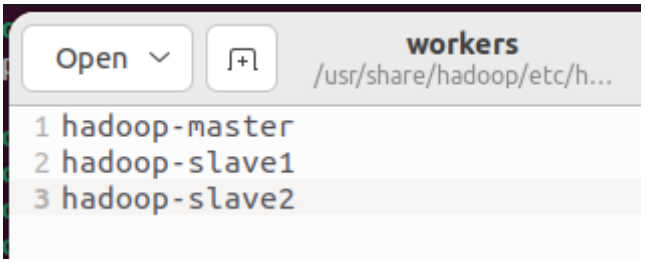
```
10   Unless required by applicable law or agreed to in writing, software
11   distributed under the License is distributed on an "AS IS" BASIS,
12   WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
13   See the License for the specific language governing permissions and
14   limitations under the License. See accompanying LICENSE file.
15 -->
16
17 <!-- Put site-specific property overrides in this file. -->
18
19 <configuration>
20
21         <property>
22                 <name>dfs.replication</name>
23                 <value>2</value>
24         </property>
25         <property>
26                 <name>dfs.datanode.data.dir</name>
27                 <value>/mnt/hadoop/data</value>
28         </property>
29         <property>
30                 <name>dfs.datanode.name.dir</name>
31                 <value>/mnt/hadoop/name</value>
32         </property>
33         <property>
34                 <name>dfs.datanode.checkpoint.dir</name>
35                 <value>/mnt/hadoop/namesecondary</value>
36         </property>
37         <property>
38                 <name>dfs.permissions</name>
39                 <value>false</value>
40         </property>
41
42
43
44 </configuration>
```

- Añade el nombre de los Slave Nodes al fichero

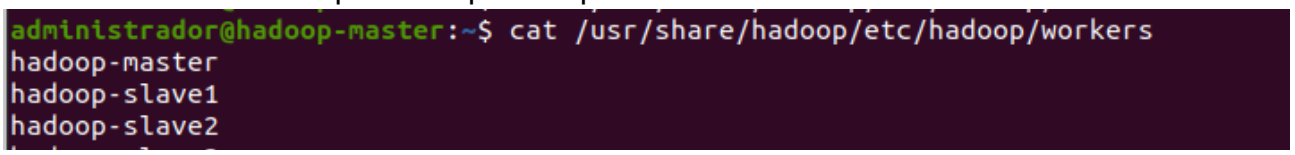
  /usr/share/hadoop/etc/hadoop/workers. Si no se añade hadoop-master, ésta máuina solo actuara en modo gestión.

```
hadoop-slave1
hadoop-slave2
```

Editamos el fichero.

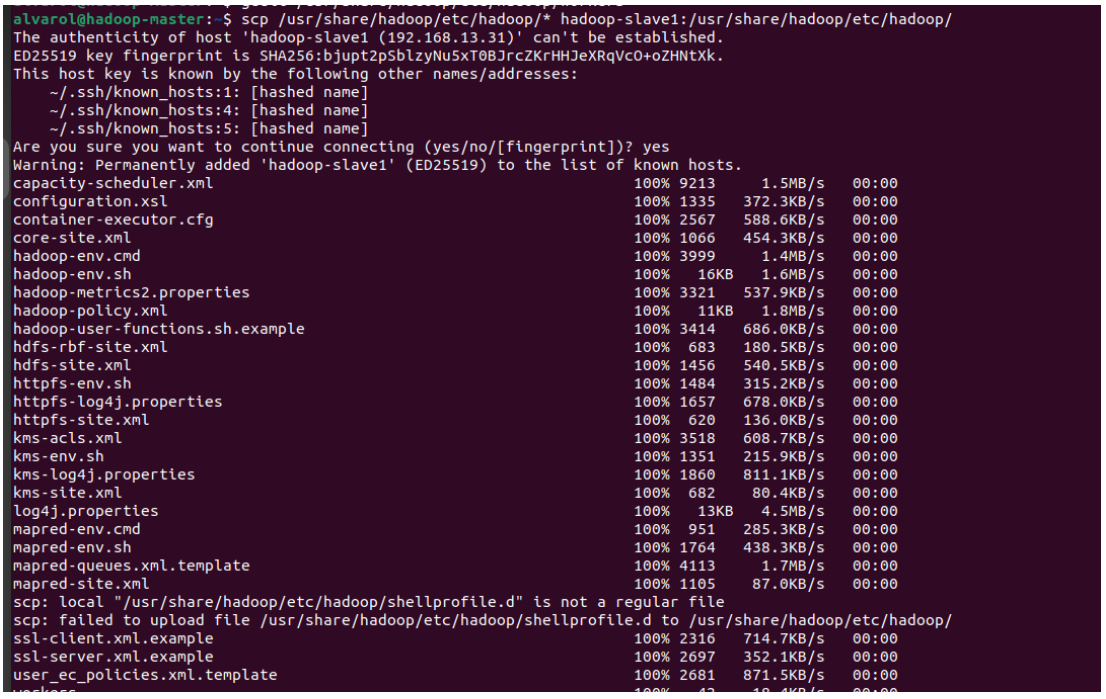


Visualizamos el fichero para comprobar que se han realizado los cambios correctamente.

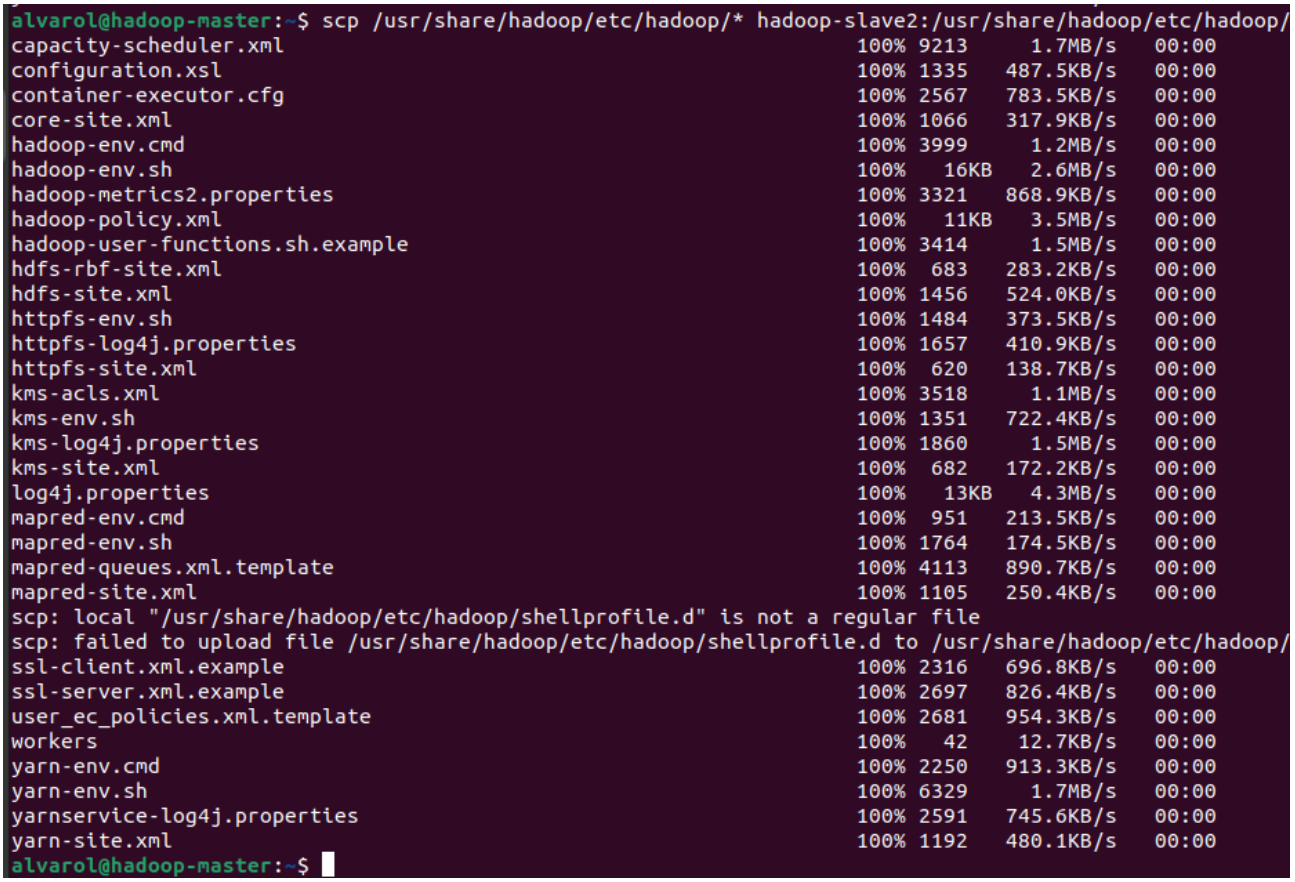


- Desde el Master Node, copia la configuración a cada uno de los nodos usando **scp**:

```
scp /usr/share/hadoop/etc/hadoop/* hadoop-slave1:/usr/share/hadoop/etc/hadoop/
```



```
scp /usr/share/hadoop/etc/hadoop/* hadoop-slave2:/usr/share/hadoop/etc/hadoop/
```



- Da formato al volumen HDFS ejecutando

```
/usr/share/hadoop/bin/hdfs namenode -format
```

- Lanza el servicio DFS ejecutando **/usr/share/hadoop/sbin/start-dfs.sh**

```
/usr/share/hadoop/sbin/start-dfs.sh
```

- Ejecuta **jps** tanto en el **Master Node** como en los **Slave Nodes** para verificar que los servicios adecuados se están ejecutando.

```
alvarol@hadoop-master:~$ /usr/share/hadoop/sbin/start-dfs.sh
Starting namenodes on [hadoop-master]
hadoop-master: Warning: Permanently added 'hadoop-master' (ED25519) to the list of known hosts.
Starting datanodes
Starting secondary namenodes [hadoop-master]
alvarol@hadoop-master:~$ jps
2448 NameNode
2744 SecondaryNameNode
2927 Jps
```

- Para terminar, conectate con el navegador al **http://hadoop-master:9870** para verificar que todos los nodos forman parte del sistema de archivos HDFS

Vemos los procesos de los esclavos:

```
alvarol@hadoop-slave1:~$ jps
2560 Jps
2490 DataNode
alvarol@hadoop-slave1:~$
```

```
alvarol@hadoop-slave2:~$ jps
2535 DataNode
2606 Jps
alvarol@hadoop-slave2:~$
```

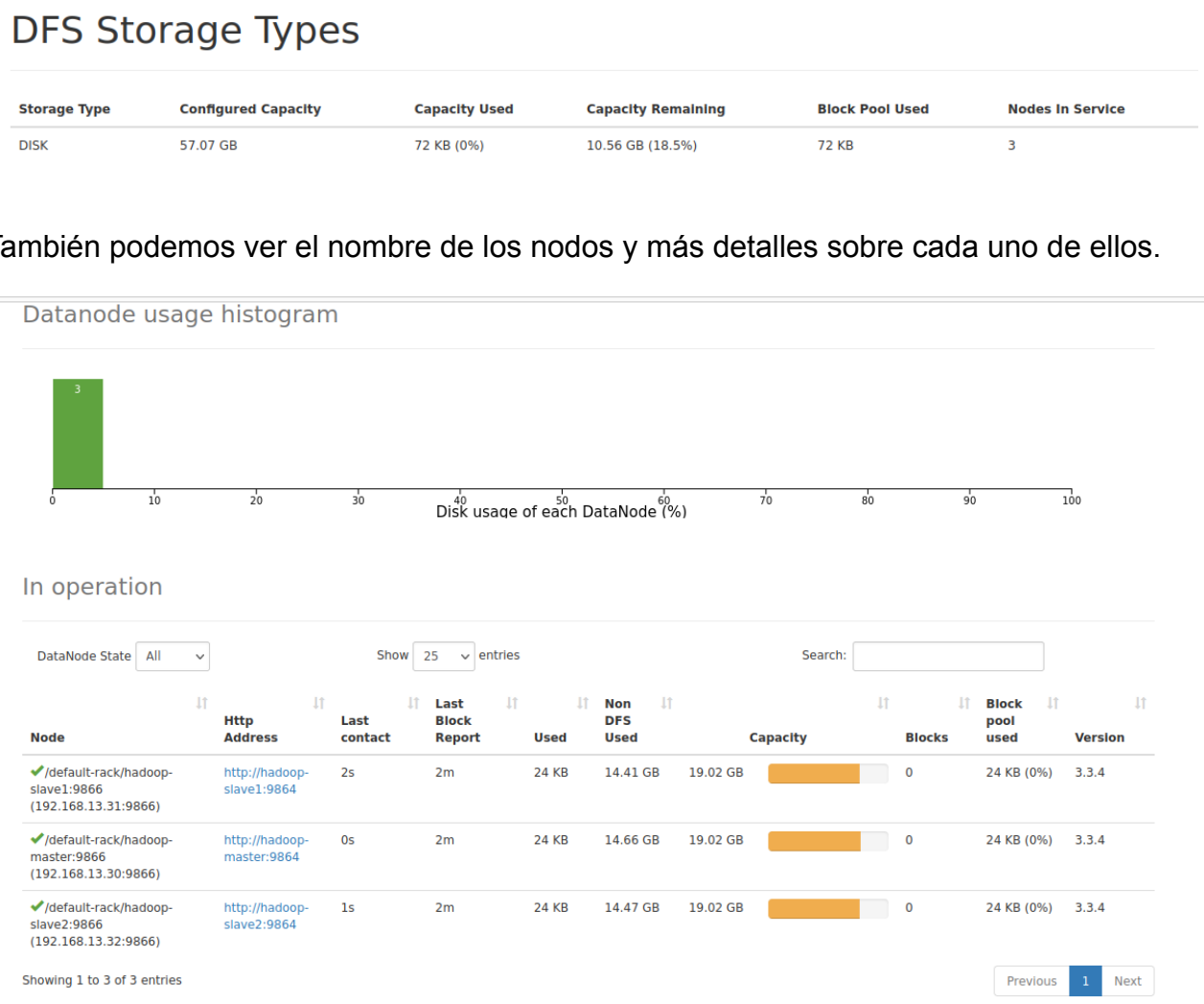
Si nos dirigimos a la web desde el master podemos observar que hay más capacidad de almacenamiento y el número nodos del cluster en este caso 3.

Configured Capacity:	57.07 GB
Configured Remote Capacity:	0 B
DFS Used:	72 KB (0%)
Non DFS Used:	43.54 GB
DFS Remaining:	10.56 GB (18.5%)
Block Pool Used:	72 KB (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)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion (including replicas)	0
Block Deletion Start Time	Wed Mar 15 19:19:40 +0100 2023

Si da algún error por el cual no se ven todos los nodos.Prueba a realizar estos comandos en todos los nodos y en el master tendría que volver hacer el -format anterior.

```
sudo rm -rf /mnt/hadoop/
sudo mkdir /mnt/hadoop/
sudo mkdir /mnt/hadoop/data
sudo mkdir /mnt/hadoop/name
sudo mkdir /mnt/hadoop/namesecondary
sudo mkdir /mnt/hadoop/tmp
sudo chown alvarol:alvarol /mnt/hadoop/ /mnt/hadoop/data/ /mnt/hadoop/name
/mnt/hadoop/namesecondary /mnt/hadoop/tmp
```

Visualizamos en la web el disco que tiene el cluster

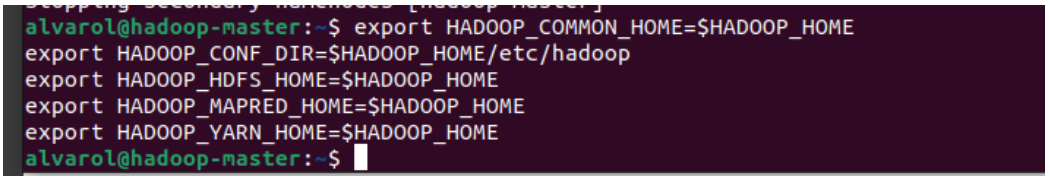


También podemos ver el nombre de los nodos y más detalles sobre cada uno de ellos.

Cluster Hadoop en configuración multi-nodo YARN

- En el Master Node, añade las siguiente variables de entorno a tu **.profile,** e inicia sesión de nuevo para que se carguen correctamente:

```
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export HADOOP_HDFS_HOME=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_YARN_HOME=$HADOOP_HOME
```



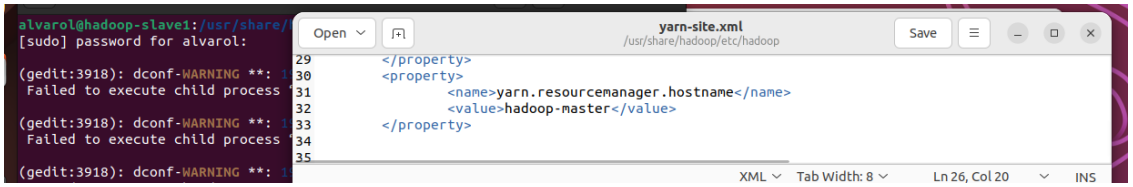
- Arranca hadoop de nuevo:

```
$ /usr/share/hadoop/sbin/start-dfs.sh
```

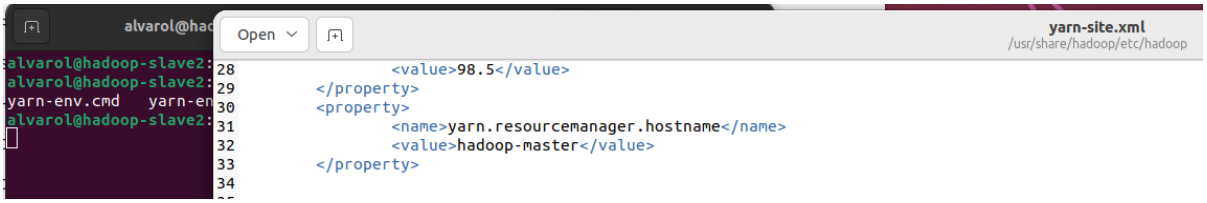
- En cada uno de los Slave Nodes, añade la siguiente propiedad al fichero **yarn-site.xml**

```
<property>
<name>yarn.resourcemanager.hostname</name>
<value>hadoop-master</value>
</property>
```

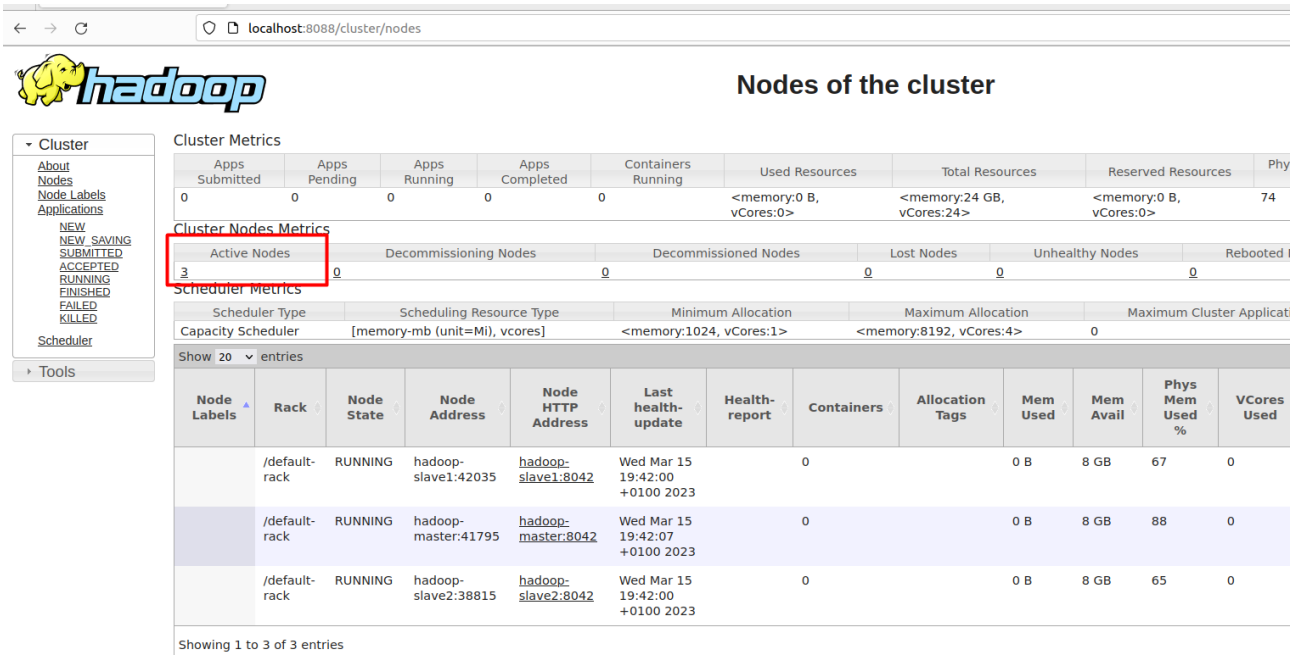
Esclavo1



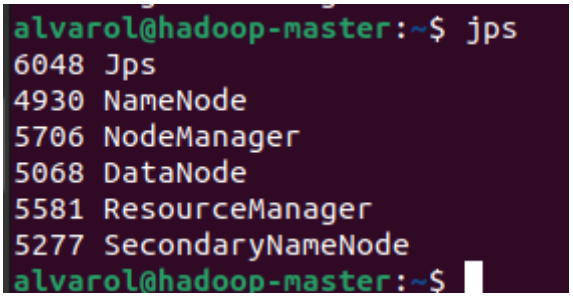
Esclavo2



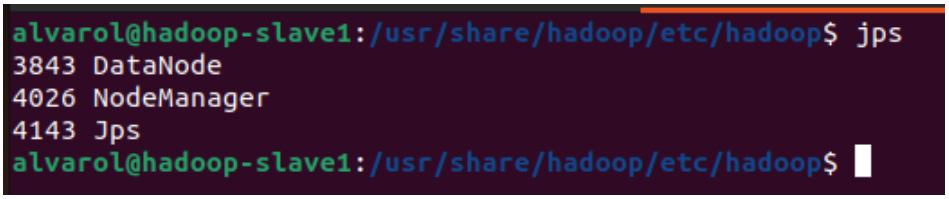
- Arranca YARN ejecutando **/usr/share/hadoop/sbin/start-yarn.sh** y verifica en **http://localhost:8088** que los nodos se han añadido al cluster correctamente



Vemos si desde el master que servicio están activos.



Y en los esclavos.



Creamos una carpeta en Hdfs y subimos El Quijote para poder realizar la prueba.

Realizamos la prueba del mapreduce

```
alvarol@hadoop-master:~$ /usr/share/hadoop/bin/hadoop jar    /usr/share/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.4.jar    grep /input /grep_example 'Sancho'
```

Observamos que se está realizando el proceso desde la web