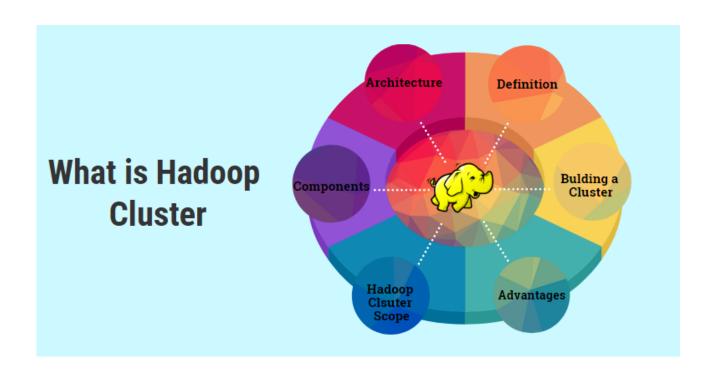
Práctica 2.2 Clúster Hadoop



Indice

Creacion de Clusters	2
Cluster Hadoop en configuración multi-nodo YARN	10

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**

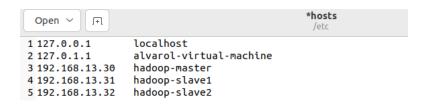
hostnamectl set-hostname hadoop-master

```
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:-$
```

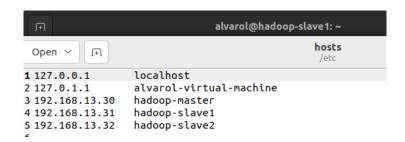
- Clona la máquina virtual tantas veces como nodos vayas a utilizar en el cluster. Al realizar la clonación, recuerda generar nuevas direcciones MAC para evitar conflictos a nivel de red.
- Comprueba las direcciones IP de cada una de las máquinas y edita el fichero
 /etc/hosts de cada máquina para que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

```
alvarolghadoop-slavei: $ sudo rm -rf /mnt/hadoop
[sudo] password for alvarol:
    alvarolghadoop-slavei: $ sudo mkdir /mnt/hadoop
    alvarolghadoop-slavei: $ sudo mkdir /mnt/hadoop/data
    alvarolghadoop-slavei: $ sudo mkdir /mnt/hadoop/name
    alvarolghadoop-slavei: $ sudo mkdir /mnt/hadoop/name
    alvarolghadoop-slavei: $ sudo mkdir /mnt/hadoop/namesecondary
    alvarolghadoop-slavei: $ sudo mkdir /mnt/hadoop/tmp
    alvarolghadoop-slavei: $ sudo mkdir /mnt/hadoop/tmp
    alvarolghadoop-slavei: $ sudo chown alvarol:alvarol /mnt/hadoop/ /mnt/hadoop/data/ /mnt/hadoop/name /mnt/hadoop/namesecondary/ /mnt/hadoop/tmp/
    alvarolghadoop-slavei: $
```

Esclavo2

```
alvarol@hadoop-slave2: $ sudo rm -rf /mnt/hadoop/
[sudo] password for alvarol:
alvarol@hadoop-slave2: $ sudo mkdir /mnt/hadoop
alvarol@hadoop-slave2: $ sudo mkdir /mnt/hadoop/data
alvarol@hadoop-slave2: $ sudo mkdir /mnt/hadoop/data
alvarol@hadoop-slave2: $ sudo mkdir /mnt/hadoop/name
alvarol@hadoop-slave2: $ sudo chown alvarol:alvarol /mnt/hadoop/ /mnt/hadoop/data/ /mnt/hadoop/name /mnt/hadoop/namesecondary/ /mnt/hadoop/tmp/
alvarol@hadoop-slave2: $
```

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

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

· En el Master Node, añade los siguientes parámetros a

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

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

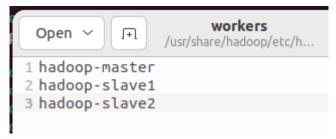
```
hdfs-site.xml
                                                                                          /usr/share/hadoop/etc/hadoop
    Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS,
    WITHOUT WARRANTIES OR 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.
17 <!-- Put site-specific property overrides in this file. -->
19 <configuration>
           operty>
                    <name>dfs.replication</name>
                   <value>2</value>
           </property>
25
    26
                   <name>dfs.datanode.data.dir</name>
                   <value>/mnt/hadoop/data</value>
           </property>
29
           operty>
                   <name>dfs.datanode.name.dir</name>
30
31
                   <value>/mnt/hadoop/name</value>
           </property>
33
                   <name>dfs.datanode.checkpoint.dir</name>
34
35
                   <value>/mnt/hadoop/namesecondary</value>
           </property>
           cproperty>
                   <name>dfs.permissions</name>
39
                    <value>false</value>
40
           </property>
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.

```
administrador@hadoop-master:~$ cat /usr/share/hadoop/etc/hadoop/workers
hadoop-master
hadoop-slave1
hadoop-slave2
```

• 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/

```
core-site.xml
hadoop-env.sh
hadoop-metrics2.properties
hadoop-policy.xml
hadoop-user-functions.sh.example
hdfs-rbf-site.xml
hdfs-site.xml
                                                                                   100% 3999
100% 16
                                                                                   100% 16KB
100% 3321
100% 11KB
100% 3414
                                                                                   100% 683
100% 1456
                                                                                                 180.5KB/s
540.5KB/s
 httpfs-env.sh
httpfs-log4j.properties
                                                                                   100% 1484
100% 1657
                                                                                                 315.2KB/s
678.0KB/s
 httpfs-site.xml
kms-acls.xml
kms-env.sh
kms-log4j.properties
kms-site.xml
                                                                                   100% 620
100% 3518
100% 1351
                                                                                                 811.1KB/s
 log4j.properties
mapred-env.cmd
                                                                                                                00:00
00:00
 user_ec_policies.xml.template
```

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

```
alvarol@hadoop-master:~$ scp /usr/share/hadoop/etc/hadoop/* hadoop-slave2:/usr/share/hadoop/etc/hadoop
capacity-scheduler.xml
                                                                        100% 9213
                                                                                       1.7MB/s
                                                                                                 00:00
                                                                                    487.5KB/s
configuration.xsl
                                                                        100% 1335
                                                                                                 00:00
container-executor.cfg
                                                                        100% 2567
                                                                                     783.5KB/s
                                                                                                 00:00
core-site.xml
                                                                        100% 1066
                                                                                     317.9KB/s
                                                                                                 00:00
                                                                                       1.2MB/s
hadoop-env.cmd
                                                                        100% 3999
                                                                                                 00:00
hadoop-env.sh
                                                                        100%
                                                                               16KB
                                                                                      2.6MB/s
                                                                                                 00:00
hadoop-metrics2.properties
                                                                        100%
                                                                                     868.9KB/s
                                                                             3321
                                                                                                 00:00
hadoop-policy.xml
                                                                        100%
                                                                                      3.5MB/s
                                                                               11KB
                                                                                                 00:00
                                                                        100% 3414
hadoop-user-functions.sh.example
                                                                                       1.5MB/s
                                                                                                 00:00
hdfs-rbf-site.xml
                                                                        100%
                                                                             683
                                                                                     283.2KB/s
                                                                                                 00:00
                                                                                    524.0KB/s
hdfs-site.xml
                                                                        100% 1456
                                                                                                 00:00
httpfs-env.sh
                                                                                    373.5KB/s
                                                                        100% 1484
                                                                                                 00:00
httpfs-log4j.properties
                                                                        100% 1657
                                                                                     410.9KB/s
                                                                                                 00:00
httpfs-site.xml
                                                                        100%
                                                                                     138.7KB/s
                                                                                                 00:00
                                                                              620
kms-acls.xml
                                                                        100% 3518
                                                                                      1.1MB/s
                                                                                                 00:00
kms-env.sh
                                                                        100% 1351
                                                                                     722.4KB/s
                                                                                                 00:00
                                                                                      1.5MB/s
kms-log4j.properties
                                                                        100%
                                                                             1860
                                                                                                 00:00
kms-site.xml
                                                                        100%
                                                                                     172.2KB/s
                                                                                                 00:00
                                                                              682
log4j.properties
                                                                        100%
                                                                               13KB
                                                                                      4.3MB/s
                                                                                                 00:00
mapred-env.cmd
                                                                        100%
                                                                              951
                                                                                    213.5KB/s
                                                                                                 00:00
mapred-env.sh
                                                                        100% 1764
                                                                                     174.5KB/s
                                                                                                 00:00
mapred-queues.xml.template
                                                                        100% 4113
                                                                                     890.7KB/s
                                                                                                 00:00
mapred-site.xml

100% 1105
scp: local "/usr/share/hadoop/etc/hadoop/shellprofile.d" is not a regular file
                                                                                     250.4KB/s
                                                                                                 00:00
scp: failed to upload file /usr/share/hadoop/etc/hadoop/shellprofile.d to /usr/share/hadoop/etc/hadoop/
ssl-client.xml.example
                                                                        100% 2316
                                                                                     696.8KB/s
                                                                                                 00:00
                                                                                     826.4KB/s
ssl-server.xml.example
                                                                        100% 2697
                                                                                                 00:00
user_ec_policies.xml.template
                                                                        100% 2681
                                                                                     954.3KB/s
                                                                                                 00:00
                                                                        100%
                                                                               42
                                                                                     12.7KB/s
                                                                                                 00:00
workers
yarn-env.cmd
                                                                        100% 2250
                                                                                     913.3KB/s
                                                                                                 00:00
yarn-env.sh
                                                                        100% 6329
                                                                                      1.7MB/s
                                                                                                 00:00
                                                                                     745.6KB/s
yarnservice-log4j.properties
                                                                        100% 2591
                                                                                                 00:00
yarn-site.xml
                                                                        100% 1192
                                                                                     480.1KB/s
                                                                                                 00:00
alvarol@hadoop-master:~$
```

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.

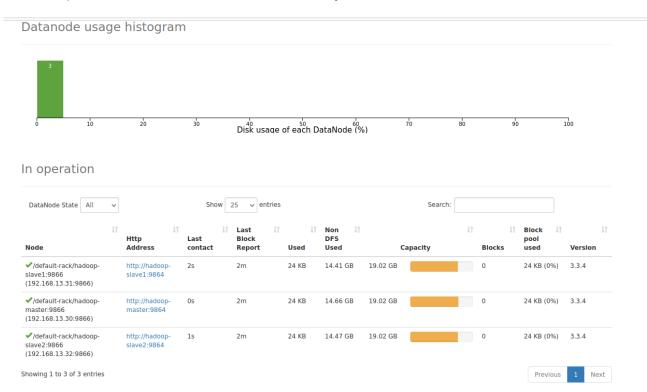
sudo rm -rf /mnt/hadoop/ sudo mkdir /mnt/hadoop/data sudo mkdir /mnt/hadoop/name sudo mkdir /mnt/hadoop/namesecondary sudo mkdir /mnt/hadoop/tmp sudo chown alvarol:alvarol /mnt/hadoop/ /mnt/hadoop/data/ /mnt/hadoop/name /mnt/hadoop/namesecondary /mnt/hadoop/tmp

Visualizamos en la web el disco que tiene el cluster

DFS Storage Types

Storage Type	Configured Capacity	Capacity Used	Capacity Remaining	Block Pool Used	Nodes In Service
DISK	57.07 GB	72 KB (0%)	10.56 GB (18.5%)	72 KB	3

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
```

```
alvarol@hadoop-master:~$ 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=$HADOOP_HOME
alvarol@hadoop-master:~$
```

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

Esclavo1

```
alvarol@hadoop-slave1:/usr/share/
[sudo] password for alvarol:

(gedit:3918): dconf-MARNING **: 1
Failed to execute child process (gedit:3918): dconf-WARNING **: 1
Failed to execute child process (gedit:3918): dconf-WARNING **: 1

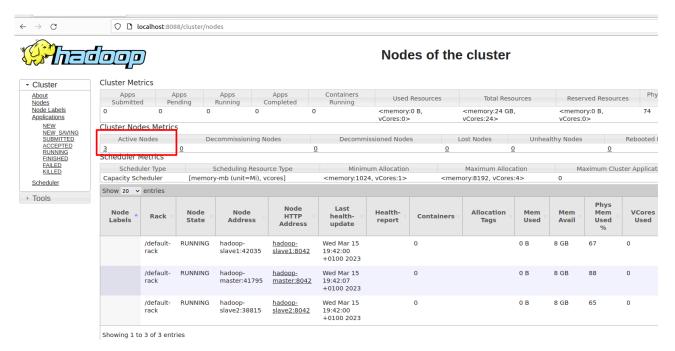
(gedit:3918): dconf-WARNING **: 1

(gedit:3918): dconf-WARNING **: 1

XML > Tab Width: 8 > Ln 26, Col 20 > INS
```

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.

```
alvarol@hadoop-master:~$ jps
6048 Jps
4930 NameNode
5706 NodeManager
5068 DataNode
5581 ResourceManager
5277 SecondaryNameNode
alvarol@hadoop-master:~$
```

Y en los esclavos.

```
alvarol@hadoop-slave1:/usr/share/hadoop/etc/hadoop$ jps
3843 DataNode
4026 NodeManager
4143 Jps
alvarol@hadoop-slave1:/usr/share/hadoop/etc/hadoop$
```

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

```
alvarol@hadoop-master:~$ /usr/share/hadoop/bin/hdfs dfs -mkdir /input
alvarol@hadoop-master:~$ /usr/share/hadoop/bin/hdfs dfs -put /home/alvarol/ElQuijote.txt /input
alvarol@hadoop-master:~$ /usr/share/hadoop/bin/hdfs dfs -ls /input
Found 1 items
-rw-r--r- 2 alvarol supergroup 2198936 2023-03-15 20:10 /input/ElQuijote.txt
alvarol@hadoop-master:~$
```

Realizamos la prueba del mapreduce



alvarolghadoop-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'

