

Actividad de refuerzo n°.7. HDFS

1. Arranca HDFS.

Este es el error que comente en clase. Lo conseguí solventar iniciando sesión en el docker desde cmd con: `docker login -u 'miusuario'`.

```
C:\> Seleccionar Símbolo del sistema

Microsoft Windows [Versión 10.0.26100.3775]
(c) Microsoft Corporation. Todos los derechos reservados.

C:\Users\alvar>cd Downloads

C:\Users\alvar\Downloads>cd docker-hadoop-master_comandos

C:\Users\alvar\Downloads\docker-hadoop-master_comandos>cd docker-hadoop-master_comandos

C:\Users\alvar\Downloads\docker-hadoop-master_comandos\docker-hadoop-master_comandos>cd docker-hadoop-master_comando
scompose up
time "2025-05-05T18:38:52+02:00" level=warning msg='oci://Users\alvar\Downloads\docker-hadoop-master_comandos\docke
r-hadoop-master_comandos\docker-hadoop-master_comandos\docker-compose.yml: the attribute `version` is obsolete, it wil
l be ignored, please remove it to avoid potential confusion'

X Running S/S
X historyserver Error failed to resolve reference "docker.io/bde2020/hadoop-nodemanager:2.0..."
X datanode Error failed to resolve reference "docker.io/bde2020/hadoop-nodemanager:2.0..."
X namenode Error failed to resolve reference "docker.io/bde2020/hadoop-nodemanager:2.0..."
X namenode Error failed to resolve reference "docker.io/bde2020/hadoop-nodemanager:2.0..."
error response from daemon: failed to resolve reference "docker.io/bde2020/hadoop-nodemanager:2.0.0-hadoop3.2.1-javab
ailed to authorize: failed to fetch auth token: unexpected status from GET request to https://auth.docker.io/token?se
repository=X3Abde2020F2Hadoop-nodemanagerX3APull&service=registry.docker.io: 401 Unauthorized

C:\Users\alvar\Downloads\docker-hadoop-master_comandos\docker-hadoop-master_comandos\docker-hadoop-master_comandos>
```

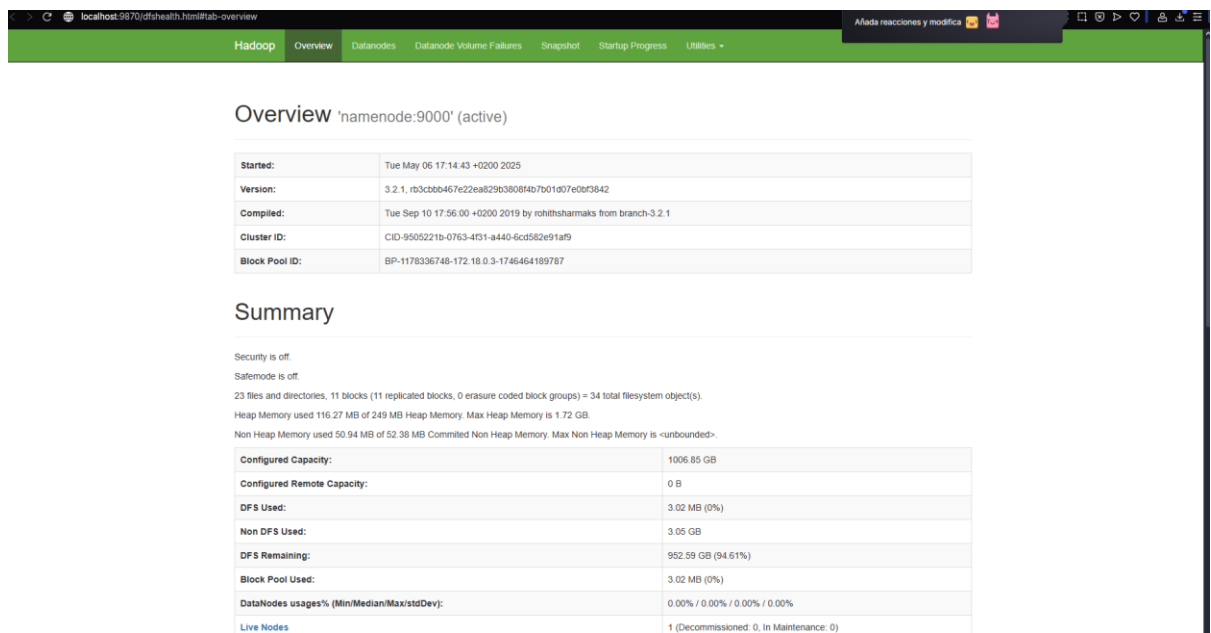
```

)
Symbolo del sistema - docker-compose up
)
resourceManager | 2025-05-06 15:15:20,870 INFO resourceManager.ResourceTrackerService: NodeManager from node 9be7cad9311d(cmPort: 43383 httpPort: 8042) registered with capability: <memory:16384, vCores:8>, assigned nodeId 9be7cad9311d:43383
resourceManager | 2025-05-06 15:15:20,874 INFO rmnode.RMNodeImpl: 9be7cad9311d:43383 Node Transitioned from NEW to RUNNING
nodemanager | 2025-05-06 15:15:20,883 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens, got key with id -269243600
nodemanager | 2025-05-06 15:15:20,884 INFO security.NMTokenSecretManagerInNM: Rolling master-key for container-tokens, got key with id -413914676
nodemanager | 2025-05-06 15:15:20,884 INFO nodemanager.NodeStatusUpdaterImpl: Registered with ResourceManager as 9be7cad9311d:43383 with total resource of <memory:16384, vCores:8>
resourceManager | 2025-05-06 15:15:20,895 INFO capacity.CapacityScheduler: Added node 9be7cad9311d:43383 clusterResource: <memory:16384, vCores:8>
resourceManager | 2025-05-06 15:15:20,922 INFO resourceManager.RMActiveServiceContext: Scheduler recovery is done. Start allocating new containers.
datanode | 2025-05-06 15:15:21,327 INFO impl.FsDatasetAsyncDiskService: Scheduling blk_1073741826_1002 replica finalizedReplica, blk_1073741826_1002, FINALIZED
datanode | getNumBytes() = 2
datanode | getBytesOnDisk() = 2
datanode | getVisibleLength()= 2
datanode | getVolume() = /hadoop/dfs/data
datanode | getBlockURI() = file:/hadoop/dfs/data/current/BP-1178336748-172.18.0.3-1746464189787/current/finalized/subdir0/subdir0/blk_1073741826 for deletion
datanode | 2025-05-06 15:15:21,328 INFO impl.FsDatasetAsyncDiskService: Deleted BP-1178336748-172.18.0.3-1746464189787 blk_1073741826_1002 URI file:/hadoop/dfs/data/current/BP-1178336748-172.18.0.3-1746464189787/current/finalized/subdir0/subdir0/blk_1073741826
View in Docker Desktop View Config Enable Watch

```

Como se puede observar tenemos el docker-compose up activo

2. Comprueba que todos los servicios de HDFS están funcionando.



Overview 'namenode:9000' (active)

Started:	Tue May 06 17:14:43 +0200 2025
Version:	3.2.1, r03c2bb467e22ea829b38084b7b01d07e0bf3842
Compiled:	Tue Sep 10 17:56:00 +0200 2019 by rohitsharmaks from branch-3.2.1
Cluster ID:	CID-9505221b-0763-4f31-a440-6cd582e91a19
Block Pool ID:	BP-117833674b-172.18.0.3-1746464189787

Summary

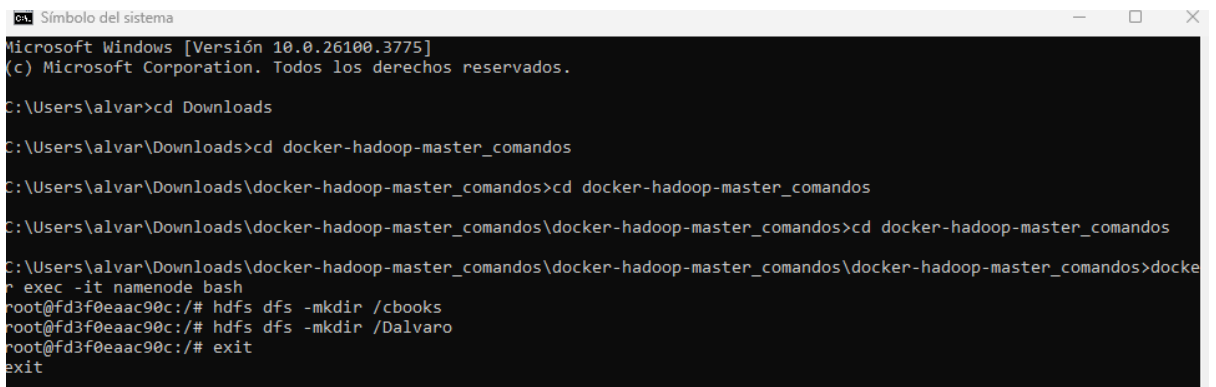
Security is off.
Safemode is off.

23 files and directories, 11 blocks (11 replicated blocks, 0 erasure coded block groups) = 34 total filesystem object(s).
Heap Memory used 116.27 MB of 249 MB Heap Memory. Max Heap Memory is 1.72 GB.
Non Heap Memory used 50.94 MB of 52.38 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	1006.85 GB
Configured Remote Capacity:	0 B
DFS Used:	3.02 MB (0%)
Non DFS Used:	3.05 GB
DFS Remaining:	952.99 GB (94.61%)
Block Pool Used:	3.02 MB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)

Servicios de HDFS funcionando.

3. Crea el directorio books en el directorio raíz / de HDFS.
4. Crea un directorio con tu nombre en el directorio del usuario ubuntu de HDFS.



```
Microsoft Windows [Versión 10.0.26100.3775]
(c) Microsoft Corporation. Todos los derechos reservados.

C:\Users\alvar>cd Downloads

C:\Users\alvar\Downloads>cd docker-hadoop-master_comandos

C:\Users\alvar\Downloads\docker-hadoop-master_comandos>cd docker-hadoop-master_comandos

C:\Users\alvar\Downloads\docker-hadoop-master_comandos\docker-hadoop-master_comandos>cd docker-hadoop-master_comandos

C:\Users\alvar\Downloads\docker-hadoop-master_comandos\docker-hadoop-master_comandos\docker-hadoop-master_comandos>docke
r exec -it namenode bash
root@fd3f0eaac90c:/# hdfs dfs -mkdir /cbooks
root@fd3f0eaac90c:/# hdfs dfs -mkdir /Dalvaro
root@fd3f0eaac90c:/# exit
exit
```

Creamos el directorio books y un directorio con mi nombre en la raíz.

5. Pasa los ficheros del directorio `./home/ubuntu/bigdata/examples/books` de la máquina virtual al directorio `/books` que has creado en HDFS.

```
C:\Users\alvar\Downloads\docker-hadoop-master_comandos\docker-hadoop-master_comandos\docker-hadoop-master_comandos>docker cp frankenstein.txt namenode:/tmp
Successfully copied 443kB to namenode:/tmp

C:\Users\alvar\Downloads\docker-hadoop-master_comandos\docker-hadoop-master_comandos\docker-hadoop-master_comandos>

root@fd3f0eaac90c:/# hdfs dfs -put /tmp/frankenstein.txt /books
2025-05-06 15:53:43,162 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
root@fd3f0eaac90c:/#
```

Ficheros en la carpeta `/books` creada. Hacemos lo mismo para Quijote.

Comprobación en interfaz gráfica:

Browse Directory

/books

Go!

Show

25

entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	root	supergroup	2.04 MB	May 05 19:47	3	128 MB	Quijote.txt	
<input type="checkbox"/>	-rw-r--r--	root	supergroup	430.7 KB	May 05 19:47	3	128 MB	frankenstein.txt	

Showing 1 to 2 of 2 entries

Previous

1

Next

Hadoop, 2019.

6. De la máquina virtual, descomprime el fichero:
`'/home/ubuntu/bigdata/examples/hdfs/fichero_result.zip'` con el comando `'unzip'` y el fichero generado `'fichero_result.txt'` pásalo al directorio que creaste con tu

nombre en HDFS (no debe quedar el 'fichero .txt' en la máquina virtual y sin borrarlo obviamente).

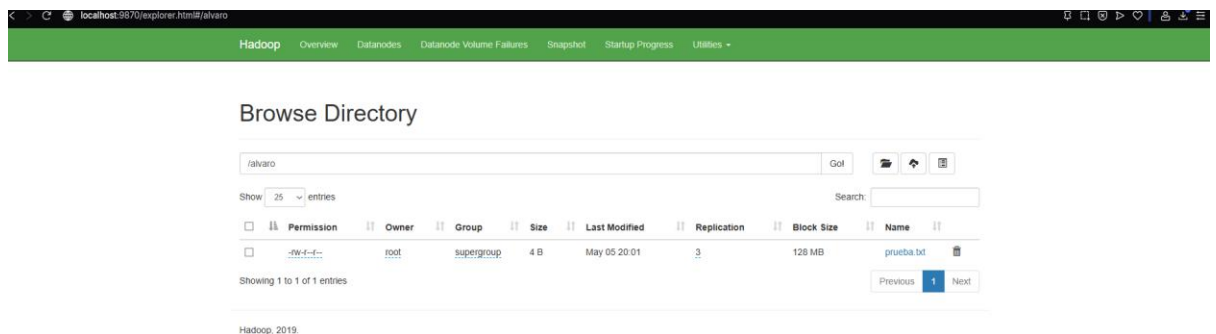
Nosotros en esta parte ya que tenemos el docker descargado por parte del profesor y facilitado para hacer el proceso vía comandos, esta parte del unzip no nos hace falta. Simplemente navegamos a la carpeta donde tenemos el archivo y realizamos el proceso.

```
:~\Downloads\docker-hadoop-master_comandos\docker-hadoop-master_comandos\docker-hadoop-master_comandos\file
o_result>docker cp prueba.txt namenode:/tmp
Successfully copied 2.05kB to namenode:/tmp

:~\Downloads\docker-hadoop-master_comandos\docker-hadoop-master_comandos\docker-hadoop-master_comandos\file
o_result>hdfs dfs -put /tmp/prueba.txt /alvaro
hdfs" no se reconoce como un comando interno o externo,
programa o archivo por lotes ejecutable.

:~\Downloads\docker-hadoop-master_comandos\docker-hadoop-master_comandos\docker-hadoop-master_comandos\file
o_result>docker exec -it namenode bash
root@fd3f0eaac90c:/# hdfs dfs -put /tmp/prueba.txt /alvaro
2025-05-05 18:01:06,541 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteH
ostTrusted = false
```

Confirmacion en interfaz gráfica:



7. Mueve el fichero 'Frankenstein.txt' del directorio '/books' de HDFS al directorio home del usuario ubuntu del propio HDFS.

```

:\Users\alvar\Downloads\docker-hadoop-master_comandos\docker-hadoop-master_comandos\docker-hadoop-master_comandos\fi
o_result>docker exec -it namenode bash
oot@fd3f0eaac90c:/# hdfs dfs -put /tmp/prueba.txt /alvaro
025-05-05 18:01:06,541 INFO sas1.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteH
stTrusted = false
oot@fd3f0eaac90c:/# hdfs dfs -cp /books/frankenstein.txt /alvaro
025-05-05 18:03:35,521 INFO sas1.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteH
stTrusted = false
025-05-05 18:03:35,644 INFO sas1.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteH
stTrusted = false
oot@fd3f0eaac90c:/#

```

Movemos el fichero al directorio /books y hacemos confirmación en la interfaz gráfica.

The screenshot shows the Hadoop web interface at localhost:9870/explorer.html#/alvaro. The 'Browse Directory' view displays a table of files in the /alvaro directory. The table has columns for Permission, Owner, Group, Size, Last Modified, Replication, Block Size, Name, and a delete icon. Two files are listed: 'frankenstein.txt' (430.7 KB, May 05 20:03) and 'prueba.txt' (4 B, May 05 20:01). Both files are owned by 'root' and belong to the 'supergroup'.

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
-rw-r--r--	root	supergroup	430.7 KB	May 05 20:03	3	128 MB	frankenstein.txt	
-rw-r--r--	root	supergroup	4 B	May 05 20:01	3	128 MB	prueba.txt	

Showing 1 to 2 of 2 entries

Previous 1 Next

Hadoop, 2019.