

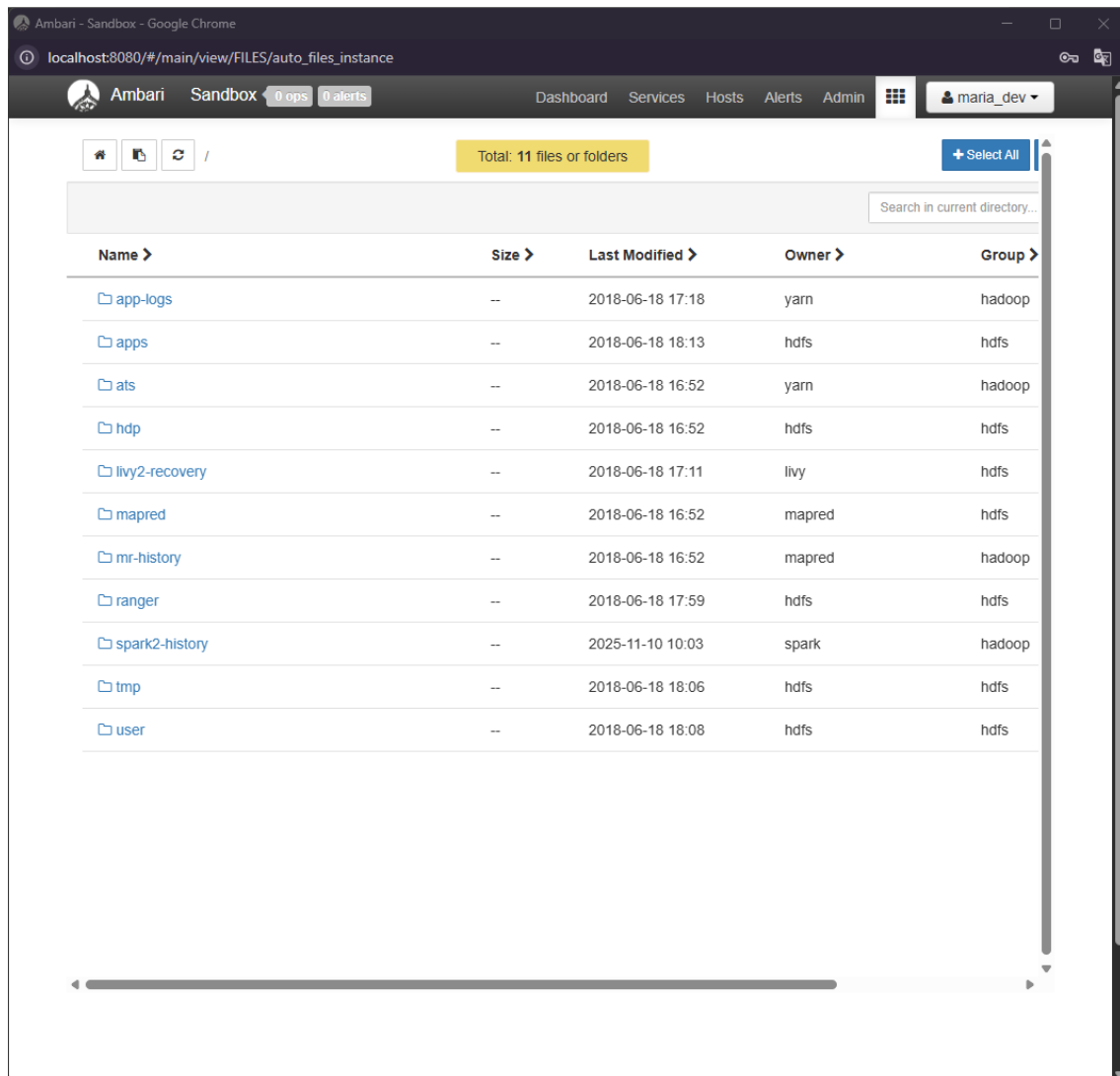
## 1. muestra la ayuda del comando para manejar el sistema de archivos de HDFS

```
PS C:\Users\Mañana> ssh maria_dev@localhost -p 2222
maria_dev@localhost's password:
Last login: Mon Nov 10 07:52:15 2025 from 172.18.0.3
[maria_dev@sandbox-hdp ~]$ hdfs dfs -help
Usage: hadoop fs [generic options]
    [-appendToFile <localsrc> ... <dst>]
    [-cat [-ignoreCrc] <src> ...]
    [-checksum <src> ...]
    [-chgrp [-R] GROUP PATH...]
    [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
    [-chown [-R] [OWNER][:[GROUP]] PATH...]
    [-copyFromLocal [-f] [-p] [-l] <localsrc> ... <dst>]
    [-copyToLocal [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
    [-count [-q] [-h] [-v] [-t <storage type>]] [-u] <path> ...]
    [-cp [-f] [-p | -p[topax]] <src> ... <dst>]
    [-createSnapshot <snapshotDir> [<snapshotName>]]
    [-deleteSnapshot <snapshotDir> <snapshotName>]
    [-df [-h] [<path> ...]]
    [-du [-s] [-h] <path> ...]
    [-expunge]
    [-find <path> ... <expression> ...]
    [-get [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
    [-getfacl [-R] <path>]
    [-getfattr [-R] {-n name | -d} [-e en] <path>]
    [-getmerge [-nl] <src> <localdst>]
    [-help [cmd ...]]
    [-ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [<path> ...]]
    [-mkdir [-p] <path> ...]
    [-moveFromLocal <localsrc> ... <dst>]
    [-moveToLocal <src> <localdst>]
    [-mv <src> ... <dst>]
    [-put [-f] [-p] [-l] <localsrc> ... <dst>]
    [-renameSnapshot <snapshotDir> <oldName> <newName>]
    [-rm [-f] [-r|-R] [-skipTrash] [-safely] <src> ...]
    [-rmdir [--ignore-fail-on-non-empty] <dir> ...]
```

## 2. muestra el contenido de la raíz de HDFS

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls /
Found 11 items
drwxrwxrwx   - yarn   hadoop           0 2018-06-18 15:18 /app-logs
drwxr-xr-x   - hdfs   hdfs            0 2018-06-18 16:13 /apps
drwxr-xr-x   - yarn   hadoop           0 2018-06-18 14:52 /ats
drwxr-xr-x   - hdfs   hdfs            0 2018-06-18 14:52 /hdp
drwx-----  - livy    hdfs            0 2018-06-18 15:11 /livy2-recovery
drwxr-xr-x   - mapred hdfs            0 2018-06-18 14:52 /mapred
drwxrwxrwx   - mapred hadoop           0 2018-06-18 14:52 /mr-history
drwxr-xr-x   - hdfs   hdfs            0 2018-06-18 15:59 /ranger
drwxrwxrwx   - spark  hadoop           0 2025-11-10 08:56 /spark2-history
drwxrwxrwx   - hdfs   hdfs            0 2018-06-18 16:06 /tmp
drwxr-xr-x   - hdfs   hdfs            0 2018-06-18 16:08 /user
```

3. visualiza dicha carpeta raíz desde el gestor de archivos del navegador. ¿En qué puerto se localiza?



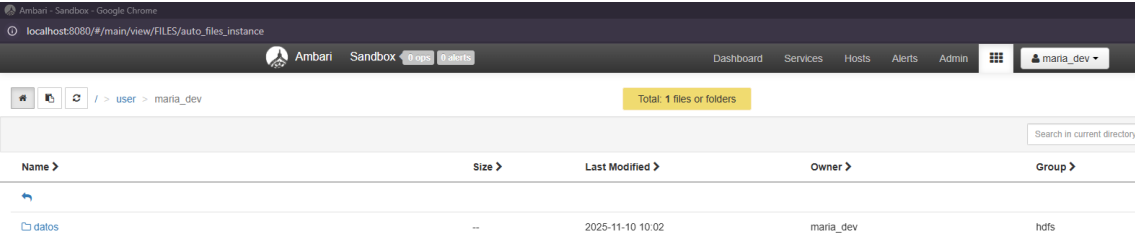
4. crea en HDFS un nuevo directorio llamado datos dentro de la carpeta /user/maria\_dev.

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -mkdir /user/maria_dev/datos
```

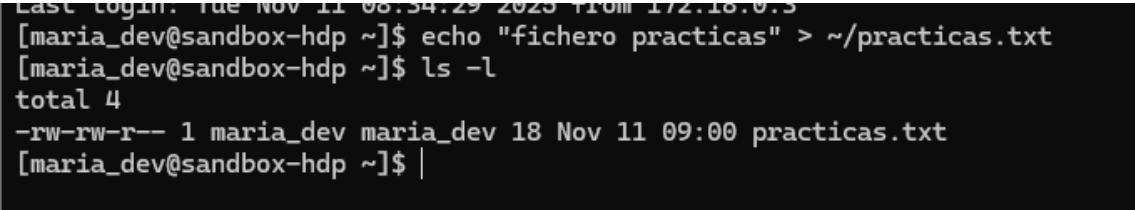
5. Comprobar que existe.

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls /user/maria_dev
Found 1 items
drwxr-xr-x - maria_dev hdfs          0 2025-11-10 09:02 /user/maria_dev/datos
```

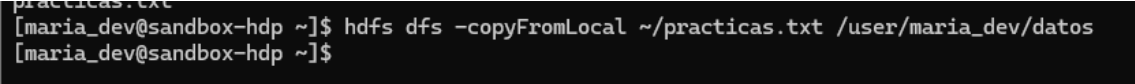
6. Mostrarlo desde el navegador.



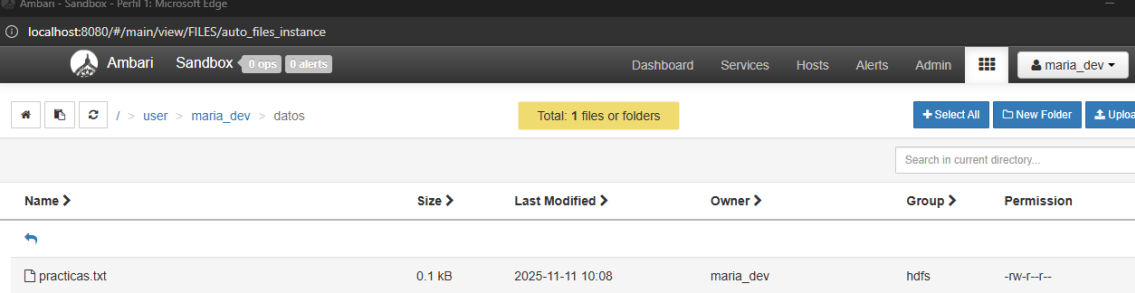
7. Crea un fichero llamado practicas.txt en tu directorio home de tu usuario en Linux con alguna frase dentro



8. Copiarlo en HDFS, en concreto al directorio datos anterior.



9. Comprueba su existencia desde la utilidad del navegador.



10. Haz clic sobre el fichero ¿Cuál es el tamaño del fichero en HDFS? ¿Cuánto ocupa realmente en HDFS? ¿Cuántas veces está replicado?

/user/maria_dev/datos							Go
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	maria_dev	hdfs	18 B	11/11/2025, 10:08:15	1	128 MB	practicas.txt

Hadoop, 2016.

El fichero son 18MB, como solo está replicado una vez, ocupa el mismo espacio.

11. Visualizar su contenido en HDFS.

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -cat /user/maria_dev/datos/practicas.txt  
fichero practicas  
[maria_dev@sandbox-hdp ~]$
```

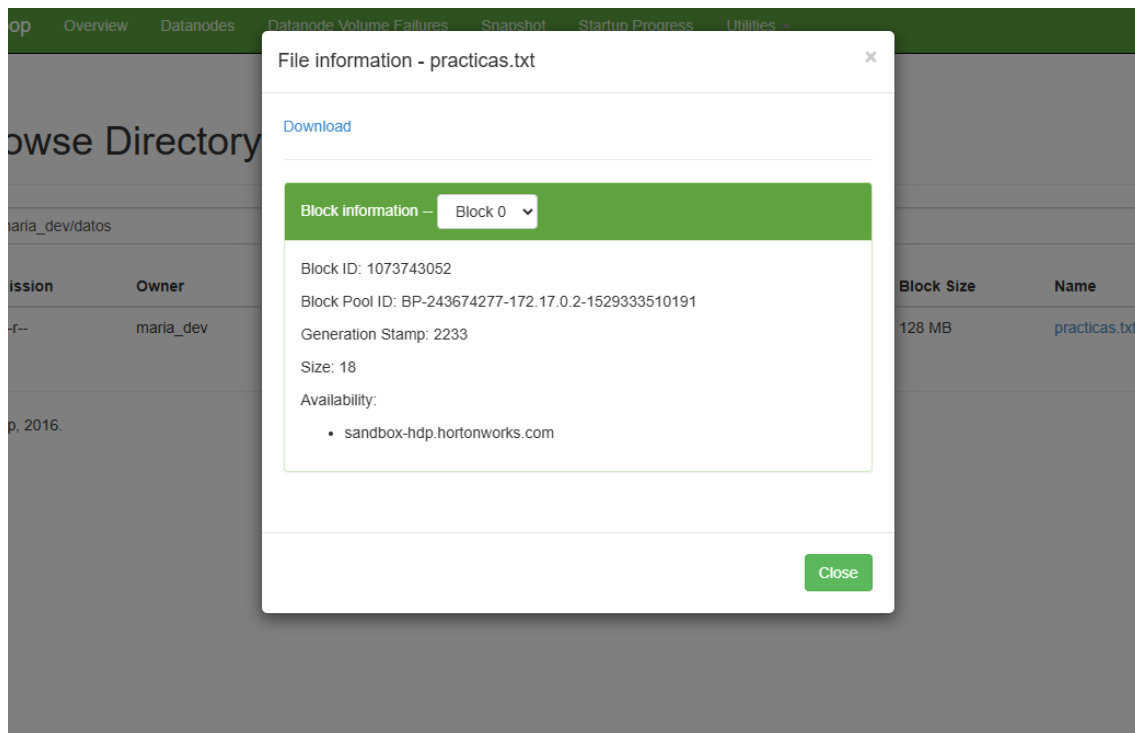
12. HDFS es un sistema de archivos distribuido que está dentro de una carpeta local de nuestro Linux. En su archivo de configuración está su ubicación. ¿Podrías localizarla? Explora su contenido desde Linux ¿Qué carpetas tiene dentro? ¿Dónde están los datos?

```
[maria_dev@sandbox-hdp ~]$ cd /etc/hadoop/conf  
[maria_dev@sandbox-hdp conf]$ ls  
capacity-scheduler.xml  hadoop-env.cmd  health_check  mapred-env.cmd  ranger-hdfs-security.xml  secure  taskcontroller.cfg  yarn.exclude  
commons-logging.properties  hadoop-env.sh  kms-acls.xml  mapred-env.sh  ranger-policymgr-ssl.xml  slaves  yarn-log4j.properties  yarn-site.xml  
configuration.xml  hadoop-metrics2.properties  kms-env.sh  mapred-queues.xml.template  ranger-policymgr-ssl-yarn.xml  ssl-client.xml  topology_mappings.data  
container-executor.cfg  hadoop-metrics.properties  kms-log4j.properties  mapred-site.xml  ranger-security.xml  ssl-client.xml.example  topology_scripts.py  
core-site.xml  hadoop-policy.xml  kms-site.xml  mapred-site.xml.template  ranger-yarn-audit.xml  ssl-server.xml  yarn-env.cmd  
dfs.exclude  hdfs-site.xml  log4j.properties  ranger-hdfs-audit.xml  ranger-yarn-security.xml  ssl-server.xml.example  yarn-env.sh  
[maria_dev@sandbox-hdp conf]$
```

```
[maria_dev@sandbox-hdp conf]$ cat hdfs-site.xml | grep hadoop  
<value>/hadoop/hdfs/data</value>  
<value>/var/lib/hadoop-hdfs/dn_socket</value>  
<value>/etc/hadoop/conf/dfs.exclude</value>  
<value>/hadoop/hdfs/journalnode</value>  
<value>/hadoop/hdfs/namesecondary</value>  
<value>org.apache.ranger.authorization.hadoop.RangerHdfsAuthorizer</value>  
<value>/hadoop/hdfs/namenode</value>  
<name>hadoop.caller.context.enabled</name>  
[maria_dev@sandbox-hdp conf]$
```

Los datos se encuentran en la carpeta 'data'

13. Si damos clic sobre el fichero en el entorno del navegador aparece el Block Id del fichero que nos indica el nombre del subdirectorio dentro del sistema de archivos local del punto anterior ¿Podrías mostrar su contenido?



14. Vamos a crear otro ejemplo con un fichero grande. Investiga como crear automáticamente desde Linux con un comando un archivo de 1GB en la carpeta home de tu usuario en Linux.

Se puede usar tanto el comando 'fallocate' como el 'truncate'

```
[maria_dev@sandbox-hdp ~]$ fallocate -l 1G ~/fichero_1GB
[maria_dev@sandbox-hdp ~]$ ls -l
total 1048584
-rw-r--r-- 1 maria_dev maria_dev 1073741824 Nov 12 10:43 fichero_1GB
-rw-rw-r-- 1 maria_dev maria_dev 18 Nov 11 09:06 practicas.txt
[maria_dev@sandbox-hdp ~]$
```

15. Copia el archivo anterior al directorio /datos de nuestro HDFS.

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -copyFromLocal ~/fichero_1GB /user/maria_dev/datos
[maria_dev@sandbox-hdp ~]$
```

16. Comprueba en la página web que ha creado múltiples bloques ¿Cuántos ha creado? ¿De qué tamaño son? Como solamente tenemos un nodo aparecen todos los bloques en el mismo, pero en un clúster real cada bloque estaría en un nodo distinto.

Ha creado 8 bloques, cada uno de 128 MB

### Browse Directory

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	maria_dev	hdfs	1 GB	12/11/2025, 11:46:55	1	128 MB	<a href="#">fichero_1GB</a>
-rw-r--r--	maria_dev	hdfs	18 B	11/11/2025, 10:08:15	1	128 MB	<a href="#">practicas.txt</a>

Hadoop, 2016.

Hadoop

Overview

Browse

/user/maria\_dev/datos

Permissions

-rw-r--r--

-rw-r--r--

Hadoop, 2016.

File information - fichero\_1GB

[Download](#)

Block information — Block 0

Block 0

Block 1

Block 2

Block 3

Block 4

Block 5

Block 6

Block 7

Block ID: 107374305

Block Pool ID: BP-2025111110081517.0.2-1529333510191

Generation Stamp: 2

Size: 134217728

Availability:

- sandbox-hdp...

Close

17. ¿Puedes localizar en el sistema de archivos local dichos bloques?

Ya que los datos se encuentran en el archivo data de hadoop, lo buscamos por el nombre del bloque, en este caso yo busco el bloque 5

```
[maria_dev@sandbox-hdp ~]$ find /hadoop/hdfs/data -name 1073743062
find: '/hadoop/hdfs/data': Permission denied
[maria_dev@sandbox-hdp ~]$ sudo find /hadoop/hdfs/data -name 1073743062
[maria_dev@sandbox-hdp ~]$ sudo find /hadoop/hdfs/data -name blk_1073743062
[maria_dev@sandbox-hdp ~]$ sudo find /hadoop/hdfs/data -name *1073743062*
/hadoop/hdfs/data/current/BP-243674277-172.17.0.2-1529333510191/current/finished/subdir0/subdir4/blk_1073743062_2245.meta
/hadoop/hdfs/data/current/BP-243674277-172.17.0.2-1529333510191/current/finished/subdir0/subdir4/blk_1073743062
[maria_dev@sandbox-hdp ~]$
```

18. Vamos a crear otro directorio llamado practicas dentro de la carpeta /user/maria\_dev.

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -mkdir /user/maria_dev/practicas
```

19. Copiamos prueba.txt desde datos a prácticas.

No sabia si era un fallo del ejercicio, pero el primer archivo se llamaba practicas.txt por lo que lo copie con ese nombre.

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -cp /user/maria_dev/datos/practicas.txt /user/maria_dev/practicas/practicas.txt
[maria_dev@sandbox-hdp ~]$
```

20. Comprobamos el contenido de prácticas.

```
/home/maria_dev
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls /user/maria_dev/practicas
Found 1 items
-rw-r--r-- 1 maria_dev hdfs      18 2025-11-12 11:11 /user/maria_dev/practicas/practicas.txt
[maria_dev@sandbox-hdp ~]$
```

21. Comprobamos el contenido de prueba.txt con un comando de HDFS.

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -cat /user/maria_dev/practicas/practicas.txt
fichero practicas
[maria_dev@sandbox-hdp ~]$
```

22. Borramos el fichero prueba.txt.

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -rm /user/maria_dev/practicas/practicas.txt
```

23. Borra el directorio prácticas.

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -rmdir /user/maria_dev/practicas
```