

# Prácticas con HDFS

- 1) Muestra la ayuda del comando para manejar el sistema de archivos de HDFS

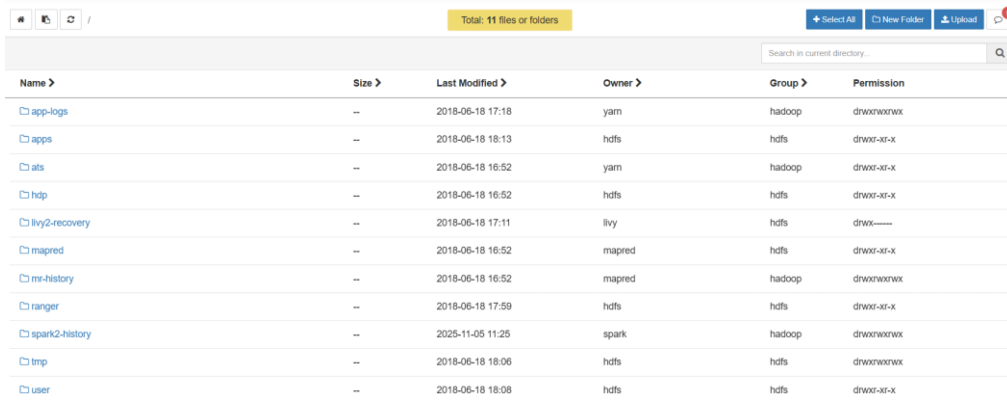
```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -help
Usage: hadoop fs [generic options]
[-appendToFile <localsrc> ... <dst>]
[-cat [-ignoreCrc] <src> ...]
[-checksum <src> ...]
[-chgrp [-R] GROUP PATH...]
[-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
[-chown [-R] [OWNER][:[GROUP]] PATH...]
[-copyFromLocal [-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> ...]
[-rm [-f] [-r|-R] [-skipTrash] [-safely] <src> ...]
```

- 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 2025-11-05 10:25 /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?

Está en el puerto 8080



Name >	Size >	Last Modified >	Owner >	Group >	Permission
app-logs	--	2018-06-18 17:18	yam	hadoop	drwxrwxrwx
apps	--	2018-06-18 18:13	hdfs	hdfs	drwxr-xr-x
ats	--	2018-06-18 16:52	yam	hadoop	drwxr-xr-x
hdp	--	2018-06-18 16:52	hdfs	hdfs	drwxr-xr-x
ivy2-recovery	--	2018-06-18 17:11	ivy	hdfs	drwx-----
mapred	--	2018-06-18 16:52	mapred	hdfs	drwxr-xr-x
mv-history	--	2018-06-18 16:52	mapred	hadoop	drwxrwxrwx
ranger	--	2018-06-18 17:59	hdfs	hdfs	drwxr-xr-x
spark2-history	--	2025-11-05 11:25	spark	hadoop	drwxrwxrwx
tmp	--	2018-06-18 18:06	hdfs	hdfs	drwxrwxrwx
user	--	2018-06-18 18:08	hdfs	hdfs	drwxr-xr-x

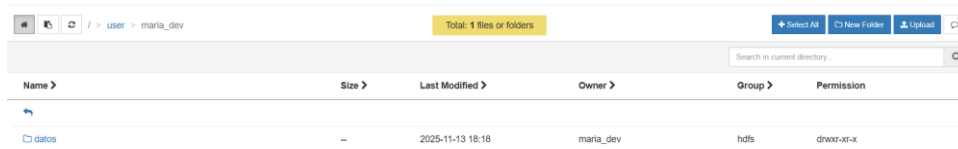
- 4) Crea en HDFS un nuevo directorio llamado datos dentro de la carpeta /user/maria\_dev.

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -mkdir /user/maria_dev/datos
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls /user/maria_dev
Found 1 items
drwxr-xr-x - maria_dev hdfs          0 2025-11-13 17:18 /user/maria_dev/datos
```

- 5) Comprobar que existe.

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -mkdir /user/maria_dev/datos
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls /user/maria_dev
Found 1 items
drwxr-xr-x - maria_dev hdfs          0 2025-11-13 17:18 /user/maria_dev/datos
```

- 6) Mostrarlo desde el navegador.



Name >	Size >	Last Modified >	Owner >	Group >	Permission
datos	--	2025-11-13 18:18	maria_dev	hdfs	drwxr-xr-x

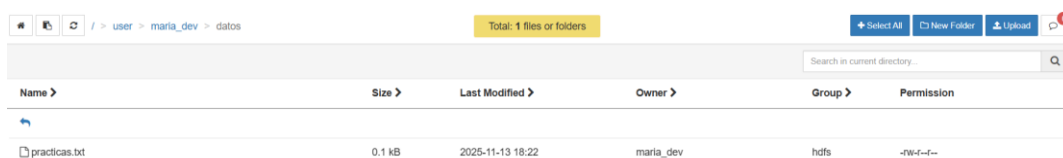
- 7) Crea un fichero llamado practicas.txt en tu directorio home de tu usuario en Linux con alguna frase dentro.

```
[maria_dev@sandbox-hdp ~]$ cd /home/maria_dev
[maria_dev@sandbox-hdp ~]$ echo "Hola profe" > practicas.txt
[maria_dev@sandbox-hdp ~]$ cat practicas.txt
Hola profe
```

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

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

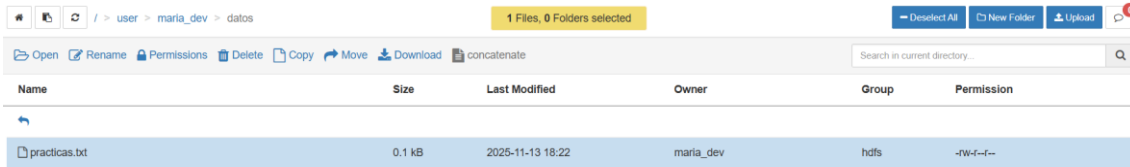
- 9) Comprueba su existencia desde la utilidad del navegador.



Name >	Size >	Last Modified >	Owner >	Group >	Permission
practicas.txt	0.1 kB	2025-11-13 18:22	maria_dev	hdfs	-rw-r--r--

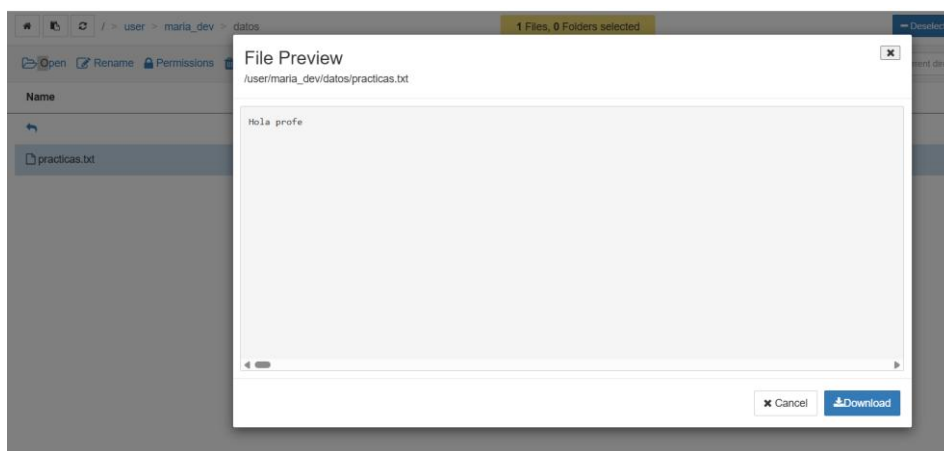
10) Haz clic sobre el fichero ¿Cuál es el tamaño del fichero en HDFS? ¿Cuánto ocupa realmente en HDFS? ¿Cuántas veces está replicado?

Pesa 0,1kB. En HDFS ocupa un bloque entero, 128 MB. Tiene 3 réplicas, el original y 2 copias.



Name	Size	Last Modified	Owner	Group	Permission
practicas.txt	0.1 kB	2025-11-13 18:22	maria_dev	hdfs	-rw-r--r--

11) Visualizar su contenido en HDFS.



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?

Visualizamos “/etc/hadoop/conf/hdfs-site.xml”:

```
[maria_dev@sandbox-hdp ~]$ cat /etc/hadoop/conf/hdfs-site.xml
```

```
<configuration>

  <property>
    <name>dfs.block.access.token.enable</name>
    <value>true</value>
  </property>

  <property>
    <name>dfs.blockreport.initialDelay</name>
    <value>120</value>
  </property>

  <property>
    <name>dfs.blocksize</name>
    <value>134217728</value>
  </property>

  <property>
    <name>dfs.client.read.shortcircuit</name>
    <value>true</value>
  </property>

  <property>
    <name>dfs.client.read.shortcircuit.stream.seek.size</name>
```

Podemos ver en esta parte que los metadatos se ubican en “/hadoop/hdfs/namenode”

```
<property>
  <name>dfs.namenode.name.dir</name>
  <value>/hadoop/hdfs/namenode</value>
  <final>true</final>
</property>
```

Y en esta que los archivos se guardan en “/hadoop/hdfs/data”

```
<property>
  <name>dfs.datanode.data.dir</name>
  <value>/hadoop/hdfs/data</value>
  <final>true</final>
</property>
```

- 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?

El block id es blk\_1073743048\_2228

```
[maria_dev@sandbox-hdp ~]$ hdfs fsck /user/maria_dev/datos/practicas.txt -files -blocks -locations
Connecting to namenode via http://sandbox-hdp.hortonworks.com:50070/fsck?ugi=maria_dev&files=1&blocks=1&locations=1&path=%2Fuser%2Fmaria_dev%2Fdatos%2Fpracticas.txt
FSCK started by maria_dev (auth:SIMPLE) from /172.18.0.2 for path /user/maria_dev/datos/practicas.txt at Thu Nov 13 18:04:11 UTC 2025
/user/maria_dev/datos/practicas.txt 11 bytes, 1 block(s): OK
0. BP-243674277-172.17.0.2-1529333510191:blk_1073743048_2228 len=11 repl=1 [DatanodeInfoWithStorage[172.18.0.2:50010,DS-ab75b94d-c6f2-4415-8639-1aaec2609e13,DISK]]

Status: HEALTHY
Total size: 11 B
Total dirs: 0
Total files: 1
Total symlinks: 0
Total blocks (validated): 1 (avg. block size 11 B)
Minimally replicated blocks: 1 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 0 (0.0 %)
Mis-replicated blocks: 0 (0.0 %)
Default replication factor: 1
Average block replication: 1.0
Corrupt blocks: 0
Missing replicas: 0 (0.0 %)
Number of data-nodes: 1
Number of racks: 1
FSCK ended at Thu Nov 13 18:04:11 UTC 2025 in 5 milliseconds

The filesystem under path '/user/maria_dev/datos/practicas.txt' is HEALTHY
```

Está aquí

```
[maria_dev@sandbox-hdp ~]$ sudo cat /hadoop/hdfs/data/current/BP-243674277-172.17.0.2-1529333510191/current/finalized/subdir0/subdir4/blk_1073743048
Hola profe
```

- 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.

Podemos usar dd o fallocate. En mi caso será fallocate sencillamente porque tarda menos.

```
[maria_dev@sandbox-hdp ~]$ cd /home/maria_dev
[maria_dev@sandbox-hdp ~]$
[maria_dev@sandbox-hdp ~]$ fallocate -l 1G ejemplo_grande.txt
```

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

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -put /home/maria_dev/ejemplo_grande.txt /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.

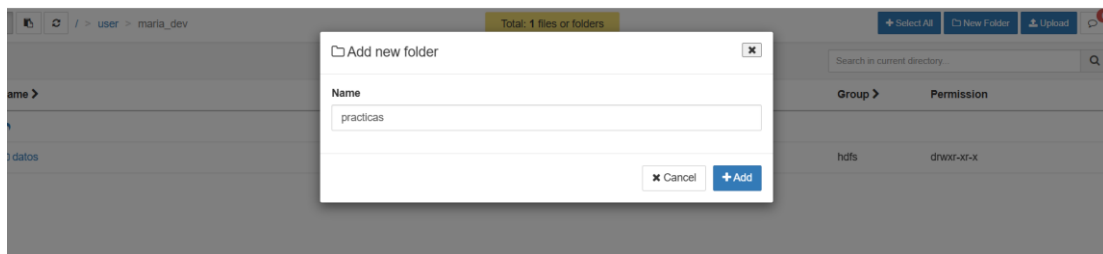
Ha creado 8 bloques, 7 de 128 Mb y el último de 64 MB

```
[maria_dev@sandbox-hdp ~]$ hdfs fsck /user/maria_dev/datos/ejemplo_grande.txt -files -blocks -locations
Connecting to namenode via http://sandbox-hdp.hortonworks.com:50070/fuse?ugi=maria_dev&files=1&blocks=1&locations=1&path=%2Fuser%2Fmaria_dev%2Fdatos%2Fejemplo_grande.tx
t
FSCK started by maria_dev (auth:SIMPLE) from /172.18.0.2 for path /user/maria_dev/datos/ejemplo_grande.txt at Thu Nov 13 18:22:00 UTC 2025
/user/maria_dev/datos/ejemplo_grande.txt 1073741824 bytes, 8 block(s): OK
0. BP-243674277-172.17.0.2-1529333510191:blk_1073743056_2237 len=134217728 repl=1 [DataNodeInfoWithStorage[172.18.0.2:50010,DS-ab75b94d-c6f2-4415-8639-1aaec2609e13,DISK
]]
1. BP-243674277-172.17.0.2-1529333510191:blk_1073743057_2238 len=134217728 repl=1 [DataNodeInfoWithStorage[172.18.0.2:50010,DS-ab75b94d-c6f2-4415-8639-1aaec2609e13,DISK
]]
2. BP-243674277-172.17.0.2-1529333510191:blk_1073743058_2239 len=134217728 repl=1 [DataNodeInfoWithStorage[172.18.0.2:50010,DS-ab75b94d-c6f2-4415-8639-1aaec2609e13,DISK
]]
3. BP-243674277-172.17.0.2-1529333510191:blk_1073743059_2240 len=134217728 repl=1 [DataNodeInfoWithStorage[172.18.0.2:50010,DS-ab75b94d-c6f2-4415-8639-1aaec2609e13,DISK
]]
4. BP-243674277-172.17.0.2-1529333510191:blk_1073743060_2241 len=134217728 repl=1 [DataNodeInfoWithStorage[172.18.0.2:50010,DS-ab75b94d-c6f2-4415-8639-1aaec2609e13,DISK
]]
5. BP-243674277-172.17.0.2-1529333510191:blk_1073743061_2242 len=134217728 repl=1 [DataNodeInfoWithStorage[172.18.0.2:50010,DS-ab75b94d-c6f2-4415-8639-1aaec2609e13,DISK
]]
6. BP-243674277-172.17.0.2-1529333510191:blk_1073743062_2243 len=134217728 repl=1 [DataNodeInfoWithStorage[172.18.0.2:50010,DS-ab75b94d-c6f2-4415-8639-1aaec2609e13,DISK
]]
7. BP-243674277-172.17.0.2-1529333510191:blk_1073743063_2244 len=134217728 repl=1 [DataNodeInfoWithStorage[172.18.0.2:50010,DS-ab75b94d-c6f2-4415-8639-1aaec2609e13,DISK
]]
```

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

```
[maria_dev@sandbox-hdp ~]$ sudo ls /hadoop/hdfs/data/current/BP-243674277-172.17.0.2-1529333510191/current/finalized/subdir0/subdir4/blk_1073743063(57..63)
/hadoop/hdfs/data/current/BP-243674277-172.17.0.2-1529333510191/current/finalized/subdir0/subdir4/blk_1073743057
/hadoop/hdfs/data/current/BP-243674277-172.17.0.2-1529333510191/current/finalized/subdir0/subdir4/blk_1073743058
/hadoop/hdfs/data/current/BP-243674277-172.17.0.2-1529333510191/current/finalized/subdir0/subdir4/blk_1073743059
/hadoop/hdfs/data/current/BP-243674277-172.17.0.2-1529333510191/current/finalized/subdir0/subdir4/blk_1073743060
/hadoop/hdfs/data/current/BP-243674277-172.17.0.2-1529333510191/current/finalized/subdir0/subdir4/blk_1073743061
/hadoop/hdfs/data/current/BP-243674277-172.17.0.2-1529333510191/current/finalized/subdir0/subdir4/blk_1073743062
/hadoop/hdfs/data/current/BP-243674277-172.17.0.2-1529333510191/current/finalized/subdir0/subdir4/blk_1073743063
```

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



19) Copiamos prueba.txt desde datos a practicas.

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

20) Comprobamos el contenido de prácticas.

/ user > maria_dev > practicas					
Total: 1 files or folders					
Search in current directory...					
Name >	Size >	Last Modified >	Owner >	Group >	Permission
practicas.txt	0.1 kB	2025-11-13 19:36	maria_dev	hdfs	-rw-r--r--

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  
Hola profe
```

### 22) Borramos el fichero prueba.txt.

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -rm /user/maria_dev/practicas/practicas.txt  
25/11/13 18:38:10 INFO fs.TrashPolicyDefault: Moved: 'hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/practicas/practicas.txt' to trash at: hdfs://sandbox-hdp.ho  
rtonworks.com:8020/user/maria_dev/.Trash/Current/user/maria_dev/practicas/practicas.txt
```

### 23) Borra el directorio practicas.

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -rm -r /user/maria_dev/practicas  
25/11/13 18:39:07 INFO fs.TrashPolicyDefault: Moved: 'hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/practicas' to trash at: hdfs://sandbox-hdp.hortonworks.com:  
8020/user/maria_dev/.Trash/Current/user/maria_dev/practicas1763059147384
```