

Prácticas con HDFS

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

```
[maria_dev@sandbox-hdp ~]$ hdfs
Usage: hdfs [--config confdir] [--loglevel loglevel] COMMAND
  where COMMAND is one of:
    dfs                run a filesystem command on the file systems supported in Hadoop.
    classpath          prints the classpath
    namenode -format    format the DFS filesystem
    secondarynamenode  run the DFS secondary namenode
    namenode            run the DFS namenode
    journalnode         run the DFS journalnode
    zkfc               run the ZK Failover Controller daemon
    datanode            run a DFS datanode
    dfsadmin            run a DFS admin client
    envvars             display computed Hadoop environment variables
    haadmin            run a DFS HA admin client
    fsck               run a DFS filesystem checking utility
    balancer           run a cluster balancing utility
    jmxget             get JMX exported values from NameNode or DataNode.
    mover              run a utility to move block replicas across
                      storage types
    oiv               apply the offline fsimage viewer to an fsimage
    oiv_legacy         apply the offline fsimage viewer to an legacy fsimage
    oev               apply the offline edits viewer to an edits file
    fetchdt           fetch a delegation token from the NameNode
    getconf            get config values from configuration
    groups             get the groups which users belong to
    snapshotDiff       diff two snapshots of a directory or diff the
                      current directory contents with a snapshot
    lsSnapshottableDir list all snapshottable dirs owned by the current user
                      Use -help to see options
    portmap           run a portmap service
    nfs3              run an NFS version 3 gateway
    cacheadmin         configure the HDFS cache
    crypto             configure HDFS encryption zones
    storagepolicies    list/get/set block storage policies
    version            print the version

Most commands print help when invoked w/o parameters.
[maria_dev@sandbox-hdp ~]$
```

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 2025-11-06 08:44 /spark2-history
drwxrwxrwx - hdfs  hdfs      0 2018-06-18 16:06 /tmp
drwxr-xr-x - hdfs  hdfs      0 2018-06-18 16:08 /user
[maria_dev@sandbox-hdp ~]$
```

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

Browse Directory

/							Go!
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxrwxrwx	yarn	hadoop	0 B	18/6/2018, 17:18:48	0	0 B	app-logs
drwxr-xr-x	hdfs	hdfs	0 B	18/6/2018, 18:13:31	0	0 B	apps
drwxr-xr-x	yarn	hadoop	0 B	18/6/2018, 16:52:05	0	0 B	ats
drwxr-xr-x	hdfs	hdfs	0 B	18/6/2018, 16:52:12	0	0 B	hdp
drwx-----	livy	hdfs	0 B	18/6/2018, 17:11:42	0	0 B	livy2-recovery
drwxr-xr-x	mapred	hdfs	0 B	18/6/2018, 16:52:10	0	0 B	mapred
drwxrwxrwx	mapred	hadoop	0 B	18/6/2018, 16:52:18	0	0 B	mr-history
drwxr-xr-x	hdfs	hdfs	0 B	18/6/2018, 17:59:35	0	0 B	ranger
drwxrwxrwx	spark	hadoop	0 B	6/11/2025, 8:56:50	0	0 B	spark2-history
drwxrwxrwx	hdfs	hdfs	0 B	18/6/2018, 18:06:11	0	0 B	tmp
drwxr-xr-x	hdfs	hdfs	0 B	18/6/2018, 18:08:07	0	0 B	user

Hadoop, 2016.

Diría que en el máster. Porque no tiene ninguna replicación.

4. Crea en HDFS un nuevo directorio llamado datos dentro de la carpeta /user/maria_dev.

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -mkdir -p /user/maria_dev/datos
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls /user/maria_dev/
Found 1 items
drwxr-xr-x  - maria_dev hdfs          0 2025-11-06 08:47 /user/maria_dev/datos
```

5. Comprobar que existe.

Esta en la imagen del ejercicio anterior

6. Mostrarlo desde el navegador.

Browse Directory

<input type="text" value="/user/aria_dev"/>							<input data-bbox="1265 526 1303 548" type="button" value="Go!"/>
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	aria_dev	hdfs	0 B	6/11/2025, 9:47:34	0	0 B	datos

Hadoop, 2016.

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

```
[aria_dev@sandbox-hdp ~]$ echo "Esta es una frase de práctica para el archivo practicas.txt." > practicas.txt
[aria_dev@sandbox-hdp ~]$ ls
practicas.txt
[aria_dev@sandbox-hdp ~]$ cat practicas.txt
Esta es una frase de práctica para el archivo practicas.txt.
```

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

```
[aria_dev@sandbox-hdp ~]$ hdfs dfs -put practicas.txt /user/aria_dev/
[aria_dev@sandbox-hdp ~]$ hdfs dfs -ls /user/aria_dev/
Found 2 items
drwxr-xr-x  - aria_dev hdfs          0 2025-11-06 08:47 /user/aria_dev/datos
-rw-r--r--  1 aria_dev hdfs        62 2025-11-06 08:53 /user/aria_dev/practicas.txt
```

9. Comprueba su existencia desde la utilidad del navegador.

Browse Directory

<input type="text" value="/user/aria_dev"/>							<input data-bbox="1265 1594 1303 1617" type="button" value="Go!"/>
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	aria_dev	hdfs	0 B	6/11/2025, 9:47:34	0	0 B	datos
-rw-r--r--	aria_dev	hdfs	62 B	6/11/2025, 9:53:28	1	128 MB	practicas.txt

Hadoop, 2016.

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

El fichero ocupa 128mb según HDFS pero pesa 62 bytes y esta replicado una vez.

11. Visualizar su contenido en HDFS.

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -cat /user/maria_dev/practicass.txt
Esta es una frase de práctica para el archivo practicas.txt.
```

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?

Aquí esta el archivo de configuración.

```
[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>
```

Esto es lo que contiene la carpeta conf de hadoop

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

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.

```
[maria_dev@sandbox-hdp ~]$ dd if=/dev/zero of=archivo_1gb.dat bs=1M count=1024
1024+0 records in
1024+0 records out
1073741824 bytes (1.1 GB) copied, 6.85391 s, 157 MB/s
[maria_dev@sandbox-hdp ~]$ ls -lh archivo_1gb.dat
-rw-rw-r-- 1 maria_dev maria_dev 1.0G Nov  6 09:09 archivo_1gb.dat
```

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

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -put archivo_1gb.dat /datos
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls /datos
-rw-r--r--  1 maria_dev hdfs 1073741824 2025-11-06 09:11 /datos
```

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.

File information - datos



[Download](#)

Block information -- Block 0 ▾

Block 0

Block ID: 107374304

Block Pool ID: BP-217.0.2-1529333510191

Generation Stamp: 2

Size: 134217728

Availability:

- sandbox-hdp. pm

Ha creado 8, con un tamaño de 128mb cada uno.

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

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 practicas.

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -cp practicas.txt practicas/prueba.txt
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls
Found 4 items
drwxr-xr-x - maria_dev hdfs      0 2025-11-06 08:47 datos
drwxr-xr-x - maria_dev hdfs      0 2025-11-11 08:59 practicas
-rw-r--r-- 1 maria_dev hdfs      62 2025-11-06 08:53 practicas.txt
-rw-r--r-- 1 maria_dev hdfs    22628 2025-11-10 07:50 u.user
^[[A^[[A[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls practicas
Found 1 items
-rw-r--r-- 1 maria_dev hdfs      62 2025-11-11 08:59 practicas/prueba.txt
```

20. Comprobamos el contenido de practicas.

Incluido en la imagen del ejercicio anterior.

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

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -cat practicas/prueba.txt
Esta es una frase de práctica para el archivo practicas.txt.
```

22. Borramos el fichero prueba.txt.

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -rm practicas/prueba.txt
25/11/11 09:01:15 INFO fs.TrashPolicyDefault: Moved: 'hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/practicas/prueba.txt' to trash at: hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/Current/user/maria_dev/practicas/prueba.txt
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls practicas
[maria_dev@sandbox-hdp ~]$
```

23. Borra el directorio practicas.

```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -rm -r practicas
25/11/11 09:03:17 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/practicas1762851797349
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls
Found 4 items
drwx----- - maria_dev hdfs      0 2025-11-11 09:01 .Trash
drwxr-xr-x - maria_dev hdfs      0 2025-11-06 08:47 datos
-rw-r--r-- 1 maria_dev hdfs      62 2025-11-06 08:53 practicas.txt
-rw-r--r-- 1 maria_dev hdfs    22628 2025-11-10 07:50 u.user
```