Prácticas BigData

1. Administración de HDFS

- Realizar un report del sistema HDFS

> **hdfs dfsadmin -report**
>
> Configured Capacity: 50940104704 (47.44 GB)
>
> Present Capacity: 44860829696 (41.78 GB)
>
> DFS Remaining: 44859387904 (41.78 GB)
>
> DFS Used: 1441792 (1.38 MB)
>
> DFS Used%: 0.00%
>
> Under replicated blocks: 0
>
> Blocks with corrupt replicas: 0
>
> Missing blocks: 0
>
>
> ---
>
> Live datanodes (1):
>
>
> Name: 127.0.0.1:50010 (localhost)
>
> Hostname: localhost
>
> Decommission Status : Normal
>
> Configured Capacity: 50940104704 (47.44 GB)
>
> DFS Used: 1441792 (1.38 MB)
>
> Non DFS Used: 6079275008 (5.66 GB)
>
> DFS Remaining: 44859387904 (41.78 GB)
>
> DFS Used%: 0.00%
>
> DFS Remaining%: 88.06%
>
> Configured Cache Capacity: 0 (0 B)
>
> Cache Used: 0 (0 B)
>
> Cache Remaining: 0 (0 B)
>
> Cache Used%: 100.00%
>
> Cache Remaining%: 0.00%
>
> Xceivers: 1
>
> Last contact: Mon Apr 20 08:17:15 CEST 2015

- Comprobar con hdfs fsck el estado del sistema de ficheros

> **hdfs fsck /**

```
Connecting to namenode via http://localhost:50070

FSCK started by hadoop (auth:SIMPLE) from /127.0.0.1 for path / at Mon Apr 20
08:19:57 CEST 2015

......Status: HEALTHY

 Total size:    1380389 B

 Total dirs:    6

 Total files:  6

 Total symlinks:                0

 Total blocks (validated):     5 (avg. block size 276077 B)

 Minimally replicated blocks:   5 (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 Mon Apr 20 08:19:57 CEST 2015 in 14 milliseconds



The filesystem under path '/' is HEALTHY
```

- También Podemos comprobar el estado de un determinado directorio, por ejemplo de prácticas

```
hdfs fsck /practicas

Connecting                    to               namenode                via
http://localhost:50070/fsck?ugi=hadoop&path=%2Fpracticas

FSCK started by hadoop (auth:SIMPLE) from /127.0.0.1 for path /practicas at Sat
Jan 06 19:55:19 CET 2018

...Status: HEALTHY

 Total size:  2647821 B

 Total dirs:  2

 Total files:  3

 Total symlinks:                0

 Total blocks (validated):    2 (avg. block size 1323910 B)

 Minimally replicated blocks:       2 (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 Sat Jan 06 19:55:19 CET 2018 in 7 milliseconds		

- Podemos ver la topología que tenemos en este momento

hdfs dfsadmin -printTopology

Rack: /default-rack

  127.0.0.1:50010 (localhost)

- Podemos comprobar si hay algún fichero abierto

hdfs dfsadmin -listOpenFiles

Client Host	Client Name	Open File Path

-