

Prácticas BigData

1. Snapshots

- Creamos un pequeño fichero en /tmp

```
echo Ejemplo de Snapshot > /tmp/f1.txt
```

- Creamos un directorio HDFS para probar

```
hadoop fs -mkdir /datos4
```

- Subimos el fichero pequeño que hayamos creado

```
hdfs dfs -put f1.txt /datos4
```

- Ejecutamos un fsck sobre el fichero. Podemos preguntar información de bloques, ficheros, nodos donde se encuentra, etc... una opción muy interesante.

```
hdfs fsck /datos4/f1.txt -files -blocks -locations
```

```
Connecting to namenode via
http://localhost:50070/fsck?ugi=hadoop&files=1&blocks=1&locations=1&path=
%2Fdatos4%2Ff1.txt
```

```
FSCK started by hadoop (auth:SIMPLE) from /127.0.0.1 for path /datos4/f1.txt
at Sat Jan 06 20:16:02 CET 2018
```

```
/datos4/f1.txt 20 bytes, 1 block(s): OK
```

```
0. BP-344905797-192.168.56.101-1515254230192:blk_1073741837_1013
len=20 Live_repl=1 [DatanodeInfoWithStorage[127.0.0.1:50010,DS-173cc83b-
694a-425e-ad0f-c4c86352e2f6,DISK]]
```

```
Status: HEALTHY
```

```
Total size: 20 B
```

```
Total dirs: 0
```

```
Total files: 1
```

```
Total symlinks: 0
```

```
Total blocks (validated): 1 (avg. block size 20 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 Sat Jan 06 20:16:02 CET 2018 in 1 milliseconds
```

The filesystem under path '/datos4/f1.txt' is HEALTHY

- Buscamos el fichero en el Sistema de ficheros del Linux a partir de su número de bloque

```
find /datos/datanode -name "blk_1073741837"
/datos/datanode/current/BP-344905797-192.168.56.101-
1515254230192/current/finalized/subdir0/subdir0/blk_1073741837
```

- Habilitamos los snapshot sobre el directorio /datos4

```
hdfs dfsadmin -allowSnapshot /datos4
Allowing snaphot on /datos4 succeeded
```

- Creamos un snapshot llamado "s1" en el directorio

```
hdfs dfs -createSnapshot /datos4 s1
Created snapshot /datos4/.snapshot/s1
```

- Comprobamos que se ha creado satisfactoriamente

```
hdfs dfs -ls /datos4/.snapshot
Found 1 items
drwxr-xr-x - hadoop supergroup      0 2018-01-06 20:20 /datos4/.snapshot/s1
```

- Si hacemos un ls, en principio debe tener lo mismo que su directorio asociado

```
hdfs dfs -ls /datos4/.snapshot/s1
Found 1 items
-rw-r--r--      1 hadoop supergroup      20 2018-01-06 20:15
/datos4/.snapshot/s1/f1.txt
```

- Vamos a borrar el fichero f1.txt

```
hdfs dfs -rm /datos4/f1.txt
Deleted /datos4/f1.txt
```

- Podemos comprobar que ya no existe

```
hdfs dfs -ls /datos4
```

- Sin embargo, con Snapshots, es muy fácil recuperarlo, sencillamente lo copiamos de nuevo a su sitio original

-

```
hadoop fs -cp /datos4/.snapshot/s1/f1.txt /datos4/
```

```
hdfs dfs -ls /datos4
```

```
Found 1 items
```

```
-rw-r--r--  1 hadoop supergroup      20 2018-01-06 20:25 /datos4/f1.txt
```