



iesgrancapitan-CEIABD-BDA / ud4-practica-1-hadoop-hdfs-Dansar...



<> Code

Issues

Pull requests 1

Actions

Projects

Wiki

Security



main ▾

ud4-practica-1-hadoop-hdfs-Dansarasix-DML / Readme.md



Dansarasix-DML update2.5

435901f · 6 months ago



178 lines (163 loc) · 7.03 KB

Preview

Code

Blame



Raw



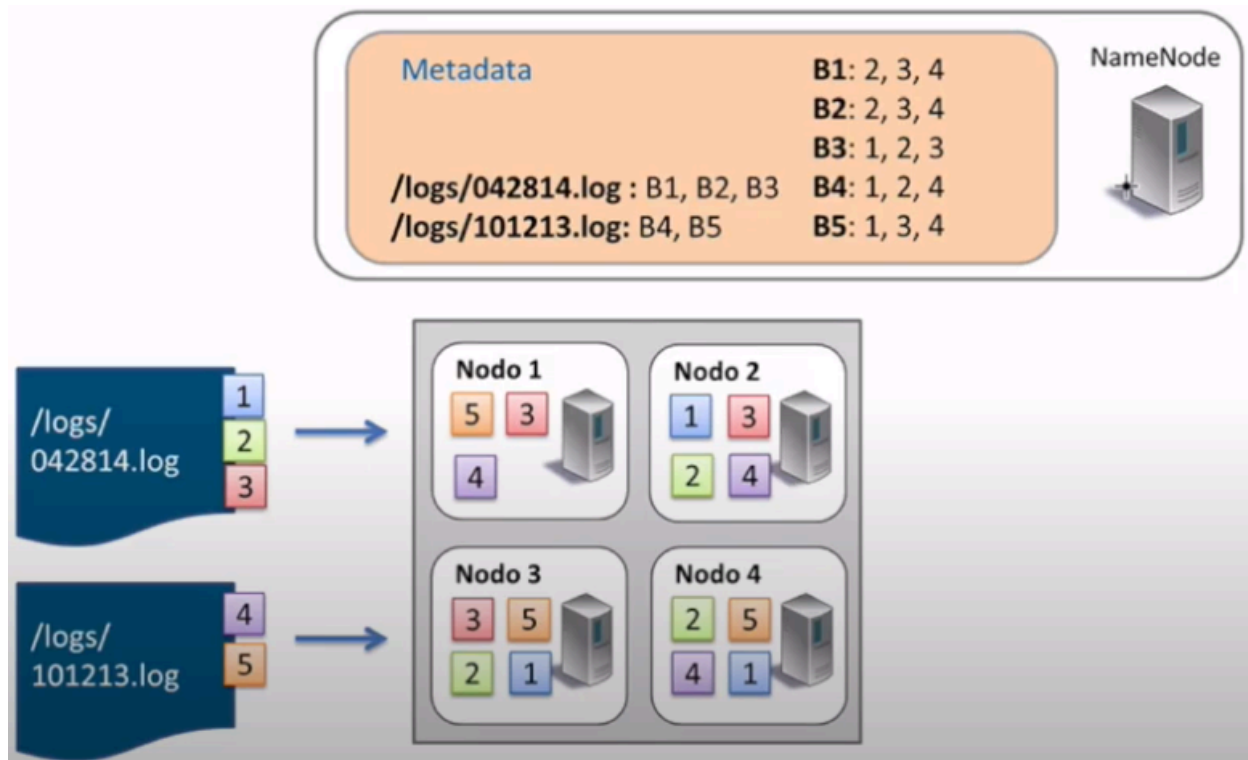
Big Data Aplicado

UD 4 - Apache Hadoop

🔗 Práctica 1 HDFS

Para los siguientes ejercicios, copia el comando y haz una captura de pantalla donde se muestre el resultado de cada acción

1. Explica paso a paso el proceso de lectura (indicando qué bloques y los datanodes empleados) que realiza HDFS si queremos leer el archivo `/logs/101213.log` (RA5075.3 / CE.3a)



Proceso de lectura HDFS

Primero el cliente consulta al **NameNode** de los bloques que corresponden el archivo `/logs/101213.log`.

El **NameNode** devuelve la siguiente información:

- El archivo se encuentra en los siguientes bloques:
 - Bloque 4:** Copiado en los **DataNodes** 1, 2, 4.
 - Bloque 5:** Copiando en los **DataNodes** 1, 3, 4.

El siguiente paso que debe hacer el cliente es elegir un **DataNode** que tenga una copia del bloque. En el caso de Hadoop lo hace de manera optimizada ya sea por cercanía ya sea local o de red. En nuestro caso sería:

- Bloque 4:** Se pueden elegir los siguientes **DataNodes**:
 - Nodo 1
 - Nodo 2
 - Nodo 4
- Bloque 5:** Se pueden elegir los siguientes **DataNodes**:
 - Nodo 1
 - Nodo 3
 - Nodo 4

El cliente inicia la tarea de lectura del archivo en cuestión. Para ello se deben recuperar los bloques en el orden correspondiente:

- Primero el Bloque 4 y luego el Bloque 5.
- Como ambos bloques están en el **DataNode** 1, se realizarán las recuperaciones en el mismo **DataNode** que es mucho más eficiente que buscar otro bloque en otro **DataNode**.

Luego por último, los bloques han sido leídos y el cliente reconstruye el archivo original.

2. En este ejercicio vamos a practicar los comandos básicos de **HDFS**. Una vez arrancado Hadoop (*RA5075.3 / CE.3a y CE.3c*):

- Crea la carpeta `/bda/ejercicios`.

```
hdfs dfs -mkdir /bda/ejercicios
```



```
hadoop@master:~$ hdfs dfs -mkdir /bda
hadoop@master:~$ hdfs dfs -mkdir /bda/ejercicios
hadoop@master:~$ |
```

- Crea o descargar un fichero de texto con tu nombre y apellidos `Apellido1Apellido2Nombre.txt` en local.

```
touch MarinLopezDaniel.txt
```



```
hadoop@master:~$ touch MarinLopezDaniel.txt
hadoop@master:~$ ls
hadoop-3.4.1.tar.gz  MarinLopezDaniel.txt
hadoop@master:~$
```

- Sube el archivo `Apellido1Apellido2Nombre.txt` a la carpeta creada.

```
hdfs dfs -put -f MarinLopezDaniel.txt /bda/ejercicios
```



```
hadoop@master:~$ hdfs dfs -put -f MarinLopezDaniel.txt /bda/ejercicios
hadoop@master:~$
```

- Crea una copia en HDFS y llámala `Apellido1Apellido2Nombre2.txt`.

```
hdfs dfs -cp /bda/ejercicios/MarinLopezDaniel.txt
/bda/ejercicios/MarinLopezDaniel2.txt
```



```
hadoop@master:~$ hdfs dfs -cp /bda/ejercicios/MarinLopezDaniel.txt /bda/ejercicios/MarinLopezDaniel2.txt
hadoop@master:~$
```

- Comprueba que se ha creado correctamente el fichero `Apellido1Apellido2Nombre2.txt`.

```
hdfs dfs -ls /bda/ejercicios
```



```
hadoop@master:~$ hdfs dfs -ls /bda/ejercicios
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2024-12-19 10:27 /bda/ejercicios/MarinLopezDaniel.txt
-rw-r--r-- 1 hadoop supergroup 0 2024-12-19 10:29 /bda/ejercicios/MarinLopezDaniel2.txt
```

- Renombra Apellido1Apellido2Nombre2.txt a Apellido1Apellido2Nombre2_copia.txt .

```
hdfs dfs -mv /bda/ejercicios/MarinLopezDaniel2.txt
/bda/ejercicios/MarinLopezDaniel2_copia.txt
```



```
hadoop@master:~$ hdfs dfs -mv /bda/ejercicios/MarinLopezDaniel2.txt /bda/ejercicios/MarinLopezDaniel2_copia.txt
hadoop@master:~$
```

- Descarga en local Apellido1Apellido2Nombre2_copia.txt con su código CRC.

```
hdfs dfs -get /bda/ejercicios/MarinLopezDaniel2_copia.txt
```



```
hadoop@master:~$ hdfs dfs -get /bda/ejercicios/MarinLopezDaniel2_copia.txt
hadoop@master:~$ ls
hadoop-3.4.1.tar.gz MarinLopezDaniel2_copia.txt MarinLopezDaniel.txt
hadoop@master:~$
```

- Adjunta una captura desde la interfaz web donde se vean ambos archivos.

Browse Directory

/bda/ejercicios

Go!

Show

25

entries

Search:

<input type="checkbox"/>	<div><div></div><div></div></div> Permission	<div><div></div><div></div></div> Owner	<div><div></div><div></div></div> Group	<div><div></div><div></div></div> Size	<div><div></div><div></div></div> Last Modified	<div><div></div><div></div></div> Replication	<div><div></div><div></div></div> Block Size	<div><div></div><div></div></div> Name	<div><div></div><div></div></div>
<input type="checkbox"/>	-rw-r--r--	hadoop	supergroup	0 B	Dec 19 11:27	1	128 MB	MarinLopezDaniel.txt	<div><div></div><div></div></div>
<input type="checkbox"/>	-rw-r--r--	hadoop	supergroup	0 B	Dec 19 11:29	1	128 MB	MarinLopezDaniel2_copia.txt	<div><div></div><div></div></div>

Showing 1 to 2 of 2 entries

Previous

1

Next

- Vuelve al terminal y elimina la carpeta con los archivos contenidos mediante un único comando.

```
hdfs dfs -rm -r /bda/ejercicios
```



```
hadoop@master:~$ hdfs dfs -rm -r /bda/ejercicios
Deleted /bda/ejercicios
hadoop@master:~$
```

3. Vamos a practicar los comandos de gestión de instantáneas y administración de HDFS. Para ello: (RA5075.4 / CE.4e y CE.4f)

- Crea la carpeta `/bda/snaps`.

```
hdfs dfs -mkdir /bda/snaps
```



```
hadoop@master:~$ hdfs dfs -mkdir /bda/snaps
hadoop@master:~$
```

- Habilita las snapshots sobre la carpeta creada.

```
hdfs dfsadmin -allowSnapshot /bda/snaps
```



```
hadoop@master:~$ hdfs dfsadmin -allowSnapshot /bda/snaps
Allowing snapshot on /bda/snaps succeeded
hadoop@master:~$
```

- Sube el archivo `Apellido1Apellido2Nombre.txt` creado antes en local a la carpeta creada.

```
hdfs dfs -put -f MarinLopezDaniel.txt /bda/snaps
```



```
hadoop@master:~$ hdfs dfs -put -f MarinLopezDaniel.txt /bda/snaps
hadoop@master:~$
```

- Crea una copia en HDFS y llámala `Apellido1Apellido2Nombre_snapshot.txt`.

```
hdfs dfs -cp /bda/snaps/MarinLopezDaniel.txt
/bda/snaps/MarinLopezDaniel_snapshot.txt
```



```
hadoop@master:~$ hdfs dfs -cp /bda/snaps/MarinLopezDaniel.txt /bda/snaps/Ma
rinLopezDaniel_snapshot.txt
hadoop@master:~$
```

- Crea una instantánea de la carpeta llamada `snap1`.

```
hdfs dfs -createSnapshot /bda/snaps snap1
```



```
hadoop@master:~$ hdfs dfs -createSnapshot /bda/snaps snap1
Created snapshot /bda/snaps/.snapshot/snap1
hadoop@master:~$
```

- Elimina ambos ficheros.

```
hdfs dfs -rm /bda/snaps/*
```



```
hadoop@master:~$ hdfs dfs -rm /bda/snaps/*
Deleted /bda/snaps/MarinLopezDaniel.txt
Deleted /bda/snaps/MarinLopezDaniel_snapshot.txt
hadoop@master:~$
```

- Comprueba que la carpeta está vacía.

```
hdfs dfs -ls /bda/snaps
```



```
hadoop@master:~$ hdfs dfs -ls /bda/snaps
hadoop@master:~$
```

- Recupera desde snap1 el archivo Apellido1Apellido2Nombre.txt .

```
hdfs dfs -cp /bda/snaps/.snapshot/snap1/MarinLopezDaniel.txt
/bda/snaps/MarinLopezDaniel.txt
```



```
hadoop@master:~$ hdfs dfs -cp /bda/snaps/.snapshot/snap1/MarinLopezDaniel.t
xt /bda/snaps/MarinLopezDaniel.txt
hadoop@master:~$
```

- Crea una nueva instantánea de la carpeta llamada snap2.

```
hdfs dfs -createSnapshot /bda/snaps snap2
```



```
hadoop@master:~$ hdfs dfs -createSnapshot /bda/snaps snap2
Created snapshot /bda/snaps/.snapshot/snap2
hadoop@master:~$
```

- Muestra el contenido de la carpeta /bda/snaps así como de sus snapshots.

```
hdfs dfs -ls /bda/snaps
hdfs dfs -ls /bda/snaps/.snapshot/snap1
hdfs dfs -ls /bda/snaps/.snapshot/snap2
```



```
hadoop@master:~$ hdfs dfs -ls /bda/snaps
Found 1 items
-rw-r--r-- 1 hadoop supergroup 0 2024-12-19 11:09 /bda/snaps/Mar
inLopezDaniel.txt
hadoop@master:~$ hdfs dfs -ls /bda/snaps/.snapshot/snap1
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2024-12-19 10:47 /bda/snaps/.sn
apshot/snap1/MarinLopezDaniel.txt
-rw-r--r-- 1 hadoop supergroup 0 2024-12-19 10:52 /bda/snaps/.sn
apshot/snap1/MarinLopezDaniel_snapshot.txt
hadoop@master:~$ hdfs dfs -ls /bda/snaps/.snapshot/snap2
Found 1 items
-rw-r--r-- 1 hadoop supergroup 0 2024-12-19 11:09 /bda/snaps/.sn
apshot/snap2/MarinLopezDaniel.txt
hadoop@master:~$
```

4. HDFS por dentro (RA5075.4 / CE.4e y CE.4f)

- Accede al archivo de configuración hdfs-site.xml y averigua la carpeta donde se almacena el namenode.

```
sudo nano hdfs-site.xml
```



```
<property>
  <name>dfs.namenode.name.dir</name>
  <value>/opt/hadoop/hadoop_data/hdfs/namenode</value>
</property>
```

- Muestra los archivos que contiene la carpeta current dentro del namenode

```
cd /opt/hadoop/hadoop_data/hdfs/namenode/current
ls
```



```
hadoop@master:/opt/hadoop/hadoop_data/hdfs/namenode/current$ ls
edits_00000000000000000001-0000000000000000009
edits_00000000000000000010-0000000000000000010
edits_00000000000000000011-0000000000000000035
edits_00000000000000000036-0000000000000000036
edits_00000000000000000037-0000000000000000041
edits_inprogress_00000000000000000042
fsimage_00000000000000000035
fsimage_00000000000000000035.md5
fsimage_00000000000000000041
fsimage_00000000000000000041.md5
seen_txid
VERSION
```

- Comprueba el id del archivo *VERSION*.

```
cat VERSION
```



```
hadoop@master:/opt/hadoop/hadoop_data/hdfs/namenode/current$ cat VERSION
#Mon Dec 23 15:16:26 UTC 2024
namespaceID=854661652
clusterID=CID-1ef0a261-95dc-43e9-b377-20c6192e6124
cTime=1734535190628
storageType=NAME_NODE
blockpoolID=BP-1639053288-127.0.1.1-1734535190628
layoutVersion=-67
```

- En los siguientes pasos vamos a realizar un checkpoint manual para sincronizar el sistema de ficheros. Para ello entramos en modo safe con el comando `hdfs dfsadmin -safemode enter`, para impedir que se trabaje con el sistema de ficheros mientras lanzamos el checkpoint.

```
hdfs dfsadmin -safemode enter
```



```
hadoop@master:/opt/hadoop/hadoop_data/hdfs/namenode/current$ hdfs dfsadmin
-safemode enter
Safe mode is ON
```

- Comprueba mediante la interfaz gráfica que el modo seguro está activo (Safe mode is ON). Ahora realiza el checkpoint con el comando `hdfs dfsadmin -saveNamespace`

```
hdfs dfsadmin -saveNamespace
```



```
hadoop@master:/opt/hadoop/hadoop_data/hdfs/namenode/current$ hdfs dfsadmin
-saveNamespace
Save namespace successful
```

- Vuelve a entrar al modo normal (saliendo del modo seguro mediante `hdfs dfsadmin -safemode leave`)

```
hdfs dfsadmin -safemode leave
```



```
hadoop@master:/opt/hadoop/hadoop_data/hdfs/namenode/current$ hdfs dfsadmin
-safemode leave
Safe mode is OFF
```

- Accede a la carpeta del namenode y comprueba que los fsimage del namenode son iguales.

```
ls -l fsimage*
```



```
hadoop@master:/opt/hadoop/hadoop_data/hdfs/namenode/current$ ls -l fsimage*
-rw-rw-r-- 1 hadoop hadoop 1589 dic 23 15:16 fsimage_00000000000000000041
-rw-rw-r-- 1 hadoop hadoop  62 dic 23 15:16 fsimage_00000000000000000041.md
5
-rw-rw-r-- 1 hadoop hadoop 1589 dic 23 15:32 fsimage_00000000000000000043
-rw-rw-r-- 1 hadoop hadoop  62 dic 23 15:32 fsimage_00000000000000000043.md
5
```