

Trabajo Nro. 3: HADOOP HDFS

Dentro del directorio /home/hadoop/scripts se crea el archivo landing.sh donde se descarga starwars.csv en /landing, se lo coloca en hdfs usando -put /ingest y luego se elimina el starwars.csv de /landing. El script es el siguiente:

```
#!/bin/bash
```

```
# Descargar el .csv con -wget
```

```
wget -P /home/hadoop/landing
```

```
https://raw.githubusercontent.com/fpineyro/homework-0/master/starwars.csv
```

```
# Se coloca el .csv en hdfs, dentro de la carpeta de /ingest
```

```
hdfs dfs -put /home/hadoop/landing/starwars.csv /ingest
```

```
# Borrar el .csv de landing
```

```
rm /home/hadoop/landing/starwars.csv
```

Luego se cambian los permisos del script landing.sh para que se pueda ejecutar. En este caso solamente se le pusieron todos los permisos al propietario, pero al grupo y otros solamente se les dio permiso de lectura.

```
hadoop@401bec58e4c6:~/scripts$ chmod 744 landing.sh
hadoop@401bec58e4c6:~/scripts$ ls -l
total 24
-rw-rw-r-- 1 hadoop hadoop 670 Feb 28 2022 derby.log
-rwxrwxr-x 1 hadoop hadoop 258 Feb 28 2022 ingest.sh
-rwxr--r-- 1 hadoop hadoop 325 Apr 18 10:32 landing.sh
```

Luego de ejecutar el script con

```
./ landing.sh
```

Se comprueba que se ejecutaron todas las acciones correctamente.

Finalmente para comprobar si starwars.csv efectivamente está en /ingest, se usa el comando:

```
hdfs dfs -ls /ingest
```

```
hadoop@401bec58e4c6:~/hadoop$ hdfs dfs -ls /ingest
Found 2 items
-rw-r--r-- 1 hadoop supergroup 154869 2024-04-18 10:32 /ingest/starwars.csv
-rw-r--r-- 1 hadoop supergroup 125981363 2024-04-17 14:14 /ingest/yellow_trip_data_2021-01.csv
hadoop@401bec58e4c6:~/hadoop$
```

Esto también se puede comprobar accediendo a la UI de Hadoop en un navegador.

Browse Directory

Show entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	hadoop	supergroup	151.24 KB	Apr 18 10:32	1	128 MB	starwars.csv	
<input type="checkbox"/>	-rw-r--r--	hadoop	supergroup	120.15 MB	Apr 17 14:14	1	128 MB	yellow_tripdata_2021-01.csv	

Showing 1 to 2 of 2 entries