

Edvai - Ejercicio Ingest - Gabriel Rey

Ingresar al directorio /home/hadoop/scripts

```
File Edit View Search Terminal Help
hadoop@0485e86b7577:/$ cd /home/hadoop/scripts/
hadoop@0485e86b7577:~/scripts$
```

Crear un script llamado landing.sh que baje el archivo <https://github.com/fpineyro/homework-0/blob/master/starwars.csv> alclear directorio temporal /home/hadoop/landing y luego lo envíe al directorio de Hadoop file system (HDFS) /ingest. Antes de finalizar el script que borre el archivo starwars.csv del directorio temporal /home/hadoop/landing

```
File Edit View Search Terminal Help
hadoop@0485e86b7577:~/scripts$ ls -l
total 196
-rw-rw-r-- 1 hadoop hadoop 670 Feb 28 2022 derby.log
-rwxrwxr-x 1 hadoop hadoop 258 Feb 28 2022 ingest.sh
drwxr-xr-x 2 hadoop hadoop 6 Feb 9 2022 spark-warehouse
-rwxrwxrwx 1 hadoop hadoop 1089 May 9 2022 start-services.sh
-rw-r--r-- 1 hadoop hadoop 181737 Mar 22 10:45 starwars.csv
-rwxrwxrwx 1 hadoop hadoop 1058 May 9 2022 transformation.py
hadoop@0485e86b7577:~/scripts$ cat > landing.sh

wget -P /home/hadoop/landing https://github.com/fpineyro/homework-0/blob/master/starwars.csv

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

rm -f /home/hadoop/landing/*.sh

hadoop@0485e86b7577:~/scripts$ ls -l
total 200
-rw-rw-r-- 1 hadoop hadoop 670 Feb 28 2022 derby.log
-rwxrwxr-x 1 hadoop hadoop 258 Feb 28 2022 ingest.sh
-rw-r--r-- 1 hadoop hadoop 184 Mar 28 09:01 landing.sh
drwxr-xr-x 2 hadoop hadoop 6 Feb 9 2022 spark-warehouse
-rwxrwxrwx 1 hadoop hadoop 1089 May 9 2022 start-services.sh
-rw-r--r-- 1 hadoop hadoop 181737 Mar 22 10:45 starwars.csv
-rwxrwxrwx 1 hadoop hadoop 1058 May 9 2022 transformation.py
hadoop@0485e86b7577:~/scripts$
```

Cambiar permisos para que el script pueda ser ejecutado

```
File Edit View Search Terminal Help
hadoop@0485e86b7577:~/scripts$ chmod 777 landing.sh
hadoop@0485e86b7577:~/scripts$ ls -l
total 200
-rw-rw-r-- 1 hadoop hadoop 670 Feb 28 2022 derby.log
-rwxrwxr-x 1 hadoop hadoop 258 Feb 28 2022 ingest.sh
-rwxrwxrwx 1 hadoop hadoop 184 Mar 28 09:01 landing.sh
drwxr-xr-x 2 hadoop hadoop 6 Feb 9 2022 spark-warehouse
-rwxrwxrwx 1 hadoop hadoop 1089 May 9 2022 start-services.sh
-rw-r--r-- 1 hadoop hadoop 181737 Mar 22 10:45 starwars.csv
-rwxrwxrwx 1 hadoop hadoop 1058 May 9 2022 transformation.py
hadoop@0485e86b7577:~/scripts$
```

Ejecutar el script para que baje el archivo starwars.csv de Github y lo envíe al directorio /ingest de HDFS

```
File Edit View Search Terminal Help
hadoop@0485e86b7577:~/scripts$ sh landing.sh
--2023-03-28 09:14:12-- https://github.com/fpineyro/homework-0/blob/master/starwars.csv
Resolving github.com (github.com)... 140.82.113.4
Connecting to github.com (github.com)|140.82.113.4|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [text/html]
Saving to: '/home/hadoop/landing/starwars.csv'

starwars.csv          [ <=>

2023-03-28 09:14:13 (27.8 MB/s) - '/home/hadoop/landing/starwars.csv' saved [181758]

hadoop@0485e86b7577:~/scripts$
```

Verificar que el archivo se encuentre en el directorio /ingest de HDFS

```
File Edit View Search Terminal Help
hadoop@0485e86b7577:~/scripts$ hdfs dfs -ls /ingest/
Found 2 items
-rw-r--r--  1 hadoop supergroup      181758 2023-03-28 09:14 /ingest/starwars.csv
-rw-r--r--  1 hadoop supergroup 125981363 2022-05-09 17:58 /ingest/yellow_tripdata_2021-01.csv
hadoop@0485e86b7577:~/scripts$
```