

Practica Ingest

1. Ingresar a la consola hadoop y luego cambiarse de usuario a hadoop:

```
hadoop@615cf53bef6c: ~  
crbdata@RubenBageta:~$ docker exec -it edvai_hadoop bash  
root@615cf53bef6c:/# su - hadoop  
hadoop@615cf53bef6c:~$ |
```

2. Ingresar al directorio /home/hadoop/scripts

```
hadoop@615cf53bef6c:~$ cd scripts  
hadoop@615cf53bef6c:~/scripts$ |
```

3. Crear un script llamado landing.sh que baje el archivo <https://github.com/fpineyro/homework-0/blob/master/starwars.csv> al 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

```
hadoop@615cf53bef6c: ~/scri  
GNU nano 4.8 landing.sh  
# Punto 3  
  
# Descarga datos desde el repositorio de Fede al directorio /home/hadoop/landing  
wget -P /home/hadoop/landing https://github.com/fpineyro/homework-0/blob/master/starwars.csv  
  
# Lleva el archivo a HDFS al directorio /ingest  
hdfs dfs -put /home/hadoop/landing/starwars.csv /ingest  
  
# Borra el archivo starwars.csv del directorio /home/hadoop/landing/  
rm /home/hadoop/landing/starwars.csv
```

4. Cambiar permisos para que el script pueda ser ejecutado

```
hadoop@615cf53bef6c:~/scripts$ chmod 555 landing.sh
hadoop@615cf53bef6c:~/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
-r-xr-xr-x 1 hadoop hadoop 398 Sep 17 16:57 landing.sh
drwxr-xr-x 2 hadoop hadoop 4096 Feb 9 2022 spark-warehouse
-rwxrwxrwx 1 hadoop hadoop 1089 May 9 2022 start-services.sh
-rwxrwxrwx 1 hadoop hadoop 1058 May 9 2022 transformation.py
hadoop@615cf53bef6c:~/scripts$ |
```

SOLO ASIGNE PERMISOS DE LECTURA Y EJECUCIÓN

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

```
hadoop@615cf53bef6c:~/scripts$ ls
derby.log ingest.sh landing.sh spark-warehouse start-services.sh transformation.py
hadoop@615cf53bef6c:~/scripts$ ./landing.sh
--2025-09-17 17:26:21-- https://github.com/fpineyro/homework-0/blob/master/starwars.csv
Resolving github.com (github.com)... 4.228.31.150
Connecting to github.com (github.com)|4.228.31.150|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [text/html]
Saving to: '/home/hadoop/landing/starwars.csv'

starwars.csv [ <=>

2025-09-17 17:26:22 (825 KB/s) - '/home/hadoop/landing/starwars.csv' saved [191382]

hadoop@615cf53bef6c:~/scripts$ |
```

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

```
hadoop@615cf53bef6c:~/scripts$ hdfs dfs -ls /ingest
Found 1 items
-rw-r--r-- 1 hadoop supergroup 191382 2025-09-17 17:26 /ingest/starwars.csv
hadoop@615cf53bef6c:~/scripts$ |
```