

# Practica Ingest

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

```
mrmatux@compujuego:~$ docker start_edvai_hadoop
docker: unknown command: docker start_edvai_hadoop

Run 'docker --help' for more information
mrmatux@compujuego:~$ docker start edvai_hadoop
edvai_hadoop
mrmatux@compujuego:~$ docker exec -it edvai_hadoop bash
root@9af7590fd1b8:/# su hadoop
hadoop@9af7590fd1b8:/$ |
```

2. Ingresar al directorio /home/hadoop/scripts

```
hadoop@9af7590fd1b8:/$ cd /home/hadoop/scripts
hadoop@9af7590fd1b8:~/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

## 1er forma para crear scripts:

```
hadoop@9af7590fd1b8:~/scripts$ nano landing.sh|
```

```
hadoop@9af7590fd1b8: x + v
GNU nano 4.8 landing.sh
#Descargando archivo
wget -P /home/hadoop/landing https://github.com/fpineyro/homework-0/blob/master/starwars
#Moviendo archivo descargado a carpeta /ingest
hdfs dfs -put /home/hadoop/landing/starwars.csv /ingest
```

Script:

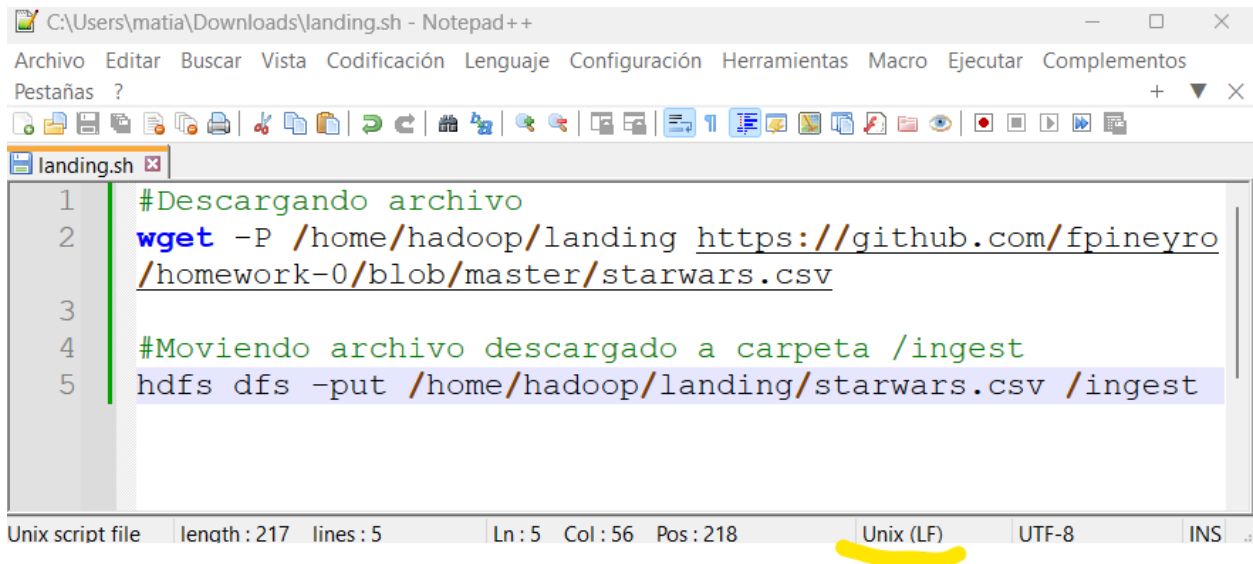
#Descargando archivo

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

#Moviendo archivo descargado a carpeta /ingest

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

## 2da forma para crear scripts:



The screenshot shows the Notepad++ editor with the file 'landing.sh' open. The script content is as follows:

```
1 #Descargando archivo
2 wget -P /home/hadoop/landing https://github.com/fpineyro/
3 /homework-0/blob/master/starwars.csv
4 #Moviendo archivo descargado a carpeta /ingest
5 hdfs dfs -put /home/hadoop/landing/starwars.csv /ingest
```

The status bar at the bottom indicates 'Unix script file', 'length: 217', 'lines: 5', 'Ln: 5 Col: 56 Pos: 218', and 'Unix (LF)' is highlighted with a yellow circle, indicating the line ending format.

Lo creamos con Notepad ++ para ponerle el formato Unix (LF) por la linea de corte. Probamos primero con el notepad común pero al ejecutar el script en hadoop venía con caracteres raros ó no interpretaba la linea y daba errores. Tratamos de arreglarlos pero siempre surgía otro...entonces mejor crear el archivo

preparador para Linux (LF) y listo.

Copia del archivo de windows a Linux:

```
mrmatux@compujuego:~$ cp /mnt/c/Users/matia/Downloads/landing.sh /home/mrmatux/landing.sh
mrmatux@compujuego:~$ ls -rtl landing.sh
-rwxr-xr-x 1 mrmatux mrmatux 226 Sep 17 17:21 landing.sh
```

De linux a la imagen edvai\_hadoop:

```
mrmatux@compujuego:~$ docker cp landing.sh edvai_hadoop:/home/hadoop/scripts
Successfully copied 2.05kB to edvai_hadoop:/home/hadoop/scripts
```

4. Cambiar permisos para que el script pueda ser ejecutado

chmod u+x para que sea ejecutado por el user, propietario.

```
hadoop@9af7590fd1b8:~/scripts$ chmod u+x landing.sh
hadoop@9af7590fd1b8:~/scripts$ ls -rtl landing.sh
-rwxrw-r-- 1 hadoop hadoop 219 Sep 17 13:52 landing.sh
hadoop@9af7590fd1b8:~/scripts$ |
```

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

Tanto con el script creado con NANO como con el enviado desde windows en formato Unix(LF), la linea de corte para Linux, ambos ejecutaron OK.

```
hadoop@9af7590fd1b8:~/scripts$ sh landing.sh
--2025-09-17 14:04:13-- 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          [ <=> ] 186.85K  --.-KB/s   in 0.1s

2025-09-17 14:04:13 (1.40 MB/s) - '/home/hadoop/landing/starwars.csv' saved [191335]

hadoop@9af7590fd1b8:~/scripts$
```

```
hadoop@9af7590fd1b8:~/scripts$ sh landing.sh
--2025-09-17 14:50:53-- 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.1'

starwars.csv.1        [ <=> ] 186.87K  --.-KB/s   in 0.1s

2025-09-17 14:50:54 (1.36 MB/s) - '/home/hadoop/landing/starwars.csv.1' saved [191351]

hadoop@9af7590fd1b8:~/scripts$
```

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

## Browse Directory

/ingest

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	186.85 KB	Sep 17 14:04	1	128 MB	starwars.csv	<div><div></div><div></div></div>

```
hadoop@9af7590fd1b8:~/scripts$ hdfs dfs -ls /ingest/star*
-rw-r--r--  1 hadoop supergroup    191335 2025-09-17 14:04 /ingest/starwars.csv
hadoop@9af7590fd1b8:~/scripts$
```