

Practica Ingest

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

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

Run 'docker --help' for more information
mrmatus@compujuego:~$ docker start edvai_hadoop
edvai_hadoop
mrmatus@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

```
hadoop@9af7590fd1b8:~/scripts$ ls -l landing.sh
-rw-rw-r-- 1 hadoop hadoop 1554 Sep 16 16:15 landing.sh
```

script:

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

#Descargando archivo

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

#Chequeamos si la descarga termina con '0' que seria OK.

\$? es el codigo de retorno de la operacion anterior

```
if [ $? -eq 0 ]; then
```

```
    descarga_estado=$?
```

```
    fecha=$(date +"%Y-%m-%d_%H-%M-%S")
```

```
    echo "Descarga exitosa! ✓"
```

```
    echo "El codigo de la descarga fue: $descarga_estado" >  
    /home/hadoop/landing/landing_ok_$fecha.txt
```

```
else
```

```
    descarga_error=$?
```

```
    echo "Descarga fallida, revisar el codigo de retorno: $descarga_error ✗"
```

```
    echo "Descarga fallida, revisar el codigo de retorno: $descarga_error" >  
    /home/hadoop/landing/$landing
```

```
fi
```

```
if hdfs dfs -test -e /ingest/starwars.csv; then
```

```
    echo "El archivo ya existe. Verificar."
```

```
else
```

#Moviendo archivo descargado a carpeta /ingest

```
move=$(hdfs dfs -put /home/hadoop/landing/starwars.csv /ingest 2>&1)
```

```
if [ $? -eq 0 ]; then
```

```
    landing=$(ls /home/hadoop/landing | grep 'landing_ok_*.txt' | sort -k6,7 |  
tail -1 | awk '{print $NF}')
```

```
    echo "La copia se hizo con exito a la carpeta /ingest"
```

```
    echo "El resultado del move fue: $move" >>  
    /home/hadoop/landing/$landing
```

#Borrando archivo temporal descargado en /landing

rm /home/hadoop/landing/starwars.csv

borrado=\$?

echo "El resultado del borrado del temporal fue: \$borrado" >>

/home/hadoop/landing/\$landing

#Muestro resultado

ruta_ingest=\$(hdfs dfs -ls -R /ingest/* 2>&1)

echo \$ruta_ingest

else

echo "Deber haber algun problema.Codigo: \$?"

fi

fi

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

#Chequeamos si la descarga termina con '0' que seria OK.
# $? es el codigo de retorno de la operacion anterior
if [ $? -eq 0 ]; then
    descarga_estado=$?
    fecha=$(date +%Y-%m-%d_%H-%M-%S)
    echo "Descarga exitosa!      ^|^e"
    echo "El codigo de la descarga fue: $descarga_estado" > /home/hadoop/landing/landing_ok_$fecha.txt
else
    descarga_error=$?
    echo "Descarga fallida, revisar el codigo de retorno: $descarga_error      ^|^l"
    echo "Descarga fallida, revisar el codigo de retorno: $descarga_error" > /home/hadoop/landing/$landing
fi

#Chequear si el archivo existe antes de moverlo
if hdfs dfs -test -e /ingest/starwars.csv; then

echo "El archivo ya existe. Verificar."

else
    #Moviendo archivo descargado a carpeta /ingest

    move=$(hdfs dfs -put /home/hadoop/landing/starwars.csv /ingest 2>&1)
    if [ $? -eq 0 ]; then
        landing=$(ls /home/hadoop/landing | grep 'landing_ok_*.txt' | sort -k6,7 | tail -1 | awk '{print $NF}')
        echo "La copia se hizo con exito a la carpeta /ingest"
        echo "El resultado del move fue: $move" >> /home/hadoop/landing/$landing
    fi

    #Borrando archivo temporal descargado en /landing
    rm /home/hadoop/landing/starwars.csv
    borrado=$?
    echo "El resultado del borrado del temporal fue: $borrado" >> /home/hadoop/landing/$landing
fi
```

```

borrado=$?
echo "El resultado del borrado del temporal fue: $borrado" >> /home/hadoop/landing/$landing
#Muestro resultado
ruta_ingest=$(hdfs dfs -ls -R /ingest/* 2>&1)
echo $ruta_ingest
else
echo "Deber haber algun problema.Codigo: $?"
fi
fi
fi

```

4. Cambiar permisos para que el script pueda ser ejecutado

Estado actual:

```

hadoop@9af7590fd1b8:~/scripts$ ls -l landing.sh
-rw-rw-r-- 1 hadoop hadoop 1602 Sep 16 16:39 landing.sh

```

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

```

hadoop@9af7590fd1b8:~/scripts$ chmod u+x landing.sh
hadoop@9af7590fd1b8:~/scripts$ ls -l landing.sh
-rwxrw-r-- 1 hadoop hadoop 1602 Sep 16 16:39 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

```
hadoop@9af7590fd1b8:~/scripts$ sh landing.sh
--2025-09-16 16:43:23-- 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.3'

starwars.csv.3          [ <=>          ] 186.77K  --.-KB/s    in 0.1s

2025-09-16 16:43:23 (1.89 MB/s) - '/home/hadoop/landing/starwars.csv.3' saved [191252]

Descarga exitosa! 
La copia se hizo con exito a la carpeta /ingest
-rw-r--r-- 1 hadoop supergroup 191187 2025-09-16 16:43 /ingest/starwars.csv -rw-r--r-- 1 hadoop supergroup 183716 2025-09-15 22:30 /ingest/yellow_tripdata_2021-01.csv
hadoop@9af7590fd1b8:~/scripts$ |
```

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

Browse Directory

/ingest Go!

Show 25 entries Search:

| | Permission | Owner | Group | Size | Last Modified | Replication | Block Size | Name | |
|--------------------------|------------|--------|------------|-----------|---------------|-------------|------------|-----------------------------|--|
| <input type="checkbox"/> | -rw-r--r-- | hadoop | supergroup | 186.72 KB | Sep 16 16:15 | 1 | 128 MB | starwars.csv | |
| <input type="checkbox"/> | -rw-r--r-- | hadoop | supergroup | 179.41 KB | Sep 15 22:30 | 1 | 128 MB | yellow_tripdata_2021-01.csv | |

Showing 1 to 2 of 2 entries

Previous 1 Next

Hadoop, 2020.

```
hadoop@9af7590fd1b8:~/scripts$ hdfs dfs -ls -R /ingest/star*
-rw-r--r--  1 hadoop supergroup    191187 2025-09-16 16:43 /ingest/starwars.csv
hadoop@9af7590fd1b8:~/scripts$
```