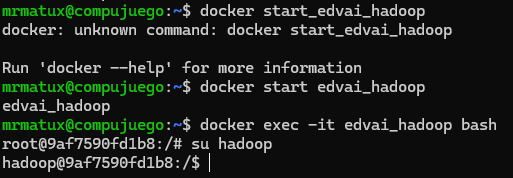
Practica Ingest

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



2. Ingresar al directorio /home/hadoop/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



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)

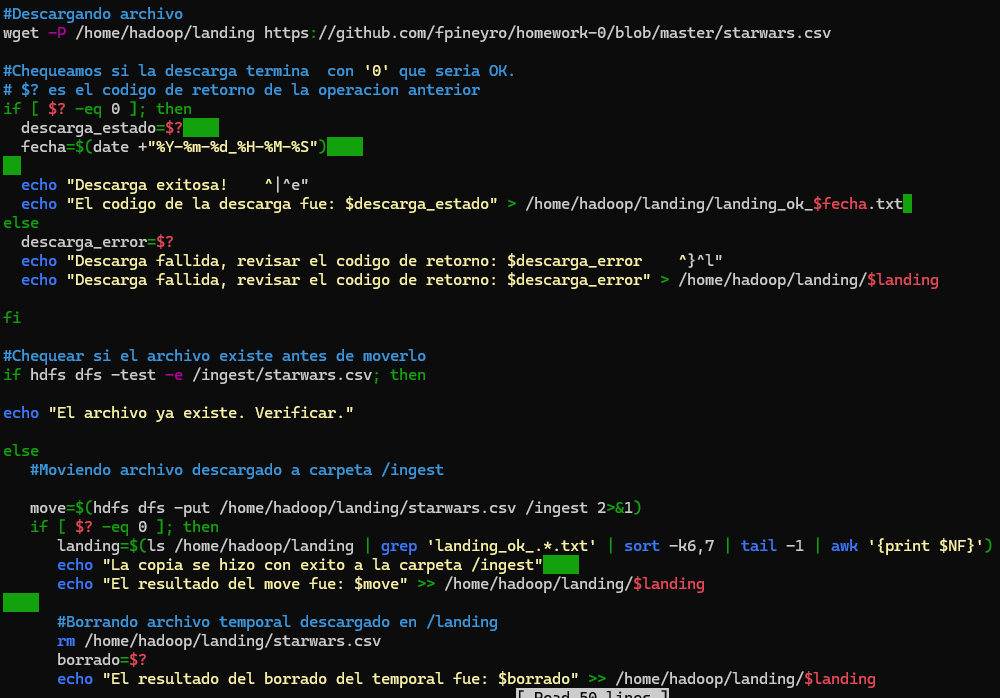
echo $ruta\_ingest

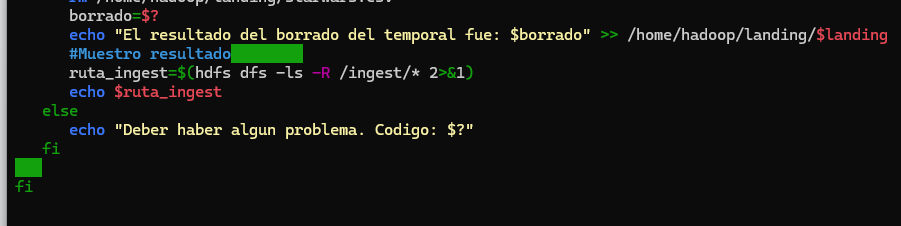
else

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

fi

fi



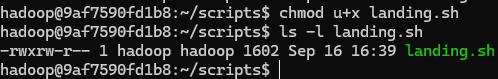


4. Cambiar permisos para que el script pueda ser ejecutado

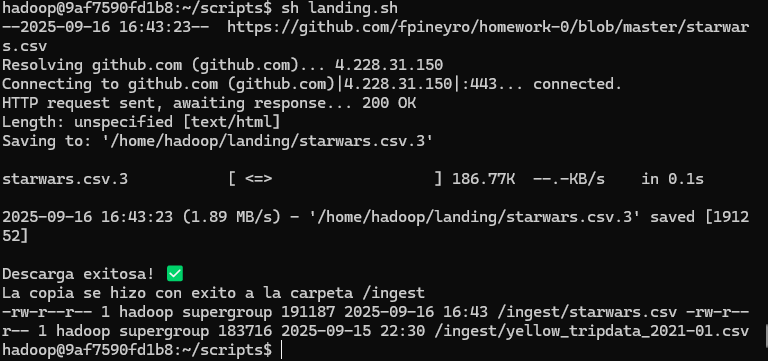
Estado actual:



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



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



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

