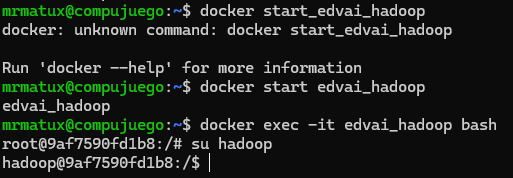
Practica Ingest

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



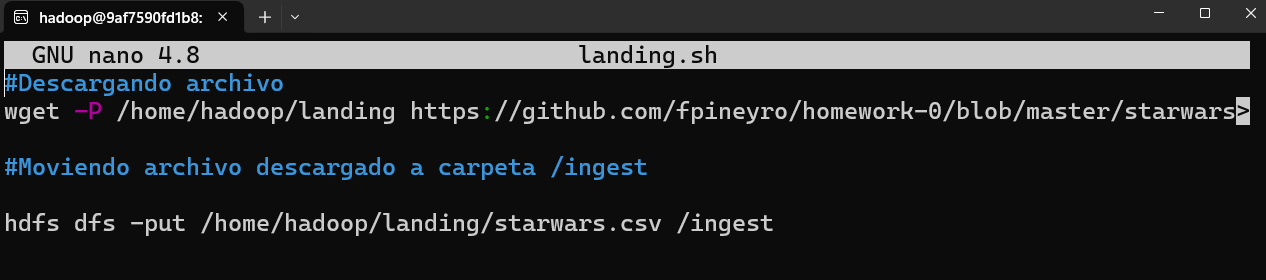
2. Ingresar al directorio /home/hadoop/scripts



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

**1er forma para crear scripts:**



****

Script:

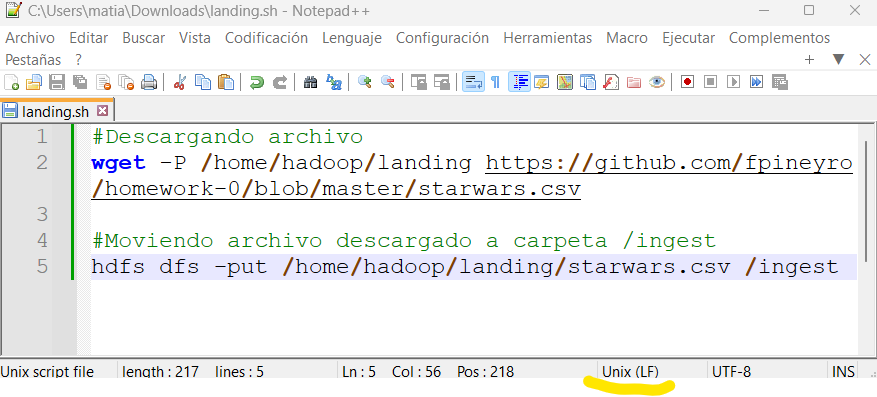
#Descargando archivo

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:**

****

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:



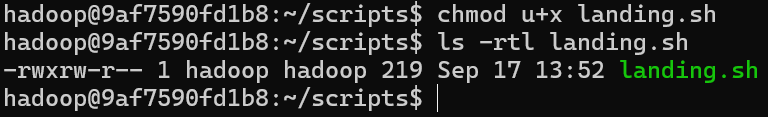


De linux a la imagen edvai\_hadoop:



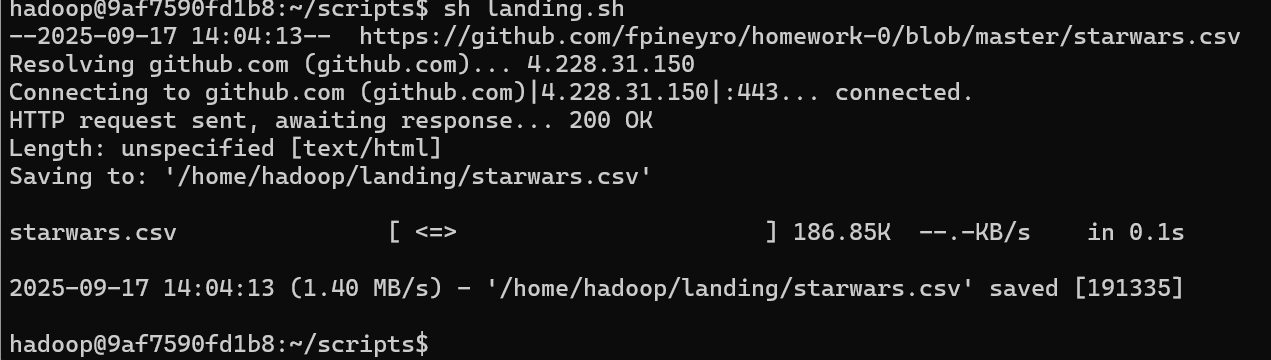
4. Cambiar permisos para que el script pueda ser ejecutado

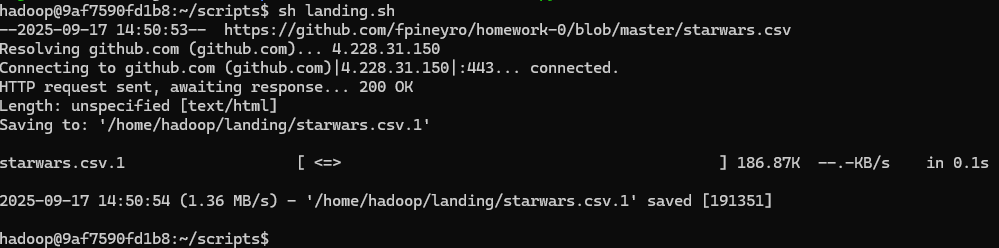
chmod u+x para que sea ejecutado por el user,propietario.



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

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.





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

