

## Práctica Ingest – Consola Hadoop (HDFS)

1)

Inicio la consola de hadoop desde la carpeta data engineering.

2)

```
cd /home/hadoop/scripts
```

El comando cd me permite ingresar un repositorio al cual dirigirme. Me dirijo al directorio scripts para crear uno nuevo.

3)

```
cat > landing.sh
```

El comando cat me permite crear un archivo de texto (en este caso en formato .sh).

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

El wget me permite descargar datos de un almacenamiento externo. Es importante especificar la ruta en la cual los queremos guardar, y el https del repositorio en el que se encuentra. En este caso los guardaremos en el directorio temporal /home/hadoop/landing.

El siguiente paso es ingresarlo en mi sistema de archivos Hadoop File System (HDFS) y para eso vamos a ejecutar los siguientes comandos:

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

siendo el /ingest el directorio de HDFS dentro del directorio principal (/) pedido en el ejercicio.

c) `cd /home/hadoop/landing`  
`rm starwars.csv`  
`ctrl + D`

Ejecuto un cd para dirigirme al directorio temporal, y consecutivamente un rm para borrar el archivo csv. Ctrl + D para terminar de escribir.

```
ls -l
```

Observo mi archivo landing.sh dentro del directorio scripts.

4)

```
chmod 777 landing.sh
```

para cambiar los permisos del script creado a “todos los permisos”.

```
ls -l
```

para verificar que los permisos se hayan cambiado correctamente.

5)

```
sh landing.sh
```

para ejecutar el script.

6)

```
hdfs dfs -ls /ingest
```

para verificar que el archivo se encuentre en la carpeta ingest.