



Módulo Profesional: Big Data Aplicado

Ingestión de datos con NiFi

Ejercicio 2: Basic Ingestion (II)

Tarea: obtener los datos de una URL y almacenarlos en el sistema de archivos.

Ejemplo:

- Utilice el siguiente conjunto de datos abiertos (Nombres de calles de Valencia):

<http://mapas.valencia.es/lanzadera/opendata/vias/csv>

- Se guarda en el sistema de archivos.



Utilice los siguientes procesadores:

- InvocarHTTP
- PutFile

Cambia “SCHEDULING” cuando invoca a HTTP URL

Ejercicio 2: Basic Ingestion (II)

Pasos previos:

Si queremos borrar el contenido de /tmp/out ejecutamos:

```
rm /tmp/out/*
```

NiFi

1.- Añadir Procesador **InvokeHTTP** y configuramos

2.- Añadir Procesador **PutFile** y configuramos

Terminal:

3.- Ejecutamos `ls /tmp/out` → comprobamos que está el fichero

4.- Ejecutamos el siguiente comando con el nombre del file:

```
cat /tmp/out/5c51fac4-65a3-4a97-8303-57d0f7fe6a7c
```

Ejercicio 2: Basic Ingestion (II)

Una vez ejecutado, podemos ejecutar el siguiente comando:

```
cat /tmp/out/5c51fac4-65a3-4a97-8303-57d0f7fe6a7c
```

```
nifi@8a4e9d01d347:/opt/nifi/nifi-current$ cat /tmp/out/5c51fac4-65a3-4a97-8303-57d0f7fe6a7c
codtipovia;codvia;codviacatastro;nomoficial;traducnooficial
PG;38850;3217;RUSSAFA;RUSSAFA
C/;38940;2450;JOS♦ GROLLO;JOSE GROLLO
C/;39000;3345;JOS♦ MARIA FUSTER;JOSE MARIA FUSTER
C/;39010;2628;JOSE MARIA HARO (MAGISTRAT);JOS♦ MARIA HARO (MAGISTRAT)
PL;39060;3143;BADAJOZ;BADAJOZ
C/;37680;1077;JACINT LABAILA;JACINTO LABAILA
C/;37710;1078;JACOMART;JACOMART
C/;37740;1079;PERIODISTA JOS♦ OMBUENA;PERIODISTE JOSE OMBUENA
PL;38460;2402;JOAQUIM DUALDE;JOAQUIN DUALDE
C/;38520;1101;JOAQUIM NAVARRO;JOAQUIN NAVARRO
C/;38640;1105;JOSEP AGUIRRE;JOSE AGUIRRE
C/;36720;1054;IFAC;IFACH
PL;36840;1057;ESGL♦SIA DEL FORN D'ALCEDO;IGLESIA HORNO DE ALCEDO
C/;36990;1062;IMPERTIN♦NCIES;IMPERTINENCIAS
C/;37080;1064;ENGINYER AUB♦N;INGENIERO AUBAN
C/;37110;1065;ENGINYER DICENTA;INGENIERO DICENTA
C/;37140;2310;ENGINYER JOS♦ SIRERA;INGENIERO JOSE SIRERA
```

Ejercicio 3: Basic Ingestion (II)

Comprobamos la salida ls /tmp/out.

```
nifi@8a4e9d01d347:/opt/nifi/nifi-current$ ls /tmp/out
DLP-357e0d92-09b4-4d9d-99e4-da6bfc1c5e7e-0.txt  DLP-357e0d92-09b4-4d9d-99e4-da6bfc1c5e7e-2.txt
DLP-357e0d92-09b4-4d9d-99e4-da6bfc1c5e7e-1.txt  DLP-357e0d92-09b4-4d9d-99e4-da6bfc1c5e7e-3.txt
nifi@8a4e9d01d347:/opt/nifi/nifi-current$
```

