



Ejemplo 3 - Filtrado de datos

En este caso, vamos a coger los datos de **ventas** (.csv adjunto), el cual tiene la siguiente estructura:

pdi_sales_small.csv

```
ProductID;Date;Zip;Units;Revenue;Country
725;1/15/1999;41540;1;115.5;Germany
787;6/6/2002;41540;1;314.9;Germany
788;6/6/2002;41540;1;314.9;Germany
```

Utilizando *Nifi*, vamos a crear un nuevo fichero CSV que contenga únicamente los datos de Francia que han realizado más de una venta.

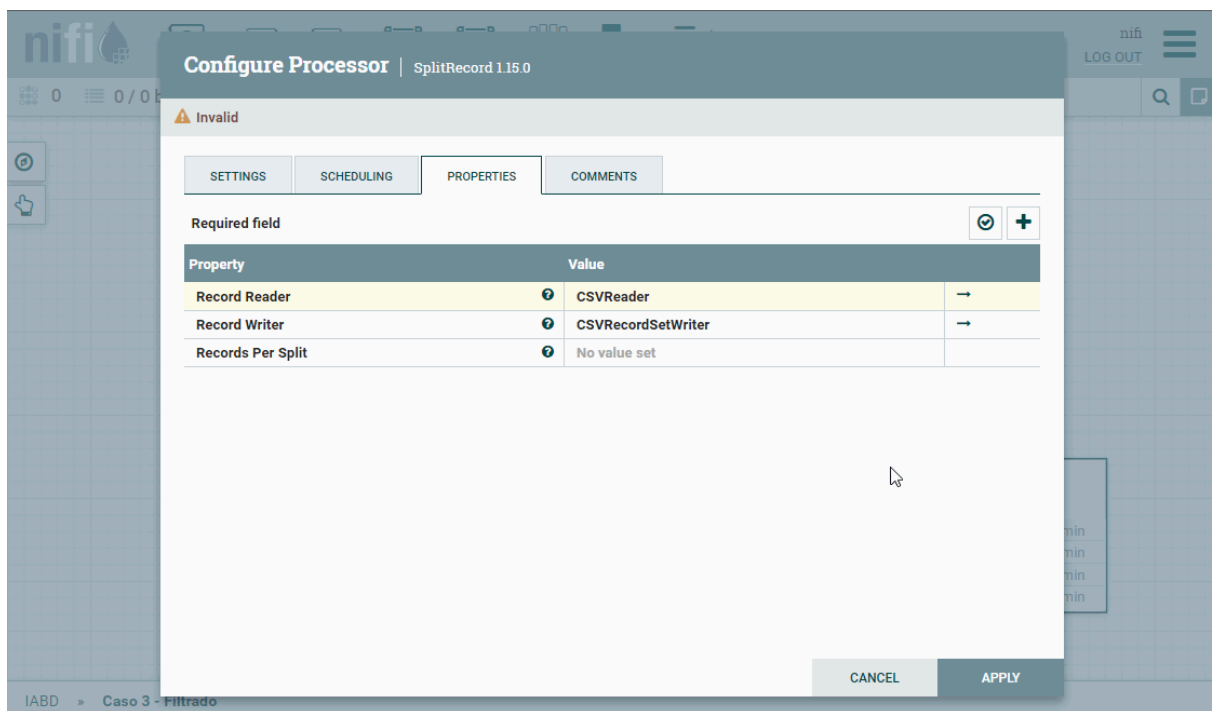
Para ello, tendremos que leer el fichero haciendo uso del procesador *GetFile*, separar cada fila en un FF mediante *SplitRecord*, filtrar los datos usando el procesador *QueryRecord* y finalmente los almacenaremos en disco gracias al procesador *PutFile*.

Lectura y división

1. Así pues, comenzamos leyendo el fichero con el procesador *GetFile*. En este caso vamos a dejar la opción *keep source file* a *true* para que no lo elimine.
2. Mediante el procesador *SplitRecord*, vamos a separar cada fila del CSV a un FF. Para ello, primero hemos de crear un *RecordReader* y un *RecordWriter* para que sepa interactuar con el CSV (*Nifi* ya tiene varios implementados que podemos utilizar). Así pues:
 - En el *Record Reader*, seleccionamos *Create new service*, y elegimos *CVSReader*.

- A su vez, en el *Record Writer* elegimos *CSVRecordSetWriter*.

3. Para configurar estos servicios, pulsaremos sobre la flecha, y veremos la pantalla de configuración. Para cada uno de ellos, tendremos otros tres iconos: la rueda para configurar, el rayo para activar/desactivar el servicio y la papelera para eliminarlo. Así, pues, tras comprobar los valores de *CSVReader* y *CSVWriter* (indicamos el ; como separador de campos tanto para la lectura como la escritura de CSV en el campo *value separator* y marcamos como *true* que el archivo contiene la primera fila con encabezados (*treat first line as header*)), pulsamos sobre el rayo para activar ambos servicios.



Configuración y activación de Split Record

4. Finalmente, en el campo *Records per Split* le indicamos **1** para que coloque cada fila en un FF.

Filtrado de FF

5. En este paso, mediante el procesador *QueryRecord* vamos a ejecutar una consulta SQL contra el FF. El resultado del nuevo FF será el resultado de la consulta. En nuestro caso, como hemos comentado antes, vamos a quedarnos con las ventas de más de una unidad realizadas en Francia. Igual que antes, configuramos los mismos *Record Reader* y *Record Writer*. Además, hemos de poner la propiedad *Include Zero Record FlowFiles* a *false* para que no vuelva a enrutar los FF que no cumplan la consulta. Finalmente, añadimos una nueva propiedad para indicar la consulta. En nuestro caso la hemos denominado *FranciaMayor1* y en el contenido ponemos la consulta:

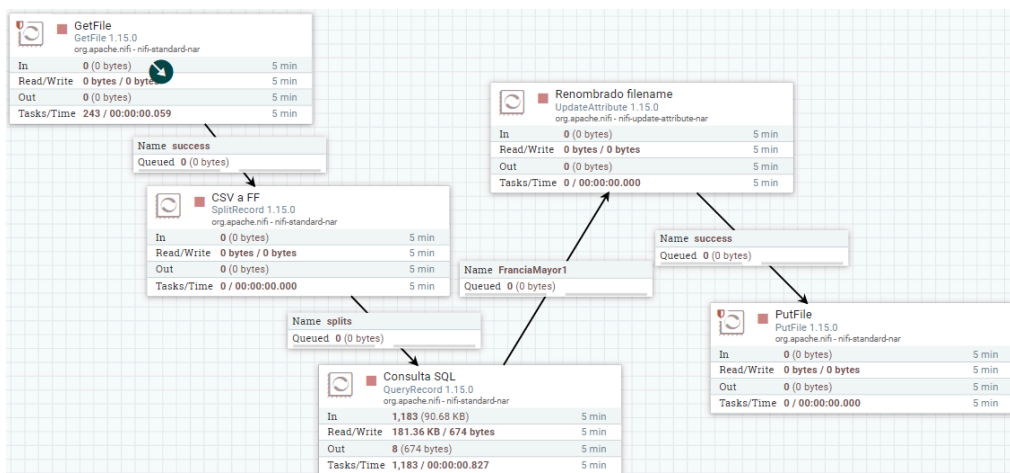
```
select * from FLOWFILE where Country = 'France' and Units > 1
```

Nota:

También podríamos haber filtrado los campos para recuperar menos contenido con una consulta similar a *Select ProductID, Date from FlowFile* Con este procesador podemos filtrar, hacer agrupaciones, cálculos, ..., del mismo modo que lo hacemos con SQL.

6. Finalmente, igual que hicimos en el caso 1, vamos a cambiarle el nombre a cada FF para generar un archivo por cada resultado mediante *UpdateAttribute* y persistimos los datos con *PutFile*.

El resultado del flujo de datos será similar a:



Flujo completo del caso 3

NOTA: Para finalizar, deberás contemplar la finalización de los procesadores *SplitRecord* y *QueryRecord* (original y failure). Puedes utilizar una conexión a un procesador *LogMessage*.

