

Seminario sobre GeoKettle.

Nuevas herramientas relacionadas con Pentaho Data Integration (PDI): GIS plug-in

Actividad 9

Enunciado del problema.

Disponemos de un **archivo de texto json** que contiene datos alfanuméricos entre los que se encuentran las coordenadas de los elementos puntuales.

```
1 [
2 ["Ademuz,Valencia, España", 40.0617367, -1.2865257, 1,1],
3 ["Ador,Valencia, España", 38.9194075, -0.2242898, 2,1],
4 ["Agullent,Valencia, España", 38.8213153, -0.5489346, 3,1],
5 ["Aielo de Malferit,Valencia, España", 38.8766144, -0.5921101, 4,1],
6 ["Aielo de Rugat,Valencia, España", 38.8816613, -0.3430091, 5,1],
7 ["Alaquàs,Valencia, España", 39.4570792, -0.4610681, 6,1],
8 ["Albaida,Valencia, España", 38.8370089, -0.5156197, 7,1],
9 ["Albal,Valencia, España", 39.3970929, -0.4121969, 8,1],
10 ["Albalat de la Ribera,Valencia, España", 39.2010974, -0.3863965, 9,1],
11 ["Albalat dels Sorells,Valencia, España", 39.542188, -0.3473705, 10,1],
```

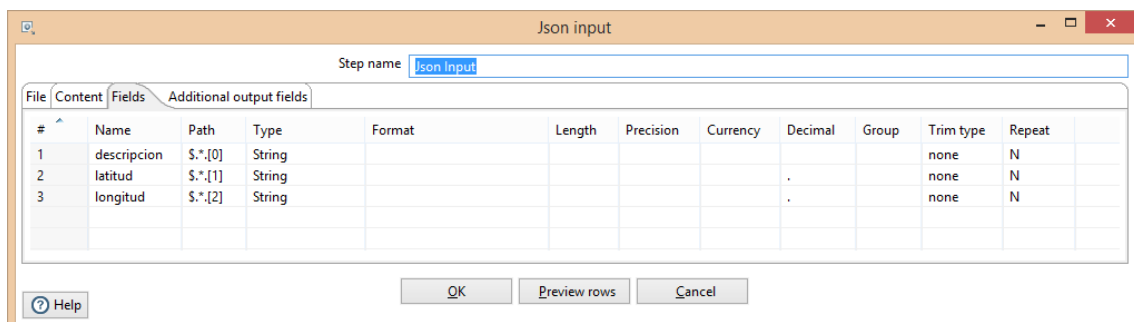
Se desea **convertir esta información en un archivo CSV** que preserve los atributos almacenados en el documento JSON.

Solución basada en PDI:

Crear una transformación que abra un fichero JSON y seleccione los atributos (por orden de aparición en el archivo JSON). Seguidamente se almacene en un archivo de texto separado por comas. También se podría almacenar en un fichero Shapefile si se usa el Plug-in PDI GIS, para lo cual la transformación se hace más compleja.



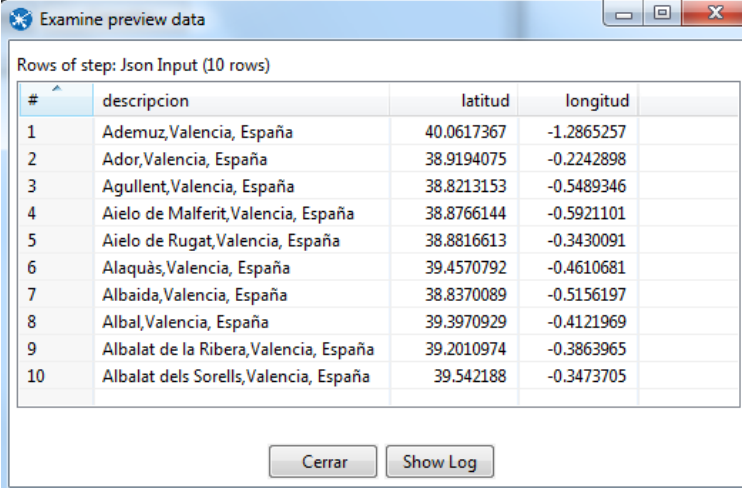
La primera etapa de la transformación, además de conectar con el fichero JSON de los datos iniciales, hay que indicar en **Fields** los patrones de búsqueda de los atributos en el documento JSON. En el ejemplo que se ha mostrado, donde como primer atributo aparece el nombre de un centro de salud de la Comunidad Valenciana, el segundo la latitud y el tercero la longitud. Para este caso con este tipo de estructura de archivo JSON el mapeado de atributos y elementos del JSON sería el que se muestra en la figura siguiente.



Hemos definido el Path en sintaxis JSONPath, como \$.*.[0] para recuperar el elemento número 0 (descripción) de cada uno de los hijos. Para más información del uso de las expresiones JSONPath podemos ver el siguiente enlace:

<http://goessner.net/articles/JsonPath/>

La Previsualización de los 10 primeros elementos del documento se muestra en la siguiente figura:

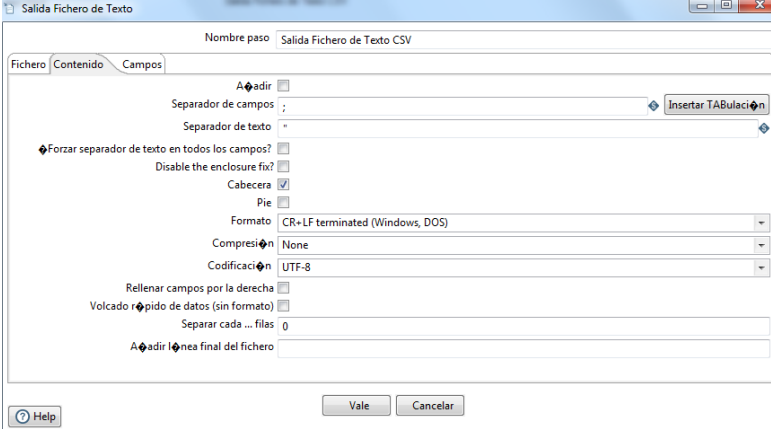


Rows of step: Json Input (10 rows)

#	descripcion	latitud	longitud
1	Ademuz,Valencia, España	40.0617367	-1.2865257
2	Ador,Valencia, España	38.9194075	-0.2242898
3	Agullent,Valencia, España	38.8213153	-0.5489346
4	Aielo de Malferit,Valencia, España	38.8766144	-0.5921101
5	Aielo de Rugat,Valencia, España	38.8816613	-0.3430091
6	Alaquàs,Valencia, España	39.4570792	-0.4610681
7	Albaida,Valencia, España	38.8370089	-0.5156197
8	Albal,Valencia, España	39.3970929	-0.4121969
9	Albalat de la Ribera,Valencia, España	39.2010974	-0.3863965
10	Albalat dels Sorells,Valencia, España	39.542188	-0.3473705

Cerrar Show Log

En el último paso de la transformación, almacenar en archivo CSV, lo único que hay que establecer es el lugar donde almacenar el archivo, cómo separar las columnas y los atributos o columnas a guardar.



Nombre paso: Salida Fichero de Texto CSV

Fichero Contenido Campos

Añadir

Separador de campos ; Insertar TABulación

Separador de texto -

Forzar separador de texto en todos los campos?

Disable the enclosure fix?

Cabecera

Pie

Formato CR+LF terminated (Windows, DOS)

Compresión None

Codificación UTF-8

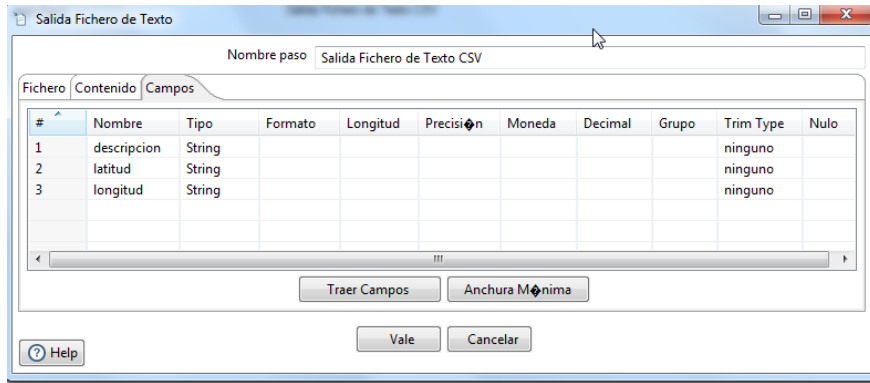
Rellenar campos por la derecha

Volcado rápido de datos (sin formato)

Separar cada ... filas 0

Añadir línea final del fichero

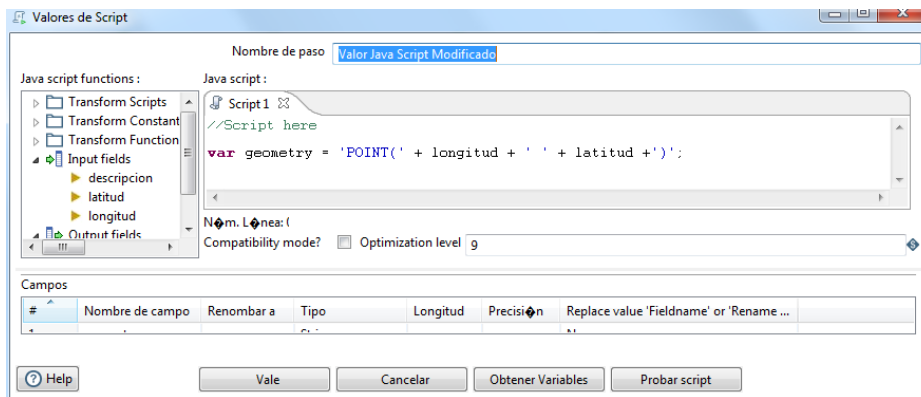
Help Vale Cancelar



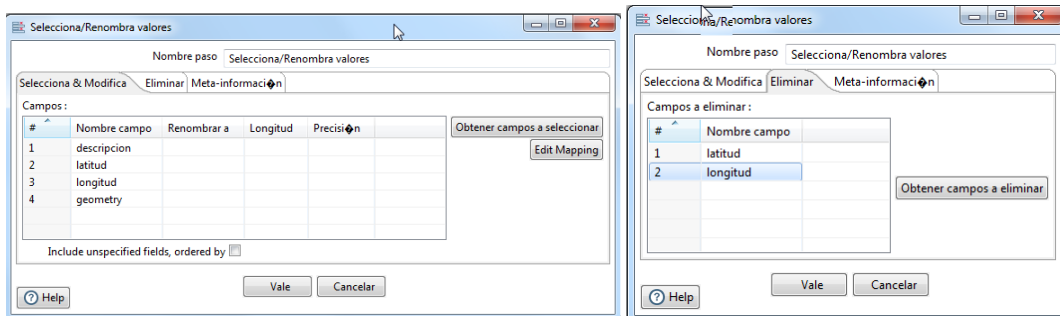
Si se usa el complemento PDI GIS plugin, la transformación a utilizar tendría los siguientes elementos.

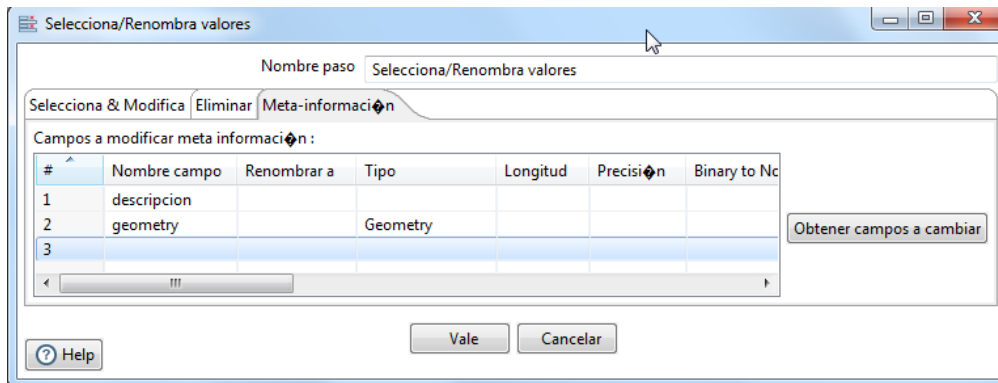


El objetivo de la primera etapa es construir una variable en JavaScript de tipo String con la sintaxis de una geometría de tipo punto a partir de las coordenadas antes leídas.



En la etapa de Selecciona, borra y renombra se han de obtener los atributos de entrada, los que se pretende eliminar (latitud y longitud como atributo) y cambiar el tipo de dato de la geometría construida.





La última etapa almacena los datos en un archivo Shapefile.

