

---

# Pentaho básico. Primera ETL

---

Big Data Aplicado

23/07/16 – IES Fernando Wirtz

Rafael Chamorro Maceiras

Fecha	Motivo del cambio
	Versión inicial

**Índice**

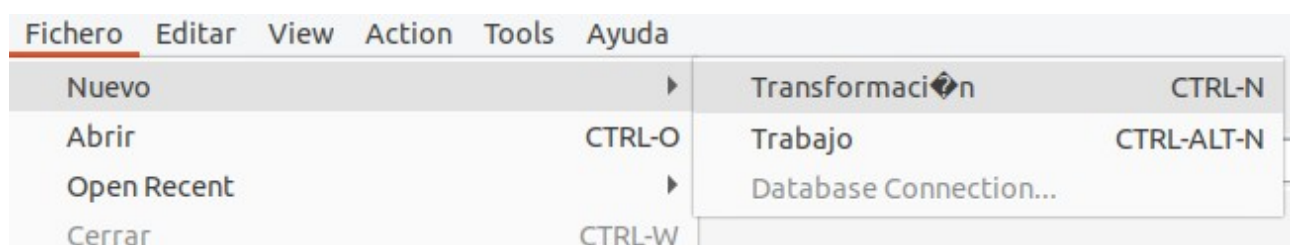
Ejemplo de creación de la primera ETL..... 2

## Ejemplo de creación de la primera ETL

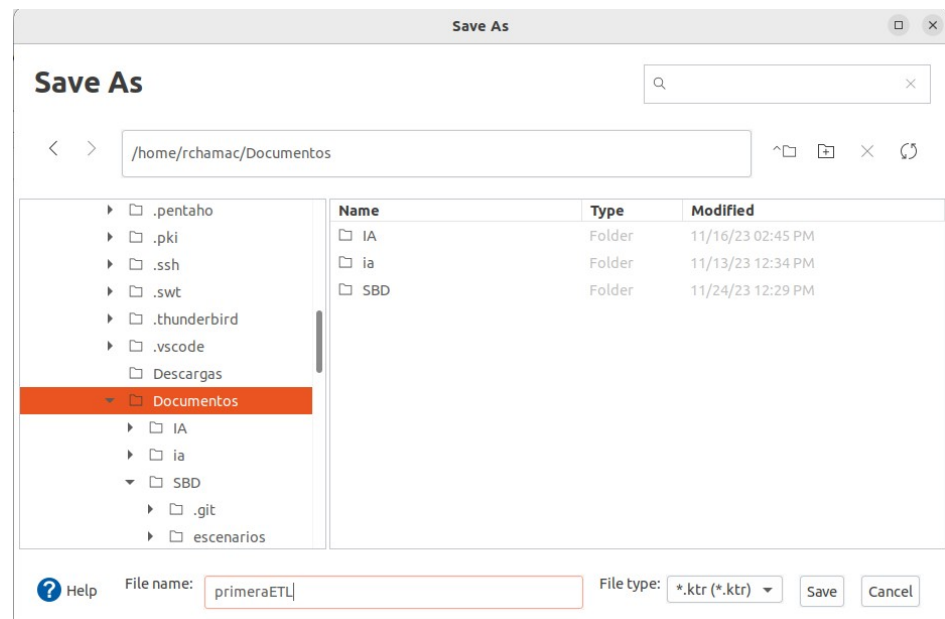
Partimos de la siguiente hoja de cálculo:

	A	B	C	D
1	Provincia	Concello	Con río	Tipo orografía
2	1	1	1	3
3	1	2	0	2
4	1	3	0	4
5	1	4	1	2
6	2	5	1	3
7	2	6	1	2
8	2	7	1	2
9	2	8	0	3
10	3	9	0	1
11	3	10	1	2
12	3	11	1	1
13	3	12	1	2
14	3	13	0	3
15	3	14	0	1
16	4	15	0	3
17	4	16	1	1
18	4	17	0	4
19	4	18	0	2
20	4	19	1	4

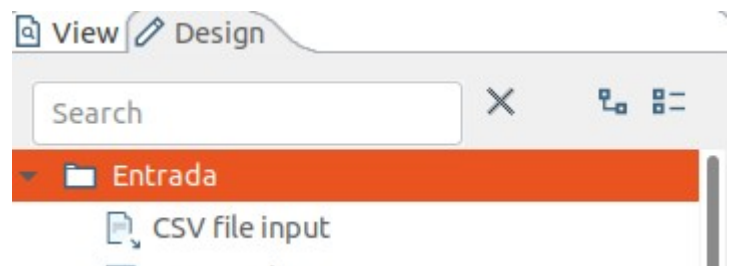
Creamos una nueva transformación



La guardamos con el nombre de primeraETL.ktr



En Diseño → Entrada buscamos la opción de Microsoft Excel Input



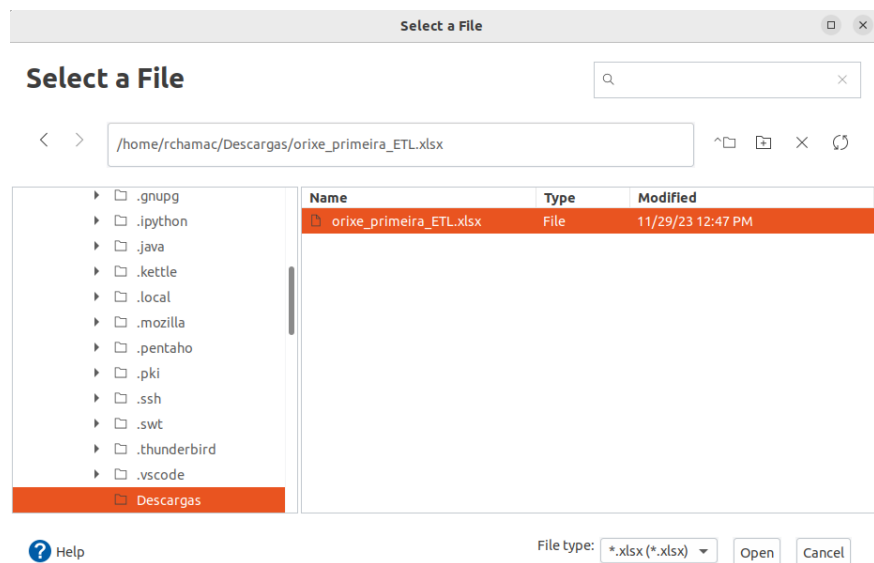
Con doble click accedemos a las opciones de la entrada.



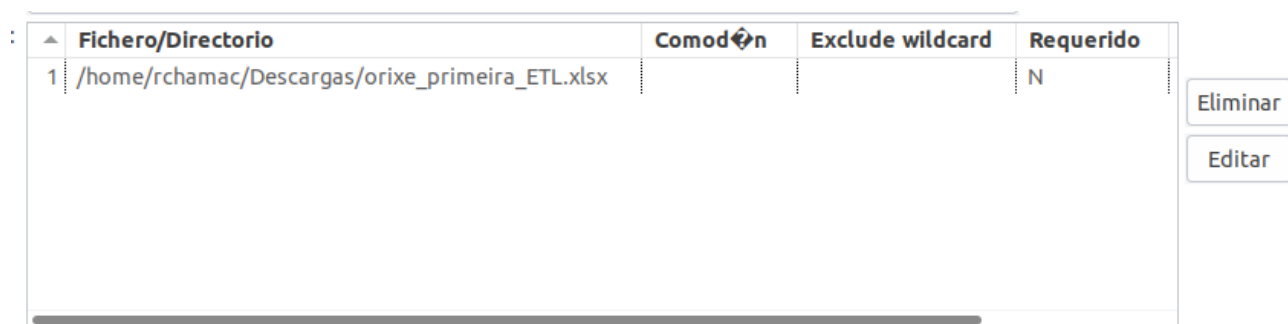
Podemos cambiar el nombre de este paso.  
Elegimos archivo de Excel 2007 XLSX.



Examinamos para localizar  
nuestro archivo de entrada  
excel.



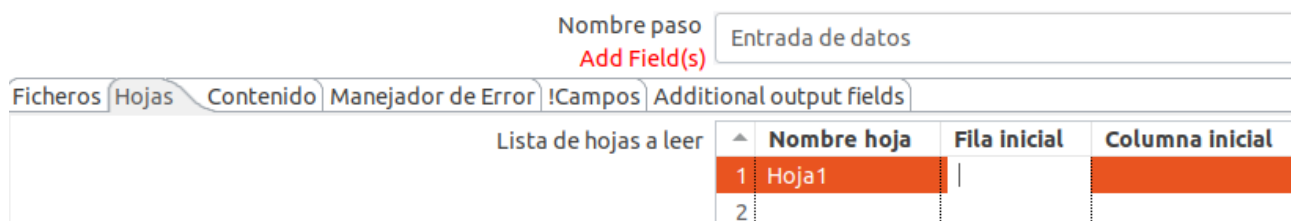
Añadimos el archivo:



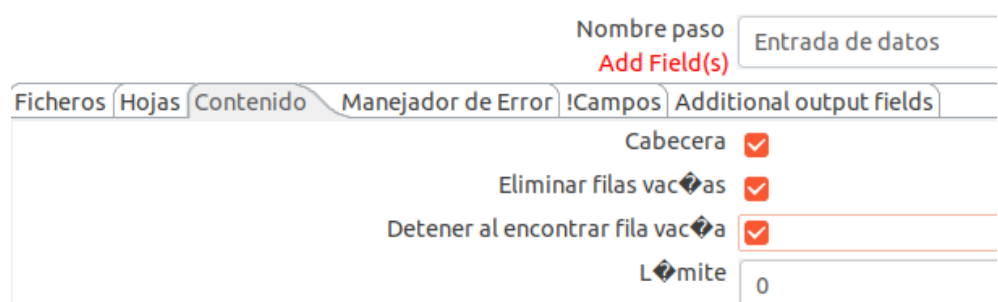
En la pestaña hojas vamos a obtener hojas y pasamos a la columna selección la/s que nos interese/n



Podemos seleccionar la fila y la columna de inicio

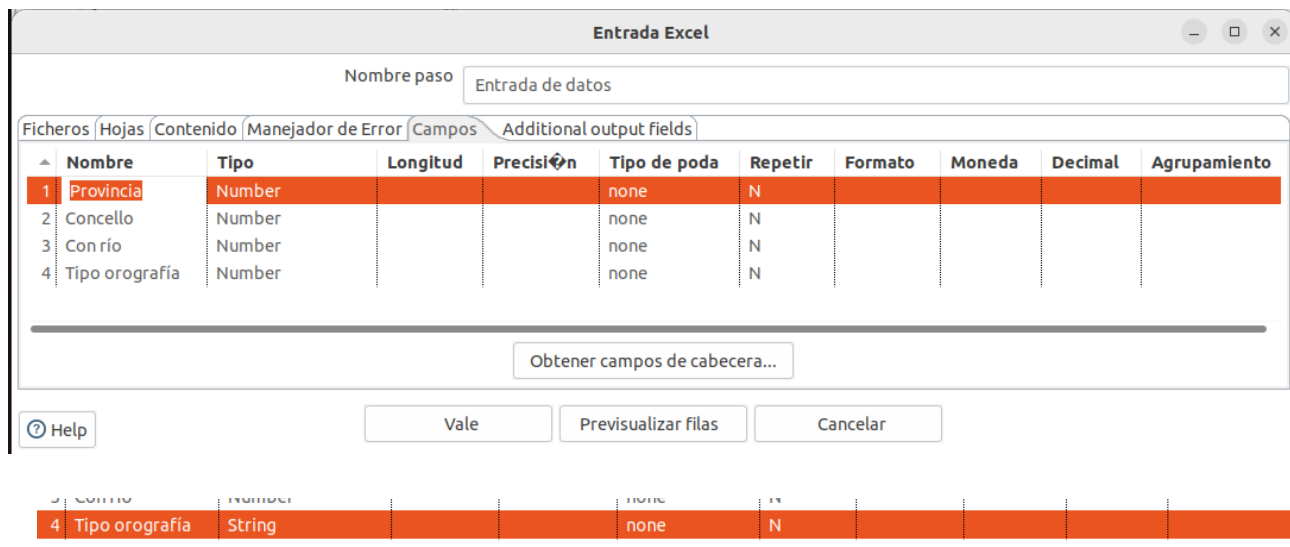


En la pestaña contenido seleccionamos que tiene una fila de encabezados, eliminar filas vacías y parar al encontrar una fila vacía



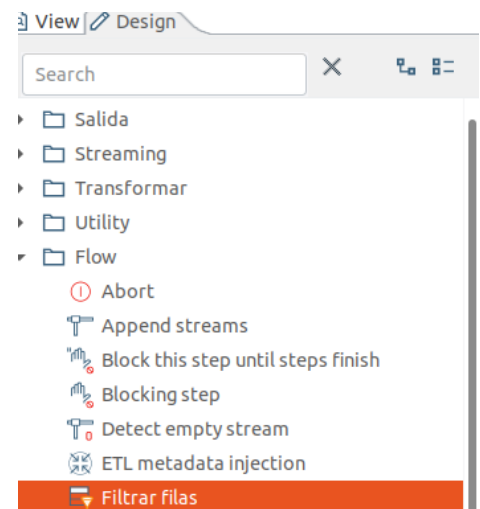
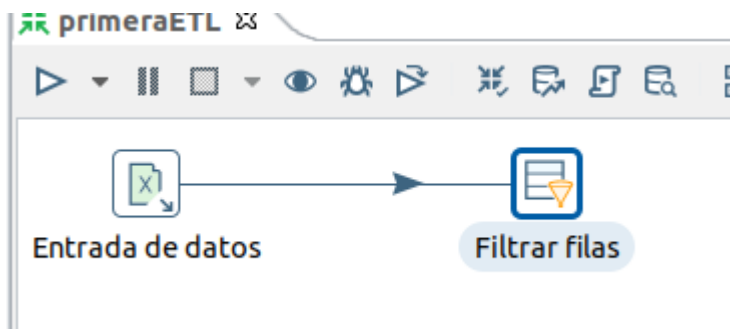
En la pestaña campos obtenemos los campos de cabecera y vemos que todos son de tipo Number.

El campo tipo de orografía lo cambiamos a String pensando en la transformación que queremos ejecutar

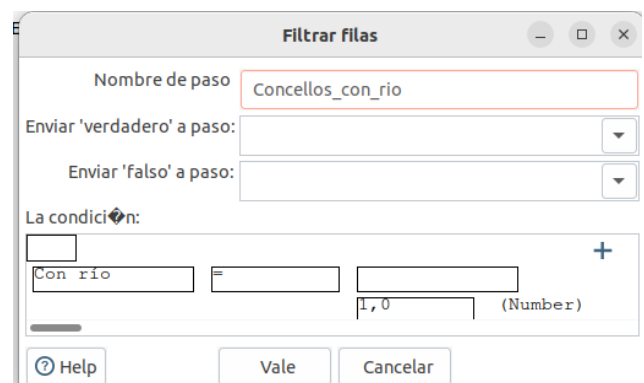


Filtrar filas:

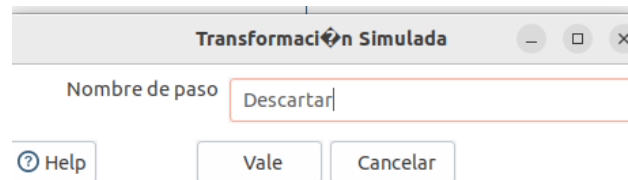
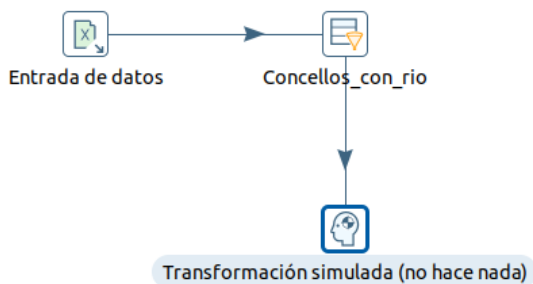
En diseño → flow buscamos la opción filtrar filas



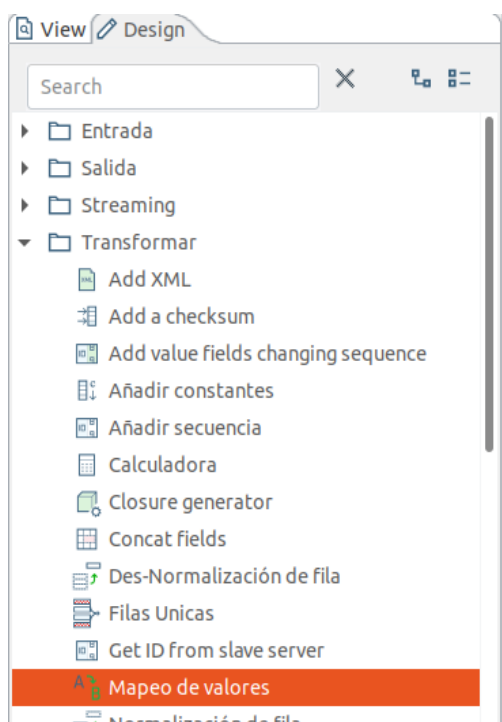
Doble click y modificamos el nombre de este paso y establecemos la condición, columna con río valor = 1



Para los datos que no cumplan la condición creamos una transformación simulada, la cuál no hace nada.

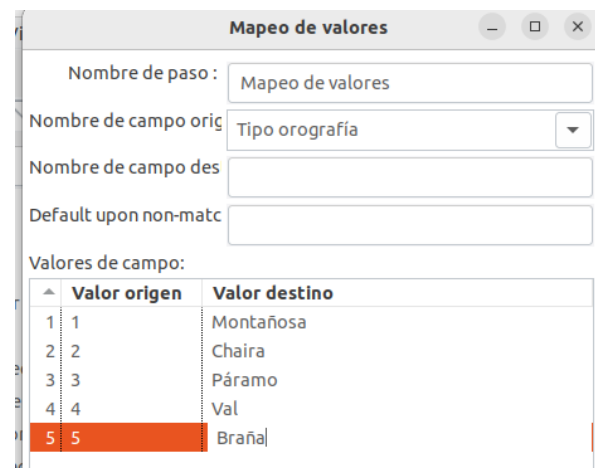


Los datos que cumplen la condición los mandamos a transformar, vamos a cambiar el valor numérico de la columna tipo de orografía por un String

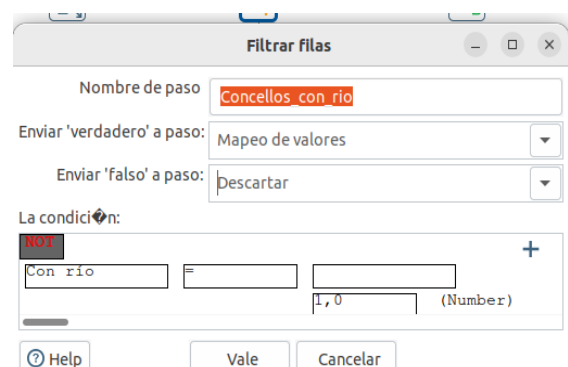


Diseño → Transformar → Mapeo de valores

Indicamos el valor de origen y el valor de destino de la transformación

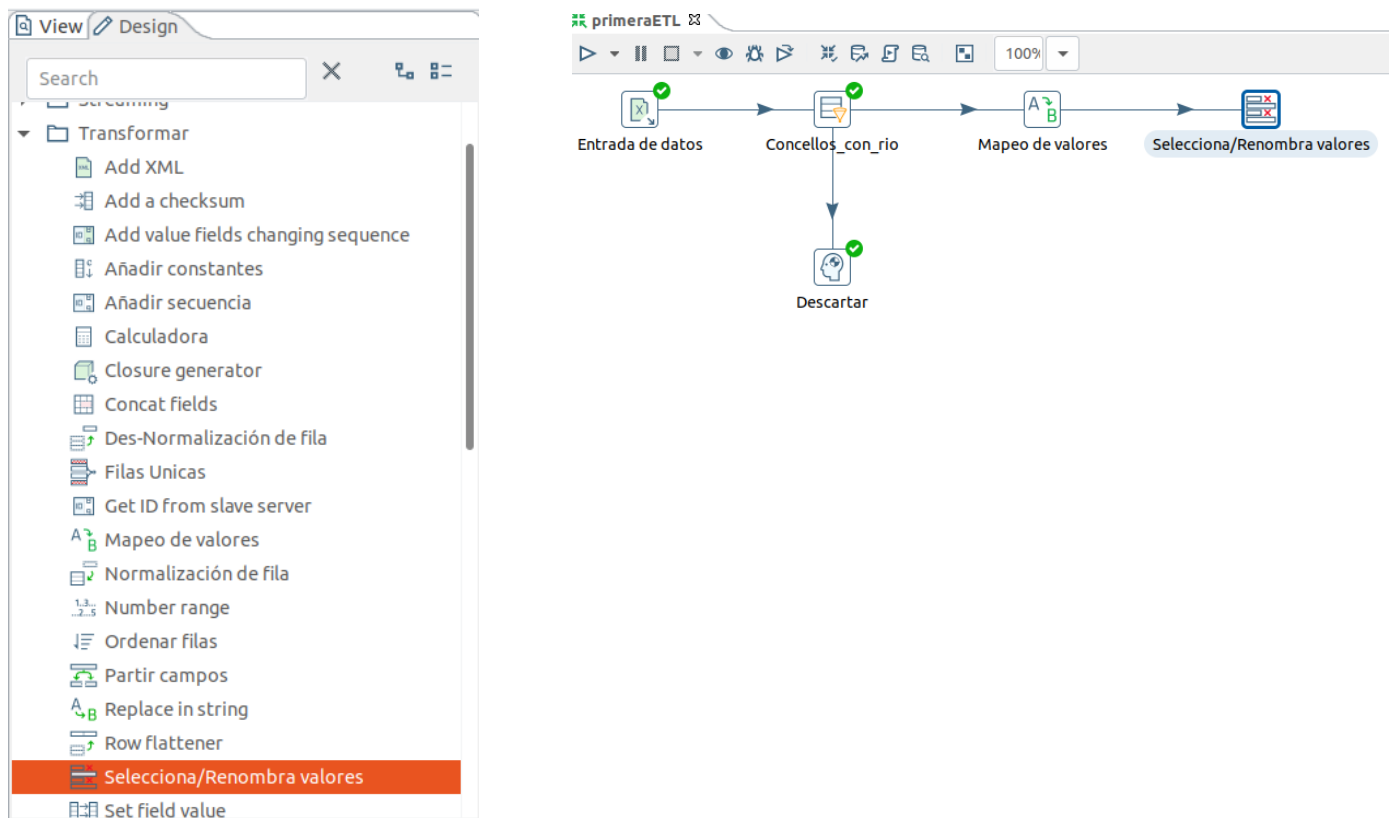


Ahora indicamos que si el valor de la condición es verdadero el valor pasa a mapeo y si es falso lo enviamos a descartar





Decidimos que valores continúan en el flujo y hacemos los cambios



Seleccionamos los campos que van a continuar en el flujo y también cambiamos el nombre de la columna 'concello'

Nombre paso

renombrar campos

Selecciona & Modifica

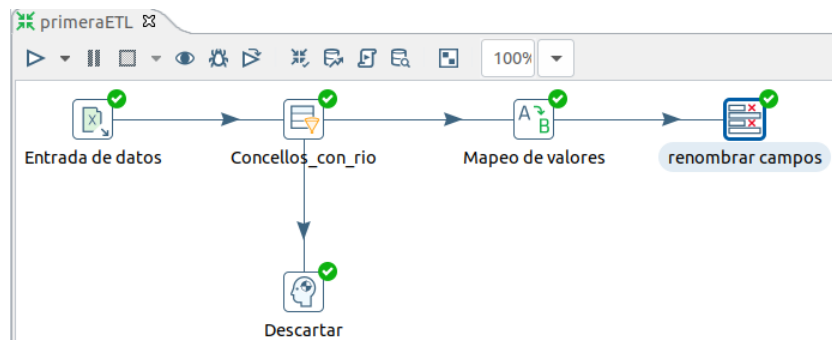
Eliminar

Meta-información

Campos :

	Nombre campo	Renombrar a	Longitud	Precisión
1	Concello	Código de municipio		
2	Tipo orografía			

## Ejecutamos la transformación



Una vez ejecutada la ETL podemos comprobar en cada paso que es lo que pasa con los datos:

En descartar nos encontramos los valores cuyo valor en la columna con rio es 0

**Execution Results**

Logging Execution History Step Metrics Performance Graph

• \${TransPreview.FirstRows.Label} • \${TransPreview.LastRows.Label}

	Provincia	Concello	Con río	Tipo orografía
1	1	2	0	2
2	1	3	0	4
3	2	8	0	3
4	3	9	0	1
5	3	13	0	3

En mapeo de valores tenemos las 4 columnas y vemos que ya se ha ejecutado la transformación en la columna tipo de orografía.

**Execution Results**

Logging Execution History Step Metrics Performance Graph

• \${TransPreview.FirstRows.Label} • \${TransPreview.LastRows.Label}

	Provincia	Concello	Con río	Tipo orografía
1	1	1	1	Páramo
2	1	4	1	Chaira
3	2	5	1	Páramo
4	2	6	1	Chaira
5	2	7	1	Chaira

En renombrar campos podemos observar que ya sólo nos quedamos con las dos columnas que nos interesaban, y vemos que el nombre de la columna concello ya se ha cambiado a Código municipio

**Execution Results**

Logging Execution History Step Metrics Performance Graph

• \${TransPreview.FirstRows.Label} • \${TransPreview.LastRows.Label}

	Código municipio	Tipo orografía
1	1	Páramo
2	4	Chaira
3	5	Páramo
4	6	Chaira
5	7	Chaira