
T2.1: Pentaho básico. A primeira ETL

Big Data Aplicado

30/11/23 – IES Fernando Wirtz

David Fernández Reboredo

Índice

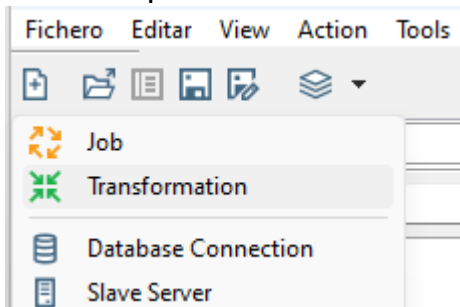
Pentaho Básico..... 3

 Creación da primeira ETL.....3

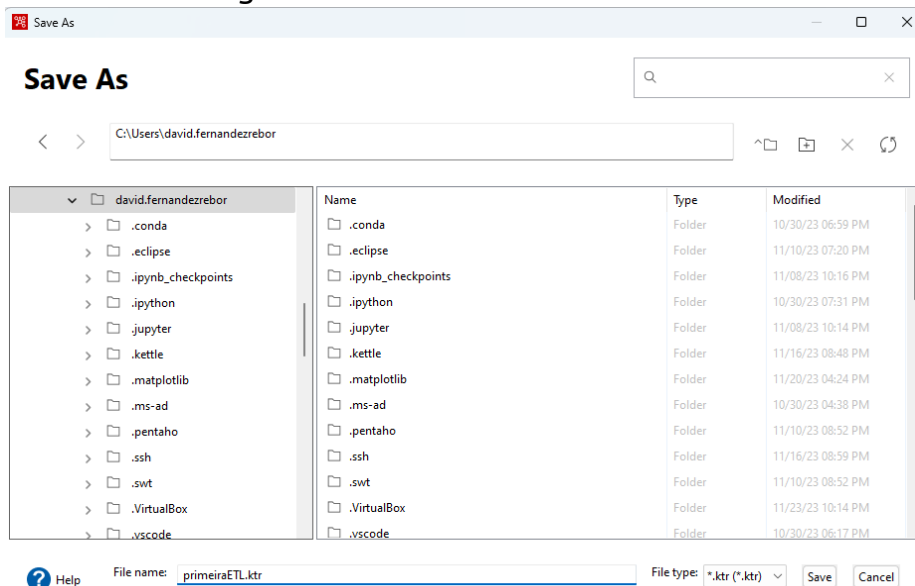
Pentaho Básico

Creación da primeira ETL

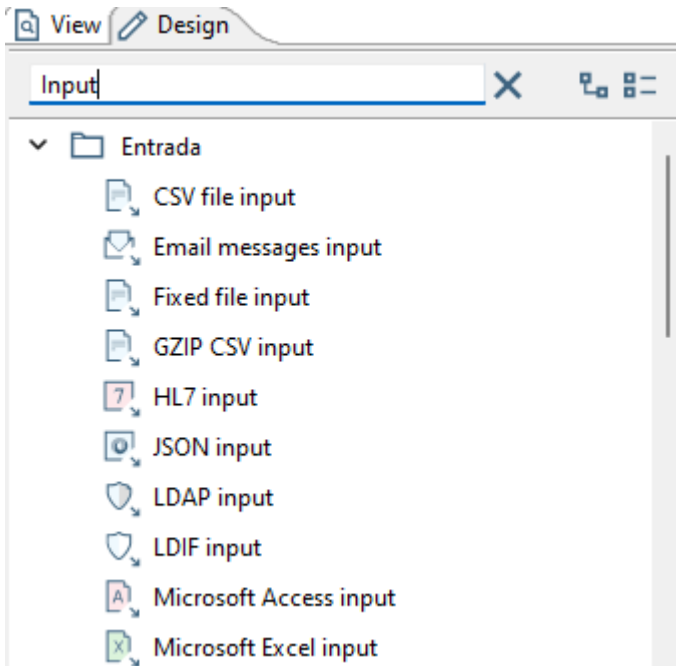
Debemos primeiramente ir a New file e Crear unha nova transformación.



Posteriormente gardaremos a ETL.



Posteriormente buscaremos en Diseño Microsoft Excel Input



Posteriormente pulsaremos no ficheiro e realizaremos os seguintes cambios:

Entrada Excel

Nombre paso: Microsoft Excel input

Add Field(s)

Ficheros | Hojas | Contenido | Manejador de Error | !Campos | Additional output fields

Spread sheet type (engine): Excel 2007 XLSX (Apache POI)

Fichero o directorio:

Expresión Regular:

Exclude Regular Expression:

Password:

Ficheros seleccionados:

#	Fichero/Directorio	Comodón	Exclude wildcard	Requerido	Include sub
1	C:/Users/david.fernandezrebor/Desktop/orixe_primeira_ETL (1).xlsx			N	N

Eliminar Editar

Aceptar nombres de fichero de pasos anteriores

Aceptar nombres de fichero de pasos anteriores ☐

Paso desde el que se obtienen los nombres de:

Campo de entrada a utilizar como nombre de fichero:

Mostrar Fichero(s)...

Help Vale Previsualizar filas Cancelar

Para que aparezca en ficheiros seleccionados debimos ir a ficheiro ou directorio buscar o noso ficheiro e posteriormente engadir.

Logo disto, na lapela de Follas teremos que seleccionar a nosa folla tal e como se amosa no seguinte enlace:

Add Field(s)

Ficheros | Hojas | Contenido | Manejador de Error | !Campos | Additional output fields

Lista de hojas a leer

#	Nombre hoja	Fila inicial	Columna inicial
1	Hoja1		

Enter list

Available items: Hoja1

Your selection:

Hoja(s)...

Agora en contido deberemos de seleccionar as seguintes caixas:

Entrada Excel

Nombre paso: Microsoft Excel input

Add Field(s)

Ficheros | Hojas | Contenido | Manejador de Error | !Campos | Additional output fields

Cabecera ☒

Eliminar filas vacías ☒

Detener al encontrar fila vacía ☒

Límite: 0

Encoding:

Result filenames:

Add filenames to result ☒

Agora prememos en Campos e en obter cabeceira e formateamola baixo a seguinte idea:

[illegible]

Logo damoslle a Vale, o seguinte paso será ir a Diseño, logo disto a Flow e a Filtrar filas(debemos facer doubleclick nela). Agora dentro do filtrar filas creado deberemos engadir unha condición .

Nombre de paso: Filtrar archivos

Enviar 'verdadero' a:

Enviar 'falso' a paso:

La condición:

Con río = 1,0 (Numb...)

Buttons: Help, Vale, Cancelar

Agora en deseño e tendo pulsado en `concellos_con_rio`(o filtrado de filas) crearemos unha transformación simulada, a cal se poderá ver en Deseño e flow, poñeremoslle o nome de paso descartar.



Transformación simulada (no hace nada)

Transformación Simulada

Nombre de paso: descartar

Help Vale Cancelar

Debemos comprobar que no filtrado estea a opción descartar para falso

Nombre de paso: concellos_con_rio

Enviar 'verdadero' a:

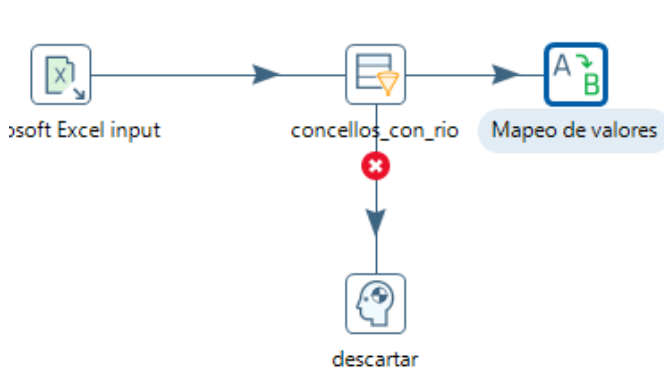
Enviar 'falso' a paso: descartar

La condición:

Con rio = 1,0 (Numb

Help Vale Cancelar

Agora e con concellos_con_rio estableceremos un mapeo de valores que se topa en Diseño/Transformación/Mapeo de valores e estableceremos como true



Nombre de paso: concellos_con_rio

Enviar 'verdadero' a: Mapeo de valores

Enviar 'falso' a paso: descartar

La condición:

Con rio = 1,0 (Numb

Help Vale Cancelar

Agora en Mapeo de valores realizaremos as seguintes insercións:

Nombre de paso: Mapeo de valores

Nombre de campo origen: Tipo orografía

Nombre de campo destino (vacío=sobreescribir):

Default upon non-matching:

Valores de campo:

#	Valor origen	Valor destino
1	1	Montañosa
2	2	Chaira
3	3	Páramo
4	4	Val
5	5	Braña

Help Vale Cancelar

Agora imos a Diseño/Transformación/Selecciona/Renombra valores e xa vamos a cambiarlle o nome de paso e lle introduciremos os seguintes valores:

Selecciona/Renombra valores

Nombre paso:

Selecciona & Modifica Eliminar Meta-información

Campos:

#	Nombre campo	Renombrar a	Longitud	Precisión
1	Concello			
2	Tipo orografía			

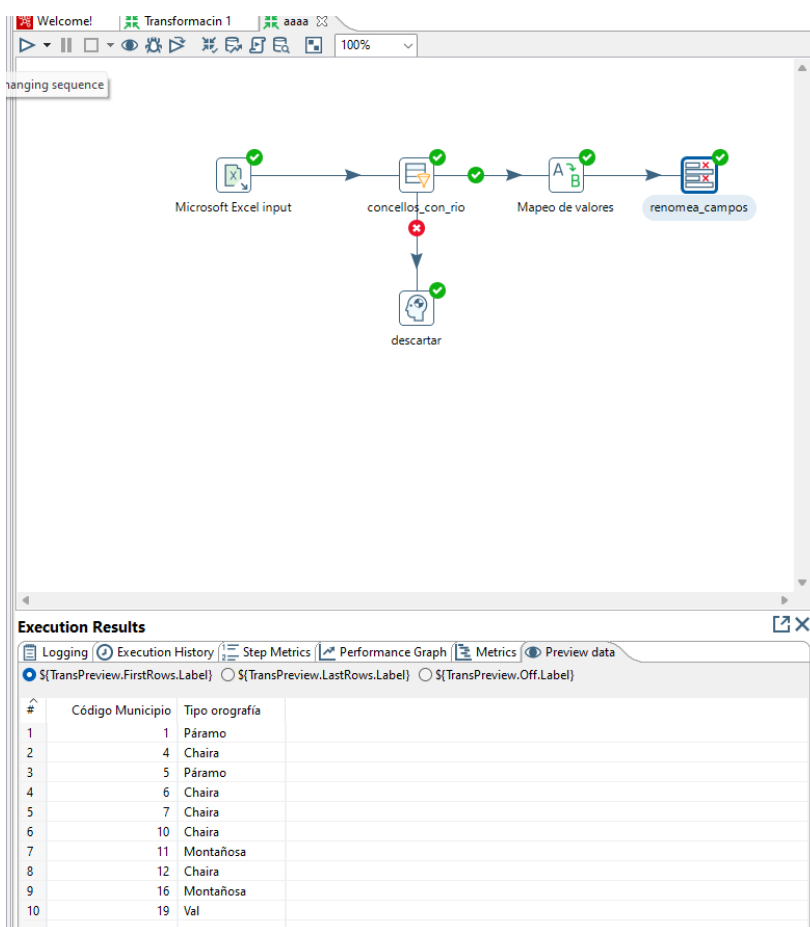
Obtener campos a seleccionar

Edit Mapping

Include unspecified fields, ordered by name ☐

Help Vale Cancelar

Logo disto executaremos a transformación:



Finalmente xa nos devolverá os datos que quixemos filtrar do Excel Orixinal