# T2.2:/Pentaho. Conectando a MySQL/MariaDB

Big Data Aplicado 08/12/23 – IES Fernando Wirtz David Fernández Reboredo

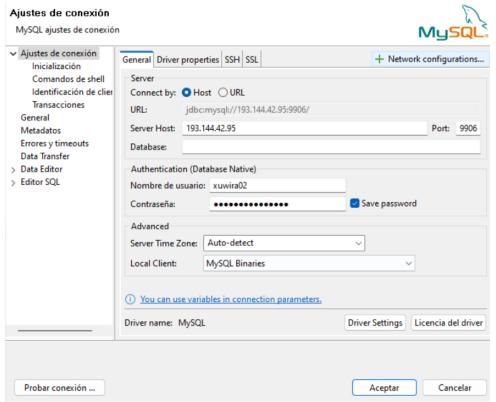
## Índice

Spoon con Base de Datos MySQL/MariaDB	3
Exercicio1	
Ampliación1	
Ampliación2	

## Spoon con Base de Datos MySQL/MariaDB

#### Exercicio1

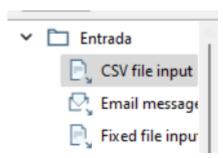
No DBeaver deberemos de realizar unha conexión co servidor MySQL do profesor coas seguintes características, en la sección SSH debemos de desmarcar todas las opciones.



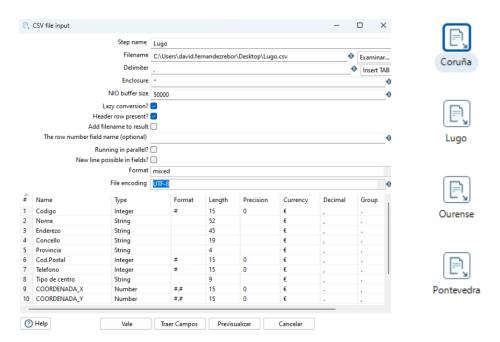
Logo disto crearemos un novo Script para crear unha nova tabla centro na base de datos chamada xuwira2.

```
show databases;
       use xuwira02;
1
P
町
     ⊖ CREATE TABLE centro(
       codigo int primary key,
          nome varchar(200),
          enderezo varchar(200),
          concello varchar(120),
          provincia varchar(50),
          cod postal int,
          tipo centro varchar(12),
          coordenada_x float,
          coordenada y float,
          titularidade varchar(1),
          ensino_concertado int,
          dependente int
```

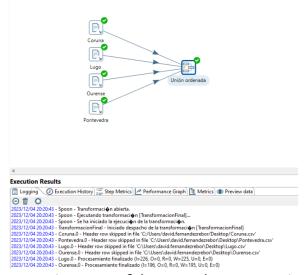
Xa con todo executado o seguinte paso será ir a Pentaho e premer na barra lateral esquerda en Entrada e posteriormente en CSV file input



Dentro da icona creada no noso espazo de traballo deberemos de seleccionar cada unha dos nosos CSV descargados previamente da web da consellería, e importante tamén cambiar a codificación xa que neste caso os arquivos CSV están en UTF-8 debemos de facer isto con cada un deles.



Logo desto na pestaña de transformacións teremos que seleccionar a unión ordenada dos campos e enlazamos cada un deles a unión creada, podemos executalo para comprobar que os cambios se realizaron correctamente.



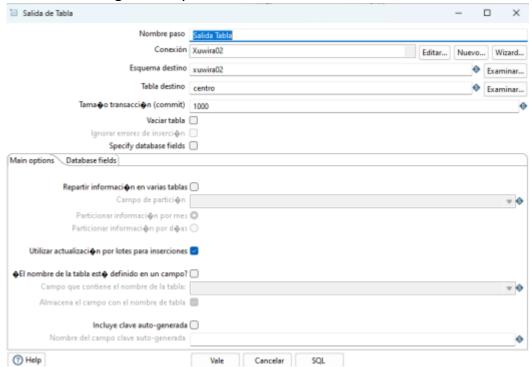
Posteriormente falasenos dun Syncronize after merge sin embargo vamolo sustituir máis adiante por unha salida a táboa, pero antes disto vamos realizar un mapeo de valores tal e como se nos indica no exercicio.

## En titularidade do centro, na base de datos MySQL meter U para pública e R para privada

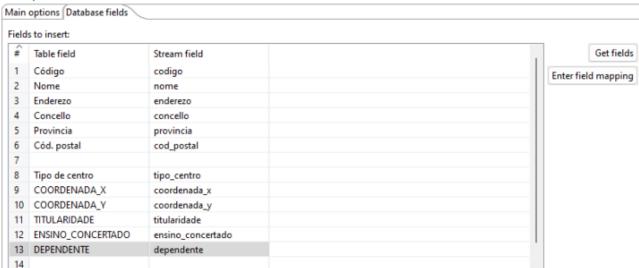
A B	Mapeo de valor	es		_	$\times$
		Nombre de pas	: Mapeo de valores		
		Nombre de campo orige	: TITULARIDADE		
Non	nbre de campo	destino (vac�o=sobreescrib	ir)		
		Default upon non-matchin			
/alor	es de campo:	beraare aport from materials	·		
#	Valor origen	Valor destino			
1	Pública	P			
2	Privada	R			
_	7 117 0 0 0				
0			Vale Cancelar		
	Help				
Ξn (	ensino_cc	ncertado meter	para si e 0 para Non.		
A B	Mapeo de valo	es		-	×
		Nombre de pas	Mapeo de valores 2		
		Nombre de campo orige	: ENSINO_CONCERTADO		
No	mbre de campo	destino (vac�o=sobreescrib	ir)		
		Default upon non-matchin	:		
Valo	res de campo:				
#	Valor origen	Valor destino			
1	Si	1			
2	Non	0			
?	) Help		Vale Cancelar		
Fin	almente e	n denendente m	eter 1 para si e 0 para Non.		
	Mapeo de valor		receive para si e a para i toni	_	×
	•	Nombre de pase			
		Nombre de campo origer			
Non	nbre de campo	destino (vac�o=sobreescrib			
		Default upon non-matching	:		
	es de campo:				
#	Valor origen	Valor destino			
1	Si	1			
2	Non	0			
<u></u>	Help		Vale Cancelar		

Logo disto para a salida seleccionaremos como xa mostramos con antelación unha saída

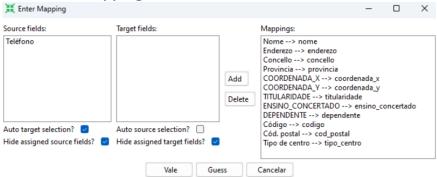
a táboa coas seguintes opcións:



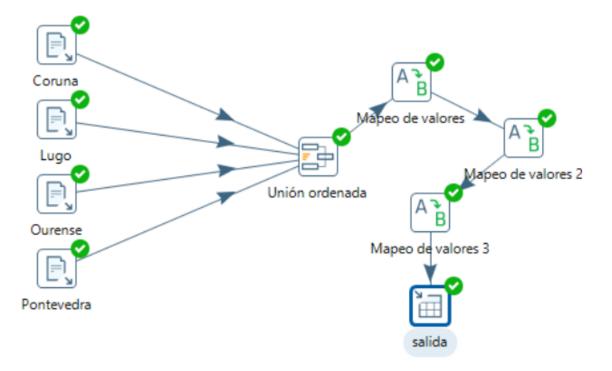
E na pestaña de Database fields deberemos de seleccionar Get Fields

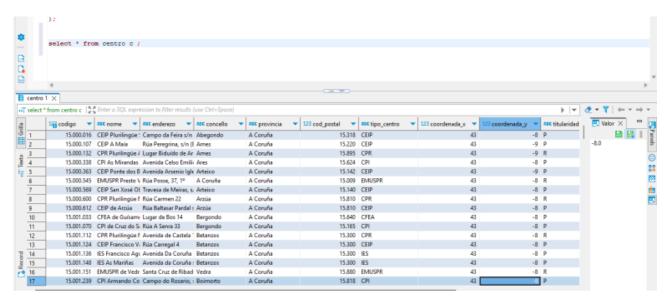


Finalmente algúns dos valores non se enlazan polo que haberá que recurrir ao boton Enter field mapping, unha vez dentro enlazaremos os datos da seguinte maneira:



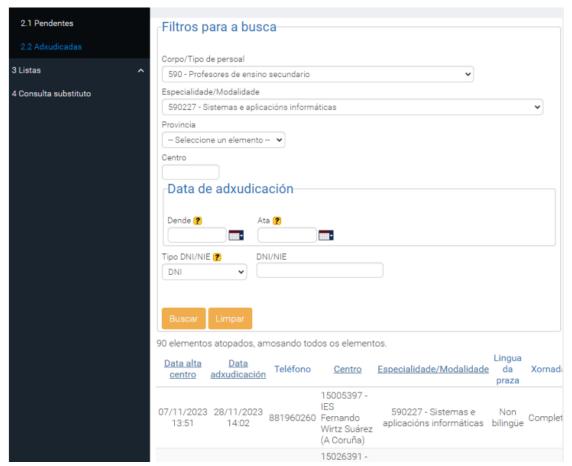
Logo disto ejecutamos e poderemos ver que na taboa da base de datos tamén se actualizou a información.



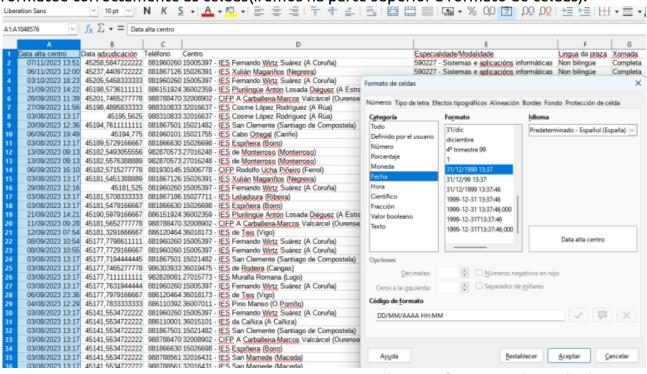


## Ampliación1

Primeiramente iremos a web da conselleria e copiaremos todos os datos para o seguinte grupo de profesores:



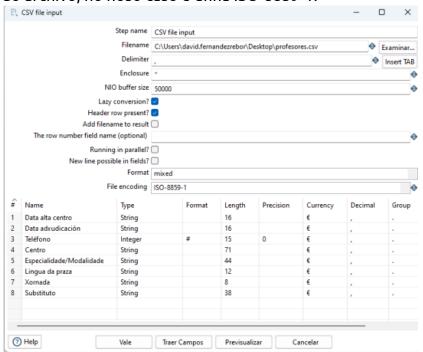
O resultado da copia deberemos de gardalo nun excel realizando cambios para que formatee correctamente as celdas(iremos na parte superior a formato de celdas):



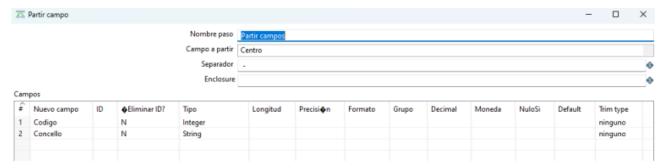
Logo entramos no DBeaver e crearemos a seguinte táboa profesores na base de datos xuwira02:

```
© CREATE TABLE profesores(
codigo int primary key,
data_alta date,
data_adxudicacion date,
telefono varchar(9),
centro varchar(200),
especialidade varchar(200),
lingua varchar(50),
xornada varchar(12),
substituto varchar(100)
);
select * from profesores;
```

Logo disto dirixirémonos a Pentaho e en entrada>csv file input seleccionaremos a táboa de excel que foi creada anteriormente, o formato deberemos comprobar a codificación do archivo, no noso caso e unha ISO-8859-1.

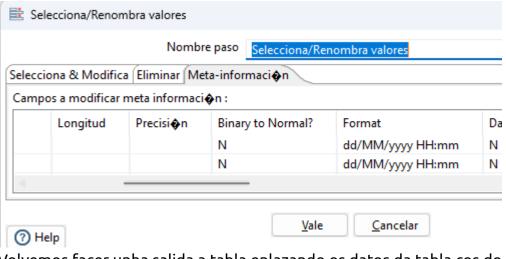


Deremos de partir o campo Concello en Codigo e Concello para iso seleccionamos a opción lateral de transformacións chamada partir campo:

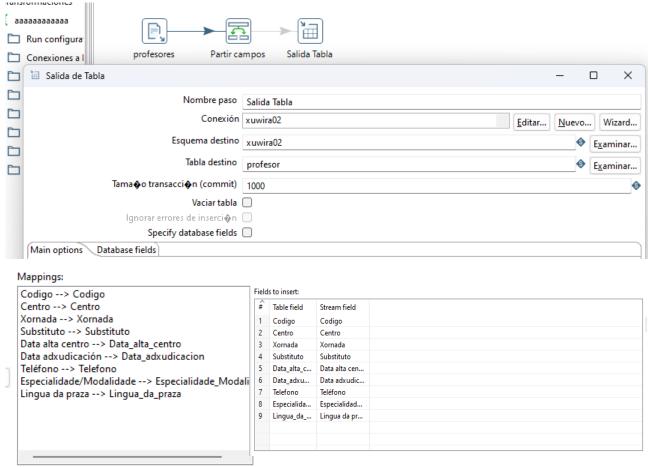


Como podemos ver na imaxe introduciuse un separador " - " para alonxar os dous campos.

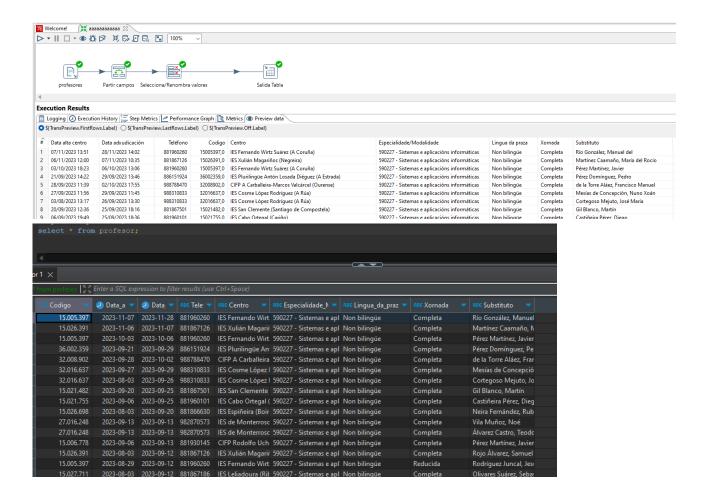
Tamen deberemos con selecciona e renombra nombres en transformacións ir a pestaña de meta información debemos cambiar el tipo de dato A par de isto crearemos no dbeaver a seguinte táboa: Facemos un selecciona/renombra valores para proporcionar a máscara:



Volvemos facer unha salida a tabla enlazando os datos da tabla cos do csv:



Executaremos o exercicio e xa se nos haberá posto na base de datos a información:



### Ampliación2

Descargaremos tabula e pulsaremos en importar, unha vez ahí seleccionaremos o arquivo pdf. Con isto comezará o proceso de carga:



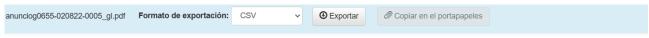
Seleccionaremos una de las tablas y daremos en repita esta selección



Posteriormente pulsaremos en vista previa y exportación de datos

■ Vista previa y exportación de datos extraídos

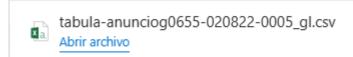
Posteriormente mostrarasenos unha previsualización na que observaremos que os datos están ben introducidos



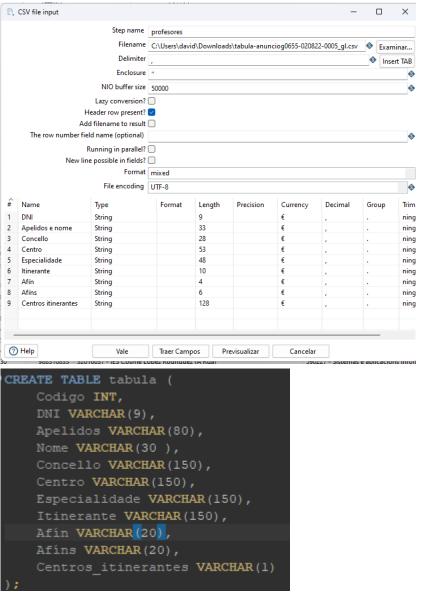
#### Vista previa de los datos tabulares extraídos

DNI	Apelidos e nome	Concello	Centro	Especialidade	Itinerante	Afín	Afíns	Centros itine
7169**	Díaz Armesto, Goretti	27031-Monforte de Lemos	27016662-CEIP Plurilingüe de Monforte de Lemos	597031-Educación Infantil	N	N		
9188**	Míguez Fernández, Pilar María	15031-Culleredo	15021809-CEIP Sofia Casanova	597038-Educación Primaria	N	N		
3192**	García Gómez, Ana Isabel	36038-Pontevedra	36006122-CEIP A Xunqueira Núm. 1	597038-Educación Primaria	N	N		
4419**	Valiño Lombardero, Encarnación	27004-Baleira	27000319-CEIP Plurilingüe Concepción López Rey	597031-Educación Infantil	N	N		
1550**	Souto Amado, José Manuel	32054-Ourense	32015207-IES O Couto	597060-Educación Especial: Pedagoxía Terapéutica	N	S	597061	
2866**	Marín Fernández, Andrea	32085-Verín	32016790-IES Castro de Baronceli	597060-Educación Especial: Pedagoxía Terapéutica	N	N		
8146**	Canosa Pasantes, Fabián	15030-A Coruña	15005701-CEIP Plurilingüe San Pedro de Visma	597038-Educación Primaria	N	N		
9997**	Vidal Gómez, Fátima	36038-Pontevedra	36006316-CEIP Daría González García	597038-Educación Primaria	N	N		
6892**	Correa Pando, Isabel	36057-Vigo	36009998-CEIP de Mosteiro- Bembrive	597033-Lingua Estranxeira: Francés	N	N		

#### Finalmente descargaremos o csv cos datos



Agora pasaremos ao pentaho para pasar os datos a base de datos e tamen crearemos a táboa no Dbeaver:

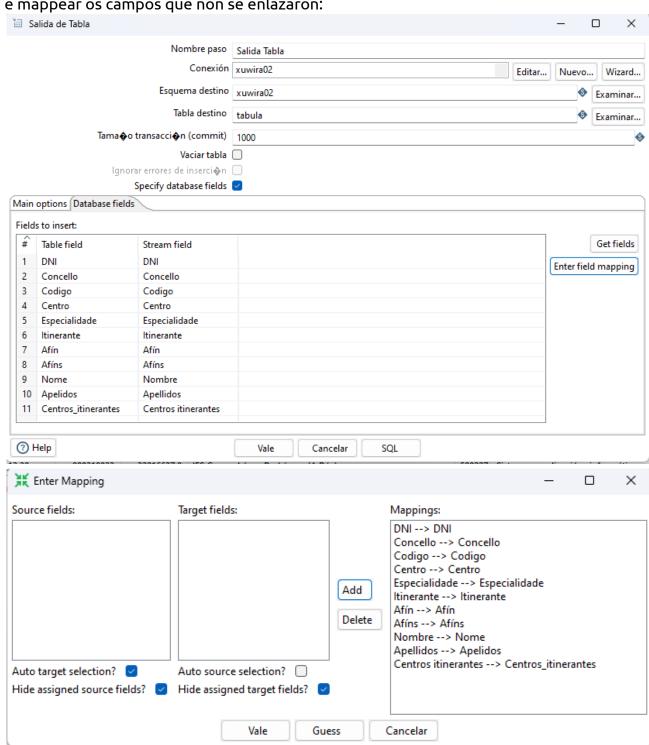


Como podemos apreciar volvimos crear un campo código para separar os campos en centro e o seu código ademais tal e como temos a táboa vamos a proceder a separar apellidos do nome con partir campo e o separador ",":

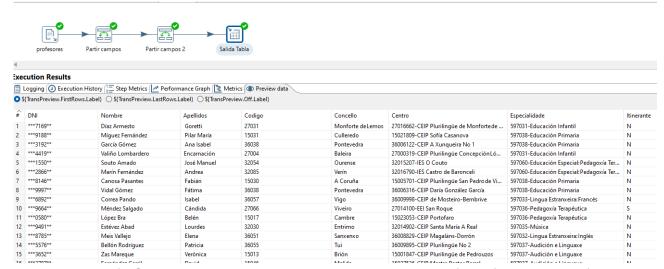




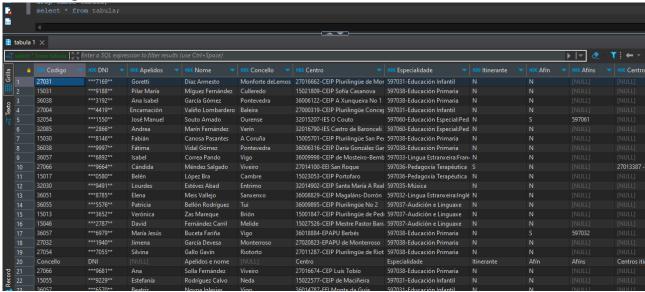
Logo disto volveremos a crear a salida a tabla pero neste caso da táboa chamada tábula e mappear os campos que non se enlazaron:



Logo disto correremos a transformación:



Vemos que todo funcionou correctamente pasamos a ver nunha select se o cambio tamén se gardou na base de datos:



Todo funcionou correctamente como podemos ver na SQL.