

Herramientas ETL.PDI(Pentaho Data Integration).

1. Crea una BD Postgresql cuyo nombre sea el nombre de la provincia que tienes asignada y tu nombre de usuario de correo UGR (en mi caso se llamará `granada_jsamos`). En el esquema `public` de esa BD crea las tablas `cuando`, `donde` `padrón` añadiendo al nombre el sufijo de tu nombre de usuario de correo UGR (en mi caso se llamará `cuando_jsamos`, `donde_jsamos` y `padron_jsamos`). La estructura de estas tablas ha de ser similar a la de las hojas correspondientes del archivo obtenido con Power Query En la actividad anterior, usa estas hojas como origen de datos (en mi caso el archivo es `granada-ETL-jsamos.xlsx`) pero, para los nombres de los campos, usa minúsculas sin tilde y, en lugar del criterio CamelCase (CamelCase), usa el carácter «`_`» como separador (`camel_case`).

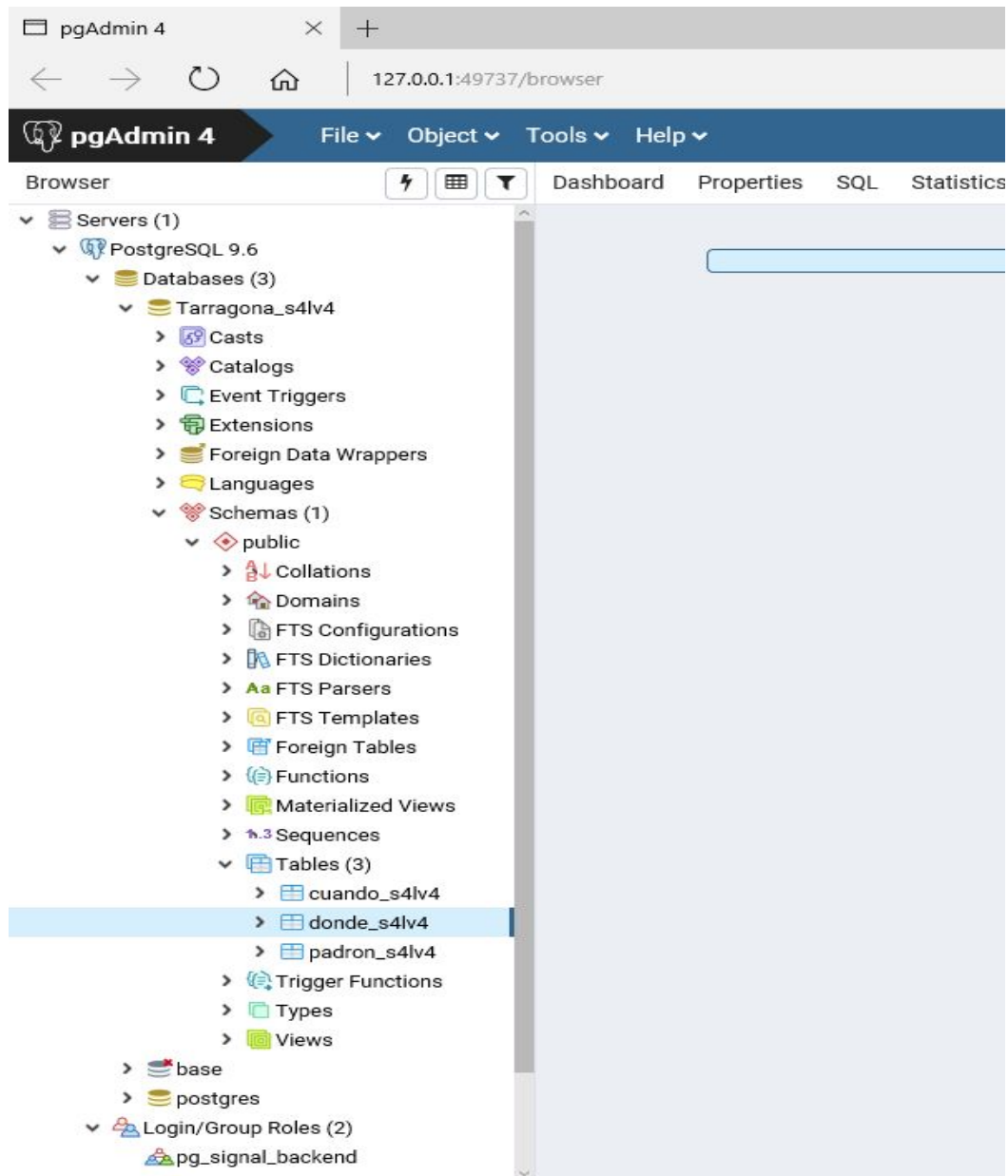
1.1. Para cada tabla define una transformación distinta cuyo nombre sea el nombre de la provincia que tienes asignada, el literal «-ETL-», tu nombre de usuario de correo UGR y el nombre de la tabla (en mi caso, por ejemplo, una transformación se llamará `granada-ETL-jsamos-cuando`).

1.2. El nombre de cada paso de las transformaciones ha de tener como sufijo tu nombre de usuario de correo UGR (por ejemplo, en mi caso, un paso se llamará `escribe-tabla-jsamos`).

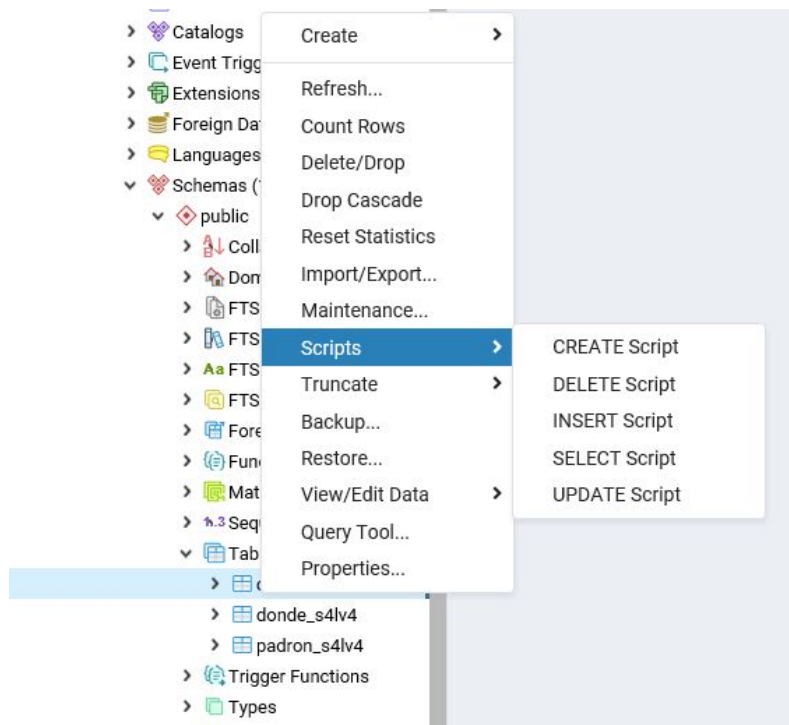
1.3. Define un trabajo («Jobs») que controle la ejecución de las transformaciones definidas previamente tanto en caso de que todo funcione bien como si se producen errores.

Para esta práctica se ha usado PgAdmins para crear la base de datos y las tablas y generar las sentencias SQL. Para la transformación y la creación del trabajo se ha usado spoon.

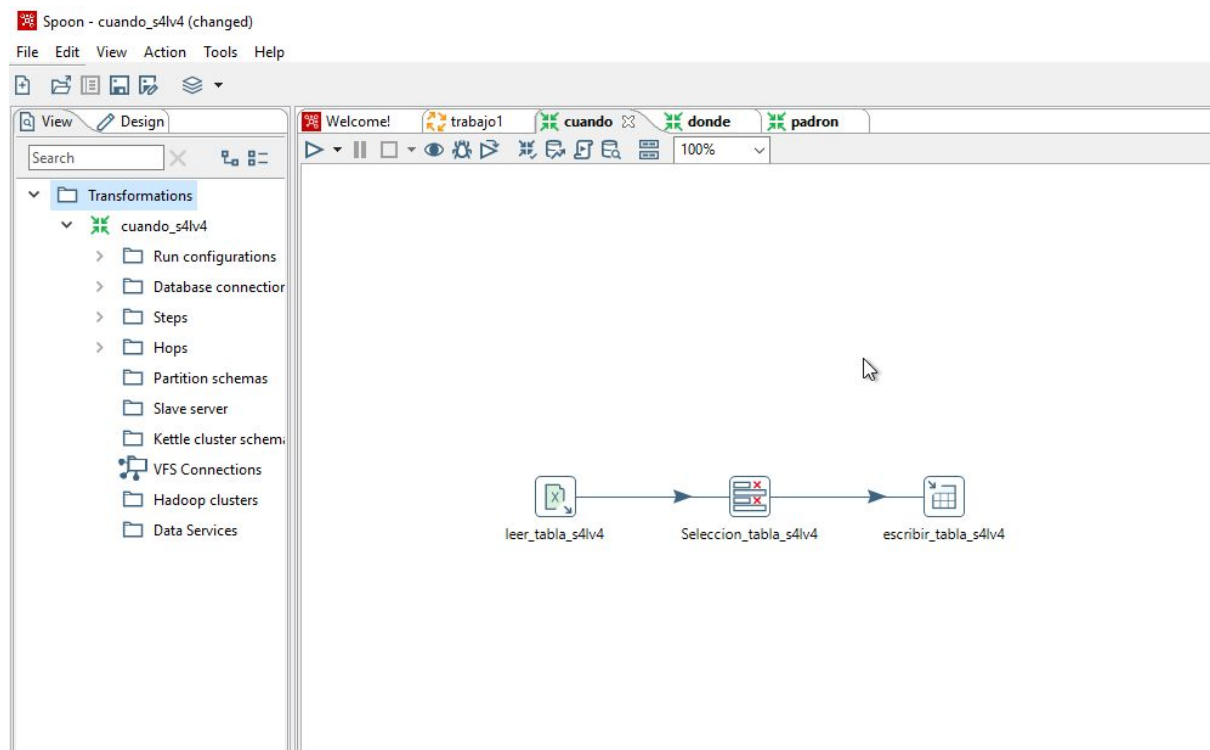
Lo primero que se ha hecho ha sido crear la base de datos con el nombre de Tarragona_s4lv4. Además se han creado los diferentes tablas que se nos pide :cuando_s4lv4,donde_s4lv4,padron_s4lv4.



Para la parte de la generación SQL nos vamos a script->Create script.



Para la creación del trabajo y las transformaciones usaremos Spoon.
Primero vamos a definir tres transformaciones que tendrán el mismo nombre que las tablas y en las transformaciones le daremos un nombre específico a cada uno de los pasos que hagamos.



Aclaración: El nombre no se actualizaba en las pestañas pero a la izquierda se ve el nombre correcto.

Ahora veremos los datos introducidos en cada tabla:

1. Cuando_s4lv4.

1.1 Leer tabla.

Microsoft Excel input

Step name: leer_tabla_s4lv4

Files | Sheets | Content | Error Handling | Fields | Additional output fields

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

File or directory: Add Browse...

Regular Expression:

Exclude Regular Expression:

Password:

Selected files:

#	File/Directory	Wildcard (RegExp)	Exclude wildcard
1	C:\Users\aulas\Downloads\fotos\Tarragona-ETL-s4lv4.xlsx		

Delete

Microsoft Excel input

Step name: leer_tabla_s4lv4

Files | Sheets | Content | Error Handling | Fields | Additional output fields

#	Name	Type	Length	Precision	Trim type	Repeat	Format	Currency	Decimal	Grouping
1	IdCuándo	Number	-1	-1	none	N				
2	año	Number	-1	-1	none	N				
3	Década	Number	-1	-1	none	N				

Microsoft Excel input

Step name: leer_tabla_s4lv4

Files | Sheets | Content | Error Handling | Fields | Additional output fields

List of sheets to read

#	Sheet name	Start row	Start column
1	Cuándo	0	0

1.2. Selección.

Select values

Step name: Seleccion_tabla_s4lv4

Select & Alter | Remove | Meta-data

Fields:

#	Fieldname	Rename to	Length	Precision
1	IdCuándo	id_cuando		
2	año			
3	Década	decada		

Get fields to select

Edit Mapping

1.3.Escribir.

Table output

Step name

escribir_tabla_s4lv4

Connection

prueba

Edit...

New...

Wizard...

Target schema

public

Browse...

Target table

cuando_s4lv4

Browse...

Commit size

1000

Truncate table

☒

Ignore insert errors

☐

Specify database fields

☐

Main options

Database fields

2.Donde_s4lv4.

2.1Leer.

Microsoft Excel input

Step name

leer_tabla_s4lv4

Files

Sheets

Content

Error Handling

Fields

Additional output fields

Spread sheet type (engine)

Excel 2007 XLSX (Apache POI)

File or directory

Add

Browse...

Regular Expression

Exclude Regular Expression

Password

Selected files:

#	File/Directory	Wildcard (RegExp)	Exclude wildcard
1	C:\Users\aulas\Downloads\fotos\Tarragona-ETL-s4lv4.xlsx		

Refresh

Microsoft Excel input

Step name

leer_tabla_s4lv4

Files

Sheets

Content

Error Handling

Fields

Additional output fields

#	Name	Type	Length	Precision	Trim type	Repeat	Format	Currency	Decimal	Grouping
1	IdDonde	Number	-1	-1	none	N				
2	CodMunicipio	Number	-1	-1	none	N				
3	Municipio	String	-1	-1	none	N				
4	CapitalMunicipio	String	-1	-1	none	N				
5	CodComarca	Number	-1	-1	none	N				
6	Comarca	String	-1	-1	none	N				

Microsoft Excel input

Step name

leer_tabla_s4lv4

Files

Sheets

Content

Error Handling

Fields

Additional output fields

List of sheets to read

#	Sheet name	Start row	Start column
1	dónde	0	0

2.2. Selección.

Step name:

Select & Alter Remove Meta-data

Fields:

#	Fieldname	Rename to	Length	Precision
1	IdDonde	id_donde		
2	CodMunicipio	cod_municipio		
3	Municipio	municipio		
4	CapitalMunicipio	capital_municipio		
5	CodComarca	cod_comarca		
6	Comarca	comarca		
7	CodProvincia	cod_provincia		
8	Provincia	provincia		
9	CapitalProvincia	capital_provincia		

Get fields to select
Edit Mapping

2.3. Escribir.

Step name:

Connection: Edit... New... Wizard...

Target schema: Browse...

Target table: Browse...

Commit size:

Truncate table: ☒

Ignore insert errors: ☐

Specify database fields: ☐

3. Padron_s4lv4.

3.1 Leer.

Step name:

Files Sheets Content Error Handling Fields Additional output fields

Spread sheet type (engine):

File or directory: Add Browse...

Regular Expression:

Exclude Regular Expression:

Password:

Selected files:

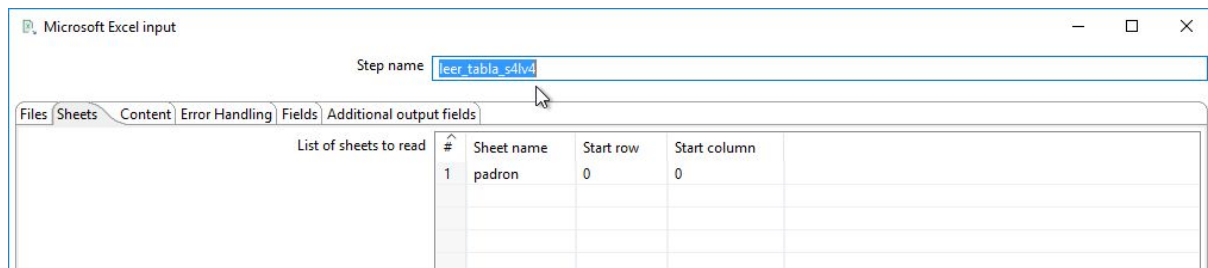
#	File/Directory	Wildcard (RegExp)	Exclude wildcard
1	C:\Users\aulas\Downloads\fotos\Tarragona-ETL-s4lv4.xlsx		

Delete

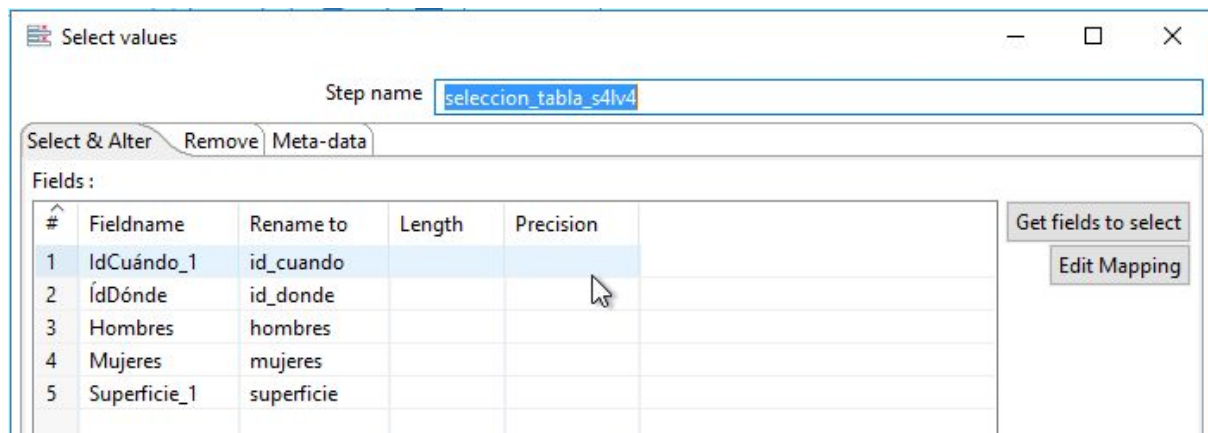
Step name:

Files Sheets Content Error Handling Fields Additional output fields

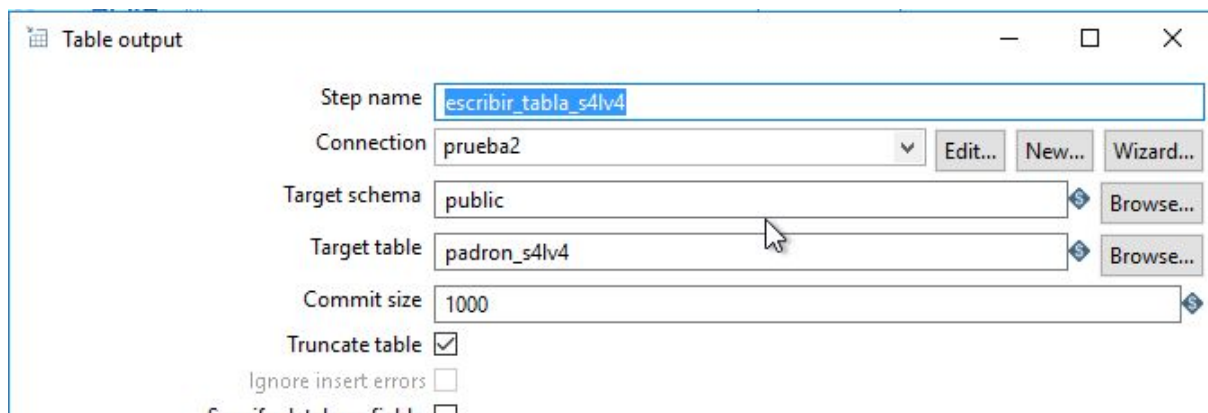
#	Name	Type	Length	Precision	Trim type	Repeat	Format	Currency	Decimal	Grouping
1	IdCuándo_1	Number	-1	-1	none	N				
2	IdDónde	Number	-1	-1	none	N				
3	Hombres	Number	-1	-1	none	N				
4	Mujeres	Number	-1	-1	none	N				
5	Superficie_1	Number	-1	-1	none	N				



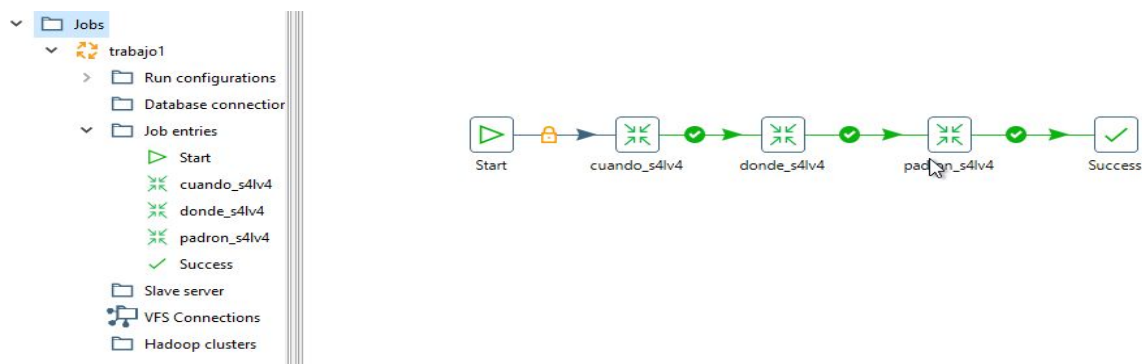
3.2. Selección.



3.3. Escribir.



Aclaración: Se ha tenido que definir tres conexiones con la base de datos por eso en la conexión hay tres nombres de prueba, prueba1 y prueba2. Ahora definimos el trabajo. El trabajo se ha definido de forma secuencial aunque no era necesario ya que ninguno depende del otro.



Ahora lo ejecutamos pulsando RUN y ya tendríamos en PgAdmins los datos de las tablas.

1.Tabla “cuando”.

The screenshot shows the PgAdmin interface. At the top, there's a toolbar with icons for file operations, search, and execution. Below the toolbar, the connection path is displayed: `public.cuando_s4lv4/Tarragona_s4lv4/postgres@PostgreSQL 9.6`. The main area is divided into two tabs: "Query Editor" and "Query History". The "Query Editor" tab is active, showing a SQL query:

```
1 SELECT * FROM public.cuando_s4lv4
2
```

Below the query editor, there are four tabs: "Data Output", "Explain", "Messages", and "Notifications". The "Data Output" tab is active, displaying the results of the query in a table format. The table has four columns: `id_cuando` (double precision), `año` (double precision), `decada` (double precision), and an empty column. The results are as follows:

	<code>id_cuando</code> double precision	<code>año</code> double precision	<code>decada</code> double precision	
1	1	1996	1990	
2	2	1998	1990	
3	3	1999	1990	
4	4	2000	2000	
5	5	2001	2000	
6	6	2002	2000	
7	7	2003	2000	
8	8	2004	2000	
9	9	2005	2000	
10	10	2006	2000	

2.Tabla “donde”.

Query Editor Query History

```
1 SELECT * FROM public.donde_s4lv4
2
```

Data Output Explain Messages Notifications

	id_donde double precision	cod_municipio double precision	municipio text	capital_municipio text	cod_comarca double precision
1	1	43001	Aiguamúrcia	Santes Creus	4301
2	6	43006	Aldover	Aldover	4309
3	7	43007	Aleixar, L'	L'Aleixar	4308
4	11	43011	Almóster	Almóster	4308
5	2	43002	Albinyana	Albinyana	4312
6	9	43009	Alforja	Alforja	4308
7	3	43003	Albiol, L'	L'Albiol	4308
8	15	43015	Arbolí	Arbolí	4308
9	17	43017	Argentera, L'	L'Argentera	4308
10	4	43004	Alcanar	Alcanar	4322

3.Tabla “padron”.

public.padron_s4lv4/Tarragona_s4lv4/postgres@PostgreSQL 9.6					
Query Editor Query History					
<pre>1 SELECT * FROM public.padron_s4lv4 2</pre>					
Data Output Explain Messages Notifications					
	id_cuando double precision	id_donde double precision	hombres double precision	mujeres double precision	superficie double precision
1	1	1	310	304	73.7537
2	2	1	313	314	73.7537
3	1	6	376	418	20.259974
4	1	2	553	524	19.568918
5	1	7	312	321	26.154494
6	1	3	74	70	20.377719
7	1	11	338	341	6.027473
8	1	4	3796	3931	47.101191
9	1	5	1798	1725	46.278446
10	1	9	597	582	38.016035