

Sistemas MultiDimensionales

Práctica 3_1. Pentaho Data Integration (PDI)

Diseño e implementación del componente ETL



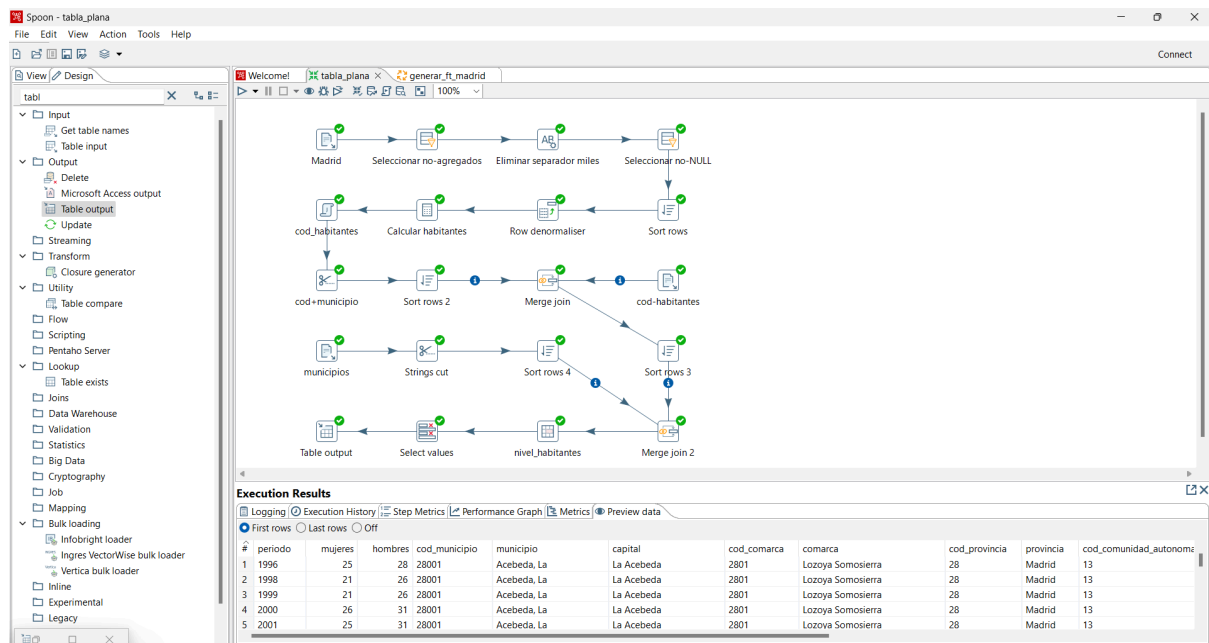
19 de mayo de 2025
Prof. José Samos Jiménez

Autor:
Ana Aragón Jerónimo

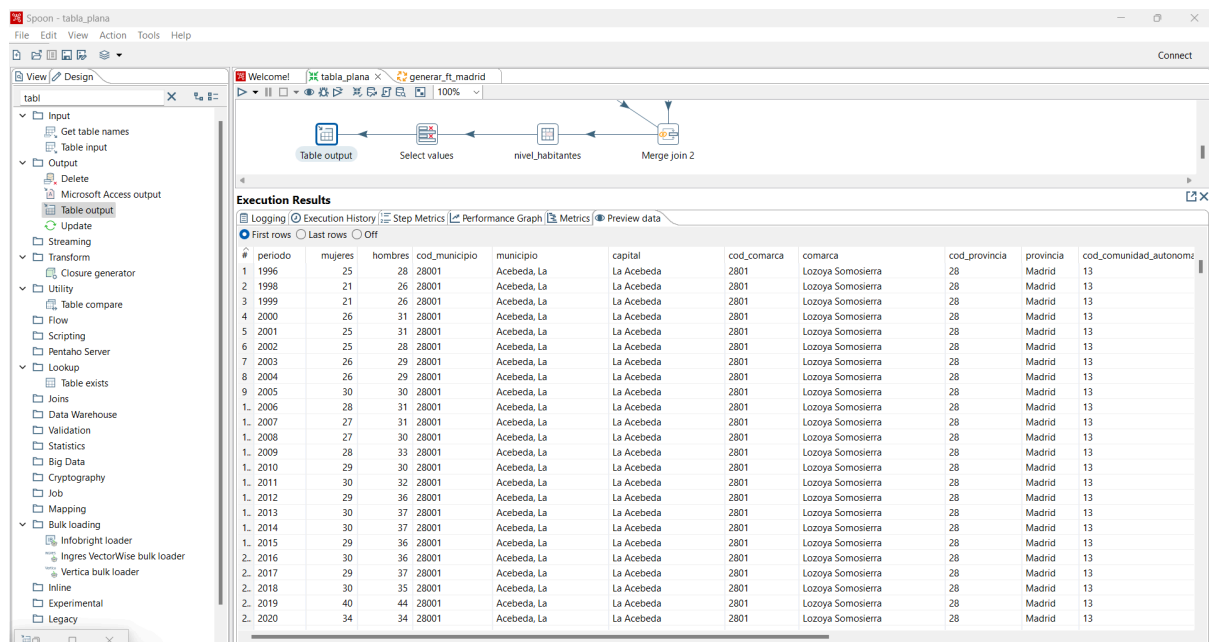
Índice

Índice.....	1
Ejercicio 3.1.....	2
Ejercicio 3.2.....	3
Ejercicio 3.3.....	5
Ejercicio 3.4.....	9
Ejercicio 3.5.....	12
Ejercicio 3.6.....	15
Ejercicio 3.7.....	18

Ejercicio 3.1

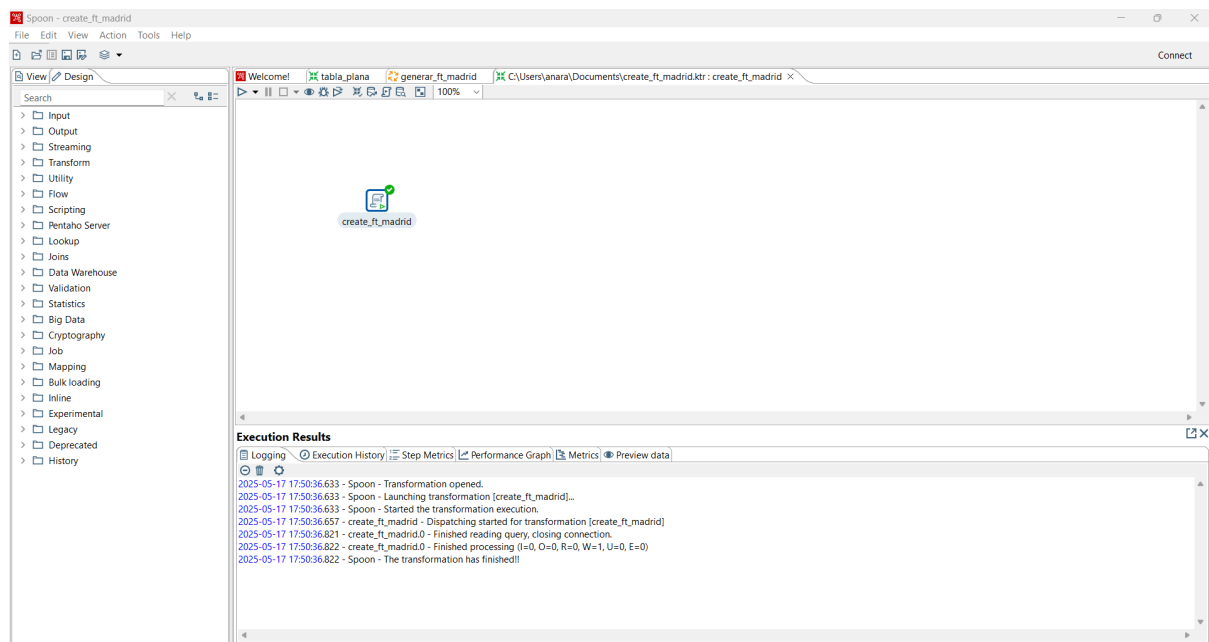


Definición del flujo de datos para la obtención de la tabla plana.

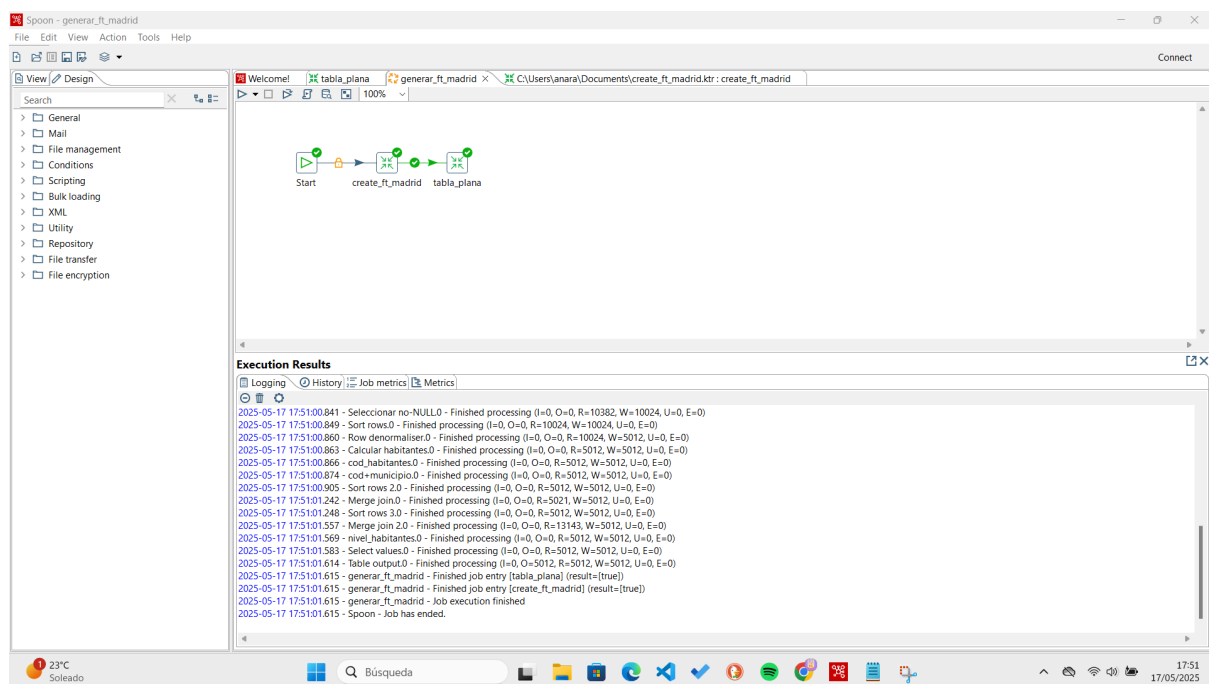


Preview data del último paso (Table output).

Ejercicio 3.2



Ejecución del script para la creación de ft_madrid.



Resultado de la ejecución del trabajo.

pgAdmin 4

Welcome public.ft_madrid/madrid-anarajer/postgres@PostgreSQL 17 X

public.ft_madrid/madrid-anarajer/postgres@PostgreSQL 17

Query Query History

1 SELECT * FROM public.ft_madrid

Data Output Messages Notifications

Showing rows: 1 to 1000 Page No: 1 of 6

	periodo character varying (17)	mujeres integer	hombres integer	cod_municipio character varying (5)	municipio character varying (100)	capital character varying (100)	cod_comarca character varying (100)	comarca character varying (100)	cod_provincia character varying (2)	provincia character varying (100)	cod_comu character v
1	1996	25	28	28001	Acebeda, La	La Acebeda	2801	Lozoya Somosierra	28	Madrid	13
2	1998	21	26	28001	Acebeda, La	La Acebeda	2801	Lozoya Somosierra	28	Madrid	13
3	1999	21	26	28001	Acebeda, La	La Acebeda	2801	Lozoya Somosierra	28	Madrid	13
4	2000	26	31	28001	Acebeda, La	La Acebeda	2801	Lozoya Somosierra	28	Madrid	13
5	2001	25	31	28001	Acebeda, La	La Acebeda	2801	Lozoya Somosierra	28	Madrid	13
6	2002	25	28	28001	Acebeda, La	La Acebeda	2801	Lozoya Somosierra	28	Madrid	13
7	2003	26	29	28001	Acebeda, La	La Acebeda	2801	Lozoya Somosierra	28	Madrid	13
8	2004	26	29	28001	Acebeda, La	La Acebeda	2801	Lozoya Somosierra	28	Madrid	13
9	2005	30	30	28001	Acebeda, La	La Acebeda	2801	Lozoya Somosierra	28	Madrid	13
10	2006	28	31	28001	Acebeda, La	La Acebeda	2801	Lozoya Somosierra	28	Madrid	13
11	2007	27	31	28001	Acebeda, La	La Acebeda	2801	Lozoya Somosierra	28	Madrid	13
12	2008	27	30	28001	Acebeda, La	La Acebeda	2801	Lozoya Somosierra	28	Madrid	13
13	2009	28	33	28001	Acebeda, La	La Acebeda	2801	Lozoya Somosierra	28	Madrid	13
14	2010	29	30	28001	Acebeda, La	La Acebeda	2801	Lozoya Somosierra	28	Madrid	13
15	2011	30	32	28001	Acebeda, La	La Acebeda	2801	Lozoya Somosierra	28	Madrid	13
16	2012	29	36	28001	Acebeda, La	La Acebeda	2801	Lozoya Somosierra	28	Madrid	13
17	2013	30	37	28001	Acebeda, La	La Acebeda	2801	Lozoya Somosierra	28	Madrid	13
18	2014	30	37	28001	Acebeda, La	La Acebeda	2801	Lozoya Somosierra	28	Madrid	13

Total rows: 5012 Query complete 00:00:00.216 CRLF Ln 1, Col 1

Primeras filas de la base de datos en PostgreSQL.

pgAdmin 4

Welcome public.ft_madrid/madrid-anarajer/postgres@PostgreSQL 17 X

public.ft_madrid/madrid-anarajer/postgres@PostgreSQL 17

Query Query History

1 SELECT * FROM public.ft_madrid

Data Output Messages Notifications

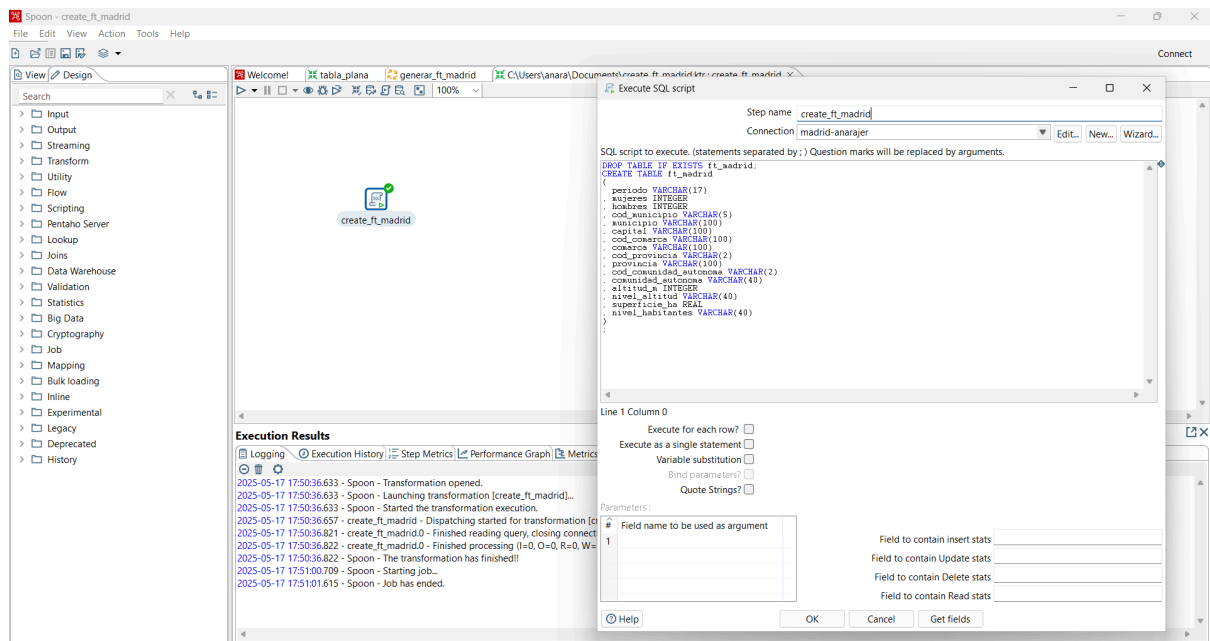
Showing rows: 5001 to 5012 Page No: 6 of 6

	periodo character varying (17)	mujeres integer	hombres integer	cod_municipio character varying (5)	municipio character varying (100)	capital character varying (100)	cod_comarca character varying (100)	comarca character varying (100)	cod_provincia character varying (2)	provincia character varying (100)	cod_comu character v
5001	2013	21607	20289	28903	Tres Cantos	Tres Cantos	2803	Area Metropolitana de Madrid	28	Madrid	13
5002	2014	21953	20593	28903	Tres Cantos	Tres Cantos	2803	Area Metropolitana de Madrid	28	Madrid	13
5003	2015	22352	20957	28903	Tres Cantos	Tres Cantos	2803	Area Metropolitana de Madrid	28	Madrid	13
5004	2016	23000	21764	28903	Tres Cantos	Tres Cantos	2803	Area Metropolitana de Madrid	28	Madrid	13
5005	2017	23697	22349	28903	Tres Cantos	Tres Cantos	2803	Area Metropolitana de Madrid	28	Madrid	13
5006	2018	24056	22694	28903	Tres Cantos	Tres Cantos	2803	Area Metropolitana de Madrid	28	Madrid	13
5007	2019	24581	23141	28903	Tres Cantos	Tres Cantos	2803	Area Metropolitana de Madrid	28	Madrid	13
5008	2020	24887	23439	28903	Tres Cantos	Tres Cantos	2803	Area Metropolitana de Madrid	28	Madrid	13
5009	2021	25073	23678	28903	Tres Cantos	Tres Cantos	2803	Area Metropolitana de Madrid	28	Madrid	13
5010	2022	25862	24325	28903	Tres Cantos	Tres Cantos	2803	Area Metropolitana de Madrid	28	Madrid	13
5011	2023	26529	24924	28903	Tres Cantos	Tres Cantos	2803	Area Metropolitana de Madrid	28	Madrid	13
5012	2024	27238	25694	28903	Tres Cantos	Tres Cantos	2803	Area Metropolitana de Madrid	28	Madrid	13

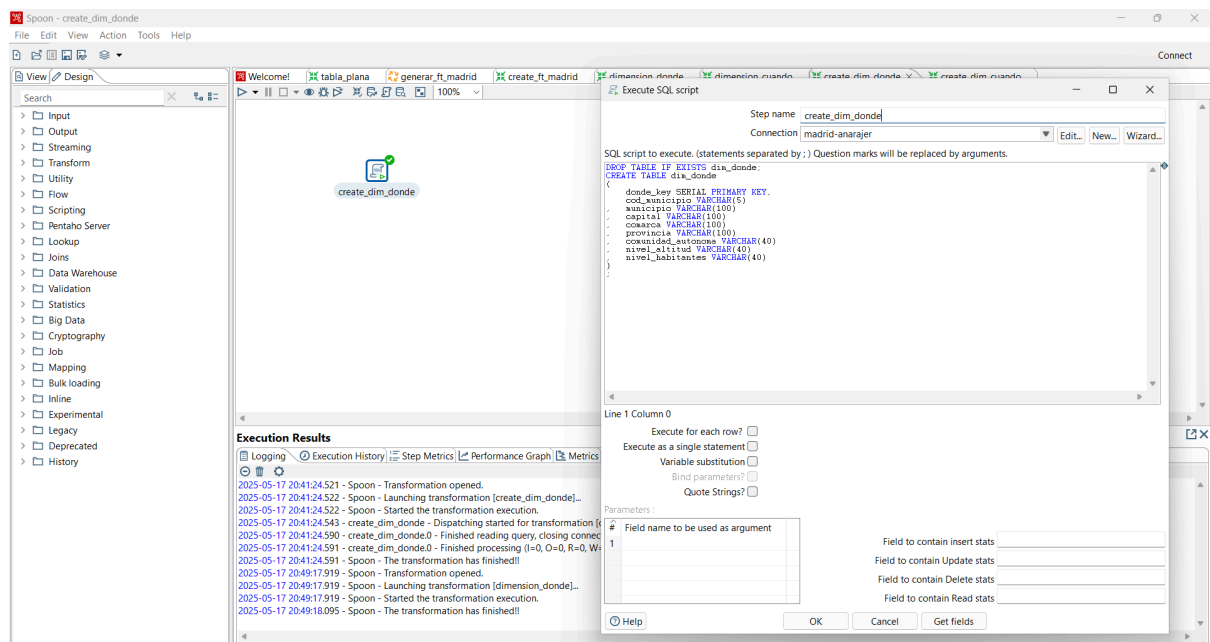
Total rows: 5012 Query complete 00:00:00.216 CRLF Ln 1, Col 1

Últimas filas de la base de datos en PostgreSQL.

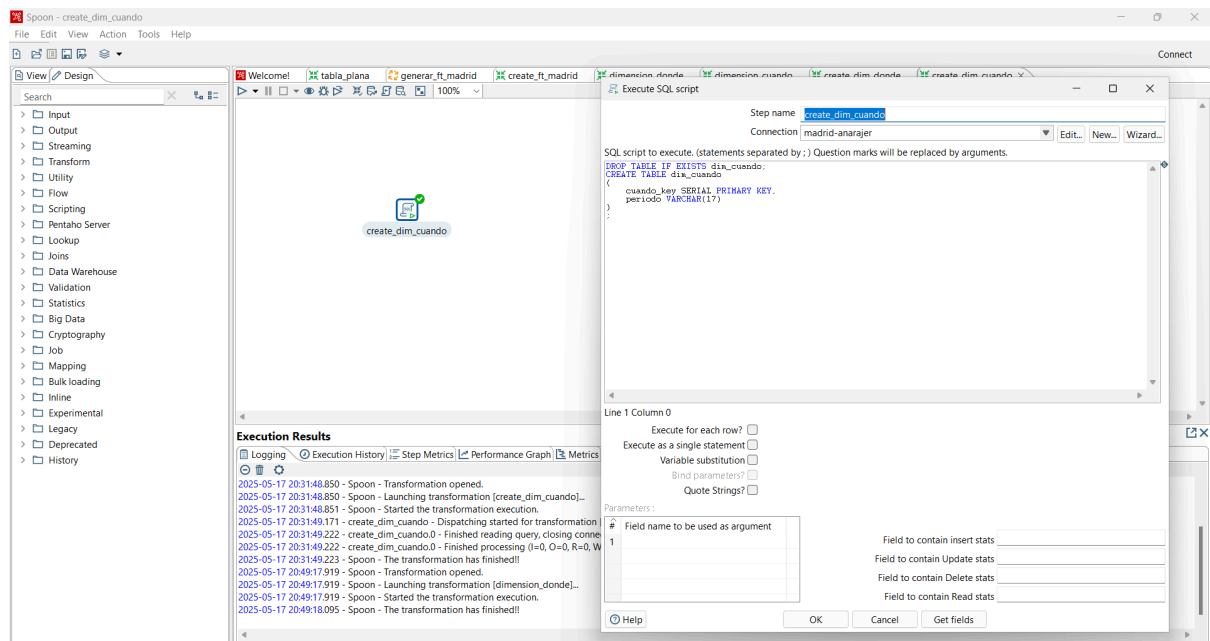
Ejercicio 3.3



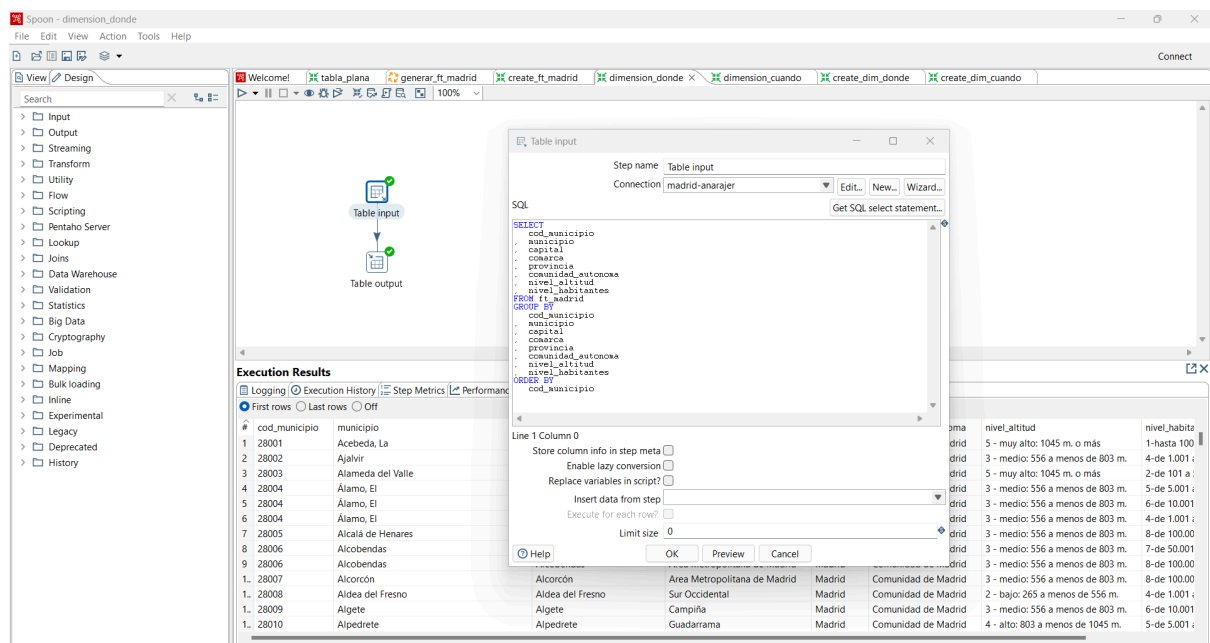
Definición de la creación de la tabla ft_madrid junto con la ejecución.



Definición de la creación de la tabla dim_donde (la dimensión DÓNDE) con las llaves primarias junto con la ejecución.



Definición de la creación de la tabla dim_cuando (dimensión CUÁNDO) con las llaves primarias junto con la ejecución.



Definición de la dimensión dónde teniendo en segundo plano el preview data con los datos de esta tabla.

Spoon - dimension_dónde

File Edit View Action Tools Help

Connect

View Design

Search

- Input
- Output
- Streaming
- Transform
- Utility
- Flow
- Scripting
- Pentaho Server
- Lookup
- Joins
- Data Warehouse
- Validation
- Statistics
- Big Data
- Cryptography
- Job
- Mapping
- Bulk loading
- Inline
- Experimental
- Legacy
- Deprecated
- History

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	cod_municipio	municipio	capital	comarca	provincia	comunidad_autonoma	nivel_altitud	nivel_habita
1	28001	Acebeda, La	La Acebeda	Lozoya Somosierra	Madrid	Comunidad de Madrid	5 - muy alto: 1045 m. o más	1-hasta 100
2	28002	Ajalvir	Ajalvir	Campaña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	4-de 1.001
3	28003	Alameda del Valle	Alameda del Valle	Lozoya Somosierra	Madrid	Comunidad de Madrid	5 - muy alto: 1045 m. o más	2-de 101 a
4	28004	Álamo, El	El Álamo	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	5-de 5.001
5	28004	Álamo, El	El Álamo	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	6-de 10.001
6	28004	Álamo, El	El Álamo	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	4-de 1.001
7	28005	Alcalá de Henares	Alcalá de Henares	Campaña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	8-de 100.00
8	28006	Alcobendas	Alcobendas	Area Metropolitana de Madrid	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	7-de 50.001
9	28006	Alcobendas	Alcobendas	Area Metropolitana de Madrid	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	8-de 100.00
1.	28007	Alcorcón	Alcorcón	Area Metropolitana de Madrid	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	8-de 100.00
1.	28008	Aldea del Fresno	Aldea del Fresno	Sur Occidental	Madrid	Comunidad de Madrid	2 - bajo: 265 a menos de 556 m.	4-de 1.001
1.	28009	Algete	Algete	Campaña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	6-de 10.001
1.	28010	Alpedrete	Alpedrete	Guadarrama	Madrid	Comunidad de Madrid	4 - alto: 803 a menos de 1045 m.	5-de 5.001
1.	28010	Alpedrete	Alpedrete	Guadarrama	Madrid	Comunidad de Madrid	4 - alto: 803 a menos de 1045 m.	6-de 10.001
1.	28011	Ambite	Ambite	Vegas	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	2-de 101 a
1.	28011	Ambite	Ambite	Vegas	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	3-de 501 a
1.	28012	Anchuelo	Anchuelo	Campaña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	4-de 1.001
1.	28012	Anchuelo	Anchuelo	Campaña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	3-de 501 a
1.	28013	Aranjuez	Aranjuez	Vegas	Madrid	Comunidad de Madrid	2 - bajo: 265 a menos de 556 m.	6-de 10.001
2.	28013	Aranjuez	Aranjuez	Vegas	Madrid	Comunidad de Madrid	2 - bajo: 265 a menos de 556 m.	7-de 50.001
2.	28014	Arganda del Rey	Arganda del Rey	Campaña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	7-de 50.001
2.	28014	Arganda del Rey	Arganda del Rey	Campaña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	6-de 10.001
2.	28015	Arroyomolinos	Arroyomolinos	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	4-de 1.001
2.	28015	Arroyomolinos	Arroyomolinos	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	6-de 10.001
2.	28015	Arroyomolinos	Arroyomolinos	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	5-de 5.001
2.	28016	Atazar, El	El Atazar	Lozoya Somosierra	Madrid	Comunidad de Madrid	4 - alto: 803 a menos de 1045 m.	1-hasta 100
2.	28016	Atazar, El	El Atazar	Lozoya Somosierra	Madrid	Comunidad de Madrid	4 - alto: 803 a menos de 1045 m.	2-de 101 a
2.	28017	Batres	Batres	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	4-de 1.001
2.	28017	Batres	Batres	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	3-de 501 a

Preview data de la dimensión dónde.

Spoon - dimension_cuando

File Edit View Action Tools Help

Connect

View Design

Search

- Input
- Output
- Streaming
- Transform
- Utility
- Flow
- Scripting
- Pentaho Server
- Lookup
- Joins
- Data Warehouse
- Validation
- Statistics
- Big Data
- Cryptography
- Job
- Mapping
- Bulk loading
- Inline
- Experimental
- Legacy
- Deprecated
- History

Table Input

Table output

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	periodo
1	1996
2	1998
3	1999
4	2000
5	2001
6	2002
7	2003
8	2004
9	2005
...	...

Definición de la dimensión cuándo junto con el preview data con los datos de esta tabla.

Spoon - generar_ft_madrid

File Edit View Action Tools Help

Search

General
Mail
File management
Conditions
Scripting
Bulk loading
XML
Utility
Repository
File transfer
File encryption

Connect

Start create_ft_madrid create_dim_donde dimension_donde create_dim_cuando dimension_cuando tabla_plana

Execution Results

Logging History Job metrics Metrics

2025-05-17 20:54:17.747 - cod_habitantes.0 - Finished processing (I=0, O=0, R=5012, W=5012, U=0, E=0)
2025-05-17 20:54:17.749 - cod_municipio.0 - Finished processing (I=0, O=0, R=5012, W=5012, U=0, E=0)
2025-05-17 20:54:17.755 - Sort rows 2.0 - Finished processing (I=0, O=0, R=5012, W=5012, U=0, E=0)
2025-05-17 20:54:18.086 - Merge join.0 - Finished processing (I=0, O=0, R=5021, W=5012, U=0, E=0)
2025-05-17 20:54:18.109 - Sort rows 3.0 - Finished processing (I=0, O=0, R=5012, W=5012, U=0, E=0)
2025-05-17 20:54:18.424 - Merge join 2.0 - Finished processing (I=0, O=0, R=13143, W=5012, U=0, E=0)
2025-05-17 20:54:18.444 - nivel_habitantes.0 - Finished processing (I=0, O=0, R=5012, W=5012, U=0, E=0)
2025-05-17 20:54:18.461 - Select values.0 - Finished processing (I=0, O=0, R=5012, W=5012, U=0, E=0)
2025-05-17 20:54:18.491 - Table output.0 - Finished processing (I=0, O=0, R=5012, W=5012, U=0, E=0)
2025-05-17 20:54:18.493 - generar_ft_madrid - Finished job entry [tabla_plana] (result=true)
2025-05-17 20:54:18.493 - generar_ft_madrid - Finished job entry [dimension_cuando] (result=true)
2025-05-17 20:54:18.493 - generar_ft_madrid - Finished job entry [dimension_donde] (result=true)
2025-05-17 20:54:18.493 - generar_ft_madrid - Finished job entry [create_dim_donde] (result=true)
2025-05-17 20:54:18.493 - generar_ft_madrid - Finished job entry [create_ft_madrid] (result=true)
2025-05-17 20:54:18.493 - generar_ft_madrid - Job execution finished
2025-05-17 20:54:18.493 - Spoon - Job has ended.

Modificación del trabajo para que se ejecute repetidas veces sin repetir datos.

pgAdmin 4

File Object Tools Edit View Window Help

Welcome public.dim_donde/madrid-anaraj/postgres@PostgreSQL 17 x

public.dim_donde/madrid-anaraj/postgres@PostgreSQL 17

No limit

Query Query History

1 SELECT * FROM public.dim_donde
2 ORDER BY donde_key ASC

Scratch Pad

Data Output Messages Notifications

Showing rows: 1 to 306 Page No: 1 of 1

donde_key [PK] Integer	cod_municipio character varying (5)	municipio character varying (100)	capital character varying (100)	comarca character varying (100)	provincia character varying (100)	comunidad_autonoma character varying (40)	nivel_altitud character varying (40)	nivel_habitantes character varying (100)
1	1 28001	Acebeda, La	La Acebeda	Lozoya Somosierra	Madrid	Comunidad de Madrid	5 - muy alto: 1045 m. o más	1-hasta 100 hab
2	2 28002	Ajalvir	Ajalvir	Campaña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	4-de 1.001 a 5.000
3	3 28003	Alameda del Valle	Alameda del Valle	Lozoya Somosierra	Madrid	Comunidad de Madrid	5 - muy alto: 1045 m. o más	2-de 101 a 500
4	4 28004	Álamo, El	El Álamo	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	5-de 5.001 a 10
5	5 28004	Álamo, El	El Álamo	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	6-de 10.001 a 5
6	6 28004	Álamo, El	El Álamo	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	4-de 1.001 a 5.000
7	7 28005	Alcalá de Henares	Alcalá de Henares	Campaña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	8-de 100.001 a
8	8 28006	Alcobendas	Alcobendas	Area Metropolitana de Madrid	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	7-de 50.001 a 1
9	9 28006	Alcobendas	Alcobendas	Area Metropolitana de Madrid	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	8-de 100.001 a
10	10 28007	Alcorcón	Alcorcón	Area Metropolitana de Madrid	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	8-de 100.001 a
11	11 28008	Aldea del Fresno	Aldea del Fresno	Sur Occidental	Madrid	Comunidad de Madrid	2 - bajo: 265 a menos de 556 m.	4-de 1.001 a 5.000
12	12 28009	Algete	Algete	Campaña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	6-de 10.001 a 5
13	13 28010	Alpedrete	Alpedrete	Guadarrama	Madrid	Comunidad de Madrid	4 - alto: 803 a menos de 1045 m.	5-de 5.001 a 10
14	14 28010	Alpedrete	Alpedrete	Guadarrama	Madrid	Comunidad de Madrid	4 - alto: 803 a menos de 1045 m.	6-de 10.001 a 5
15	15 28011	Ambite	Ambite	Vegas	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	2-de 101 a 500
16	16 28011	Ambite	Ambite	Vegas	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	3-de 501 a 1.000
17	17 28012	Anchuelo	Anchuelo	Campaña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	4-de 1.001 a 5.000
18	18 28012	Anchuelo	Anchuelo	Campaña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	3-de 501 a 1.000

Total rows: 306 Query complete 00:00:00.110 CRLF Ln 1, Col 1

Tabla de la dimensión dónde en la base de datos de PostgreSQL.

Query: `SELECT * FROM public.dim_cuando ORDER BY cuando_key ASC`

	cuando_key [PK] integer	periodo character varying (17)
1	1	1996
2	2	1998
3	3	1999
4	4	2000
5	5	2001
6	6	2002
7	7	2003
8	8	2004
9	9	2005
10	10	2006
11	11	2007
12	12	2008
13	13	2009
14	14	2010
15	15	2011
16	16	2012
17	17	2013
18	18	2014

Total rows: 28 Query complete 00:00:00.070

Tabla de la dimensión cuándo en la base de datos de PostgreSQL.

Ejercicio 3.4

Step name: `create_fact_padron`

SQL script to execute. (statements separated by ;)

```

DROP TABLE IF EXISTS fact_padron;
CREATE TABLE fact_padron (
  cuando_key INTEGER,
  donde_key INTEGER,
  hechos DOUBLE PRECISION,
  PRIMARY KEY (cuando_key, donde_key),
  FOREIGN KEY (cuando_key) REFERENCES dim_cuando(cuando_key),
  FOREIGN KEY (donde_key) REFERENCES dim_donde(donde_key)
);

```

Execution Results:

```

2025-05-18 18:16:23.895 - Spoon - Started the transformation execution.
2025-05-18 18:16:24.580 - create_fact_padron - Dispatching started for transformation
2025-05-18 18:16:24.628 - Carte - Installing timer to purge stale objects after 1440 min
2025-05-18 18:16:24.733 - create_fact_padron - Finished reading query, closing connection
2025-05-18 18:16:24.756 - create_fact_padron - Finished processing (I=0, O=0, R=0, V=0)
2025-05-18 18:16:24.758 - Spoon - The transformation has finished!!
2025-05-18 18:16:34.761 - create_fact_padron - Shutting down heartbeat signal for create_fact_padron
2025-05-18 18:27:58.046 - Spoon - Transformation opened.
2025-05-18 18:27:58.047 - Spoon - Launching transformation [Hechos Padron]...
2025-05-18 18:27:58.047 - Spoon - Started the transformation execution.
2025-05-18 18:27:59.674 - Spoon - The transformation has finished!!

```

Definición de la creación de la tabla `fact_padron` (TABLA DE HECHOS) con las llaves externas junto con la ejecución.

The screenshot shows the Spoon - Hechos Padron interface. The top menu bar includes File, Edit, View, Action, Tools, and Help. The left sidebar contains a search bar and a list of components under the 'Design' tab, including Input, Output, Automatic documentation output, Delete, Insert / update, JSON output, LDAP output, Microsoft Access output, Microsoft Excel writer, Pentaho reporting output, Properties output, RSS output, S3 file output, SQL file output, Salesforce delete, Salesforce insert, Salesforce update, Salesforce upsert, Serialize to file, Synchronize after merge, Table output, Text file output, Update, XML output, Streaming, Transform, Utility, Flow, Scripting, Pentaho Server, Lookups, Joins, Data Warehouse, and Validation.

The main workspace displays a data flow diagram with the following components: Table Input, lookup cuando, lookup donde, Select values, and Table output. The flow is as follows: Table Input → lookup cuando → lookup donde → Select values → Table output.

The 'Execution Results' tab is active, showing a table with the following data:

#	cundo_key	donde_key	mujeres	hombres
1	1	1	25	28
2	2	1	21	26
3	3	1	21	26
4	4	1	26	31
5	5	1	25	31
6	6	1	25	28
7	7	1	26	29
8	8	1	26	29
9	9	1	30	30
10	10	1	28	31
11	11	1	27	31
12	12	1	27	30

Definición de la tabla de hechos, fact_padron, junto el preview data del mismo.

The screenshot shows the Spoon - generar_ft_madrid interface. The top menu bar includes File, Edit, View, Action, Tools, and Help. The left sidebar contains a search bar and a list of components under the 'Design' tab, including General, Mail, File management, Conditions, Scripting, Bulk loading, XML, Utility, Repository, File transfer, and File encryption.

The main workspace displays a data flow diagram with the following components: Start, create_ft_madrid, create_dim_cundo, create_dim_donde, create_fact_padron, fact_padron, dimension_donde, dimension_cundo, and tabla_plana. The flow is as follows: Start → create_ft_madrid → create_dim_cundo → create_dim_donde → create_fact_padron → fact_padron → dimension_donde → dimension_cundo → tabla_plana.

The 'Execution Results' tab is active, showing a log of the execution process. The log includes the following entries:

- 2025-05-18 18:54:42.897 - fact_padron - Unspatching started for transformation [fact_padron]
- 2025-05-18 18:54:42.899 - Table output.0 - Connected to database [madrid-anarjer] (commit=1000)
- 2025-05-18 18:54:42.993 - Table input.0 - Finished reading query, closing connection
- 2025-05-18 18:54:42.993 - Table input.0 - Finished processing (I=5012, O=0, R=0, W=5012, U=0, E=0)
- 2025-05-18 18:54:43.330 - lookup cuando.0 - Finished processing (I=5012, O=0, R=5012, W=5012, U=0, E=0)
- 2025-05-18 18:54:43.403 - lookup donde.0 - Finished processing (I=5012, O=0, R=5012, W=5012, U=0, E=0)
- 2025-05-18 18:54:43.408 - Select values.0 - Finished processing (I=0, O=0, R=5012, W=5012, U=0, E=0)
- 2025-05-18 18:54:43.429 - Table output.0 - Finished processing (I=0, O=5012, R=5012, W=5012, U=0, E=0)
- 2025-05-18 18:54:43.430 - generar_ft_madrid - Finished job entry [fact_padron] (result=true)
- 2025-05-18 18:54:43.430 - generar_ft_madrid - Finished job entry [dimension_donde] (result=true)
- 2025-05-18 18:54:43.430 - generar_ft_madrid - Finished job entry [dimension_cundo] (result=true)
- 2025-05-18 18:54:43.430 - generar_ft_madrid - Finished job entry [create_fact_padron] (result=true)
- 2025-05-18 18:54:43.430 - generar_ft_madrid - Finished job entry [create_dim_donde] (result=true)
- 2025-05-18 18:54:43.430 - generar_ft_madrid - Finished job entry [create_dim_cundo] (result=true)
- 2025-05-18 18:54:43.430 - generar_ft_madrid - Finished job entry [create_ft_madrid] (result=true)
- 2025-05-18 18:54:43.430 - Spoon - Job has ended.

Modificación del trabajo generar_ft_madrid para no generar duplicados tras haber añadido los pasos para la tabla de hechos.

pgAdmin 4

Welcome public.fact_padron/madrid-anaraj/postgres@PostgreSQL 17

Query Query History

```
1 SELECT * FROM public.fact_padron
2 ORDER BY cuando_key ASC, donde_key ASC
```

Data Output Messages Notifications

Showing rows: 1 to 1000 Page No: 1 of 6

	cuando_key [PK] integer	donde_key [PK] integer	hombres double precision	mujeres double precision
1	1	1	28	25
2	1	2	842	780
3	1	3	93	73
4	1	4	1991	1887
5	1	7	81302	82084
6	1	9	40979	42052
7	1	10	69960	71505
8	1	11	634	602
9	1	12	6293	6042
10	1	14	2611	2600
11	1	16	144	146
12	1	17	327	269
13	1	20	18804	20096
14	1	22	14491	14733
15	1	24	1368	1341
16	1	26	52	44
17	1	28	322	317
18	1	30	1406	1301

Total rows: 5012 Query complete 00:00:00.153 CRLF Ln 1, Col 1

Tabla de la tabla de hechos en la base de datos de PostgreSQL.

pgAdmin 4

Welcome public.fact_padron... public.dim_donde/... public.dim_cuando/madrid-anaraj/postgres@PostgreSQL 17

Query Query History

```
1 SELECT * FROM public.dim_cuando
2 ORDER BY cuando_key ASC
```

Data Output Messages Notifications

Showing rows: 1 to 28 Page No: 1 of 1

	cuando_key [PK] integer	periodo character varying (17)
1	1	1996
2	2	1998
3	3	1999
4	4	2000
5	5	2001
6	6	2002
7	7	2003
8	8	2004
9	9	2005
10	10	2006
11	11	2007
12	12	2008
13	13	2009
14	14	2010
15	15	2011
16	16	2012
17	17	2013
18	18	2014

Total rows: 28 Query complete 00:00:00.069 CRLF Ln 1, Col 1

Tabla de la dimensión cuándo en la base de datos de PostgreSQL.

pgAdmin 4

Welcome public.cuando_ma... x public.donde_madrid/madrid_olap/postgres@PostgreSQL 17 x

public.donde_madrid/madrid_olap/postgres@PostgreSQL 17

Query Query History

```
1 SELECT * FROM public.donde_madrid
2 ORDER BY donde_key ASC
```

Data Output Messages Notifications

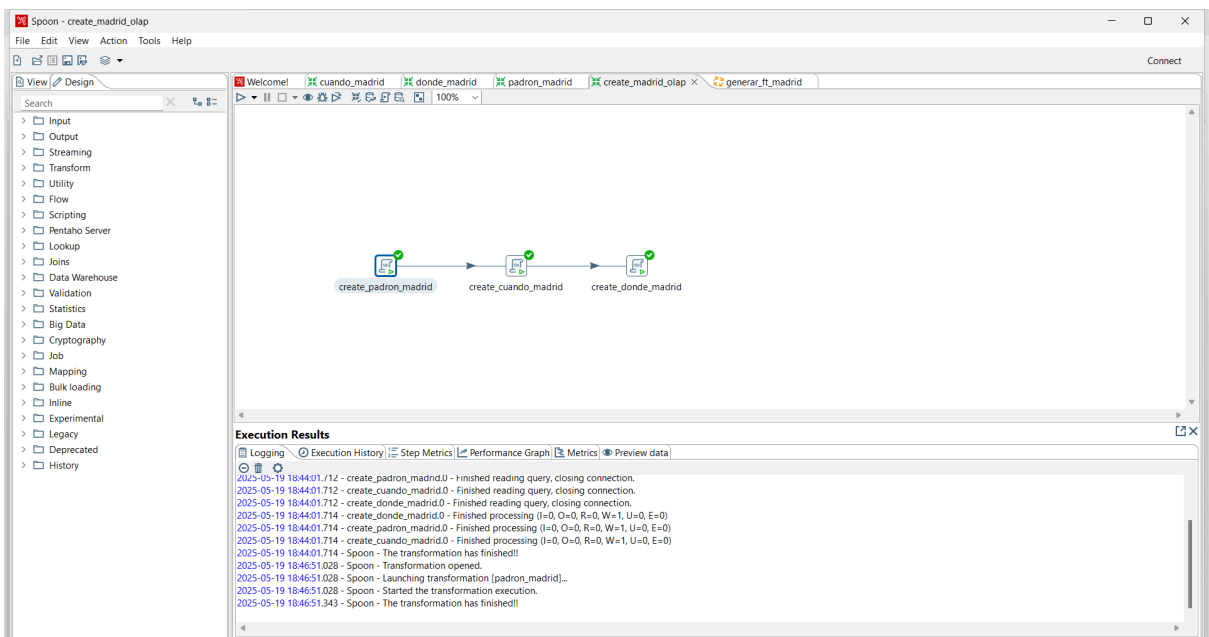
Showing rows: 1 to 306 Page No: 1 of 1

	donde_key [PK] integer	cod_municipio character varying (5)	municipio character varying (100)	capital character varying (100)	comarca character varying (100)	provincia character varying (100)	comunidad_autonoma character varying (40)	nivel_altitud character varying (40)	nivel_habitantes character varying (40)
1	1	28001	Acebeda, La	La Acebeda	Lozoya Somosierra	Madrid	Comunidad de Madrid	5 - muy alto: 1045 m. o más	1-hasta 100 ha
2	2	28002	Aljávir	Aljávir	Campiña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	4-de 1.001 a 5.000
3	3	28003	Alameda del Valle	Alameda del Valle	Lozoya Somosierra	Madrid	Comunidad de Madrid	5 - muy alto: 1045 m. o más	2-de 101 a 500
4	4	28004	Álamo, El	El Álamo	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	4-de 1.001 a 5.000
5	5	28004	Álamo, El	El Álamo	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	6-de 10.001 a 50.000
6	6	28004	Álamo, El	El Álamo	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	5-de 5.001 a 10.000
7	7	28005	Alcalá de Henares	Alcalá de Henares	Campiña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	8-de 100.001 a 500.000
8	8	28006	Alcobendas	Alcobendas	Area Metropolitana de Madrid	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	8-de 100.001 a 500.000
9	9	28006	Alcobendas	Alcobendas	Area Metropolitana de Madrid	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	7-de 50.001 a 100.000
10	10	28007	Alcorcón	Alcorcón	Area Metropolitana de Madrid	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	8-de 100.001 a 500.000
11	11	28008	Aldea del Fresno	Aldea del Fresno	Sur Occidental	Madrid	Comunidad de Madrid	2 - bajo: 265 a menos de 556 m.	4-de 1.001 a 5.000
12	12	28009	Algete	Algete	Campiña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	6-de 10.001 a 50.000
13	13	28010	Alpedrete	Alpedrete	Guadarrama	Madrid	Comunidad de Madrid	4 - alto: 803 a menos de 1045 m.	6-de 10.001 a 50.000
14	14	28010	Alpedrete	Alpedrete	Guadarrama	Madrid	Comunidad de Madrid	4 - alto: 803 a menos de 1045 m.	5-de 5.001 a 10.000
15	15	28011	Ambite	Ambite	Vegas	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	3-de 501 a 1.000
16	16	28011	Ambite	Ambite	Vegas	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	2-de 101 a 500
17	17	28012	Anchuelo	Anchuelo	Campiña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	3-de 501 a 1.000

Total rows: 306 Query complete 00:00:00.196 CRLF Ln 1, Col 1

Tabla de la dimensión dónde en la base de datos de PostgreSQL.

Ejercicio 3.5



Definición de las creaciones de la tabla de hechos y de las dimensiones cuándo y dónde, con la conectividad a la nueva base de datos, madrid_olap.

Execution Results

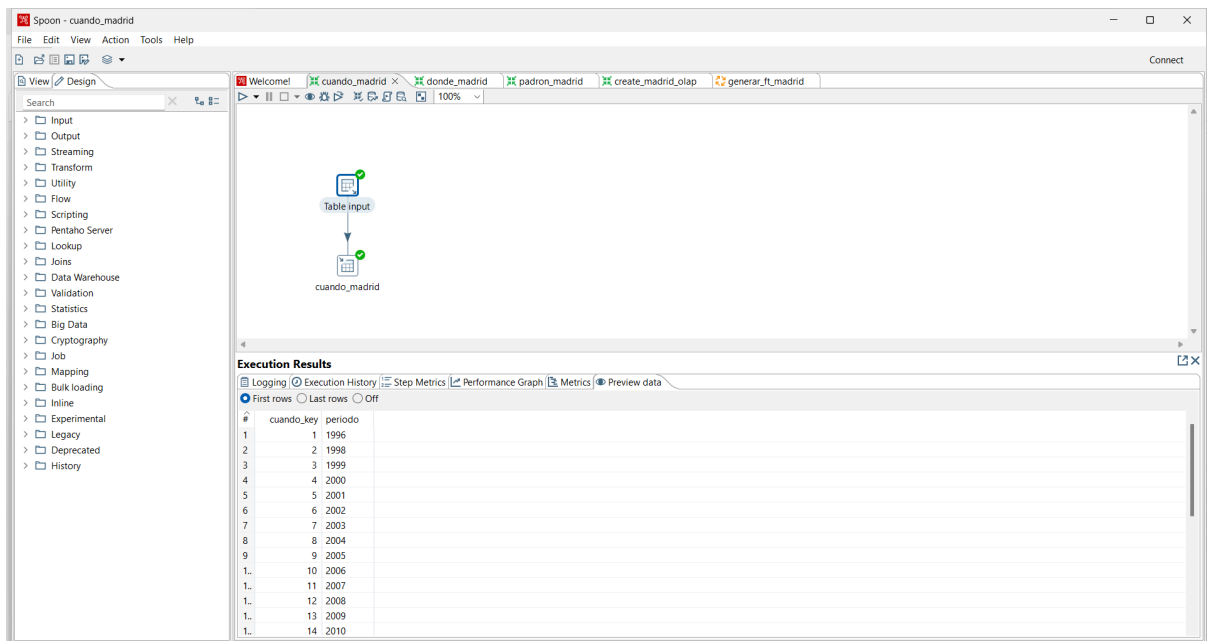
#	cuando_key	donde_key	hombres	mujeres
1	1	1	28.0	25.0
2	2	1	26.0	21.0
3	3	1	26.0	21.0
4	4	1	31.0	26.0
5	5	1	31.0	25.0
6	6	1	28.0	25.0
7	7	1	29.0	26.0
8	8	1	29.0	26.0
9	9	1	30.0	30.0
1.	10	1	31.0	28.0
1.	11	1	31.0	27.0
1.	12	1	30.0	27.0
1.	13	1	33.0	28.0
1.	14	1	30.0	29.0
1.	15	1	32.0	30.0
1.	16	1	36.0	29.0
1.	17	1	37.0	30.0
1.	18	1	37.0	30.0
1.	19	1	36.0	29.0
2.	20	1	36.0	30.0
2	21	1	37.0	29.0

Definición de la tabla de hechos en la nueva base de datos, madrid_olap, junto el preview data tras ejecutar.

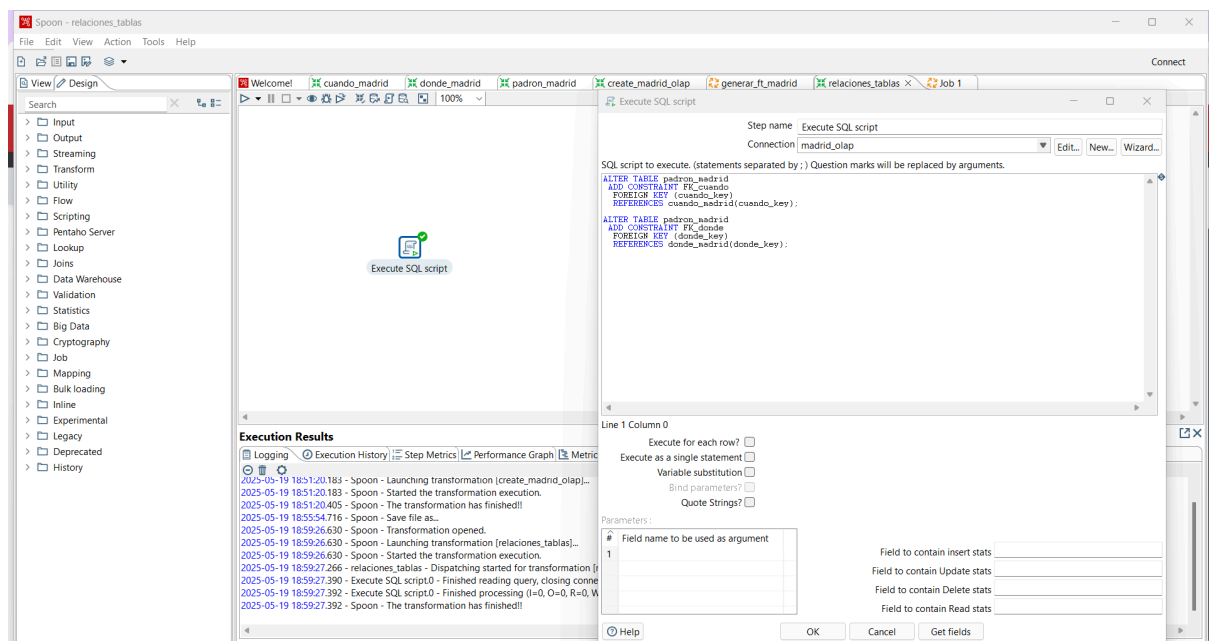
Execution Results

#	donde_key	cod_municipio	municipio	capital	comarca	provincia	comunidad_autonoma	nivel_altitud
1	1	28001	Acebeda, La	La Acebeda	Lozoya Somosierra	Madrid	Comunidad de Madrid	5 - muy alto: 1045 m. o más
2	2	28002	Ajalvir	Ajalvir	Campiña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 n
3	3	28003	Alameda del Valle	Alameda del Valle	Lozoya Somosierra	Madrid	Comunidad de Madrid	5 - muy alto: 1045 m. o más
4	4	28004	Alamo, El	El Alamo	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 n
5	5	28004	Alamo, El	El Alamo	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 n
6	6	28004	Alamo, El	El Alamo	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 n
7	7	28005	Alcalá de Henares	Alcalá de Henares	Campiña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 n
8	8	28006	Alcobendas	Alcobendas	Area Metropolitana de Madrid	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 n

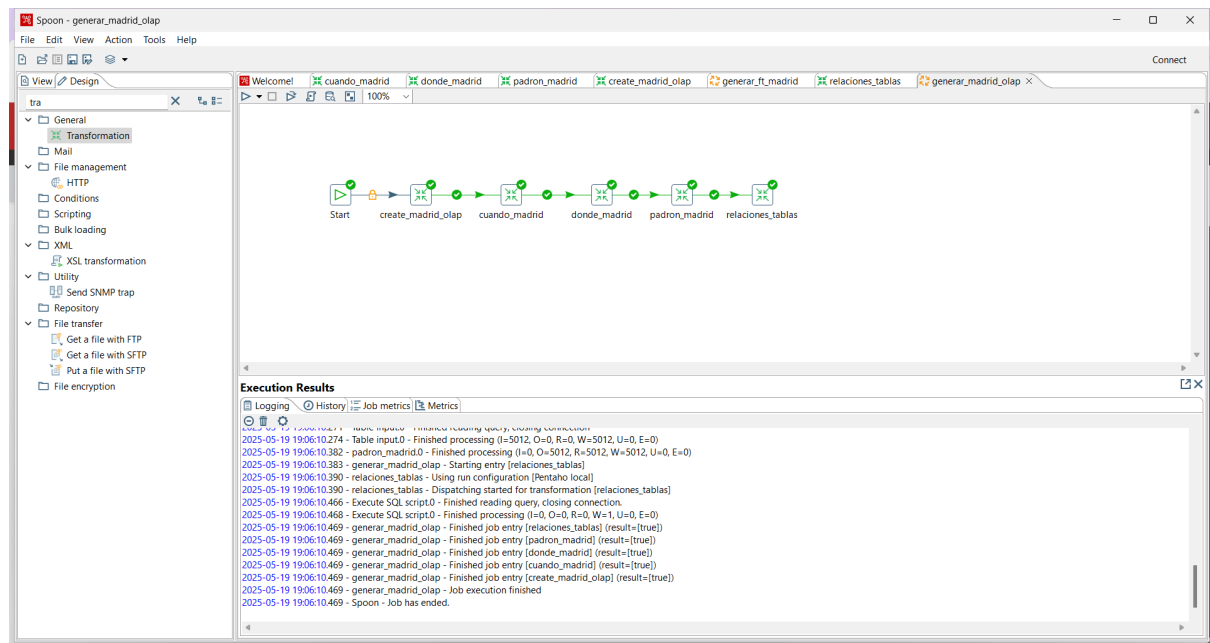
Definición de la dimensión dónde en la nueva base de datos, madrid_olap, junto el preview data tras ejecutar.



Definición de la dimensión cuándo en la nueva base de datos, madrid_olap, junto el preview data tras ejecutar.

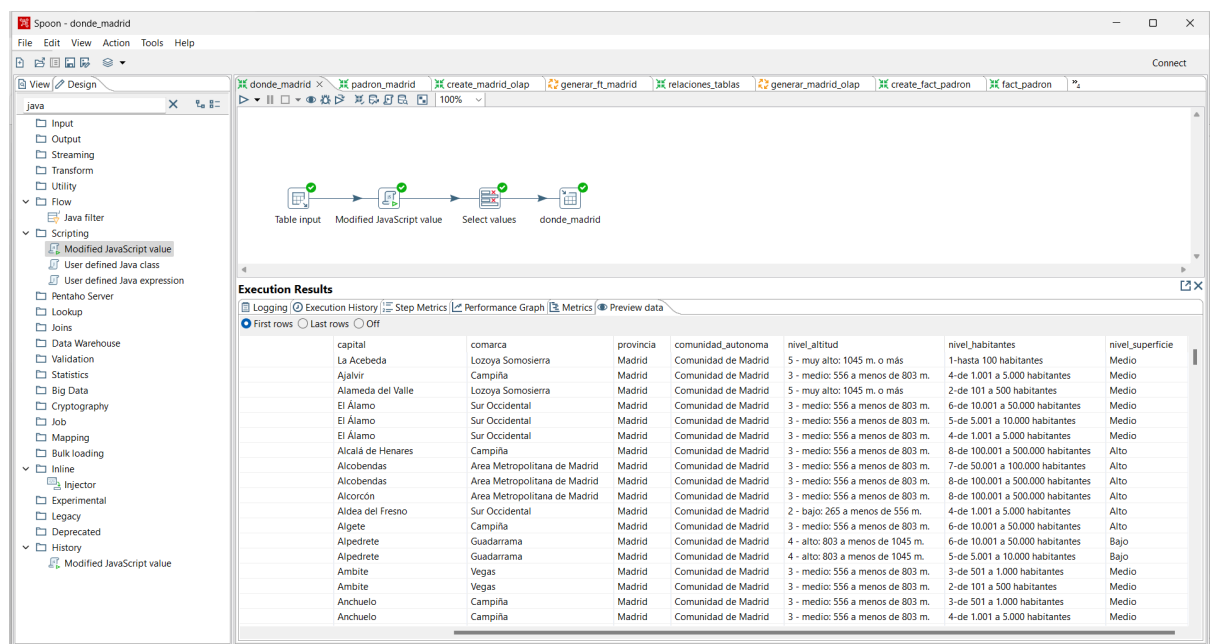


Ejecución del script con las relaciones entre las distintas tablas.



Trabaja en la nueva base de datos, madrid_olap.

Ejercicio 3.6



Para mostrar el nivel_superficie, se ha añadido en la base de datos madrid_anarajer, de la que se recogen los datos para esta base de datos, madrid_olap. Se incluye en esta base de datos en la transformación create_madrid_olap, concretamente en el script de dim_donde. También se añade en la transformación de la dimensión dónde, donde_madrid.

Además, el script usado en donde_madrid para sacar los niveles de superficie es el siguiente:


```

if (superficie_ha < 1500) {
nivel_superficie = "Bajo";
} else if (superficie_ha < 3000) {
nivel_superficie = "Medio";
} else {
nivel_superficie = "Alto";
}

```

The screenshot shows the Apache NiFi web console interface. On the left is a sidebar with a tree view of components categorized under 'Input', 'Output', 'Streaming', 'Transform', 'Utility', 'Flow', 'Scripting', 'Pentaho Server', 'Lookup', 'Joins', 'Data Warehouse', 'Validation', 'Statistics', 'Big Data', 'Cryptography', 'Job', 'Mapping', 'Bulk loading', 'Inline', 'Experimental', 'Legacy', 'Deprecated', and 'History'. The 'Scripting' category is expanded, showing 'Modified JavaScript value', 'User defined Java class', and 'User defined Java expression'. The main canvas displays a data flow with three components: 'Table input', 'Modified JavaScript value', and 'cuando_madrid', connected sequentially. Below the canvas, the 'Execution Results' tab is active, showing a table of data rows.

#	cuando_key	periodo	decenio
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
1.	10	2006	2000
1.	11	2007	2000
1.	12	2008	2000
1.	13	2009	2000
1.	14	2010	2010
1.	15	2011	2010
1.	16	2012	2010
1.	17	2013	2010
1.	18	2014	2010
1.	19	2015	2010
2.	20	2016	2010
2.	21	2017	2010

Se ha hecho de la misma manera para los decenios, incluyendo en la transformación de creación de las tablas la nueva variable, así como en la transformación de la dimensión cuándo, cuándo_madrid.

Además, el script usado en cuando_madrid para sacar los niveles de superficie es el siguiente:

```
decenio = Math.floor(periodo / 10) * 10;
```

He hecho uso de Math.floor porque decenio se ha almacenado como tipo String.

pgAdmin 4

Welcome public.cuando_madrid/madrid_olap/postgres@PostgreSQL 17 X

public.cuando_madrid/madrid_olap/postgres@PostgreSQL 17

Query Query History

```
1 SELECT * FROM public.cuando_madrid
2 ORDER BY cuando_key ASC
```

Data Output Messages Notifications

Showing rows: 1 to 28 Page No: 1 of 1

	cuando_key [PK] integer	periodo character varying (17)	decenio character varying (4)
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
11	11	2007	2000
12	12	2008	2000
13	13	2009	2000
14	14	2010	2010
15	15	2011	2010
16	16	2012	2010
17	17	2013	2010
18	18	2014	2010

Total rows: 28 Query complete 00:00:00.129 CRLF Ln 1, Col 1

Tabla de la dimensión cuándo en madrid_olap con los decenios.

pgAdmin 4

Welcome public.donde_madrid/madrid_olap/postgres@PostgreSQL 17 X

public.donde_madrid/madrid_olap/postgres@PostgreSQL 17

Query Query History

```
1 SELECT * FROM public.donde_madrid
2 ORDER BY donde_key ASC
```

Data Output Messages Notifications

Showing rows: 1 to 306 Page No: 1 of 1

	icipo character varying (100)	capital character varying (100)	comarca character varying (100)	provincia character varying (100)	comunidad_autonoma character varying (40)	nivel_altitud character varying (40)	nivel_superficie character varying (40)	nivel_habitantes character varying (40)
1	beda, La	La Acebeda	Lozoya Somosierra	Madrid	Comunidad de Madrid	5 - muy alto: 1045 m. o más	Medio	1-hasta 100 habitantes
2	vir	Ajalvir	Campiña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	Medio	4-de 1.001 a 5.000 habitantes
3	meda del Valle	Alameda del Valle	Lozoya Somosierra	Madrid	Comunidad de Madrid	5 - muy alto: 1045 m. o más	Medio	2-de 101 a 500 habitantes
4	mo, El	El Álamo	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	Medio	6-de 10.001 a 50.000 habitantes
5	mo, El	El Álamo	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	Medio	5-de 5.001 a 10.000 habitantes
6	mo, El	El Álamo	Sur Occidental	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	Medio	4-de 1.001 a 5.000 habitantes
7	álá de Henares	Alcalá de Henares	Campiña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	Alto	8-de 100.001 a 500.000 habitantes
8	obendas	Alcobendas	Area Metropolitana de Madrid	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	Alto	7-de 50.001 a 100.000 habitantes
9	obendas	Alcobendas	Area Metropolitana de Madrid	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	Alto	8-de 100.001 a 500.000 habitantes
10	orcón	Alcorcón	Area Metropolitana de Madrid	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	Alto	8-de 100.001 a 500.000 habitantes
11	aa del Fresno	Aldea del Fresno	Sur Occidental	Madrid	Comunidad de Madrid	2 - bajo: 265 a menos de 556 m.	Alto	4-de 1.001 a 5.000 habitantes
12	ete	Algete	Campiña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	Alto	6-de 10.001 a 50.000 habitantes
13	edrete	Alpedrete	Guadarrama	Madrid	Comunidad de Madrid	4 - alto: 803 a menos de 1045 m.	Bajo	6-de 10.001 a 50.000 habitantes
14	edrete	Alpedrete	Guadarrama	Madrid	Comunidad de Madrid	4 - alto: 803 a menos de 1045 m.	Bajo	5-de 5.001 a 10.000 habitantes
15	bite	Ambite	Vegas	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	Medio	3-de 501 a 1.000 habitantes
16	bite	Ambite	Vegas	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	Medio	2-de 101 a 500 habitantes
17	huelo	Anchuelo	Campiña	Madrid	Comunidad de Madrid	3 - medio: 556 a menos de 803 m.	Medio	3-de 501 a 1.000 habitantes

Total rows: 306 Query complete 00:00:00.157 CRLF Ln 1, Col 1

Tabla de la dimensión dónde en madrid_olap con el nivel_superficie.

Ejercicio 3.7

The screenshot shows the Spoon IDE interface for a job named 'generar_ft_madrid'. The 'Design' view displays a simple flow: Start → create_ft_madrid → tabla_plana. The 'Execution Results' view shows a log of the job's execution, including timestamps and status messages for each step.

Execution Results

- 2025-05-19 20:22:12.903 - Sort rows 2.0 - Finished processing (I=0, O=0, R=5012, W=5012, U=0, E=0)
- 2025-05-19 20:22:13.239 - Merge join 0 - Finished processing (I=0, O=0, R=5021, W=5012, U=0, E=0)
- 2025-05-19 20:22:13.247 - Sort rows 3.0 - Finished processing (I=0, O=0, R=5012, W=5012, U=0, E=0)
- 2025-05-19 20:22:13.558 - Merge join 2.0 - Finished processing (I=0, O=0, R=13143, W=5012, U=0, E=0)
- 2025-05-19 20:22:13.560 - nivel_habitantes 0 - Finished processing (I=0, O=0, R=5012, W=5012, U=0, E=0)
- 2025-05-19 20:22:13.563 - Select values 0 - Finished processing (I=0, O=0, R=5012, W=5012, U=0, E=0)
- 2025-05-19 20:22:13.589 - Table output 0 - Finished processing (I=0, O=5012, R=5012, W=5012, U=0, E=0)
- 2025-05-19 20:22:13.590 - generar_ft_madrid - Finished job entry [tabla_plana] (result=true)
- 2025-05-19 20:22:13.590 - generar_ft_madrid - Finished job entry [create_ft_madrid] (result=true)
- 2025-05-19 20:22:13.590 - generar_ft_madrid - Job execution finished
- 2025-05-19 20:22:13.591 - Spoon - Job has ended.

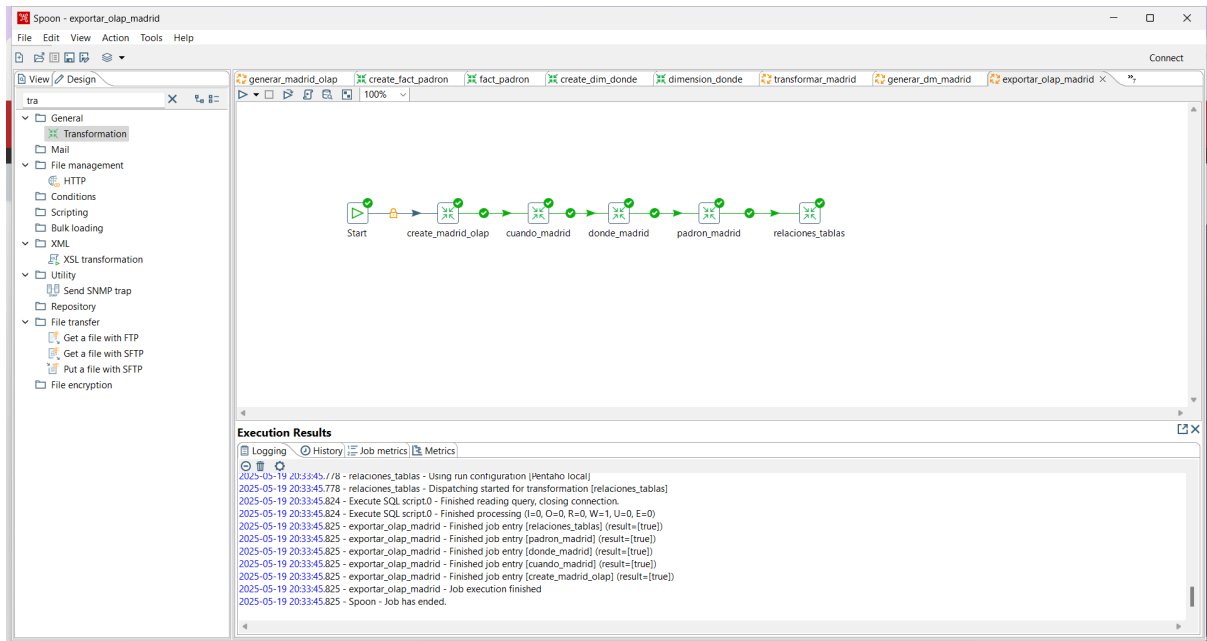
Trabajo generar_ft_madrid junto con su ejecución.

The screenshot shows the Spoon IDE interface for a job named 'generar_dm_madrid'. The 'Design' view displays a complex flow involving multiple steps: Start → create_fact_padron → create_dim_cuando → create_dim_donde → fact_padron → dimension_donde → dimension_cuando. The 'Execution Results' view shows a detailed log of the job's execution, including timestamps and status messages for each step.

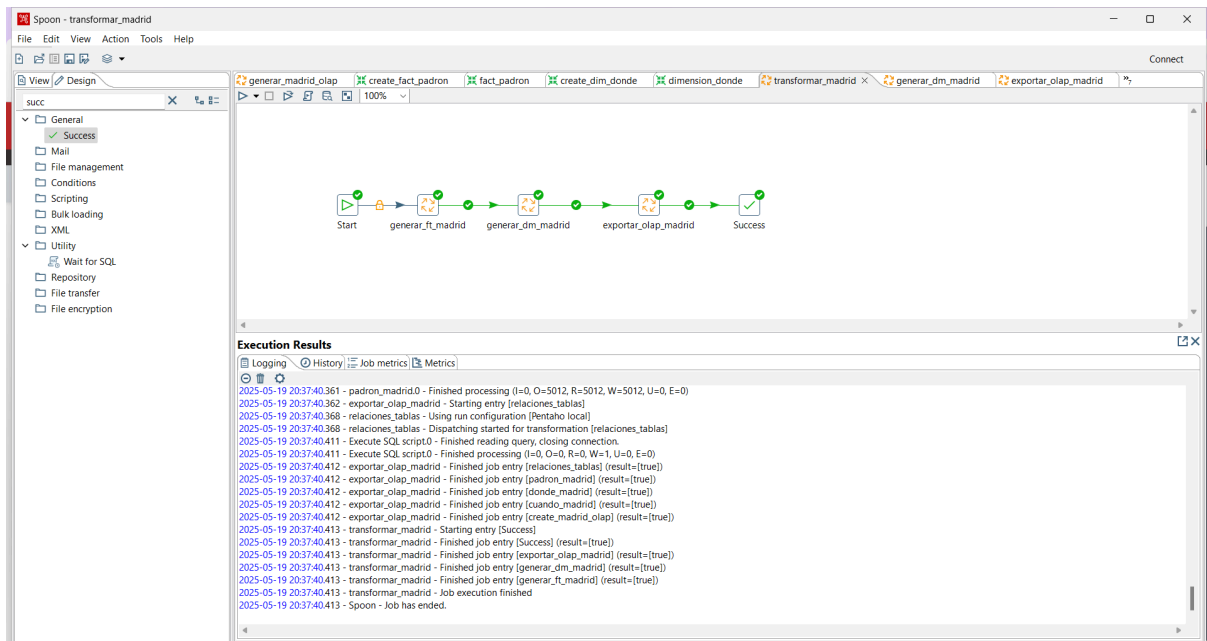
Execution Results

- 2025-05-19 20:27:45.821 - Table output 0 - Finished processing (I=0, O=306, R=306, W=306, U=0, E=0)
- 2025-05-19 20:27:45.823 - generar_dm_madrid - Starting entry [fact_padron]
- 2025-05-19 20:27:45.829 - fact_padron - Using run configuration [Pentaho local]
- 2025-05-19 20:27:45.830 - fact_padron - Dispatching started for transformation [fact_padron]
- 2025-05-19 20:27:45.830 - Table output 0 - Connected to database [madrid-anarjer] (commit=1000)
- 2025-05-19 20:27:46.032 - Table input 0 - Finished reading query, closing connection
- 2025-05-19 20:27:46.033 - Table input 0 - Finished processing (I=5012, O=0, R=0, W=5012, U=0, E=0)
- 2025-05-19 20:27:46.368 - lookup cuando 0 - Finished processing (I=5012, O=0, R=5012, W=5012, U=0, E=0)
- 2025-05-19 20:27:46.459 - lookup donde 0 - Finished processing (I=5012, O=0, R=5012, W=5012, U=0, E=0)
- 2025-05-19 20:27:46.463 - Select values 0 - Finished processing (I=0, O=0, R=5012, W=5012, U=0, E=0)
- 2025-05-19 20:27:46.476 - Table output 0 - Finished processing (I=0, O=5012, R=5012, W=5012, U=0, E=0)
- 2025-05-19 20:27:46.478 - generar_dm_madrid - Finished job entry [fact_padron] (result=true)
- 2025-05-19 20:27:46.478 - generar_dm_madrid - Finished job entry [dimension_donde] (result=true)
- 2025-05-19 20:27:46.478 - generar_dm_madrid - Finished job entry [dimension_cuando] (result=true)
- 2025-05-19 20:27:46.478 - generar_dm_madrid - Finished job entry [create_dim_cuando] (result=true)
- 2025-05-19 20:27:46.478 - generar_dm_madrid - Finished job entry [create_fact_padron] (result=true)
- 2025-05-19 20:27:46.478 - generar_dm_madrid - Job execution finished
- 2025-05-19 20:27:46.478 - Spoon - Job has ended.

Trabajo generar_dm_madrid junto con su ejecución.



Trabajo exportar_olap_madrid junto con su ejecución.



Trabajo transformar_madrid, que contiene a los trabajos mencionados anteriormente, junto con una correcta ejecución.