



Universitat Oberta
de Catalunya

Máster universitario de Ciencia de Datos

Prueba de Evaluación Continua – PEC2

**Diseño y uso de bases de datos analíticas – Diseño
multidimensional y toma de contacto con la creación de
procesos ETL**

Autor:

Mario Ubierna San Mamés

Índice de Contenido

Índice de Contenido	3
Índice de tablas	4
Índice de ilustraciones	5
1. Parte teórica.....	7
1.1. Pregunta 1	7
1.2. Pregunta 2	8
1.3. Pregunta 3	11
1.4. Pregunta 4	12
2. Parte práctica	14
2.1. Pregunta 5	14
3. Bibliografía	25

Índice de tablas

Tabla 1 - Dimensiones del hecho FACT_Composition.	7
--	---

Índice de ilustraciones

Ilustración 1 - Diseño lógico pregunta 1.	8
Ilustración 2 - Diseño físico del hecho FACT_WineOrders.	9
Ilustración 3 Modelo conceptual pregunta 3.	11
Ilustración 4 - Diseño conceptual FACT_Sarampion.	12
Ilustración 5 - Diseño conceptual FACT_Sarampion desnormalizado.	12
Ilustración 6 - Creación de la tabla STG_Denuncias_Infracciones.	14
Ilustración 7 - Comprobación de la creación de la tabla.	15
Ilustración 8 - Selección de la hoja "Datos_tratados".	16
Ilustración 9 - Especificación del comienzo de la fila de lectura.	16
Ilustración 10 - Lectura de los campos a partir de "Get fields from header row" ...	17
Ilustración 11 - Definición del nombre de los campos y su tipo.	17
Ilustración 12 - Desmarcamos la casilla de la cabecera.	18
Ilustración 13 - Visualización de los datos.	18
Ilustración 14 – Normalización de los datos.	19
Ilustración 15 - Ordenación de los campos.	20
Ilustración 16 - Conexión a la base de datos.	21
Ilustración 17 - Target table.	21
Ilustración 18 - Truncate table.	21
Ilustración 19 - Configuración de Table Output.	22
Ilustración 20 - Ejecución de la transformación.	22
Ilustración 21 - Visualización de la pestaña step metrics.	22
Ilustración 22 - Consulta SQL para saber el número de registros.	23
Ilustración 23 - Número de registros cargados en la base de datos.	23
Ilustración 24 - Consulta SQL top 10 registros.	23
Ilustración 25 - Top 10 registros.	24

Ilustración 26 - Top 10 registros de la tabla Staging..... 24

1. Parte teórica

1.1. Pregunta 1

Disponemos de la tabla de hechos FACT_Composition que almacena información sobre las composiciones de alimentos, nutriente a nutriente, formando un esquema con forma de estrella con sus dimensiones.

Dibujad el diagrama del modelo lógico correspondiente a la tabla de hechos FACT_Composition, con los siguientes atributos descriptores de las dimensiones detalladas a continuación:

Dimensiones	Atributos descriptores
DIM_Country	Código, nombre_país, nombre_region
DIM_Date	Año, mes
DIM_Food	Código, familia, nombre, especie
DIM_Type	Código, nombre
DIM_Processing	Código, nombre
DIM_Component	Código, nombre, unidad, grupo

Tabla 1 - Dimensiones del hecho FACT_Composition.

Dado el hecho y cada una de las dimensiones, obtendríamos el siguiente modelo lógico, cabe destacar que se ha definido el tipo de los datos según su nombre y esto puede corresponderse o no con el diseño físico ya que no tenemos información suficiente:

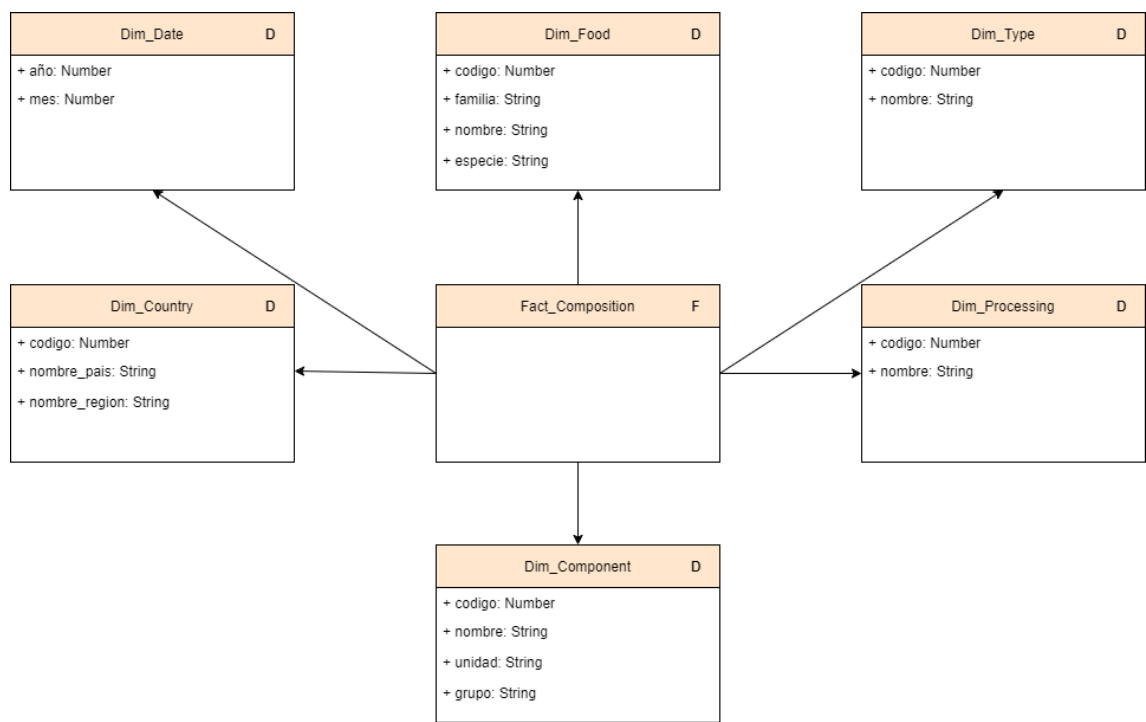


Ilustración 1 - Diseño lógico pregunta 1.

1.2. Pregunta 2

Disponemos del diseño físico siguiente:

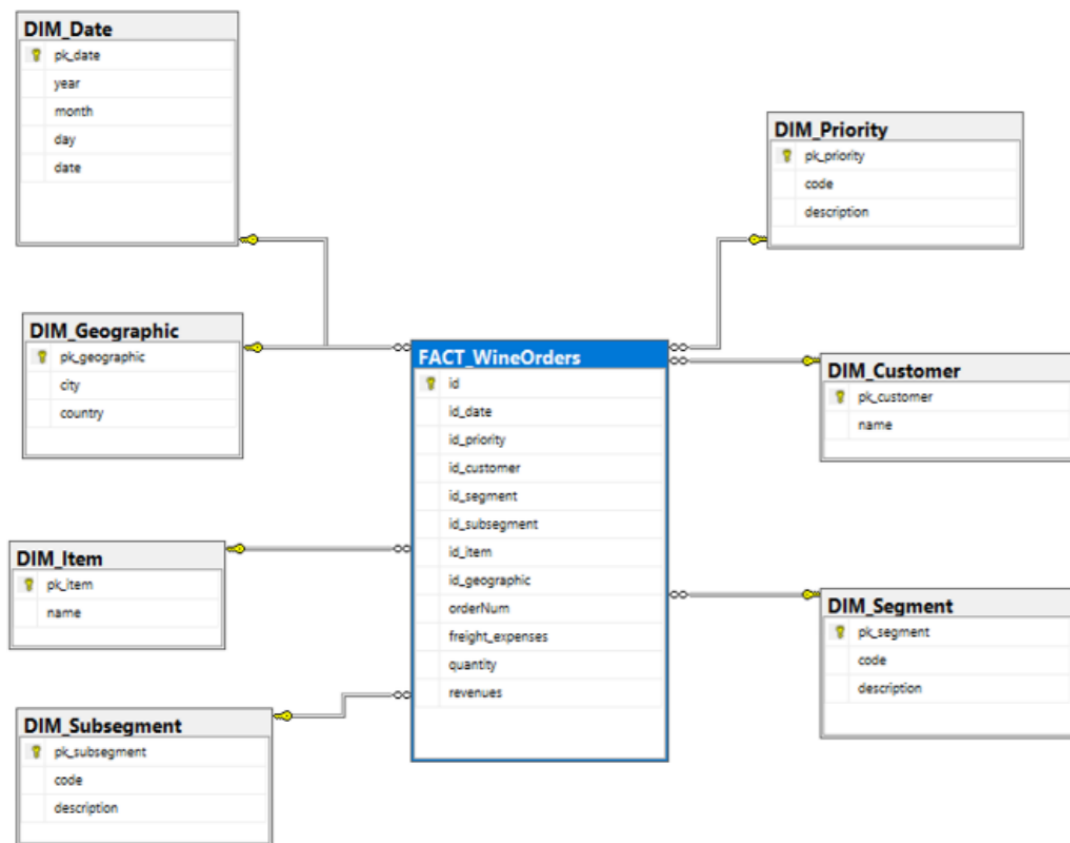


Ilustración 2 - Diseño físico del hecho FACT_WineOrders.

Este modelo está basado en información sobre pedidos de venta de botellas de vino. Nos indican el significado de los siguientes campos:

- **orderNum**: Número de pedido (numérico).
- **freight_expenses**: Gastos de envío (numérico).
- **quantity**: Cantidad solicitada en el pedido (numérico).
- **revenues**: Importe de ese pedido (numérico).

A partir de esta información, indicad si las siguientes afirmaciones son correctas o no, justificando tus respuestas.

- a) El campo *[orderNum]* se podría definir como una dimensión degenerada.

Esta afirmación es correcta, hay veces en las que podemos tener una dimensión que no tiene atributos pero es útil para identificar las instancias de un hecho, es decir, nos puede resultar interesante el identificar las ventas y agrupar todas aquellas que se han hecho con el mismo pedido. Cabe destacar que si fuera una dimensión degenerada no tendríamos atributos en dicha dimensión, sino que tendríamos un atributo en el hecho y éste pertenecería a la clave.

- b) El diagrama físico no es correcto porque en la tabla de hecho faltaría añadir también como PKs los campos que se relacionan con las dimensiones: *[id_date]*, *[id_priority]*, *[id_customer]*, *[id_segment]*, *[id_subsegment]*, *[id_item]* y *[id_geographic]*.

Esta afirmación es falsa, tal y como viene definida en la teoría la clave primaria de un hecho está formada exclusivamente por la concatenación de las claves foráneas correspondientes a una Base del Hecho. Por lo tanto, es cierto que la clave primaria del hecho estaría formada por: *[id_date]*, *[id_priority]*, *[id_customer]*, *[id_segment]*, *[id_subsegment]*, *[id_item]* y *[id_geographic]*, ya que éstas son las claves foráneas a la Base del Hecho, pero no tiene sentido que el atributo *[id]* pertenezca a la clave primaria, ya que de pertenecer a la clave primaria permitiría que no hubiera valores únicos, es decir, podríamos tener la misma fecha, prioridad, cliente, segmento, subsegmento, artículo y lugar, lo cual haría carecer de sentido la clave primaria.

- c) Se dispone de 7 dimensiones (*DIM_Date*, *DIM_Priority*, *DIM_Customer*, *DIM_Segment*, *DIM_Subsegment*, *DIM_Item* y *DIM_Geographic*) y de 4 medidas (*orderNum*, *freight_expenses*, *quantity* y *revenues*).

Esta afirmación es correcta, tenemos 7 dimensiones que podemos apreciar en el diseño (*DIM_Date*, *DIM_Priority*, *DIM_Customer*, *DIM_Segment*, *DIM_Subsegment*, *DIM_Item* y *DIM_Geographic*) y tenemos “4” medidas, establezco el cuatro entre comillas porque una medida que es *orderNum* podría considerarse una medida como tal si queremos mostrar el número de la orden, concatenar órdenes... Pero tal y como hemos visto en el apartado “a” podría ser una dimensión degenerada, pero para ello en el diseño físico este atributo tendría que pertenecer a la clave primaria, y por el diseño que nos han proporcionado este atributo al no pertenecer a la clave primaria no lo podemos considerar como una dimensión degenerada.

- d) El diagrama físico es correcto porque el campo *[id]* de la tabla de hechos corresponde con una clave subrogada y éstas siempre se deben definir en las tablas de hechos.

Esta afirmación es falsa, es verdad que podemos tener claves subrogadas que permitan reducir el tamaño de la tabla del hecho, pero estas claves subrogadas nunca se definen en el hecho lo hacemos en las dimensiones ya que la clave primaria del hecho, tal y como hemos visto, sería la combinación de cada una de las claves foráneas correspondientes a las dimensiones.

- e) Ninguna de las anteriores es correcta.

Esta afirmación es falsa, tanto el apartado “a” como “c” son correctos.

1.3. Pregunta 3

Disponemos de dos modelos conceptuales diseñados a partir de un conjunto de fuentes con información correspondiente al sector vitivinícola.

Ambos modelos se refieren a un único *data mart*, dado que principalmente se basan en una única área temática. La información que se almacena el *data mart* está compuesta de las siguientes tablas de hechos:

- **FACT_WineRatings:** Datos de calificaciones y reseñas de vino.
- **FACT_WineStatistics:** Datos con diferentes estadísticas relacionadas con el sector vitivinícola.

A partir de estos dos modelos:

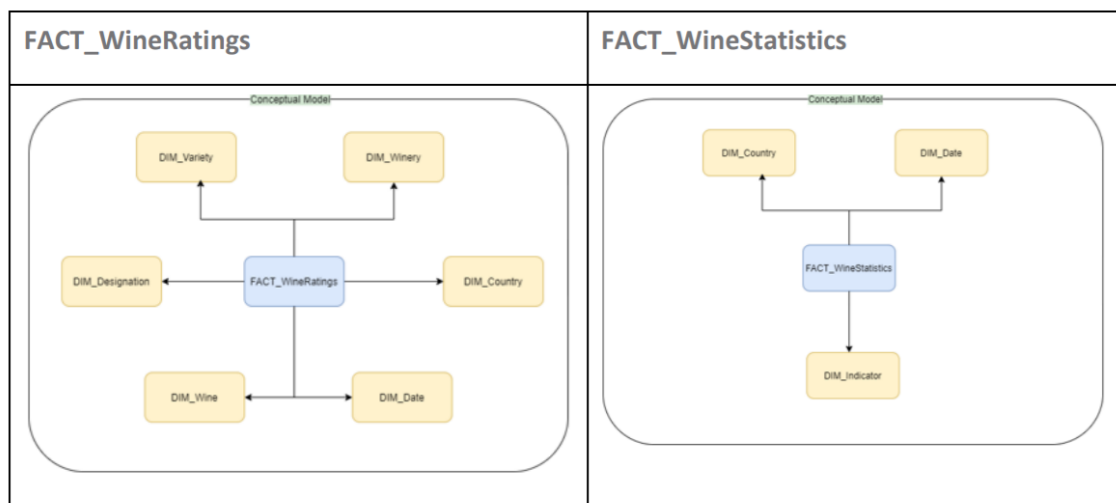


Ilustración 3 Modelo conceptual pregunta 3.

Justificad brevemente si puede existir alguna dimensión conformada en el modelo planteado.

Antes que nada tenemos que definir qué es una dimensión conformada:

- Una dimensión conformada es aquella dimensión que es utilizada en más de una Estrella, eso no significa que se tenga que usar la misma dimensión tal cual, puede variar el grado de agregación de las dimensiones compartidas en diferentes Estrellas.

Por lo tanto, tal y como podemos apreciar en las Estrellas proporcionadas, sí que tenemos dimensiones comunes (*DIM_Country* y *DIM_Date*), pero que tengamos dimensiones comunes no significa como bien hemos mencionado que deban tener el

mismo grado de agregación, por ejemplo la fecha en ratings puede ser mensual mientras que en statistics sea diaria.

1.4. Pregunta 4

Disponemos de la tabla de hechos *FACT_SARAMPION*, basada en información sobre la cobertura de inmunización. Su modelo de datos es el siguiente:

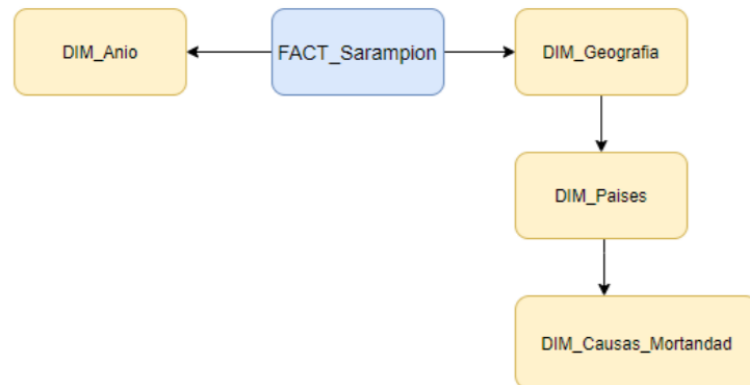


Ilustración 4 - Diseño conceptual FACT_Sarampion.

De acuerdo con este modelo, indicad si las siguientes afirmaciones son correctas o no, justificando brevemente todas las respuestas.

- a) Si en el diagrama del modelo conceptual se desnormaliza la información de las causas principales de mortandad en la dimensión *DIM_PAISES*, el diagrama del diseño conceptual de la tabla de hechos *FACT_SARAMPION* sería:

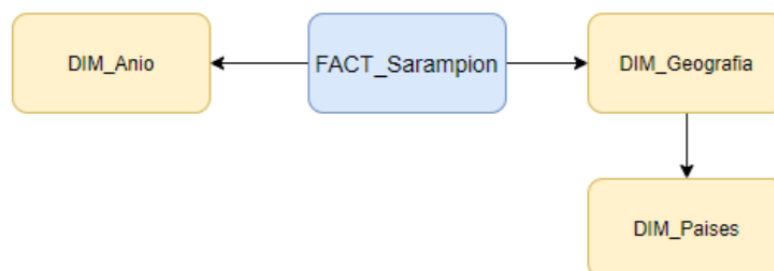


Ilustración 5 - Diseño conceptual FACT_Sarampion desnormalizado.

Esta afirmación es correcta, cuando desnormalizamos una dimensión lo que buscamos es reducir la complejidad del diseño y así mejorar el rendimiento de las consultas, y para conseguirlo lo que hacemos es “introducir” los atributos de una dimensión en otra, reduciendo el número de dimensiones [1].

- b) El diseño conceptual presenta el mayor nivel de abstracción ya que es el más alejado a la representación física del modelo.

Esta afirmación es correcta, el diseño conceptual lo que busca es identificar de primera mano qué hechos y qué dimensiones vamos a tener en el modelo, una vez que eso está claro, se reduce un poco la abstracción con el diseño lógico, gracia este diseño definimos cuáles van a ser las métricas de los hechos y los atributos de las dimensiones. Finalmente, llegamos al modelo físico en el cual se reduce completamente la abstracción del modelo, buscamos definir el cómo se relacionan las diferentes dimensiones, los hechos, cómo van a ser las tablas de la base de datos, los campos que la componen, el tipo de dato de cada campo, todo ello con el fin de mejorar el rendimiento de las consultas.

- c) La representación gráfica de su correspondiente diseño físico es un diagrama en copo de nieve.

Esta afirmación es correcta, cuando las dimensiones se relacionan entre ellas en vez de relacionarse directamente con el hecho, lo que estamos haciendo es generar una “especialización”, y cuando hacemos eso la representación gráfica del modelo va a ser del tipo copo de nieve debido a las diferentes ramificaciones que van surgiendo.

- d) Todas las anteriores son correctas.

Esta afirmación es correcta, todas las afirmaciones como bien hemos visto son correctas.

2. Parte práctica

2.1. Pregunta 5

A partir del fichero “ACUMULADO-DENUNCIAS-INFRACCIONES.xlsx” hoja “Datos_tratados”, se debe diseñar, implementar y ejecutar los procesos de extracción, transformación y carga para la Transformación IN_DENUNCIAS-INFRACCIONES, siguiendo y completando las siguientes cuestiones:

- a) Completad y ejecutad el siguiente comando SQL para la creación de la tabla intermedia donde se almacenará los datos del origen “ACUMULADODENUNCIAS-INFRACCIONES.xlsx” hoja “Datos_tratados”. Sustituid las XXX, por el valor apropiado:

```
CREATE TABLE [dbo].[STG_Denuncias_Infracciones](  
    [provincia] [varchar](100) NULL,  
    [identificados_ertzaintza] [float] NULL,  
    [detenidos_ertzaintza] [float] NULL,  
    [denuncias_ertzaintza] [float] NULL,  
    [vehic_intercept_ertzaintza] [float] NULL,  
    [identificados_ppll] [float] NULL,  
    [detenidos_ppll] [float] NULL,  
    [denuncias_ppll] [float] NULL,  
    [vehic_intercept_ppll] [float] NULL,  
    [fecha] [datetime] NULL  
    ) ON [PRIMARY]  
GO
```

Ilustración 6 - Creación de la tabla STG_Denuncias_Infracciones.

Una vez finalizado el script para la creación de la tabla, ejecutamos dicho *script* y comprobamos que ha creado la tabla en la base de datos:

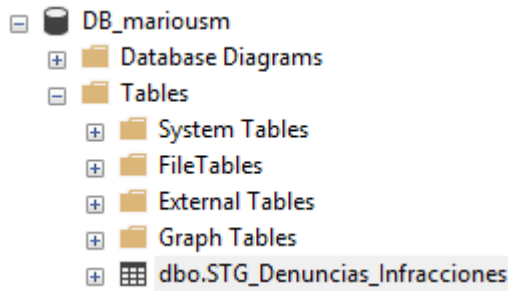


Ilustración 7 - Comprobación de la creación de la tabla.

- b) Lectura de los ficheros.xlsx. Completad la siguiente información del paso “File Input”:

Nombre: <Lectura del fichero>

Componente: Microsoft Excel Input

Descripción: Permite cargar datos de entrada provenientes de un fichero Excel, <Este proceso es el encargo de hacer la lectura del fichero Excel, para ello le vamos a indicar que tiene que leer el fichero “ACUMULADO-DENUNCIAS-INFRACCIONES.xlsx” hoja “Datos_tratados”, una vez establecido eso la lectura la tiene que comenzar a realizar a partir de la fila 5 y debemos nosotros modificar el nombre de los atributos y comprobar si los ha leído de forma correcta>

Parámetros:

- **Files / File or directory:** <F:\Mario\PEC2\data\ACUMULADO-DENUNCIAS-INFRACCIONES.xlsx>
- **Sheets (Hojas)** mediante “get sheet names”, hoja “Datos_tratados”, con fila inicial “start row” = 5, podemos observar la siguiente captura con la configuración:

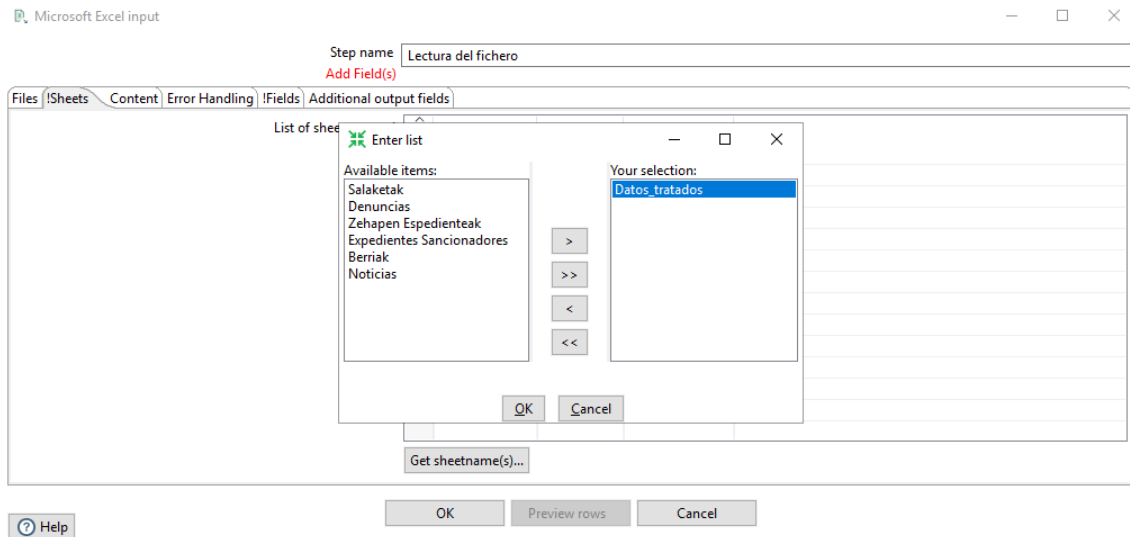


Ilustración 8 - Selección de la hoja "Datos_tratados".

Finalmente, indicamos desde que fila tiene que comenzar a leer en nuestro caso la cinco:

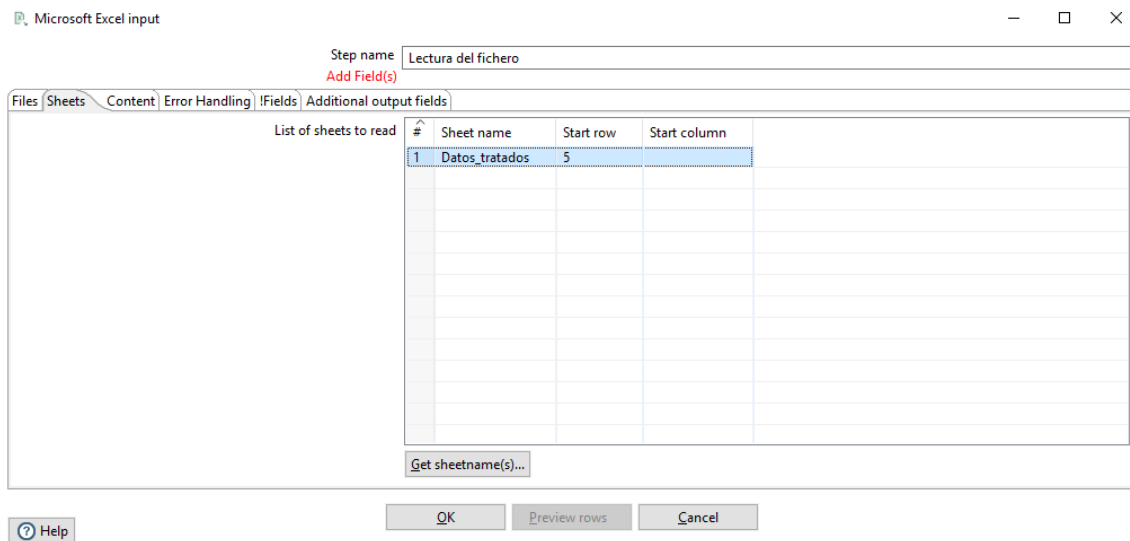


Ilustración 9 - Especificación del comienzo de la fila de lectura.

- **Fields (campos)** mediante el botón “Get fields from header row...” se obtienen todos los campos del fichero, así como el tipo, formato y longitud de los datos.

En la siguiente captura podemos ver cómo seleccionamos los campos a partir de “get fields”:

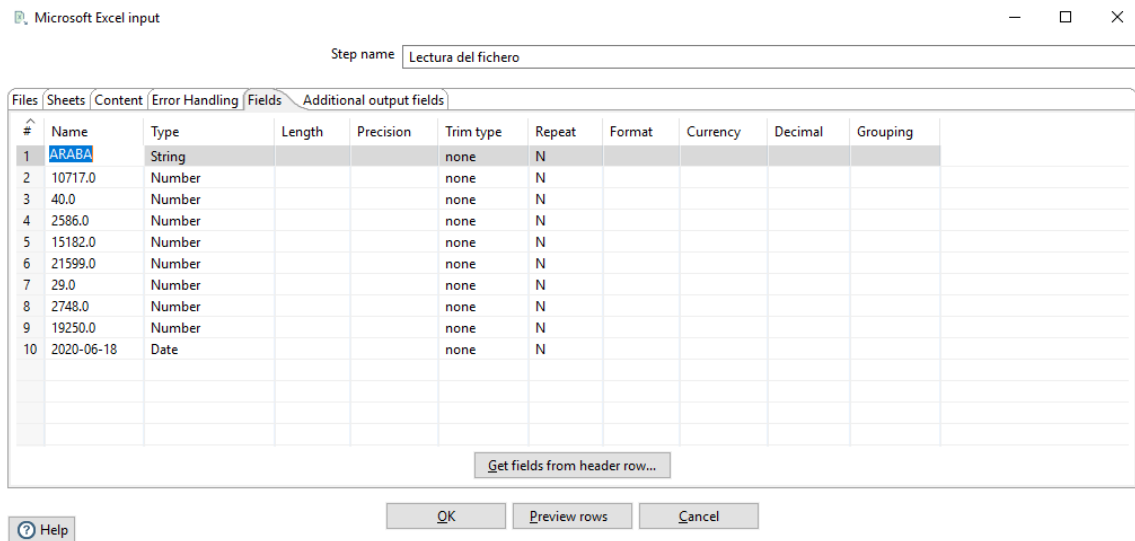


Ilustración 10 - Lectura de los campos a partir de "Get fields from header row".

Como podemos apreciar no detecta bien el nombre de los campos, es por ello que lo modificamos de forma manual y comprobamos que el tipo de dato se corresponda:

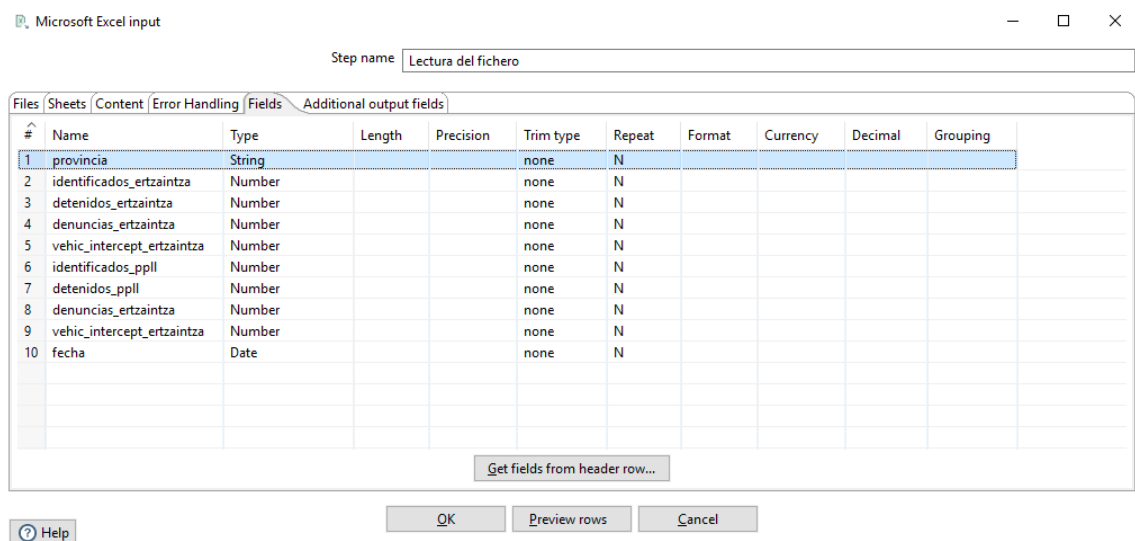


Ilustración 11 - Definición del nombre de los campos y su tipo.

- **Preview:** botón <Preview rows> (Previsualizar filas).

Antes de mostrar las filas debemos indicar que ignore la cabecera del fichero, para que así lea todos los datos y no se salte el primero. Nos dirigimos a la pestaña "Content" y desmarcamos la casilla "Header", tal y como observamos en la siguiente captura:

Microsoft Excel input

Step name: Lectura del fichero

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

Header ☐
 No empty rows ☒
 Stop on empty row ☐
 Limit: 0
 Encoding: ▼

Result filenames

Add filenames to result ☒

OK Preview rows Cancel

Help

Ilustración 12 - Desmarcamos la casilla de la cabecera.

Ahora sí que podemos visualizar los datos y ver que se corresponden con los del fichero Excel:

Examine preview data

Rows of step: Lectura del fichero (219 rows)

#	provincia	identificados_ertaintza	detenidos_ertaintza	denuncias_ertaintza	vehic_intercept_ertaintza	identificados_ppil	detenidos_ppil	denuncias_ppil	vehic_intercept_ppil	fecha
1	ARABA	10717.0	40.0	2586.0	15182.0	21599.0	29.0	2748.0	19250.0	2020/06/18 00:00:00.000
2	BIZKAIA	29955.0	228.0	6249.0	56139.0	27160.0	65.0	9209.0	50539.0	2020/06/18 00:00:00.000
3	GIPUZKOA	26051.0	62.0	4884.0	17246.0	27069.0	40.0	4473.0	37797.0	2020/06/18 00:00:00.000
4	ARABA	10708.0	40.0	2586.0	15180.0	21598.0	29.0	2748.0	19250.0	2020/06/17 00:00:00.000
5	BIZKAIA	29867.0	228.0	6249.0	56009.0	27138.0	65.0	9209.0	50518.0	2020/06/17 00:00:00.000
6	GIPUZKOA	26004.0	62.0	4882.0	17183.0	27059.0	40.0	4472.0	37791.0	2020/06/17 00:00:00.000
7	ARABA	10705.0	40.0	2585.0	15176.0	21598.0	29.0	2748.0	19250.0	2020/06/16 00:00:00.000
8	BIZKAIA	29751.0	228.0	6249.0	55859.0	27124.0	65.0	9209.0	50468.0	2020/06/16 00:00:00.000
9	GIPUZKOA	25942.0	62.0	4882.0	17173.0	27038.0	38.0	4465.0	37782.0	2020/06/16 00:00:00.000
10	ARABA	10704.0	40.0	2585.0	15176.0	21593.0	29.0	2746.0	19250.0	2020/06/15 00:00:00.000
11	BIZKAIA	29674.0	228.0	6247.0	55754.0	27106.0	65.0	9203.0	50436.0	2020/06/15 00:00:00.000
12	GIPUZKOA	25872.0	62.0	4879.0	17103.0	27022.0	38.0	4465.0	37770.0	2020/06/15 00:00:00.000
13	ARABA	10700.0	40.0	2585.0	15173.0	21573.0	29.0	2739.0	19250.0	2020/06/14 00:00:00.000
14	BIZKAIA	29597.0	228.0	6246.0	55674.0	27063.0	65.0	9202.0	50428.0	2020/06/14 00:00:00.000
15	GIPUZKOA	25775.0	62.0	4873.0	17003.0	27015.0	38.0	4465.0	37754.0	2020/06/14 00:00:00.000
16	ARABA	10652.0	40.0	2572.0	15134.0	21572.0	29.0	2738.0	19250.0	2020/06/13 00:00:00.000
17	BIZKAIA	29477.0	228.0	6236.0	55470.0	27055.0	65.0	9201.0	50386.0	2020/06/13 00:00:00.000
18	GIPUZKOA	25672.0	62.0	4862.0	16924.0	27004.0	38.0	4464.0	37744.0	2020/06/13 00:00:00.000
19	ARABA	10618.0	40.0	2569.0	15110.0	21571.0	29.0	2737.0	19250.0	2020/06/12 00:00:00.000
20	BIZKAIA	29308.0	228.0	6220.0	55348.0	27011.0	65.0	9176.0	50339.0	2020/06/12 00:00:00.000
21	GIPUZKOA	25656.0	62.0	4840.0	16786.0	26988.0	38.0	4464.0	37730.0	2020/06/12 00:00:00.000
22	ARABA	10591.0	40.0	2567.0	15098.0	21570.0	29.0	2736.0	19250.0	2020/06/11 00:00:00.000
23	BIZKAIA	29169.0	227.0	6214.0	55121.0	26979.0	65.0	9168.0	50283.0	2020/06/11 00:00:00.000
24	GIPUZKOA	25556.0	62.0	4829.0	16711.0	26951.0	38.0	4457.0	37713.0	2020/06/11 00:00:00.000
25	ARABA	10590.0	40.0	2566.0	15098.0	21565.0	29.0	2731.0	19250.0	2020/06/10 00:00:00.000
26	BIZKAIA	29137.0	227.0	6212.0	55082.0	26939.0	65.0	9154.0	50254.0	2020/06/10 00:00:00.000
27	GIPUZKOA	25490.0	62.0	4804.0	16660.0	26939.0	38.0	4454.0	37702.0	2020/06/10 00:00:00.000
28	ARABA	10579.0	40.0	2564.0	15093.0	21537.0	29.0	2723.0	19250.0	2020/06/09 00:00:00.000
29	BIZKAIA	29070.0	227.0	6208.0	55019.0	26898.0	65.0	9137.0	50206.0	2020/06/09 00:00:00.000
30	GIPUZKOA	25323.0	62.0	4781.0	16450.0	26913.0	38.0	4447.0	37691.0	2020/06/09 00:00:00.000
31	ARABA	10557.0	40.0	2558.0	15064.0	21527.0	29.0	2712.0	19249.0	2020/06/08 00:00:00.000
32	BIZKAIA	28971.0	227.0	6204.0	54904.0	26848.0	65.0	9128.0	50185.0	2020/06/08 00:00:00.000
33	GIPUZKOA	25207.0	62.0	4771.0	16351.0	26899.0	38.0	4443.0	37676.0	2020/06/08 00:00:00.000

Ilustración 13 - Visualización de los datos.

- c) Asegurad la homogeneidad de los datos mediante la normalización de los valores de los campos tipo `<String>`. Convirtiendo a mayúsculas y eliminando los espacios en blanco al inicio y al final de cada cadena. Completad la siguiente información del paso *"String Operation"*:

Nombre: `<Normalización>`

Componente: `<String operations>`

Descripción: <Permite transformar los datos que son de tipo *String*, es decir, realizar operaciones sobre la cadena de caracteres>. Asegurad la homogeneidad de los datos mediante la normalización de los valores de los campos tipo <*String*>. Convirtiendo a mayúsculas y eliminando los espacios en blanco al inicio y al final de cada cadena.

Parámetros:

String operations

Step name:

The fields to process:

#	In stream field	Out stream field	Trim type	Lower/Upper	Padding	Pad char	Pad Length	InitCap	Escape	Digits	Remove Special character
1	provincia		both	upper	none			N	None	none	none

Ilustración 14 – Normalización de los datos.

- d) Ordenación ascendente de todos los campos según su colocación en la tabla Staging.

Completad la siguiente información del paso “Row Order”:

Nombre: <Ordenación>

Componente: <Sort rows>, cabe destacar que el paso “Row Order” ya no existe por contra tenemos el paso mencionado anteriormente “Sort rows”.

Descripción: <Permite realizar la ordenación de los campos>. Ordenación ascendente de todos los campos según su colocación en la tabla Staging.

Parámetros:

[illegible]

Ilustración 15 - Ordenación de los campos.

e) Cargad la información transformada en la tabla de base de datos. Completar la siguiente información del paso “Table Output”:

Nombre: <Carga en la base de datos>

Componente: Table Output

Descripción: <Permite cargar los datos de entrada en la base de datos>, <Cargamos la información transformada en la tabla de base de datos>.

Parámetros:

- **Connection:** <PEC2>

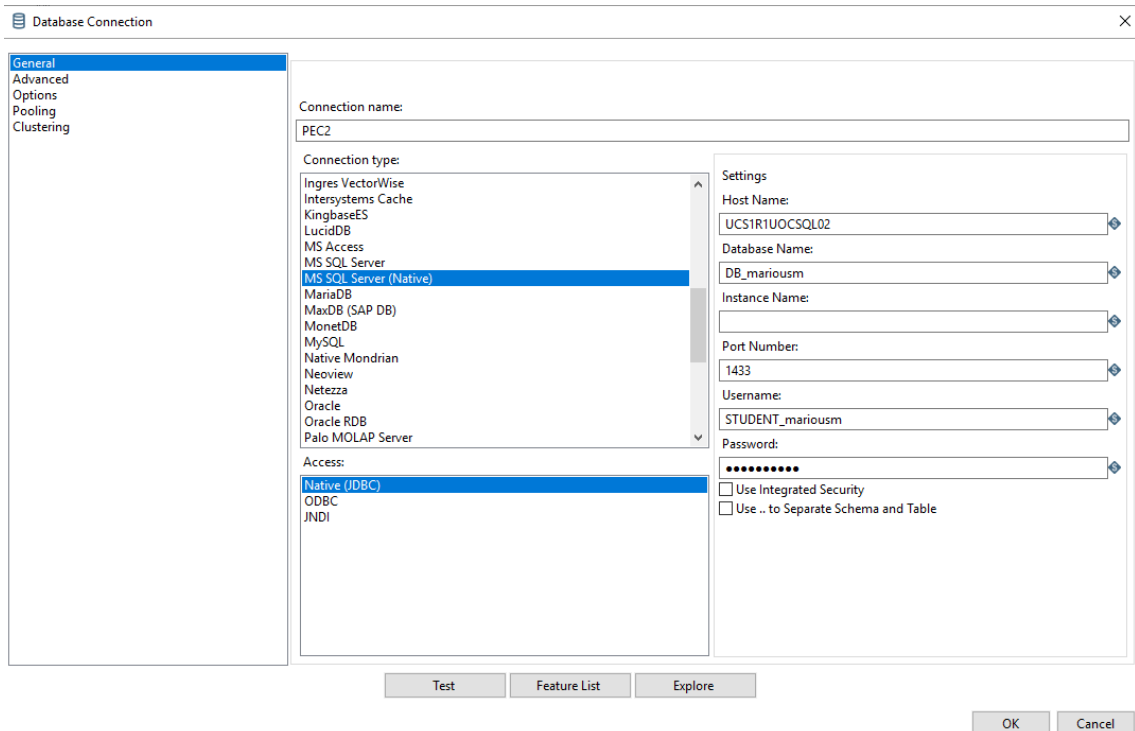


Ilustración 16 - Conexión a la base de datos.

- **Target table:** <dbo.STG_Denuncias_Infracciones>, para seleccionar la target table, tenemos que seleccionar primero dbo al pulsa sobre el botón “Browse...” y finalmente, sobre el botón “Browse...” de target table seleccionamos la tabla STG_Denuncias_Infracciones.

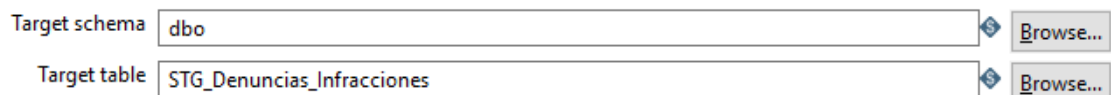


Ilustración 17 - Target table.

- **Truncate table:** <Pulsamos sobre la casilla truncate>

Truncate table ☒

Ilustración 18 - Truncate table.

Tal y como podemos observar en la siguiente ilustración, este componente queda configurado de la siguiente manera:

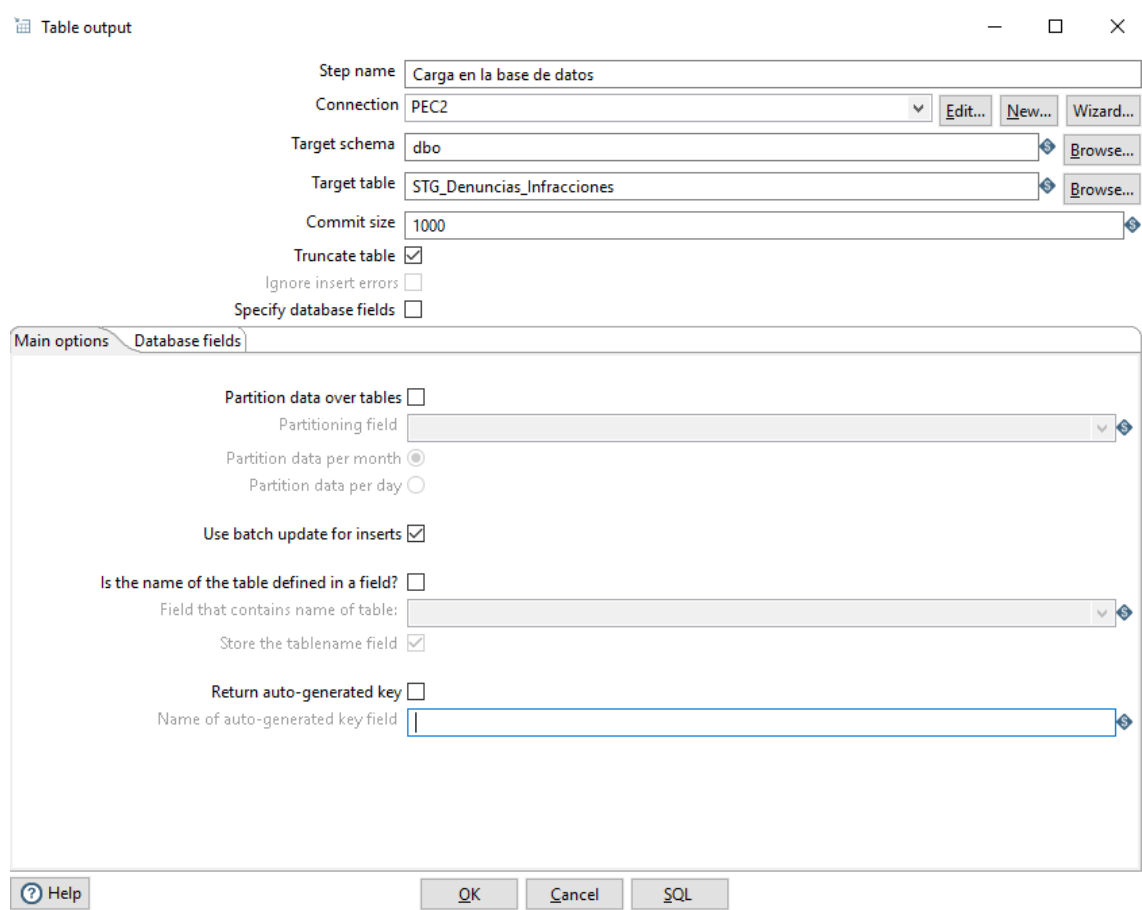


Ilustración 19 - Configuración de Table Output.

- f) Capturad la pantalla de la transformación completa, incluyendo la pestaña informativa de ejecución “step metrics”.

Una vez que ya tenemos toda la transformación completa, la ejecutamos y vemos que se ejecuta de forma correcta, tal y como vemos en la siguiente captura:



Ilustración 20 - Ejecución de la transformación.

Por otro lado, nos genera la siguiente información en “step metrics”:

Logging Execution History Step Metrics Performance Graph Metrics Preview data													
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Lectura del fichero	0	0	219	219	0	0	0	0	Finished	0.3s	711	-
2	Normalización	0	219	219	0	0	0	0	0	Finished	0.3s	689	-
3	Ordenación	0	219	219	0	0	0	0	0	Finished	0.4s	626	-
4	Carga en la base de datos	0	219	219	0	219	0	0	0	Finished	0.4s	512	-

Ilustración 21 - Visualización de la pestaña step metrics.

- g) Realizad una Consulta en la Base de datos, que devuelva el número de registros de la tabla cargada. ¿Coincide con el número de registros procesados en cada paso, mostrados en “step metrics”?

La consulta SQL que se ha realizado para saber el número de registros es la siguiente:

```
SELECT count(provincia)
FROM [DB_mariouism].[dbo].[STG_Denuncias_Infracciones]
```

Ilustración 22 - Consulta SQL para saber el número de registros.

La ejecución de la anterior consulta nos proporciona este resultado:

(No column name)
219

Ilustración 23 - Número de registros cargados en la base de datos.

Tal y como vemos en el anterior apartado, el número de registros escritos en la base de datos son 219, el cual coincide con el de la anterior ilustración, por lo tanto se ha cargado toda la información de forma correcta.

- h) Realizad la consulta en la Base de Datos y capturad el resultado del Top 10 de registros sin ordenar, ¿coinciden con los 10 primeros registros ordenados ascendentemente de todos los campos según su colocación en la tabla *Staging*?

Realizamos la consulta a la base de datos de la siguiente forma:

```
SELECT TOP (10) [provincia]
, [identificados_ertzaintza]
, [detenidos_ertzaintza]
, [denuncias_ertzaintza]
, [vehic_intercept_ertzaintza]
, [identificados_pp11]
, [detenidos_pp11]
, [denuncias_pp11]
, [vehic_intercept_pp11]
, [fecha]
FROM [DB_mariouism].[dbo].[STG_Denuncias_Infracciones]
```

Ilustración 24 - Consulta SQL top 10 registros.

La ejecución de la anterior consulta nos proporciona el siguiente resultado:

Results Messages									
provincia	identificados_ertzaintza	detenidos_ertzaintza	denuncias_ertzaintza	vehic_intercept_ertzaintza	identificados_ppl	detenidos_ppl	denuncias_ppl	vehic_intercept_ppl	fecha
ARABA	3007	5	965	7251	5237	10	1119	6574	2020-04-07 00:00:00.000
ARABA	3172	7	987	7637	5609	12	1163	7037	2020-04-08 00:00:00.000
ARABA	3296	9	1022	7990	5982	13	1213	7490	2020-04-09 00:00:00.000
ARABA	3549	9	1091	8245	6315	13	1247	7839	2020-04-10 00:00:00.000
ARABA	3721	9	1110	8461	7166	14	1287	8617	2020-04-11 00:00:00.000
ARABA	3817	10	1144	8545	7404	14	1333	8836	2020-04-12 00:00:00.000
ARABA	3936	11	1193	8720	7689	15	1410	9142	2020-04-13 00:00:00.000
ARABA	4024	11	1220	8847	8025	15	1458	9608	2020-04-14 00:00:00.000
ARABA	4074	11	1240	8978	8309	15	1493	9840	2020-04-15 00:00:00.000
ARABA	4122	11	1267	9086	8701	15	1531	10129	2020-04-16 00:00:00.000

Ilustración 25 - Top 10 registros.

Ahora, comprobamos con los 10 primeros registros ordenados ascendentemente en la tabla *Staging*:

Execution Results

Logging

Execution History

Step Metrics

Performance Graph

Metrics

Preview data

☒ First rows

☐ Last rows

☐ Off

#	provincia	identificados_ertzaintza	detenidos_ertzaintza	denuncias_ertzaintza	vehic_intercept_ertzaintza	identificados_ppll	detenidos_ppll	denuncias_ppll	vehic_intercept_ppll	fecha
1	ARABA	3007.0	5.0	965.0	7251.0	5237.0	10.0	1119.0	6574.0	2020/04/07 00:00:00.000
2	ARABA	3172.0	7.0	987.0	7637.0	5609.0	12.0	1163.0	7037.0	2020/04/08 00:00:00.000
3	ARABA	3296.0	9.0	1022.0	7990.0	5982.0	13.0	1213.0	7490.0	2020/04/09 00:00:00.000
4	ARABA	3549.0	9.0	1091.0	8245.0	6315.0	13.0	1247.0	7839.0	2020/04/10 00:00:00.000
5	ARABA	3721.0	9.0	1110.0	8461.0	7166.0	14.0	1287.0	8617.0	2020/04/11 00:00:00.000
6	ARABA	3817.0	10.0	1144.0	8545.0	7404.0	14.0	1333.0	8836.0	2020/04/12 00:00:00.000
7	ARABA	3936.0	11.0	1193.0	8720.0	7689.0	15.0	1410.0	9142.0	2020/04/13 00:00:00.000
8	ARABA	4024.0	11.0	1220.0	8847.0	8025.0	15.0	1458.0	9608.0	2020/04/14 00:00:00.000
9	ARABA	4074.0	11.0	1240.0	8978.0	8309.0	15.0	1493.0	9840.0	2020/04/15 00:00:00.000
10	ARABA	4122.0	11.0	1267.0	9086.0	8701.0	15.0	1531.0	10129.0	2020/04/16 00:00:00.000

Ilustración 26 - Top 10 registros de la tabla *Staging*.

Tal y como podemos comprobar, ambos resultados son idénticos.

3. Bibliografía

- [1] «Modelo Dimensional», *BI Geek Blog*, oct. 09, 2017. <https://blog.bi-geek.com/modelo-dimensional/> (accedido abr. 20, 2021).