

Caso práctico: Almacén de datos para el análisis del parque de vehículos

PR4 – Análisis de datos

La cuarta y última práctica del curso, **PR4 – Análisis de datos** consta de dos partes: una primera parte destinada a la evaluación del contenido teórico y una segunda parte práctica destinada a la evaluación de los conocimientos adquiridos mediante un conjunto de ejercicios a realizar en el entorno de prácticas de la asignatura.

La parte teórica tiene una ponderación del 30% sobre la nota total de la práctica. Se evalúa mediante un **cuestionario**, disponible en el apartado *Contenidos del aula*. Su objetivo es comprobar la correcta comprensión del tercer módulo de la asignatura. Debe realizarse después de haber estudiado el módulo 4 y haber leído detenidamente las indicaciones de realización del mismo.

La parte práctica también pondrá el 70% de la nota total de la práctica, se lleva a cabo en la máquina virtual (VDI) mediante la realización de todos los ejercicios que se detallan en el apartado “Enunciado parte práctica”. Tened en cuenta que deberéis utilizar las especificaciones técnicas, tales como el nombre de las dos bases de datos PostgreSQL y SQL Server ya facilitadas y utilizadas en anteriores prácticas.

Se recomienda consultar la documentación disponible en el apartado *Recursos de Aprendizaje PR3* del espacio *Contenidos del aula*. Dichos recursos pretenden facilitar la realización de los ejercicios y resolver posibles dudas. También se incluye la **documentación técnica**, imprescindible para trabajar en el entorno virtual (VDI), que está preparado con todas las herramientas necesarias, instaladas y configuradas. Recordad que ante cualquier duda y/o problema técnico relativo al funcionamiento del entorno o del software instalado debéis consultar al profesor del aula 76.596.-

Laboratorio de Soporte a las bases de datos analíticas. Únicamente se evaluarán los ejercicios cuya resolución esté acompañada de comentarios; las imágenes o gráficas sin ningún tipo de explicación no serán evaluadas.

Parte práctica

En esta cuarta práctica (PR4), contextualizada en el caso práctico “**Almacén de datos para el análisis del parque de vehículos**”, se utilizarán diferentes componentes de la FIC (Factoría de Información Corporativa) para construir los distintos elementos necesarios para almacenar los datos en una base de datos analítica y poder llevar a cabo diferentes actividades de análisis.

Tras la realización de las PRs anteriores, en esta práctica se consolidarán las diferentes técnicas aprendidas para llevar a cabo el análisis de los datos mediante diversas visualizaciones y ejercicios prácticos.

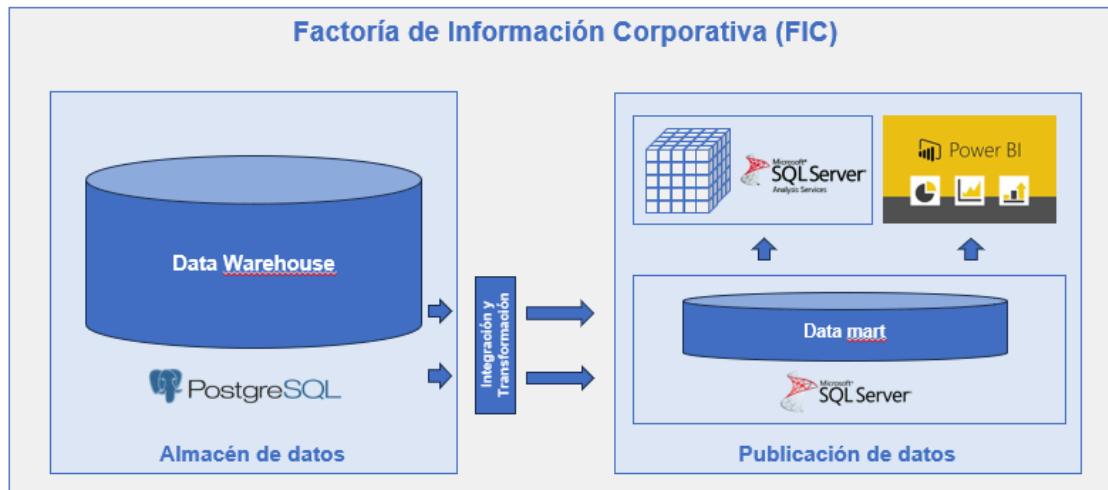
Contexto

Con una plataforma multifuncional, que incluye tanto el software para la gestión del parque de vehículos como una Factoría de Información Corporativa (FIC) orientada a dar respuesta a las distintas necesidades de análisis, ha surgido la necesidad de incorporar un analista de datos. Este profesional, utilizando las herramientas disponibles en la plataforma, deberá ser capaz de analizar la actividad y evolución del parque de vehículos gestionado por la DGT.

El analista de datos dispondrá de las siguientes herramientas para acceder y analizar los datos almacenados en los sistemas de información:

- Acceso al *Data Warehouse* corporativo soportado por el sistema de gestión de base de datos *PostgreSQL*, al cual se accede en formato autoconsumo y mediante consultas SQL.
- Acceso al *Data mart*, el cual está especializado en los datos de gestión aeroportuaria, soportado por un servidor *Microsoft Analysis Services* y al cual

se accede mediante un cubo multidimensional (OLAP) y mediante la herramienta *Power BI*.



En ambos entornos (el almacén de datos y la publicación de datos) los datos están estructurados mediante modelos multidimensionales, habiendo definido para ello, hechos (*facts*), dimensiones de análisis (*dimensions*), métricas y atributos que permitan tener el nivel de granularidad suficiente para la presentación de los resultados.

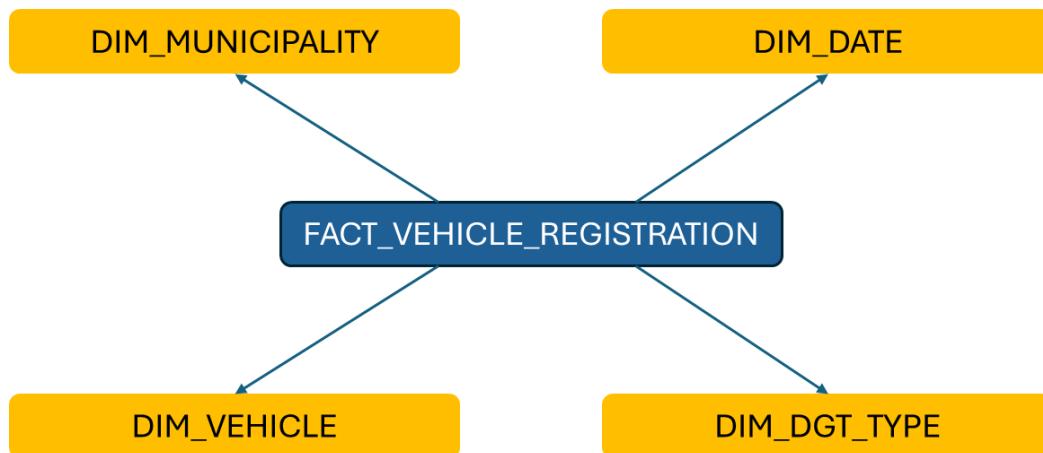
Se ha creado una tabla de hechos haciendo referencia a los datos de matriculación de vehículos.

Tabla de hechos	Descripción
FACT_VEHICLE_REGISTRATION	Registro de todas las matriculaciones de vehículos, incluyendo fecha, vehículo y municipio

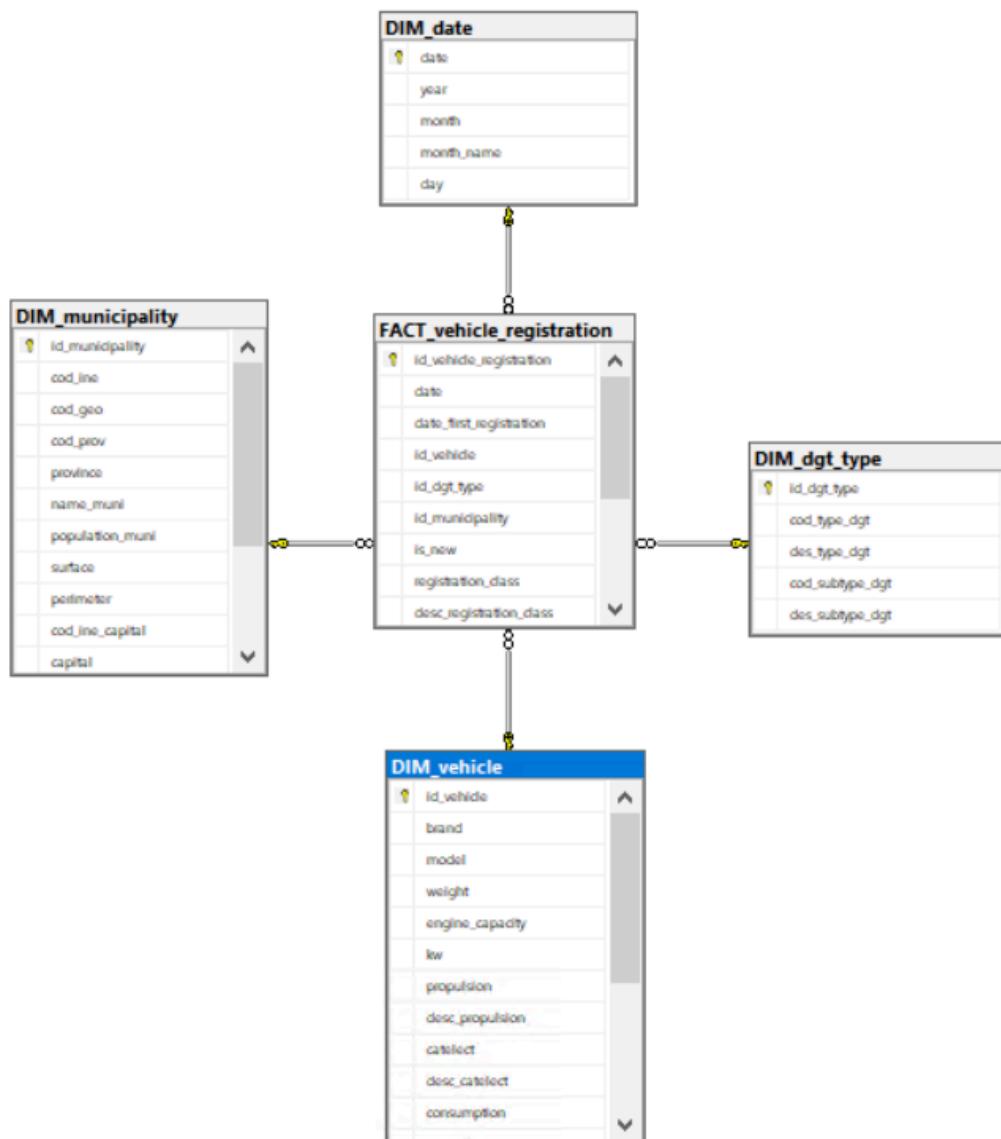
Para que las métricas de estas tablas de hechos puedan ser analizadas desde las diferentes perspectivas, se han construido las siguientes dimensiones:

Dimensiones	Descripción
DIM_DATE	Información temporal (año, mes, día) de la matriculación.
DIM_MUNICIPALITY	Información geográfica (municipio, provincia, población, etc.)
DIM_VEHICLE	Detalles del vehículo (marca, modelo, tipo, cilindrada, etc.)
DIM_DGT_TYPE	Tipo de vehículo (turismo, motocicleta, etc.) según clasificación de la Dirección General de Tráfico (DGT). Está compuesto por un tipo y un subtipo.

Simplificando el caso de uso, el diseño conceptual para la tabla de hechos (FACT_VEHICLE_REGISTRATION) y sus dimensiones se obtiene el siguiente **diseño conceptual en estrella**:



Que tendrá la siguiente representación física:



Ejercicio 1. Análisis de datos mediante autoconsumo (10%)

Objetivo:

Practicar los diferentes conocimientos obtenidos sobre sentencias SQL, poniendo el foco en aquellas destinadas al análisis de los datos mediante técnicas de autoconsumo.

Indicaciones para el desarrollo del ejercicio:

Este ejercicio se realizará en **PostgreSQL**. Para disponer de un conjunto de datos destinado al análisis, y en caso de no haberlo hecho previamente, es necesario ejecutar el script de creación de tablas y carga de registros en la base de datos PostgreSQL: **PR4_DDL+Data_PostgreSQL.sql**, facilitado junto al enunciado.

Dada la volumetría de los datos, la ejecución del script podría demorarse varios minutos.

1.1. Autodescubrimiento de la base de datos

Se solicitan las sentencias SQL que den respuesta a las siguientes cuestiones:

- 1) Obtener la información existente en el esquema dbo mediante una consulta SQL sobre la vista *pg_tables* (vista que proporciona el catálogo de las tablas almacenadas en la base de datos).
- 2) Combinar la consulta anterior con una consulta a la tabla *pg_stat_user_tables* utilizando JOIN para obtener el total de registros de cada tabla. Mostrar los resultados ordenados por total de registros de forma ascendente.
- 3) Utilizando el esquema dbo y la vista *information_schema.columns*, obtener de la tabla *fact_vehicle_registration* los nombres de las columnas, los tipos de datos de cada columna y si éstas admiten nulos o no.

- 4) Ejecutar la siguiente consulta de base de datos y explicar los resultados por cada una de las columnas obtenidas.

```
SELECT
    attname, null_frac, n_distinct, most_common_vals
FROM
    pg_stats
WHERE
    tablename = 'dim_vehicle' AND
    schemaname = 'dbo' AND
    attname IN ('id_vehicle', 'brand', 'model', 'co2_emissions')
```

En esta actividad se evaluará que el estudiantado sea capaz de crear correctamente las consultas SQL solicitadas y mostrar los resultados obtenidos. De manera especial, en la consulta 4 se valorará que se explique correctamente el significado de los resultados obtenidos para cada columna y su interpretación en función de los datos de la tabla, demostrando un buen entendimiento del análisis de estadísticas de base de datos.

Ejercicio 2. Análisis de datos multidimensional (25%)

Objetivo:

Conocer la aplicación para diseñar cubos multidimensionales (OLAP) con el objetivo de incorporar información procedente de un almacén de datos y analizarla mediante consultas multidimensionales.

Indicaciones para el desarrollo del ejercicio:

Este ejercicio se realizará en **Microsoft SQL Server**. A continuación, se detallarán los diferentes pasos necesarios para la creación de un cubo que almacene la información de la tabla de hechos *FACT_vehicle_registration*, y sus dimensiones vinculadas, con el objetivo de poder realizar un análisis mediante las herramientas de análisis multidimensional facilitadas en el entorno virtual.

Para disponer de un conjunto de datos destinado al análisis, y en caso de no haberlo hecho previamente, es necesario ejecutar el script de creación de tablas y carga de registros en la base de datos SQL Server: **PR4_DDL+Data_SQLS.sql**, facilitado junto al enunciado.

Dada la volumetría de los datos a cargar, la ejecución del script podría demorarse varios minutos.

2.1. Creación del proyecto, configuración del destino y origen de los datos

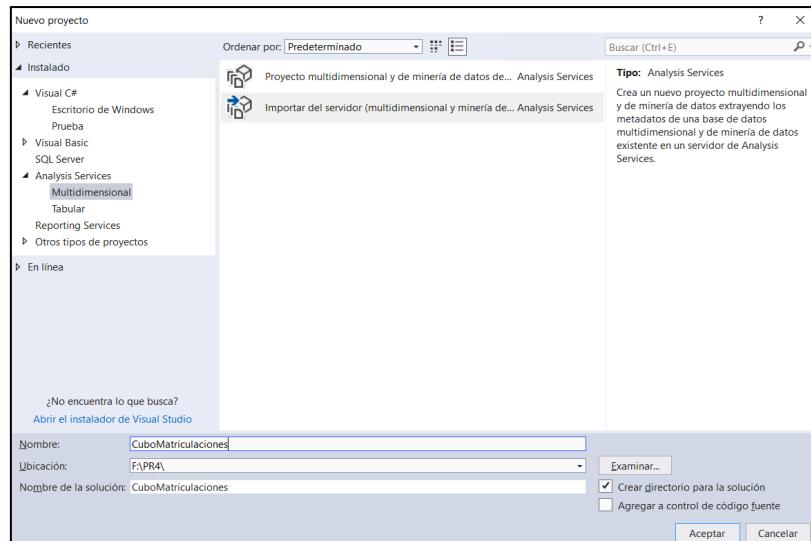
Mediante la herramienta Visual Studio, se solicita crear un proyecto nuevo de tipo **Importar del servidor (multidimensional y minería de datos)**, que permita construir cubos analíticos y analizar los datos a través de ellos.

En primer lugar, se abrirá **Microsoft Visual Studio 2017**, disponible en el escritorio del entorno virtual.

Nota: Es posible que Visual Studio solicite credenciales de inicio de sesión. En ese caso, se deben introducir las credenciales de acceso al campus virtual de la UOC (usuario y contraseña). Durante este proceso, pueden aparecer mensajes de advertencia indicando que hay errores de script en la página; se pueden ignorar pulsando el botón “No”.

Una vez abierto Visual Studio, se debe ir a **Archivo > Nuevo > Proyecto**, y seleccionar **Analysis Services > Multidimensional > Importar del servidor (Multidimensional y minería de datos)**

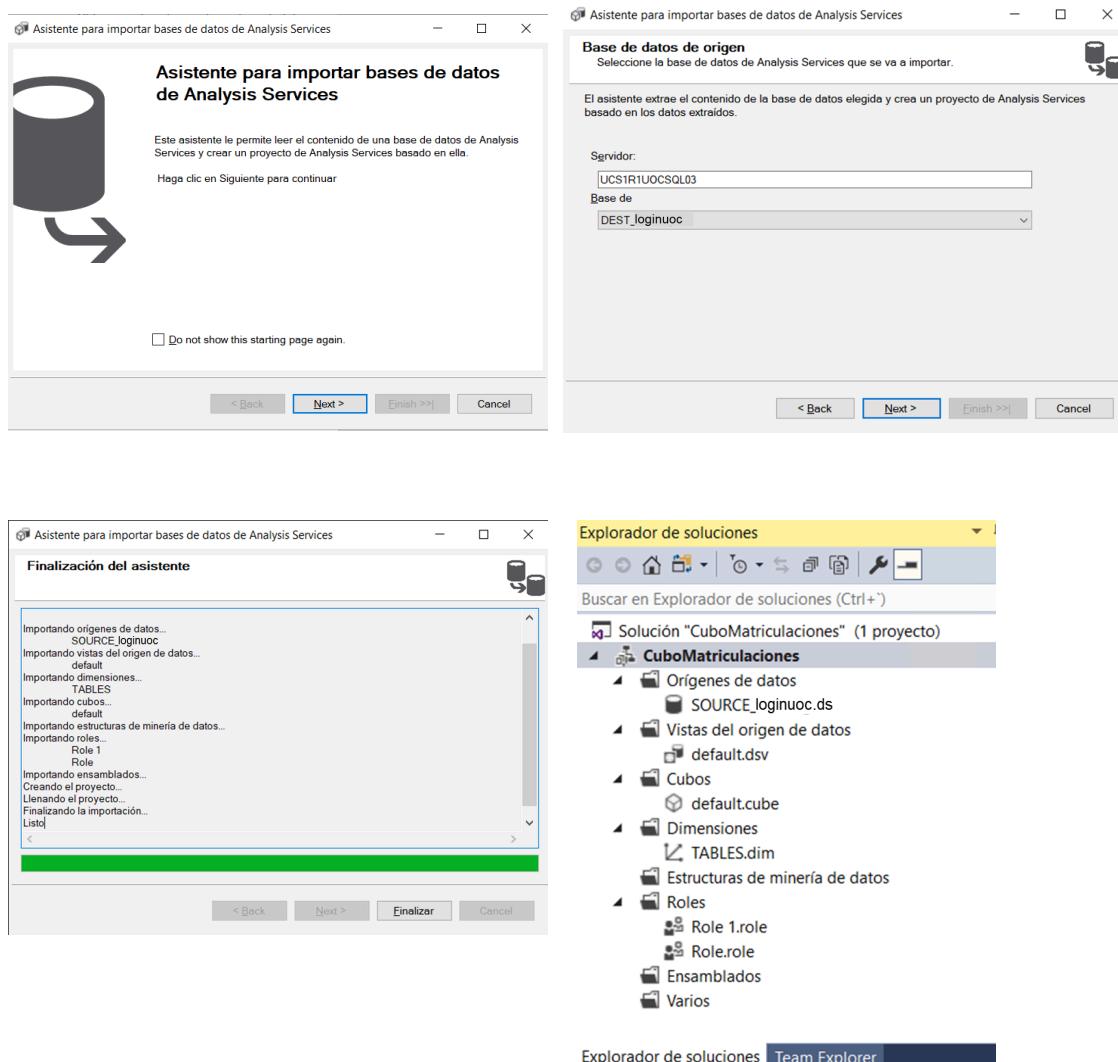
A continuación, se debe asignar un nombre representativo al proyecto y seleccionar una ubicación dentro de la unidad **F:** donde se desee guardar.



A partir de este punto, se mostrará un asistente cuyos pasos deberán seguirse en orden para completar la creación del proyecto.

Uno de los pasos del asistente mostrará el diálogo **Base de datos de origen**, donde se deberá indicar el servidor de *Analysis Services* proporcionado en el laboratorio. En el campo **Base de datos**, se introducirá *DEST_loginuoc*, siendo *loginuoc* el nombre de usuario en la UOC.

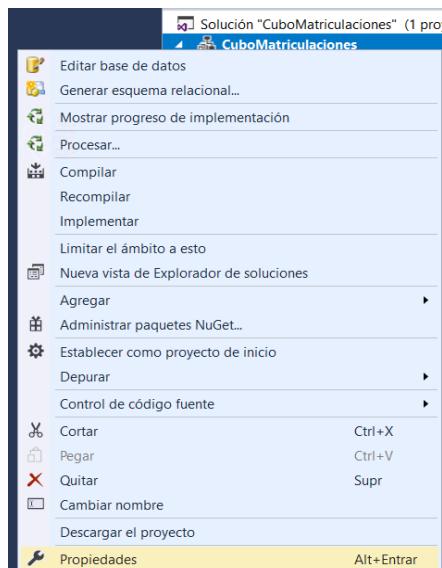
Una vez finalizado el asistente, se podrán visualizar los componentes del proyecto en la parte derecha de la aplicación, en el **Explorador de soluciones**.



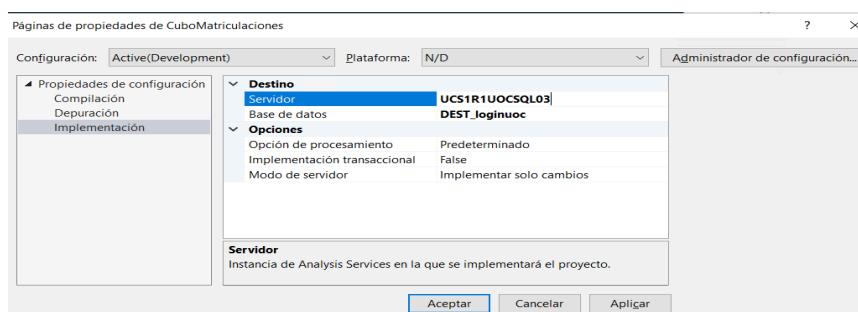
Importante: No se debe eliminar el cubo **default** obtenido tras la importación.

A continuación, se editarán las propiedades del proyecto para indicar el destino del cubo que se va a implementar. Para ello:

1. Hacer clic derecho sobre el nombre del proyecto y seleccionar **Propiedades**.

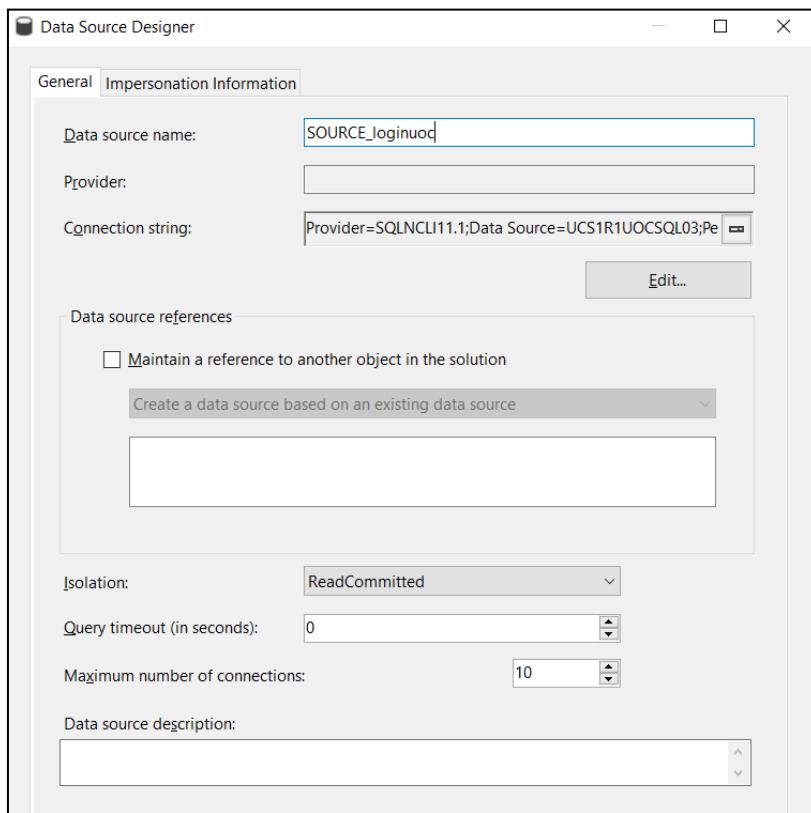


2. En el diálogo que aparece, acceder a la opción **Implementación**.
3. Actualizar el atributo **Servidor** con el nombre del servidor de *Analysis Services* indicado en el laboratorio.
4. Asegurarse de que el nombre de la base de datos sea *DEST_loginuoc*, donde *loginuoc* corresponde al nombre de usuario en la UOC.

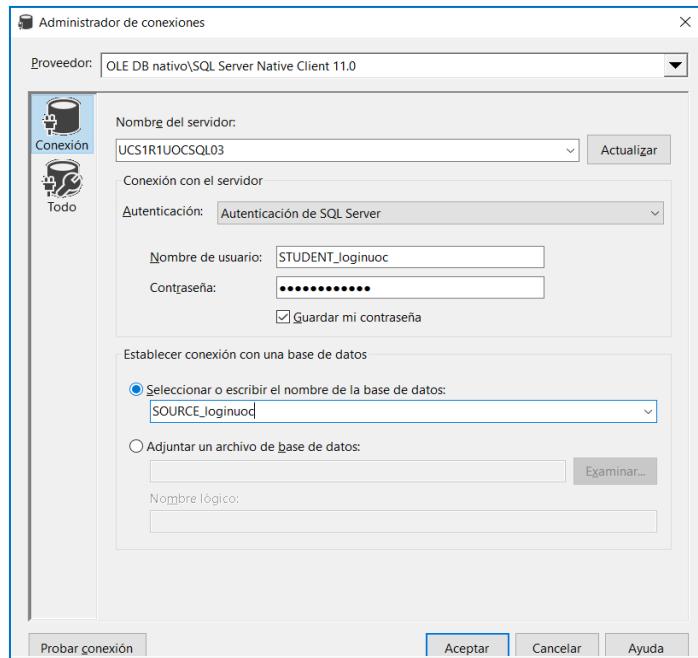


Después, se configurará el origen de los datos que alimentarán el cubo. Para ello:

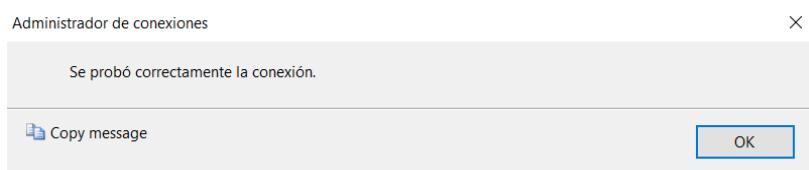
1. Dentro del explorador de soluciones, localizar el componente situado dentro de la carpeta **origen de datos** y hacer doble clic para abrir la ventana de configuración.
2. Pulsar el botón **Edit...** para abrir la ventana emergente.



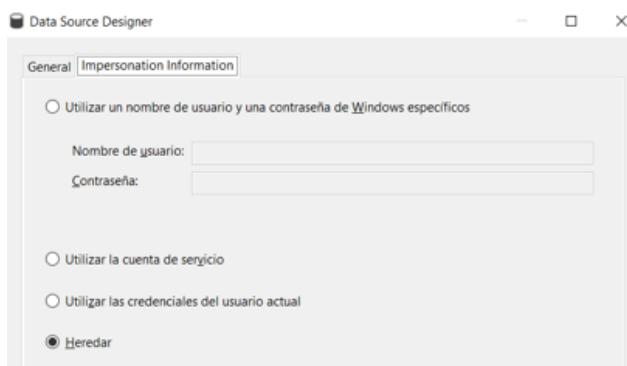
3. Se mostrará un nuevo diálogo para definir la conexión, en el cual se deberá verificar que el nombre del servidor es correcto. A continuación, se introducirán tanto el nombre de usuario (*STUDENT_loginuoc*, donde *loginuoc* es el nombre de usuario del campus virtual de la UOC) como la contraseña correspondiente.



4. Pulsar el botón **Probar conexión** para comprobar que la información es correcta.



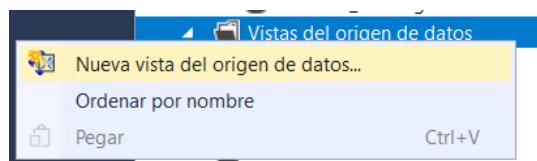
5. Finalmente, seleccionar la pestaña **Impersonation Information** y asegurarse de que esté activada la opción **Heredar**.



2.2. Crear una vista de origen de datos para el modelo de estrella de la tabla `FACT_Vehicle_Registration`, así como crear y procesar el cubo correspondiente

El siguiente paso consistirá en crear la vista `View_Vehicle_registrations`, destinada a incorporar los datos del modelo estrella, así como en la creación y el procesamiento del cubo `Cube_vehicle_registrations`, orientado a la explotación analítica de dicha información.

Para ello, será necesario hacer clic derecho sobre la carpeta **Vistas del origen de datos** en el **Explorador de soluciones** y seleccionar la opción **Nueva vista del origen de datos**.



A continuación, se mostrará un asistente cuyos pasos deberán seguirse en el orden indicado:

1. En el diálogo **Seleccionar un origen de datos**, aceptar el valor predeterminado y pulsar el botón **Next**.
2. En el diálogo **Coincidencia de nombres**, aceptar los valores propuestos y pulsar el botón **Next**.
3. En el diálogo **Seleccionar tablas y vistas**, seleccionar todas las tablas que van a formar parte del modelo: `DIM_Date`, `DIM_dgt_type`, `DIM_municipality`, `DIM_vehicle` y `FACT_vehicle_registration`.
4. En la pantalla final, asignar el nombre `View_Vehicle_registrations` a la nueva vista y pulsar el botón **Finalizar** para generar el componente.

The screenshot displays four windows from the 'Asistente para vistas del origen de datos' (Data Source Wizard) in Microsoft SQL Server Management Studio.

- Step 1: Asistente para vistas del origen de datos**

Shows the welcome screen with a cylinder icon and text about creating a new view of the data source. It includes a checkbox for 'No volver a mostrar esta página' (Do not show this page again).
- Step 2: Seleccionar un origen de datos**

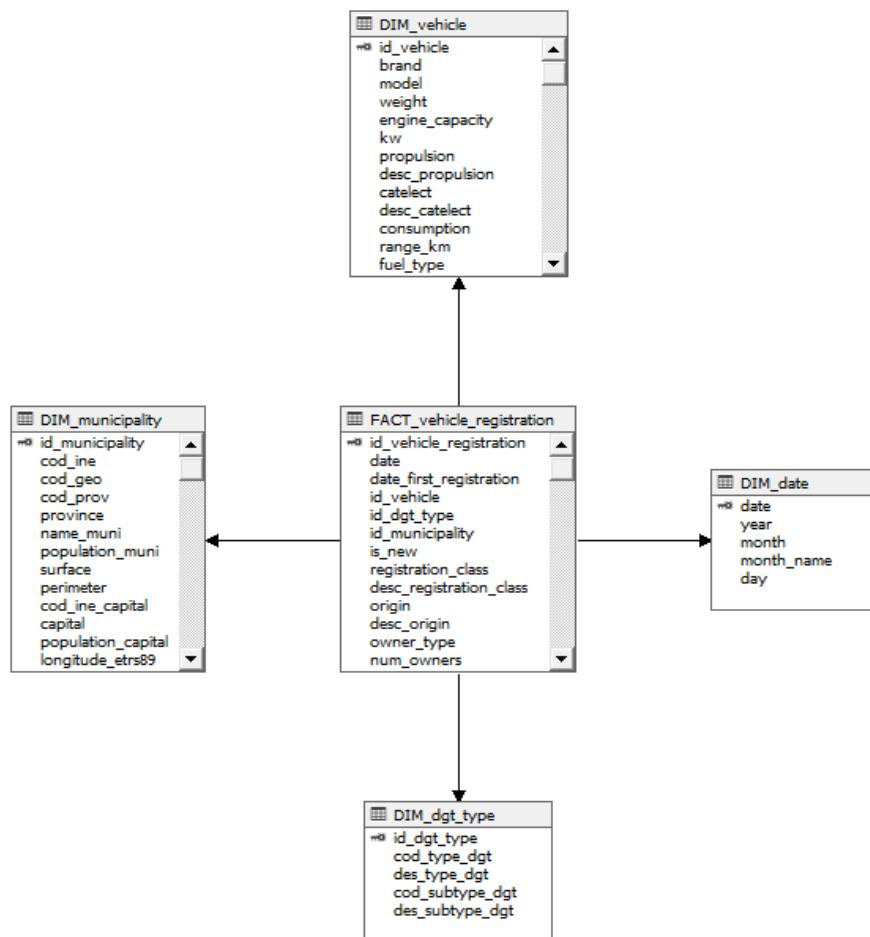
Shows the 'Relational data sources' list with 'SOURCE_ccalongec' selected. Properties for this source are listed on the right, including:
 - Data Source: UCSI1IUOCSQL03
 - Initial Catalog: SOURCE_ccalongec
 - Persist Sec.: True
 - Provider: SQLNCLI11.1
 - User ID: STUDENT_ccalon...
- Step 3: Coincidencia de nombres**

Shows the 'Matching names' step. It has a checked checkbox for 'Crear relaciones lógicas por columnas coincidentes' (Create logical relationships for matching columns). Below it, there are three radio button options:
 - Mismo nombre que el de la clave principal (Same name as the primary key)
 - Mismo nombre que el nombre de tabla de destino (Same name as the destination table name)
 - Nombre de la tabla de destino + nombre de la clave principal (Name of the destination table + primary key name)A note below says: 'El nombre de columna de clave externa de origen es igual que el nombre de la columna de clave principal de la tabla de destino. Por ejemplo: Order.CustomerID y Customer.CustomerID'
- Step 4: Seleccionar tablas y vistas**

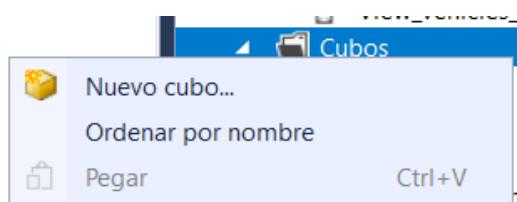
Shows the 'Select tables and views' step. It lists available objects (Nombre, Tipo) and included objects (Nombre, Tipo). Buttons for moving objects between lists are shown, along with a 'Filter' field and a 'Show system objects' checkbox.
- Step 5: Finalización del asistente**

Shows the finalization step where you can enter a name for the new view. The entered name is 'View_vehicles_registrations'. A preview window shows the selected tables: 'View_vehicles_registrations' (DIM_date, DIM_dgt_type, DIM_municipality, DIM_vehicle, FACT_vehicle_registration).

Una vez generada la vista, al hacer doble clic sobre ella en el Explorador de soluciones, se podrá visualizar el diagrama con las relaciones entre las tablas que la componen.

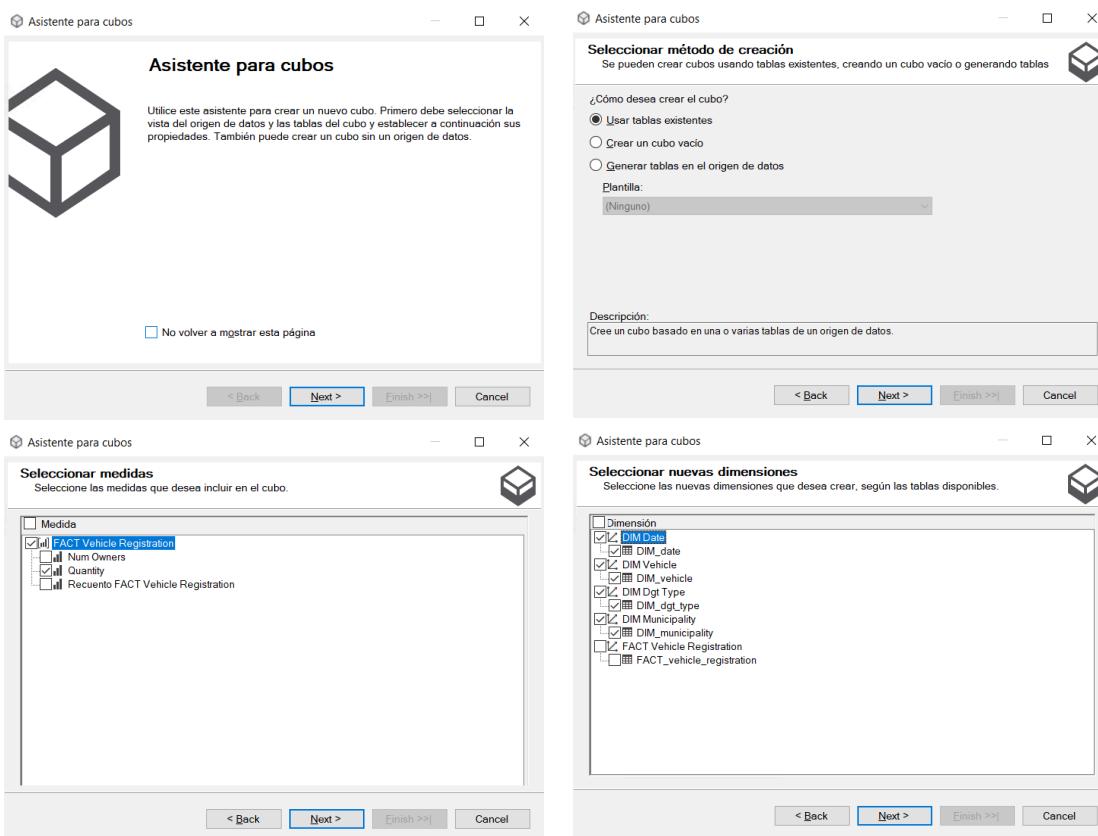


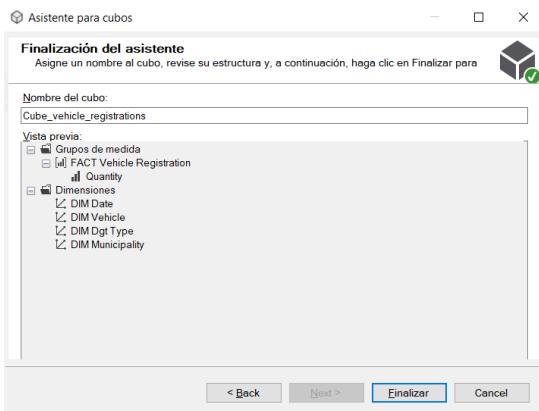
Tras la generación de este componente, se procederá a la creación del cubo. Para ello, hay que ubicarse en la carpeta **Cubos** dentro del **Explorador de soluciones**, y seleccionar la opción **Nuevo cubo**.



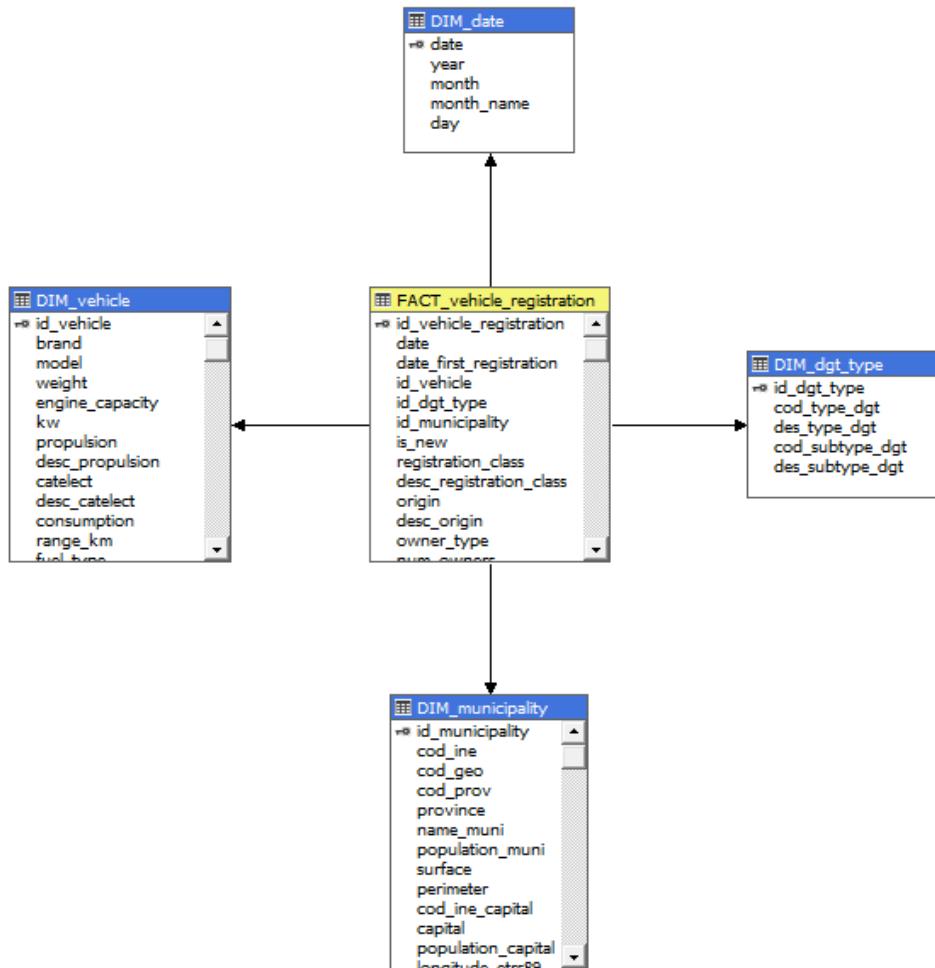
A continuación, se mostrará un asistente en el que se deberán seguir los siguientes pasos:

1. En el diálogo **Selección del método de creación**, seleccionar **Usar tablas existentes** y pulsar el botón **Next**.
2. En el diálogo **Seleccionar tablas y grupos de medida**, seleccionar la vista de origen de datos *View_Vehicle_registrations* creada en el paso anterior. A continuación, seleccionar la tabla *FACT_vehicle_registration* y pulsar el botón **Next**.
3. En el diálogo **Seleccionar medidas**, marcar únicamente la medida *Quantity* y pulsar el botón **Next**.
4. En el diálogo **Seleccionar nuevas dimensiones**, seleccionar todas las tablas cuyo nombre comienza por *DIM_* y excluir la tabla *FACT_vehicle_registration*. A continuación, pulsar el botón **Next**.
5. Finalmente, asignar el nombre *Cube_vehicle_registrations* al nuevo cubo, y pulsar el botón **Finalizar**.





Tras realizar estos pasos, se mostrará el diagrama del cubo, el cual también se puede visualizar haciendo doble clic sobre el componente en el *Explorador de soluciones*.



2.3. Jerarquías, dimensiones y atributos

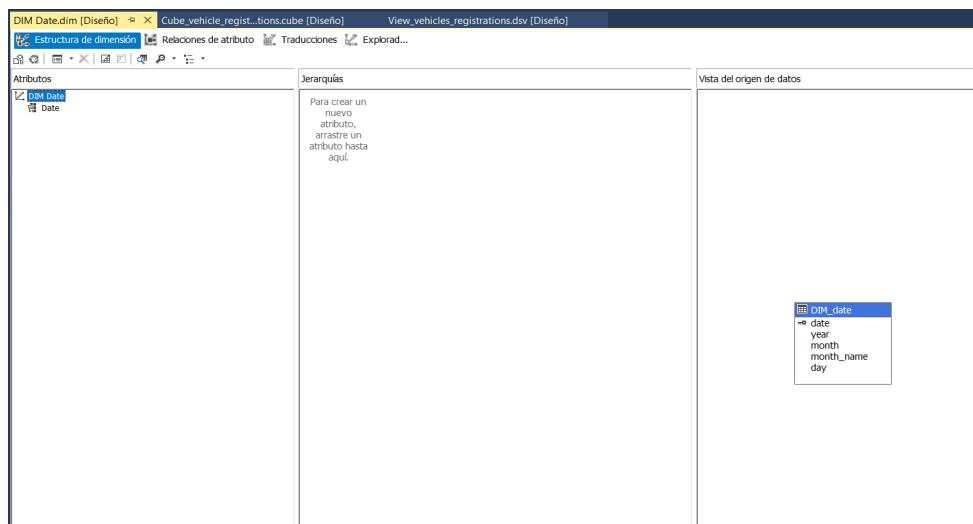
Como último paso en la configuración del cubo OLAP, es necesario configurar las dimensiones del modelo, especificando sus atributos y jerarquías. Esta configuración es esencial para permitir el filtrado, la navegación jerárquica y el análisis eficaz de la información.

Para ello, se debe hacer doble clic sobre la dimensión correspondiente en la carpeta *Dimensiones* del *Explorador de soluciones*, o bien seleccionar la opción *Editar dimensión* desde el menú contextual. Esto abrirá la dimensión en la pestaña *Estructura de dimensión*, donde se podrán definir los atributos y construir jerarquías según corresponda.

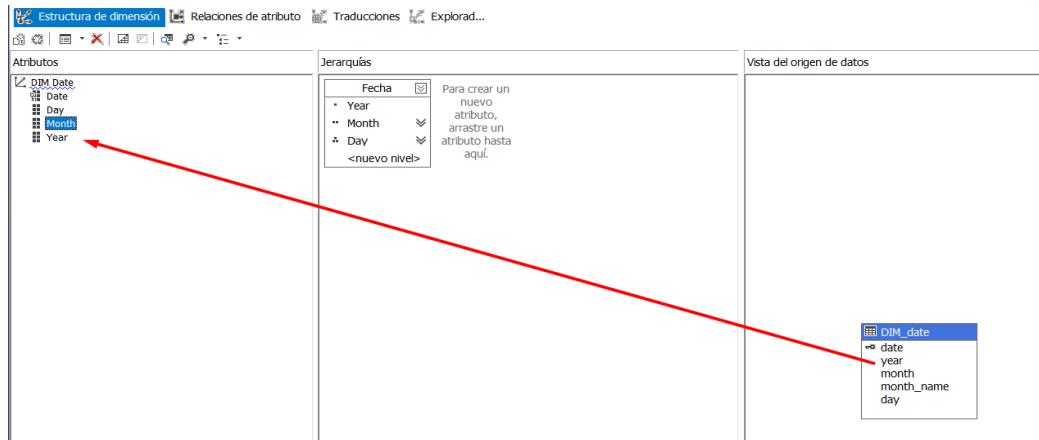
2.3.1. Configuración de la dimensión DIM_DATE

Esta dimensión representa la dimensión temporal, que suele ser una de las dimensiones más comunes en los cubos OLAP. En este caso se realizarán los siguientes pasos:

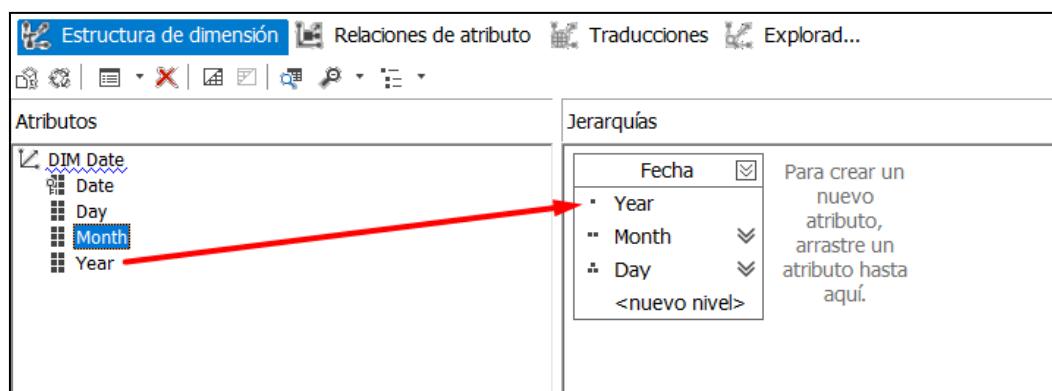
1. En la carpeta *Dimensiones* del *Explorador de soluciones*, se debe hacer doble clic sobre la dimensión **DIM_DATE**. A continuación, se abrirá la ventana de diseño de la dimensión.



2. Desde la parte derecha del diseñador, se debe hacer clic y arrastrar, uno a uno, los atributos *year*, *month* y *day* al área de *Atributos* situada en la parte izquierda del diseñador.



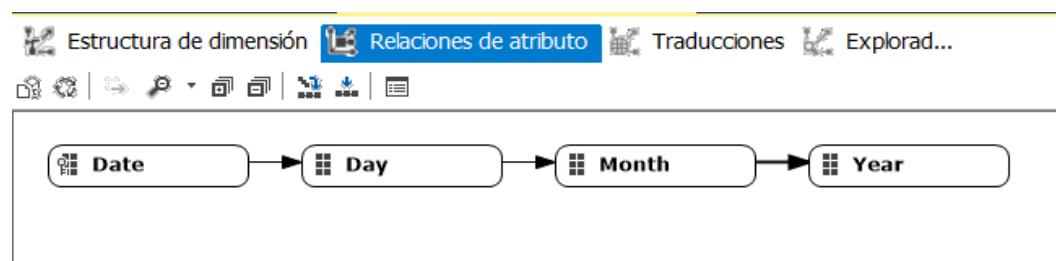
3. Se creará una jerarquía para mostrar los elementos de la dimensión en forma de árbol: *Año* > *Mes* > *Día*. Para ello, se deben arrastrar, uno a uno, los atributos *Year*, *Month* y *Day* al área *Jerarquías*, comenzando por *Year* y colocando el resto de niveles en el orden correspondiente. A continuación, se debe hacer clic derecho sobre el nombre de la jerarquía y, mediante la opción *Cambiar nombre*, se procederá a renombrarla como *Fecha*.



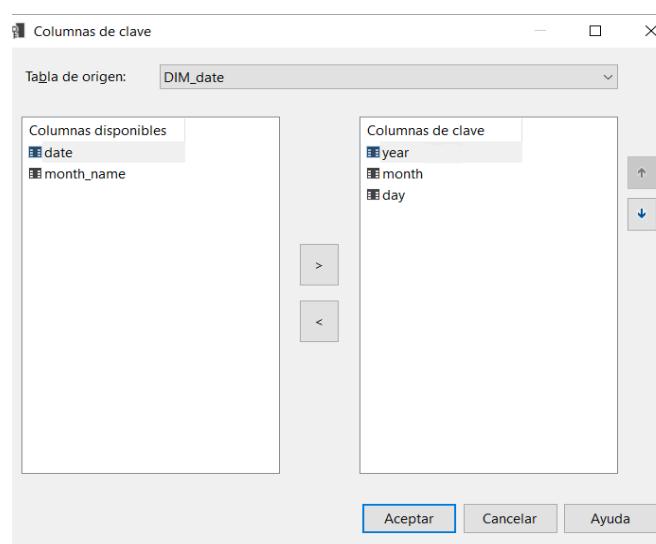
4. Como parte de la configuración de la jerarquía, será necesario configurar las *Relaciones de atributo*. Para ello, se debe hacer clic en la pestaña *Relaciones de atributo* y, utilizando el botón *Nueva relación de atributo* o editando manualmente las flechas de relación, verificar que se ha definido la jerarquía

Date > Day > Month > Year, asegurando que los niveles estén ordenados del más detallado al más general.

A continuación, se debe cambiar el tipo de relación entre atributos para que sea *rígida* (es decir, los datos no cambian con el tiempo). Esto se realiza haciendo clic con el botón derecho sobre la flecha correspondiente y seleccionando la opción *Tipo de relación*.



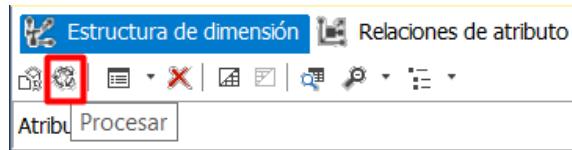
5. También es necesario garantizar que cada nivel de la jerarquía esté correctamente configurado en cuanto a la columna clave (*KeyColumn*) y los valores a mostrar (*NameColumn*). Una configuración incorrecta en este aspecto puede provocar errores durante el procesamiento de la dimensión. Para ello, en la pestaña *Estructura de dimensión*, se debe seleccionar el atributo *Day* y modificar las siguientes propiedades:
 - **KeyColumns:** se deben añadir nuevas columnas de clave de forma que correspondan, en este orden, a *Year*, *Month* y *Day*.



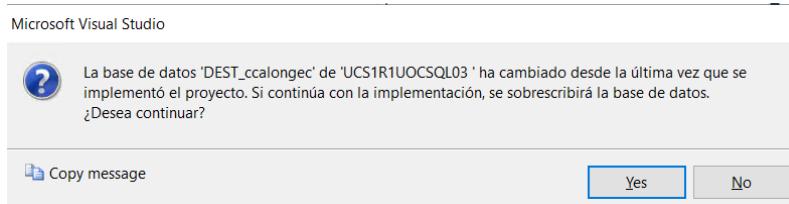
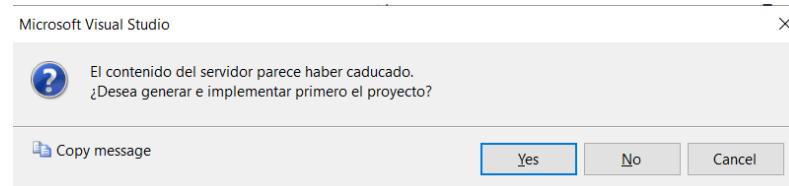
- **NameColumn:** se debe seleccionar la columna *Day*, para que este valor sea el que se muestre en los elementos de la jerarquía.
 - **OrderBy:** se debe establecer la opción *Key*. Esta configuración es importante para asegurar que los elementos de la jerarquía aparecen ordenados correctamente de forma cronológica.
6. A continuación, se editará el atributo *Month* para asignarle las siguientes propiedades:
- **KeyColumns:** se deben establecer las columnas *Year* y *Month*, en ese orden.
 - **NameColumn:** se debe seleccionar la columna *Month*.
 - **OrderBy:** se debe configurar la opción *Key*, con el fin de garantizar el orden cronológico de los valores.
7. Para el atributo *Year*, se mantendrán los valores predeterminados de **KeyColumns**, **NameColumn** y **OrderBy**.
8. Finalmente, se debe establecer la propiedad **AttributeHierarchyVisible** en *False* para los atributos *Year*, *Month* y *Day*, ya que únicamente se requiere que se muestren dentro de la jerarquía de tres niveles definida previamente, y no como filtros independientes en el cubo

Propiedades	
Day	DimensionAttribute
AttributeHierarchyDisplayFolder	
AttributeHierarchyEnabled	True
AttributeHierarchyOptimizedState	FullyOptimized
AttributeHierarchyOrdered	True
AttributeHierarchyProcessingState	Unprocessed
AttributeHierarchyVisible	False
CustomRollupColumn	(ninguno)
CustomRollupPropertiesColumn	(ninguno)
DataEncodingHint	AutomaticEncoding

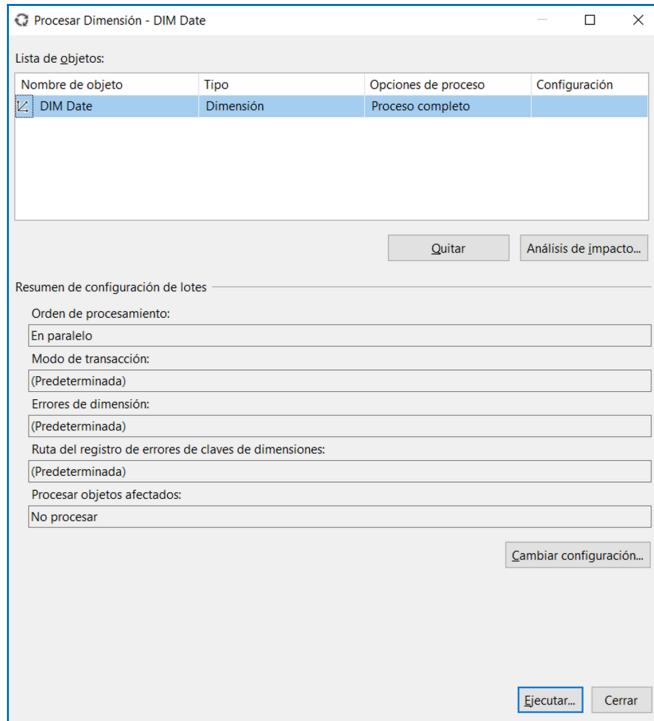
Una vez realizadas estas configuraciones, se debe pulsar el botón **Procesar**, lo que permitirá ejecutar los cálculos necesarios para generar los elementos de la dimensión.



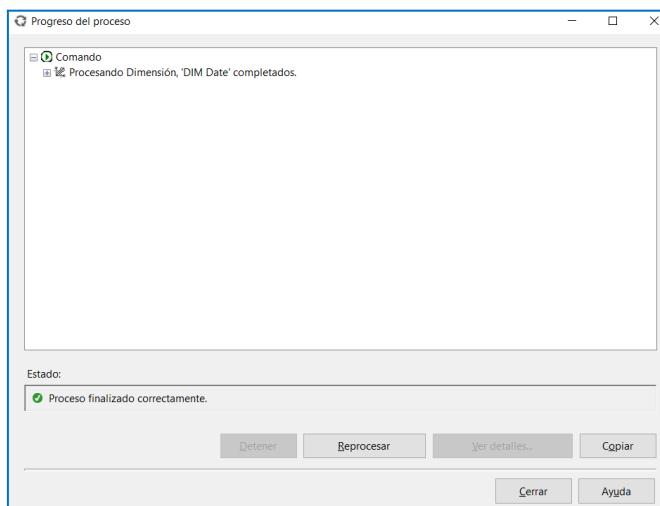
Al pulsar el botón, se mostrarán varios mensajes de advertencia que deberán aceptarse. Estos mensajes hacen referencia a la sincronización entre el proyecto local y el servidor de *Analysis Services*, y es necesario confirmarlos para continuar con la implementación.



A continuación, se mostrará un cuadro de diálogo con un resumen de la operación. En dicho cuadro, se debe pulsar el botón *Ejecutar* para iniciar el proceso.



Una vez finalizada la ejecución, se mostrará otro cuadro de diálogo con el resumen del procesamiento, indicando que se ha completado correctamente.:



Finalmente, tras cerrar las ventanas del procesamiento de la dimensión, se debe hacer clic en la pestaña *Explorador de dimensiones*. Esta pestaña permitirá visualizar la jerarquía creada y navegar por sus valores

The screenshot shows a user interface for managing a data dimension. At the top, there are several tabs: 'Estructura de dimensión' (Structure), 'Relaciones de atributo' (Attribute relationships), 'Traducciones' (Translations), and 'Explorad...' (Explore...). Below the tabs are search and filter tools, including a dropdown for 'Jerarquía' (Hierarchy) set to 'Fecha' and another for 'Idioma' (Language) set to 'Predeterminado' (Default).

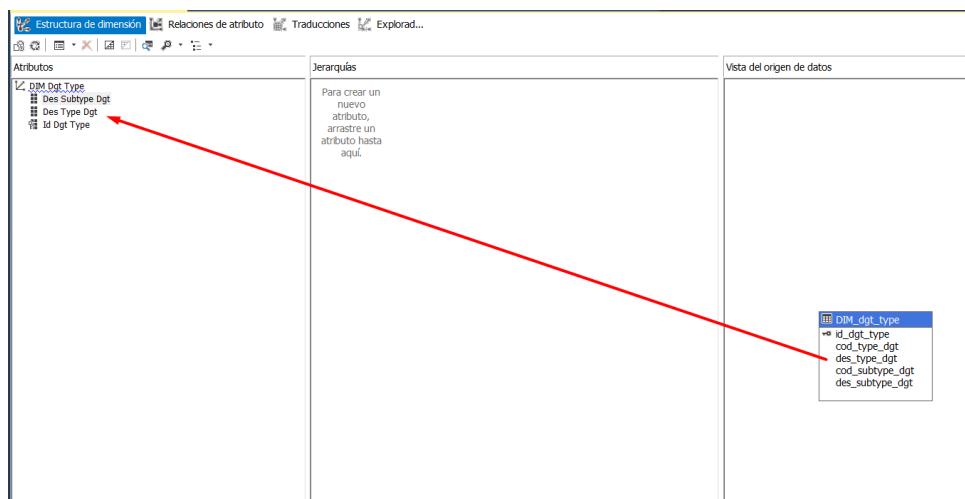
The main area displays a hierarchical tree structure under 'Nivel actual: (All)'. The tree starts with 'All', which is expanded to show '2008', then '1', then '10' sub-levels. Each level is represented by a circular icon with a number inside.

- All
- 2008
 - 1
 - 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8
 - 9
 - 10

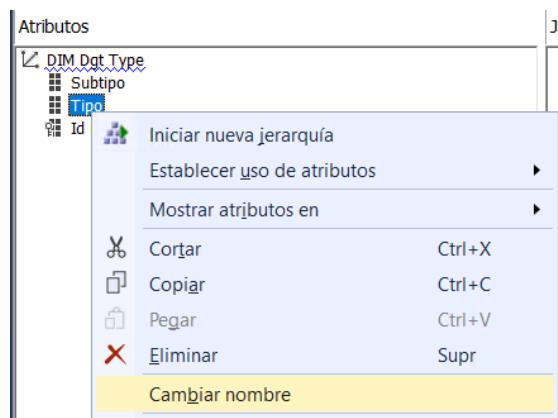
2.3.2. Configuración de la dimensión DIM_DGT_TYPE

Esta dimensión permite analizar los datos por una clasificación de vehículos definida por la Dirección General de Tráfico (DGT). La dimensión debe configurarse para mostrar una jerarquía de Tipo y Subtipo, realizando los pasos mostrados a continuación:

1. Hacer doble clic en la dimensión **DIM_dgt_type** del *Explorador de Soluciones* para abrir el diseñador de la dimensión.
2. Arrastrar los atributos *des_type_dgt* y *des_subtype_dgt* al área de atributos de la dimensión.



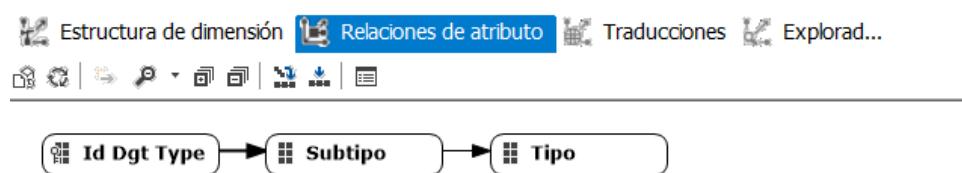
3. Haciendo clic derecho en el nombre de cada atributo, cambiar el nombre a *Tipo* y *Subtipo* respectivamente.



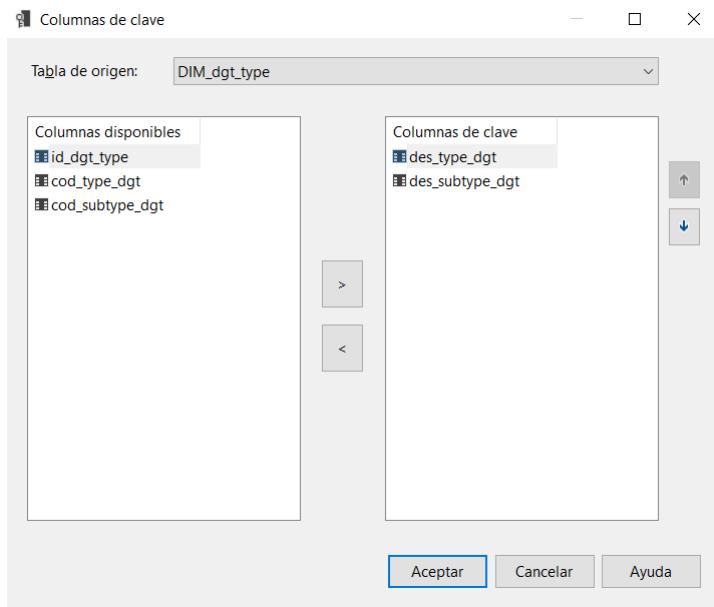
4. Arrastrar ambos atributos al área de Jerarquía, para formar una jerarquía de dos niveles:



5. Configurar las relaciones de atributo para que tengan el orden adecuado: *Id Dgt Type > Subtipo > Tipo*



6. Se configurará la propiedad **KeyColumns** del atributo *Subtipo* para tener como columnas clave los atributos *des_type_dgt* y *des_subtype_dgt*. Su propiedad **NameColumn** se cambiará para tener el atributo *des_subtype_dgt*.



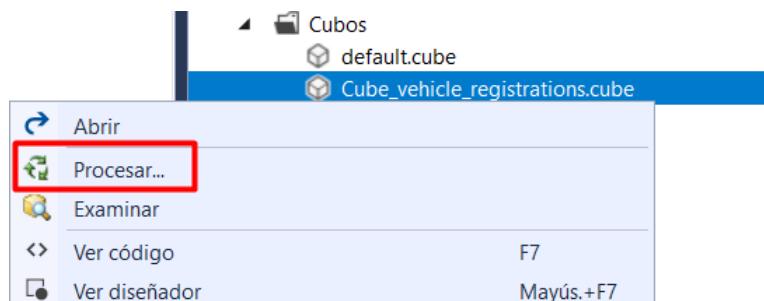
7. Se configurará la propiedad *AttributeHierarchyVisible* a *False* para los atributos *Id dgt type*, *Subtipo* y *Tipo*, de forma que sólo se muestre la jerarquía que se ha definido en los puntos anteriores.

Una vez realizados estos pasos, se procesará la dimensión y se explorarán sus datos:

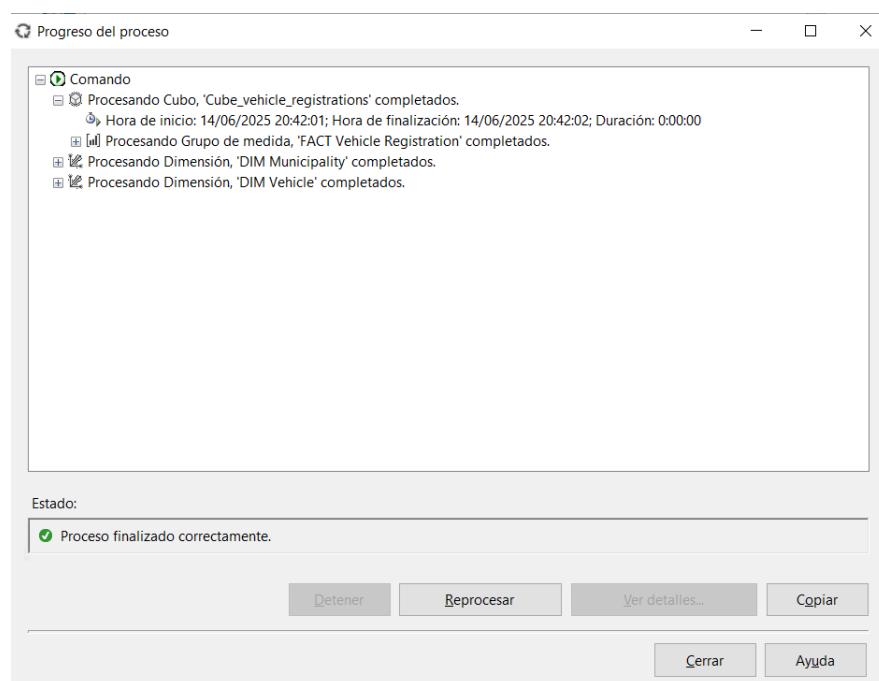
- (All)
- AUTOBUSES
 - AUTOBÚS
 - AUTOBÚS ARTICULADO
 - AUTOBÚS LABORATORIO
 - AUTOBÚS MIXTO
 - AUTOBÚS SANITARIO
 - AUTOBÚS TALLER
 - BIBLIOBÚS
 - CAMIONES
 - CICLOMOTORES
 - FURGONETAS
 - MOTOCICLETAS
 - OTROS vehiculos
 - R Y S
 - TRACTORES INDUSTRIALES
 - TURISMOS
 - Unknown

2.4. Procesamiento y despliegue del cubo

Una vez realizadas estas configuraciones, se desplegará el cubo en *Analysis Services*. Para ello, primero se hará clic derecho el cubo *Cube_vehicle_registrations* en el *Explorador de Soluciones* y se pulsará el botón **Procesar**.

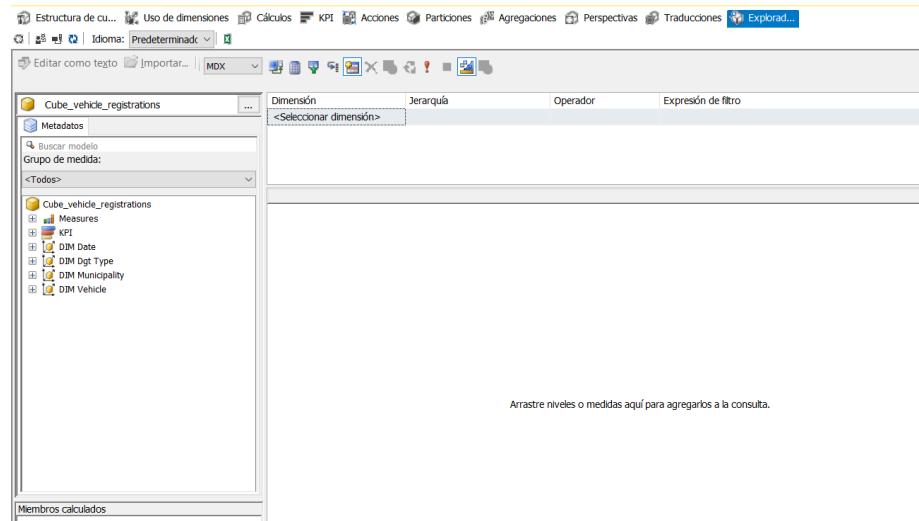


Tras aceptar los mensajes de advertencias y ejecutar el proceso, se obtendrá un mensaje indicando que el cubo se ha procesado correctamente.

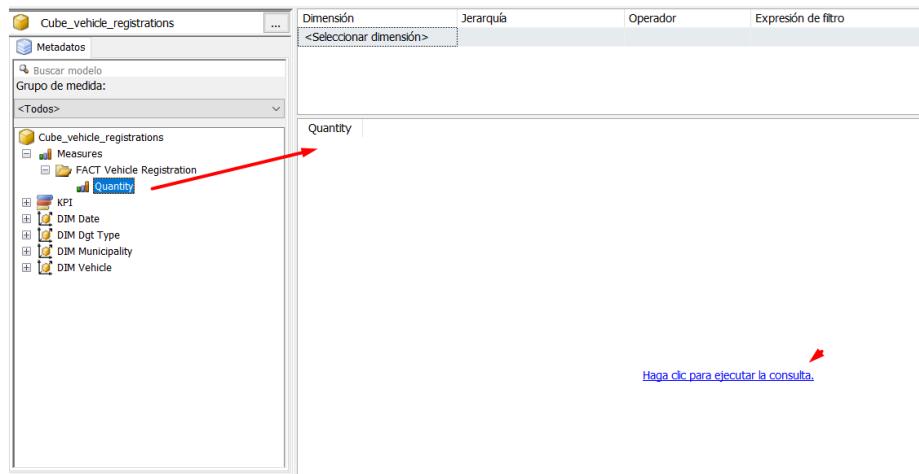


Tras ello, se deben cerrar las ventanas correspondientes al proceso de implementación del cubo y hacer doble clic sobre el cubo para abrir su diseñador. A

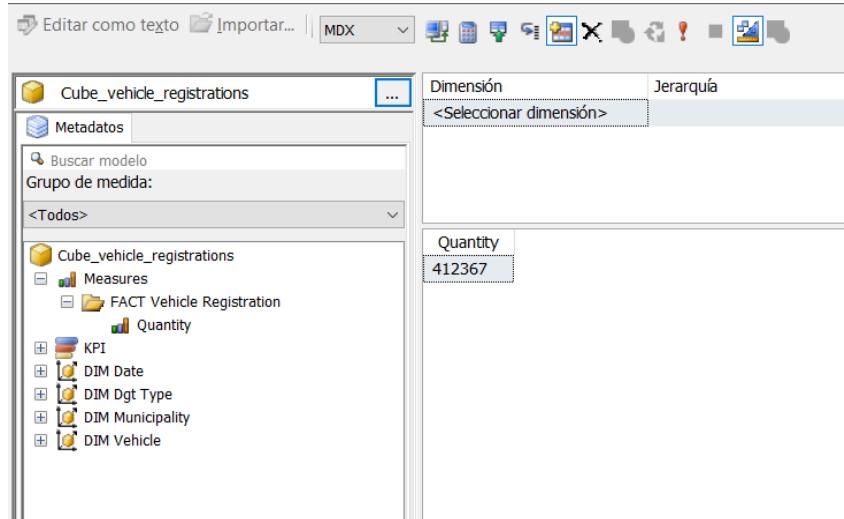
continuación, se debe seleccionar la opción **Explorador** para visualizar y navegar por el cubo implementado.



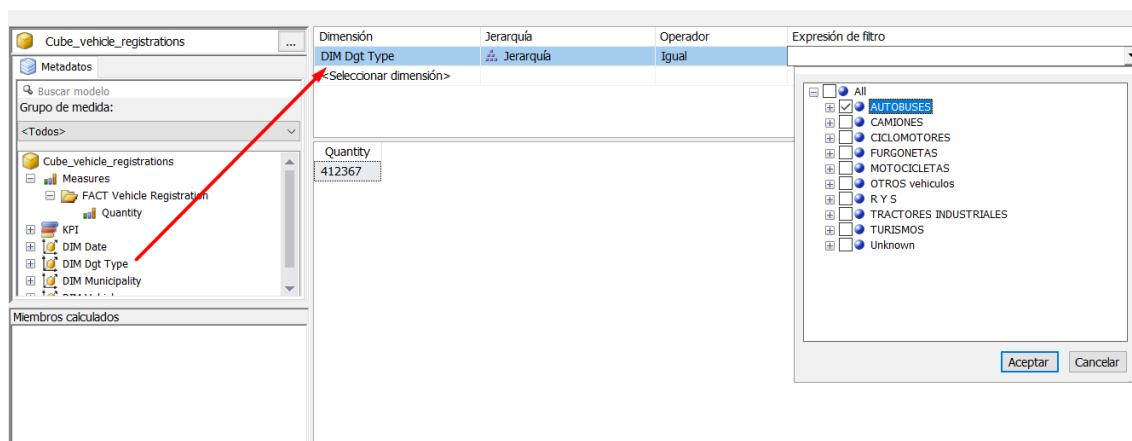
Se realizará una prueba de navegación, arrastrando la medida **Quantity** al área de resultados y pulsando clic en el enlace del área de resultados para ejecutar la consulta:



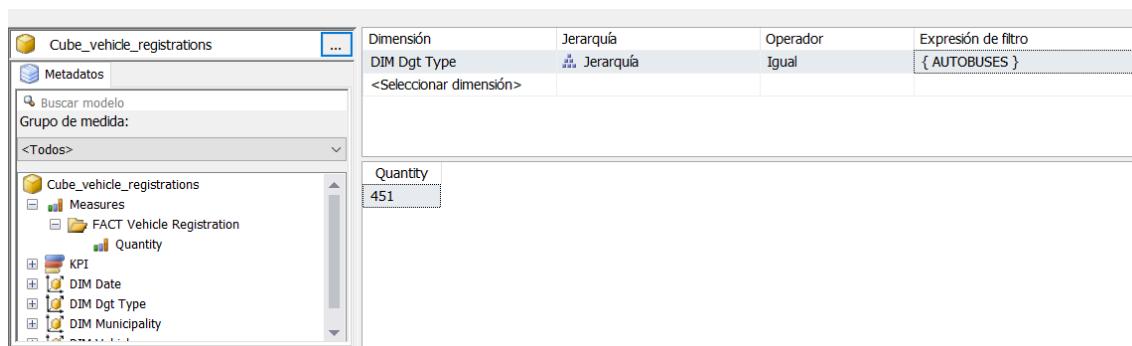
Finalmente, se mostrarán los resultados:



Para comprobar el funcionamiento de los filtros, se debe arrastrar la dimensión *DIM_dgt_type* al área de filtros y seleccionar el valor "AUTOBUSES". Esta acción actualizará automáticamente la cantidad mostrada en función del filtro aplicado.



Dimensión	Jerarquía	Operador	Expresión de filtro
DIM Dgt Type	Jerarquía	Igual	{AUTOBUSES}



Dimensión	Jerarquía	Operador	Expresión de filtro
DIM Dgt Type	Jerarquía	Igual	{AUTOBUSES}

2.5. Ejercicios de configuración del cubo OLAP

Siguiendo los pasos anteriores, se requiere realizar los siguientes ejercicios para acabar de completar la configuración del cubo:

1) Configurar la dimensión DIM Vehicles, con las siguientes características:

- Debe incluir los atributos de dimensión: *brand*, *model*, *desc_categ*, *desc_fuel_type*, *desc_propulsion*, *desc_eco_label_type* y *seats*. Estos atributos deberán renombrarse como *Marca*, *Modelo*, *Categoría Eléctrica*, *Alimentación*, *Propulsión*, *Etiqueta ECO* y *Plazas*, respectivamente.
- Debe tener una jerarquía de dos niveles formada por *Marca* y *Modelo*.
- El atributo *Id Vehicle* no debe configurarse como jerarquía.

2) Configurar la dimensión DIM_Municipality, con las siguientes características:

- Debe incluir los atributos de dimensión: *province*, *name_muni* y *population_muni*. Estos atributos se deben renombrar a *Provincia*, *Municipio* y *Habitantes* respectivamente.
- Debe tener una jerarquía de dos niveles formada por *Provincia* y *Municipio*.
- Los atributos individuales *Id Municipality*, *Provincia* y *Municipio* no deben configurarse como jerarquía.

3) Procesar el cubo completo tras los cambios realizados y verificar que el proceso ha finalizado correctamente. Explicar por qué es necesario realizar este paso.

2.6. Ejercicios de explotación del cubo OLAP

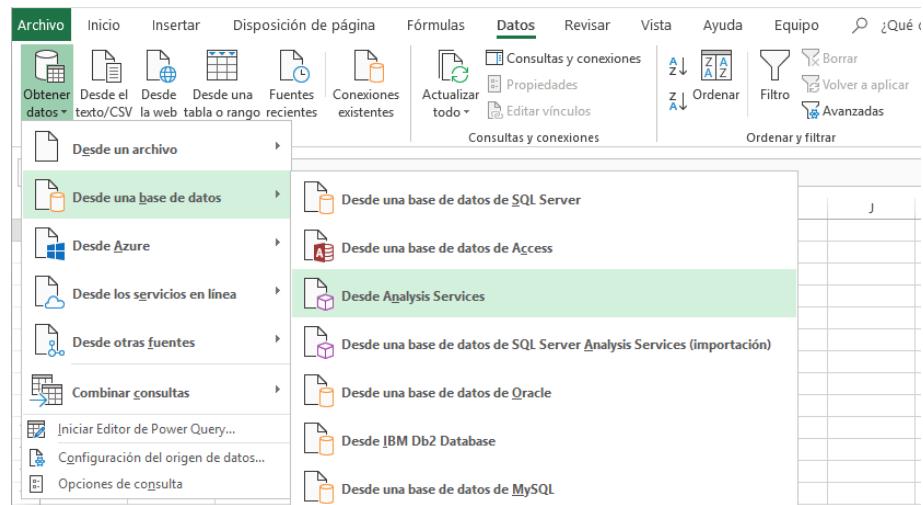
Utilizando el explorador de cubos disponible en el diseñador de *Visual Studio*, se deben realizar las siguientes consultas:

- 1) Listar el total de matriculaciones por año.
- 2) Listar el total de matriculaciones en la provincia de Madrid desglosado por tipo de vehículo.
- 3) Listar el total de coches matriculados con etiqueta ECO = *Etiqueta 0 emisiones* desglosados por marca y modelo, del año 2020 en adelante y que sean Turismos.
- 4) Listar el total de matriculaciones de autobuses de propulsión Eléctrica desglosados por Municipio, Marca y Modelo.
- 5) Listar el total de vehículos de 4 o 5 plazas matriculados en el 2023 desglosados por mes y etiqueta ECO.

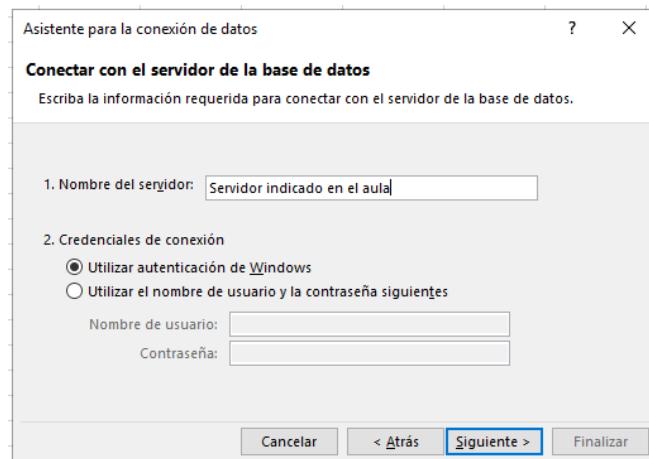
Además de realizar consultas desde el explorador de cubos, a continuación se solicita construir algunas consultas utilizando la herramienta ofimática **Microsoft Excel**:

Para ello, primero se configurará la herramienta para que se conecte al cubo:

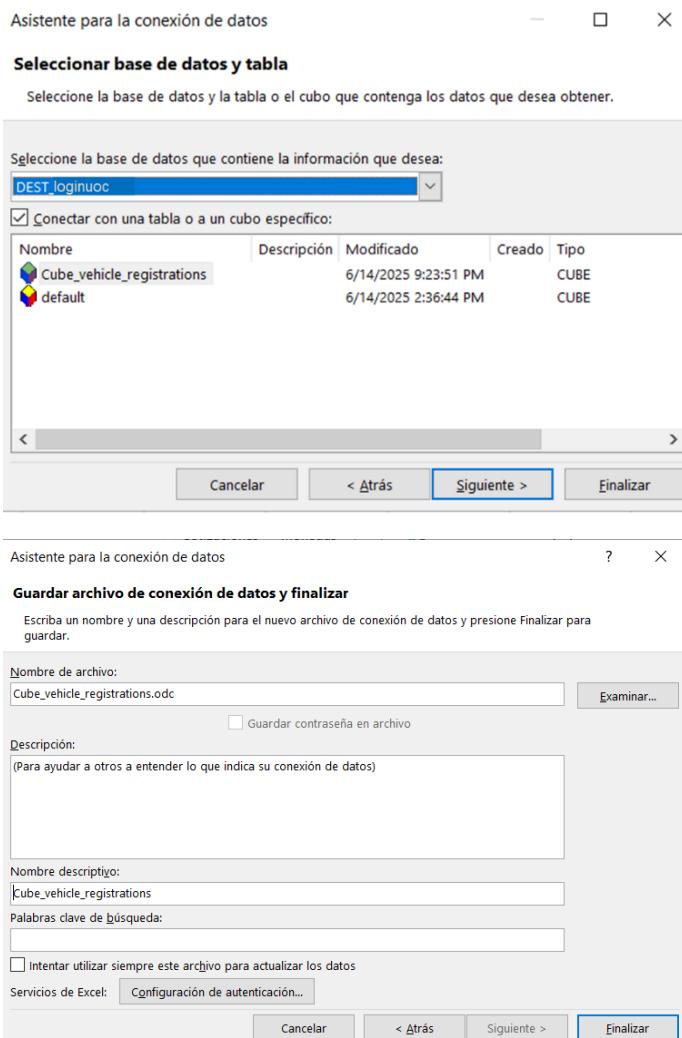
- Desde MS Excel crear una conexión con Analysis Services



- Se deberá indicar el nombre del servidor y mantener la opción *Utilizar autenticación de Windows* como método de conexión.



- Se seleccionará el cubo *Cube_vehicle_registrations* y se guardará la conexión creada.



- A partir de este punto ya se pueden crear informes o gráficos dinámicos.

Se solicita construir en MS Excel las siguientes consultas:

- 6) Listar el total de matriculaciones por marca y tipo de propulsión.
- 7) Crear un gráfico de barras verticales en el que se pueda observar la evolución de las matriculaciones por año.
- 8) Crear una matriz con las matriculaciones por provincia (filas) y año (columnas), para los vehículos tipo Turismos.

En esta actividad se evaluará la creación correcta de los diferentes componentes del cubo OLAP (proyectos, vistas de origen de datos, cubos, dimensiones, atributos y

jerarquías) y la realización de consultas multidimensionales. Será necesario mostrar evidencia mediante capturas de pantalla o similar de la configuración de los componentes del cubo creados, acompañadas de un breve resumen que explique el proceso realizado. Además, en la explotación del cubo, se valorará la presentación de las consultas realizadas en un formato similar, donde se pueda apreciar de forma clara las medidas seleccionadas, los filtros aplicados y los resultados obtenidos.

Ejercicio 3. Análisis de datos mediante Dashboards (35%)

Objetivo:

Adquirir conocimientos básicos sobre la aplicación **Microsoft Power BI**, disponible en el entorno VDI, con el objetivo de diseñar informes y gráficos utilizando modelos multidimensionales en memoria (*in-memory*).

Indicaciones para el desarrollo del ejercicio:

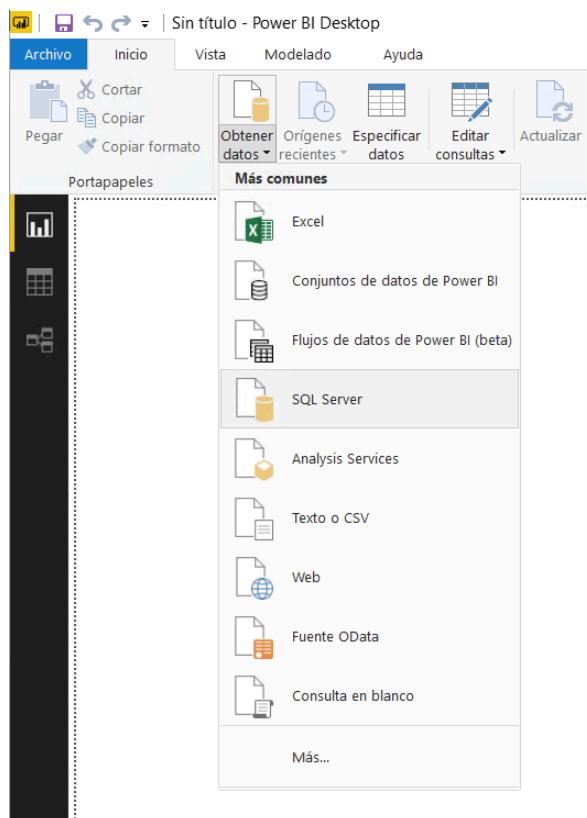
Para disponer de un conjunto de datos destinado al análisis, y en caso de no haberlo hecho previamente, es necesario ejecutar el script de creación de tablas y carga de registros en la base de datos SQL Server: **PR4_DDL+Data_SQLS.sql**, facilitado junto al enunciado.

Dada la volumetría de los datos a cargar, la ejecución del script podría tardar entre 10 y 15 minutos.

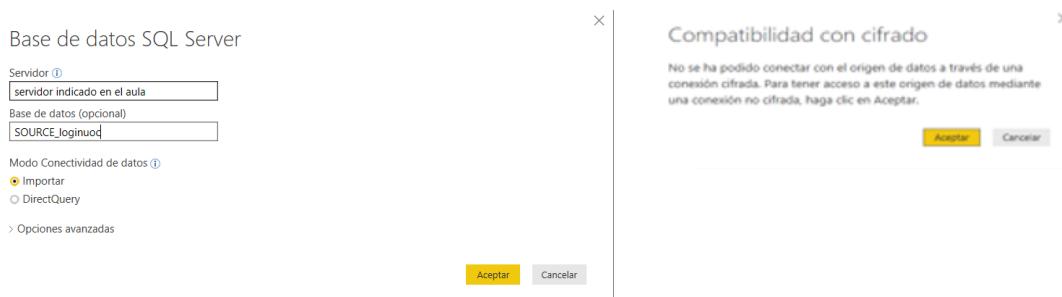
El ejercicio se desarrollará íntegramente con la herramienta **Power BI Desktop**, disponible en el escritorio del entorno virtual.

3.1. Configurar un origen de datos (DataSource) conectado a SOURCE_loginuoc

Para indicar que el origen de los datos sea la base de datos de **Microsoft SQL Server**, se debe seleccionar la opción *Obtener datos* y, a continuación, elegir *SQL Server* como tipo de origen.



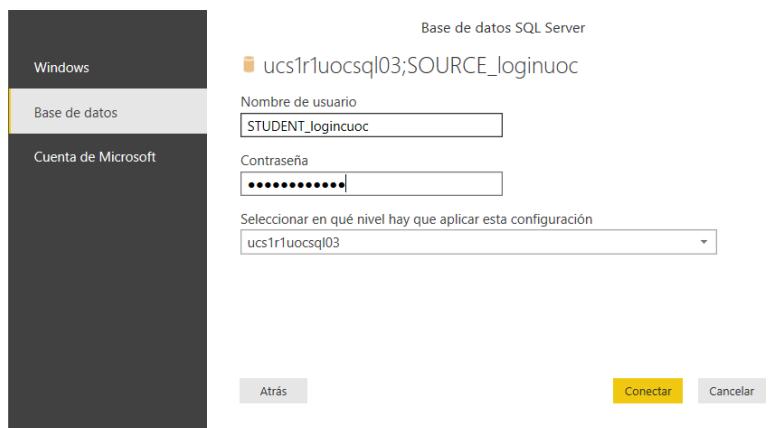
En el campo **Servidor**, se debe indicar el nombre de la instancia de **Microsoft SQL Server**, y en **Base de datos**, se debe especificar **SOURCE_loginuoc**. A continuación, se debe dejar seleccionada la opción *Importar* como modo de conectividad de datos y aceptar la compatibilidad con cifrado si se solicita.



En el siguiente diálogo se informarán de las credenciales de acceso a los datos:

- Seleccionar la opción *Base de Datos*.

- En el campo *Nombre de usuario*, escribir **STUDENT_loginuoc**, donde *loginuoc* corresponde al nombre de usuario del campus de la UOC.
- En el campo *Contraseña*, introducir la contraseña correspondiente a las credenciales de acceso a **SQL Server**.

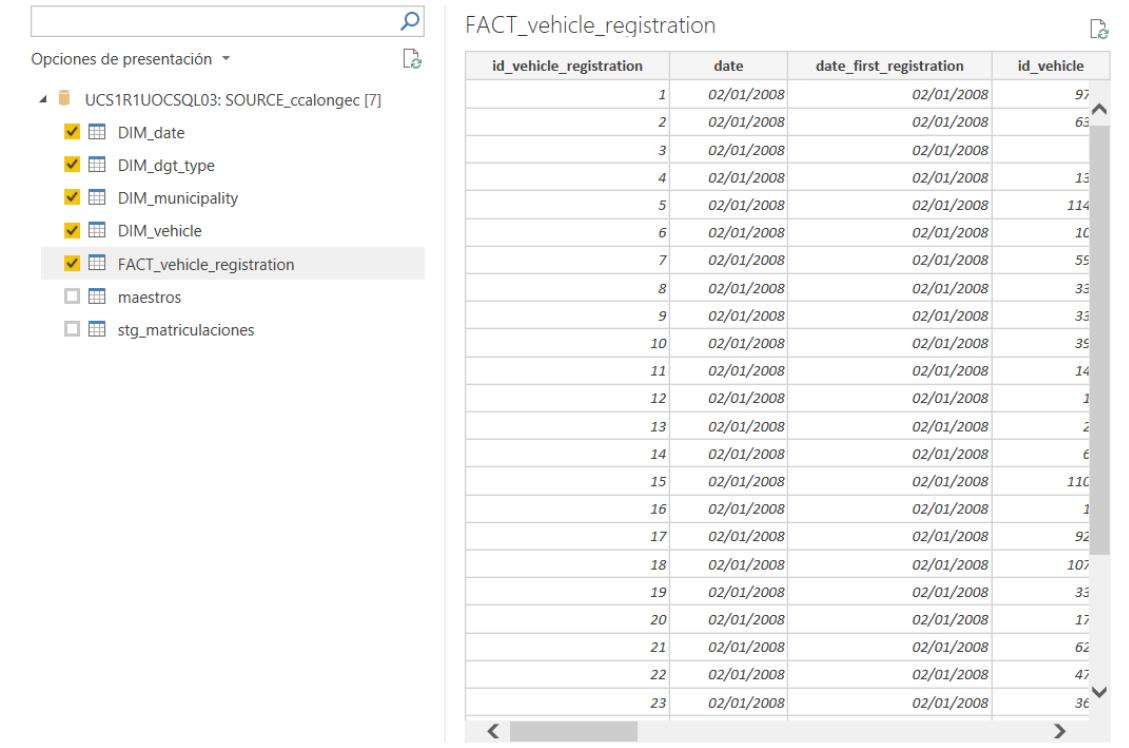


- Al hacer clic en *Conectar*, es posible que se muestre un mensaje de Compatibilidad con cifrado, el cual deberá aceptarse.



A continuación, se solicitará la selección de las tablas a importar. Se deben marcar las siguientes: **DIM_date**, **DIM_dgt_type**, **DIM_municipality**, **DIM_vehicle** y **FACT_vehicle_registration**. Una vez seleccionadas, se debe pulsar el botón *Cargar*.

Navegador



The screenshot shows the Power BI Navigator interface. On the left, there is a sidebar with a search bar and a 'Opciones de presentación' dropdown. Below these are several items listed with checkboxes: 'UCS1R1UOCSQL03: SOURCE_ccalongec [7]' (unchecked), 'DIM_date' (checked), 'DIM_dgt_type' (checked), 'DIM_municipality' (checked), 'DIM_vehicle' (checked), 'FACT_vehicle_registration' (checked, highlighted in grey), 'maestros' (unchecked), and 'stg_matriculaciones' (unchecked). At the bottom of the sidebar are buttons for 'Seleccionar tablas relacionadas', 'Cargar' (highlighted in yellow), 'Editar', and 'Cancelar'. The main area displays a table titled 'FACT_vehicle_registration' with four columns: 'id_vehicle_registration', 'date', 'date_first_registration', and 'id_vehicle'. The table contains 23 rows, each with the value '02/01/2008' in the 'date' and 'date_first_registration' columns, and various values in the other columns. The last row is highlighted with a yellow background.

id_vehicle_registration	date	date_first_registration	id_vehicle
1	02/01/2008	02/01/2008	97
2	02/01/2008	02/01/2008	63
3	02/01/2008	02/01/2008	
4	02/01/2008	02/01/2008	13
5	02/01/2008	02/01/2008	114
6	02/01/2008	02/01/2008	10
7	02/01/2008	02/01/2008	55
8	02/01/2008	02/01/2008	35
9	02/01/2008	02/01/2008	35
10	02/01/2008	02/01/2008	35
11	02/01/2008	02/01/2008	14
12	02/01/2008	02/01/2008	1
13	02/01/2008	02/01/2008	2
14	02/01/2008	02/01/2008	6
15	02/01/2008	02/01/2008	110
16	02/01/2008	02/01/2008	1
17	02/01/2008	02/01/2008	92
18	02/01/2008	02/01/2008	107
19	02/01/2008	02/01/2008	35
20	02/01/2008	02/01/2008	17
21	02/01/2008	02/01/2008	62
22	02/01/2008	02/01/2008	47
23	02/01/2008	02/01/2008	36

Una vez cargada la información, en el área *Relaciones*, ubicada en el panel izquierdo de la herramienta, se podrá visualizar un diagrama que muestra las distintas tablas junto con sus atributos, así como las relaciones entre ellas detectadas automáticamente por la herramienta.



Creación de una columna personalizada

Para facilitar el análisis, se realizará un cambio en el modelo de datos a partir del campo *owner_type*, que contiene los valores "PF" y "PJ". Se creará una nueva columna con valores más descriptivos: "Persona Física" y "Persona Jurídica".

Para ello, se seleccionará la opción **Datos** en el panel izquierdo. En el panel derecho, se hará clic con el botón derecho sobre la tabla **FACT_vehicle_registration** y se seleccionará la opción **Nueva columna**:

The screenshot shows a Microsoft Power BI Data View window. At the top, there's a ribbon with a 'New Column' button highlighted by a red box. On the right side, a 'CAMPOS' (Fields) pane is open, showing a tree structure of tables: DIM_date, DIM_dgt_type, DIM_municipality, and FACT_vehicle_registration, which is also highlighted by a red box. Below the pane are several context menu options: Nueva medida, Nueva columna, Nueva medida rápida, Actualizar datos, Editar consulta, Copiar tabla, Cambiar nombre, Eliminar, Ocultar en la vista de informes, Marcar como tabla de fechas, and Mostrar todos.

A continuación, se habilitará en la parte superior el editor de fórmulas, donde se introducirá la siguiente expresión:

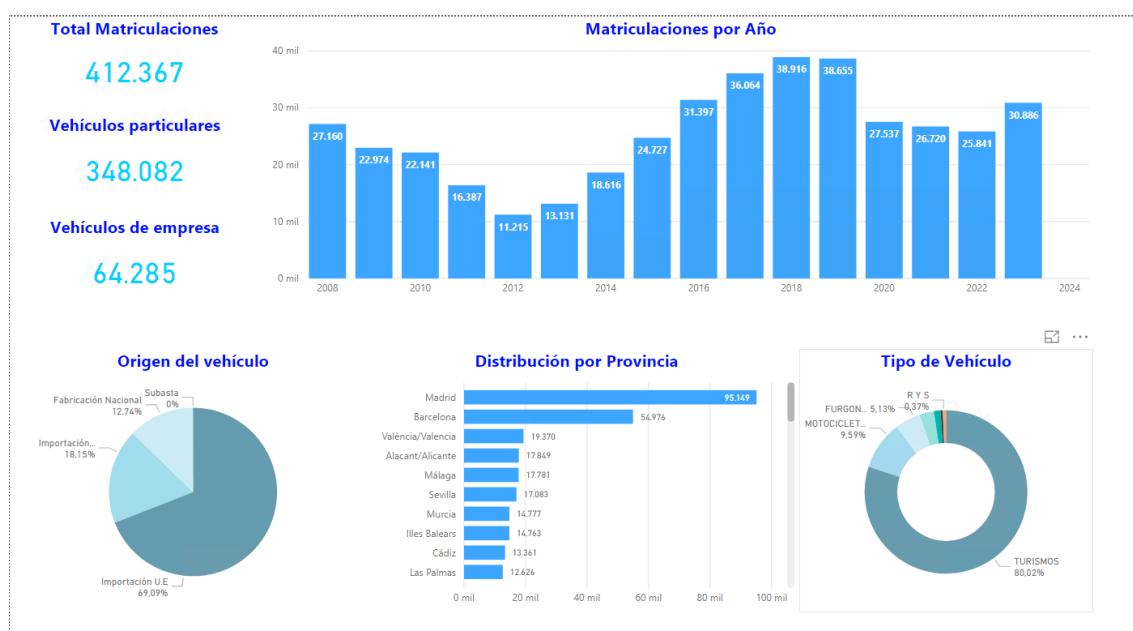
```
tipo_propietario = SWITCH(FACT_vehicle_registration[owner_type];
    "PF"; "Persona Física";
    "PJ"; "Persona Jurídica";
    "Desconocido"
)
```

Una vez aceptada la fórmula, se generará la nueva columna con los valores correspondientes, que podrá utilizarse en las visualizaciones del informe:

The screenshot shows the same Microsoft Power BI Data View window after the formula has been accepted. The table now includes a new column at the end named 'tipo_propietario'. The first five rows of the table show the formula definition, and the subsequent rows show the resulting values for each row. The values in the 'tipo_propietario' column are: Persona Física, Persona Física, Persona Física, Persona Física, and Persona Física respectively.

3.2. Construcción de la pestaña Matriculaciones

Para construir la pestaña **Matriculaciones**, es necesario incorporar distintos elementos gráficos que permitan representar los datos relacionados con las matriculaciones de vehículos. A través de estos componentes se facilitará el análisis y la interpretación de la información.



Construcción del indicador Total Matriculaciones

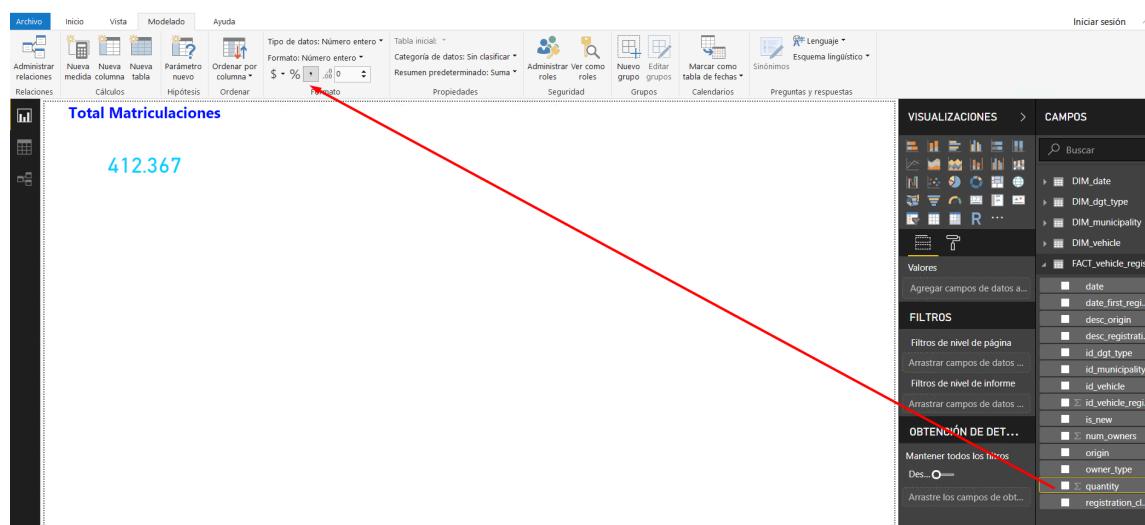
Tras renombrar la pestaña con el nombre *Matriculaciones*, se creará una tarjeta para mostrar el total de matriculaciones registradas.



Se seguirán los siguientes pasos para configurar la visualización:

- Seleccionar la medida *Quantity*, disponible en la tabla *FACT_vehicle_registration*.

- En el área de *Formato* de la visualización, desactivar la opción **Etiqueta de categoría**.
- Activar la opción **Título** y establecer como texto la etiqueta “*Total matriculaciones*”.
- Modificar el formato de la medida *Quantity*. Para ello, seleccionar la medida (sin marcar su casilla) y acceder a la pestaña **Modelado > Tipo de datos**, activando la opción **Separador de miles**.



- Finalmente, ajustar la tipografía y el color de fuente del título y del valor mostrado, dejando el componente preparado para su uso en el informe.

Total Matriculaciones

412.367

A continuación, se solicita la construcción del resto de componentes necesarios para completar la pestaña **Matriculaciones**:

- 1) Creación del indicador *Vehículos particulares*.
- 2) Creación del indicador *Vehículos de empresa*.
- 3) Creación del gráfico *Matriculaciones por Año*.
- 4) Creación del gráfico *Origen del Vehículo*.

5) Creación del gráfico *Distribución por Provincia*.

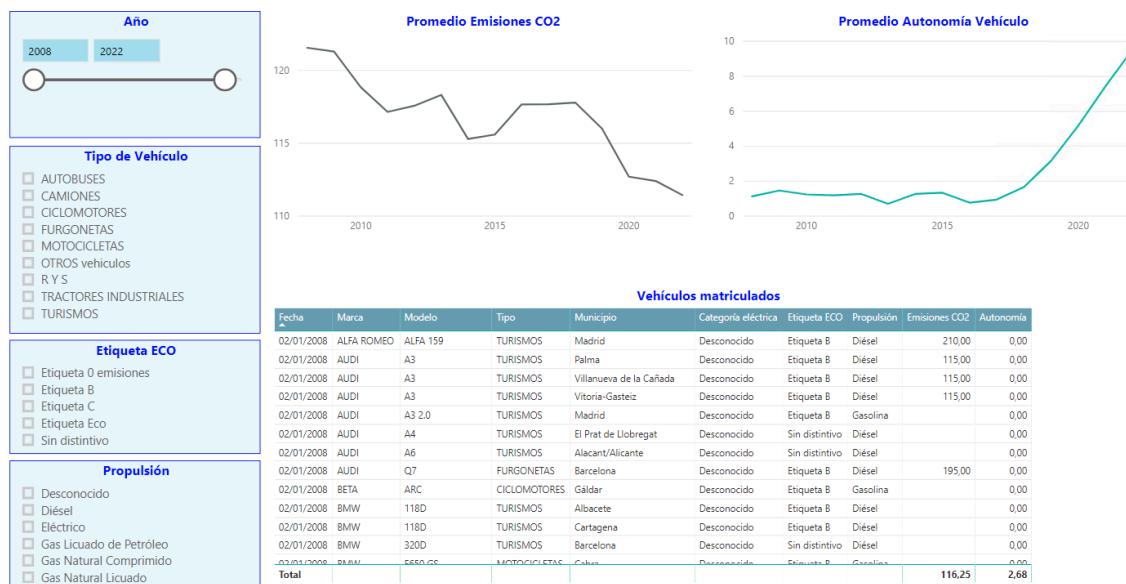
6) Creación del gráfico *Tipo de Vehículo*.

Una vez completada la construcción, se deberá responder a las siguientes preguntas:

- a) Observando el gráfico *Matriculaciones por año*, ¿qué interpretación puede darse a la notable disminución de matriculaciones a partir de 2008 y de nuevo a partir de 2019?
- b) Utilizando los componentes visuales creados en el dashboard, aplicar un filtro para el año **2023** de forma que todos los elementos visuales se actualicen automáticamente. Incluir una **captura de pantalla** con el filtro aplicado.

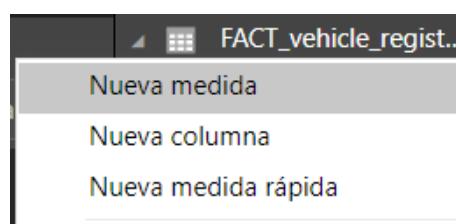
3.3. Construcción de la pestaña *Eficiencia Energética*

De manera análoga a la construcción de la pestaña *Matriculaciones*, se creará otra pestaña que permita analizar los datos de vehículos matriculados desde la perspectiva de *eficiencia energética*.



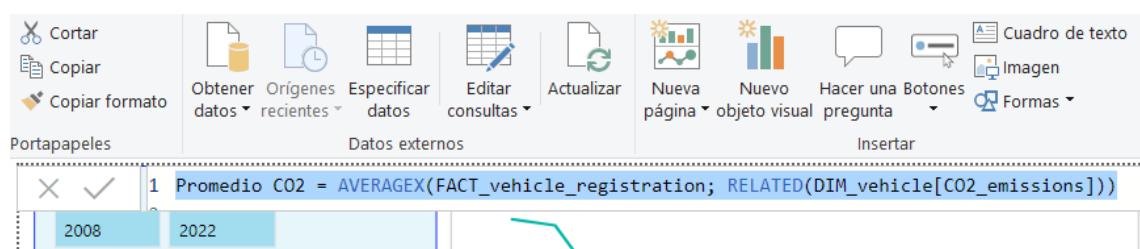
Esta pestaña requiere la creación de medidas para calcular el promedio de las emisiones de CO₂ y de la autonomía eléctrica del vehículo. A continuación, se detallan los pasos para crear la medida correspondiente al promedio de emisiones de CO₂:

- En el área de Campos, seleccionar la tabla *FACT_vehicle_registration*, hacer clic derecho sobre ella y elegir la opción **Nueva medida**.



- En la parte superior de la pantalla se habilitará un panel de fórmula. A continuación, introducir la siguiente expresión DAX:

```
Promedio CO2 = AVERAGEX(FACT_vehicle_registration;  
RELATED(DIM_vehicle[CO2_emissions]))
```



Para construir el dashboard, se requiere configurar los siguientes componentes:

- 1) De forma similar a la medida *Promedio CO2*, crear otra medida llamada *Promedio Autonomía* que realice el promedio sobre el atributo *range_km*.
- 2) Creación del área de filtros
 - a. Creación del filtro “Año”
 - b. Creación del filtro “Tipo de Vehículo”
 - c. Creación del filtro “Etiqueta ECO”

- d. Creación del filtro “Propulsión”
- 3) Creación del gráfico de líneas “Promedio Emisiones CO2”
- 4) Creación del gráfico de líneas “Promedio Autonomía Vehículo”
- 5) Creación de la tabla “Vehículos matriculados”

Una vez completada la construcción, se deberá responder a las siguientes preguntas:

- a) Observando los gráficos de líneas sobre el promedio de emisiones de CO2 y el promedio de autonomía de los vehículos, ¿qué interpretación puede darse al hecho de que ambas tendencias sean inversas?
- b) Utilizando los filtros del dashboard, investigar qué tipo de vehículos generan más emisiones de CO2.

En esta actividad se evaluará la configuración correcta de los orígenes de datos, la creación y ajuste del modelo de datos y la construcción de los dashboards solicitados en Power BI, incluyendo todos los componentes gráficos y filtros requeridos. Se valorará positivamente la aplicación de un formato visual coherente de los componentes similar al propuesto por el enunciado.

Será necesario mostrar evidencia mediante capturas de pantalla o similar tanto de la configuración de los orígenes, como del modelo de datos final y de los dashboards construidos. Además, se deberá aportar una breve explicación resumida de cómo se ha realizado cada una de estas configuraciones, sin necesidad de describir en detalle todos los pasos individuales, pero indicando en cada caso las medidas y visualizaciones utilizadas, los filtros aplicados y los resultados obtenidos.

En cuanto a las preguntas planteadas, se deberá responder adecuadamente, proporcionando una explicación clara y bien fundamentada.

Criterios de evaluación

Parte teórica (30%)

- Se detallan al inicio del cuestionario accesible desde el apartado *Contenidos del aula, Entrega cuestionario PR4.*

Parte práctica (70%)

- Se valorará mediante un documento en formato Word o PDF que muestre evidencias de cumplimiento de los pasos indicados y aportados por el estudiante al resolver los ejercicios solicitados.
- Los aspectos a tener en cuenta están descritos al final de cada ejercicio a realizar.
- En el aula está disponible la rúbrica de corrección.

Nota final:

La nota final se calculará de acuerdo con esta fórmula:

$$NF = \text{Nota Cuestionario (30\%)} + \text{Nota Ejercicios Prácticos (70\%)}$$

dónde:

$$\text{Nota Ejercicios Prácticos} = \text{Ejercicio 1 (10\%)} + \text{Ejercicio 2 (25\%)} + \text{Ejercicio 3 (35\%)}$$

Formato y fecha de entrega

La solución de la PR4 se entregará, a través del espacio “**Contenidos del aula**”, en dos partes:

1. **Entrega Cuestionario PR4**
2. **Entrega práctica PR4.** Envío del documento solución de la parte práctica en formato Word o PDF a través del espacio “**Contenidos**” del aula. El nombre del fichero a entregar debe seguir el patrón:
BDA_PR4_Apellido1_Apellido2_Nombre.extensión.

IMPORTANTE: Tal y como indica el plan docente:

- Es responsabilidad única del estudiante asegurarse que envía el cuestionario y entrega el documento con la solución en el espacio del aula habilitado para tal fin.
- Entregas realizadas fuera de los canales indicados se considerarán como No presentadas.
- Entregas pasadas las 23:59 h de la fecha límite no serán aceptadas y por tanto, no podrán ser evaluadas.

La fecha máxima de entrega es el **13/01/2026** a las 23:59 horas.