



**Universitat Oberta
de Catalunya**

Máster universitario de Ciencia de Datos

Práctica 3

**Diseño y uso de bases de datos analíticas – Explotación
de datos.**

Autor:

Mario Ubierna San Mamés

Índice de Contenido

Índice de Contenido	3
Índice de tablas	5
Índice de ilustraciones	6
1. Introducción	8
1.1. Presentación	8
1.2. Descripción	8
2. Creación del modelo OLAP	10
2.1. Creación de la estructura física del modelo	10
2.2. Creación del proyecto.....	10
2.3. Vistas del origen de datos	16
2.3.1. Vista Llamadas112.....	17
2.3.2. Vista Mediciones	19
2.4. Creación de cubos	22
2.4.1. Cubo Llamadas112	23
2.4.2. Cubo Mediciones.....	26
2.5. Jerarquías y dimensiones	33
2.5.1. DIM_Ambito_Geografico	33
2.5.2. DIM_Fecha	33
2.5.3. DIM_Grupo_Edad.....	34
2.5.4. DIM_Medicion.....	34
2.5.5. DIM_Tipologia	34
3. Implementación de la solución	35
4. Explotación de la solución	40

Índice de tablas

No se encuentran elementos de tabla de ilustraciones.

Índice de ilustraciones

Ilustración 1 - Base de datos DB_mariouism.	10
Ilustración 2 - Creación del proyecto multidimensional.	11
Ilustración 3 - Configuración origen de los datos.	12
Ilustración 4 - Propiedades del proyecto Visual Studio.	13
Ilustración 5 - Configuración destino de datos en Visual Studio.	14
Ilustración 6 - Ventana origen de datos.	15
Ilustración 7 - Modificación conexión de datos.	16
Ilustración 8 - Creación de nueva vista del origen de datos.	17
Ilustración 9 - Selección origen de datos vista.	17
Ilustración 10 - Selección tablas para la vista Llamadas112.	18
Ilustración 11 - Diseño vista vLlamadas112.	19
Ilustración 12 - Selección tablas para la vista Mediciones.	20
Ilustración 13 - Nombre de la vista vMediciones.	21
Ilustración 14 - Diseño vista vMediciones.	22
Ilustración 15 - Creación de un cubo.	23
Ilustración 16 - Uso de tablas existentes en el cubo.	23
Ilustración 17 - Selección vista del origen de datos cLlamadas112.	24
Ilustración 18 - Selección métricas de cLlamadas112.	24
Ilustración 19 - Selección dimensiones de cLlamadas112.	25
Ilustración 20 - Resumen de cLlamadas112.	25
Ilustración 21 - Ventana diseño cLlamadas112.	26
Ilustración 22 - Selección vista del origen de datos cMediciones.	27
Ilustración 23 - Selección métricas de cMediciones.	27
Ilustración 24 - Selección dimensiones de cMediciones.	28
Ilustración 25 - Resumen de cMediciones.	28

Ilustración 26 - Ventana diseño cMediciones.....	29
Ilustración 27 - Eliminación de DIM_Ambito_Geografico_1	29
Ilustración 28 - Eliminación de DIM_Fecha_1.	30
Ilustración 29 - Agregar dimensiones a cMediciones.	30
Ilustración 30 - Selección de dimensiones de cMediciones.	31
Ilustración 31 - Ventana diseño cMediciones.....	31
Ilustración 32 - Relación de DIM_Ambito_Geografico.	32
Ilustración 33 - Relación de DIM_Fecha.	32
Ilustración 34 - Relaciones dimensiones cMediciones.	33
Ilustración 35 - DIM_Ambito_Geografico.....	33
Ilustración 36 - DIM_Fecha.	34
Ilustración 37 - DIM_Grupo_Edad.	34
Ilustración 38 - DIM_Medicion.	34
Ilustración 39 - DIM_Tipologia.....	34
Ilustración 40 - Implementar la solución.	35
Ilustración 41 - Implementación de la solución.....	36
Ilustración 42 - Creación de la dimensión temporal.....	37
Ilustración 43 - Asignación del tipo para el día.....	37
Ilustración 44 - Asignación del tipo para el mes.	38
Ilustración 45 - Asignación del tipo para el año.....	38
Ilustración 46 - Asignación del tipo para la fecha.....	39

1.Introducción

1.1. Presentación

A partir de la solución oficial de la segunda práctica (PRA2) el estudiante debe implementar los cubos multidimensionales necesarios para la explotación de la información y el posterior análisis de datos. De este modo se facilitará la toma de decisiones de los usuarios potenciales.

Así pues, esta actividad tiene el objetivo de implementar un modelo multidimensional online analytical processing (OLAP) para el análisis multidimensional de la información con el fin de responder a las preguntas definidas en el análisis de requerimientos.

Adicionalmente, se facilitará junto a este enunciado el fichero «*export_DW_COVID.sql*», que contiene los scripts de generación y carga de todas las tablas planteadas en la solución, para que el estudiantado pueda partir de la misma base.

1.2. Descripción

Más concretamente esta tercera parte del caso práctico consiste en diseñar un modelo OLAP para el análisis multidimensional de la información disponible en el almacén de datos que permita dar respuesta a las siguientes cuestiones:

- Análisis de las provincias con mayor porcentaje de movilidad según datos móviles.
- Análisis del porcentaje de la población que evitaba las aglomeraciones según la comunidad autónoma.
- Análisis del promedio de sanciones por habitante.
- Evolución de las llamadas de urgencia al 112 en Cataluña por tipología de llamada.

- Análisis de las llamadas de urgencia frente al porcentaje de la población que evitaba las aglomeraciones entre los meses comprendidos entre marzo y junio de 2020 en Cataluña, desglosado por provincia.
- Determinación del día de la semana con menor número de denuncias.
- Análisis de las diez fechas (*top ten*) con mayor número de llamadas de urgencia al 112 con tipología de tránsito registrada, tanto en época de *COVID* como antes.

2. Creación del modelo OLAP

En este apartado vamos a crear tanto el proyecto en *Visual Studio*, como observar el origen de datos, la creación de los cubos y la decisión sobre jerarquías/dimensiones/atributos relacionados.

2.1. Creación de la estructura física del modelo

Lo primero que debemos de hacer es ejecutar el *script* que se nos ha proporcionado “*export_DW_COVID.sql*”, ya que haciendo esto obtenemos el diseño tanto de los hechos como de las dimensiones junto a sus datos.

Cabe destacar que para ejecutar el *script* lo hemos tenido que modificar añadiendo las dos primeras líneas siguientes, para que así haga uso de la base de datos correcta:

```
USE [DB_marioum]  
GO
```

Ilustración 1 - Base de datos DB_marioum.

2.2. Creación del proyecto

Una vez que tenemos todos los datos que vamos a hacer uso de ellos para explotarlos, podemos crear el proyecto en *Visual Studio*.

Para ello abrimos el programa y creamos un nuevo proyecto de tipo “*Analysis Services Multidimensional*”:

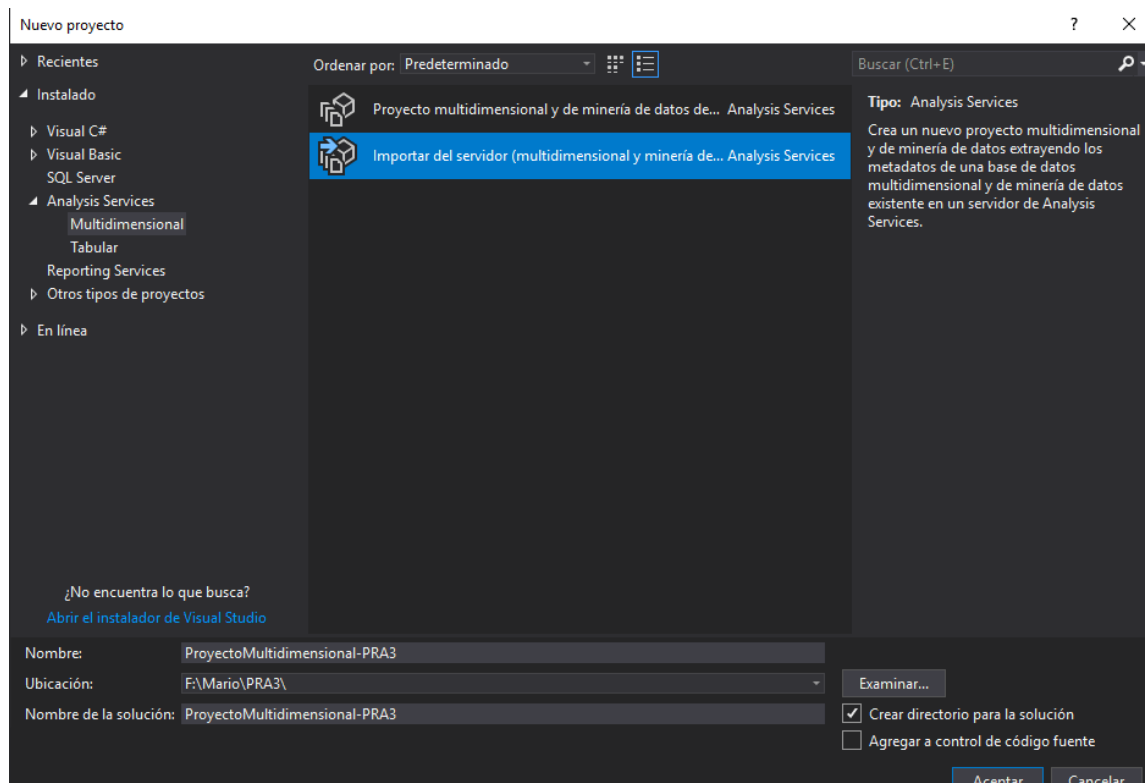
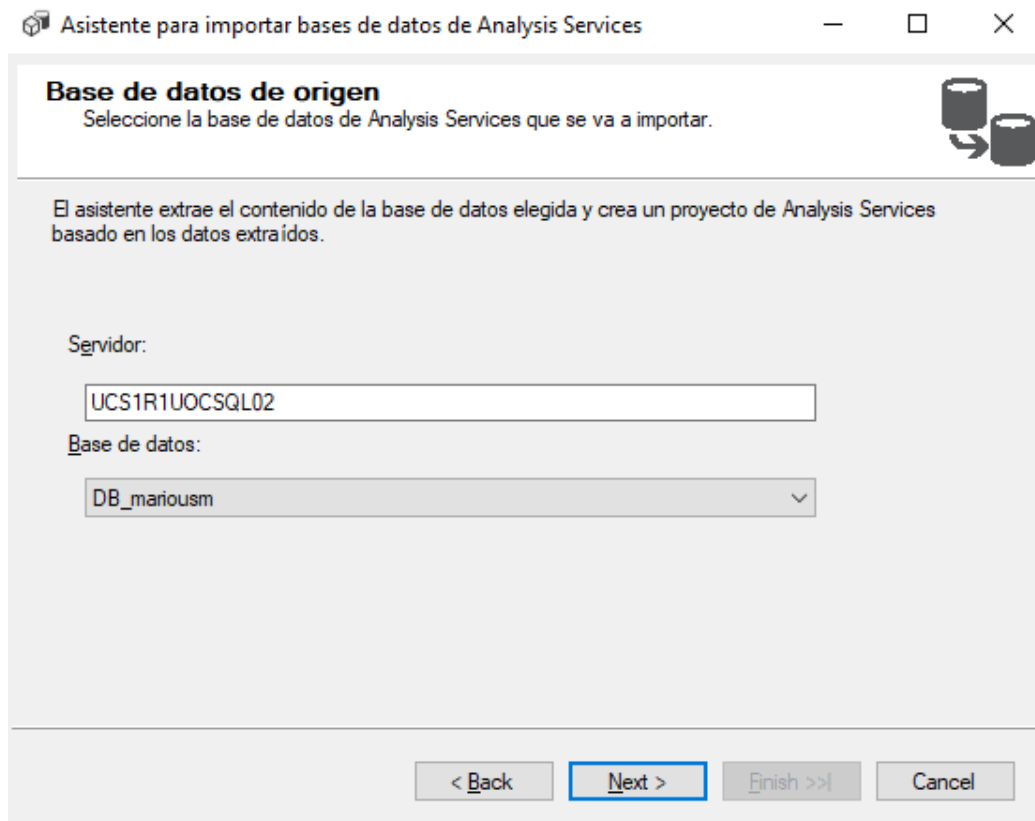


Ilustración 2 - Creación del proyecto multidimensional.

Una vez que hemos seleccionado la opción “Importar del servidor” y establecemos el nombre del proyecto junto con su ubicación, nos aparecerá un asistente para configurar la base de datos a usar.

Para ello indicamos el servidor y en el menú desplegable de la base de datos, seleccionamos la nuestra, es decir, “DB_mariouism”:



Asistente para importar bases de datos de Analysis Services

Base de datos de origen
Seleccione la base de datos de Analysis Services que se va a importar.

El asistente extrae el contenido de la base de datos elegida y crea un proyecto de Analysis Services basado en los datos extraídos.

Servidor:
UCS1R1UOCSQL02

Base de datos:
DB_mariouism

< Back **Next >** Finish >> Cancel

Ilustración 3 - Configuración origen de los datos.

Una vez importada la estructura de la base de datos a nuestro proyecto, hay que configurar el destino de los datos. Para ello, nos dirigimos a las propiedades del proyecto:

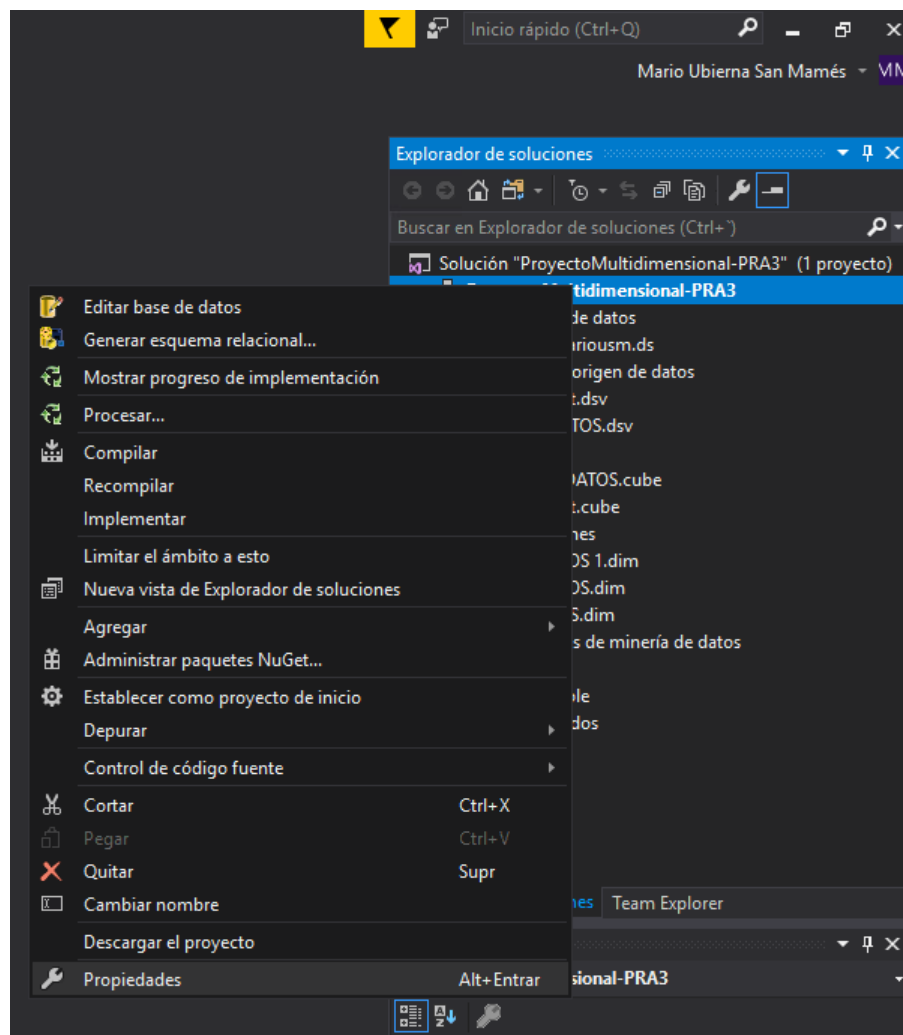


Ilustración 4 - Propiedades del proyecto Visual Studio.

En las propiedades nos vamos al apartado implementación, y en él cambiamos el nombre del servidor tal y como se muestra en la siguiente captura:

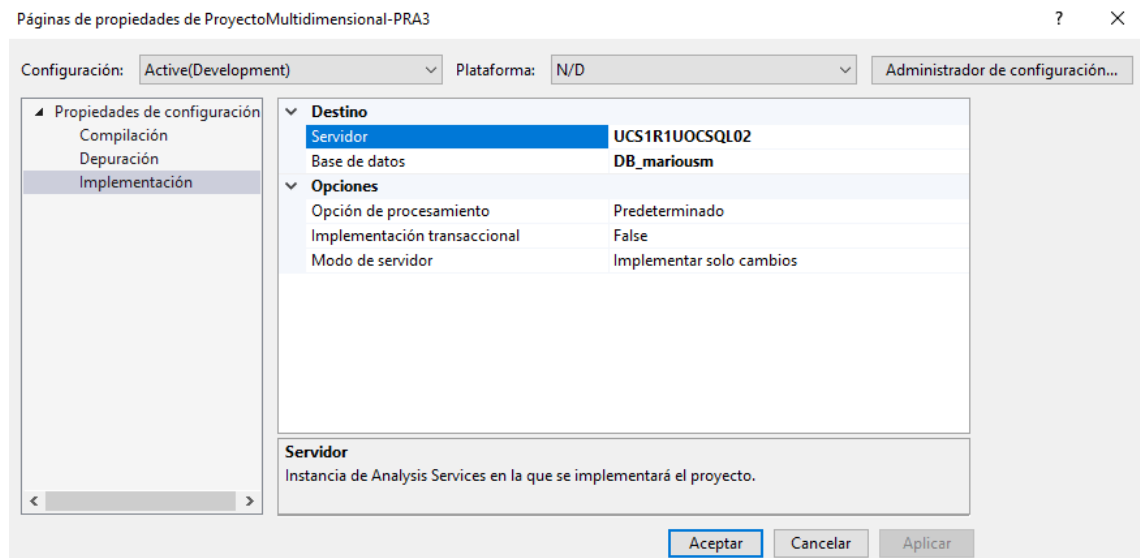


Ilustración 5 - Configuración destino de datos en Visual Studio.

Finalmente hay que configurar el origen de los datos, para ello sobre el nombre de la base de datos en el explorador de soluciones, hacemos doble click y nos mostrará la siguiente ventana:

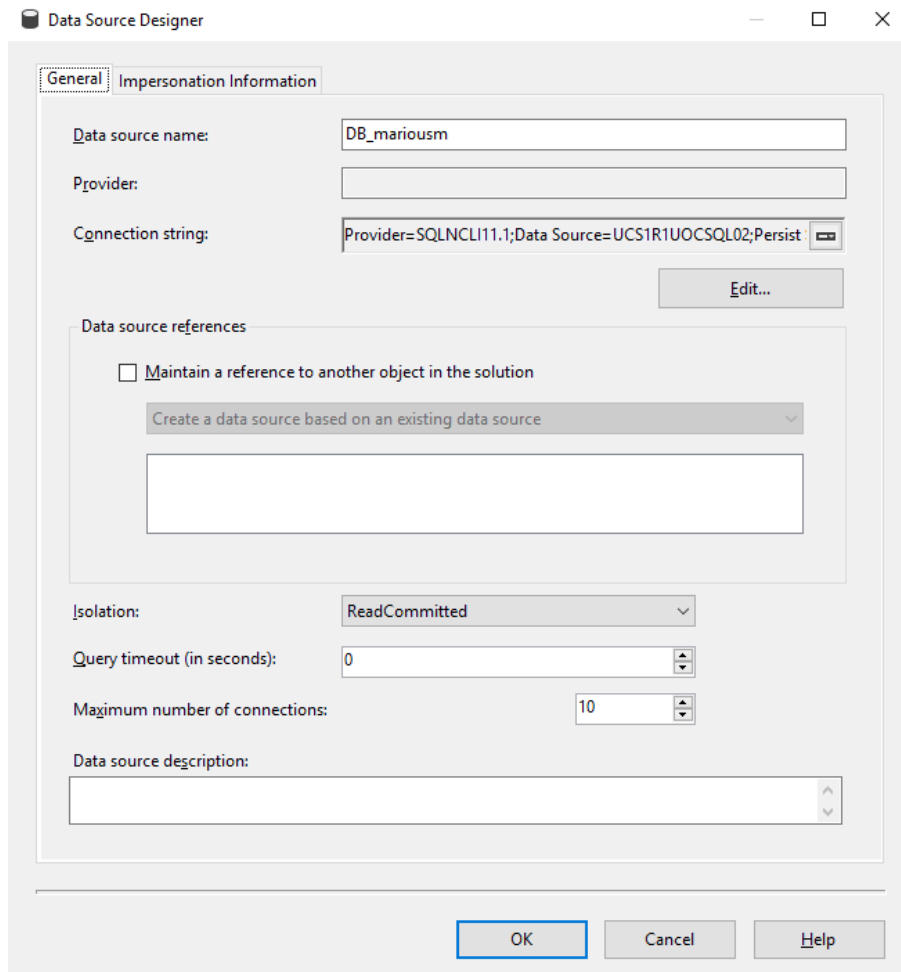


Ilustración 6 - Ventana origen de datos.

Pulsamos sobre el botón “Edit” para cambiar la conexión a la base de datos, una vez hemos pulsado veremos la siguiente pantalla:

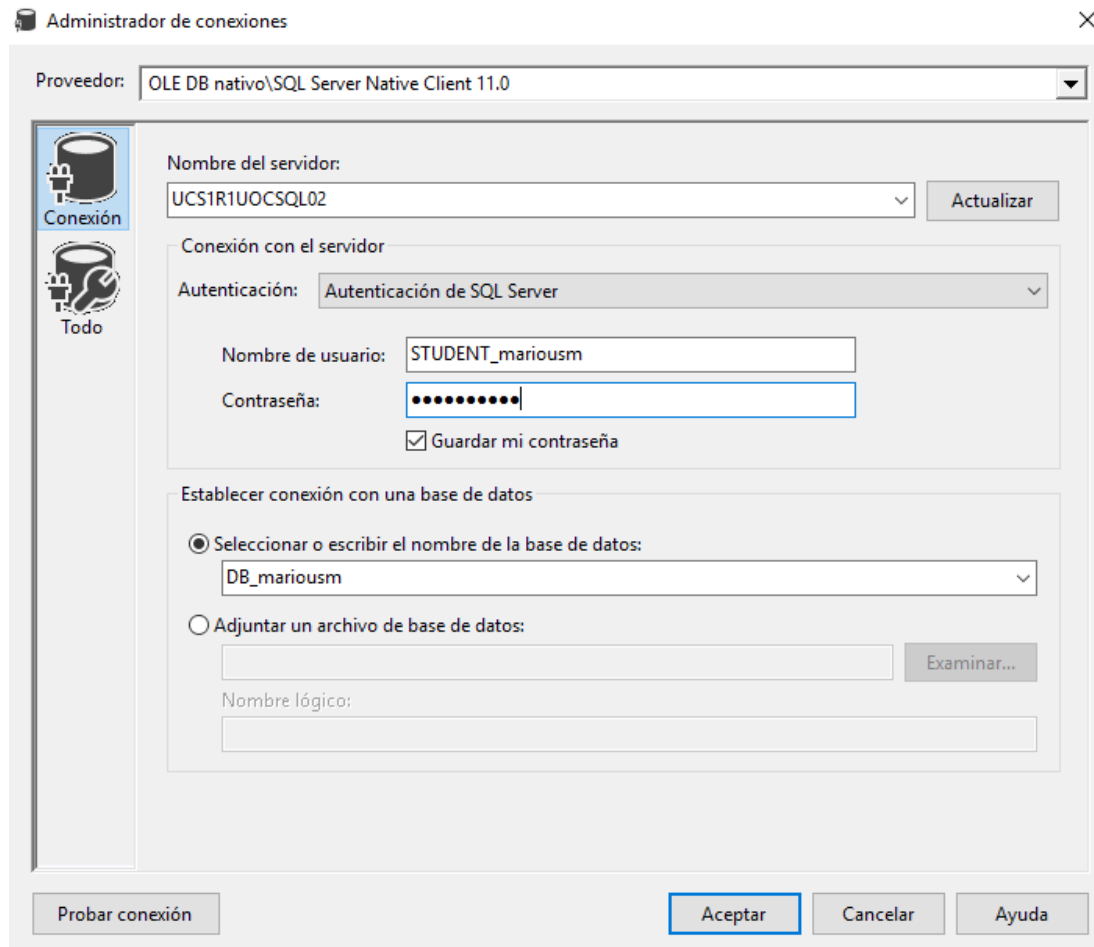


Ilustración 7 - Modificación conexión de datos.

Introducimos el servidor que se nos ha proporcionado, el nombre del usuario y la contraseña que usamos para acceder a *Microsoft SQL Server Management Studio 2017*.

Una vez realizados todos estos pasos podemos dar por concluido este apartado.

2.3. Vistas del origen de datos

Una vez que ya tenemos configurado nuestro proyecto, podemos comenzar a definir las diferentes vistas respecto al origen de los datos.

En nuestro caso vamos a tener dos:

- Vista de Llamadas112: nos va a permitir realizar el acceso a la tabla correspondiente de las llamadas y así hacer un análisis de las mismas.
- Vista de Mediciones: en ella vamos a poder realizar el acceso a la tabla de métricas y hacer un profundo análisis de las mismas.

Para crear las dos vistas debemos acceder al explorador de soluciones del proyecto y hacer click derecho sobre la carpeta “Vistas del origen de datos”, posteriormente pulsaremos sobre la opción “Nueva vista del origen de datos”:

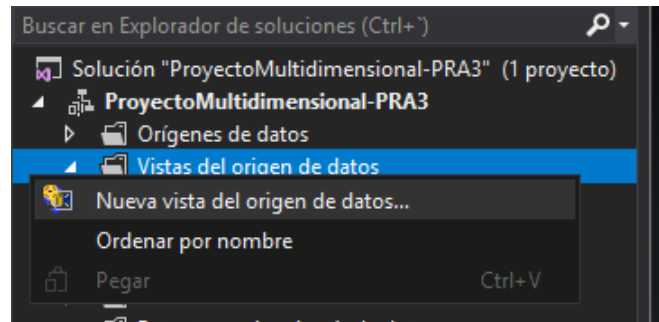


Ilustración 8 - Creación de nueva vista del origen de datos.

Una vez que hemos accedido al asistente para la creación de la vista, nos aparecerá la siguiente ventana en la que tenemos que seleccionar el origen de datos:

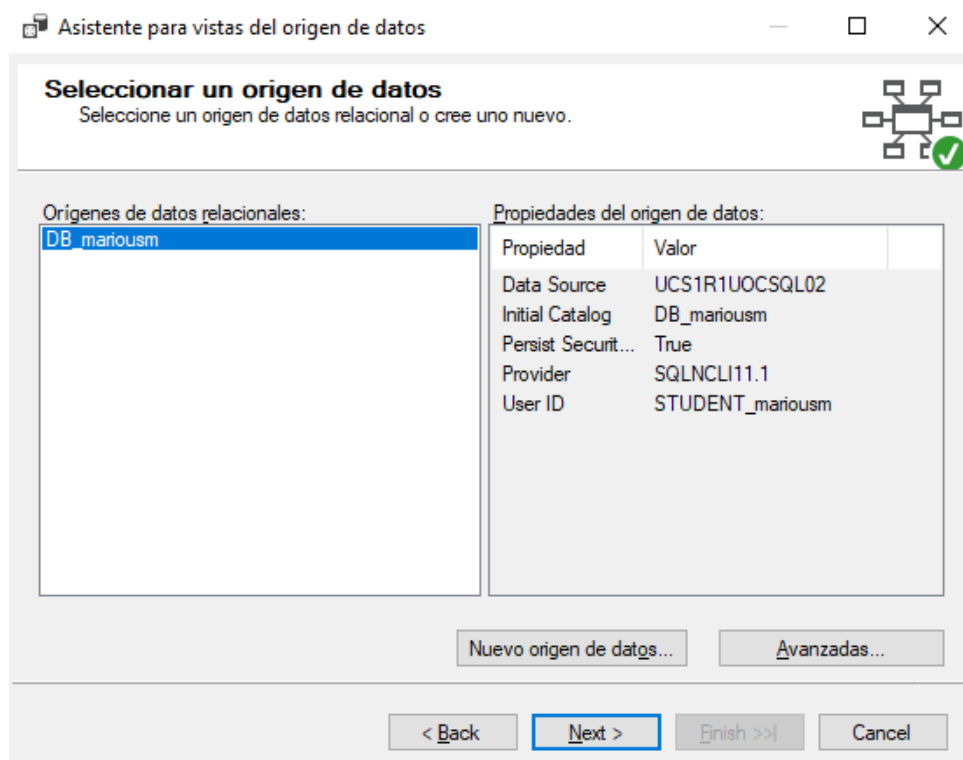


Ilustración 9 - Selección origen de datos vista.

2.3.1. Vista Llamadas112

Posteriormente seleccionamos tanto la tabla del hecho como de las dimensiones correspondientes, en nuestro caso el hecho es “FACT_Llamadas112” y las dimensiones: “DIM_Ambito_Geografico”, “DIM_Fecha”, “DIM_Tipologia”:

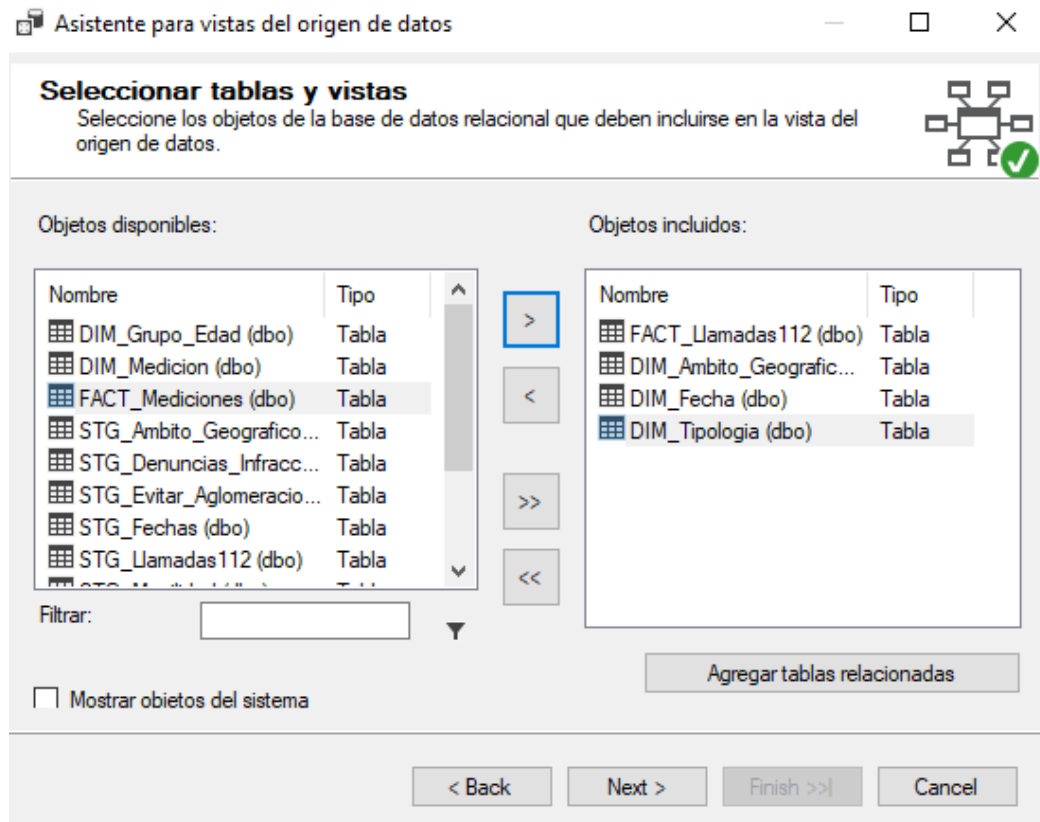


Ilustración 10 - Selección de tablas para la vista Llamadas112.

Pulsamos el botón “Next”, indicamos el nombre de la vista “vLlamadas112” y finalizamos el asistente.

Al hacer doble click sobre la vista nos muestra el diseño de la misma:

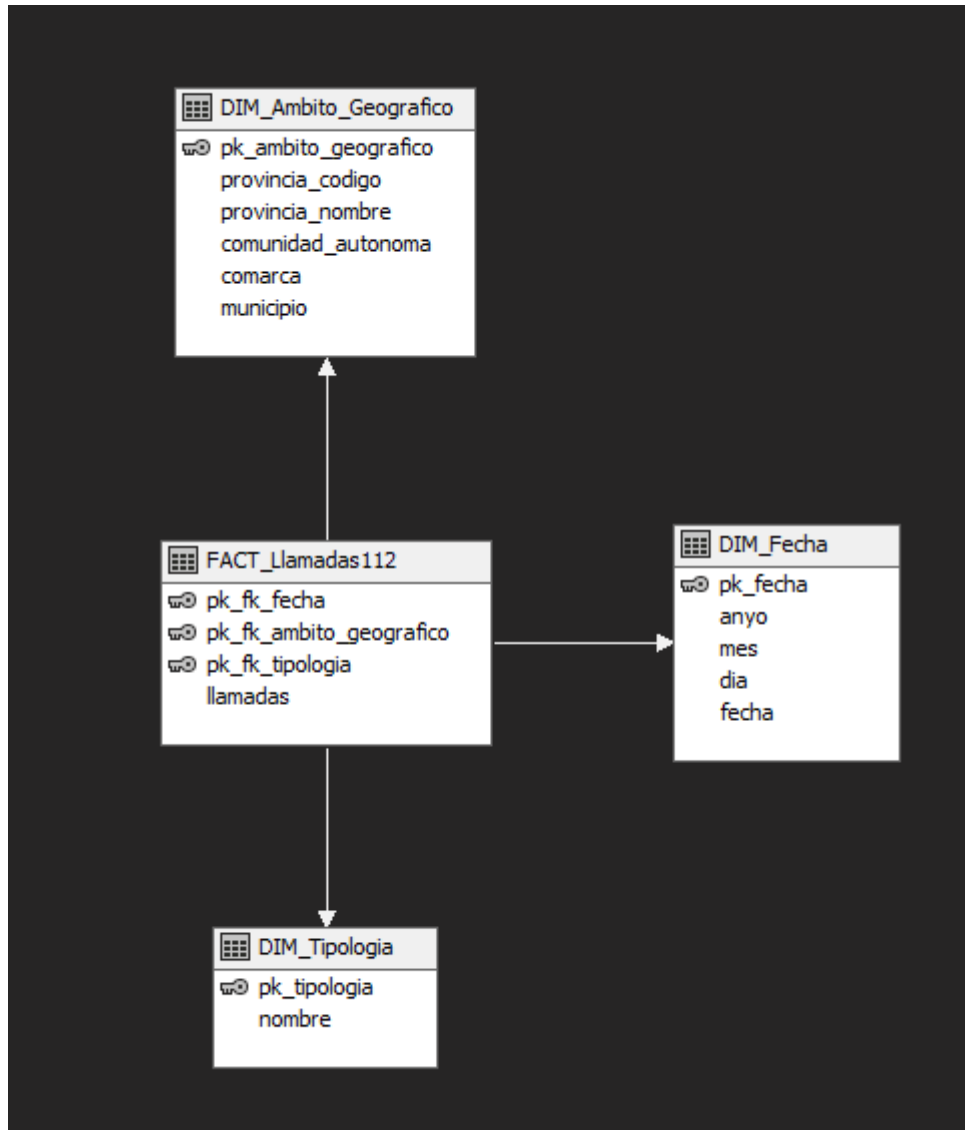


Ilustración 11 - Diseño vista vLlamadas112.

2.3.2. Vista Mediciones

A continuación, vamos a realizar el mismo proceso que el detallado en la creación de la vista anterior, pero en este caso vamos a crear una vista con el fin de explotar la información sobre las denuncias, el porcentaje de población que evitaba aglomeraciones...

Esta vista tiene en cuenta el hecho “`FACT_Mediciones`” y las dimensiones: “`DIM_Ambito_Geografico`”, “`DIM_Fecha`”, “`DIM_Grupo_Edad`” y “`DIM_Medicion`”:

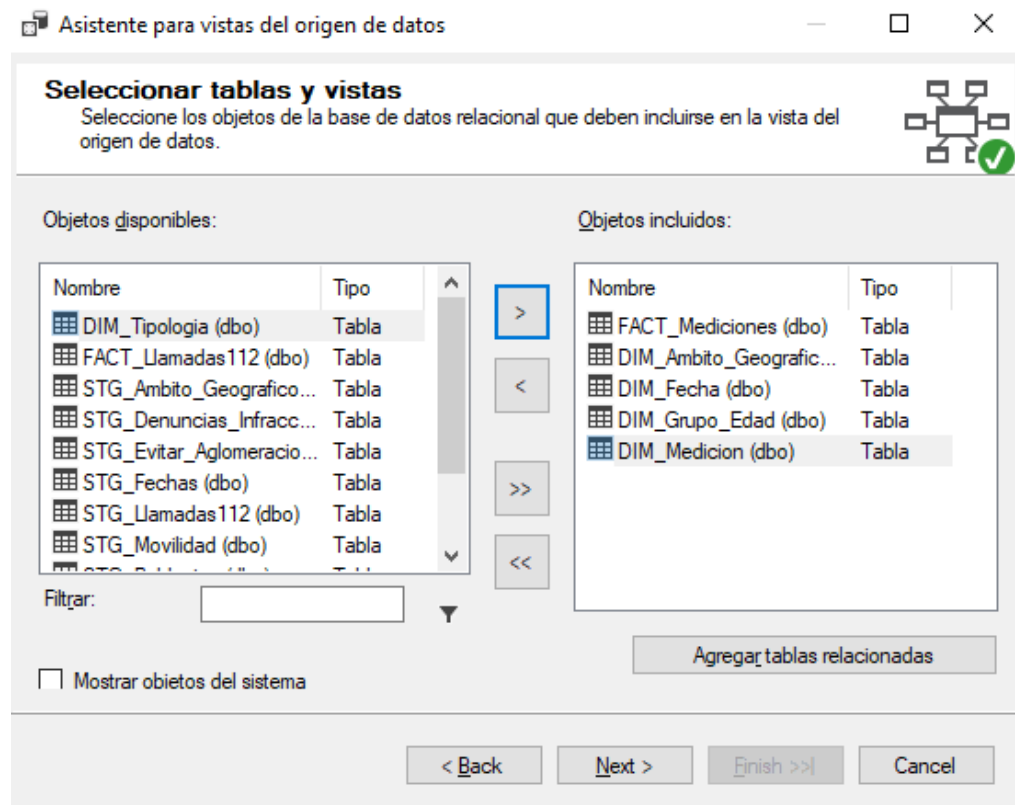


Ilustración 12 - Selección de tablas para la vista Mediciones.

Pulsamos el botón “Next”, indicamos el nombre de la vista “vMediciones” y finalizamos el asistente:

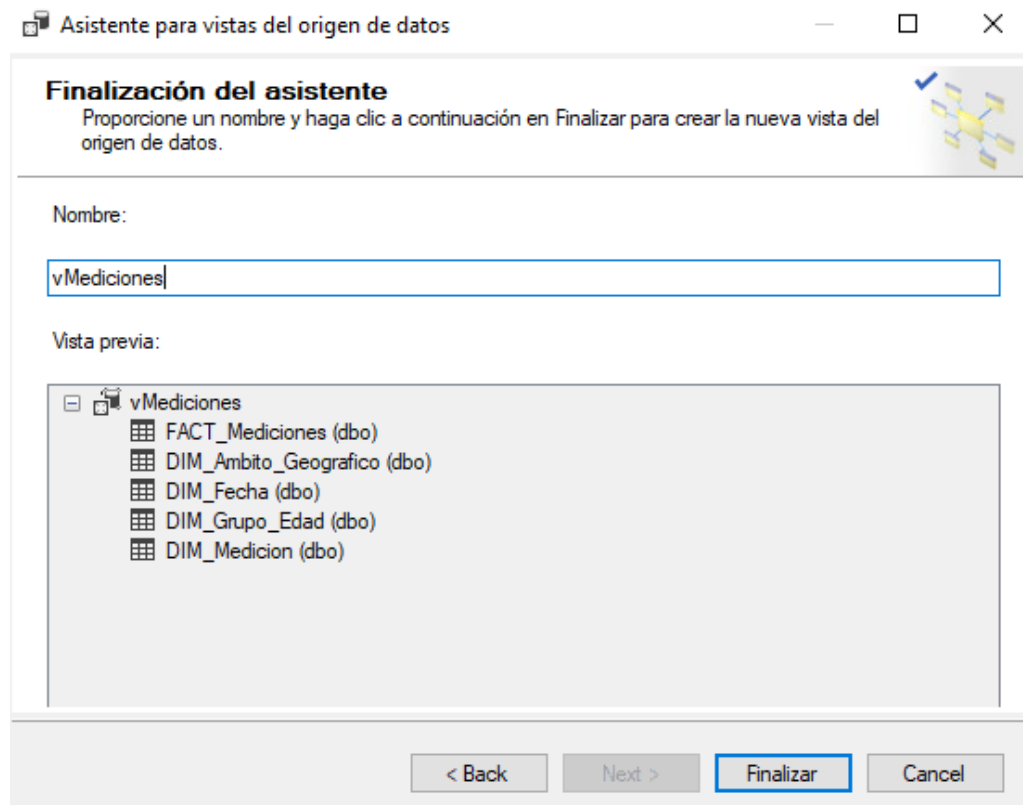


Ilustración 13 - Nombre de la vista vMediciones.

Al hacer doble click sobre la vista nos muestra el diseño de la misma, comprobamos que todo está de forma correcta:

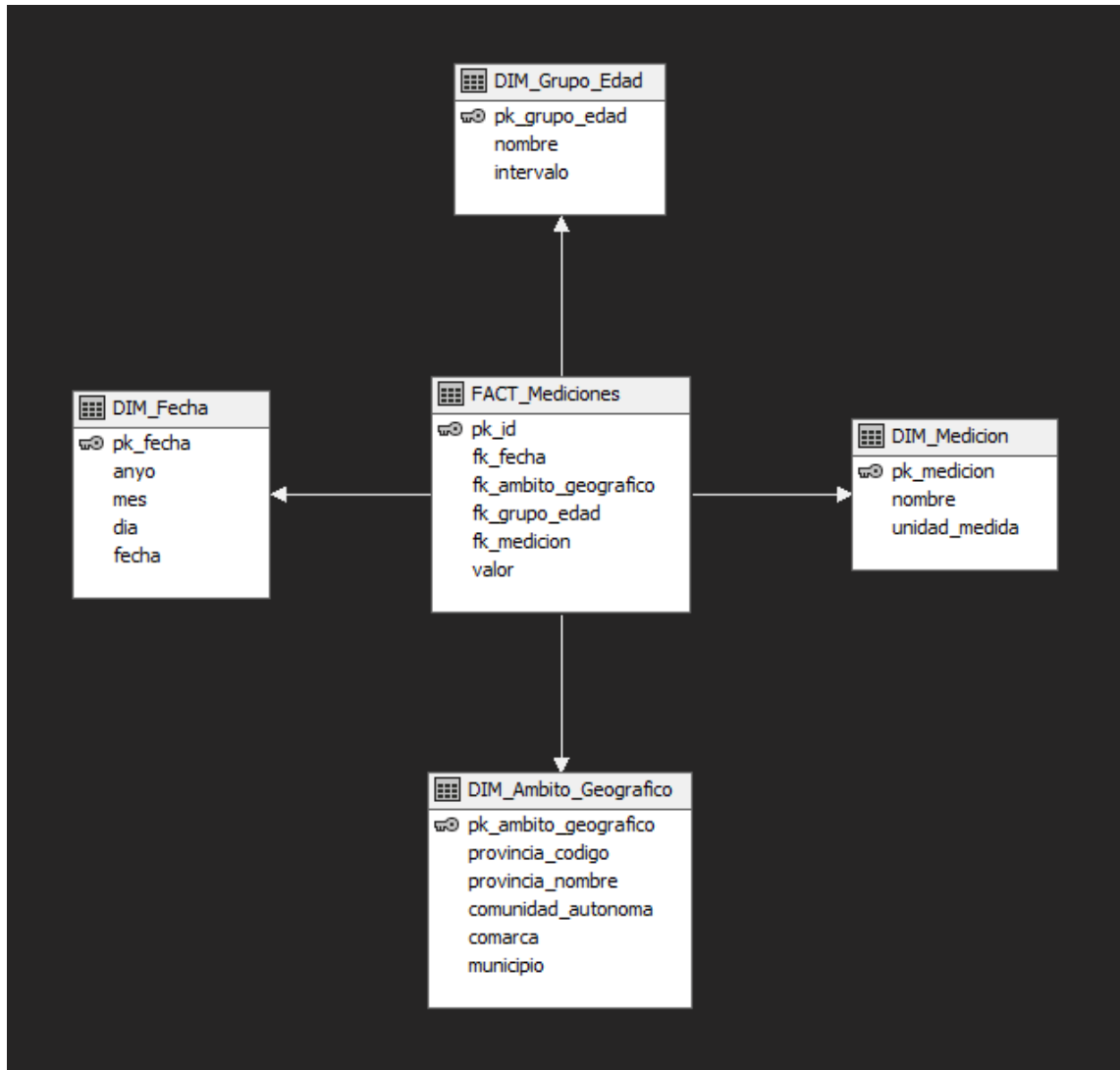


Ilustración 14 - Diseño vista vMediciones.

2.4. Creación de cubos

Una vez creadas las diferentes vistas para así poder realizar el análisis de los datos, tenemos que crear cada uno de los cubos correspondientes.

Para crear un cubo nos debemos dirigir al explorador de soluciones, y sobre la carpeta “Cubos” pulsamos botón derecho, posteriormente seleccionamos la opción “Nuevo cubo”:

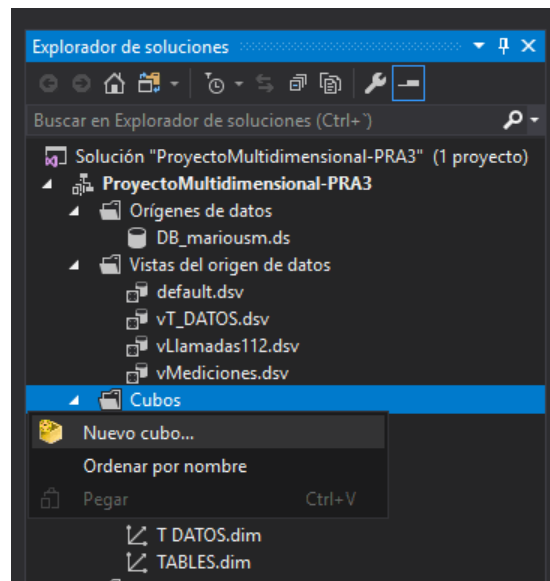


Ilustración 15 - Creación de un cubo.

Una vez seleccionada la opción “Nuevo cubo” nos aparecerá el siguiente asistente, donde debemos indicar que vamos a hacer uso de tablas existentes:

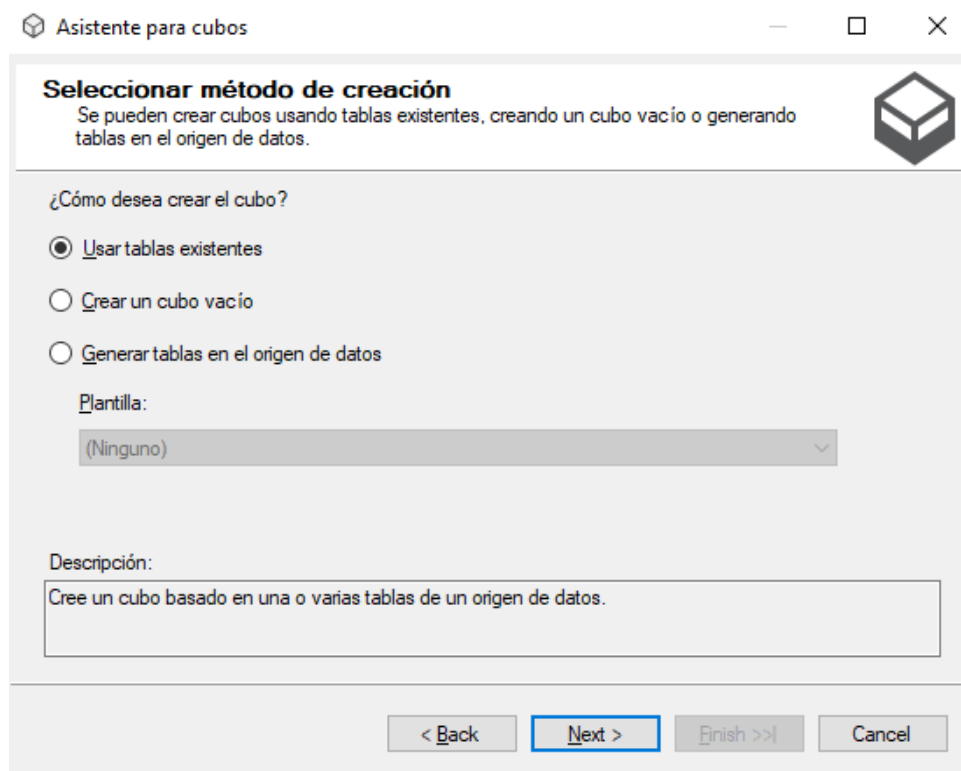


Ilustración 16 - Uso de tablas existentes en el cubo.

2.4.1. Cubo Llamadas112

Una vez que hemos realizado los pasos anteriores, indicamos que queremos la vista correspondiente a las llamadas al 112, es decir, “vLlamadas112”:

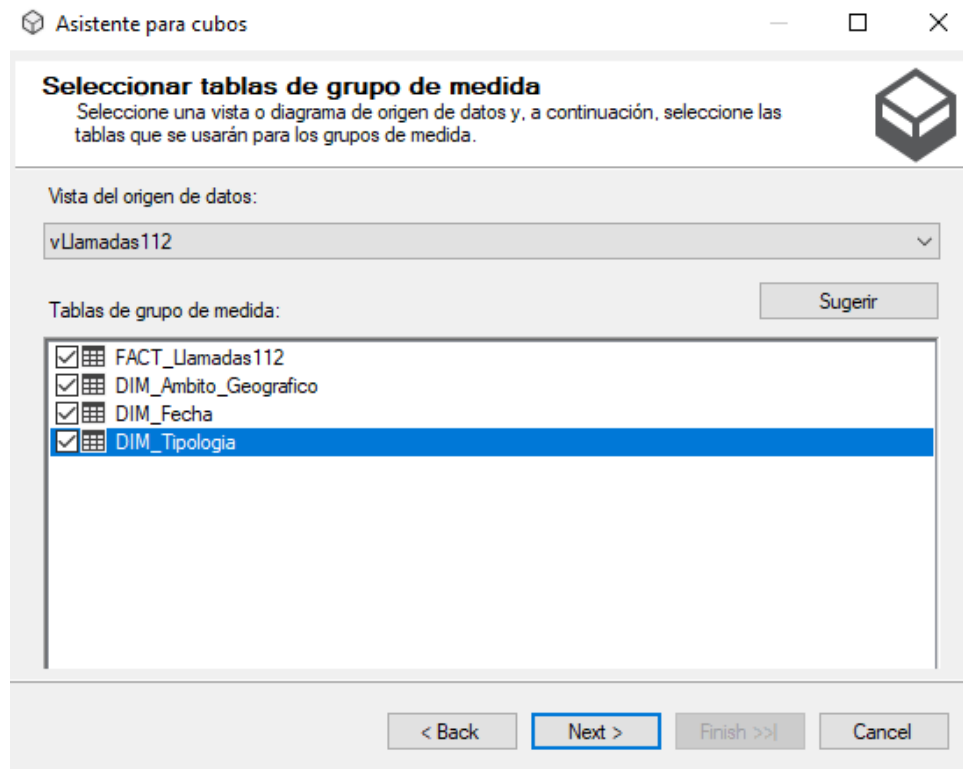


Ilustración 17 - Selección vista del origen de datos cLlamadas112.

Posteriormente, seleccionamos las métricas:

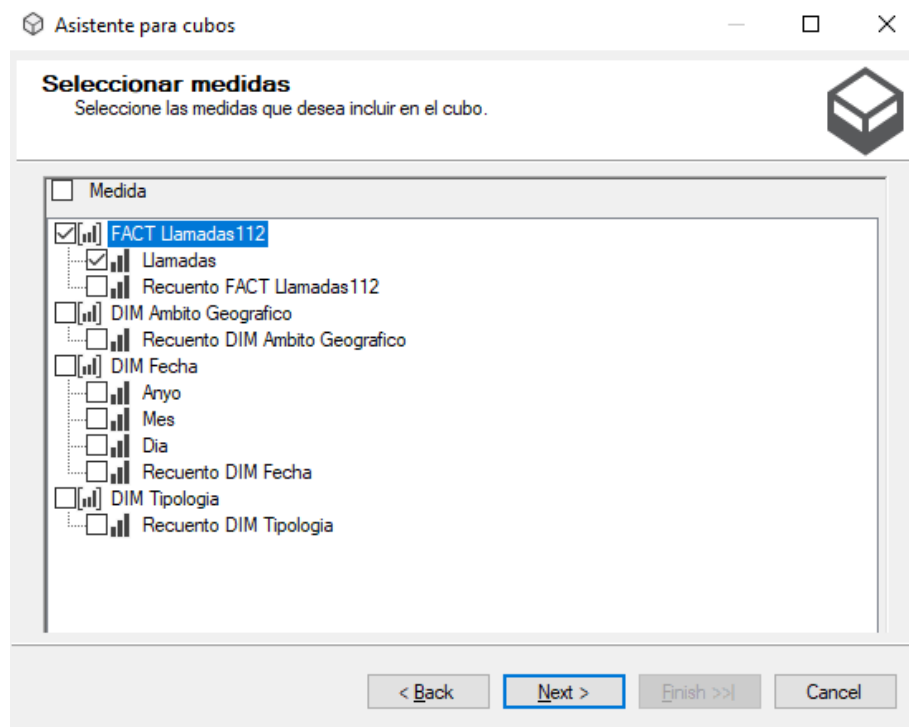


Ilustración 18 - Selección métricas de cLlamadas112.

Por último indicamos las dimensiones del cubo:

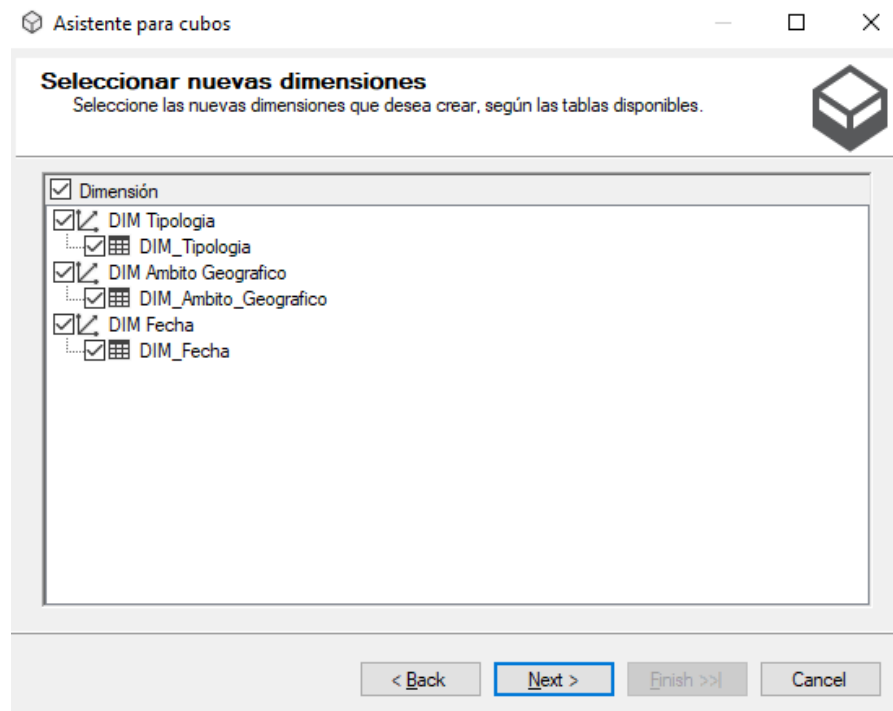


Ilustración 19 - Selección dimensiones de cLlamadas112.

Al terminar el asistente obtenemos el siguiente resumen, también indicamos el nombre del cubo:

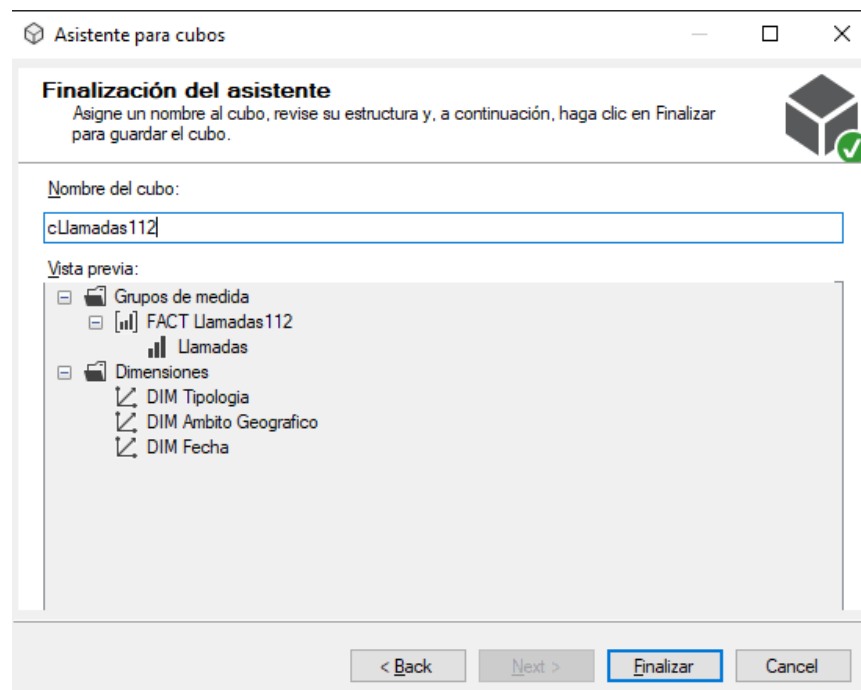


Ilustración 20 - Resumen de cLlamadas112.

Finalmente, cuando hemos creado el cubo tenemos la siguiente ventana de diseño de “cLlamadas112”, a la izquierda observamos las métricas y dimensiones, y a la derecha la vista del origen de datos:

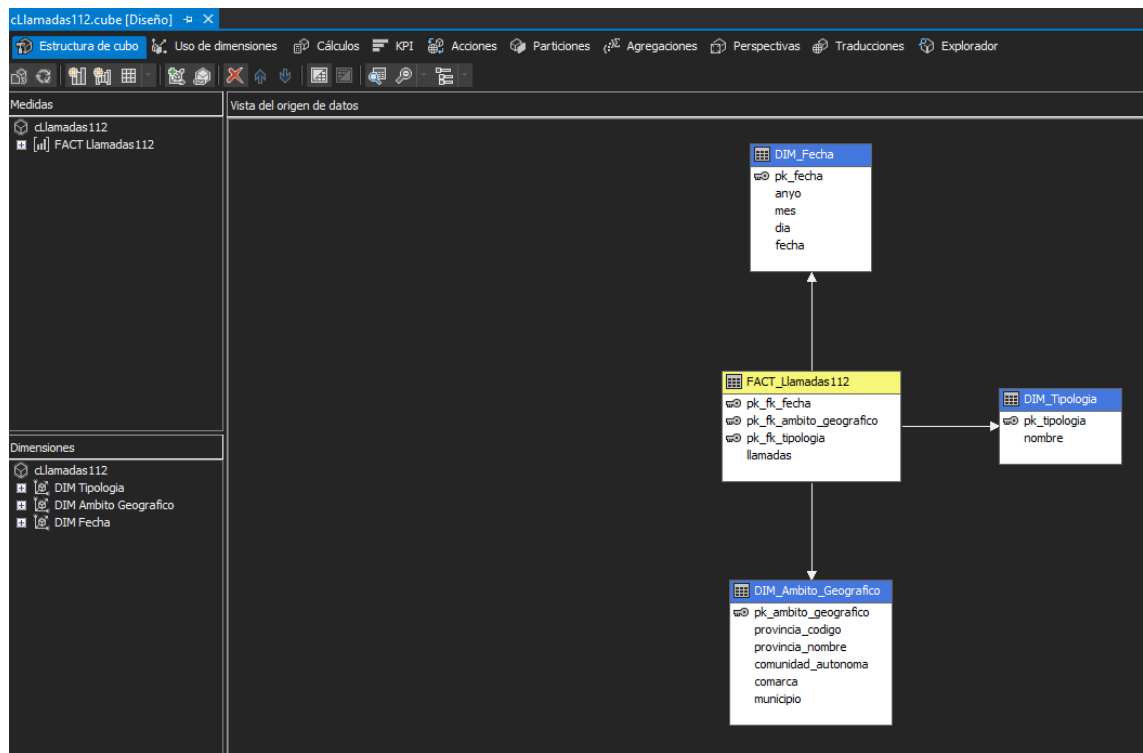


Ilustración 21 - Ventana diseño cLlamadas112.

2.4.2. Cubo Mediciones

Para crear el cubo de mediciones seguimos los mismos pasos que en el cubo anterior. Una vez que llegamos al asistente para crear el cubo, indicamos que la vista que queremos es la de “vMediciones”:

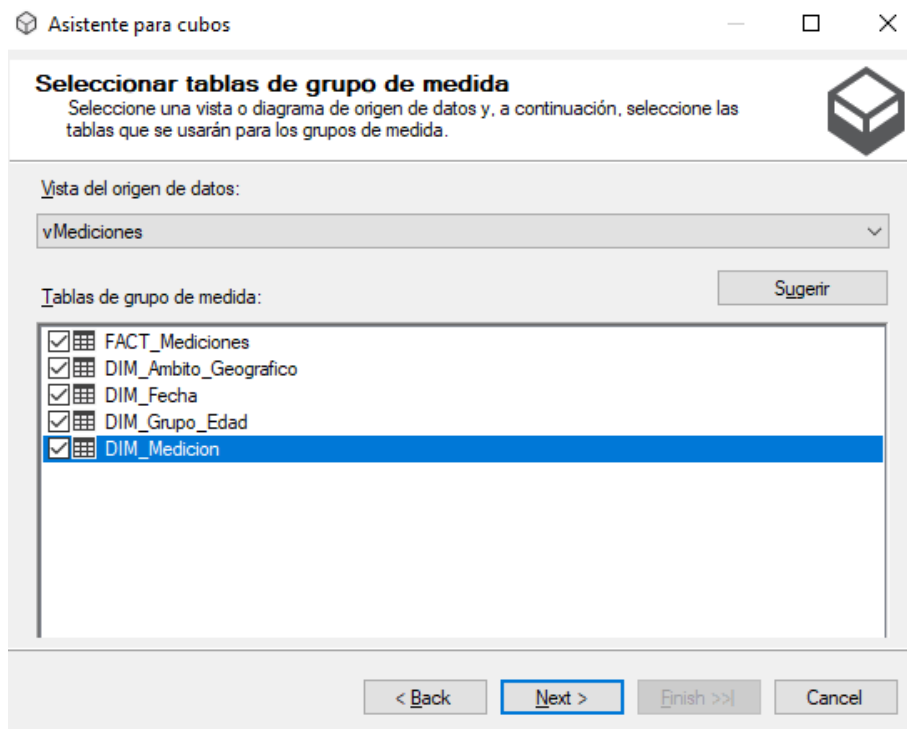


Ilustración 22 - Selección vista del origen de datos cMediciones.

Posteriormente, seleccionamos las métricas:

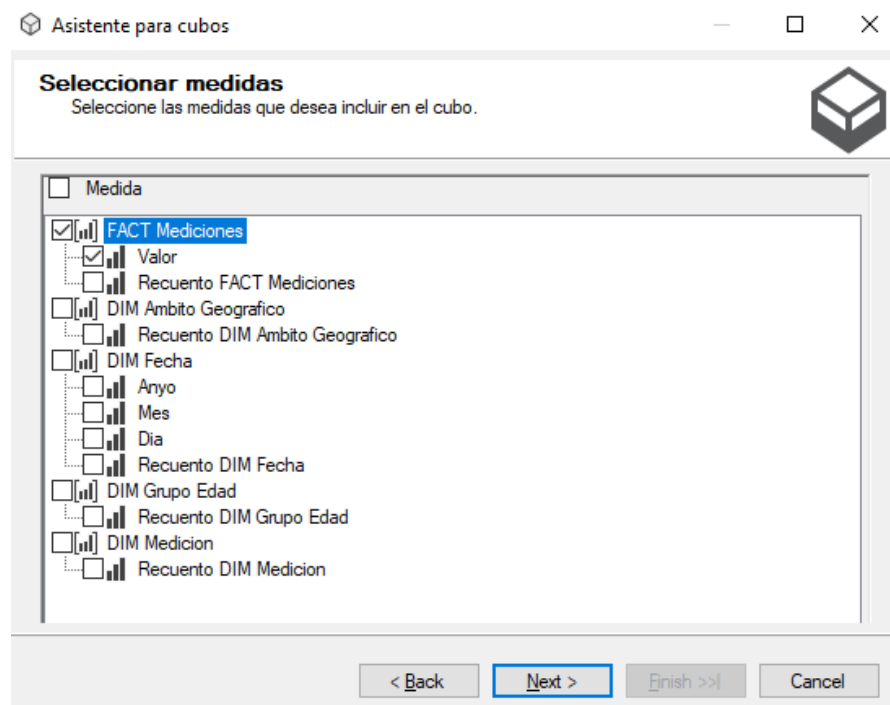


Ilustración 23 - Selección métricas de cMediciones.

Por último, indicamos las dimensiones del cubo:

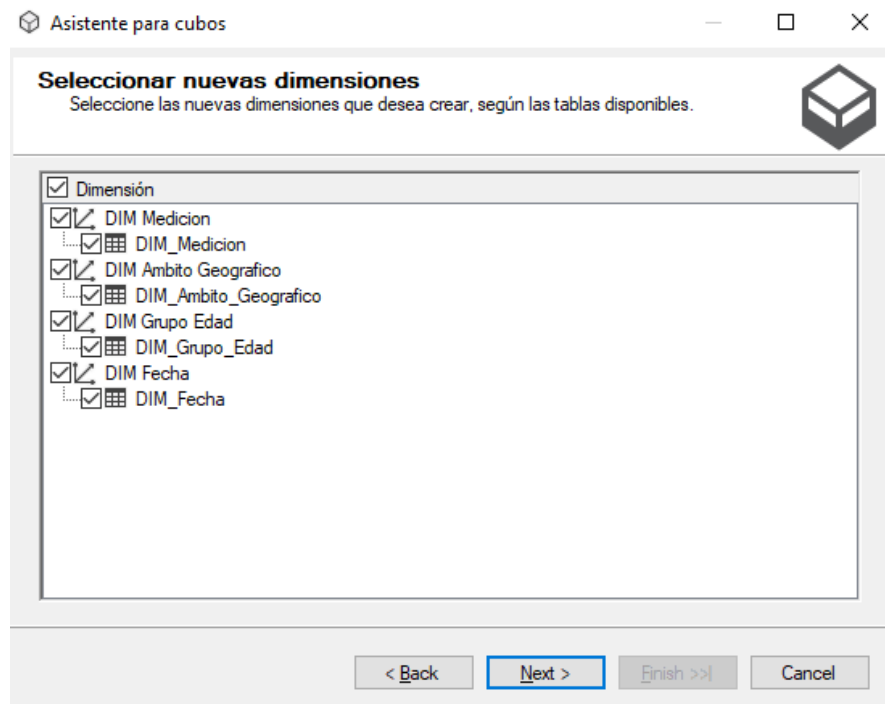


Ilustración 24 - Selección dimensiones de cMediciones.

Al terminar el asistente obtenemos el siguiente resumen, también indicamos el nombre del cubo:

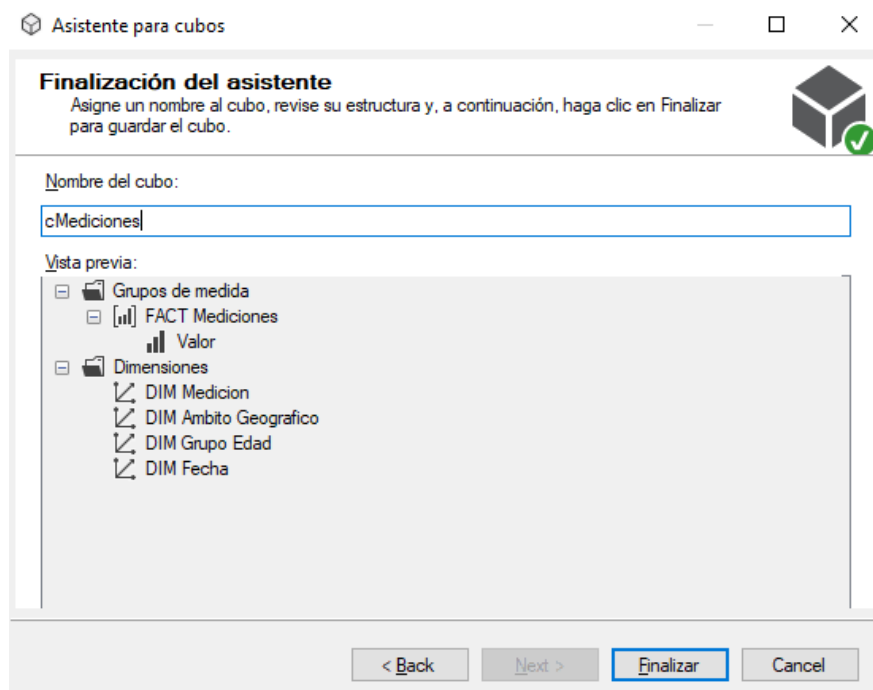


Ilustración 25 - Resumen de cMediciones.

Finalmente, cuando creamos el cubo nos salta la ventana de diseño del mismo, a la izquierda observamos las métricas y las dimensiones, y a la derecha la vista del origen de datos:

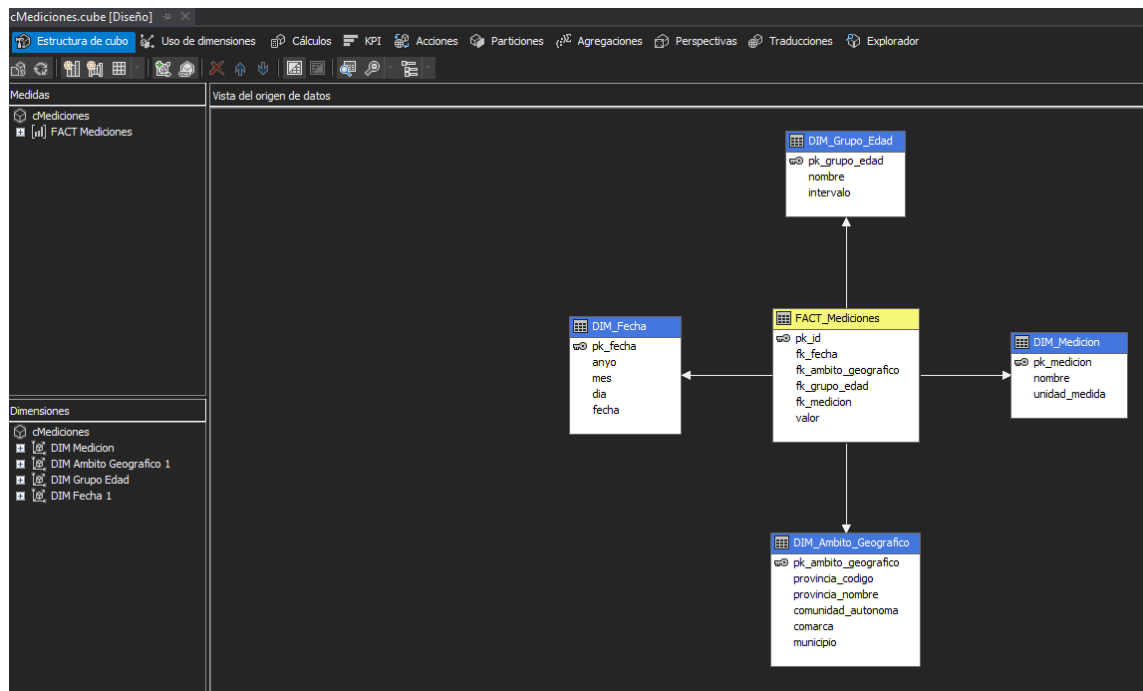


Ilustración 26 - Ventana diseño cMediciones.

Tal y como podemos observar en la anterior ilustración, para las dimensiones comunes “DIM_Ambito_Geografico” y “DIM_Fecha”, nos ha vuelto a crear la misma dimensión, es decir, tenemos duplicadas ambas dimensiones. Esto no es correcto, ya que tenemos que usar las mismas dimensiones en todos los cubos.

Por lo tanto, eliminamos dichas dimensiones que ha creado nuevas, para ello en el explorador de soluciones nos dirigimos a la carpeta “Dimensiones” y en ella eliminamos tanto “DIM_Ambito_Geografico_1” como “DIM_Fecha_1”:

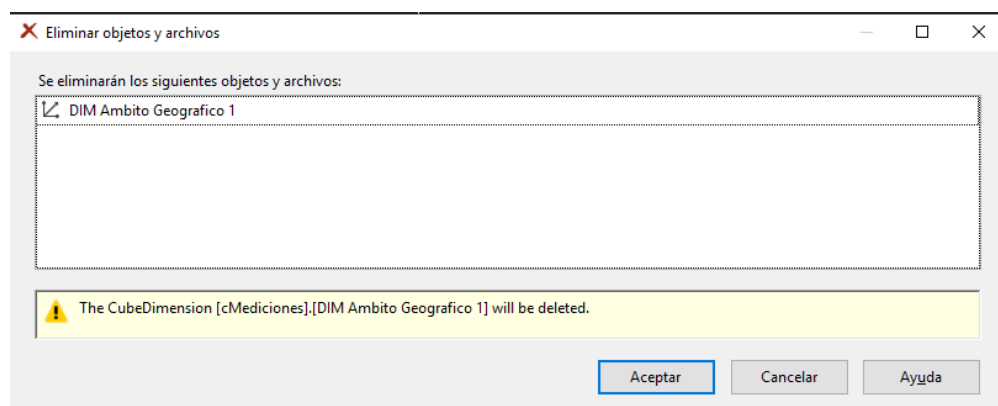


Ilustración 27 - Eliminación de DIM_Ambito_Geografico_1

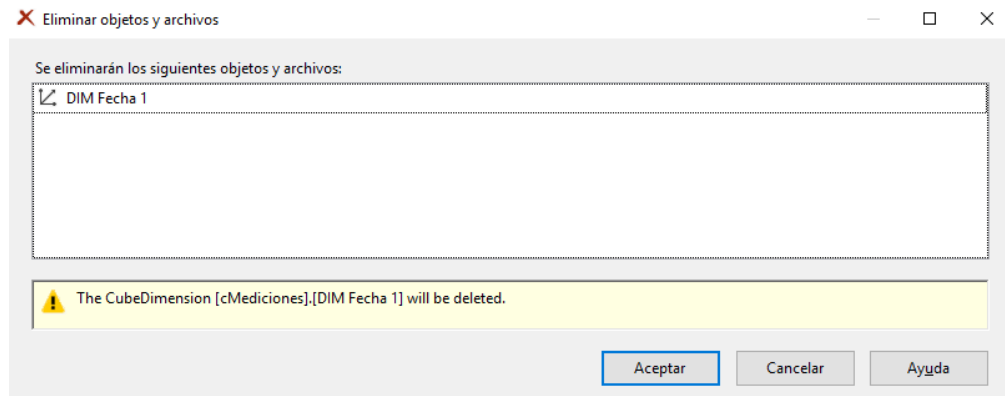


Ilustración 28 - Eliminación de DIM_Fecha_1.

Al eliminar dichas dimensiones se eliminan también del cubo, por lo que tenemos que añadir de nuevo dichas dimensiones, pero las que ya habíamos usado en la creación del cubo anterior. Para ello, en el diseño del cubo seleccionamos “Agregar dimensión al cubo”:

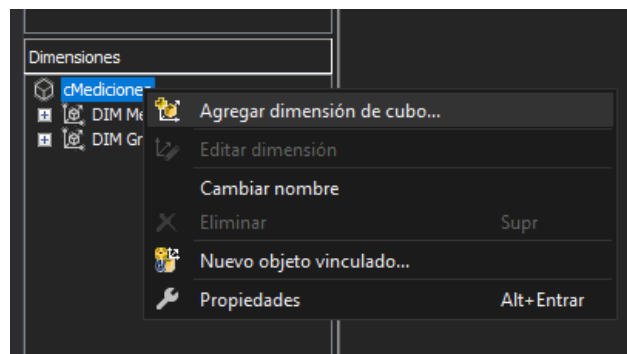


Ilustración 29 - Agregar dimensiones a cMediciones.

Seleccionamos la dimensión “DIM_Ambito_Geografico” y “DIM_Fecha”:

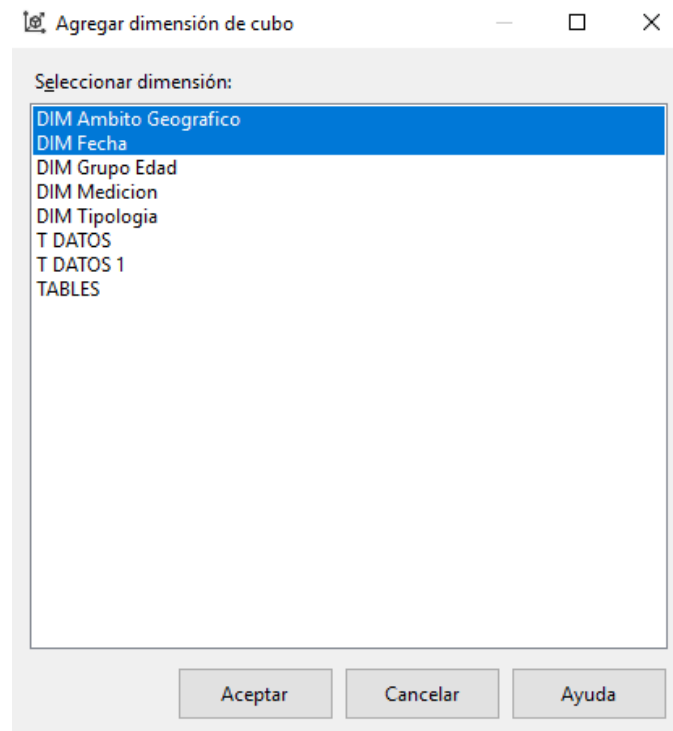


Ilustración 30 - Selección de dimensiones de cMediciones.

Finalmente, el cubo “cMediciones” nos queda de la siguiente forma:

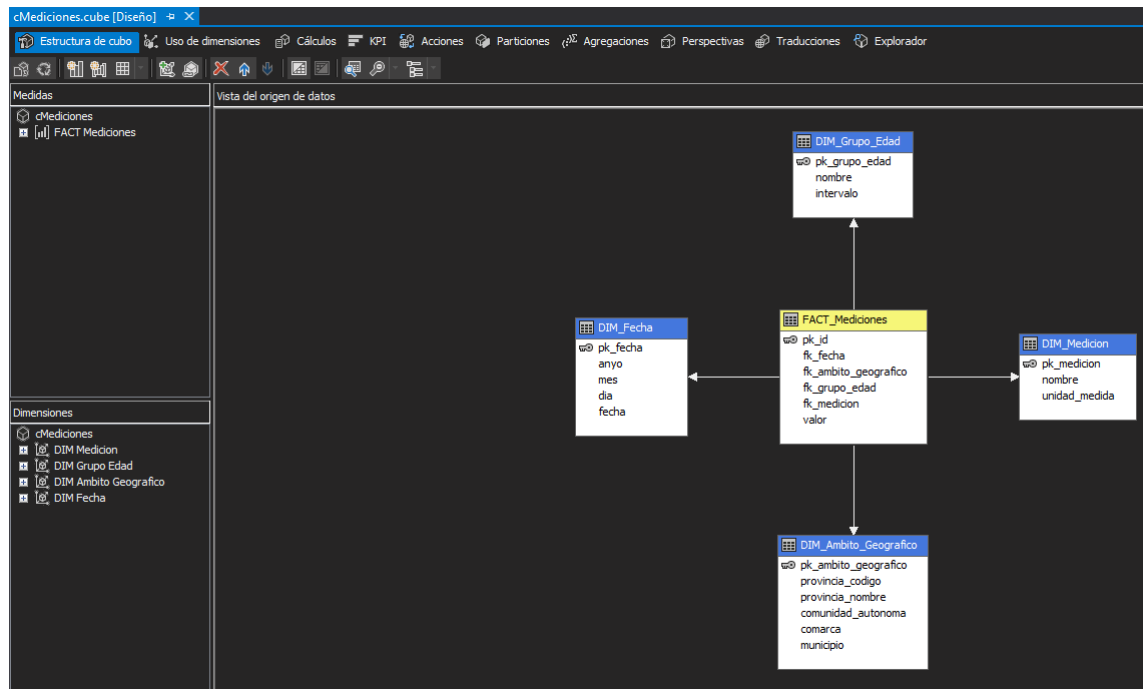



Ilustración 31 - Ventana diseño cMediciones.

Al borrar las dimensiones y añadirlas de nuevo hemos perdido las relaciones de ambas dimensiones, por lo que nos dirigimos a “Uso de dimensiones” y establecemos la relación de ambas:

Definir relación

Seleccionar tipo de relación: Normal

La tabla de dimensiones está unida directamente a la tabla de hechos.



Atributo de granularidad: Pk Ambito Geografico

Tabla de dimensiones: DIM_Ambito_Geografico

Tabla de grupos de medida: FACT_Mediciones

Relación:

Columnas de dimensión	Columnas de grupo de medida
pk_ambito_geografico	fk_ambito_geografico

Avanzadas...


Aceptar Cancelar Ayuda

Ilustración 32 - Relación de DIM_Ambito_Geografico.

Definir relación

Seleccionar tipo de relación: Normal

La tabla de dimensiones está unida directamente a la tabla de hechos.



Atributo de granularidad: Pk Fecha

Tabla de dimensiones: DIM_Fecha

Tabla de grupos de medida: FACT_Mediciones

Relación:

Columnas de dimensión	Columnas de grupo de medida
pk_fecha	fk_fecha

Avanzadas...

Aceptar Cancelar Ayuda

Ilustración 33 - Relación de DIM_Fecha.

Por lo tanto, las relaciones de todas las dimensiones nos quedarían de la siguiente forma:

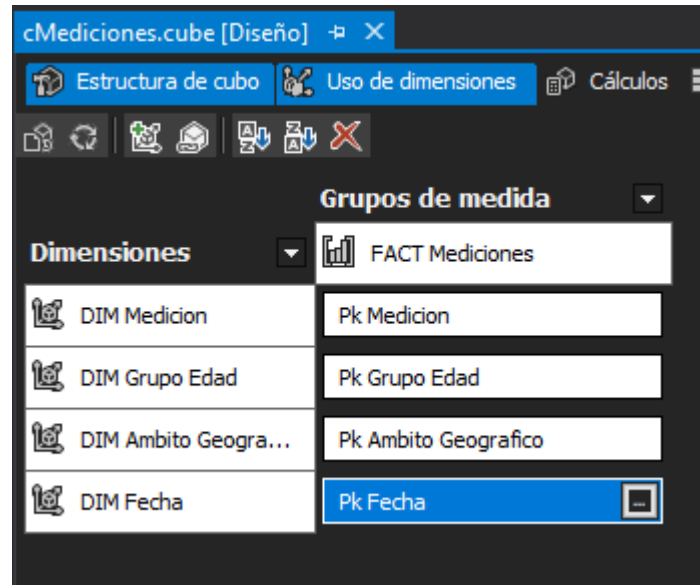


Ilustración 34 - Relaciones dimensiones cMediciones.

2.5. Jerarquías y dimensiones

En este apartado vamos a definir tanto las jerarquías y dimensiones que tenemos en nuestro modelo de datos.

2.5.1. DIM_Ambito_Geografico

Para definir cada una de las dimensiones nos vamos a la carpeta “Dimensiones” dentro del explorador de soluciones, y hacemos doble click sobre la dimensión que vamos a definir, en nuestro caso “DIM_Ambito_Geografico”.

Para editar los atributos de la dimensión solamente tenemos que arrastrar desde la vista del origen de datos hacia la dimensión en sí, tal y como se observa en la siguiente ilustración:

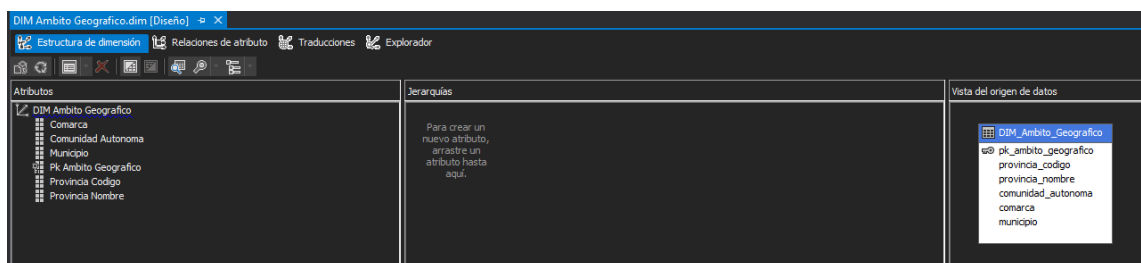


Ilustración 35 - DIM_Ambito_Geografico.

2.5.2. DIM_Fecha

Realizamos el mismo proceso para la dimensión “DIM_Fecha”, arrastramos los atributos desde la vista del origen de datos a los atributos de la dimensión:

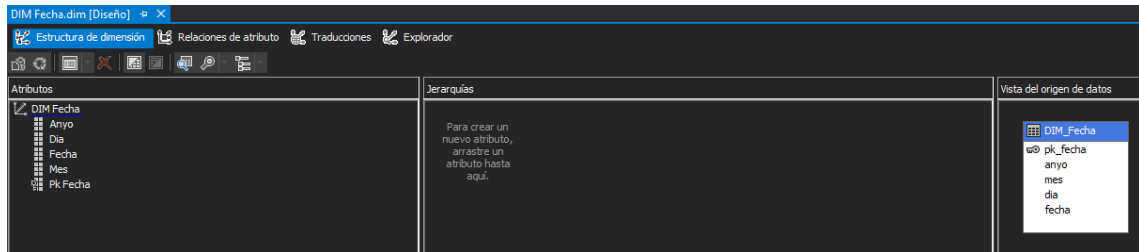


Ilustración 36 - DIM_Fecha.

2.5.3. DIM_Grupo_Edad

Para la dimensión “DIM_Grupo_Edad” arrastramos al igual que antes todos los atributos:

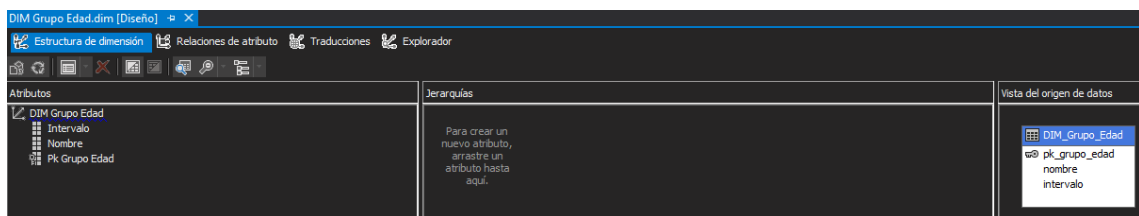


Ilustración 37 - DIM_Grupo_Edad.

2.5.4. DIM_Medicion

Realizamos el mismo proceso para “DIM_Medicion”:

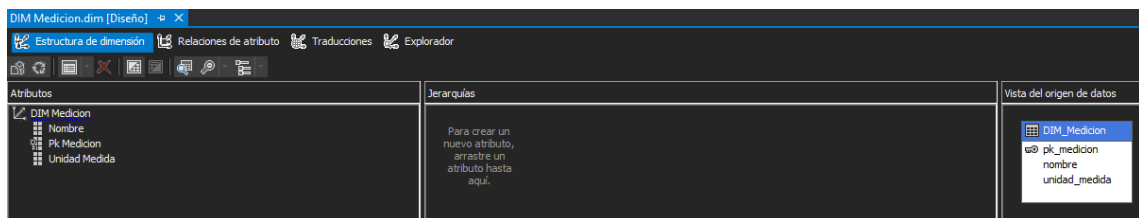


Ilustración 38 - DIM_Medicion.

2.5.5. DIM_Tipologia

Finalmente, definimos la dimensión “DIM_Tipologia” a partir de los atributos de la vista del origen de datos:

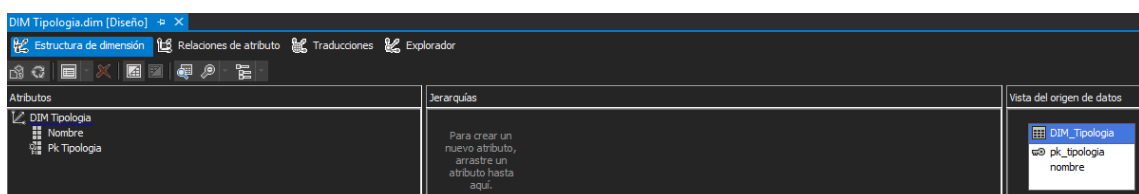


Ilustración 39 - DIM_Tipologia.

3. Implementación de la solución

Antes de comenzar con la explotación de los datos, tenemos que implementar la solución definida en el punto anterior.

Para poder implementarla, no dirigimos al menú “Compilar” y seleccionamos la opción “Implementar solución”:

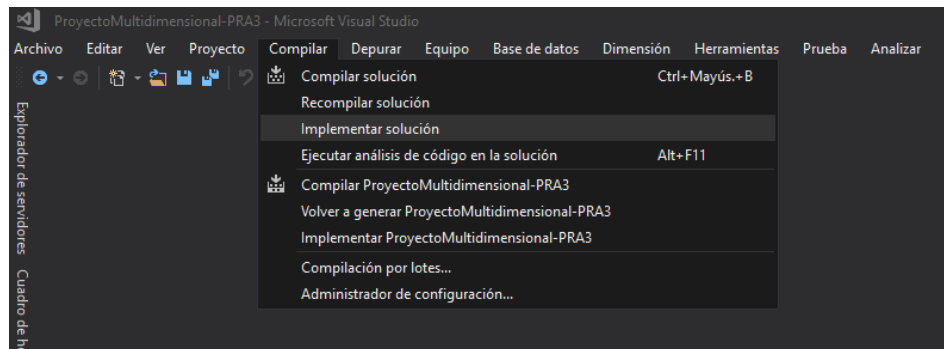


Ilustración 40 - Implementar la solución.

Como resultado de implementar la solución obtenemos la siguiente ventana:

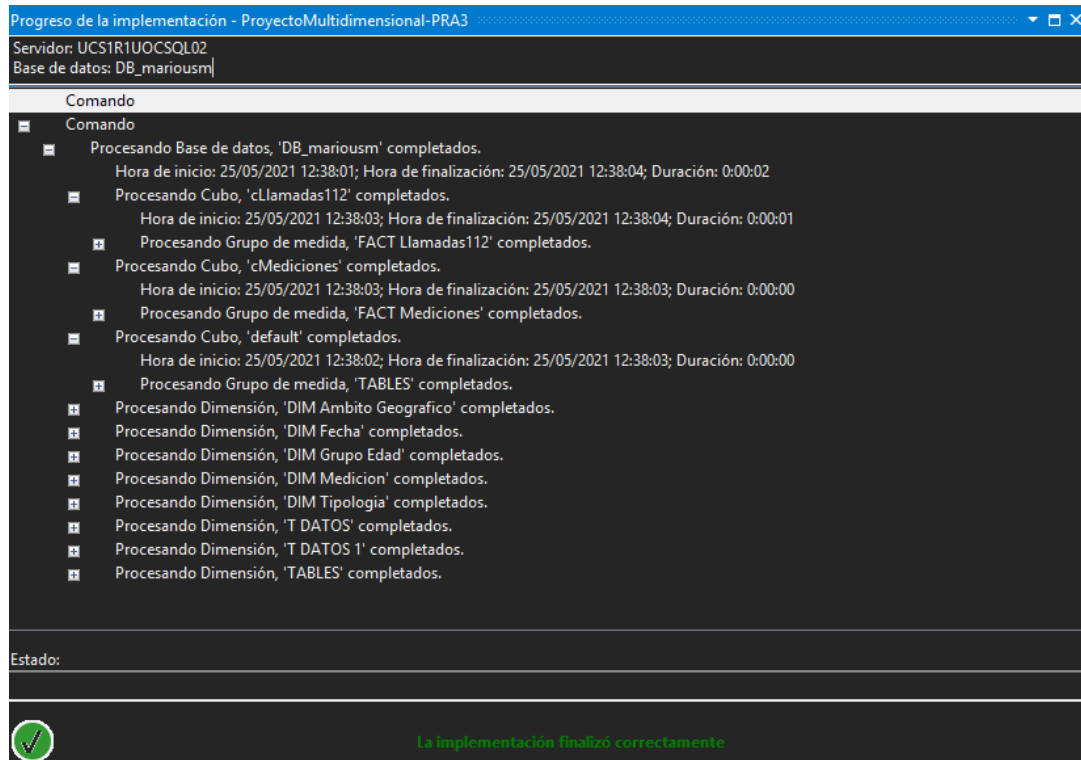


Ilustración 41 - Implementación de la solución.

Cabe destacar que hay un cubo “default” y tres dimensiones: “T DATOS”, “T DATOS 1” y “TABLES” que no son necesarias para la resolución de la práctica. Sin embargo, para no perder permisos hemos decidido mantener todos los elementos que teníamos cuando hemos creado el proyecto desde cero.

Finalmente, vemos las recomendaciones que nos sugiere *Visual Studio*:

- Definir una dimensión temporal.

Para ello, nos dirigimos a la dimensión “DIM Fecha”, la cual se encuentra dentro de la carpeta “Dimensiones” en el explorador de soluciones, e indicamos que dicha dimensión a partir de tu propiedad “Type” sea “Time” en vez de “Regular”:

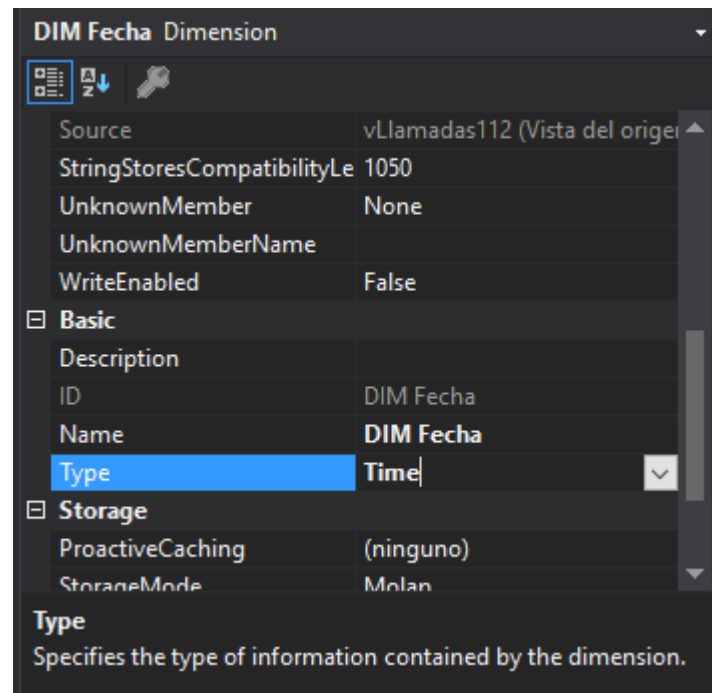


Ilustración 42 - Creación de la dimensión temporal.

Al tener los días, los meses, los años, e incluso la fecha en sí, modificamos los atributos de la dimensión para indicar que dichos campos son de tipo “days”, “months”, “years” y “date” respectivamente:

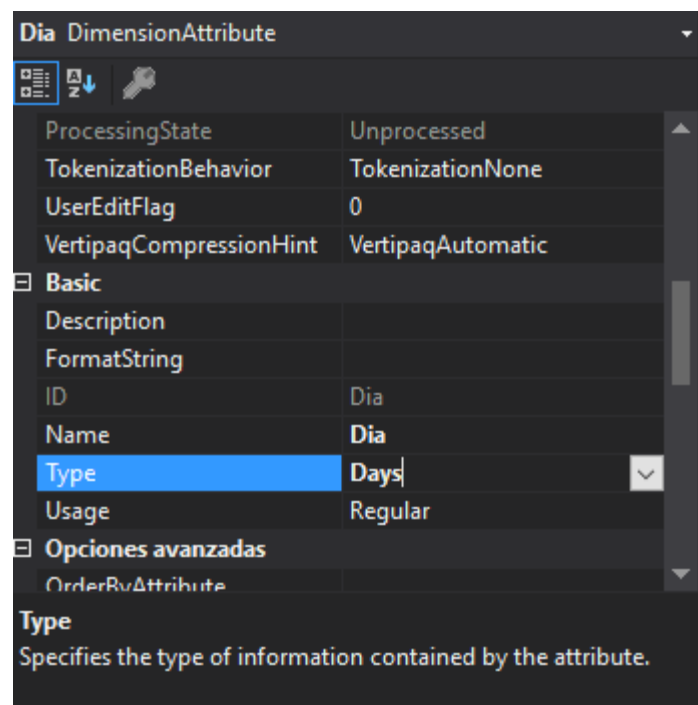


Ilustración 43 - Asignación del tipo para el día.

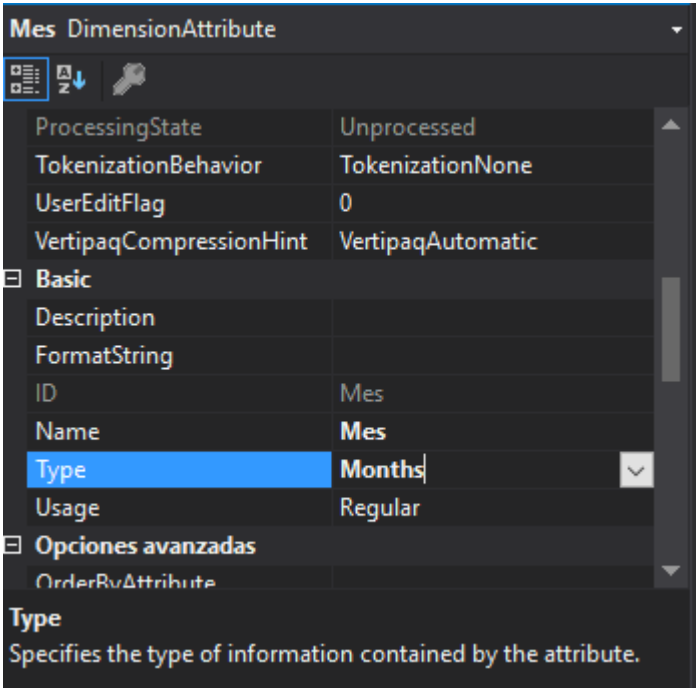


Ilustración 44 - Asignación del tipo para el mes.

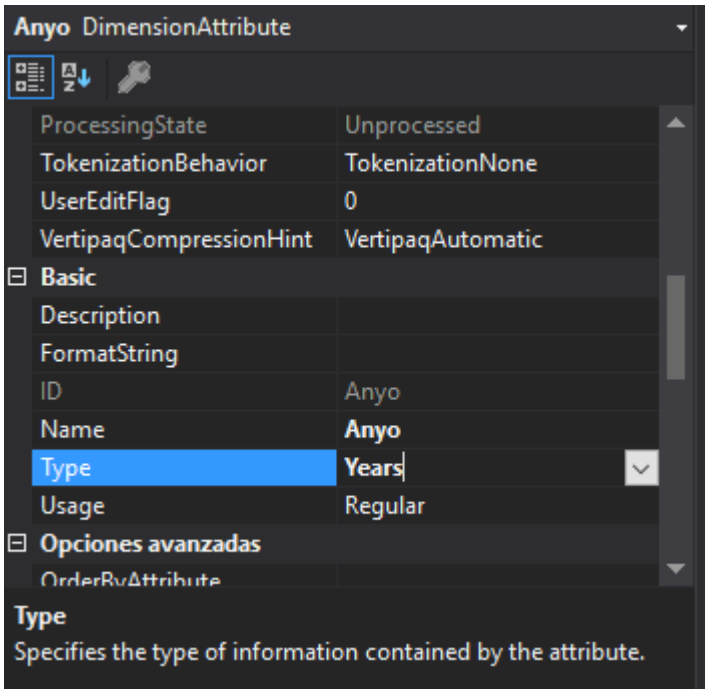


Ilustración 45 - Asignación del tipo para el año.

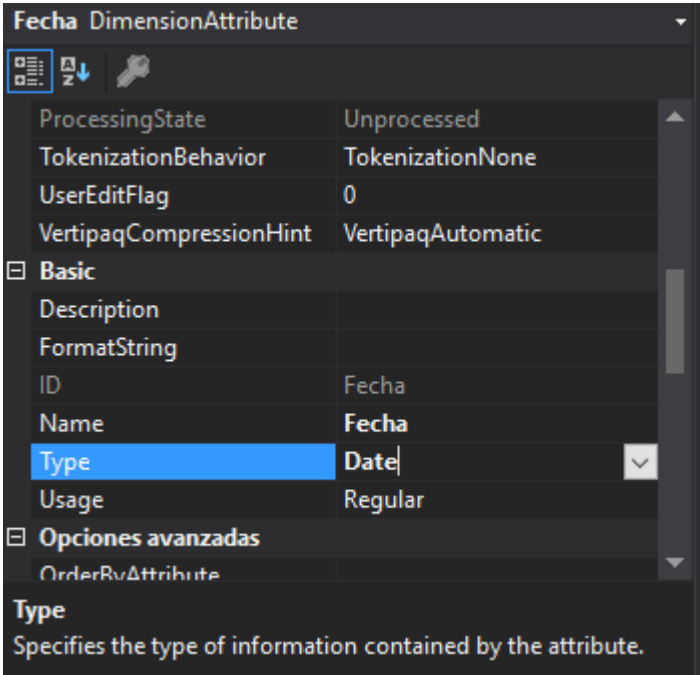


Ilustración 46 - Asignación del tipo para la fecha.

4.Explotación de la solución

a