



**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	9
1.1. Presentación	9
1.2. Descripción	9
2. Creación del modelo OLAP	11
2.1. Creación de la estructura física del modelo	11
2.2. Creación del proyecto.....	11
2.3. Vistas del origen de datos	17
2.3.1. Vista Llamadas112.....	18
2.3.2. Vista Mediciones	20
2.4. Creación de cubos	23
2.4.1. Cubo Llamadas112	24
2.4.2. Cubo Mediciones.....	27
2.5. Jerarquías y dimensiones	34
2.5.1. DIM_Ambito_Geografico	34
2.5.2. DIM_Fecha	38
2.5.3. DIM_Grupo_Edad.....	41
2.5.4. DIM_Medicion.....	41
2.5.5. DIM_Tipologia	42
3. Implementación de la solución	43
4. Explotación de la solución	50

4.1.	Análisis de las provincias con mayor porcentaje de movilidad según datos móviles	50
4.2.	Análisis del porcentaje de la población que evitaba aglomeraciones según la comunidad autónoma	53
4.3.	Análisis del promedio de sanciones por habitantes.....	56
4.4.	Evolución de las llamadas de urgencia al 112 en Cataluña por tipología de llamada	56

Índice de tablas

No se encuentran elementos de tabla de ilustraciones.

Índice de ilustraciones

Ilustración 1 - Base de datos DB_mariouism.	11
Ilustración 2 - Creación del proyecto multidimensional.	12
Ilustración 3 - Configuración origen de los datos.	13
Ilustración 4 - Propiedades del proyecto Visual Studio.	14
Ilustración 5 - Configuración destino de datos en Visual Studio.	15
Ilustración 6 - Ventana origen de datos.	16
Ilustración 7 - Modificación conexión de datos.	17
Ilustración 8 - Creación de nueva vista del origen de datos.	18
Ilustración 9 - Selección origen de datos vista.	18
Ilustración 10 - Selección tablas para la vista Llamadas112.	19
Ilustración 11 - Diseño vista vLlamadas112.	20
Ilustración 12 - Selección tablas para la vista Mediciones.	21
Ilustración 13 - Nombre de la vista vMediciones.	22
Ilustración 14 - Diseño vista vMediciones.	23
Ilustración 15 - Creación de un cubo.	24
Ilustración 16 - Uso de tablas existentes en el cubo.	24
Ilustración 17 - Selección vista del origen de datos cLlamadas112.	25
Ilustración 18 - Selección métricas de cLlamadas112.	25
Ilustración 19 - Selección dimensiones de cLlamadas112.	26
Ilustración 20 - Resumen de cLlamadas112.	26
Ilustración 21 - Ventana diseño cLlamadas112.	27
Ilustración 22 - Selección vista del origen de datos cMediciones.	28
Ilustración 23 - Selección métricas de cMediciones.	28
Ilustración 24 - Selección dimensiones de cMediciones.	29
Ilustración 25 - Resumen de cMediciones.	29

Ilustración 26 - Ventana diseño cMediciones.....	30
Ilustración 27 - Eliminación de DIM_Ambito_Geografico_1	30
Ilustración 28 - Eliminación de DIM_Fecha_1.	31
Ilustración 29 - Agregar dimensiones a cMediciones.	31
Ilustración 30 - Selección de dimensiones de cMediciones.	32
Ilustración 31 - Ventana diseño cMediciones.....	32
Ilustración 32 - Relación de DIM_Ambito_Geografico.	33
Ilustración 33 - Relación de DIM_Fecha.	33
Ilustración 34 - Relaciones dimensiones cMediciones.	34
Ilustración 35 - DIM_Ambito_Geografico.....	34
Ilustración 36 - Jerarquía DIM_Ambito_Geografico.	35
Ilustración 37 - KeyColumns provincia_nombre.....	35
Ilustración 38 NameColumn provincia_nombre.....	36
Ilustración 39 - KeyColumns comarca.....	36
Ilustración 40 - NameColumn comarca.	37
Ilustración 41 - KeyColumns municipio.....	37
Ilustración 42 - NameColumn municipio.	38
Ilustración 43 - DIM_Fecha.	38
Ilustración 44 - Jerarquía DIM_Fecha.	39
Ilustración 45 - KeyColumns mes.....	39
Ilustración 46 - NameColumn mes.....	40
Ilustración 47 - KeyColumns dia.....	40
Ilustración 48 - NameColumn dia.	41
Ilustración 49 - DIM_Grupo_Edad.	41
Ilustración 50 - DIM_Medicion.	42
Ilustración 51 - DIM_Tipologia.....	42
Ilustración 52 - Implementar la solución.	43
Ilustración 53 - Implementación de la solución.....	44
Ilustración 54 - Creación de la dimensión temporal.....	45
Ilustración 55 - Asignación del tipo para el día.....	46
Ilustración 56 - Asignación del tipo para el mes.	46
Ilustración 62 - Asignación del tipo para el año.....	47

Ilustración 63 - Asignación del tipo para la fecha.	47
Ilustración 57 - Relación de atributos DIM Ambito Geografico.	48
Ilustración 58 - Relación de atributos DIM Fecha.	48
Ilustración 59 - Establcer la relación de flexible a rígida.	48
Ilustración 60 - Relaciones rígidas en DIM Ambito Geografico.	49
Ilustración 61 - Relaciones rígidas en DIM Fecha.	49
Ilustración 64 - Análisis de las provincias con mayor porcentaje de movilidad.	50
Ilustración 65 - Creación de la media.	51
Ilustración 66 - Creación de la media.	51
Ilustración 67 - Análisis de las provincias con mayor porcentaje de movilidad.	52
Ilustración 68 - Análisis de las provincias con mayor porcentaje de movilidad.	52
Ilustración 69 - Análisis del porcentaje de población que evitaba aglomeraciones según la CCAA.	53
Ilustración 70 - Creación número de filas.	53
Ilustración 71 - Creación número de filas.	54
Ilustración 72- Análisis del porcentaje de población que evitaba aglomeraciones según la CCAA.	54
Ilustración 73 - Creación media población aglomeración.	55
Ilustración 74 - Análisis del porcentaje de población que evitaba aglomeraciones según la CCAA.	56
Ilustración 75 - Evolución de las llamadas según la tipología.	57
Ilustración 76 - Evolución de las llamadas según la tipología.	58
Ilustración 77 - Evolución de las llamadas según la tipología.	59
Ilustración 78 - Evolución de las llamadas según la tipología.	60

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*”:

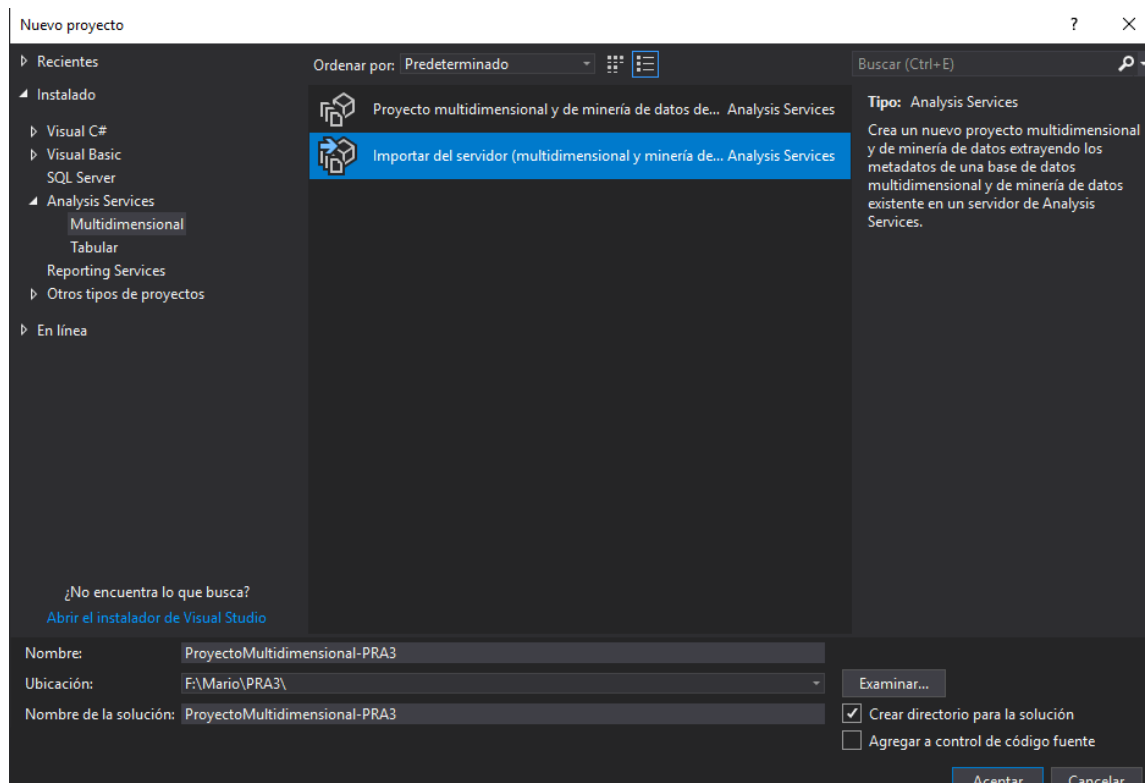


Ilustración 2 - Creación del proyecto multidimensional.

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”:

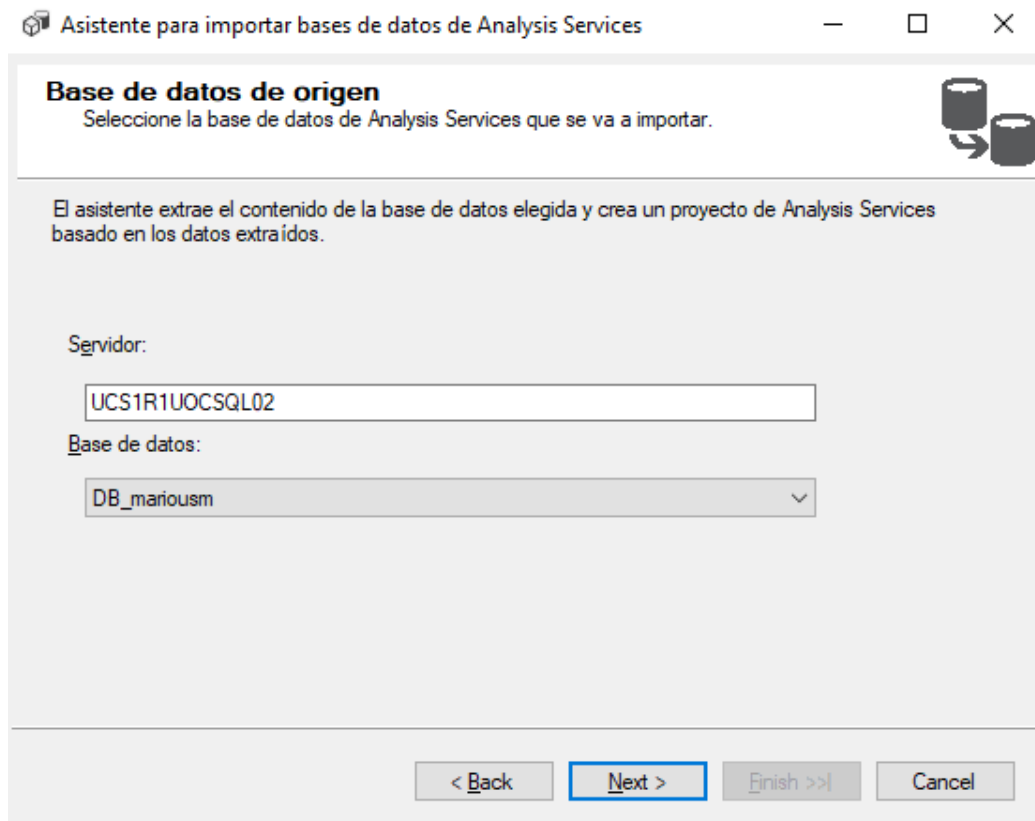


Ilustración 3 - Configuración origen de los datos.

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:

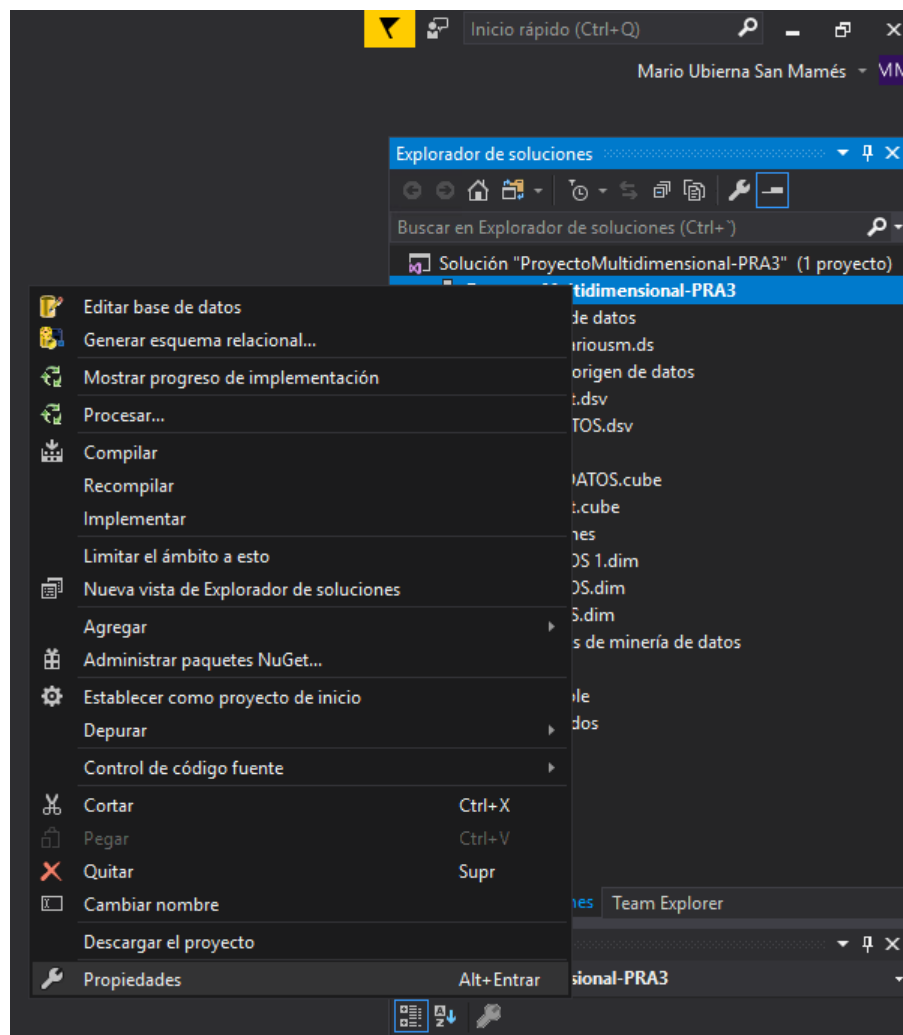


Ilustración 4 - Propiedades del proyecto Visual Studio.

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:

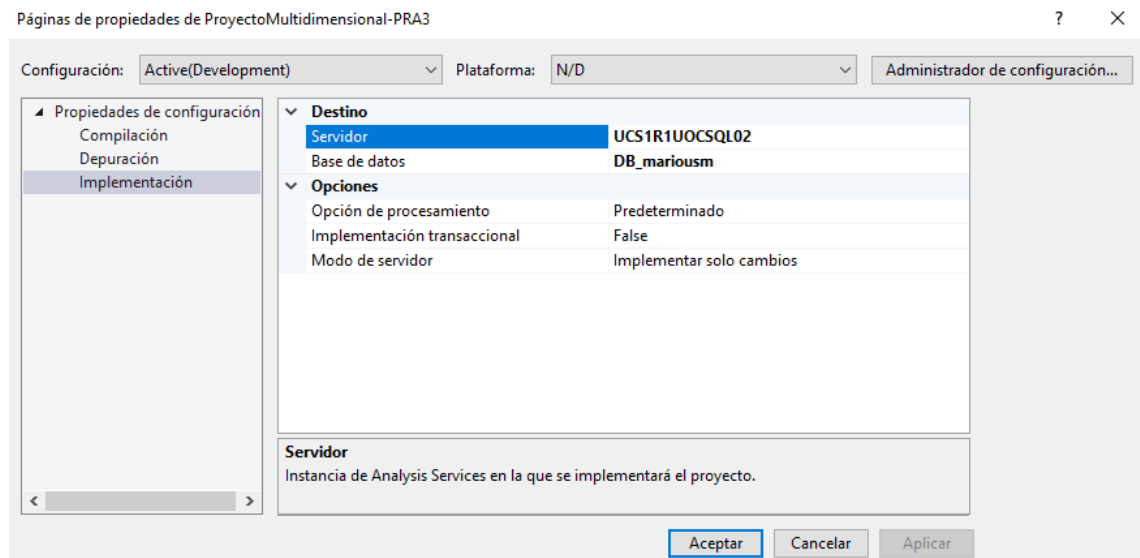


Ilustración 5 - Configuración destino de datos en Visual Studio.

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:

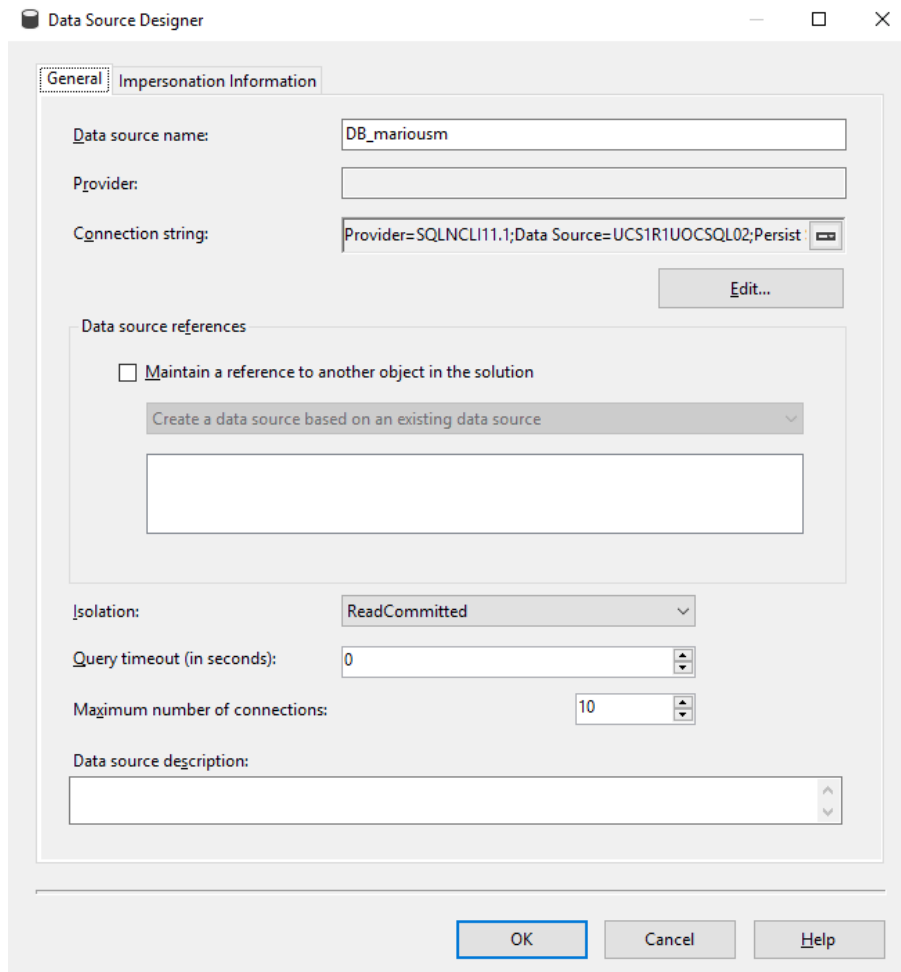


Ilustración 6 - Ventana origen de datos.

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

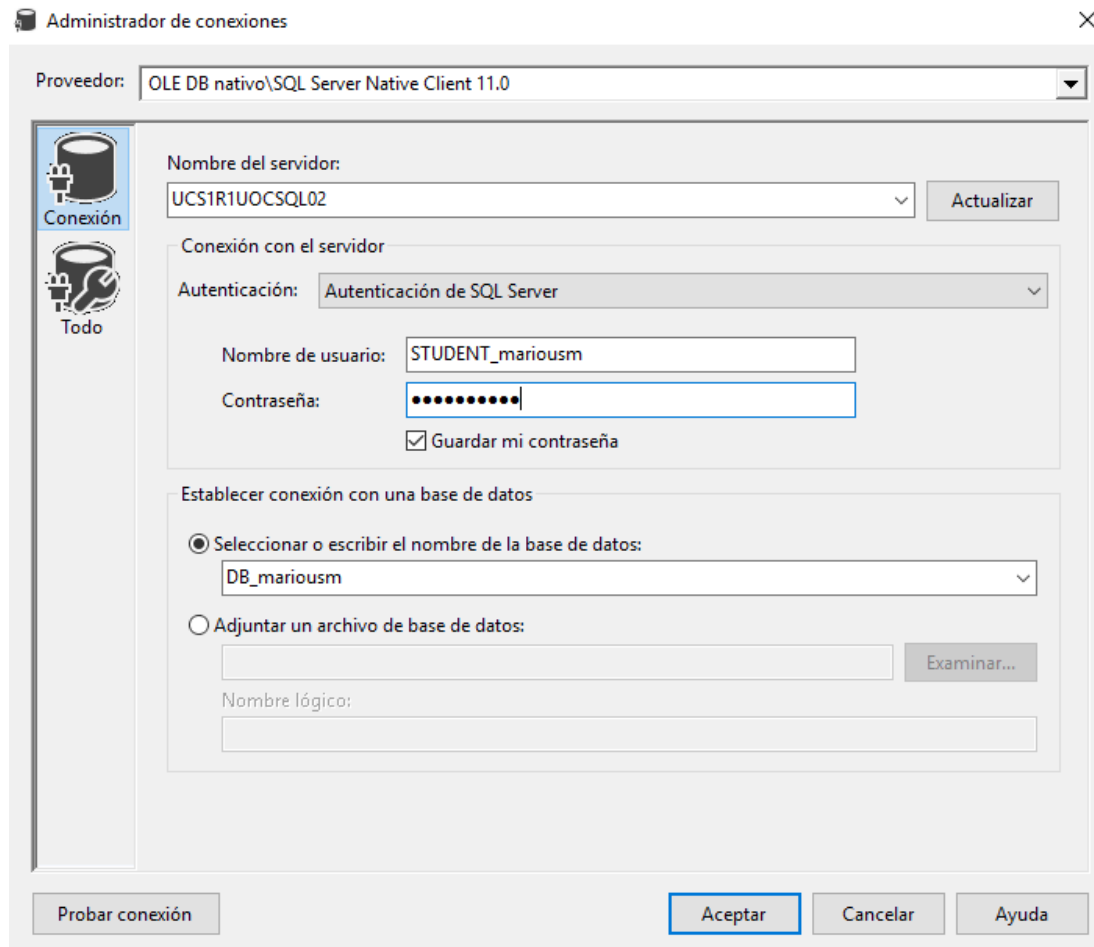


Ilustración 7 - Modificación conexión de datos.

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”:

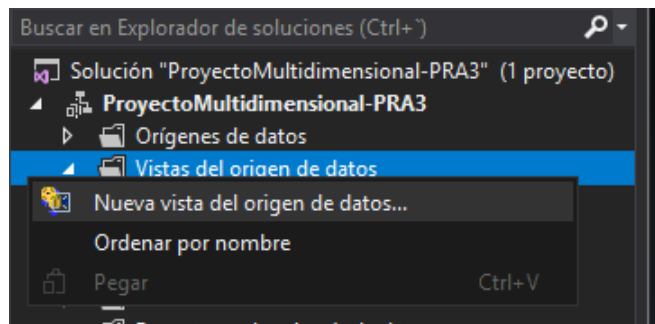


Ilustración 8 - Creación de nueva vista del origen de datos.

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:

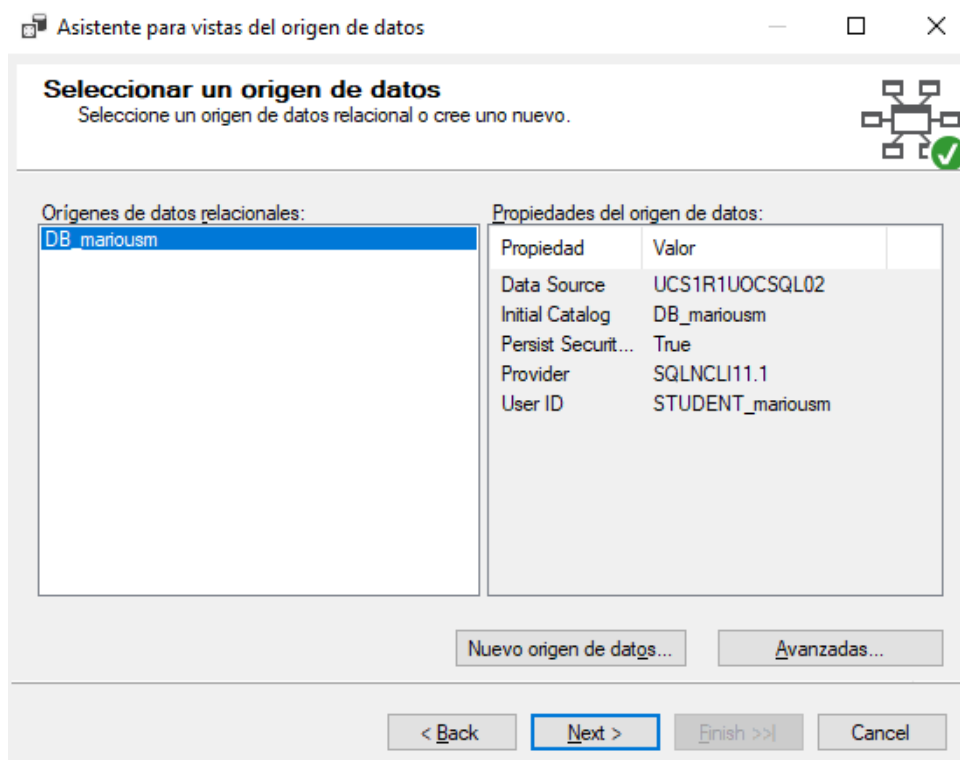


Ilustración 9 - Selección origen de datos vista.

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”:

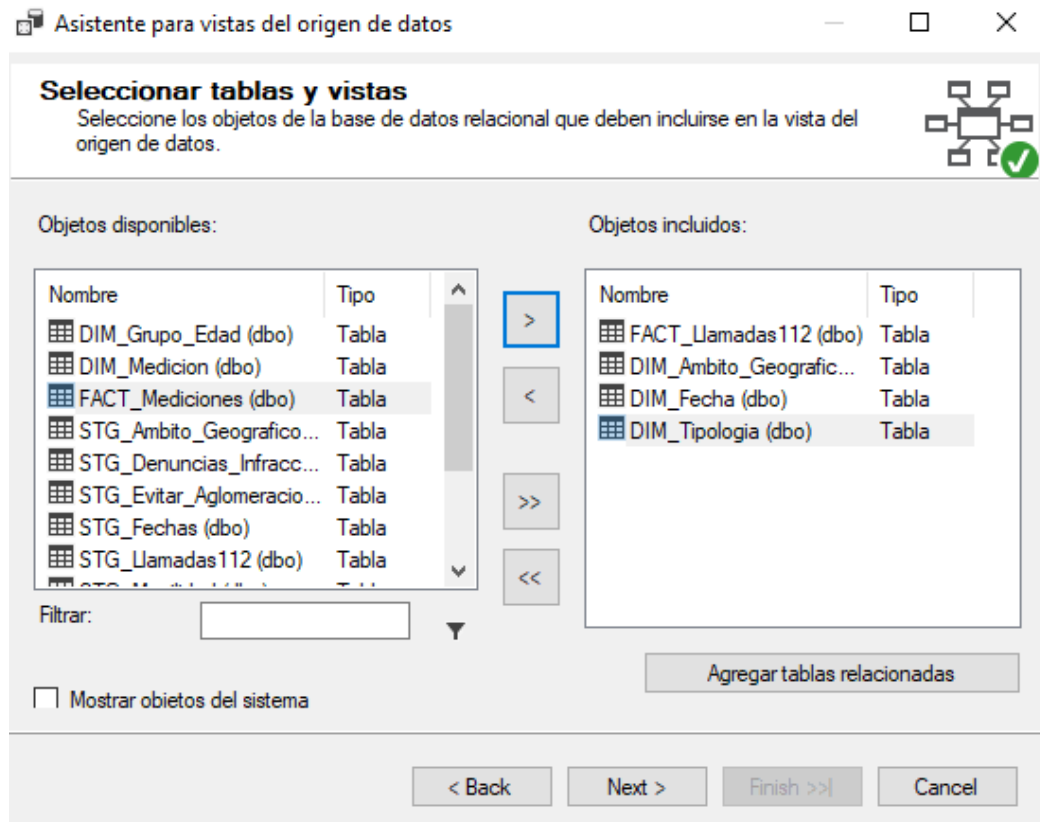


Ilustración 10 - Selección de tablas para la vista Llamadas112.

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:

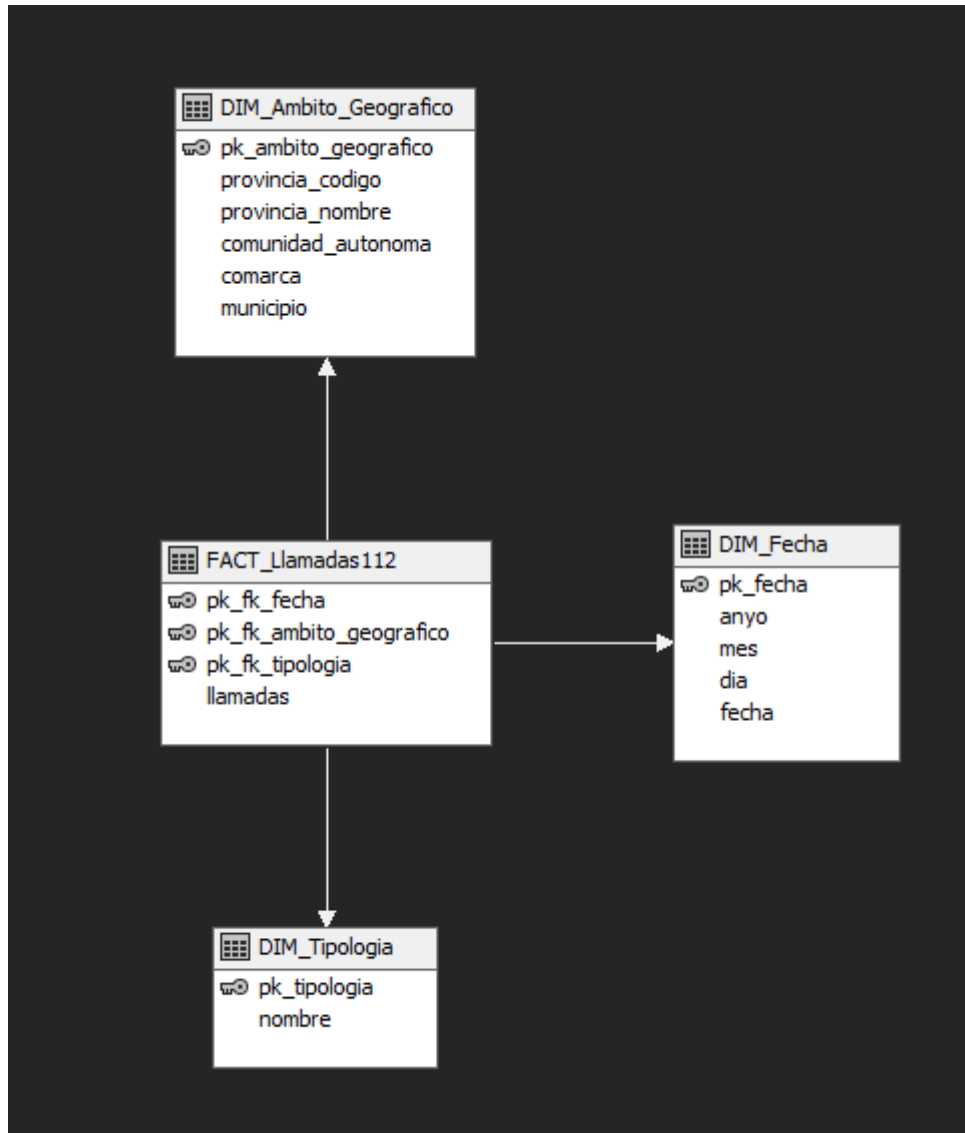


Ilustración 11 - Diseño vista vLlamadas112.

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**”:

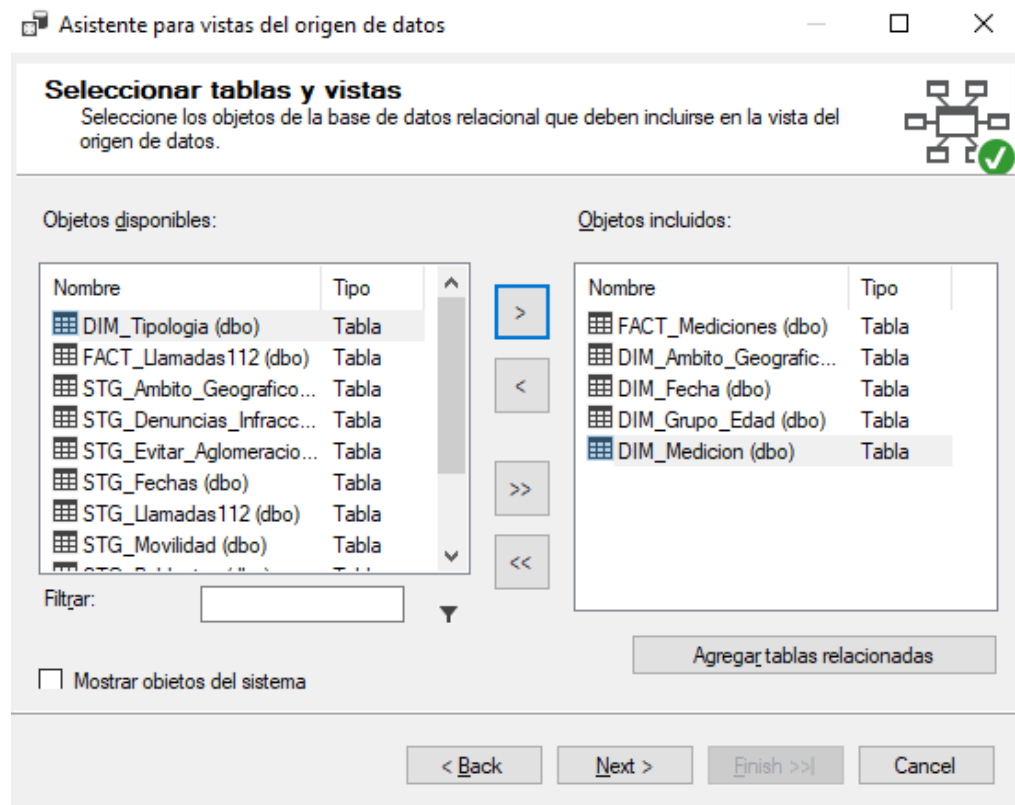


Ilustración 12 - Selección de tablas para la vista Mediciones.

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

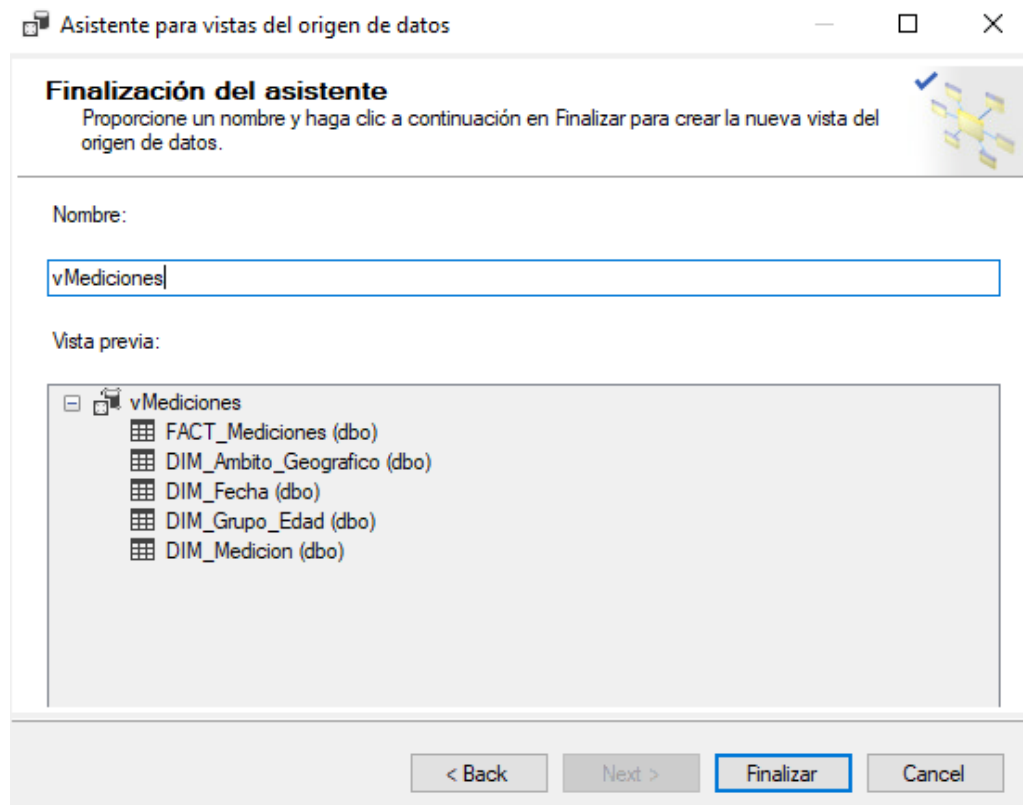


Ilustración 13 - Nombre de la vista vMediciones.

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

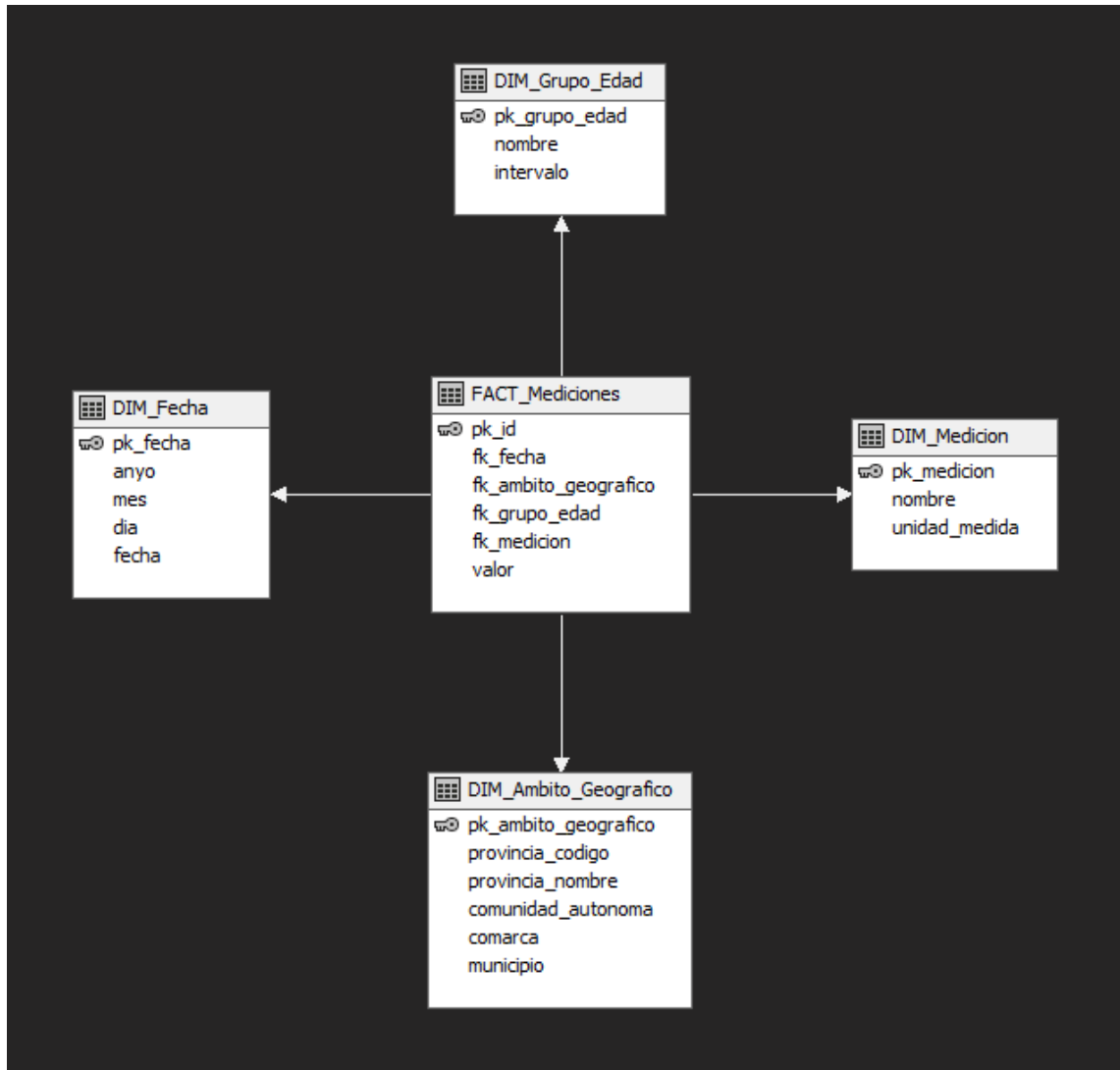


Ilustración 14 - Diseño vista vMediciones.

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”:

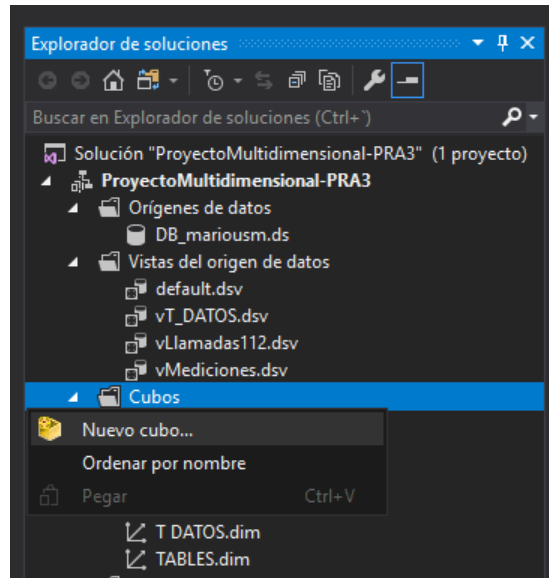


Ilustración 15 - Creación de un cubo.

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

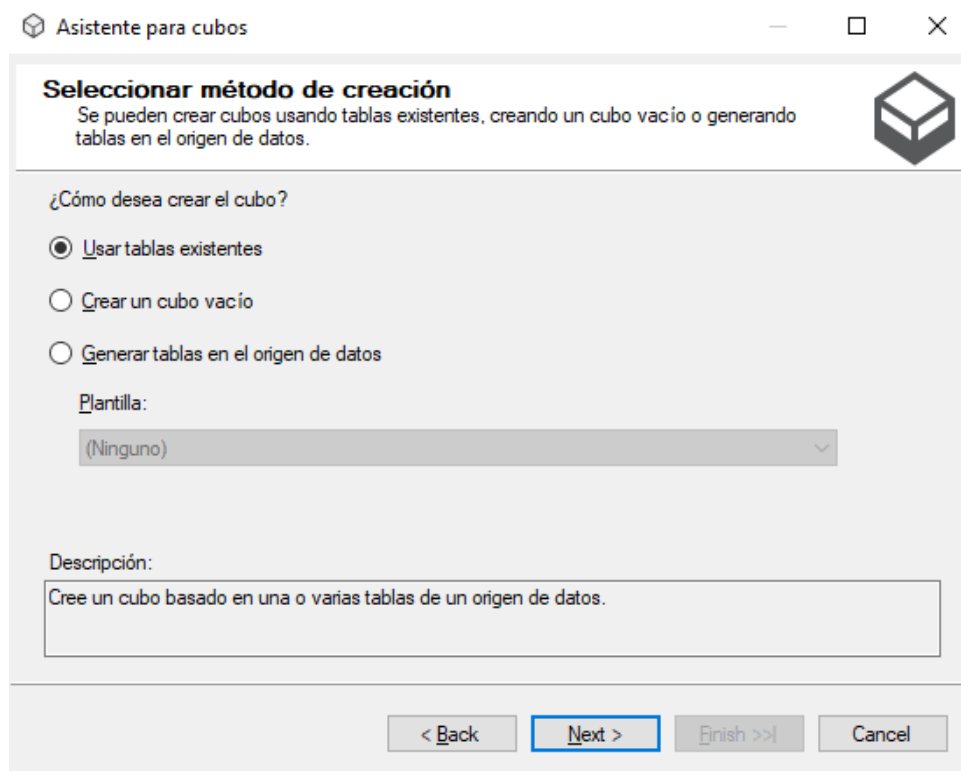


Ilustración 16 - Uso de tablas existentes en el cubo.

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”:

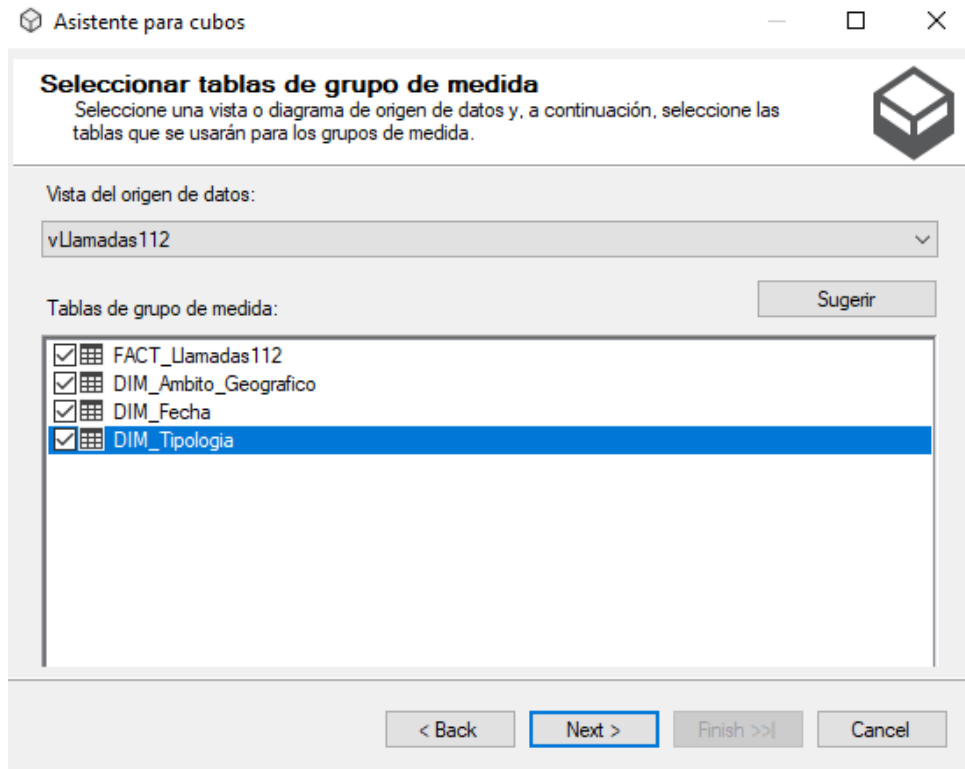


Ilustración 17 - Selección vista del origen de datos cLlamadas112.

Posteriormente, seleccionamos las métricas:

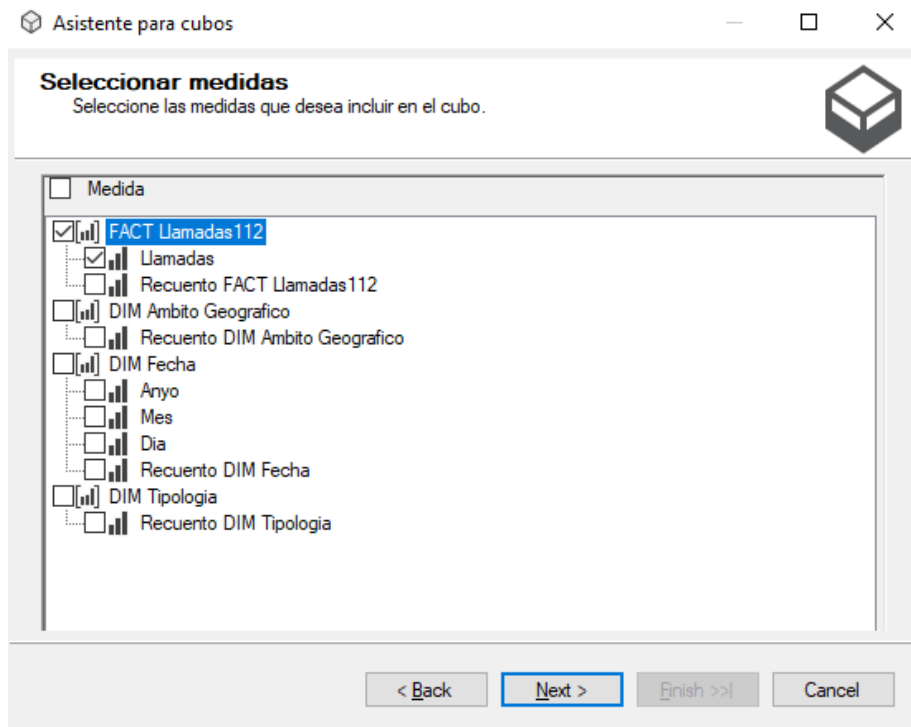


Ilustración 18 - Selección métricas de cLlamadas112.

Por último indicamos las dimensiones del cubo:

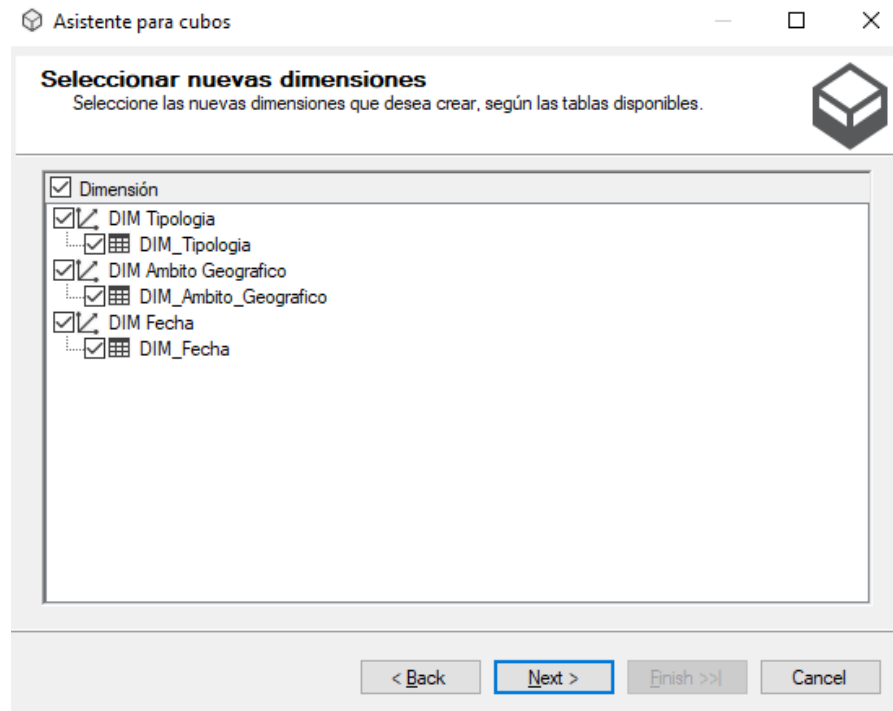


Ilustración 19 - Selección dimensiones de cLlamadas112.

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

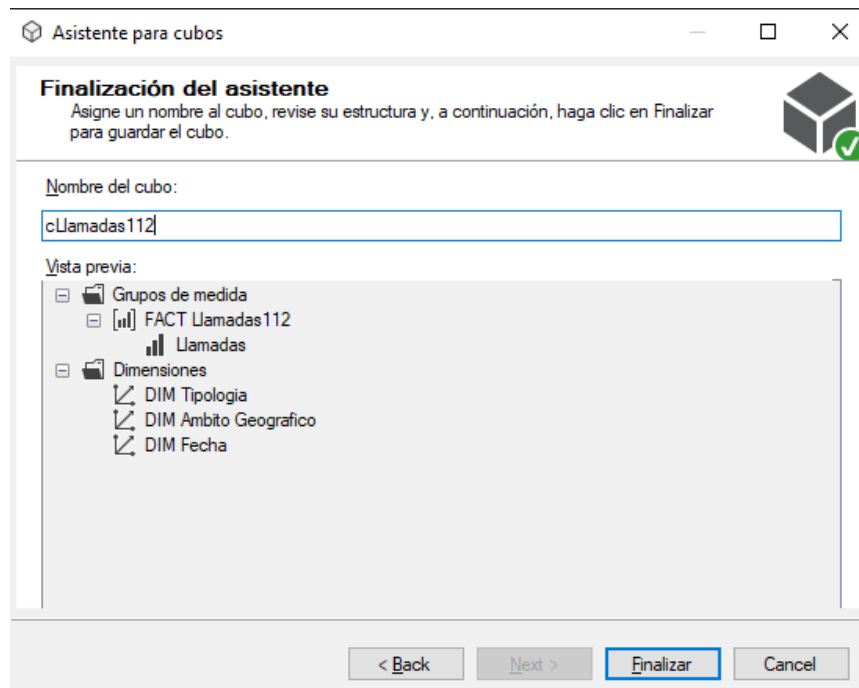


Ilustración 20 - Resumen de cLlamadas112.

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:

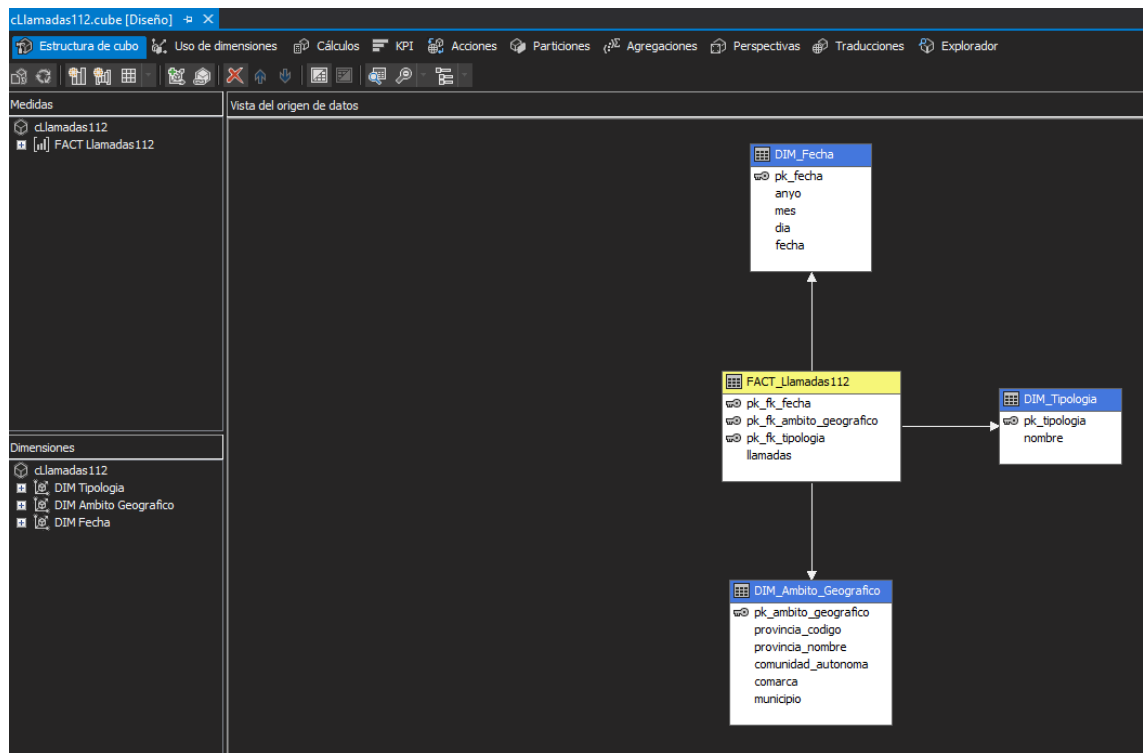


Ilustración 21 - Ventana diseño cLlamadas112.

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”:

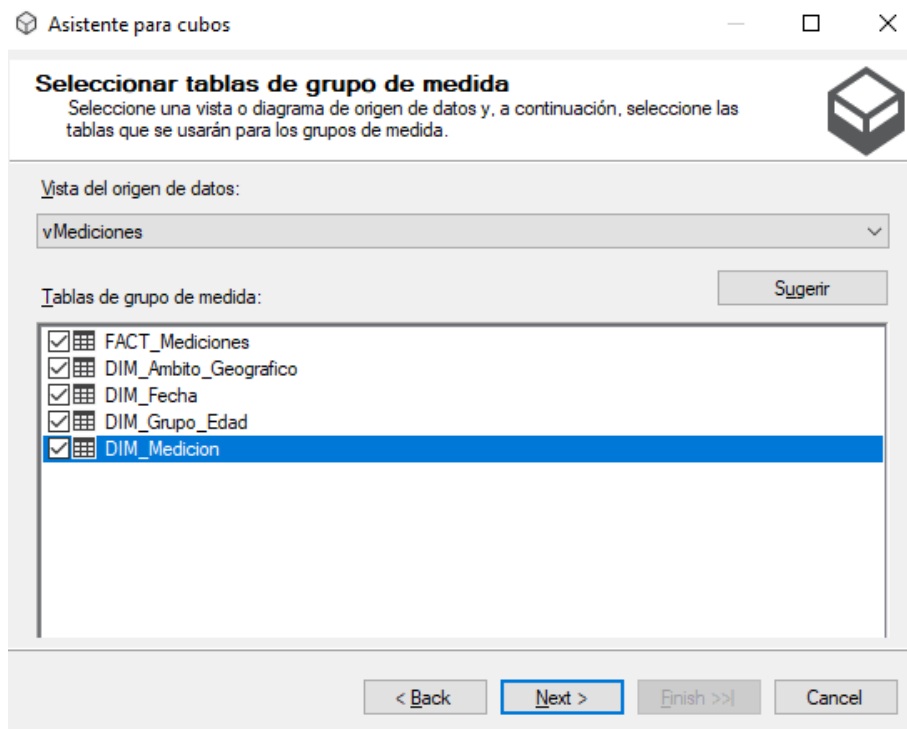


Ilustración 22 - Selección vista del origen de datos cMediciones.

Posteriormente, seleccionamos las métricas:

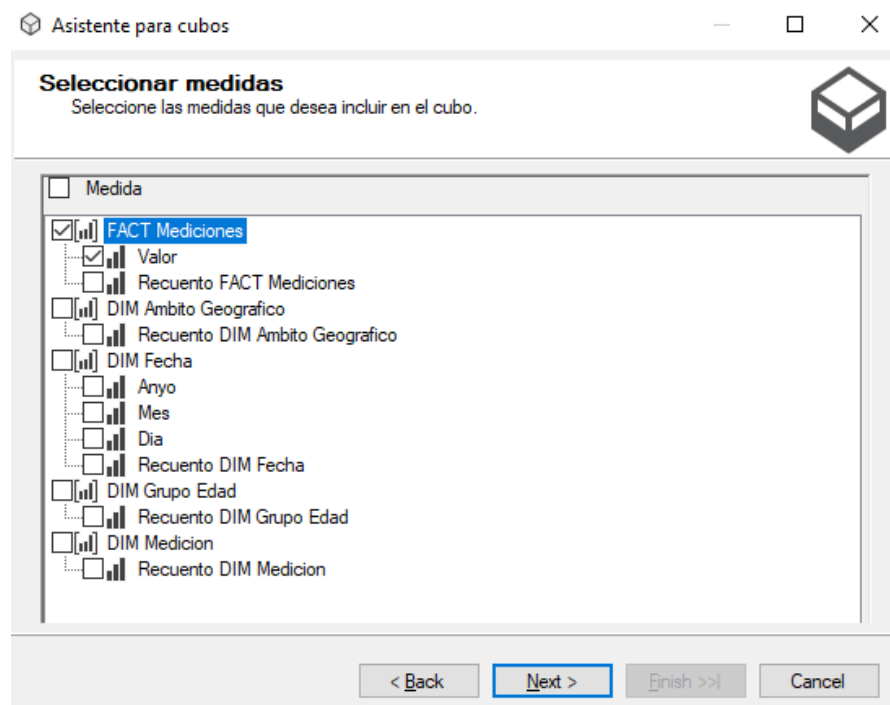


Ilustración 23 - Selección métricas de cMediciones.

Por último, indicamos las dimensiones del cubo:

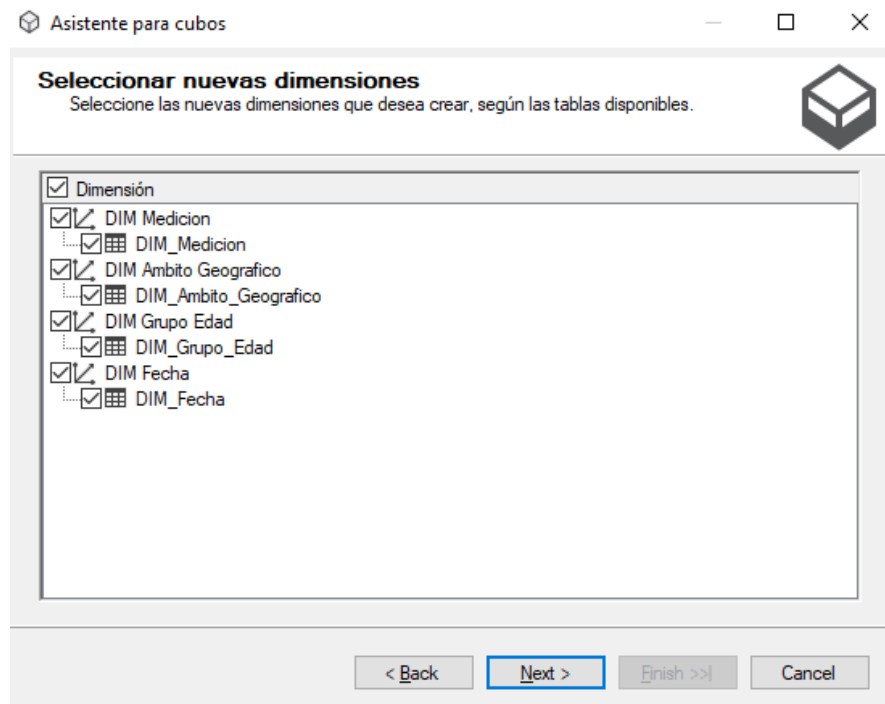


Ilustración 24 - Selección dimensiones de cMediciones.

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

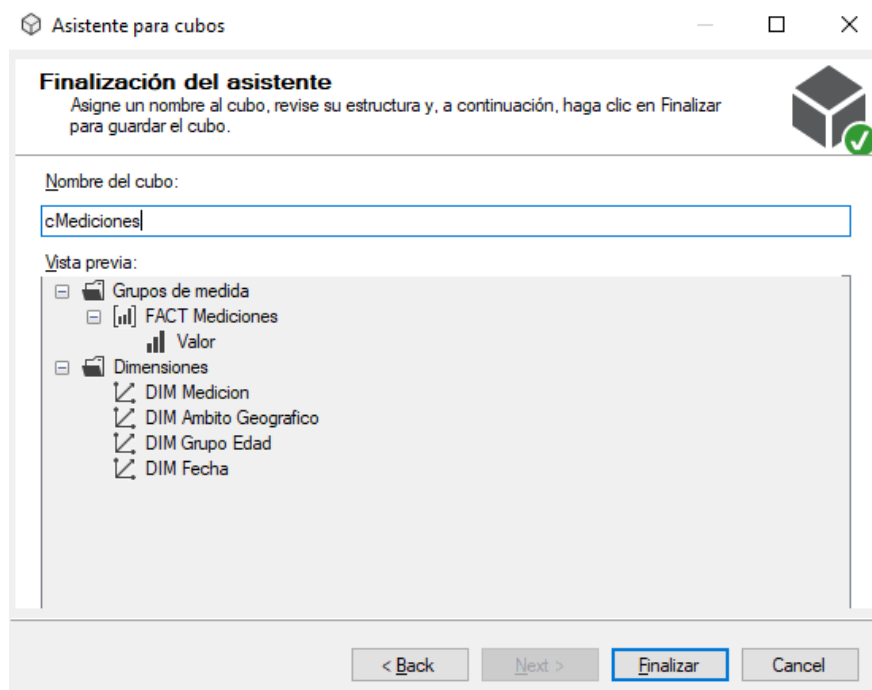


Ilustración 25 - Resumen de cMediciones.

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:

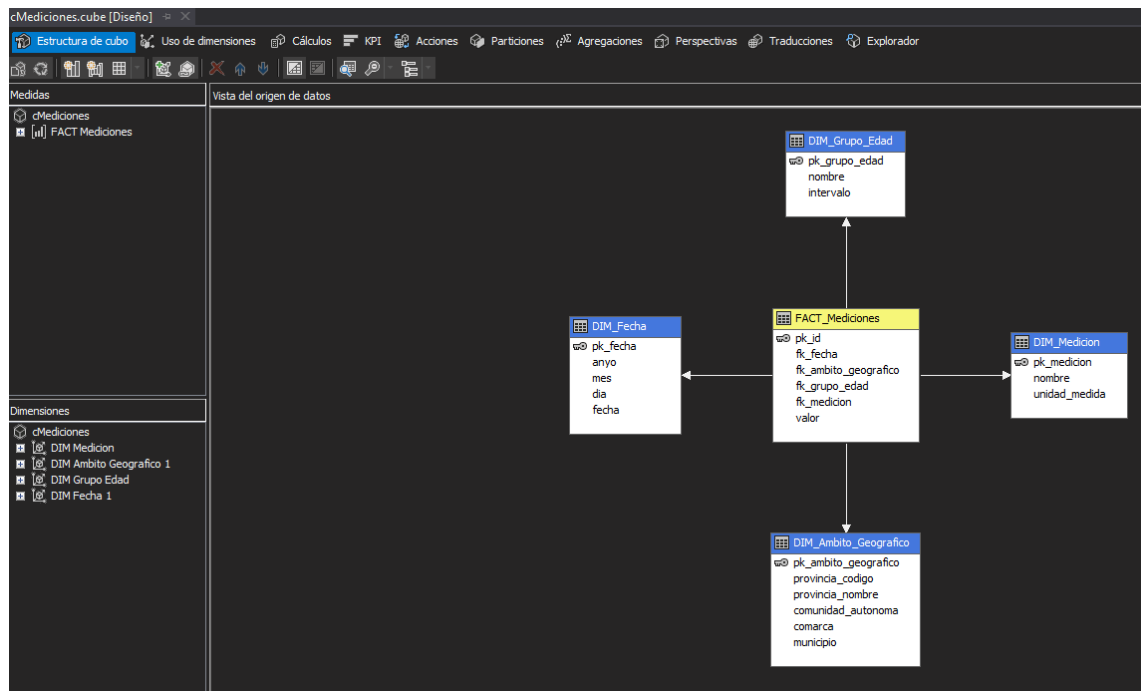


Ilustración 26 - Ventana diseño cMediciones.

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”:

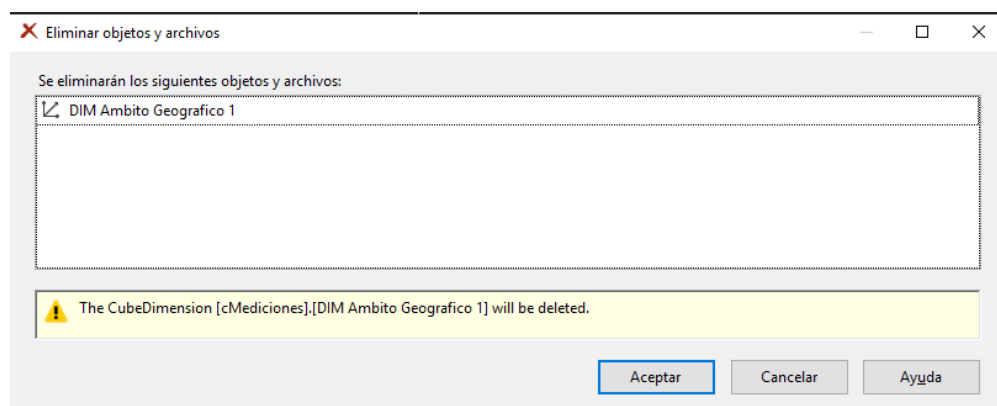


Ilustración 27 - Eliminación de DIM_Ambito_Geografico_1

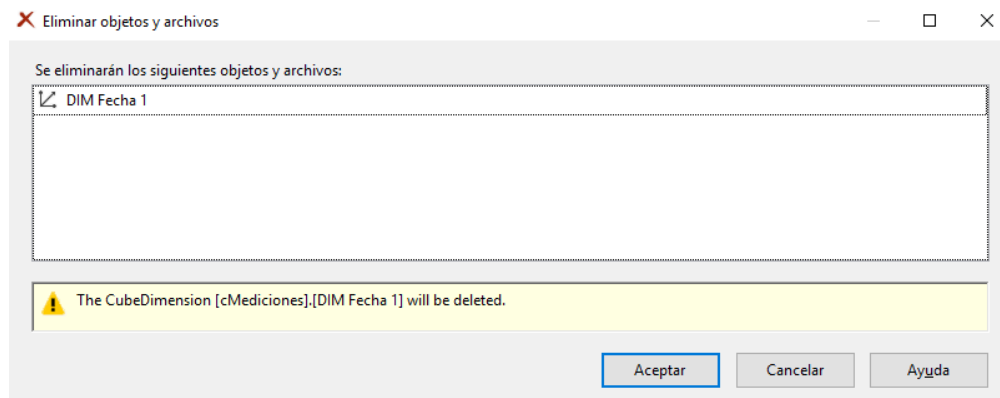


Ilustración 28 - Eliminación de DIM_Fecha_1.

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”:

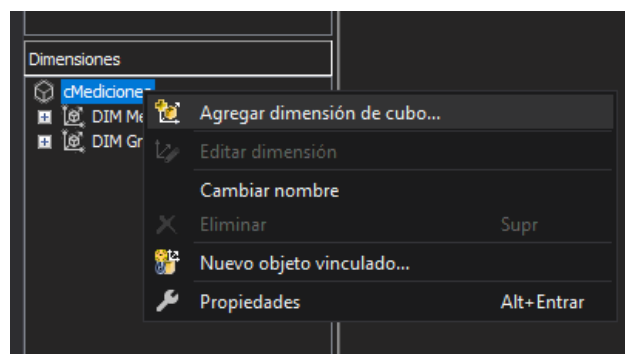


Ilustración 29 - Agregar dimensiones a cMediciones.

Seleccionamos la dimensión “DIM_Ambito_Geografico” y “DIM_Fecha”:

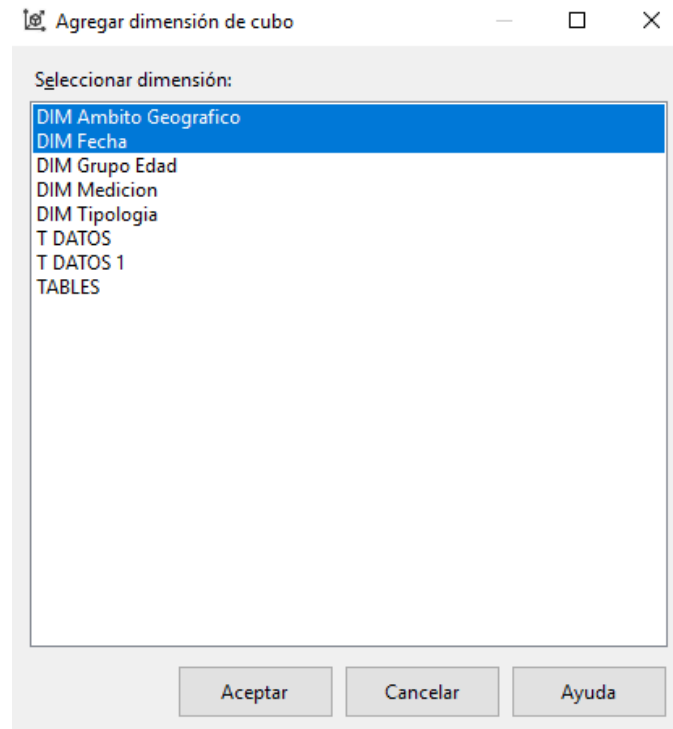


Ilustración 30 - Selección de dimensiones de cMediciones.

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

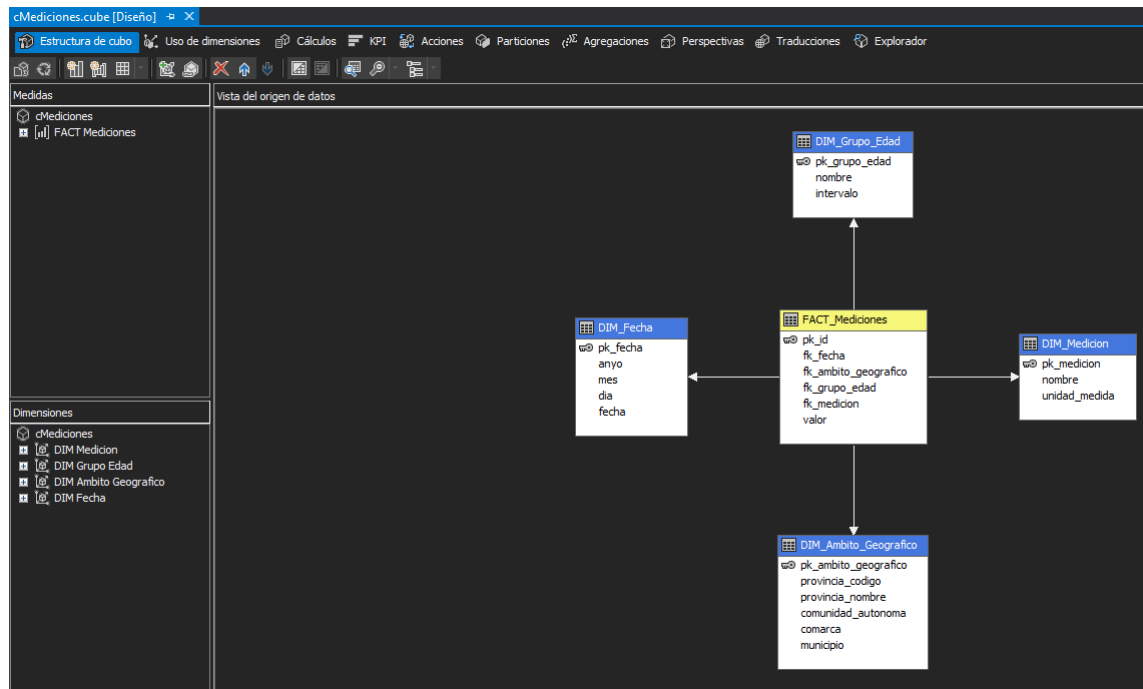



Ilustración 31 - Ventana diseño cMediciones.

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:

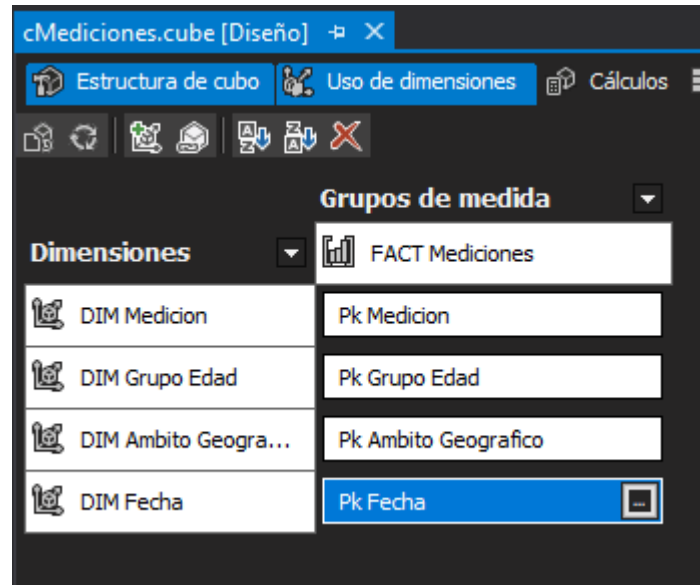


Ilustración 34 - Relaciones dimensiones cMediciones.

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:

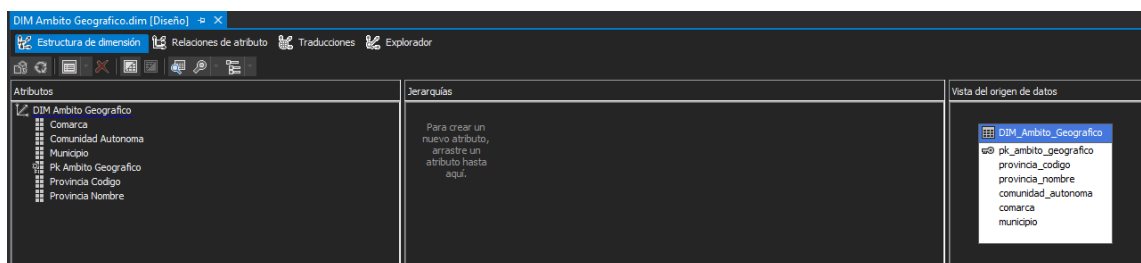


Ilustración 35 - DIM_Ambito_Geografico.

Por otro lado, definimos una jerarquía para que nos resulte luego más fácil realizar las consultas respecto al ámbito geográfico. La jerarquía va a ser “Comunidad autónoma” > “Provincia nombre” > “Comarca” > “Municipio”:

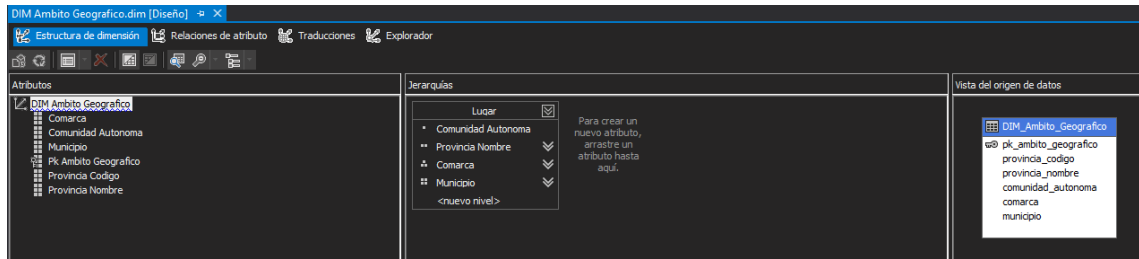


Ilustración 36 - Jerarquía DIM_Ambito_Geografico.

Al definir la jerarquía tenemos que cambiar tanto el atributo “KeyColumns” como el atributo “NameColumn” de los campos pertenecientes a la jerarquía con la excepción de la jerarquía más alta, en nuestro caso el atributo “Comunidad Autonoma”.

Por lo tanto, vamos a definir dichas propiedades para los atributos “Provincia Nombre”, “Comarca” y “Municipio” que se encuentran dentro de la dimensión:

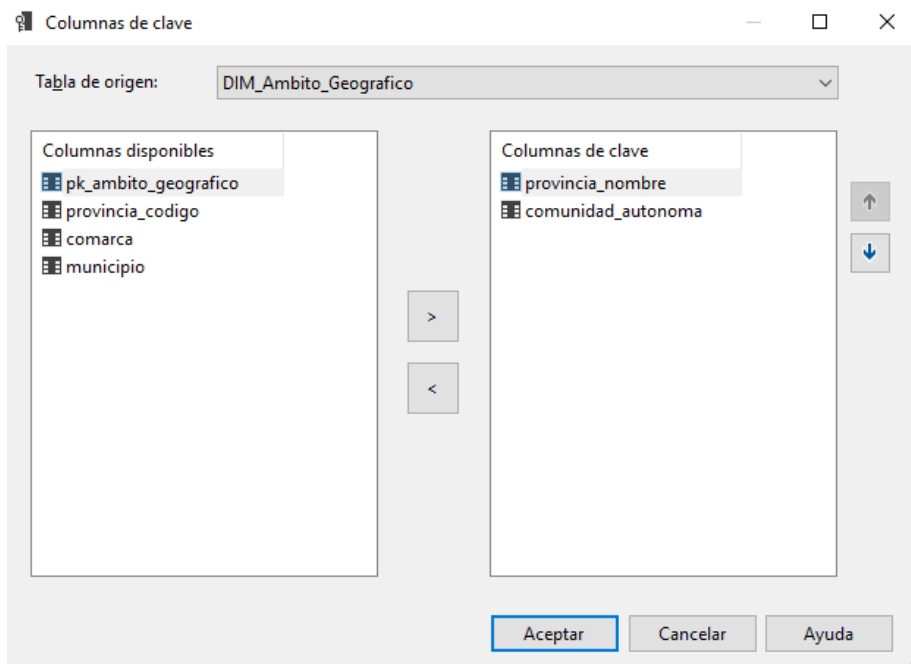


Ilustración 37 - KeyColumns provincia_nombre.

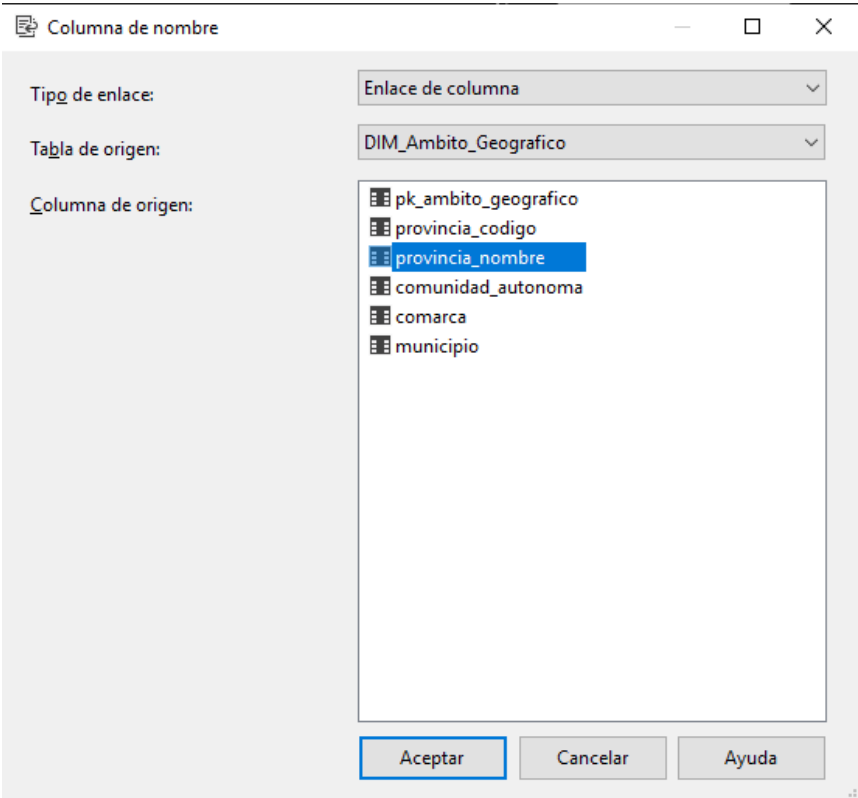


Ilustración 38 NameColumn provincia_nombre.

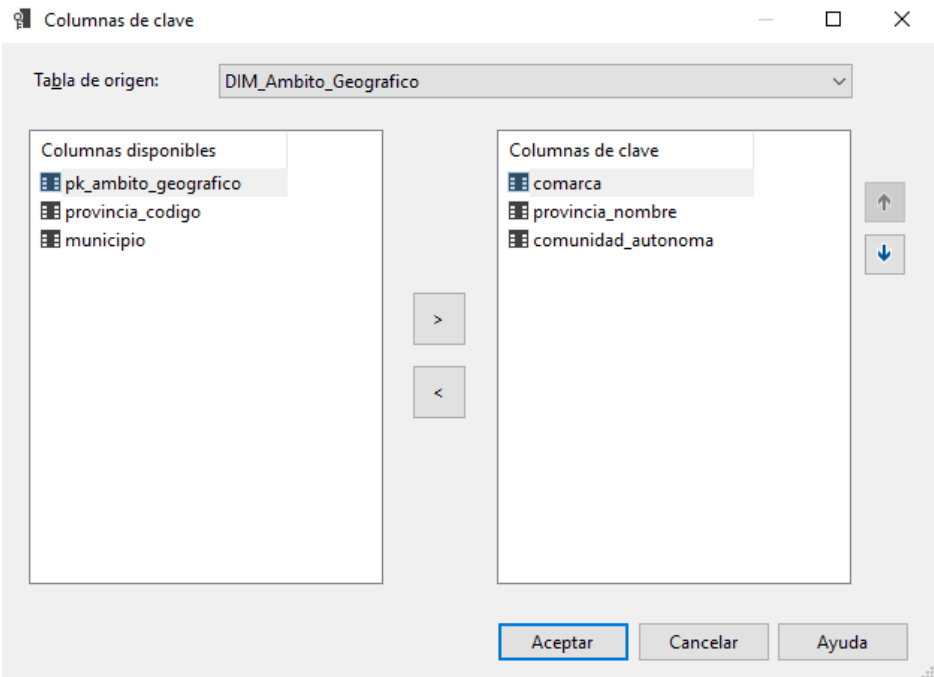


Ilustración 39 - KeyColumns comarca.

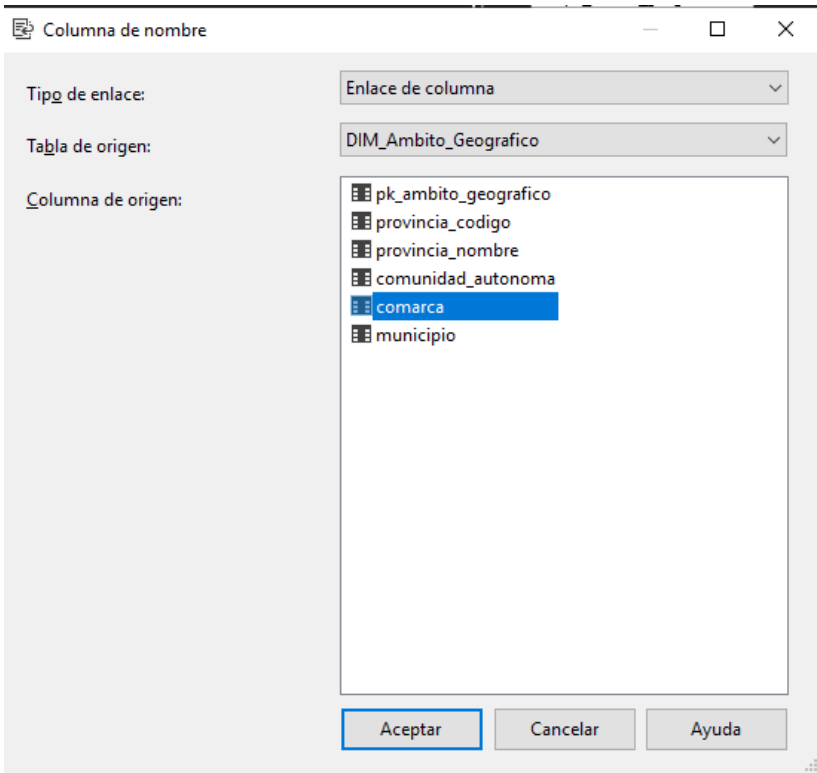


Ilustración 40 - NameColumn comarca.

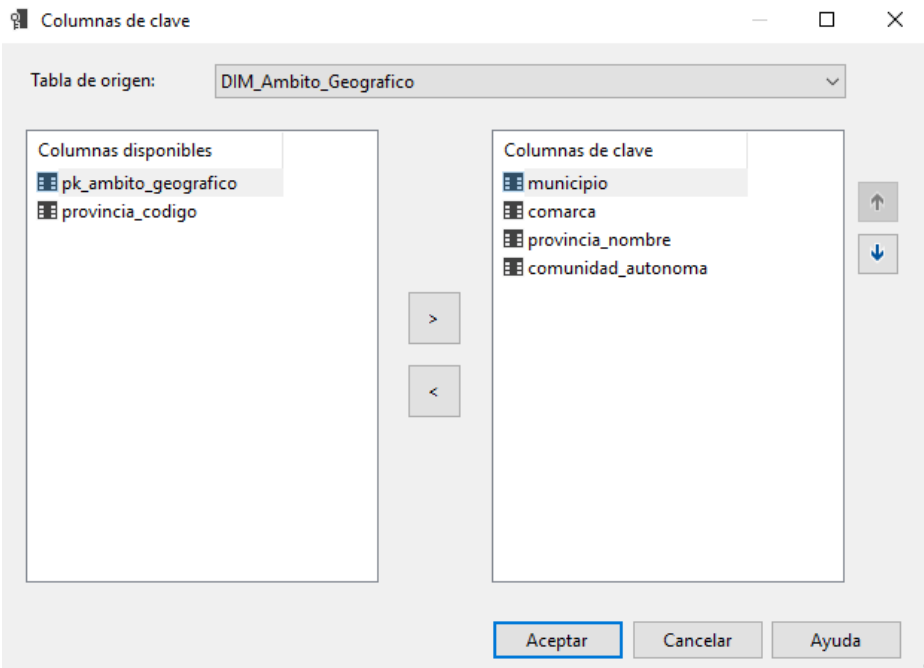


Ilustración 41 - KeyColumns municipio.

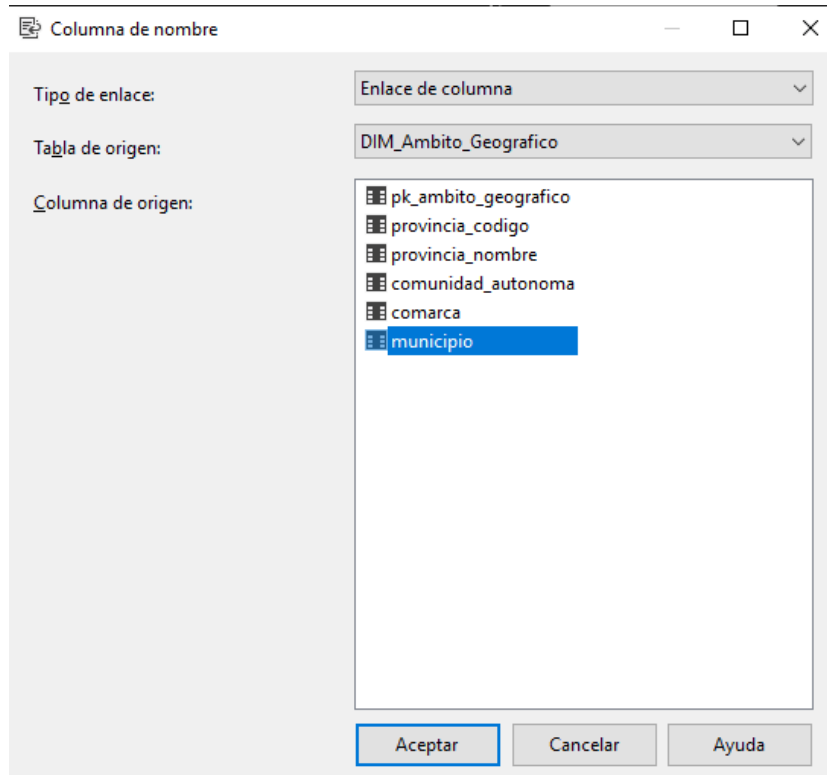


Ilustración 42 - NameColumn municipio.

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:

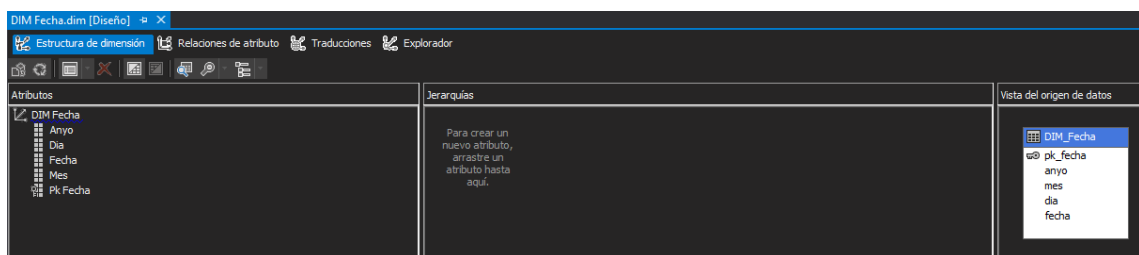


Ilustración 43 - DIM_Fecha.

Al igual que sucedía con la dimensión anterior, definimos una jerarquía para facilitar la construcción de las consultas después. La jerarquía va a ser “Anyo” > “Mes” > “Dia”:

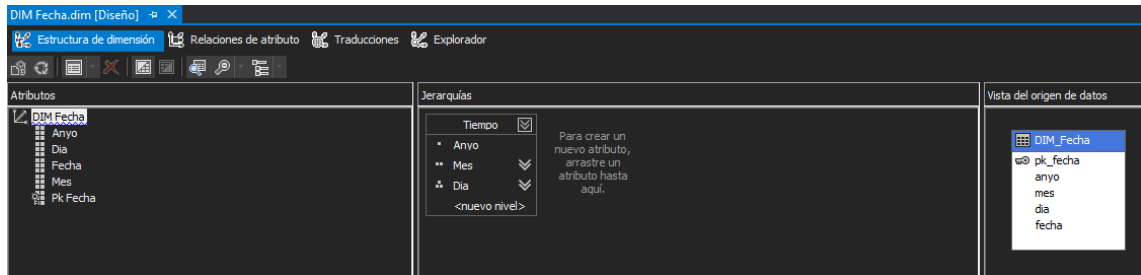


Ilustración 44 - Jerarquía DIM_Fecha.

Cuando definimos la jerarquía tenemos que cambiar las propiedades “*KeyColumns*” y “*NameColumn*” al igual que sucedía con la anterior dimensión.

Por lo tanto, vamos a definir dichas propiedades para los atributos “*Mes*” y “*Dia*” que se encuentran dentro de la dimensión:

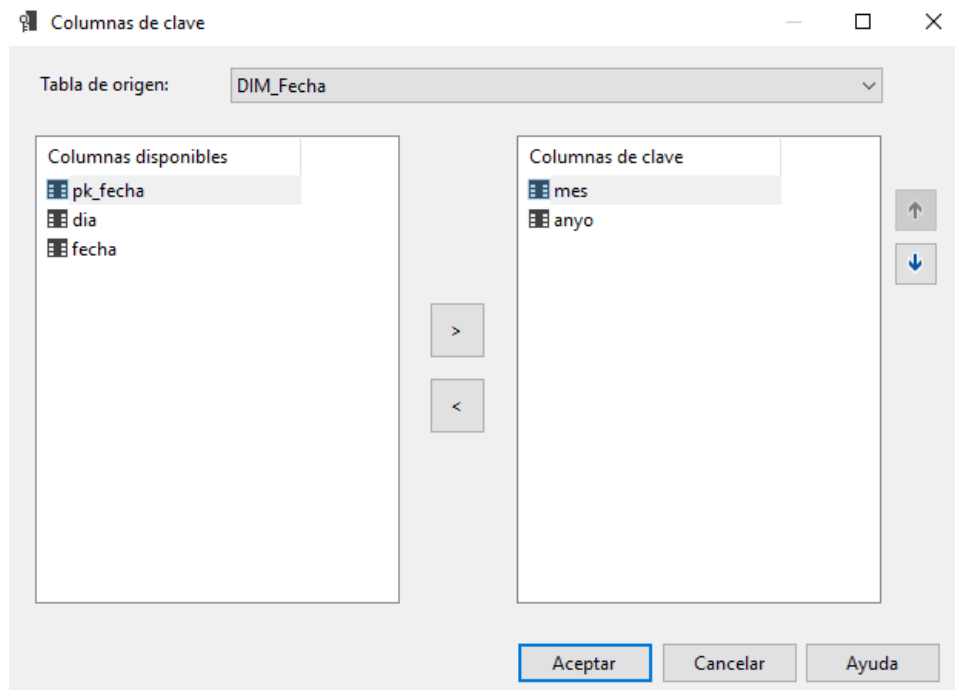


Ilustración 45 - KeyColumns mes.

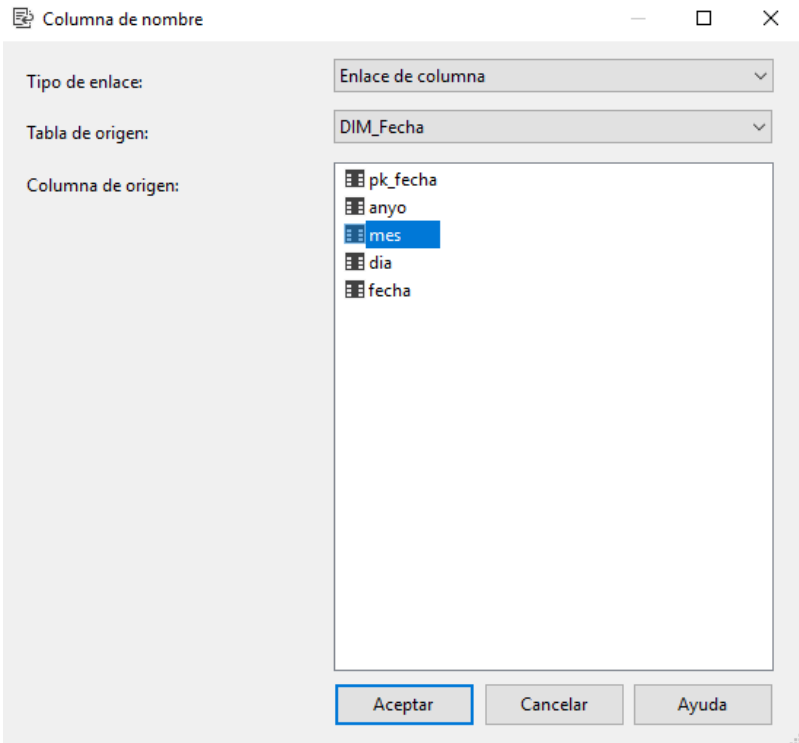


Ilustración 46 - NameColumn mes.

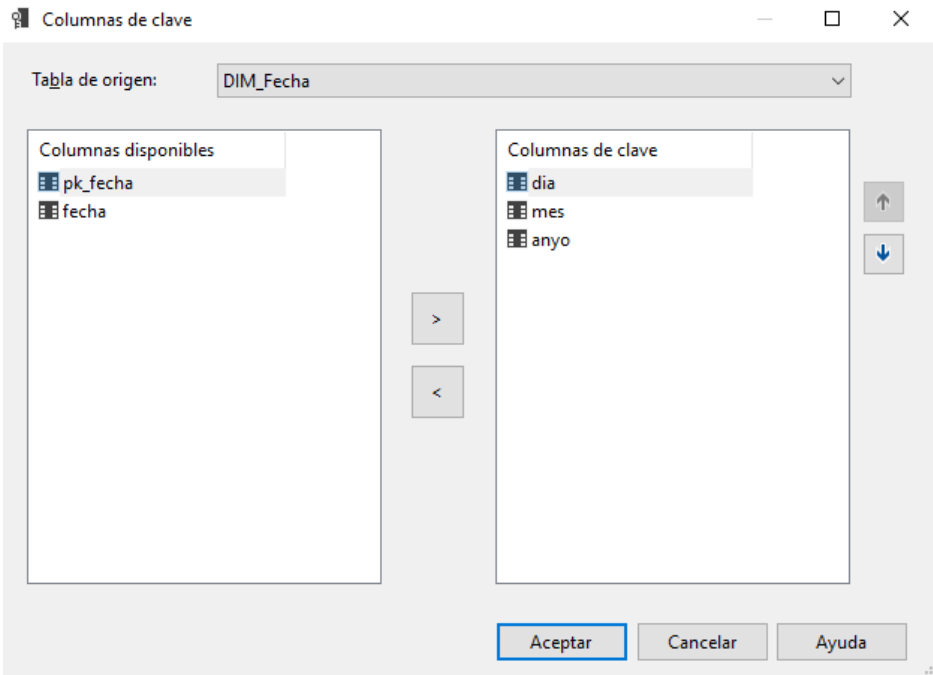
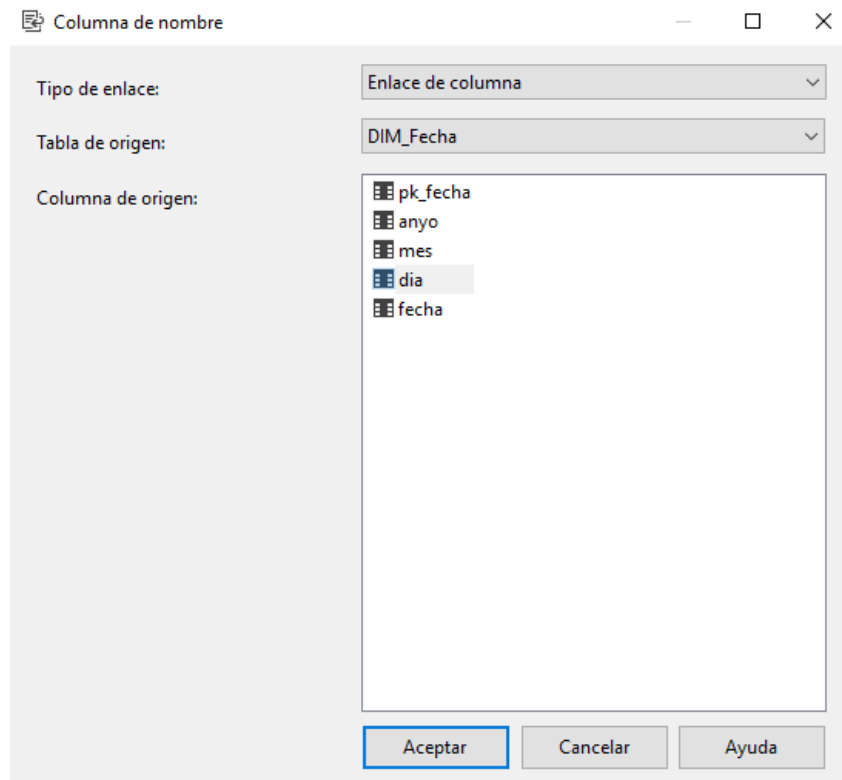
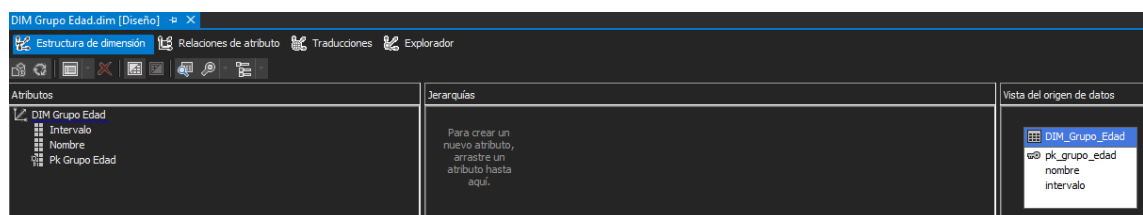


Ilustración 47 - KeyColumns dia.

*Ilustración 48 - NameColumn dia.*

2.5.3. DIM_Grupo_Edad

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

*Ilustración 49 - DIM_Grupo_Edad.*

2.5.4. DIM_Medicion

Realizamos el mismo proceso para “DIM_Medicion”:

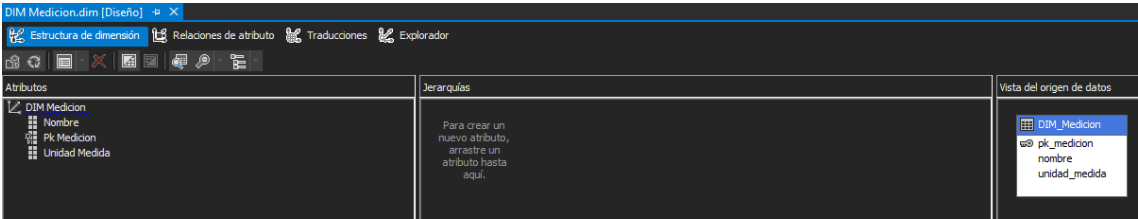


Ilustración 50 - DIM_Medicion.

2.5.5. DIM_Tipologia

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

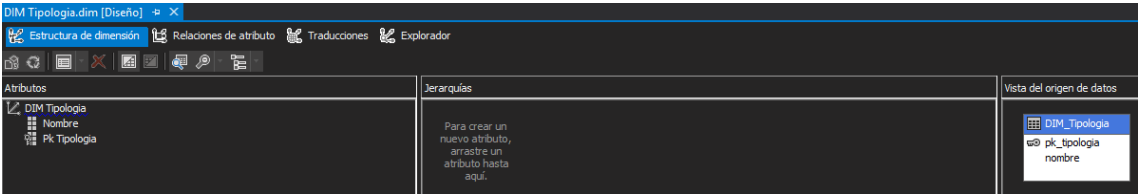


Ilustración 51 - DIM_Tipologia.

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, nos dirigimos al menú “Compilar” y seleccionamos la opción “Implementar solución”:

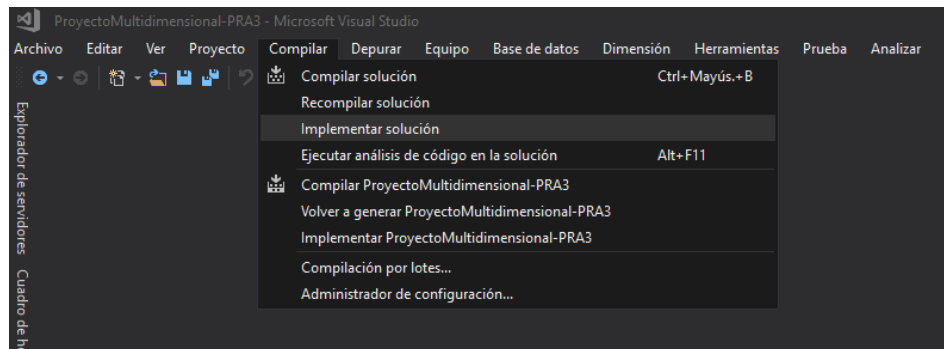


Ilustración 52 - Implementar la solución.

Como resultado de implementar la solución obtenemos la siguiente ventana:

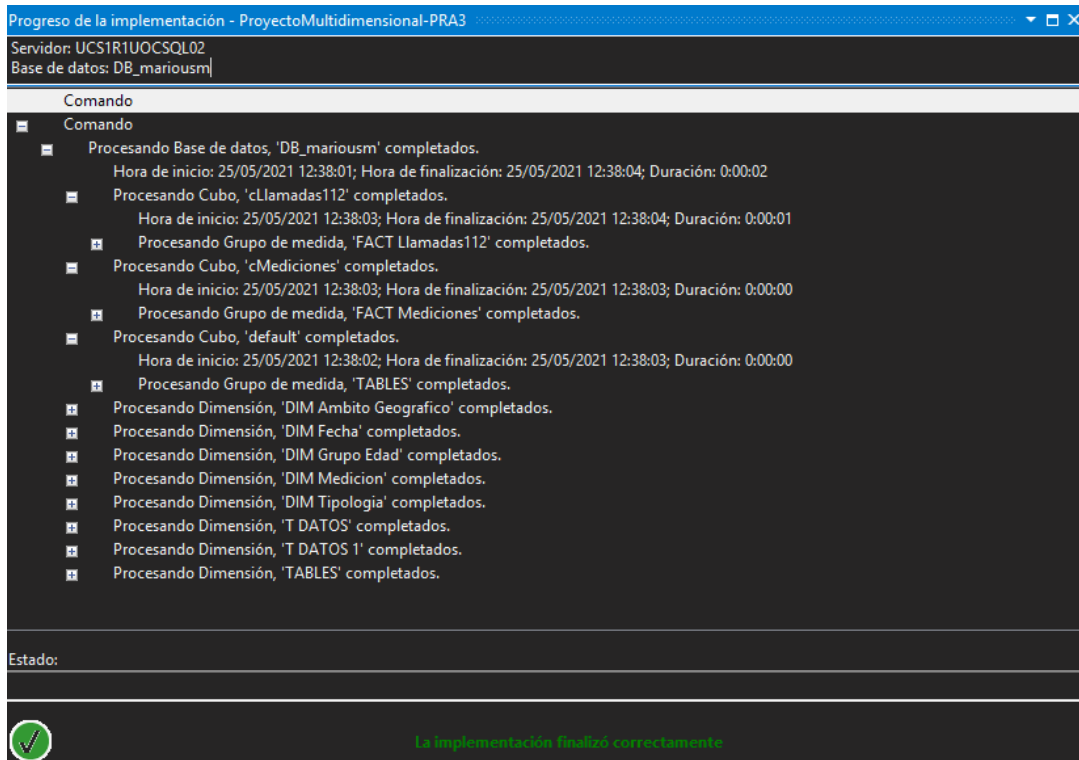


Ilustración 53 - Implementación de la solución.

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.
- Uso de relaciones rígidas en lugar de flexibles, tanto en “DIM Fecha” como “DIM Ambito Geografico”.
- Ocultación de atributos si se usan jerarquías, tanto en “DIM Fecha” como “DIM Ambito Geografico”.
- No existe una relación entre atributos entre uno o más niveles de jerarquía, tanto en “DIM Fecha” como “DIM Ambito Geografico”.
- Crear jerarquías en dimensiones secundarias no principales, esto sucede para “DIM Grupo Edad”, “DIM Medicion”, “DIM Tipologia”.

Para solventar la dimensión temporal 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”:

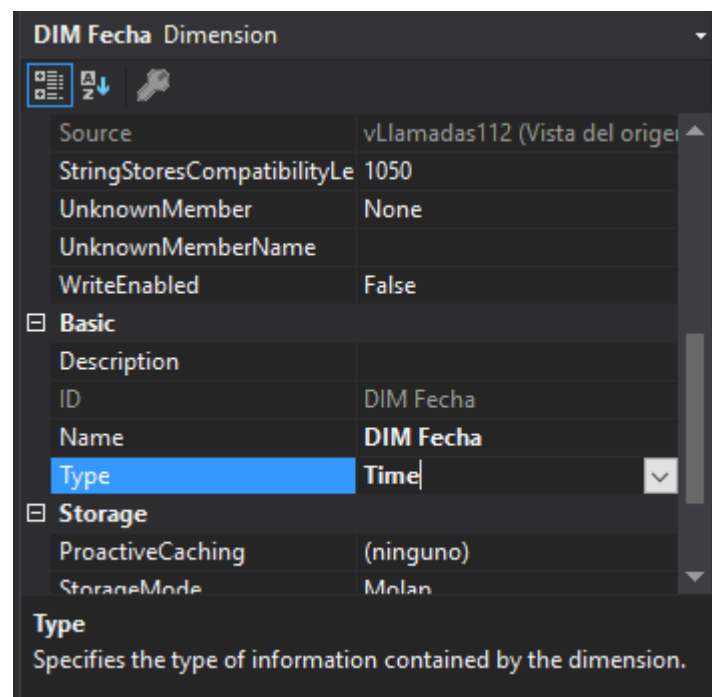


Ilustración 54 - Creación de la dimensión temporal.

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:

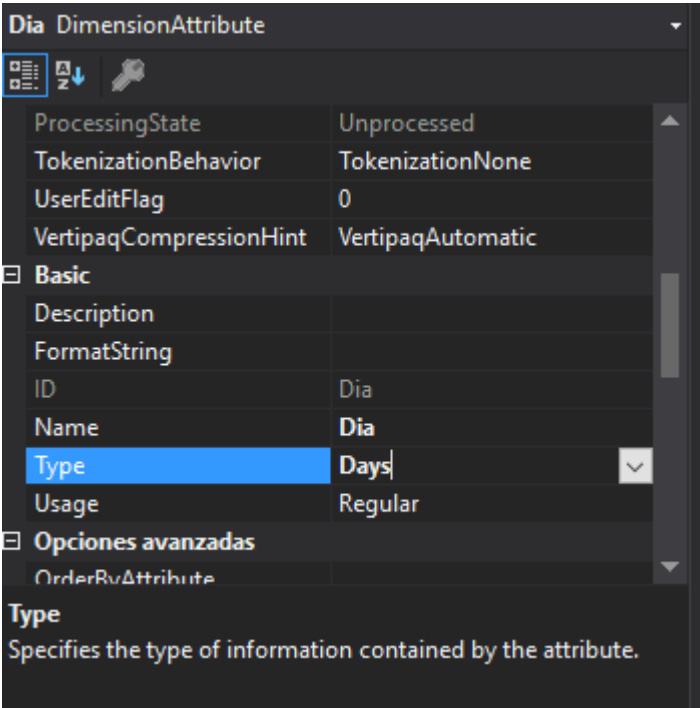


Ilustración 55 - Asignación del tipo para el día.

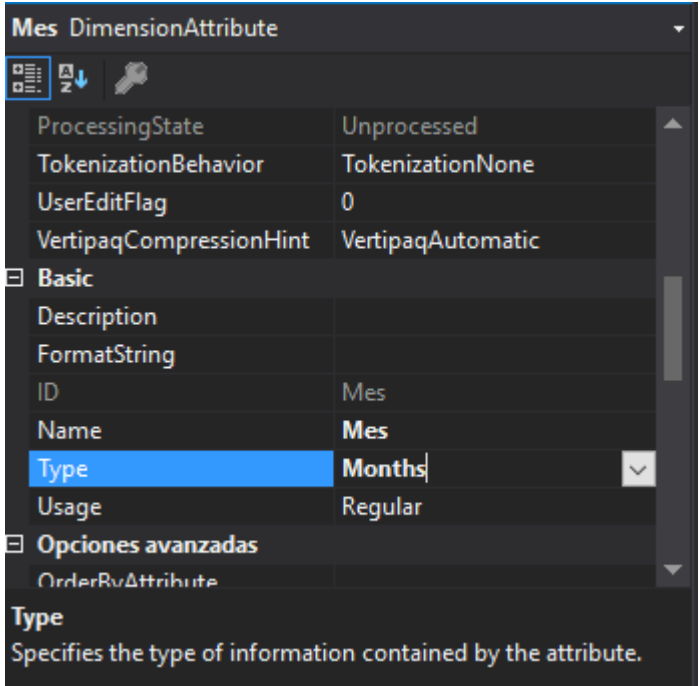


Ilustración 56 - Asignación del tipo para el mes.

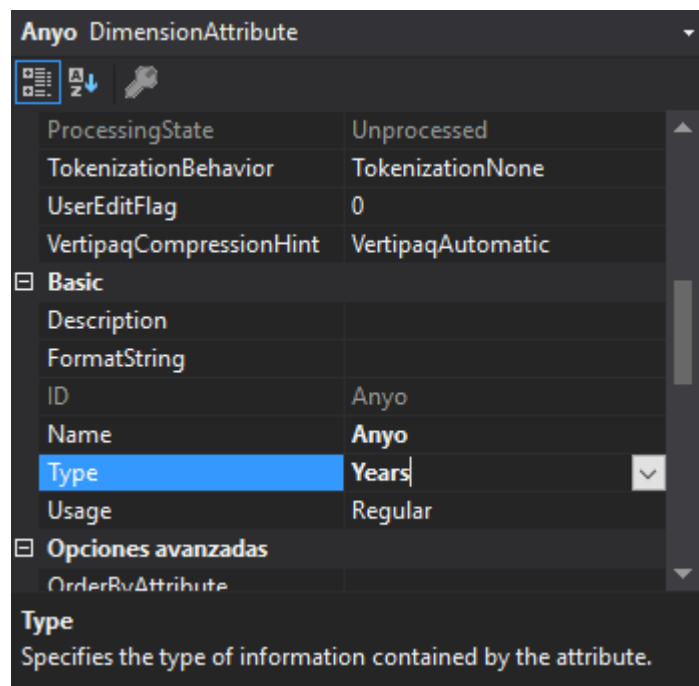


Ilustración 57 - Asignación del tipo para el año.

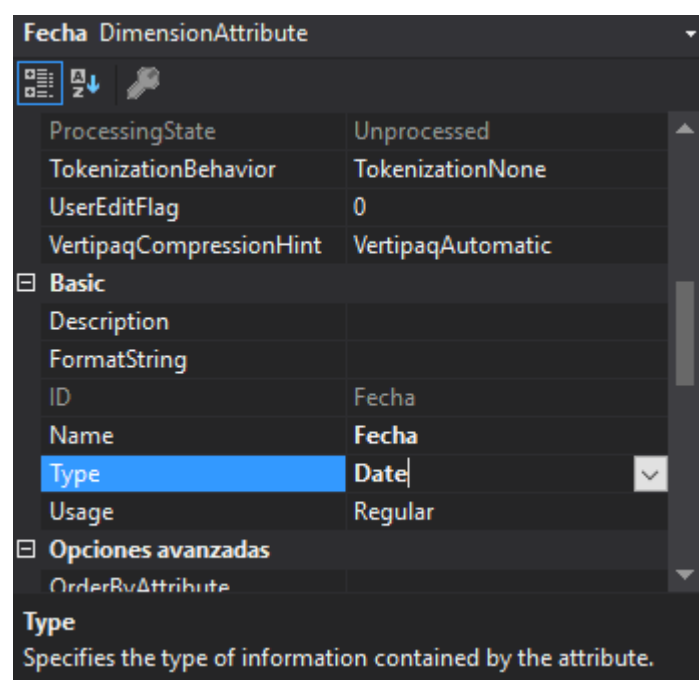


Ilustración 58 - Asignación del tipo para la fecha.

Ahora tenemos que solventar los *warnings* relativos a las dimensiones “DIM Fecha” y “DIM Ambito Geografico”. Lo primero de todo es analizar la recomendación de ocultar atributos si se usan jerarquías, para ello podemos cambiar la visibilidad de los atributos,

pero en nuestro caso no vamos a realizar este cambio ya que así podemos hacer uso de ellos directamente en la explotación de los datos.

La siguiente recomendación es definir una relación entre atributos entre uno o más niveles de jerarquía, para ello definimos dicha relación tanto en la dimensión “DIM Fecha” como “DIM Ambito Geografico”, como aporte cabe destacar que en diseño de las dimensiones se ha explicado la jerarquía y el cómo la hemos definido:



Ilustración 59 - Relación de atributos DIM Ambito Geografico.



Ilustración 60 - Relación de atributos DIM Fecha.

Finalmente, para terminar con las advertencias de estas dimensiones tenemos que definir relaciones rígidas en vez de flexibles entre los atributos de las jerarquías, a modo de ejemplo se muestra el cómo se define para un atributo:

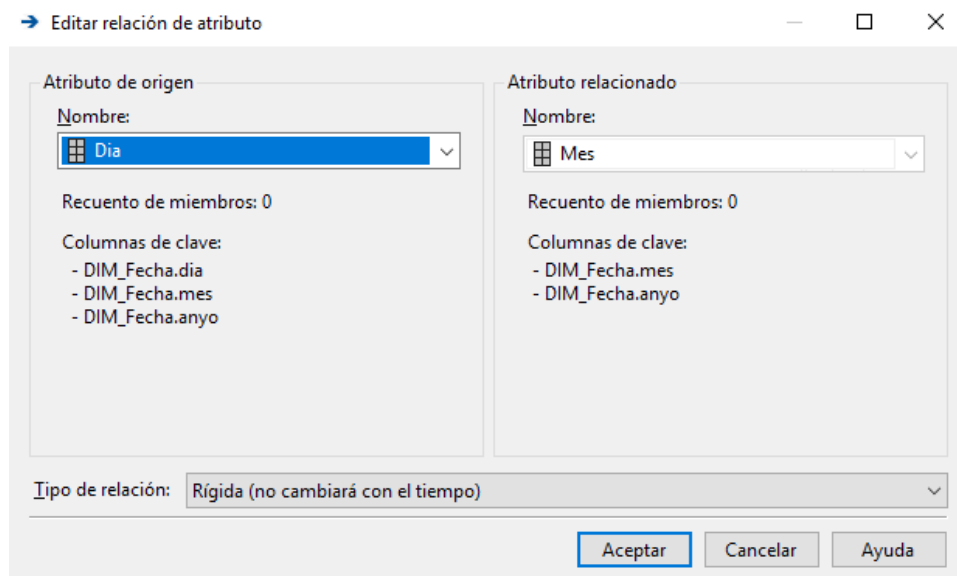


Ilustración 61 - Establcer la relación de flexible a rígida.

Como podemos apreciar, una vez seleccionada la relación modificamos abajo el tipo de relación.

Para terminar con este punto mostramos el cómo quedan finalmente las relaciones rígidas tanto en “DIM Fecha” como “DIM Ambito Geografico”:



Ilustración 62 - Relaciones rígidas en DIM Ambito Geografico.

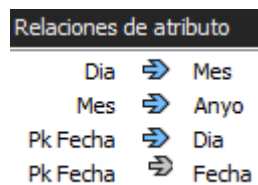


Ilustración 63 - Relaciones rígidas en DIM Fecha.

Cabe destacar, que solo hemos definidos como relaciones rígidas aquellas en los que los atributos pertenecían a la jerarquía, ya que los que no pertenecen no es necesario definir dicha relación como rígida.

Para terminar con una correcta implementación, cabe mencionar que nos salta un *warning* para las dimensiones “DIM Grupo Edad”, “DIM Medicion” y “DIM Tipologia”, esta recomendación es la de definir una jerarquía, pero en nuestro caso no nos hace falta ya que dentro de cada dimensión tenemos relaciones uno a uno y no uno a varios, por lo que todos los valores son únicos y podemos hacer uso de ellos directamente para la explotación de los datos.

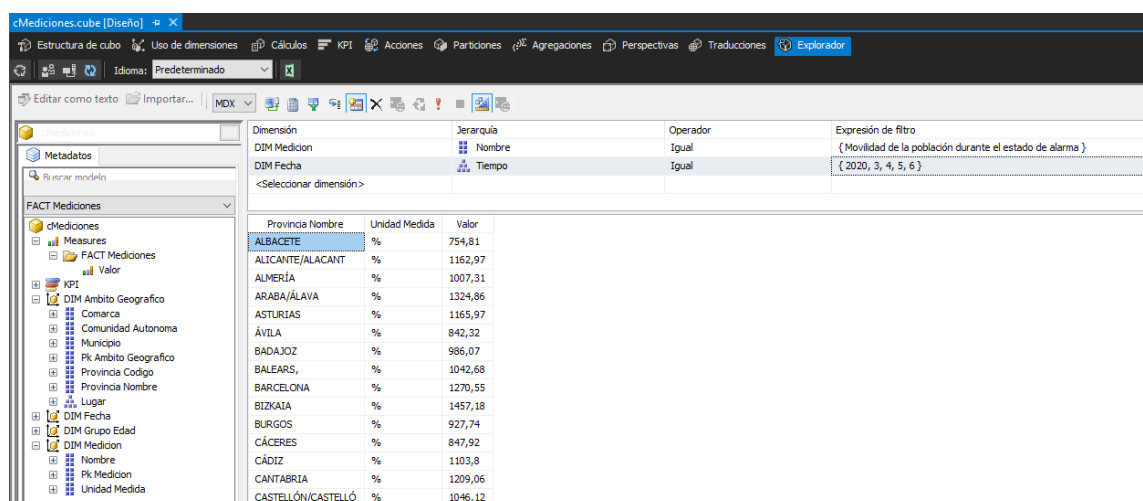
4. Explotación de la solución

Una vez que tenemos el diseño de los cubos y los hemos creado, podemos hacer uso de ellos para obtener información sobre los datos.

Para realizar cada una de las consultas que se van a ver a continuación nos tenemos que dirigir a la pestaña “Explorador”, la cual se encuentra dentro de la ventana del diseño del cubo correspondiente.

4.1. Análisis de las provincias con mayor porcentaje de movilidad según datos móviles

Para realizar esta consulta lo primero que debemos de indicar es el filtrado de la dimensión “DIM_Medidas” con el atributo “Movilidad de la población durante el estado de alarma”, para así definir que lo que vamos a analizar es el porcentaje de movilidad de la población durante el estado de alarma. Además, debemos de filtrar las fechas, es decir, que solo vamos a estudiar la movilidad durante el estado de alarma, para ello seleccionamos el año 2020 y los meses 3, 4, 5 y 6:



The screenshot shows the QlikView Explorer interface. On the left, the 'FACT Mediciones' model is expanded, showing dimensions like 'DIM Ambito Geografico' and 'DIM Fecha'. The main area displays a table with the following data:

Provincia Nombre	Unidad Medida	Valor
ALBACETE	%	754,81
ALICANTE/ALACANT	%	1162,97
ALMERÍA	%	1007,31
ARABA/ÁLAVA	%	1324,86
ASTURIAS	%	1165,97
ÁVILA	%	842,32
BADAJOS	%	986,07
BALEARS,	%	1042,68
BARCELONA	%	1270,55
BIZKAIA	%	1457,18
BURGOS	%	927,74
CÁCERES	%	847,92
CÁDIZ	%	1103,8
CANTABRIA	%	1209,06
CASTELLÓN/CASTELLÓ	%	1046,12

Ilustración 64 - Análisis de las provincias con mayor porcentaje de movilidad.

Al ejecutar la anterior consulta, vemos que el porcentaje que obtenemos es mayor que el 100%, esto se debe a que está haciendo un sumatorio y no está haciendo la media, por lo que creamos un nuevo campo a partir del campo “Valor”, y éste se encarga de hacer la media:

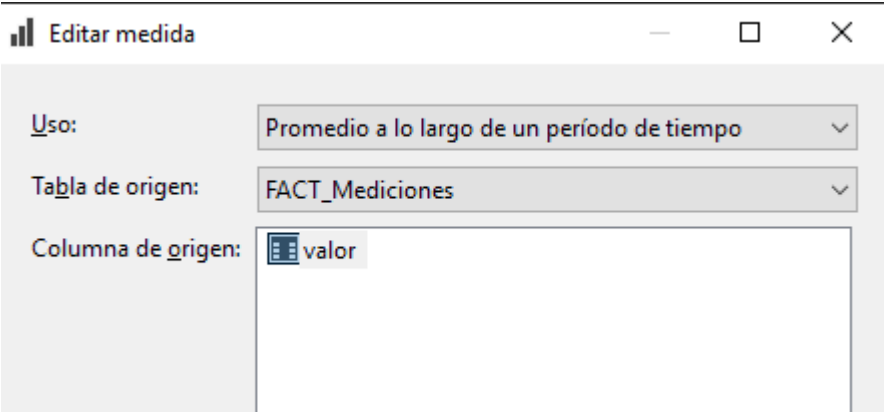


Ilustración 65 - Creación de la media.

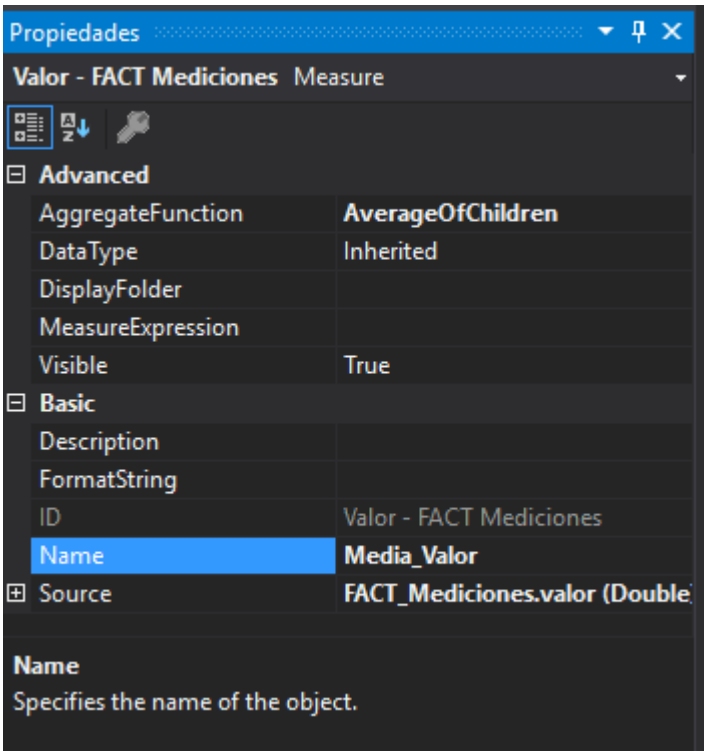


Ilustración 66 - Creación de la media.

Ahora volvemos a crear la consulta con el mismo filtrado que antes tanto para el ámbito geográfico, como para la fecha, con la excepción de que vamos a visualizar el nuevo campo creado “Media_Valor” en vez de “Valor”:

4.2. Análisis del porcentaje de la población que evitaba aglomeraciones según la comunidad autónoma

Lo primero que debemos hacer es filtrar la medición para que solo tenga en cuenta el atributo “Porcentaje de la población que evitaba aglomeraciones”, por otro lado, seleccionamos tanto la comunidad como las métricas para obtener la información:

Comunidad Autónoma	Valor	Media_valor
ANDALUCÍA	2204,67	2204,67
ARAGON	839,17	839,17
CANARIAS	575,24	575,24
CANTABRIA	266,97	266,97
CASTILLA Y LEÓN	2491,45	2491,45
CASTILLA-LA MANCHA	1355,06	1355,06
CATALUÑA	1096,8	1096,8
COMUNIDAD DE MA...	269,6	269,6
COMUNIDAD FORAL ...	287,9	287,9
COMUNIDAD VALEN...	824,37	824,37
EXTREMADURA	546,15	546,15
GALICIA	1047,69	1047,69
ISLAS BALEARES	275,54	275,54
LA RIOJA	237,85	237,85
PAÍS VASCO	767,24	767,24
PRINCIPADO DE AST...	255,32	255,32
REGIÓN DE MURCIA	293,37	293,37

Ilustración 69 - Análisis del porcentaje de población que evitaba aglomeraciones según la CCAA.

Como podemos apreciar, sí que nos devuelve la información pero la media no la está realizando correctamente, ya que dicha métrica realiza el promedio a lo largo de un periodo de tiempo, en nuestro caso, dicho periodo es siempre el mismo 30 de septiembre del 2020.

Por lo tanto, necesitamos hacer el cálculo del valor entre (número de provincias de la CCAA * 6 (tenemos 6 grupos de edad)), para ello definimos una nueva métrica que va a ser el número de filas, ésta nos dará como resultado el valor de número de provincias de la CCAA * 6 grupos de edad:

Editar medida

Uso: Número de filas

Tabla de origen: FACT_Mediciones

Columna de origen: valor

Ilustración 70 - Creación número de filas.

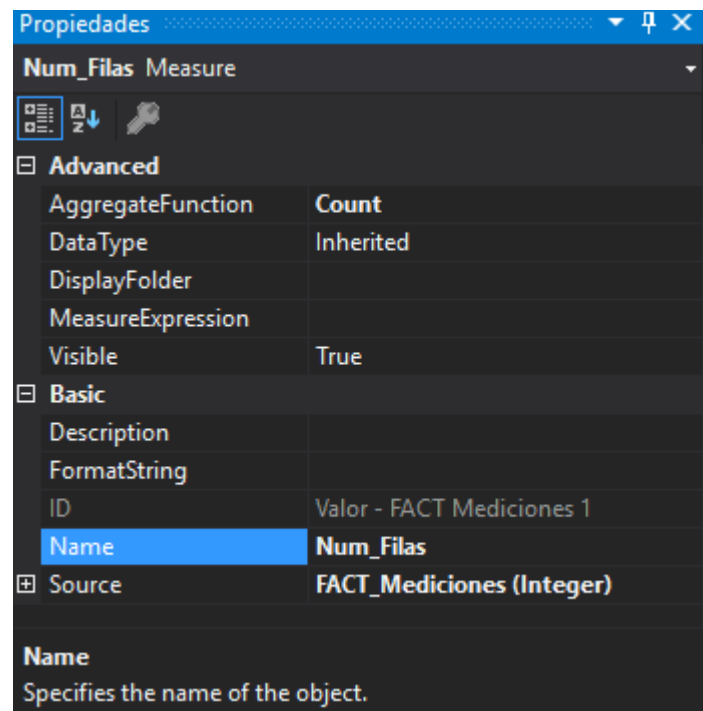


Ilustración 71 - Creación número de filas.

De tal forma que, si volvemos a ejecutar la consulta anterior, obtenemos el siguiente resultado:

Comunidad Autonoma	Valor	Media_Valor	Num_Filas
ANDALUCÍA	220...	2204,67	48
ARAGON	839...	839,17	18
CANARIAS	575...	575,24	12
CANTABRIA	266...	266,97	6
CASTILLA Y LEÓN	249...	2491,45	54
CASTILLA-LA MANCHA	135...	1355,06	30
CATALUÑA	109...	1096,8	24
COMUNIDAD DE MA...	269,6	269,6	6
COMUNIDAD FORAL ...	287,9	287,9	6
COMUNIDAD VALEN...	824...	824,37	18
EXTREMADURA	546...	546,15	12
GALICIA	104...	1047,69	24
ISLAS BALEARES	275...	275,54	6
LA RIOJA	237...	237,85	6
PAÍS VASCO	767...	767,24	18
PRINCIPADO DE AST...	255...	255,32	6
REGIÓN DE MURCIA	293...	293,37	6

Ilustración 72- Análisis del porcentaje de población que evitaba aglomeraciones según la CCAA.

Ahora lo que nos faltaría es dividir el “Valor” entre “Num_Filas”, de esa forma tendríamos unos datos coherentes relativos al porcentaje de población que evitaba aglomeraciones por cada comunidad. Para ello, creamos un miembro calculado el cual llamamos “Media_Pob_Aglo”:

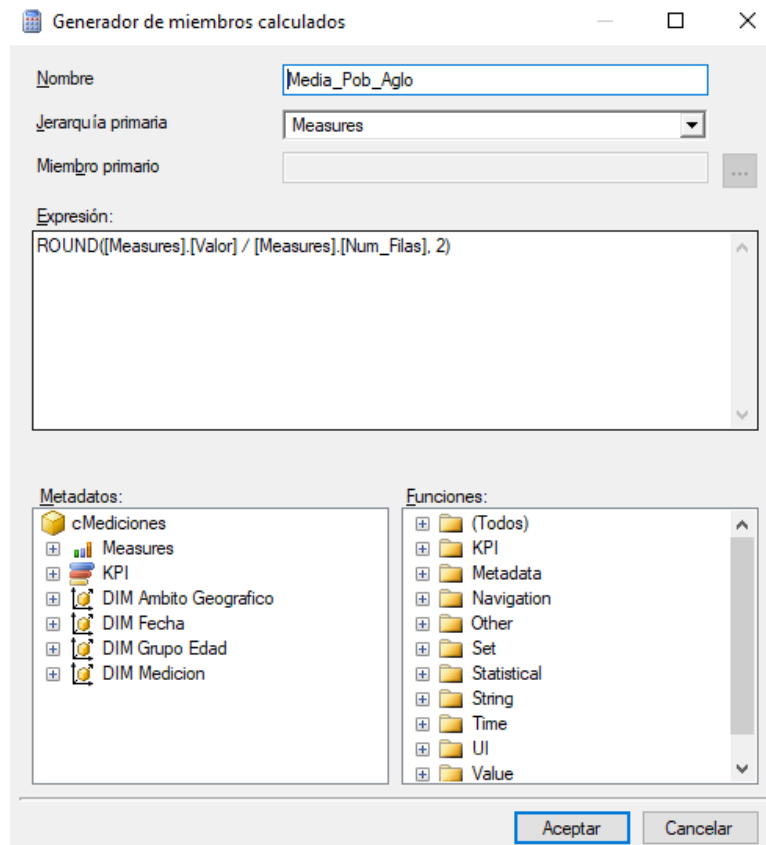


Ilustración 73 - Creación media población aglomeración.

Cuando ya tenemos el nuevo campo creado lo incluimos en el diseño de la consulta, de esta forma nos proporciona el siguiente resultado:

The screenshot shows the 'cMediciones.cube [Diseño]' window. The left pane displays the metadata tree with 'Mediciones' expanded, showing 'FACT Mediciones' and 'Medidas'. The main area shows a cube design with dimensions 'DIM Medicion' and 'Jerarquía' (Nombre). The 'Operador' is set to 'Igual' and the 'Expresión de filtro' is '{ Porcentaje de la población que evitaba aglomeraciones }'. Below this, a table displays the results for various autonomous communities (CCAA).

Comunidad Autónoma	Unidad Medida	Valor	Num_Filas	Media_Pob_Aglo
ANDALUCÍA	%	2204,67	48	45,93
ARAGON	%	839,17	18	46,62
CANARIAS	%	575,24	12	47,94
CANTABRIA	%	266,97	6	44,5
CASTILLA Y LEÓN	%	2491,45	54	46,14
CASTILLA-LA MANCHA	%	1355,06	30	45,17
CATALUÑA	%	1096,8	24	45,7
COMUNIDAD DE MADRID	%	269,6	6	44,93
COMUNIDAD FORAL DE NAVARRA	%	287,9	6	47,98
COMUNIDAD VALENCIANA	%	824,37	18	45,8
EXTREMADURA	%	546,15	12	45,51
GALICIA	%	1047,69	24	43,65
ISLAS BALEARES	%	275,54	6	45,92
LA RIOJA	%	237,85	6	39,64
PAÍS VASCO	%	767,24	18	42,62
PRINCIPADO DE ASTURIAS	%	255,32	6	42,55
REGIÓN DE MURCIA	%	293,37	6	48,9

Ilustración 74 - Análisis del porcentaje de población que evitaba aglomeraciones según la CCAA.

Del análisis de la anterior consulta podemos concluir que está bastante nivelado el porcentaje de población que evitaba aglomeraciones según la comunidad autónoma. Cabe destacar que la comunidad autónoma que más porcentaje de población evitaba las aglomeraciones era la Región de Murcia con un 48.9%, sin embargo, La Rioja es la peor comunidad con un 39.64%, estos datos son significativos porque vemos que da igual si la comunidad autónoma es más grande o no, ya que La Rioja y la Región de Murcia son similares en cuanto a extensión y sus porcentajes son muy dispares.

Por otro lado, aunque no hay mucha diferencia entre las comunidades sí que podemos ver una ligera tendencia a que cuanto menor población en la comunidad autónoma menor es también el porcentaje de población que evitaba aglomeraciones.

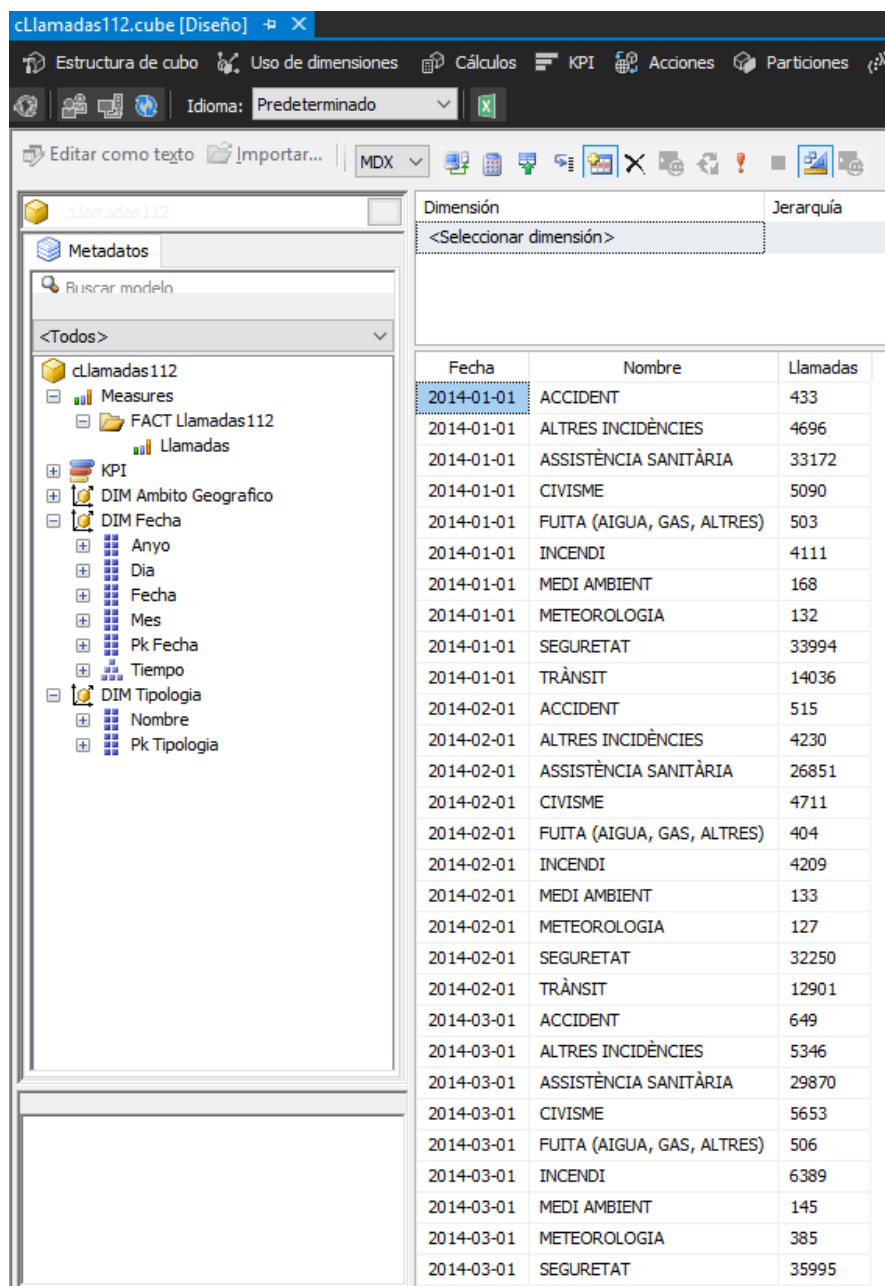
4.3. Análisis del promedio de sanciones por habitantes

A

4.4. Evolución de las llamadas de urgencia al 112 en Cataluña por tipología de llamada

Para realizar un análisis evolutivo lo que debemos de tener en cuenta es la dimensión del tiempo. Por otro lado, nos pide que se quiere analizar el número de llamadas según la tipología de la misma.

Por lo tanto, a la hora de hacer la consulta en el cubo vamos a añadir la fecha, el nombre de la tipología y la medida para obtener el número de llamadas:



The screenshot shows the QlikView interface for the cube 'cLlamadas112.cube'. The left pane displays the metadata tree with the following structure:

- cLlamadas112
 - Measures
 - FACT Llamadas112
 - Llamadas
 - KPI
 - DIM Ambito Geografico
 - DIM Fecha
 - Anyo
 - Dia
 - Fecha
 - Mes
 - Pk Fecha
 - Tiempo
 - DIM Tipologia
 - Nombre
 - Pk Tipologia

The right pane shows the selected dimensions and a table of data. The dimensions are 'Fecha' and 'Nombre'. The table displays the number of calls ('Llamadas') for each date and category.

Fecha	Nombre	Llamadas
2014-01-01	ACCIDENT	433
2014-01-01	ALTRES INCIDÈNCIES	4696
2014-01-01	ASSISTÈNCIA SANITÀRIA	33172
2014-01-01	CIVISME	5090
2014-01-01	FUITA (AIGUA, GAS, ALTRES)	503
2014-01-01	INCENDI	4111
2014-01-01	MEDI AMBIENT	168
2014-01-01	METEOROLOGIA	132
2014-01-01	SEGURETAT	33994
2014-01-01	TRÀNSIT	14036
2014-02-01	ACCIDENT	515
2014-02-01	ALTRES INCIDÈNCIES	4230
2014-02-01	ASSISTÈNCIA SANITÀRIA	26851
2014-02-01	CIVISME	4711
2014-02-01	FUITA (AIGUA, GAS, ALTRES)	404
2014-02-01	INCENDI	4209
2014-02-01	MEDI AMBIENT	133
2014-02-01	METEOROLOGIA	127
2014-02-01	SEGURETAT	32250
2014-02-01	TRÀNSIT	12901
2014-03-01	ACCIDENT	649
2014-03-01	ALTRES INCIDÈNCIES	5346
2014-03-01	ASSISTÈNCIA SANITÀRIA	29870
2014-03-01	CIVISME	5653
2014-03-01	FUITA (AIGUA, GAS, ALTRES)	506
2014-03-01	INCENDI	6389
2014-03-01	MEDI AMBIENT	145
2014-03-01	METEOROLOGIA	385
2014-03-01	SEGURETAT	35995

Ilustración 75 - Evolución de las llamadas según la tipología.

Como podemos apreciar en la anterior consulta, al incluir el campo fecha nos muestra la evolución del número de llamadas según su tipología para todos los meses que hay desde el 2014 hasta el 2020. Esto nos da como resultado una consulta un poco grande, y como en el enunciado no se pide si la evolución de las llamadas se tiene que estudiar con granularidad de mes o año, a decisión personal se decide estudiar la evolución con granularidad anual ya que nos da un análisis más general, de tal forma sustituimos el campo “Fecha” por “Anyo”:

Anyo	Nombre	Llamadas
2014	ACCIDENT	8533
2014	ALTRES INCIDÈNCIES	71901
2014	ASSISTÈNCIA SANITÀRIA	376123
2014	CIVISME	85664
2014	FUITA (AIGUA, GAS, ALTRES)	7560
2014	INCENDI	60477
2014	MEDI AMBIENT	3220
2014	METEOROLOGIA	6683
2014	SEGURETAT	455698
2014	TRÀNSIT	190412
2015	ACCIDENT	8638
2015	ALTRES INCIDÈNCIES	74831
2015	ASSISTÈNCIA SANITÀRIA	411565
2015	CIVISME	86672
2015	FUITA (AIGUA, GAS, ALTRES)	8597
2015	INCENDI	76085
2015	MEDI AMBIENT	3658
2015	METEOROLOGIA	3609
2015	SEGURETAT	474487
2015	TRÀNSIT	202976
2016	ACCIDENT	9433
2016	ALTRES INCIDÈNCIES	79646
2016	ASSISTÈNCIA SANITÀRIA	449425
2016	CIVISME	101391
2016	FUITA (AIGUA, GAS, ALTRES)	9217
2016	INCENDI	82999
2016	MEDI AMBIENT	3708
2016	METEOROLOGIA	3106
2016	SEGURETAT	498314

Ilustración 76 - Evolución de las llamadas según la tipología.

Anyo	Nombre	Llamadas
2016	TRÀNSIT	223972
2017	ACCIDENT	10773
2017	ALTRES INCIDÈNCIES	86325
2017	ASSISTÈNCIA SANITÀRIA	480785
2017	CIVISME	111845
2017	FUITA (AIGUA, GAS, ALTRES)	9930
2017	INCENDI	93054
2017	MEDI AMBIENT	4419
2017	METEOROLOGIA	5162
2017	SEGURETAT	554716
2017	TRÀNSIT	244363
2018	ACCIDENT	12423
2018	ALTRES INCIDÈNCIES	95147
2018	ASSISTÈNCIA SANITÀRIA	524448
2018	CIVISME	110933
2018	FUITA (AIGUA, GAS, ALTRES)	10407
2018	INCENDI	69598
2018	MEDI AMBIENT	6597
2018	METEOROLOGIA	7597
2018	SEGURETAT	620167
2018	TRÀNSIT	255293
2019	ACCIDENT	12067
2019	ALTRES INCIDÈNCIES	98236
2019	ASSISTÈNCIA SANITÀRIA	555227
2019	CIVISME	113692
2019	FUITA (AIGUA, GAS, ALTRES)	9885
2019	INCENDI	93566
2019	MEDI AMBIENT	6881
2019	METEOROLOGIA	7972

Ilustración 77 - Evolución de las llamadas según la tipología.

2019	SEGURETAT	698847
2019	TRÀNSIT	259845
2020	ACCIDENT	10351
2020	ALTRES INCIDÈNCIES	73824
2020	ASSISTÈNCIA SANITÀRIA	594384
2020	CIVISME	147858
2020	FUITA (AIGUA, GAS, ALTRES)	8262
2020	INCENDI	58871
2020	MEDI AMBIENT	6075
2020	METEOROLOGIA	14355
2020	SEGURETAT	555191
2020	TRÀNSIT	157044

Ilustración 78 - Evolución de las llamadas según la tipología.

Si analizamos los datos de forma general todas las tipologías de llamadas presentan una tendencia al alza, es decir, a medida que va pasando el tiempo se producen un mayor número de llamadas en casi todas las tipologías.

Al estar analizando el impacto del *COVID* en nuestra sociedad, tenemos que fijarnos más detalladamente en las incidencias que se producen en el año 2020, ya que es en éste cuando comienza el problema de dicha enfermedad. Si observamos las incidencias producidas en este año vemos que la gran mayoría van a la baja, es decir, comparando el año 2020 con los demás suele haber un menor número de incidencias. Sin embargo, en este año hay más incidencias respecto a la asistencia sanitaria y de civismo, lo cual tiene todo el sentido del mundo ya que fue un momento difícil a nivel sanitario y también a nivel de comportamiento de la sociedad.