

Universidad Nacional Autónoma de México

Facultad de Ciencias, 2024 - 1

Almacenes y Minería de Datos

Práctica 04. Integración de Datos

PostgresandoesoSQLazos

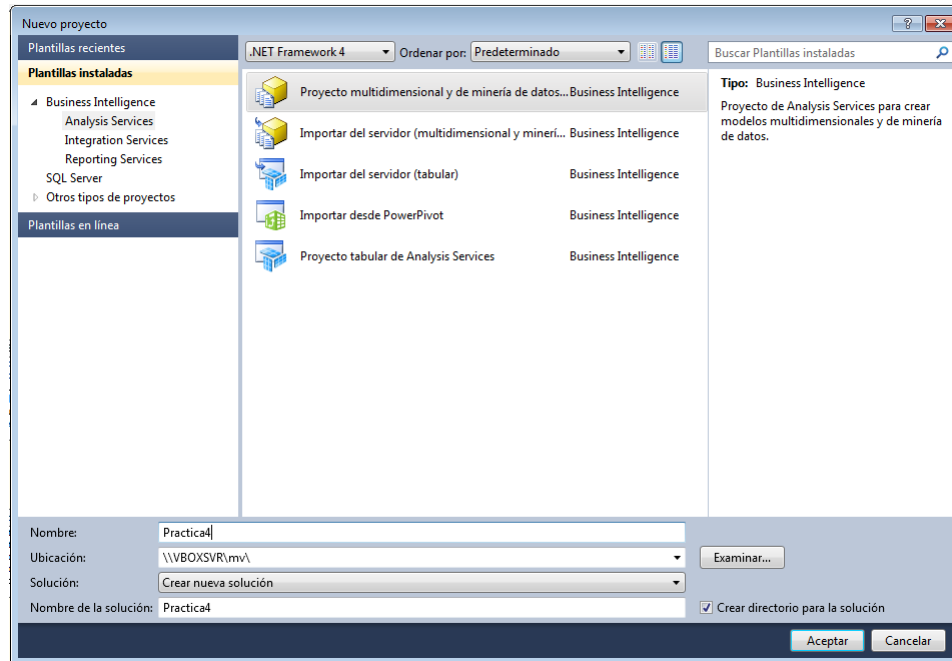


Integrantes

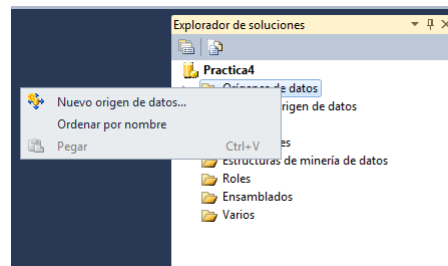
Adrian Aguilera Moreno	421005200
Marco Antonio Rivera Silva	318183583
Sebastián Alejandro Gutiérrez Medina	318287021
Israel Hernández Dorantes	318206604
Alejandra Ortega García	420002495

1. Procedimiento realizado

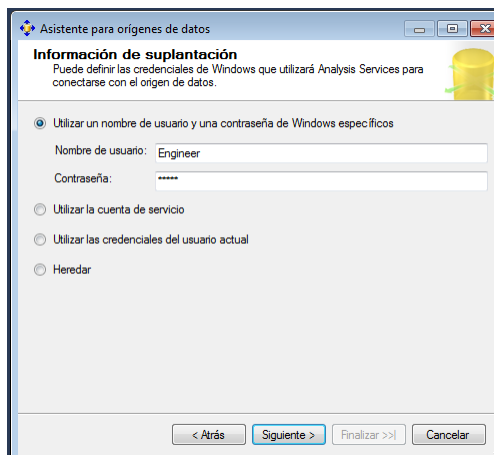
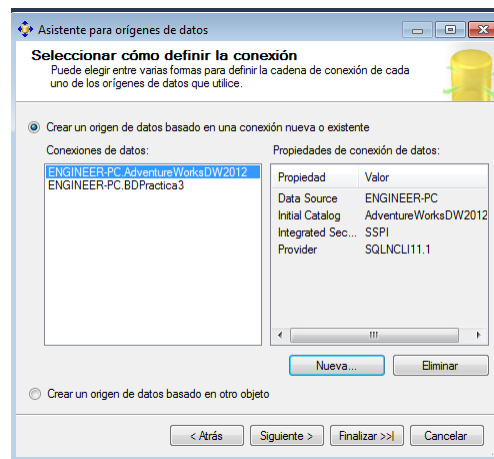
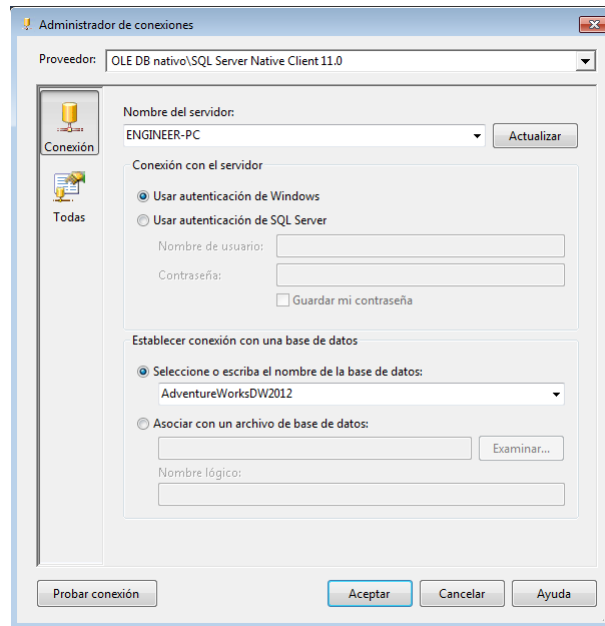
Primero creamos un proyecto multidimensional y de minería de datos, lo llamamos *Practica4*.



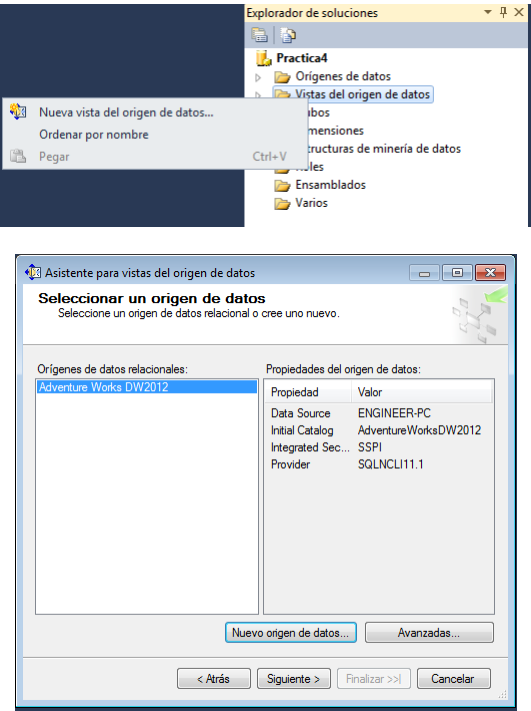
Después comenzamos a agregar un origen de datos para el proyecto.



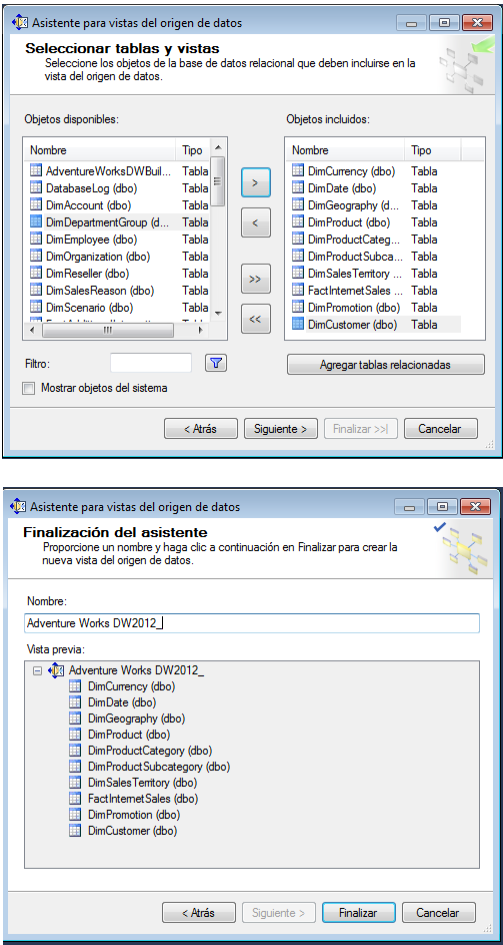
Configuramos la conexión con la BD *AdventureWorksDW2012*.



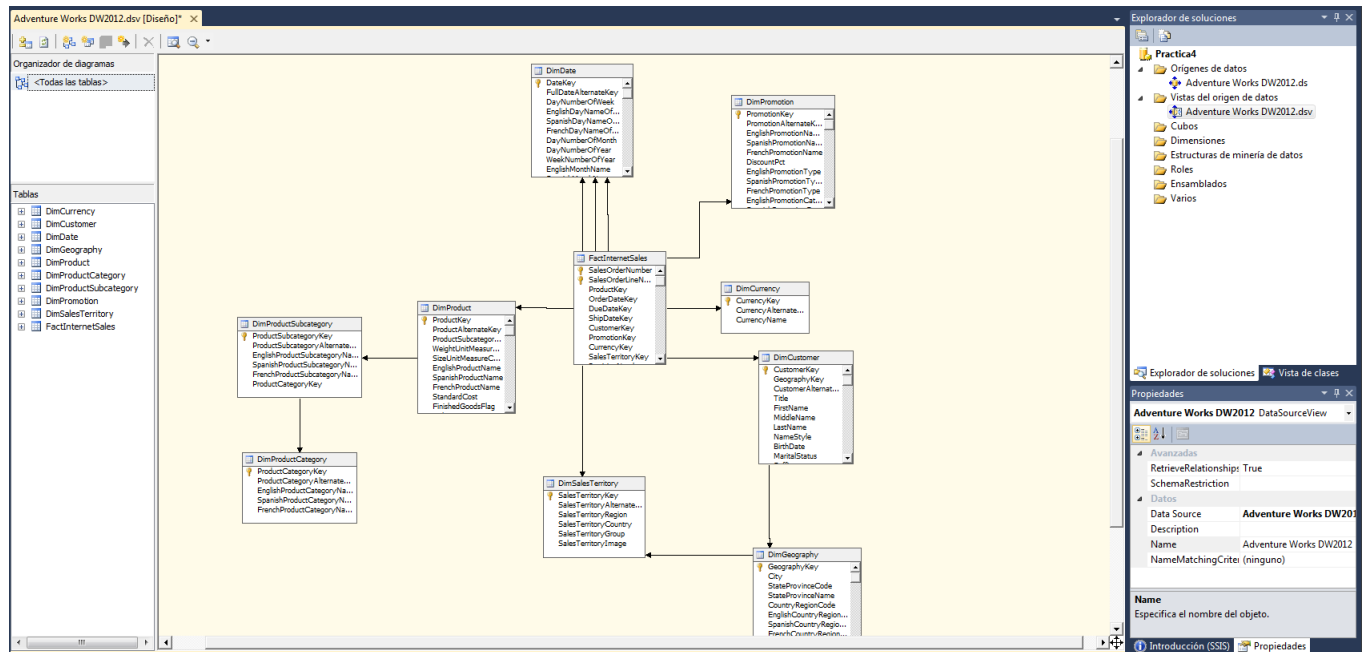
Continuamos con la creación de una vista del origen de datos.



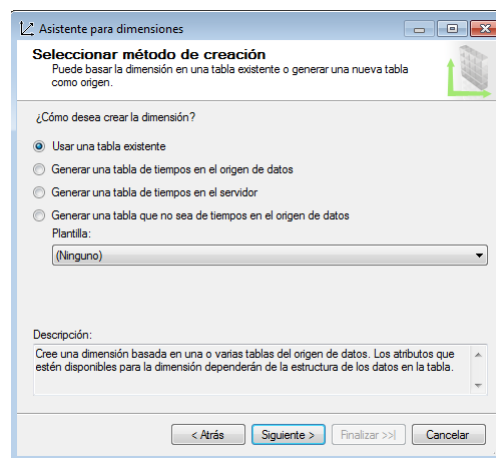
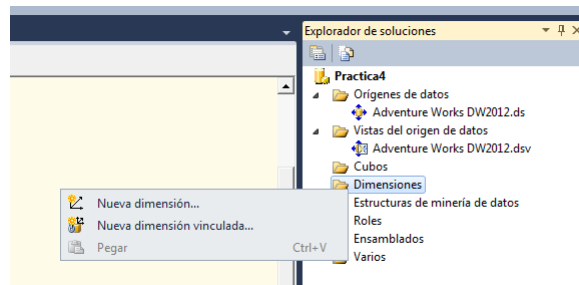
Seleccionamos las tablas que vamos a utilizar.

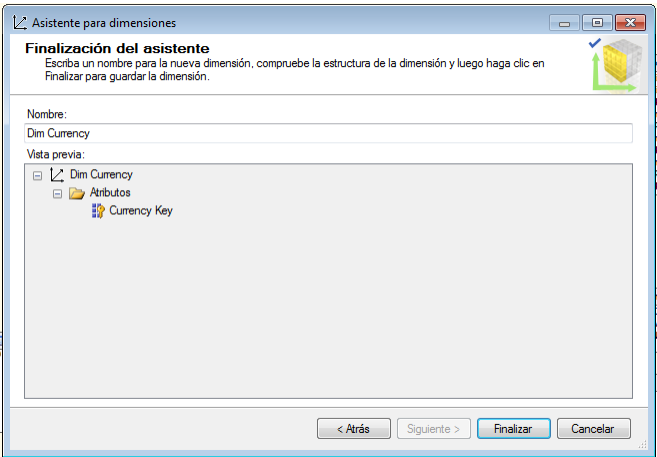
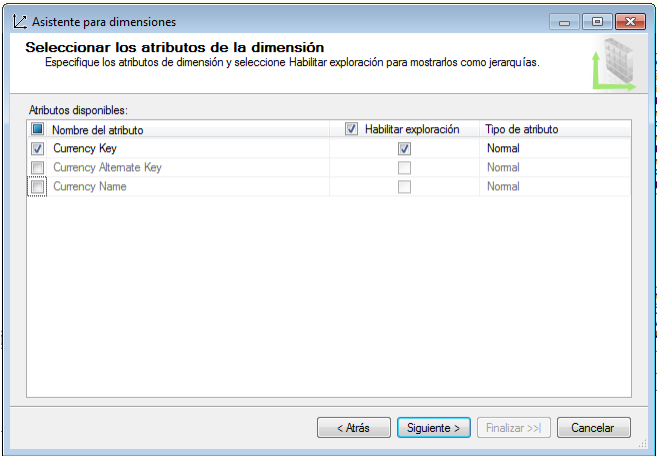


Resultado de configurar la vista.

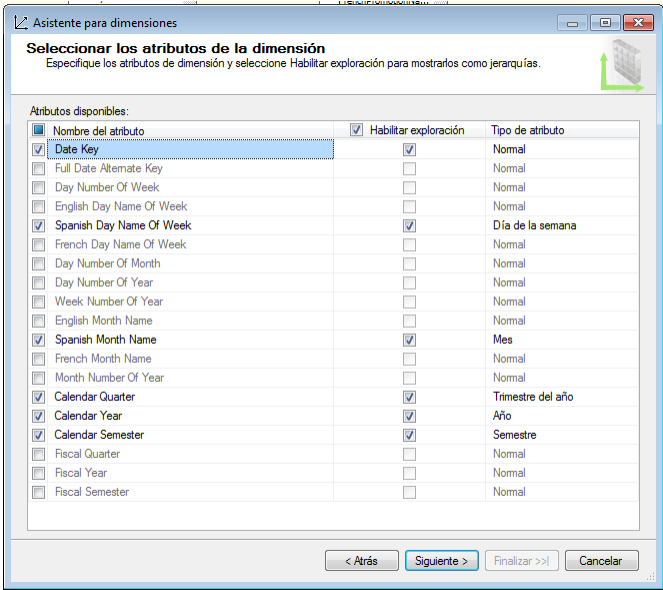


Empezamos a agregar las dimensiones utilizando las tablas existentes, comenzando con la dimensión *Currency*

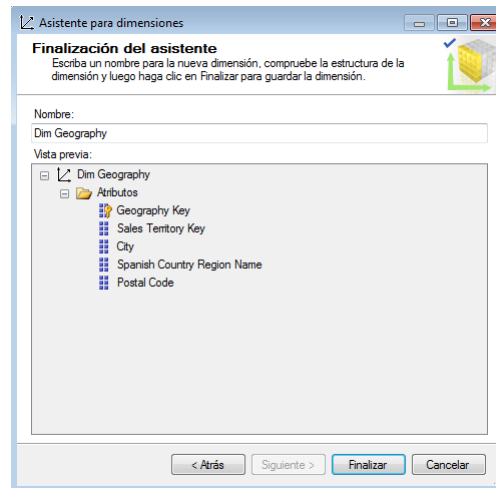




Después con la dimensión Date.



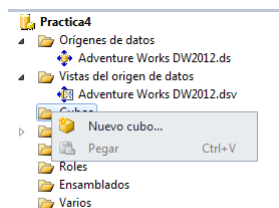
La siguiente dimensión fue la de Geography



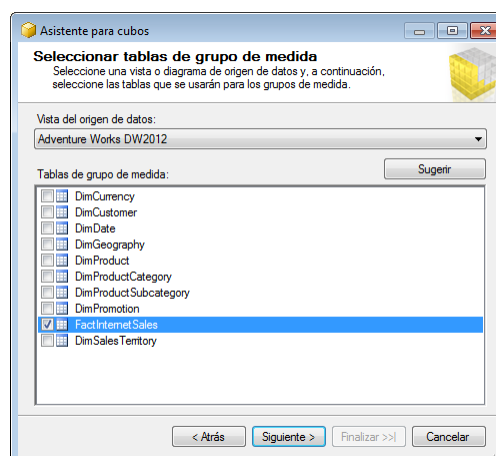
Las dimensiones configuradas fueron :

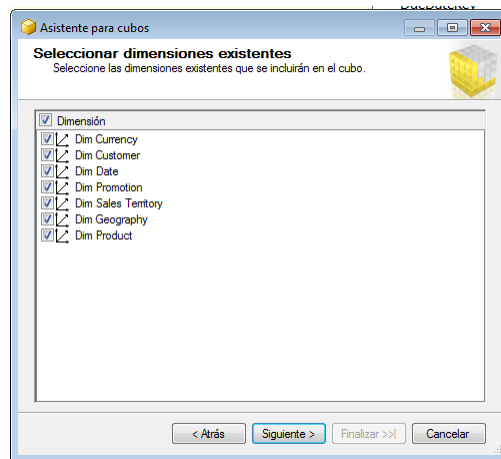
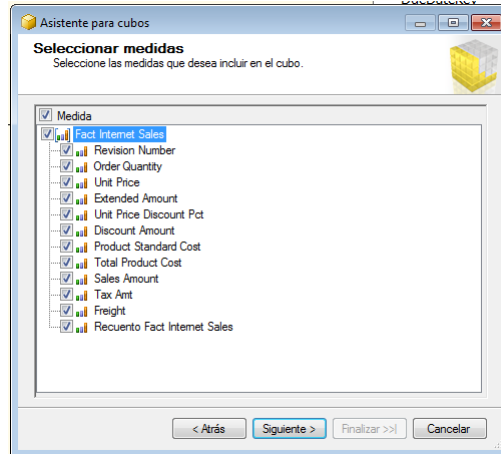
- **Currency**
- **Customer**
- **Date**
- **Promotion**
- **Sales Territory**
- **Geography**
- **Product**

Lo siguiente fue crear el cubo.

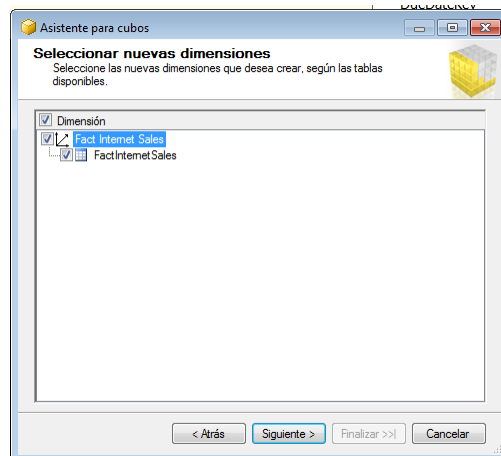


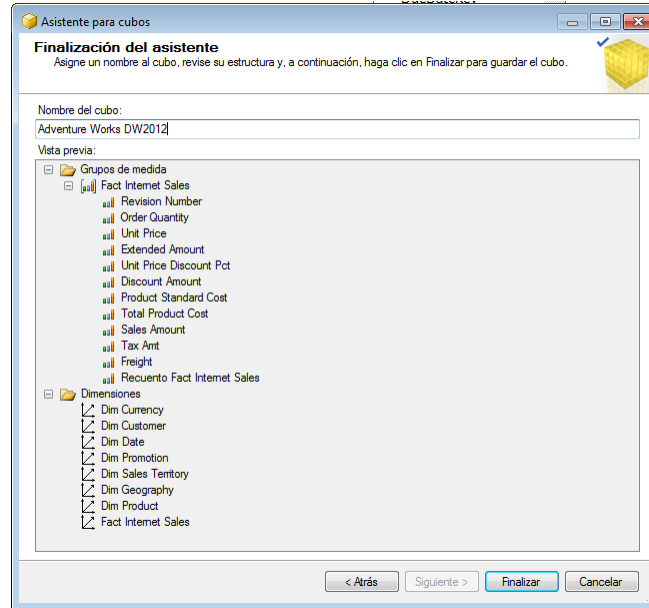
Utilizamos la tabla *FactInternetSales* como la tabla de hechos



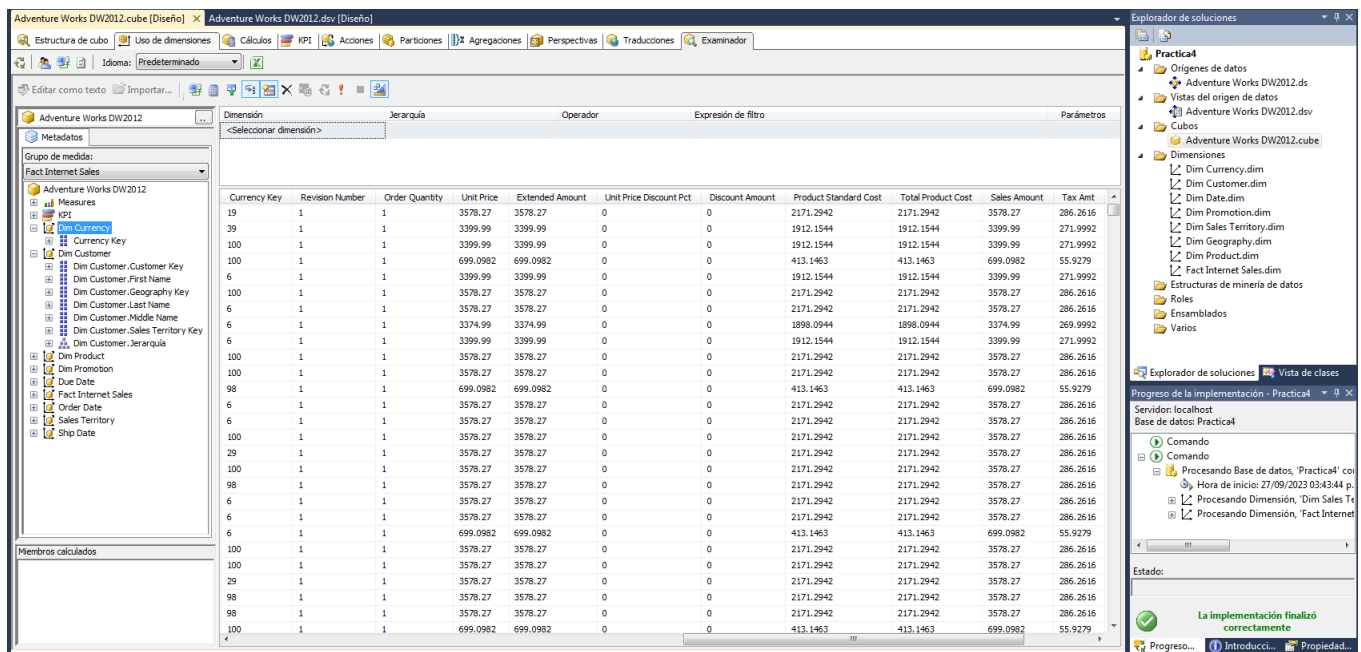


Al crear el cubo también se crea la dimensión *Fact Internet Sales*.

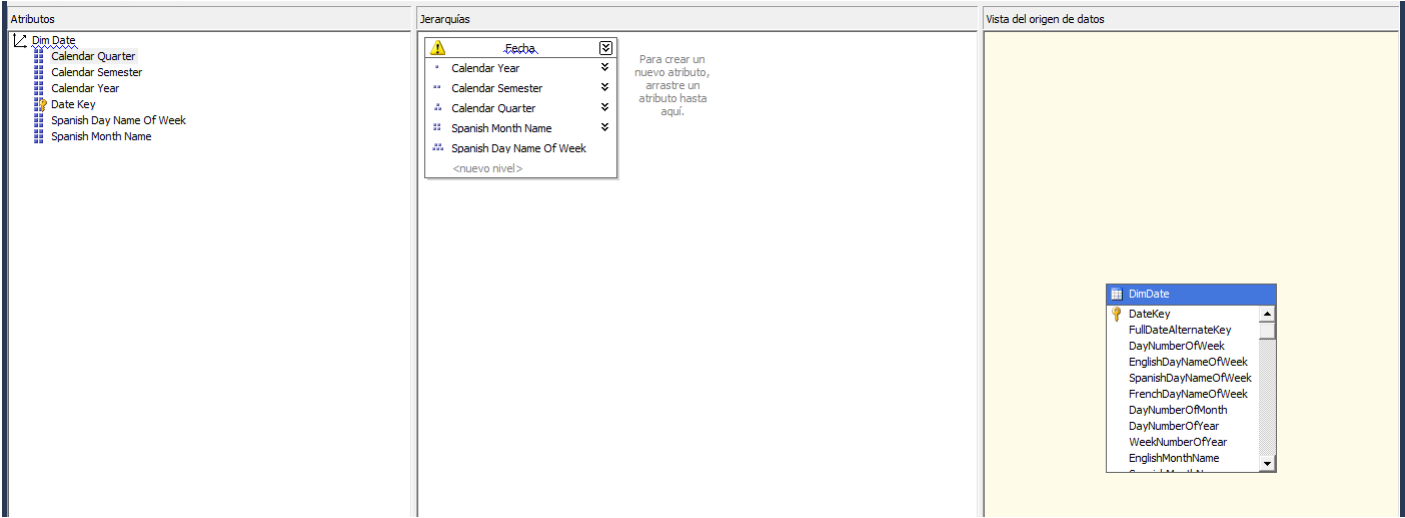




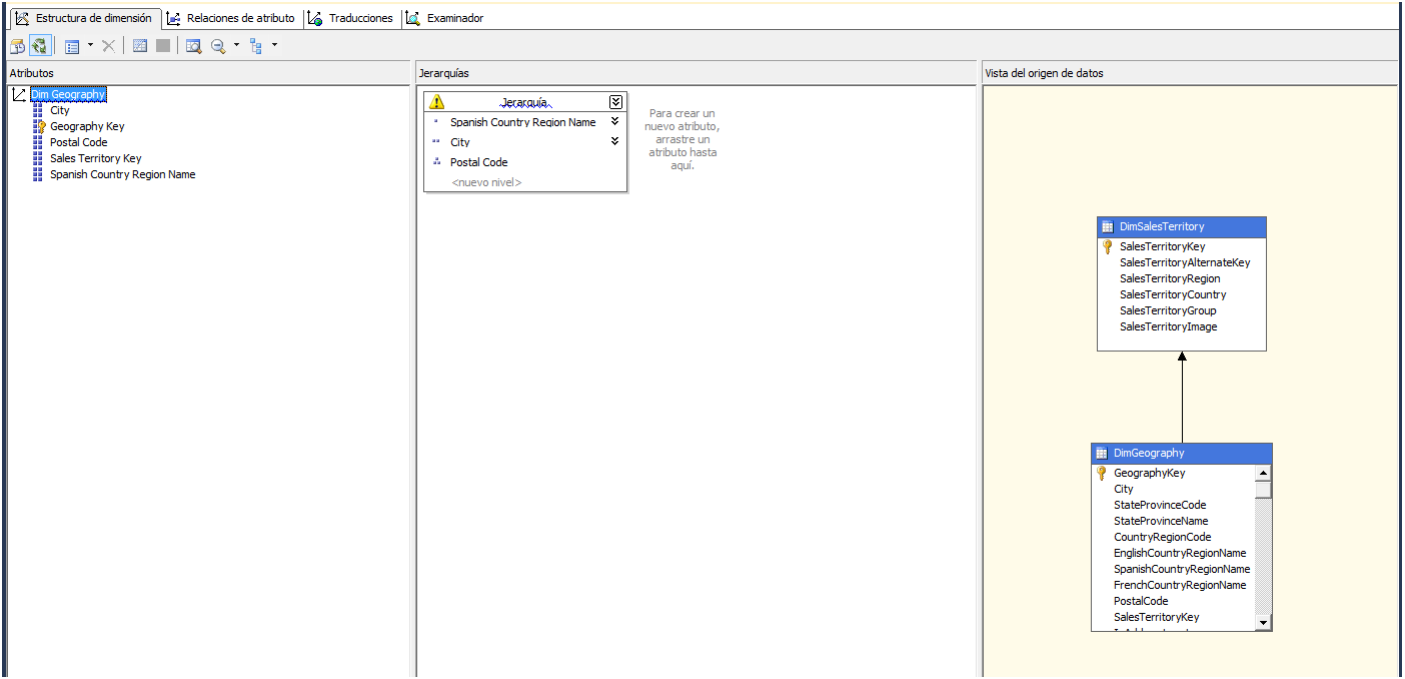
El relato de la creación del cubo lo podemos observar en el examinador del mismo.



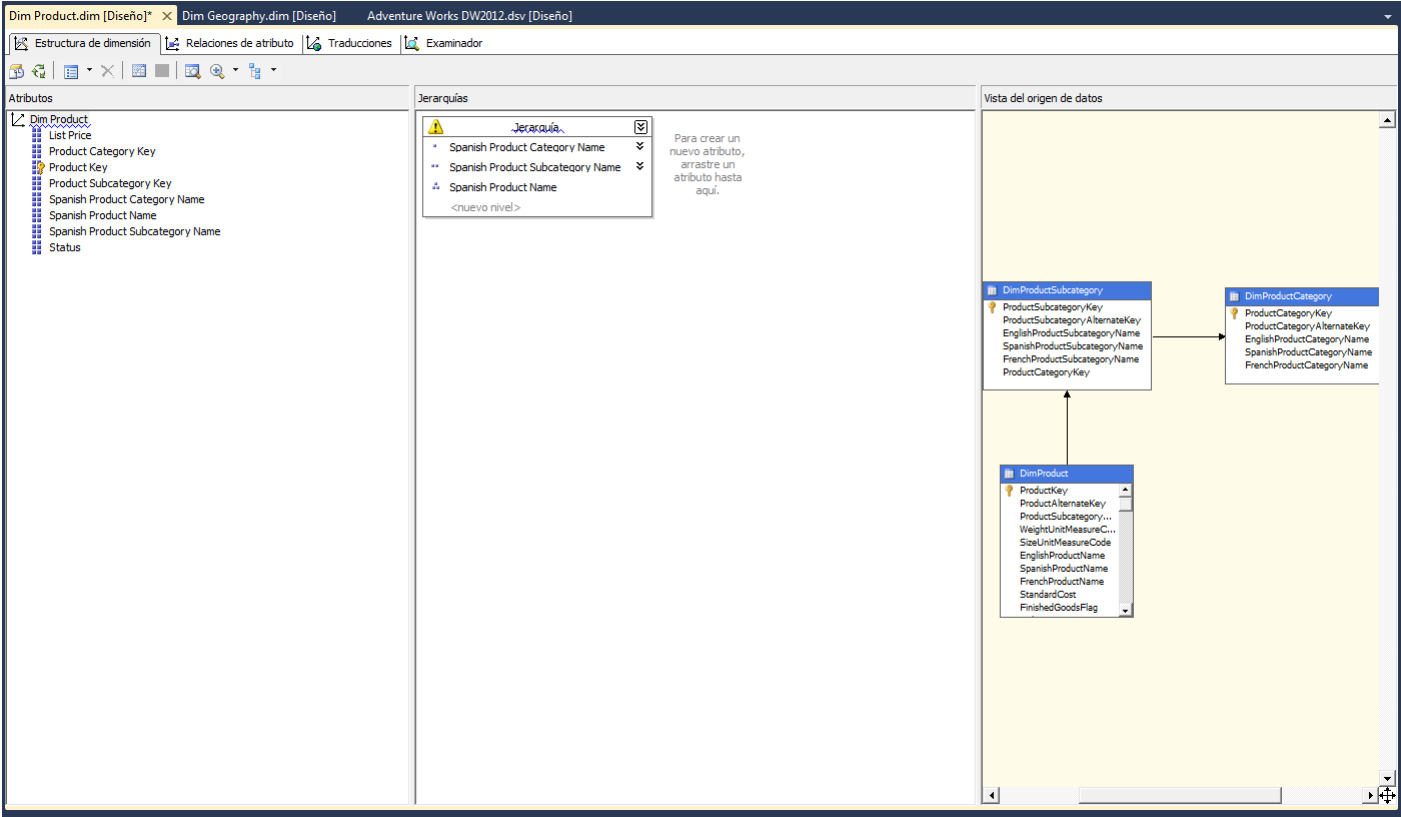
El siguiente paso fue crear jerarquías en las dimensiones donde tuvieran sentido.
La primera fue en la dimensión fue *Date*, con el año, semestre, trimestre, mes y día de la semana, en ese orden.



La siguiente jerarquía fue en la dimensión *Geography* entre el nombre de la región del país, la ciudad y el código postal, en ese orden



Y por ultimo la jerarquía en la dimensión *Product* entre la categoría, subcategoría y nombre del producto.



Para finalizar obtenemos una consulta entre todas las dimensiones, desafortunadamente era demasiada extensa y no fue posible mostrarla toda en pantalla, y con cualquier modificación mandaba error de llenado de memoria.

Dimensión	Jerarquía	Operador	Expresión de filtro	Parámetro
<Seleccionar dimensión>				
Sales Order Number	Currency Key	Last Name	Middle Name	First Name
SO43697	1	Adams		Adam
				Spanish Product Category Name
				Spanish Product Subcategory Name
				Spanish Product Name
				Discount Pct
				Calendar Year
				Calendar Semester
				Calendar Quarter

Selección de hechos

Para la selección de hechos realizamos la siguiente consulta:

Spanish Product Name	Total Product Cost	Discount Amount	Tax Amt	Order Quantity	Sales Amount
Calcomines para carre...	95380.6922999971	0	18734...	13135	234181.229...
Calcomines para carre...	907.821000000001	0	194.184	270	2427.29999...
Calcomines para carre...	1001.9654	0	214.3216	298	2679.01999...
Camiseta clásica, G	4631.05499999998	0	990.60...	195	12382.5
Camiseta clásica, M	4726.05099999998	0	1010.92	199	12636.5
Camiseta clásica, P	3989.83199999998	0	853.44...	168	10668
Carretera: 150, roja,...	610133.670200001	0	80439...	281	1005493.87
Carretera: 150, roja,...	731726.145400001	0	96470...	337	1205876.99
Carretera: 150, roja,...	655730.848400001	0	86451...	302	1080637.54
Carretera: 150, roja,...	640531.789000001	0	84447...	295	1055589.65
Carretera: 150, roja,...	729554.851200001	0	96183...	336	1202298.72
Carretera: 250, negr...	391170.812000002	0	50270...	271	628377.262...
Carretera: 250, negr...	430343.236100003	0	55296...	298	691206.262...
Carretera: 250, negr...	455734.954900003	0	58752...	319	734401.199...

En donde seleccionamos:

1. **Total Product Cost** para poder obtener el costo total de los productos.
2. **Discount Amount** para poder obtener importe del descuento que se hizo, si es que hubo alguno.
3. **Tax Amt** para poder obtener el importe del impuesto de los productos.
4. **Order Quantity** para poder obtener la cantidad de órdenes, pedidos o compras que se hicieron de cada producto.
5. **Sales Amount** para poder obtener el total de ventas de cada producto.

Utilizando estos 5 hechos o medidas nos facilitarán la tarea de administrar de una mejor manera los productos que se venden.

2. Preguntas

- ¿Que es un sistema de almacenamiento ROLAP?

Así cómo las estrategias OLAP, el sistema ROLAP (Procesamiento Analítico en Línea Relacional) nos sirve para almacenar y procesar datos. Es una herramienta OLAP construida a partir de una base de datos relacional en dónde tiene suma importancia la tabla de hechos que es dónde se almacena la historia de la información relevante para la organización.

Sustituye el uso de cubos (o modelo multidimensional) de datos por una base de datos relacional, lo que en principio, puede ser más sencillo para alguien que sólo conoce el modelo relacional. Es un modelo de procesamiento de datos utilizado en sistemas de gestión de bases de datos y en la creación de almacenes de datos para análisis de negocios. Su estructura esta dada por:

- Estructura Relacional.
 - Se basa en el modelo Estrella y Copo de Nieve.
 - Consultas SQL.
 - Optimización de Rendimiento.
 - Flexibilidad.
 - Requerimientos de Almacenamiento.
 - Escalabilidad.
 - Integración con Bases de Datos Relacionales.
- ¿Cuáles son las principales diferencias que hay entre ROLAP y MOLAP?

ROLAP	MOLAP
En ROLAP, los datos se almacenan en bases de datos relacionales. Los datos analíticos se organizan en tablas relacionales utilizando lenguaje SQL.	En MOLAP, los datos se almacenan en una estructura multidimensional, generalmente en cubos OLAP. Los datos se organizan en dimensiones y medidas, lo que facilita un acceso rápido y eficiente a los datos analíticos.
Tiene un rendimiento más lento en comparación con MOLAP, ya que las consultas deben ser traducidas en SQL y ejecutadas en una base de datos relacional.	Ofrece un rendimiento más rápido, ya que los datos se precálculan y se almacenan en una estructura multidimensional, lo que permite respuestas más rápidas a las consultas analíticas.
Requiere más espacio de almacenamiento debido a la estructura de base de datos relacional y la redundancia de datos en tablas.	Utiliza menos espacio de almacenamiento, ya que los datos se almacenan de manera más eficiente en cubos multidimensionales sin redundancia significativa.
Ofrece una mayor flexibilidad en términos de manejo de datos complejos y relaciones entre ellos, lo que es adecuado para escenarios donde los datos son altamente normalizados.	Es más adecuado para datos que se pueden modelar de manera efectiva en una estructura multidimensional. Puede no ser tan flexible para relaciones de datos complejas.
Puede requerir un mayor esfuerzo de mantenimiento debido a la complejidad de las bases de datos relacionales subyacentes.	Tiende a requerir menos mantenimiento, ya que los datos se precálculan y se almacenan en una estructura que está diseñada para consultas analíticas.
Es más escalable en términos de capacidad para manejar grandes volúmenes de datos y cargas de trabajo complejas.	Puede ser menos escalable en comparación con ROLAP para casos de uso con requisitos extremadamente altos de rendimiento y escalabilidad.

- Explica a grandes rasgos, como se implementaría un sistema de almacenamiento ROLAP.

Se debe diseñar una estructura de base de datos relacional que refleje las necesidades de la organización, debemos extraer datos como bases de datos transaccionales, sistemas externos u otras fuentes de datos. Estos datos se consolidan y se almacenan en la base de datos relacional. Luego, podemos usar esquemas como Copo de Nieve o Estrella. Necesitamos plantear algunas consultas, pueden ser lo suficientemente sencillas o complejas, al gusto del experto en datos. También se establecen permisos de seguridad en la base de datos relacional para controlar el acceso a los datos. Esto garantiza que solo usuarios autorizados puedan realizar consultas y acceder a la información. Cuando requerimos actualizaciones, podemos usar ETL para duplicar y actualizar. Por último los usuarios utilizan herramientas de Business Intelligence (BI) para interactuar con la base de datos relacional y realizar análisis. Estas herramientas permiten crear informes, gráficos y dashboards para el análisis de datos.

- ¿En que se basa el sistema de almacenamiento HOLAP?

HOLAP (Hybrid Online Analytical Processing). Es una combinación de los modelos MOLAP y ROLAP, se basa en la idea de que no es necesario elegir uno de estos modelos puros, sino que se pueden aprovechar las ventajas de ambos enfoques. A continuación listamos algunos conceptos relacionados:

- Almacenamiento híbrido.
- Capacidad de elección.
- Transparencia para los usuarios.
- Optimización de consultas.