

**UNIVERSIDAD DE COSTA RICA
ESCUELA DE CIENCIAS DE LA
COMPUTACIÓN E INFORMÁTICA**

Almacenes de Datos

Elaborado por
Dra. Elzbieta Malinowski G.
y modificado por el
Dr. Luis Gustavo Esquivel Q.

Versión 02-2024



Documentos de Almacenes de Datos by [Dr. Luis Gustavo Esquivel Quirós](#)
is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Costa Rica License](#).
Permissions beyond the scope of this license may be available at <http://creativecommons.org>.



OBJETIVO GENERAL DE LA PRÁCTICA

Familiarizarse con la estructura y componentes básicos de las herramientas que permiten crear, procesar y manipular los cubos OLAP en el servidor y en la herramienta cliente.

OBJETIVOS ESPECÍFICOS DE LAS PRÁCTICAS

El estudiante aprenderá:

- Crear cubos OLAP con dimensiones, jerarquías y medidas a partir de las fuentes de datos expresados en forma multidimensional y en forma convencional (tablas normalizadas para bases de datos operacionales/transaccionales).
- Incluir los elementos adicionales calculados a partir de los atributos existentes en los orígenes de datos.
- Modificar las características de despliegue para asegurar las agregaciones correctas y para mejorar la comprensión de datos durante el proceso de análisis.
- Aplicar los conceptos avanzados del modelo multidimensional para la creación y manipulación de cubos OLAP.
- Utilizar cubos OLAP en las herramientas del cliente.

METODOLOGÍA

Se especificarán los pasos de ejecución de cada práctica incluyendo los pantallazos en los casos necesarios. Los estudiantes deben realizar cada una de las prácticas y presentar los proyectos implementados por medio de la manipulación de datos en cubos OLAP (servidores OLAP y aplicación cliente, cuando se especifica).

NOTA

En esta práctica se usará SSDT (SQL Server Data Tools) de SQL Server. Antes de realizar la práctica deben asegurarse de que los servicios de SQL Server requeridos para esta herramienta están corriendo. En versiones anteriores de SQL Server SDT es conocido como SSAS (SQL [Server Analysis Services](#)).

LABORATORIO OLAP

DESCRIPCIÓN

El almacén de datos AdventureWorksDW2022 de Microsoft incluye los datos referentes a ventas por Internet. La figura 1 presenta un pequeño extracto de este almacén que se utilizará en esta práctica. Deben crear el modelo conceptual correspondiente, así como un usuario que tenga privilegios sobre el almacén de datos y justificar la asignación de permisos que realice sobre este.

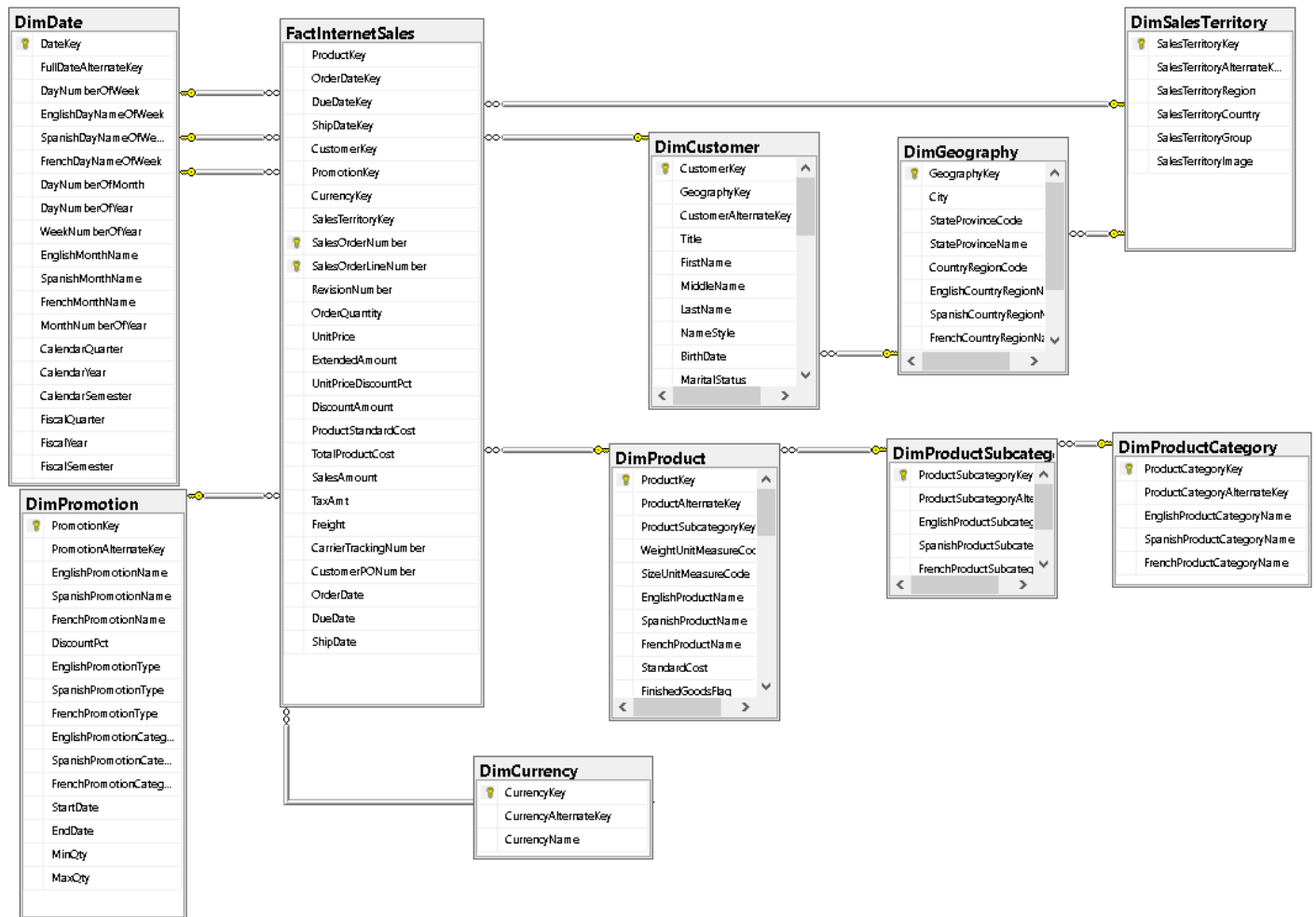


Figura 1: Extracto de AdventureWorksDW2022

PASOS:

Previos a la práctica

Familiarizarse con el diagrama de AdventureWorksDW2022 presente en la figura 1.

- Considerando el diagrama facilitado usted debe crear un esquema conceptual del mercado de datos que incluya todas las dimensiones presentes en la figura 1 y, además, las siguientes jerarquías y medidas con las funciones de agregación:

Dimensión	Jerarquía
Product	Product name – Subcategory – Category
Customer	Customer last name – Gender
	Customer last name – Martial Status
	Customer last name – Total children
Promotion	Promotion key – English promotion category
Order sate	Day number of month – Month name – Year
Sale Territory	Region – Country - Group

Medida	Función de agregación
Order quantity	Sum
Extended amount	
Sales amount	
Total product cost	
Discount amount	
Product standard cost	No aditiva
Unit price	
Unit price discount Pct	
	Max

Durante la práctica

I. Creación del proyecto y origen de datos

- Desde Visual Studio crear un nuevo proyecto de Analysis Services *Multidimensional* llamado OLAPPractical que pertenece a la solución PracticasOLAP (ver figura 2).

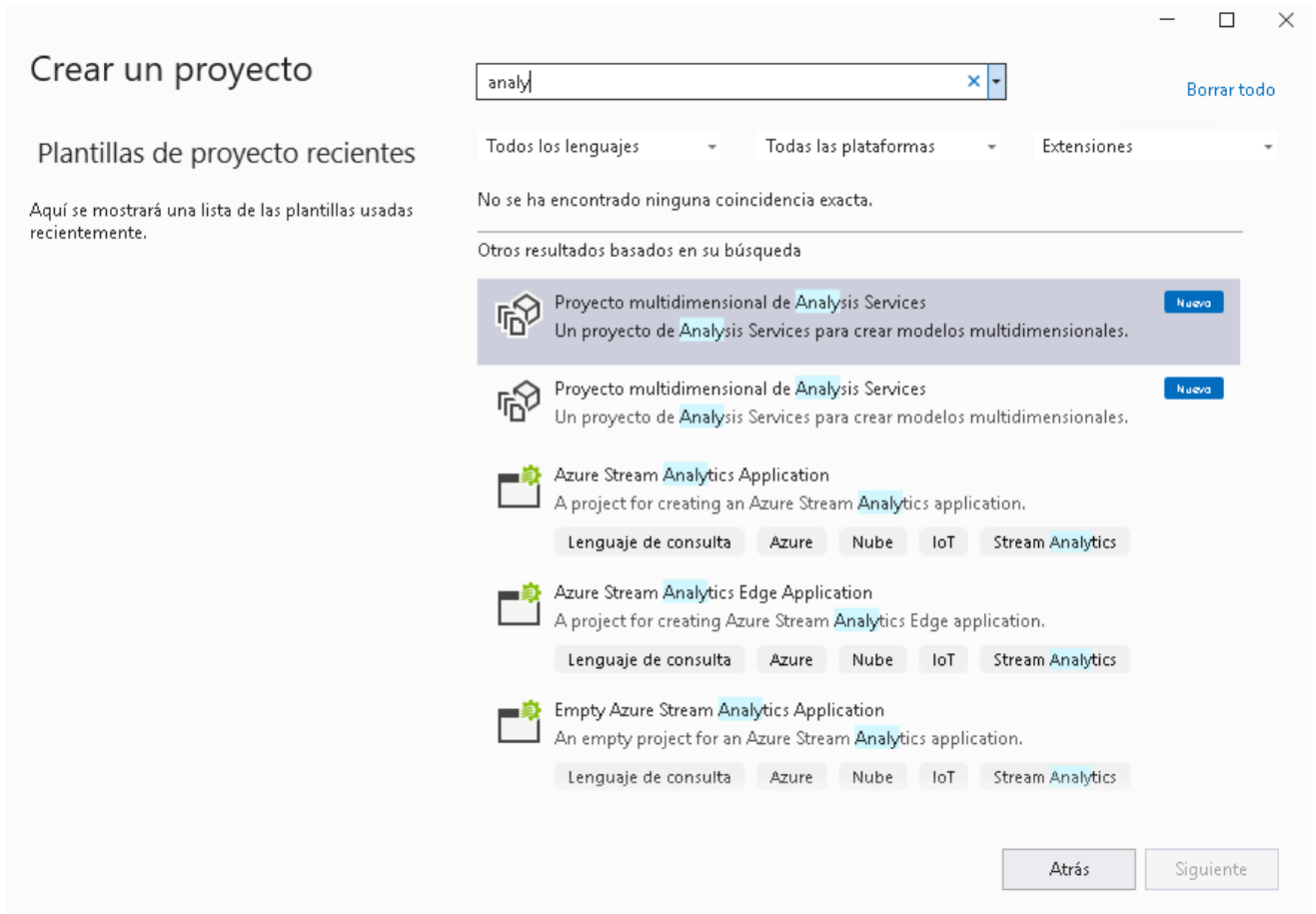


Figura 2: Creación de proyecto OLAP

- 2) Establecer un nuevo origen de datos por medio de wizard (click derecho en Data Sources y seguir los pasos, (puede obtener más información sobre las diferentes opciones de la Impersonation Information en <http://msdn.microsoft.com/en-us/library/ms187597.aspx>) La conexión es importante, debido a que, por falta de derechos, pueden tener problemas de compilar el cubo. También es importante asegurarse que el usuario de bases de datos utilizado pueda leer la base de datos (AdventureWorksDW2022) a utilizar.

II. Creación de una nueva vista de datos

- 1) Crear una vista de datos de origen (clic derecho en Data Source View) usando la conexión definida anteriormente e incluir todas las tablas especificadas en la figura 1 (resultado en la figura 3).

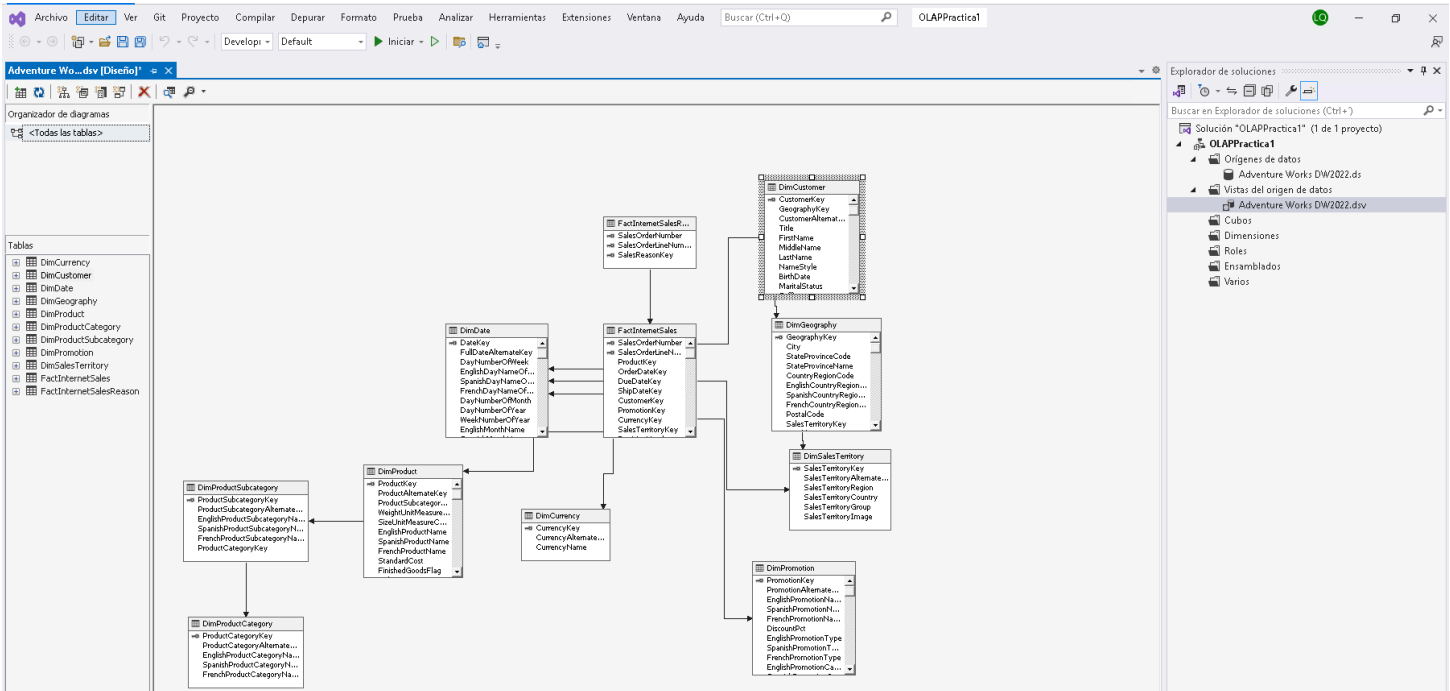


Figura 3: Vista creada

2) Incluir los siguientes campos calculados (Named calculations, clic derecho en la tabla de DimCustomer) que son expresiones SQL representadas como atributos:

- Campo “FullName” como concatenación de “FirstName”, “MiddleName” y “LastName” separados por espacios usando la siguiente expresión:

$$FirstName + ' ' + ISNULL(MiddleName + ',') + LastName$$
- Campo “MaritalStatusName” indicando el valor “Single” si el campo “MaritalStatus” es igual a “S” y “Married” en el otro caso usando la expresión:

$$CASE \text{ WHEN } MaritalStatus = 'S' \\ \text{ THEN 'Single'} \\ \text{ ELSE 'Married'} \\ \text{ END}$$

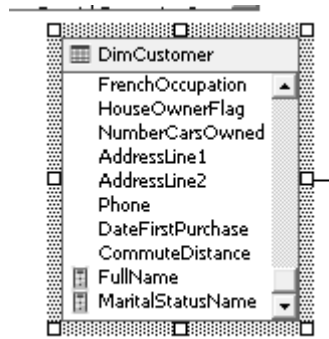


Figura 4: Inclusión de campos calculados

- 3) Explorar la distribución de datos de algunas dimensiones (posicionarse en la tabla interés y con clic derecho seleccionar Explore data).

III. Creación y procesamiento del cubo

- 1) Crear un cubo usando el ayudante, se deben incluir las dimensiones y la tabla de hechos especificadas en la vista de origen de datos (ver figura 3).
 - Para las medidas seleccionar todas excepto “Revision Number” y “Fact internet Sales Count”.
 - Para las dimensiones seleccionar todas y desmarcar la opción de “FactInternetSales” como la dimensión.

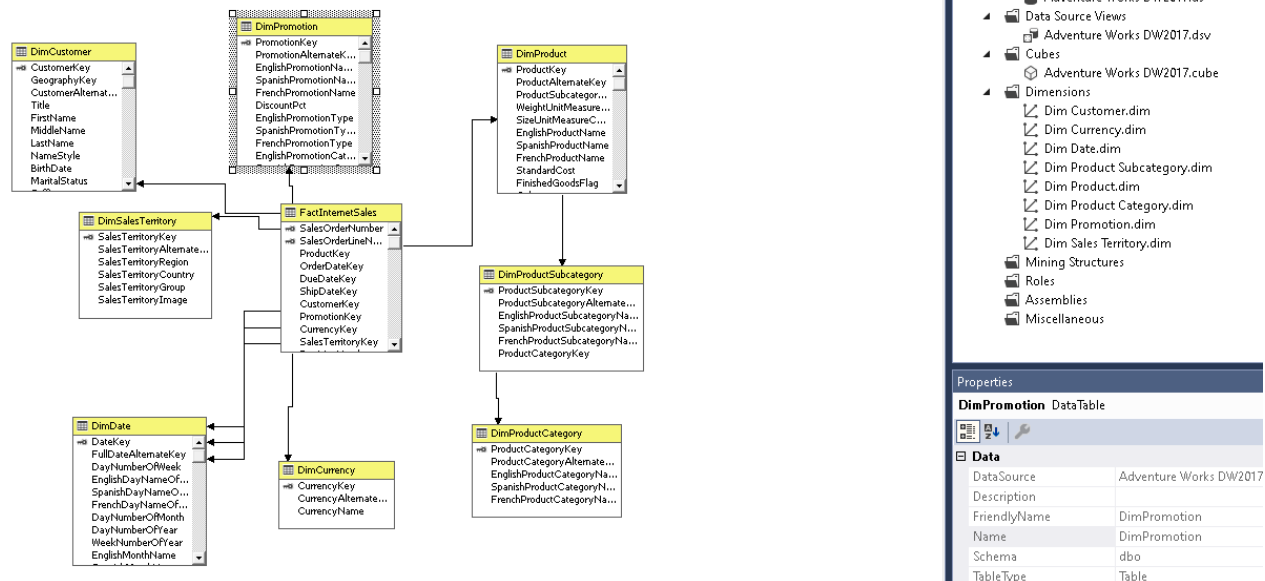


Figura 5: Cubo

- Verificar los componentes de cubo como se muestran en la figura 5.

- 2) Procesar el cubo: Existen varias opciones de hacerlo, por ejemplo, usando la opción de clic derecho en el nombre del cubo y seleccionando Process.
- 3) En la pestaña de Cube structure abrir el árbol de medidas para verificar la existencia de todas las medidas definidas para este cubo.
- 4) Pasar al panel de dimensiones (Dimensions) y comparar la existencia de tres dimensiones de tiempo (¿cuáles?) contra solo una dimensión de tiempo como tabla ("Dim Date") en el panel de "Solution Explorer". ¿Por qué el ayudante incluyó estas tres dimensiones?
- 5) Entrar a la opción de Browser, colocar la medida "Sales Amount" en el centro de la pantalla. Abrir la dimensión de "Product". Arrastrar el atributo de "Product Category Key" y a la par "Product Subcategory Key". Ejecutar consulta para ver los resultados. Puede notar que no está formada la jerarquía y los nombres no son significativos para usuarios de toma de decisiones.
- 6) Puede ver la estructura de esta dimensión y sus datos en la pestaña de Cube Structure, el panel de Dimensions y en la opción de Edit de la dimensión "Product". Verificar falta de la definición de jerarquías para esta dimensión. Entrar a la pestaña de Browser de esta dimensión y ver los datos que se despliegan. ¿Son estos datos significativos para usuarios de toma de decisiones?

IV. Cambios en las opciones por defecto en las dimensiones y medida

- 1) Para crear la jerarquía de la dimensión "Product", entrar al editor de esta dimensión, arrastrar del panel Data Source View hacia el panel de Attributes los atributos indicados en la figura 6. Además, arrastrar los atributos al panel de jerarquías como se muestra en la misma figura.

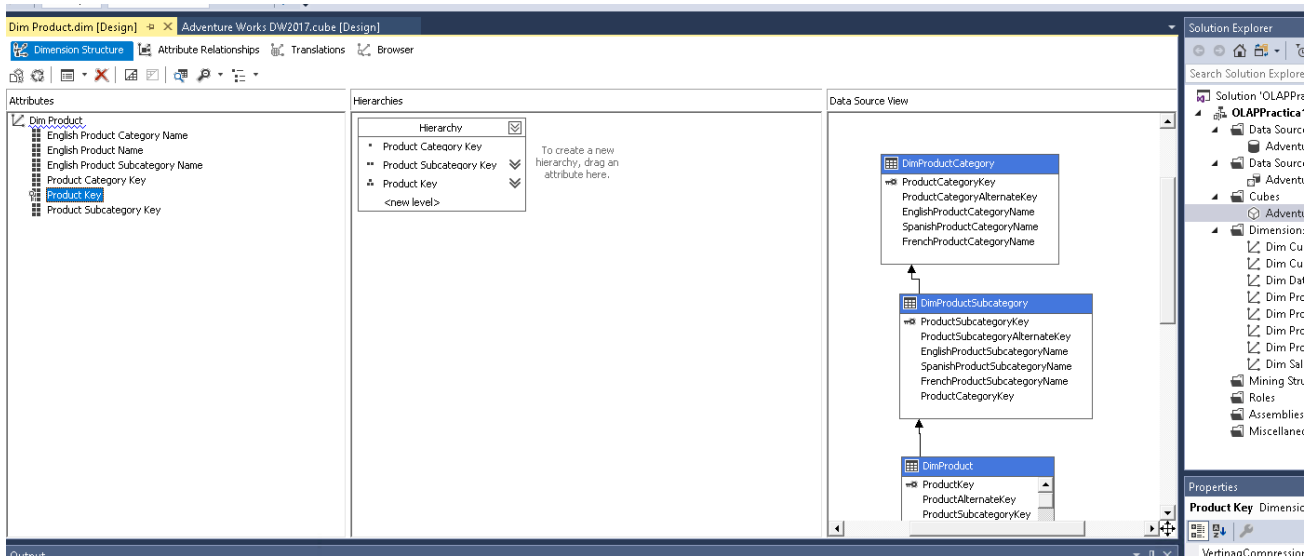


Figura 6: Jerarquía de la dimensión "Product"

- 2) Verificar y modificar (si fuese necesario) las relaciones entre atributos en la pestaña correspondiente de Attribute Relationships (ver figura 7).

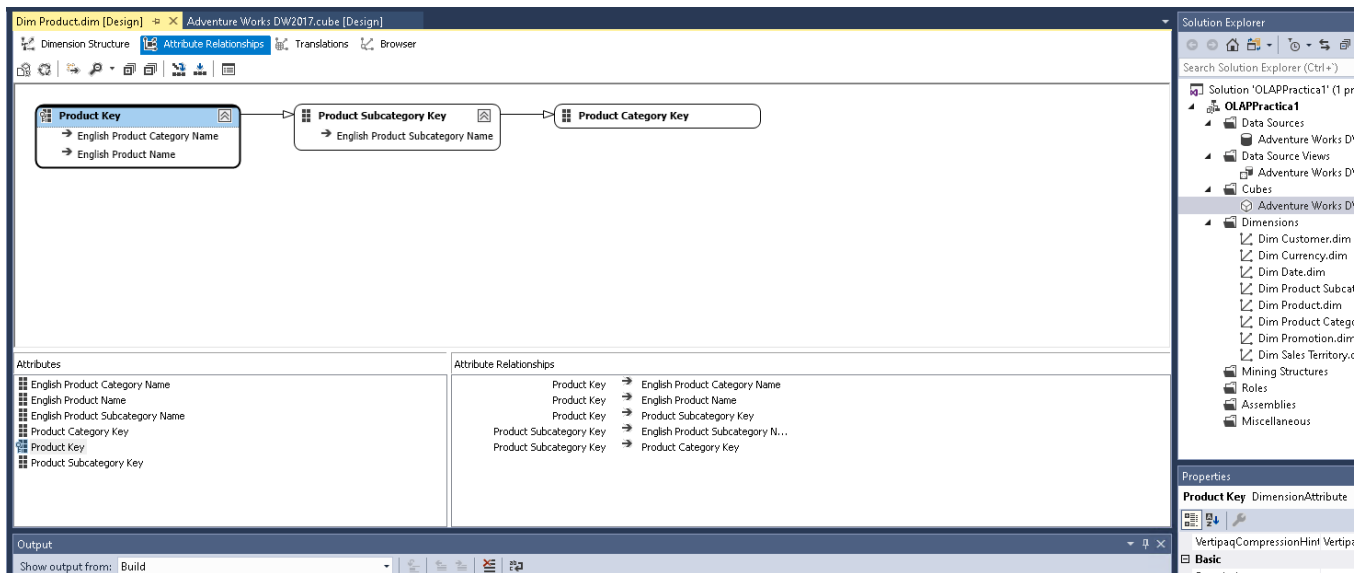


Figura 7: Relaciones de atributos de dimensión "Product"

- 3) Para cambiar los valores que se despliegan en la dimensión "Product" por datos más significativos, por ejemplo, nombres, posicionarse en "Product Category Key" en el panel de Attributes y con clic derecho seleccionar Properties. Cambiar la opción de NameColumn poniendo nombre de "EnglishProductCategoryName" como se muestra en la figura 8.

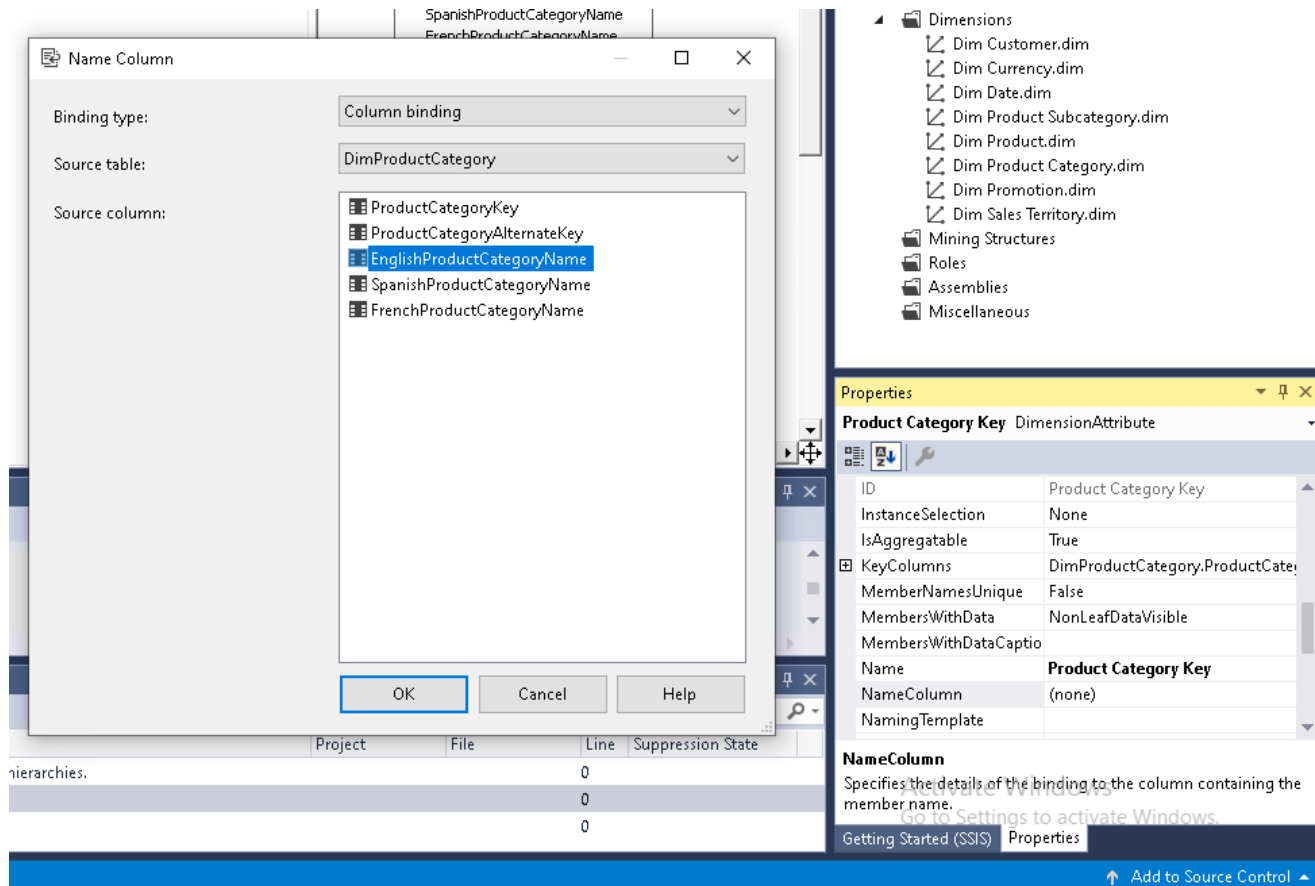


Figura 8: Cambio de la opción de "NameColumn".

- 4) Repetir este proceso para los demás atributos que forman esta jerarquía.
- 5) Volver a procesar la dimensión "Product", reconectar al servidor de OLAP y verificar los cambios ocurridos en el Browser de la dimensión "Product" en la jerarquía que acaba de crear.
- 6) Entrar al Browser del cubo, colocar la jerarquía de producto en las filas y la medida de "Sales Amount" en las celdas de este cubo.
- 7) Presionar en el Browser del cubo el ícono de Excel como se muestra en la figura 9 para acceder Excel y manipular el cubo dentro de la hoja electrónica. Efectuar la operación de drill-down sobre la jerarquía de productos. ¿Puede ser este despliegue de datos útil y satisfactorio para las personas encargadas de análisis de datos y toma de decisiones?

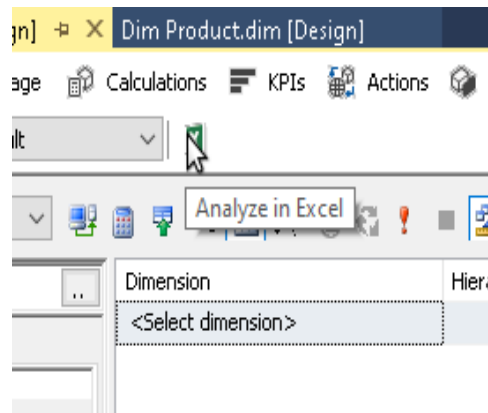


Figura 9: Ícono de Excel.

- 8) Modificar la dimensión “Customer” e incorporar varios atributos creando las jerarquías como se indica en la figura 10. Asegurarse que en lugar de desplegar la llave del cliente se despliega su apellido.

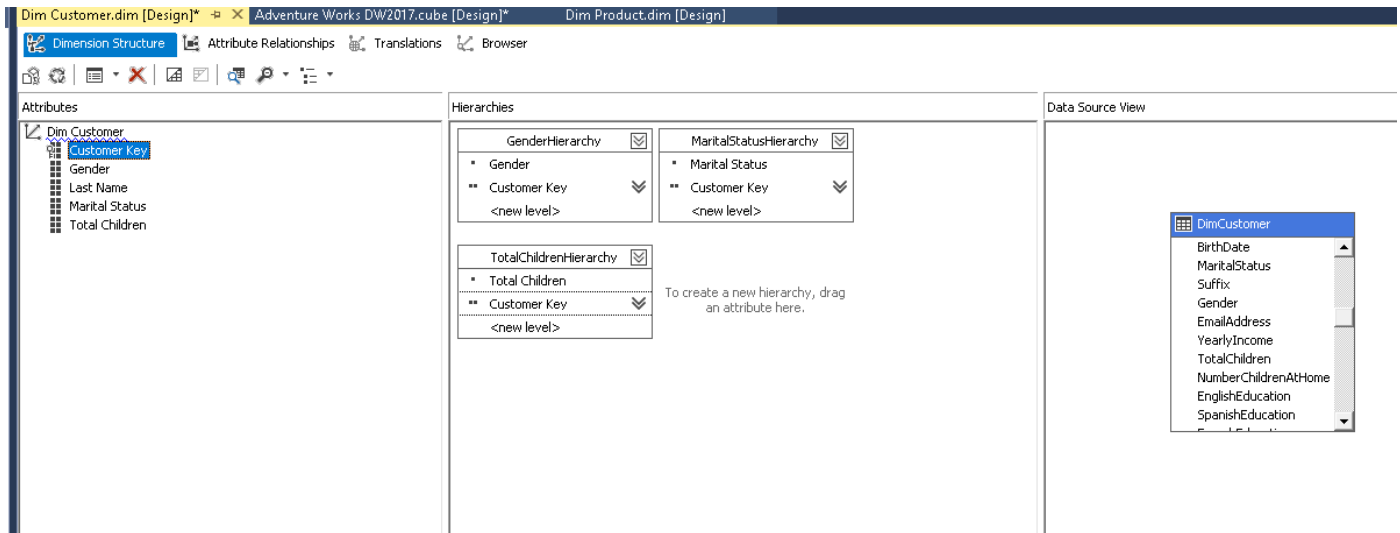


Figura 10: Dimensión "Customer"

- 9) En forma parecida modificar la dimensión de “Promotion” creando la jerarquía con los niveles: “Promotion key” – “English promotion category” (se puede usar “Spanish promotion category”).

VI. Cambios en las opciones por defecto en las dimensiones de tiempo

- 1) Entrar al editor de la dimensión “Date”, establecer la jerarquía de “Date key” – “English month name” – “Calendar Year” y modificar el valor de despliegue por valores más significativos, por ejemplo, en lugar de “Date key” usar “DayNumberOfMonth”. Revisar y corregir (si es necesario) las relaciones creadas en la pestaña de Attribute Relationships. Procesar esta dimensión. ¿Qué situación ha ocurrido?
- 2) La dimensión de tiempo requiere modificar su tipo de Regular al tipo Time. Para esto se debe entrar a la opción de editar la dimensión de tiempo de interés, sus propiedades y en las características básicas cambiar el tipo como se puede ver en la figura 11.

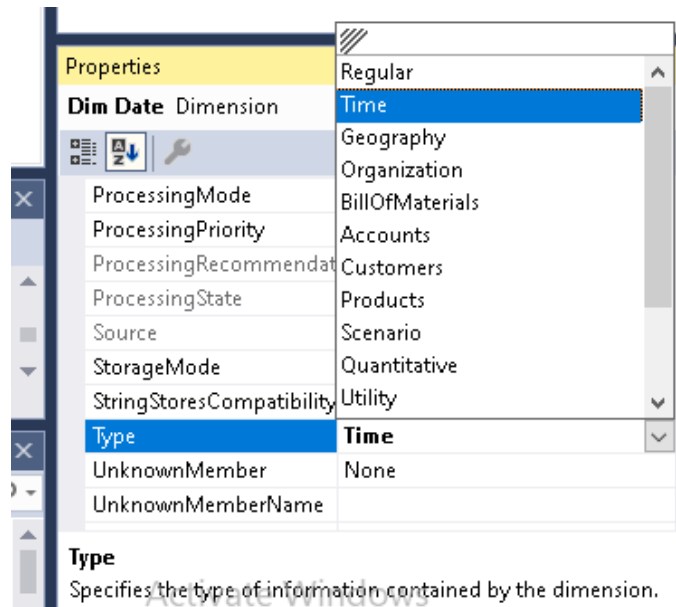


Figura 11: Cambio de tipo regular a tipo Time

- 3) Se debe establecer la llave compuesta para cada nivel de jerarquía ya que el crear un nivel actúa parecido a un tipo de entidad débil en el modelo ER y requiere completar la llave parcial. Para convertir esta llave parcial en llave primaria se debe crear una llave compuesta. Para esto se debe posicionar en el atributo de interés, por ejemplo, “Calendar Year” y en sus propiedades modificar la propiedad de “KeyColumns” estableciendo como llave la combinación de los tres atributos como se puede ver en la figura 12. También se debe indicar el valor para NameColumn para despliegues más significativos.

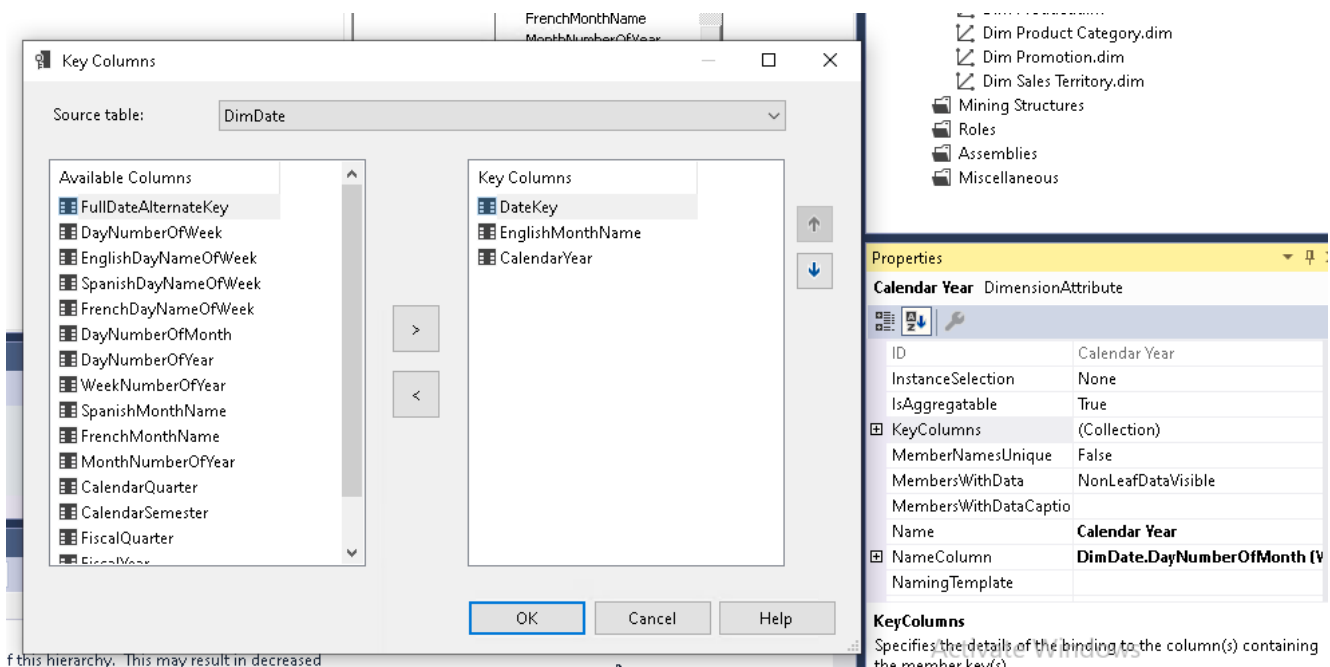


Figura 12: Cambio de llave para “Calendar Year”

- 4) Procesar la dimensión y verificar sus datos (cada vez que se cambia una dimensión se debe volver a procesar y verificar sus datos).
- 5) En el diseño del cubo entrar al panel de “Measures” y con clic derecho verificar las propiedades de algunas de las medidas, especialmente la propiedad que se refiere a “AggregationFunction”. Observe que por defecto las medidas tienen asignada la función de agregación Sum (ver figura 13).

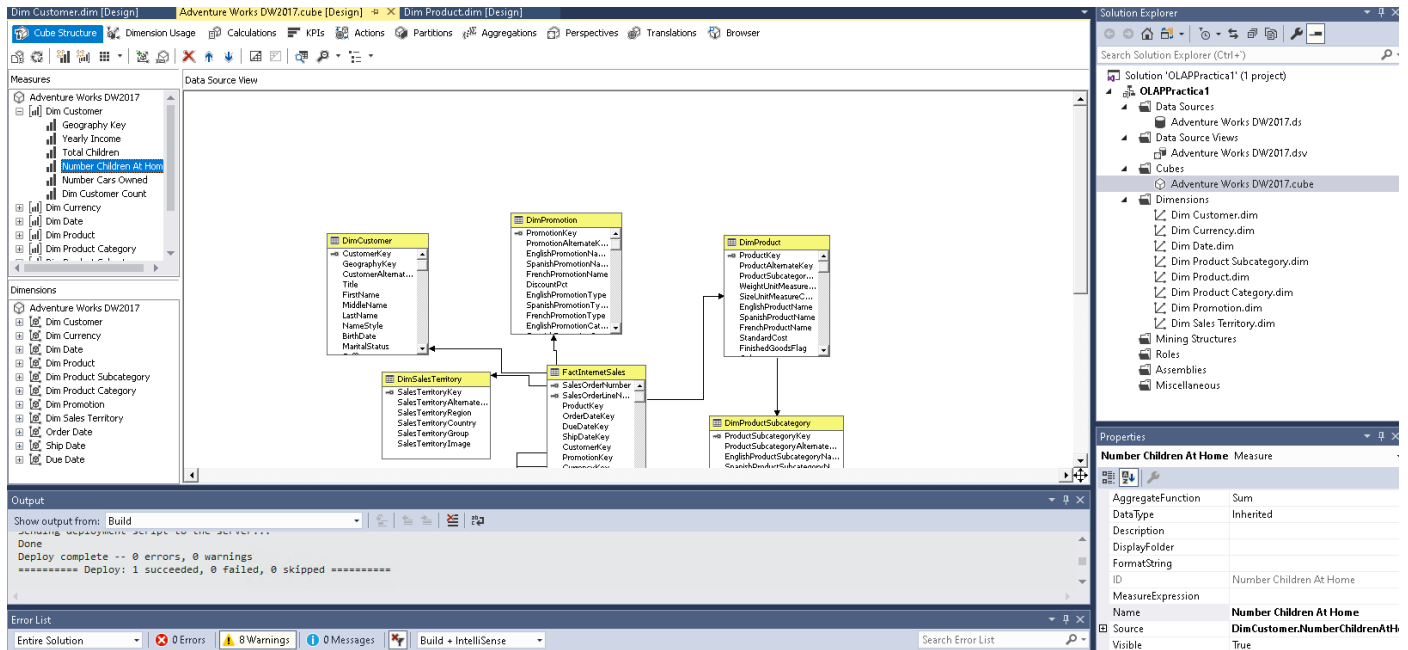


Figura 13: Panel de “Measures”

- 6) Modificar “AggregationFunction” de acuerdo con lo siguiente:
 1. La medida “Unit Price Discount” usa la función Max.
 2. La medida de “ProductStandardCost” es la medida semi-aditiva con la función de “AverageOfChildren”.
- 7) Procesar el cubo, entrar a Browser y pasar a la hoja Excel. Arrastrar a las filas y columnas algunas jerarquías de las diferentes dimensiones y en la parte central algunas medidas. Utilizar más que tres jerarquías y por lo menos dos medidas. Aplicar las operaciones de drill-down, roll-up, slice-and-dice y pivot para familiarizarse con las características de herramientas OLAP. (Recuerde documentar cualquier modificación que realice como por ejemplo si crea o modifica alguna otra dimensión o jerarquía).
- 8) Investigue y conecte en su cubo OLAP con Power BI. Documente todo lo realizado y muestre evidencia de su resultado final.