



**SALESIANOS UNIVERSIDAD DON BOSCO**

**FACULTAD DE INGENIERÍA**

**ESCUELA DE COMPUTACIÓN**

**CICLO 02-2020**

**"SEGUNDO DESAFÍO PRÁCTICO"**

**GRUPO DE LABORATORIO:**

01

**CARRERA:**

INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN.

**PRESENTADO POR:**

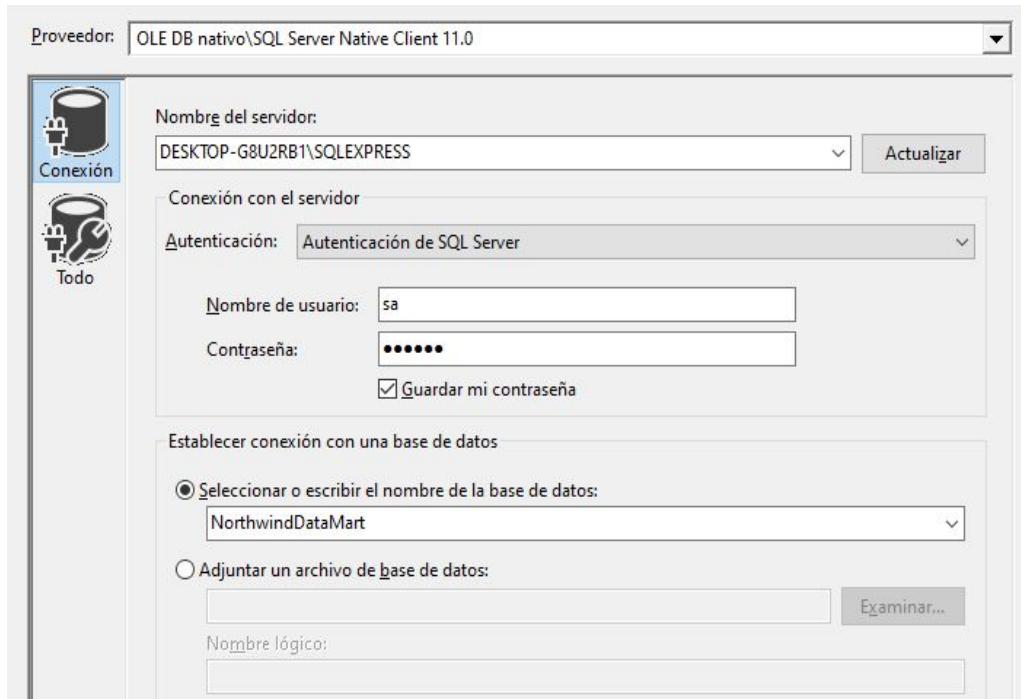
<b>Carnet</b>	<b>Nombre</b>	<b>Apellido</b>
VC190544	Francisco José	Valle Cornejo
AV190086	César Adilson	Ayala Vásquez

**DOCENTE:**

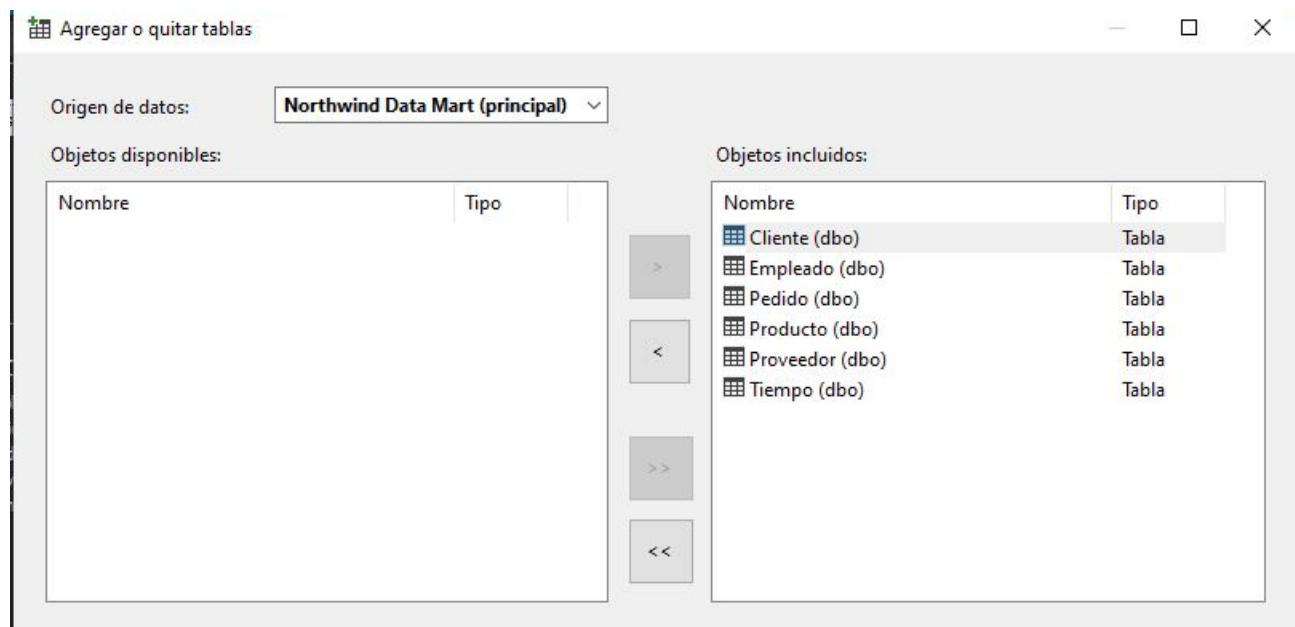
Alexander Alberto Sigüenza Campos

## EJERCICIO 1

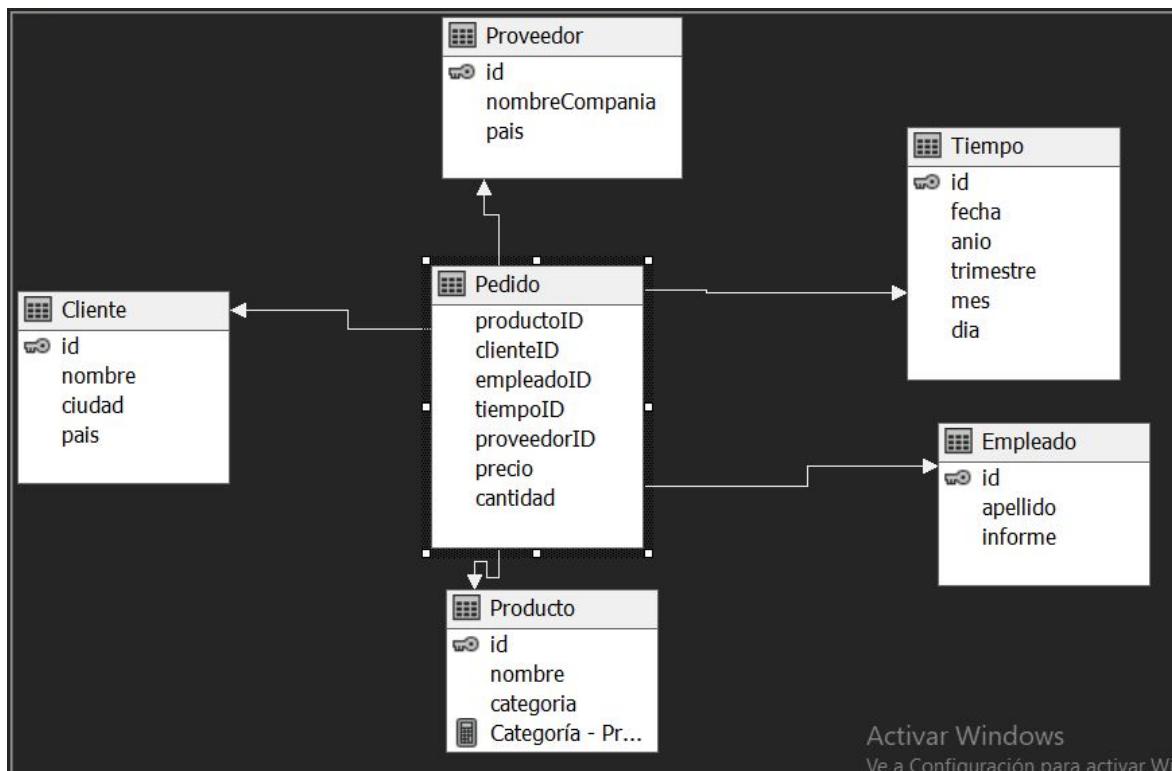
**Paso 1:** Una vez iniciado Visual Studio, vamos a definir el origen de datos para nuestro procedimiento. En este caso, el origen será la base de datos Northwind Data Mart, la configuración quedará de la siguiente manera:



**Paso 2:** Definir una vista que incluya las tablas que deseamos analizar, en nuestro caso, escogeremos todas las tablas para un mayor detalle al diseñar e implementar el cubo. Nos quedará de la siguiente manera:



La vista puede ser visualizada de esta manera:



**Paso 3:** Crear un nuevo Cubo, y definiremos como tabla hechos a la tabla “pedidos”, posteriormente definir las dimensiones: Proveedor, tiempo, empleado, producto y cliente.

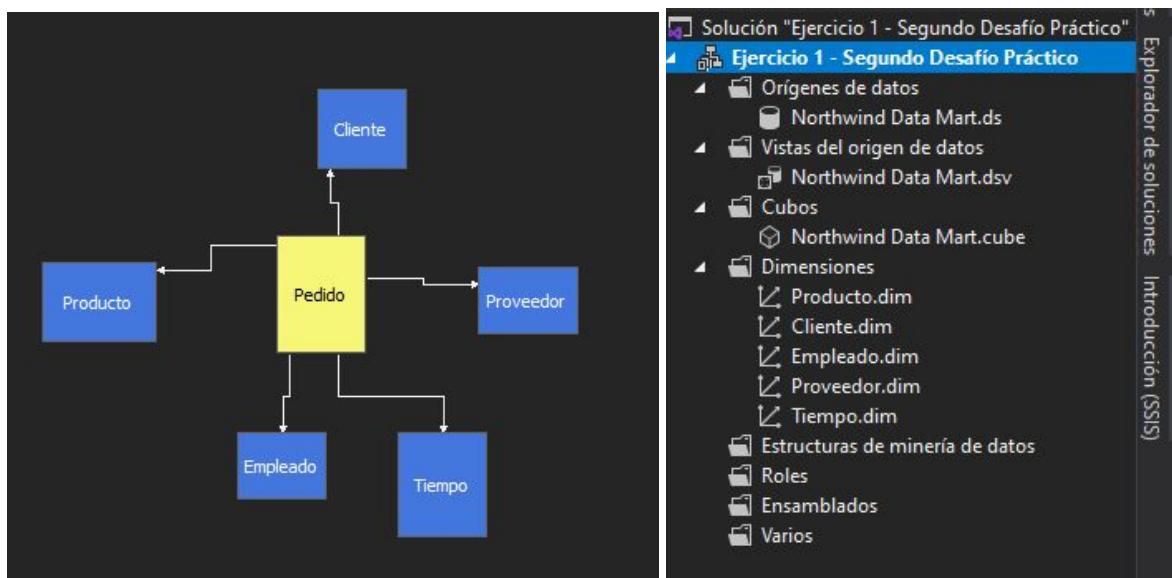
En la parte izquierda, se muestra la configuración de dimensiones para un cubo:

- Medida**: Pedido (selección activa).
- Precio**
- Cantidad**
- Recuento Pedido**

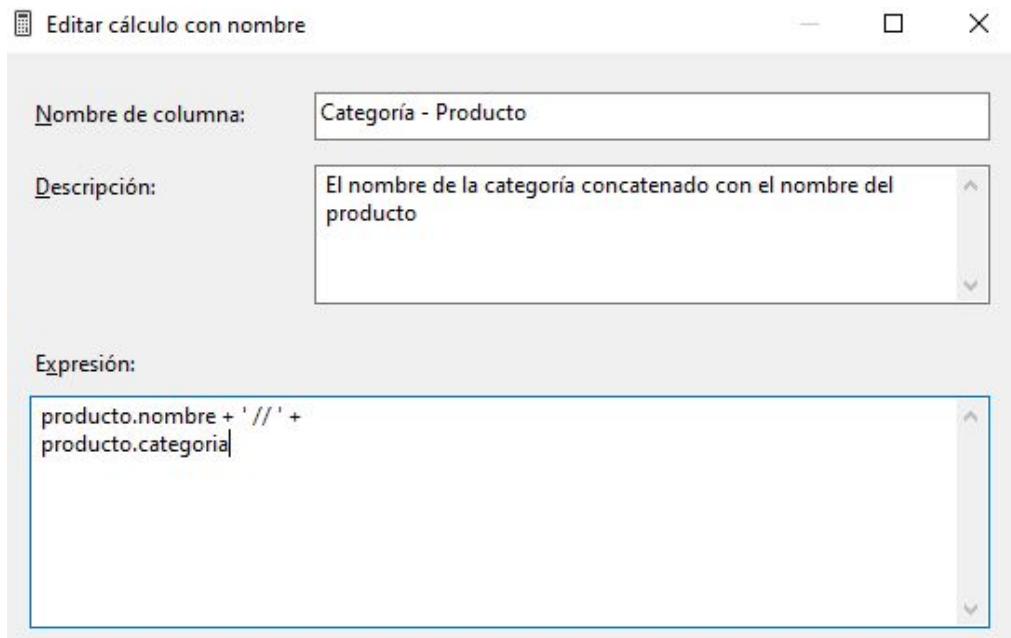
En la parte derecha, se detallan las dimensiones seleccionadas:

- Dimensión**
  - Producto**
  - Cliente**
  - Empleado**
  - Proveedor**
  - Tiempo**

**Paso 4:** El nombre del cubo será el que el programa nos da por defecto, luego de construir el cubo podemos visualizarlo mejor con un zoom del 50%. Quedando nuestra tabla hechos con color amarillo y todas las dimensiones de color azul.



**Paso 5:** Antes de la implementación, crearemos un campo nuevo Nombre/Categoría de la tabla productos, para ello, nos dirigimos a la vista creada en el paso 2, click derecho en la tabla producto y “Nuevo cálculo con nombre...”. Configurar los campos como se muestra en la imagen:



La expresión para concatenar los campos es la siguiente:

***producto.nombre + ' // ' + producto.categoría***

**Paso 6:** Cambiar las dimensiones para que las consultas se guien más allá de las llaves primarias

The screenshot shows the Analysis Services Management Studio. On the left, there are two panes: 'Atributos' (Attributes) and 'Jerarquias' (Hierarchies). The 'Atributos' pane contains a dimension named 'Cliente' with attributes: Ciudad, Id, Nombre, and País. The 'Jerarquias' pane shows a hierarchy for 'Residencia' with levels País and Ciudad, and an option for 'Nuevo nivel' (New level). Below these are three smaller panes: 'Atributos' for 'Producto' (with attributes Categoría, Id, Nombre), 'Atributos' for 'Empleado' (with attributes Apellido, Id), and 'Atributos' for 'Proveedor' (with attributes Id, País).

### Paso 7: Procesar el cubo y examinar el cubo aplicando la consulta pedida

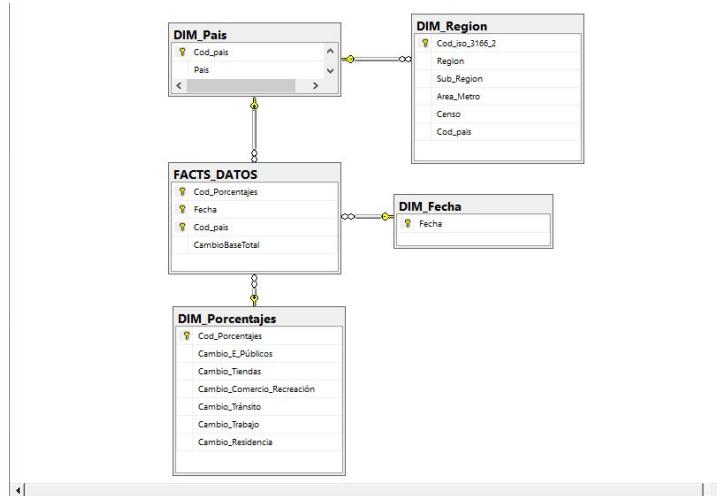
- Comando
  - + Procesando Dimensión, 'Cliente' completados.
  - + Procesando Dimensión, 'Empleado' completados.
  - Procesando Cubo, 'Northwind Data Mart' completados.
    - Hora de inicio: 4/10/2020 18:39:18; Hora de finalización: 4/10/2020 18:39:20; Duración: 0:00:01
    - + Procesando Grupo de medida, 'Pedido' completados.
  - + Procesando Dimensión, 'Producto' completados.
  - + Procesando Dimensión, 'Proveedor' completados.

### Paso 8: Ingresar la consulta pedida en el desafío un cubo en donde pueda visualizar nombre y país del cliente, apellido del empleado, país del proveedor, y nombre y categoría del producto. Estos dos últimos concatenados en una misma columna

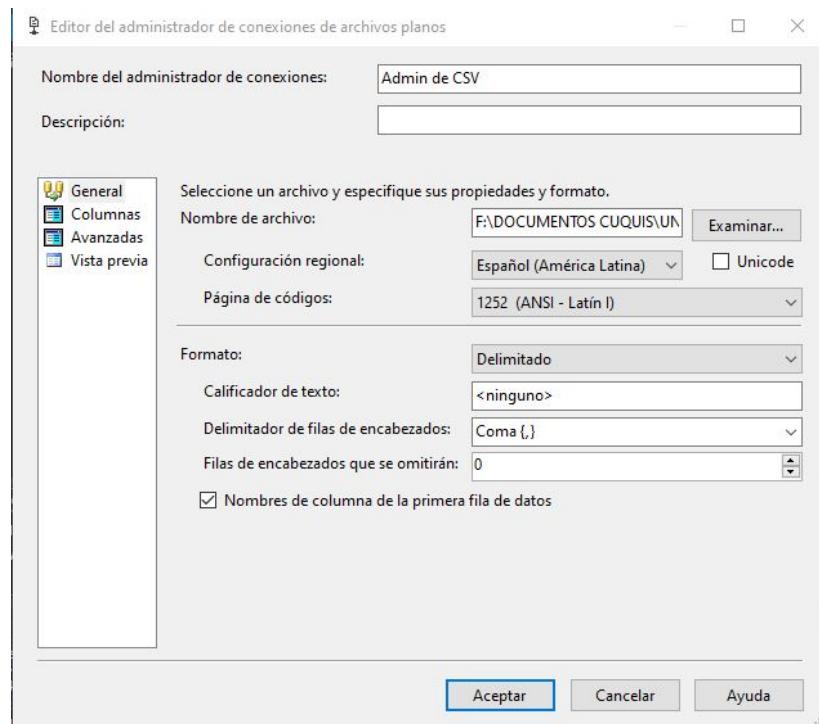
Nombre	País	Apellido	País	Categoría - Producto	Precio
Alejan...	Spain	Davolio	Au...	Perth Pastries // Meat/Poultry	262
Alejan...	Spain	Davolio	Ca...	Tourtière // Meat/Poultry	29.5
Alejan...	Spain	Davolio	Ge...	Nord-Ost Matjeshering // Seafood	207
Alejan...	Spain	Peacock	Brazil	Guaraná Fantástica // Beverages	21.6
Alejan...	Spain	Peacock	Fra...	Camembert Pierrot // Dairy Products	340
Alejan...	Spain	Peacock	Ge...	Nord-Ost Matjeshering // Seafood	150.09
Alejan...	Spain	Peacock	Italy	Ravioli Angelo // Grains/Cereals	31.2
Alejan...	Spain	Peacock	UK	Teatime Chocolate Biscuits // Confections	7.3
Alejan...	Spain	Peacock	USA	Steeleye Stout // Beverages	57.6
Alexan...	Ge...	Bucha...	Au...	Vegie-spread // Condiments	263.4
Alexan...	Ge...	Bucha...	Finl...	Leinenkugel's Blue Moon Beer	100

## EJERCICIO 2

**Paso 1:** Tras analizar los documentos separados con comas brindados en los recursos se ha definido usar un modelo de copo de nieve para su base de datos, con 3 tablas dimensiones principales y una relaciona a la tabla País



**Paso 2:** Se comienza con la creación del ETL, se usará un modelo de dos DataFlow para asegurar que las tablas Dimensiones ya tengan sus llaves primarias cuando la tabla Hechos comience a llenarse, primero se crea el origen a un archivo plano y se agrega que la separación de columnas sea por medio de la coma (,)



**Paso 3:** Crear una columna Derivada que lea cada fila y verifique que no hay NULL en sitios que no debería, cambiando los porcentajes base de NULL a “0” y la clave de Subregiones de NULL a “UNK”

Nombre de columna d...	Columna derivada	Expresión	Tipo de datos	Longitud	Precisión
retail_and_recreation_p...	Reemplazar 'retail_and...	retail_and_recreation_percent_change_from_baseli...	string [DT_STR]	50	
grocery_and_pharmac...	Reemplazar 'grocery_a...	grocery_and_pharmacy_percent_change_from_bas...	string [DT_STR]	50	
parks_percent_change_...	Reemplazar 'parks_per...	parks_percent_change_from_baseline == "" ? 0 : (D...	string [DT_STR]	50	
transit_stations_percen...	Reemplazar 'transit_sta...	transit_stations_percent_change_from_baseline == ...	string [DT_STR]	50	
workplaces_percent_ch...	Reemplazar 'workplace...	workplaces_percent_change_from_baseline == "" ? ...	string [DT_STR]	50	
residential_percent_ch...	Reemplazar 'residential...	residential_percent_change_from_baseline == "" ? ...	string [DT_STR]	50	
sub_region_1	Reemplazar 'sub_region...	sub_region_1 == "" ? "UNK" : sub_region_1	string [DT_STR]	50	
iso_3166_2_code	Reemplazar 'iso_3166_2...	iso_3166_2_code == "" ? "UNK" : iso_3166_2_code	string [DT_STR]	50	

**Paso 4:** Convertir todos los datos posibles haciendo que concuerden con los de la base de datos

Configure las propiedades utilizadas para convertir el tipo de datos de una columna de entrada a otro tipo. Configure la longitud, la precisión, la escala y la página de códigos de la columna en función del tipo de datos al que se convertirá la columna.

Columna de entrada	Alias de salida	Tipo de datos	Longitud	Precisión	Escala	Página de códigos
country_region_code	Copia de country_regi...	string [DT_STR]	2			1252 (AN)
country_region	Copia de country_regi...	string [DT_STR]	50			1252 (AN)
sub_region_1	Copia de sub_region_1	string [DT_STR]	50			1252 (AN)
sub_region_2	Copia de sub_region_2	string [DT_STR]	50			1252 (AN)
metro_area	Copia de metro_area	string [DT_STR]	50			1252 (AN)
iso_3166_2_code	Copia de iso_3166_2_c...	string [DT_STR]	10			1252 (AN)
census_fips_code	Copia de census_fips...	string [DT_STR]	50			1252 (AN)
date	Copia de date	date [DT_DATE]				
retail_and_recreation_percent_change_from_baseline	Copia de retail_and_re...	float [DT_R4]				
grocery_and_pharmacy_percent_change_from_baseline	Copia de grocery_and...	float [DT_R4]				
parks_percent_change_from_baseline	Copia de parks_perce...	float [DT_R4]				
transit_stations_percent_change_from_baseline	Copia de transit_stati...	float [DT_R4]				
workplaces_percent_change_from_baseline	Copia de workplaces_...	float [DT_R4]				

**Configurar la salida de errores...** **Aceptar** **Cancelar** **Ayuda**

**Paso 5:** Los porcentajes son archivos del mismo tipo y siempre están enlazados entre sí, al pensar en separarlos en su propia Tabla se tiene que crear un dato que no existe que será su columna de ID, para esto usamos sus propios datos sumándose, multiplicándose y dividiéndose (El divisor se suma con constantes para que nunca sea cero) creando así una llave muy poco común, pero que aún es posible que se repita, por lo que se usa otra Derived Column para concatenarle su Fecha, Su país y su subRegión.

Nombre de columna derivada	Columna derivada	Expresión	Tipo de datos
ID_Porcentajes	<agregar como columna>	(([Copia de retail_and_recreation_percent_change_from_baseline]+[Copia de grocery_and_pharmacy_percent_change_from_baseline])*([Copia de parks_percent_change_from_baseline]+[Copia de transit_stations_percent_change_from_baseline]))/(1.265656+([Copia de	float [DT_R4]

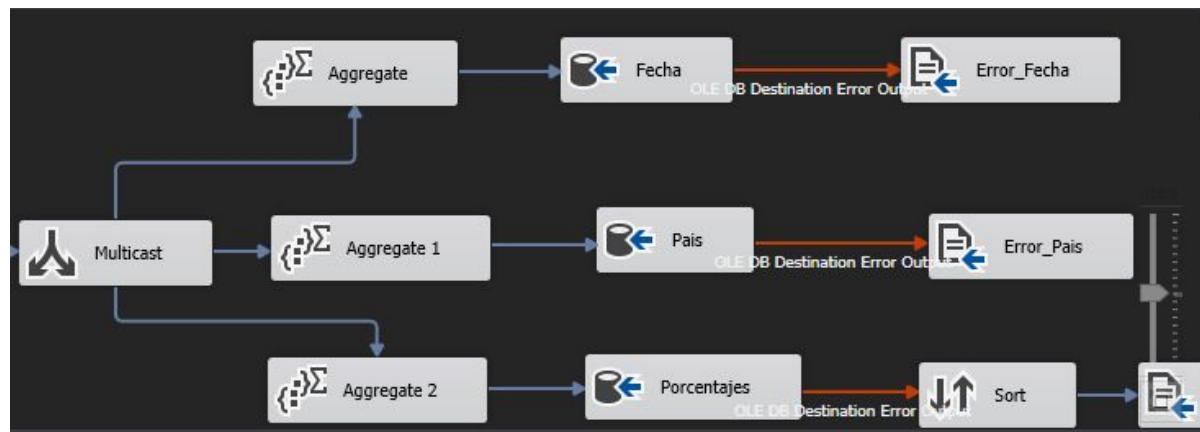
(([Copia de retail\_and\_recreation\_percent\_change\_from\_baseline]+[Copia de grocery\_and\_pharmacy\_percent\_change\_from\_baseline])\*([Copia de parks\_percent\_change\_from\_baseline]+[Copia de transit\_stations\_percent\_change\_from\_baseline]))/(1.265656+([Copia de

*workplaces\_percent\_change\_from\_baseline]+[Copia  
residential\_percent\_change\_from\_baseline]))*

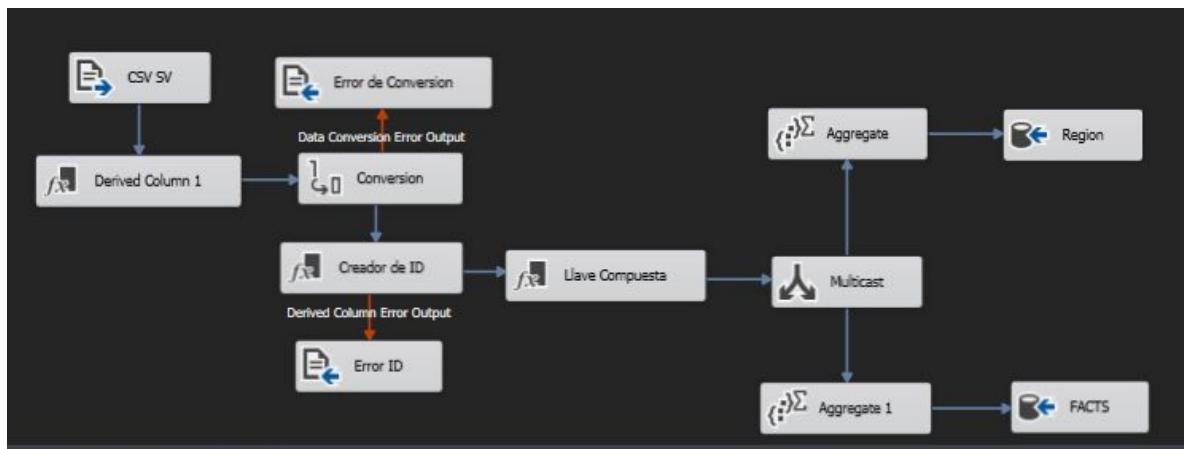
Nombre de columna d...	Columna derivada	Expresión	Tipo de datos	L...
Copia de ID_Porcentajes	<agregar como colum...	(DT_STR,50,1252)(country_region_code + iso_3166_...	string [DT_STR]	5

*(DT\_STR,50,1252)(country\_region\_code + iso\_3166\_2\_code + date + (DT\_STR,50,1252)ID\_Porcentajes)*

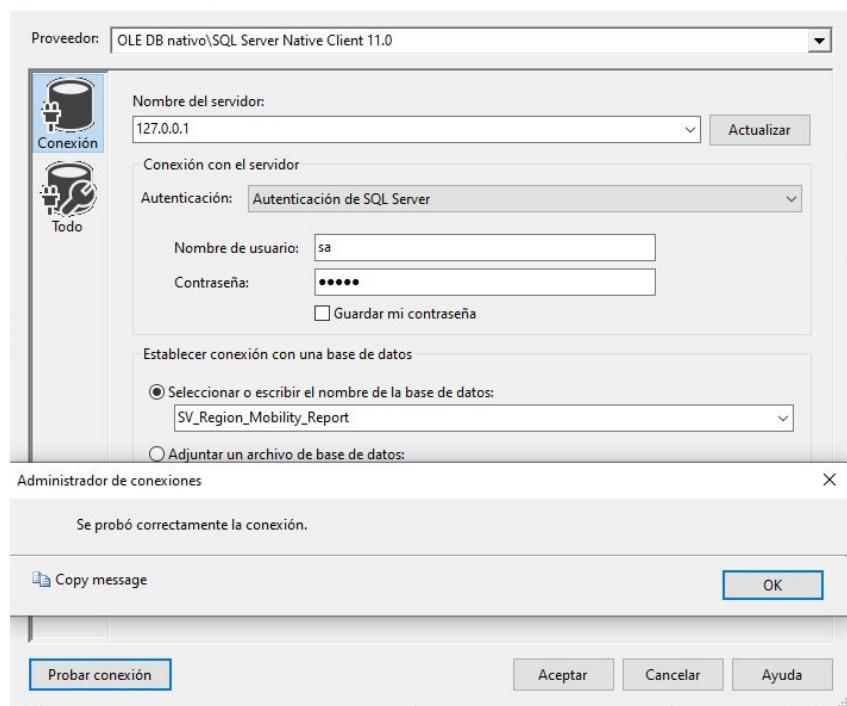
**Paso 6:** Con el uso de Multicast se pasa esta información a distintos Aggregate para que agrupen la información según la tabla lo necesite



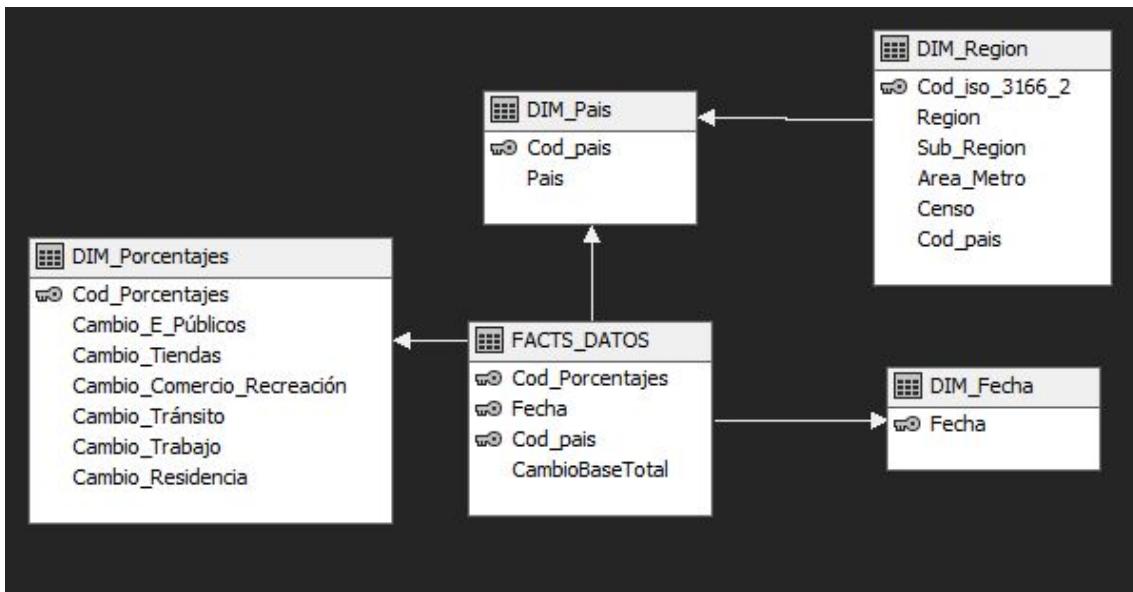
**Paso 7:** Se copia la estructura casi completa del ETL Anterior, pero a partir del multicast se cambian los aggregate para que nuevamente, este se adecue a las asignaciones que piden la Tabla Regiones y la Tabla Hechos



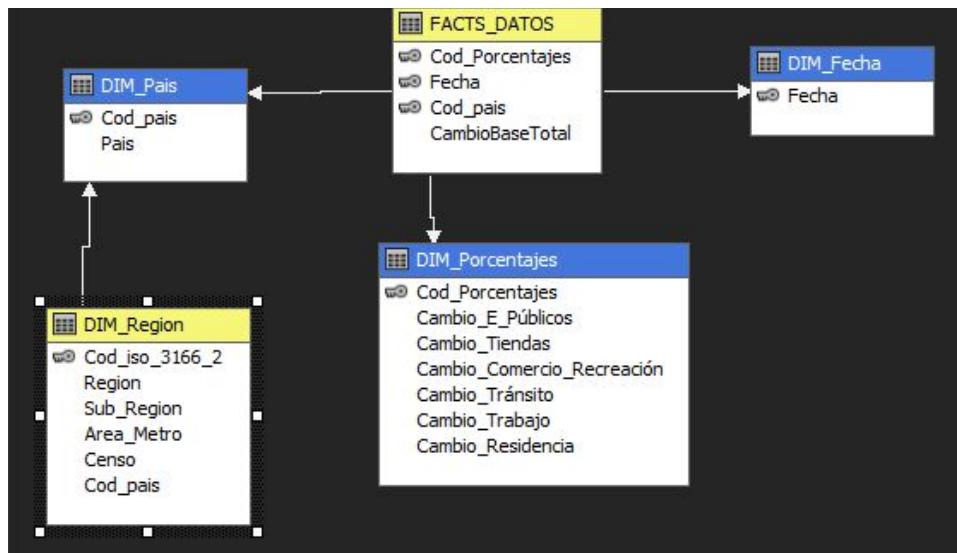
**Paso 8:** Se crea un proyecto de Analysis Services Multidimensional y se selecciona la base de datos creada como Origen de Datos, en este caso se uso la Autenticación de SQL Server y el servidor 127.0.0.1 (localhost)



**Paso 9:** Crear la vista tomando el origen y todas las tablas contenidos en este



**Paso 10:** Crear el cubo tomando la tabla Fact y la DimRegion como puntos de medida, esta última agregada únicamente para hacer consultas con la dimensión de Región las cuales son más específicas que hacerlas únicamente por país



**Paso 11:** Se crean jerarquías para la Dimensión Región para que la Región, SubRegión, Metro y Censo sean juntos un mismo dato, y otra jerarquía en los porcentajes, están únicamente para agruparlos y hacer consultas de todos sin necesidad de traerlos uno por uno



**Paso 12:** Procesar y Examinar el cubo para hacer consultas

Region	Fecha	Cambio Base Total
Ahuachapán Department	2020-02-15	87
Ahuachapán Department	2020-02-16	28
Ahuachapán Department	2020-02-17	28
Ahuachapán Department	2020-02-18	13
Ahuachapán Department	2020-02-19	-38
Ahuachapán Department	2020-02-20	-16
Ahuachapán Department	2020-02-21	121
Ahuachapán Department	2020-02-22	-1
Ahuachapán Department	2020-02-23	-32
Ahuachapán Department	2020-02-24	-14

Al hacer consultas como estas podemos ver detalles curiosos como que a partir de mediados de marzo el cambio base total daba picos negativos tan alto, y para mayores detalles podemos traer la jerarquía de porcentajes

Fecha	Cambio Trabajo	Cambio Comercio Recreac...	Cambio Tiendas	Cambio Tránsito	Cambio E Públicos
2020-02-17	3	-1	2	4	1
2020-02-17	4	1	6	-1	-1
2020-02-17	4	-7	2	0	0
2020-02-17	5	0	0	0	-2
2020-02-17	5	0	4	1	-3
2020-02-17	5	3	0	0	-7
2020-02-17	6	-1	2	4	-3
2020-02-17	7	0	0	0	-1
2020-02-17	7	0	0	0	-9
2020-02-17	9	0	0	0	-2
2020-02-17	9	-1	-2	0	-14

Podemos ver cómo se mantenía una estabilidad en febrero que muchas veces se mantenía con números de cero o pequeños, siendo casi todos los cambios en tránsito, sitios públicos y el trabajo.

Fecha	Cambio Trabajo	Cambio Comercio Recreac...	Cambio Tiendas	Cambio Tránsito	Cambio E Públicos	Cambio
2020-03-30	-57	-60	-49	-90	-40	0
2020-03-30	-57	-65	-41	0	-46	0
2020-03-30	-57	-66	-47	-60	-44	27
2020-03-30	-59	-66	-42	-69	-50	27
2020-03-30	-64	-69	-45	-64	-60	32
2020-03-30	-64	-70	-45	-64	-63	29
2020-03-30	-68	-71	-48	-63	-70	35
2020-03-30	-68	-75	-51	-58	-66	38
2020-03-31	-50	-49	0	0	-56	0
2020-03-31	-51	-53	0	-40	-59	0
2020-03-31	-52	0	0	0	-53	0

Esta tendencia de estabilidad cambia en finales de Marzo cuando podemos encontrar picos negativos en la mayoría de los rubros estudiados