Universidad Don Bosco

Facultada de Ingeniería

Escuela de computación

Asignatura: Data Warehouse y Minería de Datos

Desafío 2 DMD

Docente: Ing. Karens Medrano

Integrantes:

Apellidos	Nombres	Carnet
Lemus Cardoza	Nelson Orlando	LC111108
López Revelo	Cristian Odir	LR161911
Barriere Campos	Gerson Daniel	BC200025

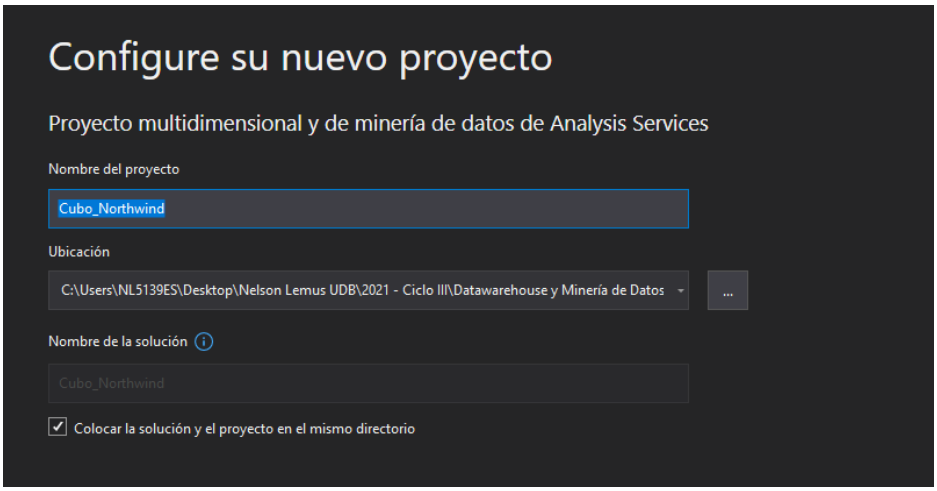
Viernes 25 de junio de 2021

Ejercicio 1

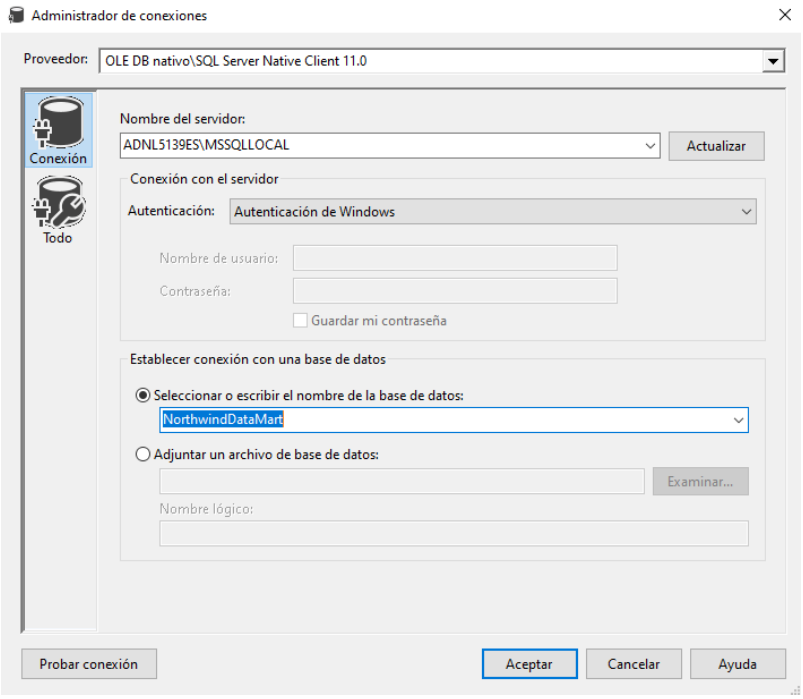
PROCESO DE ETL: Análisis multidimensional OLAP (Datamart de Northwind)

- **Elabore un cubo en donde pueda visualizar nombre y país del cliente, apellido del empleado, país del proveedor, y nombre y categoría del producto.**
- **Crear un nuevo campo concatenando la categoría del producto con el nombre del producto.**

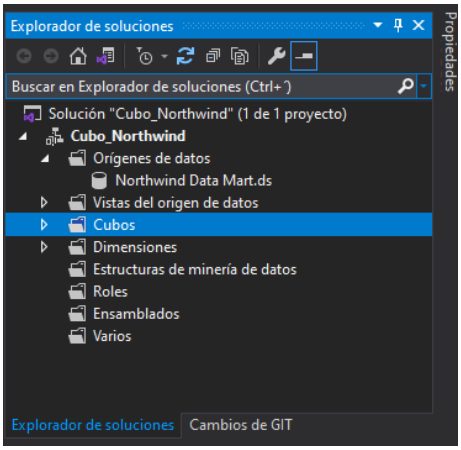
1. Creamos un proyecto llamado Cubo_Northwind en “Proyecto multidimensional y de minería de datos de Analysis Services”.



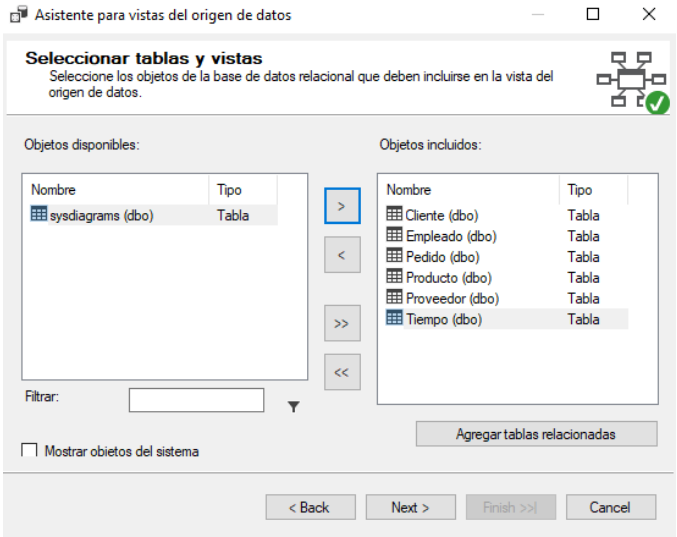
2. Iniciamos la conexión hacia SQL y seleccionamos la base de datos NortwindDataMart

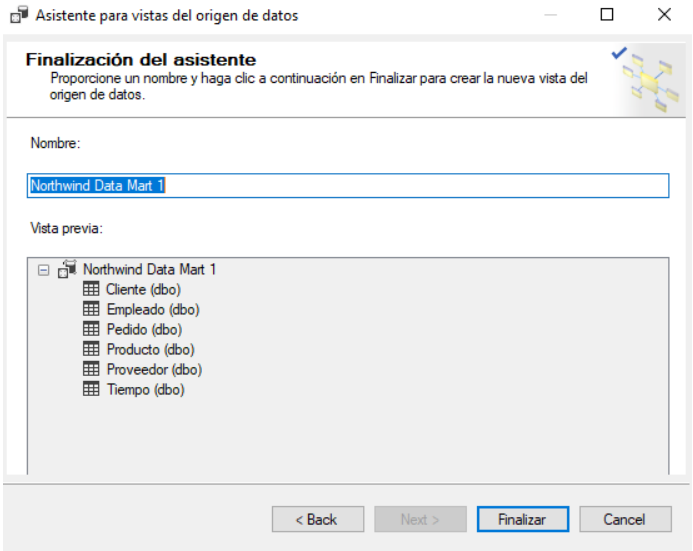


3. Hasta el momento el proyecto queda según imagen.

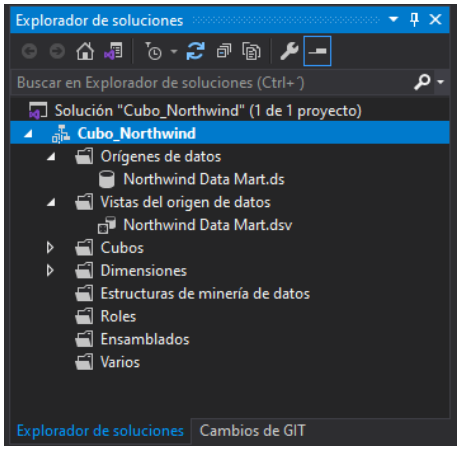


4. Ahora creamos la vistas en la conexión del proyecto, seleccionamos las tablas "Cliente, Empleado, Pedido, Producto, Proveedor y Tiempo" para incluirlos en Objetos, luego siguiente y finalizamos el asistente.

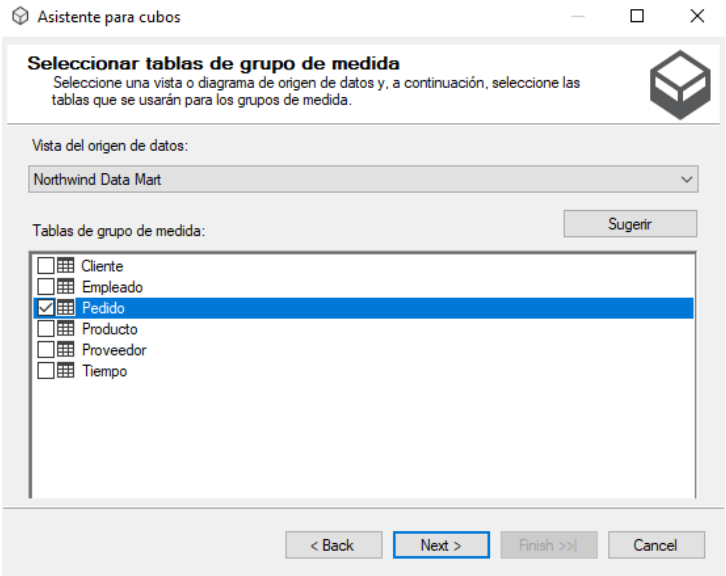


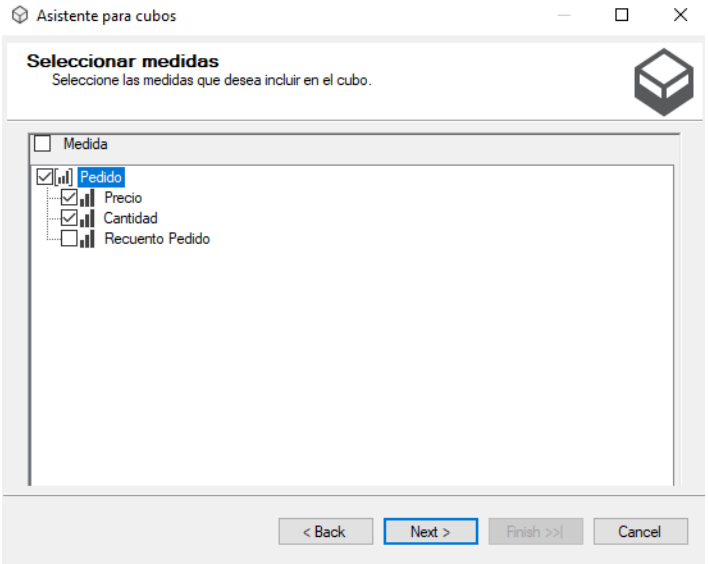


5. Hasta el momento el proyecto queda según imagen.

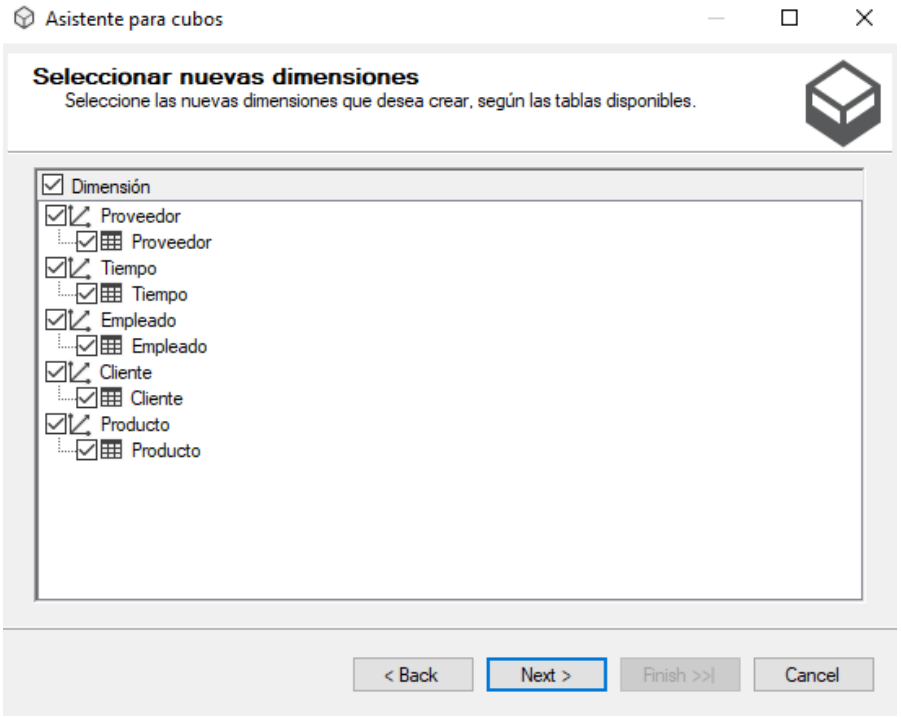


6. Ahora creamos el cubo y seleccionamos la tabla Pedido como medida ya que contiene los datos para conectar a las demás tablas.

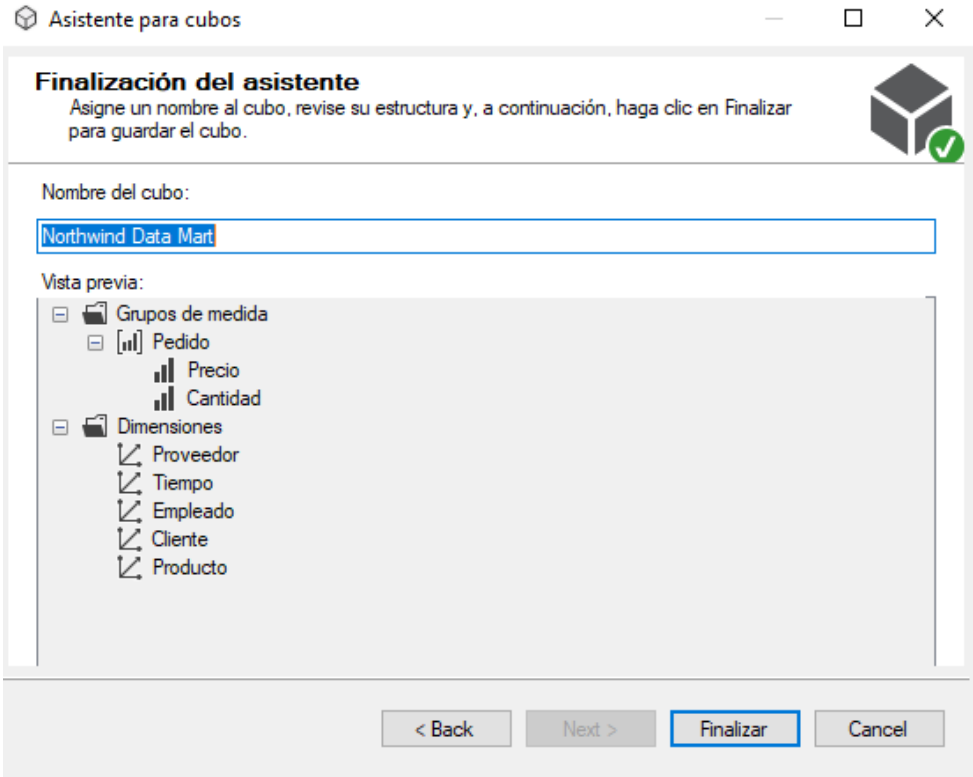




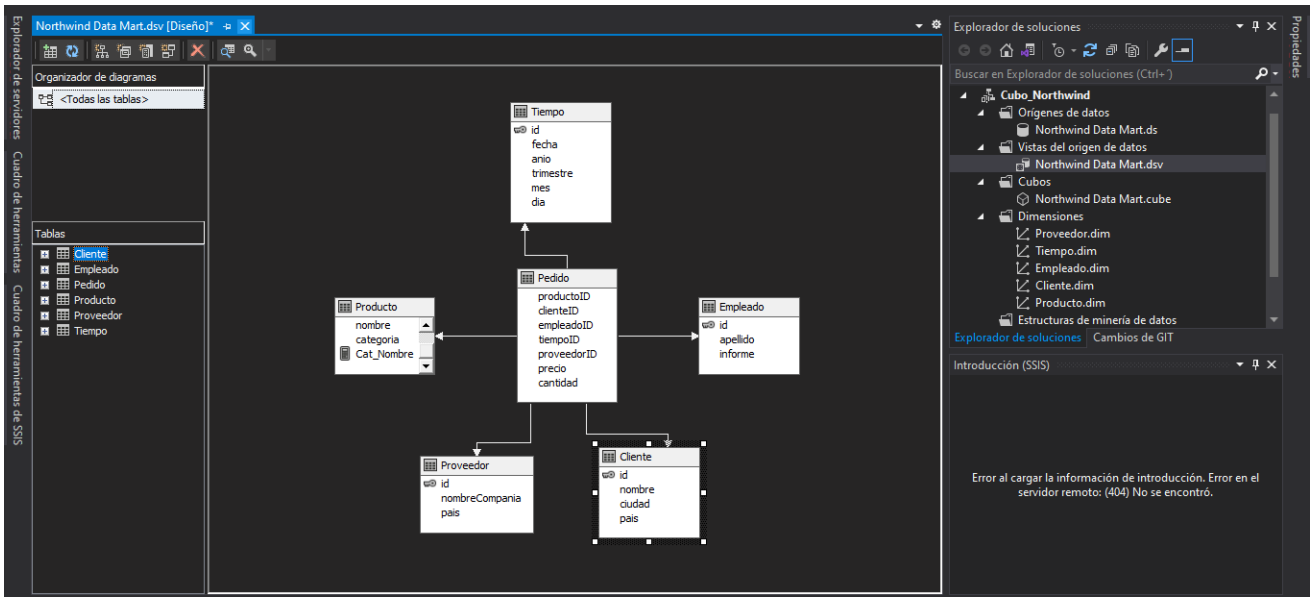
7. Seleccionamos las dimensiones que están conformadas por las tablas Proveedor, Tiempo, Empleado, Cliente y Producto.



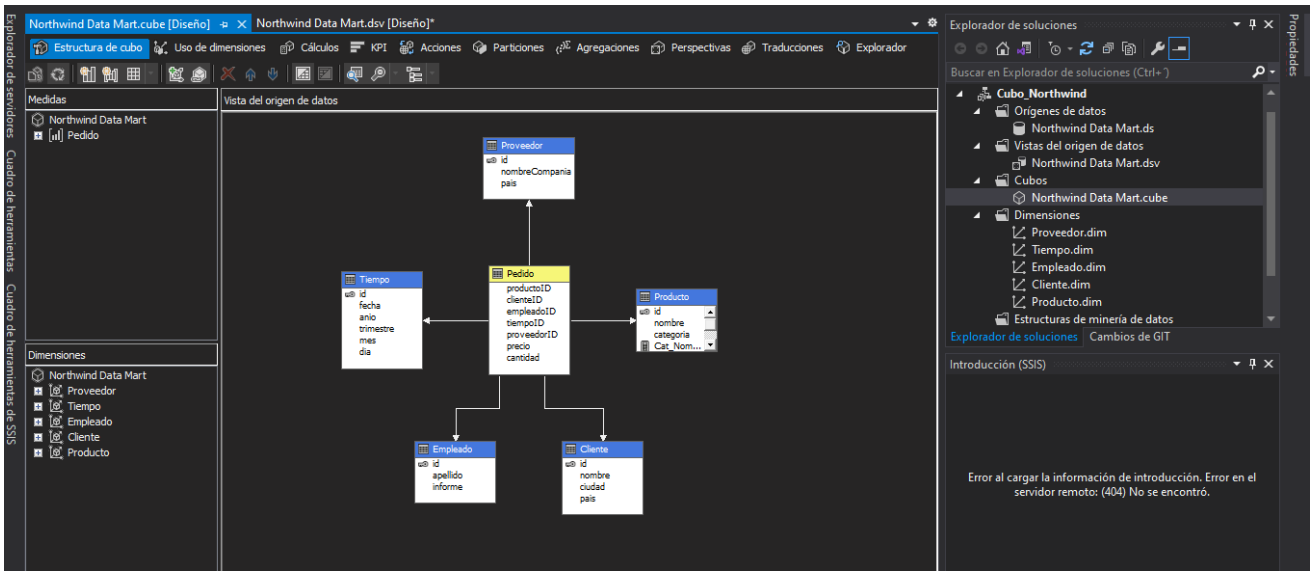
8. Finalizamos el asistente asegurando que Pedido es la medida y las dimensiones las tablas del paso anterior.



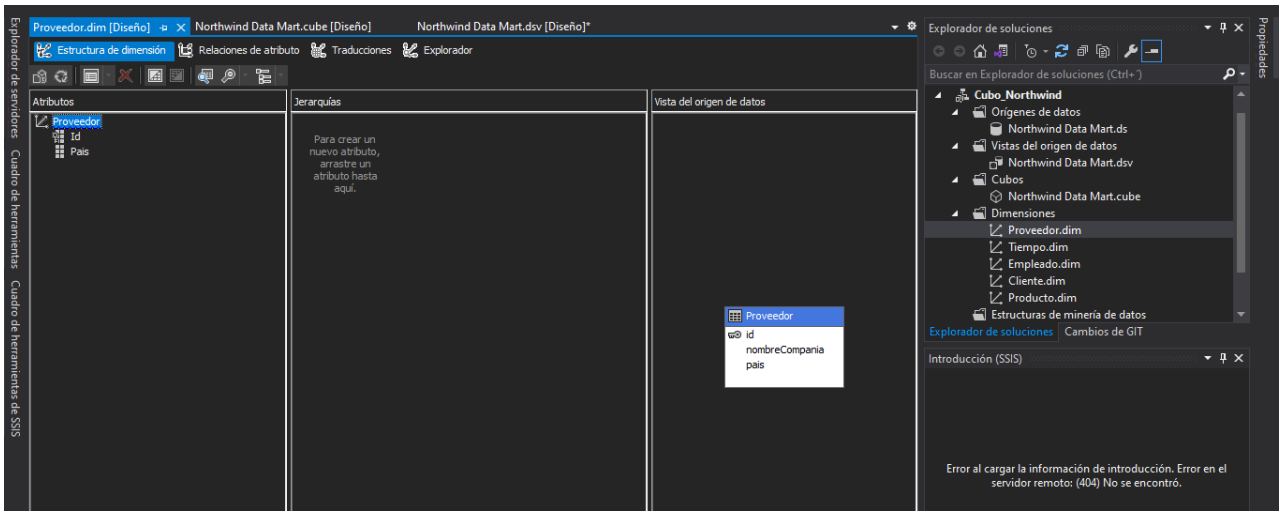
9. La vista del proyecto queda de la siguiente manera, el modelo estrella donde la tabla Pedido es la tabla de hechos y conecta al resto de tablas.



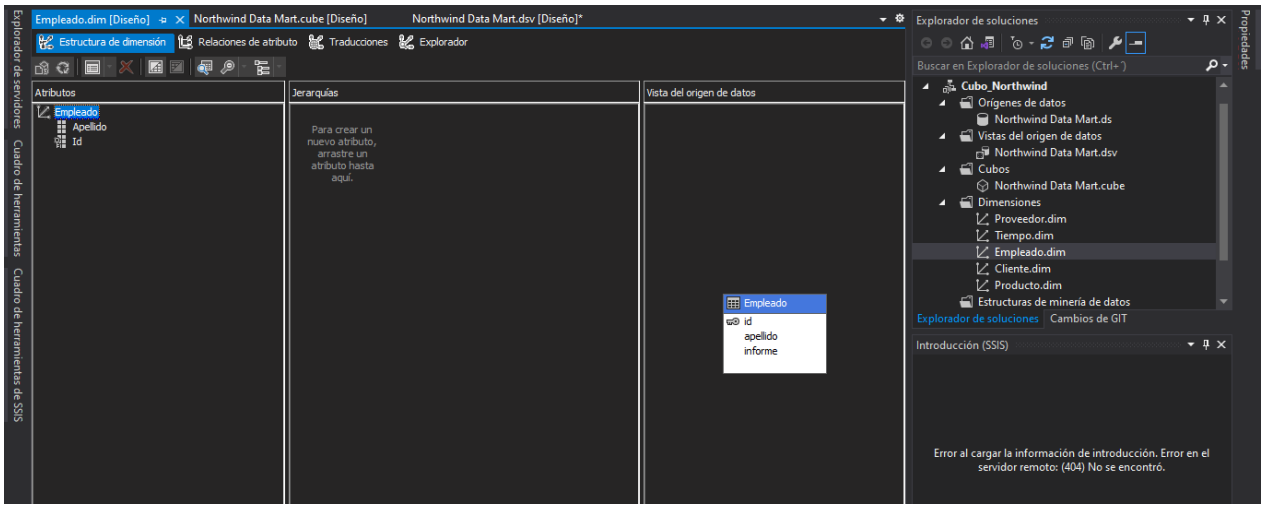
10. El cubo del proyecto queda de la siguiente manera, la tabla pedido como medida que contiene los datos de los pedidos.



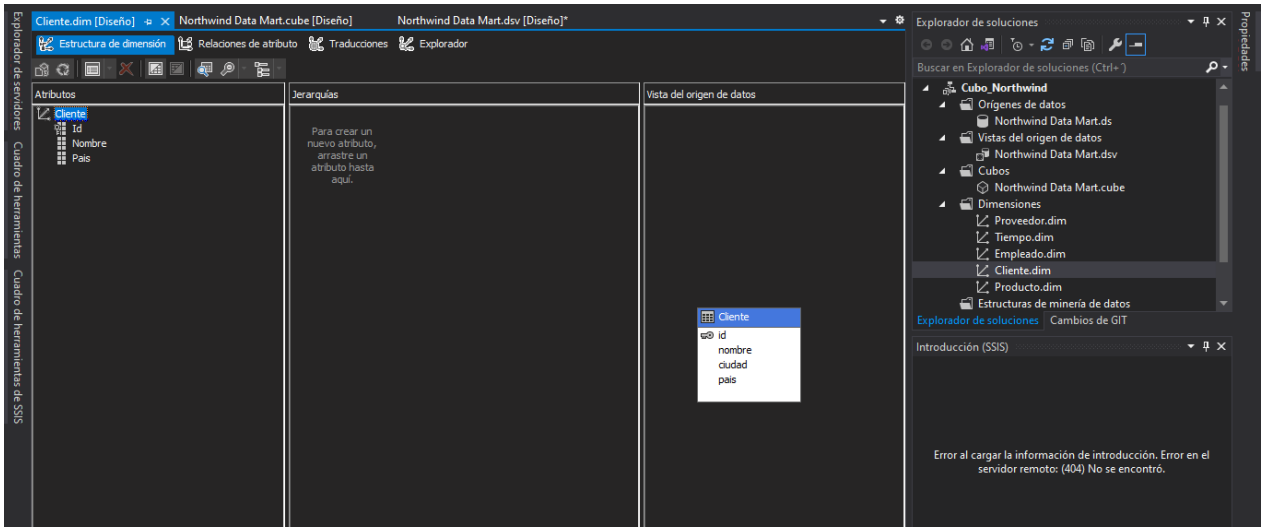
11. Ahora damos doble click a cada dimesion para agregar los atributos con los que exploraremos los datos. Agregando País Proveedor



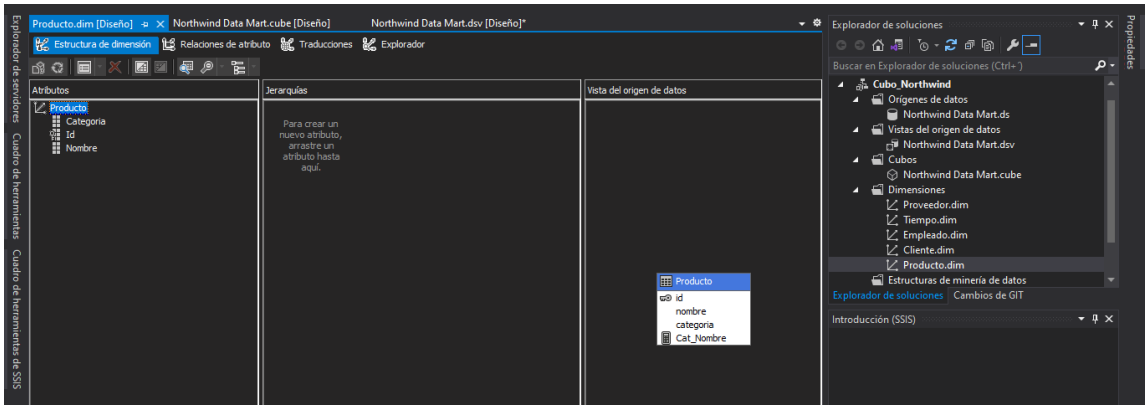
12. Agregando Apellido Empleado.



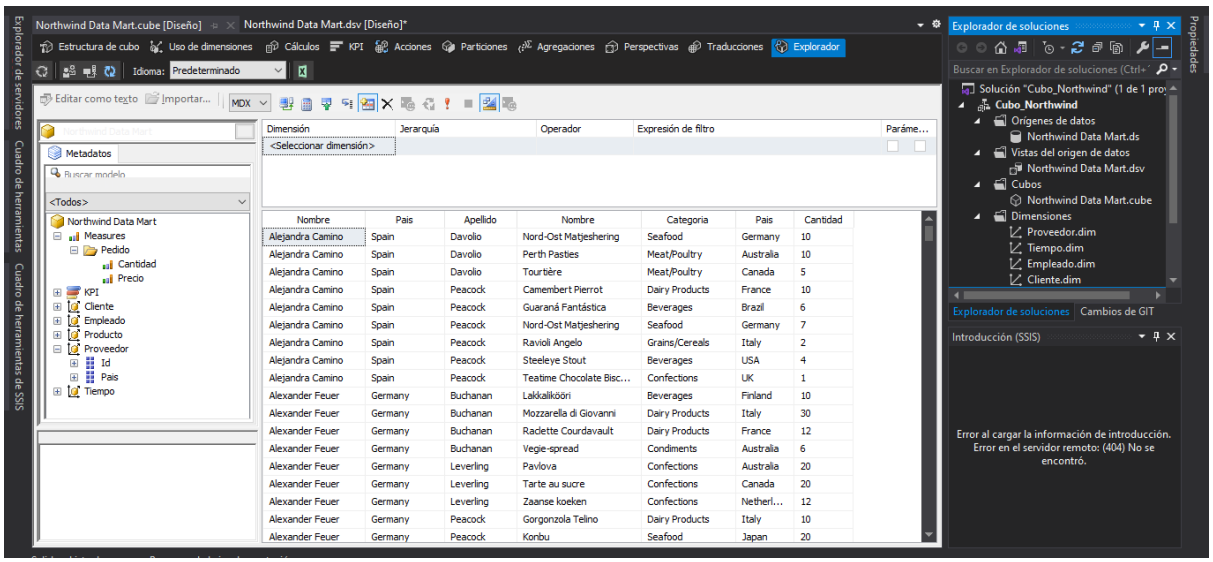
13. Agregando Nombre y País Cliente.



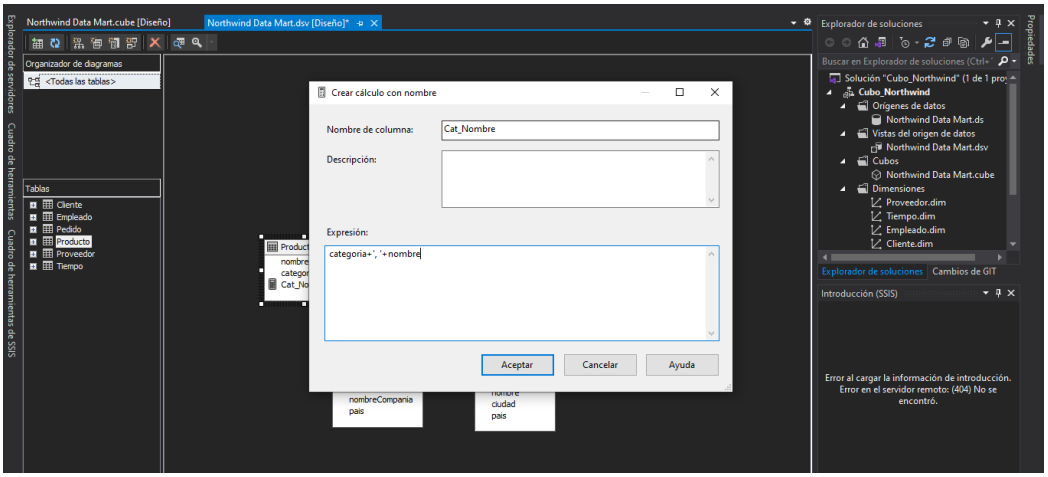
14. Agregando ID y Categoría de Producto.



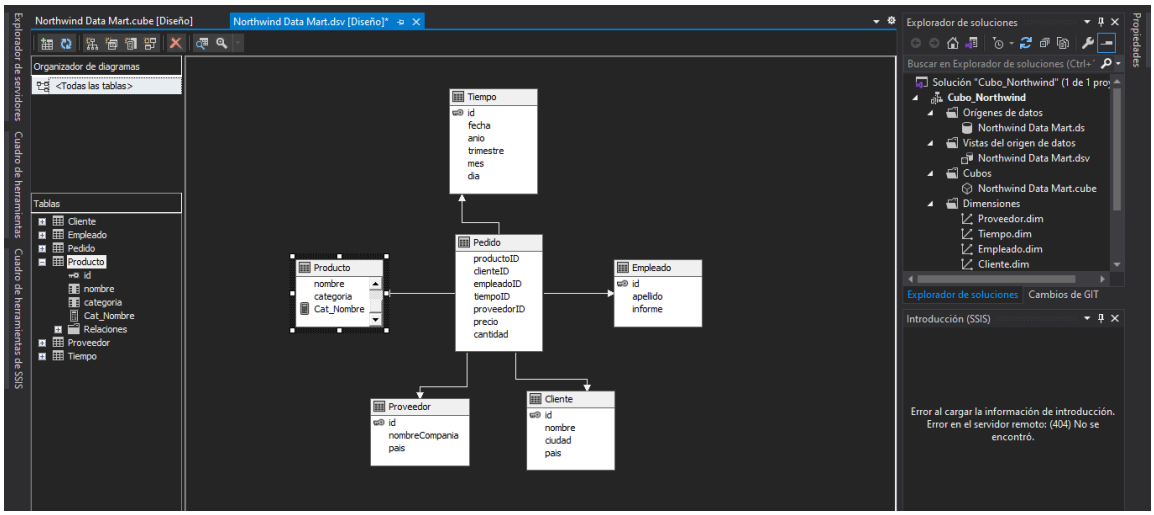
15. Ahora exploramos los datos para visualizar los datos de Nombre Cliente, País Cliente, Apellido Empleado, Nombre Categoria, Categoria (Tipo), País de Proveedor, agregamos la medida cantidad para visualizar la consulta con datos generados.



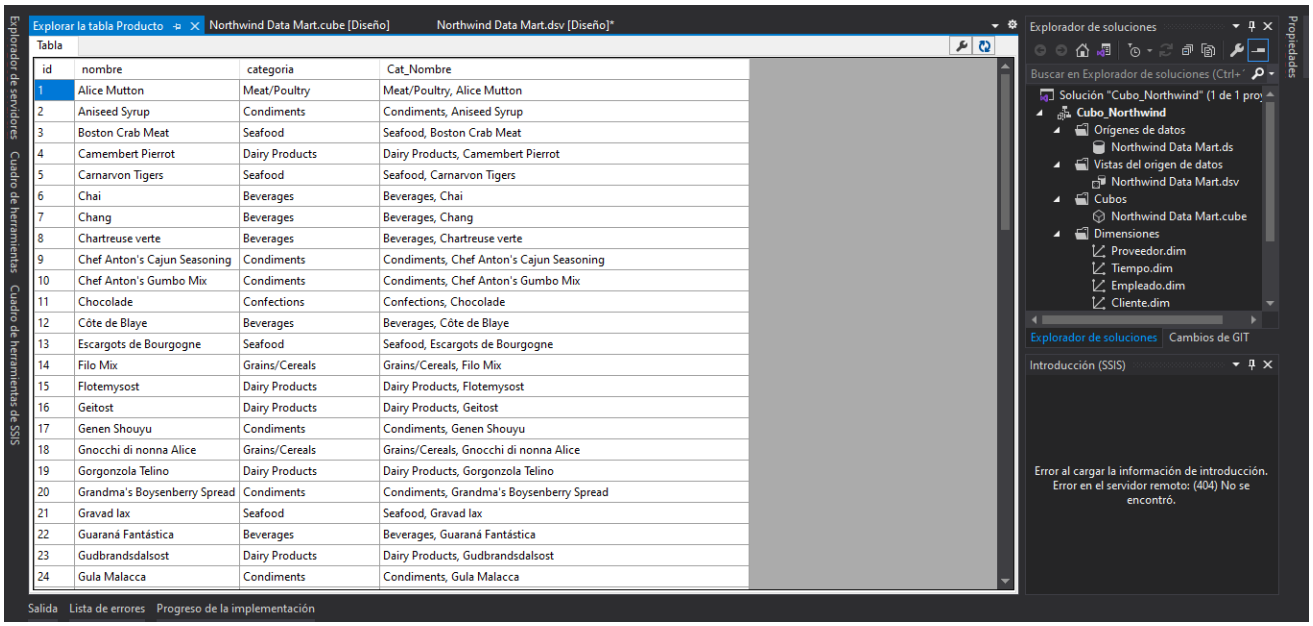
16. Ahora vamos a agregar el campo calculado en la tabla pedido para unir categoría con el nombre y nombraremos el campo como Cat_Nombre.



17. Visualizamos que el campo queda agregado como atributo de la dimensión productos.



18. Exploramos los datos de la dimensión Producto y confirmamos que el campo agregado Cat_Nombre se ejecutó de forma correcta.



id	nombre	categoria	Cat_Nombre
1	Alice Mutton	Meat/Poultry	Meat/Poultry, Alice Mutton
2	Aniseed Syrup	Condiments	Condiments, Aniseed Syrup
3	Boston Crab Meat	Seafood	Seafood, Boston Crab Meat
4	Camembert Pierrot	Dairy Products	Dairy Products, Camembert Pierrot
5	Carnarvon Tigers	Seafood	Seafood, Carnarvon Tigers
6	Chai	Beverages	Beverages, Chai
7	Chang	Beverages	Beverages, Chang
8	Chartreuse verte	Beverages	Beverages, Chartreuse verte
9	Chef Anton's Cajun Seasoning	Condiments	Condiments, Chef Anton's Cajun Seasoning
10	Chef Anton's Gumbo Mix	Condiments	Condiments, Chef Anton's Gumbo Mix
11	Chocolade	Confections	Confections, Chocolade
12	Côte de Blaye	Beverages	Beverages, Côte de Blaye
13	Escargots de Bourgogne	Seafood	Seafood, Escargots de Bourgogne
14	Filo Mix	Grains/Cereals	Grains/Cereals, Filo Mix
15	Flotemysost	Dairy Products	Dairy Products, Flotemysost
16	Geitost	Dairy Products	Dairy Products, Geitost
17	Genen Shouyu	Condiments	Condiments, Genen Shouyu
18	Gnocchi di nonna Alice	Grains/Cereals	Grains/Cereals, Gnocchi di nonna Alice
19	Gorgonzola Telino	Dairy Products	Dairy Products, Gorgonzola Telino
20	Grandma's Boysenberry Spread	Condiments	Condiments, Grandma's Boysenberry Spread
21	Gravad lax	Seafood	Seafood, Gravad lax
22	Guaraná Fantástica	Beverages	Beverages, Guaraná Fantástica
23	Gudbrandsdalsost	Dairy Products	Dairy Products, Gudbrandsdalsost
24	Gula Malacca	Condiments	Condiments, Gula Malacca

Ejercicio 2

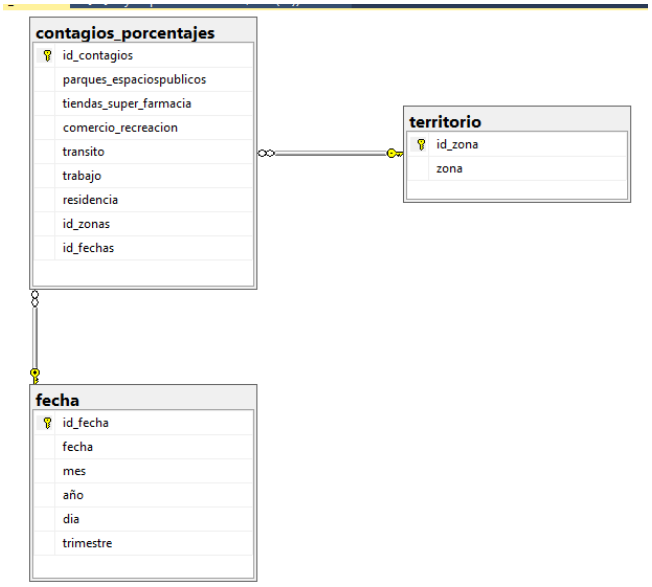
PROCESO DE ETL: Análisis para El Salvador COVID19

19. **Abrimos SQL Server y creamos la siguiente base de datos, esta representara nuestro modelo tipo estrella.**

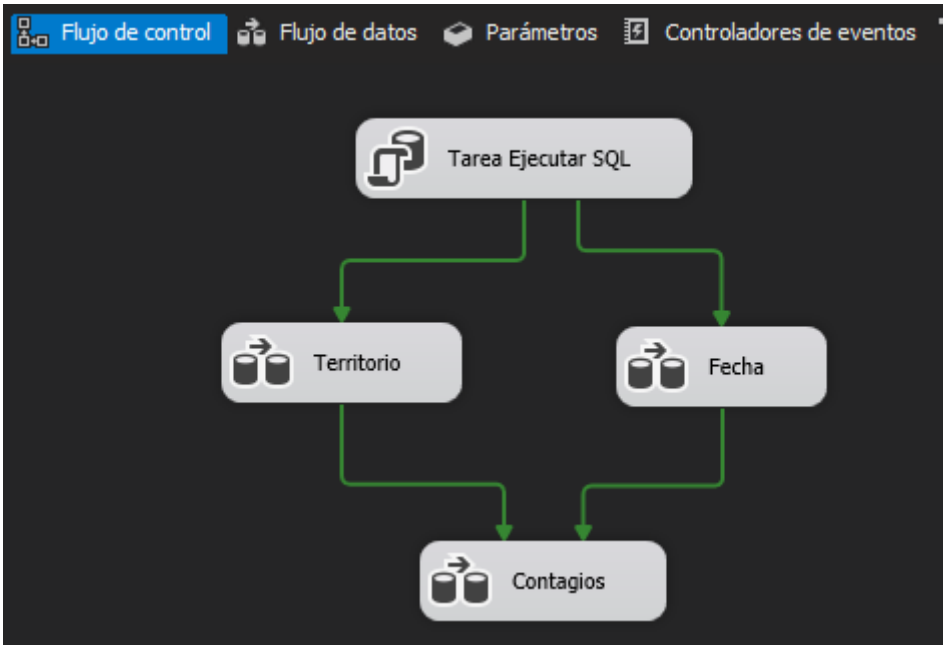
```sql
create database covid19;
use covid19

create table territorio
(
id_zona int identity(1,1) primary key,
zona varchar(50)
)
go
create table fecha
(
id_fecha int identity(1,1) primary key,
fecha date,
mes varchar(15),
año int,
dia int,
trimestre int
)
go
create table contagios_porcentajes
(
id_contagios int identity(1,1) primary key,
parques_espaciospublicos int,
tiendas_super_farmacia int,
comercio_recreacion int,
transito int,
trabajo int,
residencia int,
id_zonas int foreign key references territorio(id_zona),
id_fechas int foreign key references fecha(id_fecha)
)
go
```

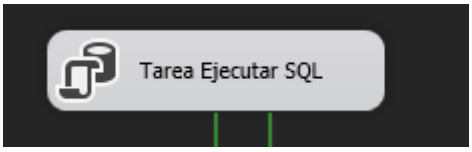
20. **Una vez ejecutado el script, nuestro diagrama se muestra de la siguiente manera, como vemos hemos segmentado la información para lograr formar nuestras tablas dimensiones y tabla hecho**



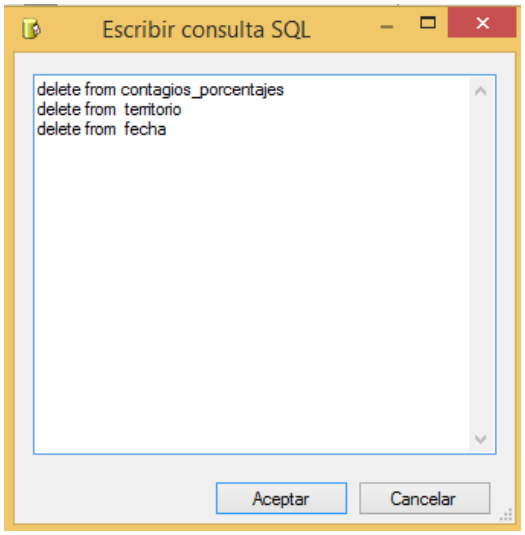
21. **Creamos un nuevo proyecto SSIS en visual studio y agregaremos los siguientes componentes a nuestro flujo de control. El orden en que se ejecutan es importante debido las relaciones de llaves foráneas y primarias que existe por lo que primero se ejecuta y llenan los registros para las tablas Territorio y Fecha y luego en Contagios**



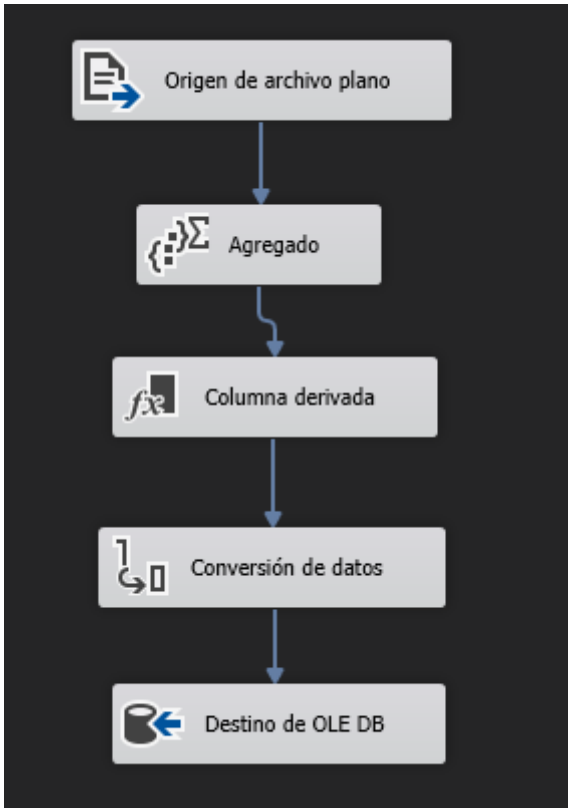
22. **Dentro de cada componente agregaremos los siguientes controles, que ayudaran a llenar las tablas, para la Tarea Ejecutar SQL será un Script que se ejecuta antes del llenado de registros a nuestras tablas**



Conjunto de resultados	
ResultSet	**Ninguno**
General	
Name	**Tarea Ejecutar SQL**
Description	**Tarea Ejecutar SQL**
Instrucción SQL	
ConnectionType	**OLE DB**
Connection	**LOPEZREVELO.covid19**
SQLSourceType	**Entrada directa**
SQLStatement	**delete from contagios_porcentajesdelete from**
IsQueryStoredProcedure	False
BypassPrepare	**True**
Opciones	
TimeOut	**0**
CodePage	**1252**
TypeConversionMode	**Permitido**



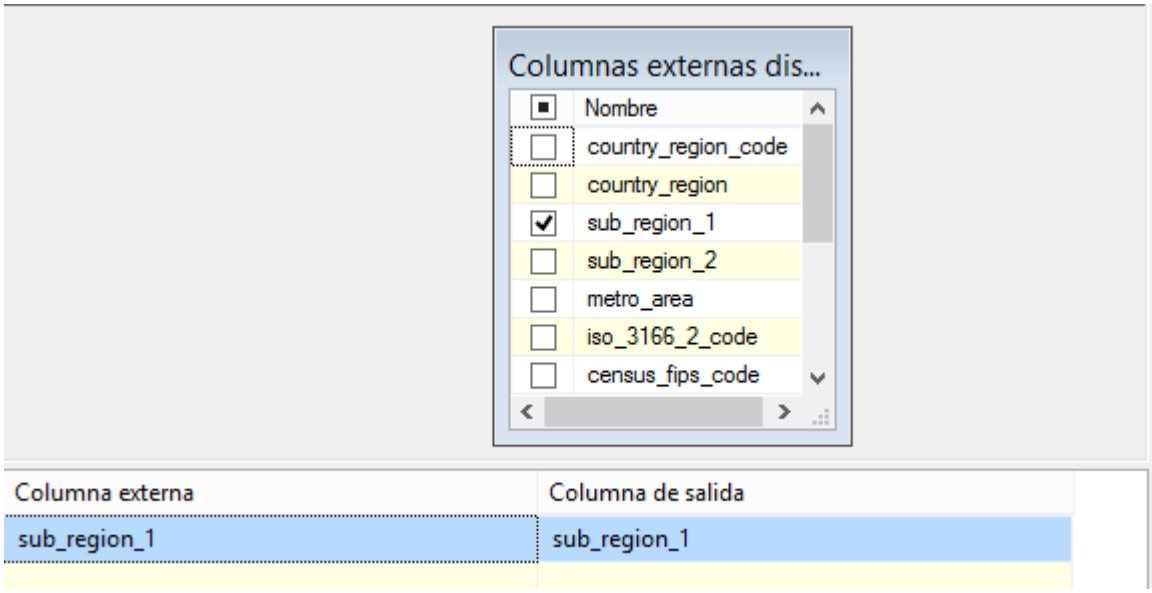
23. Dentro de territorio crearemos el siguiente esquema.



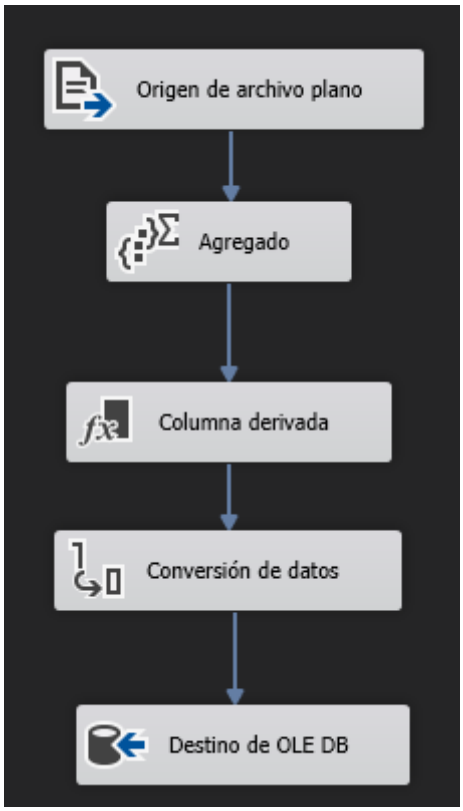
24. Dentro de columna derivada agregamos lo siguiente.

Nombre de columna d...	Columna derivada	Expresión	Tipo de datos
zona	<agregar como colum...	(DT_WSTR,50)sub_region_1 == "" ? "El Salvador" : sub_region_1	cadena Unicode [[

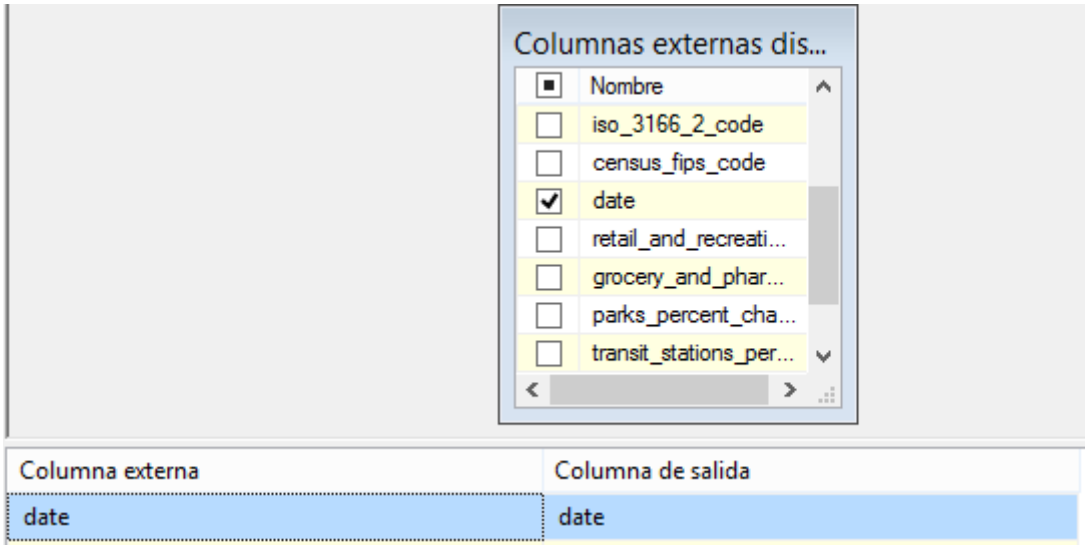
25. En origen de archivo seleccionaremos solamente el siguiente campo



26. Para el componenete fecha, creamos el siguiente esquema



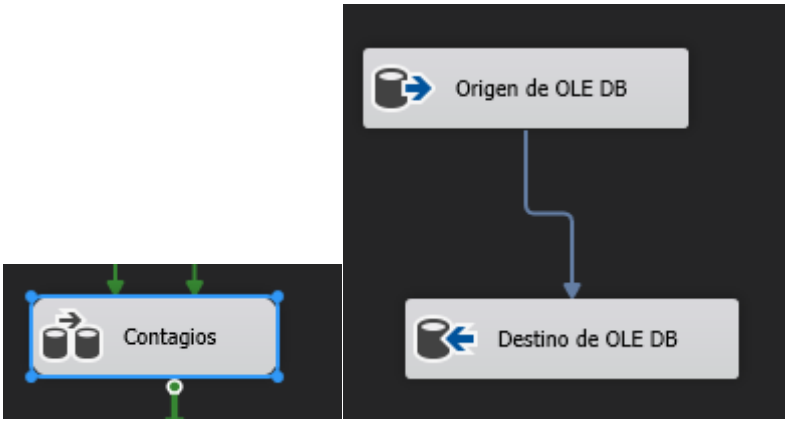
27. Dentro de Origen de archivo plano seleccionamos solamente el date.



28. En columna derivada agregamos la siguiente configuración.

Nombre de columna d...	Colu...	Expresión	Tipo de datos	Longitud
año	<agr...	DATEPART("YEAR",(DT_DBDATE)date)	entero de cuatro bytes ...	
trimestre	<agr...	DATEPART("QUARTER",(DT_DBDATE)date)	entero de cuatro bytes ...	
mes	<agr...	(DT_WSTR,15)DATEPART("MONTH",(DT_DBDATE)...	cadena Unicode [DT_...	15
dia	<agr...	DATEPART("DAY",(DT_DBDATE)date)	entero de cuatro bytes ...	
fecha_formateada	<agr...	(DT_DATE)date	fecha [DT_DATE]	

29. Dentro del componente Contagios, creamos el esquema que se muestra.



30. Ahora es importante haber hecho un vaciado de la información del archivo .csv en una base auxiliar para poder llenar nuestra tabla de relaciones. Dentro de Origen de OLE DB debemos agregar la siguiente sentencia, que tomara exclusivamente los campos que nosotros necesitaremos.

LOPEZREVELO.covid19

Modo de acceso a datos:

Comando SQL

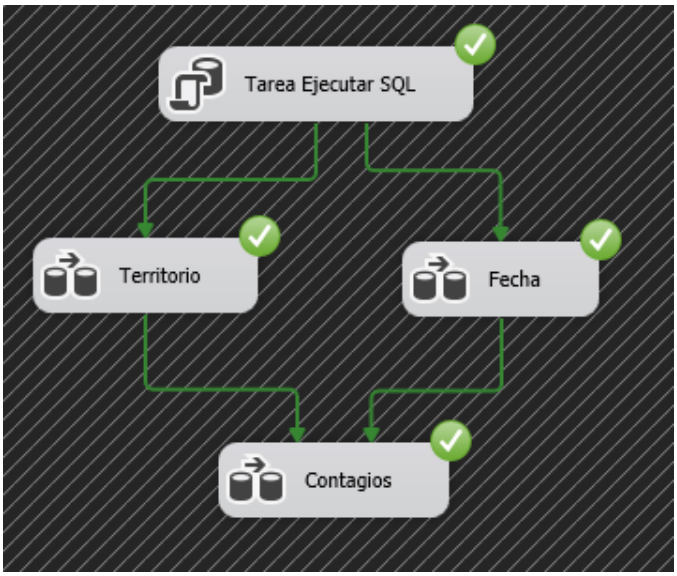
Texto de comando SQL:

```
SELECT
dt.parks_percent_change_from_baseline,
dt.grocery_and_pharmacy_percent_change_from_baseline,
dt.retail_and_recreation_percent_change_from_baseline,
dt.transit_stations_percent_change_from_baseline,
dt.workplaces_percent_change_from_baseline,
dt.residential_percent_change_from_baseline,
t.id_zona, f.id_fecha
FROM covid.dbo.dato dt INNER JOIN covid19.dbo.fecha f ON
(dt.fecha=f.fecha)
                                                    INNER
JOIN covid19.dbo.territorio t ON(dt.sub_region_1=t.zona)
```

15. Hacemos las respectivas asignaciones dentro de Destino OLE DB

Columna de entrada	Columna de destino
<omitir>	id_contagios
parks_percent_change_from_baseline	parques_espaciospublicos
grocery_and_pharmacy_percent_change_fro...	tiendas_super_farmacia
retail_and_recreation_percent_change_from_...	comercio_recreacion
transit_stations_percent_change_from_baseline	transito
workplaces_percent_change_from_baseline	trabajo
residential_percent_change_from_baseline	residencia
id_zona	id_zonas
id_fecha	id_fechas

16. Y finalmente ejecutamos nuestro proyecto



17. Verificamos si los datos se agregaron correctamente a nuestra base

```
select * from fecha
```

100 %

Resultados | Mensajes

	id_fecha	fecha	mes	año	dia	trimestre
1	453	2020-07-10	7	2020	10	3
2	454	2020-07-20	7	2020	20	3
3	455	2020-07-30	7	2020	30	3
4	456	2020-07-01	7	2020	1	3
5	457	2020-07-11	7	2020	11	3
6	458	2020-07-21	7	2020	21	3
7	459	2020-07-31	7	2020	31	3

```
select * from territorio
```

100 %

Resultados | Mensajes

	id_zona	zona
1	31	El Salvador
2	32	Santa Ana Department
3	33	Ahuachapán Department
4	34	La Unión Department
5	35	San Miguel Department
6	36	San Salvador Department
7	37	Usulután Department
8	38	La Paz Department
9	39	Sonsonate Department
10	40	La Libertad Department
11	41	Cabañas Department
12	42	Cuscatlán Department

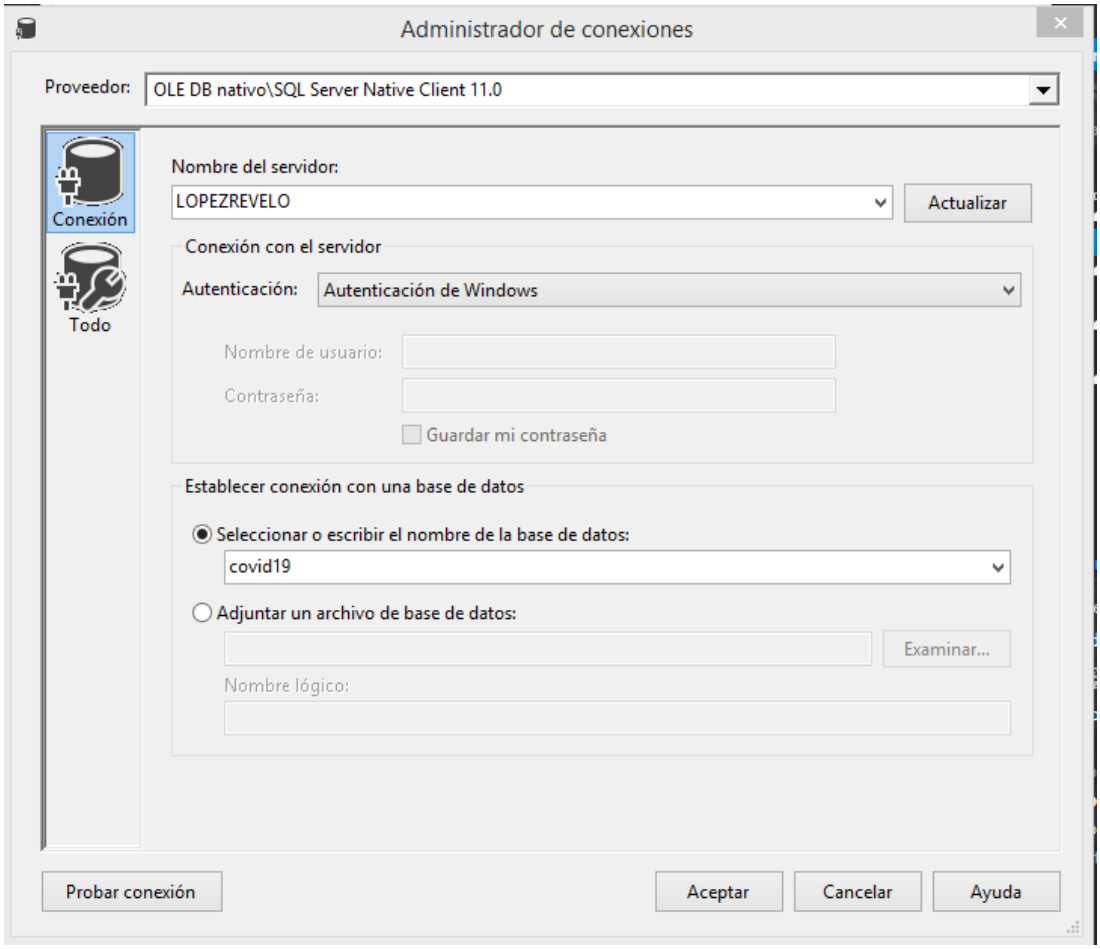
```
select * from contagios_porcentajes
```

Resultados | Mensajes

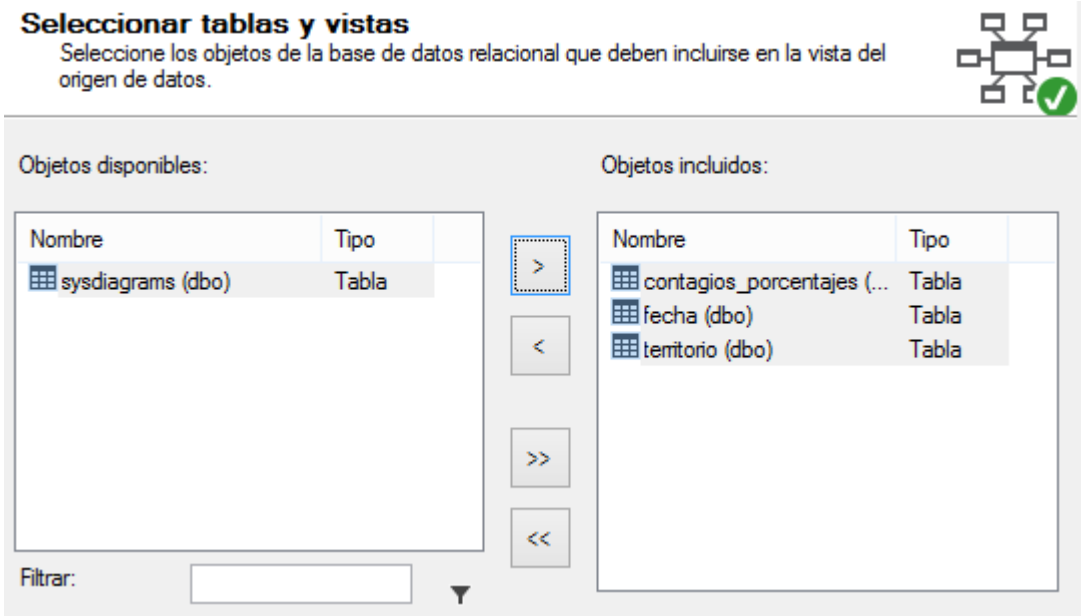
id_contagios	parques_espaciospublicos	tiendas_super_farmacia	comercio_recreacion	transito	trabajo	residencia	id_zonas	id_fechas
3391	0	5	4	-1	4	-1	31	489
3392	1	6	4	1	0	0	31	491
3393	-3	4	0	1	5	0	31	493
3394	0	0	0	3	4	-1	31	495
3395	-1	0	0	2	3	-1	31	497
3396	-4	-1	-1	3	3	0	31	484
3397	-4	2	0	2	8	-1	31	485
3398	-3	1	0	-3	4	0	31	486
3399	0	1	0	-3	1	0	31	487
3400	-3	1	-2	0	5	0	31	488
3401	-3	-1	0	-1	4	-1	31	490
3402	3	-3	-2	-1	4	0	31	492

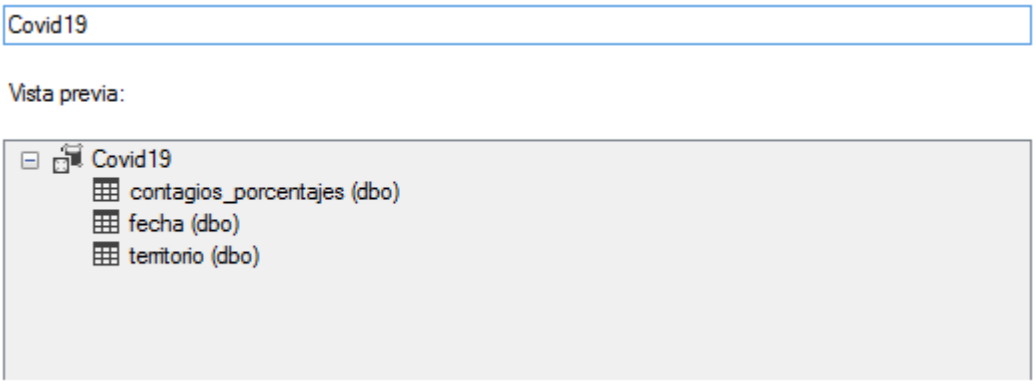
CREACION DE CUBO.

1. Creamos un nuevo proyecto Multidimensional en Visual Studio, seguido de esto agregaremos un nuevo origen de datos, que se conectare a nuestra base creada en el ETL.

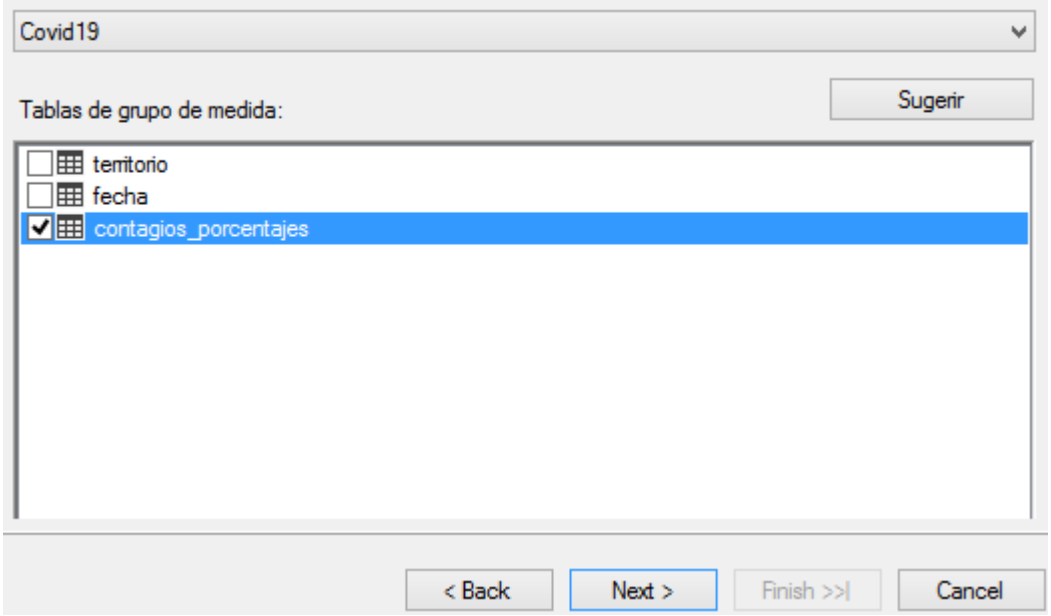


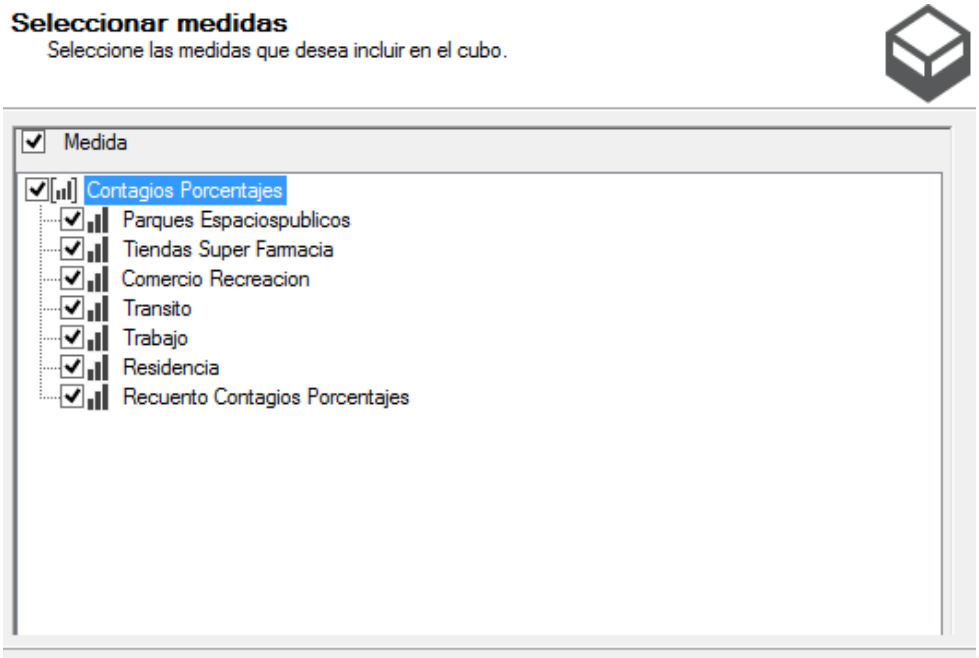
2. Creamos una nueva vista, seleccionando las tablas necesarias para nuestro cubo

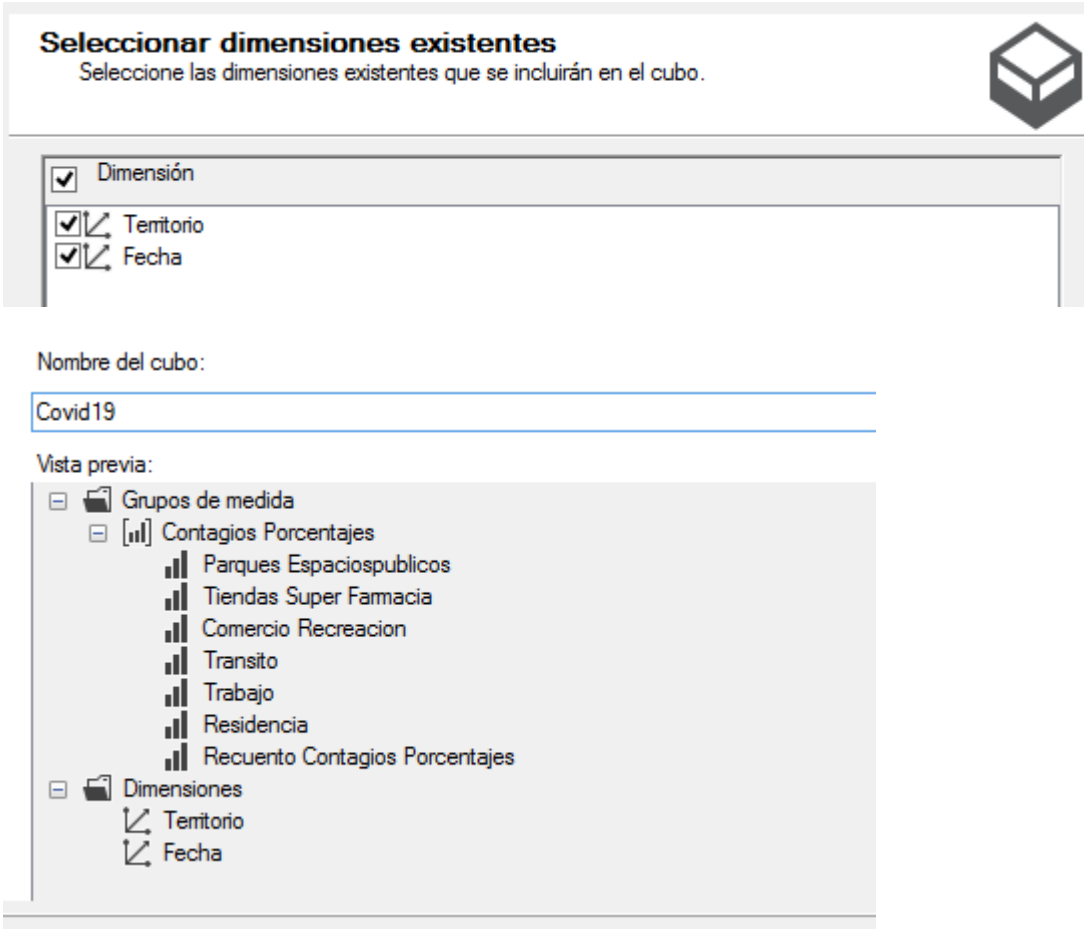




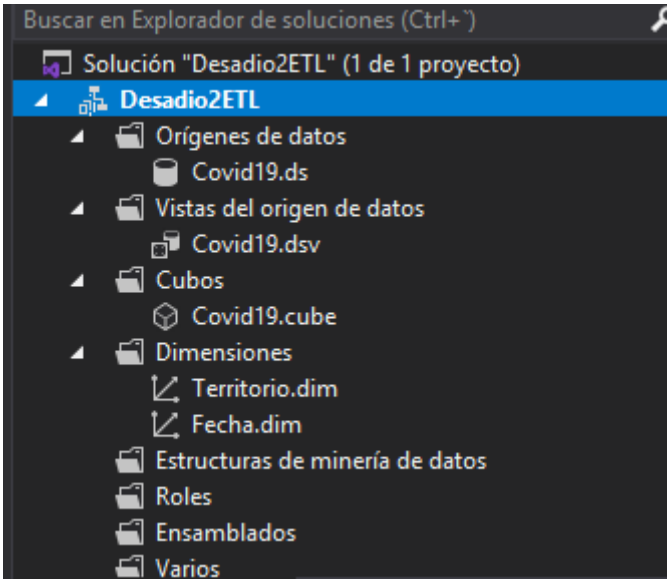
3. Ahora crearemos un nuevo cubo y seleccionaremos la tabla contagios_porcentajes como la tabla de medidas.



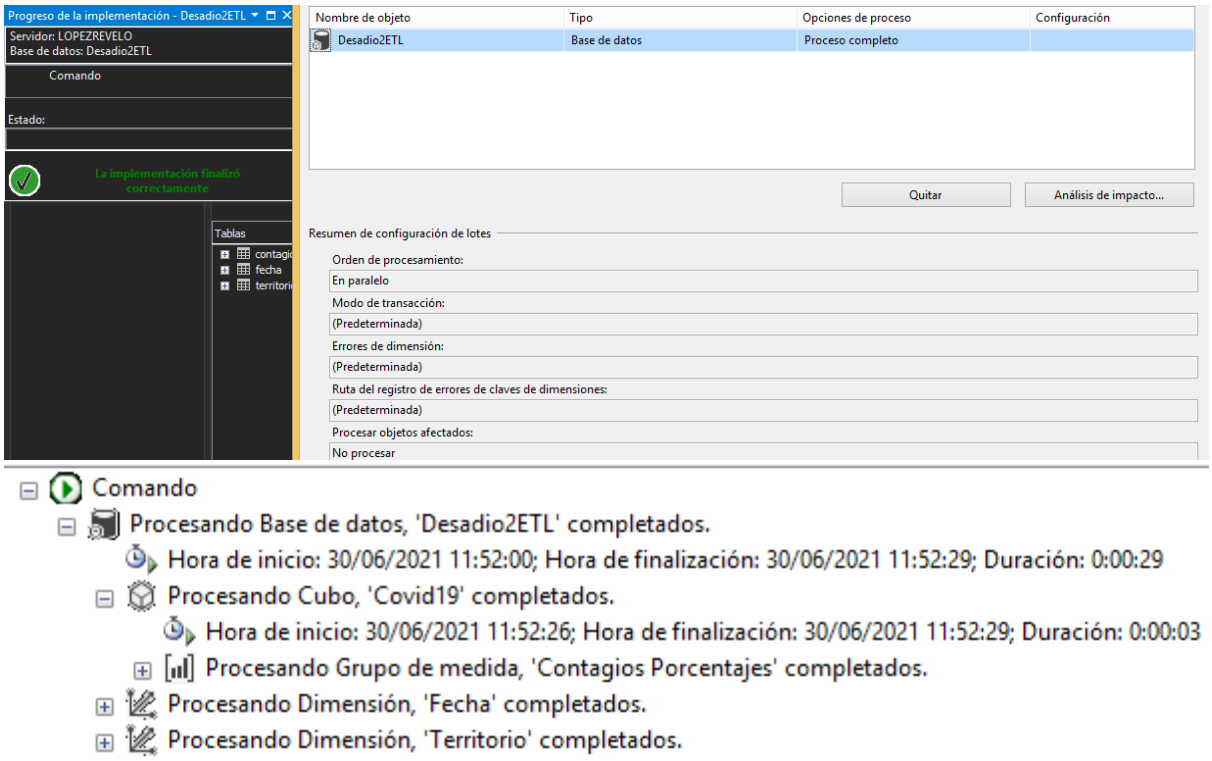




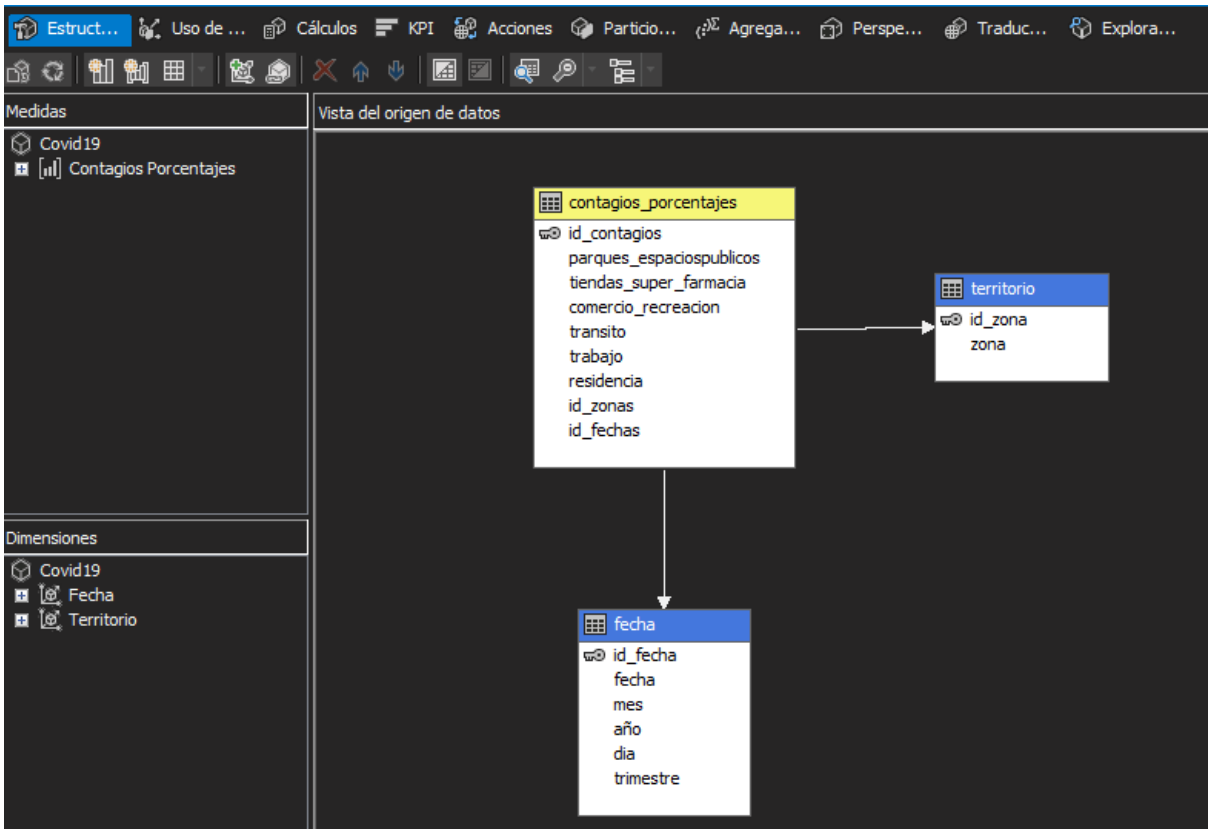
4. Una vez creado todo tendría que verse de la siguiente manera, en la columna Explorador de solución.



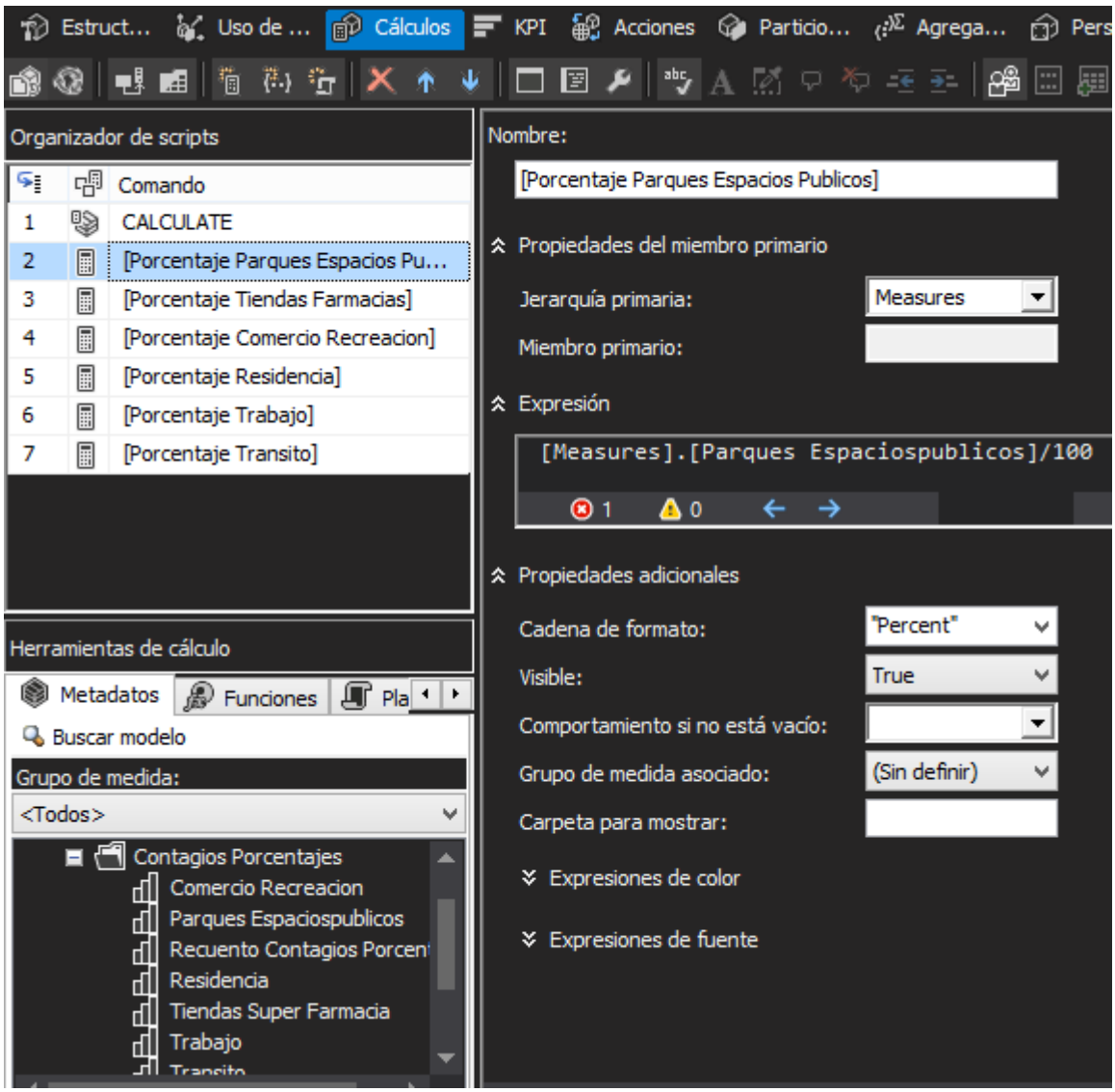
5. Le damos procesar al proyecto y luego ejecutamos.



6. Ahora damos click derecho sobre el cubo y seleccionamos examinar nos mostrara la siguiente área de trabajo



7. Nos dirigimos a Cálculos donde vamos a convertir los datos de las medidas en valores de porcentajes

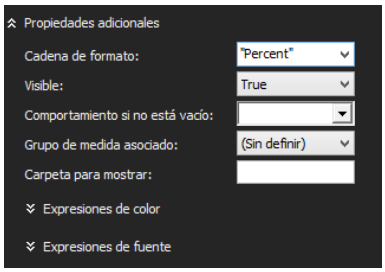


Agregamos las siguientes expresiones según corresponda la medida

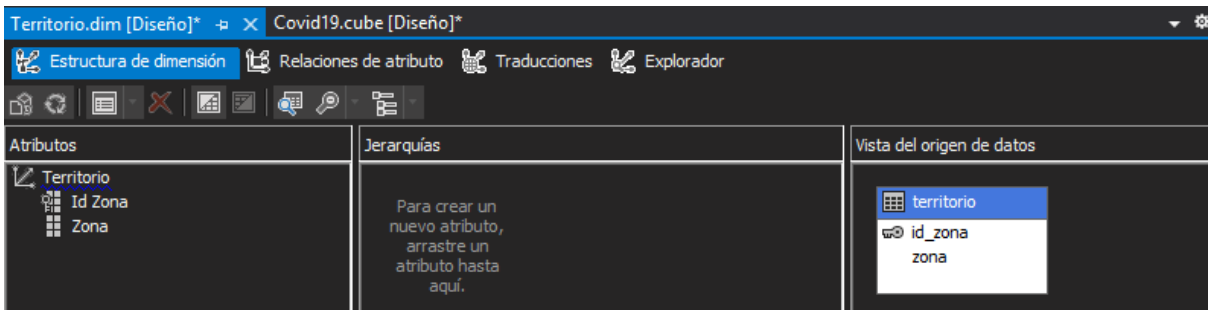
```
[Measures].[Tiendas Super Farmacia]/100
```

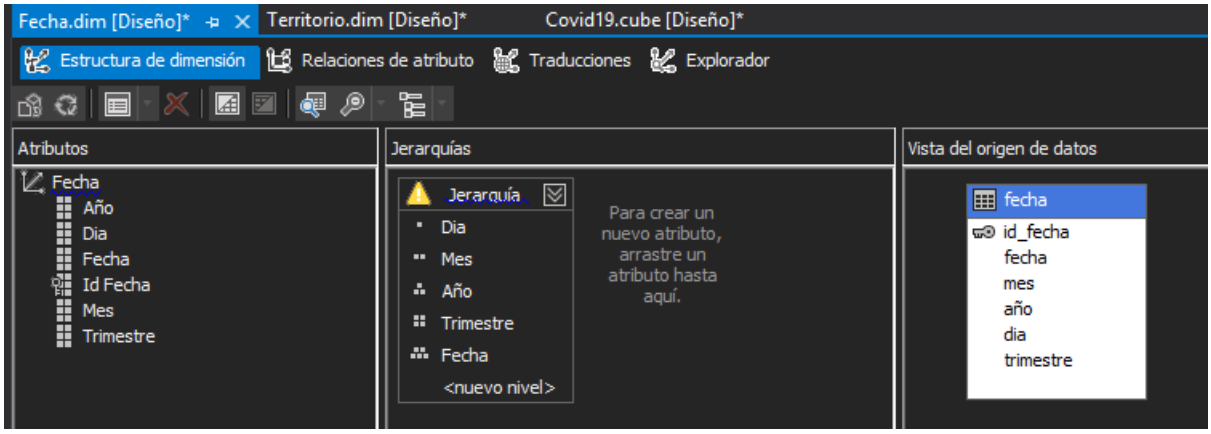
```
[Measures].[Comercio Recreacion]/100
```

```
[Measures].[Trabajo]/100
```

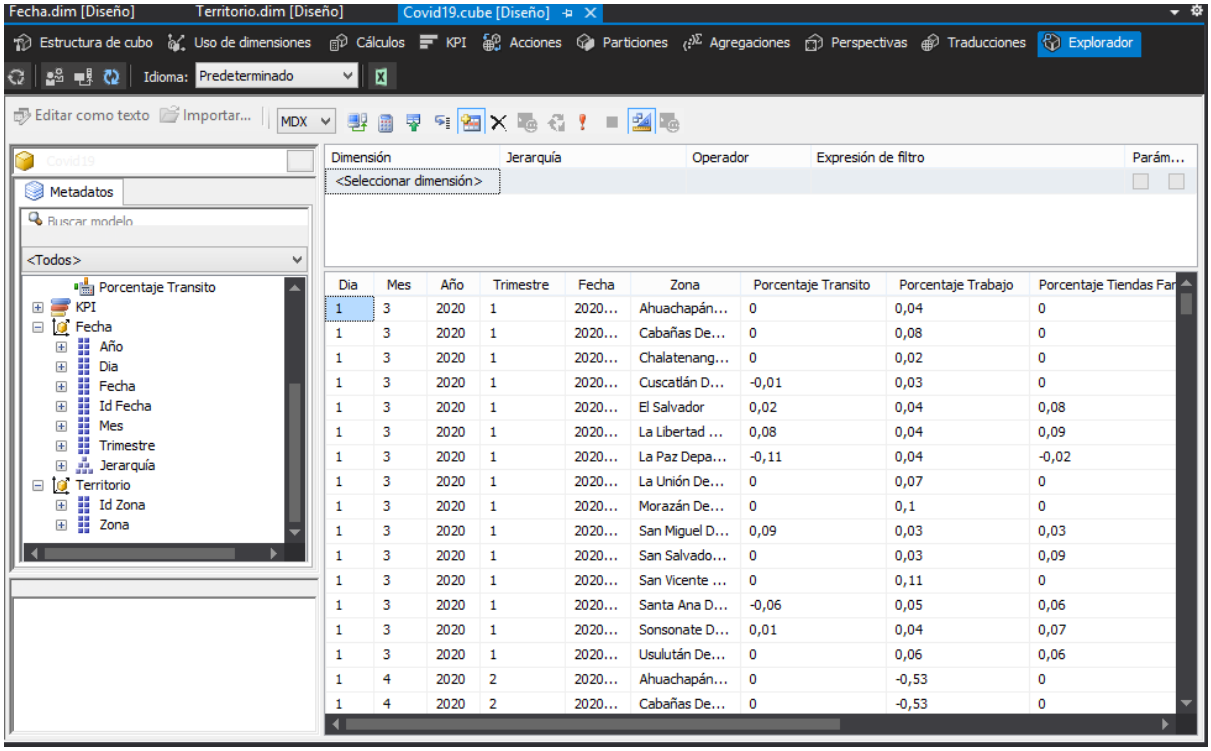


10. Abrimos la carpeta de dimensiones, editaremos y modificaremos las dos y de la siguiente manera



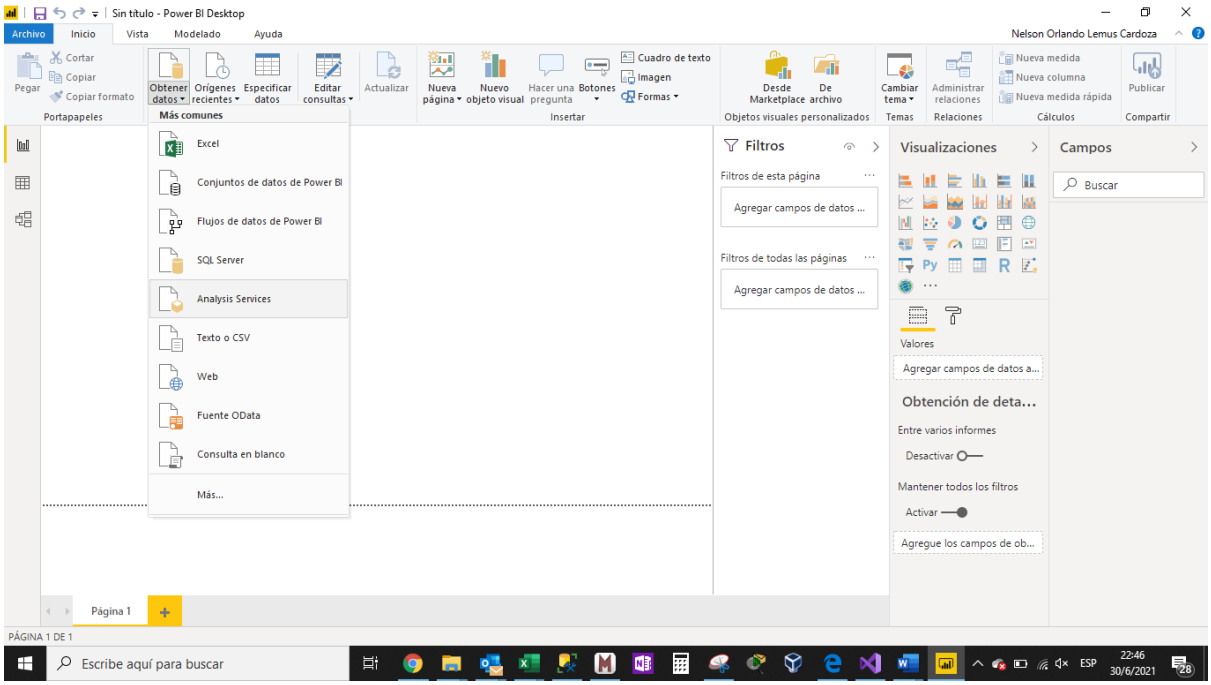


11. Ahora verificaremos la funcionalidad de nuestro cubo

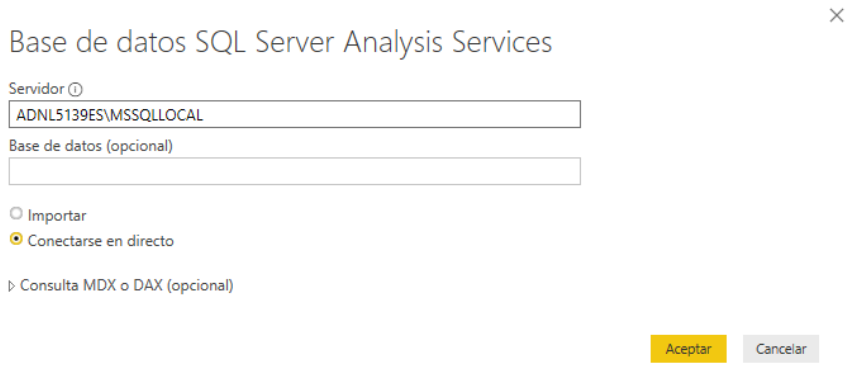


Dia	Mes	Año	Trimestre	Fecha	Zona	Porcentaje Transito	Porcentaje Trabajo	Porcentaje Tiendas Far
1	3	2020	1	2020...	Ahuachapán...	0	0,04	0
1	3	2020	1	2020...	Cabañas De...	0	0,08	0
1	3	2020	1	2020...	Chalatenang...	0	0,02	0
1	3	2020	1	2020...	Cuscatlán D...	-0,01	0,03	0
1	3	2020	1	2020...	El Salvador	0,02	0,04	0,08
1	3	2020	1	2020...	La Libertad ...	0,08	0,04	0,09
1	3	2020	1	2020...	La Paz Depa...	-0,11	0,04	-0,02
1	3	2020	1	2020...	La Unión De...	0	0,07	0
1	3	2020	1	2020...	Morazán De...	0	0,1	0
1	3	2020	1	2020...	San Miguel D...	0,09	0,03	0,03
1	3	2020	1	2020...	San Salvado...	0	0,03	0,09
1	3	2020	1	2020...	San Vicente ...	0	0,11	0
1	3	2020	1	2020...	Santa Ana D...	-0,06	0,05	0,06
1	3	2020	1	2020...	Sonsonate D...	0,01	0,04	0,07
1	3	2020	1	2020...	Usulután De...	0	0,06	0,06
1	4	2020	2	2020...	Ahuachapán...	0	-0,53	0
1	4	2020	2	2020...	Cabañas De...	0	-0,53	0

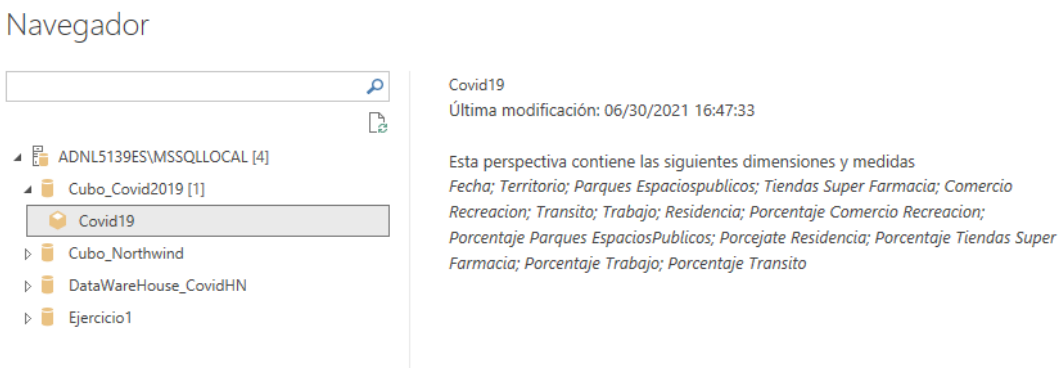
12. Analizando la data en PowerBI, damos click en la opción obtener datos y luego en analysis service para conectar a nuestro cubo desde SQL.



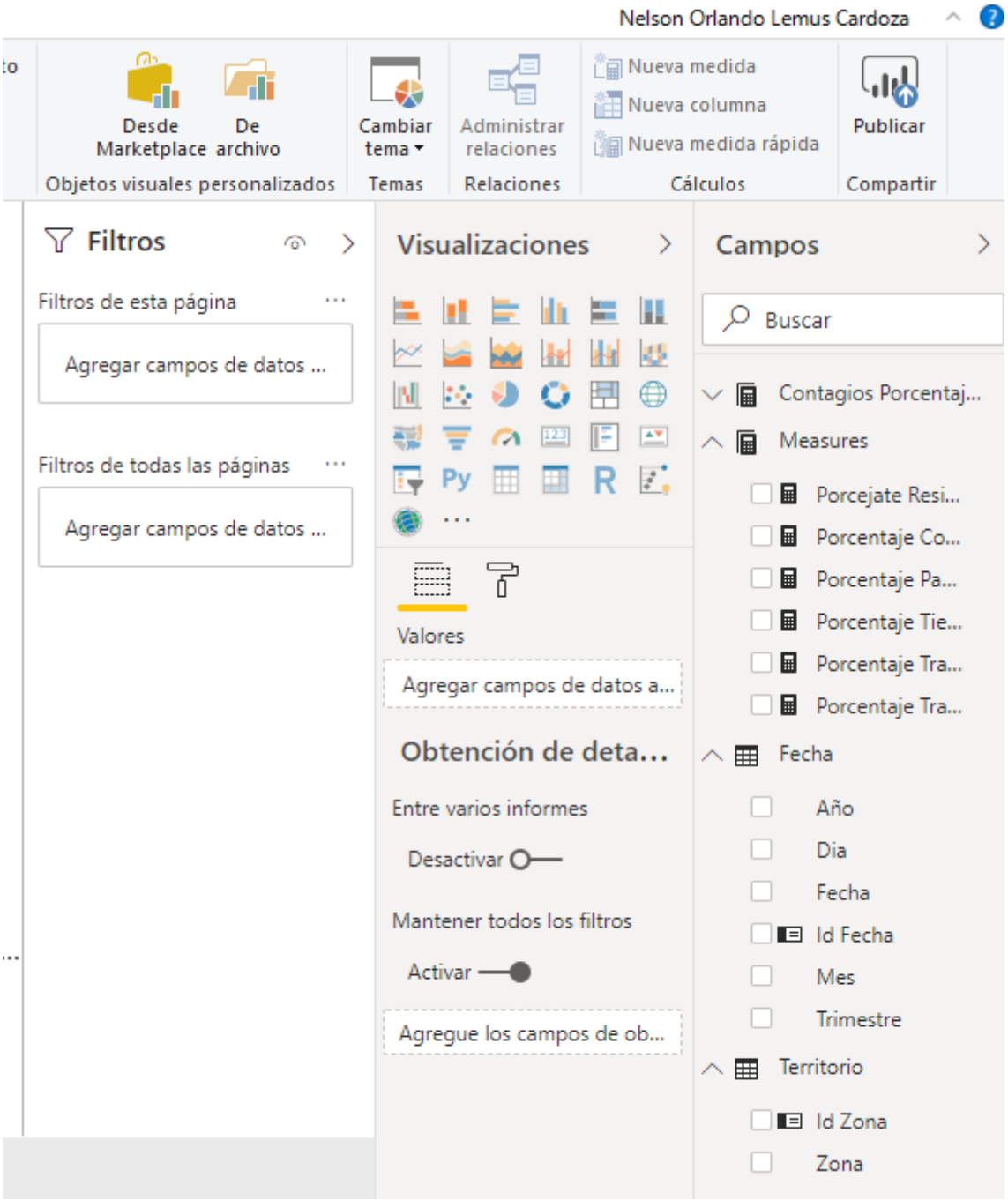
Colocamos el nombre de nuestro servidor



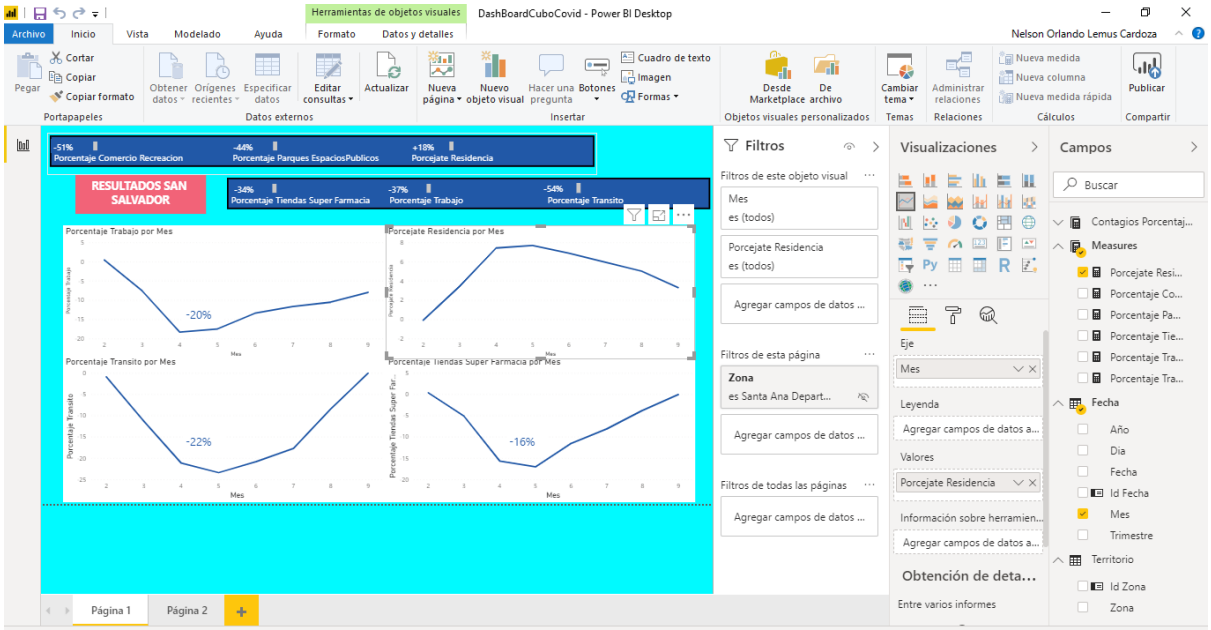
Luego seleccionamos el cubo Covid19 para establecer la conexión.



Al realizar la conexión se muestran las medidas y las dimesiones del cubo como objetos seleccionables para analizar la data, también las visualizaciones como tablas, graficas de pastel, barra o lineales.

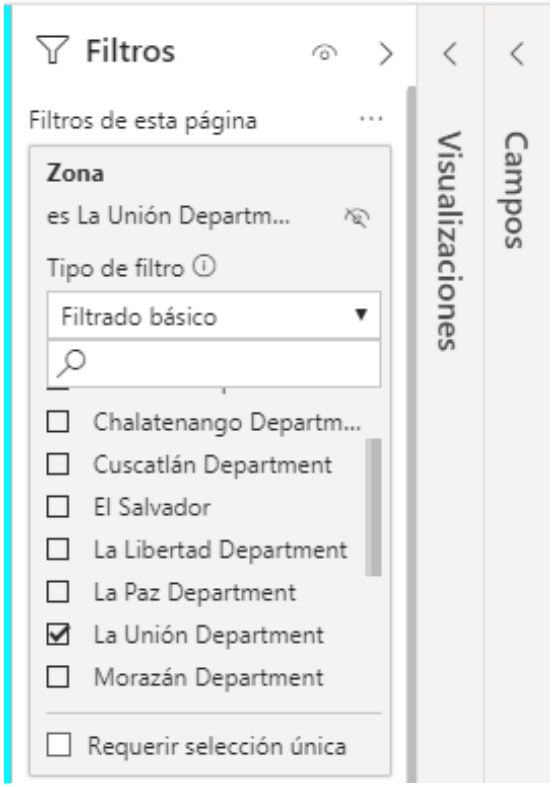


13. Agregamos 2 tarjetas de varias filas (Cuadros azules) para seleccionar 3 campos en cada tarjeta y visualizar los datos de promedios de cada una de las movilidades, así vemos como fue afectado por departamento las restricciones del Covid19. Los valores negativos nos dicen que la movilidad bajo y los valores positivos nos dicen que la movilidad aumento.

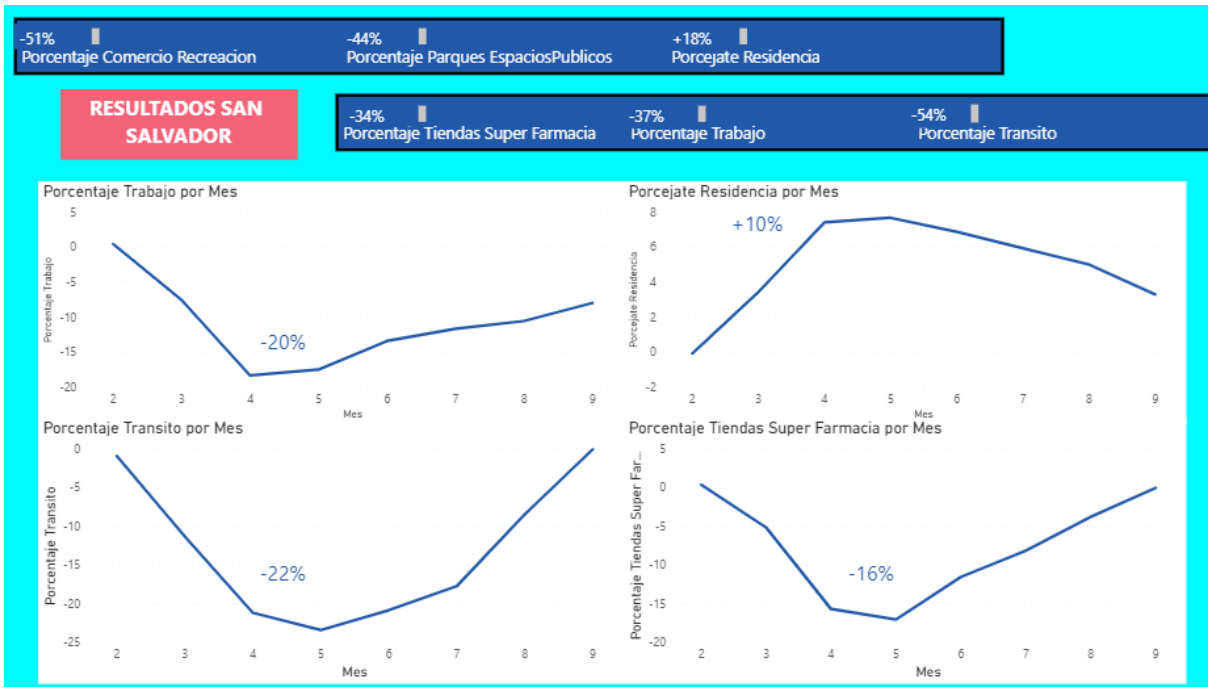


13. Luego agregamos cuatro graficas de línea para comparar mes vs Trabajo, Residencia, Transito y Farmacias. Y ocupamos cuatro tarjetas simples para mostrar el % donde la grafico hizo pico.

14. Finalmente se agrego el filtro de zona para poder seleccionar el departamento que queramos ver (cuadro de color rojo) al darle click podemos seleccionar un departamento diferente y así visualizar la data para cada uno y estudiar los impactos donde Covid19 tuvo su mayor apogeo y afecto la movilidad en El Salvador.

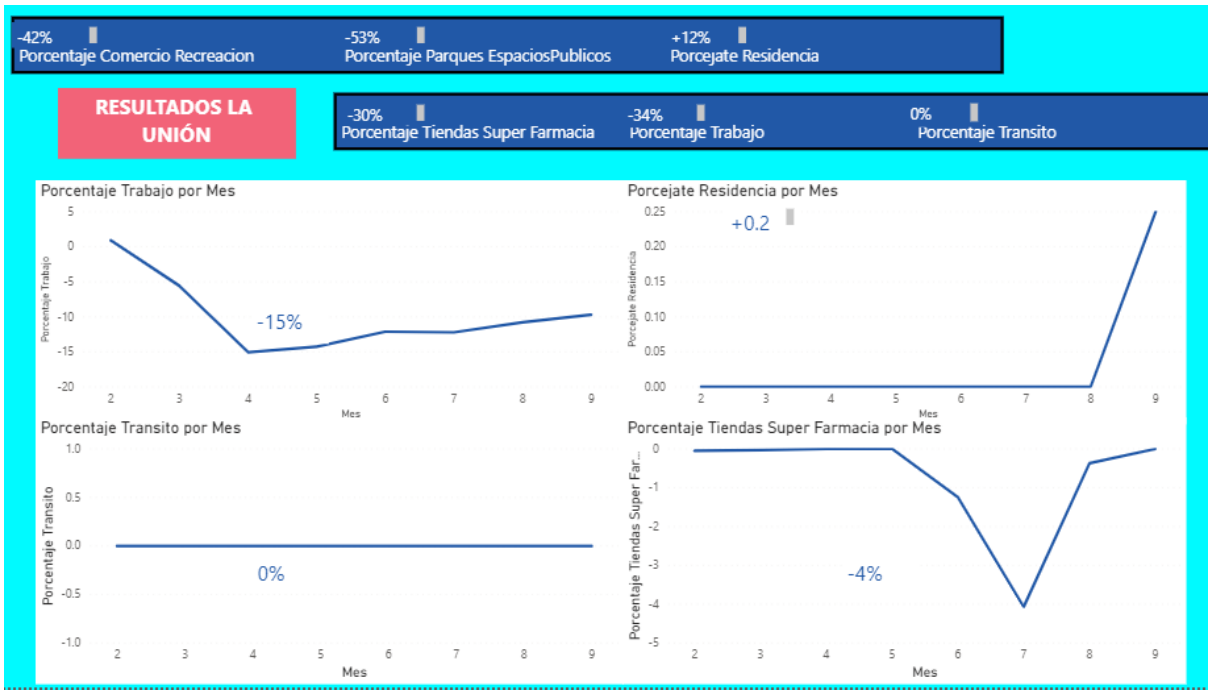


15. Resultados San Salvador.



16. Resultados Santa Ana.

17. Resultados La Unión.



17. Resultados Sonsonate.