



UNIVERSIDAD DON BOSCO

Asignatura

Datawarehouse y Minería de Datos G01T

Catedrático

Ing. Herson Miguel Serrano Chacón

Actividad

Desafío 2

Fecha de entrega

01/06/2023

Estudiante

Cruz Mejía, Josué Esaú

CM221973

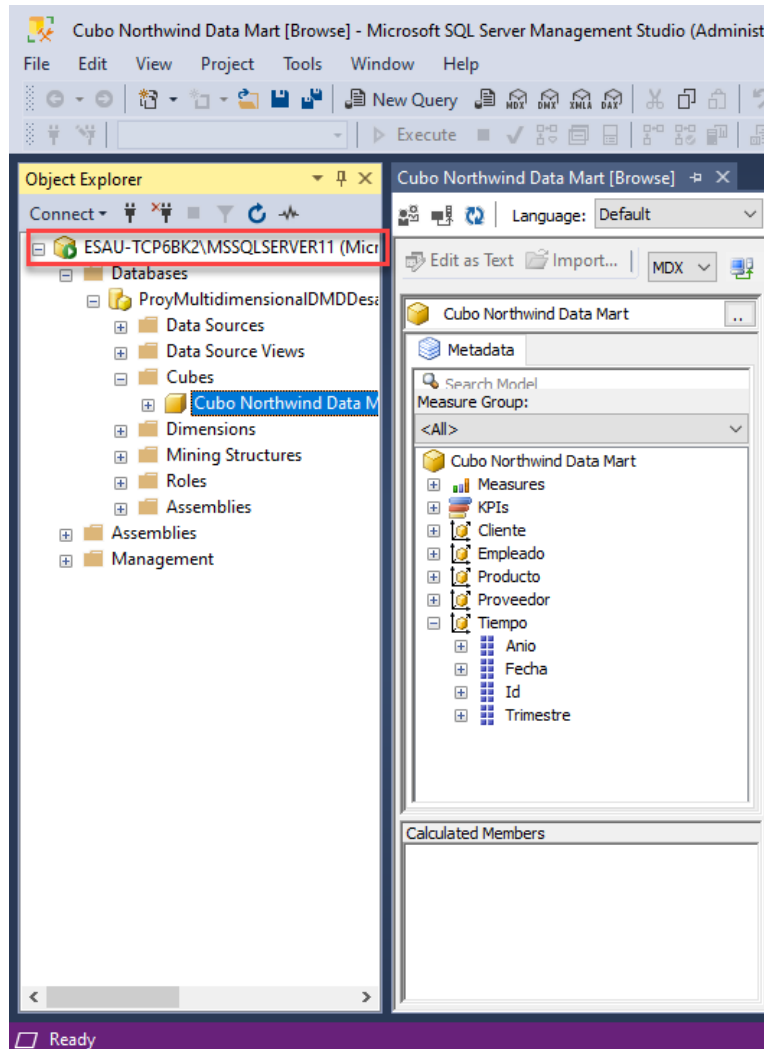


LINK REPOSITORIO: <https://github.com/EsauCruz/DMDDesafio2.git>

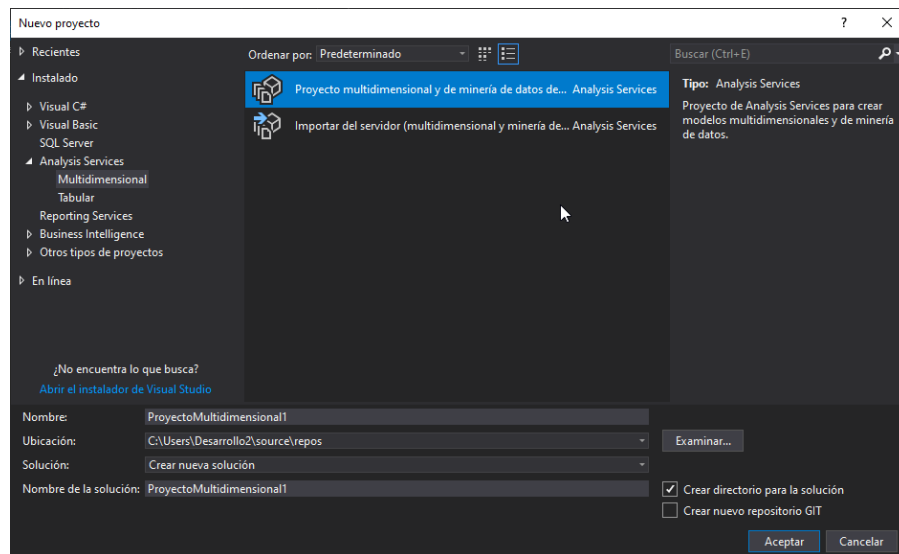
Parte I: Utilizando el Datamart de Northwind.

1) Indicación: Elabore un cubo en donde pueda visualizar nombre y país del cliente, apellido del empleado, país del proveedor, y nombre y categoría del producto.

Para realizar un cubo, se debe verificar que se cuente con el servidor instalado y ejecutándose, en este caso el SQL Server Analysis Services.



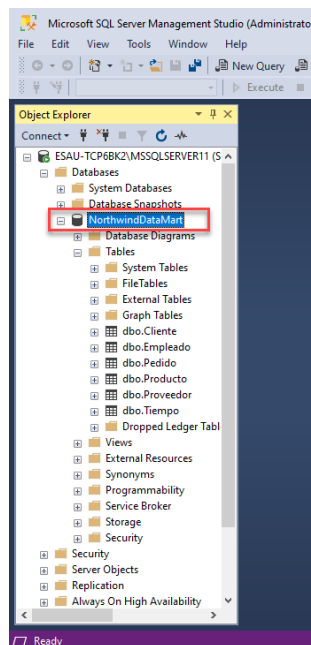
Luego, se debe verificar que en la instalación de Visual Studio que se está utilizando, esté instalado el complemento a utilizar para desarrollar en dicha plataforma el cubo deseado, ya que, en este caso, el Management Studio solamente se utilizará para crear la base de datos, desde el asistente de recuperación de base de datos, con el archivo .bak proporcionado para este desafío. Además, también se podrá visualizar el cubo desde esa interfaz, pero el desarrollo y creación de este, será trabajado únicamente desde Visual Studio, por su mayor alcance en reconocimiento de errores de ejecución.



Una vez se tienen listas las herramientas a utilizar, ya se empiezan a realizar los pasos necesarios para crear el cubo especificado

Restaurando el archivo Datamart de Northwind en el Management Studio

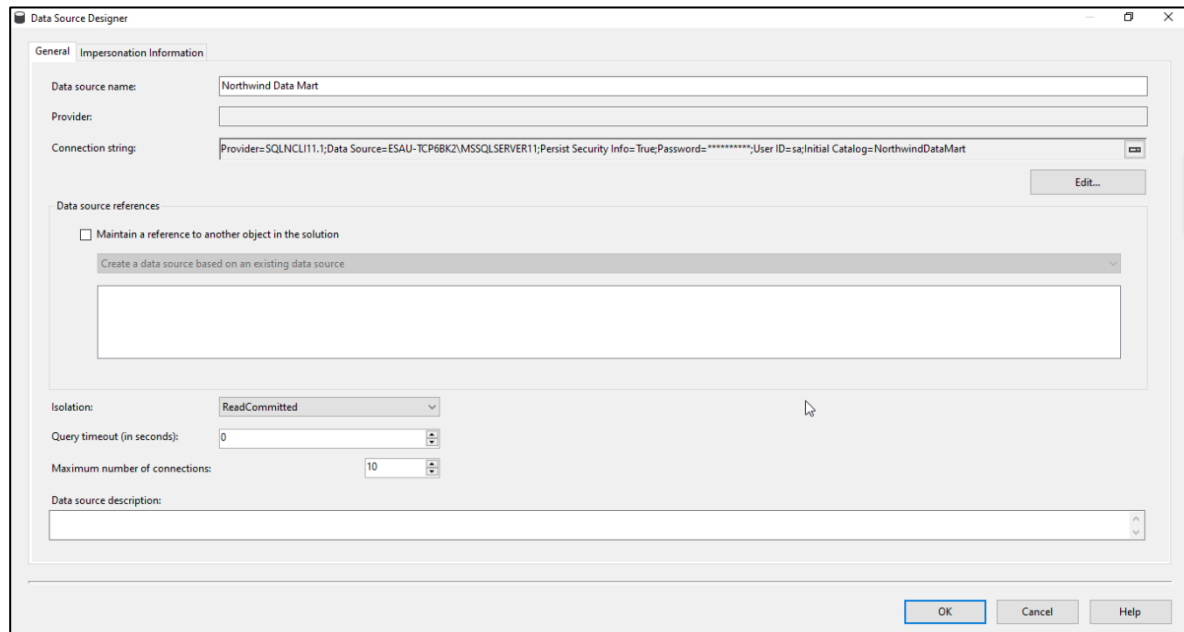
- Clic derecho en carpeta de bases de datos > Restaurar base
- Clic en “dispositivo” > clic en los 3 puntos
- Buscar el archivo .bak que previamente se debió colocar manualmente en la carpeta Backup, de la instalación específica de SQL Server a trabajar
- Clic en “verificar media”, si todo está bien clic en “finalizar”



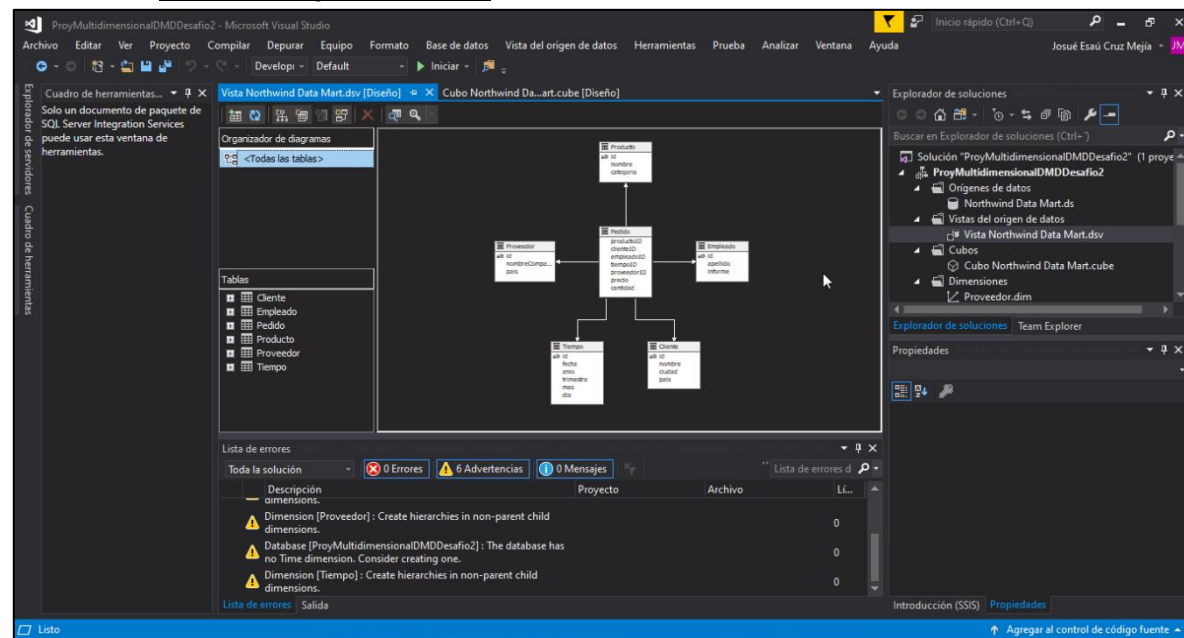


Una vez ya se tiene la base de datos a utilizar, que en este caso es una parte de la base de datos Northwind; ya se puede empezar el proyecto en Visual Studio.

Definiendo “origen de datos”

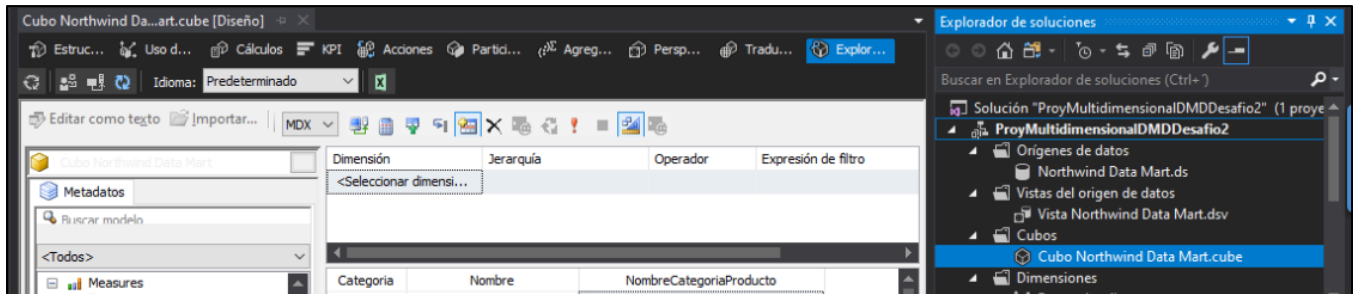


Definiendo “vista de origen de datos”



Definiendo “cubo”

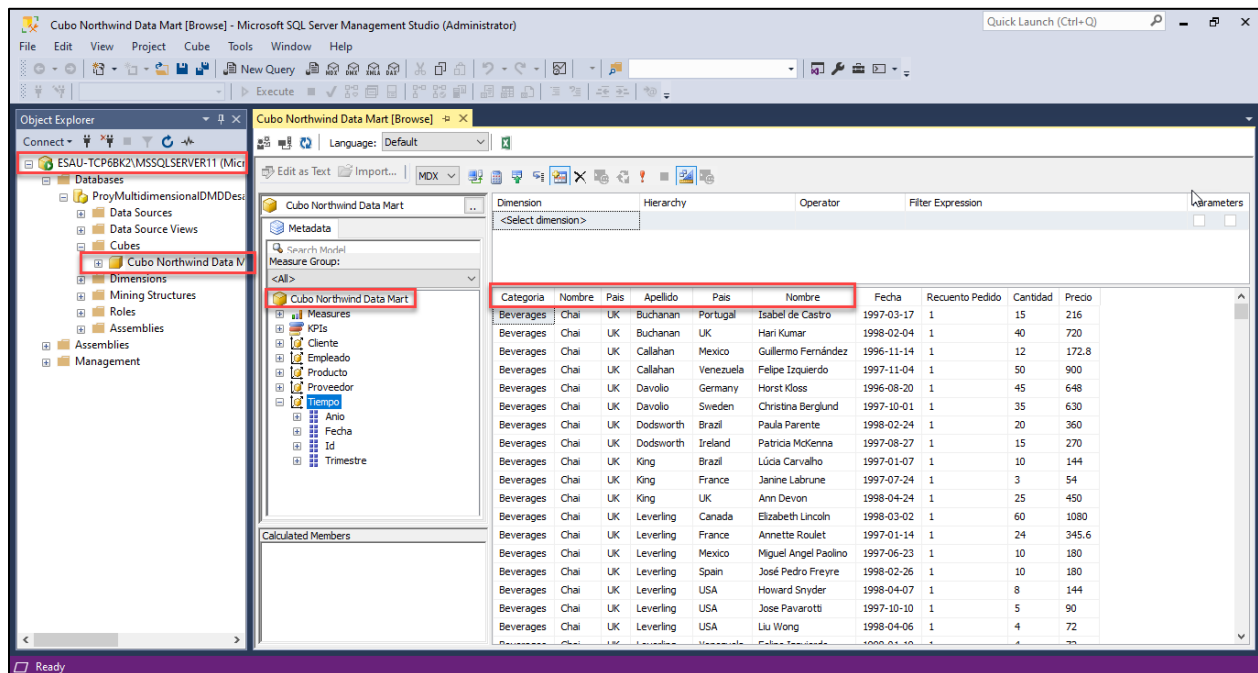
- Clic derecho en carpeta “cubos” > Nuevo cubo > Clic en “tablas existentes”
- En vista del origen de datos seleccionar la que se hizo anteriormente
- En tablas de grupo de medida seleccionar “Pedido” y “Tiempo”
- En “medida” verificar que todo esté seleccionado
- En “dimensión” verificar que todo esté seleccionado
- Colocar nombre al cubo y “finalizar”



Una vez ya se tiene creado el cubo, se debe dar clic derecho a este, y seleccionar “Procesar” para que este sea ejecutado y publicado en el servidor de Analysis Services.

Nota: verificar en las propiedades del proyecto que en la “implementación” esté el nombre del servidor de SQL Server Analysis Services, porque por defecto aparece “localhost”

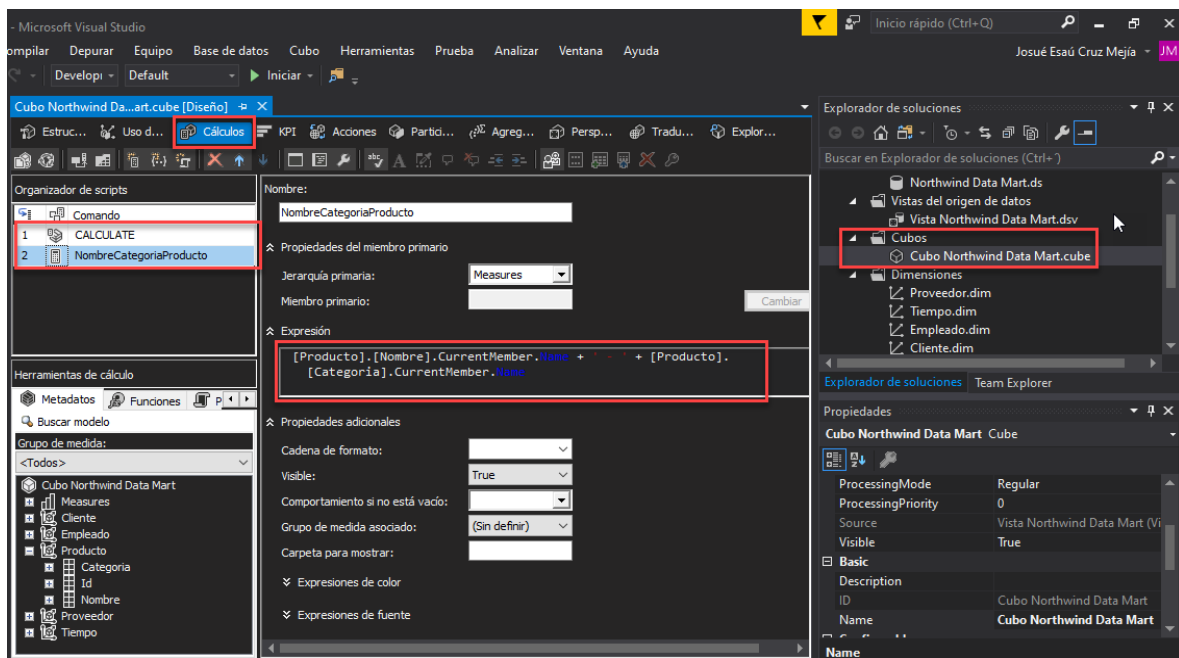
Cubo que muestra los datos solicitados en el numeral 1)



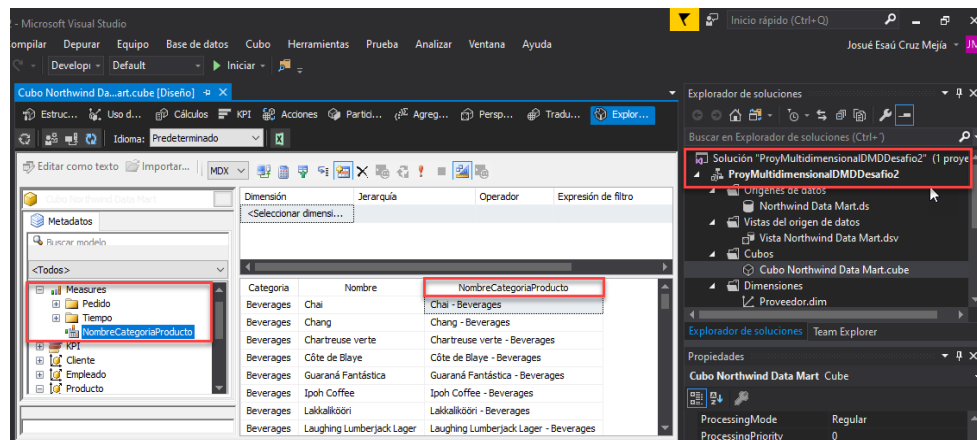
2) Indicación: Crear un nuevo campo concatenando la categoría del producto con el nombre del producto.

Para crear un capo nuevo y que sea calculado, se siguen los pasos:

- Abrir el cubo que se realizó anteriormente
- Clic en la pestaña “Cálculos”
- Clic derecho en la opción “CALCULATE”
- Colocar la expresión que consiga lo que se pide
- Guardar los cambios y procesar de nuevo el cubo para que se tenga el nuevo campo a disposición al hacer las consultas en el cubo



En el cubo, en la pestaña “Explorador” ya se tiene el campo calculado para utilizarlo



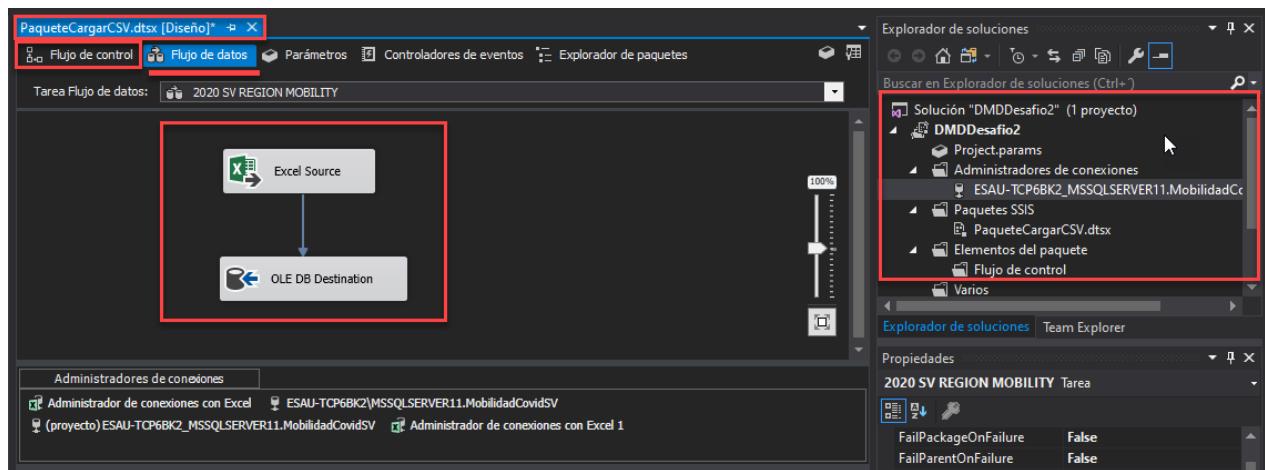


Parte I: Descubre cómo han cambiado los desplazamientos de tu comunidad debido al COVID-19.

1) Indicación: Se les proporciona la información en archivos csv, un archivo por cada país, se debe elegir 1 país para cargar la información a una base de datos y realizar el análisis

Se selecciona el archivo “2020_SV_Region_Mobility_Report.csv” para que sean enviados a SQL Server Managment Studio, a través de un ETL creado en Visual Studio.

Se crea un flujo de control llamado “PaqueteCargarCSV” que tiene dentro (en flujo de datos) 2 objetos. Asistente de orígenes: para extraer los datos del archivo CSV y en Asistente de destinos: para depositar los datos en la base SQL Server



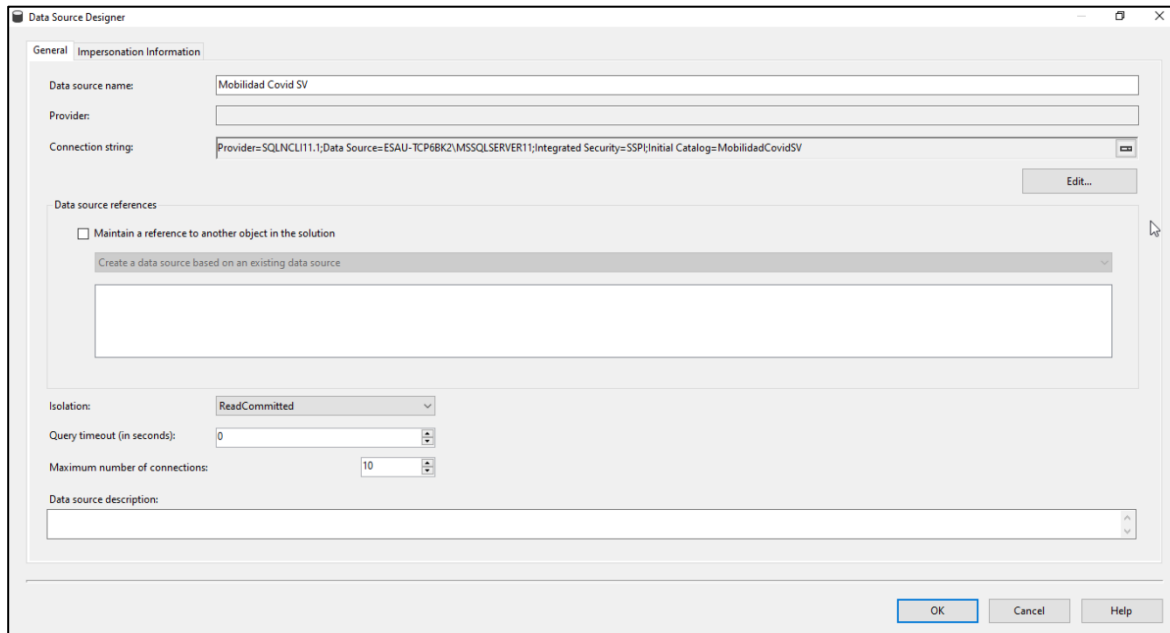
Una vez ejecutado el ETL, ya se cuenta con los datos en una base de datos SQL Server

	country_region_code	country_region	sub_region_1	sub_region_2	metro_area	iso_3166_2_code	census_fips_code	date	retail_and_recreation_percent_change_from_baseline
1	SV	El Salvador	NULL	NULL	NULL	NULL	NULL	2020-02-15 00:00:00.000	4
2	SV	El Salvador	NULL	NULL	NULL	NULL	NULL	2020-02-16 00:00:00.000	4
3	SV	El Salvador	NULL	NULL	NULL	NULL	NULL	2020-02-17 00:00:00.000	0
4	SV	El Salvador	NULL	NULL	NULL	NULL	NULL	2020-02-18 00:00:00.000	0
5	SV	El Salvador	NULL	NULL	NULL	NULL	NULL	2020-02-19 00:00:00.000	0
6	SV	El Salvador	NULL	NULL	NULL	NULL	NULL	2020-02-20 00:00:00.000	-1
7	SV	El Salvador	NULL	NULL	NULL	NULL	NULL	2020-02-21 00:00:00.000	0
8	SV	El Salvador	NULL	NULL	NULL	NULL	NULL	2020-02-22 00:00:00.000	0
9	SV	El Salvador	NULL	NULL	NULL	NULL	NULL	2020-02-23 00:00:00.000	0
10	SV	El Salvador	NULL	NULL	NULL	NULL	NULL	2020-02-24 00:00:00.000	-2

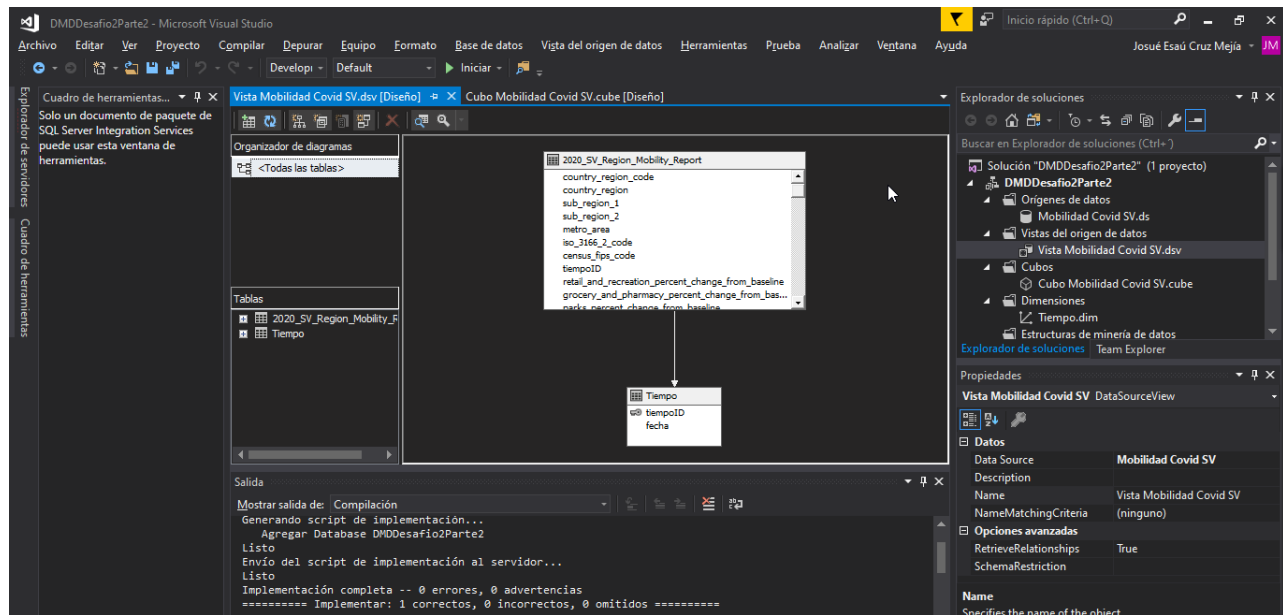


2) Indicación: Elaborar un cubo en donde pueda visualizar el análisis que han realizado del archivo seleccionado. Que criterios tomar para el análisis, será responsabilidad de cada uno ver qué información es relevante para el análisis. En otras palabras no es el docente quién está pidiendo un análisis concreto, son ustedes lo que van a determinar que análisis realizar. Recuerden todos los países tienen departamentos, fechas, lugares, etc.

Definiendo “origen de datos”



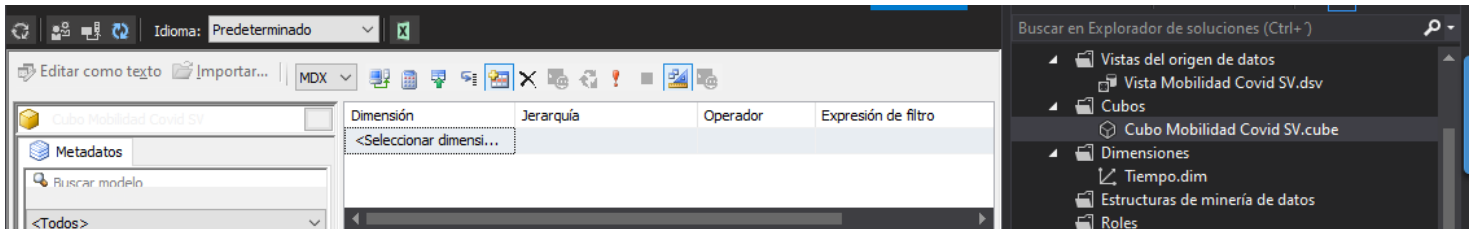
Definiendo “vista de origen de datos”





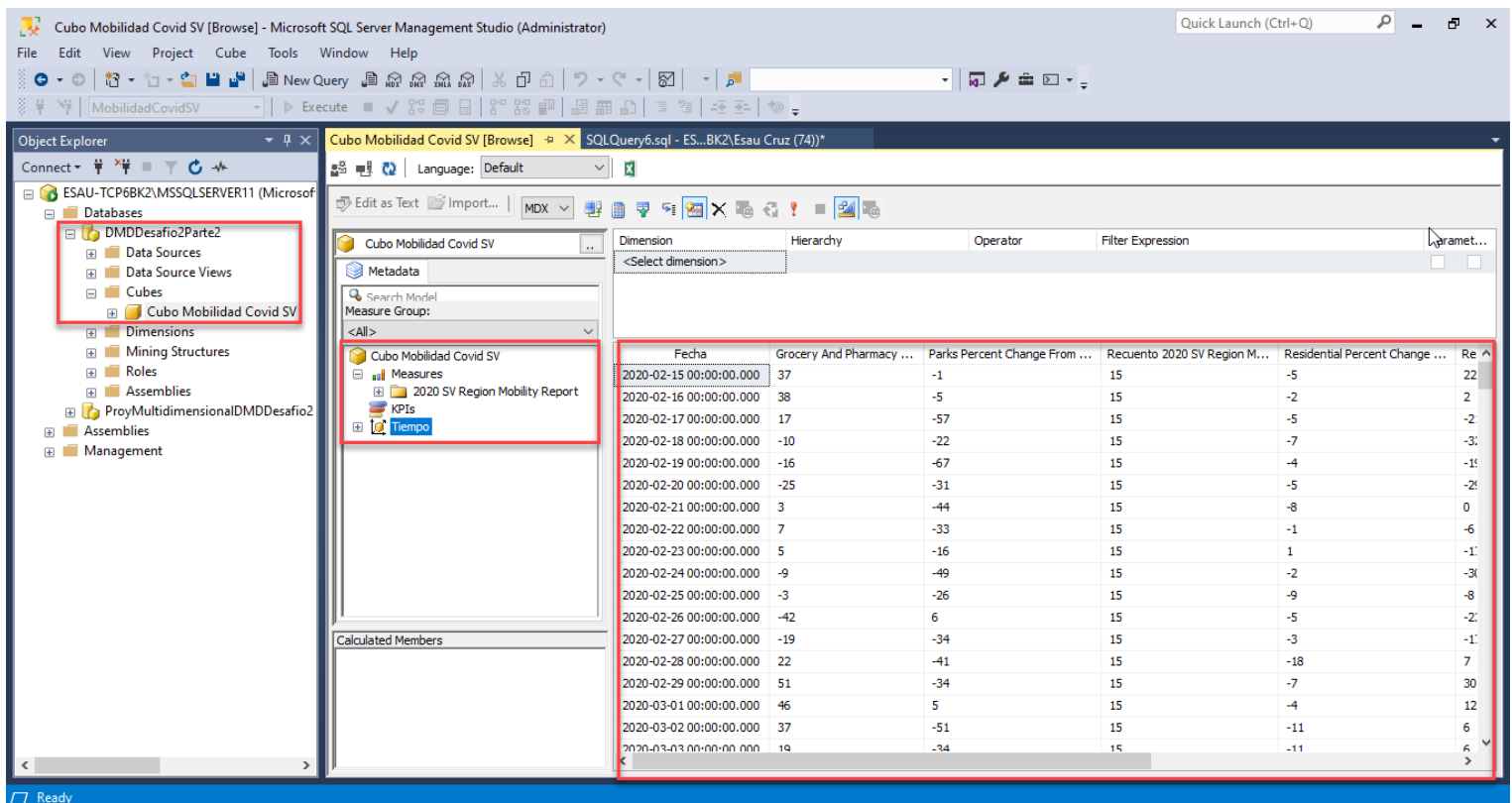
Definiendo “cubo”

- Clic derecho en carpeta “cubos” > Nuevo cubo > Clic en “tablas existentes”
- En vista del origen de datos seleccionar la que se hizo anteriormente
- En tablas de grupo de medida seleccionar “2020_SV_Region_Mobility_Report” y “Tiempo”
- En “medida” verificar que todo esté seleccionado
- En “dimensión” verificar que todo esté seleccionado
- Colocar nombre al cubo y “finalizar”



Una vez ya se tiene creado el cubo, se debe dar clic derecho a este, y seleccionar “Procesar” para que este sea ejecutado y publicado en el servidor de Analysis Services.

Cubo que muestra los datos solicitados en el numeral 2)





Análisis de resultados

3) Indicación: ¿Qué indicadores nos dicen los valores negativos?

Estos valores negativos indican un cambio negativo en la movilidad de las personas en los lugares de trabajo en comparación con el período de referencia establecido por Google.

Cuando se dice que hay un cambio negativo en la movilidad de los lugares de trabajo, significa que ha habido una disminución en la cantidad de personas que se desplazan hacia los lugares de trabajo en comparación con el período de referencia. Esto, puede ser el resultado de factores, como medidas de confinamiento, restricciones de viaje, adopción del trabajo remoto u otras circunstancias que limiten la actividad laboral presencial.

El valor negativo en esta columna indica un cambio en relación con un punto de referencia específico y no proporciona información absoluta sobre la cantidad de personas que están trabajando en un lugar determinado. Además, los datos se basan en la información agregada y anónima recopilada por Google, por lo que es necesario interpretarlos con precaución y tener en cuenta el contexto y las limitaciones de los datos.

Un análisis interesante, es lo que ocurrió en el mes de agosto (vacación), que todos los lugares externos a las viviendas tuvieron movilidad negativa, es decir, en promedio hubo muchas menos visitas que en el comparativo, ya que las personas se quedaron en casa

Fecha	Workplaces ...	Grocery ...	Parks ...	Residential ...	Retail And...	Transit Stations Pe...
2020-08-01 00:00:00.000	-414	-204	-625	97	-551	-437
2020-08-02 00:00:00.000	-408	-262	-724	55	-638	-473
2020-08-03 00:00:00.000	-684	-223	-558	131	-440	-434
2020-08-04 00:00:00.000	-701	-156	-535	125	-438	-410
2020-08-05 00:00:00.000	-776	-243	-563	144	-487	-440
2020-08-06 00:00:00.000	-914	-302	-622	166	-632	-496
2020-08-07 00:00:00.000	-620	-230	-590	134	-546	-439
2020-08-08 00:00:00.000	-443	-278	-651	104	-587	-487
2020-08-09 00:00:00.000	-405	-313	-749	59	-681	-518
2020-08-10 00:00:00.000	-611	-248	-593	117	-474	-450
2020-08-11 00:00:00.000	-616	-210	-543	114	-457	-406
2020-08-12 00:00:00.000	-603	-234	-540	116	-473	-403
2020-08-13 00:00:00.000	-619	-240	-555	121	-522	-420
2020-08-14 00:00:00.000	-534	-226	-601	119	-538	-429
2020-08-15 00:00:00.000	-374	-181	-582	88	-522	-439
2020-08-16 00:00:00.000	-369	-233	-685	52	-614	-423
2020-08-17 00:00:00.000	-596	0	0	111	0	0
2020-08-18 00:00:00.000	-585	0	0	110	0	0
2020-08-19 00:00:00.000	-603	0	0	115	0	0
2020-08-20 00:00:00.000	-609	0	0	116	0	0
2020-08-21 00:00:00.000	-528	0	0	117	0	0
2020-08-22 00:00:00.000	-404	0	0	95	0	0
2020-08-23 00:00:00.000	-380	0	0	56	0	0
2020-08-24 00:00:00.000	-558	0	0	104	0	0
2020-08-25 00:00:00.000	-567	0	0	113	0	0
2020-08-26 00:00:00.000	-560	0	0	114	0	0
2020-08-27 00:00:00.000	-581	0	0	105	0	0
2020-08-28 00:00:00.000	-501	0	0	100	0	0
2020-08-29 00:00:00.000	-351	0	0	72	0	0
2020-08-30 00:00:00.000	-297	0	0	40	0	0
2020-08-31 00:00:00.000	-534	0	0	111	0	0