

# Datawarehouse y Minería de Datos

**Docente**: Karens Medrano

Estudiante: Eduardo Ezequiel López Rivera

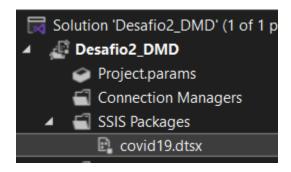
Carné: LR230061

Ciclo: 01-2025

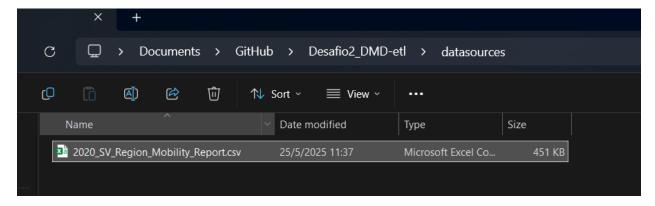
Desafío #2

Creación de proceso ETL (75%)

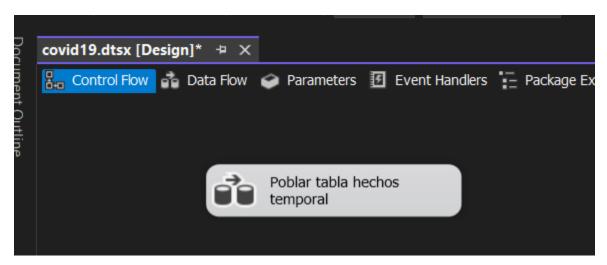
Creamos un proyecto en Visual Studio 2022 de tipo Integration Services Project, renombramos nuestro paquete SSIS a covid19.dtsx



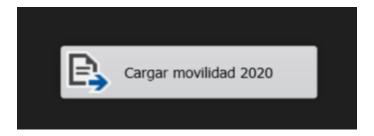
Descargamos el informe de movilidad de COVID-19 de El Salvador del año 2020, obtenemos el siguiente archivo en formato .CSV



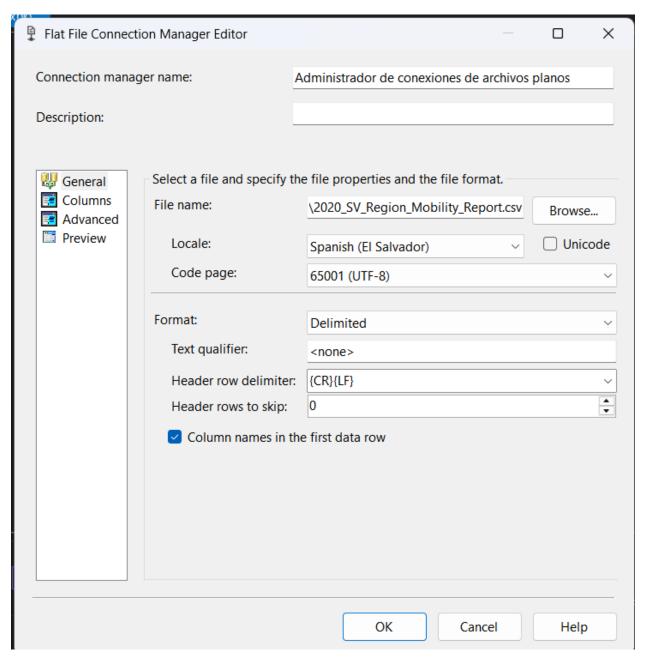
En nuestro Control Flow, creamos un Data Flow Task, el cual nos servirá para ordenar la data del CSV, la nombramos "Poblar tabla hechos temporal"



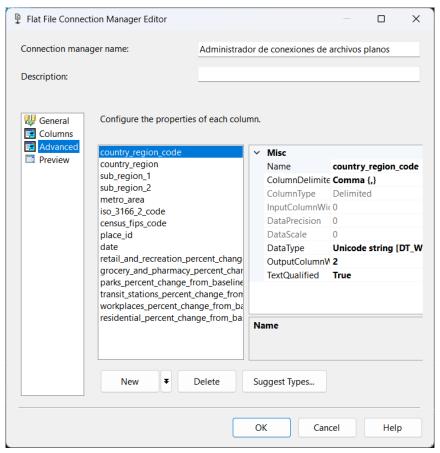
Agregamos un Flat File Source, el cual llamaremos "Cargar movilidad 2020".

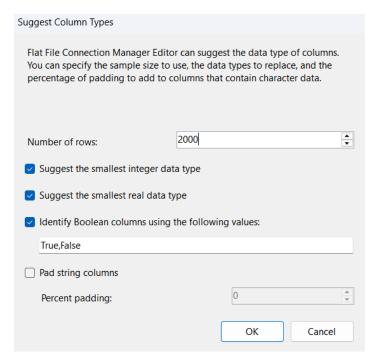


Configuramos el Flat File Source con los siguientes parámetros (Ruta del archivo .CSV, Code page en 65001 UTF-8 para que aparezcan las tildes de forma correcta)

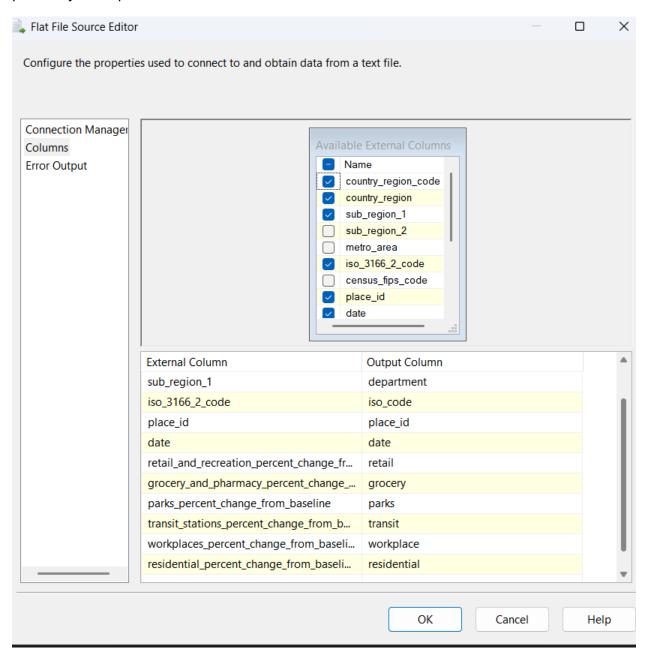


Configuramos los tipos de datos en la pestaña Columns, todos los strings los convertimos a Unicode string y utilizando la herramienta "Suggest Types" obtenemos un OutputColumnWidth para cada campo y un tipado de datos sugerido para los demás campos, calculado de los primeros 2000 registros del .CSV

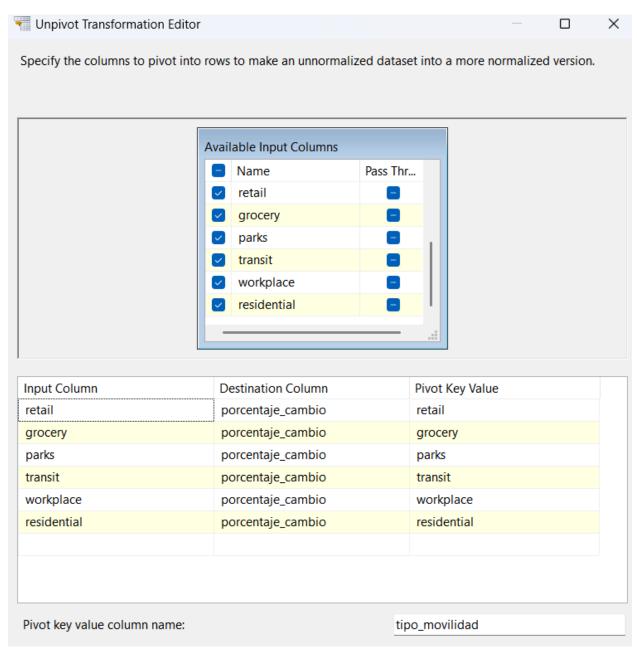




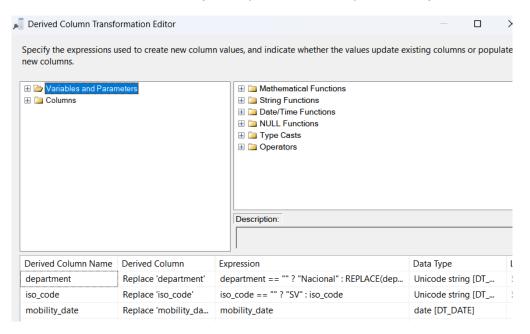
En el Flat File Source editor nos deshacemos de algunas columnas que vienen vacías porque no aplican para nuestro país, así mismo, cambiamos de nombre algunas columnas para mayor simplicidad.



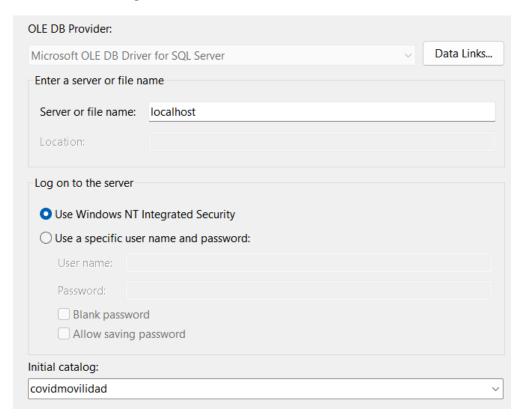
Conectamos un "Unpivot" para transformar las columnas que tienen datos de movilidad en una sola y sus respectivos valores en otra.



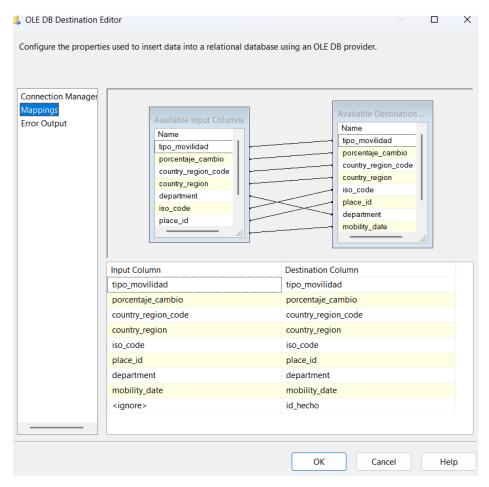
Conectamos un "Derived column" para limpiar los datos de la columna Department, poblar los datos vacíos en iso\_code, y reemplazamos date por mobility\_date.



Finalmente agregamos un OLE DB Destination para guardar la data en nuestra base de datos en SQL Server. Configuramos la conexión a la base de datos.

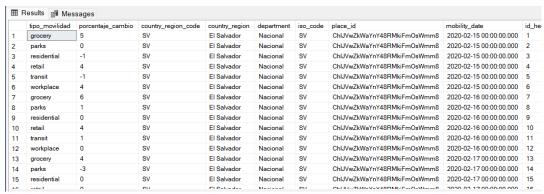


# Creamos la tabla "hechos" en nuestra base de datos y enlazamos las columnas correspondientes

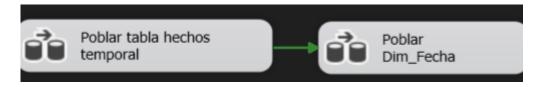


#### Ejecutamos con éxito el flujo. (15%)

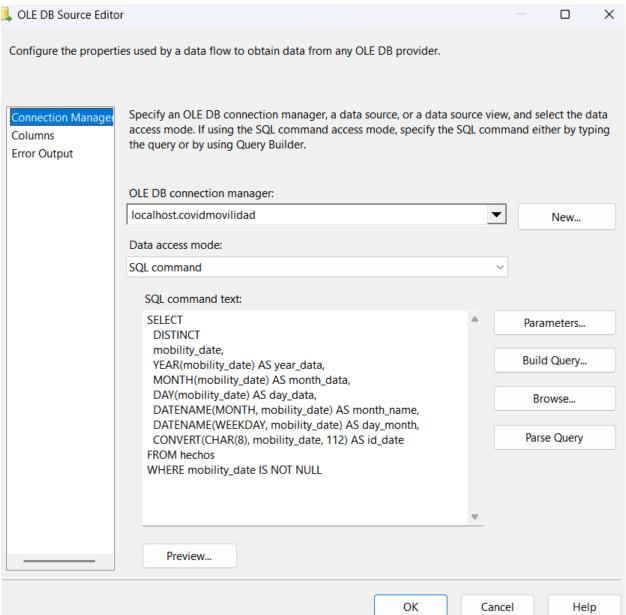




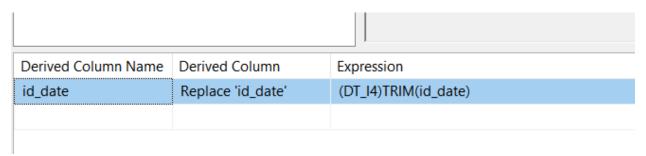
Procederemos a crear 3 dimensiones a partir de nuestra tabla de hechos temporal, comenzamos con Dim\_Fecha, creamos otro Data Flow Task, le llamaremos "Poblar Dim\_Fecha"



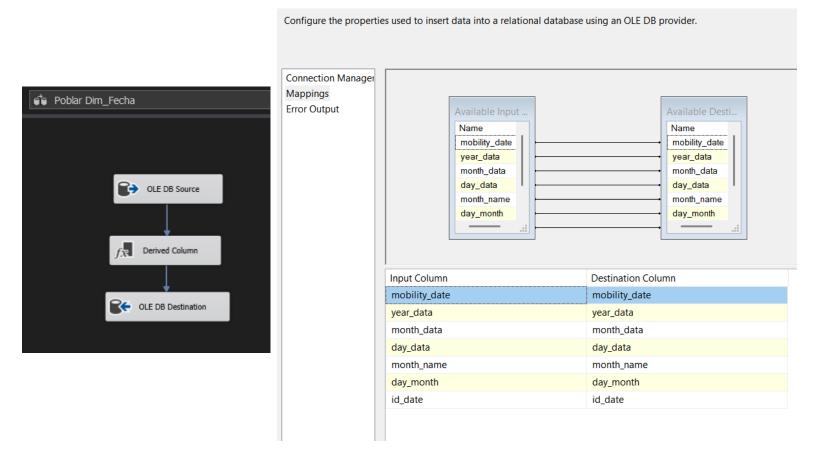
Colocamos un OLE DB Source dentro del flujo y obtenemos las fechas de los registros sin repetir con la siguiente consulta SQL y les damos un formato segmentado.



#### Convertimos la columna id\_date recién creada a integer



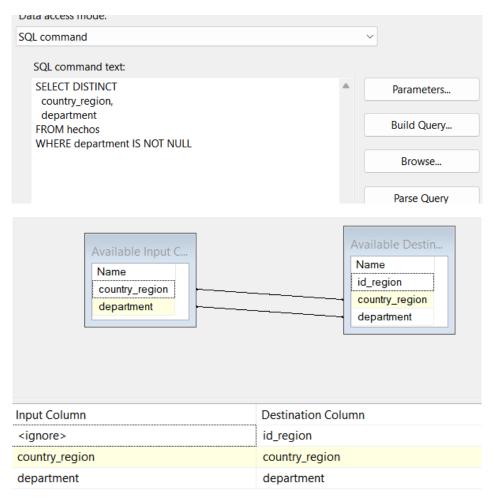
Conectamos un OLE DB Destination y creamos la tabla Dim\_Fecha, mapeamos las columnas. (25%)

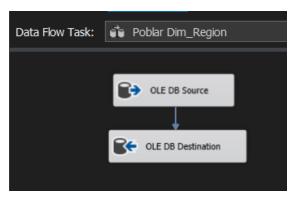


Creamos otro flujo para poblar Dim\_Region. (35%)



Conservamos una estructura muy simple, obtenemos datos mediante una consulta, en la misma consulta cambiamos un poco las columnas, y al crear la tabla Dim\_Region, creamos un campo id\_region primary key.

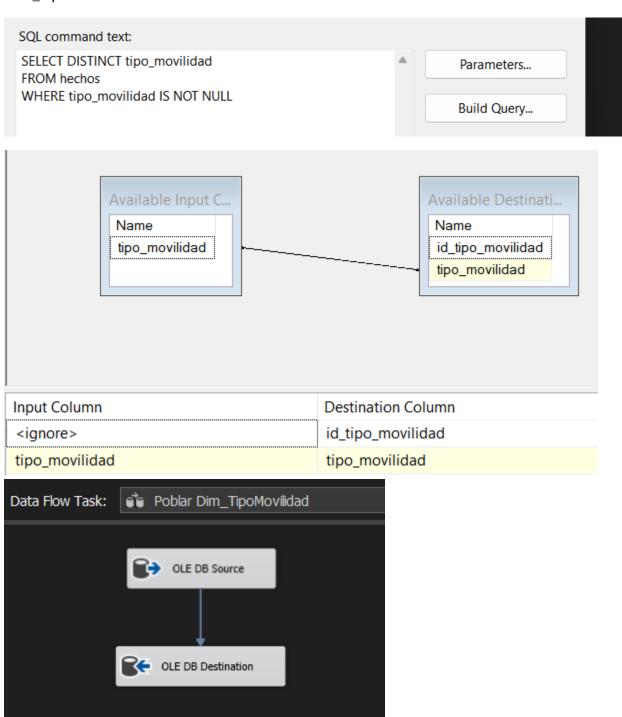




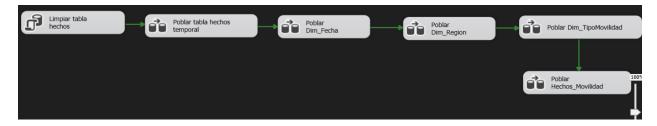
Ahora creamos otro flujo para poblar Dim\_TipoMovilidad. (45%)



Obtenemos la información con una consulta SQL, y esta la almacenamos en la tabla Dim\_TipoMovilidad.



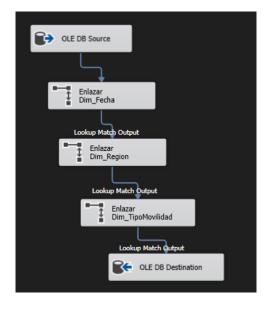
Finalmente, crearemos el flujo Hechos\_Movilidad, el cual creara una tabla que se conectara a las dimensiones previamente creadas. (Adicionalmente se agregar un Execute SQL task para limpiar los datos de las tablas cada vez que ejecutamos el proceso ETL) (50%)



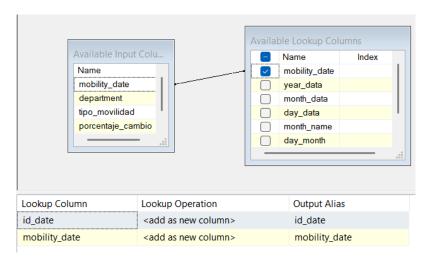
En nuestro nuevo flujo llamado Poblar Hechos\_Movilidad, creamos un OLE DB Source con la siguiente consulta



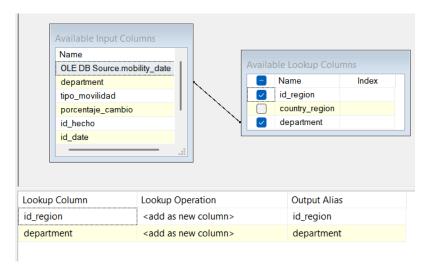
Conectamos 3 Lookups y un OLE DB Destination para guardar la tabla, cada uno es para conectar las dimensiones, iremos indagando paso a paso en cada una de las dimensiones.



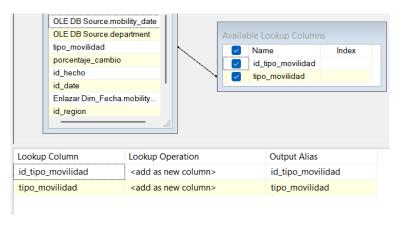
Conectamos mobility\_date de la tabla "hechos" con mobility\_date de Dim\_Fecha para enlazarlos. (55%)



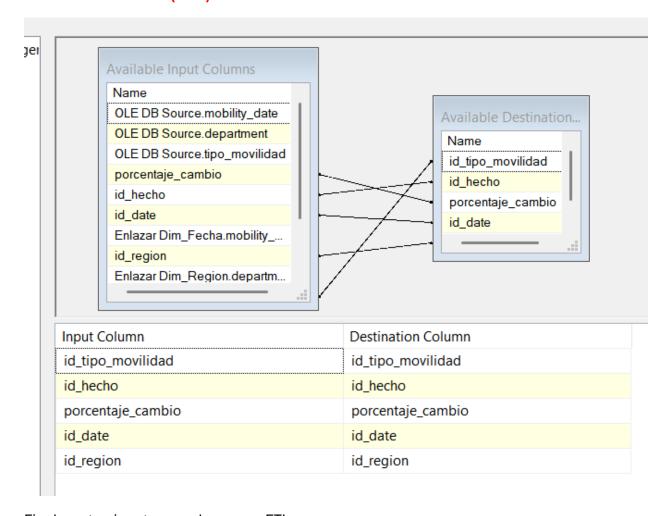
Conectamos department de la tabla "hechos" con department de Dim\_Region. (60%)



Conectamos tipo\_movilidad de la tabla "hechos" con tipo\_movilidad de Dim\_TipoMovilidad. (65%)



Creamos la tabla "Hechos\_Movilidad" y solo conservamos las columnas correspondientes a las llaves foráneas. (75%)



Finalmente ejecutamos el proceso ETL

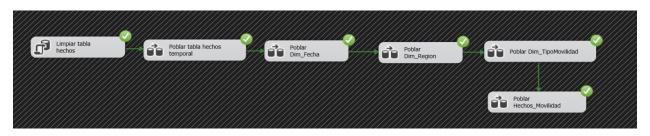
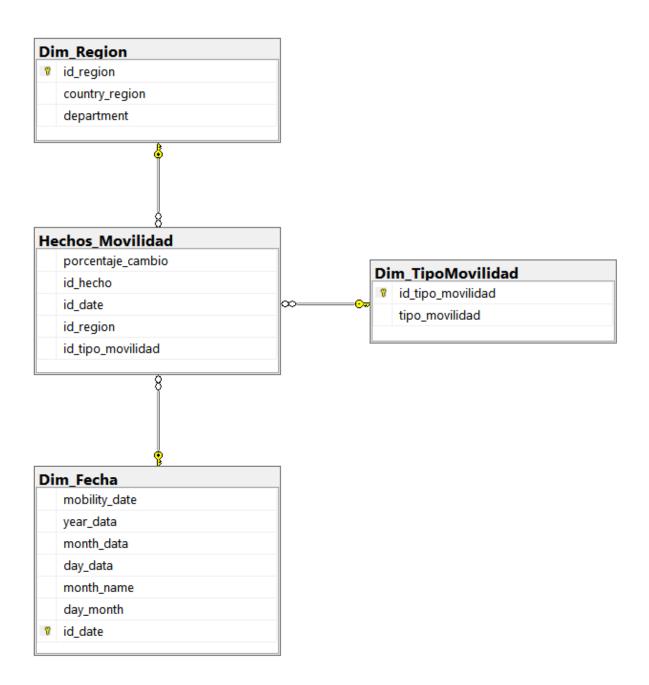


Diagrama de las tablas relacionadas.

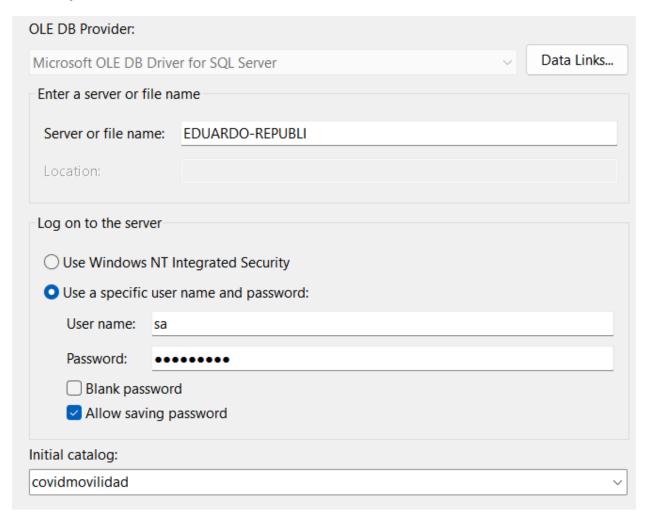


Creación de cubo OLAP (25%)

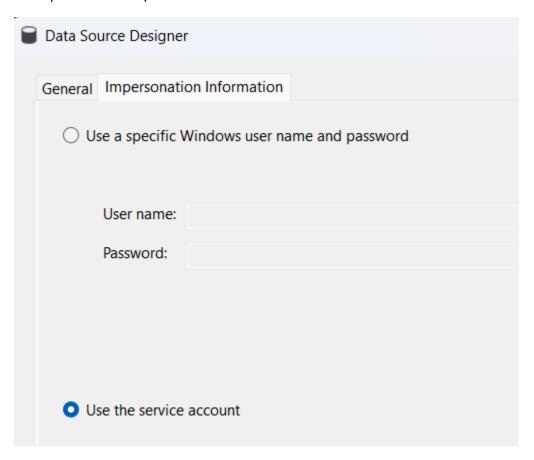
Creamos un proyecto de Analysis Services. En primer lugar, crearemos el Data Source Covidmovilidad.ds (80%)



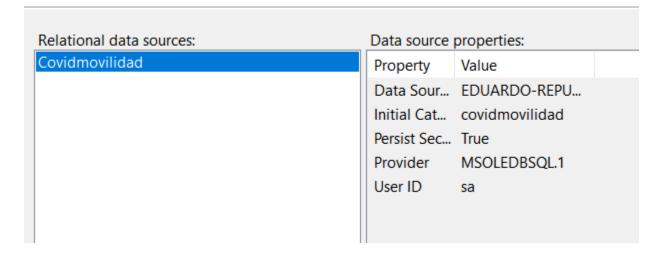
Nos conectamos a la base de datos que poblamos con nuestro proceso ETL con el usuario sa de SQL Server.



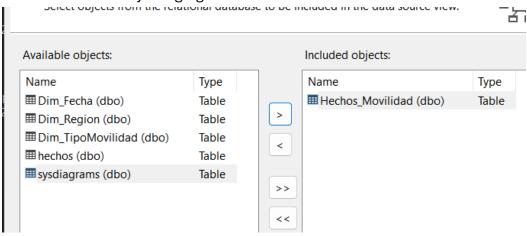
Le especificamos que utilice la cuenta de servicio.



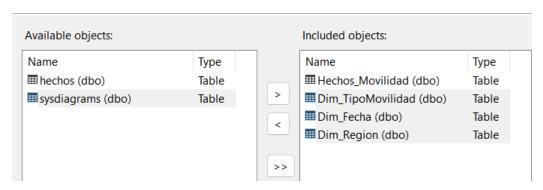
A continuación, crearemos un Data Source View del Data Source que acabamos de crear. **(82.5%)** 



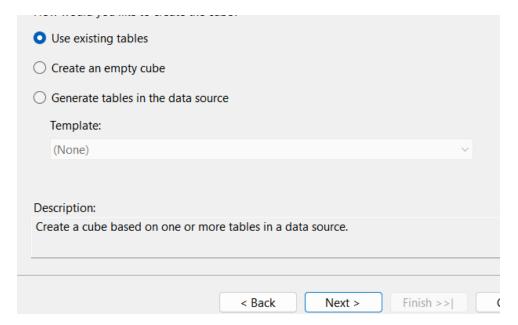
Escogemos que tablas se incluirán, al escoger la tabla de hechos, podemos presionar en Add Related Tables y se agregaran las tablas de dimensiones automáticamente.



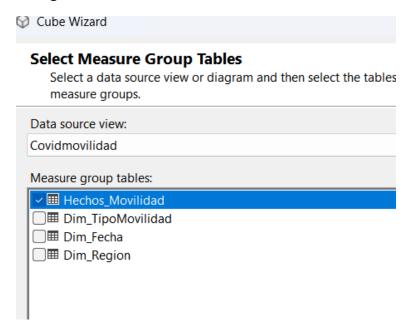
#### Al presionar el botón:



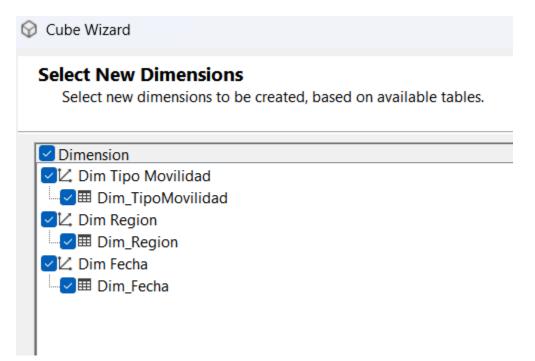
A continuación, crearemos el cubo. Le indicamos que usaremos tablas existentes. (85%)



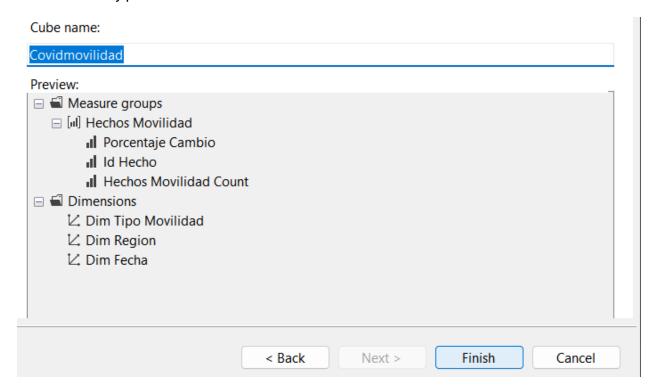
Escogemos la tabla de hechos.



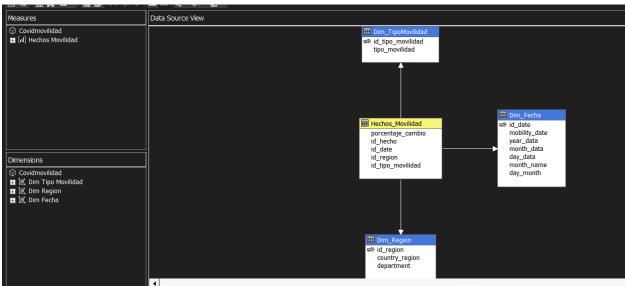
Se crean las dimensiones, basadas en las tablas restantes.



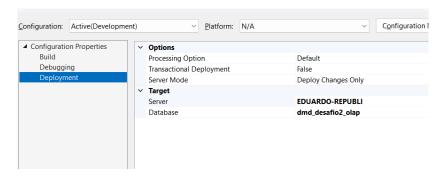
## Confirmamos y presionamos finalizar.



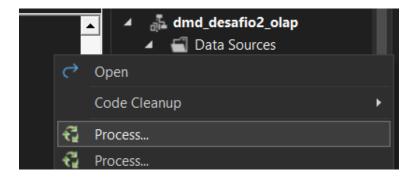
#### Automáticamente se crea el cubo con sus respectivas dimensiones. (90%)



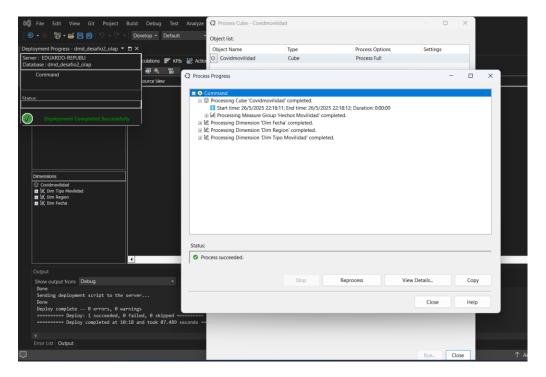
Antes de procesar el cubo, nos vamos a las propiedades de nuestro proyecto, en el apartado de Deployment, en Server reemplazamos "localhost" por el nombre de nuestro servidor de Analysis Services. (92.5%)



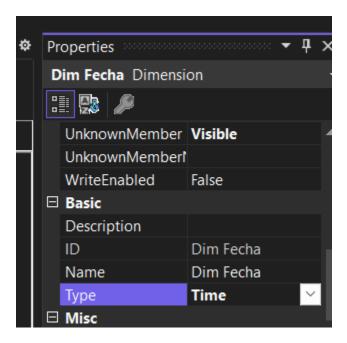
Procedemos a procesar el cubo.



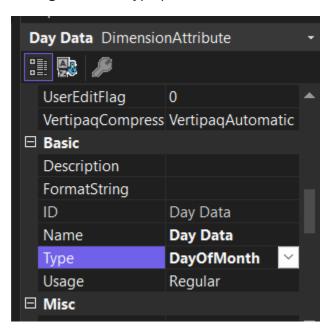
Confirmamos que todo se ejecuto correctamente. Ahora tenemos acceso a mas funcionalidades de nuestro cubo.



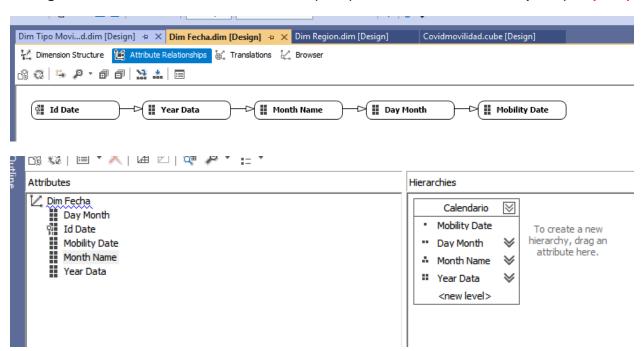
Definimos Dim\_Fecha como nuestra dimensión de tiempo. (95%)



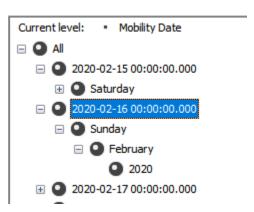
Configuramos el type para cada uno de los campos de Dim\_Fecha



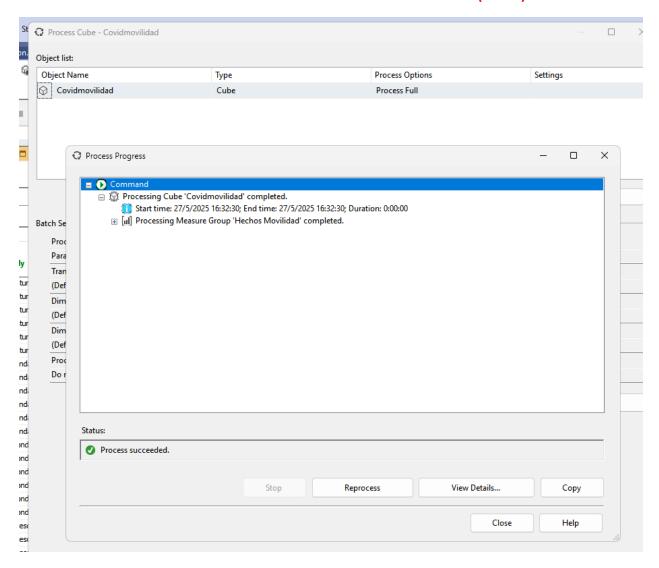
Configuramos las relaciones entre atributos para posteriormente crear una jerarquía (98%)



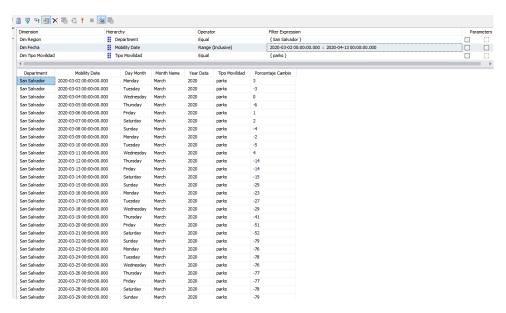
Al procesar el cubo, vemos que las fechas aparecen así:



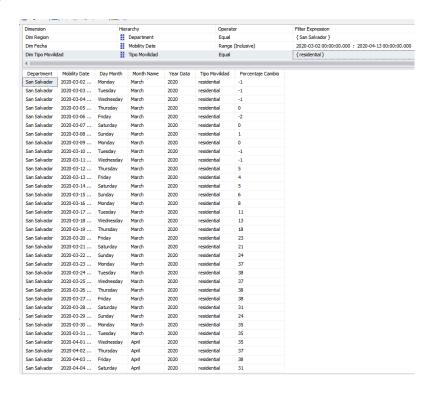
## Procesamos el cubo con la dimensión actualizada correctamente. (100%)



Ahora se nos habilita el explorador en nuestro cubo, en el cual podemos realizar una serie de consultas, por ejemplo, veremos los cambios de movilidad con respecto a los parques, desde el 2 de marzo del 2020 (antes de la cuarentena) y el 13 de abril (durante la cuarentena)



Podemos ver que hay un notorio aumento de valores negativos, los cuales reflejan que hubo muy poca movilidad en parques, ahora apliquemos el mismo filtro, pero para áreas residenciales.



Pudimos notar en el anterior que hay cambios en números positivos, lo cual refleja el aumento de movilidad en lugares residenciales (por claros motivos de cuarentena). Finalmente, verifiquemos las estadísticas de tránsito a nivel nacional durante el periodo de tiempo especificado anteriormente.

