

Proyecto Fase 1

Herramientas que se usaron

- Microsoft Visual Studio 2022
- SQL Server 2022

Descripción de donde se aplicó cada una de las fases del proceso de ETL

Se crearon bloques de secuencias para llevar de una forma más ordenada cada uno de los pasos a seguir y quedo de la siguiente manera:



Pasos que hace nuestro programa

1. Elimina datos de todas las tablas, incluyendo las tablas pivot y tablas de dimensiones y hechos en el Data Warehouse.
2. Llenado de tablas temporales, en este paso se procede a extraer la información de los archivos proporcionados y con esa información se llena las tablas temporales/pivote.
3. Llenado de tablas de dimensión, en este proceso se procede a realizar la limpieza y transformación de la información que se encuentra en nuestras tablas temporales/pivote y se guarda la data lo mas limpia en nuestras dimensiones.
4. Por último procedemos a llenar nuestras tablas de hechos, una vez tenemos nuestras dimensiones listas se procede a enlazar cada una de ellas.

Extracción de datos

Para la extracción de datos se decidió usar directamente las herramientas que proporciona SSIS, se uso el componente Flat File Source quedando el flujo de la siguiente manera:



Los archivos con numeración 01 fueron cargados en nuestra primer base de datos en SQL Server con el nombre seminario2_base1, y los archivos con numeración 02 fueron cargados a la segunda base de datos que también esta en SQL Server con el nombre seminario2_base2.

Transformación y limpieza de datos

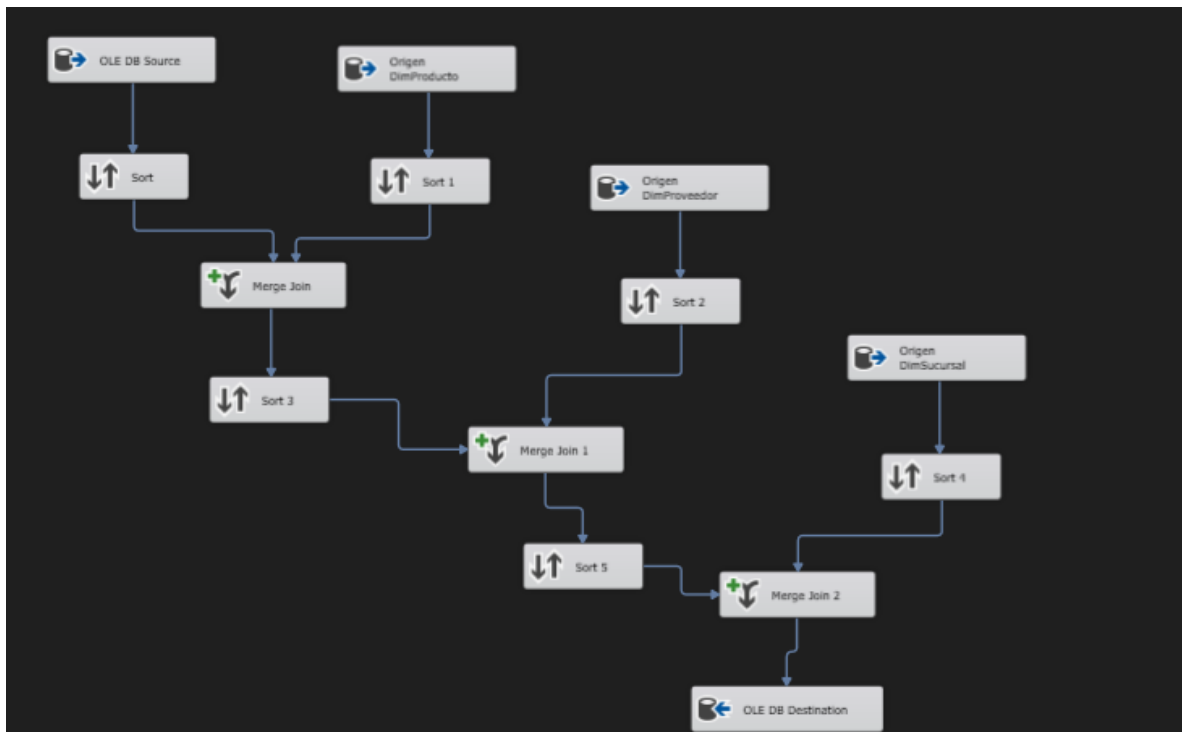
Una vez teníamos ya la extracción de los datos se procedió a transformar cada uno de ellos a su respectivo tipo de dato y a realizar la limpieza, el flujo del mismo quedo de la siguiente manera:



En este flujo se omiten duplicados y también se suprimen los objetos que no traen ID o que sea nulo esto para evitar complicaciones ya que ese identificador funciona como primary key para las dimensiones. También dentro de la limpieza se quita cualquier carácter extraño y se realizan ciertas transformaciones a cadenas de texto para que sean uniformes.

Carga de datos

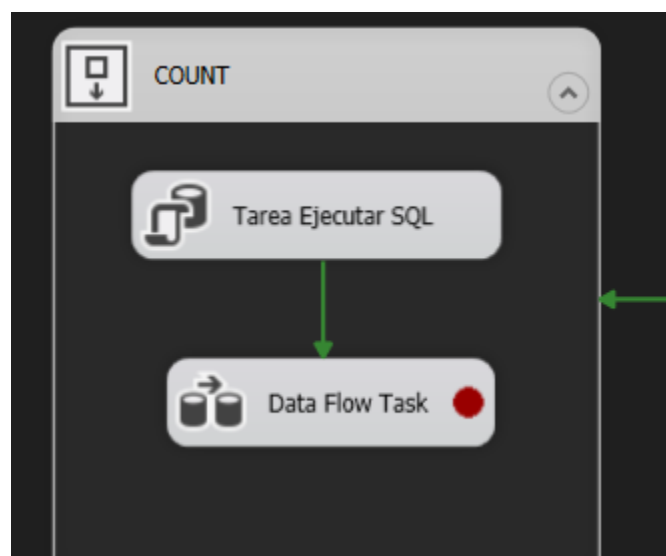
En el paso anterior ya se explicó la forma en que se le hizo la carga a nuestras dimensiones ahora toca cargar nuestra tabla de hechos, el flujo del mismo queda de la siguiente manera:



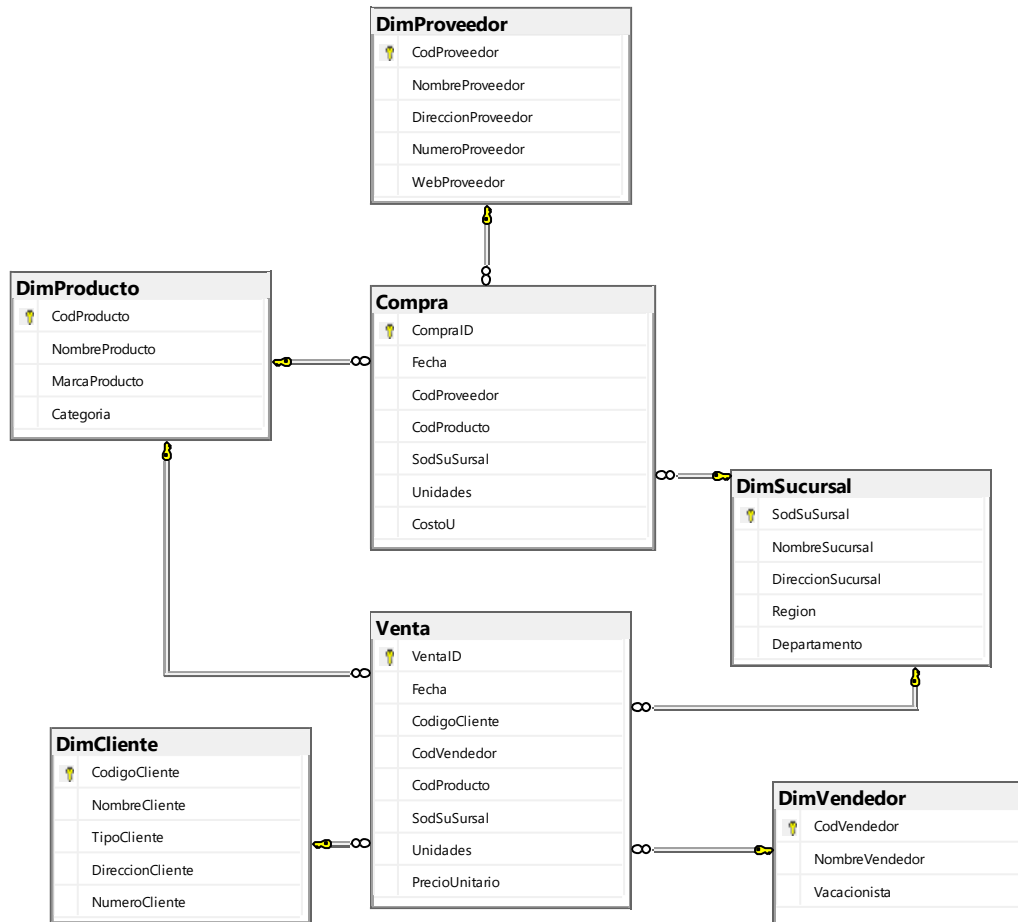
En esta parte con ayuda de los componentes de ordenamiento y los merge join se logro llenar las tablas de hechos que son Compras y Ventas.

Count

En nuestro proceso final del proyecto se solicitaba realizar una consulta para ver la cantidad de registros en todas las tablas de nuestro DataWarehouse por lo que el flujo para ello quedo de la siguiente manera:



Modelo usado para el DataWarehouse



Se usó el modelo **constelación** para este almacén de datos debido a su capacidad para manejar eficazmente la complejidad de los datos transaccionales y facilitar el análisis detallado. En este modelo, las tablas de hechos contienen las métricas clave o los eventos centrales que se desean analizar, mientras que las tablas dimensionales proporcionan contexto y detalles adicionales para esos eventos.

Tablas de dimensiones

- **Producto:** Describe los productos disponibles para compra o venta, incluyendo su código, nombre, marca y categoría
- **Proveedor:** Proporciona detalles sobre los proveedores, como su código, nombre y dirección.
- **Sucursal:** Ofrece detalles sobre las sucursales, como su código, nombre, dirección y ubicación geográfica
- **Cliente:** Contiene información sobre los clientes, como su código, nombre y tipo.
- **Vendedor:** Ofrece datos relacionados con los vendedores, incluyendo su código, nombre y estado de vacaciones.

Tablas de hechos

- **Compra:** Registra las transacciones de compra, con métricas como unidades, costos y fechas.
- **Venta:** Almacena las transacciones de venta, incluyendo información como unidades vendidas, precios y fechas.