

Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ciencias y Sistemas
Seminario de Sistemas 2



Documentación

SG-Food

Nombre: Douglas Alexander Soch Catalán
Carnet: 201807032

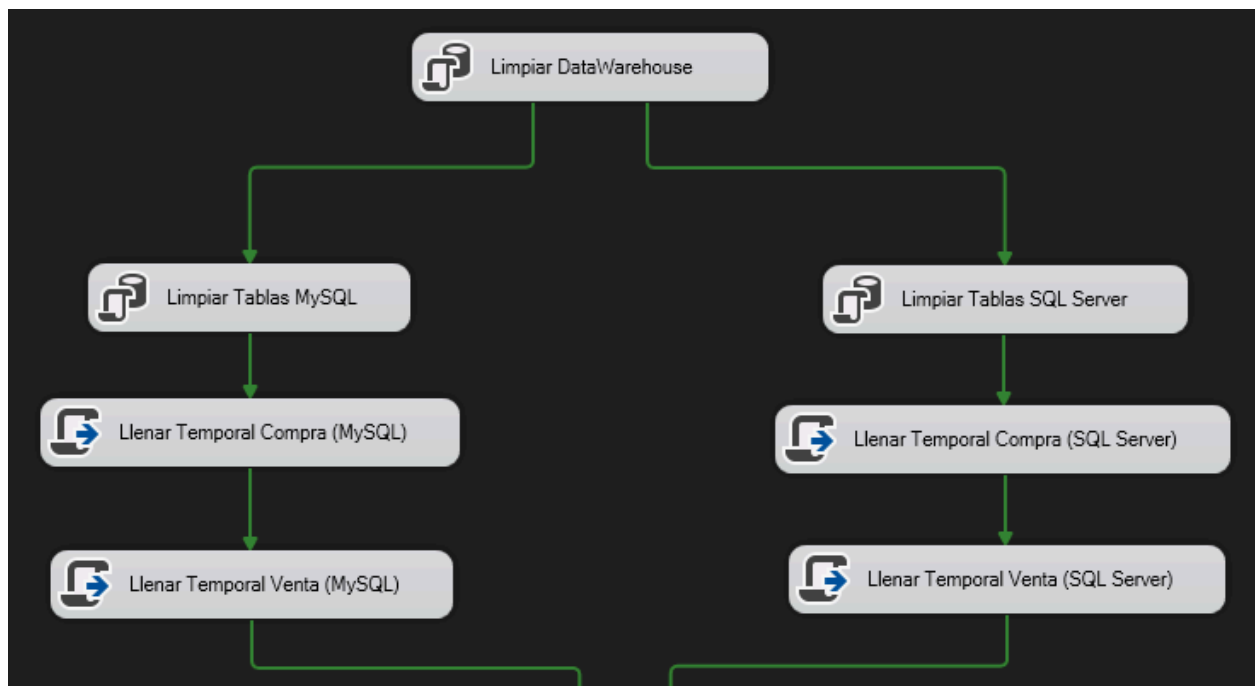
Herramientas utilizadas

- MySQL
- SQL Server 2022
- Visual Studio 2022

Fases ETL

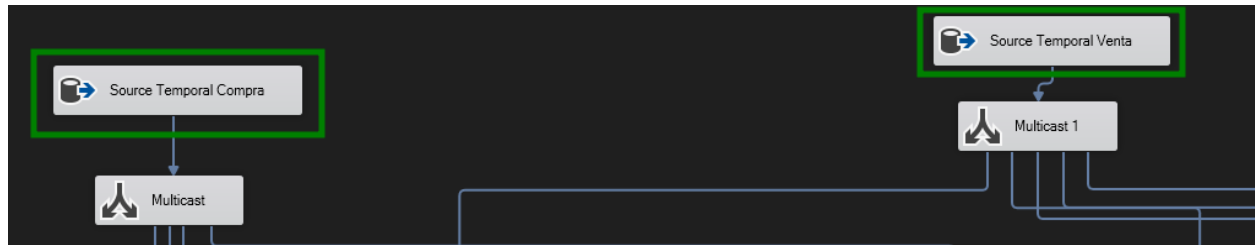
Inicialización

Antes de embarcarnos en las diversas etapas del proceso ETL, es fundamental dar inicio al procedimiento de carga de información en tablas temporales. Este proceso preliminar en tablas temporales asegura la integridad y calidad de los datos, brindando una sólida base para las etapas subsiguientes del ETL. Cabe destacar que se utilizaron dos gestores de bases de datos para almacenar la información temporalmente los cuales son MySQL y SQL Server.



Extracción (E)

Debido a que en la etapa de inicialización se cargó la información en las tablas temporales, ahora se realiza la extracción de información de la siguiente manera:



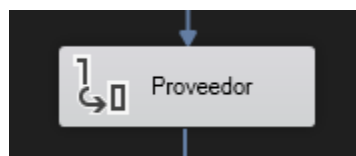
Cabe recalcar que la extracción de información de cada medio definido en el proyecto (base de datos, archivo) es el mismo para todos los casos.

En esta sección se utilizó un componente llamado Multicast para un mejor manejo de la información extraída.

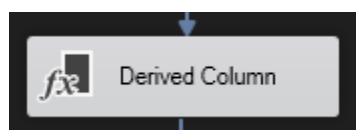
Transformación (T)

Para esta etapa se realizaron varios procesos para transformar la información extraída los cuales son los siguientes:

- Conversión de la información: En este componente se busca definir el tipo de dato que representa el atributo obtenido.

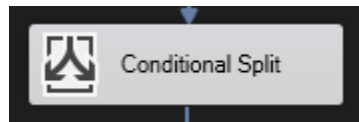


- Columnas derivadas: En este componente, se generan nuevas columnas con valores actualizados según las necesidades específicas. En esta instancia, se ha establecido la norma de asignar el valor NULL a los datos vacíos, formatear el atributo de fecha y convertir algunos datos a tipos enteros o decimales. En situaciones donde un valor no cumple con los requisitos de conversión o la definición de NULL, se excluye directamente, ya que no es considerado como un valor válido para el atributo previamente definido.



- Separación condicional: Este componente desempeña la función crucial de definir las reglas que determinan la idoneidad de un valor para su incorporación en el Datawarehouse. Específicamente, se ha establecido que un valor de tipo string se considera no válido si es NULL, y un valor numérico no se incluirá si es

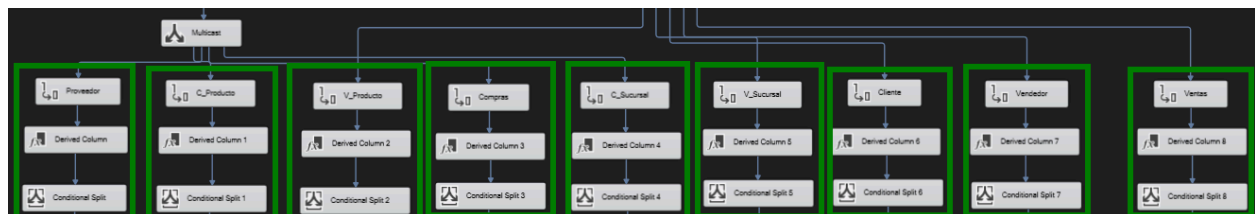
menor a 0. Estas pautas garantizan la integridad de los datos que ingresan al Datawarehouse, asegurando que solo los valores que cumplen con criterios específicos sean incorporados en el proceso.



- Ordenamiento: Este componente fue empleado con el propósito de evitar la entrada de registros duplicados en las tablas dimensionales, asegurando así que no se repitan y garantizando una integración precisa de la información.

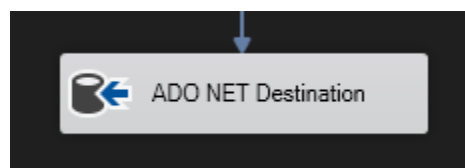


Cada uno de los componentes fue utilizado para construir el Datawarehouse, por lo cual se obtiene el siguiente diagrama.



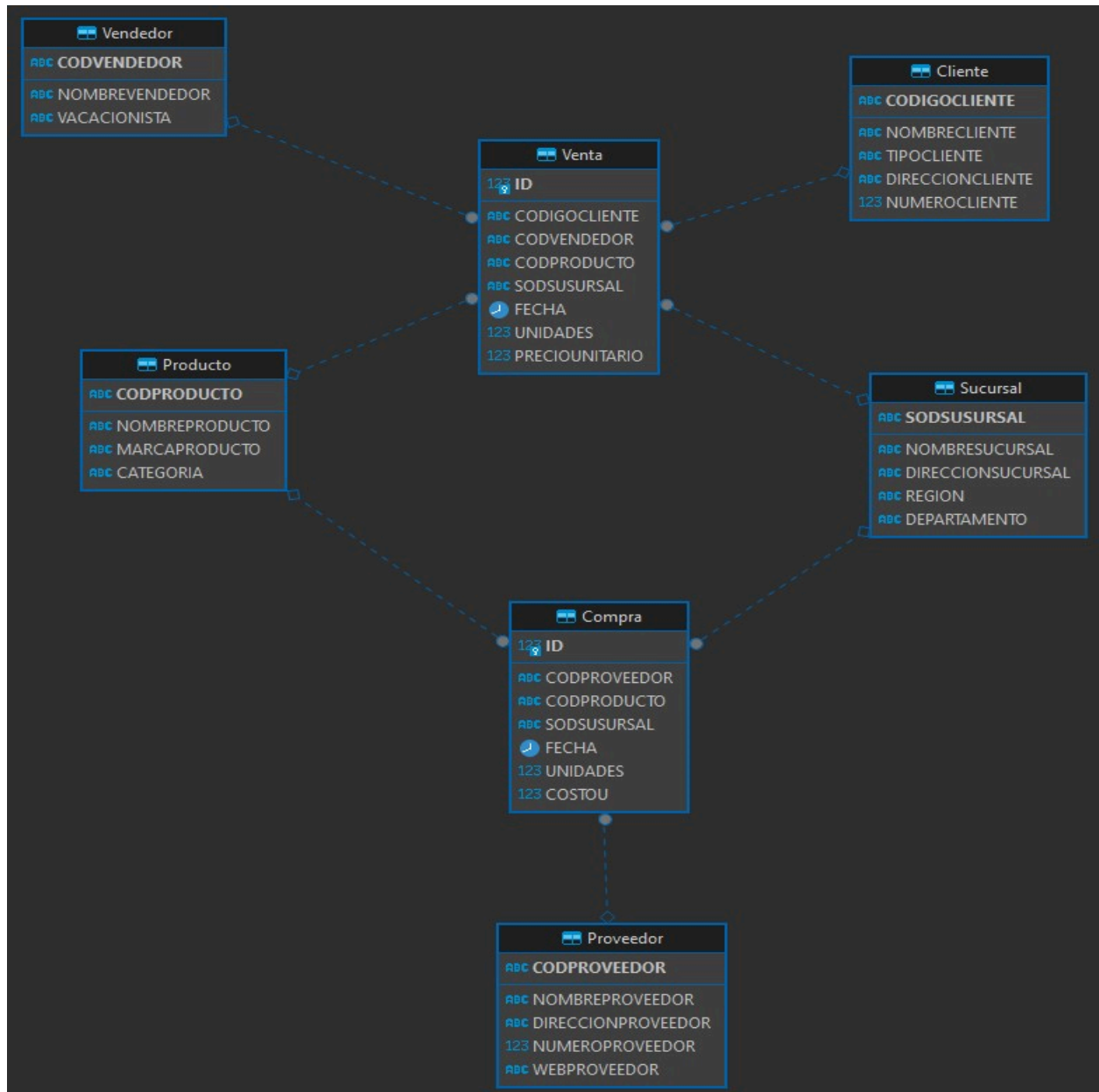
Carga (L)

En esta fase, se llevó a cabo la carga de información en el Datawarehouse, utilizando este componente de manera sistemática en cada etapa de transformación.



Modelo

Para el Datawarehouse obtenido se ha utilizado el modelo estrella debido a que es el más sencillo en su estructura ya que consta únicamente de una tabla central de hechos y varias tablas que están unidas únicamente a la tabla de hechos. Por lo cual, se obtuvo el siguiente diagrama.



Se han definido las siguientes tablas debido a la existencia de dimensiones que se relacionan con más de una tabla de hechos.

Tablas de hechos:

- Compra: Registra la información relacionada con las transacciones de compra realizadas, proporcionando detalles sobre los productos adquiridos, los proveedores involucrados y otros aspectos relevantes.
- Venta: Contiene datos sobre las transacciones de venta efectuadas, incluyendo información detallada acerca de los productos vendidos, los clientes y vendedores involucrados como también otros elementos pertinentes.

Tablas dimensionales:

- Vendedor: Almacena información específica sobre los vendedores.
- Cliente: Registra detalles clave sobre los clientes, como información demográfica, y cualquier otro dato útil para comprender y atender mejor a la clientela.
- Producto: Contiene datos detallados sobre los productos ofrecidos, incluyendo marca y categoría.
- Sucursal: Almacena información sobre las distintas sucursales o ubicaciones de la empresa, facilitando la gestión logística y el seguimiento de actividades específicas asociadas con cada localización.
- Proveedor: Registra información sobre los proveedores que suministran productos a la empresa, incluyendo detalles de contacto, y otras variables relevantes para la gestión de la cadena de suministro.