

**Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ciencias y Sistemas
Seminario de Sistemas 2 Sección A
Ing. Luis Alberto Vettorazzi España
Aux. Freddy Alejandro Monterroso**



Proyecto #1 – Fase 1

**Daniel Estuardo Chicas Carías
201807079**

Proyecto para la empresa “FoodVital”

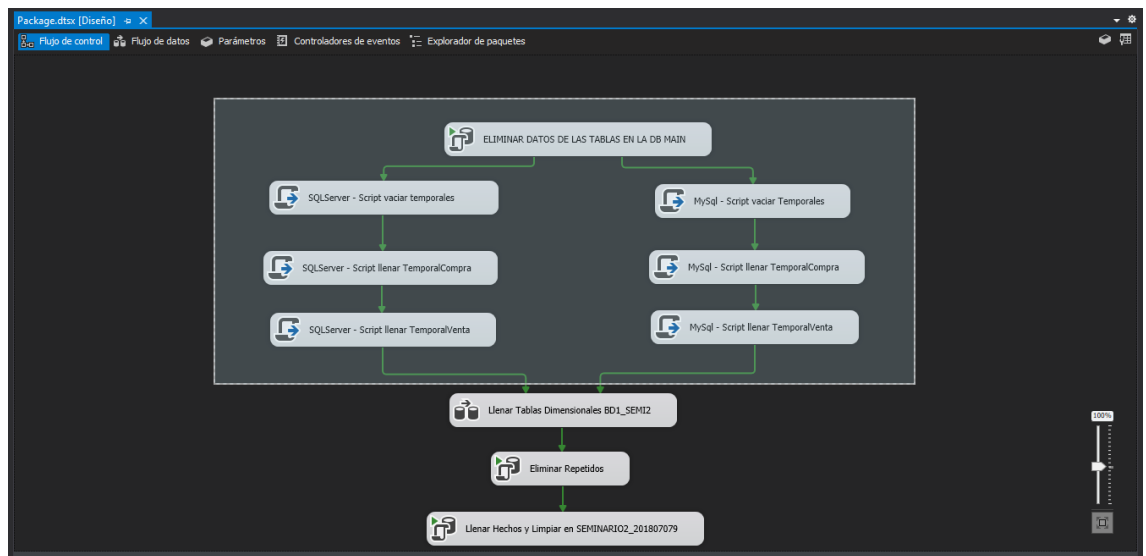
¿Qué es ETL?

El proceso ETL es un conjunto de pasos utilizado en la integración de datos para mover y consolidar datos de múltiples fuentes a un almacén de datos. Las fases del proceso ETL son las siguientes:

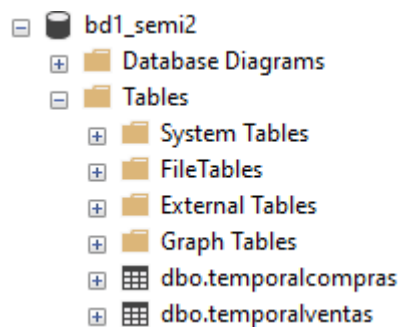
- **Extracción (Extract):**

En esta fase, los datos se extraen de diversas fuentes, que pueden incluir bases de datos, archivos, sistemas en línea, etc.

El objetivo es recopilar la información necesaria de las fuentes para su posterior procesamiento. En el proyecto se realizó en la parte de carga a las tablas temporales de las bases de datos de mysql y sql server donde únicamente se encuentran tablas pivote:



Sql Server:



MySQL



- **Transformación (Transform):**

Durante esta fase, los datos extraídos se transforman en el formato necesario para el análisis y almacenamiento en el almacén de datos.

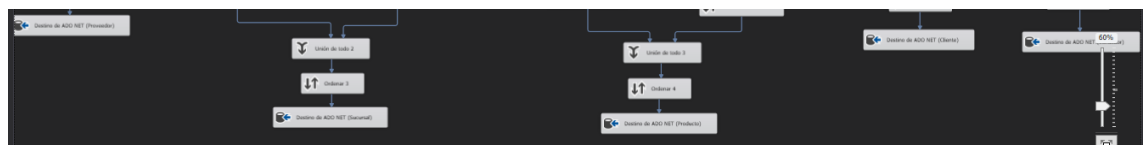
Las transformaciones pueden incluir limpieza de datos, conversión de formatos, cálculos, derivación de nuevas variables, y otras manipulaciones para preparar los datos. En este proyecto se utilizó la sección “Llenar Tablas Dimensionales”:



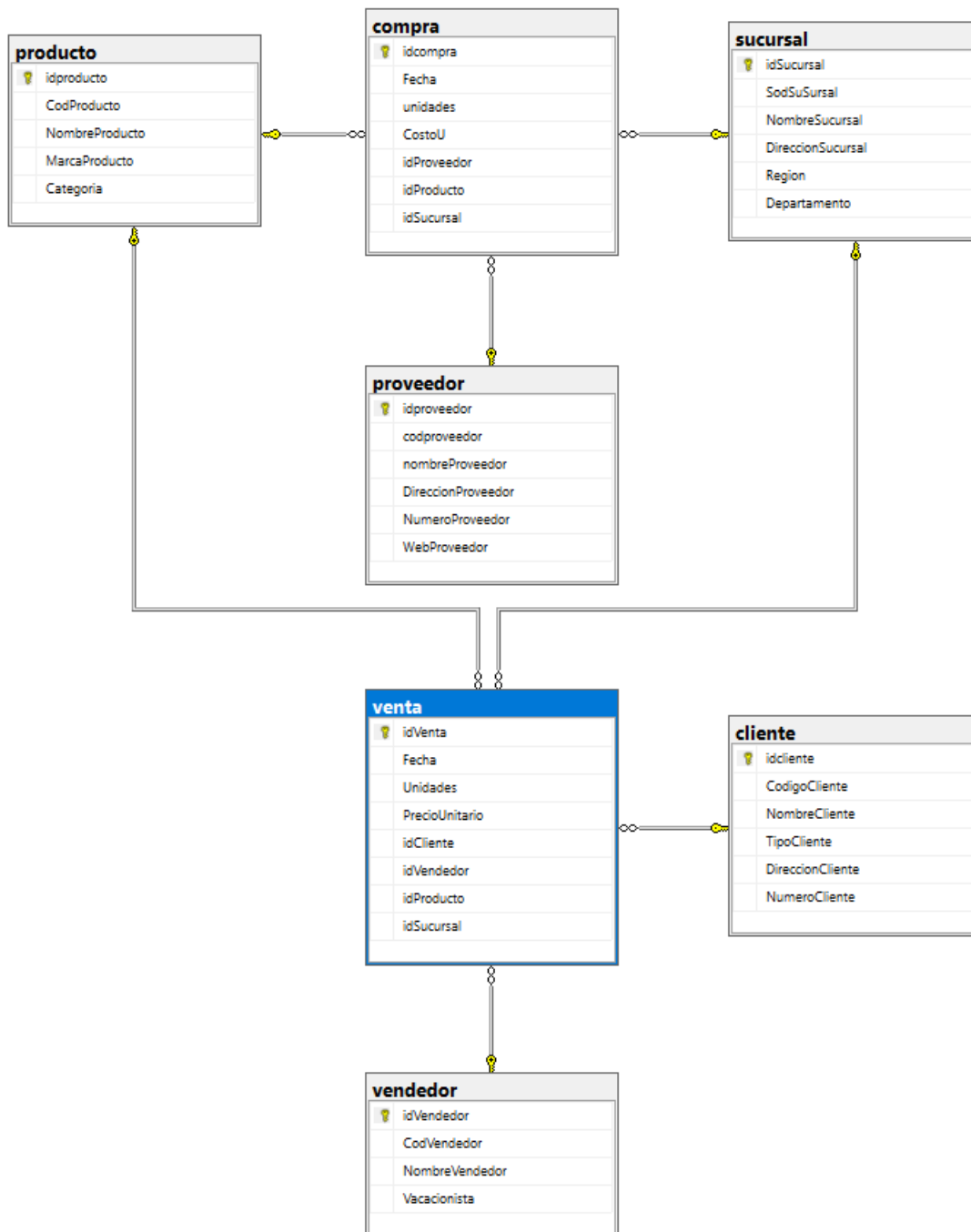
- **Carga (Load):**

En la fase de carga, los datos transformados se cargan en el almacén de datos, que puede ser un almacén de datos relacional, un datawarehouse, un data lake, o cualquier otro sistema de almacenamiento diseñado para la analítica.

La carga puede ser incremental (solo datos nuevos o modificados) o completa, dependiendo de los requisitos del sistema. Para el proyecto esta parte se cumple en la última sección “Llenar Hechos y limpiar en base principal”, ya que se termina la carga de datos a la base de datos.



Modelo de Datawarehouse utilizado: **Galaxia**



¿Por qué utilicé este modelo? Elegí el modelo galaxia por sus beneficios: facilita el análisis combinado de datos de diversas fuentes, reduce costos al compartir dimensiones, y mejora el rendimiento de consultas. Implementé el modelo creando tablas de hechos para compras y ventas, tablas de dimensiones para datos compartidos, y establecí relaciones para análisis conjunto. Este enfoque permite generar informes sobre compras y ventas de manera eficiente, cumpliendo con los requisitos que se establezcan.