



UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA
ESCUELA DE CIENCIAS Y SISTEMAS
SEMINARIO DE SISTEMAS 2
SECCIÓN A
ING. LUIS VETTORAZZI
AUX. ESCARLETH VELASCO

SMART ETL

Proyecto Fase 1

Introducción

En la era digital en la que nos encontramos, la cantidad de información generada día a día es abrumadora. En este contexto, el análisis y procesamiento de datos se ha vuelto una tarea crucial para cualquier organización que quiera tomar decisiones informadas y basadas en evidencia. Es en este contexto donde surge la necesidad de implementar un proceso de Extracción, Transformación y Carga (ETL, por sus siglas en inglés), el cual permite integrar y consolidar información de diversas fuentes para ser analizada y utilizada en la toma de decisiones.

El presente proyecto tiene como objetivo implementar un proceso de ETL, con el fin de integrar y consolidar información relevante de diversas fuentes para su posterior análisis. En el siguiente documento se detalla el proceso de ETL implementado y se presenta el modelo de DataWarehouse diseñado para almacenar y acceder a la información de manera eficiente.

Proceso ETL

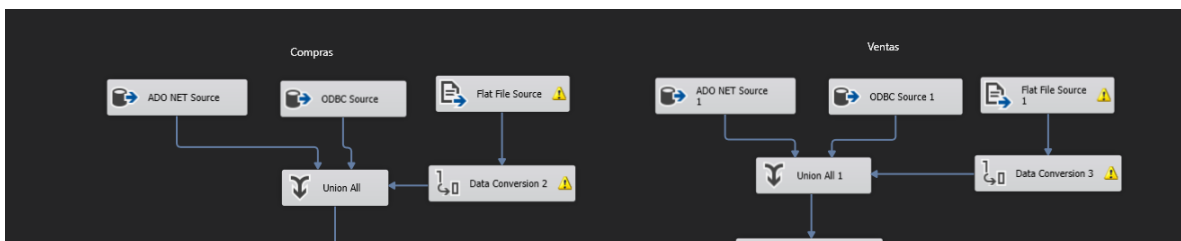
Este proyecto involucró un proceso de ETL (Extract, Transform, Load) para integrar datos de diversas fuentes y prepararlos para su análisis en un DataWarehouse. A continuación, se detalla dónde se aplicó cada una de las fases:

Extract (Extracción)

En esta fase se obtuvieron los datos de diferentes fuentes:

1. Base de datos SQL Server
2. Base de datos MySQL
3. Archivo CSV

Los datos se extrajeron utilizando SSIS.



Transform (Transformación)

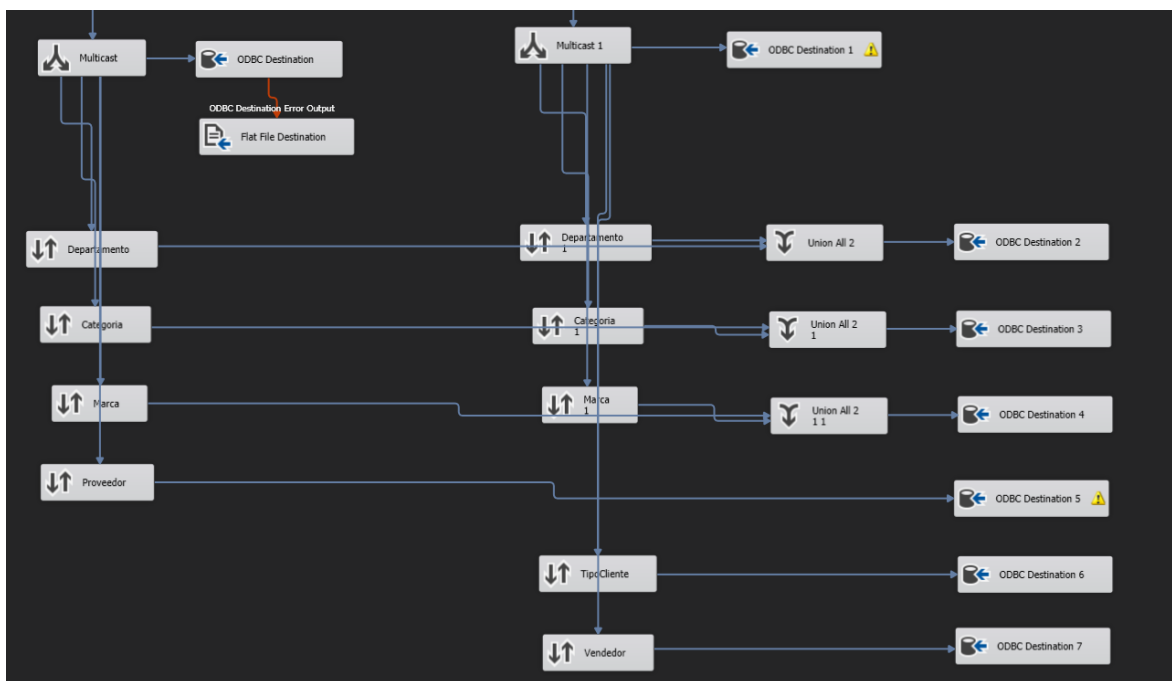
En esta fase se limpiaron, integraron y transformaron los datos para que puedan ser utilizados en el análisis. Se aplicaron técnicas de limpieza y normalización de datos, así como transformaciones para unificar formatos y asegurar la consistencia de los datos. Las transformaciones fueron:

1. Limpieza de datos nulos
2. Limpieza de números según su conjunto



Load (Carga)

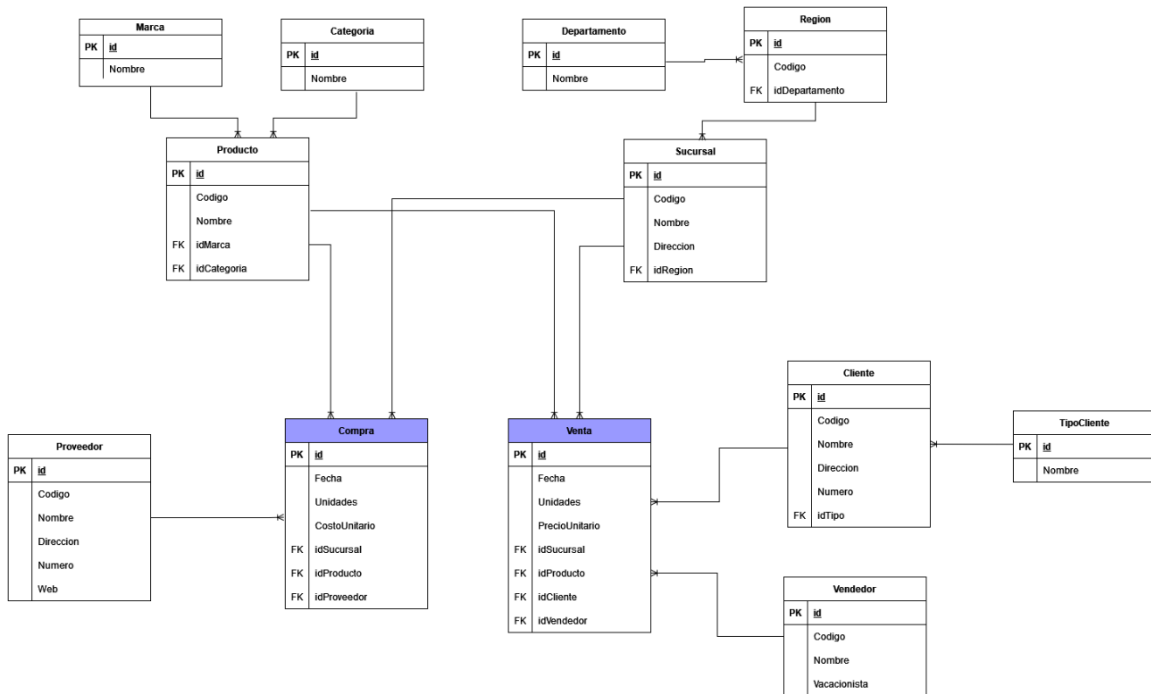
En esta fase se cargaron los datos limpios y transformados en el DataWarehouse. Se utilizó SSIS para este proceso.



Modelo para el DataWarehouse

Para el DataWarehouse se implementó un modelo constelación (starflake schema) debido a que las tablas de hechos compartían una o más tablas de dimensiones.

El modelo es el siguiente:



Las tablas de hechos son:

1. Compra
2. Ventas

Las tablas de dimensiones son:

1. Marca
2. Categoría
3. Producto
4. Departamento
5. Región
6. Sucursal
7. Proveedor
8. TipoCliente
9. Cliente
10. Vendedor

Se observan que las dimensiones producto y sucursal son compartidas entre las tablas de hechos.