Proceso de transformación de datos y carga en el data mart final

Elaborado por

Sol Elena Serna Vargas

Docente

Antonio Jesús Valderrama

Institución Universitaria Digital de Antioquia

Base de datos II

Medellín 30 de septiembre de 2025

INTRODUCCION

En este trabajo tiene como objetivo el proceso ETL (Extracción, Transformación y Carga) utilizando consultas SQL en MySQL. El propósito fue organizar la información de la base de datos Jardinería y dejarla lista para el análisis dentro de un Data Mart estructurado.

El proceso lo dividí en tres etapas principales:

* Extracción desde la base de datos origen.
* Transformación en Staging (limpieza y normalización).
* Carga en un Data Mart.

De esta manera logré que los datos quedaran organizados de forma eficiente, permitiendo realizar consultas analíticas más fáciles y rápidas, como identificar el producto más vendido, analizar ventas por clientes o revisar resultados en diferentes periodos de tiempo.

1. Preparación

Para este trabajo inicialmente revise el modelo Estrella definido en el proyecto de análisis de la base de datos Jardinería para comprender la estructura y las relaciones entre las tablas de dimensiones y la tabla de hechos.

Verifique la disponibilidad y consistencia de la base de datos de staging previamente creada para el proyecto.

1. Extracción de datos (Origen → Staging)

Ya había creado la base de datos staging, en la entrega anteriror que actúa como zona intermedia donde copié la información de la BD Jardinería.

Creé las tablas vacías en staging con la misma estructura.

Inserté los datos directamente con consultas SQL.

* Se copiaron las tablas pedido, detalle\_pedido y pago.
* Inserté los datos directamente con consultas SQL.
* Verifiqué que las tablas tuvieran la misma cantidad de registros que en la base original.

Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.

1. Transformación de datos (Staging)

En staging realicé las transformaciones necesarias para preparar la información:

* Normalización: pasé los estados de los pedidos a mayúsculas para evitar inconsistencias.
* Corrección de valores inválidos: puse en NULL los precios negativos.
* Enriquecimiento: creación y cálculo de la columna montoTotal = cantidad \* precio\_unidad en detalle\_pedido.

Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.

1. Carga al Data Mart (Modelo Estrella)

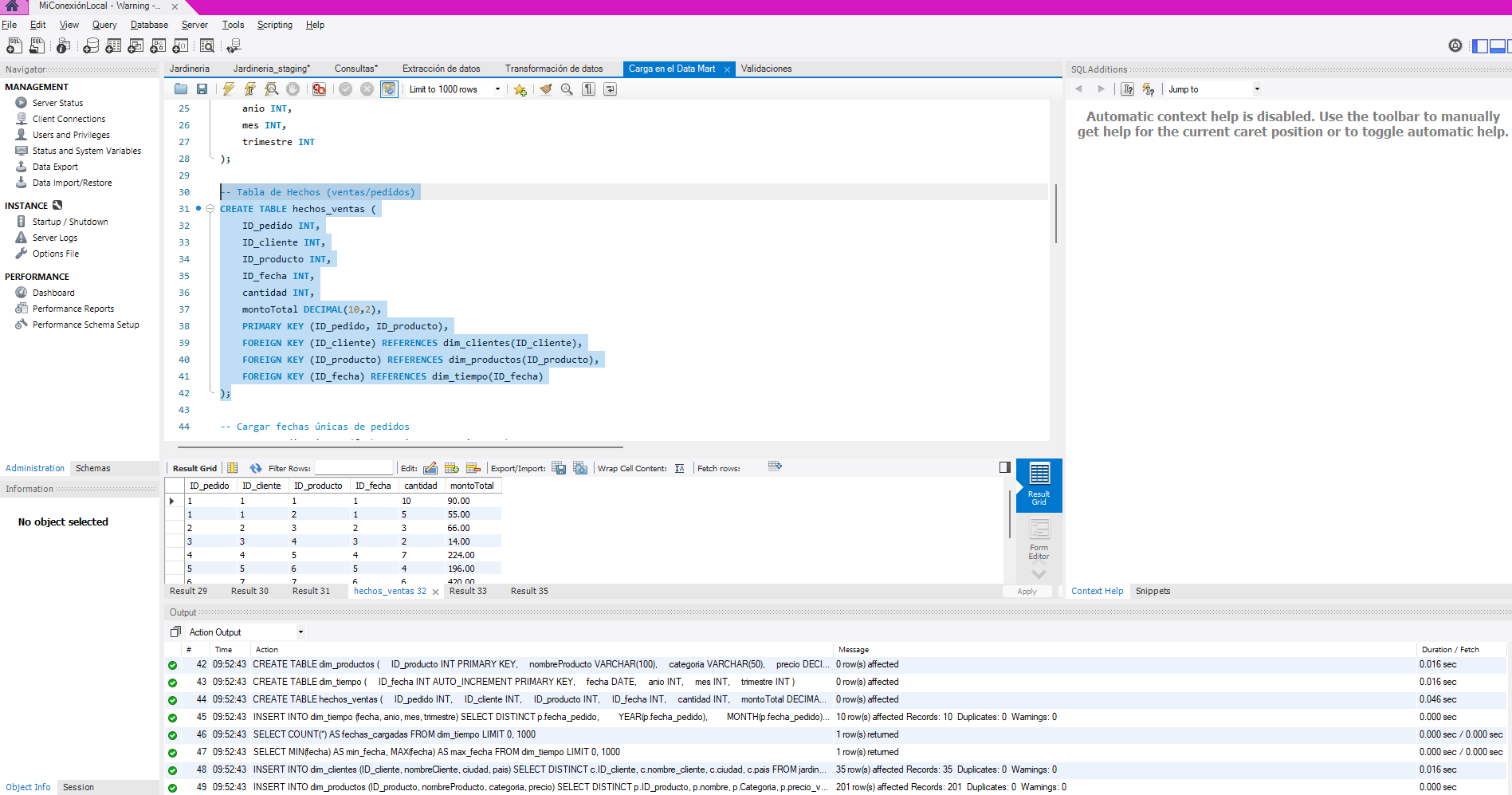
Luego diseñé un Data Mart bajo el modelo estrella, que incluye:

Dimensiones:

* dim\_clientes (información de clientes)
* dim\_productos (información de productos)
* dim\_tiempo (fechas de pedidos)

Tabla de hechos:

* hechos\_ventas (ventas/pedidos con relaciones hacia las dimensiones)



Después cargué los datos de staging hacia el Data Mart con consultas INSERT SELECT.

Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.

Validaciones

Finalmente realicé verificaciones para confirmar la calidad de los datos cargados y para asegurar la consistencia:

* Conteo de registros en cada dimensión y en la tabla de hechos.
* Validación de claves foráneas para validar que no hubiera clientes, productos o fechas inexistentes en la tabla de hechos
* Revisión de algunas filas de hechos\_ventas.

Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.

Conclusiones

Con este proceso logré:

* Implementar un ETL básico usando solo SQL en MySQL.
* Asegurar la limpieza y consistencia de los datos en staging.
* Aplicar técnicas de limpieza, normalización y enriquecimiento de datos para asegurar calidad.
* Construir un Data Mart con modelo estrella, lo que facilita el análisis de la información.
* Dejar listo un esquema para realizar consultas analíticas como: ventas por cliente, ventas por producto, ventas mensuales o trimestrales.

En resumen, este ejercicio me permitió entender cómo se organiza un flujo ETL completo y cómo el modelo estrella ayuda a estructurar los datos para análisis de negocio.

Anexos

Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.