**Evidencia de aprendizaje 3. Proceso de transformación de datos y carga en el data mart final**

Cristian Andréi Rendón Alcaraz

Kevin Saldarriaga Vélez

Santiago Agudelo Escobar

**Ingeniería de Software y datos**

Antonio Jesús Valderrama Jaramillo

**Docente**

Bases de Datos II-PREICA2502B010064

**Curso**

**Institución Universitaria Digital de Antioquia**

**Antioquia**

**2025**

**Evidencia de aprendizaje 3. Proceso de transformación de datos y carga en el data mart final**

**Introducción**

Este informe documenta el proceso completo de **Extracción, Transformación y Carga (ETL)** implementado para la creación del Data Mart **dm\_jardineria**, diseñado bajo un **modelo dimensional tipo Estrella**. El objetivo principal es consolidar información dispersa del área de operaciones en un repositorio optimizado para análisis de negocio y toma de decisiones estratégicas.

**Objetivos**

Los principales objetivos dentro la elaboración de la base de datos es centralizar datos operacionales en un esquema analítico, también buscamos facilitar consultas de inteligencia de negocio, de ese mismo modo mejorar la calidad de los datos mediante limpieza y transformación para luego enriquecer la información con métricas calculadas y de este modo obtener y optimizar el rendimiento de consultas analíticas.

**Arquitectura del Data Mart**

El diseño implementado sigue el paradigma **Estrella (Star Schema)**, que se caracteriza por:

**Componentes principales:**

**Tabla de Hechos Central**

* hecho\_ventas: Contiene las métricas cuantitativas del negocio (cantidad, precio, total de línea)

**Tablas Dimensionales (4 dimensiones)**

* dim\_producto: Información sobre productos, categorías y proveedores
* dim\_cliente: Datos demográficos y de ubicación de clientes
* dim\_empleado: Información de la fuerza de ventas
* dim\_tiempo: Dimensión temporal para análisis cronológicos

**Etapas del proceso ETL**

|  |  |
| --- | --- |
| Etapa | Descripción |
| Creación del Data Mart. | Definición de la estructura dimensional (tablas de dimensión y hecho). |
| Transformación y Carga (T+L). | Extracción desde **staging\_jardineria**, aplicación de transformaciones y carga en **dm\_jardineria**. |
| Análisis de negocio. | Consultas en SQL para responder preguntas estratégicas. |
| Validación de calidad. | Verificación de integridad, conteos, y cálculos |

Texto

El contenido generado por IA puede ser incorrecto.

Enriquecimientos: Cálculo del margen (%).

Manejo de nulos: **COALESCE** para evitar errores si **precio\_proveedor** es **NULL.**

Normalización: Uso de **LEFT JOIN** para no perder productos sin categoría

Texto

El contenido generado por IA puede ser incorrecto.

Limpieza: Se selecciona solo los campos relevantes para el análisis.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

El contenido generado por IA puede ser incorrecto.

Normalización: Unificación del nombre completo

Enriquecimiento: Agrega contenido geográfico del empleado.

Texto

El contenido generado por IA puede ser incorrecto.

Generación dinámica: Solo fechas con pedidos reales.

Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.

Integración: Combina detalles de pedido, pedido y cliente.

Cálculo: **total\_linea** se precalcula para optimizar consultas.

Clave de empleado: Se obtiene del cliente (**ID\_empleado\_rep\_ventas**).

Producto más vendido (por unidades).

Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.

Interfaz de usuario gráfica, Aplicación, Word

El contenido generado por IA puede ser incorrecto.

Producto más vendido (por ingresos).

Texto

El contenido generado por IA puede ser incorrecto.

Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.

Cliente con mayor volumen de compras.

Texto

El contenido generado por IA puede ser incorrecto.

Interfaz de usuario gráfica

El contenido generado por IA puede ser incorrecto.

Empleado con más ventas (en valor).

Texto

El contenido generado por IA puede ser incorrecto.

Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.

Ventas por trimestre.

Tabla

El contenido generado por IA puede ser incorrecto.

Validación de calidad de datos.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

El contenido generado por IA puede ser incorrecto.

**Conclusión**

Este script implementa un flujo ETL completo y robusto que transforma datos operativos en un Data Mart. El modelo dimensional facilita consultas rápidas y análisis profundos, mientras que las transformaciones garantizan calidad, coherencia y valor empresarial.