

EVIDENCIA DE APRENDIZAJE 3



PROCESO DE TRANSFORMACIÓN DE DATOS Y CARGA EN EL DATA MART FINAL

PRESENTADO POR:
JOSÉ DAVID CARDONA MAZO
CARLOS ESTEBAN TRUJILLO PAZ
DANIELA LOZANO MORALES
JUAN FELIPE PARRA PALACIOS
DIEGO BARROS

INGENIERÍA DE SOFTWARE Y DATOS
BASE DE DATOS II
PREICA2502B010064

DOCENTE:
ANTONIO JESÚS VALDERRAMA JARAMILLO

INSTITUCION UNIVERSITARIA DIGITAL DE ANTIOQUIA
MEDELLIN, COLOMBIA

SEPTIEMBRE 2025

1.Preparación

Corregimos el modelo estrella presentado en la primera actividad. En la versión inicial, los datos fueron cargados directamente desde la fuente sin pasar por el staging. Ahora, en esta actualización, los datos que almacenamos provienen directamente del staging, mejorando la calidad y organización del flujo de datos.

Además, como parte de esta actividad, hemos construido un Data Mart orientado al análisis de ventas, utilizando un modelo en estrella a nivel de línea de pedido. Este modelo facilita un análisis más detallado de las ventas, permitiendo realizar consultas específicas a nivel de cada transacción.

- Creación de la base de datos de staging (jardineria_stg) con tablas para oficinas, empleados, categorías, productos, clientes, pedidos, detalle de pedidos y pagos.
- Creación de la base de datos de Data Mart (jardineria_dw) con el esquema 'dw'.
- Diseño de la tabla de hechos FactVentas con PK y FKs hacia las dimensiones.
- Se agregaron constraints y relaciones para garantizar integridad referencial.

2. Planteamiento del Problema

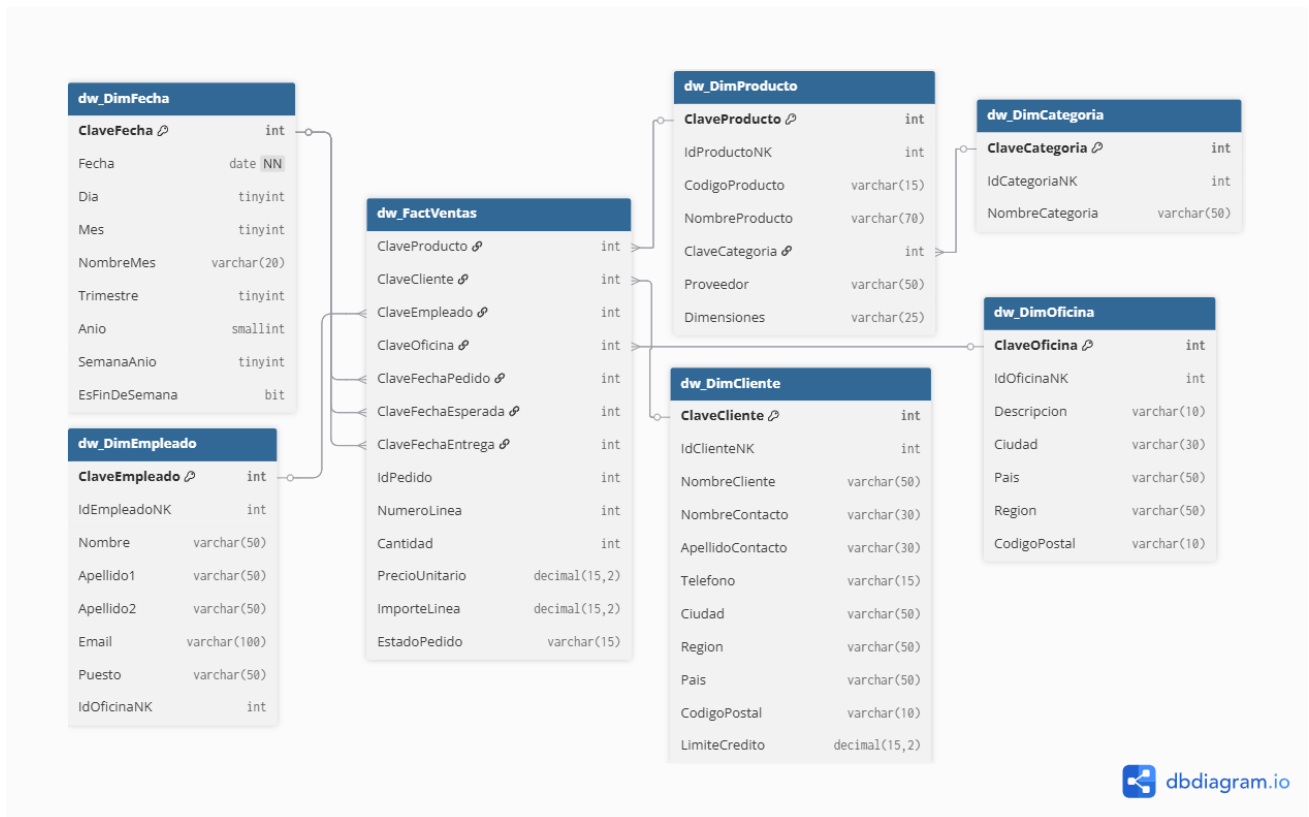
La base transaccional de Jardinería, al estar fuertemente normalizada, resulta poco eficiente para realizar análisis complejos, ya que la información se

encuentra distribuida en múltiples tablas, con riesgos de redundancia y dificultades para la integración. Aunque la creación del Staging representó un primer avance, se detectó la necesidad de aplicar transformaciones adicionales y de consolidar la información en un Data Mart dimensional que soporte consultas analíticas rápidas y optimizadas.

3. Objetivos

- Implementar un proceso de transformación de datos que asegure información limpia, normalizada y coherente.
- Construir y poblar un Data Mart de ventas basado en un modelo en estrella, con un nivel de detalle a nivel de línea de pedido.
- Ejecutar consultas SQL que permitan la carga eficiente desde Staging hacia el Data Mart final.
- Aplicar constraints y relaciones que garanticen la integridad referencial entre tablas.
- Verificar la consistencia de la información mediante validaciones 1:1 entre Staging y Data Mart.
- Demostrar el valor práctico del Data Mart con consultas analíticas relevantes, como la identificación del producto más vendido o el análisis por cliente y región.

La siguiente gráfica muestra el diseño de las dimensiones y la estructura de la base de datos en modelo estrella:



4.Extracción de datos

Llevamos a cabo la extracción de datos desde la base de datos de origen Jardinería hacia la base de datos de Staging, utilizando el script *CargaInfoStaging*, como parte del proceso de preparación para el modelo de datos. Esta etapa fue fundamental porque permitió centralizar la información en un entorno intermedio controlado, asegurando que los registros se trasladaran de manera ordenada y listos para ser sometidos a verificaciones posteriores.



Como parte del proceso de carga al staging, realizamos las siguientes acciones:

- Aplicamos transformaciones de limpieza utilizando funciones como LTRIM, RTRIM y NULLIF para normalizar los valores.
- Registramos cada carga mediante batch_id y extract_ts, lo que permitió asegurar la trazabilidad del proceso.

- Verificamos el contenido del staging tras la extracción ejecutando el script VerificacionStaging para garantizar la consistencia de los datos.

Como resultado de esta verificación:

- Se identificaron y corrigieron orfandades en las tablas de pedidos y detalles de pedido.
- Se validaron y trataron valores negativos o nulos no permitidos en los campos críticos.

5.Transformación de datos

Se prepararon los datos para la construcción del modelo en estrella. En esta etapa, se realizó la carga completa de los datos necesarios para alimentar el modelo.

Primero, se ejecutó el script CargaInfoDIMFECHA para poblar la dimensión de fecha. Posteriormente, se utilizaron los scripts contenidos en DimensionesEstrella para cargar el resto de las dimensiones.

- Se poblaron las fechas en la tabla DimFecha, utilizando el rango de fechas detectado en el staging.
- Se realizó un proceso de MERGE tipo SCD-1 para mantener actualizadas las dimensiones de Cliente, Producto, Categoría, Oficina y Empleado.
- Se resolvieron las claves sustitutas y se establecieron correctamente las relaciones entre las distintas dimensiones.

6.Carga a tablas de hechos

Se ejecutó el script CargaTablaHechos para poblar la tabla de hechos FactVentas, con las siguientes acciones:

- Se resolvieron las claves de fechas mediante un JOIN a la tabla DimFecha.
- Se utilizó un MERGE para realizar operaciones de upsert, evitando duplicados en la tabla de hechos.
- Se garantizó la integridad de los datos mediante una restricción UNIQUE(IdPedido, NumeroLinea).

Validaciones 1:1 entre DW y Staging

Se ejecutó el script Validacion1_1-STG-DW con el objetivo de asegurar la consistencia total entre el Data Warehouse (DW) y el área de staging, realizando las siguientes acciones:

- Se eliminaron registros huérfanos presentes en el DW y se insertaron los registros faltantes para asegurar la integridad.
- Se compararon los conteos y los importes entre staging y DW hasta lograr una igualdad exacta 1:1.

Conclusión

La implementación de esta tercera fase permitió completar el ciclo ETL, asegurando que los datos de Jardinería pasaran por Staging y llegaran al Data Mart final en condiciones óptimas. Con este proceso se logró garantizar la calidad de los

datos, obteniendo información limpia, coherente y libre de redundancias; se optimizó el análisis al disponer de un modelo en estrella que facilita consultas más rápidas y precisas; y se generó un valor empresarial tangible, al posibilitar la obtención de indicadores clave como la identificación del producto más vendido, fundamentales para apoyar la toma de decisiones estratégicas.

Referencias

- Amazon Web Services, Inc. (n.d.). ¿Qué es ETL?. Recuperado de: <https://aws.amazon.com/es/what-is/etl/>
- IBM. (n.d.). Qué es el proceso de extracción, transformación y carga (ETL). Recuperado de: <https://www.ibm.com/es-es/topics/etl>