

S30 - Evidencia de aprendizaje 3. Proceso de transformación de datos y carga en el data mart final

Jonatan Dair Avila Agamez

Juan Fernando Cataño Higueta Jhan

Camilo Moreno Delgado

Base de datos II

Profesor Antonio de Jesús Valderrama

30 de sept. de 25

Universidad Digital de Antioquia

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

El presente trabajo se compone dentro del proceso de modelado dimensional aplicado a la base de datos Jardinería, con el propósito de diseñar e implementar un Data Mart orientado al análisis integral de ventas, clientes, productos y recursos humanos. A través de la construcción de un modelo estrella que organiza la información en tablas de dimensiones y una tabla de hechos, se busca no solo describir su estructura, sino también aprovechar las oportunidades que ofrece para generar conocimiento estratégico en la toma de decisiones.

La presente evidencia tiene como finalidad documentar el proceso completo de creación del Data Mart Jardinería, a partir de la base de datos transaccional original y del entorno de staging. Este proceso contempla las fases del ciclo ETL (Extract, Transform, Load), garantizando la correcta extracción de los datos, su transformación mediante técnicas de limpieza y normalización, y su carga final en un esquema dimensional. De esta forma, se asegura la coherencia y calidad de la información para un uso analítico en contextos reales de inteligencia de negocios.

1. Objetivos

Objetivo general:

Construir un Data Mart en modelo estrella basado en la BD Jardinería_staging para análisis de ventas.

Objetivos específicos:

- Analizar el modelo estrella de la base Jardinería.
- Diseñar las tablas de dimensiones y la tabla de hechos.
- Aplicar técnicas de transformación de datos.
- Poblar el Data Mart con datos consistentes.

- Validar la carga de información mediante consultas SQL.

2. Planteamiento del problema

La base de datos Jardinería_staging es un sistema transaccional que registra clientes, productos, pedidos, empleados y oficinas. Sin embargo, este esquema no está optimizado para el análisis de información. Se requiere construir un Data Mart que permita centralizar métricas de negocio y facilitar la toma de decisiones.

3. Análisis del problema

El análisis permitió identificar las entidades principales: cliente, producto, empleado, oficina, y los pedidos/detalles de pedido. Estas entidades se organizaron bajo un modelo estrella, donde las dimensiones contienen datos descriptivos y la tabla de hechos concentra las métricas de ventas.

4. Propuesta de solución

La solución planteada fue construir un Data Mart denominado Jardineria_DM, que integra dimensiones y una tabla de hechos de ventas. Se utilizaron sentencias SQL para la creación de tablas y carga de datos desde la base de datos origen jardineria_staging.

5. Desarrollo

A continuación, se presentan los scripts de creación y carga de cada dimensión y de la tabla de hechos.

Para dar solución al problema planteado, se diseñó un Data Mart denominado Jardineria_DM, construido a partir de la base de datos de staging previamente implementada.

Las tablas de Jardineria_DM fueron modeladas bajo un esquema estrella, compuesto por cinco dimensiones principales (dim_cliente, dim_producto, dim_empleado, dim_oficina, dim_tiempo) y una tabla de hechos (fact_ventas). Este modelo no solo conserva las claves de negocio de la base de datos original, sino que además introduce claves sustitutas (surrogate keys) que facilitan la integración y consistencia de los datos en el entorno analítico.

Posteriormente, se desarrollaron consultas INSERT...SELECT para extraer la información desde la base de datos jardineria (origen) y transformarla en campos normalizados y enriquecidos dentro de cada tabla dimensional. Estas transformaciones incluyeron la concatenación de nombres completos de empleados, la unificación de direcciones de oficinas y el cálculo de métricas como la tardanza en la entrega de pedidos.

Finalmente, se procedió a poblar la tabla de hechos fact_ventas, que integra los identificadores de todas las dimensiones y concentra las métricas de negocio, como cantidades vendidas, precios unitarios y valores totales. La carga fue validada mediante consultas de verificación (SELECT COUNT(*), TOP 10) contrastando los resultados entre la base de datos de origen y el Data Mart, garantizando así la integridad y consistencia de la información.

Descripción del Análisis Base de Datos Staging

El diagrama inicial de la base de datos Jardineria_Staging con sus respectivas tablas y relaciones aplicados al modelo.

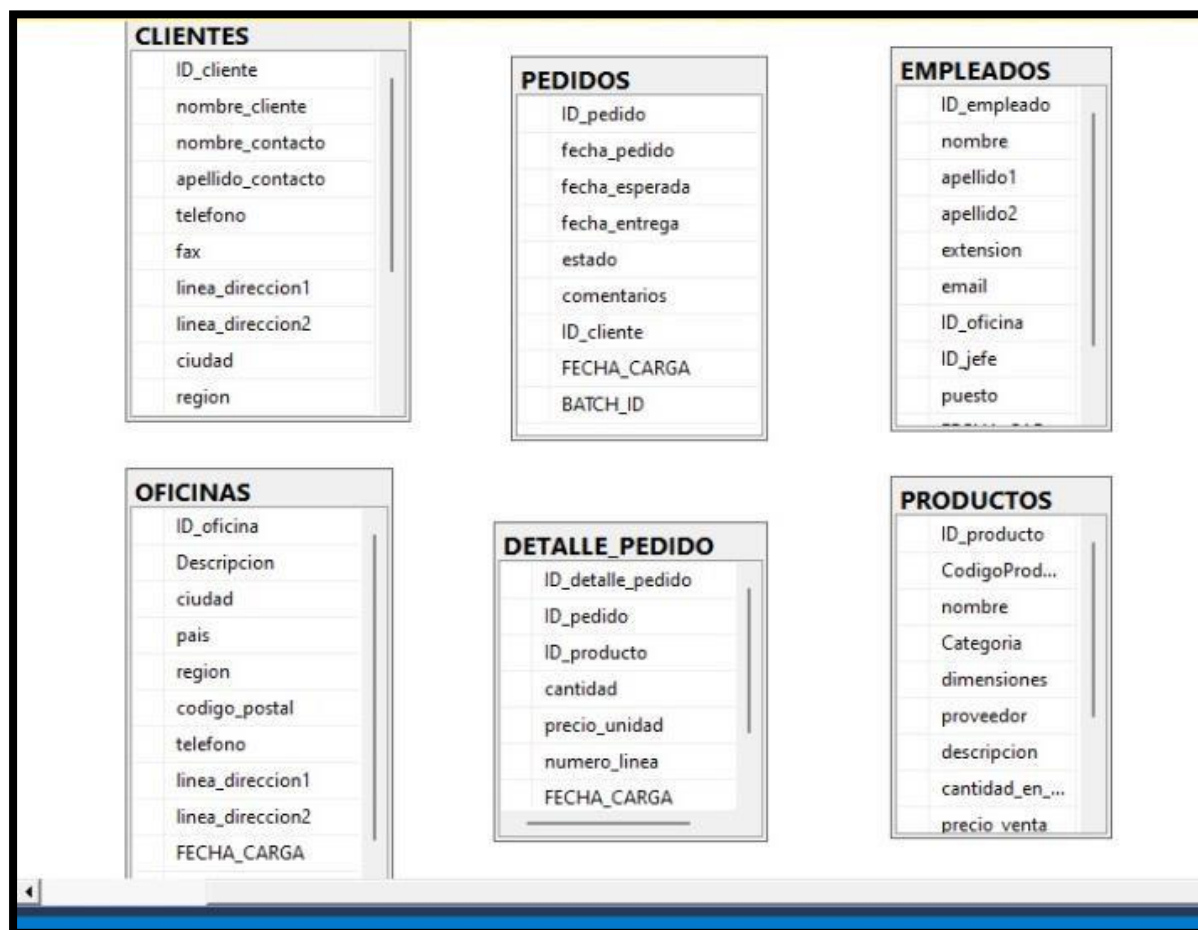


Imagen 1

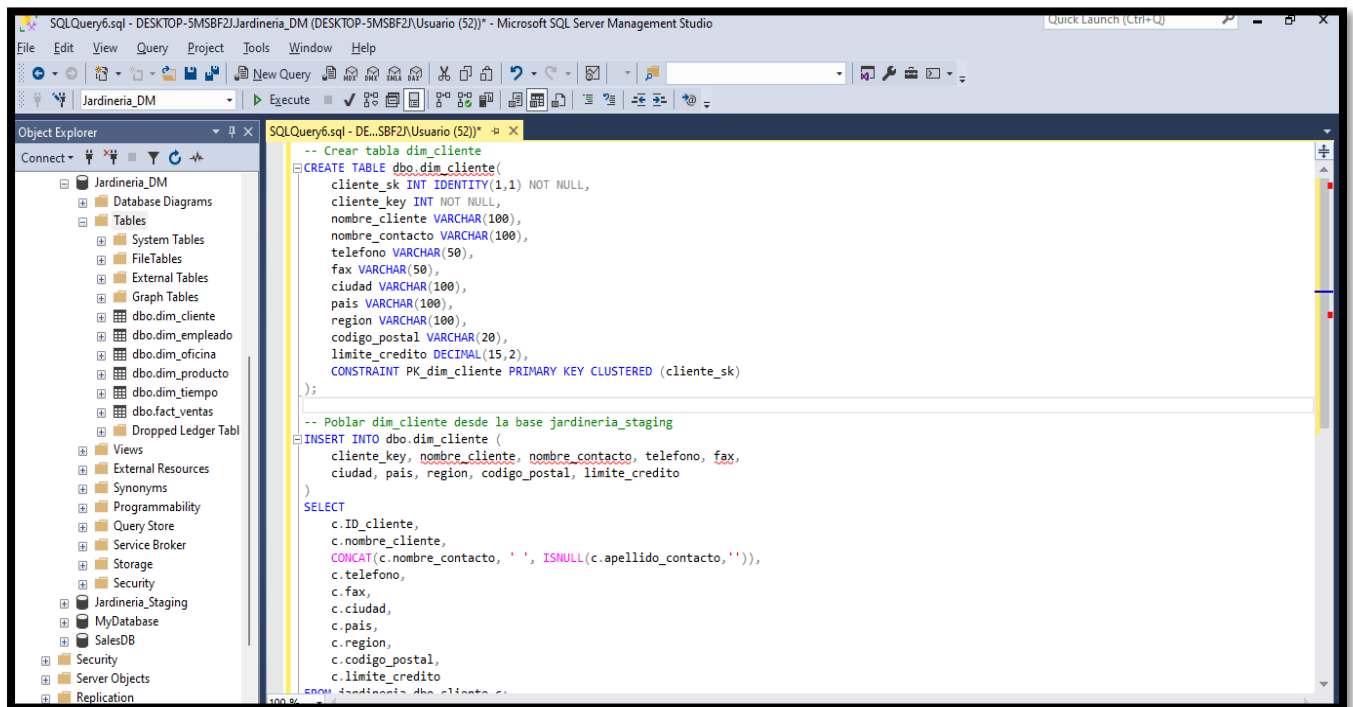
Diagrama modelo aplicando el Staging a la Base de datos Jardinería

6. Dimensiones:

Creación de la tabla y carga de datos en todas las tablas:

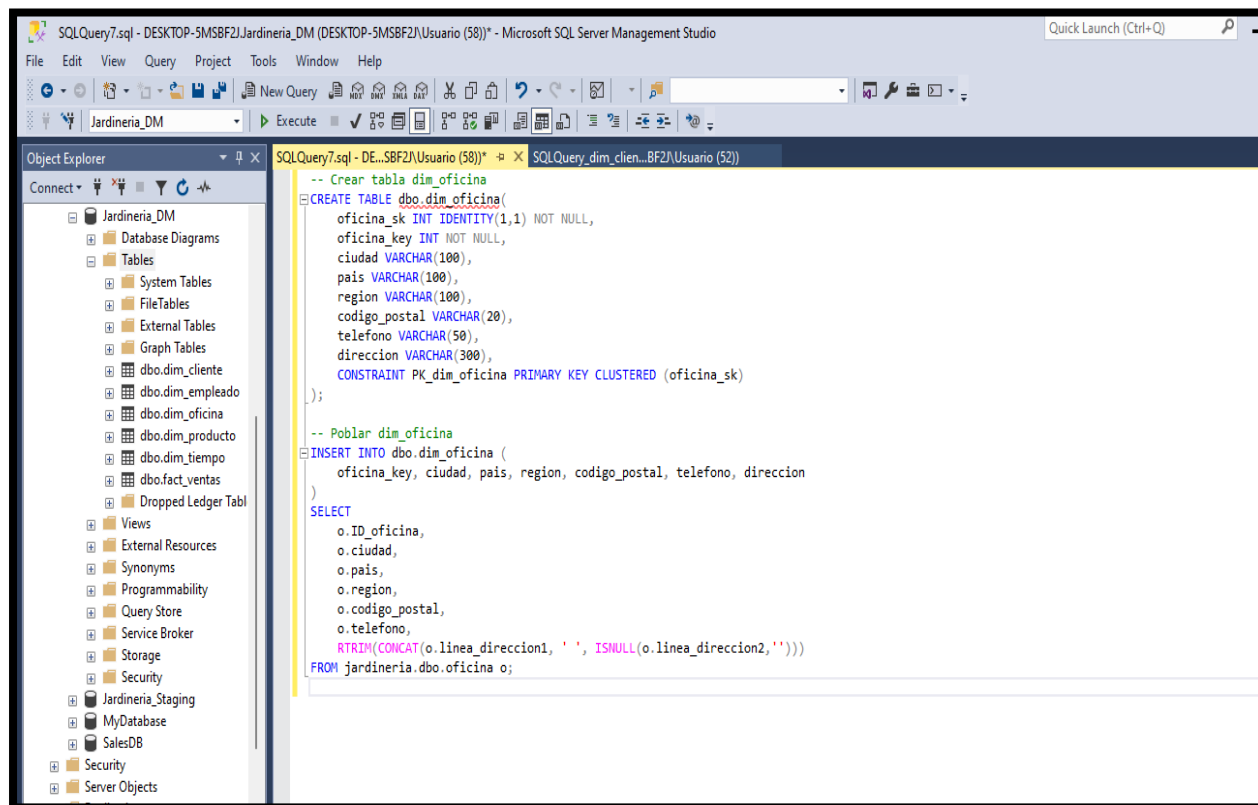
6.1 Dimensión Cliente

Imagen 2



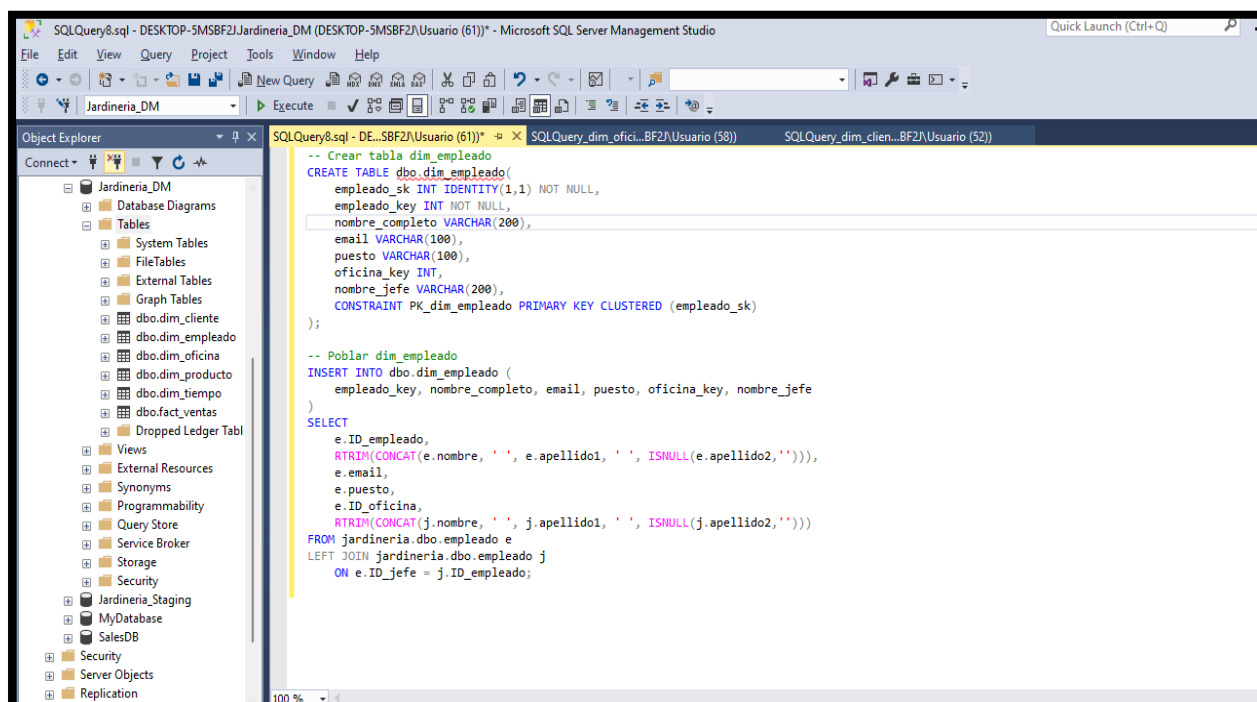
6.2. Dimensión Oficina

Imagen 3



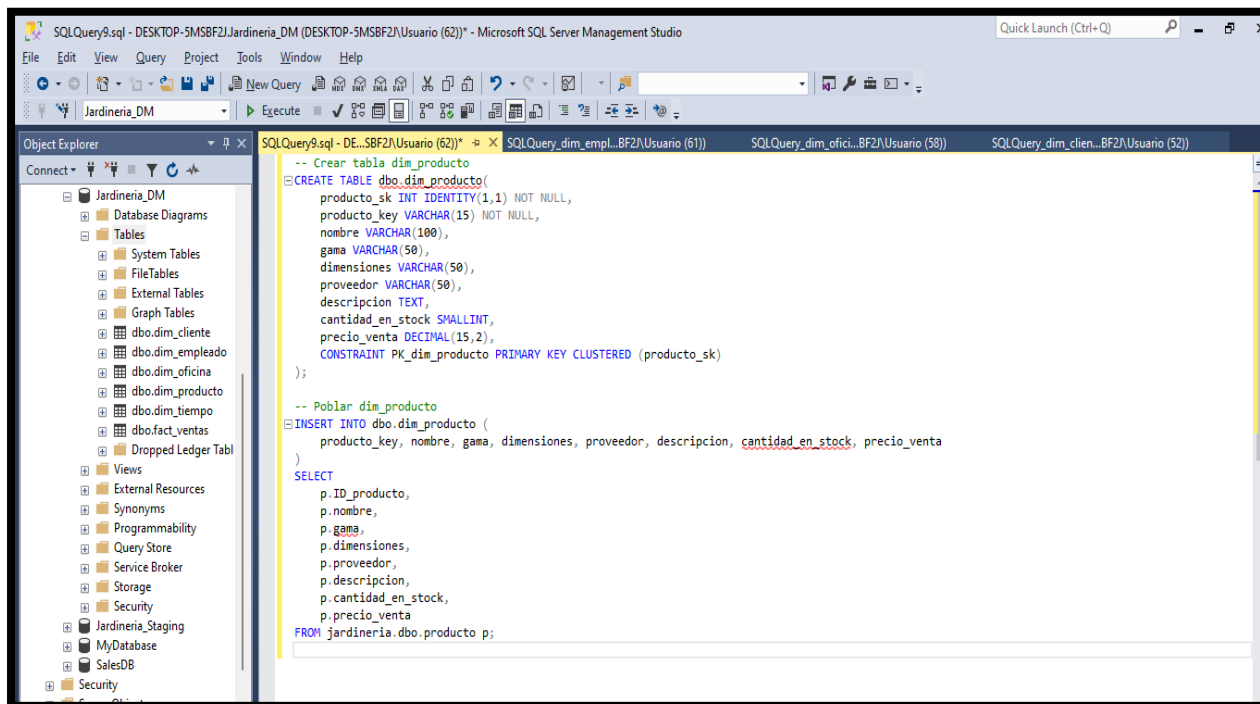
6.3. Dimensión Empleado

Imagen 4



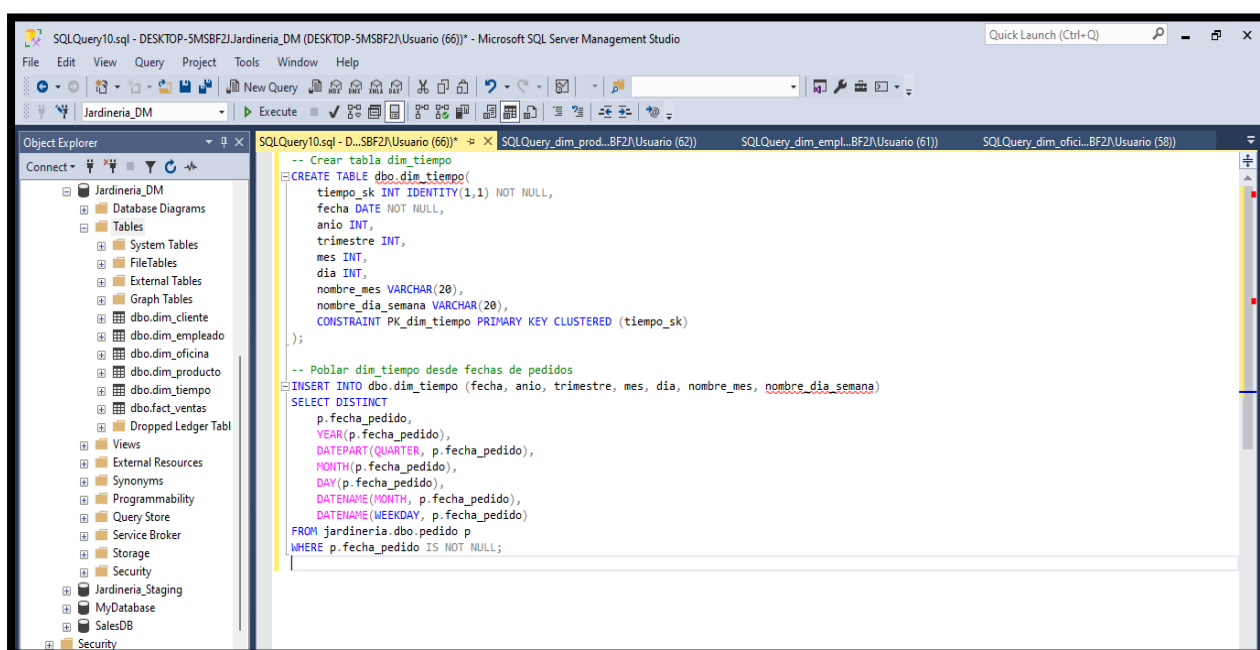
6.4. Dimensión Producto

Imagen 5



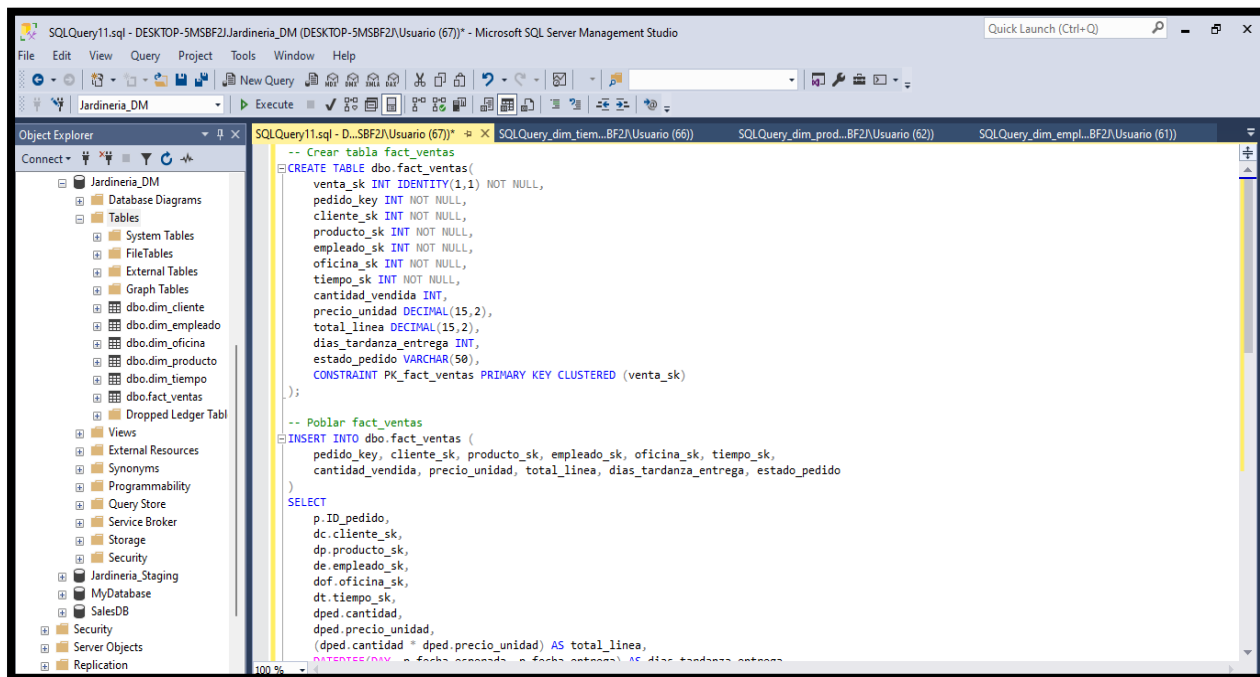
6.5. Dimensión Tiempo

Imagen 6



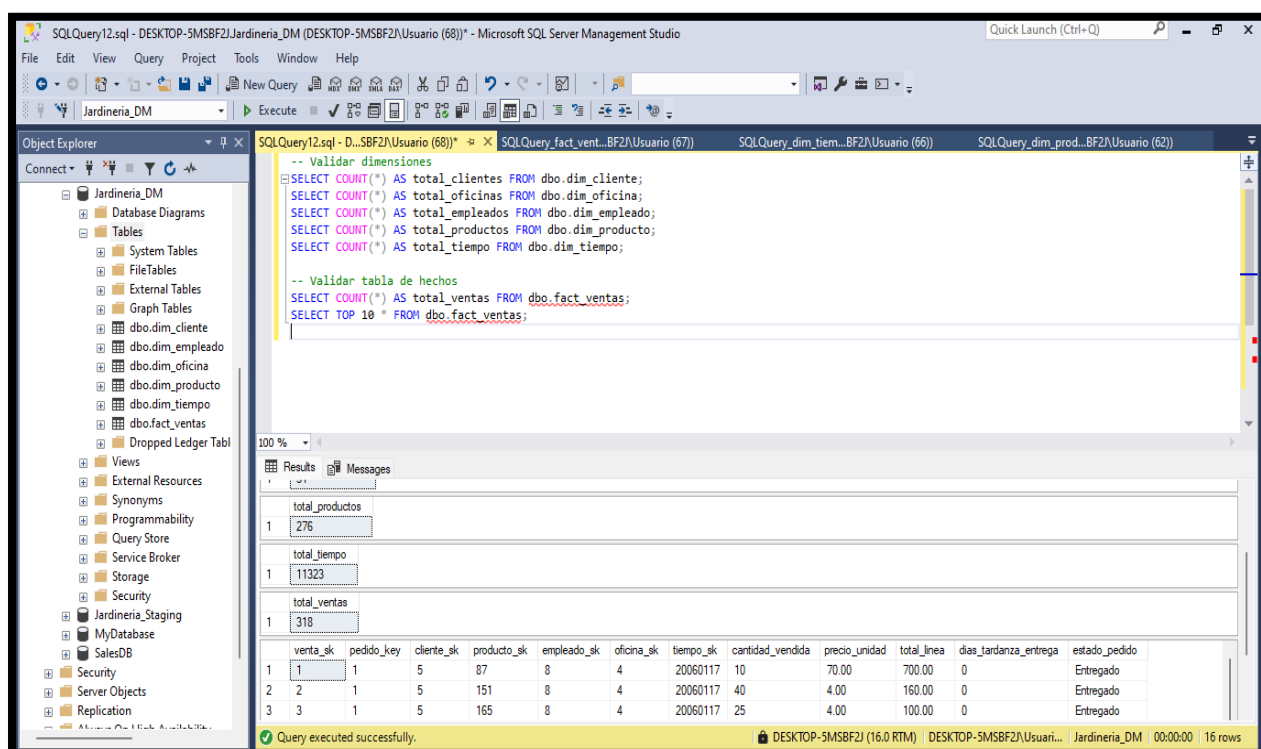
6.6. Tabla de Hechos: fact_ventas

Imagen 7



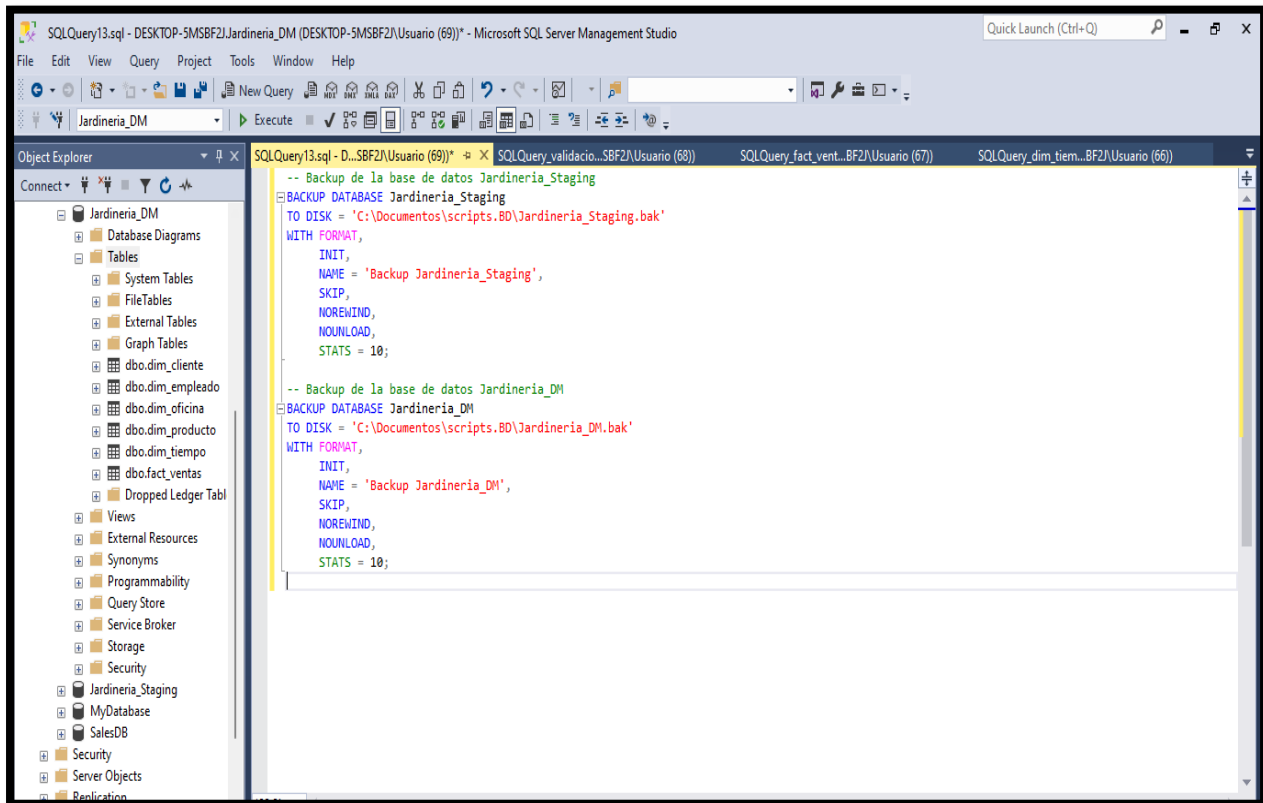
7. Validación de datos cargados

Imagen 8



8. Backups:

Imagen 9



9. Conclusiones:

Al desarrollar este proyecto, comprendimos con mayor profundidad cómo un modelo dimensional puede transformar la forma en que se organiza y analiza la información de un negocio como el de jardinería. Lo más valioso fue entender que el diseño del modelo estrella no es un mero ejercicio académico, sino una práctica esencial para que los datos fluyan de manera coherente y se conviertan en conocimiento útil para la toma de decisiones.

El proceso de extracción, transformación y carga (ETL) nos permitió experimentar en la práctica la importancia de la limpieza y normalización de datos, desde la concatenación de nombres hasta el cálculo de métricas como la tardanza en las entregas. Cada transformación demostró que los pequeños detalles en el diseño pueden marcar una gran diferencia en la calidad de los reportes finales.

La creación de la base de datos Jardineria_DM permitió centralizar la información en un entorno analítico preparado para la explotación en herramientas de Business Intelligence. La integridad de los datos fue garantizada gracias al uso de claves sustitutas, relaciones bien definidas y validaciones de carga con consultas de prueba.

Finalmente, este ejercicio no solo fortaleció nuestras habilidades técnicas en SQL Server y modelado dimensional, sino también nuestra capacidad de análisis crítico. No se trató únicamente de poblar tablas, sino de pensar en cómo los datos reflejan los procesos reales de una organización y cómo, al estructurarlos adecuadamente, se convierten en un recurso estratégico para la gestión y la competitividad empresarial.

Anexos

Script SQL de creación de la base de datos Jardinería_MD y su carga de datos de ejemplo.

Script SQL de creación de los INSERT...SELECT de Jardineria_Staging

Backup de ambas bases de datos (.bak).

Imagen del diagrama de la base staging.

Referencias bibliográficas

- Coronel, C., & Morris, S. (2018). Database Systems: Design, Implementation, & Management. Cengage Learning.
- Elmasri, R., & Navathe, S. (2016). Fundamentals of Database Systems. Pearson.
- Hernández, R., Fernández, C., & Baptista, P. (2014). Metodología de la investigación. McGraw Hill.
- Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. Wiley.
- Microsoft Docs. (2025). SQL Server Documentation. <https://learn.microsoft.com/sql/>
- APA. (2020). Publication Manual of the American Psychological Association (7th ed.). American Psychological Association.