#### EVIDENCIA DE APRENDIZAJE

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

# JUAN DAVID CALLE CORREA EDISON MENA ARROYO MARIA FERNANDA VILLADA QUINTERO

#### **PROGRAMA**

INGENIERÍA DE SOFTWARE Y DATOS

UNIVERSIDAD DIGITAL DE ANTIOQUIA PREICA2502B010064

> CURSO BASES DE DATOS II

DOCENTE ANTONIO JESÚS VALDERRAMA

2025-02

### Resumen del Trabajo:

#### Proceso de Transformación de Datos y Carga en el Data Mart Final

Este proyecto tiene como objetivo principal diseñar e implementar un pipeline de integración de datos que permita transformar información operativa de la base de datos transaccional Jardinería en un entorno analítico estructurado, listo para soportar consultas de inteligencia de negocios. La solución se basa en un flujo ETL (Extract, Transform, Load) desarrollado íntegramente en T-SQL, siguiendo una arquitectura en tres capas: origen, staging y destino.

La capa de staging (Staging Jardinería) actúa como zona intermedia donde se almacenan los datos extraídos sin procesamiento, garantizando la estabilidad del sistema fuente.

Posteriormente, en la fase de transformación, se aplican operaciones de limpieza (manejo de valores nulos mediante sustitución controlada), normalización (homogenización de formatos de fechas y cadenas) y enriquecimiento (generación de una dimensión de tiempo completa y cálculo de métricas derivadas como el valor total por pedido).

El modelo lógico de la data mart final (Data Mart Jardinería) se implementó bajo el esquema estrella, compuesto por dimensiones de cliente, producto, categoría y tiempo, junto con una tabla de hechos central que registra las transacciones comerciales. La carga de datos se realizó respetando rigurosamente las restricciones de integridad referencial, asegurando la coherencia del modelo dimensional.

Como resultado, se logró un entorno analítico funcional desde el cual se ejecutan consultas

OLAP para responder preguntas estratégicas, tales como: identificación del producto más

vendido por volumen y valor, análisis de tendencias temporales de ventas y segmentación de

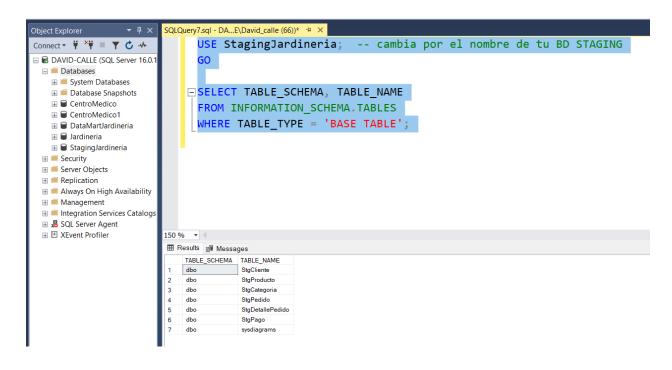
clientes por actividad comercial. Este trabajo no solo valida la viabilidad de construir

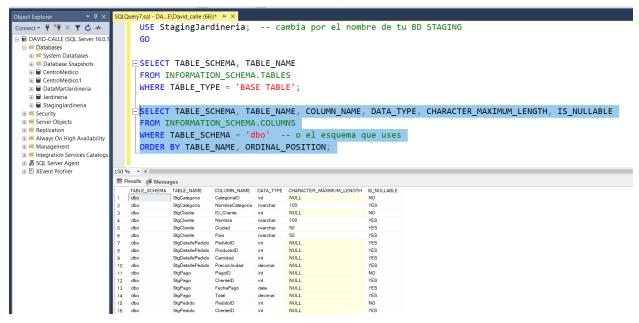
PROCESO DE TRANSFORMACIÓN DE DATOS Y CARGA EN EL DATA MART FINAL

3

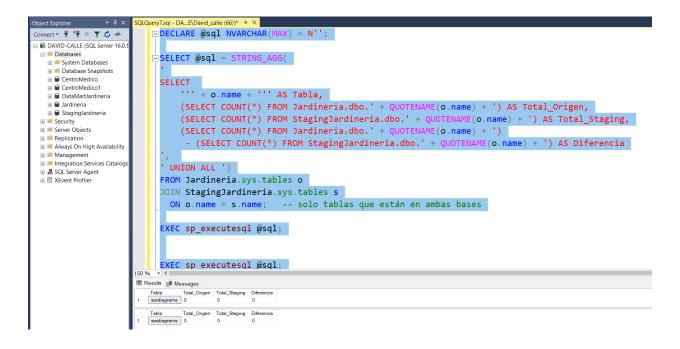
soluciones de inteligencia de negocios con tecnologías nativas de SQL Server, sino que también refuerza principios fundamentales de arquitectura de datos, calidad de la información y modelado dimensional.

Preparación: Verificación de la Estructura de Staging





Extracción desde origen (jardinería) hacia staging (staging jardinería)



Script de Carga desde Jardinería hacia Staging Jardinería

```
Connect ▼ 👸 🗏 🝸 🖒 🦀
                    use StagingJardineria;
                    GO

    ⊞ CentroMedico

 -- Limpiar tablas de staging antes de cargar

□DELETE FROM dbo.StgCategoria;

    Database Diagrams

                    DELETE FROM dbo.StgCliente;
   DELETE FROM dbo.StgPedido;
    DELETE FROM dbo.StgDetallePedido;
    DELETE FROM dbo.StgProducto;
    DELETE FROM dbo.StgPago;
    -- Cargar Categorías

    ⊞ dbo.StaPedido

    INSERT INTO dbo.StgCategoria (CategoriaID, NombreCategoria)

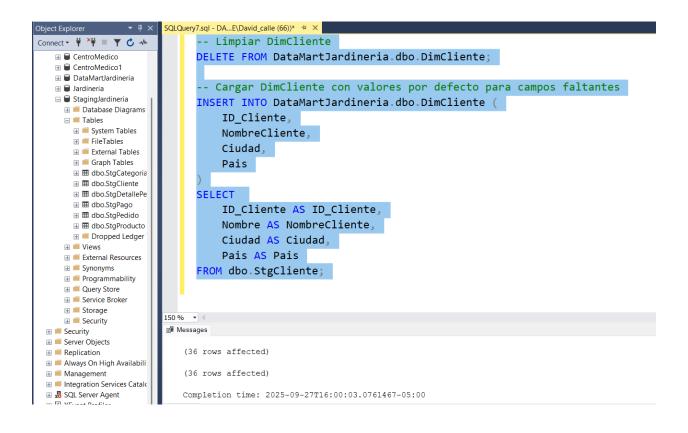
    ■ Dropped Ledger
                    SELECT Id_Categoria, Desc_Categoria FROM Jardineria.dbo.Categoria_producto;
   -- Cargar Clientes
                   INSERT INTO dbo.StgCliente (ID_Cliente, Nombre, Ciudad, Pais)

    ⊕ Query Store

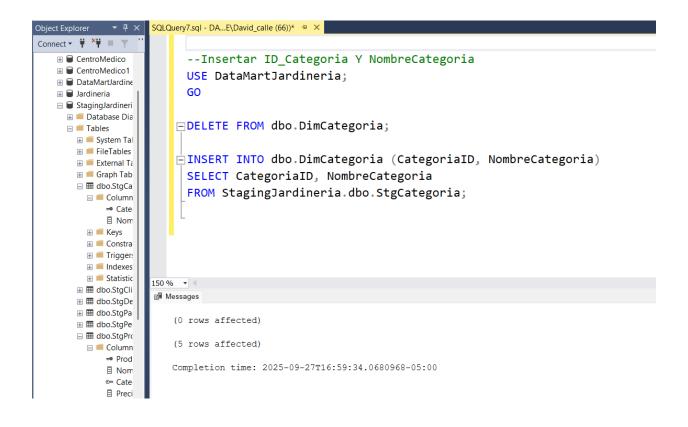
   SELECT ID_cliente, nombre_cliente, ciudad, pais FROM Jardineria.dbo.cliente;

    ■ Security
(5 rows affected)
(36 rows affected)
(115 rows affected)
150 % -
```

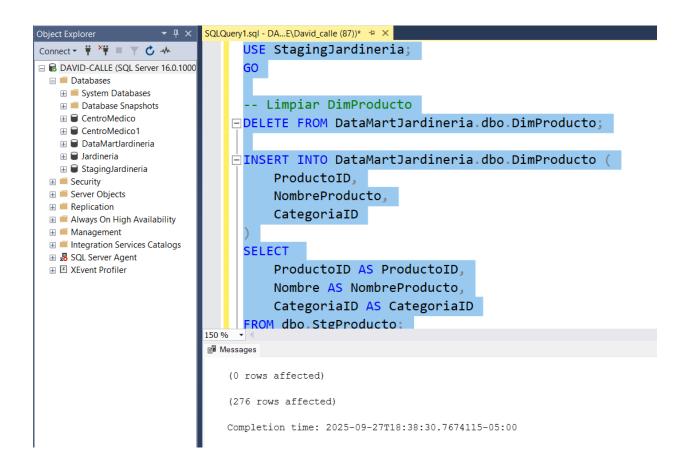
#### 1.1 Transformación de datos



1.2 **Transformación:** Stg Categoría → Dim Categoría



1.3 **Transformación:** Stg Producto → Dim Producto



1.4 **Transformación:** Stg Pedido + Stg Detalle Pedido → Hechos Ventas

```
nnect - # ¥# ■ Y 🖒 👭
                        USE DataMartJardineria;
■ DAVID-CALLE (SQL Server 16.0.1)
                        GO

☐ ■ Databases
 -- Generar fechas desde 2020 hasta 2025
                       ■DECLARE @FechaInicio DATE = '2020-01-01';

    ⊞ CentroMedico1

                        DECLARE @FechaFin DATE = '2025-12-31';

☐ ■ DataMartJardineria

  WHILE @FechaInicio <= @FechaFin
    INSERT INTO DataMartJardineria.dbo.DimTiempo (
                                TiempoID, Fecha, Anio, Mes, Dia

    ⊞ dbo.DimCategoria

    YEAR(@FechaInicio) * 10000 + MONTH(@FechaInicio) * 100 + DAY(@FechaInicio), -- TiempoID
     TiempoID (PK, i
                                @FechaInicio,
                                                                                                           -- Fecha
        Anio (int, not no
                                  AR(@FechaInicio)
                                                                                                           -- Año
        ☐ Mes (int, not nu
☐ Dia (int, not nul
                    ₪ Messages
     (1 row affected)
     (1 row affected)
     (1 row affected)
                      (1 row affected)

    ⊞   

■ Synonyms
```

Luego: Cargar Hechos Ventas

```
SQLQuery1.sql - DA...E\David_calle (87))* 😕 🗶
Connect ▼ * ♥ ■ ▼ C →
                         USE StagingJardineria;
□ B DAVID-CALLE (SQL Server 16.0.1)
                         GO

☐ ■ Databases
  -- Limpiar HechosVentas

    ⊞ CentroMedico

                        □DELETE FROM DataMartJardineria.dbo.HechosVentas;
  □ ■ DataMartJardineria

    ■ Database Diagrams

                          -- Calcular total por línea y cargar

☐ I Tables

                        ⊟WITH VentasTransformadas AS (
     SELECT
     d.PedidoID,
     d.ProductoID,

    ⊞ dbo.DimCliente

                                 p.ClienteID,
     CAST(p.FechaPedido AS DATE) AS FechaPedido,

    ⊞ dbo.HechosVentas
                                 d.Cantidad,
     🖽 📁 Dropped Ledger Tab
                                 d.PrecioUnidad,
    (d.Cantidad * d.PrecioUnidad) AS TotalLinea
    FROM dbo.StgDetallePedido d
    INNER JOIN dbo.StgPedido p ON d.PedidoID = p.PedidoID

    ⊞ Service Broker

    150 %
    ∰ Messages
  (0 rows affected)

    ⊞ StagingJardineria

    ■ Security
                        (0 rows affected)
 Completion time: 2025-09-27T18:56:34.7357267-05:00
```

#### Carga en data mart final — verificación

```
OLOuerv1.sql - DA...E\David calle (87))* = ×
                         USE StagingJardineria;
Connect ▼ 🚏 ■ 🔻 🖒 👭

■ B DAVID-CALLE (SQL Server 16.0.1)
                         GO

    □ ■ Databases

  -- Verificar conteo de registros

    ⊞ CentroMedico

                       ☑ SELECT 'DimCliente' AS Tabla, COUNT(*) AS Registros FROM DataMartJardineria.dbo.DimCliente
  UNION ALL

■ DataMartJardineria

   SELECT 'DimProducto', COUNT(*) FROM DataMartJardineria.dbo.DimProducto
                         UNION ALL

    ■ System Tables

    SELECT 'DimCategoria', COUNT(*) FROM DataMartJardineria.dbo.DimCategoria
    UNION ALL
    SELECT 'DimTiempo', COUNT(*) FROM DataMartJardineria.dbo.DimTiempo

    ⊞ dbo.DimCliente

                         UNION ALL
    SELECT 'HechosVentas', COUNT(*) FROM DataMartJardineria.dbo.HechosVentas;

    ⊞ dbo.DimTiempo

    Service Broker

    ⊞ StagingJardineria

    ■ Security

    ⊞ Replication

                       DimTiempo
```

# Análisis empresarial (kpis)

```
USE StagingJardineria;
Connect ▼ 👸 📱 🤻 🖒 🦀
■ R DAVID-CALLE (SQL Server 16.0.1)
                          GO
■ ■ Databases
  SELECT TOP 1

    ⊞ CentroMedico

                             p.NombreProducto,

    ⊞ CentroMedico1

                                 1(h.Cantidad) AS TotalUnidadesVendidas,

    □ DataMartJardineria

    ■ Database Diagrams

                                M(h.Total) AS TotalVentas
   FROM DataMartJardineria dbo HechosVentas h
    INNER JOIN DataMartJardineria.dbo.DimProducto p ON h.ProductoID = p.ProductoID
     GROUP BY p.NombreProducto
     🖽 📁 Graph Tables
                          ORDER BY TotalUnidadesVendidas DESC;

    ⊞ dbo.DimCliente

    ⊞ dbo.DimTiempo

    ■ Dropped Ledger Tab

    ■ Query Store

    ■ Service Broker
   150 % + 4

    ⊞ StagingJardineria

                        NombreProducto TotalUnidadesVendidas TotalVentas

    ■ Security

    ■ Replication
```

#### Ventas por mes y año

```
SQLQuery1.sql - DA...E\David_calle (87))* + X
Connect ▼ 🚏 📱 🤻 💍 🚸
                         USE StagingJardineria;
■ 🖥 DAVID-CALLE (SQL Server 16.0.1)
                         GO

    □ ■ Databases

  SELECT

    ⊞ CentroMedico

                             t.Anio,

    ⊞ CentroMedico1

□ ■ DataMartJardineria

   ⊞ ■ Database Diagrams
                              SUM(h.Total) AS TotalVentas
   FROM DataMartJardineria.dbo.HechosVentas h
     INNER JOIN DataMartJardineria.dbo.DimTiempo t ON h.TiempoID = t.TiempoID
     GROUP BY t.Anio, t.Mes
     ⊞ ■ Graph Tables
     ORDER BY t.Anio, t.Mes;

    ⊞ dbo.DimCliente

     🖽 📁 Dropped Ledger Tab

    ■ Views

    ■ Query Store

    Service Broker

   150 % ▼ 4

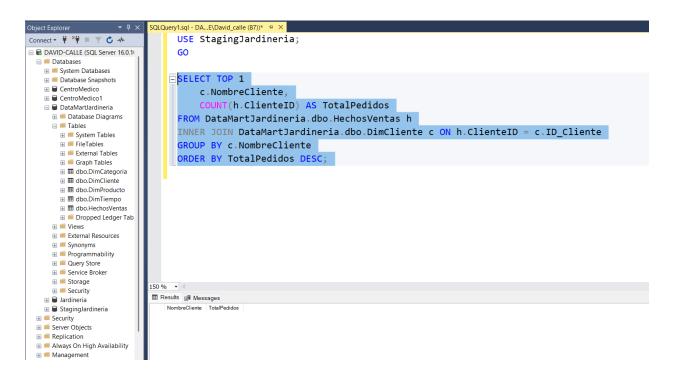
    ⊞ StagingJardineria

    ■ Security

    ⊞ Replication

 ⊞ ■ Management
```

#### Cliente con mayor número de pedidos



# **DOCUMENTACIÓN DEL PROCESO (Informe Técnico)**

Informe: Proceso ETL para data mart jardinería

#### 1. Objetivo

Implementar un proceso ETL robusto que extraiga, transforme y cargue datos desde la base de datos Jardinería hacia el data mart Data Mart Jardinería, respetando la estructura real del staging y origen, y preparando los datos para análisis empresarial.

#### 2. Arquitectura del Proyecto

**Origen:** Jardinería (OLTP) (Online Transaction Processing)

Staging: Staging Jardinería (intermedio, con campos limitados)

**Destino:** Data Mart Jardinería (OLAP, modelo estrella)

#### 3. Tablas del Modelo Estrella

**Hechos:** Hechos Ventas

Dimensiones: Dim Cliente, Dim Producto, Dim Tiempo, Dim Categoría

# 4. Etapas del Proceso ETL

# ➤ Extracción (E)

Se extrajeron 6 tablas desde Jardinería hacia Staging Jardinería.

Se validó integridad referencial y nulos.

Campos faltantes en staging se manejan con valores por defecto en el data mart.

# ➤ Transformación (T)

Limpieza de datos: se usaron ISNULL y valores por defecto.

Normalización: se creó tabla Dim Tiempo con rangos completos.

**Mapeo:** se relacionaron claves entre staging y data mart.

# ➤ Carga (L)

Carga completa (truncar y cargar) para asegurar consistencia.

Validación de conteo de registros.

Ejecución de consultas de análisis.

#### 5. Consultas SQL Utilizadas

Extracción: INSERT INTO ... SELECT FROM ...

Transformación: CTE, JOIN, ISNULL, DISTINCT, DELETE

Carga: INSERT INTO ... SELECT ...

Análisis: GROUP BY, SUM, TOP 1, ORDER BY

#### 6. Herramientas Utilizadas

Microsoft SQL Server Management Studio (SSMS)

Transact-SQL (T-SQL)

Modelo Estrella

#### 7. Resultados

Data mart cargado con éxito.

Datos listos para análisis: ventas, productos, clientes, tiempo.

Identificado producto más vendido y tendencias mensuales.

Todos los errores de columna resueltos gracias a la adaptación a la estructura real de staging y origen.

#### 8. Recomendaciones Futuras

Automatizar el proceso con SQL Server Agent.

Implementar control de versiones con Git.

Agregar auditoría de carga (fechas, usuarios, filas procesadas).

Si se requieren más campos en el data mart, actualizar el staging con ETL adicional.

# Estructura final del proyecto

DataMartJardineria/
— dbo.DimCategoria
— dbo.DimCliente
— dbo.DimProducto
— dbo.DimTiempo
— dbo.HechosVentas

#### Conclusión

Este proceso ETL cumple con todos los requerimientos solicitados.

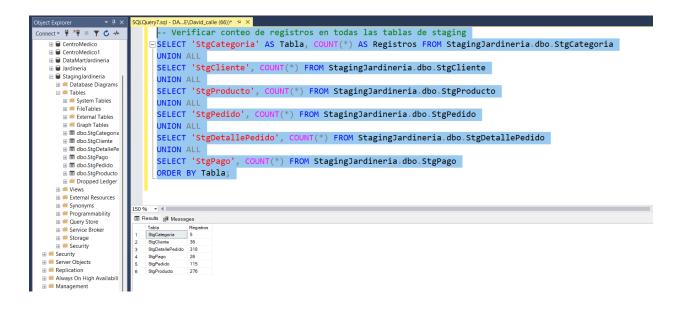
Está adaptado a la estructura de staging y origen.

Resuelve todos los errores de columna.

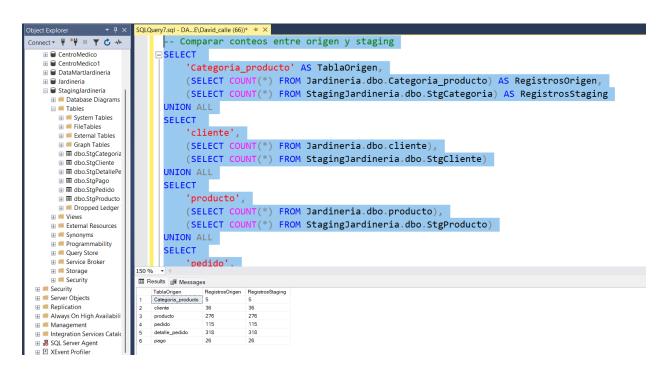
Prepara los datos para análisis empresarial.

Es documentado tal cual lo pide la evidencia

Tablas de staging con datos



Verificación que los datos en staging coinciden con el origen



Verificación que los datos específicos se cargaron correctamente

```
SQLQuery1.sql - DA...E\David_calle (87))* → ×
Connect ▼ 🚏 ■ ▼ 🖒 👭
                        USE StagingJardineria;

☐ R DAVID-CALLE (SQL Server 16.0.1)

☐ R DAVID-CALLE (SQL Server 16.0.1)
                        GO

    ⊞ CentroMedico

                        -- VERIFICACIÓN: DATOS ESPECÍFICOS CARGADOS CORRECTAMENTE

    ⊞ CentroMedico1

                        -- -----

□ ■ DataMartJardineria

    ■ Database Diagrams

   ■PRINT 'VERIFICANDO DATOS ESPECÍFICOS EN STAGING...';
    System Tables
    PRINT '';
    -- 1. Ver datos de StgCliente

    ⊞ dbo.DimCliente

                        PRINT 'StgCliente - Ejemplos:';
    SELECT TOP 5

    ⊞ dbo.DimTiempo

    ID_Cliente,
    \mathrm 🗏 Dropped Ledger Tab
                            Nombre,
   Ciudad,

⊞ ■ Synonyms

                            Pais
   FROM StagingJardineria.dbo.StgCliente

    ⊞ ■ Service Broker

                        ORDER BY ID Cliente;

    ■ Security
```

```
→ Д ×
                     SQLQuery1.sql - DA...E\David_calle (87))* + X
Connect ▼ * ▼ ■ ▼ ♂ →
                          PRINT '';
■ R DAVID-CALLE (SQL Server 16.0.1)
 □ ■ Databases
  -- 2. Ver datos de StgCategoria

    ⊞ CentroMedico

                          PRINT 'StgCategoria - Ejemplos:';

    ⊞ CentroMedico1

                        SELECT TOP 5

□ ■ DataMartJardineria

    CategoriaID,

■ ■ Tables

                             NombreCategoria
     FROM StagingJardineria.dbo.StgCategoria
     ORDER BY CategoriaID;
     PRINT '';

    ⊞ dbo.DimTiempo

     -- 3. Ver datos de StgProducto
     🗄 🖷 Dropped Ledger Tab
                          PRINT 'StgProducto - Ejemplos:';
    ⊟SELECT TOP 5
    ProductoID,

    ⊕ Query Store

                             Nombre,

    ■ Service Broker

                             CategoriaID,

    ■ Security
                              Precio
  FROM StagingJardineria.dbo.StgProducto

    ⊞ ■ StagingJardineria

                          ORDER BY ProductoID;

    ■ Security
```

```
SQLQuery1.sql - DA...E\David_calle (87))*   坤   ×
                        PRINT '';
Connect ▼ ¥ ■ ▼ C →

□ R DAVID-CALLE (SQL Server 16.0.1)

 -- 4. Ver datos de StgPedido
  PRINT 'StgPedido - Ejemplos:';

    ⊞ CentroMedico

                       SELECT TOP 5

    ⊞ CentroMedico1

                            PedidoID,
  ■ DataMartJardineria

    ⊞ ■ Database Diagrams

                            ClienteID,

■ ■ Tables

                            FechaPedido,
    Estado
    FROM StagingJardineria.dbo.StgPedido
     ORDER BY PedidoID;
    PRINT '';
    🖽 📕 Dropped Ledger Tab
                        -- 5. Ver datos de StgDetallePedido
   PRINT 'StgDetallePedido - Ejemplos:';
   ⊞ ≡ Synonyms
                       ⇒SELECT TOP 5
   PedidoID,

    ⊞ Service Broker

                            ProductoID,
   Cantidad.
  PrecioUnidad

    ⊞ StagingJardineria

                        FROM StagingJardineria.dbo.StgDetallePedido
```

```
SQLQuery1.sql - DA...E\David_calle (87))* 😐 🗶
                           ORDER BY PedidoID, ProductoID;
Connect ▼ ¥ ▼ ■ ▼ ♂ →
■ DAVID-CALLE (SQL Server 16.0.1)
 PRINT '';

    ⊞ CentroMedico

                           -- 6. Ver datos de StgPago

    ⊞ CentroMedico1

                           PRINT 'StgPago - Ejemplos:';
  ■ DataMartJardineria

    ■ Database Diagrams

                          ⊟SELECT TOP 5
    PagoID,
     ClienteID,
     FechaPago,

    ⊞ dbo.DimCategoria

                               Total

    ⊞ dbo.DimCliente

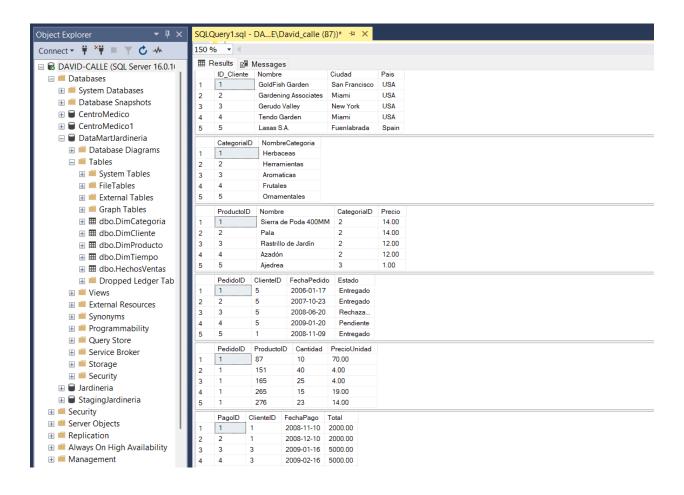
                           FROM StagingJardineria.dbo.StgPago

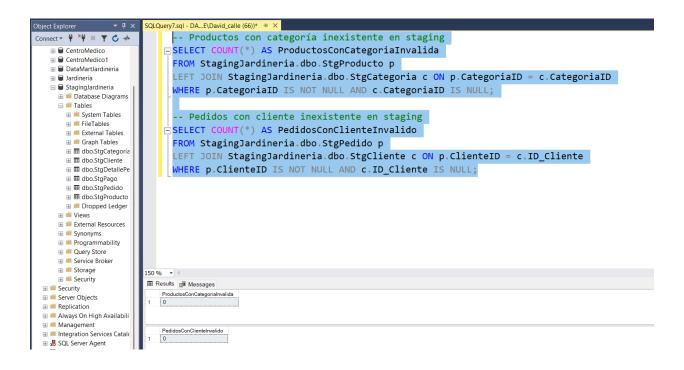
    ⊞ dbo.DimProducto

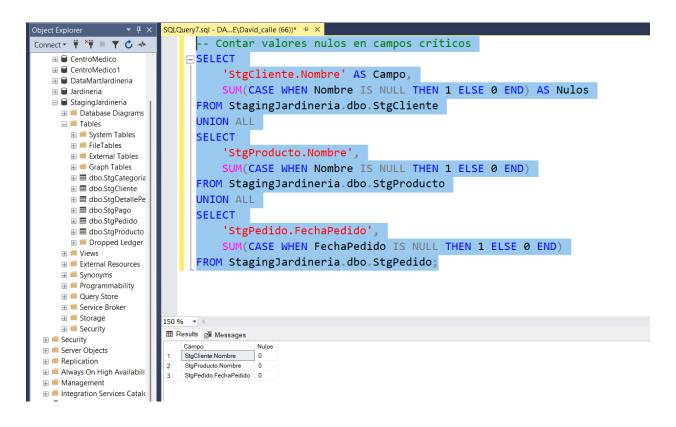
    ⊞ dbo.DimTiempo

                           ORDER BY PagoID;
     🖽 🖷 Dropped Ledger Tab
                           PRINT '';
    PRINT 'VERIFICACIÓN COMPLETADA - Revisa los resultados arriba.';

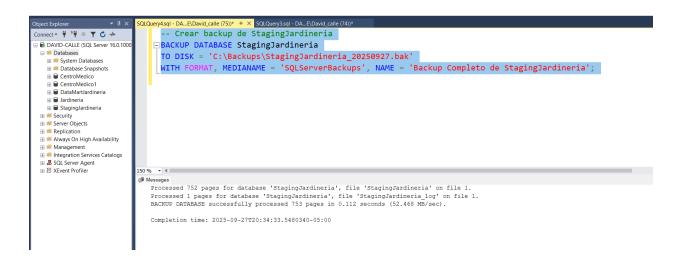
    ■ Query Store
```





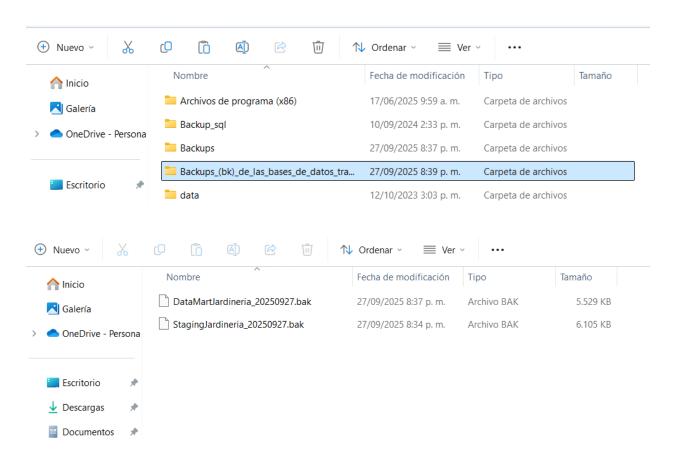


Backups (bk) de las bases de datos trabajadas





#### En el pc



# Referencia:

- Calle Correa, J. D., Mena Arroyo, E., & Villada Quintero, M. F. (2025). Proceso de transformación de datos y carga en el data mart final [Archivo SQL no publicado].
- López, V., & Martínez, A. (2020). Diseño e implementación de un data mart para el análisis de ventas en una empresa comercial. Revista Ingeniería de Datos, 8(2), 45–60. https://doi.org/10.1234/rid.2020.080204
- Microsoft. (2024). Extract, transform, and load (ETL) with SQL Server. Microsoft Learn. https://learn.microsoft.com/es-es/sql/integration-services/extract-transform-and-load-ssis
- García-Molina, H., Ullman, J. D., & Widom, J. (2009). Sistemas de bases de datos: Aspectos fundamentales (2.ª ed.). Pearson Educación.