

EVIDENCIA DE APRENDIZAJE

MODELO ESTRELLA DE UN DATA MART

JUAN DAVID CALLE CORREA

EDINSON 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

Introducción

En las organizaciones modernas, la información se ha convertido en un recurso estratégico. Las bases de datos operacionales almacenan los registros del día a día, pero cuando se requiere analizar la información de forma histórica y responder preguntas clave del negocio, es necesario contar con un Data Mart.

Un Data Mart permite organizar los datos bajo un esquema diseñado específicamente para el análisis, siendo el modelo estrella el más utilizado por su sencillez y eficiencia.

En este trabajo se construye un modelo estrella a partir de la base de datos Jardinería, con el fin de responder a tres preguntas de negocio:

- 1 - ¿Cuál es el producto más vendido?**
- 2 - ¿Cuál es la categoría con mayor cantidad de productos?**
- 3 - ¿En qué año se registraron más ventas?**

Objetivos

Desarrollar un modelo estrella que permita organizar y analizar la información de ventas de la base de datos Jardinería, de manera que la empresa pueda responder de forma ágil y confiable a preguntas estratégicas relacionadas con productos, categorías y periodos de mayor impacto comercial.

Objetivo general

Construir un modelo estrella que permita analizar las ventas de la base de datos Jardinería y responder las principales preguntas de negocio relacionadas con productos, categorías y años de mayor impacto.

Objetivos específicos

Diseñar una tabla de hechos que integre la información de ventas.

Crear tablas de dimensiones para organizar los datos en torno a productos, clientes, empleados, oficinas y tiempo.

Proveer un modelo que facilite consultas de análisis de ventas de forma eficiente.

Planteamiento del problema

La base de datos Jardinería contiene información detallada sobre clientes, empleados, oficinas, pedidos, productos, pagos y categorías. Sin embargo, está normalizada para la operación transaccional y no facilita directamente el análisis estratégico.

El problema principal es que, para responder a preguntas como “**¿qué producto se vendió más en 2008?**” o “**qué gama de productos concentra más ventas?**”, se requieren consultas complejas con múltiples uniones entre tablas.

La solución es reorganizar los datos en un esquema estrella que concentre las ventas en una tabla de hechos y agrupe la información complementaria en dimensiones.

Análisis del problema

La tabla detalle pedido es la base ideal para la tabla de hechos, ya que contiene las ventas realizadas, los productos involucrados, cantidades y precios.

Las tablas producto, gama producto, cliente, empleado y oficina son candidatas a convertirse en dimensiones.

La dimensión tiempo debe construirse a partir de la fecha del pedido y de los pagos.

De esta forma, se organiza la información en torno al proceso de ventas, facilitando el análisis histórico y comparativo.

Propuesta del modelo estrella

Tabla de Hechos: hecho ventas

Contendrá la información cuantitativa de las ventas.

Descripción del Modelo Estrella Propuesto

El modelo estrella diseñado para el Data Mart de la empresa Jardinería se centra en el análisis de las ventas como proceso principal. En el centro del esquema se encuentra la tabla de hechos “Hechos Ventas”, la cual registra las transacciones realizadas por la empresa. Esta tabla almacena los valores cuantitativos que serán objeto de análisis, como la cantidad de productos vendidos, el precio unitario y el total de cada venta.

Alrededor de la tabla de hechos se encuentran las tablas de dimensiones, que aportan el contexto descriptivo necesario para interpretar los datos de las ventas:

DimProducto: describe las características de cada producto vendido, incluyendo su nombre, precio y la categoría a la que pertenece.

DimCategoría: organiza los productos dentro de categorías más amplias, facilitando el análisis por líneas de negocio o familias de productos.

DimCliente: permite conocer a los compradores, considerando información como nombre, ciudad y país, lo que apoya el análisis de mercado y segmentación.

DimTiempo: proporciona una perspectiva temporal de las ventas, permitiendo desglosar la información por días, meses, trimestres y años.

La relación entre la tabla de hechos y las dimensiones se da mediante llaves primarias y foráneas, lo que garantiza la integridad referencial. De esta manera, el modelo facilita la construcción de consultas analíticas para responder preguntas estratégicas, tales como:

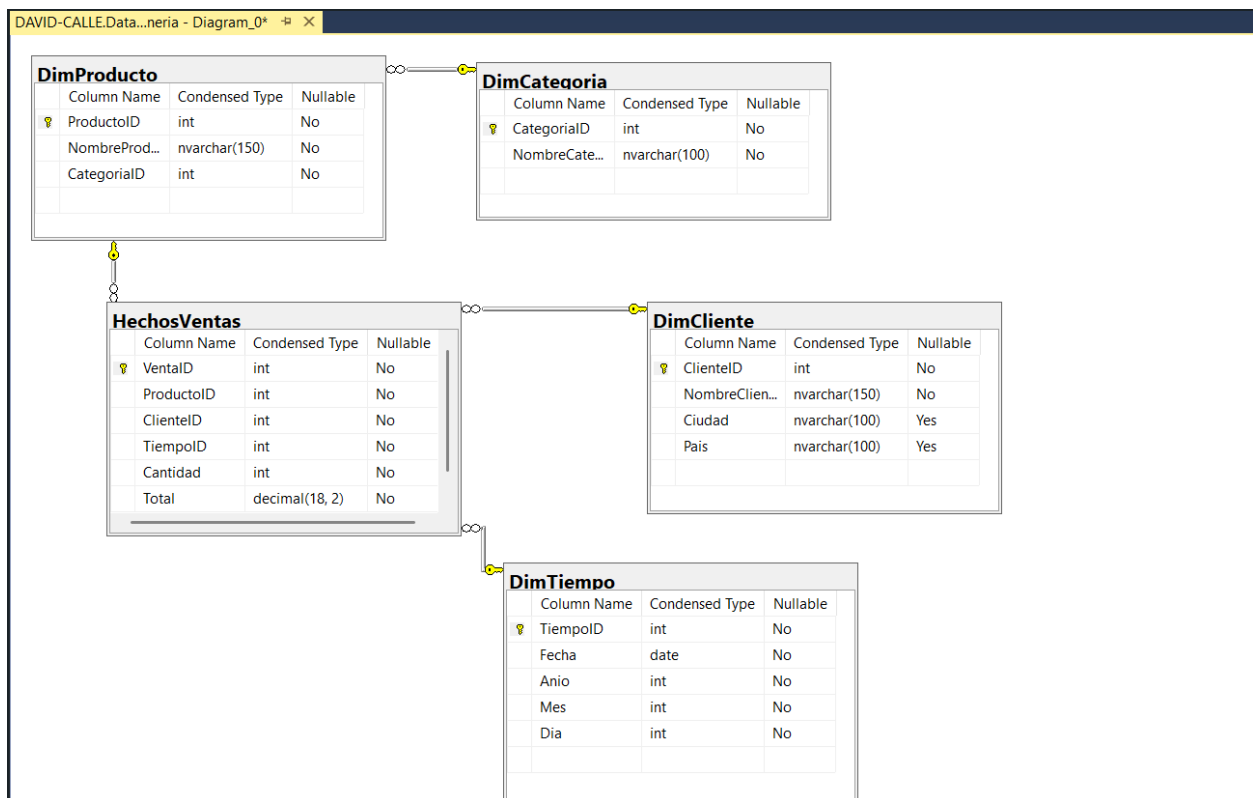
✓ ¿Cuál fue el producto más vendido en un periodo determinado?

✓ ¿Qué categoría generó más ingresos?

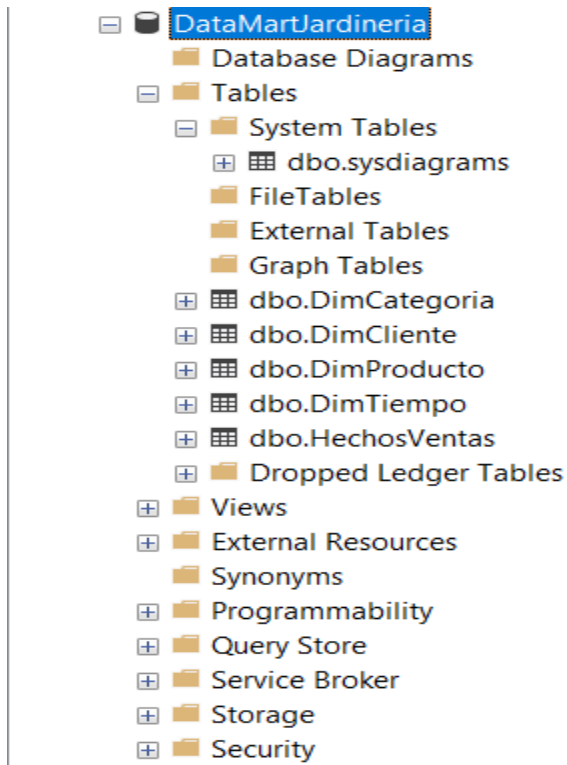
✓ ¿En qué año se alcanzó el mayor volumen de ventas?

Este modelo estrella proporciona una estructura clara, flexible y optimizada para el análisis de información, permitiendo a la organización tomar decisiones fundamentadas en datos.

Diseño (Imagen) del modelo estrella donde se puedan observar las dimensiones, la tabla de hechos, sus campos, tipos de datos y relaciones.



Bases de Datos Creadas



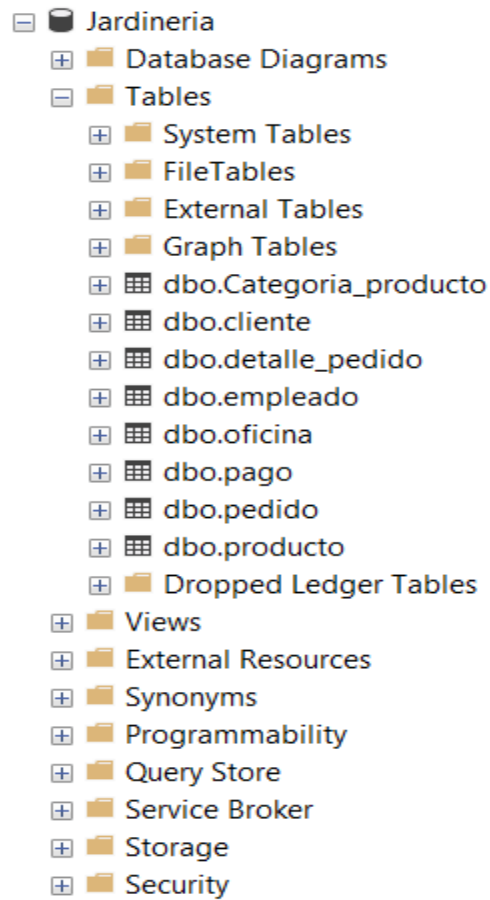


Tabla de hechos principal

Campos:

- id_detalle (PK)
- id_pedido (FK → dimensión tiempo)
- id_producto (FK → dimensión producto)
- id_cliente (FK → dimensión cliente)
- id_empleado (FK → dimensión empleada)
- id_oficina (FK → dimensión oficina)
- cantidad
- precio_venta
- importe_total (cantidad × precio_venta)

Lista de dimensiones propuestas

1. Dimensión Producto (dim_producto)

Esta dimensión almacena información de cada producto que maneja la empresa, incluyendo su nombre, dimensiones físicas, proveedor y la categoría a la que pertenece.

Columna, Tipo de dato sugerido, Función

id_producto, INT, **clave primaria, Identifica de forma única cada producto**

nombre, VARCHAR (100), **Nombre del producto**

gama, INT, **clave foránea a dim_gama, Categoría o gama del producto**

dimensiones, VARCHAR (50), **Medidas físicas o características del producto**

proveedor, VARCHAR (100), **Nombre del proveedor**

2. Dimensión Gama o Categoría (dim_gama)

Contiene información sobre las categorías de los productos, lo que permite agruparlos para análisis de ventas por tipo o línea.

Columna, Tipo de dato sugerido, Función

id_gama, INT, **clave primaria, Identificador único de la categoría**

descripción, VARCHAR (100), **Nombre o descripción de la categoría**

3. Dimensión Cliente (dim_cliente)

Guarda datos de los clientes, necesarios para segmentar ventas por ubicación geográfica, teléfono y nombre del comprador.

Columna, Tipo de dato sugerido, Función

id_cliente, INT, **clave primaria, Identificador único del cliente**

nombre_cliente, VARCHAR (100), **Nombre completo del cliente**

teléfono, VARCHAR (20), **Número de contacto**

ciudad, VARCHAR (50), **Ciudad de residencia**

país, VARCHAR (50), **País del cliente**

4. Dimensión Empleado (dim_empleado)

Se centra en la información de los empleados que participan en las ventas o gestión de pedidos.

Columna, Tipo de dato sugerido, Función

id_empleado, INT, **clave primaria, Identificador único del empleado**

nombre, VARCHAR (50), **Primer nombre del empleado**

apellido, VARCHAR (50), **Apellido del empleado**

puesto, VARCHAR (50), **Cargo que ocupa en la empresa**

5. Dimensión Oficina (dim_oficina)

Almacena los datos de las oficinas o sucursales donde se realizan operaciones.

Columna, Tipo de dato sugerido, Función

id_oficina, INT, **clave primaria, Identificador único de la oficina**

ciudad, VARCHAR (50), **Ciudad donde se encuentra la oficina**

país, VARCHAR (50), **País de la oficina**

región, VARCHAR (50), **Región geográfica o zona de la oficina**

6. Dimensión Tiempo (dim_tiempo)

Esta dimensión permite analizar ventas y pagos en distintos periodos, desglosando por año, mes y trimestre.

Columna, Tipo de dato sugerido, Función

id_tiempo, INT, **clave primaria, Identificador único del registro temporal**

fecha, DATE, **Fecha exacta del pedido o pago**

año, INT, **Año de la transacción**

mes, INT, **Mes de la transacción**

trimestre, INT, **Trimestre correspondiente**

Diseño de las Relaciones

La tabla de hechos **Hechos_Ventas** se conecta con cada dimensión a través de llaves foráneas:

HechosVentas.IdProducto → DimProducto.IdProducto

HechosVentas.IdCategoria → DimCategoria.IdCategoria

HechosVentas.IdCliente → DimCliente.IdCliente

HechosVentas.IdTiempo → DimTiempo.IdTiempo

Estas relaciones garantizan la integridad referencial y permiten realizar consultas analíticas para responder preguntas clave, como:

✓ **¿Cuál fue el producto más vendido?**

✓ ¿Qué categoría tuvo más ventas?

✓ ¿En qué año se alcanzaron mayores ingresos?

¿Cuál es el producto más vendido?

SQLQuery1.sql - DA...E\David_calle (73))*

```
--1. ¿Cuál es el producto más vendido?
SELECT TOP 1
    dp.NombreProducto AS Producto,
    SUM(hv.Cantidad) AS Total_Vendido
FROM dbo.HechosVentas hv
INNER JOIN dbo.DimProducto dp ON hv.ProductoID = dp.ProductoID
GROUP BY dp.NombreProducto
ORDER BY SUM(hv.Cantidad) DESC;
```

150 %

Results Messages

Producto	Total_Vendido
----------	---------------

¿Cuál es la categoría con mayor cantidad de productos?

SQLQuery1.sql - DA...E\David_calle (73))*

```
--¿Cuál es la categoría con mayor cantidad de productos?
SELECT TOP 1
    dc.NombreCategoria AS Categoria,
    COUNT(dp.ProductoID) AS Cantidad_Productos
FROM dbo.DimProducto dp
INNER JOIN dbo.DimCategoria dc ON dp.CategoriaID = dc.CategoriaID
GROUP BY dc.NombreCategoria
ORDER BY COUNT(dp.ProductoID) DESC;
```

150 %

Results Messages

Categoria	Cantidad_Productos
-----------	--------------------

¿En qué año se registraron más ventas?

SQLQuery1.sql - DA...E\David_calle (73))*

```
--¿En qué año se registraron más ventas?
SELECT TOP 1
    dt.Anio,
    SUM(hv.Total) AS Total_Ventas
FROM dbo.HechosVentas hv
INNER JOIN dbo.DimTiempo dt ON hv.TiempoID = dt.TiempoID
GROUP BY dt.Anio
ORDER BY SUM(hv.Total) DESC;
```

150 %

Results Messages

Anio	Total_Ventas
------	--------------

¿Cuál es el cliente que más ha comprado?

SQLQuery1.sql - DA...E\David_calle (73))*

```
--¿Cuál es el cliente que más ha comprado?
SELECT TOP 1
    dc.NombreCliente AS Cliente,
    SUM(hv.Total) AS Total_Compras
FROM dbo.HechosVentas hv
INNER JOIN dbo.DimCliente dc ON hv.ClienteID = dc.ClienteID
GROUP BY dc.NombreCliente
ORDER BY SUM(hv.Total) DESC;
```

150 %

Results Messages

Cliente	Total_Compras

¿Cuál es el país con más ventas?

SQLQuery1.sql - DA...E\David_calle (73))*

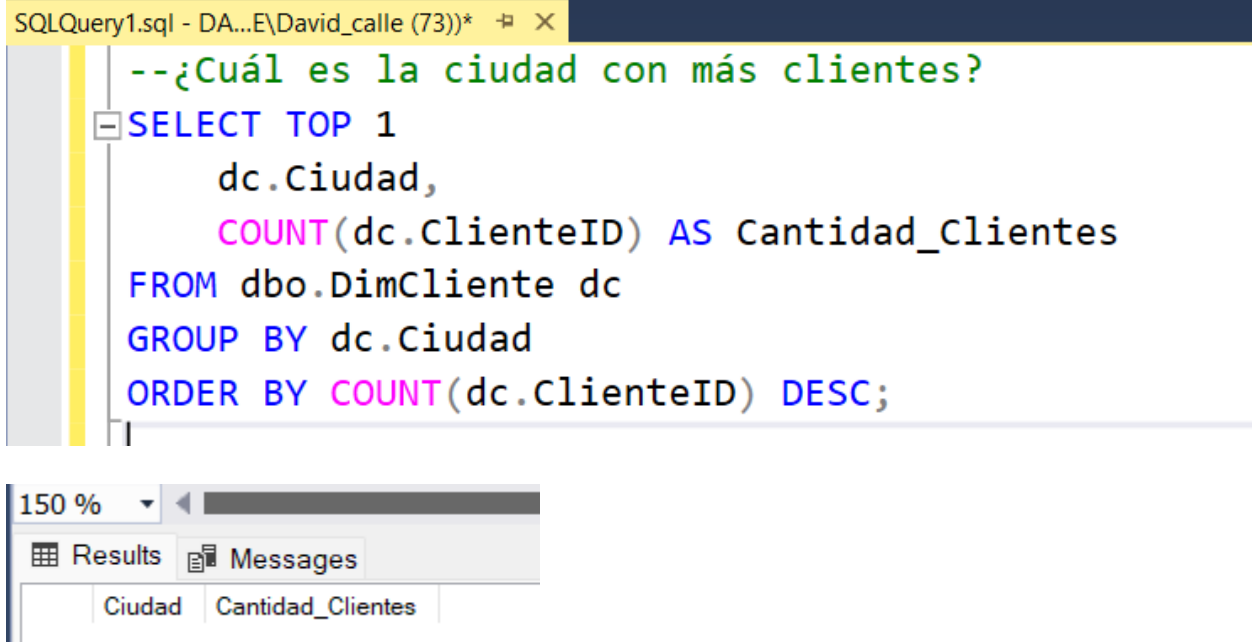
```
--¿Cuál es el país con más ventas?
SELECT TOP 1
    dc.Pais,
    SUM(hv.Total) AS Total_Ventas
FROM dbo.HechosVentas hv
INNER JOIN dbo.DimCliente dc ON hv.ClienteID = dc.ClienteID
GROUP BY dc.Pais
ORDER BY SUM(hv.Total) DESC;
```

150 %

Results Messages

Pais	Total_Ventas

¿Cuál es la ciudad con más clientes?



The screenshot shows a SQL query window titled "SQLQuery1.sql - DA...E\David_calle (73))". The query is written in green and blue text. Below the query, the results are displayed in a table with two columns: "Ciudad" and "Cantidad_Clientes". The results are shown at 150% zoom. The "Results" tab is active, and the "Messages" tab is also visible.

```
--¿Cuál es la ciudad con más clientes?  
SELECT TOP 1  
    dc.Ciudad,  
    COUNT(dc.ClienteID) AS Cantidad_Clientes  
FROM dbo.DimCliente dc  
GROUP BY dc.Ciudad  
ORDER BY COUNT(dc.ClienteID) DESC;
```

Ciudad	Cantidad_Clientes
--------	-------------------

Conclusiones

Organización clara de los datos:

El modelo estrella permite centralizar la información de ventas en una tabla de hechos y complementarla con dimensiones que aportan contexto, haciendo que el análisis sea mucho más intuitivo y estructurado.

Análisis eficiente y rápido:

Con esta estructura, la empresa puede obtener respuestas precisas a preguntas clave, como qué productos se venden más, cuáles categorías generan mayores ingresos y en qué periodos se registran más ventas, sin depender de consultas complejas sobre múltiples tablas normales.

Seguridad en la integridad de los datos:

La relación mediante claves primarias y foráneas asegura que los registros de ventas estén correctamente vinculados con sus dimensiones, evitando inconsistencias y garantizando que los resultados de los análisis sean confiables.

Flexibilidad para futuras necesidades:

El modelo diseñado permite agregar nuevas dimensiones o medidas según cambien los requerimientos de la empresa, lo que facilita la escalabilidad del Data Mart sin modificar la estructura existente.

Valor estratégico para la toma de decisiones:

Al tener los datos organizados de manera que reflejen tanto lo cuantitativo como lo descriptivo, la empresa puede basar sus decisiones en información histórica confiable, optimizando la planificación de ventas, inventarios y estrategias comerciales.

Simplicidad y claridad visual:

El esquema estrella facilita la comprensión de los datos para cualquier usuario del negocio, permitiendo que incluso quienes no son expertos en bases de datos puedan interpretar la información y generar reportes útiles.

Bibliografía

- Coronado, M. A. (2020). Bases de datos relacionales: Diseño y modelado con ejemplos prácticos. Ciudad de México: Alfaomega.
- García, L., & Torres, P. (2019). Modelos de datos y almacenamiento en sistemas de información. Bogotá: Editorial Universidad Nacional.
- Hernández, J. (2021). Inteligencia de negocios: Fundamentos y aplicaciones prácticas. Madrid: Ediciones Pirámide.
- López, R., & Martínez, C. (2020). Data Warehouse y Minería de Datos: Teoría y práctica aplicada. Barcelona: Marcombo.
- Romero, O., & Abelló, A. (2018). Almacenes de datos y modelado multidimensional. Barcelona: Universitat Politècnica de Catalunya.
- Sánchez, F. (2019). Sistemas de información empresarial y toma de decisiones. Lima: Fondo Editorial PUCP.
- Silberschatz, A., Korth, H., & Sudarshan, S. (2019). Fundamentos de bases de datos (7.^a ed.). México: McGraw-Hill.