

Modelo De Estrella Data Mart.

Luisa Fernanda Gómez Quiroz

Simón Arbey Castaño Ríos

Juan Pablo Gonzalez Gil

Docente: Antonio Jesús Valderrama Jaramillo



Institución Universitaria Digital de Antioquia.

Facultad de Ingenierías.

Ingeniería en Software y Datos.

Medellín, Colombia.

Bases de Datos II.

29 de agosto de 2025

Tabla de contenido

Introducción	3
1 Objetivos	4
1.1 Objetivo general	4
1.2 Objetivos específicos.....	4
2 Planteamiento del problema	5
3 Análisis del problema.....	6
3.1 Identificación de tablas relevantes	6
3.2 Relaciones principales.....	6
4 Propuesta solución del modelo estrella	7
4.1 TABLA DE HECHOS – Fact_ventas	7
5 Diseño del modelo estrella	8
6 Consultas realizadas	9
7 Conclusiones	10
8 Anexos.....	10
9 Bibliografía.....	10

Introducción

En el mundo de la inteligencia de negocios, las empresas se enfrentan a un reto clave: convertir la gran cantidad de datos que generan a diario en información clara y útil para tomar mejores decisiones. Para lograrlo, es esencial integrar, limpiar y organizar esos datos en estructuras que faciliten su análisis.

Una forma muy común de hacerlo es a través de data marts, que son como pequeños almacenes de datos enfocados en un área específica del negocio.

Dentro de estas soluciones, uno de los modelos más populares es el modelo en estrella (star schema). Este se compone de una tabla de hechos en el centro, donde se guardan las métricas más importantes del negocio —por ejemplo, precios, cantidades vendidas o fechas de transacción— y varias tablas de dimensiones alrededor, que aportan información descriptiva y de contexto. Su nombre viene de que, al verlo gráficamente, la estructura se parece a una estrella. Este diseño no solo hace que las consultas sean más rápidas y sencillas, sino que también facilita el análisis de la información.

En este proyecto, aplicaremos el modelo en estrella a la base de datos Jardinería, que contiene información sobre clientes, pedidos, productos y categorías.

1 Objetivos

1.1 Objetivo general

El propósito de este proyecto es **diseñar y construir un modelo en estrella** (*star schema*) para un *data mart* basado en la base de datos **Jardinería**. Con él, podremos analizar las ventas y responder preguntas clave para el negocio, como:

- ¿Cuál es el producto que más se vende?
- ¿Qué categoría reúne la mayor cantidad de productos?
- ¿En qué año se registró el mayor volumen de ventas?

Entonces, buscamos transformar los datos en respuestas claras que ayuden a tomar decisiones estratégicas.

1.2 Objetivos específicos

Para alcanzar este objetivo general, seguiremos una serie de pasos:

1. **Explorar la base de datos** transaccional de Jardinería para identificar las tablas y campos más relevantes.
2. **Seleccionar la información clave** que formará parte de la tabla de hechos y de las tablas de dimensiones.
3. **Diseñar la tabla de hechos**, donde se registrarán todas las ventas o transacciones de la empresa.
4. **Definir las dimensiones** que complementarán el análisis, añadiendo atributos descriptivos que faciliten la interpretación de los datos.
5. **Establecer las relaciones** entre la tabla de hechos y las dimensiones, asegurando la integridad de la información y optimizando el rendimiento de las consultas.
6. **Presentar el modelo final** mediante un diagrama y una descripción detallada de cada tabla, incluyendo sus campos y tipos de datos.

2 Planteamiento del problema

La base de datos operativa **Jardinería** guarda información detallada sobre productos, categorías, clientes, empleados, oficinas, pedidos y todos sus detalles asociados. Su diseño normalizado funciona muy bien para las operaciones diarias, ya que asegura la integridad de las transacciones. Sin embargo, no es la mejor opción cuando se trata de analizar la información.

Incluso consultas que parecen sencillas —como saber cuál es el producto más vendido, qué categoría reúne más productos o en qué año se registraron más ventas— requieren combinar varias tablas (por ejemplo, `detalle_pedido_id`, `pedido_id` y `producto_sk`) y realizar cálculos complejos. Esto no solo aumenta el tiempo de respuesta, sino que también impacta el rendimiento y complica la generación de reportes para la gerencia.

Para superar esta limitación, la empresa necesita un modelo que convierta los datos operativos en información analítica de fácil acceso. La solución es diseñar un **data mart** con un **esquema en estrella** (*star schema*), donde una **tabla de hechos** central concentre las métricas de ventas y un conjunto de **tablas de dimensiones** aporte el contexto descriptivo.

Este enfoque permitirá realizar consultas más simples y rápidas, facilitar el cálculo de agregados y el análisis en diferentes niveles de detalle, y ofrecer una base sólida para responder preguntas clave del negocio sin comprometer el rendimiento de la base de datos transaccional.

3 Análisis del problema

3.1 Identificación de tablas relevantes

- **-DimProducto:** contiene información de los artículos disponibles para la venta (*producto_id*, *nombre_producto*, *codigo_producto*, *precio_venta*, *stock*, etc.).
- **-Dim_categoría:** agrupa los productos en categorías (*Categoria_id*, *descripcion*, *imagen*, etc.).
- **-Fact_Ventas:** almacena los pedidos realizados por los clientes (*pedido_id*, *tiempo*, *estado*, *cliente_sk*, etc.).
- **-detalle_pedido_id:** registra el detalle de cada producto incluido en un pedido (*detalle_pedido_id*, *-producto_id*, *cantidad_vendida*, *precio_unitario*).
- **-Dimcliente:** almacena información de los clientes y su representante de ventas (*cliente_id*, *-cliente_sk*, *nombre_cliente*, *país*, *ciudad*, etc.).

3.2 Relaciones principales

- *-detalle_pedido_id* → *producto* (por *producto_id*): permite identificar qué productos forman parte de cada pedido.
- *-detalle_pedido_id* → *pedido* (por *pedido_id*): asocia cada línea de detalle con el pedido correspondiente.
- *-producto_sk* → *categoría_sk* (por *categoria_sk*): organiza los productos en sus respectivas categorías.
- *-pedido_id* → *cliente* (por *cliente_sk*): vincula cada pedido con el cliente que lo realizó.

4 Propuesta solución del modelo estrella

4.1 TABLA DE HECHOS – Fact_ventas

Representa cada línea de pedido registrada en la empresa.

- **Campos propuestos:**

- venta_sk (PK, surrogate key)
- pedido_id (referencia operativa)
- detalle_pedido_id (referencia operativa)
- producto_sk (FK a DimProducto)
- categoria_sk (FK a DimCategoria)
- tiempo_sk (FK a DimTiempo)
- cliente_sk (FK a DimCliente)
- cantidad_vendida (INT)
- precio_unitario (DECIMAL)
- total_linea (DECIMAL)

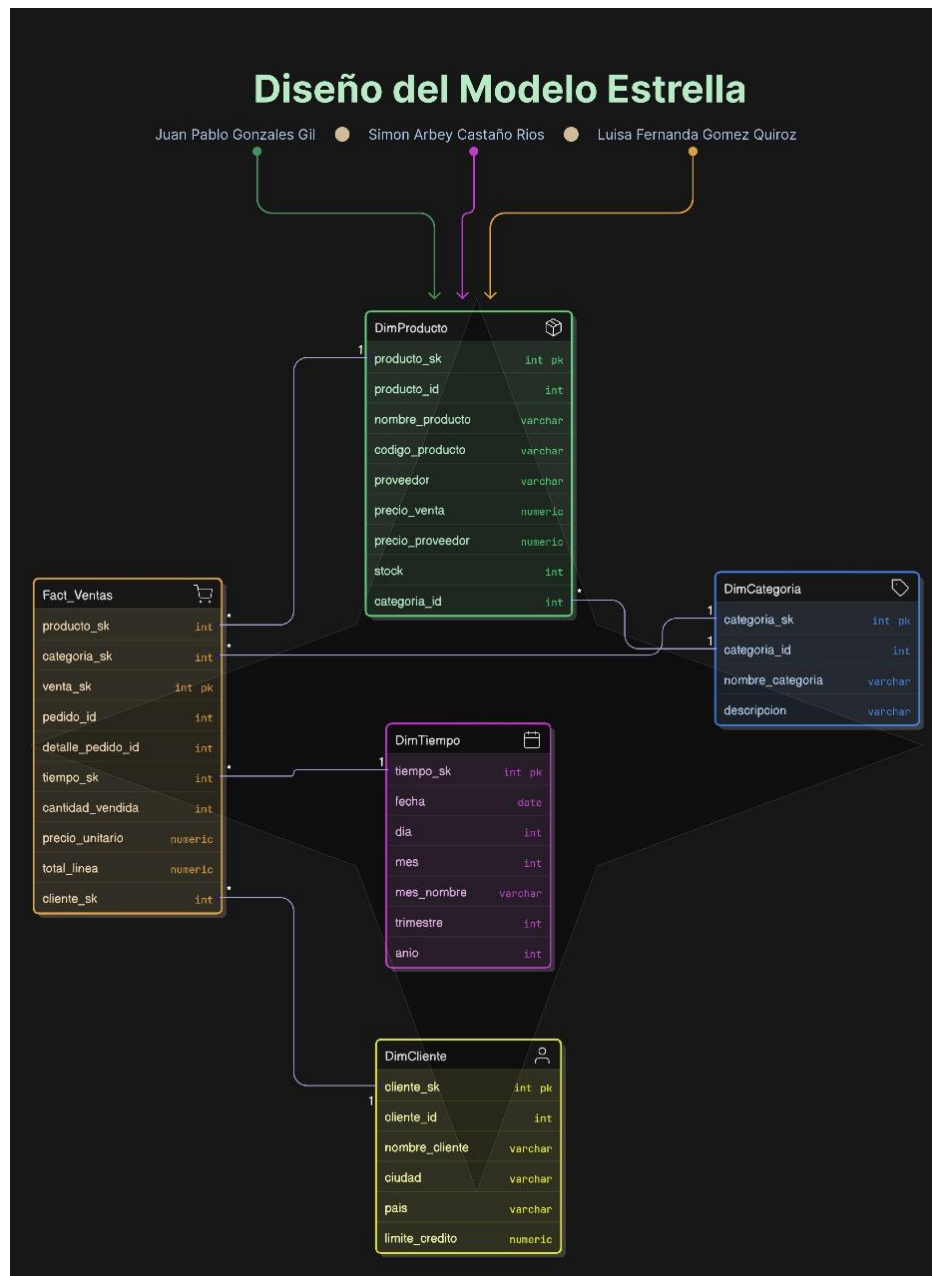
- **Dimensiones:**

- DimProducto : - producto_sk (PK) - producto_id (operacional) - nombre_producto - codigo_producto - proveedor - precio_venta - precio_proveedor - stock - categoria_id (FK a DimCategoria)
- DimCategoria: - categoria_sk (PK) - categoria_id (operacional) - nombre_categoria - descripcion
- DimTiempo: - tiempo_sk (PK) - fecha - dia - mes - mes_nombre - trimestre - anio
- DimCliente: - cliente_sk (PK) - cliente_id (operacional) - nombre_cliente - ciudad - pais - limite_credito

- **Relaciones:**

- Fact_Ventas.producto_sk → DimProducto.producto_sk
- DimProducto.categoria_id → DimCategoria.categoria_id
- Fact_Ventas.tiempo_sk → DimTiempo.tiempo_sk
- Fact_Ventas.cliente_sk → DimCliente.cliente_sk

5 Diseño del modelo estrella



6 Consultas realizadas

- **Producto más Vendido**

```
--Producto más vendido
SELECT TOP 1
    p.nombre,
    SUM(dp.cantidad) AS total_vendido
FROM detalle_pedido dp
INNER JOIN producto p ON dp.id_producto = p.id_producto
GROUP BY p.nombre
ORDER BY total_vendido DESC;
```

- **Categoría con más productos**

```
--Categoría con más productos
SELECT TOP 1
    c.Desc_Categoria AS nombre_categoria,
    COUNT(p.ID_producto) AS cantidad_productos
FROM producto p
INNER JOIN Categoria_producto c
    ON p.Categoria = c.Id_Categoria
GROUP BY c.Desc_Categoria
ORDER BY cantidad_productos DESC;
```

- **Año con más venta**

```
--Año con más ventas
SELECT TOP 1
    YEAR(p.fecha_pedido) AS anio,
    SUM(dp.cantidad * dp.precio_unidad) AS total_ventas
FROM pedido p
INNER JOIN detalle_pedido dp
    ON p.ID_pedido = dp.ID_pedido
GROUP BY YEAR(p.fecha_pedido)
ORDER BY total_ventas DESC;
```

7 Conclusiones

- -El modelo estrella diseñado facilita el análisis de ventas.
- -La tabla de hechos representa de manera precisa las transacciones.
- -Las dimensiones propuestas permiten obtener información agregada de productos, categorías, tiempo y clientes.
- -La integridad referencial asegura calidad en los datos.
- -El modelo responde de forma directa a los tres objetivos planteados

8 Anexos

- **-Script de creación de la base de datos:** El archivo CreacionJardineria.sql contiene la definición completa de las tablas y los datos de la base de datos original proporcionada para este ejercicio.
- **-Imagen del modelo:** se adjunta el diagrama del modelo estrella en formato PNG como parte de este documento.
- **-Implementación:** Para implementar el data mart se puede crear la estructura de tablas en un sistema gestor de bases de datos (por ejemplo, MySQL o SQL Server), extraer y transformar los datos desde la base de datos Jardinería hacia las dimensiones y la tabla de hechos, y finalmente cargar los registros. Esta tarea se puede automatizar mediante procesos de ETL.

9 Bibliografía

- Observatorio Business School. (2024). *Cómo utilizar el modelo estrella en una base de datos*. Recuperado el 29 de agosto de 2025, de <https://www.obsbusiness.school/blog/como-utilizar-el-modelo-estrella-en-una-base-de-datos>
- Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. Wiley.
- Script fuente: Base de datos Jardinería (archivo SQL)