

MODELO ESTRELLA DE UN DATA MART

Felipe Fernández Rodríguez

Docente:

Antonio Jesús Valderrama Jaramillo

Curso:

Bases de Datos II

Programa Ingeniería en Software y Datos

Institución Universitaria Digital de Antioquia

2025

INTRODUCCIÓN

El presente documento tiene como objetivo describir el diseño e implementación de un modelo estrella (Star Schema) para un data mart de ventas basado en la base de datos transaccional Jardinería. Este modelo permite analizar de manera eficiente las operaciones comerciales de la empresa, enfocándose en identificar los productos más vendidos, las categorías con mayor número de ventas y los períodos de mayor actividad comercial por año.

El modelo estrella se caracteriza por separar los datos en una tabla de hechos, que contiene las métricas cuantitativas de las ventas, y varias dimensiones, que proporcionan contexto a los hechos a través de atributos descriptivos como información del producto, cliente, categoría, empleado y tiempo. Esta estructura facilita el análisis y la generación de reportes, optimizando el rendimiento de las consultas analíticas frente a la base de datos transaccional original.

La implementación de este data mart busca proporcionar a los analistas y tomadores de decisiones de la empresa una herramienta confiable y escalable para evaluar el rendimiento de ventas, identificar tendencias de consumo y apoyar estrategias de negocio basadas en datos históricos precisos.

OBJETIVOS

Objetivo General: Construir un modelo estrella (Star Schema) para la base de datos transaccional Jardinería, lo que permitirá la generación de un data mart de ventas haciendo más eficientes los análisis y más óptimas las diferentes consultas sobre y precios, productos, categorías y períodos de mayor actividad comercial de la empresa.

Objetivos específicos:

- Diagramar un modelo de estrella, identificando la tabla de hechos y sus respectivas tablas dimensionales, generando así un data mart de la base de datos Jardinería.
- Identificar los productos y años que generan mayor volumen de ventas.
- Determinar la categoría con más productos vendidos.
- Generar reportes y consultas analíticas, en forma eficiente y coherente.

PLANTEAMIENTO DEL PROBLEMA

La empresa Jardinería gestiona sus operaciones comerciales a través de una base de datos transaccional que almacena información detallada sobre productos, clientes, pedidos, pagos y empleados. Si bien esta base de datos es eficiente para el registro de operaciones diarias, presenta limitaciones significativas cuando se requiere analizar el desempeño de ventas y generar reportes estratégicos.

Entre los problemas identificados se encuentran:

- Dificultad para identificar tendencias de ventas, la estructura no normalizada de la base de datos hace que consultas complejas sobre productos, categorías o períodos específicos sean lentas y propensas a errores.
- Limitaciones para la toma de decisiones estratégicas, ya que la información disponible no permite responder rápidamente a preguntas críticas, como cuáles son los productos más vendidos, cuáles categorías tienen mayor aceptación o qué años presentan mayores ventas.
- La base de datos transaccional no está optimizada para consolidar información de diversas dimensiones (producto, cliente, tiempo, empleado) de manera que facilite el análisis comparativo y la generación de indicadores clave de desempeño (KPI).
- La obtención de información a través de consultas ad-hoc puede generar inconsistencias y requiere un esfuerzo significativo del personal, afectando la confiabilidad de los reportes.

Frente a estos problemas, surge la necesidad de construir un modelo estrella que permita consolidar la información en un data mart de ventas, optimizando el análisis de datos históricos, garantizando la integridad de la información y facilitando la generación de reportes estratégicos que apoyen la toma de decisiones.

ANÁLISIS DEL PROBLEMA

1. Limitaciones de la Base de Datos Transaccional

- La base de datos actual está diseñada para registrar operaciones diarias, cumpliendo su función operativa. Sin embargo, presenta varias restricciones para el análisis.
- Las consultas para obtener información consolidada requieren múltiples joins entre tablas (detalle_pedido, pedido, producto, cliente, empleado y Categoria_producto), lo que genera consultas complejas y lentas.
- La estructura sin normalizar dificulta la obtención de informes multidimensionales, como comparar ventas por categoría, producto o periodo de tiempo.
- La información temporal no está optimizada para análisis histórico; por ejemplo, agrupar ventas por mes, trimestre o año implica cálculos adicionales en cada consulta.

2. Impacto en la Toma de Decisiones

- La dificultad para consolidar datos puede impedir identificar rápidamente cuáles productos son los más demandados, afectando la planificación de inventarios y la estrategia comercial.
- Sin un análisis eficiente, la empresa podría no detectar qué categorías tienen mejor desempeño y cuáles requieren atención o promoción.

- La falta de una dimensión de tiempo organizada limita la capacidad de evaluar tendencias y patrones de consumo, afectando decisiones estratégicas sobre inversiones, campañas y proyecciones.

3. Necesidad de un Modelo Estrella

Para superar estas limitaciones, es necesario diseñar un modelo estrella, donde:

- La tabla de hechos contenga las métricas clave de ventas (cantidad, precio, total).
- Las dimensiones proporcionen contexto descriptivo sobre productos, categorías, clientes, tiempo y empleados.
- Se facilite el análisis multidimensional, reduciendo la complejidad y el tiempo de respuesta de las consultas.

El modelo estrella permitirá consolidar información, mejorar la eficiencia en el análisis de ventas y garantizar la integridad de los datos, ofreciendo a la empresa una herramienta confiable para la toma de decisiones estratégicas.

PROPUESTA DE SOLUCIÓN

Para superar las limitaciones del sistema transaccional y facilitar el análisis de ventas en la empresa Jardinería, se propone implementar un modelo estrella (Star Schema). Este modelo permite organizar la información de manera que las métricas de ventas estén centralizadas en una tabla de hechos, mientras que los atributos descriptivos relevantes se estructuran en dimensiones que proporcionan contexto a los análisis.

1. Tabla de Hechos: FactVentas

La tabla de hechos constituye el núcleo del modelo estrella y contiene las métricas cuantitativas de cada transacción de venta.

Campos principales:

- ID_fact_venta: Identificador único de cada venta (clave surrogate).
- ID_producto: Referencia al producto vendido.
- ID_cliente: Referencia al cliente que realizó la compra.
- ID_tiempo: Referencia a la fecha de la venta.
- ID_categoria: Referencia a la categoría del producto.
- ID_empleado (opcional): Referencia al representante de ventas que gestionó la transacción.
- Cantidad: Número de unidades vendidas.
- Precio_unidad: Precio de venta por unidad.
- Total_venta: Monto total de la transacción ($\text{cantidad} \times \text{precio_unidad}$).

Esta tabla permite medir y analizar las ventas desde diferentes perspectivas, facilitando consultas agregadas por producto, categoría, cliente, empleado o período de tiempo.

2. Dimensiones

- DimProducto: Contiene información descriptiva de los productos, como código, nombre, proveedor, dimensiones, precio de venta y precio del proveedor, así como la categoría a la que pertenece.
- DimCategoría: Permite analizar las ventas por categorías de productos. Incluye el nombre de la categoría, descripción textual y representación gráfica.
- DimCliente: Contiene datos del cliente, incluyendo nombre, contacto, ubicación geográfica (ciudad, región, país), teléfono, fax y límite de crédito.
- DimTiempo: Permite analizar las ventas en función del tiempo. Contiene fecha completa, día, mes, trimestre y año, lo que facilita la identificación de tendencias temporales.
- DimEmpleado (opcional): Contiene información del representante de ventas, como nombre, apellidos, email, puesto y oficina de trabajo.

3. Integración entre Hechos y Dimensiones

La tabla de hechos se conecta con cada dimensión mediante claves foráneas, asegurando la integridad referencial y la coherencia de los datos.

Este esquema facilita el análisis multidimensional, permite consultas rápidas sobre ventas por producto, categoría o tiempo, y simplifica la generación de reportes estratégicos para la toma de decisiones.

Lista de Dimensiones Propuestas:

DimProducto

Descripción: Contiene información descriptiva de los productos vendidos.

Columnas y tipos de datos:

Columna	Tipo de dato	Descripción
ID_producto	INT	Clave primaria, identificador del producto
CodigoProducto	VARCHAR(15)	Código del producto
nombre	VARCHAR(70)	Nombre del producto
proveedor	VARCHAR(50)	Nombre del proveedor
dimensiones	VARCHAR(25)	Medidas del producto
precio_venta	NUMERIC(15,2)	Precio de venta del producto
precio_proveedor	NUMERIC(15,2)	Precio de compra del producto
ID_categoria	INT	Clave foránea a DimCategoria

DimCategoria

Descripción: Permite analizar las ventas por categorías de productos.

Columnas y tipos de datos:

Columna	Tipo de dato	Descripción
ID_categoria	INT	Clave primaria, identificador de la categoría
Desc_Categoria	VARCHAR(50)	Nombre de la categoría
descripcion_texto	TEXT	Descripción textual de la categoría
imagen	VARCHAR(256)	Ruta o URL de la imagen

		representativa
--	--	----------------

DimCliente

Descripción: Contiene información del cliente para análisis comercial y geográfico.

Columnas y tipos de datos:

Columna	Tipo de dato	Descripción
ID_cliente	INT	Clave primaria, identificador del cliente
nombre_cliente	VARCHAR(50)	Nombre de la empresa
nombre_contacto	VARCHAR(30)	Nombre del contacto
apellido_contacto	VARCHAR(30)	Apellido del contacto
telefono	VARCHAR(15)	Teléfono principal
fax	VARCHAR(15)	Número de fax
linea_direccion1	VARCHAR(50)	Dirección principal
linea_direccion2	VARCHAR(50)	Dirección secundaria (opcional)
ciudad	VARCHAR(50)	Ciudad del cliente
region	VARCHAR(50)	Región o departamento
pais	VARCHAR(50)	País
codigo_postal	VARCHAR(10)	Código postal
limite_credito	NUMERIC(15,2)	Límite de crédito disponible

DimTiempo

Descripción: Permite análisis temporal de las ventas.

Columnas y tipos de datos:

Columna	Tipo de dato	Descripción
ID_tiempo	INT	Clave primaria, identificador de la fecha
fecha	DATE	Fecha completa de la venta
dia	INT	Día del mes
mes	INT	Mes
trimestre	INT	Trimestre del año
año	INT	Año

DimEmpleado (opcional)

Descripción: Información sobre los representantes de ventas.

Columnas y tipos de datos:

Columna	Tipo de dato	Descripción
ID_empleado	INT	Clave primaria, identificador del empleado
nombre	VARCHAR(50)	Nombre
apellido1	VARCHAR(50)	Primer apellido
apellido2	VARCHAR(50)	Segundo apellido (opcional)
email	VARCHAR(100)	Correo electrónico
puesto	VARCHAR(50)	Cargo del empleado
ID_oficina	INT	Clave foránea a DimOficina

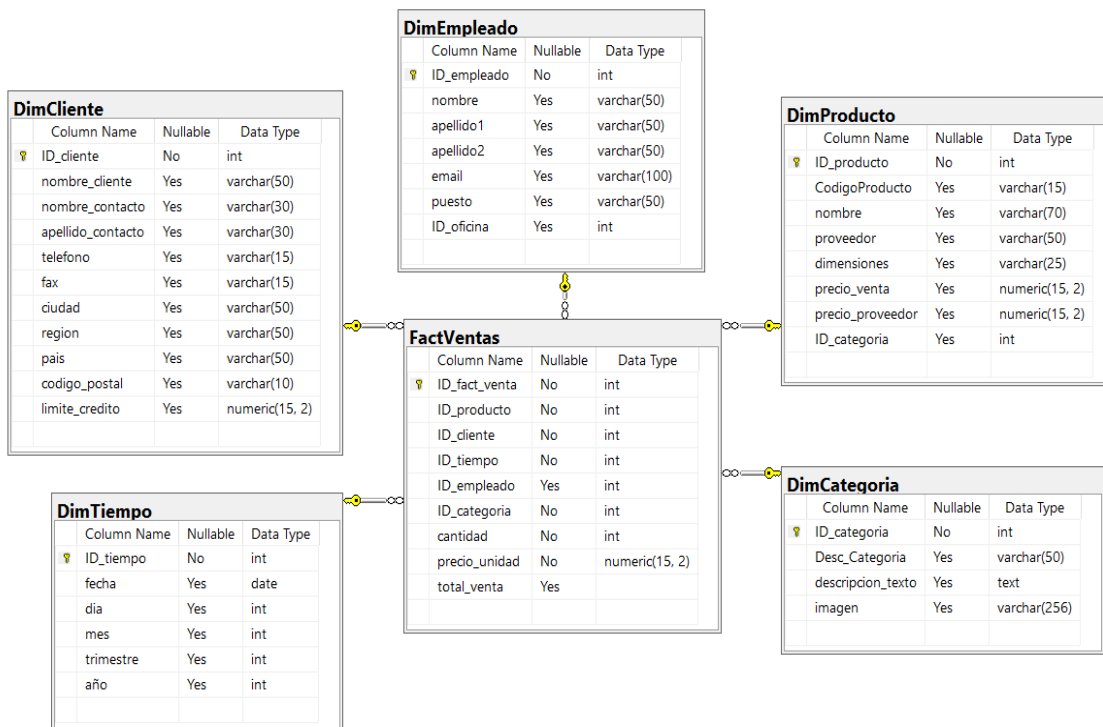
Tabla de Hechos: FactVentas

Descripción:

La tabla FactVentas centraliza las métricas de ventas de la empresa Jardinería. Contiene datos cuantitativos de cada transacción y se conecta con las dimensiones para análisis multidimensional.

Campos y tipos de datos:

Columna	Tipo de dato	Descripción
ID_fact_venta	INT IDENTITY(1,1) PRIMARY KEY	Identificador único de la venta (clave surrogate)
ID_producto	INT NOT NULL	Clave foránea hacia DimProducto, indica el producto vendido
ID_cliente	INT NOT NULL	Clave foránea hacia DimCliente, indica el cliente que realizó la compra
ID_tiempo	INT NOT NULL	Clave foránea hacia DimTiempo, indica la fecha de la venta
ID_categoria	INT NOT NULL	Clave foránea hacia DimCategoria, indica la categoría del producto
ID_empleado	INT NULL	Clave foránea hacia DimEmpleado, representa al empleado que gestionó la venta (opcional)
cantidad	INT NOT NULL	Número de unidades vendidas en la transacción
precio_unidad	NUMERIC(15,2) NOT NULL	Precio de venta por unidad del producto
total_venta	NUMERIC(15,2) PERSISTED	Total de la venta calculado como cantidad * precio_unidad



CONCLUSIONES

La implementación del modelo estrella permite consolidar los datos de ventas de la empresa Jardinería en una estructura diseñada para análisis multidimensional, lo que facilita identificar rápidamente los productos más vendidos, las categorías con mayor volumen y los períodos de mayor actividad comercial.

Mejora en la eficiencia de las consultas al centralizar las métricas de ventas en la tabla de hechos y vincularlas a dimensiones descriptivas, reduciendo las consultas complejas sobre la base de datos transaccional, logrando tiempos de respuesta más rápidos y reportes más precisos.

Las relaciones entre la tabla de hechos y las dimensiones aseguran que toda venta registrada esté asociada a un producto, cliente, categoría y período temporal válidos, garantizando la confiabilidad de la información para la toma de decisiones, afianzando la integridad y consistencia de los datos.

El modelo estrella permite realizar análisis desde diferentes perspectivas (producto, categoría, cliente, empleado y tiempo), proporcionando información útil para la planificación estratégica, control de inventarios y evaluación del desempeño de ventas.

La estructura propuesta es flexible y puede integrarse con nuevas dimensiones o métricas adicionales en el futuro permitiendo su escalabilidad lo que le permite adaptarse a las necesidades crecientes de la empresa sin afectar la consistencia del modelo.

BIBLIOGRAFÍA

Databricks. (s. f.). *Understanding Star Schema*. En *Databricks Glossary*. Recuperado de <https://www.databricks.com/glossary/star-schema>

Sharma, M. K. (2025, 11 julio). *Star Schema in Data Warehouse modeling*. En *GeeksforGeeks*. Recuperado de <https://www.geeksforgeeks.org/dbms/star-schema-in-data-warehouse-modeling/>

Microsoft. (2024, 30 diciembre). *Understand star schema and the importance for Power BI*. En *Microsoft Learn*. Recuperado de <https://learn.microsoft.com/en-us/power-bi/guidance/star-schema>

Wikipedia. (2025, 15 agosto). *Star schema*. En *Wikipedia*. Recuperado de https://en.wikipedia.org/wiki/Star_schema

Visual-Flow. (2024, 4 diciembre). *What Is the Star Schema Data Model?*. Recuperado de <https://visual-flow.com/blog/what-is-the-star-schema-data-model>

TechTarget. (2024, 4 marzo). *What Is a Star Schema and how does it work?* En *TechTarget*. Recuperado de <https://www.techtarget.com/searchdatamanagement/definition/star-schema>

MotherDuck. (s. f.). *The Star Schema: Making Your Data Warehouse Shine*. Recuperado de <https://motherduck.com/learn-more/star-schema-data-warehouse-guide/>

IBM. (s. f.). (2024, 19 noviembre). *Data marts*. En *IBM Informix Documentation*. Recuperado de <https://www.ibm.com/docs/en/informix-servers/15.0.0?topic=aqts-data-marts>

Wikipedia. (2025, 15 agosto). *Data mart*. En *Wikipedia*. Recuperado de https://en.wikipedia.org/wiki/Data_mart

Wikipedia. (2025, 25 agosto). *Dimensional modeling*. En *Wikipedia*. Recuperado de https://en.wikipedia.org/wiki/Dimensional_modeling

YouTube. (s. f.). *Data Modeling Tutorial: Star Schema (aka Kimball Approach)*.
YouTube. Recuperado de <https://www.youtube.com/watch?v=gRE3E7VUzRU>