

Evidencia de Aprendizaje

1. Modelo estrella de un Data Mart

Integrantes

BRAYAN SANTIAGO JARAMILLO AMÉZQUITA

ALEICER VESGA RUEDA

Docente

Antonio Jesus Valderrama Jaramillo

Bases de Datos II

PREICA2502B010064

Ingeniería de software y datos

Institución Universitaria Digital de Antioquia

Medellín, Antioquia

02 de septiembre de 2025

Introducción

El presente documento realiza un análisis de la base de datos "Jardinería" y propone el diseño de un modelo estrella. Este modelo está optimizado para el Procesamiento Analítico en Línea, facilitando consultas rápidas y eficientes que permitirán a la empresa comprender mejor su rendimiento de ventas y tomar decisiones estratégicas basadas en datos.

Objetivos

- **Objetivo General:** Diseñar un modelo estrella para un Data Mart de ventas a partir de la base de datos transaccional "Jardinería".
- **Objetivos Específicos:**
 1. Analizar la estructura y relaciones de la base de datos "Jardinería".
 2. Identificar el proceso de negocio a modelar y definir la granularidad de la tabla de hechos.
 3. Diseñar las dimensiones que proporcionarán contexto a los datos de ventas.
 4. Estructurar la tabla de hechos que contendrá las métricas clave de rendimiento (KPIs).
 5. Documentar la solución propuesta de manera clara y estructurada.

Planteamiento del Problema

La empresa "Jardinería" registra sus operaciones en una base de datos MySQL normalizada. Si bien este diseño es óptimo para la inserción, actualización y eliminación de datos del día a día (garantizando la integridad y evitando la redundancia), resulta ineficiente para responder preguntas de negocio complejas.

Preguntas como: “¿Cuál fue el total de ventas de la categoría 'Herramientas' en el último trimestre, desglosado por país de oficina y representante de ventas?”, requerirían la unión de múltiples tablas (detalle_pedido, pedido, producto, Categoria_producto, cliente, empleado, oficina), resultando en consultas lentas y complejas de escribir. Esta lentitud dificulta la capacidad de los gerentes y analistas para explorar los datos de forma interactiva y obtener insights oportunos para la toma de decisiones.

Análisis del Problema

1. Análisis de la Base de Datos Jardinería

La base de datos relacional proporcionada consta de 8 tablas principales que giran en torno al proceso de ventas:

- Tablas de Transacciones: pedido, detalle_pedido y pago forman el núcleo de las operaciones. detalle_pedido es la tabla más granular, ya que registra cada producto dentro de un pedido, su cantidad y el precio unitario.
- Tablas Maestras (Maestros): cliente, producto, empleado y oficina contienen datos descriptivos. Estas tablas describen quién compra, qué se vende, quién vende y desde dónde se vende.
- Tablas de Catálogo: Categoria_producto clasifica los productos, añadiendo un nivel más de detalle para el análisis.

Las relaciones son de tipo uno a muchos (ej. un cliente puede tener muchos pedidos, un pedido puede tener muchos detalle_pedido). Los empleados (empleado) están asignados a una oficina y a su vez son representantes de ventas del cliente.

2. Identificación de Campos para el Modelo Estrella

Para construir el modelo estrella, se debe identificar un proceso de negocio central. En este caso, el proceso es claramente el de Ventas.

- Tabla de Hechos: La información para la tabla de hechos provendrá principalmente de detalle_pedido, ya que contiene las métricas cuantitativas que queremos analizar: cantidad

y precio_unidad. La granularidad de la tabla de hechos será "cada línea de producto dentro de un pedido".

- Dimensiones: Las dimensiones proporcionarán el contexto a los hechos. Se identifican a partir de las tablas maestras que responden a las preguntas "quién, qué, cuándo y dónde":
 - Cuándo: Las fechas de los pedidos (pedido.fecha_pedido).
 - Qué: La información de los productos (producto y Categoria_producto).
 - Quién (cliente): Los datos del cliente que realiza la compra (cliente).
 - Quién (vendedor): Los datos del empleado que gestiona la venta (empleado).
 - Dónde: La ubicación desde la que se gestiona la venta (oficina).

Propuesta de la solución

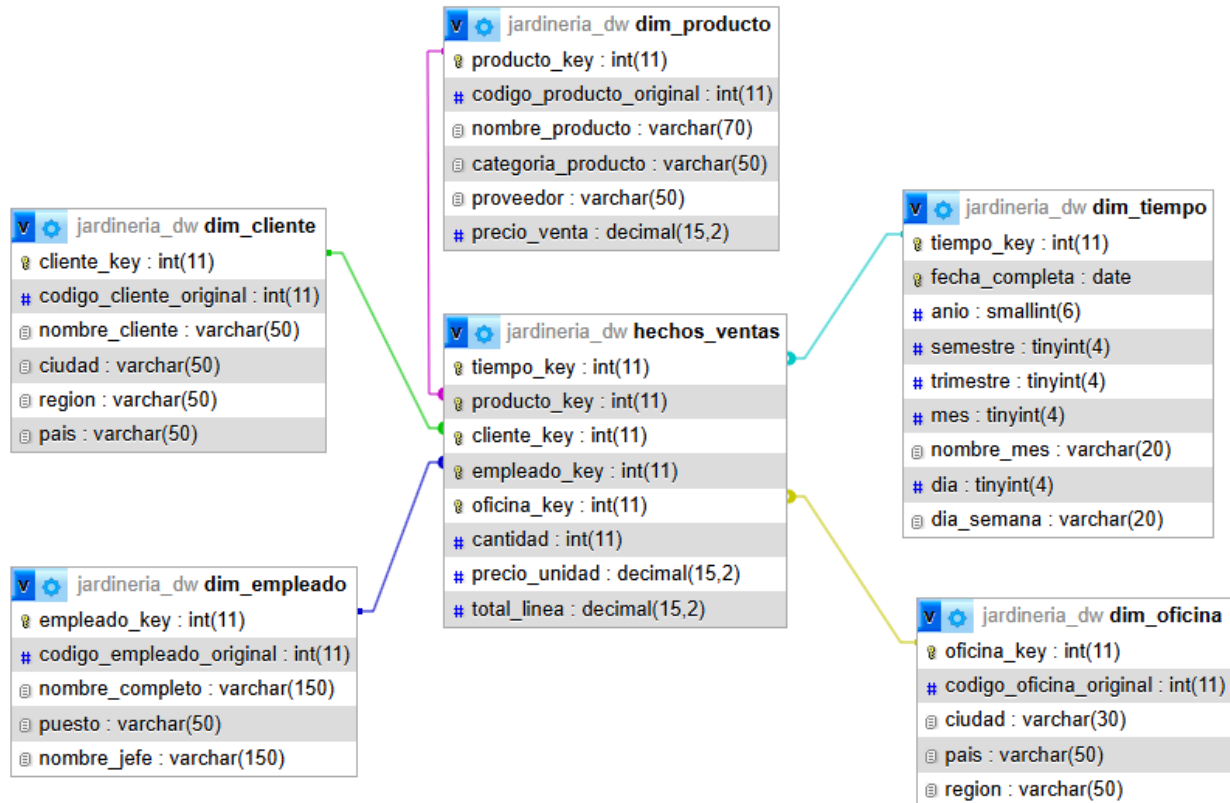
Se propone la creación de un modelo estrella compuesto por una tabla de hechos central (hechos_ventas) y cinco dimensiones (dim_tiempo, dim_producto, dim_cliente, dim_empleado, dim_oficina).

Descripción del Modelo Estrella Propuesto

El modelo centralizará todas las métricas de ventas (cantidad, precio, total) en la tabla hechos_ventas. Cada registro en esta tabla representará una línea de un producto vendido en un pedido específico. Esta tabla de hechos estará conectada directamente a las tablas de dimensiones a través de claves foráneas. Las dimensiones almacenarán los atributos descriptivos de forma desnormalizada para optimizar la velocidad de las consultas. Por ejemplo, la dimensión de producto contendrá no solo el nombre del producto, sino también el nombre de su categoría, evitando un join adicional en tiempo de consulta.

Figura 1

Diseño del modelo estrella



Nota. Imagen tomada de phpMyAdmin.

Lista de Dimensiones Propuestas

1. Dimensión Tiempo (dim_tiempo): Contiene atributos de fecha para analizar las ventas a lo largo del tiempo.
2. Dimensión Producto (dim_producto): Agrupa toda la información relevante de los productos y sus categorías.
3. Dimensión Cliente (dim_cliente): Almacena los datos descriptivos de los clientes.

4. Dimensión Empleado (dim_empleado): Contiene la información de los representantes de ventas y su jerarquía.
5. Dimensión Oficina (dim_oficina): Describe la ubicación geográfica desde donde operan los empleados.

Detalle de las Dimensiones

1. **Dimensión Tiempo** (dim_tiempo) Esta tabla se genera a partir de la columna fecha_pedido de la tabla pedido.

Tabla 1

Dimensión Tiempo

Nombre de la Columna	Tipo de Dato	Descripción
tiempo_key	INT (PK)	Clave subrogada para la dimensión de tiempo (ej: 20250829).
fecha_completa	DATE	La fecha completa (ej: '2025-08-29').
anio	SMALLINT	Año de la fecha.
semestre	TINYINT	Semestre del año (1 o 2).
trimestre	TINYINT	Trimestre del año (1, 2, 3 o 4).
mes	TINYINT	Número del mes (1-12).
nombre_mes	VARCHAR(20)	Nombre del mes (ej: 'Enero').
dia	TINYINT	Día del mes (1-31).
dia_semana	VARCHAR(20)	Nombre del día de la semana (ej: 'Lunes').

2. **Dimensión Producto** (dim_producto) Esta tabla se construye a partir de la unión de las tablas producto y Categoria_producto.

Tabla 2
Dimensión Producto

Nombre de la Columna	Tipo de Dato	Descripción
producto_key	INT (PK)	Clave subrogada para el producto.
codigo_producto_original	INT	ID original del producto en la base de datos transaccional.
nombre_producto	VARCHAR(70)	Nombre del producto.
categoria_producto	VARCHAR(50)	Descripción de la categoría del producto.
proveedor	VARCHAR(50)	Nombre del proveedor.
precio_venta	DECIMAL(15,2)	Precio de venta unitario del producto.

3. Dimensión Cliente (dim_cliente) Esta tabla se alimenta directamente de la tabla cliente.

Tabla 3
Dimensión Cliente

Nombre de la Columna	Tipo de Dato	Descripción
cliente_key	INT (PK)	Clave subrogada para el cliente.
codigo_cliente_original	INT	ID original del cliente en la base de datos transaccional.
nombre_cliente	VARCHAR(50)	Nombre de la empresa cliente.
ciudad	VARCHAR(50)	Ciudad del cliente.
region	VARCHAR(50)	Región del cliente.
pais	VARCHAR(50)	País del cliente.

4. Dimensión Empleado (dim_empleado) Se construye a partir de la tabla empleado, incluyendo una auto-referencia para obtener el nombre del jefe.

Tabla 4

Dimensión Empleado

Nombre de la Columna	Tipo de Dato	Descripción
empleado_key	INT (PK)	Clave subrogada para el empleado.
codigo_empleado_original	INT	ID original del empleado en la base de datos transaccional.
nombre_completo	VARCHAR(150)	Nombre y apellidos del empleado.
puesto	VARCHAR(50)	Cargo del empleado (ej: 'Representante Ventas').
nombre_jefe	VARCHAR(150)	Nombre completo del jefe directo del empleado.

5. Dimensión Oficina (dim_oficina) Se alimenta directamente de la tabla oficina.

Tabla 5

Dimensión Oficina

Nombre de la Columna	Tipo de Dato	Descripción
oficina_key	INT (PK)	Clave subrogada para la oficina.
codigo_oficina_original	INT	ID original de la oficina en la base de datos transaccional.
ciudad	VARCHAR(30)	Ciudad donde se ubica la oficina.
pais	VARCHAR(50)	País de la oficina.
region	VARCHAR(50)	Región a la que pertenece la oficina.

Detalle de la Tabla de Hechos

Tabla de Hechos (hechos_ventas) Es el centro del modelo y se construye a partir de detalle_pedido, pedido y las relaciones necesarias para conectar con las dimensiones.

Tabla 6

Detalle de la Tabla de Hechos

Nombre de la Columna	Tipo de Dato	Descripción
tiempo_key	INT (FK)	Clave foránea que referencia a dim_tiempo.
producto_key	INT (FK)	Clave foránea que referencia a dim_producto.
cliente_key	INT (FK)	Clave foránea que referencia a dim_cliente.
empleado_key	INT (FK)	Clave foránea que referencia a dim_empleado.
oficina_key	INT (FK)	Clave foránea que referencia a dim_oficina.
cantidad	INT	Cantidad de unidades vendidas en la línea del pedido.
precio_unidad	DECIMAL(15,2)	Precio por unidad en la transacción.
total_linea	DECIMAL(15,2)	Métrica calculada (cantidad * precio_unidad).

Conclusiones

Se realiza todo el análisis del requisito dado, se estructura la solución propuesta a través de un modelo estrella para evitar consultas extensas por parte de los encargados de analítica del mercado, con esta solución se logra condensar la información importante y de necesidad próxima para un análisis más eficaz.

Rendimiento Mejorado: Las consultas analíticas se ejecutarán de manera significativamente más rápida, permitiendo un análisis de datos interactivo y fluido.

Simplicidad: La estructura de estrella es más fácil de entender para los usuarios de negocio, lo que fomenta la auto-suficiencia en la generación de informes y reduce la dependencia del departamento de TI.

Visión Integral del Negocio: Facilitará la creación de dashboards y reportes que crucen información de ventas, productos, clientes y geografía, ofreciendo una visión completa del rendimiento comercial.

Toma de Decisiones Estratégica: Con acceso a información consolidada y fiable, la dirección podrá identificar tendencias, patrones de compra y oportunidades de mercado con mayor eficacia, fundamentando sus decisiones en datos concretos.

Bibliografía

- *IBM InfoSphere Data Architect*. (2021, marzo 8). Ibm.com.
<https://www.ibm.com/docs/es/ida/9.1.2?topic=schemas-star>
- *Star schema*. (2022, abril 21). Databricks.
<http://databricks.com/glossary/star-schema>