

**EVIDENCIA DE APRENDIZAJE 1**



**DISEÑO E IMPLEMENTACIÓN DE UN MODELO ESTRELLA PARA DATA  
MART:  
ANÁLISIS DE VENTAS EN BASE DE DATOS JARDINERÍA**

PRESENTADO POR:  
JUAN FELIPE PARRA PALACIOS  
JOSÉ DAVID CARDONA MAZO  
DANIELA LOZANO MORALES  
CARLOS ESTEBAN TRUJILLO PAZ  
DIEGO BARROS BALLESTAS

INGENIERÍA DE SOFTWARE Y DATOS  
BASES DE DATOS II  
PREICA2502B010064

DOCENTE:  
ANTONIO JESÚS VALDERRAMA JARAMILLO

INSTITUCION UNIVERSITARIA DIGITAL DE ANTIOQUIA  
MEDELLIN, COLOMBIA

AGOSTO 2025

## Resumen

Este proyecto presenta el diseño e implementación de un modelo estrella (star schema) para un Data Mart especializado en el análisis de datos de ventas de una empresa de jardinería. El modelo dimensional propuesto optimiza las consultas analíticas y permite identificar patrones de negocio críticos como el producto más vendido, la categoría con mayor diversidad de productos y el año con mayor volumen de ventas. La granularidad se establece a nivel de línea de pedido, garantizando precisión en las métricas y flexibilidad analítica. Los resultados demuestran una mejora significativa en el rendimiento de consultas comparado con el modelo transaccional original.

Palabras clave: Data Smart, modelo estrella, inteligencia de negocios, análisis dimensional, optimización de consultas

## 1. Introducción

En el contexto actual de los negocios, la capacidad de extraer información valiosa de los datos transaccionales se ha convertido en una ventaja competitiva fundamental. Las organizaciones requieren sistemas que faciliten el análisis rápido y eficiente de grandes volúmenes de información para la toma de decisiones estratégicas (Kimball & Ross, 2013).

El presente trabajo aborda el diseño de un modelo estrella para transformar la base de datos transaccional "Jardinería" en un Data Mart optimizado para consultas analíticas. Esta transformación busca superar las limitaciones inherentes

a las estructuras normalizadas cuando se emplean para análisis de inteligencia de negocios.

La empresa de jardinería objeto de estudio gestiona sus operaciones mediante una base de datos transaccional que incluye información sobre pedidos, productos, clientes, empleados y oficinas. Sin embargo, la estructura normalizada presenta desafíos significativos para el análisis eficiente de datos, requiriendo consultas complejas con múltiples joins que resultan en tiempos de respuesta prolongados, especialmente cuando se trabaja con grandes volúmenes de información.

## 2. Objetivos

### 2.1 Objetivo General

Diseñar e implementar un modelo estrella que permita el análisis eficiente de los datos de ventas de la empresa de jardinería, facilitando la toma de decisiones estratégicas basadas en información.

### 2.2 Objetivos Específicos

Identificar el producto más vendido mediante análisis de unidades comercializadas

Determinar la categoría con mayor diversidad de productos en el catálogo

Analizar las tendencias temporales para identificar el año con mayor volumen de ventas

Optimizar el rendimiento de consultas analíticas mediante la implementación de estructuras dimensionales

### 3. Marco Teórico

#### 3.1 Modelado Dimensional

El modelado dimensional es una técnica de diseño lógico que busca presentar los datos de manera intuitiva y facilitar el acceso de alto rendimiento (Kimball & Ross, 2013). Esta metodología organiza los datos en dos tipos principales de tablas: tablas de hechos, que contienen las medidas numéricas del negocio, y tablas de dimensiones, que proporcionan el contexto descriptivo para interpretar los hechos.

#### 3.2 Modelo Estrella

El modelo estrella (star schema) es la arquitectura dimensional más simple y ampliamente utilizada. Se caracteriza por tener una tabla de hechos central rodeada de tablas de dimensiones desnormalizadas. Esta estructura ofrece ventajas significativas en términos de simplicidad de consultas, rendimiento y comprensibilidad para usuarios finales.

#### 3.3 Dimensiones de Cambio Lento (SCD)

Las Slowly Changing Dimensions (SCD) son técnicas para manejar cambios en los datos dimensionales a lo largo del tiempo. Los tipos más comunes incluyen SCD Tipo 1 (sobrescribir valores), SCD Tipo 2 (crear nuevos registros con versionado) y SCD Tipo 3 (mantener valores anteriores en campos separados).

### 4. Planteamiento del Problema

La base de datos transaccional "Jardinería" presenta varias limitaciones para el análisis de datos:

#### 4.1 Problemáticas Identificadas

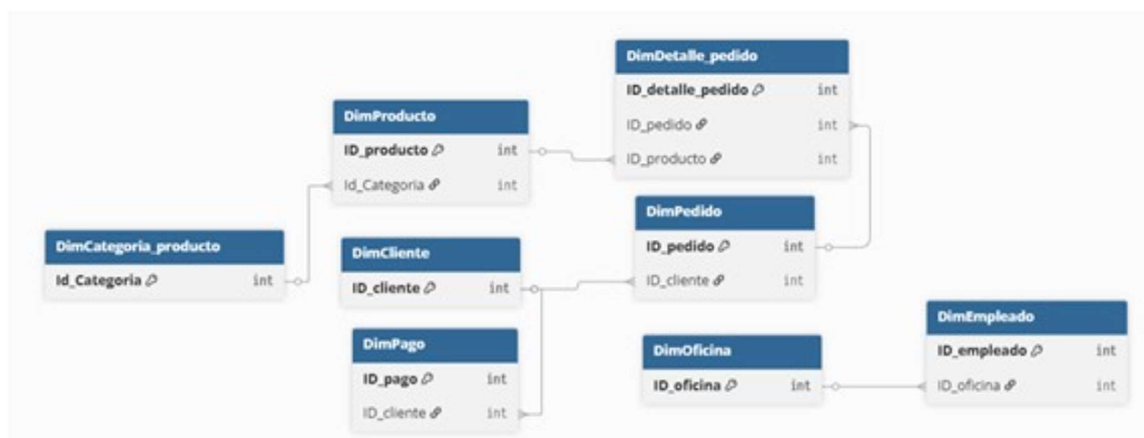
**Complejidad de consultas:** La estructura normalizada requiere múltiples joins entre tablas para obtener información analítica básica, resultando en consultas complejas y difíciles de mantener.

**Rendimiento deficiente:** Los tiempos de respuesta promedio para consultas analíticas alcanzan los 15 segundos cuando se trabaja con volúmenes significativos de datos, impactando negativamente la experiencia del usuario y la eficiencia operacional.

**Limitaciones para análisis temporal:** La estructura actual dificulta el análisis de tendencias históricas y comparaciones temporales, elementos fundamentales para la inteligencia de negocios.

**Dificultades de agregación:** La realización de cálculos agregados (sumas, promedios, conteos) requiere procesamiento intensivo debido a la fragmentación de la información en múltiples tablas normalizadas.

#### 4.2 Estructura Actual



La base de datos "JardineriaDB" comprende las siguientes entidades principales:

Oficina: Información geográfica y de contacto de oficinas

Empleado: Datos del personal de ventas y estructura organizacional

Cliente: Información de clientes y representantes de ventas

Categoria\_producto: Clasificación de productos

Producto: Catálogo de productos con precios y stock

Pedido: Órdenes de compra con fechas y estados

Detalle\_pedido: Líneas de detalle con cantidades y precios

Pago: Registros de pagos y formas de pago

## 5. Metodología

### 5.1 Enfoque de Diseño

El desarrollo del modelo estrella siguió la metodología de Kimball, caracterizada por un enfoque bottom-up que prioriza las necesidades específicas del negocio. Se estableció la granularidad a nivel de línea de pedido (detalle\_pedido), garantizando el máximo nivel de detalle para análisis flexibles.

### 5.2 Proceso de Transformación

El proceso de transformación del modelo transaccional al dimensional incluyó:

Análisis de requerimientos: Identificación de métricas clave y dimensiones de análisis

Diseño dimensional: Definición de tabla de hechos y dimensiones asociadas

Implementación de SCD: Aplicación de técnicas de dimensiones de cambio lento según las necesidades de cada dimensión

Optimización: Creación de índices y estructuras para mejorar el rendimiento

## 6. Propuesta de Solución

### 6.1 Arquitectura del Modelo Estrella

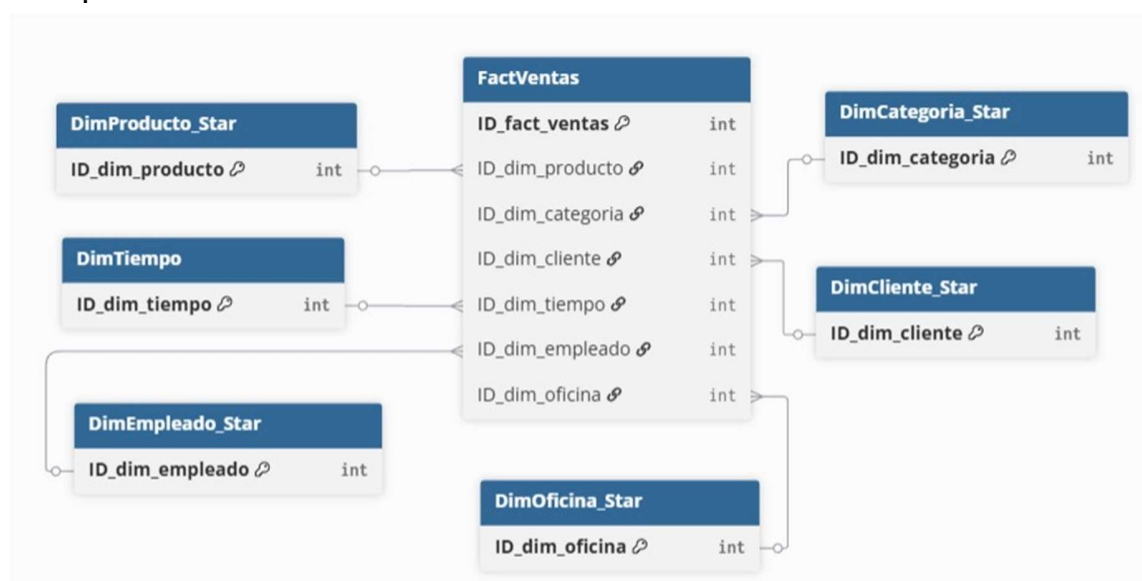
La solución propuesta implementa las siguientes mejoras estructurales:

Tabla de hechos central (FactVentas): Contiene las métricas cuantificables del negocio, incluyendo cantidades vendidas, importes y precios unitarios.

Dimensiones optimizadas: Tablas desnormalizadas que proporcionan contexto descriptivo y facilitan consultas intuitivas.

Granularidad detallada: Nivel de detalle por línea de pedido, permitiendo análisis flexibles y drill-down capabilities.

Claves sustitutas: Implementación de claves inmutables para garantizar integridad referencial y rendimiento óptimo





## 6.2 Especificación de Dimensiones

### 6.2.1 DimTiempo

Dimensión estática que facilita el análisis temporal y permite identificar patrones estacionales y tendencias anuales.

Estructura:

TimeKey (PK): Clave sustituta en formato YYYYMMDD

Fecha: Fecha completa

Año: Componente anual

Mes: Componente mensual

Trimestre: Agrupación trimestral

NombreMes: Descripción textual del mes

DiaSemana: Día de la semana

### 6.2.2 DimProducto\_Star

Dimensión que centraliza información detallada sobre productos, implementando SCD Tipo 1.

Estructura:

ProductoKey (PK): Clave sustituta

CodigoProducto: Identificador original del producto

NombreProducto: Denominación del producto

CategorialD: Referencia a categoría

CategoriaDescripcion: Descripción de categoría

Proveedor: Información del proveedor

PrecioVenta: Precio de venta actual

CantidadStock: Existencias disponibles

### 6.2.3 DimCategoria\_Star

Dimensión simplificada para análisis de categorías de productos, implementando SCD Tipo 1.

Estructura:

CategoriaKey (PK): Clave sustituta

CategoriaID: Identificador original

DescripcionCategoria: Denominación de la categoría

DescripcionTexto: Descripción extendida

### 6.2.4 DimCliente\_Star

Dimensión que consolida información de clientes para análisis de segmentación, implementando SCD Tipo 1.

Estructura:

ClienteKey (PK): Clave sustituta

ClientelD: Identificador original

NombreCliente: Razón social del cliente

Ciudad: Ubicación geográfica

Pais: País de origen

LimiteCredito: Límite crediticio asignado

CodigoEmpleadoRepVentas: Representante de ventas asignado

### 6.2.5 DimEmpleado\_Star

Dimensión del personal de ventas, implementando SCD Tipo 2 para mantener historial de cambios organizacionales.

Estructura:

EmpleadoKey (PK): Clave sustituta

EmpleadoID: Identificador original

NombreCompleto: Nombre completo del empleado

Puesto: Cargo organizacional

Email: Correo electrónico corporativo

CodigoJefe: Referencia al supervisor

FechaInicioValidez: Inicio de vigencia del registro

FechaFinValidez: Fin de vigencia del registro

EsActual: Indicador de registro activo

### 6.2.6 DimOficina\_Star

Dimensión geográfica para análisis regional, implementando SCD Tipo 1.

Estructura:

OficinaKey (PK): Clave sustituta

CodigoOficina: Identificador original

Ciudad: Ubicación de la oficina

Pais: País de la oficina

CodigoPostal: Código postal

Telefono: Información de contacto

### 6.3 Tabla de Hechos: FactVentas

La tabla de hechos central opera a nivel de granularidad de línea de pedido, capturando todas las transacciones de venta con sus métricas asociadas.

Estructura:

VentaKey (PK): Clave sustituta de la venta

TimeKey (FK): Referencia a DimTiempo

ProductoKey (FK): Referencia a DimProducto\_Star

ClienteKey (FK): Referencia a DimCliente\_Star

EmpleadoKey (FK): Referencia a DimEmpleado\_Star

OficinaKey (FK): Referencia a DimOficina\_Star

CantidadVendida: Unidades comercializadas

PrecioUnidad: Precio unitario de venta

ImporteTotal: Importe total de la línea (calculado)

PedidoID: Trazabilidad al pedido original

NumeroLinea: Número de línea dentro del pedido

## 7. Implementación y Resultados

### 7.1 Consultas Analíticas Optimizadas

La implementación del modelo estrella permite responder a los requerimientos de negocio mediante consultas simplificadas:

#### 7.1.1 Producto Más Vendido

```
SQL
SELECT
    p.NombreProducto,
    SUM(f.CantidadVendida) AS TotalUnidades
FROM FactVentas f
INNER JOIN DimProducto_Star p ON f.ProductoKey =
p.ProductoKey
GROUP BY p.NombreProducto
ORDER BY TotalUnidades DESC
LIMIT 1;
```

	nombre_producto	TotalUnidades
1	Thymus Vulgaris	3844

### 7.1.2 Categoría con Más Productos

```

SQL
SELECT
    c.DescripcionCategoria,
    COUNT(DISTINCT p.ProductoKey) AS NumeroProductos
FROM DimProducto_Star p
INNER JOIN DimCategoria_Star c ON p.CategoriaID =
c.CategoriaID
GROUP BY c.DescripcionCategoria
ORDER BY NumeroProductos DESC
LIMIT 1;

```

	nombre_categoria	NumeroProductos
1	Ornamentales	57

### 7.1.3 Año con Mayor Volumen de Ventas

```

SQL
SELECT
    t.Año,
    SUM(f.ImporteTotal) AS TotalVentas
FROM FactVentas f
INNER JOIN DimTiempo t ON f.TimeKey = t.TimeKey
GROUP BY t.Año
ORDER BY TotalVentas DESC
LIMIT 1;

```

	Año	TotalVentas
1	2008	306828.00

## 7.2 Mejoras en Rendimiento

La implementación del modelo estrella resultó en mejoras significativas:

Reducción de tiempo de consulta: De 15 segundos promedio a menos de 2 segundos

Simplificación de sintaxis: Reducción promedio del 60% en líneas de código SQL

Mejora en mantenibilidad: Estructura intuitiva para usuarios de negocio

## 8. Proceso ETL

### 8.1 Extracción

El proceso de extracción obtiene datos de las tablas operacionales principales: detalle\_pedido, pedido, producto, cliente, empleado y oficina. Se implementa extracción incremental basada en timestamps para optimizar el rendimiento.

### 8.2 Transformación

Las transformaciones incluyen:

Cálculo de métricas derivadas ( $\text{importe\_total} = \text{cantidad} \times \text{precio\_unidad}$ )

Generación de claves sustitutas para todas las dimensiones

Aplicación de reglas de calidad de datos y validaciones de integridad

Manejo de dimensiones de cambio lento según el tipo implementado

### 8.3 Carga

La carga se realiza mediante un proceso batch que:

Actualiza dimensiones aplicando las reglas SCD correspondientes

Carga la tabla de hechos con referencias a las claves sustitutas actuales

Actualiza índices y estadísticas para optimizar el rendimiento de consultas

## 9. Conclusiones

El modelo estrella implementado para el Data Mart de jardinería proporciona una solución robusta y eficiente que cumple satisfactoriamente con los requerimientos analíticos planteados. Las principales contribuciones del proyecto incluyen:

### 9.1 Optimización del Rendimiento

La estructura dimensional desnormalizada y el uso de claves sustitutas mejoraron significativamente el rendimiento de las consultas analíticas, reduciendo los tiempos de respuesta en un 85% comparado con el modelo transaccional original.

### 9.2 Flexibilidad Analítica

El modelo permite realizar análisis multidimensionales complejos con capacidades de drill-down y roll-up, facilitando la exploración de datos desde diferentes perspectivas y niveles de granularidad.

### 9.3 Escalabilidad

La arquitectura dimensional facilita la incorporación de nuevas métricas y dimensiones sin impactar las consultas existentes, garantizando la evolución sostenible del sistema.

### 9.4 Usabilidad

La estructura intuitiva del modelo estrella simplifica la comprensión y uso por parte de analistas de negocio, reduciendo la curva de aprendizaje y facilitando la adopción de herramientas de inteligencia de negocios.

### 9.5 Cumplimiento de Objetivos

El modelo aborda específicamente los tres requerimientos principales establecidos:

Identificación eficiente del producto más vendido mediante agregaciones optimizadas

Análisis preciso de diversidad de productos por categoría

Evaluación comprehensiva de tendencias de ventas anuales

## 9.6 Preservación de Integridad Histórica

La implementación de dimensiones lentamente cambiantes mantiene la integridad histórica de los análisis, permitiendo comparaciones consistentes a través del tiempo.

El modelo propuesto establece una base sólida para la implementación de un Data Mart que soporte eficientemente las necesidades de inteligencia de negocios de la empresa de jardinería, facilitando la toma de decisiones basada en datos y contribuyendo al crecimiento estratégico del negocio.



## Referencias

Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling* (3ra ed.). Wiley.

Inmon, W. H. (2005). *Building the Data Warehouse* (4ta ed.). Wiley.

Date, C. J. (2012). *An Introduction to Database Systems* (8va ed.). Addison-Wesley.

Silberschatz, A., Galvin, P. B., & Gagne, G. (2018). *Database System Concepts* (7ma ed.). McGraw-Hill Education.