

Evidencia de aprendizaje 1. Modelo estrella de un Data Mart.

Carlos Dueiner Castaño Rodríguez

Facultad de Ingeniería y ciencias agropecuarias

Bases de datos II

Docente: Víctor Hugo Mercado

Ingeniería de Software y Datos

IU Digital de Antioquía

Marzo 2025

## Introducción

En el contexto actual donde la inteligencia empresarial (Business intelligence) es el motor de crecimiento de una organización, los datos se convierten en la fuente de dicha inteligencia, siendo el eje mediante el cual, tras su recolección, limpieza, organización y selección, se toman las decisiones que marcarán fundamentalmente el rumbo de dicha organización. De este modo, la manera en la que se presentan estos datos tras ser recolectados, es vital para lograr la obtención de información valiosa que ayude a tomar las medidas correspondientes frente a las necesidades o demandas de la empresa.

Es claro que las aplicaciones o plataformas manejadas por una organización están soportadas por una base de datos tradicional, evitando redundancias, por lo que están altamente normalizadas, estas, junto con otras fuentes de utilidad para la organización arrojan estos datos esenciales, dichos datos usualmente son de grandes volúmenes y necesitan ser procesados de manera correcta y ordenada de manera que generen información de valor que soporten las decisiones tomadas para el crecimiento esperado.

En este contexto, uno de los modelos más utilizados para estructurar estos datos y lograr los objetivos indicados es el modelo estrella, el cual permite consultar y analizar dichos datos por medio de una tabla central de hechos vista o analizada desde distintas dimensiones, que a su vez son otras tablas conectadas a esta que aportan detalles relevantes adicionales a los hechos. Para el propósito de esta actividad se diseñará un modelo estrella para un data mart que permita dar respuesta a tres categorías de la base de datos llamada **Jardinería**, dichas categorías son:

- Identificar el producto más vendido
- Identificar la categoría con más productos
- Identificar el año con más ventas.

## **Objetivos**

### **Objetivo general:**

La idea es que con la ejecución de esta actividad, creando el data mart de la base de datos jardinería, se demuestre la comprensión de los temas relacionados al modelo de estrella, definiendo y diferenciando la tabla de hechos y su relación con las dimensiones desde la que será evaluada, identificando los campos y tipos de datos que son relevantes en mi base de datos tradicional que nos ayudarán a lograr los requerimientos que senos han realizado, asimismo como la creación de estos mismos campos pero ahora en la tabla de hechos y las dimensiones correspondientes. De este modo se podrá reconocer la adquisición de habilidades de detalle y análisis que son esenciales para la correcta resolución de la problemática abordada.

### **Objetivos específicos:**

- Lograr identificar correctamente la tabla de hechos y las dimensiones realmente necesarias para abordar la problemática de estudio.
- Detallar de forma correcta todos los campos y su tipo de dato de la tabla de hechos y las tablas de dimensiones que serán indispensables para lograr dar respuesta a cada requerimiento.
- Establecer las relaciones lógicas entre la tabla de hechos y las dimensiones, que permitan cumplir con las categorías solicitadas: identificar el producto más vendido, la categoría con más productos y el año con más ventas, utilizando la base de datos Jardinería como fuente de datos.
- Diseñar de manera elegante y entendible la gráfica del modelo de estrella donde los anteriores objetivos se vean plasmados de forma correcta.

## **Planteamiento del problema**

El software de la empresa Jardinería, está soportado por una base de datos transaccional, la cual tiene almacenada información relevante acerca de su servicio en las siguientes tablas: producto, categoría\_producto, cliente, pedido, detalle\_pedido, pago, empleado y oficina, estas tablas están correctamente normalizadas y relacionadas de manera apropiada para el sostenimiento de la aplicación que administra la empresa, almacenando así las operaciones diarias; pero dicha base de datos no está diseñada para realizar análisis profundos y eficientes que conlleven a la toma de decisiones estratégicas en la búsqueda del exitoso crecimiento de la organización.

Frente a esta realidad, se ha planteado la necesidad de crear un data mart, usando el modelo de estrella, en el que se pueda dar respuesta a los siguientes requerimientos específicamente:

- Identificar el producto más vendido
- Señalar la categoría con más productos
- Traer el año con más ventas

Para ello, debe hacerse un análisis profundo que permita identificar las tablas de la base de datos transaccional que conformarán la tabla de hechos y a su vez, relacionarla con las tablas de dimensiones con las que se logrará dar respuesta a cada requerimiento, identificando los campos que deben estar presente en cada tabla; de modo que no solamente se de respuesta a los tres requerimientos planteados, sino que también se logre dar una base sólida para posibles futuras categorías adicionales que puedan necesitarse en pro del crecimiento de la jardinería; por ejemplo, en el caso que no solo pregunten por el año con más ventas, sino el mes, semana o día del año con más ventas.

## **Análisis del problema**

Como es sabido, las bases de datos transaccionales son excelentes para servir como soporte de almacenamiento y consistencia de los datos que por medio de la plataforma se ingresa, haciendo posible la ejecución de distintos procesos para los cuales la aplicación fue diseñada en beneficio de la organización, no obstante, este tipo de base de datos transaccionales, aunque por medio del lenguaje de consulta DQL es posible extraer información de utilidad para ciertas decisiones a tomar, estos no nos dan un detalle profundo o amplio de distintos enfoques desde los que podríamos soportar decisiones de gran importancia dentro de la empresa.

Si se intentara trabajar solo con consultas SQL, se requeriría hacer las mismas consultas cada vez que se requiere nueva información o se actualizan los datos, lo cual genera una mayor carga operativa y mayor extensión en el tiempo requerido para integrar los datos manualmente, además dado que es muy manual dicho proceso, podría desaprovecharse información de gran relevancia, maximizando la posibilidad de tomar decisiones sin mucho fundamento que podrían generar confusión en lugar de dar soluciones.

Para superar dichas limitaciones, se hace necesaria la creación de un data mart, en el que podamos por medio de un modelo de estrella, reorganizar todo este gran volumen de datos para facilitar la identificación de métricas clave, consolidando las ventas en una tabla de hechos y sus correspondientes descripciones en las tablas de dimensiones, con esto logramos:

- Reducir la complejidad de cada consulta
- Permitir acceso de manera directa a información que es clave.
- Escalabilidad, pues no solo se da respuesta a las preguntas actuales sino a posibles consultas futuras que vayan surgiendo a medida que crece el negocio.

## Propuesta de la solución

### Descripción del modelo de estrella:

El modelo de estrella creado, tiene una estructura que se basa en una tabla de hechos central y varias tablas de dimensiones que nos dan contexto y hace posible el análisis de los datos. Cada tabla de dimensión tiene su clave primaria que pasa como clave foránea en la table de hechos, haciendo así una relación de uno a muchos, conectando las dimensiones con la tabla de hechos. De este modo es posible analizar cada una de las ventas desde distintas dimensiones, además, es posible agruparlas según cada dimensión y dar respuesta a los requerimientos que el dueño del producto demanda para su análisis y toma de decisiones.

**Tabla de hechos: Llamada FAC\_Ventas;** Contiene su clave principal id\_venta y las claves foráneas de cada una de las tablas de dimensiones. Además, cuenta con otros campos propios como la forma de pago, la cantidad y el total de la venta

FAC_Ventas		
PK	id_venta	int
FK	id_tiempo	int
FK	id_categoria	int
FK	id_producto	int
FK	id_sucursal	int
FK	id_cliente	int
	forma_de_pago	varchar(20)
	cantidad	int
	total_venta	decimal(10,2)

**Tabla de Dimensión de tiempo; DIM\_Tiempo:** En esta tabla, encontramos campos relacionados a detalles sobre el tiempo asociado a las ventas, cuya clave principal es el `id_tiempo` y demás campos como `fecha`, `año`, `semestre`, `mes_texto`, `num_semana`, `fin_de_semana` (booleano), `día_semana`, y `hora`.

DIM_Tiempo		
PK	id_tiempo	int
	fecha	date
	año	int
	semestre	int
	mes_texto	varchar(20)
	num_semana	int
	fin_de_semana	int
	día_semana	Varchar(20)
	día_semana_txt	bool
	Hora	time

**Tabla de dimensión de categoría; DIM\_Categoría\_producto:** Proporciona información sobre las categorías a las que pertenece cada producto vendido, incluyendo el `id_categoria` como clave primaria, además del atributo `nombre`.

DIM_Categoria_producto		
PK	id_categoria	int
	nombre_categoria	varchar(100)

**Tabla de dimensión de producto; DIM\_Producto:** Aquí se proporciona detalles sobre los productos vendidos y cuenta con su clave primaria id\_producto, además de atributos de valor como nombre\_producto, precio, marca, stock\_actual y proveedor\_principal

DIM_Producto		
PK	id_producto	int
	nombre_producto	varchar(50)
	precio	decimal(10,2)
	marca	varchar(50)
	stock_actual	int
	proveedor_principal	varchar(100)

Además de estas tablas, con las cuales de por si ya es posible dar respuesta a las categorías que el usuario de la Jardinería demanda, se han añadido dos tablas adicionales de gran importancia, dado el caso que en el futuro se demanden más categorías, dentro de estas tablas tenemos:

**Tabla de dimensión de sucursal; DIM\_Sucursal;** Contiene detalles de las sucursales donde se realizan las ventas. Incluye: su clave primaria id\_sucursal y demás atributos como nombre\_sucursal, país, ciudad y dirección.

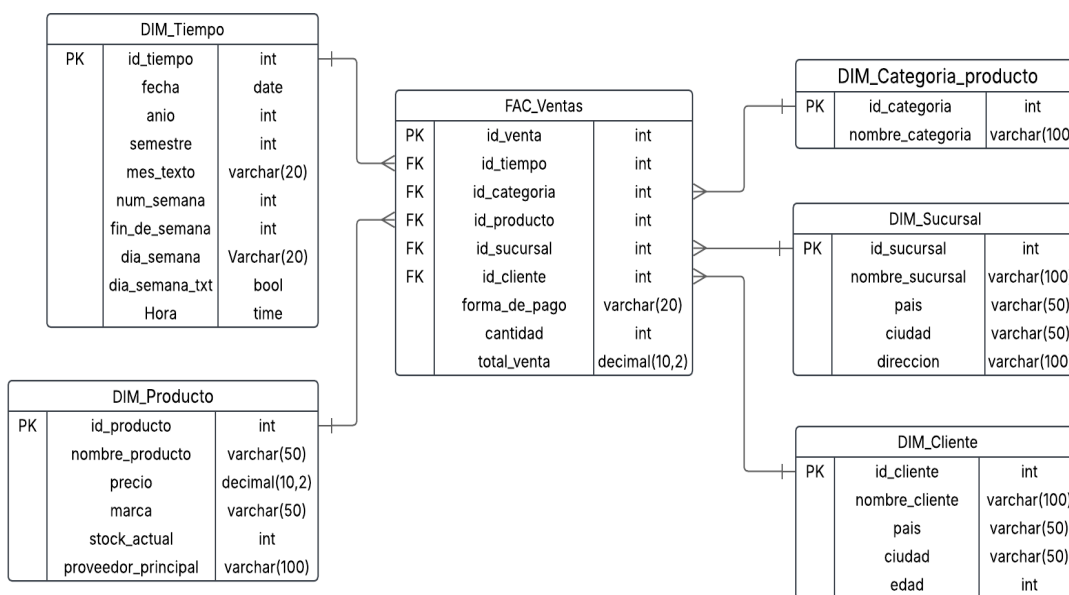
DIM_Sucursal		
PK	id_sucursal	int
	nombre_sucursal	varchar(100)
	pais	varchar(50)
	ciudad	varchar(50)
	direccion	varchar(100)



**Tabla de dimensión de cliente; DIM\_Cliente:** Proporciona información sobre los clientes. Incluyendo la clave principal id\_cliente y demás atributos como nombre\_cliente, país, ciudad y edad.

DIM_Cliente		
PK	id_cliente	int
	nombre_cliente	varchar(100)
	pais	varchar(50)
	ciudad	varchar(50)
	edad	int

A continuación, se comparte la gráfica del modelo de estrella, tanto la imagen, como su respectivo link de la herramienta Lucidchart.



[https://lucid.app/lucidchart/c6a1db98-295d-4e54-bde1-3f3c343e90aa/edit?viewport\\_loc=-90%2C-55%2C1862%2C787%2C0\\_0&invitationId=inv\\_987a1269-c012-4257-96b8-1ec8f577ce5f](https://lucid.app/lucidchart/c6a1db98-295d-4e54-bde1-3f3c343e90aa/edit?viewport_loc=-90%2C-55%2C1862%2C787%2C0_0&invitationId=inv_987a1269-c012-4257-96b8-1ec8f577ce5f)

**Bibliografía:**

Herramienta externa Lucidchart: <https://www.lucidchart.com>

Guía para la construcción del modelo de estrella: Zerpa, H., García, R., & Izquierdo, H. (2020). DATAMART BASADO EN EL MODELO STAR PARA LA IMPLEMENTACIÓN DE INDICADORES CLAVE DE DESEMPEÑO COMO SALIDA DE BIG DATA. *Universidad Ciencia Y Tecnología* , 24 (102), 47-54. <https://doi.org/10.47460/uct.v24i102.342>

