

## Evidencia de Aprendizaje 1. Modelo Estrella de un Data Mart

Presentado por:

Damián Pérez Salvatierra

Catalina Vinasco Arredondo

Marleny Del Pilar Ochoa Tapias

Yilma Valeria Espitia Sanabria

Grupo: 76

Docente:

Víctor Hugo Mercado

Bases de Datos II

Institución Universitaria Digital de Antioquia

Ingeniería de Software y Datos

21 de abril de 2024

## Introducción

Las empresas generan gran cantidad de datos cada día, siendo estos un activo muy valioso. Para su análisis es fundamental utilizar la inteligencia de negocio y la organización para dar facilidad al momento de tomar decisiones. El modelo estrella es una estructura dentro del Data Warehouse, diseñada para optimizar y facilitar el análisis de datos en entornos empresariales. Un Data Mart convierte la información sin procesar en contenido estructurado y significativo para un departamento empresarial específico (AWS). Los datos se organizan en una tabla central de hechos que contiene las métricas del objeto a analizar y tablas de dimensiones que asemejan los brazos de una estrella, representando cada uno de los ejes en un espacio multidimensional.

Una de las grandes ventajas de un modelo estrella Data Mart es la eficiencia en el análisis, flexibilidad, mejora del rendimiento empresarial, optimización del proceso de toma de decisiones y compatibilidad con herramientas de análisis.

Esta clase de modelo facilita la exploración de datos, la generación de informes e identificación de tendencias y patrones clave que pueden ayudar a mejorar el rendimiento empresarial, debido a que permite optimizar el proceso de uso de la información, a través de la agrupación de factores de interés que inciden en un hecho en particular. (Zerpa, Garcia, & Izquierdo, 2020).

En el proceso de búsqueda de información en la base de datos proporcionada, es necesario identificar los aspectos de mayor relevancia. Es por esto que, en el siguiente trabajo nos enfocaremos en la construcción de un modelo estrella Data Mart utilizando la base de datos de jardinería y se llevará a cabo el análisis de las categorías para identificar el producto más vendido, el año de mayores ventas y la categoría con más productos.

## **Objetivo**

Construir un modelo estrella para un *data mart* que permita analizar y responder a tres categorías específicas: 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.

## **Objetivos Específicos:**

- Identificar dimensiones relevantes, que contribuyan al análisis de ventas.
- Desarrollar el proceso ETL minuciosamente, haciendo una adecuada limpieza de datos de la base de datos jardinería.
- Revisar el diseño actual de la base de datos, eliminando duplicidad y mejorando la eficiencia del almacenamiento de datos.
- Desarrollar tablas de hechos y dimensiones donde se visualice el modelo estrella, teniendo como respaldo herramientas de administración de bases de datos.

## **Planteamiento del problema:**

Si bien la base de datos de Jardinería se encuentra la relación entre varias tablas relacionadas al cliente, pagos, empleados, pedidos y productos, estas tablas no proporcionan la información adecuada para lograr responder las preguntas del negocio.

En primer lugar, la falta de conexión entre estas tablas y la dispersión de los datos de la base de datos de Jardinería hace que sea una tarea compleja el lograr identificar con la granularidad necesaria para saber cuál es el producto más vendido y cuál es la categoría con más productos.

En segundo lugar, la falta de un modelo estrella dificulta la correcta interpretación y análisis de la información disponible no solo de las preguntas del negocio, sino también la identificación de posibles problemas, relaciones, patrones e incluso oportunidades que llegarán a presentar entre los datos.

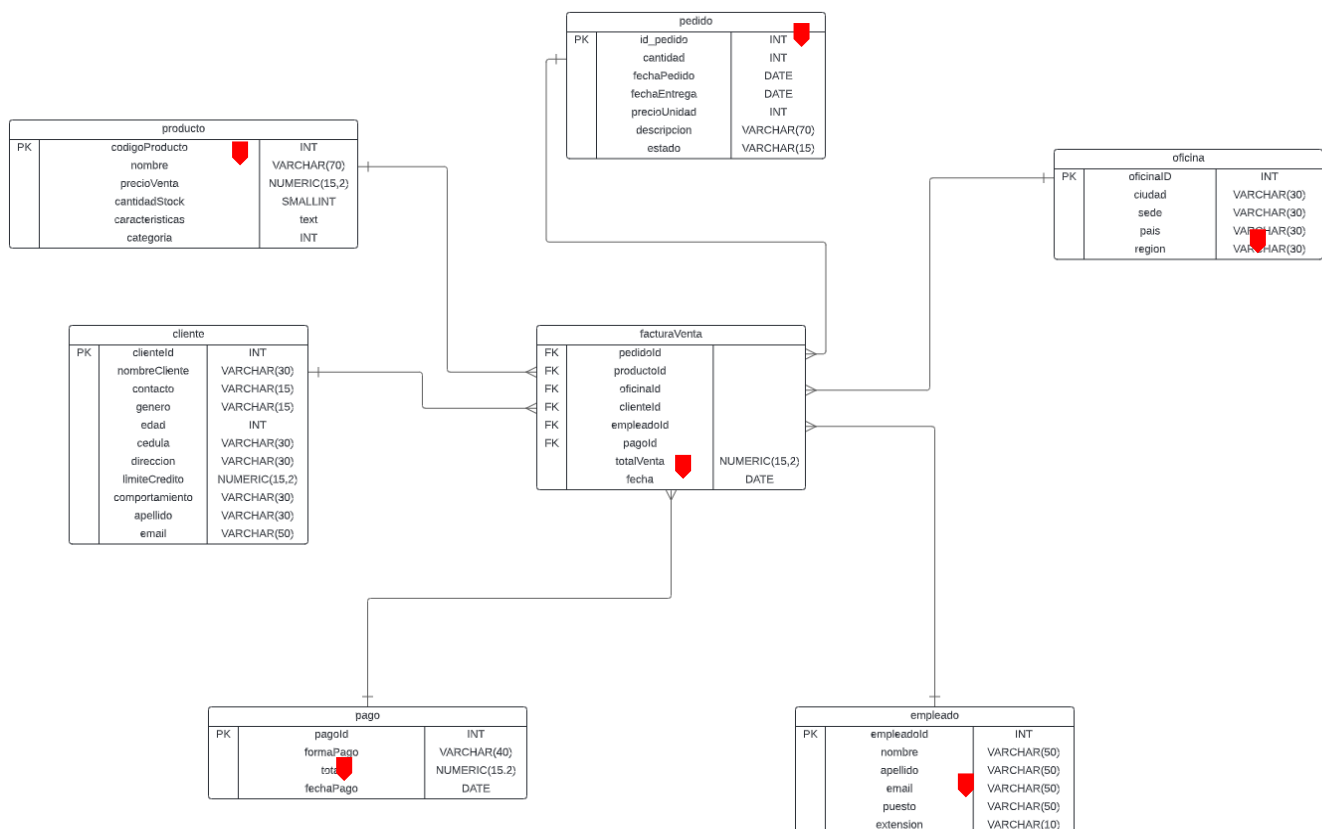
## Análisis del Problema:

De acuerdo con el análisis minuciosamente realizado dentro de nuestro grupo, pudimos observar que, gracias al modelo de estrella, hay una forma de identificar las tres categorías solicitadas en la evidencia solicitada. Ya que el problema radica en la falta de conexión entre las tablas y la dispersión de datos en la base de datos de Jardinería, al implementar un modelo estrella con FacturaVentas y dimensiones resolvería estas deficiencias.

## Modelo Estrella

Descripción del modelo estrella:

Se parte de una tabla de hechos FacturaVentas, relacionada a seis dimensiones, pedido, producto, cliente, oficina, empleado, pago; las cuales están relacionadas cada dimensión de 1 a muchos con la tabla de hechos. Dichas relaciones y modelo nos permitirán entre otras, resolver las preguntas propuestas en el objetivo.



## Lista de dimensiones propuestas

Pudimos identificar seis dimensiones que son las siguientes:

### **Producto:**

producto		
PK	codigoProducto	INT
	nombre	VARCHAR(70)
	precioVenta	NUMERIC(15,2)
	cantidadStock	SMALLINT
	caracteristicas	text
	categoria	INT

### **Cliente**

cliente		
PK	clienteId	INT
	nombreCliente	VARCHAR(30)
	contacto	VARCHAR(15)
	genero	VARCHAR(15)
	edad	INT
	cedula	VARCHAR(30)
	direccion	VARCHAR(30)
	limiteCredito	NUMERIC(15,2)
	comportamiento	VARCHAR(30)
	apellido	VARCHAR(30)
	email	VARCHAR(50)

### **Pedido**

pedido		
PK	id_pedido	INT
	cantidad	INT
	fechaPedido	DATE
	fechaEntrega	DATE
	precioUnidad	INT
	descripcion	VARCHAR(70)
	estado	VARCHAR(15)

### **Oficina**

oficina		
PK	oficinaID	INT
	ciudad	VARCHAR(30)
	sede	VARCHAR(30)
	pais	VARCHAR(30)
	region	VARCHAR(30)

### **Pago**

pago		
PK	pagoid	INT
	formaPago	VARCHAR(40)
	total	NUMERIC(15.2)
	fechaPago	DATE

### **Empleado**

empleado		
PK	empleadoid	INT
	nombre	VARCHAR(50)
	apellido	VARCHAR(50)
	email	VARCHAR(50)
	puesto	VARCHAR(50)
	extension	VARCHAR(10)

### **Tabla de Hechos**

facturaVenta		
FK	pedidoid	
FK	productoid	
FK	oficinaid	
FK	clienteid	
FK	empleadoid	
FK	pagoid	
	totalVenta	NUMERIC(15,2)
	fecha	DATE

### **Tablas**

```
CREATE TABLE pedido (  
  id_pedido INT,  
  cantidad INT,  
  fechaPedido DATE,  
  fechaEntrega DATE,  
  precioUnidad INT,  
  descripcion VARCHAR(70),  
  estado VARCHAR(15),
```

```
PRIMARY KEY (id_pedido)
);

CREATE TABLE producto (
    codigoProducto INT,
    nombre VARCHAR(70),
    precioVenta NUMERIC(15,2),
    cantidadStock SMALLINT,
    características text,
    categoria INT,
    PRIMARY KEY (codigoProducto)
);

CREATE TABLE oficina (
    oficinaID INT,
    ciudad VARCHAR(30),
    sede VARCHAR(30),
    pais VARCHAR(30),
    region VARCHAR(30),
    PRIMARY KEY (oficinaID)
);

CREATE TABLE cliente (
    clienteID INT,
    nombreCliente VARCHAR(30),
    contacto VARCHAR(15),
    genero VARCHAR(15),
    edad INT,
    cedula VARCHAR(30),
    direccion VARCHAR(30),
    limiteCredito NUMERIC(15,2),
```

comportamiento VARCHAR(30),  
apellido VARCHAR(30),  
email VARCHAR(50),  
PRIMARY KEY (clienteld)

);

CREATE TABLE empleado (

empleadold INT,  
nombre VARCHAR(50),  
apellido VARCHAR(50),  
email VARCHAR(50),  
puesto VARCHAR(50),  
extension VARCHAR(10),  
PRIMARY KEY (empleadold)

);

CREATE TABLE pago (

pagold INT,  
formaPago VARCHAR(40),  
total NUMERIC(15.2),  
fechaPago DATE,  
PRIMARY KEY (pagold)

);

CREATE TABLE FacturaVenta (

pedidold INT,  
productold INT,  
oficinald INT,  
clienteld INT,



empleadold INT,  
pagold INT,  
totalVenta NUMERIC(15,2),  
fecha DATE,

CONSTRAINT FK\_fact\_ventaPedidold FOREIGN KEY (pedidold) REFERENCES  
pedido(id\_pedido),

CONSTRAINT FK\_fact\_ventaProductold FOREIGN KEY (productold) REFERENCES  
producto(codigoProducto),

CONSTRAINT FK\_fact\_ventaOficinald FOREIGN KEY (oficinald) REFERENCES  
oficina(oficinaID),

CONSTRAINT FK\_fact\_ventaClienteld FOREIGN KEY (clienteld) REFERENCES  
cliente(clienteld),

CONSTRAINT FK\_fact\_ventaPagold FOREIGN KEY (pagold) REFERENCES  
pago(pagold),

CONSTRAINT FK\_fact\_ventaEmpleadold FOREIGN KEY (empleadold) REFERENCES  
empleado(empleadold)

);

## Conclusiones

Para una tarea simple o necesitar información específica es beneficioso un sistema data mart; es como una minibiblioteca que en el caso de la jardinería permite contener datos específicos de la empresa, en este caso con el fin de identificar el producto más vendido, la categoría con más productos y el año con más ventas.

La implementación del data mart de facturación de ventas permitió a la jardinería almacenar la data de ventas, esto es fundamental para poder identificar los años con más ventas y tener un análisis; este sistema permite almacenar toda la información de forma consolidada.

Esto permite la eficiencia en el acceso a datos teniendo una estructura optimizada ( permitiendo un acceso rápido y eficiente a los datos), las dimensiones implementadas permite la comprensión del modelo; el diseño del modelo estrella nos permite un análisis y/o la capacidad de explorar los datos desde diversas perspectivas con el fin de obtener información valiosa para la toma de decisiones en la inteligencia de negocios, resaltando la importancia de mantener la integridad y la calidad de los datos garantizando la fiabilidad de los análisis realizados en el data mart.

La implementación de un data mart basado en el modelo estrella me proporciona una aplicación efectiva como herramienta para la inteligencia de negocios, permitiendo a las organizaciones anticipar y/o identificar oportunidades de mejora y toma de decisiones basadas en datos para optimizar el rendimiento empresarial.

## **Anexos**

Enlace del modelo estrella: [https://lucid.app/lucidchart/ed320200-aa47-44a0-a079-c94b5701f72a/edit?viewport\\_loc=1821%2C185%2C1973%2C1056%2C0\\_0&invitationId=inv\\_b42c10d2-24ac-4ab1-a613-624ff29454f5](https://lucid.app/lucidchart/ed320200-aa47-44a0-a079-c94b5701f72a/edit?viewport_loc=1821%2C185%2C1973%2C1056%2C0_0&invitationId=inv_b42c10d2-24ac-4ab1-a613-624ff29454f5)

## **Bibliografía**

Base de datos Jardinería

Iniciar sesión en Canvas. (s. f.).

<https://iudigital.instructure.com/courses/15580/files/4737073/download>

AWS. (s.f.). *¿ Que es un data mart?- Explicación de los data marts- AWS*. Obtenido de

<https://aws.amazon.com/es/what-is/data-mart/>

Zerpa, H., Garcia, R., & Izquierdo, H. (2020). Datamart Basado en el modelo estrella para la implementación de indicadores clave de desempeño como salida del big data.