

Modelo estrella de un Data Mart

Jefree Tamayo Herrera

Facultad de Ingeniería de Software y Datos, Institución Universitaria Digital De  
Antioquia

Antonio Jesús Valderrama

2 de septiembre del 2025

Un esquema de estrella es un tipo de esquema de base de datos relacional que consta de una sola tabla de hechos central redondeada de tablas de dimensiones. Las tablas dimensionales describen entidades empresariales (son las cosas que se modelan). Las entidades pueden incluir productos, personas, lugares y conceptos, incluido el propio tiempo. Una tabla de dimensiones contiene una columna de clave (o columnas) que actúan como identificador único y otras columnas. Otras columnas admiten la agrupación de los datos. Las tablas de hechos almacenan observaciones o eventos y pueden ser pedidos de venta, saldos de existencias, tipos de cambio, temperaturas, etc. Una tabla de hechos contiene columnas clave de dimensiones que se relacionan con las tablas de dimensiones y columnas de medida numéricas, las columnas de clave de dimensiones determinan la dimensionalidad de una tabla de hechos.

Esta técnica de modelado es mejor utilizarla cuando los usuarios tienen una comprensión clara de los datos requeridos, por ejemplo la cantidad de productos vendidos por estado, cuando los datos son estructurados y cuantitativos con algunos atributos categóricos, también cuando quieren los datos de manera rápida y sencilla, sin crear múltiples uniones, el rendimiento de las consultas es la máxima prioridad. Los esquemas de tipo estrella cuentan con diferentes ventajas frente a otros esquemas, como que estos esquemas son fáciles de entender e implementar. Su estructura desnormalizada reduce el número de uniones necesarias para recuperar datos. Esto simplifica y conduce a una agregación de datos y generación de informes mas rápidos. Los esquemas en estrella permiten un análisis de datos intuitivo y directo. Los usuarios pueden comprender fácilmente las relaciones y jerarquías entre las dimensiones. Aunque no todo es color de

rosa, también existen algunas desventajas de este modelo. La desnormalización puede causar redundancia de datos. Los atributos dimensionales suelen repetirse en varios registros dentro de una tabla de dimensionales, lo que puede provocar problemas de calidad de los datos. Dado que los datos se duplican en la normalización, los cambios frecuentes también pueden hacer que ciertas tablas muestren información desactualizada. También agregar datos redundantes aumenta los costos de computación y almacenamiento. Esto puede ser especialmente problemático cuando se manejan grandes conjuntos de datos.

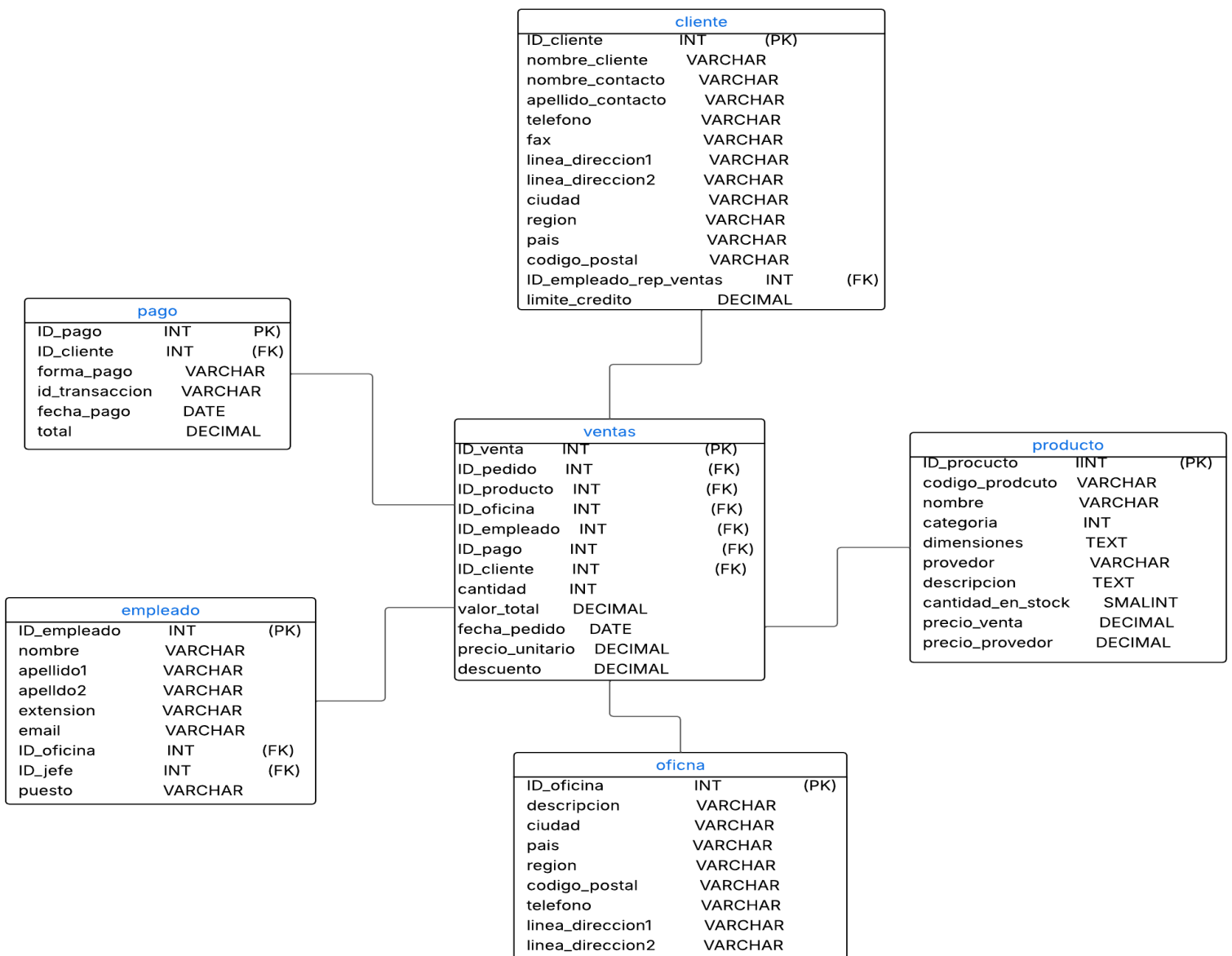
Dentro de nuestro estudio tenemos los siguientes objetivos

- Como primer objetivo tenemos es que identificar cual fue el producto mas vendido
- Como segundo objetivo tenemos ver la categoría con mas productos
- Y como tercer objetivo tenemos ver el año con mas ventas realizadas

En una jardinería enfrentan dificultades para gestionar y analizar la información de sus ventas debido a que los datos se encuentran distribuidos en diferentes tablas, esto no facilita en análisis de la información, a falta de un esquema de datos diseñado para el análisis limita la capacidad de la jardinería para identificar los patrones de consumo como identificar el producto más vendido, la categoría con mas productos y el año con mas ventas. Ante esta situación, surge la necesidad de crear e implementar un modelo de datos estrella creando una nueva tabla de ventas que concentre la información y organice las

dimensiones principales, de esta manera la jardinería podrá satisfacer las necesidades requeridas.

En este caso he diseñado un modelo de datos estrella donde la tabla de hechos lleva como nombre fecha y se conecta con las dimensiones más importantes de la base de datos como cliente, producto, empleado, oficina, pago, al ser un esquema de estrella, la tabla de hechos se encuentra en el medio y las tablas de dimensiones la rodean



Podemos apreciar que tenemos diferentes tablas de dimensiones, las mas importantes para llevar acabo nuestro análisis, tenemos la tabla de oficina que cuenta con sus diferentes atributos o características, también tenemos la tabla empleado, cliente, pago y producto y en la parte central del grafico podemos apreciar la tabla de hechos que creamos llamada ventas para facilitar la realización del análisis, esta tabla cuenta con lo id de las tablas de dimensiones y se le agregan más medidas de hechos como valor\_total que es de tipo DECIMAL, fecha\_pedido que es de tipo DATE, precio\_unitario que es de tipo DECIMAL y un descuento que también es de tipo DECIMAL

```
CREATE TABLE ventas (
  id_ventas INT AUTO_INCREMENT PRIMARY KEY,
  id_producto INT NOT NULL,
  id_cliente INT NOT NULL,
  id_tiempo DATE NOT NULL,
  cantidad INT NOT NULL,
  precio_unitario DECIMAL(10,2) NOT NULL,
  total_venta DECIMAL(10,2) NOT NULL,
  FOREIGN KEY (id_producto) REFERENCES producto(ID_producto),
  FOREIGN KEY (id_cliente) REFERENCES cliente(ID_cliente)
);
```

Aquí creamos en nuestra base de datos la tabla de hechos llamada venta, que tiene como datos los IDS de las tablas de dimensiones para que nos proporcione la información para satisfacer nuestros objetivos

```
INSERT INTO ventas (id_producto, id_cliente, id_tiempo, cantidad, precio_unitario, total_venta)
SELECT dp.ID_producto,
p.ID_cliente,
pe.fecha_pedido,
dp.cantidad,
dp.precio_unidad, (dp.cantidad * dp.precio_unidad) AS total_venta
FROM detalle_pedido dp
JOIN pedido pe ON dp.ID_pedido = pe.ID_pedido
JOIN cliente p ON p.ID_cliente = p.ID_cliente ;
```

Aquí poblanos nuestra nueva tabla de ventas con la información transaccional de las tabla de dimensiones de nuestro esquema de estrella

```
SELECT p.nombre, SUM(v.cantidad) AS total_vendido
FROM ventas v
JOIN producto p ON v.id_producto
GROUP BY p.nombre
ORDER BY total_vendido DESC
LIMIT 1;
```

Hacemos la consulta para saber cual es el producto mas vendido y obtenemos el siguiente resultado

Result Grid		Filter Rows:		Export:		Wrap Cell Content:		Fetch rows:	
	nombre	total_vendido							
▶	Cerezo	2152080							

Ahora pasamos a hacer la consulta para que cumplamos el segundo objetivo que es saber la categoría con mas productos

```
SELECT c.Desc_categoria, COUNT(p.id_producto) AS total_productos
FROM categoria_producto c
JOIN producto p ON p.Categoria
GROUP BY c.Desc_categoria
ORDER BY total_productos DESC
LIMIT 1;
```

Y obtenemos como resultado lo siguiente:

Result Grid	Filter Rows:	Export:	Wrap Cell Content:	Fetch rows:
Desc_categoria	total_productos			
Ornamentales	276			

Y por último hacemos la consulta para nuestro tercer objetivo que es saber cual fue el año con mas ventas

```
SELECT YEAR (v.id_tiempo) AS anio, SUM(v.total_venta) AS total_ventas
FROM ventas v
GROUP BY anio
ORDER BY total_ventas DESC
LIMIT 1;
```

Y obtenemos el siguiente resultado

Result Grid	Filter Rows:	Export:	Wrap Cell Content:	Fetch rows:
anio	total_ventas			
2008	4268088.00			

Así que podemos concluir que el producto que mas se vendió fue el cerezo con un total de 2150080 millones de unidades vendidas, también podemos decir que la categoría con mas productos, con un total de 276 productos es la categoría de ornamentales y por ultimo el año en que hubieron mas ventas fue en año 2008 con 4268088 millones de ventas.

**IBM (International Business Machines Corporation).** (s. f.). *Modelado dimensional: Esquemas de estrella*. En *InfoSphere Data Architect 9.1.2 documentation*. Recuperado de IBM Docs: <https://www.ibm.com/docs/es/ida/9.1.2?topic=schemas-star>

**Microsoft Learn.** (2024, diciembre 27). *Descripción de un esquema de estrella e importancia para Power BI*. Recuperado de Microsoft Learn: <https://learn.microsoft.com/es-es/power-bi/guidance/star-schema>

**Astera Marketing Team.** (2024, 25 de marzo). *¿Qué es el esquema de estrella? Ventajas y desventajas*. En *Astera Knowledge Center*. Recuperado de Astera: <https://www.astera.com/es/knowledge-center/star-schema/>