

Actividad: Creación de una base de datos de Staging

Estudiantes:

Miguel Ángel Piedrahita Gallego

Julián Esteban Valencia Gómez

Alejandro Muñoz Sierra

Bases de Datos II

(VICTOR HUGO MERCADO)

Ingeniería en Desarrollo de Software y datos

Universidad Digital de Antioquia

2025

Introducción

En el ámbito del análisis de datos, los modelos estrella son una estructura fundamental en los data marts para facilitar consultas eficientes y mejorar la toma de decisiones. En este contexto, se ha diseñado un modelo estrella basado en la base de datos **Jardinería**, con el propósito de consolidar información clave sobre productos, categorías y ventas. Este modelo permite una exploración estructurada y optimizada de los datos, proporcionando una visión clara del desempeño comercial de los productos y sus tendencias a lo largo del tiempo.

Objetivo

El objetivo principal de este modelo estrella es facilitar el análisis de ventas de la base de datos **Jardinería**, permitiendo responder tres preguntas clave:

1. **Identificar el producto más vendido**, proporcionando información valiosa sobre la demanda de los artículos.
2. **Determinar la categoría con mayor cantidad de productos**, lo que ayuda a comprender la diversificación del catálogo.
3. **Establecer el año con más ventas**, para evaluar tendencias y patrones de crecimiento en las ventas.

Este modelo proporciona una estructura de datos eficiente para la generación de informes estratégicos y la optimización de decisiones comerciales en el sector de la jardinería.

Planteamiento del problema

En la actualidad, las empresas que manejan un gran volumen de ventas requieren herramientas analíticas eficientes para tomar decisiones estratégicas basadas en datos. La empresa Jardinería, que gestiona la venta de productos relacionados con jardinería, enfrenta el reto de analizar sus datos transaccionales de manera óptima para identificar tendencias clave del negocio.

Actualmente, la base de datos Jardinería almacena información sobre clientes, pedidos, productos y pagos, pero estos datos se encuentran estructurados en un esquema transaccional que dificulta la generación de reportes eficientes. En particular, la empresa necesita responder tres preguntas clave para mejorar su estrategia comercial y de inventario:

¿Cuál es el producto más vendido?

¿Cuál es la categoría con más productos?

¿Cuál ha sido el año con más ventas?

Para solucionar este problema, se propone construir un modelo estrella de un Data Mart, donde los datos sean organizados de manera eficiente para facilitar su análisis. Este Data Mart permitirá la generación de reportes rápidos y precisos, mejorando la toma de decisiones en la empresa Jardinería.

Análisis del problema

Tomando como base el planteamiento del problema, se llevó a cabo un análisis detallado que permitió evaluar las tres preguntas planteadas. Estas interrogantes fueron fundamentales para el desarrollo del modelo estrella, proporcionando palabras claves de las cuales se pudieron abstraer las dimensiones y la tabla de hechos. A través de este enfoque, se logró estructurar un modelo de análisis que facilita la toma de decisiones estratégicas y optimiza la gestión comercial.

Descripción del modelo estrella propuesto

Tras analizar el modelo entidad-relación y las preguntas clave planteadas, llegamos a la conclusión de que la **tabla de hechos** se construye a partir de la unión entre las tablas "**Detalle_pedido**" y "**Pedido**". Consideramos que estas dos tablas son fundamentales para la solución, ya que contienen la información central sobre las transacciones realizadas.

Para responder la pregunta "**Producto más vendido**", identificamos que la tabla "**Producto**" es clave, ya que su atributo "**Codigo_producto**" permite rastrear los artículos registrados en cada pedido. Esto nos permite determinar con precisión cuáles son los productos con mayor demanda.

En cuanto a la pregunta "**Categoría con más productos**", analizamos la tabla "**Productos**", donde encontramos el atributo "**Categoría**". A partir de esta información, pudimos identificar la categoría con la mayor diversidad de artículos dentro del catálogo.

Finalmente, para responder la pregunta "**El año con más ventas**", observamos que en la **tabla de hechos** existe el atributo "**estado**", el cual indica si un pedido ha sido completado. Cuando el estado es "**true**", significa que el pedido fue pagado y entregado, mientras que un estado "**false**" indica que el pedido fue anulado, no pagado, devuelto o perdido. A partir de esta distinción, pudimos determinar el año con el mayor volumen de ventas efectivas.

Diseño del modelo estrella

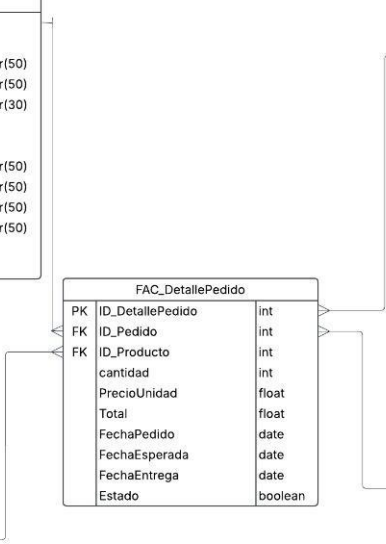
| DIM_Cliente | | |
|-------------|------------------------|-------------|
| PK | ID_Cliente | int |
| FK | ID_Empleado_Rep_Ventas | int |
| | Nombre_Cliente | Varchar(50) |
| | Nombre_contacto | Varchar(50) |
| | apellido_contacto | Varchar(30) |
| | Telefono | int |
| | Fax | int |
| | Linea_Direccion1 | Varchar(50) |
| | Ciudad | Varchar(50) |
| | Region | Varchar(50) |
| | Pais | Varchar(50) |
| | Codigo_postal | int |
| | Limite_Credito | int |

| DIM_Empleado | | |
|--------------|-------------|-------------|
| PK | ID_Empleado | int |
| FK | ID_Jefe | int |
| FK | ID_Oficina | int |
| | Nombre | Varchar(50) |
| | Apellido_1 | Varchar(20) |
| | Apellido_2 | Varchar(20) |
| | Extension | int |
| | Email | Varchar(50) |
| | Puesto | Varchar(25) |

| FAC_DetallePedido | | |
|-------------------|------------------|---------|
| PK | ID_DetallePedido | int |
| FK | ID_Pedido | int |
| FK | ID_Producto | int |
| | cantidad | int |
| | PrecioUnidad | float |
| | Total | float |
| | FechaPedido | date |
| | FechaEsperada | date |
| | FechaEntrega | date |
| | Estado | boolean |

| DIM_Tiempo | | |
|------------|--------------|------|
| PK | ID_DimTiempo | date |
| | Fecha | date |
| | Hora | time |

| DIM_Producto | | |
|--------------|-----------------|--------------|
| PK | ID_DimProducto | int |
| | CodigoProducto | int |
| | Nombre | varchar(20) |
| | Categoria | varchar(20) |
| | Proveedor | varchar(20) |
| | Descripcion | varchar(150) |
| | CantidadEnStock | int |
| | PrecioVenta | float |



Para la construcción del modelo estrella, se han de
cuales permitirán un análisis más detallado de las

- **Dimensión Tiempo:** Permite analizar las ventas en diferentes períodos (año, mes, día).
- **Dimensión Producto:** Facilita la identificación del producto más vendido a través de su código.
- **Dimensión Cliente:** Proporciona información sobre los clientes que realizan las compras, permitiendo analizar patrones de consumo y segmentación de mercado.
- **Dimensión Empleado:** Permite evaluar el desempeño de los empleados involucrados en la gestión de los pedidos y su impacto en las ventas.

Descripción del análisis realizado a la base de datos Jardinería y cómo estos se trasladaron a la base de datos Staging:

Se empieza realizando un análisis de nuestro modelo estrella, donde podemos evidenciar las dimensiones y, con base en esto, se comienzan a validar las tablas y los datos necesarios para su creación. En este caso, se inicia el proceso de extracción de producto, para lo cual necesitamos dos orígenes: uno que es producto y otro complementario, que es categoría producto.

Antes de iniciar este proceso, procedimos con las conexiones a la base de datos desde donde extraeremos la información y con la creación de la base de datos de staging.

Continuamos con la extracción de cliente, la cual realizamos como corrección de la primera entrega. En este caso, solo necesitaremos los datos y la tabla del cliente principalmente. De la misma manera, se realiza la extracción de empleado, donde tomamos un solo origen y un solo destino.

Seguimos con la extracción de tiempo, para lo cual realizamos un análisis de dónde se encuentran los datos necesarios para alimentar la dimensión. Nos damos cuenta de que esta información se encuentra en la tabla pedido, ya que allí se almacenan los datos de fecha pedido, fecha esperada y fecha de entrega.

Por último, realizamos la extracción de la fact, la cual es detalle pedido. Gracias a que ya logramos la extracción de las demás dimensiones, en este caso solo necesitaremos la tabla de origen detalle pedido, y con la ayuda de las demás tablas podremos finalizar nuestro proceso de extracción e iniciar el proceso de transformación.

Conclusión

El desarrollo de este **modelo estrella** para el **Data Mart** de la base de datos **Jardinería** permitió estructurar y optimizar el análisis de ventas mediante una organización eficiente de la información. A través de la unión de las tablas "**Detalle_pedido**" y "**Pedido**", se estableció una **tabla de hechos** que sirve como la base del análisis, facilitando la identificación de tendencias clave en las ventas.

Gracias a esta estructura, se puede responder de manera efectiva las tres preguntas fundamentales del estudio: **el producto más vendido, la categoría con mayor cantidad de productos y el año con más ventas**. La tabla "**Producto**" proporcionó la clave para determinar los artículos con mayor demanda, mientras que el atributo "**Categoría**" dentro de la tabla "**Productos**" permitió analizar la diversidad del catálogo. Asimismo, el atributo "**estado**" en la tabla de hechos nos permitió identificar los pedidos completados y, con ello, el año con el mayor volumen de ventas efectivas.

Además, la inclusión de las dimensiones **Cliente** y **Empleado** enriquece el análisis al proporcionar información adicional sobre los consumidores y el personal involucrado en el proceso de ventas. Esto permite realizar estudios más detallados sobre el comportamiento de los clientes y la eficiencia del equipo de trabajo.

En conclusión, el modelo diseñado no solo facilita la consulta y el análisis de datos, sino que también proporciona una base sólida para la toma de decisiones estratégicas en el sector de la jardinería. Su implementación permite una visión más clara del comportamiento de las ventas, apoyando la planificación comercial, la optimización del inventario y la mejora de la relación con los clientes y empleados.