

Modelo estrella de un Data Mart

Yusep Alexander Ruiz Agudelo

Ingeniería en Desarrollo y Datos

Trabajo realizado para

Bases de Datos II

Profesor

Victor Hugo Mercado

Institución Universitaria Digital de Antioquia

Facultad De Ingenierías

Medellín

2024

Tabla de Contenido

INTRODUCCIÓN	2
OBJETIVOS	4
OBJETIVO GENERAL	4
OBJETIVOS ESPECÍFICOS.....	4
PLANTEAMIENTO DEL PROBLEMA	5
ANÁLISIS DEL PROBLEMA	6
PROPUESTA DE LA SOLUCIÓN.....	7
DESCRIPCIÓN DEL MODELO ESTRELLA PROPUESTO.	7
<i>Tabla De Hechos</i>	7
<i>Tablas De Dimensiones</i>	8
DISEÑO (IMAGEN) DEL MODELO ESTRELLA	10
LISTA DE DIMENSIONES PROPUESTAS.....	11
TABLA DE HECHOS PROPUESTA	13
CONCLUSIONES	14
BIBLIOGRAFÍA	15

Introducción

“El esquema de estrella es un enfoque de modelado maduro ampliamente adoptado por los almacenes de datos relacionales. Requiere que los modeladores clasifiquen las tablas del modelo como dimensiones o hechos.” (Microsoft, 2023)

Una buena definición que se puede obtener de una tabla de hechos, de una forma clara según (OpenIA, 2024) sería:

una tabla central en un modelo de datos que contiene métricas o medidas cuantitativas relacionadas con los eventos de negocio. Estas tablas registran datos numéricos que son relevantes para el análisis, como ventas, pedidos o transacciones. Cada fila en una tabla de hechos representa un evento individual y contiene claves foráneas que se relacionan con las tablas de dimensiones, lo que permite segmentar y analizar los datos en diferentes contextos. Las tablas de hechos suelen tener un gran volumen de registros, ya que están diseñadas para capturar información a un nivel de detalle específico, conocido como granularidad.

(Microsoft, 2023) dice que las tablas denominadas dimensiones “describen entidades empresariales (las cosas que se modelan)”, a su vez (OpenIA, 2024) informa que “Estas tablas contienen atributos descriptivos que permiten categorizar y segmentar los datos de la tabla de hechos. Por ejemplo, una tabla de dimensiones puede incluir información sobre productos, clientes o fechas”.

Por medio de un análisis detallado de la base de datos de Jardinería, se identificará las relaciones que existen entre cada tabla y la relevancia de los datos que se encuentran en ellas. Posterior al análisis, se definirán los campos necesarios para la construcción de la tabla de

hechos y las dimensionales, proporcionando un diseño claro y estructurado para la construcción del modelo estrella.

Aplicar este modelo brindará la posibilidad de identificar el producto más vendido, la categoría con más productos y el año con más ventas, proporcionando una cantidad de información vital para la toma de decisiones y un mejor entendimiento del comportamiento del mercado.

Objetivos

Objetivo General

Elaborar un modelo estrella para un data mart que permita analizar y responder cuál es el producto más vendido, la categoría con más productos y el año con más ventas basado en la información de la base de datos de Jardinería.

Objetivos Específicos

- Analizar la estructura de la base de datos para determinar cómo será construida la tabla de hechos y las tablas de dimensiones para el modelo estrella.
- Definir los campos que son necesarios para la construcción de la tabla de hechos y dimensionales.
- Diseñar la tabla de hechos donde se evidencie las transacciones y ventas.
- Elaborar las tablas de dimensiones que sean relevantes para complementar la información de la tabla de hechos.

Planteamiento Del Problema

La base de datos de Jardinería contiene información sobre productos, clientes, pedidos, pagos y empleados, organizada en múltiples tablas interrelacionadas. Sin embargo, la estructura actual se encuentra en un modelo relacional normalizado, generando desafíos en el análisis de los datos. Se está presentando una dificultad a la hora de acceder y analizar la información de una forma eficiente, limitando la capacidad en la compañía para la toma de decisiones.

Para abordar esta problemática, es necesario implementar un modelo estrella que organice los datos con una estructura más accesible y a su vez analizable. Esto brindaría a la compañía la capacidad de simplificar el acceso a la información y a su vez obtener los resultados que desean como la identificación del producto más vendido, la categoría con más productos y el año con más ventas.

Análisis Del Problema

Teniendo en cuenta la necesidad de obtener las ventas o transacciones de la empresa, el análisis propuesto es el siguiente:

Se analizó la base de datos de Jardinería y se llega a la conclusión de que la tabla más importante es *detalle_pedido*, ya que contiene toda la información de las transacciones que se generaron, almacenando el número del pedido, el producto, unidades vendidas y precio. Esta tabla será considerada la tabla de hechos del ejercicio y se llamará *fac_ventas*.

Ahora bien, para realizar las tablas de dimensiones, se tomará en cuenta las siguientes:

- **dim_fecha:** Esta tabla se conformará de la columna *fecha_pedido* en la tabla *pedido*.
- **dim_geografia:** Permitirá almacenar la información geográfica.
- **dim_producto:** Contiene las características del producto y a su vez la descripción de la categoría.
- **dim_empleado:** Contiene la información básica del empleado junto con su jefe.
- **dim_oficina:** Contiene la descripción de la oficina.
- **dim_cliente:** Contiene la información básica del cliente.

Propuesta De La Solución

Descripción Del Modelo Estrella Propuesto.

Tabla De Hechos

fac_ventas: Esta tabla contendrá la información acerca de las ventas realizadas en la empresa.

Campos de la tabla de hechos:

- **id_fac_ventas** (PK)
- **id_dim_geo_empleado** (FK con la tabla dim_geografia)
- **id_dim_geo_cliente** (FK con la tabla dim_geografia)
- **id_dim_geo_oficina** (FK con la tabla dim_geografia)
- **id_dim_producto** (FK con la tabla dim_producto)
- **id_dim_fecha** (FK con la tabla dim_fecha)
- **id_dim_empleado** (FK con la tabla dim_empleado)
- **id_dim_cliente** (FK con la tabla dim_cliente)
- **id_dim_oficina** (FK con la tabla dim_oficina)
- **cantidad** (cantidad vendida)
- **precio_unidad** (precio por unidad)
- **valor_total** (valor total calculado)

Tablas De Dimensiones

Tabla de fecha: Contendrá las fechas en las que se realizaron las ventas.

Campos de la dimensión ***dim_fecha:***

- id_dim_fecha (PK)
- año
- mes
- día
- semestre
- num_semana

Tabla de geografía: Contendrá la información geográfica de los clientes, los empleados y la oficina.

Campos de la dimensión ***dim_geografia:***

- id_dim_geografia(PK)
- ciudad
- país
- región

Tabla de producto: Contendrá la información del nombre del producto, la descripción de la categoría y el precio.

Campos de la dimensión *dim_producto*:

- id_dim_producto (PK)
- nombre
- precio_venta
- desc_categoria

Tabla de empleado: Contendrá información del empleado.

Campos de la dimensión *dim_empleado*:

- id_dim_empleado (PK)
- nombre
- puesto
- nombre_jefe

Tabla de oficina: Contendrá las fechas en las que se realizaron las ventas.

Campos de la dimensión *dim_oficina*:

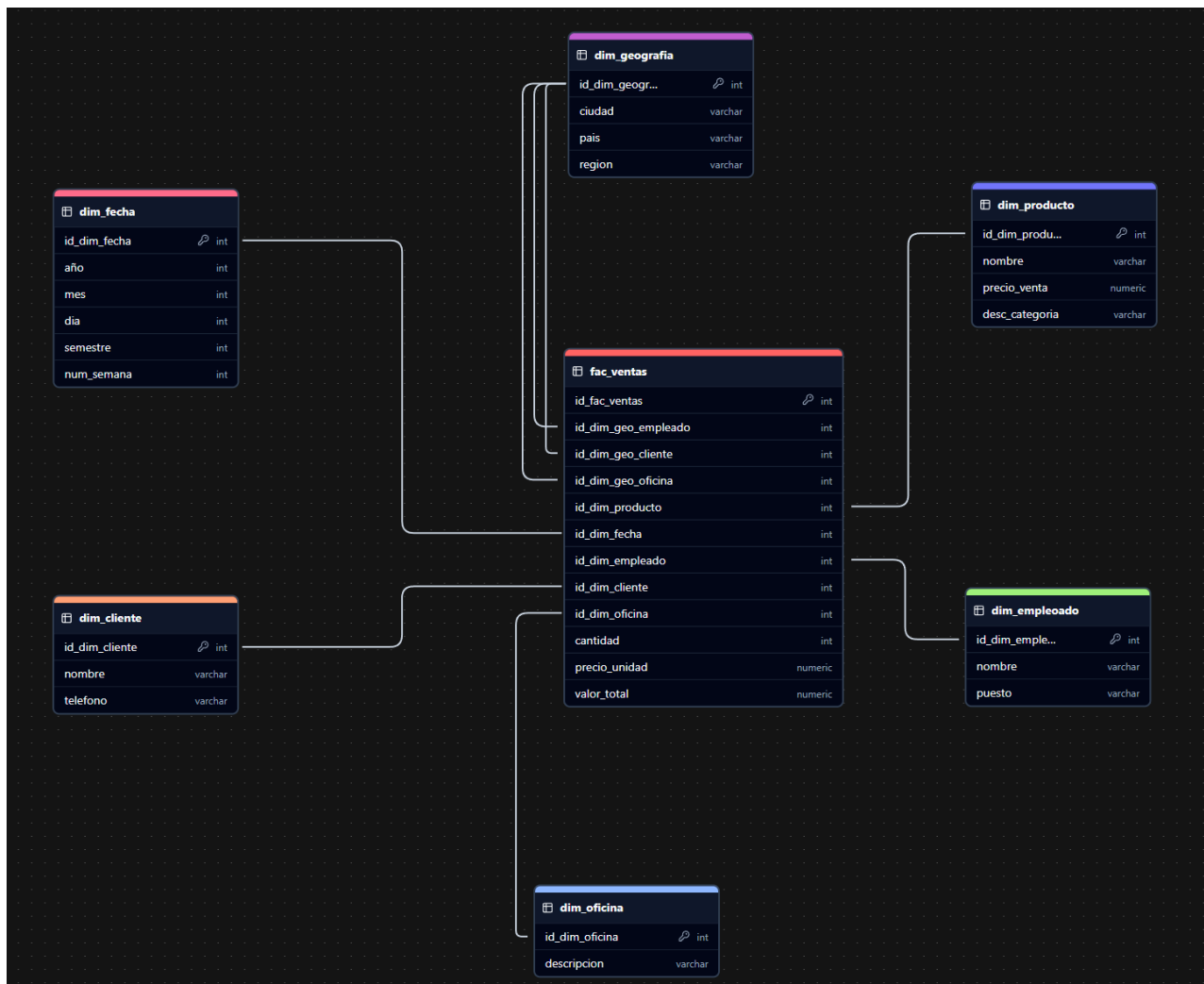
- id_dim_oficina (PK)
- descripción

Tabla de cliente: Contendrá la información del cliente.

Campos de la dimensión *dim_cliente*:

- id_dim_cliente (PK)
- nombre
- telefono

Diseño (Imagen) Del Modelo Estrella



Lista De Dimensiones Propuestas.

Tabla de fecha: Contendrá las fechas en las que se realizaron las ventas.

Campos de la dimensión *dim_fecha*:

- id_dim_fecha INT - PK
- año INT
- mes INT
- día INT
- semestre INT
- num_semana INT

Tabla de geografía: Contendrá la información geográfica de los clientes, los empleados y la oficina.

Campos de la dimensión *dim_geografia*:

- id_dim_geografia INT - PK
- ciudad VARCHAR(30)
- país VARCHAR(50)
- región VARCHAR(50)

Tabla de producto: Contendrá la información del nombre del producto, la descripción de la categoría y el precio.

Campos de la dimensión *dim_producto*:

- id_dim_producto PK
- nombre VARCHAR(70)
- precio_venta NUMERIC(15,2)
- desc_categoria VARCHAR(50)

Tabla de empleado: Contendrá información del empleado.

Campos de la dimensión *dim_empleado*:

- id_dim_empleado PK
- nombre VARCHAR(50)
- puesto VARCHAR(50)
- nombre_jefe VARCHAR(50)

Tabla de oficina: Contendrá las fechas en las que se realizaron las ventas.

Campos de la dimensión *dim_oficina*:

- id_dim_oficina PK
- descripción VARCHAR(10)

Tabla de cliente: Contendrá la información del cliente.

Campos de la dimensión ***dim_cliente***:

- id_dim_cliente PK
- nombre VARCHAR(50)
- teléfono VARCHAR(15)

Tabla De Hechos Propuesta

fac_ventas: Esta tabla contendrá la información acerca de las ventas realizadas en la empresa.

Campos de la tabla de hechos:

- id_fac_ventas INT - PK
- id_dim_geo_empleado INT - FK con la tabla dim_geografia
- id_dim_geo_cliente INT - FK con la tabla dim_geografia
- id_dim_geo_oficina INT - FK con la tabla dim_geografia
- id_dim_producto INT - FK con la tabla dim_producto
- id_dim_fecha INT - FK con la tabla dim_fecha
- id_dim_empleado INT - FK con la tabla dim_empleado
- id_dim_cliente INT - FK con la tabla dim_cliente
- id_dim_oficina INT - FK con la tabla dim_oficina
- cantidad INT
- precio_unidad NUMERIC(15,2)
- valor_total NUMERIC(15,2)

Conclusiones

El modelo estrella permite simplificar las consultas, ya que al tener menos tablas y menos relaciones no es necesario aplicar muchos JOINS al momento de realizar la sentencia sql para obtener la información deseada.

Las dimensiones permiten organizar fácilmente el análisis desde diferentes perspectivas, ya que contiene los datos organizados.

Los modelos de datos en estrella son una manera de organizar la información, utilizada en el análisis de datos, con una estructura simple, clara, conformada por una tabla de hechos y *n* cantidad de tablas denominadas dimensiones que aportan contexto a la tabla principal.

En comparación de los modelos relacionales, donde siempre suelen estar normalizados los datos para evitar redundancias, los modelos en estrella no siguen estas reglas, buscando un mejor rendimiento y comprensión.

Bibliografía

- Microsoft. (22 de Marzo de 2023). *Descripción de un esquema de estrella e importancia para Power BI*. Obtenido de Microsoft: <https://learn.microsoft.com/es-es/power-bi/guidance/star-schema>
- OpenIA. (01 de Septiembre de 2024). Obtenido de ChatGPT: <https://chatgpt.com/>