

S25 - EVIDENCIA DE APRENDIZAJE 2. BASE DE DATOS STAGING

EULICER ZAPATA ORREGO

DAWIN DE JESUS SALAZAR OVIEDO

ERNELEY POSSO

DOCENTE:

VICTOR HUGO MERCADO

INSTITUCIÓN UNIVERSITARIA DIGITAL DE ANTIOQUIA

BASES DE DATOS II

MEDELLÍN

2024-2

INTRODUCCIÓN

Este trabajo, se enfoca en la construcción de un modelo estrella para una base de datos, un enfoque clave en la creación de data marts que facilita la organización y el análisis de grandes volúmenes de datos. Utilizando nuestros conocimientos previos y con el apoyo de herramientas tecnológicas como Lucidchart, se desarrolla el modelo estrella que nos permite estructurar los datos de manera eficiente y comprensible. Este modelo no solo es fundamental para entender la dinámica de los datos, sino que también nos ayuda en la realización de análisis detallados y responder a preguntas clave del negocio, cómo identificar las categorías con más productos, determinar el año con mayores ventas, y reconocer los productos más vendidos, entre otros. Al construir y comprender este modelo, se logra una base sólida para la toma de decisiones informadas y estratégicas en la gestión de datos

OBJETIVOS

- Diseñar un modelo estrella para una data mart que nos permita la identificación de los datos relevantes para el negocio.
- Construir un modelo estrella que permita visualizar la información clave de manera clara y estructurada, apoyando el análisis y la toma de decisiones estratégicas para el negocio.
- Optimizar la estructura y accesibilidad de la información.

1. PLANTEAMIENTO DEL PROBLEMA

Actualmente, muchas organizaciones necesitan conocer las estadísticas y datos clave de su negocio para crear estrategias que les permitan mantener su competitividad en el mercado. Sin embargo, empresas como la que estamos analizando cuentan con bases de datos relacionales que, aunque almacenan y conservan la información adecuadamente, no son suficientemente efectivas para la extracción y análisis de datos relevantes. En el caso de la empresa de jardinería, surge la necesidad de identificar patrones clave que les permitan determinar cuáles son las categorías más vendidas, los años con mayores ventas y los productos más populares.

Para abordar este problema, se propone la construcción de un modelo estrella que organice la información de manera que facilite la obtención de respuestas a estas categorías. El modelo estrella es una estructura óptima para data marts, ya que simplifica la consulta de datos complejos y proporciona un marco claro y accesible para el análisis. Con el desarrollo de este modelo, se busca no solo mejorar la organización de la información, sino también permitir a la empresa tomar decisiones basadas en un análisis más preciso y comprensible. Posteriormente, estos conocimientos se utilizarán para optimizar sus operaciones y estrategias de ventas.

2. ANÁLISIS DEL PROBLEMA

La empresa de jardinería enfrenta un desafío común en muchas organizaciones que manejan grandes volúmenes de datos: la ausencia de una estructura de datos adecuada para realizar un análisis efectivo de la información y extraer datos valiosos para la toma de decisiones. Entre las causas subyacentes se encuentran:

Estructura de datos transaccional: La base de datos está diseñada principalmente para el registro de transacciones, lo que resulta óptimo para la gestión de datos operativos, pero inadecuado para el análisis estratégico.

Dificultad en la identificación de patrones: Sin un modelo estrella, la identificación de patrones, tendencias y relaciones se convierte en un proceso manual que consume numerosos recursos y tiempo.

Limitaciones en la toma de decisiones: La falta de una estructura de datos optimizada para el análisis afecta directamente la capacidad de la empresa para tomar decisiones informadas. Esto puede llevar a decisiones basadas en suposiciones o datos incompletos, lo que impacta negativamente en la competitividad y eficiencia de la organización.

3. PROPUESTA DE SOLUCIÓN

- 3.1 Descripción del modelo estrella propuesto.

Este modelo estrella consta de una tabla central de hechos y varias tablas de dimensiones que se conectan a esta tabla central, formando una estructura que facilita la consulta y el análisis de los datos.

FAC_HECHOS_VENTA

La tabla de hechos, denominada FAC_HECHOS_VENTA, es el núcleo del modelo estrella y donde se almacenarán los datos cuantitativos que se analizarán, como ventas, cantidades y precios. Esta tabla incluirá las siguientes claves foráneas que la relacionan con las dimensiones:

id_tiempo_fecha_pedido, id_tiempo_fecha_pago, id_tiempo_fecha_entrega:
Referencia a la dimensión de tiempo (DIM_TIEMPO), lo cual permite analizar las ventas en diferentes momentos y comparar el rendimiento a lo largo de diferentes periodos.

id_empleado: Referencia a la dimensión de empleados (DIM_EMPLEADO), proporcionando información sobre el personal involucrado en cada transacción.

id_cliente: Referencia las ventas con la dimensión de clientes (DIM_CLIENTE), lo cual es clave para identificar patrones de compra y segmentar a los clientes.

id_ubicacion_cliente, id_ubicacion_oficina: Referencias que conectan con la dimensión de ubicación (DIM_UBICACION), permitiendo análisis por región o área geográfica tanto del cliente como de las oficinas.

id_pedidos: referencia con la dimensión de pedidos (DIM_PEDIDOS), proporcionando un desglose detallado de los pedidos, productos involucrados y sus estados.

id_productos: referencia con la dimensión de productos (DIM PRODUCTOS), esencial para identificar los productos más vendidos y las categorías de productos con mayor desempeño.

id_oficina: Referencia con la dimensión de oficina (DIM_OFICINA) donde se permite saber en cual oficina se realizó la venta del producto.

Productos: Se refiere a uno o más productos que han sido vendidos.

Precio: Indica el precio de venta del producto facturado.

Estado: Describe la situación del pedido (puede ser rechazado, entregado o pendiente).

Tablas de Dimensiones:

DIM_TIEMPO: Contiene información detallada sobre el tiempo (hora, día, mes, año), lo cual permite realizar análisis temporales, como identificar el año con más ventas.

DIM_EMPLEADO: Incluye datos sobre los empleados, como nombres y otros atributos relevantes para identificar la contribución de cada empleado a las ventas.

DIM_CLIENTE: Almacena información sobre los clientes, permitiendo segmentaciones y análisis de patrones de compra.

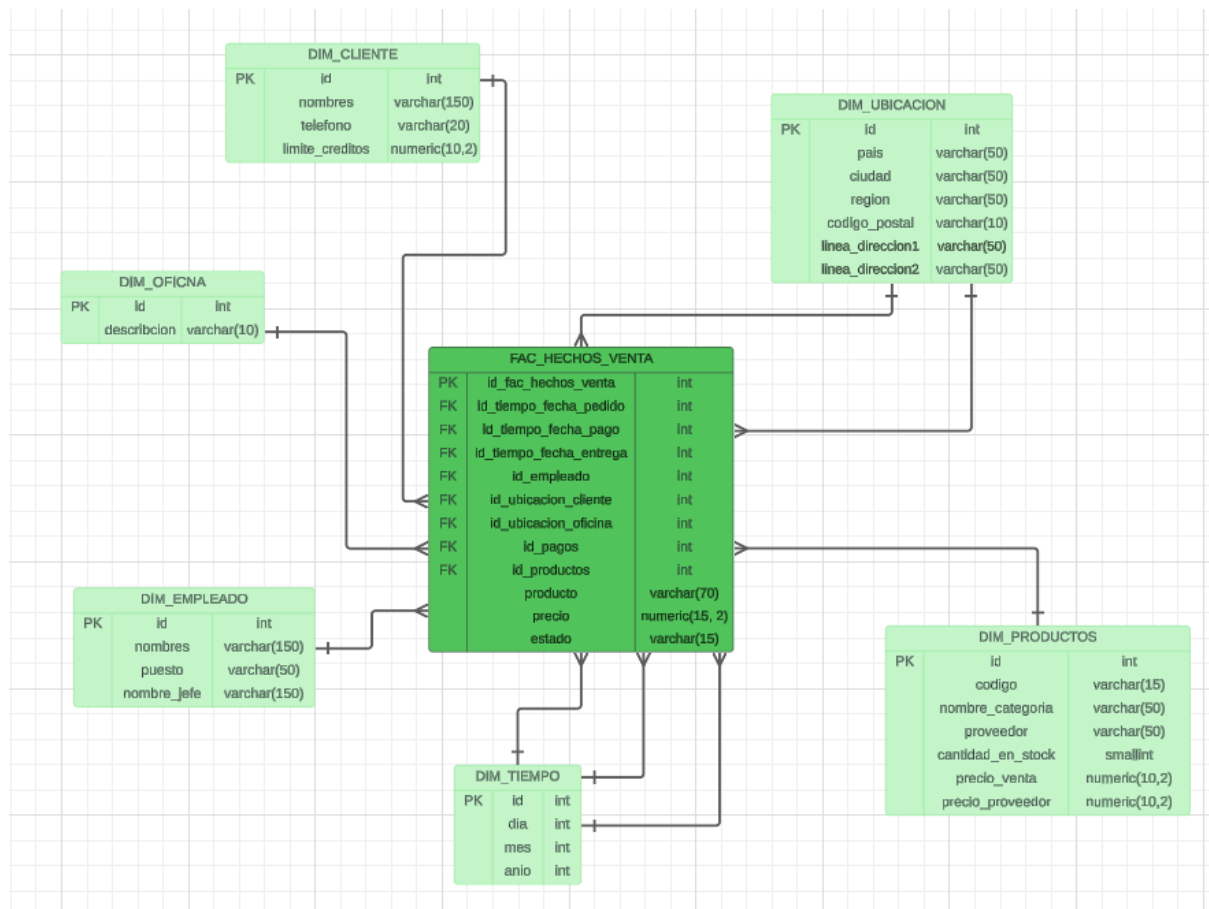
DIM_UBICACION: Proporciona datos geográficos que permiten analizar el rendimiento en diferentes regiones y ajustar estrategias locales.

DIM_PEDIDOS: Detalla los pedidos, productos incluidos, cantidades, precios y su estado, lo cual es esencial para evaluar el desempeño de los productos.

DIM_PRODUCTOS: Almacena información sobre los productos, incluyendo códigos, categorías, proveedores y precios. Esta dimensión es clave para identificar los productos y categorías con mayor éxito en ventas.

DIM_PRODUCTOS: Contiene información relevante sobre la oficina en la cual se hizo la venta, esta dimensión permite saber cual es la oficina en la cual se efectuó la venta.

- **3.2 Diseño (Imagen) del modelo estrella donde se puedan observar las dimensiones, la tabla de hechos, sus campos, tipos de datos y relaciones.**



- 3.3 Detalla la tabla de hechos, con sus campos y tipos de datos.

FAC_HECHOS_VENTA

id_fac_hechos_venta INT UNSIGNED AUTO_INCREMENT PRIMARY KEY,

id_tiempo_fecha_pedido INT NOT NULL,

id_tiempo_fecha_pago INT NOT NULL,

id_tiempo_fecha_entrega INT NOT NULL,

id_empleado INT, id_cliente INT NOT NULL,

id_ubicacion_cliente INT NOT NULL,

id_ubicacion_oficina INT NOT NULL,

id_pagos INT,

id_productos INT NOT NULL,

id_oficina INT NOT NULL,

producto VARCHAR(70) NOT NULL,

precio NUMERIC(15, 2) NOT NULL,

estado VARCHAR(15) NOT NULL;

FOREIGN KEY (id_tiempo_fecha_pedido) REFERENCES DIM_TIEMPO(id),

FOREIGN KEY (id_tiempo_fecha_pago) REFERENCES DIM_TIEMPO(id),

FOREIGN KEY (id_tiempo_fecha_entrega) REFERENCES DIM_TIEMPO(id),

FOREIGN KEY (id_empleado) REFERENCES DIM_EMPLEADO(id),

FOREIGN KEY (id_cliente) REFERENCES DIM_CLIENTE(id),

FOREIGN KEY (id_ubicacion_cliente) REFERENCES DIM_UBICACION(id),

FOREIGN KEY (id_ubicacion_oficina) REFERENCES DIM_UBICACION(id),

FOREIGN KEY (id_pagos) REFERENCES DIM_PAGOS(id),

FOREIGN KEY (id_productos) REFERENCES DIM_PRODUCTOS(id) ,

FOREIGN KEY (id_oficina) REFERENCES DIM_OFICINA(id) ;

- 3. 4 Lista de dimensiones propuestas.

Cada dimensión debe tener especificadas las columnas y qué tipos de datos van a almacenar.

DIM_TIEMPO

id INT UNSIGNED AUTO_INCREMENT PRIMARY KEY,

dia INT,

mes INT,

año INT;

trimestre INT;

DIM_EMPLEADO

id INT UNSIGNED AUTO_INCREMENT PRIMARY KEY,
nombres VARCHAR(150) NOT NULL ,
puesto VARCHAR(50) NOT NULL,
Nombre_jefe VARCHAR(150) NOT NULL;

DIM_CLIENTE (

id INT UNSIGNED AUTO_INCREMENT PRIMARY KEY,
nombres VARCHAR(150) NOT NULL,
telefono VARCHAR(20),
limite_credits NUMERIC(10, 2),

DIM_UBICACION

id INT UNSIGNED AUTO_INCREMENT PRIMARY KEY,
país VARCHAR(50),
ciudad VARCHAR(50),
region VARCHAR(50),
codigo_postal VARCHAR(10);
linea_direccion1 VARCHAR(50),
linea_direccion2 VARCHAR(50)

DIM_PRODUCTOS

id INT UNSIGNED AUTO_INCREMENT PRIMARY KEY,
codigo VARCHAR(15) NOT NULL,
nombre_categoria VARCHAR(50) NOT NULL,
proveedor VARCHAR(50) NOT NULL,
cantidad_en_stock SMALLINT NOT NULL,
precio_venta NUMERIC(10, 2) NOT NULL,
precio_proveedor NUMERIC(10, 2) NOT NULL;

DIM_OFICINA (

id INT UNSIGNED AUTO_INCREMENT PRIMARY KEY,
descripcion VARCHAR(10) NOT NULL,

4.DESCRIPCIÓN DEL ANÁLISIS REALIZADO A LOS DATOS *JARDINERÍA* Y CÓMO ESTOS SE TRASLADARON A LA BASE DE DATOS *STAGING*.

En el análisis de los datos, se seleccionaron campos clave de las tablas fuente en la base de datos de Jardinería, que fueron mapeados a las tablas de la base de datos staging. A continuación, se describe el traslado de los datos para cada una de las tablas involucradas:

Para la tabla *categoria_producto* se toman los siguientes datos:

Tabla en staging: *categoriaproductoST*

Se trasladaron los datos de esta tabla para categorizar los productos en el modelo estrella, lo cual es clave para agrupar los productos según sus categorías y facilitar el análisis de las ventas por categoría.

Campos trasladados:

- id: Se mantiene como identificador único para cada categoría.

- desc_Categoria: Este campo es necesario para asociar el nombre de la categoría en la dimensión productos.

Para la tabla *cliente* tomamos los siguientes datos:

Tabla en staging: clienteST

Los datos de los clientes son fundamentales para poder analizarlos como una dimensión en el modelo estrella, lo que permitirá segmentar y entender mejor el comportamiento de los clientes.

Campos trasladados:

- id: Se trasladó como identificador único de cada cliente.
- nombre_cliente: Este campo es necesario para la dimensión cliente.
- telefono: Se incluye para proporcionar información adicional sobre el cliente.
- linea_dirección1, linea_dirección2: Ambos campos se trasladaron para completar la información de ubicación del cliente, relevante para análisis geográficos.
- ciudad, región, país, código postal: Estos campos se movieron para proporcionar una vista completa de la ubicación del cliente.
- limite_creditos: Se trasladó para incorporar datos financieros relevantes sobre los clientes en la dimensión cliente.

Para la tabla *detalle_pedido* se toman los siguientes datos:

Tabla en staging: detallepedidoST

Esta tabla es crucial para conectar la información de productos, pedidos y precios en la tabla de hechos. Proporciona los detalles necesarios para entender qué productos fueron vendidos, en qué cantidad y a qué precio.

Campos trasladados:

- id_detalle_pedido: Identificador único para cada detalle de pedido, esencial para la integración con la tabla de hechos.
- id_pedido: Se trasladó para vincular los detalles del pedido con la tabla de hechos.
- id_producto: Se incluyó para identificar qué producto fue vendido, vinculándolo con la dimensión de productos.
- cantidad: Se trasladó para calcular el inventario y analizar la cantidad de productos vendidos.
- precio_unidad: Se pasó a la base de datos staging para poder calcular los ingresos y márgenes por producto.

Para la tabla EmpleadoST se toman los siguientes datos:

- La tabla EmpleadoST es fundamental para la creación de la dimensión de empleado, que proporciona la información clave sobre el personal que involucra la operación en los procesos de ventas y operaciones. Esta tabla captura los detalles necesarios para segmentar y analizar el rendimiento por empleado, como son el nombre, apellidos, puesto y la oficina a la que pertenecen. Estos datos permiten realizar análisis específicos por empleado, como son las ventas generadas o el desempeño por oficinas.

Campos Traslados:

- ID_empleado: Identificador único para cada empleado, indispensable para integrar la dimensión de empleado con la tabla de hechos.
- nombre: Se incluyó para identificar el empleado, proporcionando un dato importante en la identificación personal.
- apellido1 y apellido2: Los apellidos del empleado son importantes para el registro y segmentación de personas.
- puesto: Este campo se trasladó para facilitar el análisis del rol y cargo del empleado en las ventas u operaciones.
- ID_oficina: Se incluyó para vincular al empleado con la oficina correspondiente, integrando el análisis con la dimensión de la ubicación.

Para la tabla UbicacionST (Oficina)se toman los siguientes datos:

- La tabla UbicacionST (Oficina) permite la construcción de la dimensión de la ubicación, que es clave importante para analizar los datos de ventas por región o ubicación geográfica. Los datos de esta tabla ayudan a segmentar las transacciones y operaciones en función de las oficinas o sucursales donde se realizaron las ventas.

Campos Traslados:

- ID_oficina: Identificador único para cada oficina, necesario para conectar la ubicación con la tabla de hechos y otros análisis geográficos.
- Descripción: Se trasladó para proporcionar el nombre o descripción de la oficina, lo cual facilita el análisis por cada punto de venta o sede.
- ciudad: Campo importante para el análisis a nivel ciudad, permitiendo segmentar las ventas por ubicación geográfica.

- pais: Información necesaria para el análisis a nivel país o región, permitiendo desglosar los datos en niveles superiores de ubicación.
- región: Este campo fue incluido para facilitar el análisis regional lo que es útil cuando las oficinas se agrupan por zonas o regiones.

5. Para la tabla ProductosST se toman los siguientes datos:

Tabla en staging: ProductosST

Los productos son importantes en una dimensión en el modelo estrella, ya que es lo que genera el mayor movimiento en la tabla de hechos.

Campos trasladados:

- ID_producto
- nombre
- Categoria int
- dimensiones
- proveedor
- descripcion
- cantidad
- precio_venta
- precio_proveedor

Para la tabla PagosST se toman los siguientes datos:

Tabla en staging: PagosST

Los pagos reflejan las transacciones con los clientes y los pedidos.

Campos trasladados:

- ID_pago
- ID_cliente
- forma_pago
- id_transaccion

- fecha_pago date
- total

Para la tabla PedidoST se toman los siguientes datos:

Tabla en staging: PedidoST

Los pedidos son importantes porque también nos afectan la tabla de hechos y se relaciona con los productos y los clientes.

Campos trasladados:

- ID_pedido
- fecha_pedido date
- fecha_esperada date
- fecha_entrega date
- estado
- comentarios
- ID_cliente

En esta parte se verifica que los registros se cargaron correctamente en cada una de las tablas de la base de datos staging

-- Transferir datos de la tabla pago

```
INSERT INTO staging.PagosST (ID_cliente, forma_pago, id_transaccion, fecha_pago, total)
SELECT ID_cliente, forma_pago, id_transaccion, fecha_pago, total
FROM jardineria.pago;
```

-- Transferir datos de la tabla producto

```
INSERT INTO staging.ProductosST (ID_producto , nombre, Categoría, dimensiones,
proveedor, descripcion, cantidad_en_stock, precio_venta, precio_proveedor)
SELECT ID_producto , nombre, Categoria, dimensiones, proveedor, descripcion,
cantidad_en_stock, precio_venta, precio_proveedor
FROM jardineria.producto;
```

-- Transferir datos de la tabla pedido

```
INSERT INTO staging. PedidoST (fecha_pedido, fecha_esperada, fecha_entrega, estado,
comentarios, ID_cliente)
SELECT fecha_pedido, fecha_esperada, fecha_entrega, estado, comentarios, ID_cliente
FROM jardineria.pedido;
```

-- Transferir datos de la tabla empleado

```
INSERT INTO staging.empleado (ID_empleado, nombre, apellido1, puesto, ID_oficina)
SELECT ID_empleado, nombre, apellido1, puesto, ID_oficina FROM jardineria.empleado;
```

-- Transferir datos de la tabla Categoría_producto

```
INSERT INTO staging. categoriaproductoST (id, desc_Categoria)
```

```
SELECT id, desc_Categoria FROM jardineria.Categoria_producto;
```

-- Transferir datos de la tabla cliente

```
INSERT INTO staging.clienteST (id, nombre_cliente,  
telefono, fax, linea_direccion1, linea_direccion2, ciudad, region, pais,  
codigo_postal, limite_credito)
```

```
SELECT id, nombre_cliente, telefono, fax, linea_direccion1, linea_direccion2, ciudad,  
region, pais, codigo_postal, limite_credito FROM jardineria.cliente;
```

-- Transferir datos de la tabla pedido

```
INSERT INTO staging.detallepedidoST (id_detalle_pedido, id_pedido,  
id_producto, cantidad precio_unidad)
```

```
SELECT id_detalle_pedido, id_pedido, id_producto, cantidad precio_unidad FROM  
jardineria.pedido;
```

Aqui podemos verificar si los datos cargados coinciden con los datos en la tabla original

-- Verificar diferencias en la tabla productos

```
SELECT * FROM jardineria.productos EXCEPT SELECT * FROM staging.productosST
```

-- Verificar diferencias en la tabla empleados

```
SELECT * FROM jardineria.empleados EXCEPT SELECT * FROM staging.empleadoST
```


CONCLUSIONES

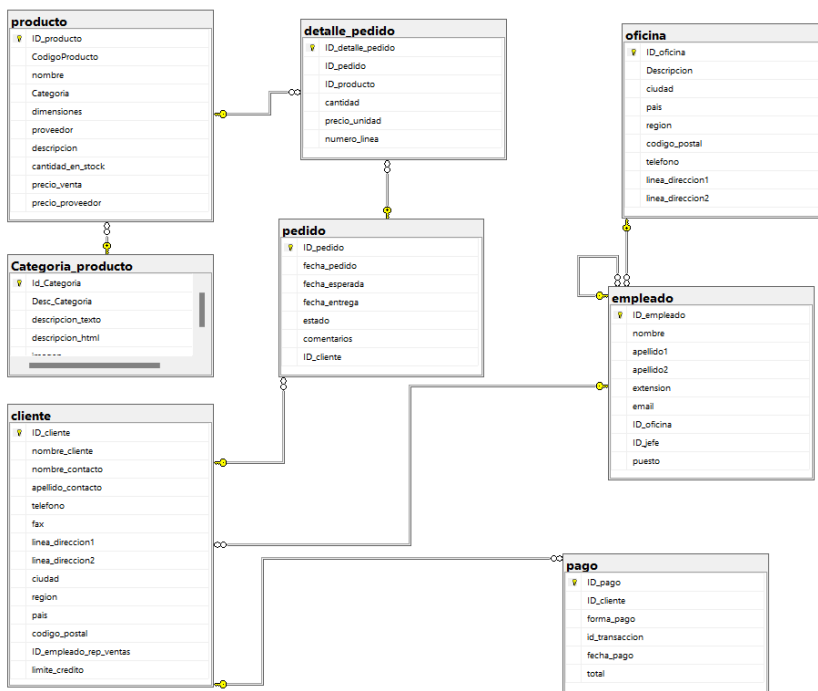
Desde un punto de vista técnico y de aprendizaje, la construcción del modelo estrella brinda una comprensión profunda de cómo se pueden optimizar los datos de una base de datos transaccional para su uso en análisis de inteligencia de negocios. Se ha aprendido a estructurar un modelo de datos que prioriza la eficiencia y la claridad, permitiendo ver de primera mano cómo una buena modelación de datos puede mejorar significativamente el rendimiento y la utilidad de una base de datos.

Además, este tipo de modelo nos ha mostrado la importancia de la organización y la estandarización de los datos, así como el valor de conectar diferentes aspectos a través de dimensiones claramente definidas. Este conocimiento no solo es crucial para equiparnos con habilidades valiosas para futuros desafíos en la gestión y análisis de datos.

El desarrollo del modelo estrella no solo nos proporciona una herramienta para transformar datos en un activo estratégico, sino que también permite consolidar nuestras habilidades en diseño de bases de datos y análisis de información, marcando un avance significativo en nuestro aprendizaje y preparación profesional.

ANEXOS :

1. Imagen del modelo relacional de la base de datos de jardinería.



2. Link del modelo estrella:

https://lucid.app/lucidchart/6b563a2b-a802-4d36-90e6-9fe7df160121/edit?viewport_1oc=-46%2C-367%2C3074%2C1338%2C0_0&invitationId=inv_fc88dbc4-5b91-43ca-b649-0e6f68059b12