

Actividad de Aprendizaje 2

Creación de una base de datos de Staging

**Asignatura
Bases de Datos II**

**GRUPO
POSICA2501B010004**

**PROFESOR
Víctor Hugo Mercado Ramos**

**ESTUDIANTE
Carlos Andrés Vega Marín**

**Institución Universitaria Digital de Antioquia
Facultad de Ingeniería y Ciencias Agropecuarias
Especialización en Analítica y Big Data
Medellín
2025**



INTRODUCCIÓN

El modelo estrella es una técnica para modelar y organizar los datos de un modelo entidad relación con el fin de permitir realizar consultas más rápidas y eficientes, al igual que el acceso a los datos cuando se realiza la construcción de tableros de visualización para la empresa.

Para este entregable trabajaremos con el modelo entidad relación de una base de datos llamada Jardinería, la cual transformaremos en un modelo dimensional en el cual se identifiquen las tablas de hechos y dimensiones para optimizar el análisis de datos y la generación de reportes.

También, se realizará la extracción de los datos de las tablas de la base de datos de Jardinería como primer paso de construcción de la ETL.

OBJETIVOS

- Diseñar un modelo estrella que permita realizar el análisis de ventas y transacciones en la base de datos Jardinería.
- Construir la tabla de hechos de tal forma que permita centralizar la información de las ventas.
- Identificar las dimensiones que se relacionarán con la tabla de hechos.
- Relacionar las dimensiones con la tabla de hechos que se defina.
- Crear una base de datos de Staging para alojar los datos extraídos de la base de datos Jardinería.
- Configurar las conexiones necesarias para la extracción de información de las tablas incluidas en el modelo estrella.

PLANTEAMIENTO DEL PROBLEMA

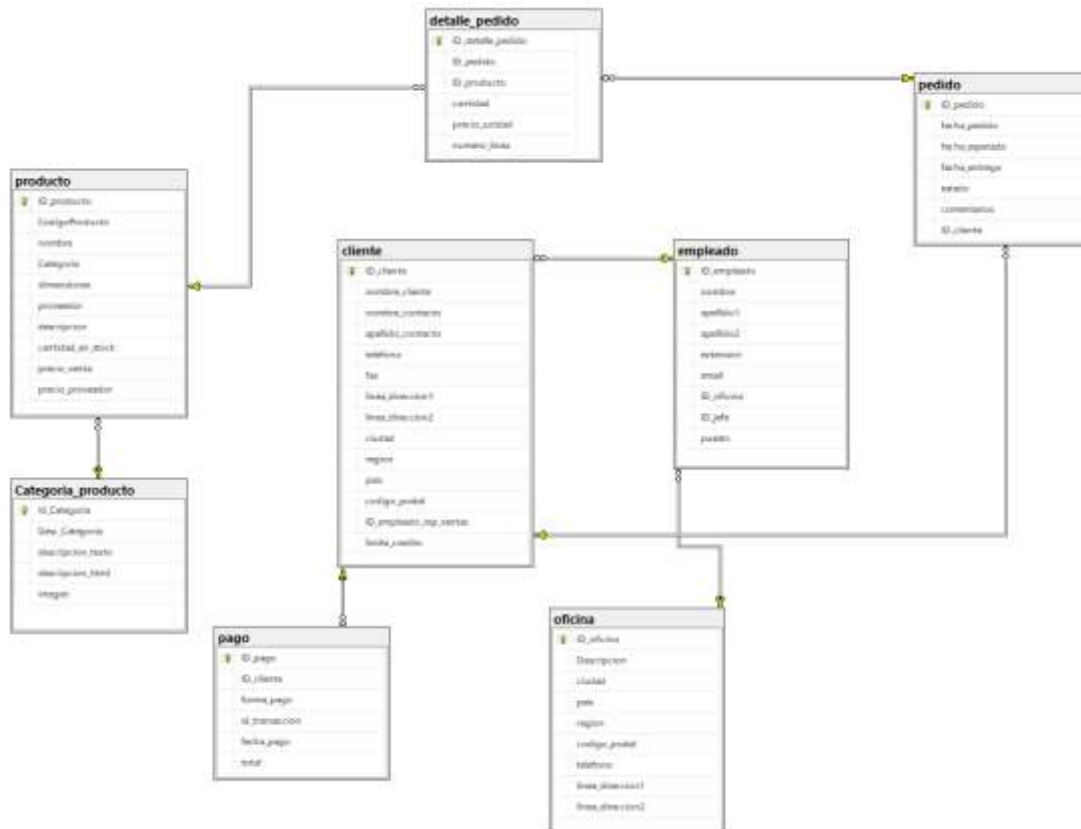
La base de datos Jardinería está diseñada sobre un modelo entidad relación cuyo objetivo es registrar las ventas, los clientes, los productos y los pagos del negocio. Sin embargo, para poder realizar un análisis y consulta eficiente de los datos se requiere de un modelo de datos más simplificado que permita ejecutar consultas que presente las ventas y transacciones con mayor rapidez y menor consumo de recursos.

Para resolver este problema, se propone diseñar un modelo dimensional tipo estrella, donde se definirán una tabla de hechos y varias tablas de dimensiones con sus respectivas medidas. Esta estructura permitirá mejorar la eficiencia de las consultas y facilitar la creación de tableros de control, optimizando así la toma de decisiones estratégicas.

Además, se requiere implementar un proceso ETL que extraiga los datos de la tabla de *Jardinería* para luego realizar las fases de transformación y carga. El primer paso será configurar las fuentes de origen y el destino de los datos, asegurando que se cuente con la información necesaria para completar el resto de los procesos de la ETL.

ANÁLISIS DEL PROBLEMA

La base de datos Jardinería contiene información clave sobre ventas, productos, categorías, oficinas, clientes, empleados y pagos. A partir del modelo entidad relación, podemos concluir lo siguiente:



- Ventas: registradas en las tablas pedido y detalle_pedido.
- Cliente: identificados en la tabla cliente.
- Producto: definidos en las tablas producto y categoria_producto.
- Empleado: relacionados con la tabla clientes y oficina.
- Pago: tiene relación con el cliente.

Limitación del modelo relacional actual: el objetivo principal del modelo entidad relación es organizar los datos de una forma normalizada al guardar la información, pero para realizar consultas y análisis de datos este modelo es lento y por tanto se debe trabajar sobre un modelo dimensional que permita optimizar recursos de procesamiento.

DESCRIPCIÓN MODELO ESTRELLA

Se propone el siguiente modelo estrella para el análisis de eficiente de los datos, a continuación se presenta la table de hechos y las dimensiones.

Tabla de Hechos: venta_transacciones

La tabla de hechos almacena las transacciones de las ventas realizadas.

| Variable | Tipo de Dato | Descripción |
|------------------|--------------|---|
| ID_venta (PK) | int | Identificador único de la venta |
| ID_pedido (FK) | int | Identificador del pedido realizado |
| ID_cliente (FK) | int | Relación con el cliente que realizó la compra |
| ID_producto (FK) | int | Relación con el producto vendido |
| ID_empleado (FK) | int | Relación con el empleado que realizó la venta |
| ID_pago (FK) | int | Relación con la forma de pago utilizada |
| fecha_venta | date | Fecha en que se realizó la venta |
| cantidad | int | Cantidad de productos vendidos |
| precio_unitario | decimal | Precio de cada producto vendido |
| total_venta | decimal | Monto total de la venta |

Tablas de Dimensiones: dim_tiempo – Dimensión de tiempo

Esta dimensión permite realizar análisis temporales de las ventas (diario, semanal, mensual, y anual).

| Variable | Tipo de Dato | Descripción |
|----------------|--------------|------------------------------------|
| ID_tiempo (PK) | int | Identificador único de la fecha |
| fecha_venta | date | Fecha específica de la transacción |
| año | int | Año en que se realizó la venta |
| mes | int | Mes en que se realizó la venta |
| semana | int | Número de semana del año |
| día | int | Día del mes |

dim_empleado – Dimensión de Empleado

| Variable | Tipo de Dato | Descripción |
|------------------|--------------|---|
| ID_empleado (PK) | int | Identificador único del empleado |
| nombre | varchar | Nombre del empleado |
| apellido1 | varchar | Primer apellido del empleado |
| apellido2 | varchar | Segundo apellido del empleado |
| extension | int | Número de extensión telefónica del empleado |
| email | varchar | Correo electrónico del empleado |
| ID_Jefe | int | Identificador del jefe del empleado |
| ID_Oficina | int | Identificador de la oficina donde trabaja el empleado |
| puesto | varchar | Puesto del empleado |

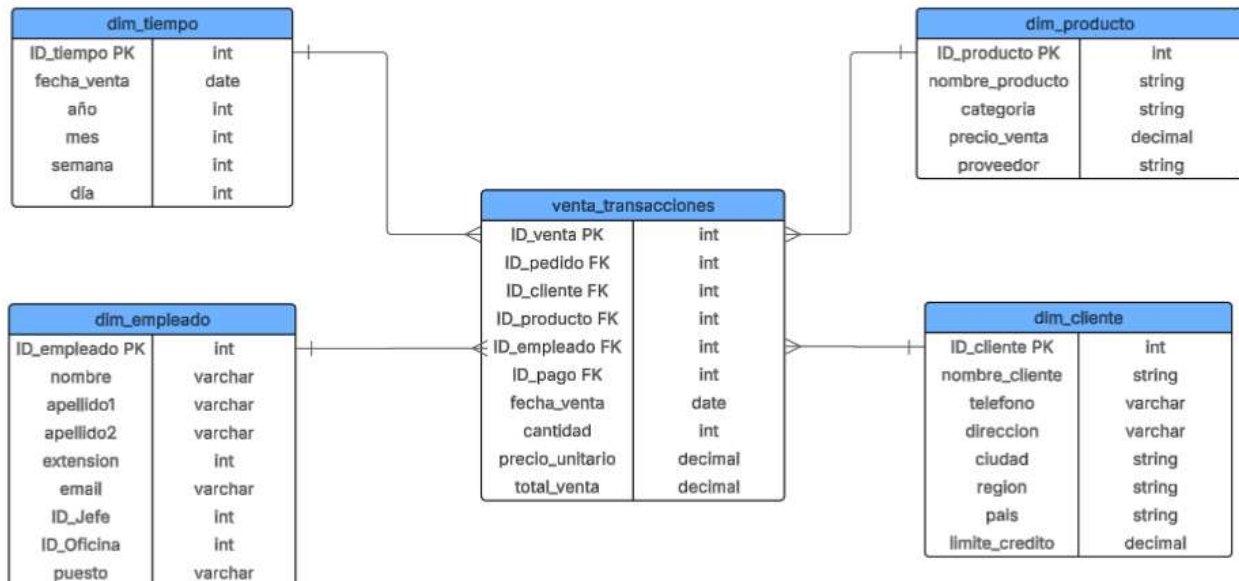
dim_cliente – Dimensión del Cliente

| Variable | Tipo de Dato | Descripción |
|-----------------|--------------|---------------------------------------|
| ID_cliente (PK) | int | Identificador único del cliente |
| nombre_cliente | varchar | Nombre del cliente |
| telefono | varchar | Teléfono del cliente |
| direccion | varchar | Dirección del cliente |
| ciudad | varchar | Ciudad de residencia del cliente |
| región | varchar | Región del cliente |
| país | varchar | País del cliente |
| limite_credito | int | Límite de crédito asignado al cliente |

dim_producto – Dimensión del Producto

| Variable | Tipo de Dato | Descripción |
|------------------|--------------|--|
| ID_producto (PK) | int | Identificador único del producto |
| nombre_producto | varchar | Nombre del producto |
| categoría | varchar | Categoría a la que pertenece el producto |
| precio_venta | int | Precio de venta del producto |
| proveedor | varchar | Proveedor del producto |

MODELO ESTRELLA IMPLEMENTADO



CONSULTAS SQL

Dimensión empleado

```
SELECT [ID_empleado]
      ,[nombre]
      ,[apellido1]
      ,[apellido2]
      ,[extension]
      ,[email]
      ,[ID_oficina]
      ,[ID_jefe]
      ,[puesto]
FROM [Jardineria].[dbo].[empleado]
```

```
CREATE TABLE "ST_Empleado" (
  "ID_empleado" int,
  "nombre" nvarchar(50),
  "apellido1" nvarchar(50),
  "apellido2" nvarchar(50),
  "extension" nvarchar(10),
  "email" nvarchar(100),
  "ID_oficina" int,
  "ID_jefe" int,
  "puesto" nvarchar(50))
```

Dimensión producto

```
SELECT [ID_producto]
      ,[nombre]
      ,[Categoria]
      ,[proveedor]
      ,[precio_venta]
FROM [Jardineria].[dbo].[producto]
```

```
CREATE TABLE "Destino_ST_Producto" (
  "ID_producto" int,
  "nombre" nvarchar(70),
  "Categoria" int,
  "proveedor" nvarchar(50),
  "precio_venta" numeric(15,2))
```


Dimensión Cliente

```
SELECT [ID_cliente]
      ,[nombre_cliente]
      ,[telefono]
      ,[linea_direccion1]
      ,[ciudad]
      ,[region]
      ,[pais]
      ,[limite_credito]
FROM [Jardineria].[dbo].[cliente]
```

```
CREATE TABLE "ST_Cliente" (
  "ID_cliente" int,
  "nombre_cliente" nvarchar(50),
  "telefono" nvarchar(15),
  "linea_direccion1" nvarchar(50),
  "ciudad" nvarchar(50),
  "region" nvarchar(50),
  "pais" nvarchar(50),
  "limite_credito" numeric(15,2))
```

Dimensión tiempo

```
SELECT DISTINCT [fecha_pedido]
FROM [Jardineria].[dbo].[pedido]
WHERE [fecha_pedido] IS NOT NULL
```

UNION ALL

```
SELECT DISTINCT [fecha_esperada]
FROM [Jardineria].[dbo].[pedido]
WHERE [fecha_esperada] IS NOT NULL
```

UNION ALL

```
SELECT DISTINCT [fecha_entrega]
FROM [Jardineria].[dbo].[pedido]
WHERE [fecha_entrega] IS NOT NULL
```

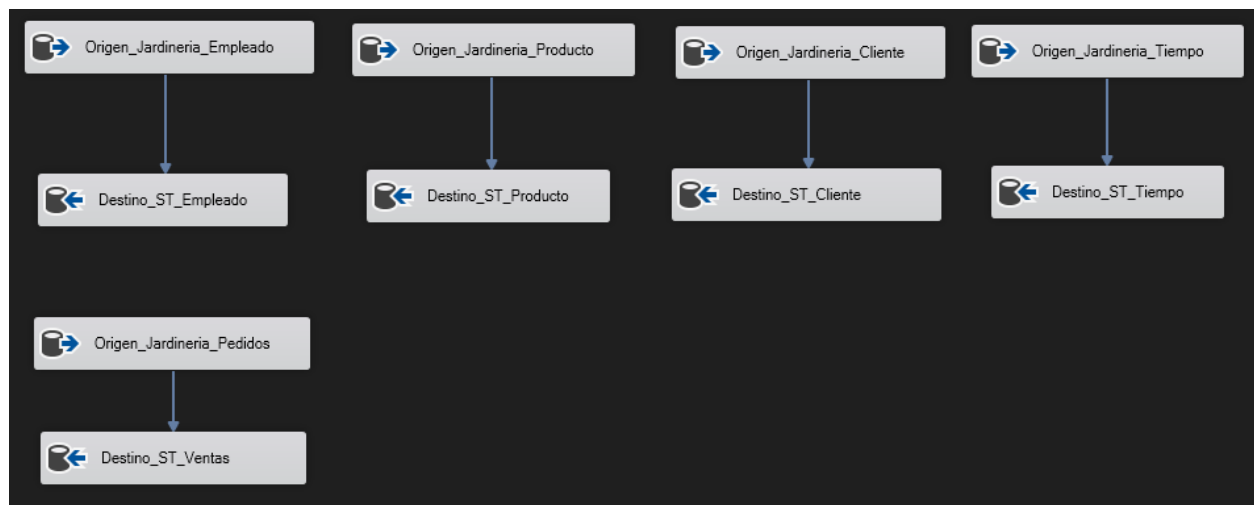
```
ORDER BY [fecha_pedido]
```

```
CREATE TABLE "ST_Tiempo" (
    "fechas_tbl_pedidos" date)
```

Dimensión Pedido

```
SELECT [ID_pedido]
    ,[fecha_pedido]
    ,[fecha_esperada]
    ,[fecha_entrega]
    ,[estado]
    ,[comentarios]
    ,[ID_cliente]
FROM [Jardineria].[dbo].[pedido]
```

```
CREATE TABLE "ST_Ventas" (
    "ID_pedido" int,
    "fecha_pedido" date,
    "fecha_esperada" date,
    "fecha_entrega" date,
    "estado" nvarchar(15),
    "comentarios" nvarchar(max),
    "ID_cliente" int)
```





Descripción del análisis.

Al ejecutar la solución, todos los orígenes de las tablas configurados en la base de datos de Jardinería se conectan con el proceso de destino en la base de datos de Staging. De esta manera, se lleva a cabo la transferencia de información correspondiente a la primera fase de la ETL, denominada extracción de datos.

CONCLUSIONES

- Transformar el modelo entidad relación de la base de datos de Jardinería a un modelo estrella optimizará la velocidad de consultas y análisis que se realicen.
- Con este tipo de modelos implementados se puede realizar de una forma más eficiente la conexión a una herramienta como Power BI para la construcción de tableros de visualización.
- Este modelo permitirá realizar consultas y análisis de las ventas y transacciones realizadas de una forma más simple y rápida.
- Para realizar la fase de extracción, primero se configura el origen de datos y luego se conecta con el destino.
- Para evitar duplicados al cargar la información, se lleva a cabo la instrucción **truncate table** antes de cada nueva carga. Dado que las tablas de origen contienen un volumen de datos significativo, este procedimiento no genera una sobrecarga de recursos.

ANEXOS

https://lucid.app/lucidchart/db974ff1-bbd6-4252-8eb5-43ff8aadb445/edit?viewport_loc=-8957%2C-2304%2C2992%2C1405%2C0_0&invitationId=inv_9aec7eaa-da42-4190-b98c-fee1a9d3d339

BIBLIOGRAFÍA.

Software Lucidchart - <https://www.lucidchart.com/pages/es>