



Nombre: Mauricio Bautista Rubiano
Carrera: Ingeniería de Software y Datos.
Grupo: PREICA2502B010064

Docente: Antonio Jesús Valderrama Jaramillo
Materia: Bases de Datos II

1. Objetivo del Proyecto

El objetivo de este trabajo era tomar la base de datos normal de la tienda de jardinería, que en lo que es un contexto hipotético académicamente hablando, se usa para las operaciones del día a día (guardar pedidos, clientes, etc.), y transformarla en un Modelo en Estrella. La idea es que este nuevo modelo sea mucho más fácil y rápido para hacer consultas de análisis, como ver qué producto se vende más o en qué mes tenemos más ventas.

Para lograr esto, utilicé un proceso conocido como ETL (Extract, Transform, Load).

2. Metodología: ¿Qué es el Proceso ETL?

El proceso ETL tiene tres grandes pasos que seguí para mover y acomodar los datos.

Extract (Extraer): Primero, saqué una copia de los datos que necesitaba de la base de datos original (jardineria). No trabajé directamente sobre la original para no arriesgarme a borrar o modificar algo por error. Esta copia la puse en una base de datos temporal que llamé jardineria_stg (la "stg" es de Staging, que es como una zona de preparación).

Transform (Transformar): Este fue el paso más importante. Aquí no solo limpié los datos, sino que cambié por completo la estructura. La base de datos original tiene muchas tablas relacionadas de formas complejas. La transformación consistió en crear la estructura del Modelo en Estrella. Este modelo es más simple:

- **Una Tabla de Hechos (Fact_Ventas):** Es el centro de la estrella. Aquí puse solo los números y las cosas que se pueden medir, como la cantidad de productos vendidos y el total en dinero.
- **Varias Tablas de Dimensiones (Dim_Cliente, Dim_Producto, etc.):** Son las puntas de la estrella. Estas tablas describen los hechos. Por ejemplo, Dim_Cliente dice quién compró, Dim_Producto dice qué se compró, y Dim_Tiempo dice cuándo se compró.

Load (Cargar): Finalmente, una vez que los datos estaban listos y la nueva estructura creada, cargué la información desde la base de datos de Staging (jardineria_stg) a la base de datos final, que llamé jardineria_dw (la "dw" es de Data Warehouse).

3. Etapas y Consultas SQL Realizadas

A continuación, detallo los pasos que seguí y el código SQL que utilicé.

Etapa 1: Creación de la Base de Datos de Staging (Extracción)

Primero creé una base de datos nueva para la preparación y copié la estructura de las tablas que iba a necesitar de la base de datos original.

Código:

```
-- Creación de la base de datos de Staging
CREATE DATABASE IF NOT EXISTS jardineria_stg;
USE jardineria_stg;

-- Copia de la estructura de las tablas necesarias
CREATE TABLE `stg_categoria_producto` LIKE `jardineria`.`categoria_producto`;
CREATE TABLE `stg_cliente` LIKE `jardineria`.`cliente`;
CREATE TABLE `stg_detalle_pedido` LIKE `jardineria`.`detalle_pedido`;
CREATE TABLE `stg_empleado` LIKE `jardineria`.`empleado`;
CREATE TABLE `stg_oficina` LIKE `jardineria`.`oficina`;
CREATE TABLE `stg_pedido` LIKE `jardineria`.`pedido`;
CREATE TABLE `stg_producto` LIKE `jardineria`.`producto`;
```

Etapa 2: Creación del Data Warehouse con el Modelo en Estrella (Transformación)

Aquí construí la estructura final. Creé las tablas de Dimensiones (las descripciones) y la tabla de Hechos (los números). A las tablas de dimensiones les puse un ID_..._SK que es una llave nueva que sirve para que el modelo funcione mejor.

```
-- Creación de la base de datos del Data Warehouse
CREATE DATABASE IF NOT EXISTS jardineria_dw;
USE jardineria_dw;

-- Creación de las Tablas de Dimensiones
CREATE TABLE Dim_Tiempo ( ID_Tiempo_SK INT NOT NULL PRIMARY KEY, ... );
CREATE TABLE Dim_Cliente ( ID_Cliente_SK INT NOT NULL AUTO_INCREMENT
PRIMARY KEY, ... );
CREATE TABLE Dim_Producto ( ID_Producto_SK INT NOT NULL AUTO_INCREMENT
PRIMARY KEY, ... );
CREATE TABLE Dim_Empleado ( ID_Empleado_SK INT NOT NULL AUTO_INCREMENT
PRIMARY KEY, ... );
```

```

CREATE TABLE Dim_Oficina ( ID_Oficina_SK INT NOT NULL AUTO_INCREMENT
PRIMARY KEY, ... );

-- Creación de la Tabla de Hechos
CREATE TABLE Fact_Ventas (
  ID_Venta_SK INT NOT NULL AUTO_INCREMENT PRIMARY KEY,
  ID_Tiempo_SK INT NOT NULL,
  ID_Cliente_SK INT NOT NULL,
  ID_Producto_SK INT NOT NULL,
  ID_Empleado_SK INT NOT NULL,
  ID_Oficina_SK INT NOT NULL,
  cantidad_vendida INT NOT NULL,
  precio_unidad DECIMAL(10, 2) NOT NULL,
  monto_total DECIMAL(20, 2) NOT NULL,

  -- Conexiones a las tablas de dimensiones
  CONSTRAINT `fk_ventas_tiempo` FOREIGN KEY (ID_Tiempo_SK) REFERENCES
Dim_Tiempo(ID_Tiempo_SK),
  ...
);

```

Etapa 3: Proceso de Carga de Datos (Load)

Este proceso lo dividí en dos partes:

1. Carga de Staging:

Llené las tablas de jardineria_stg con los datos de la base de datos original.

```

USE jardineria_stg;
-- Vacíé las tablas por si tenían algo
TRUNCATE TABLE stg_cliente;
...
-- Llené las tablas con los datos originales
INSERT INTO stg_cliente SELECT * FROM jardineria.cliente;

```

2. Carga del Data Warehouse:

Este es el paso final. Primero llené las tablas de Dimensiones y, por último, la tabla de Hechos, que conecta todo.

```
USE jardineria_dw;

-- Llenado de las Dimensiones (ejemplo con Dim_Cliente)
INSERT INTO Dim_Cliente (ID_Cliente_Original, nombre_cliente, ciudad, region, pais)
SELECT ID_cliente, nombre_cliente, ciudad, region, pais FROM
jardineria_stg.stg_cliente;

-- (Se repite un proceso similar para las otras dimensiones)

-- Llenado de la Tabla de Hechos
INSERT INTO Fact_Ventas (
    ID_Tiempo_SK, ID_Cliente_SK, ID_Producto_SK, ID_Empleado_SK, ID_Oficina_SK,
    cantidad_vendida, precio_unidad, monto_total
)
SELECT
    dt.ID_Tiempo_SK, dc.ID_Cliente_SK, dp.ID_Producto_SK, de.ID_Empleado_SK,
    do.ID_Oficina_SK,
    det.cantidad, det.precio_unidad, (det.cantidad * det.precio_unidad) AS monto_total
FROM
    jardineria_stg.stg_detalle_pedido AS det
    JOIN jardineria_stg.stg_pedido AS ped ON det.ID_pedido = ped.ID_pedido
    JOIN jardineria_stg.stg_cliente AS cli ON ped.ID_cliente = cli.ID_cliente
    JOIN jardineria_stg.stg_empleado AS emp ON cli.ID_empleado_rep_ventas =
emp.ID_empleado
    JOIN jardineria_stg.stg_oficina AS ofi ON emp.ID_oficina = ofi.ID_oficina
    JOIN jardineria_dw.Dim_Tiempo AS dt ON ped.fecha_pedido = dt.fecha
    JOIN jardineria_dw.Dim_Cliente AS dc ON cli.ID_cliente = dc.ID_Cliente_Original
    JOIN jardineria_dw.Dim_Producto AS dp ON det.ID_producto =
dp.ID_Producto_Original
    JOIN jardineria_dw.Dim_Empleado AS de ON emp.ID_empleado =
de.ID_Empleado_Original
    JOIN jardineria_dw.Dim_Oficina AS do ON ofi.ID_oficina = do.ID_Oficina_Original;
Nota: Durante este último paso, encontré un error de "collations". Lo solucioné
asegurándome de que todas las bases de datos usaran el mismo formato de texto
(utf8mb4_spanish_ci) para que se pudieran comparar correctamente.
```

4. Conclusión

El proceso se completó con éxito. Se extrajeron los datos de la base de datos original, se transformaron a un Modelo en Estrella y se cargaron en un nuevo Data Warehouse.

Ahora, la base de datos jardineria_dw queda lista para realizar consultas de análisis de datos de forma eficiente. Por ejemplo, si queremos saber el total de ventas por país, creo que la consulta es directa y rápida, uniendo la tabla de hechos con la dimensión de oficina, algo que en la base de datos original habría sido más complicado.