



IU Digital
de Antioquia



Evidencia de aprendizaje 3 y 4

Proceso de transformación de datos y carga en el data mart final

Proyecto integrador-Repositorio de todas las actividades

Maria Johana Arenas Usuga

Eduardo Enrique Guzmán Ledezma

PREICA2401B010076_3

Universidad Digital de Antioquia

Ingeniería de Software y Datos

Bases de Datos II (VICTOR HUGO MERCADO) - PREICA2401B010076

Mayo de 2024

Introducción

En las organizaciones de hoy, la capacidad para tomar decisiones informadas y de forma estratégica, se ha convertido en un factor determinante para su éxito y sostenibilidad. En este contexto, cada vez más el análisis de datos desempeña un papel crucial, permitiendo a las empresas extraer información valiosa para comprender mejor su desempeño, identificar oportunidades y anticipar tendencias. Para lograr esto, herramientas como los Data Mart son elementos fundamentales, proporcionando entornos especializados para el análisis de datos específicos y relevantes para áreas particulares de la empresa.

En las anteriores entregas se elaboró la construcción de un Modelo Estrella para un Data Mart y la creación de su base de datos Staging o zona de descarga como estadio intermedio para La base de datos "Jardinería" se selecciona como origen para este propósito, proporcionando datos de forma detallada que permite determinar ciertas categorías de información, tales como el producto más vendido, la categoría con más productos y el año con más ventas.

En el presente informe, se aborda la metodología para los procesos de extracción, transformación y carga de los datos de análisis del DataMart orientado a la gestión de la base de datos denominada "Jardinería".

El objetivo de este ejercicio es proporcionar una guía práctica para la construcción y utilización de un Modelo Estrella en un entorno de Data Mart, para la base de datos propuesta, destacando el potencial de la herramienta para analizar y responder a

consultas específicas que pueden llegar a apoyar una gestión efectiva de la información en un entorno empresarial.

Objetivos

Objetivo General

Diseñar un Modelo Estrella para un Data Mart basado en la base de datos "Jardinería", que permita el análisis y la respuesta a tres categorías específicas de consultas: la identificación del producto más vendido, la categoría con más productos y el año con más ventas.

Objetivos Específicos

1. Revisar la estructura de la base de datos Jardinería para identificar las tablas relevantes y sus relaciones, con los campos necesarios para construir el modelo estrella, incluyendo la tabla de hechos y las dimensiones pertinentes, de acuerdo a las categorizaciones propuestas.
2. Identificar y diseñar las dimensiones relevantes que se relacionarán con la tabla de hechos, asegurando la integridad referencial, de acuerdo a las categorizaciones propuestas.
3. Realizar la transformación de las dimensiones trabajadas en la base de datos Staging que son dimensión Cliente, dimensión Productos y dimensión Tiempo, relacionarla con la tabla de Hechos Fac Ventas y posteriormente la realizar la carga al Data Mart.
4. Ejecutar las consultas y validar que los datos queden almacenados correctamente en la Base de datos DMVentasJardineria.

Planteamiento del Problema

Se requiere realizar la transformación y posterior carga de la información base de datos denominada "Staging", la cual se ha creado a partir de la extracción de datos de la base de datos original llamada "Jardinería". El objetivo principal de este proyecto es transformar estos datos en un modelo de datos estrella y, posteriormente, crear un Data Mart de Ventas, que permita analizar y responder a tres categorías específicas en el contexto de la base de datos "Jardinería". Estas categorías incluyen la identificación del producto más vendido, la categoría con más productos y el año con más ventas. La complejidad de este problema radica en la necesidad de diseñar una estructura de datos que sea eficiente, escalable y capaz de proporcionar respuestas precisas a consultas complejas en un entorno empresarial dinámico y competitivo como el de la industria de la jardinería.

En general, en contextos administrativos y gerenciales, existe la necesidad de comprender y analizar diversos aspectos relacionados con los productos, las ventas y las tendencias del mercado. En ese sentido, identificar el producto más vendido, la categoría con más productos y el año con más ventas son consultas fundamentales que pueden proporcionar información valiosa para la toma de decisiones estratégicas, la planificación de inventario y la optimización de recursos.

Análisis del Problema

La base de datos “Staging” está compuesta por ocho tablas u entidades: 1. oficina, 2. empleado, 3. Categoría_producto, 4. cliente, 5. pedido, 6. producto, 7. detalle_pedido y 8. Pago. Por lo que se puede observar en la estructura de estas tablas, la base de datos almacena información relacionada con sus productos (tabla producto), descripciones y stocks y sus respectivas categorías (tabla categoría_producto), así como el registro de pedidos (tabla pedidos) con sus respectivos detalles (tabla detalle_pedido) donde dejan trazabilidad de cantidades, precios y números de línea; igualmente discriminan las oficinas donde se desarrollan sus actividades (tabla oficina) y sus empleados (tabla empleado). En lo relacionado con las ventas, por lo que se puede evidenciar con la información disponible, registran los pedidos relacionando información detallada y los respectivos pagos en tablas adicionales relacionadas, siendo la tabla pago la encargada de registrar la transacción monetaria y su respectiva forma de pago.

Entendiendo que se busca dar respuesta a tres interrogantes en particular para este ejercicio, la información relacionada con las oficinas donde se realizan los movimientos, así como los empleados que las ejecutan, no resultan relevantes para el diseño específico de una tabla de hechos específica para este Data Mart. Los pedidos y los detalles de los pedidos, se relacionan con los clientes que solicitan esos productos, así como las descripciones de las categorías se encuentran relacionadas en la tabla productos-

En consecuencia, se seleccionan las tablas que se relacionan directamente con las categorizaciones solicitadas para este modelo y que corresponden a: producto, cliente, tiempo, las cuales harán parte del modelo estrella que se presenta a continuación.

Propuesta de la solución.

Descripción del modelo estrella y base de datos “staging” propuesto.

Para el diseño de la solución propuesta, se tomaron en consideración las tres preguntas o categorías específicas señaladas para la solución del modelo: 1. identificar el producto más vendido, 2. la categoría con más productos y 3. el año con más ventas, utilizando la base de datos Jardinería como fuente de datos. En ese sentido, se asocian las entidades producto, cliente, pedido y detalle_pedido y se agrega la dimensión tiempo como forma de responder a la categorización del año con más ventas.

En lo relacionado con la categoría con más productos, se define en este caso, a aquellos cuyas ventas fueron realizadas mas no con las que se encuentran en stock en bodega y para el caso de determinar el año con más ventas, existe la posibilidad de establecer tanto las cantidades vendidas con las descripciones de dichos productos, así como los valores asociados a las ventas realizadas, por lo que con la misma tabla de hechos propuesta, se puede responder a ese requerimiento sea cual fuere el criterio de selección, a saber: cantidades o valores vendidos.

La base de datos Staging (***Staging_076_3***) que nos permitirá ejecutar los procesos de extracción y transformación para luego hacer la carga de los datos en el DataMart se construyó con la herramienta Visual Studio como una solución SQL Integration Server (***ETL_PREICA2401B010076_3_J***).

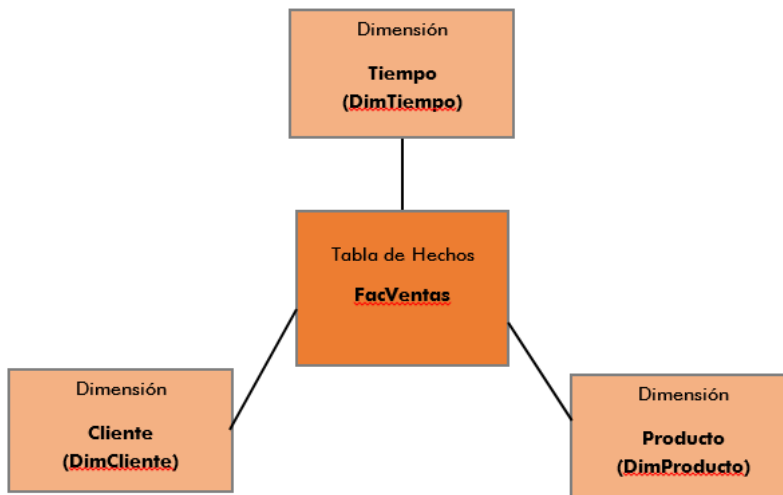


Fig. 1. Modelo estrella con dimensiones

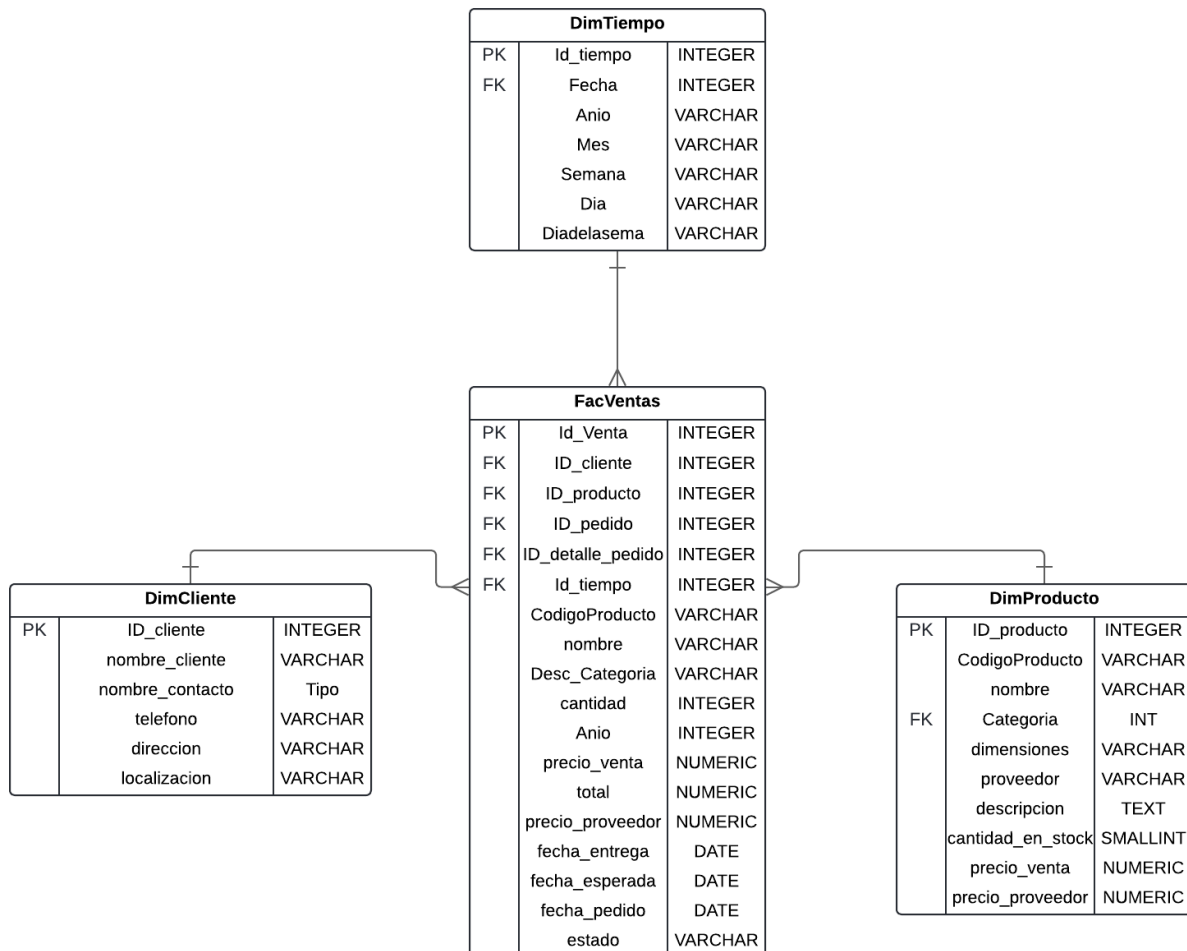


Fig. 2. Modelo estrella con dimensiones, tabla de hechos (FacVentas), campos, tipos de datos y relaciones.

https://lucid.app/lucidchart/914618e3-5d87-40a3-8207-2572e8a53175/edit?viewport_loc=-1888%2C-844%2C3328%2C1546%2C0_0&invitationId=inv_f4b1f866-1351-4c5c-989b-3159b9b1c83f

En relación con la base de datos “staging”, utilizando la herramienta Visual Studio como una solución SQL Integration Server (**ETL_PREICA2401B010076_3_J**), se analizaron los siguientes orígenes y destinos de datos, con el propósito de dar estructura a las dimensiones propuestas:

ETL_PREICA2401B010076_3_J

Extraccion_076_3_J

Origen	Destino	Descripción
Origen_Cliente_J	Destino_TbCliente	Datos de clientes
Origen_Producto_J	Destino_TbProducto	Datos de productos
Origen_Categoria_J	Destino_TbCategoria	Datos de detalle de categorías
Origen_Pedido_J	Destino_TbPedido	Datos de pedidos por cliente
Origen_DetalleP_J	Destino_TbDetalleP	Datos detallados de pedidos
Origen_Tiempo_J	Destino_TbTiempo	Datos de tiempo

Descripción de la transformación de la base de datos

“staging” y carga en el Data Mart Final “DMVentasJardineria”.

Base de datos Staging: Actúa como una zona de preparación donde se depositan los datos extraídos de la base original “Jardinería”. Aquí se realiza la limpieza de la información y las transformaciones de las dimensiones clientes, productos y tiempo, las cuales son esenciales para el análisis de la tabla de hechos FacVentas y así poder asegurar la calidad de los datos antes de integrarlos en el modelo estrella y cargarlos al Data Mart final.

En relación con la base de datos se toma como origen la información de la tabla “Staging_076_3”, en esta se hizo toda la transformación de los registros necesarios, luego se lleva la carga los registros en el Data Mart final llamado “DMVentasJardineria”,

utilizando la herramienta Visual Studio como una solución SQL Integration Server
(**ETL_PREICA2401B010076_3_J**).

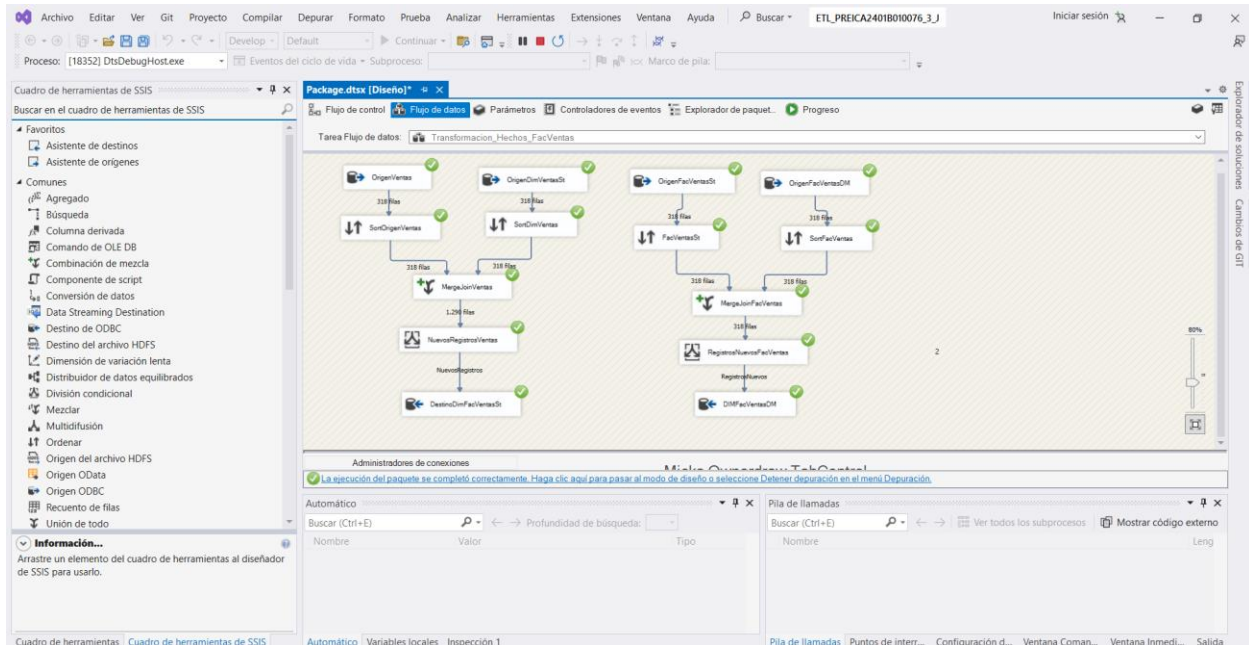


Fig.3. Transformación y carga de la tabla de hechos FacVentas.

```
USE [Staging_076_3]
GO

/***** Object: Table [dbo].[TbCliente]    Script Date: 25/05/2024 5:50:07 p. m. *****/
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[TbCliente](
    [ID_cliente] [int] IDENTITY(1,1) NOT NULL,
    [ID_cliente_o] [int] NULL,
    [nombre_cliente] [nvarchar](50) NULL,
    [nombre_contacto] [nvarchar](30) NULL,
    [telefono] [nvarchar](15) NULL,
    [linea_direccion1] [nvarchar](50) NULL,
    [linea_direccion2] [nvarchar](50) NULL,
    [ciudad] [nvarchar](50) NULL,
    [region] [nvarchar](50) NULL,
    [pais] [nvarchar](50) NULL,
    PRIMARY KEY CLUSTERED
    (
        [ID_cliente] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON, OPTIMIZE_FOR_SEQUENTIAL_KEY = OFF) ON [PRIMARY]
) ON [PRIMARY]
GO
```

Fig4. Script para creación de las tablas de cada una de las dimensiones (tiempo, cliente y producto).

```
SELECT TOP (1000) [ID_FacVentas]
,[ID_pedido_o]
,[ID_cliente_o]
,[ID_producto_o]
,[CodigoProducto]
,[cantidad]
,[precio_proveedor]
,[Precio_venta]
,[precio_unidad]
,[fecha_esperada]
,[Fecha_Entrega]
,[Estado]
,[ID_Tiempo]
FROM [staging_076_3].[dbo].[DimFacVentasSt]
```

110 %

Results Messages

ID_FacVentas	ID_pedido_o	ID_cliente_o	ID_producto_o	CodigoProducto	cantidad	precio_proveedor	Precio_venta	precio_unidad	fecha_esperada	Fecha_Entrega	Estado	ID_Tiempo	
256	256	97	36	73	FR-54	14	7.00	126.00	9.00	2007-04-24	2007-04-24	ENTREGADO	4
257	257	97	36	180	OR-156	10	8.00	100.00	10.00	2007-04-24	2007-04-24	ENTREGADO	4
258	258	98	36	50	FR-33	14	6.00	112.00	8.00	2008-03-30	2008-03-30	ENTREGADO	10
259	259	98	36	107	FR-85	6	56.00	420.00	70.00	2008-03-30	2008-03-30	ENTREGADO	10
260	260	98	36	101	FR-8	18	4.00	108.00	6.00	2008-03-30	2008-03-30	ENTREGADO	10
261	261	98	36	75	FR-56	16	6.00	128.00	8.00	2008-03-30	2008-03-30	ENTREGADO	10
262	262	98	36	80	FR-60	8	25.00	256.00	32.00	2008-03-30	2008-03-30	ENTREGADO	10
263	263	99	36	251	OR-227	30	51.00	1820.00	64.00	2009-04-06	2009-05-07	PENDIENTE	61
264	264	99	36	181	OR-157	15	8.00	150.00	10.00	2009-04-06	2009-05-07	PENDIENTE	61
265	265	100	36	87	FR-67	10	56.00	700.00	70.00	2008-11-09	2009-01-09	RECHAZADO	25
266	266	100	36	117	FR-94	40	25.00	360.00	9.00	2008-11-09	2009-01-09	RECHAZADO	25
267	267	100	36	109	FR-87	20	17.00	440.00	22.00	2008-11-09	2009-01-09	RECHAZADO	25
268	268	101	36	87	FR-67	10	56.00	700.00	70.00	2009-01-29	2009-01-31	ENTREGADO	48
269	269	101	36	13	AR-009	150	0.00	150.00	1.00	2009-01-29	2009-01-31	ENTREGADO	48
270	270	101	36	87	FR-67	10	56.00	700.00	70.00	2009-01-29	2009-01-31	ENTREGADO	48
271	271	101	36	10	AR-006	50	0.00	50.00	1.00	2009-01-29	2009-01-31	ENTREGADO	48
272	272	102	36	4	30310	23	11.00	278.00	12.00	2009-02-27	2009-02-27	PENDIENTE	30

Fig5. Tabla de hechos FacVentas consultada desde el Sql, origen Tabla Staging con carga de registros en Data Mart.

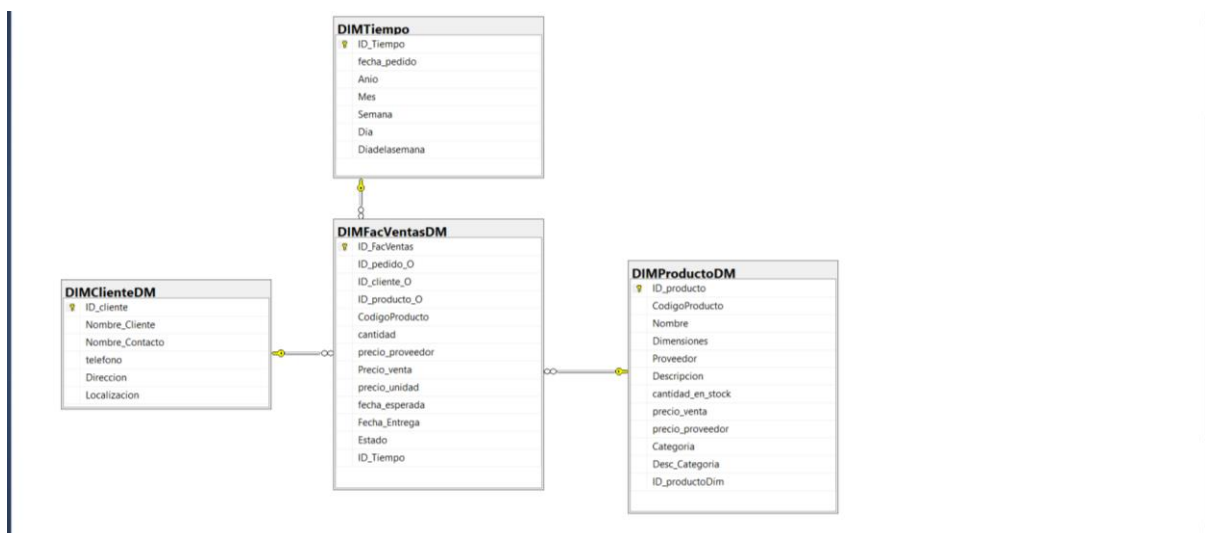


Fig6. Diagrama Data Mart Modelo Estrella.

Conclusiones

Como conclusiones generales del ejercicio, se puede destacar la importancia de entender los requerimientos que se exigen al momento de diseñar un Data Mart, toda vez que de esto depende completamente la selección de las entidades y los datos de dichas entidades que conformarán la tabla de hechos. En este caso en particular, la revisión de la estructura de la base de datos Jardinería para identificar las tablas relevantes y sus relaciones, de acuerdo a las categorizaciones propuestas determinó las dimensiones de la tabla de hechos. Un análisis parcial o equivocado puede llevar a la selección de información irrelevante o a omitir aquella que si responde a las solicitudes de categorización requeridas.

Por otro lado, se evidenció que la correcta construcción de la base de datos “Staging”, está directamente relacionada con una adecuada identificación de las dimensiones que se requieren, toda vez que los campos requeridos deben ser trasladados a los orígenes de datos de tal forma que respondan a los requerimientos exigidos para la construcción del DataMart.

El proceso de limpiar y transformar los datos de la base de datos Jardinería y su carga en un Data Mart basado en un modelo estrella ha sido un ejercicio integral de gestión de datos. Este esfuerzo ha mejorado significativamente la calidad de los datos y ha proporcionado una estructura robusta para el análisis de ventas, apoyando así las necesidades de información y análisis de la empresa. La metodología empleada y las herramientas utilizadas establecen una base sólida para futuras iniciativas de inteligencia de negocios y análisis de datos.