

Evidencia de Aprendizaje 3. Proceso ETL Jardinería

Dairo A. Ceballos y Flor L. Carmona

Institución Universitaria Digital de Antioquia

Especialización en Analítica y Big Data

Victor Hugo Mercado

15 de diciembre de 2024

Introducción

Teniendo en cuenta la necesidad de los negocios de sistemas de datos enfocados a la toma de decisiones basados en la información, para los datos de Jardinería se crea un Data Mart, luego se hace necesario el uso de tablas temporales mientras se prepara la información a las tablas finales de manera que los datos sean óptimos y finalmente se establece el Data Mart definitivo, el cual permita realizar consultas mediante la extracción, transformación y carga de los datos que tiene el negocio.

Objetivo

Completar el proceso ETL (Extracción, Carga y Transformación) de los datos previamente extraídos en la base de datos Staging, para crear el Data Mart final, el cual debe contener los datos consistentes.

Planteamiento del Problema

Para el proyecto Jardinería, es importante registrar la información de los pedidos, pagos, productos, categorías de productos, clientes, oficinas, empleados de la empresa de jardinería. Inicialmente se plantea un Data Mart que funcione para el planteamiento del negocio. Después de extraer los datos en una base de datos temporal (Staging), se continua con la transformación y carga de la información contenida en el Data Mart definitivo, necesario para realizar las consultas del negocio.

Análisis del Problema

Después de identificar de la información es relevante de Jardinería para migrarla a la base de datos Staging, de manera que permita enfocar las acciones a responder a las necesidades del

negocio, es necesario crear un Data Mart a partir de extracción, transformación y carga de los datos, teniendo en cuenta que los datos cargados cumplan las reglas de integridad y permita su posterior acceso a consultas.

Propuesta de Solución

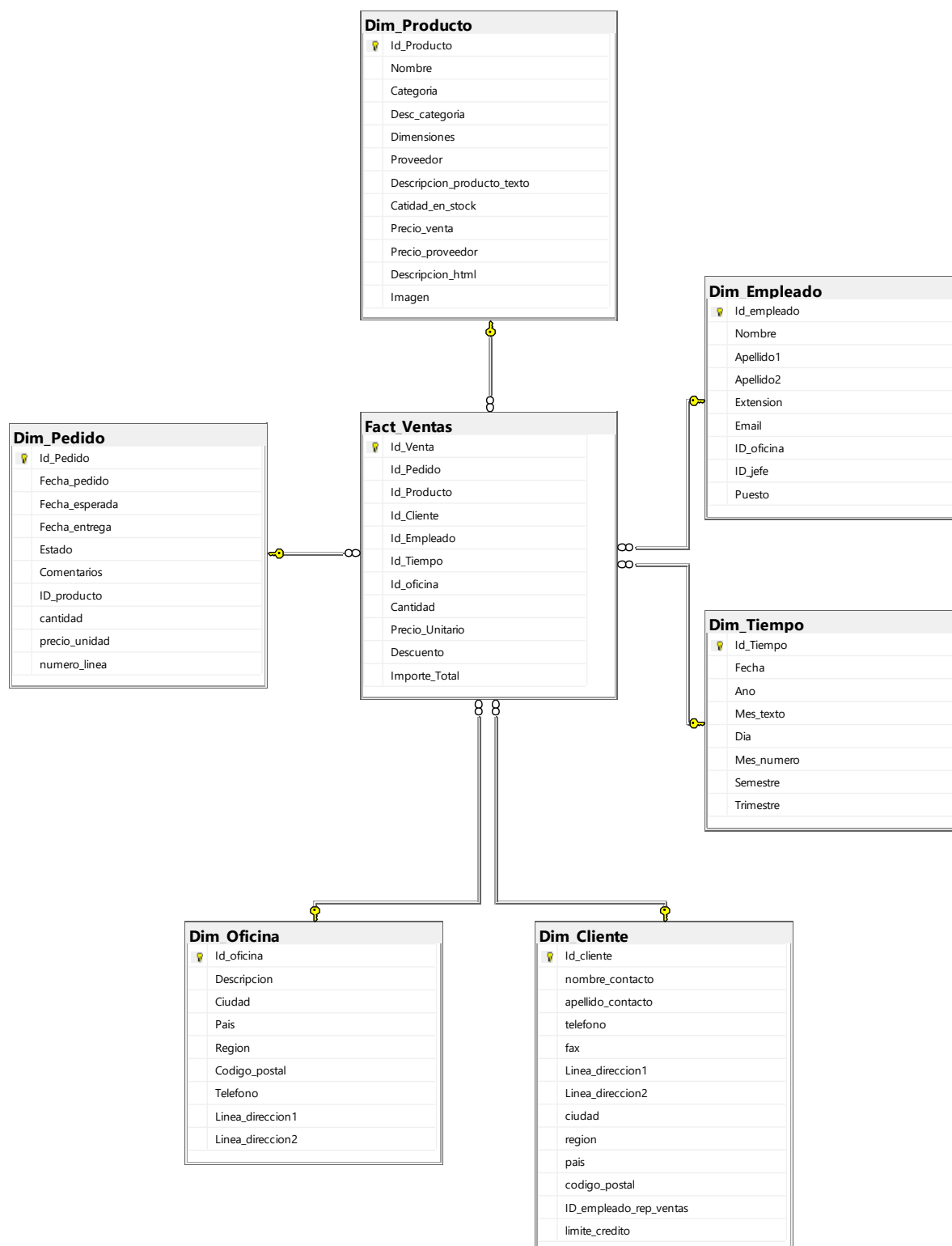
Modelo Estrella

Una vez realizadas las correcciones al modelo estrella, de acuerdo a las sugerencias realizadas a la Evidencia de Aprendizaje 1 se tiene la tabla hechos Venta y las tablas de dimensiones: Cliente, Empleado, Oficina, Tiempo, Pedido y Producto. Modelo mostrado en la imagen 1.

Se verifica que la información disponible satisfaga la necesidad del modelo planteado, de manera que no hayan campos vacíos en las dimensiones y se permitan referenciar las dimensiones con la tabla de hechos.

Imagen 1.

Modelo Estrella



Extracción a Staging

Se realizan consultas SQL para la creación de una base de datos temporal (Staging) en el proceso de extracción de los datos, además de la verificación de integridad de estos.

Imagen 2.

Creación de Staging

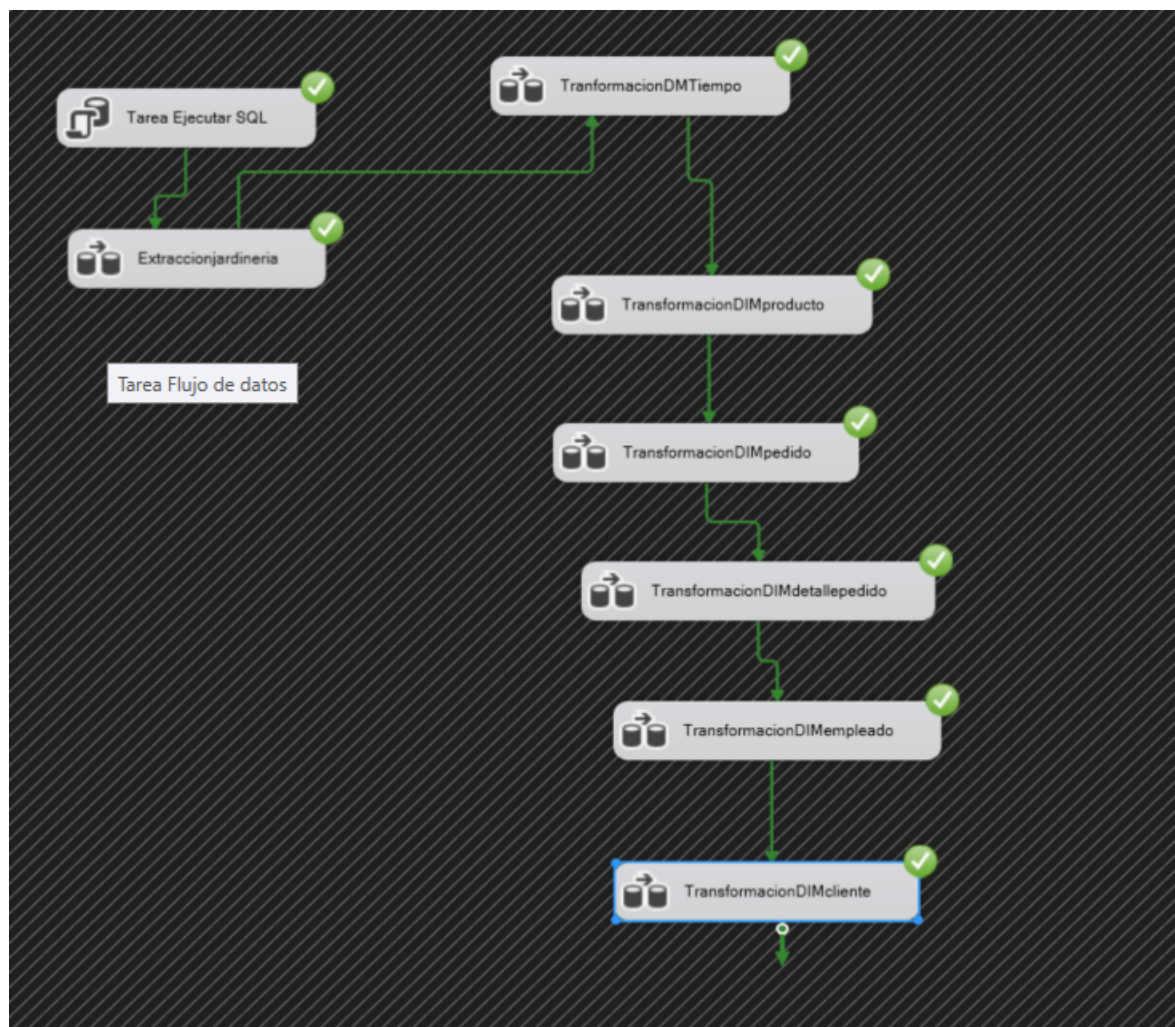


Transformación

Se realiza la revisión de los datos que van a llegar a la base de datos definitiva y se realizan las transformaciones requeridas, de forma que los datos cargados sean consistentes.

Imagen 3

Proceso de construcción Data Mart



Carga de datos Data Mart

Se realiza la carga al Data Mart, que será la base de datos definitiva siguiendo el código:

Creacion de DIMproductoADONET

```
SELECT [ID_producto],
```

```
[CodigoProducto],
```

```
,[nombre]
```

```
,p.[Categoria]as nameCategoria
```

,[dimensiones]

,[proveedor]

,[descripcion]

,[cantidad_en_stock]

,[precio_venta]

,[precio_proveedor]

FROM [jardineriaADO_NET].[dbo].[productoADO_NET] as p left join
[Categoria_productoADO_NET] as ps on p.Categoria=ps.Id_Categoria

DIMpedidoADONET

SELECT [ID_pedido]

,[fecha_pedido]

,[fecha_esperada]

,[fecha_entrega]

,[estado]

,[comentarios]

,c.[ID_cliente] as namecliente

FROM [jardineriaADO_NET].[dbo].[DestinopedidoADO_NET] as c left join
[clienteADO_NET] as cs on c.iD_cliente=cs.ID_cliente

DIMdetallepedido

```
SELECT [ID_detalle_pedido]

,[ID_pedido]

,d.[ID_producto] as codeproducto

,[cantidad]

,[precio_unidad]

,[numero_linea]

FROM [jardineriaADO_NET].[dbo].[detalle_pedidoADO_NET] as d left join
[productoADO_NET] as ds on d.ID_producto=ds.ID_producto
```

Transformacion empleado

```
SELECT [ID_empleado]

,[nombre]

,[apellido1]

,[apellido2]

,[extension]

,[email]

,o.[ID_oficina] as codeoficina

,[ID_jefe]

,[puesto]
```



```
FROM [jardineriaADO_NET].[dbo].[empleadoADO_NET] as o left join [oficinaADO_NET]
as os on o.ID_oficina=os.ID_oficina
```

Trasnformacion cliente

```
SELECT [ID_cliente]
```

```
,[nombre_cliente]
```

```
,[nombre_contacto]
```

```
,[apellido_contacto]
```

```
,[telefono]
```

```
,[fax]
```

```
,[linea_direccion1]
```

```
,[linea_direccion2]
```

```
,[ciudad]
```

```
,[region]
```

```
,[pais]
```

```
,[codigo_postal]
```

```
,[ID_empleado_rep_ventas]
```

```
,[limite_credito]
```

```
FROM [jardineriaADO_NET].[dbo].[clienteADO_NET] as e left join [empleadoADO_NET] as
es on e.ID_empleado_rep_ventas=es.ID_empleado
```

Conclusiones

La base de datos Staging, creada para almacenar de manera temporal los datos, es fundamental para la creación de una base definitiva, una vez completado el proceso de transformación y carga de los datos.

El Data Mart definitivo no solo satisface el modelo estrella planteado, sino que garantiza la integridad de los datos, una vez realizadas las transformaciones requeridas. Además, permite la ejecución de consultas necesarias para el negocio de Jardinería.

Anexos

Script Construcción del Modelo SQL

```
CREATE DATABASE JardineriaST
```

```
use JardineriaST
```

```
CREATE TABLE [Dim_Oficina] (
```

```
    [Id_oficina] INT,
```

```
    [Descripcion] VARCHAR(50),
```

```
    [Ciudad] VARCHAR(50),
```

```
    [Pais] VARCHAR(50),
```

```
    [Region] VARCHAR(50),
```

```
    [Codigo_postal] INT,
```

```
    [Telefono] INT,
```

```
[Linea_direccion1] VARCHAR(50),  
  
[Linea_direccion2] VARCHAR(50),  
  
PRIMARY KEY ([Id_oficina]));  
  
CREATE TABLE [Dim_Empleado] (  
  
[Id_empleado] INT,  
  
[Nombre] VARCHAR(50),  
  
[Apellido1] VARCHAR(50),  
  
[Apellido2] VARCHAR(50),  
  
[Extension] INT,  
  
[Email] VARCHAR(50),  
  
[ID_oficina] INT,  
  
[ID_jefe] INT,  
  
[Puesto] VARCHAR(50),  
  
PRIMARY KEY ([Id_empleado]));  
  
CREATE TABLE [Dim_Tiempo] (  
  
[Id_Tiempo] INT,  
  
[Fecha] DATE,  
  
[Ano] INT,
```

[Mes_texto] VARCHAR(20),

[Dia] INT,

[Mes_numero] INT,

[Semestre] INT,

[Trimestre] INT,

PRIMARY KEY ([Id_Tiempo]));

CREATE TABLE [Dim_Cliente] (

[Id_cliente] INT,

[nombre_contacto] VARCHAR(50),

[apellido_contacto] VARCHAR(50),

[telefono] VARCHAR(50),

[fax] INT,

[Linea_direccion1] VARCHAR(50),

[Linea_direccion2] VARCHAR(50),

[ciudad] VARCHAR(50),

[region] VARCHAR(50),

[pais] VARCHAR(50),

[codigo_postal] INT,

```
[ID_empleado_rep_ventas] INT,  
  
[limite_credito] INT,  
  
PRIMARY KEY ([Id_cliente]));  
  
CREATE TABLE [Dim_Producto] (  
  
[Id_Producto] INT,  
  
[Nombre] VARCHAR(50),  
  
[Categoria] INT,  
  
[Desc_categoria] TEXT,  
  
[Dimensiones] VARCHAR(50),  
  
[Proveedor] VARCHAR(50),  
  
[Descripcion_producto_texto] TEXT,  
  
[Cantidad_en_stock] SMALLINT,  
  
[Precio_venta] NUMERIC (15,2),  
  
[Precio_proveedor] NUMERIC(15,2),  
  
[Descripcion_html] VARCHAR(256),  
  
[Imagen] INT,  
  
PRIMARY KEY ([Id_Producto]) );  
  
CREATE TABLE [Dim_Pedido] (
```

[Id_Pedido] INT,

[Fecha_pedido] DATE,

[Fecha_esperada] DATE,

[Fecha_entrega] DATE,

[Estado] VARCHAR(50),

[Comentarios] TEXT,

[ID_producto] INT,

[cantidad] SMALLINT,

[precio_unidad] NUMERIC (15,2),

[numero_linea] INT,

PRIMARY KEY ([Id_Pedido]));

CREATE TABLE [Fact_Ventas] (

[Id_Venta] INT IDENTITY(1,1) PRIMARY KEY,

[Id_Pedido] INT,

[Id_Producto] INT,

[Id_Cliente] INT,

[Id_Empleado] INT,

[Id_Tiempo] INT,

```

[Id_oficina] INT,

[Cantidad] INT,

[Precio_Unitario] NUMERIC(15,2),

[Descuento] NUMERIC(5,2), -- Agregar una columna para descuentos si aplica

[Importe_Total] AS (Cantidad * Precio_Unitario * (1 - Descuento)), -- Calcular el importe
total

FOREIGN KEY (Id_Pedido) REFERENCES Dim_Pedido(Id_Pedido),

FOREIGN KEY (Id_Producto) REFERENCES Dim_Producto(Id_Producto),

FOREIGN KEY (Id_Cliente) REFERENCES Dim_Cliente(Id_Cliente),

FOREIGN KEY (Id_Empleado) REFERENCES Dim_Empleado(Id_Empleado),

FOREIGN KEY (Id_Tiempo) REFERENCES Dim_Tiempo(Id_Tiempo),

FOREIGN KEY (Id_oficina) REFERENCES Dim_Oficina(Id_oficina));

CREATE INDEX [IX_Fact_Ventas_Total_productos] ON [Fact_Ventas] (cantidad);

```

Script utilizado para realizar las consultas

```

SELECT Id_Categoria, Desc_Categoria, descripcion_texto, descripcion_html, imagen

FROM   Categoria_producto

SELECT [ID_detalle_pedido]

      ,[ID_pedido]

      ,[ID_producto]

```

,[cantidad]

,[precio_unidad]

,[numero_linea]

FROM [jardineria].[dbo].[detalle_pedido]

SELECT [ID_producto]

,[CodigoProducto]

,[nombre]

,[Categoria]

,[dimensiones]

,[proveedor]

,[descripcion]

,[cantidad_en_stock]

,[precio_venta]

,[precio_proveedor]

FROM [jardineria].[dbo].[producto]

SELECT [ID_empleado]

,[nombre]

,[apellido1]

,[apellido2]

,[extension]

,[email]

,[ID_oficina]

,[ID_jefe]

,[puesto]

FROM [jardineria].[dbo].[empleado]

SELECT [ID_cliente]

,[nombre_cliente]

,[nombre_contacto]

,[apellido_contacto]

,[telefono]

,[fax]

,[linea_direccion1]

,[linea_direccion2]

,[ciudad]

,[region]

,[pais]

,[codigo_postal]

,[ID_empleado_rep_ventas]

,[limite_credito]

FROM [jardineria].[dbo].[cliente]

SELECT [ID_oficina]

,[Descripcion]

,[ciudad]

,[pais]

,[region]

,[codigo_postal]

,[telefono]

,[linea_direccion1]

,[linea_direccion2]

FROM [jardineria].[dbo].[oficina]

SELECT [ID_pedido]

,[fecha_pedido]

,[fecha_esperada]

,[fecha_entrega]

,[estado]

,[comentarios]

,[ID_cliente]

FROM [jardineria].[dbo].[pedido]

SELECT distinct [fecha_pedido]

FROM [jardineria].[dbo].[pedido]

union

SELECT distinct [fecha_esperada]

FROM [jardineria].[dbo].[pedido]

union

SELECT distinct [fecha_entrega]

FROM [jardineria].[dbo].[pedido]

order by 1