

Electiva II Bases de datos II

Evidencia de aprendizaje 3. Proceso de transformación de datos y carga en el data mart final

Martha Cecilia Rúgeles

Camila Gallego Franco

Institución Universitaria Digital de Antioquia

Especialización en Analítica y Big Data

Víctor Hugo Mercado

30 de marzo de 2025

INTRODUCCIÓN

Según IUDigital (s.f.), La minería de datos tuvo sus inicios en la década de 1950, cuando las áreas de informática comenzaron a generar resúmenes de información. Sin embargo, estos eran extensos y de gran volumen. En la década de 1960, aunque ya existían sistemas de gestión de bases de datos, estos seguían siendo poco flexibles y no se integraban entre los distintos departamentos de una empresa, lo que dificultaba su manejo.

A finales de los años 80 surgió el termino de Data Warehouse, que permitió resolver muchos de los problemas previos en el tratamiento de datos. Gracias a esta tecnología, las tareas comenzaron a automatizarse y centralizarse, facilitando la gestión y el acceso eficiente a la información.

En la actualidad, los Data Warehouses desempeñan un papel clave en la analítica de datos, ya que centralizan la información de todas las áreas de una empresa e integran datos provenientes de múltiples fuentes. A partir de este concepto surge el Data Mart, un modelo enfocado en un área específica. Para su construcción, es esencial el proceso de ETL, que permite extraer datos desde diversas fuentes, como archivos de Excel o archivos planos, y trasladarlos a una base temporal denominada staging. En esta fase, los datos son transformados y preparados antes de su carga final en el Data Warehouse. Este proceso se ha convertido en una herramienta fundamental para la gestión de información y la generación de conocimiento, permitiendo a las empresas optimizar el uso de sus datos y facilitar la toma de decisiones estratégicas.

En este proyecto se desarrollará un Data Mart a partir de un modelo relacional de una base de datos de jardinería. Para su construcción, se diseñará un modelo estrella en Lucidchart y se implementará un proceso ETL que transformará los datos hacia dicho modelo, utilizando herramientas como SQL Server Management Studio y Visual Studio.

OBJETIVOS

Objetivo General:

Implementar un proceso ETL (Extracción, transformación y carga) que permita transformar una base de datos relacional de jardinería en un Data Mart basado en un modelo estrella, utilizando herramientas como SQL Server Management Studio y Visual Studio, con el objetivo de optimizar la organización y análisis de los datos para la toma de decisiones.

Objetivos Específicos:

- Diseñar un modelo estrella en Lucidchart para organizar y estructurar eficientemente los datos del Data Mart.
- Desarrollar un proceso de extracción de datos utilizando SQL Server Management Studio y Visual Studio, empleando una base de staging para la integración y preparación de la información.
- Implementar procesos de transformación de datos en SQL Server Management Studio, mediante consultas que incluyan la creación de columnas derivadas, conversión de datos y ordenamiento, para asegurar su correcta estandarización.
- Cargar los datos transformados en el Data Mart, garantizando su correcta integración y disponibilidad para el análisis mediante consultas SQL en SQL Server.
- Estructurar y optimizar el modelo estrella en el Data Mart para facilitar un análisis eficiente y la toma de decisiones basadas en los datos.

PLANTEAMIENTO DEL PROBLEMA

En la actualidad, las bases de datos relacionales son ampliamente utilizadas para el almacenamiento y gestión de información en diversos sectores, incluida la jardinería. Sin embargo, debido a la complejidad y volumen de los datos, su análisis puede volverse poco eficiente en un modelo relacional tradicional.

Para optimizar el acceso y la interpretación de la información, es necesario implementar un Data Mart, el cual permite estructurar los datos de manera más eficiente para su análisis. Para lograrlo, se propone la aplicación de un proceso ETL en el que se inicia con la construcción de un modelo estrella, facilitando el conocimiento de los datos.

ANÁLISIS DEL PROBLEMA

Las bases de datos relacionales son fundamentales para estructurar, organizar y gestionar información, permitiendo establecer conexiones entre distintos conjuntos de datos. Sin embargo, cuando se requiere realizar análisis más avanzados para proyectos de inteligencia de negocios con grandes volúmenes de datos, su rendimiento puede verse afectado.

Para optimizar estos procesos, es necesario utilizar herramientas como el Data Mart, que permite estructurar la información de un área específica y mejorar el acceso a los datos. Esto se logra mediante un proceso ETL (Extracción, Transformación, Carga), el cual sigue una metodología estructurada para garantizar que la información sea precisa, accesible y útil para el análisis.

En este contexto, el modelo de estrella se convierte en un componente clave, ya que complementa al modelo relacional al facilitar la identificación de las variables más relevantes y permitir una organización eficiente de los datos. Gracias a esta estructura, es posible optimizar las consultas y extraer información clave para la toma de decisiones estratégicas.

PROPUESTA DE LA SOLUCIÓN:

- Correcciones a la entrega 1.

Para la primera entrega, se diseñó un modelo de estrella en Lucidchart, donde la tabla de hechos representa las ventas en jardinería, complementada por cinco tablas de dimensiones: cliente, tiempo, producto, categoría y empleado.

Para la segunda fase, se desarrolló un nuevo modelo de estrella en SQL Server Management Studio, incorporando dimensiones adicionales como pago y oficina, lo que permitirá generar consultas más detalladas.

Cada dimensión contiene atributos clave que optimizan el proceso de consulta y análisis de información. Por ejemplo, la tabla cliente almacena datos como el nombre del cliente, la tabla producto incluye el nombre y precio de los artículos, la tabla tiempo registra la fecha de la venta, mientras que las dimensiones categoría y empleado contienen información relevante sobre la clasificación de productos y los cargos de los empleados.

Imagen del modelo estrella entrega 1:

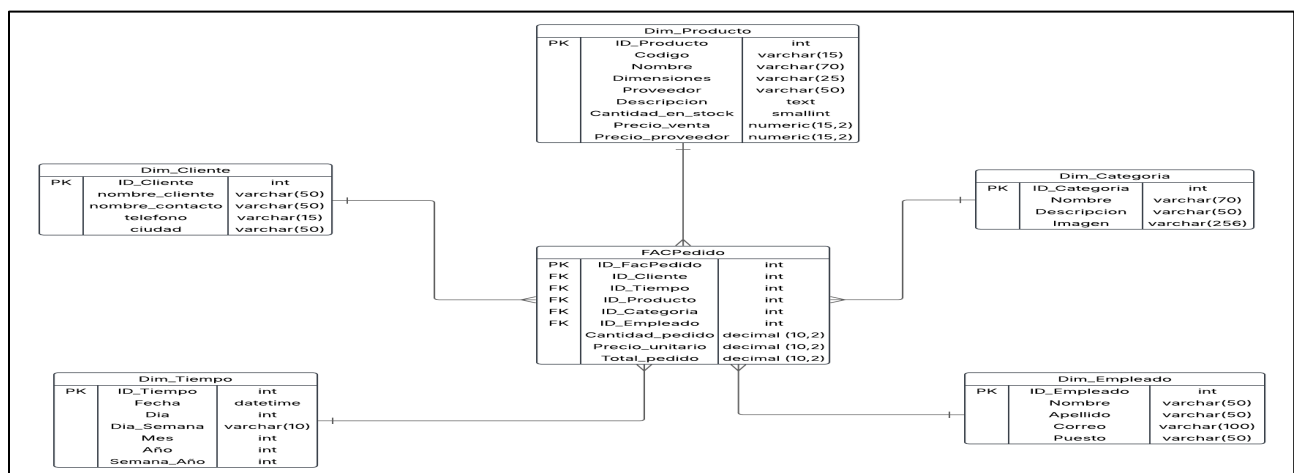
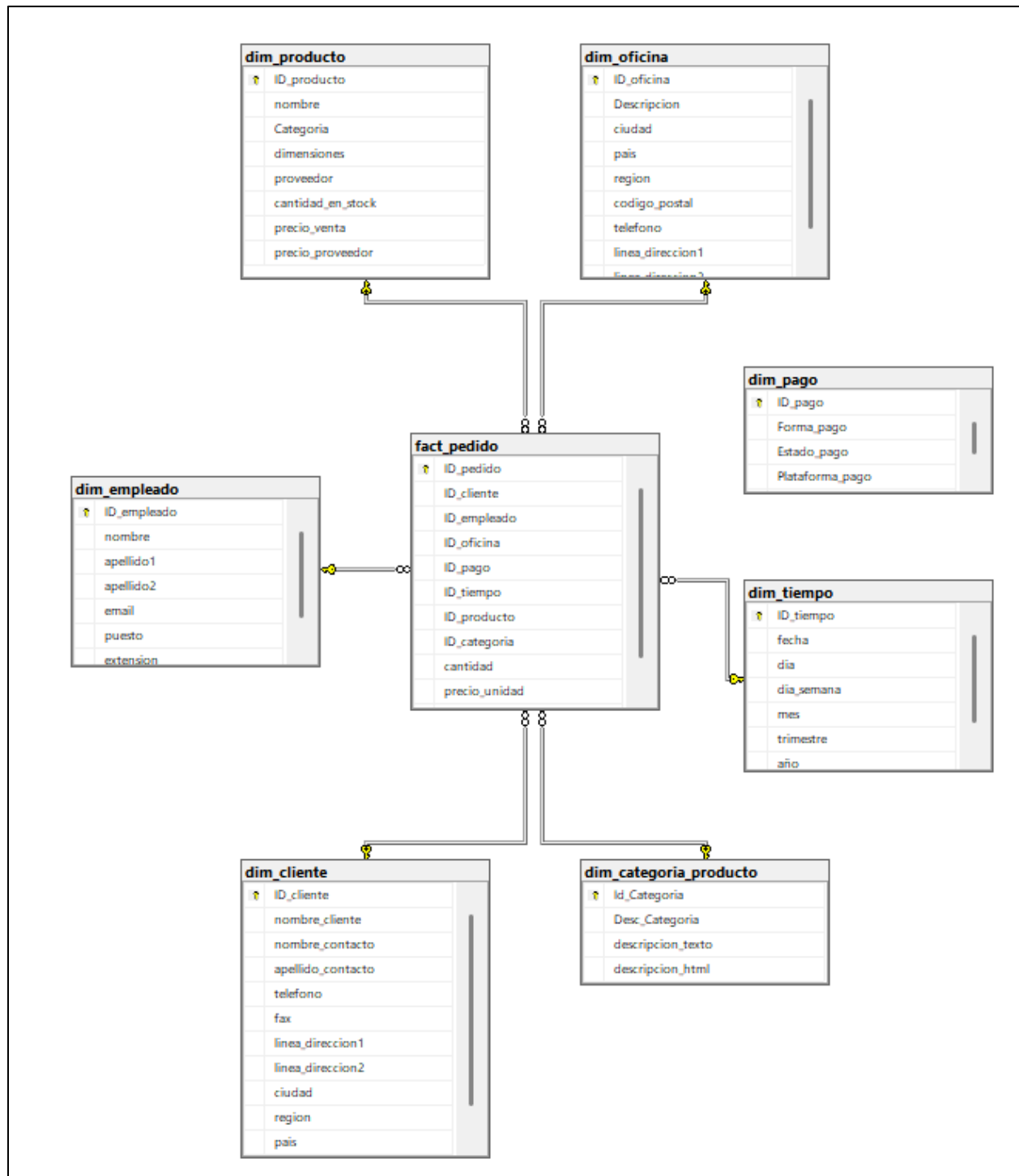


Imagen del modelo estrella entrega 2:



- Descripción del análisis realizado a los datos Jardinería y cómo estos se trasladaron a la base de datos Staging.

El análisis de los datos de la base de datos Jardinería se enfocó en identificar y sintetizar las tablas más relevantes para la construcción del modelo estrella. Durante este proceso, se seleccionaron las entidades clave que conforman el dominio de jardinería, como cliente, producto, pago, empleado, entre otras, que posteriormente serán fundamentales en el Data Mart.

En primer lugar, las tablas de la base de datos Jardinería fueron creadas y se poblaron en una base de datos en SQL Server. Estas tablas fueron estructuradas de forma que permitan una fácil integración y acceso para análisis posteriores.

Una vez que las tablas fueron creadas y poblaron la base de datos, se estableció la conexión entre SQL Server y Visual Studio para llevar a cabo el proceso de extracción. A través de Visual Studio, extrayendo los datos desde las tablas de la base de datos Jardinería hacia una base intermedia Staging.

El proceso de extracción implicó la implementación de flujos de control y flujos de datos que conectaron las tablas de origen con las tablas de destino. Para cada tabla, se definieron los orígenes y destinos específicos, asegurando que los datos se trasladaran de manera ordenada y precisa. Finalmente, los datos fueron cargados en la base de datos Staging,

- Descripción del análisis realizado en el proceso de transformación y carga

En el proceso de transformación en Visual Studio, mediante la herramienta "Derivada Columna", se reemplazaron los valores NULL. Además, se crearon columnas adicionales calculadas y se aplicaron funciones de fecha. Por otro lado, los datos se ordenaron utilizando el ID de cada tabla.

También se emplearon combinaciones de mezcla y división condicional para incorporar registros nuevos al destino.

En la carga desde la base de datos staging, se definió un proceso similar, que incluyó el ordenamiento de los datos y la incorporación de registros nuevos.

- Consultas SQL proceso de transformación

```
CREATE TABLE [dbo].[DimClienteST](  
  
    [ID_cliente] [int] NOT NULL,  
  
    [nombre_cliente] [nvarchar](50) NULL,  
  
    [telefono] [nvarchar](15) NULL,  
  
    [linea_direccion1] [nvarchar](50) NULL,  
  
    [ciudad] [nvarchar](50) NULL,  
  
    [region] [nvarchar](50) NULL,  
  
    [limite_credito] [numeric](15, 2) NULL,
```

```
CREATE TABLE [dbo].[DimEmpleadoST](  
  
    [ID_empleado] [int] NOT NULL,  
  
    [nombre] [nvarchar](50) NULL,  
  
    [email] [nvarchar](100) NULL,  
  
    [puesto] [nvarchar](50) NULL,  
  
    [extension] [nvarchar](10) NULL,
```

```
    [tipo_empleado] [nvarchar](13) NULL,  
  
CREATE TABLE [dbo].[DimOficinaST](  
  
    [ID_oficina] [int] NOT NULL,  
  
    [Descripcion] [nvarchar](10) NULL,  
  
    [ciudad] [nvarchar](30) NULL,  
  
    [pais] [nvarchar](50) NULL,  
  
    [region] [nvarchar](50) NULL,  
  
    [telefono] [nvarchar](20) NULL,  
  
    [linea_direccion1] [nvarchar](50) NULL,
```

```
CREATE TABLE [dbo].[DimPagoST](  
  
    [ID_pago] [int] NOT NULL,  
  
    [Forma_pago] [nvarchar](50) NULL,  
  
    [Estado_pago] [nvarchar](20) NULL,  
  
    [Plataforma_pago] [nvarchar](50) NULL,  
  
    [Categoria_pago] [nvarchar](20) NULL,  
  
    [Pago_Detallado] [nvarchar](126) NULL,
```

```
CREATE TABLE [dbo].[DimProductoST](  
  
    [ID_producto] [nvarchar](15) NOT NULL,
```

[nombre] [nvarchar](70) NULL,
[dimensiones] [nvarchar](25) NULL,
[proveedor] [nvarchar](50) NULL,
[cantidad_en_stock] [smallint] NULL,
[precio_venta] [numeric](15, 2) NULL,
[precio_proveedor] [numeric](15, 2) NULL,
[desc_Categoria] [nvarchar](255) NULL,

CREATE TABLE [dbo].[DimTiempoST](

[IDDimTiempo] [int] IDENTITY(1,1) NOT NULL,
[Fecha] [date] NULL,
[Dia] [int] NULL,
[Mes] [int] NULL,
[Anio] [int] NULL,
[Dia_semana] [int] NULL,
[Trimestre] [int] NULL,
[Dia_anio] [int] NULL,

CREATE TABLE [dbo].[HechosPedidoST](

[ID_pedido] [int] NOT NULL,

[ID_cliente] [int] NULL,

[ID_empleado] [int] NULL,

[ID_oficina] [int] NULL,

[ID_pago] [int] NULL,

[ID_producto] [nvarchar](15) NULL,

[IDDimTiempo] [int] NULL,

[cantidad] [int] NULL,

[precio_unidad] [numeric](15, 2) NULL,

[total] [numeric](26, 2) NULL,

- Consultas SQL proceso de carga

CREATE TABLE [dbo].[DimCliente](

[ID_cliente] [int] NOT NULL,

[nombre_cliente] [nvarchar](50) NULL,

[telefono] [nvarchar](15) NULL,

[linea_direccion1] [nvarchar](50) NULL,

[ciudad] [nvarchar](50) NULL,

[region] [nvarchar](50) NULL,

[limite_credito] [numeric](15, 2) NULL,

PRIMARY KEY CLUSTERED

(

CREATE TABLE [dbo].[DimEmpleado](

[ID_empleado] [int] NOT NULL,

[nombre] [nvarchar](50) NULL,

[email] [nvarchar](100) NULL,

[puesto] [nvarchar](50) NULL,

[extension] [nvarchar](10) NULL,

[tipo_empleado] [nvarchar](13) NULL,

PRIMARY KEY CLUSTERED

(

CREATE TABLE [dbo].[DimOficina](

[ID_oficina] [int] NOT NULL,

[Descripcion] [nvarchar](10) NULL,

[ciudad] [nvarchar](30) NULL,

[pais] [nvarchar](50) NULL,

[region] [nvarchar](50) NULL,

[telefono] [nvarchar](20) NULL,

[linea_direccion1] [nvarchar](50) NULL,

PRIMARY KEY CLUSTERED

(

CREATE TABLE [dbo].[DimPago](

[ID_pago] [int] NOT NULL,

[Forma_pago] [nvarchar](50) NULL,

[Estado_pago] [nvarchar](20) NULL,

[Plataforma_pago] [nvarchar](50) NULL,

[Categoria_pago] [nvarchar](20) NULL,

[Pago_Detallado] [nvarchar](126) NULL,

PRIMARY KEY CLUSTERED

(

CREATE TABLE [dbo].[DimProducto](

[ID_producto] [nvarchar](15) NOT NULL,

[nombre] [nvarchar](70) NULL,

[dimensiones] [nvarchar](25) NULL,

[proveedor] [nvarchar](50) NULL,

[cantidad_en_stock] [smallint] NULL,

[precio_venta] [numeric](15, 2) NULL,

[precio_proveedor] [numeric](15, 2) NULL,

[desc_Categoria] [nvarchar](255) NULL,

PRIMARY KEY CLUSTERED

(

CREATE TABLE [dbo].[DimTiempo](

[IDDimTiempo] [int] NOT NULL,

[Fecha] [date] NULL,

[Dia] [int] NULL,

[Mes] [int] NULL,

[Anio] [int] NULL,

[Dia_semana] [int] NULL,

[Trimestre] [int] NULL,

[Dia_anio] [int] NULL,

PRIMARY KEY CLUSTERED

(

CREATE TABLE [dbo].[FacPedido](

[ID_pedido] [int] NOT NULL,

[ID_cliente] [int] NULL,

[ID_empleado] [int] NULL,

[ID_oficina] [int] NULL,

[ID_pago] [int] NULL,

[ID_producto] [nvarchar](15) NULL,

[IDDimTiempo] [int] NULL,

[cantidad] [int] NULL,

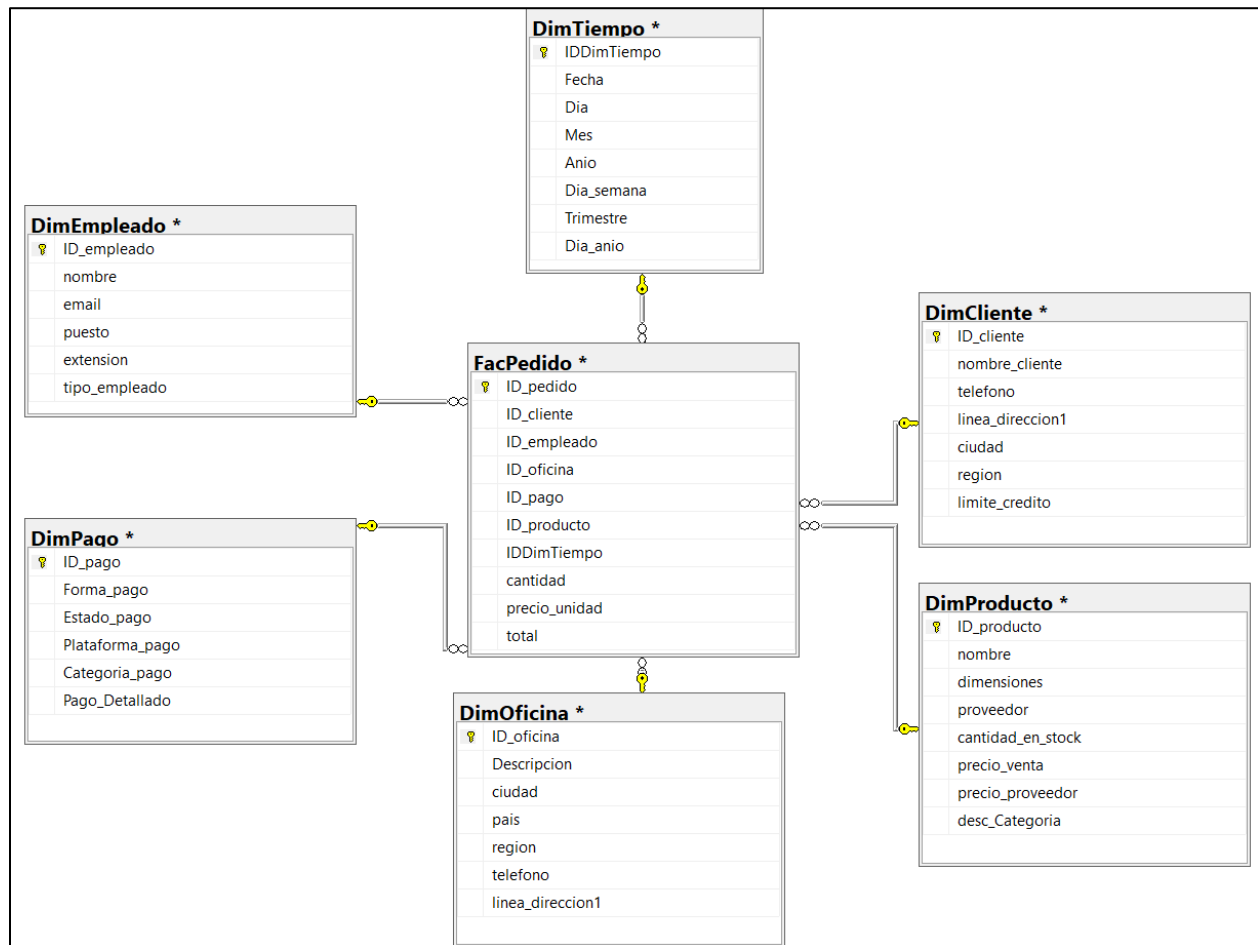
[precio_unidad] [numeric](15, 2) NULL,

[total] [numeric](26, 2) NULL,

PRIMARY KEY CLUSTERED

(

- Modelo estrella Datamart



ANEXOS

- Consultas necesarias para crear la base de datos Staging

```
CREATE DATABASE [StagingBD2POS20251]
```

```
CONTAINMENT = NONE
```

```
ON PRIMARY
```

```
( NAME = N'StagingBD2POS20251', FILENAME = N'C:\Program Files\Microsoft SQL  
Server\MSSQL16.SQLEXPRESS\MSSQL\DATA\StagingBD2POS20251.mdf' , SIZE =  
8192KB , MAXSIZE = UNLIMITED, FILEGROWTH = 65536KB )
```

```
LOG ON
```

```
( NAME = N'StagingBD2POS20251_log', FILENAME = N'C:\Program Files\Microsoft SQL  
Server\MSSQL16.SQLEXPRESS\MSSQL\DATA\StagingBD2POS20251_log.ldf' , SIZE =  
8192KB , MAXSIZE = 2048GB , FILEGROWTH = 65536KB )
```

```
WITH CATALOG_COLLATION = DATABASE_DEFAULT, LEDGER = OFF
```

```
GO
```

```
IF (1 = FULLTEXTSERVICEPROPERTY('IsFullTextInstalled'))
```

```
begin
```

```
EXEC [StagingBD2POS20251].[dbo].[sp_fulltext_database] @action = 'enable'
```

```
end
```

```
GO
```

```
ALTER DATABASE [StagingBD2POS20251] SET ANSI_NULL_DEFAULT OFF
```

```
GO
```

```
ALTER DATABASE [StagingBD2POS20251] SET ANSI_NULLS OFF
```

```
GO
```

ALTER DATABASE [StagingBD2POS20251] SET ANSI_PADDING OFF

GO

ALTER DATABASE [StagingBD2POS20251] SET ANSI_WARNINGS OFF

GO

ALTER DATABASE [StagingBD2POS20251] SET ARITHABORT OFF

GO

ALTER DATABASE [StagingBD2POS20251] SET AUTO_CLOSE OFF

GO

ALTER DATABASE [StagingBD2POS20251] SET AUTO_SHRINK OFF

GO

ALTER DATABASE [StagingBD2POS20251] SET AUTO_UPDATE_STATISTICS ON

GO

ALTER DATABASE [StagingBD2POS20251] SET CURSOR_CLOSE_ON_COMMIT OFF

GO

ALTER DATABASE [StagingBD2POS20251] SET CURSOR_DEFAULT GLOBAL

GO

ALTER DATABASE [StagingBD2POS20251] SET CONCAT_NULL_YIELDS_NULL OFF

GO

ALTER DATABASE [StagingBD2POS20251] SET NUMERIC_ROUNDABORT OFF

GO

ALTER DATABASE [StagingBD2POS20251] SET QUOTED_IDENTIFIER OFF

GO

ALTER DATABASE [StagingBD2POS20251] SET RECURSIVE_TRIGGERS OFF

GO

ALTER DATABASE [StagingBD2POS20251] SET DISABLE_BROKER

GO

ALTER DATABASE [StagingBD2POS20251] SET AUTO_UPDATE_STATISTICS_ASYNC

OFF

GO

ALTER DATABASE [StagingBD2POS20251] SET

DATE_CORRELATION_OPTIMIZATION OFF

GO

ALTER DATABASE [StagingBD2POS20251] SET TRUSTWORTHY OFF

GO

ALTER DATABASE [StagingBD2POS20251] SET ALLOW_SNAPSHOT_ISOLATION OFF

GO

ALTER DATABASE [StagingBD2POS20251] SET PARAMETERIZATION SIMPLE

GO

ALTER DATABASE [StagingBD2POS20251] SET READ_COMMITTED_SNAPSHOT OFF

GO

ALTER DATABASE [StagingBD2POS20251] SET HONOR_BROKER_PRIORITY OFF

GO

ALTER DATABASE [StagingBD2POS20251] SET RECOVERY SIMPLE

GO

ALTER DATABASE [StagingBD2POS20251] SET MULTI_USER

GO

ALTER DATABASE [StagingBD2POS20251] SET PAGE_VERIFY CHECKSUM

GO

ALTER DATABASE [StagingBD2POS20251] SET DB_CHAINING OFF

GO

ALTER DATABASE [StagingBD2POS20251] SET FILESTREAM(

NON_TRANSACTED_ACCESS = OFF)

GO

ALTER DATABASE [StagingBD2POS20251] SET TARGET_RECOVERY_TIME = 60

SECONDS

GO

ALTER DATABASE [StagingBD2POS20251] SET DELAYED_DURABILITY = DISABLED

GO

ALTER DATABASE [StagingBD2POS20251] SET

ACCELERATED_DATABASE_RECOVERY = OFF

GO

ALTER DATABASE [StagingBD2POS20251] SET QUERY_STORE = ON

GO

ALTER DATABASE [StagingBD2POS20251] SET QUERY_STORE (OPERATION_MODE =

READ_WRITE, CLEANUP_POLICY = (STALE_QUERY_THRESHOLD_DAYS = 30),

DATA_FLUSH_INTERVAL_SECONDS = 900, INTERVAL_LENGTH_MINUTES = 60,

MAX_STORAGE_SIZE_MB = 1000, QUERY_CAPTURE_MODE = AUTO,

```
SIZE_BASED_CLEANUP_MODE = AUTO, MAX_PLANS_PER_QUERY = 200,  
WAIT_STATS_CAPTURE_MODE = ON)
```

```
GO
```

```
ALTER DATABASE [StagingBD2POS20251] SET READ_WRITE
```

```
GO
```

- Script de las tablas

```
CREATE TABLE [dbo].[CategoriaST](  
    [Id_Categoria] [int] NULL,  
    [desc_Categoria] [nvarchar](50) NULL,  
    [descripcion_texto] [nvarchar](max) NULL,  
    [descripcion_html] [nvarchar](max) NULL  
) ON [PRIMARY] TEXTIMAGE_ON [PRIMARY]  
GO
```

```
CREATE TABLE [dbo].[ClienteST](  
    [ID_cliente] [int] NULL,  
    [nombre_cliente] [nvarchar](50) NULL,  
    [nombre_contacto] [nvarchar](30) NULL,  
    [apellido_contacto] [nvarchar](30) NULL,  
    [telefono] [nvarchar](15) NULL,  
    [fax] [nvarchar](15) NULL,  
    [linea_direccion1] [nvarchar](50) NULL,  
    [ciudad] [nvarchar](50) NULL,
```

```
        [region] [nvarchar](50) NULL,  
        [pais] [nvarchar](50) NULL,  
        [limite_credito] [numeric](15, 2) NULL  
    ) ON [PRIMARY]  
GO
```

```
CREATE TABLE [dbo].[EmpleadoST](  
    [ID_empleado] [int] NULL,  
    [nombre] [nvarchar](50) NULL,  
    [apellido1] [nvarchar](50) NULL,  
    [apellido2] [nvarchar](50) NULL,  
    [email] [nvarchar](100) NULL,  
    [puesto] [nvarchar](50) NULL,  
    [extension] [nvarchar](10) NULL  
    ) ON [PRIMARY]  
GO
```

```
CREATE TABLE [dbo].[OficinaST](  
    [ID_oficina] [int] NULL,  
    [Descripcion] [nvarchar](10) NULL,  
    [ciudad] [nvarchar](30) NULL,  
    [pais] [nvarchar](50) NULL,  
    [region] [nvarchar](50) NULL,
```

```
[codigo_postal] [nvarchar](10) NULL,  
  
[telefono] [nvarchar](20) NULL,  
  
[linea_direccion1] [nvarchar](50) NULL,  
  
[linea_direccion2] [nvarchar](50) NULL
```

```
) ON [PRIMARY]
```

```
GO
```

```
CREATE TABLE [dbo].[PagoST](
```

```
[ID_pago] [int] NULL,  
  
[Forma_pago] [nvarchar](50) NULL,  
  
[Estado_pago] [nvarchar](20) NULL,  
  
[Plataforma_pago] [nvarchar](50) NULL,  
  
[Categoria_pago] [nvarchar](20) NULL
```

```
) ON [PRIMARY]
```

```
GO
```

```
CREATE TABLE [dbo].[ProductoST](
```

```
[ID_producto] [nvarchar](15) NULL,  
  
[nombre] [nvarchar](70) NULL,  
  
[dimensiones] [nvarchar](25) NULL,
```



```
[proveedor] [nvarchar](50) NULL,  
  
[cantidad_en_stock] [smallint] NULL,  
  
[precio_venta] [numeric](15, 2) NULL,  
  
[precio_proveedor] [numeric](15, 2) NULL
```

```
) ON [PRIMARY]
```

```
GO
```

```
CREATE TABLE [dbo].[TiempoST](
```

```
    [ID_tiempo] [int] NULL,  
  
    [fecha] [date] NULL,  
  
    [dia] [int] NULL,  
  
    [dia_semana] [int] NULL,  
  
    [mes] [int] NULL,  
  
    [trimestre] [int] NULL,  
  
    [año] [int] NULL
```

```
) ON [PRIMARY]
```

```
GO
```

```
CREATE TABLE [dbo].[PedidoST](
```

```
    [ID_pedido] [int] NULL,
```

[ID_cliente] [int] NULL,

[ID_empleado] [int] NULL,

[ID_oficina] [int] NULL,

[ID_pago] [int] NULL,

[ID_tiempo] [int] NULL,

[ID_producto] [nvarchar](15) NULL,

[ID_categoria] [int] NULL,

[cantidad] [int] NULL,

[precio_unidad] [numeric](15, 2) NULL,

[total] [numeric](15, 2) NULL

) ON [PRIMARY]

GO

BIBLIOGRAFIA

IUDigital. (s.f.). Tema 2. *Minería de datos: evolución*. Institución Universitaria Digital de Antioquia.[https://iudigital.instructure.com/courses/21521/pages/tema-2-mineria-de-datos-evolucion?module_item_id=1377832].