

Evidencia de aprendizaje 3

Andres Felipe Osorno Quiroz

andres.osorno@est.iudigital.edu.co

Mayo 2024

Institución Universitaria Digital de Antioquia
Ingeniería de Software y Datos
Bases de Datos II (VICTOR HUGO MERCADO)

Introducción

En el ámbito empresarial actual, la capacidad para analizar datos y extraer información significativa se ha convertido en un factor crítico para la toma de decisiones efectiva. Las organizaciones buscan constantemente formas de aprovechar sus recursos de datos para obtener una comprensión más profunda de su negocio y mejorar su rendimiento.

En este contexto, la implementación de un data mart se presenta como una solución eficaz para proporcionar acceso rápido y específico a conjuntos de datos relevantes para el análisis y la toma de decisiones. Un data mart bien diseñado, centrado en un modelo estrella, permite a las empresas concentrarse en áreas específicas de interés, como ventas, inventario, o recursos humanos, y realizar análisis detallados para obtener información valiosa.

El presente trabajo tiene como objetivo diseñar un modelo estrella para un data mart basado en la base de datos Jardinería. Esta base de datos, que contiene información sobre productos, clientes, pedidos y más, servirá como fuente de datos para nuestro modelo. Nos centraremos en la creación de un modelo estrella que nos permita analizar y responder a tres categorías específicas: identificar el producto más vendido, la categoría con más productos y el año con más ventas.

A lo largo de este documento, exploraremos la estructura de la base de datos Jardinería, identificaremos las tablas relevantes y sus relaciones, diseñaremos la estructura del modelo estrella, y discutiremos las implicaciones y beneficios de este enfoque para el análisis de datos en el contexto empresarial.

Objetivos

Diseñar un modelo estrella para un data mart basado en la base de datos Jardinería que permita analizar y responder a tres categorías específicas: identificar el producto más vendido, la categoría con más productos y el año con más ventas.

Planteamiento del Problema

En un entorno empresarial altamente competitivo, la capacidad para analizar y comprender los datos de manera efectiva es esencial para el éxito y la supervivencia de las organizaciones. La empresa Jardinería posee una base de datos que contiene información valiosa sobre productos, clientes, pedidos y más, pero carece de una estructura adecuada para realizar análisis detallados

y responder a preguntas específicas relacionadas con las ventas, las categorías de productos y el rendimiento a lo largo del tiempo.

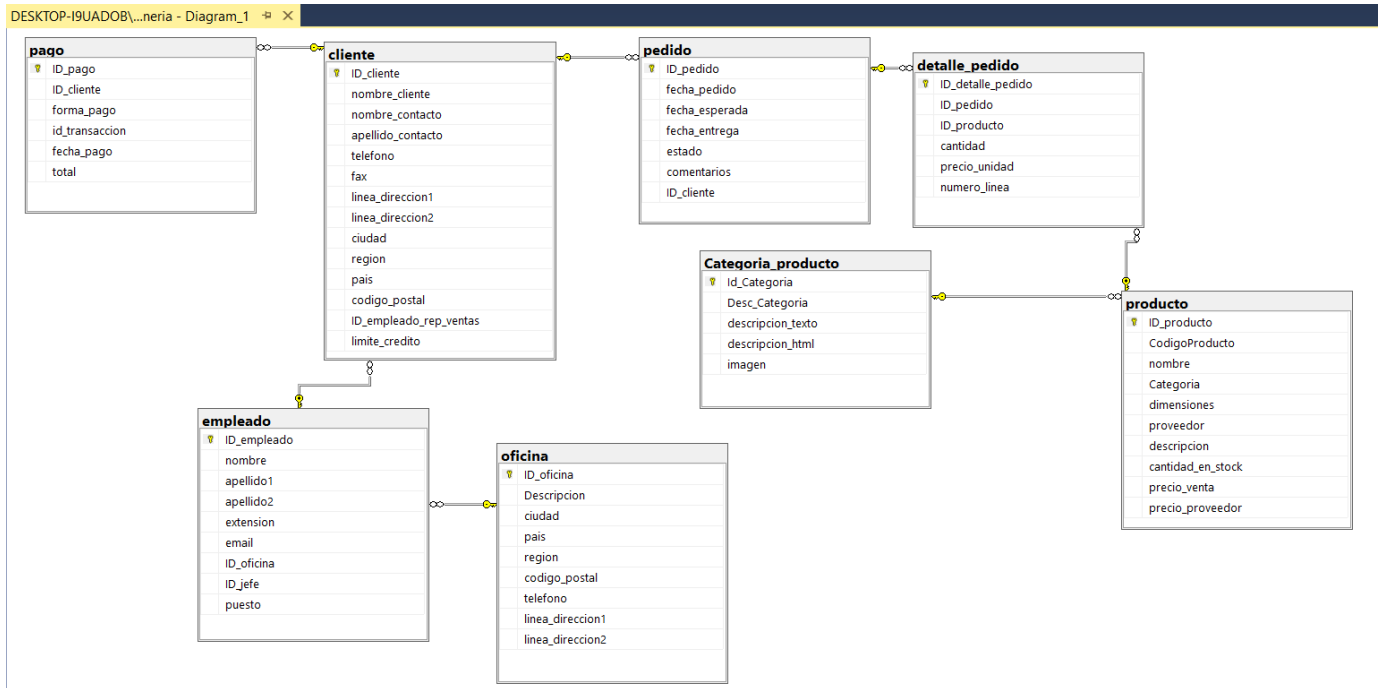
La falta de un modelo de datos optimizado y orientado al análisis dificulta la capacidad de la empresa Jardinería para identificar tendencias, tomar decisiones informadas y aprovechar al máximo su potencial de negocio. Por lo tanto, surge la necesidad de diseñar e implementar un modelo estrella para un data mart que permita a la empresa Jardinería realizar análisis detallados y responder a preguntas clave de manera eficiente y efectiva.

Análisis del Problema

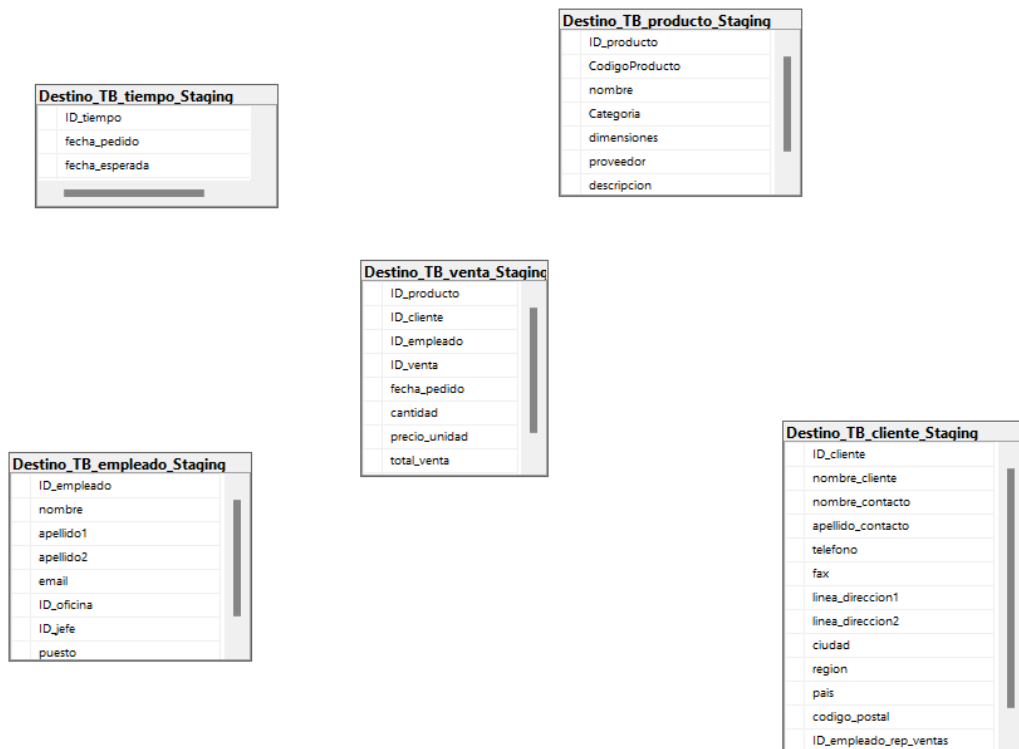
El análisis detallado de la base de datos Jardinería revela la presencia de tablas relevantes que contienen información sobre productos, clientes, pedidos y más. Sin embargo, la falta de una estructura de datos optimizada para el análisis dificulta la capacidad de la empresa para obtener información valiosa y tomar decisiones estratégicas basadas en datos.

La ausencia de un modelo estrella que vincule una tabla de hechos centrada en las ventas con dimensiones relevantes, como productos, categorías y tiempo, limita la capacidad de la empresa para identificar tendencias, entender el comportamiento del cliente y optimizar el rendimiento del negocio.

Vista general de la Base de Datos “Jardinería”



Vista general del modelo estrella



Se puede observar también en la extracción de jardinería en Staging.



Descripción del modelo estrella propuesto.

Este modelo estrella propuesto tiene la estructura adecuada para analizar y responder a las tres categorías planteadas:

- **Identificar el producto más vendido**
- **La categoría con más productos**
- **El año con más ventas**

A continuación, la descripción del modelo.

1. Tabla de Hechos (**hecho_ventas**):

- Esta tabla representa los eventos de interés del negocio, que en este caso son las ventas de productos. Cada fila en esta tabla corresponde a una transacción de venta específica.

2. Dimensión de Producto (**dimension_producto**):

- Contiene detalles sobre los productos vendidos, como el nombre del producto, la categoría a la que pertenece, el proveedor, la cantidad en stock y el precio de venta.

3. Dimensión de Tiempo (**dimension_tiempo**):

- Almacena toda la información de los tiempos en que los pedidos fueron comprados, la información de cuando se espera que el pedido llegue, la fecha de entrega y cuantos días de retraso estos tuvieron.

4. Dimensión de Cliente (**dimension_cliente**):

- Contiene información sobre los clientes que realizaron los pedidos, incluyendo el nombre del cliente, información de contacto, dirección y detalles de ubicación.

5. Dimensión de Empleado (dimension_empleado):

- Contiene los datos del empleado de la empresa, tales como nombre completo en una celda, email corporativo, y el encargado de dicho empleado.

Extracción de Datos:

Se identifican las tablas y campos necesarios para el análisis y se extraen los datos relevantes. Las tablas extraídas incluyen:

Ventas: Contiene datos de transacciones de ventas, como ID de producto, ID de cliente, fecha de pedido, cantidad, precio, etc.

Clientes: Información sobre clientes, como nombre, contacto, dirección, límite de crédito, etc.

Productos: Detalles de los productos, como código, nombre, categoría, dimensiones, proveedor, etc.

Empleados: Datos sobre los empleados, como nombre, apellidos, email, oficina, jefe, puesto, etc

Tiempo: Datos sobre el tiempo, fecha de entrega, fecha de pedido, retrasos.

Transformación de Datos (Transform):

Durante la etapa de transformación, se aplican diversas reglas y cálculos para limpiar, enriquecer y estructurar los datos. Las transformaciones incluyeron:

Cálculo de Totales y Derivados:

Cálculo del total de ventas multiplicando la cantidad por el precio unitario.

Cálculo de la duración de entregas restando la fecha de pedido de la fecha de entrega.

Enriquecimiento de Datos:

Extracción del año, mes y día de la semana de las fechas de pedido para facilitar análisis temporales.

Creación de campos categorizados, como clasificar ventas en categorías (Bajo, Medio, Alto) según su monto total.

Limpieza y Validación de Datos:

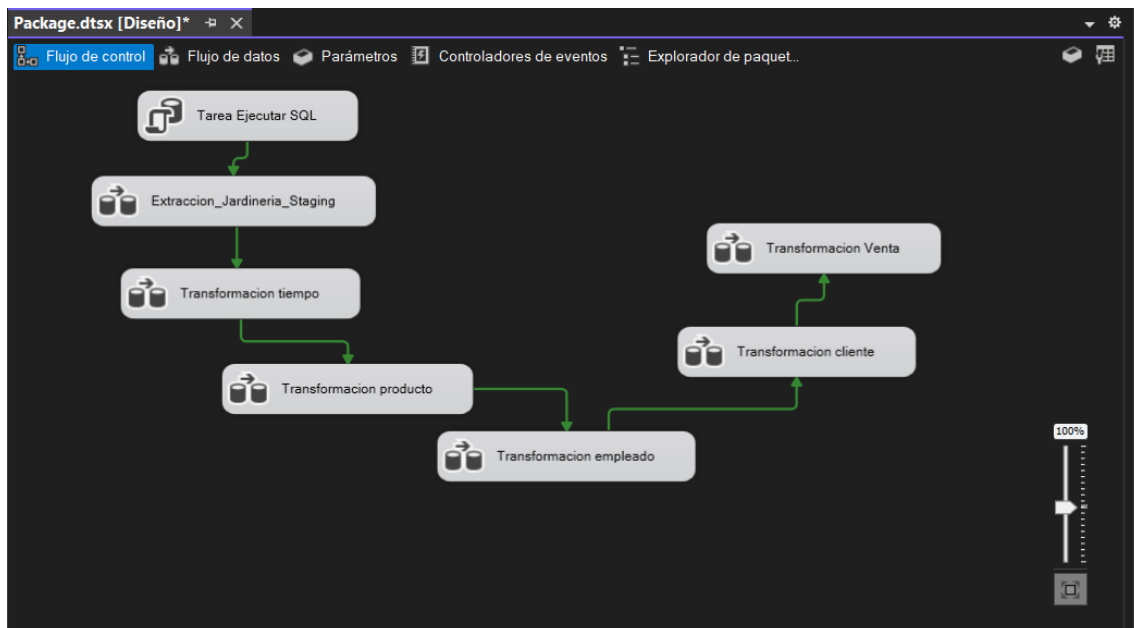
Validación de datos nulos o inconsistentes.

Normalización de formatos de datos, como convertir direcciones a un formato estandarizado.

Creación de Nuevas Columnas:

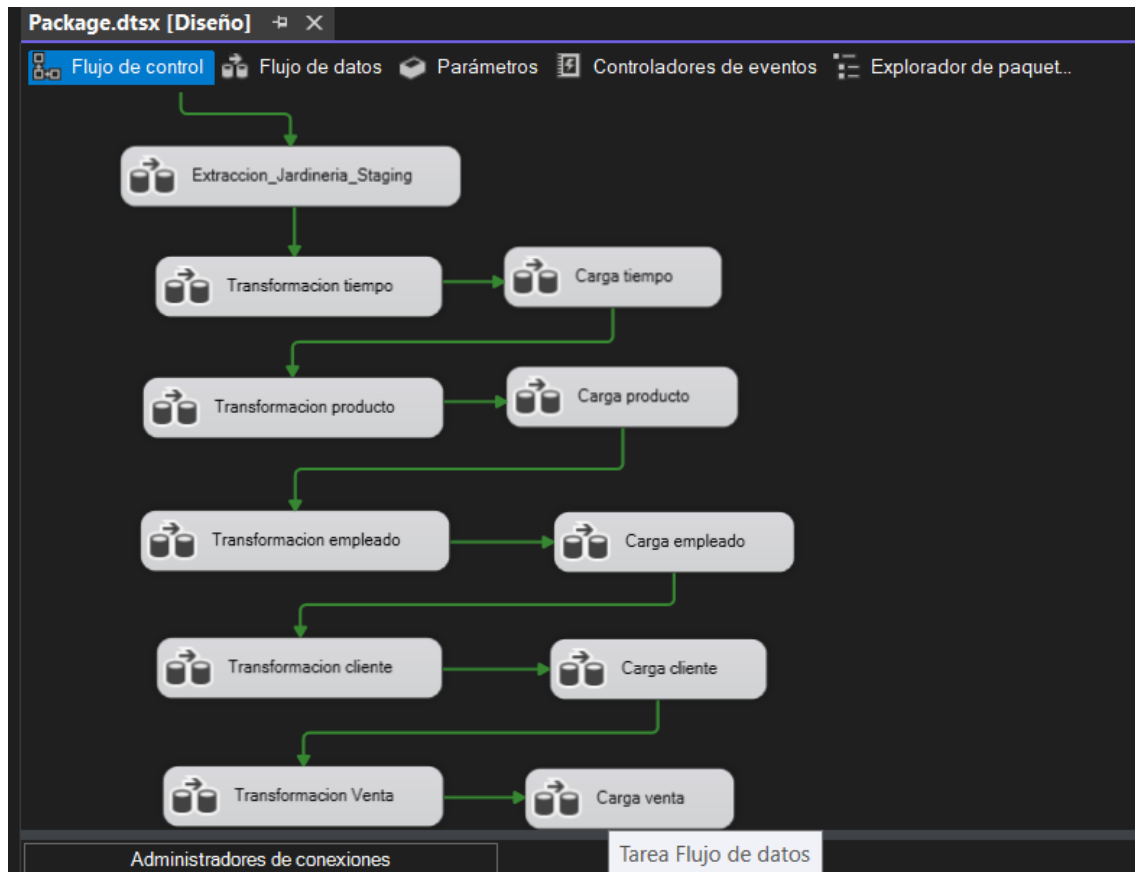
Concatenación de campos de dirección en un solo campo completo.

Creación de campos booleanos, como indicar si un pedido fue entregado a tiempo.



Carga de Datos (Load)

La etapa de carga implica insertar los datos transformados en la base de datos de destino. Se crean tablas dimensionales o de hechos según el esquema de la base de datos de destino.



Conclusiones

El modelo estrella diseñado para el data mart de Jardinería ofrece una estructura organizada y eficiente para almacenar y analizar los datos de ventas de productos. La implementación de este proceso en SSIS ha facilitado la integración y el manejo de datos, mejorando la capacidad de generar informes y obtener insights valiosos para la toma de decisiones. Este enfoque asegura que los datos sean precisos, consistentes.