

Evidencia de aprendizaje No. 4

Proyecto integrador-Repositorio de todas las actividades

Bases de Datos II

PREICA2401B010094

Eder David Sánchez Usuga



Víctor Hugo Mercado

Mayo 2024

Introducción

La capacidad de analizar datos de ventas es una necesidad para cualquier empresa que desee optimizar sus operaciones y estrategias de mercado en la era actual de los negocios. La empresa Jardinería, que vende productos de jardinería, ha reconocido la importancia de cambiar su manejo de datos para permitir un análisis más profundo y fácil de acceder a su desempeño en ventas.

Objetivos

El principal objetivo de este proyecto es construir un modelo estrella para un Data Mart que permita a Jardinería analizar y responder eficazmente a tres categorías específicas:

1. Identificar el producto más vendido.
2. Determinar la categoría con más productos.
3. Establecer el año con más ventas.

Además, desarrollar una base de datos Staging es crucial para este objetivo, ya que actuará como un área de preparación para los datos extraídos de las fuentes operativas. La base de datos Staging permitirá realizar transformaciones, limpieza, y otras operaciones de procesamiento necesarias para que los datos lleguen al Data Mart en un formato óptimo y listo para el análisis.

Planteamiento del Problema

La empresa de jardinería actualmente enfrenta grandes desafíos para acceder y analizar datos de ventas. La estructura compleja de su base de datos hace que las consultas sean lentas y complicadas, lo que impide una toma de decisiones ágil y basada en datos. Además, la variedad y el volumen constante de los datos requieren una solución escalable y robusta que pueda adaptarse a las necesidades cambiantes de las empresas.

Análisis del Problema

El análisis inicial de la base de datos de Jardinería revela que la información está distribuida en varias tablas relacionales que incluyen datos sobre productos, categorías de productos, pedidos, detalles de pedidos, y más. Esta estructura relacional es ideal para transacciones, pero no para análisis y reportes rápidos debido a la necesidad de realizar múltiples joins complejos y operaciones sobre grandes volúmenes de datos.

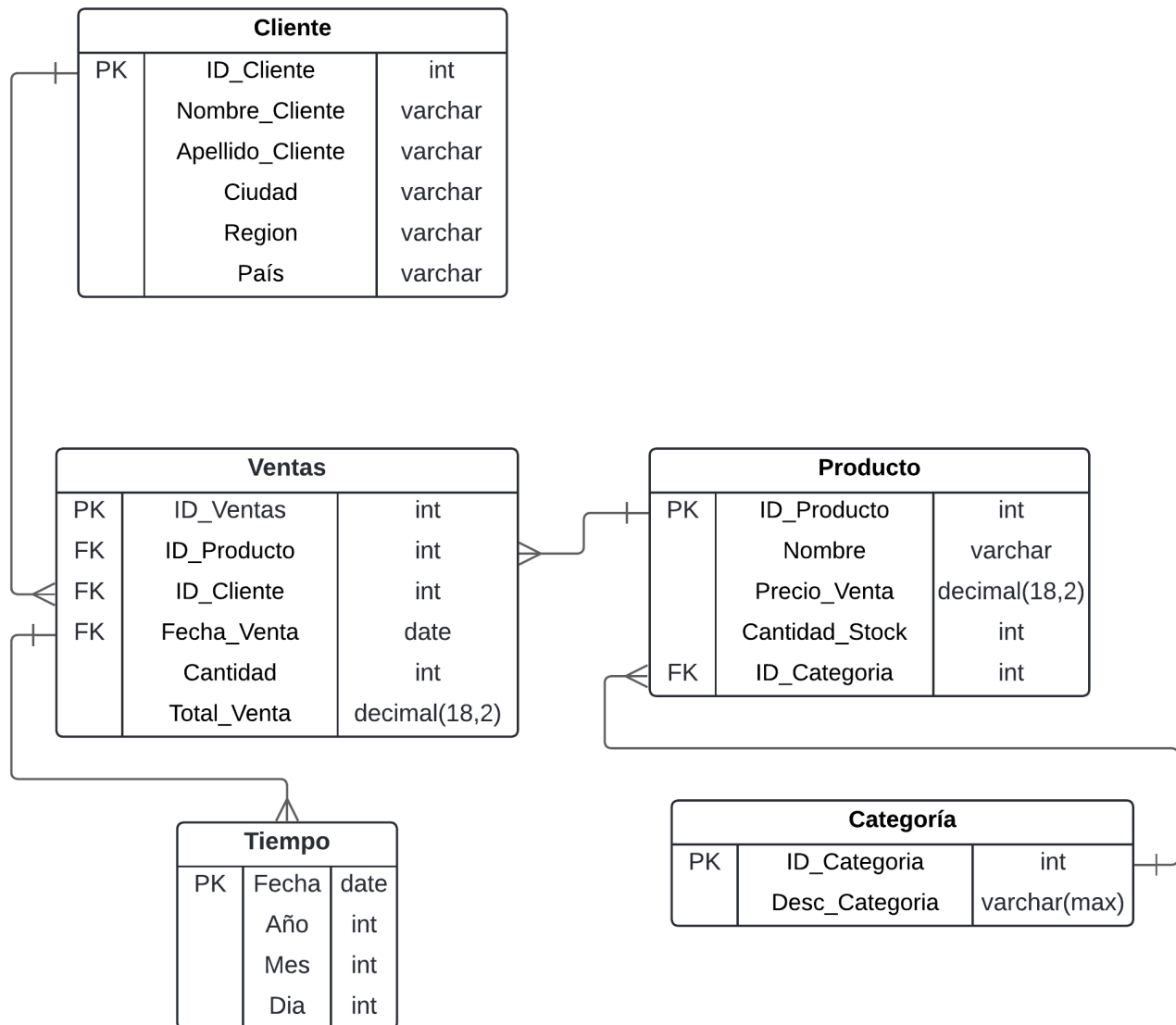
Para superar estos desafíos, se propone la creación de un Data Mart utilizando el modelo estrella, que simplifica las consultas y mejora el rendimiento al reducir la complejidad de las relaciones. Este modelo permitirá a Jardinería enfocarse en indicadores clave de rendimiento y facilitar la toma de decisiones estratégicas con base en insights claros y accesibles.

El diseño de este modelo estrella implicará la creación de una tabla de hechos central, que capture las transacciones de ventas, vinculada a dimensiones específicas que reflejen aspectos críticos de los datos, como productos, categorías, y fechas. Esta estructura asegurará que los datos estén organizados de manera que las preguntas clave de negocio puedan ser respondidas de manera eficiente y efectiva.

Además, se incluirá la elaboración de una base de datos Staging como paso intermedio crucial antes de la alimentación del Data Mart. La base de datos Staging servirá como un área de transición donde los datos extraídos se limpiarán, transformarán y consolidarán. Este proceso no solo prepara los datos para el análisis, sino que también minimiza los riesgos de afectar el rendimiento y la integridad de las bases de datos operacionales durante el proceso de extracción de datos. Al utilizar una base de datos Staging, Jardinería puede asegurar que los datos sean consistentes, actualizados y listos para ser utilizados en el Data Mart, proporcionando así una plataforma robusta para análisis avanzados y reportes dinámicos.

Propuesta de la solución

El modelo estrella para la empresa de jardinería está destinado a mejorar el análisis de ventas y brindar un acceso rápido y eficiente a los datos esenciales para la toma de decisiones. El modelo se centra en una tabla de hechos principal que registra todas las transacciones de venta y está conectada a una serie de dimensiones que describen los elementos pertinentes de cada venta.



Lista de dimensiones propuestas.

1. Dimensión Producto

- ID_producto (Int): Identificador único del producto.
- nombre_producto (Varchar (MAX)): Nombre del producto.
- precio_venta (Decimal(18,2)): Precio de venta del producto.

2. Dimensión Categoría

- ID_categoria (Int): Identificador único de la categoría.
- Desc_Categoria (Varchar (50)): Descripción de la categoría.

3. Dimensión Cliente

- ID_cliente (Int): Identificador único del cliente.
- nombre_cliente (Varchar (50)): Nombre del cliente.
- apellido_cliente (Varchar (50)): Apellido del cliente
- ciudad_cliente (Varchar (50)): Ciudad del cliente.
- region_cliente (Varchar (50)): Región del cliente.
- pais_cliente (Varchar(50)): País del cliente.

4. Dimensión Tiempo

- fecha (Date): Fecha exacta del pedido.
- año (Int): Año extraído de la fecha del pedido.
- mes (Int): Mes extraído de la fecha del pedido.
- día (Int): Día extraído de la fecha del pedido.

Detalla la tabla de hechos

- ID_venta (Int): Identificador único para cada venta
- fecha_venta (Date): La fecha en que se realizó el pedido.
- ID_producto (Int): FK identificador del producto en la dimensión Producto.
- ID_categoria (Int): FK identificador de la categoría del producto en la dimensión Categoría.
- ID_cliente (Int): FK identificador del cliente en la dimensión Cliente.
- cantidad (Int): La cantidad de productos vendidos en cada transacción.
- total_venta (Decimal(18,2)). Representa el valor total de cada venta.

Creación de Base de Datos Staging

Estructura de las tablas Staging:

Tabla Producto
<pre>CREATE TABLE ProductoStaging (ID_Producto INT NOT NULL PRIMARY KEY, ID_Categoria INT NOT NULL, Nombre NVARCHAR(70) NOT NULL, Precio_Venta NUMERIC(15, 2) NOT NULL, Cantidad_Stock SMALLINT NOT NULL);</pre>
Tabla Cliente
<pre>CREATE TABLE ClienteStaging (ID_Cliente INT NOT NULL PRIMARY KEY, Nombre_Cliente NVARCHAR(50) NOT NULL, Apellido_Cliente NVARCHAR(50) NOT NULL, Ciudad NVARCHAR(50) NOT NULL, Region NVARCHAR(50) NOT NULL, Pais NVARCHAR(50) NOT NULL);</pre>
Tabla Ventas
<pre>CREATE TABLE VentaStaging (ID_Venta INT NOT NULL, ID_Producto INT NOT NULL, ID_Cliente INT NOT NULL, Fecha_Venta DATE NOT NULL, Cantidad INT NOT NULL, Total_Venta numeric(26, 2) NOT NULL);</pre>
Tabla Tiempo
<pre>CREATE TABLE TiempoStaging (Fecha DATE NOT NULL PRIMARY KEY, Año INT NOT NULL, Mes INT NOT NULL, Dia INT NOT NULL);</pre>

Consultas

Transferir datos de ventas a Staging

```
SELECT D.ID_pedido ID_Ventas, D.ID_producto ID_Producto, ID_cliente Id_Cliente,
fecha_pedido Fecha_Venta, cantidad Cantidad, (cantidad * precio_venta) Total_Venta
FROM jardineria.dbo.detalle_pedido D
INNER JOIN jardineria.dbo.pedido P ON D.ID_pedido = P.ID_pedido
INNER JOIN jardineria.DBO.producto T ON D.ID_producto = T.ID_producto;
```

Transferir datos de productos a Staging

```
SELECT ID_producto ID_Producto, Id_Categoria ID_Categoria, Nombre, Precio_Venta,
cantidad_en_stock Cantidad_Stock
FROM jardineria.dbo.producto P
INNER JOIN jardineria.dbo.Categoria_producto C ON P.Categoria = C.Id_Categoria;
```

Transferir datos de cliente a Staging

```
SELECT ID_cliente ID_Cliente, nombre_cliente Nombre_Cliente, apellido_contacto
Apellido_Cliente, ciudad Ciudad, region Region, pais Pais
FROM jardineria.dbo.Cliente;
```

Transferir datos de tiempo a Staging

```
SELECT fecha_pedido Fecha, YEAR(fecha_pedido) Año, MONTH(fecha_pedido) Mes,
DAY(fecha_pedido) Día
FROM jardineria.dbo.pedido;
```

Introducción:

Este informe describe el proceso de transformación y carga de datos desde la base de datos origen hacia la base de datos de staging y posteriormente hacia el data mart. El objetivo principal es preparar los datos para análisis, asegurando que cumplan con las necesidades analíticas y de reporte de Jardinería.

Preparación:

Revisión del Modelo Estrella:

El modelo estrella está compuesto por una tabla de hechos central y varias tablas de dimensiones. La tabla de hechos captura transacciones de ventas, mientras que las tablas de dimensiones contienen datos sobre productos, categorías, y fechas.

Relación entre tablas de dimensiones y tabla de hechos:

DimProducto (contiene información de productos)

DimCategoria (contiene información de categorías)

DimFecha (contiene información temporal)

FactVentas (contiene las transacciones de ventas)

Verificación de la Base de Datos de Staging:

La base de datos de staging se ha creado para almacenar datos transformados y limpiados antes de cargarlos en el data mart.

Tablas de staging: ClienteStaging, ProductoStaging, CategoriaStaging, PedidoStaging, DetallePedidoStaging, TiempoStaging.

Extracción de Datos:

Consultas SQL para la Extracción:

Extracción de Clientes:

```
INSERT INTO Jardineria_staging.ClienteStaging (ID_Cliente, Nombre_Cliente,
Apellido_Cliente, Ciudad, Region, Pais)
SELECT ID_Cliente, Nombre_Cliente, Apellido_Cliente, Ciudad, Region, Pais
FROM jardineria.dbo.Cliente;
```

Extracción de Productos:

```
INSERT INTO Jardineria_staging.ProductoStaging (ID_Producto, ID_Categoria,
Nombre, Precio_Venta, Cantidad_Stock)
SELECT ID_Producto, ID_Categoria, Nombre, Precio_Venta, Cantidad_Stock
FROM jardineria.dbo.Producto;
```

Transformación de Datos:

Técnicas de Transformación Aplicadas:

Limpieza de datos: Normalización de nombres de productos.

Normalización de datos: Asegurar que los valores numéricos estén en el formato correcto.

Enriquecimiento de datos: Agregar datos adicionales si es necesario.

Herramientas Utilizadas:

SQL Server Management Studio (SSMS): Para ejecutar consultas SQL y administrar las bases de datos.

Visual Studio 2022 y Microsoft Integration Services (SSIS): Para diseñar y ejecutar paquetes ETL que automatizan el proceso de extracción, transformación y carga de datos.

Resumen:

Este proyecto involucra la extracción, transformación y carga de datos desde la base de datos jardineria hasta Jardineria_staging, utilizando SQL y SSIS. El modelo estrella se utiliza para facilitar el análisis de datos y responder a preguntas clave del negocio, como identificar el producto más vendido, la categoría con más productos y el año con más ventas.

Referencias:

Modelo Estrella: Estructura del modelo estrella definida en el proyecto de análisis.

Técnicas de ETL: Aplicación de técnicas de extracción, transformación y carga de datos utilizando SQL y SSIS.

Conclusiones

Los datos de ventas se pueden acceder de manera más fácil y centralizada mediante el diseño del modelo estrella propuesto para el Data Mart de Jardinería. Esto hace que las consultas sean menos complicadas y mejore significativamente la eficiencia en la toma de decisiones.

Las dimensiones seleccionadas (Producto, Categoría, Cliente y Tiempo) permiten un análisis completo y diversificado de las ventas, lo que permite respuestas rápidas a preguntas importantes sobre el rendimiento del producto, las tendencias de los clientes y las variaciones en las ventas que ocurren a lo largo del año o la temporada.

Y por último el modelo está diseñado para ser escalable y robusto, asegurando que pueda adaptarse al crecimiento en el volumen de datos y a los cambios en las necesidades empresariales sin degradar el rendimiento.

El proceso descrito asegura que los datos estén listos para análisis en el data mart, proporcionando información valiosa para la toma de decisiones empresariales en Jardinería. El uso de SSIS y consultas SQL garantiza la automatización y precisión en la manipulación de los datos.

Anexos

Modelo Estrella

Bibliografía

<https://iudigital.instructure.com/courses/15609/files/4737082/download>

