



Evidencia de aprendizaje 3

Proceso de transformación de datos

Juan Esteban Atehortua Sanchez	PREICA2401B010076
Juan Camilo Villada González	PREICA2401B010076
Duvan Alexis Henao Gomez	PREICA2401B010093
Juliana Maria Peña Suarez	PREICA2401B010076

Bases de Datos II

PREICA2401B010076

Victor Hugo Mercado

21/04/2024

Índice

Introducción	3
Objetivos	4
Planteamiento del problema	5
Análisis del problema	6
Propuesta de solución	7
Descripción del modelo estrella propuesto	7
Lista de dimensiones propuestas	8
Conclusiones	11
Anexos	12

Introducción

En el mundo actual, donde los datos son un recurso valioso, es esencial tener una estructura de datos eficiente y efectiva. En este contexto, los modelos de data mart, como el modelo estrella, se han vuelto cada vez más populares debido a su capacidad para facilitar el análisis de datos y la toma de decisiones.

Este documento se centra en la construcción de un modelo estrella para un data mart 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, para lograr esto, utilizaremos la base de datos Jardinería como fuente de datos.

A lo largo de este trabajo, se realizará un análisis exhaustivo de la base de datos Jardinería para identificar las tablas más relevantes y sus relaciones. A partir de este análisis, se identificarán los campos necesarios para construir el modelo estrella, incluyendo la tabla de hechos y las dimensiones pertinentes.

Posteriormente, se diseñará la estructura de la tabla de hechos que represente las ventas o transacciones de la empresa, así como las dimensiones relevantes que se relacionarán con la tabla de hechos.

Finalmente se presentará una propuesta de solución que incluirá una descripción detallada del modelo estrella propuesto, un diseño visual del modelo realizado en una herramienta de diagramación, una lista de dimensiones propuestas con especificaciones detalladas de las columnas y los tipos de datos que se almacenarán y detalles que nos parezcan relevantes a nivel de la tabla de hechos.

Objetivos

- Proporcionar una guía útil para aquellos interesados en la construcción de modelos estrella para data marts. Contribuyendo a la comprensión y aplicación efectiva de estos modelos en el análisis de datos y la toma de decisiones, esto diseñando y construyendo un modelo de data mart en forma de estrella utilizando la base de datos Jardinería para facilitar el análisis de ventas y la toma de decisiones estratégicas.
- Determinar cuál es el producto que ha generado más ventas en un periodo de tiempo específico, lo que podría proporcionar información sobre la popularidad de ciertos productos entre los clientes.
- Identificar el año en que la empresa ha registrado mayor volumen de ventas, lo que podría indicar tendencias de crecimiento o estacionalidad en las ventas y ayudar en la planificación y toma de decisiones futuras.
- Identificar los campos necesarios para construir el modelo de data mart en forma de estrella.
- Diseñar la estructura de la tabla de hechos y las dimensiones relevantes.
- Identificar los campos necesarios para construir el modelo de data mart en forma de estrella.
- Diseñar la estructura de la tabla de hechos y las dimensiones relevantes.

Planteamiento del problema

La empresa de jardinería se encuentra en la necesidad de analizar su historial de ventas con el objetivo de mejorar su rendimiento y tomar decisiones estratégicas informadas. Para lograrlo, es necesario identificar el producto estrella de su catálogo, determinar la categoría con mayor cantidad de productos y establecer el año en el que se han registrado las ventas más altas. Estos datos proporcionarán una visión completa de la dinámica de ventas de la empresa, permitiendo desarrollar estrategias efectivas para maximizar sus ingresos y mantener su competitividad en el mercado.

Análisis del problema

El negocio de jardinería enfrenta un problema crítico: la falta de una herramienta adecuada para analizar y comprender la valiosa información contenida en la base de datos Jardinería. Esta situación impide la identificación de patrones, tendencias y oportunidades clave para optimizar las estrategias de negocio y la toma de decisiones.

Factores que contribuyen al problema

- **Falta de un modelo de datos:** No existe un modelo estructurado que facilite el acceso, la organización y el análisis de la información en la base de datos Jardinería.
- **Dificultad en el análisis de datos:** La gran cantidad de datos sin procesar y la falta de herramientas adecuadas dificultan la extracción de insights relevantes para el negocio.
- **Toma de decisiones basada en intuición:** Sin un análisis profundo de los datos, las decisiones estratégicas se basan en suposiciones o información incompleta, lo que aumenta el riesgo de errores.

Las consecuencias de este problema son considerables y afectan negativamente el desempeño del negocio, por eso analizaremos la estructura y relaciones entre tablas de la base de datos Jardinería, esto nos ayudará a entender cómo se almacenan los datos y dar solución a las incógnitas.

Propuesta de solución

Descripción del modelo estrella propuesto

El modelo estrella es un modelo de datos que se utiliza en los sistemas de información de gestión para organizar los datos de manera que sean más accesibles y fáciles de analizar. En este caso nuestro modelo estrella propuesta consta de cuatro dimensiones (Productos, Clientes, Ubicación y Tiempo) y una tabla de hechos (Ventas).

Diseño (Imagen)

Se creó una nueva base de datos staging, se definieron las tablas de dimensiones según el modelo estrella

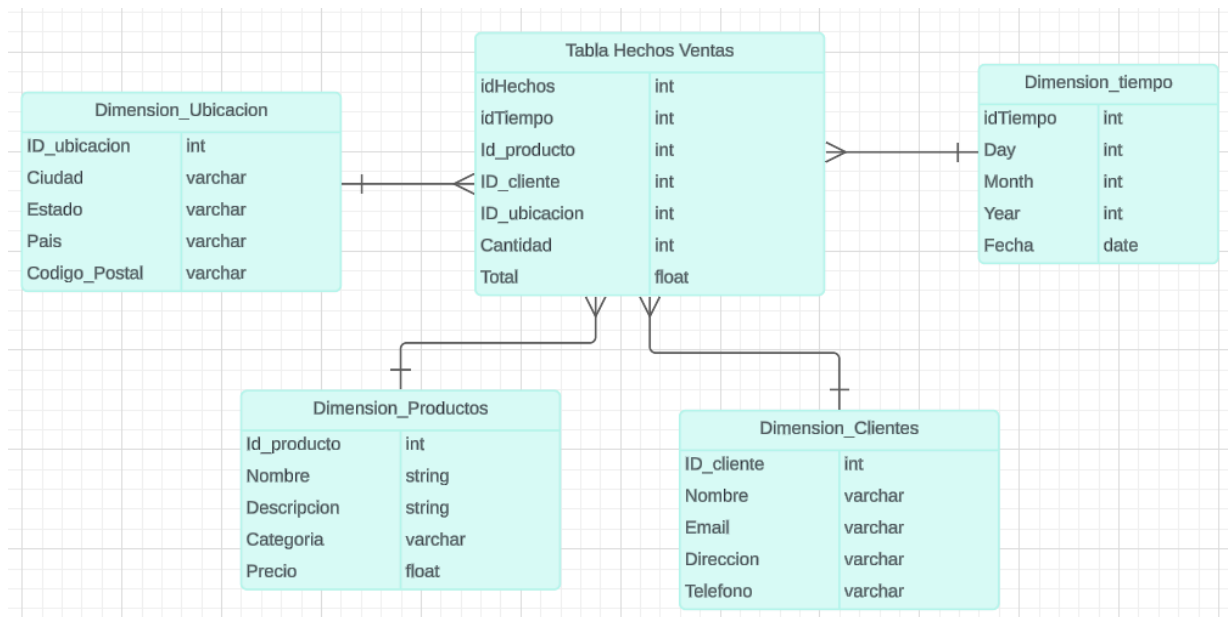
Carga de datos

los datos de la base de datos jardinería se transfirieron a la base de datos de staging

Pruebas post-migración

se verificó la integridad de los datos en la base de datos staging, se evaluó el rendimiento de las consultas

Figura 1. Diagrama de Modelo Estrella. Imagen creada por el autor



Lista de dimensiones propuestas

Dimensiones:

1. Productos

- ID_producto (int)
- Nombre (varchar)
- Descripcion (varchar)
- Categoría (varchar)
- Precio (float)

2. Clientes

- ID_cliente (int)
- Nombre (varchar)

- Email (varchar)
- Direccion (varchar)
- Teléfono (varchar)

3. Ubicación

- ID_ubicacion (int)
- Ciudad (varchar)
- Estado (varchar)
- Pais (varchar)
- Codigo_postal (varchar)

4. Tiempo

- ID_tiempo (int)
- Day (int)
- Month (int)
- Year (int)
- Fecha (date)

Hechos:

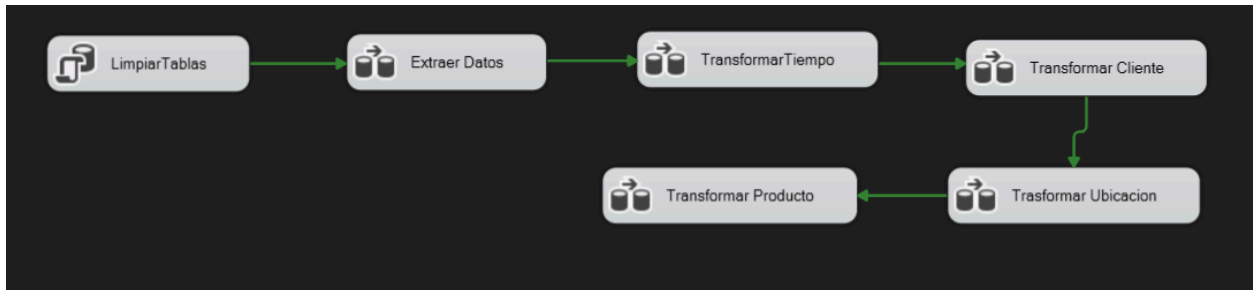
1. Ventas

- ID_venta (int)
- ID_producto (int)
- ID_cliente (int)
- ID_ubicacion (int)

- Cantidad (int)
- Total (float)

TRANSFORMACIÓN DE DATOS

Luego de la extracción de datos de la base Jardinería, se almacenaron en una base Staging , donde luego se procede a transformar la información, para que quede óptima para enviar al data mart, como se evidencia en la imagen.



En el proceso de transformación de datos se organizó para que no quedaran campos vacíos en la descripción de productos, se hizo el desglose de las fechas.

Conclusiones

- En el contexto actual, donde los datos son altamente valorados, la eficiencia en la estructura de datos es crucial para el análisis efectivo y la toma de decisiones.
- Este enfoque proporciona una estructura sólida para el análisis de datos de venta de jardinería, por ello se empleó los modelos data mart, como el modelo estrella, ya que se han vuelto cada vez más populares debido a su capacidad para facilitar el análisis de datos, mejorar el rendimiento de las consultas y toma de decisiones estratégicas.
- Con el desarrollo de este proyecto y la construcción del modelo estrella pudimos identificar que para la realización de este es necesario granular la información, determinar y establecer específicamente tablas y dimensiones según la necesidad.
- Realizar el proceso de transformación de datos y así darnos cuenta del procesamiento y optimización(limpieza de datos) de datos para su posterior uso.

Anexos

- [Enlace a diagrama http://t.ly/SqhCq](http://t.ly/SqhCq)

Bibliografia