



BASES DE DATOS II

Evidencia de Aprendizaje 3. Modelo Estrella de un Data Mart

GRUPO

PREICA2402B010070

ESTUDIANTE

DAWIN DE JESUS SALAZAR OVIEDO

Email: dawin.salazar@est.iudigital.edu.co

PROFESOR:

VICTOR HUGO MERCADO.

**INSTITUCIÓN UNIVERSITARIA DIGITAL DE ANTIOQUIA
FACULTAD DE INGENIERÍA Y CIENCIAS AGROPECUARIAS
INGENIERÍA DE SOFTWARE Y DATOS.**

MEDELLÍN

Octubre 12 2024

INTRODUCCIÓN

En este trabajo se presenta el diseño e implementación de un modelo dimensional basado en un esquema estrella para un sistema de pedidos y ventas. El enfoque se centra en la transformación y carga de datos mediante el uso de SQL Server Integration Services (SSIS), permitiendo una gestión eficiente de la información. El objetivo principal del modelo es facilitar el análisis de datos relacionado con los pedidos y las ventas de productos, así como la categorización de estos. Se busca identificar el producto más vendido, la categoría con mayor cantidad de productos y el año con más ventas.

Para facilitar este análisis, se ha desarrollado una tabla de hechos que almacena las transacciones de ventas, junto con dimensiones clave que brindan contexto a los datos. La transformación de datos se ha llevado a cabo con el fin de garantizar la calidad y consistencia de la información, permitiendo así una carga efectiva en el modelo dimensional.

OBJETIVO GENERAL

Desarrollar un modelo dimensional en esquema estrella para un data mart que permita analizar y responder de manera eficiente a tres aspectos clave: identificar el producto más vendido, determinar la categoría con mayor cantidad de productos y establecer el año con mayor volumen de ventas, utilizando como fuente de datos la base de datos de Jardinería. Para lograrlo, se implementará un proceso de transformación y carga de datos mediante SQL Server Integration Services (SSIS), garantizando la calidad y consistencia de la información en el modelo dimensional.

OBJETIVOS ESPECÍFICOS:

1. Diseñar e implementar una tabla de hechos que permita almacenar de manera estructurada la información de las transacciones de ventas, incluyendo detalles como productos vendidos, cantidades, precios y fechas, con el fin de centralizar los datos necesarios para realizar análisis detallados y precisos sobre las ventas.
2. Analizar los datos permitiendo el desglose de la información que facilite el grado de granularidad.
3. Construir las dimensiones adecuadas que permitan desglosar los datos en diferentes niveles de granularidad, como producto, categoría y tiempo, facilitando el análisis específico de cada dimensión y permitiendo obtener insights más profundos sobre el comportamiento de ventas en distintos contextos
4. Diseñar la estructura del modelo dimensional y crear una tabla de hechos que almacene las transacciones de ventas y dimensiones que proporcionen el contexto necesario para el análisis de datos, incluyendo dimensiones como producto, categoría y tiempo.
5. Implementar el proceso de transformación de datos utilizando SQL Server Integration Services (SSIS) para transformar y cargar los datos de la base de datos de Jardinería, asegurando la limpieza y la normalización de la información antes de su carga en el data mart.

Desarrollo.

Construir 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, utilizando la base de datos Jardinería como fuente de datos.

Extracción de datos desde la base de datos origen hacia la base de datos de Staging:

Utilizar consultas SQL para extraer datos relevantes de la base de datos origen y cargarlos en las tablas correspondientes de la base de datos de staging.

Verificar la integridad y consistencia de los datos extraídos para asegurar que cumplan con los requisitos del modelo Estrella.

Transformación de datos según las necesidades analíticas:

Aplicar técnicas de transformación de datos, como limpieza, normalización y enriquecimiento, para preparar los datos de acuerdo con las necesidades analíticas específicas.

Realizar la transformación de los datos utilizando consultas SQL y herramientas de ETL (Extract, Transform, Load) según sea necesario para garantizar la calidad y coherencia de los datos.

Carga de registros en el Data Mart final:

Diseñar consultas SQL o scripts de carga para insertar los registros transformados desde la base de datos de staging en las tablas del data mart final.

Ejecutar las consultas de carga y verificar que los datos se hayan insertado correctamente en el data mart final.

Documentación y presentación:

Documentar todo el proceso de transformación de datos y carga en un informe detallado que incluya una descripción de las etapas realizadas, las consultas SQL utilizadas.

Presentar el informe de manera clara y concisa, asegurándose de incluir referencias al modelo Estrella y las técnicas de ETL aplicadas.

1. Para Construir el Modelo Estrella, iniciamos con la construcción de la tabla de hechos: llamada FAC_Hechos_Pedido y su análisis correspondiente para identificar el producto más vendido, la categoría con más productos, y año con más ventas:

La tabla de hechos, contiene los datos relacionados con los pedidos y las ventas, cantidad vendida, precio unitario, fecha de las transacciones.

el total de ventas se calcula con la cantidad vendida y el precio unidad,

Las dimensiones proporcionan el contexto a la tabla de hechos para un análisis del modelo.

El proceso de transformación y carga de datos implica varios pasos fundamentales para garantizar que la información se integre correctamente en el modelo dimensional.

1. **Extracción:** Los datos se extraen de la base de datos de Jardinería, donde se recopila información relevante sobre pedidos, ventas y productos.
2. **Transformación:** En esta etapa, se aplican diversas transformaciones a los datos para asegurar su calidad y consistencia. Esto incluye la limpieza de datos, la normalización (ajuste de formatos y tipos de datos) y la agregación (suma de cantidades, cálculos de totales, etc.) para preparar los datos para el análisis.
3. **Carga:** Una vez transformados, los datos se cargan en el data mart utilizando SQL Server Integration Services (SSIS). Se asegura que los datos se inserten correctamente en las tablas de hechos y dimensiones del modelo dimensional.

Este proceso no solo facilita la organización de la información, sino que también mejora la eficiencia y la precisión del análisis de datos en el sistema de pedidos y ventas.

FAC_HECHOS_PEDIDO

PRIMARY KEY (ID_pedido), INT

PRIMARY KEY (ID_producto), INT

FOREIGN KEY (ID_pedido) REFERENCES pedido (ID_pedido),

FOREIGN KEY (ID_producto) REFERENCES producto (ID_producto)

FOREIGN KEY (ID_oficina) REFERENCES oficina (ID_oficina),

FOREIGN KEY (ID_empleado_rep_ventas) REFERENCES empleado (ID_empleado)

cantidad_vendida INTEGER

precio_unidad NUMERIC

total_venta INT, (cantidad_vendida * precio_unidad)

fecha_pedido date NOT NULL,

ANÁLISIS DE LAS DIMENSIONES:

DIMENSIÓN TIEMPO: Esta tabla permite el análisis de fechas de las transacciones agrupadas por mes, año, semestre, trimestre.

DIM_TIEMPO

PRIMARY KEY ID_fecha DATE

Año INT

Mes TINYINT

Día TINYINT

Semestre TINYINT

Trimestre TINYINT

DIMENSIÓN PRODUCTO: Esta tabla contiene los productos que se venden en el negocio. Proporcionando información de cada producto.

DIM_PRODUCTO

PRIMARY KEY ID_producto INTEGER

CódigoProducto VARCHAR(15)

nombre VARCHAR(70)

Categoría INT

dimensiones VARCHAR(25)
cantidad_en_stock SMALLINT
precio_venta NUMERIC(15,2)
FOREIGN KEY (Categoría) REFERENCES Categoria_producto (Id_Categoria)

DIMENSIÓN CATEGORÍA: Esta tabla permite el análisis por categorías de productos, proporcionando la información de las categorías de los productos mas vendidos.

DIM_CATEGORIA
PRIMARY KEY Id_Categoria INT
Desc_Categoria VARCHAR(50)
descripcion_texto TEXT,

DIMENSIÓN EMPLEADO: Esta tabla permite el análisis del empleado que más venden, tipo de productos.

DIM_EMPLEADO
PRIMARY KEY ID_empleado INT
nombre VARCHAR(50)
apellido1 VARCHAR(50)
apellido2 VARCHAR(50)
puesto VARCHAR(50)
FOREIGN KEY (ID_oficina) REFERENCES oficina (ID_oficina)

DIMENSIÓN OFICINA: Esta tabla permite el análisis por tienda geográficamente, permite el análisis por ventas y ubicación física.

DIM_OFICINA
ID_oficina INT
Descripcion VARCHAR(10)
ciudad VARCHAR(30)
pais VARCHAR(50)
region VARCHAR(50)

codigo_postal VARCHAR(10)

SCRIPT BASE DE DATOS:

```
CREATE TABLE `DIM_TIEMPO` (  
  `id_fecha` DATE,  
  `anio` INT,  
  `mes` INT,  
  `dia` INT,  
  `semestre` INT,  
  `trimestre` INT,  
  PRIMARY KEY (`id_fecha`)  
);
```

```
CREATE TABLE `DIM_CATEGORIA` (  
  `id_categoria` INT,  
  `Desc_Categoria` VARCHAR,  
  `descripcion_texto` TEXT,  
  PRIMARY KEY (`id_categoria`)  
);
```

```
CREATE TABLE `DIM_PRODUCTO` (  
  `id_producto` INT,  
  `codigoProducto` VARCHAR,  
  `nombre` VARCHAR,  
  `categoria` INT,  
  `dimensiones` VARCHAR,  
  `cantidad_en_stock` SMALLINT,
```

```
`precio_venta` NUMERIC,  
`(Categoría) REFERENCES Categoria_producto (Id_Categoria)` Type,  
PRIMARY KEY (`id_producto`)  
);
```

```
CREATE TABLE `DIM_EMPLEADO` (  
  `ID_empleado` INT,  
  `nombre` VARCHAR,  
  `apellido1` VARCHAR,  
  `apellido2` VARCHAR,  
  `puesto` VARCHAR,  
  `(ID_oficina) REFERENCES oficina (ID_oficina)` INT,  
  PRIMARY KEY (`ID_empleado`)  
);
```

```
CREATE TABLE `DIM_OFICINA` (  
  `ID_oficina` INT,  
  `Descripcion` VARCHAR,  
  `ciudad` VARCHAR,  
  `pais` VARCHAR,  
  `region` VARCHAR,  
  `codigo_postal` VARCHAR,  
  PRIMARY KEY (`ID_oficina`)  
);
```

```

CREATE TABLE `FAC_HECHOS_PEDIDO` (
  `id_Pedido` INT,
  `id_Producto` INT,
  `(ID_pedido) REFERENCES pedido (ID_pedido)` INT,
  `(ID_producto) REFERENCES producto (ID_producto)` INT,
  `(ID_oficina) REFERENCES oficina (ID_oficina)` INT,
  `(ID_empleado_rep_ventas) REFERENCES empleado (ID_empleado)` INT,
  `cantidad_vendida` INT,
  `precio_unidad` INT,
  `total_venta` INT,
  `fecha_pedido` DATE,
  PRIMARY KEY (`id_Pedido`, `id_Producto`),
  FOREIGN KEY (`(ID_empleado_rep_ventas) REFERENCES empleado
(ID_empleado)`) REFERENCES `DIM_PRODUCTO`(`(Categoría) REFERENCES
Categoria_producto (Id_Categoria)`),
  FOREIGN KEY (`(ID_oficina) REFERENCES oficina (ID_oficina)`) REFERENCES
`DIM_CATEGORIA`(`descripcion_texto`)
);

```

LINK DIAGRAMA DIMENSIONAL

https://lucid.app/lucidchart/9eef1425-2d2f-4dd1-bfa9-5b31d78bf819/edit?viewport_loc=-13678%2C-2420%2C22994%2C10254%2C0_0&invitationId=inv_7f03927c-11ba-4070-bb71-f2fbef7a866f

GGGG



Referencias

(n.d.). Astera.

<https://www.astera.com/es/knowledge-center/dimensional-modeling-guide/>

(n.d.). Modelado Dimensional Hechos y Dimensiones.

<https://verneacademy.com/blog/articulos-data/hechos-y-dimensiones-modelado-dimensional/>

(n.d.). Learn Microsoft.

<https://learn.microsoft.com/es-es/power-query/dataflows/best-practices-for-dimensional-model-using-dataflows>

IBM. (n.d.). *Modelado Dimensional*.

<https://www.ibm.com/docs/es/ida/9.1.2?topic=modeling-dimensional>