

Institución Universitaria Digital de Antioquia

Ingeniería de Software y Datos

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

Bases de Datos II (VICTOR HUGO MERCADO)

PREICA2401B010076

PREICA2401B010093

Juan David Peñaloza Hernández

Rafael de Jesus Molina Haeckermann

Danilo Ocampo Mira

Melina Muñoz Marin

Victor Hugo Mercado

Mayo – 2024

Introducción

Business Intelligence (BI) revela información para tomar decisiones estratégicas. Las herramientas de inteligencia empresarial analizan datos históricos y actuales y presentan conclusiones en formatos visuales intuitivos. (*Qué Es Business Intelligence (BI) En La Gestión De Proyectos [2024]* • Asana, 2024)

En este documento se presenta una propuesta de diseño de un modelo estrella para un data mart que permita analizar y responder a diversas consultas sobre las ventas de productos en una empresa de jardinería. El modelo propuesto se basa en las tablas proporcionadas por la empresa, y tiene como objetivo principal facilitar el análisis de datos para la toma de decisiones estratégicas.

Objetivos

El objetivo principal de este trabajo es diseñar un modelo estrella que permita analizar y responder a consultas específicas sobre las ventas de productos en la empresa de jardinería. Los objetivos específicos son:

1. Identificar las dimensiones y la tabla de hechos necesarias para el modelo estrella.
2. Desarrollar un diseño claro y comprensible del modelo estrella.
3. Proponer consultas de ejemplo que puedan ser respondidas utilizando el modelo propuesto.
- 4.

Planteamiento del Problema

La empresa de jardinería necesita mejorar su capacidad de análisis de datos para tomar decisiones más informadas sobre su negocio. Actualmente, carece de un modelo de datos adecuado que le permita analizar las ventas de productos de manera eficiente y precisa.

Dada la falta de un modelo de datos adecuado, la empresa de jardinería enfrenta desafíos significativos en la comprensión de sus ventas de productos. Sin un sistema estructurado para analizar estos datos, se dificulta la identificación de tendencias, la evaluación del rendimiento de los productos, la detección de patrones de compra de los clientes y la toma de decisiones estratégicas basadas en información concreta.

Un modelo de datos sólido permitiría a la empresa recopilar, organizar y analizar datos relevantes sobre las ventas de sus productos. Esto incluiría información

detallada sobre qué productos se venden con más frecuencia, en qué áreas geográficas se realizan las ventas más exitosas, cuáles son los períodos de mayor demanda, y otros factores que impactan en el rendimiento de ventas.

Además, al mejorar la capacidad de análisis de datos, la empresa podría identificar oportunidades de crecimiento, optimizar su inventario, desarrollar estrategias de marketing más efectivas y mejorar la satisfacción del cliente al ofrecer productos y servicios más alineados con sus necesidades y preferencias.

En resumen, la implementación de un modelo de datos adecuado proporcionará a la empresa de jardinería una base sólida para mejorar su comprensión del mercado, optimizar sus operaciones y tomar decisiones más informadas y estratégicas para impulsar el éxito del negocio.

Análisis del Problema

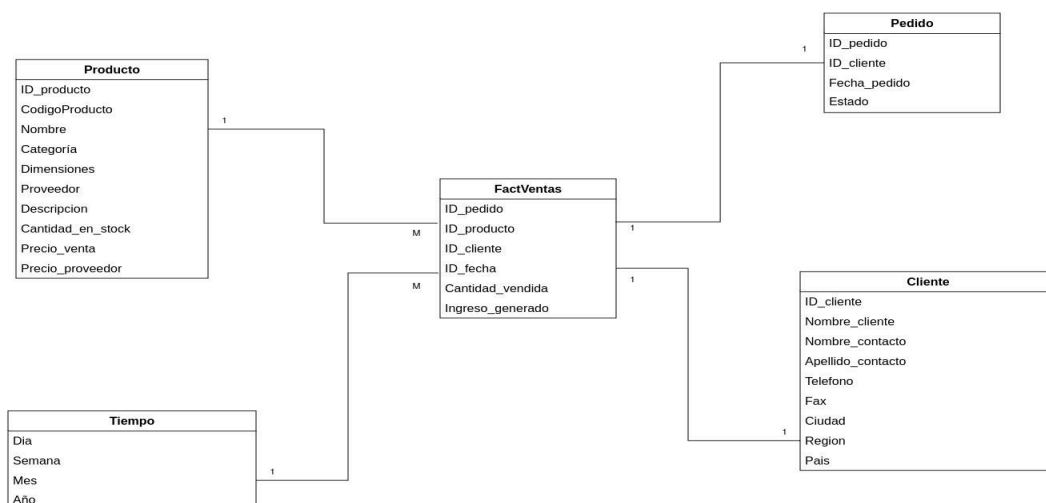
El principal problema radica en la falta de una estructura de datos adecuada para analizar las ventas de productos. Las tablas proporcionadas por la empresa no están optimizadas para este fin, lo que dificulta el análisis y la obtención de información relevante.

Propuesta de la Solución

Descripción del Modelo Estrella Propuesto

El modelo estrella propuesto consta de una tabla de hechos central, FactVentas, rodeada de varias tablas de dimensiones que proporcionan contexto para los datos de ventas. Las dimensiones propuestas incluyen Producto, Tiempo, Cliente y Pedido.

Diseño del Modelo Estrella



<https://drive.google.com/file/d/17p2yRtWCmYE9ZI98MHpVhDgseJMTTzb7/view?usp=sharing>

Lista de Dimensiones Propuestas

Dimensión Producto

ID_producto (Entero)
CodigoProducto (Texto)
Nombre (Texto)
Categoría (Texto)
Dimensiones (Texto)
Proveedor (Texto)
Descripcion (Texto)
Cantidad_en_stock (Entero)
Precio_venta (Decimal)
Precio_proveedor (Decimal)

Dimensión Tiempo

Dia (Entero)
Mes (Texto)
Año (Entero)
Semana (Entero)

Dimension Cliente

ID_cliente (Entero)
Nombre_cliente (Texto)
Nombre_contacto (Texto)
Apellido_contacto (Texto)
Telefono (Texto)
Fax (Texto)
Ciudad (Texto)
Region (Texto)
Pais (Texto)

Dimensión Pedido

ID_pedido(Entero)
ID_cliente(Entero)
Fecha_Pedido(Fecha)
Estado(Entero)

Tabla de Hechos (FactVentas)

ID_pedido (Entero)
ID_producto (Entero)
ID_cliente (Entero)
ID_Fecha (Fecha)
Cantidad_vendida (Entero)
Ingreso_generado (Decimal)

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 jardinería es un paso crucial en el proceso de integración de datos, que implica comprender la estructura y la calidad de los datos disponibles en las diferentes tablas, incluyendo Producto, Tiempo, Cliente y Pedido. Este análisis abarca aspectos como la integridad de los datos, la consistencia de los formatos, la presencia de datos faltantes o duplicados, y la identificación de posibles relaciones entre las tablas.

Proceso de Traslado a la Base de Datos Staging

Creación del Proyecto en Visual Studio: Se inicia el proceso creando un nuevo proyecto de Integration Services en Visual Studio.

Conexión a la Base de Datos de Origen (Jardinería): Se establecen conexiones a la base de datos de jardinería desde donde se extraerán los datos. Se configuran conexiones individuales para cada una de las tablas (Producto, Tiempo, Cliente, Pedido).

Configuración de Componentes de Extracción: Se utilizan componentes de SSIS para extraer datos de las tablas de origen. Consultas SQL o vistas pueden ser empleadas como origen de datos para cada tabla.

Transformación de Datos: Los datos extraídos pueden requerir transformaciones, como limpieza, conversión de tipos de datos o aplicaciones de reglas de negocio. Esto se lleva a cabo mediante transformaciones SSIS como "Derived Column", "Lookup", o "Merge Join".

Carga en la Base de Datos Staging: Después de las transformaciones necesarias, los datos se cargan en las tablas correspondientes de la base de datos Staging. Para cada tabla, se utilizan componentes de destino SSIS.

Ejecución y Monitoreo del Paquete SSIS: Una vez configurado el flujo de trabajo de SSIS, se ejecuta el paquete para transferir los datos de jardinería a la base de datos Staging. Durante la ejecución, se monitorea el progreso y se verifica la correcta transferencia de datos.

Validación de los Datos en la Base de Datos Staging: Posterior a la carga, se realiza una validación para asegurar que los datos se hayan trasladado correctamente. Esto implica comparar recuentos de filas, verificar la integridad referencial y revisar datos específicos para confirmar su coherencia.

BK y documentos de la queries

<https://drive.google.com/drive/folders/1Y6XO1doV04ax49HcrxM433DzqRbcZBRK?usp=sharing>

Conclusiones

El diseño del modelo estrella propuesto proporciona una estructura clara y eficiente para analizar las ventas de productos en la empresa de jardinería. La implementación de este modelo permitirá a la empresa obtener información valiosa para la toma de decisiones estratégicas.

Registro data mart.

Descripción de Etapas.

- Se revisó el modelo estrella para comprender la estructura y relaciones entre las tablas de dimensiones y la tabla de hechos. Se verificó la disponibilidad y consistencia de la base de datos de staging (`stagin`).
- Se utilizaron consultas SQL para extraer datos relevantes desde la base de datos `jardineria` hacia las tablas correspondientes en `stagin`.

- Se aplicaron técnicas de limpieza, normalización y cálculo de campos derivados para preparar los datos según las necesidades analíticas. Por ejemplo, se normalizaron los nombres de clientes y se calcularon ingresos generados en ventas.

Carga en el Data Mart.

Se diseñaron consultas SQL para cargar los datos transformados desde `stagin` hacia las tablas del Data Mart final `DataMart`. Se aseguró la integridad y consistencia de los datos durante el proceso.

Consultas SQL Utilizadas.

- ETL

Extracción de datos: se realizó la extracción de datos de las tablas de 'stagin' de las tablas: BD_CLIENT, TB_PRODUCTOS, TB_time y factVentas.

Transformación: en esta fase se puede realizar limpieza de datos y agregar datos, en nuestro caso no realizamos ninguna transformación específica, peor a

continuación un ejemplo en query:

```
SQLQuery2.sql - D:\IMIOCSF\USER (65))* -> X
USE DataMart
UPDATE DataMart.DimTiempo
SET fecha = DATEADD(day, dia - 1, DATEFROMPARTS(año, MONTH(CAST('01 ' + mes AS DATETIME)), 1));
--en este caso "fehca" sale en rojo porque en nuestra tabla no esta definida,
-- pero es un ejemplo de transformación, de como lo podriamos realizar
```

Carga: por último, insertamos los datos transformados en las tablas correspondientes del Data Mart.

```
SQLQuery2.sql - D:\IMIOCSF\USER (65))* -> X
USE DataMart
INSERT INTO DataMart.DimTiempo (id, dia, semana, mes, año)
SELECT id, dia, semana, mes, año
FROM [stagin].dbo.TB_time;
```

- verificamos la existencia de tablas en el data mart:

Query

```
SQLQuery1.sql - D:\IMIOCSF\USER (55))* -> X
USE DataMart;
SELECT name AS Table_Name
FROM sys.tables
ORDER BY name;
```

Respuesta

Results		Messages
	Table_Name	
1	DimCliente	
2	DimProducto	
3	DimTiempo	
4	FactVentas	

- verificamos la existencia de datos en el data mart:

Query

```
SQLQuery1.sql - D...IMIOCSF\USER (55))* -p X
USE DataMart
-- Verificar el número de filas en la tabla DimCliente
SELECT COUNT(*) AS TotalFilas
FROM DataMart.DimCliente;
```

Respuesta

Results		Messages
	TotalFilas	
1	36	

Query

```
SQLQuery1.sql - D...IMIOCSF\USER (55))* -p X
USE DataMart
-- Verificar el número de filas en la tabla DimProducto
SELECT COUNT(*) AS TotalFilas
FROM DataMart.DimProducto;
```

Respuesta

Results		Messages
	TotalFilas	
1	5	

Query

```
SQLQuery1.sql - D...\IMIOCSF\USER (55))* -p X
USE DataMart
-- Verificar el número de filas en la tabla DimTiempo
SELECT COUNT(*) AS TotalFilas
FROM DataMart.DimTiempo;
```

Respuesta

	TotalFilas
1	3

Query

```
SQLQuery1.sql - D...\IMIOCSF\USER (55))* -p X
USE DataMart
-- Verificar el número de filas en la tabla FactVentas
SELECT COUNT(*) AS TotalFilas
FROM DataMart.FactVentas;
```

Respuesta

	TotalFilas
1	7

- verificamos datos específicos en el data mart:

Query

```
SQLQuery1.sql - D...\IMIOCSF\USER (55))* -p X
USE DataMart
-- Validar la existencia de datos de ventas en la tabla FactVentas
SELECT TOP 10 *
FROM DataMart.FactVentas;
```

Respuesta

	ID_detalle_pedido	ID_producto	ID_cliente	fecha_pedido	cantidad_venta	ingreso_generado
1	1	11679	7	2009-01-06	5	70.00
2	2	11679	9	2009-01-20	5	70.00
3	3	11679	5	2008-12-30	14	196.00
4	4	11679	4	2008-03-10	1	14.00
5	5	11679	19	2009-03-12	5	70.00
6	6	11679	13	2009-01-05	3	42.00
7	7	11679	30	2009-04-06	12	168.00

Query

```
SQLQuery1.sql - D:\IMIOCSF\USER (55))*
USE DataMart
-- Validar la existencia de datos de ventas en la tabla DimCliente
SELECT TOP 10 *
FROM DataMart.DimCliente;
```

Respuesta

	id_cliente	nombre_cliente	nombre_contacto	apellido_contacto	telefono	fax	linea_direccion1	linea_direccion2	ciudad	region	pais	codigo_postal	ID_empleado_rep_ventas	limite_credito
1	1	GOLDFISH GARDEN	DANIEL G	GOLDFISH	5556901745	5556901746	False Street 52 2 A	NULL	San Francisco	NULL	USA	24006	19	3000.00
2	2	GARDENING ASSOCIATES	ANNE	WRIGHT	5557410345	5557410346	Wall-e Avenue	NULL	Miami	Miami	USA	24010	19	6000.00
3	3	GERUDO VALLEY	LINK	FLAUTE	5552323129	5552323128	Oaks Avenue n°22	NULL	New York	NULL	USA	85495	22	12000.00
4	4	TENDO GARDEN	AKANE	TENDO	55591233210	55591233211	Null Street n°69	NULL	Miami	NULL	USA	696969	22	600000.00
5	5	LASAS S.A.	ANTONIO	LASAS	34916540145	34914851312	C/Leganés 15	NULL	Fuenlabrada	Madrid	Spain	28945	8	154310.00
6	6	BERAGUA	JOSE	BERMEJO	654987321	916549872	C/pintor segundo	Getafe	Madrid	Madrid	Spain	28942	11	20000.00
7	7	CLUB GOLF PUERTA DEL HIERRO	PACO	LOPEZ	62456810	919535678	C/sinesio delgado	Madrid	Madrid	Madrid	Spain	28930	11	40000.00
8	8	NATURAGUA	GUILLERMO	RENGIFO	689234750	916428956	C/majadahonda	Boadilla	Madrid	Madrid	Spain	28947	11	32000.00
9	9	DARADISTRIBUCIONES	DAVID	SERRANO	675598001	916421756	C/szoree	Fuenlabrada	Madrid	Madrid	Spain	28946	11	50000.00
10	10	MADRILEÑA DE RIEGOS	JOSE	TACAÑO	655983045	916689215	C/Lagañas	Fuenlabrada	Madrid	Madrid	Spain	28943	11	20000.00

- verificamos la consultas de análisis en el data mart:

Query

```
SQLQuery1.sql - D:\IMIOCSF\USER (55))* -p X
USE DataMart
-- Consulta para obtener el total de ingresos generados por cliente
SELECT
    dc.nombre_cliente,
    SUM(fv.ingreso_generado) AS TotalIngresos
FROM DataMart.DimCliente dc
JOIN DataMart.FactVentas fv ON dc.ID_cliente = fv.ID_cliente
GROUP BY dc.nombre_cliente;
```

Respuesta

	nombre_cliente	TotalIngresos
1	AMERICH GOLF MANAGEMENT SL	70.00
2	CAMPOHERMOSO	168.00
3	CLUB GOLF PUERTA DEL HIERRO	70.00
4	DARADISTRIBUCIONES	70.00
5	DARDENA S.A.	42.00
6	LASAS S.A.	196.00
7	TENDO GARDEN	14.00

Conclusiones finales.

En conclusión, la creación del Data Mart nos permitió centralizar y organizar los datos de diferentes fuentes en un único repositorio. Con las tablas DimCliente, DimProducto, DimTiempo y FactVentas, ahora es posible realizar análisis detallados sobre las ventas un gran factor que nos ayuda a identificar patrones de comportamiento (tanto en clientes como productos) para un buen marketing y el proceso ETL nos permitió que los datos fueran precisos, garantiza que los datos sean fiables.

Repositorio en GitHub:

<https://github.com/dev-juanda01/integrador-bd2-iudigital>

Bibliografía

Qué es Business Intelligence (BI) en la gestión de proyectos [2024] • Asana. (2024, February 10). Asana. Retrieved April 21, 2024, from <https://asana.com/es/resources/business-intelligence>

Qué es la inteligencia empresarial | Microsoft Power BI. (n.d.). Power BI. Retrieved April 21, 2024, from <https://powerbi.microsoft.com/es-es/what-is-business-intelligence/>