

S30 - Evidencia de aprendizaje 2. Creación de una base de datos de Staging

Juan David Maya Tobon

CC 1.035.831.285

Grupo 94

Docente

Victor Hugo Mercado Ramos

Bases de Datos II

IU Digital de Antioquia

Ingeniería de Software y Datos

11 de mayo de 2024

Introducción

En el ámbito de la gestión de bases de datos, el modelo estrella emerge como un pilar fundamental en el diseño multidimensional, especialmente valioso para estructurar datos en entornos empresariales complejos, este enfoque adquiere una importancia notable en sistemas de información que demandan un análisis profundo de datos, como las transacciones comerciales de empresas, donde la organización eficiente y el acceso ágil a la información se vuelven críticos.

En este estudio, profundizaremos en la comprensión y aplicación del modelo estrella dentro del contexto de una base de datos enfocada en el sector de la jardinería, explorando sus elementos clave y evaluaremos su capacidad para mejorar la gestión de datos empresariales en este sector específico, destacando su potencial para impulsar decisiones estratégicas informadas y optimizar procesos de análisis de información.

Además, examinaremos el proceso de traslado de datos hacia una base de datos de preparación (staging), un componente esencial en la arquitectura de datos para garantizar la integridad y consistencia de la información antes de su posterior análisis y utilización.

Objetivos

Diseñar e implementar la estructura de un modelo estrella en la base de datos de la empresa de jardinería, optimizando la gestión y análisis de datos organizacionales.

Establecer relaciones claras entre la tabla de hechos de ventas y las dimensiones pertinentes, como clientes, productos, tiempo y pagos, garantizando la integridad referencial y la consistencia de la información.

Mejorar la eficiencia en la consulta y extracción de datos mediante la implementación de índices y optimización de consultas, permitiendo un acceso rápido y preciso a la información requerida para análisis y reportes.

Establecer una estructura organizada y coherente en la base de datos, facilitando el análisis y la gestión de la información relacionada con ventas, clientes, productos y transacciones.

Diseñar e implementar un proceso eficiente de traslado de datos hacia una base de datos staging, con el fin de garantizar la integridad, calidad y consistencia de la información, optimizando así los tiempos de carga y preparación para su posterior análisis y utilización en el contexto de un sistema de gestión empresarial.

Desarrollar un flujo de trabajo automatizado y seguro para el traslado de datos a una base de datos staging, con el propósito de asegurar la correcta transformación, limpieza y preparación de los datos provenientes de diversas fuentes, facilitando así su integración y disponibilidad para su posterior análisis y procesamiento en un entorno empresarial.

Planteamiento del problema

La necesidad imperante de optimizar la gestión de información en la base de datos de la empresa de jardinería, surge tras la ausencia de una estructura organizada, relaciones claras entre los datos y la presencia de inconsistencias; aspectos que dificultan el análisis de ventas, seguimiento de clientes y toma de decisiones estratégicas. Dichos factores, tras ser analizados plantean el objetivo de desarrollar e implementar un modelo estrella que permita gestionar de manera eficiente los datos de ventas, clientes, productos y transacciones; con el fin de mejorar la precisión y actualización de la información, facilitando así, la toma de decisiones fundamentadas en datos confiables y oportunos.

Análisis del problema

La implementación del modelo estrella surge de la necesidad de mejorar la gestión de la información en la base de datos de la empresa de jardinería, donde actualmente, la estructura de la base de datos presenta ciertas deficiencias en términos de organización y relaciones de datos; aspectos que dificultan el análisis integral de las ventas, el seguimiento detallado de los clientes y la toma de decisiones estratégicas; lo que impacta directamente la capacidad de la empresa para obtener información precisa y oportuna sobre sus operaciones comerciales, afectando a su vez, la eficiencia en la toma de decisiones y el rendimiento general del negocio.

Al adoptar el modelo estrella, se busca establecer una estructura más robusta y coherente en la base de datos, ya que este modelo se caracteriza por tener una tabla de hechos central que registra las transacciones comerciales clave, como las ventas, y dimensiones relacionadas que contienen atributos descriptivos detallados, como información de clientes, productos y tiempos; lo que permite una organización más eficiente de los datos y una relación clara entre ellos, facilitando así la extracción de información relevante para análisis y reportes.

Con la implementación del modelo estrella, se espera mejorar significativamente la capacidad de la empresa para analizar y comprender sus datos comerciales, lo que incluye la capacidad de identificar tendencias de ventas, segmentar clientes de manera más precisa, evaluar el rendimiento de productos específicos y realizar proyecciones financieras más precisas; en última instancia, el objetivo es mejorar la toma de decisiones estratégicas y optimizar los procesos internos, aumentando la competitividad y el éxito general de la empresa.

Propuesta de la solución

Descripción del modelo estrella propuesto

El modelo estrella propuesto para la base de datos JardineriaModeloEstrella, es una estructura organizativa avanzada, que sigue una metodología de diseño multidimensional; la cual es crucial para gestionar eficientemente grandes volúmenes de datos, especialmente en el contexto de las complejas transacciones de ventas de una empresa de jardinería.

Dimensiones Propuestas

Dimensión Tiempo: Esta dimensión aborda el aspecto temporal de las transacciones de ventas, proporcionando campos como fecha_pedido, anio, mes, trimestre y dia_semana, estos atributos permiten análisis temporales detallados, como tendencias de ventas estacionales y comportamiento de compra a lo largo del tiempo.

Dimensión Cliente: Aquí se recopila información detallada sobre los clientes, incluyendo ID_cliente, nombre_cliente, datos de contacto, dirección y límite de crédito, lo que facilita la segmentación de clientes, seguimiento de relaciones comerciales y evaluación del rendimiento de ventas por cliente.

Dimensión Producto: Esta dimensión describe los productos vendidos, proporcionando datos como ID_producto,CodigoProducto, nombre, categoría, dimensiones, proveedor, precios y stock, estos detalles son esenciales para analizar la popularidad de productos, tendencias de inventario y relación con proveedores.

Dimensión Categoría Producto: En esta dimensión se registra información clave relacionada con las categorías de productos presentes en las transacciones, incluyendo el ID de categoría y el nombre de la categoría. Esta información es fundamental para el análisis de

ventas, seguimiento de la demanda por categoría de producto y gestión eficiente del inventario.

La tabla de hechos (hechos_pedido): Actúa como el núcleo central que integra las dimensiones mencionadas anteriormente, debido a que registra detalles clave de cada transacción de pedido, como ID_cliente, ID_producto, ID_pago, ID_tiempo, cantidad, precio_unitario y total; la implementación de restricciones de integridad referencial garantiza la consistencia y coherencia de los datos entre las dimensiones y la tabla de hechos.

Este diseño del modelo estrella proporciona una estructura sólida y eficiente para realizar consultas complejas y análisis multidimensionales; lo que permite a la empresa de jardinería, comprender mejor sus operaciones comerciales, identificar tendencias de ventas, analizar el comportamiento del cliente y tomar decisiones estratégicas informadas basadas en datos precisos y oportunos.

Lista de dimensiones propuestas

Dimensión Tiempo

Campos: fecha_pedido, anio, mes, trimestre, dia_semana.

Motivo de uso: Permite analizar las ventas y transacciones en función del tiempo, como identificar tendencias estacionales, patrones de compra mensuales o comparar el desempeño trimestral.

Dimensión Cliente

Campos: 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.

Motivo de uso: Facilita el análisis de ventas por cliente, permite segmentar clientes por ubicación geográfica, gestión de crédito, y asignar representantes de ventas.

Dimensión Producto

Campos: ID_producto, CodigoProducto, nombre, Categoria, dimensiones, proveedor, descripcion, cantidad_en_stock, precio_venta, precio_proveedor.

Motivo de uso: Permite analizar el desempeño de productos en ventas, gestionar el inventario, conocer proveedores y categorizar los productos para análisis detallados.

Dimensión Pago

Campos: ID_pago, ID_cliente, forma_pago, id_transaccion, fecha_pago, total.

Motivo de uso: Permite analizar los pagos asociados a las transacciones, identificar patrones de pago, formas de pago preferidas por los clientes y gestionar registros financieros.

Descripción del análisis de datos realizado para el sector de Jardinería y el proceso de traslado a la base de datos staging

En el contexto del sector de Jardinería, se realizó un análisis exhaustivo de datos relacionados con transacciones comerciales, productos, clientes y otros aspectos relevantes para la gestión empresarial en este sector. El objetivo principal fue entender y mejorar la eficiencia operativa, la toma de decisiones y la planificación estratégica mediante el uso inteligente de los datos.

El análisis de datos incluyó las siguientes etapas:

Identificación de fuentes de datos: Se identificaron las fuentes de datos clave, que podrían incluir sistemas de ventas, inventarios, información de clientes, entre otros.

Extracción de datos: Se extrajeron los datos relevantes de estas fuentes, utilizando técnicas adecuadas como consultas SQL u otras herramientas de extracción de datos.

Limpieza y transformación de datos: Los datos extraídos fueron sometidos a un proceso de limpieza y transformación para asegurar la consistencia, integridad y calidad de los mismos. Esto incluyó eliminar duplicados, corregir errores y homogeneizar formatos.

Análisis exploratorio: Se realizó un análisis exploratorio de los datos para identificar patrones, tendencias y relaciones entre variables, lo cual permitió obtener información valiosa sobre el comportamiento del mercado, preferencias de los clientes y desempeño de productos.

Definición de la estructura de la base de datos staging: Basado en los hallazgos del análisis y los requerimientos del negocio, se diseñó la estructura de la base de datos staging. Esto implicó definir las tablas, columnas y relaciones necesarias para almacenar los datos de manera óptima para su posterior análisis.

**El traslado de datos a la base de datos staging
se llevó a cabo de la siguiente manera:**

Creación de la base de datos staging: Se creó la base de datos staging en un entorno adecuado para almacenar temporalmente los datos antes de su carga final en el sistema de gestión.

Carga inicial de datos: Los datos limpios y transformados fueron cargados inicialmente en la base de datos staging utilizando procesos de carga masiva o inserción de registros.

Validación de datos: Se realizaron pruebas de validación para asegurar que los datos cargados en la base de datos staging fueran consistentes y completos, y cumplieran con los estándares establecidos.

Preparación para análisis: Una vez que los datos estuvieron en la base de datos staging, se realizaron operaciones adicionales de preparación según las necesidades específicas del análisis a realizar, como la agregación de datos, creación de vistas o índices, entre otros.

Integración con herramientas de análisis: Finalmente, los datos preparados en la base de datos staging estuvieron listos para ser utilizados con herramientas de análisis de datos como SQL Server Analysis Services (SSAS), , permitiendo a los usuarios realizar análisis avanzados y generar informes interactivos.

Consultas verificación traslado base de datos Staiging

Consulta base de datos DestinoTbClienteJr

Connect

DESKTOP-SISTEMA (SQL Server 16.0.1000.6 - DESKTOP-SISTEMA\Er

Databases

System Databases

Database Snapshots

jardineria

JardineriaModeloEstrella

jardineriaStaging

Security

Server Objects

Replication

Always On High Availability

Management

Integration Services Catalogs

SQL Server Agent (Agent XPs disabled)

XEvent Profiler

-- Tabla Cliente

SELECT

ID_cliente,

nombre_cliente,

telefono,

ciudad,

pais,

ID_empleado_rep_ventas

FROM

jardineriaStaging.dbo.DestinoTbClienteJr

ORDER BY

nombre_cliente ASC;

100 %

Results

Messages

ID_cliente	nombre_cliente	telefono	ciudad	pais	ID_empleado_rep_ventas
1	Agrojardin	675432926	Getafe	Spain	30
2	Aloha	916485852	Canarias	Spain	12
3	Americh Golf Management SL	964493072	Barcelona	Spain	12
4	Beragua	654987321	Madrid	Spain	11
5	Campohermoso	645925376	Fuenlabrada	Spain	30
6	Camunas Jardines S.L.	34914873241	San Lorenzo del Escorial	Spain	8
7	Club Golf Puerta del hierro	62456810	Madrid	Spain	11
8	DaraDistribuciones	675598001	Madrid	Spain	11
9	Dardena S.A.	34912453217	Madrid	Spain	8

Consulta base de datos DestinoTbEmpleadoJr

Connect

DESKTOP-SISTEMA (SQL Server 16.0.1000.6 - DESKTOP-SISTEMA\Er

Databases

System Databases

Database Snapshots

jardineria

JardineriaModeloEstrella

jardineriaStaging

Security

Server Objects

Replication

Always On High Availability

Management

Integration Services Catalogs

SQL Server Agent (Agent XPs disabled)

XEvent Profiler

-- Tabla Empleado

SELECT

ID_empleado,

nombre,

apellido1,

email,

ID_oficina

FROM

jardineriaStaging.dbo.DestinoTbEmpleadoJr

ORDER BY

ID_empleado ASC;

100 %

Results

Messages

ID_empleado	nombre	apellido1	email	ID_oficina
1	Marcos	Magaña	marcos@jardineria.es	8
2	Ruben	López	rlopez@jardineria.es	8
3	Alberto	Soria	asoria@jardineria.es	8
4	Maria	Solís	msolis@jardineria.es	8
5	Felipe	Rosas	frosas@jardineria.es	8
6	Juan Carlos	Ortiz	cortiz@jardineria.es	8
7	Carlos	Soria	csoria@jardineria.es	4
8	Mariano	López	mlopez@jardineria.es	4
9	Lucio	Campoamor	lcampoamor@jardineria.es	4
10	Hilario	Rodriguez	hrodriguez@jardineria.es	4
11	Emmanuel	Magaña	manu@jardineria.es	1
12	José Manuel	Martinez	jmmart@hotmail.es	1
13	David	Palma	dpalma@jardineria.es	1
14	Oscar	Palma	opalma@jardineria.es	1

Consulta base de datos DestinoTbProductoJr

Connect

DESKTOP-SISTEMA (SQL Server 16.0.1000.6 - DESKTOP-SISTEMA\Er

Databases

System Databases

Database Snapshots

jardineria

JardineriaModeloEstrella

jardineriaStaging

Security

Server Objects

Replication

Always On High Availability

Management

Integration Services Catalogs

SQL Server Agent (Agent XPs disabled)

XEvent Profiler

--Consultas para la generacion de bases de datos Staging

-- Tabla Producto

SELECT

ID_producto,

nombre,

Categoria,

cantidad_en_stock,

precio_venta,

precio_proveedor

FROM

jardineriaStaging.dbo.DestinoTbProductoJr

ORDER BY

ID_producto ASC;

100 %

Results Messages

	ID_producto	nombre	Categoria	cantidad_en_stock	precio_venta	precio_proveedor
1	1	Sierra de Poda 400MM	2	15	14.00	11.00
2	2	Pala	2	15	14.00	13.00
3	3	Rastrillo de Jardín	2	15	12.00	11.00
4	4	Azadón	2	15	12.00	11.00
5	5	Ajedrea	3	140	1.00	0.00
6	6	Lavándula Dentata	3	140	1.00	0.00
7	7	Mejorana	3	140	1.00	0.00
8	8	Melissa	3	140	1.00	0.00
9	9	Mentha Sativa	3	140	1.00	0.00
10	10	Petrosilium Hortense (Peregil)	3	140	1.00	0.00
11	11	Salvia Mix	3	140	1.00	0.00
12	12	Thymus Citriodra (Tomillo limón)	3	140	1.00	0.00
13	13	Thymus Vulgaris	3	140	1.00	0.00
14	14	Santolina Chamacynarus	3	140	1.00	0.00

Consulta base de datos DestinoTbOficinaJr

Connect

DESKTOP-SISTEMA (SQL Server 16.0.1000.6 - DESKTOP-SISTEMA\Er

Databases

System Databases

Database Snapshots

jardineria

JardineriaModeloEstrella

jardineriaStaging

Security

Server Objects

Replication

Always On High Availability

Management

Integration Services Catalogs

SQL Server Agent (Agent XPs disabled)

XEvent Profiler

-- Tabla Oficina

SELECT

ID_oficina,

Descripcion,

ciudad,

telefono

FROM

jardineriaStaging.dbo.DestinoTbOficinaJr;

100 %

Results Messages

	ID_oficina	Descripcion	ciudad	telefono
1	1	BCN-ES	Barcelona	+34 93 3561182
2	2	BOS-USA	Boston	+1 215 837 0825
3	3	LON-UK	Londres	+44 20 78772041
4	4	MAD-ES	Madrid	+34 91 7514487
5	5	PAR-FR	Paris	+33 14 723 4404
6	6	SFC-USA	San Francisco	+1 650 219 4782
7	7	SYD-AU	Sydney	+61 2 9264 2451
8	8	TAL-ES	Talavera de la Reina	+34 925 867231
9	9	TOK-JP	Tokyo	+81 33 224 5000
10	10	BCN-ES	Barcelona	+34 93 3561182

Conclusiones

La implementación del modelo estrella en la base de datos de la empresa de jardinería, ha demostrado ser altamente efectiva para mejorar la eficiencia en el análisis de datos; debido a que su estructura multidimensional facilita la realización de consultas complejas y el análisis detallado de información, lo que permite a la empresa obtener insights valiosos sobre sus operaciones y clientes.

Gracias al modelo estrella, se ha observado una mejora significativa en la capacidad de la empresa para tomar decisiones estratégicas basadas en datos precisos y actualizados, gracias a la integración de dimensiones como tiempo, cliente, producto y pago se ha proporcionado una visión holística de las transacciones comerciales, lo que ha permitido identificar tendencias, patrones y oportunidades de mejora.

El diseño del modelo estrella, ha contribuido a una mayor agilidad en la generación de informes y análisis ad hoc; además, la estructura optimizada de la base de datos ha reducido los tiempos de respuesta en la obtención de información, lo que ha mejorado la capacidad de la empresa para responder rápidamente a las necesidades del negocio y a los cambios en el mercado.

Referencias

- García, J. (2013). *DISEÑO DATA WAREHOUSE: HECHOS Y DIMENSIONES; MODELO ESTRELLA VS COPO DE NIEVE*. Obtenido de mundodb: <https://mundodb.es/disenodata-warehouse-hechos-y-dimensiones-modelo-estrella-vs-copo-de-nieve>
- Rodriguez Afaro, F. (2019). *El modelo de estrella. El pilar fundamental del Business Intelligence*. Obtenido de datamanagement: <https://datamanagement.es/2019/06/27/business-intelligence-modelo-estrella/>
- Sánchez, J., & Canelo , C. (2017). *MODELO DE DATA WAREHOUSE CON APLICACION DE INTELIGENCIA DE NEGOCIOS PARA LAS PYMES*. Obtenido de Universidad Nacional Jorge Basadre Grohmann: <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwi-zryUsuaFAxWNSDABHShGDvU4RhAWegQIDxAB&url=https%3A%2F%2Frevistas.unjbg.edu.pe%2Findex.php%2Fcyd%2Farticle%2Fdownload%2F737%2F749%2F1384&usg=AOvVaw0On9M4FBKuUhy1Avd1W>
- Zerpa , H., García , R., & Izquierdo , H. (2020). *DATAMART BASADO EN EL MODELO ESTRELLA PARA LA IMPLEMENTACIÓN DE INDICADORES CLAVE DE DESEMPEÑO COMO SALIDA DEL BIG DATA*. Obtenido de Universidad Nacional Experimental Politécnica: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwj9KHtsOaFAxUGSDABHd-sBtAQFnoECDcQAQ&url=https%3A%2F%2Ftunexpo.autanabooks.com%2Findex.php%2Ftunexpo%2Farticle%2Fdownload%2F342%2F618%2F&usg=AOvVaw0xm28ke_Kas10lQrvxcXl