

EA4. CARGA AL DATA MART FINAL Y REPOSITORIO DE TODAS LAS

ACTIVIDADES

Jhon Alexis Machado Rodríguez (94)

C.C: 1017143658

Julián José Martínez Camacho (93)

C.C: 1081907956

Ingeniería de Software y Datos

Institución Universitaria Digital de Antioquia

Semestre 7°

Bases de Datos II

Víctor Hugo Mercado

25 de mayo del 2024

ÍNDICE

EA3. PROCESO DE TRANSFORMACIÓN DE DATOS.....	1
INTRODUCCIÓN.....	3
OBJETIVOS.....	5
PLANTEAMIENTO DEL PROBLEMA Y ANÁLISIS.....	6
MODELO ESTRELLA	7
DESCRIPCIÓN DEL ANÁLISIS REALIZADO A LOS DATOS	8
ANEXOS	9
CONCLUSIONES.....	9
REFERENCIAS	10

Repositorio de todas las Actividades en Git:

<https://github.com/Alexis-Machado/REPOSITORIO-ACTIVIDADES-BD-II.git>

INTRODUCCIÓN

Elaborar un modelo en forma de estrella cuando estamos manejando datos es crucial para comprender y utilizar eficazmente la información que se encuentra en una base de datos. Este método, reconocido por su habilidad para facilitar la complejidad de los datos y proporcionar una perspectiva ordenada y nítida, toma forma visual de una estrella, simplificando así el proceso de análisis y toma de decisiones.

En esta actividad, combinaremos la construcción de un modelo estrella con la creación de una base de datos de staging basada en el modelo relacional de la base de datos Jardinería. El objetivo principal es crear consultas SQL que nos permitan trasladar los datos relevantes de la base de datos Jardinería a una nueva estructura de almacenamiento temporal, denominada base de datos de staging.

Antes de proceder con la creación de la base de datos de staging, es fundamental analizar los datos almacenados en la base de datos Jardinería para identificar qué información es relevante y debe ser transferida. Este proceso de revisión y selección de datos es crucial para garantizar que la base de datos de staging contenga únicamente los datos necesarios para el posterior procesamiento y análisis.

Una vez que identifiquemos los datos relevantes, procederemos a diseñar la estructura de tablas que conformarán la base de datos de staging. En esta parte, debemos asegurar una organización eficiente de los datos, facilitando su posterior transformación y análisis, para construir las consultas SQL necesarias para extraer los registros de la base de datos Jardinería.

y transferirlos a la nueva base de datos de staging, las consultas tienen que ser ejecutadas y validadas para garantizar que los datos se almacenen correctamente.

El análisis de datos requiere una estructura organizada para extraer información valiosa. Este proyecto se centra en transformar datos desde la base de datos staging a dimensiones. El objetivo es preparar los datos para análisis, permitiendo identificar el producto más vendido y realizar otros análisis empresariales relevantes.

Primero, se revisará el modelo Estrella de la base de datos Jardinería para entender la estructura y relaciones entre las tablas de dimensiones y la tabla de hechos. También se verificará la disponibilidad y consistencia de la base de datos de staging.

Luego, se extraerán datos relevantes desde la base de datos staging utilizando consultas SQL, asegurando su integridad y consistencia. En la base de datos de staging, se aplicarán técnicas de transformación de datos, como limpieza, normalización y enriquecimiento, utilizando consultas SQL o herramientas ETL para pasar los datos a las diferentes dimensiones correctamente.

La presente y última actividad se enfoca en la creación de una estructura de carpetas clara y eficiente en un repositorio de Git, lo cual facilitará la gestión de los archivos y scripts relacionados con el proceso de carga de datos. Esta estructura permitirá mantener un flujo de trabajo ordenado, separando los componentes clave del proyecto, como los datos en la base de datos de staging y el data mart final, el proceso de carga y la documentación relevante. Al establecer una organización adecuada desde el inicio, se sentarán las bases para un proceso de carga de datos efectivo y un análisis posterior preciso, lo que conducirá a una toma de decisiones.

OBJETIVOS

- Al hacer el Modelo Estrella de la base de datos Jardinería, buscamos extraer información importante que nos permita tomar decisiones efectivas. Por ejemplo, determinar el año con más ventas, identificar el producto más vendido y analizar la categoría con mayor cantidad de productos. Además, al tener este modelo estrella, buscamos simplificar la manipulación de la base de datos para agregar nuevas dimensiones o datos según lo requieran las actividades, lo que proporciona flexibilidad y adaptabilidad a las necesidades cambiantes de las actividades comerciales y de análisis.
- El objetivo principal de esta actividad es construir consultas SQL que permitan la creación de una base de datos staging a partir de la base de datos Jardinería. Esta base de datos staging nos facilitará el proceso de análisis de datos al proporcionar una estructura organizada y optimizada para consultas y procesos posteriores ya que al realizar un análisis de los datos almacenados en la base de datos Jardinería identificaremos patrones, tendencias y relaciones relevantes que puedan ser útiles para la toma de decisiones estratégicas en los diferentes ámbitos como el comercial.
- Desarrollar un proceso robusto y eficiente de transformación de datos desde la base de datos staging a las dimensiones. Este proceso tiene como finalidad preparar los datos de manera adecuada para su posterior análisis, permitiendo así la identificación del producto más vendido y la obtención de otras ideas relevantes que respalden la toma de decisiones empresariales estratégicas.

- Implementar una estructura de carpetas organizada y un proceso robusto para la carga de datos transformados en el Data Mart final, asegurando la integridad, calidad y disponibilidad de la información para análisis posteriores que respalden la toma de decisiones empresariales estratégicas.

PLANTEAMIENTO DEL PROBLEMA Y ANÁLISIS

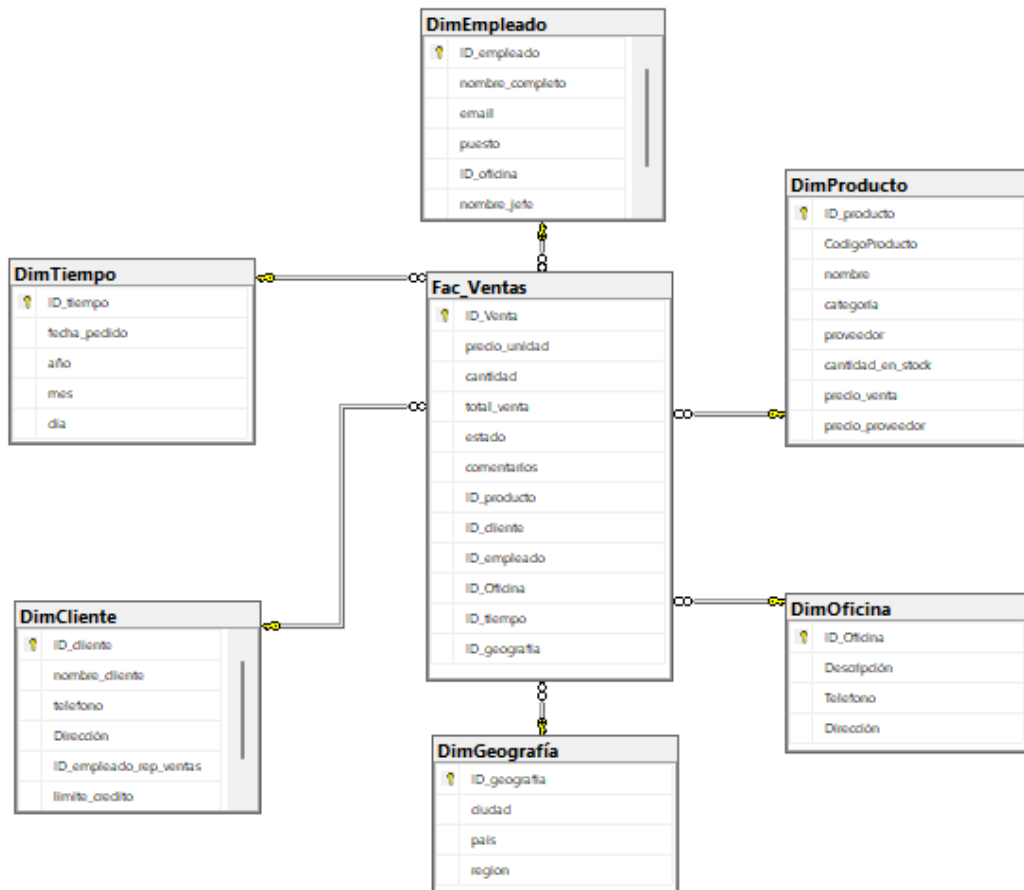
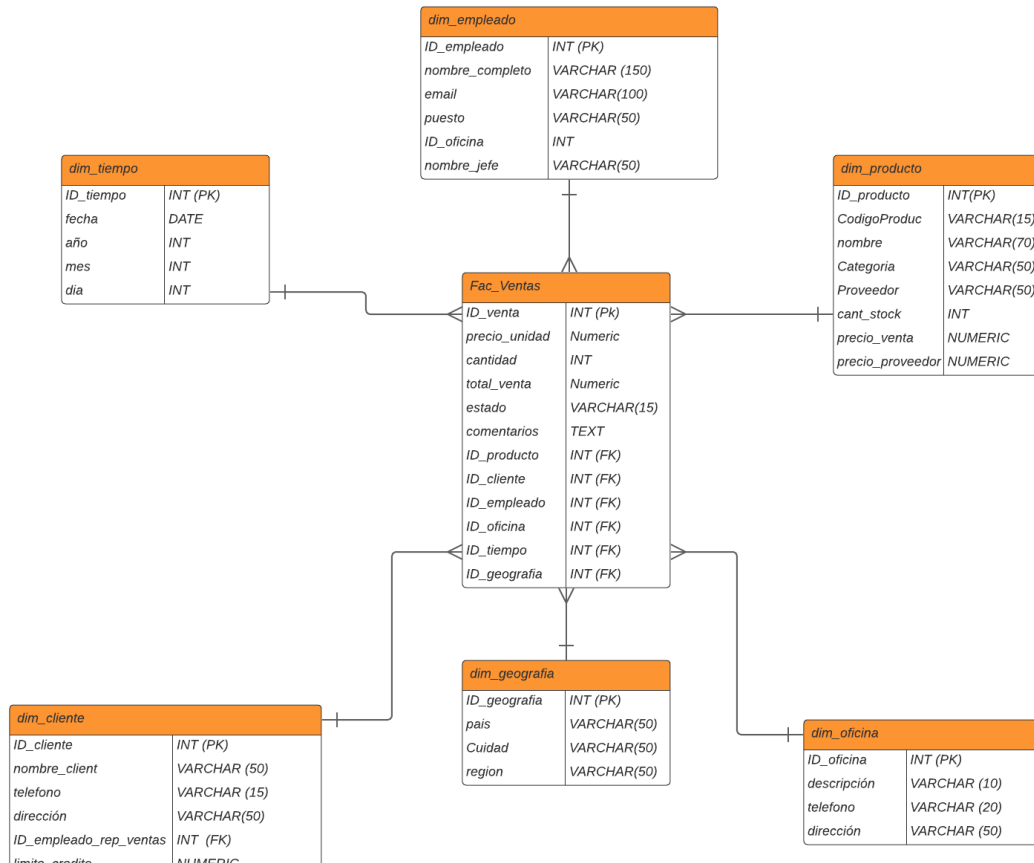
La empresa Jardinería enfrenta dificultades para realizar análisis detallados y tomar decisiones estratégicas debido a la estructura no optimizada de su base de datos. Esta situación impide identificar el producto más vendido, la categoría con mayor cantidad de productos y el año con mayor volumen de ventas, lo cual es crucial para entender patrones de ventas, preferencias de clientes y tendencias del mercado.

La estructura actual también dificulta análisis avanzados, como la segmentación de clientes, la venta cruzada y la evaluación del desempeño de productos o regiones. Para solucionar esto, se necesita implementar un proceso de transformación de datos, extrayendo información de la base de datos staging y organizándola en las dimensiones. En esta etapa, los datos se limpian, transforman y enriquecen según los requisitos analíticos.

El análisis en la base de datos de staging proporciona información valiosa para la toma de decisiones estratégicas en gestión de inventarios, desarrollo de productos, estrategias de precios y optimización de ventas y distribución.

Para solucionar esto, se requiere implementar un proceso robusto de transformación de datos desde la base de datos de staging al Data Mart final, organizando la información en dimensiones y aplicando técnicas de limpieza, transformación y enriquecimiento de datos según los requisitos analíticos.

MODELO ESTRELLA



DESCRIPCIÓN DEL ANÁLISIS REALIZADO A LOS DATOS

Para desarrollar un proceso efectivo de transformación de datos y carga en el Data Mart final, se realizó un análisis exhaustivo de la base de datos origen y la base de datos de staging. El objetivo principal fue comprender la estructura de los datos, identificar relaciones clave y determinar los datos relevantes para el análisis posterior.

En primer lugar, se analizaron todas las tablas de la base de datos de staging, incluyendo oficina, empleado, cliente, entre otras, para comprender la organización y las relaciones existentes entre los datos.

Se ejecutaron consultas SQL para explorar las relaciones entre las tablas, identificar patrones y detectar posibles inconsistencias en los datos.

Además, se realizó una validación exhaustiva de la integridad y calidad de los datos, revisando la presencia de valores nulos, duplicados y otros errores potenciales, con el fin de asegurar que los datos fueran confiables y de alta calidad para su posterior transformación y análisis.

Finalmente, se realizó un análisis previo, para facilitar el proceso de transformación de datos y la carga en el Data Mart final.

Captura de Pantalla como prueba:



ANEXOS

Los anexos estarán junto con este documento en el archivo comprimido y además todo estará en el repositorio de Git

CONCLUSIONES

Aunque mi compañero Julián y mi persona tuvimos algunas dificultades con demasiadas confusiones y errores que nos tocó hasta comenzar de cero en varias ocasiones gracias a ello logramos aprender mucho consultar y analizar de diferentes formas y aspectos para realizar esta actividad, por lo cual fue enriquecedora y la culminamos satisfactoriamente.

El desarrollo de un proceso efectivo de carga de datos desde la base de datos de staging al Data Mart final ha sido fundamental para habilitar el análisis detallado de la información empresarial en Jardinería. Este proceso ha superado los desafíos iniciales relacionados con la complejidad y la falta de organización de los datos originales, facilitando la extracción, transformación y preparación de los datos para su posterior análisis. Gracias a esto, se ha logrado obtener información valiosa, lo cual es crucial para la toma de decisiones estratégicas. Además, la implementación de este proceso sienta las bases para futuras expansiones y mejoras en el análisis de datos, asegurando que Jardinería esté mejor preparada para mantenerse competitiva en un entorno comercial impulsado por los datos.

REFERENCIAS

PowerData, R. (2024). *Staging: la salvaguarda de los procesos ETL*. Retrieved from <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/312643/Staging-la-salvaguarda-de-los-procesos-ETL>

Alfaro, F. R. (2021). *El modelo de estrella. El pilar fundamental del Business Intelligence*. Retrieved from <https://datamanagement.es/2019/06/27/business-intelligence-modelo-estrella/>

Cardona, G. (2023). *Qué es el Modelo Estrella: Aglaia Consulting*. Retrieved from <https://aglaia.es/blog/power-bi/que-es-el-modelo-estrella-en-power-bi/>

Zerpa, H., Garcia, R., & Izquierdo, H. (2020). Datamart basado en el modelo estrella para la implementación de indicadores clave de desempeño como salida del Big Data. *Universidad Ciencia y Tecnología*, 24(102), 47-54.

¿Qué es un data mart? (2023). Retrieved from <https://www.ibm.com/mx-es/topics/data-mart>

Universidad Autónoma del Estado de Hidalgo. (n.d.). Retrieved from <https://www.uaeh.edu.mx/scige/boletin/icea/n9/r1.html>