

EVIDENCIA DE APRENDIZAJE

Creación de una base de datos de Staging

Elaborado por:

Jeisson Estevens Araque Ramírez.

Docente:

Víctor Hugo Mercado

Institución Universitaria Digital de Antioquia

Ingeniería de Software y Datos

Bases de Datos II

2024

Introducción

En el contexto de la empresa de jardinería, la creación de una base de datos staging es un paso necesario para el éxito del proceso de business intelligence (BI). Esta base de datos actúa como un área de preparación intermedia, donde los datos se transforman, depuran y optimizan antes de ser cargados en el sistema de BI. A través de esta etapa, se garantiza que los datos sean coherentes, confiables y listos para su análisis.

El siguiente trabajo detalla la creación de una base de datos staging donde tomando el modelo estrella diseñado en la primera entrega, se diseñan las tablas y se llevan los campos que se consideran relevantes para la realización del proceso de business intelligence de la empresa de jardinería

Objetivos

- Analizar los datos en el origen para identificar los campos relevantes que deben trasladarse a la base de datos destino.
- Diseñar las tablas y crear la base de datos staging.
- Almacenar correctamente los datos en la base de datos destino.

Planteamiento del problema

En el contexto de la empresa de jardinería, nos enfrentamos a un desafío fundamental: optimizar el proceso de business intelligence (BI) mediante el proceso de extracción de datos del origen y la creación de una base de datos destino (staging) que actuará como un puente eficiente entre los datos operativos y el análisis de BI.

Análisis del problema

Para realizar un modelo de business intelligence efectivo y descriptivo es necesario identificar los campos relevantes de cada tabla para ser trasladados a la base de datos staging la cual fue creada para recibir los datos y organizarlos en diferentes tablas para luego crear las dimensiones y la tabla de hechos.

Identificación de Campos Relevantes:

El primer paso es determinar qué campos son verdaderamente relevantes para nuestro proceso de BI. No todos los datos recopilados son igualmente importantes.

Debemos examinar cada tabla y seleccionar los atributos que aportan valor estratégico.

Clientes:

Almacena información sobre los clientes, como su ID, nombre, teléfono, país y ciudad.

Empleado:

Contiene detalles sobre los empleados de cada oficina, como el ID del empleado, el nombre, el apellido, el cargo, la extensión y el email.

Fechas:

Contiene información sobre las fechas relevantes, como el ID de fecha, la fecha en sí misma, el día, el mes y el año.

Productos:

Almacena información sobre los productos disponibles para la venta, incluyendo su ID, la categoría a la que pertenecen, el nombre, el precio de venta, el proveedor y el código del producto.

Oficinas:

Representa las oficinas donde se realizan las ventas Incluye campos como ID de oficina nombre de la oficina, país, ciudad, teléfono, dirección y la descripción de la oficina.

Diseño de la Base de Datos Staging:

La base de datos staging es un área intermedia donde los datos se preparan antes de ser cargados en el sistema de BI.

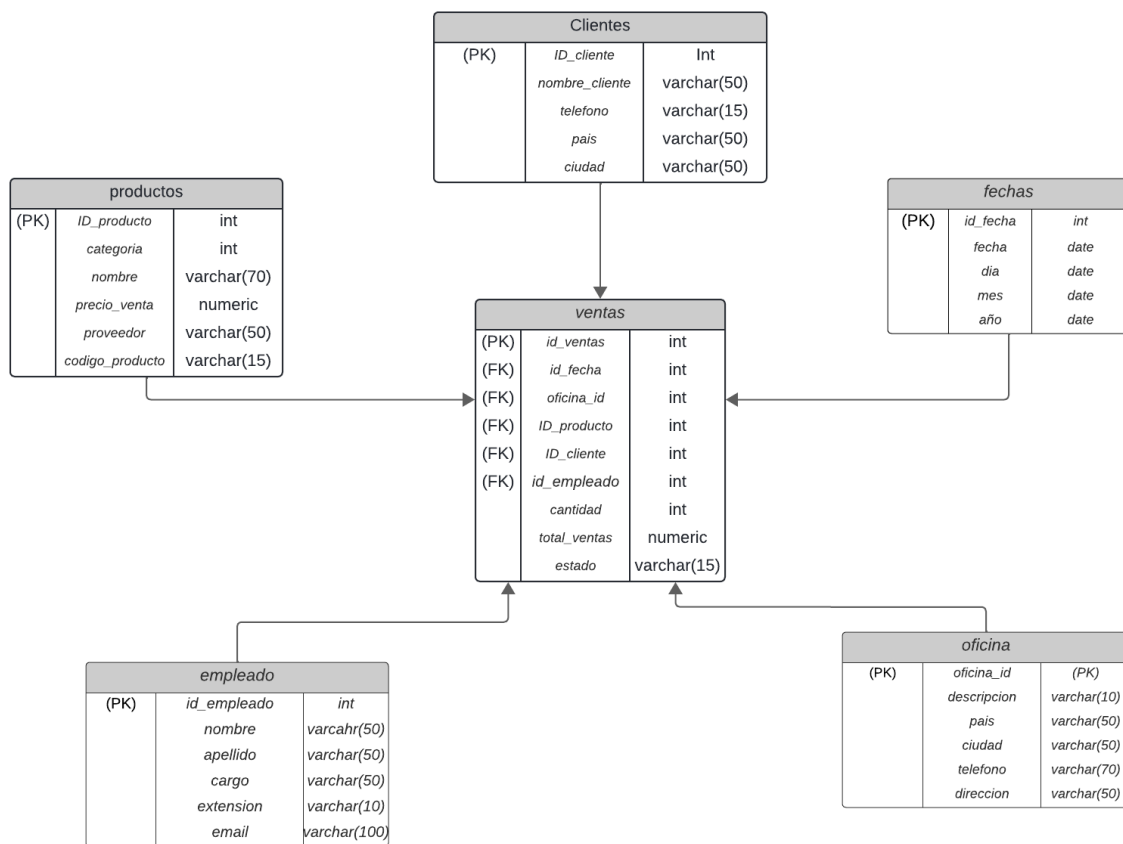
Su diseño debe ser eficiente y escalable. Debe permitir la transformación, limpieza y optimización de los datos.

El problema consiste en cómo estructurar esta base de datos para que cumpla su función de manera efectiva.

Para esta entrega no se incluye la tabla de hechos ya que será diseñada en proceso de transform.

Correcciones de la entrega 1

La corrección que se hizo fue eliminar la dimensión de formas de pago ya que no aportaba nada al modelo y se decidió incluir la dimensión de empleado ya que se puede relacionar directamente con la oficina y sacar datos de ventas de cada empleado. También se agregaron campos relevantes como estado en la tabla de hechos y el código del producto.



Descripción del análisis realizado a los datos Jardinería y cómo estos se trasladaron a la base de datos Staging.

Se identificaron los datos que generan valor al modelo de cada tabla en la base de datos origen(jardinería) y se realizan las consultas SQL correspondientes para luego ser llevados a la base de datos destino(stagin), la forma como se realizó este proceso fue usando la herramienta visual studio, donde se crea un flujo de datos con conexión a ambas bases de datos origen y destino luego se debe poner un origen en el cual se realiza una consulta a cada tabla para extraer los campos identificados como relevantes, luego en el destino se crea la tabla con los mismos campos del origen y se agregan las claves primarias, esto se realiza por cada una de las tablas en donde están los campos que queremos extraer luego se ejecuta el proceso y se verifica que los campos están correctamente en la base de datos destino.

Anexos

Backup de la base de datos staging.

Archivo del proceso de extracción.

Consultas SQL a la base de datos origen jardinería.

Bibliografía.

Mercado, Víctor (2024). Notas de clases Bases de datos II.IUDIIGITAL. Medellín.

https://iudigital.instructure.com/courses/15252/pages/unidad-3-diseno-de-procesos-etl?module_item_id=915662