**Evidencia de aprendizaje 2. Creación de una base de datos de Staging**

Andrés Camilo Arango Calle

**Docente:**

Victor Hugo Mercado Ramos

**Curso:**

Bases de Datos II

Programa Ingeniería de Software y Datos

Facultad Ingenierías y Ciencias Agropecuarias

Institución Universitaria Digital de Antioquia

2024

**Introducción**

Una base de datos de Staging desempeña un papel crucial en el proceso de desarrollo de software y gestión de datos. Actuando como un ambiente controlado, ofrece un espacio designado para realizar pruebas exhaustivas y validar cambios antes de que se desplieguen en la base de datos de producción. Al simular las condiciones de producción, pero en un entorno seguro, estas bases de datos permiten a los equipos de desarrollo detectar y corregir errores potenciales, mientras optimizan el rendimiento de las aplicaciones, facilitando así una transición sin problemas hacia la producción.

Además de su función en la garantía de calidad y estabilidad del software, las bases de datos de Staging desempeñan un papel vital en la gestión de riesgos. Al separar el proceso de prueba y desarrollo del entorno de producción, se reduce el impacto de posibles errores en los datos y procesos empresariales críticos. Esta separación también fomenta la innovación, ya que proporciona un espacio seguro para explorar nuevas funcionalidades y tecnologías sin comprometer la seguridad de los datos en producción.

**Objetivos**

1. Validar los datos antes de la carga en el modelo dimensional para garantizar su integridad y calidad.
2. Facilitar la transformación y ETL de los datos, permitiendo realizar las manipulaciones necesarias antes de cargarlos en el modelo dimensional.
3. Probar y validar el modelo dimensional en un entorno separado, asegurando su correcto funcionamiento antes de implementarlo en producción.

**Planeamiento y análisis del problema**

En el presente escenario, nos encontramos con el desafío de crear una base de datos de Staging a partir de una fuente de datos que consta de múltiples tablas. El objetivo principal es identificar y elegir los datos más pertinentes de cada tabla de origen para luego convertirlos en dimensiones que serán utilizadas para alimentar un modelo dimensional.

Al analizar este problema, nos enfrentamos a varios aspectos importantes:

* La base de datos original está compuesta por diversas tablas con diferentes tipos de datos y relaciones entre ellas. Esta complejidad agrega un nivel adicional de dificultad al proceso de selección y transformación de los datos, ya que implica la identificación de las tablas y campos más relevantes para el modelo dimensional.
* Es crucial determinar qué datos son más significativos para el modelo dimensional. Esto requiere una evaluación detallada de los requisitos del negocio y las necesidades de los usuarios finales para determinar qué información es esencial para la toma de decisiones y el análisis de datos.
* Una vez que se han identificado los datos relevantes, es necesario establecer un proceso eficiente para seleccionar y transformar estos datos en dimensiones. Esto implica la implementación de técnicas de limpieza, normalización y enriquecimiento de datos para garantizar su integridad y calidad en la base de datos de Staging.

**Solución**

1. **Correcciones entrega 1**

**Diagrama, Esquemático

Descripción generada automáticamente**

* **Análisis de dimensiones:**
  + **Dimensión Cliente:**
    - Id\_cliente (PK)
    - Nombre
    - Apellido
    - Teléfono
    - Fax
    - Dirección
    - Id\_geografia(FK)
    - Limite\_credito
  + **Dimensión Geografia:**
    - Id\_geografia(PK)
    - País
    - Ciudad
    - Región
    - Código\_postal
  + **Dimensión Oficina:**
    - Id\_oficina (PK)
    - Descripción
    - Id\_geografia(FK)
    - Telefono
    - Dirección
  + **Dimensión Empleado:**
    - id\_empleado (PK)
    - nombre
    - apellido
    - extensión
    - email
    - cargo
  + **Dimensión Producto:**
    - Id\_producto (PK)
    - Código
    - Nombre
    - Categoría
    - Altura
    - Ancho
    - Largo
    - Descripción
    - Cantidad\_en\_stock
    - Precio\_venta
    - Precio\_proveedor
    - Imagen\_Producto (Esta dimensión resulta de la combinación de las tablas Producto y Categoría Producto, y uno de las columnas que tenía la tabla Categoría Producto era “imagen”
  + **Dimensión Tiempo**
    - Id\_tiempo (PK)
    - Fecha
    - Año
    - Mes
    - Día
    - Día\_semana
    - Trimestre
    - Semana
* **Tabla de hechos (Pedidos):**
  + Id\_cliente (FK)
  + Id\_oficina (FK)
  + Id\_empleado (FK)
  + Id\_producto (FK)
  + Fecha\_pedido (FK)
  + Fecha\_entrega (FK)
  + Id\_pedido (PK)
  + Estado
  + Comentarios
  + Forma\_pago
  + Numero\_transacción
  + Total\_compra
  + Tiempo\_Entrega

**Se eliminaron los campos Edad (Dimensión cliente) y Costo\_envio (Tabla de Hechos), ya que efectivamente no se pueden obtener con los datos disponibles en la base de datos de origen.**

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

* **Tabla cliente:** Se descartaron dos columnas de esta tabla, ya que nos interesa la información únicamente de la persona que compra, por lo que se descarta el nombre y apellido del contacto, además también se prefirió no utilizar la columna de fax, puesto que en la actualidad no es de mucho uso.
* **Tabla Categoria\_producto:** Se descartaron las columnas descripción\_html e imagen, ya que no son utilizados.
* **Tabla Detalle\_pedido:** Se descarta la columna numero\_linea, ya que no es relevante como detalle del pedido para una tabla de hechos futura de Pedido.
* **Tabla Empleado:** Todas las columnas son relevantes, ya que tienen información personal y de contacto del empleado.
* **Tabla Oficina:** Se descarta únicamente la columna de línea\_direccion2, puesto que sólo es relevante la dirección principal, junto con los otros datos de ubicación y contacto de las oficinas.
* **Tabla Pago:** Se seleccionaron todas las columnas, debido a que es importante la información del pago para una futura tabla de hechos de Pedido.
* **Tabla Pedido:** Se seleccionaron todas las columnas, debido a que es importante toda la información del pedido para una futura tabla de hechos de Pedido.
* **Tabla Producto:** Se seleccionaron todas las columnas, debido a que es importante toda la información del producto para una futura dimensión de Producto en conjunto con la tabla Categoria\_producto
* **Tabla Tiempo:** Se selecciona únicamente la columna fecha\_pedido de la tabla Pedido, para la dimensión futura de tiempo, que permita reemplazar las fechas que estarían en las demás dimensiones

1. **Script de las consultas para crear la base de datos Staging**

--TableCliente

SELECT C.ID\_cliente,C.nombre\_cliente, C.telefono, C.linea\_direccion1, C.linea\_direccion2, C.ciudad, C.region, C.pais, C.codigo\_postal, C.ID\_empleado\_rep\_ventas, C.limite\_credito

FROM jardineria.dbo.cliente C

ORDER BY C.ID\_cliente ASC

--TableCategoria\_producto

SELECT CP.Id\_Categoria,CP.Desc\_Categoria,CP.descripcion\_texto

FROM jardineria.dbo.Categoria\_producto CP

--TableDetalle\_pedido

SELECT DP.ID\_detalle\_pedido,DP.ID\_pedido,DP.ID\_producto,DP.cantidad,DP.precio\_unidad

FROM jardineria.dbo.detalle\_pedido DP

--TableEmpleado

SELECT E.ID\_empleado,E.nombre,E.apellido1,E.apellido2,E.extension,E.email,E.ID\_oficina,E.ID\_jefe,E.puesto

FROM jardineria.dbo.empleado E

ORDER BY E.ID\_empleado ASC

--TableOficina

SELECT O.ID\_oficina,O.Descripcion, O.ciudad, O.pais,O.region,O.codigo\_postal,O.telefono,O.linea\_direccion1

FROM jardineria.dbo.oficina O

ORDER BY O.ID\_oficina ASC

--TablePago

SELECT P.ID\_pago,P.ID\_cliente,P.forma\_pago,P.id\_transaccion,P.fecha\_pago,P.total

FROM jardineria.dbo.pago P

--TablePedido

SELECT Pd.ID\_pedido,Pd.fecha\_pedido,Pd.fecha\_esperada,Pd.fecha\_entrega,Pd.estado,Pd.comentarios,Pd.ID\_cliente

FROM jardineria.dbo.pedido Pd

--TableProducto

SELECT Pr.ID\_producto,Pr.CodigoProducto,Pr.nombre,Pr.Categoria,Pr.dimensiones,Pr.proveedor,Pr.descripcion,Pr.cantidad\_en\_stock,Pr.precio\_venta,Pr.precio\_proveedor

FROM jardineria.dbo.producto Pr

--Table Tiempo

SELECT T.fecha\_pedido

FROM jardineria.dbo.pedido T

ORDER BY 1 ASC

**Anexos**

**Enlace Modelo estrella:** <https://lucid.app/lucidchart/3c255804-6a57-4eea-996e-ef63ef2ba8a2/edit?viewport_loc=-1321%2C-1471%2C4743%2C1913%2C0_0&invitationId=inv_89f14898-1944-41db-a0e9-67e5baee3515>