EVIDENCIA DE APRENDIZAJE 2. CREACIÓN DE UNA BASE DE DATOS DE STAGING

Por

Docente

INSTITUCIÓN UNIVERSITARIA DIGITAL DE ANTIOQUIA MEDELLÍN

2024

**Introducción**

**Objetivos**

**Planteamiento Del Problema**

En contadas ocasiones se podrán realizar modificaciones en las bases de datos para incorporar metadatos que contengan esa información. En otras situaciones, los datos se presentarán en formatos no estructurados. Una base de datos staging ofrece la posibilidad de añadir columnas a las tablas de datos, que pueden incluir desde una simple fecha hasta una clave externa vinculada a otra tabla con metadatos más detallados.

El objetivo es desarrollar una base de datos staging basada en el modelo relacional de la base de datos Jardinería. Este proyecto tiene como propósito mejorar el proceso de análisis de datos y facilitar la toma de decisiones estratégicas.

Es imprescindible identificar qué datos de la base de datos Jardinería son relevantes para ser trasladados a la base de datos staging.

El reto radica en diseñar una estructura de tablas en la base de datos staging que represente fielmente el modelo relacional de la base de datos Jardinería, garantizando la integridad de los datos y la eficiencia en su manejo.

Además, es necesario elaborar consultas SQL eficientes que extraigan los datos relevantes de la base de datos Jardinería y los carguen correctamente en las tablas de la base de datos staging, asegurando la consistencia y la coherencia de los datos transferidos.

**Análisis Del Problema**

Primero, es necesario llevar a cabo un análisis exhaustivo de los datos almacenados en la base de datos Jardinería para identificar la información relevante que se incorporará en la base de datos staging. Este paso incluye entender la estructura de la base de datos Jardinería y seleccionar los campos y tablas esenciales para el proceso.

A continuación, se deben diseñar consultas SQL que extraigan los registros necesarios de la base de datos Jardinería y los inserten en las tablas correspondientes de la base de datos staging. Estas consultas deben estar optimizadas para maximizar el rendimiento del proceso de carga y, si es necesario, incluir transformaciones o tareas de limpieza de datos.

Finalmente, se debe validar la información en la base de datos staging para garantizar que los datos hayan sido cargados y almacenados correctamente. Esto implica realizar consultas de verificación y comparación entre ambas bases de datos, detectando y corrigiendo cualquier discrepancia o error en los datos transferidos.

**Análisis de los datos almacenados en la base de datos Jardinería:**

1. Análisis de los datos de la base de datos Jardinería  
   Examinar los datos almacenados en la base de datos Jardinería para determinar cuáles son relevantes y deben ser transferidos a la base de datos Staging.
2. Construcción de la base de datos Staging  
   Crear la base de datos Staging, estableciendo su infraestructura inicial.
3. Diseño de la estructura de tablas en la base de datos Staging  
   Definir la organización de las tablas que conformarán la base de datos Staging, asegurando que reflejen de manera precisa los datos seleccionados de Jardinería.
4. Elaboración de consultas para la transferencia de datos  
   Diseñar y desarrollar consultas SQL que permitan extraer los registros de la base de datos Jardinería y cargarlos en las tablas de la base de datos Staging.
5. Ejecución y validación de las consultas  
   Ejecutar las consultas desarrolladas y verificar que los datos hayan sido transferidos y almacenados correctamente en la base de datos Staging.
6. Creación de respaldos (BK) de ambas bases de datos

Realizar respaldos de seguridad tanto de la base de datos Jardinería como de la base de datos Staging, garantizando la protección de la información.

**Propuesta de la solución con**

Modelo estrella - primera entrega



**Corrección**

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

**1:**

**Creación de la base de datos Staging.**



La base de datos staging actúa como un repositorio temporal diseñado para almacenar información relevante extraída de la base de datos de origen, con el propósito de prepararla antes de transferirla al datamart final.

**2:**

**Seleccionar los atributos a llevar de cada tabla a BDStaging.**

Basándose en el modelo en estrella, se realiza la selección de los atributos necesarios para cada una de las tablas que conformarán las dimensiones. Este proceso se lleva a cabo mediante consultas en SQL Server, asegurando que los datos seleccionados correspondan a los atributos clave y relevantes para la estructura dimensional del modelo.









:



Crear los flujos de datos agregando un dataflow la cual llamaremos: ExtraccionJardineria\_StagingJardineria.

* 
* Incorporar un origen de ADO.NET

Configurar un origen de ADO.NET que permita extraer los datos de cada tabla seleccionada en la base de datos Jardinería, asegurando la conexión adecuada y el formato correcto.

* Establecer un destino de ADO.NET

Configurar un destino de ADO.NET para cada tabla extraída, definiendo las propiedades necesarias para cargar los datos en la ubicación deseada, como la base de datos staging o el datamart.

* Estructura resultante

Cada tabla extraída de la base de datos Jardinería tendrá su origen y destino configurados mediante ADO.NET, asegurando una transferencia estructurada y eficiente de los datos.





En cada destino ADO.NET se especificará la sentencia SQL necesaria para crear la tabla correspondiente en la base de datos Staging. Esto asegura que las tablas se generen con la estructura adecuada antes de cargar los datos. Las asignaciones entre los orígenes y destinos se podrán visualizar gráficamente, mostrando la relación entre las tablas de la base de datos de origen (Jardinería) y sus respectivas contrapartes en la base de datos Staging.



Al ejecutar cada una de las secuencias SQL correspondientes a las tablas seleccionadas, la base de datos Staging se configura con las diez tablas previstas. Cada tabla contiene los atributos extraídos y asignados de acuerdo con el análisis previo, reflejando la estructura diseñada para optimizar el manejo de los datos.

-

* Anexos: debe agregar los BK de las dos bases de datos y el documento del script de las consultas para crear la base de datos Staging.

**Link de los backup y la ETL en visual studio comunity** [https://drive.google.com/drive/folders/1ByIs0aI17wulpfkRFIQbi39ijuhrtjOd?us](https://drive.google.com/drive/folders/1ByIs0aI17wulpfkRFIQbi39ijuhrtjOd?usp=sharing) [p=sharing](https://drive.google.com/drive/folders/1ByIs0aI17wulpfkRFIQbi39ijuhrtjOd?usp=sharing)

**Conclusiones**

La creación de una base de datos staging es un paso fundamental para garantizar la calidad, integridad y disponibilidad de los datos que apoyan la toma de decisiones estratégicas y el éxito operativo.

Este proceso incluye la selección cuidadosa de los datos relevantes de la base de datos Jardinería, seguida de su traslado a la base de datos staging. Durante esta etapa, se emplean consultas SQL diseñadas para mantener la consistencia y coherencia de la información. Asimismo, se aborda el desafío de diseñar una estructura de tablas precisa y realizar una validación detallada de los datos transferidos.

Finalmente, una base de datos staging bien implementada no solo mejora el análisis de datos, sino que también permite una adaptación flexible a futuros cambios en la estructura de la base de datos. Esto garantiza la utilidad y relevancia de los datos a lo largo del tiempo.

**Bibliografía**

Calzada, J. M. (2020, 16 noviembre). *Seis razones por la que usar staging*.

<https://www.certia.net/seis-razones-por-las-que-usar-staging/>