

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

Kevin Andrés Argumedo Cadavid

Astrid Carolina Hernández Torres

Institución Universitaria Digital de Antioquia

Facultad de ingeniería

Ingeniería de software y datos

Base de datos II

Antonio Jesús Valderrama Jaramillo

Urabá Antioquia

16/09/2025

Introducción.

La implementación de una base de datos staging surge de la necesidad de procesar y preparar la información más relevante proveniente de la base de datos jardinería. Este proceso permite transformar los datos en información mejor estructurada para ser analizada según sea requerida.

Una base de datos staging nos permite preparar los datos de una fuente primaria que no ha sido procesada y extraer únicamente la información necesaria para el trabajo específico.

En este trabajo se presenta el diseño y la implementación de esta solución, tomando la idea de negocio de un data mart creado anteriormente.

Objetivo general

Garantizar mediante una base de datos Staging que la información de la base de datos Jardinería que se utilizará para obtener la información que se va a limpiar y validar se mantenga con una calidad de datos minimizando errores antes de su uso.

Objetivos específicos.

- Diseñar la estructura de tablas que estarán en la base de datos Staging.
- Construir las consultas que permitan traer los registros de Jardinería a la Base de Datos Staging.
- Ejecutar las consultas y validar que los datos queden almacenados correctamente en la Base de datos Staging.
- Construir el BK de ambas bases de datos

Análisis del problema

El manejar grandes volúmenes de datos presenta diversos inconvenientes dependiendo del origen de estos. El principal desafío encontrado es la calidad de los datos que estamos manejando, factor crítico que impacta directamente en la efectividad de los análisis y la generación de informes. Entre más limpios estén nuestros datos, mejor y más detallados serán nuestros informes.

Propuesta de solución

Para el desarrollo del staging se utiliza la base de datos "jardinería", previamente empleada para la creación de un data mart. Considerando los requerimientos específicos del data mart existente, por ello se selecciona información de las tablas: **producto**, **detalle_pedido** Y **pedido**.

La información en bruto de estas tres tablas, se procesa mediante procesos ETL (Extract, Transform, Load), para generar en tres tablas staging optimizadas.

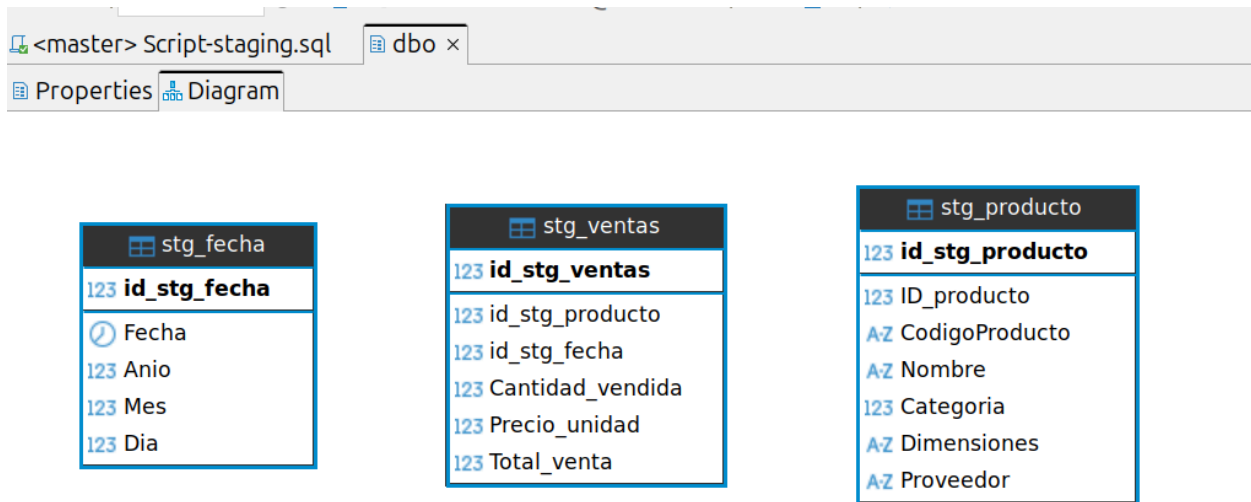
- **stg_fecha**
- **stg_ventas**
- **stg_producto**

Este enfoque selectivo garantiza que solo la información relevante sea usada para el data mart, optimizando el rendimiento, así como la precisión de los análisis posteriores.

Script Staging link:

https://drive.google.com/drive/folders/1xWDSUMrNr1Ca0XywF_829gTsHXkkazCH

Diagrama Staging



Comparación de datos

```

-- Productos originales vs staging
SELECT
    'Conteo' AS Descripción,
    (SELECT COUNT(*) FROM Jardinerial.dbo.producto) AS Original,
    (SELECT COUNT(*) FROM StagingJardineria.dbo.stg_producto) AS Staging;

-- Fechas originales vs staging
SELECT COUNT(DISTINCT CAST(fecha_pedido AS DATE)) AS OriginalFechas FROM Ja
  
```

Results 1 x

ELECT 'Conteo' AS Descripción, (SELECT COUNT(*) FF | Enter a SQL expression to filter results (use Ctrl+

A-Z Descripción	123 Original	123 Staging
Conteo	276	265

Consulta de información Cargada

-- Ver datos cargados

```
SELECT 'Productos' AS tabla, COUNT(*) AS registros FROM stg_producto
UNION ALL
SELECT 'Fechas', COUNT(*) FROM stg_fecha
UNION ALL
SELECT 'Ventas', COUNT(*) FROM stg_ventas;
```

-- Validar visualmente algunas filas

```
SELECT TOP 10 * FROM stg_producto;
SELECT TOP 10 * FROM stg_fecha;
SELECT TOP 10 * FROM stg_ventas;
```

Results 1 ×

SELECT 'Productos' AS tabla, COUNT(*) AS registros FF | Enter a SQL expression to filter results (use Ctrl+Space)

	AZ tabla	123 registros
1	Productos	265
2	Fechas	57
3	Ventas	226

-- Validar visualmente algunas filas

```
SELECT TOP 10 * FROM stg_producto;
SELECT TOP 10 * FROM stg_fecha;
SELECT TOP 10 * FROM stg_ventas;
```

-- Productos originales vs staging

```
SELECT COUNT(*) AS Original FROM jardineria.dbo.productos;
```

Results 1 ×

SELECT TOP 10 * FROM stg_producto | Enter a SQL expression to filter results (use Ctrl+Space)

	123 id stg producto	123 ID producto	AZ CodigoProducto	AZ Nombre	123 Categoria	AZ Dimensiones	AZ Proveedor
1	1	1	11679	Sierra de Poda 400MM	2	0,258	HiperGarden Tools
2	2	2	21636	Pala	2	0,156	HiperGarden Tools
3	3	3	22225	Rastrillo de Jardín	2	1,064	HiperGarden Tools
4	4	4	30310	Azadón	2	0,168	HiperGarden Tools
5	5	15	FR-1	Expositor Citricos Mix	4	100-120	Frutales Talavera S.A
6	6	16	FR-10	Limonero 2 años injerto	4		NaranjasValencianas.com
7	7	17	FR-100	Nectarina	4	8/10	Frutales Talavera S.A
8	8	18	FR-101	Nogal	4	8/10	Frutales Talavera S.A
9	9	19	FR-102	Olea-Olivos	4	8/10	Frutales Talavera S.A
10	10	20	FR-103	Olea-Olivos	4	10/12	Frutales Talavera S.A

Conclusión

Los objetivos planteados se concluyeron con éxito, se diseñó una estructura la cual permitió la creación de la base de datos staging, tomando como referencia el modelo de negocio de nuestro anterior trabajo Data Mart, de este mismo no fue requerido realizar modificaciones a su estructura, puesto que ya se tenían seleccionada las tablas fundamentales y necesarias para alcanzar los objetivos propuestos, de esta manera nuestro staging mantuvo una estructura simple y enfocada a suministrar los datos necesarios a nuestra Data Mart.

Bibliografía

Secoda. (2024, 16 de septiembre). What is a Data Staging Area? Secoda.

<https://www.secoda.co/glossary/data-staging-area>