

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

Felipe Fernández Rodríguez

Docente:

Antonio Jesús Valderrama Jaramillo

Curso:

Bases de Datos II

Ingeniería en Software y Datos

Institución Universitaria Digital de Antioquia

2025

INTRODUCCIÓN

El presente trabajo tiene como propósito explicar la construcción de una base de datos Staging a partir de la información de la base de datos Jardinería. Una base de datos Staging puede entenderse como un espacio intermedio donde los datos se organizan y preparan para su posterior análisis. En este caso, la intención no es reemplazar la base original, sino crear una estructura que facilite responder preguntas clave sobre el negocio de la jardinería, como por ejemplo: ¿cuál es el producto más vendido?, ¿qué categoría de productos cuenta con mayor variedad?, o ¿en qué año se registraron más ventas?

El trabajo se centra en mostrar cómo la información dispersa en diferentes tablas de la base de datos original puede transformarse en un repositorio simplificado y ordenado. De esta manera, se logra un punto de partida claro y confiable para llevar a cabo análisis que apoyen la toma de decisiones. El lector podrá comprender la importancia de contar con una base intermedia que actúe como “puente” entre los datos operativos del día a día y los procesos de análisis que permiten identificar tendencias y oportunidades dentro de la empresa.

OBJETIVOS

Objetivo General: Construir consultas SQL para la creación de una base de datos Staging, basada en el modelo relacional de la base de datos Jardinería, que permita su análisis respondiendo a tres categorías específicas: identificación el producto más vendido, la categoría con más productos y el año con más ventas.

Objetivos específicos:

- Diseñar la estructura de la base de datos Staging de acuerdo con la información proveniente de la base de datos Jardinería.
- Implementar consultas SQL poblando la base de datos Staging con los datos relevantes para el análisis.
- Analizar los datos almacenados en la base de datos Staging respondiendo a las categorías definidas del proyecto.

PLANTEAMIENTO DEL PROBLEMA

La base de datos Jardinería fue diseñada con un enfoque operacional, es decir, orientada a registrar y gestionar las transacciones del día a día de la empresa, tales como pedidos, pagos, clientes y productos. Sin embargo, este tipo de estructura, aunque eficiente para las operaciones cotidianas, no está optimizada para realizar análisis que apoyen la toma de decisiones estratégicas.

Cuando se busca responder a preguntas clave, como cuál es el producto más vendido, qué categoría de productos concentra la mayor cantidad de artículos o en qué año se registraron más ventas, la base de datos relacional presenta limitaciones. La información se encuentra dispersa en múltiples tablas, con relaciones que requieren procesos complejos para obtener resultados claros y útiles. Esto genera dificultades en la extracción y análisis de datos, así como un mayor tiempo de procesamiento y riesgo de errores en la interpretación.

Ante esta situación, surge la necesidad de implementar una base de datos Staging, que actúe como un repositorio intermedio donde la información se organiza y simplifica para fines analíticos. Este espacio permitirá consolidar los datos relevantes de la base Jardinería y transformarlos en un formato que facilite la generación de consultas más directas y eficientes. De esta manera, la empresa podrá disponer de información confiable y accesible para identificar patrones de ventas y apoyar la toma de decisiones basadas en evidencia.

ANÁLISIS DEL PROBLEMA

La base de datos Jardinería cumple adecuadamente su función operativa al registrar la información de clientes, empleados, productos, pedidos y pagos. No obstante, al estar construida bajo un modelo relacional y normalizado, presenta ciertas dificultades cuando se pretende utilizar para fines de análisis. En este tipo de bases, la información está distribuida en múltiples tablas, lo que implica que para obtener resultados concretos se deben realizar consultas complejas con varias uniones y cálculos, incrementando el tiempo de procesamiento y la posibilidad de inconsistencias.

Al plantear preguntas específicas, como identificar el producto más vendido, reconocer la categoría con mayor número de productos o determinar el año con más ventas, se evidencia que la base original no está diseñada para responder de forma directa a estas necesidades analíticas. La dispersión de los datos, la ausencia de un entorno enfocado en el análisis y el riesgo de sobrecargar la base operativa son factores que limitan su utilidad en la generación de conocimiento estratégico.

En este contexto, la creación de una base de datos Staging se presenta como una alternativa necesaria. Este espacio intermedio permite centralizar los datos relevantes, reducir la complejidad de las consultas y preparar la información para ser utilizada en procesos analíticos. Así, la empresa puede contar con una herramienta que facilite la obtención de indicadores clave de manera ágil, confiable y sin afectar el rendimiento de la base de datos transaccional.

PROPUESTA DE SOLUCIÓN

Para dar respuesta a las necesidades de análisis planteadas, se propone la creación de una base de datos Staging que actúe como un espacio intermedio entre la base de datos operativa Jardinería y el análisis de la información. Esta base no busca sustituir el sistema transaccional existente, sino complementarlo al organizar y simplificar los datos relevantes para la toma de decisiones.

La solución consiste en diseñar un esquema de Staging que concentre únicamente la información necesaria para responder a tres categorías específicas: identificar el producto más vendido, reconocer la categoría con mayor número de productos y determinar el año con más ventas. Para ello, se seleccionaron tablas y campos específicos de la base Jardinería que permiten realizar dichos análisis:

- Tabla Stg_Categoria:

Campos seleccionados:

- Id_Categoria: Identificador único de la categoría (PK).
- Desc_Categoria: Nombre o descripción de la categoría.

Estos campos permiten almacenar las diferentes categorías de productos disponibles. Permite clasificar los productos y sirve para responder a la pregunta: ¿Cuál es la categoría con más productos?.

- Tabla Stg_Producto:

Campos seleccionados:

- ID_producto: Identificador único del producto (PK).
- nombre: Nombre del producto.
- ID_categoria: Llave foránea que enlaza con Stg_Categoria (FK).
- precio_venta: Precio unitario de venta del producto.

Contiene la información detallada de los productos comercializados. Su relación con la tabla de ventas permite identificar el producto más vendido y asociarlo con su categoría.

- Tabla Stg_Tiempo:

Campos seleccionados:

- ID_tiempo: Identificador único de tiempo (PK).
- fecha: Fecha exacta de la venta.
- anio: Año en el que se realizó la venta.

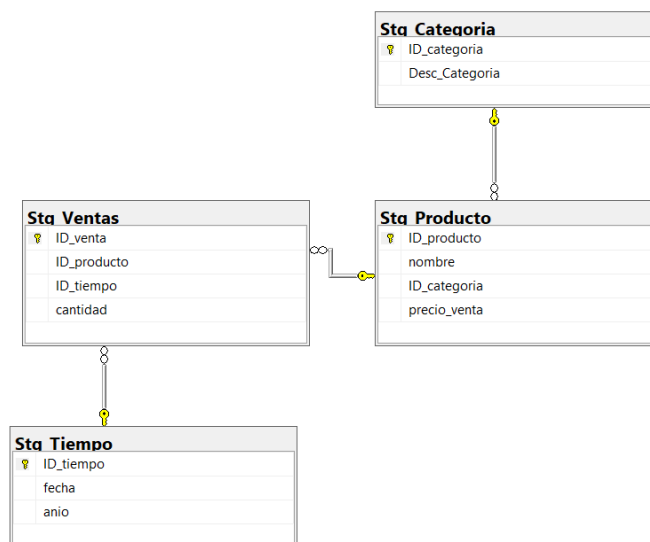
Estos campos representan la dimensión del tiempo, necesaria para analizar las ventas a lo largo de los años. Es clave para responder a la pregunta: ¿Cuál es el año con más ventas?.

- Tabla Stg_Ventas:

Campos seleccionados:

- ID_venta: Identificador único de la venta (PK, autoincremental).
- ID_producto: Llave foránea que enlaza con Stg_Producto (FK).
- ID_tiempo: Llave foránea que enlaza con Stg_Tiempo (FK).
- cantidad: Número de unidades vendidas del producto.

Registra las transacciones de ventas. Relaciona los productos con el tiempo y permite calcular volúmenes de ventas. Es la tabla central para medir y responder las tres categorías: producto más vendido, categoría con más productos (al cruzar con productos) y año con más ventas.

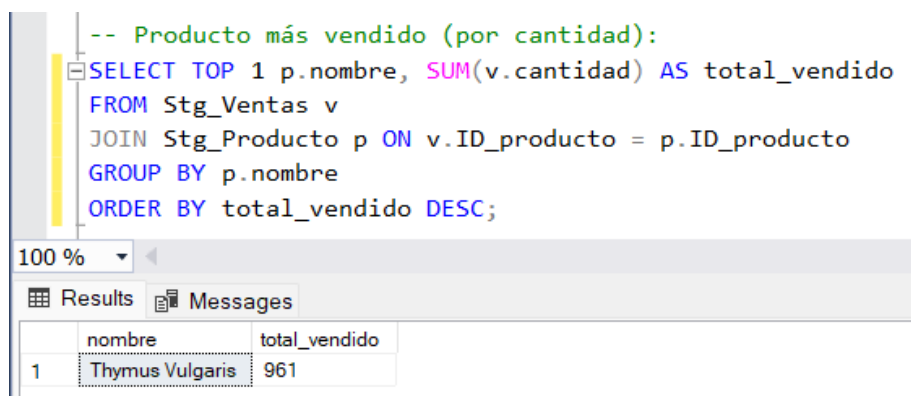


Con esta selección de tablas y campos, la base de datos Staging facilita la ejecución de consultas SQL directas y comprensibles, evitando la complejidad de la base transaccional original. De esta forma, la empresa puede obtener indicadores clave como el producto más vendido, la categoría con mayor variedad de artículos y el año con más ventas, apoyando así la toma de decisiones estratégicas basadas en información confiable y bien estructurada.

RESULTADOS DEL ANÁLISIS

A partir de la información organizada en la base de datos Staging, se obtuvieron los resultados correspondientes a las tres categorías definidas en el proyecto.

En primer lugar, al analizar el comportamiento de las ventas se identificó que el producto más vendido corresponde a Thymus Vulgaris, con un total de 961 unidades comercializadas. Este resultado refleja la alta demanda de esta especie dentro de la oferta de productos de la empresa.



The screenshot shows a SQL query editor with the following text:

```
-- Producto más vendido (por cantidad):  
SELECT TOP 1 p.nombre, SUM(v.cantidad) AS total_vendido  
FROM Stg_Ventas v  
JOIN Stg_Producto p ON v.ID_producto = p.ID_producto  
GROUP BY p.nombre  
ORDER BY total_vendido DESC;
```

Below the query, there is a 'Results' tab showing a single row of data:

	nombre	total_vendido
1	Thymus Vulgaris	961

En segundo lugar, al examinar la distribución de los artículos por categoría se determinó que la categoría con mayor número de productos es la de Ornamentales, con un total de 154 referencias registradas. Esto evidencia la amplia variedad de productos que ofrece la empresa en esta línea, lo que podría explicarse por la importancia de este tipo de plantas en el mercado de jardinería.

```
-- Categoría con más productos:
SELECT TOP 1 c.Desc_Categoria, COUNT(p.ID_producto) AS total_productos
FROM Stg_Producto p
JOIN Stg_Categoria c ON p.ID_categoria = c.ID_categoria
GROUP BY c.Desc_Categoria
ORDER BY total_productos DESC;
```

100 %

Results Messages

	Desc_Categoria	total_productos
1	Ornamentales	154

Por último, el análisis revela que 2009 fue el año más exitoso en términos de volumen de ventas para la empresa Jardinería, con un total de 5.136 unidades vendidas. Este hallazgo representa una oportunidad para profundizar en las causas que motivaron este resultado y tomar decisiones estratégicas para replicar o mejorar este desempeño en años futuros.

```
-- Año con más ventas:
SELECT TOP 1 t.anio, SUM(v.cantidad) AS total_vendido
FROM Stg_Ventas v
JOIN Stg_Tiempo t ON v.ID_tiempo = t.ID_tiempo
GROUP BY t.anio
ORDER BY total_vendido DESC;
```

100 %

Results Messages

	anio	total_vendido
1	2009	5136

En conjunto, estos resultados demuestran cómo la base de datos Staging permitió responder de manera clara y precisa a los objetivos planteados, facilitando la obtención de indicadores clave para el análisis y la toma de decisiones estratégicas.

INTERPRETACIÓN DE LOS RESULTADOS

Los resultados obtenidos a partir de la base de datos Staging ofrecen información valiosa para la empresa de jardinería, pues permiten transformar los registros transaccionales en conocimiento estratégico.

En cuanto al producto más vendido, *Thymus Vulgaris* destaca como el de mayor demanda, lo cual sugiere que la empresa debe prestar especial atención a su gestión de inventarios para garantizar disponibilidad constante. Asimismo, este hallazgo puede orientar campañas de marketing específicas, aprovechar su popularidad para atraer nuevos clientes o incluso diversificar productos relacionados que complementen su venta.

Respecto a la categoría con más productos, el predominio de los Ornamentales refleja una estrategia comercial centrada en la diversidad de esta línea. Sin embargo, la amplitud de referencias puede implicar retos en la gestión de inventarios y en la rotación de productos. Este resultado invita a evaluar cuáles de estos productos ornamentales presentan mayores niveles de venta y cuáles podrían estar generando costos sin representar beneficios significativos.

Finalmente, el hecho de que el año 2009 haya concentrado el mayor volumen de ventas abre la posibilidad de analizar los factores que incidieron en ese desempeño. Pudo estar relacionado con condiciones del mercado, campañas comerciales específicas o una mayor disponibilidad de productos clave. Comprender las razones detrás de este éxito puede ayudar a replicar estrategias efectivas en períodos futuros.

En síntesis, los resultados no solo responden a las categorías planteadas en el proyecto, sino que también ofrecen insumos prácticos para la toma de decisiones empresariales, tales como mejorar la

gestión de inventarios, orientar esfuerzos de marketing y diseñar planes estratégicos que permitan optimizar las ventas y la rentabilidad de la organización.

CONCLUSIONES

La creación de una base de datos Staging a partir de la información de la base Jardinería permitió evidenciar la importancia de contar con un espacio intermedio de organización y depuración de datos antes de realizar análisis estratégicos. A través de este proceso, se logró simplificar la estructura original, orientada a la operación diaria, en un modelo más ágil y enfocado en el análisis de información relevante.

En primer lugar, se demostró que la selección de tablas y campos específicos posibilitó enfocar los esfuerzos en responder preguntas concretas: identificar el producto más vendido, reconocer la categoría con mayor número de productos y determinar el año con más ventas. Esto permitió trabajar con datos más claros, reduciendo la complejidad de las consultas y facilitando su interpretación.

En segundo lugar, se comprobó que una base de datos Staging no solo optimiza la eficiencia de las consultas, sino que también evita la sobrecarga de la base transaccional, garantizando que el sistema operativo siga cumpliendo su función principal sin verse afectado por las tareas analíticas.

Finalmente, el trabajo realizado muestra cómo una adecuada estrategia de preparación de datos puede transformar la información en conocimiento útil para la empresa. La base de datos Staging se convierte así en un paso fundamental hacia la implementación de un modelo de análisis más robusto, como un Data Warehouse, que permita ampliar las posibilidades de exploración de datos y fortalecer la toma de decisiones en la organización.

BIBLIOGRAFÍA

- Action. (s. f.). *Data Staging: A critical step in ETL process*. Action. Recuperado el 14 de septiembre de 2025, de <https://www.action.com/data-staging/>
- Atlan. (2023, 10 de junio). *Data staging area: Definition, examples, benefits & more!* Atlan. Recuperado de <https://atlan.com/what-is/data-staging-area/>
- Baremon, E. (2021, 22 de mayo). *The role of staging tables in database administration*. Baremon.eu. Recuperado de <https://www.baremon.eu/staging-tables-in-database-administration/>
- ER/Studio. (s. f.). *Staging area in data warehousing*. ER/Studio Glossary. Recuperado el 14 de septiembre de 2025, de <https://erstudio.com/glossary/staging-area/>
- GeeksforGeeks. (2023, 4 de abril). *What is a data staging area in data warehouse?* GeeksforGeeks. Recuperado de <https://www.geeksforgeeks.org/software-testing/what-is-a-data-staging-area-in-data-warehouse/>
- Hevo Data. (2023, 7 de febrero). *Data staging area: Definition, types, and benefits*. Hevo Data. Recuperado de <https://hevodata.com/learn/data-staging-area/>
- Munchy Bytes. (2022, 15 de septiembre). *Optimizing data processes with staging areas in business intelligence*. Medium. Recuperado de <https://medium.com/munchy-bytes/optimizing-data-processes-with-staging-areas-in-business-intelligence-353bfb1e4bfe>
- ScienceDirect. (s. f.). *Data staging area – An overview*. ScienceDirect Topics. Recuperado el 14 de septiembre de 2025, de <https://www.sciencedirect.com/topics/computer-science/data-staging-area>
- Simitsis, A., Dayal, U., & Castellanos, M. (2009). *Data staging strategies and their impact on the execution of ETL workflows*. ACM Digital Library. <https://doi.org/10.1145/1552280.1592459>
- Wang, X., & Zhou, L. (2014). *Two-level data staging ETL for transaction data*. arXiv. <https://arxiv.org/abs/1409.1636>

Wikipedia. (2023, 15 de marzo). *Área de stage (datos)*. En *Wikipedia*. Recuperado de

https://es.wikipedia.org/wiki/%C3%81rea_de_stage_%28datos%29

Zuar. (2022, 16 de noviembre). *Data staging guide: Overview & best practices*. Zuar. Recuperado de

<https://www.zuar.com/blog/complete-guide-to-data-staging/>