



Creación de una base de datos de Staging

Presentado por:

Juan Esteban Casadiego & Evelyn Cerro Acuña

Ingeniería de software y datos, Institución Universitaria Digital de Antioquia

Bases de Datos II

Entregado a:

Antonio Jesús Valderrama

Barranquilla, Atlántico

2025

Tabla de contenido

Introducción.....	3
Objetivo general.....	4
Objetivos específicos.....	5
Planteamiento del problema.....	6
Análisis del problema.....	7
Propuesta de la solución.....	8
Lista de dimensiones propuestas.....	9
Descripción del análisis realizado a los datos Jardinería y cómo estos se trasladaron a la base de datos Staging.....	10
Conclusiones.....	11
Anexos.....	12
Bibliografía.....	13

Introducción

En el contexto actual, donde la calidad y la confiabilidad de la información determinan la efectividad de los procesos analíticos y de inteligencia de negocios, contar con entornos intermedios de preparación de datos resulta esencial. Entre estos, las bases de datos Staging se han consolidado como un componente crítico en los procesos de integración y migración de información, al permitir la depuración, validación y organización de los datos antes de su carga en sistemas de análisis. De acuerdo con Kimball y Caserta (2004), “el área de staging constituye el espacio de trabajo donde los datos son extraídos, transformados y preparados para su posterior explotación analítica, garantizando su consistencia y trazabilidad”.

El presente trabajo tiene como objetivo diseñar y construir una base de datos Staging a partir de la base transaccional Jardinería, que sirva como repositorio temporal para asegurar la integridad y confiabilidad de los datos antes de ser utilizados en procesos de análisis o migración hacia un data warehouse.

Se realizó un análisis exhaustivo de la estructura de la base de datos Jardinería, identificando las tablas relevantes, así como sus relaciones de dependencia. A partir de este estudio se definió la estructura de la base Staging, replicando las claves primarias y foráneas, y añadiendo columnas técnicas para el control de cargas. Posteriormente, se desarrollaron y ejecutaron consultas SQL que trasladaron la información de la base transaccional hacia la base Staging, validando la consistencia de los datos mediante comparaciones de conteo e integridad referencial.

Este trabajo se justifica en la necesidad de contar con un entorno controlado que actúe como filtro de calidad de datos, lo cual facilita posteriores procesos de análisis, evita errores en la información consolidada y mejora la gestión estratégica de la organización. Como señalan Coronel y Morris (2017), “los procesos de preparación de datos en entornos intermedios son indispensables para garantizar que la información final en un almacén de datos sea confiable, precisa y útil para la toma de decisiones”.

Objetivo general

Diseñar y construir una base de datos Staging a partir del modelo relacional de la base Jardinería, garantizando la consistencia, integridad y respaldo de la información.

Objetivos específicos

- A.** Analizar los datos almacenados en la base Jardinería para determinar su relevancia y consistencia.
- B.** Diseñar la estructura de tablas con sus respectivas claves primarias y foráneas en la base Staging.
- C.** Desarrollar consultas SQL que permitan trasladar los datos desde la base Jardinería hacia Staging.
- D.** Validar la correcta inserción de datos y la integridad referencial en Staging.
- E.** Ejecutar respaldos (BK) de ambas bases de datos para asegurar la preservación de la información.

Planteamiento del problema

En la actualidad, las organizaciones requieren infraestructuras que garanticen la confiabilidad de la información utilizada en procesos analíticos y estratégicos. De acuerdo con Kimball y Caserta (2004), “los entornos de staging son esenciales para preparar y transformar los datos antes de ser integrados en almacenes analíticos”, lo que permite asegurar la calidad de la información y reducir riesgos de inconsistencia.

En el caso de la empresa *Jardinería*, su base de datos transaccional almacena información de clientes, pedidos, productos, categorías, pagos y empleados, entre otros elementos. Sin embargo, esta base fue diseñada para registrar operaciones diarias, no para realizar procesos de depuración y validación que garanticen datos listos para el análisis estratégico.

Esta situación genera limitaciones al momento de consolidar la información, ya que es posible encontrar redundancias, registros incompletos o dependencias que dificultan la migración directa hacia un esquema analítico. Según Coronel y Morris (2017), “la implementación de entornos intermedios de preparación de datos permite mejorar la consistencia, integridad y confiabilidad de la información que será utilizada para la toma de decisiones”.

En consecuencia, la empresa enfrenta restricciones para disponer de un flujo de datos controlado que asegure la calidad de la información antes de ser utilizada en procesos de análisis o migración hacia un data warehouse.

Por lo tanto, surge la necesidad de diseñar e implementar una **base de datos Staging** a partir de la información contenida en *Jardinería*. Este entorno intermedio permitirá trasladar los datos de manera ordenada, validar su integridad, aplicar controles de consistencia y preparar la información para su explotación analítica, contribuyendo así a optimizar la gestión estratégica y la competitividad de la organización.

Análisis del problema

El análisis de la base de datos *Jardinería* permite identificar que, aunque su diseño transaccional es adecuado para registrar operaciones diarias, presenta dificultades cuando se requiere garantizar la calidad y consistencia de los datos para su posterior explotación analítica. Esto se debe a que la información se encuentra distribuida en múltiples tablas relacionadas (clientes, pedidos, productos, categorías, pagos, empleados, oficinas y detalle de pedidos), lo que puede generar registros huérfanos o inconsistencias en el traslado hacia entornos analíticos.

En primer lugar, la verificación de integridad entre las tablas mostró dependencias críticas, como la relación de clientes con representantes de ventas o la correspondencia de los pedidos con sus clientes. Si no se controla este tipo de relaciones al momento de migrar los datos, existe el riesgo de cargar información incompleta o duplicada en sistemas posteriores.

En segundo lugar, la carga de empleados presenta una complejidad particular debido a la relación jerárquica con el campo *ID_jefe*. Esta dependencia autorreferenciada puede provocar errores en la inserción de registros en entornos de staging si no se aplica una estrategia de control de claves foráneas.

Finalmente, la ausencia de un entorno intermedio genera dificultades para detectar a tiempo posibles inconsistencias en las tablas de pagos o en los detalles de pedidos, donde los datos deben coincidir estrictamente con los clientes, productos y órdenes existentes. Cargar estos registros de forma directa a un almacén de datos implicaría exponerlo a errores de integridad referencial que afectarían la confiabilidad de los análisis.

Estos factores evidencian que la base de datos transaccional no está preparada para actuar como fuente directa de análisis estratégico. De ahí la necesidad de construir una **base de datos Staging**, que funcione como área de preparación, depuración y validación de la información. Este entorno permitirá garantizar la consistencia de los registros, mantener las dependencias entre tablas y asegurar que los datos trasladados hacia modelos analíticos sean confiables y útiles para la toma de decisiones estratégicas.

Propuesta de la solución

La construcción de una base de datos *Staging* para la empresa Jardinería tiene como propósito central garantizar la calidad, consistencia e integridad de los datos antes de ser utilizados en procesos de análisis estratégico o migrados a un data warehouse. Este entorno intermedio actúa como una capa de preparación en la que la información transaccional es depurada, organizada y validada.

La estructura de la base *Staging* se compone de tablas que replican el modelo transaccional original, incorporando claves primarias y foráneas que aseguran la integridad referencial. Adicionalmente, se incluyeron campos técnicos como *ETL_LoadDate* y *ETL_Source*, que permiten auditar el momento y la procedencia de cada carga.

Este modelo simplifica y controla el proceso de migración de datos, ya que permite:

- Validar la relación entre clientes, pedidos, productos, empleados y pagos, evitando inconsistencias.
- Manejar dependencias jerárquicas complejas, como la relación de empleados con sus jefes directos, mediante reglas de carga controladas.
- Detectar y excluir registros huérfanos o duplicados antes de que lleguen a un entorno analítico.
- Generar respaldos consistentes de la información, asegurando la trazabilidad de cada carga.

De esta manera, la base de datos Staging no solo resuelve las limitaciones del modelo transaccional, sino que también prepara la información de manera óptima para futuros procesos de análisis multidimensional o para la implementación de un modelo estrella en un data mart. En consecuencia, la empresa Jardinería podrá disponer de datos confiables, organizados y listos para apoyar la toma de decisiones estratégicas.

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

El análisis de la base de datos Jardinería permitió identificar que su estructura estaba compuesta por tablas transaccionales orientadas al registro de operaciones diarias, entre las que se destacan: oficina, empleado, cliente, pedido, producto, categoría de producto, detalle_pedido y pago. Cada una de estas tablas cumple un rol dentro del ciclo de negocio de la empresa: desde la gestión de clientes y sus pedidos, hasta el control de productos, ventas y pagos realizados.

La revisión inicial tuvo como propósito determinar qué datos eran relevantes para el proceso de migración y, al mismo tiempo, verificar la existencia de posibles inconsistencias. Para ello, se ejecutaron consultas de validación con el fin de identificar registros huérfanos o relaciones incompletas, tales como clientes sin representante de ventas, pedidos sin cliente asociado, productos sin categoría o detalles de pedidos sin referencia a productos o pedidos válidos. Los resultados evidenciaron que la información cumplía con las reglas de integridad en el origen, lo que permitió avanzar con confianza en la construcción del entorno Staging.

Posteriormente, se diseñó la base `jardineria_stg`, replicando la estructura de las tablas originales y definiendo en cada caso sus claves primarias y foráneas, con el fin de mantener la integridad referencial. Se añadieron además columnas técnicas como `ETL_LoadDate` y `ETL_Source`, destinadas a registrar la fecha de carga y la procedencia de los datos, aportando trazabilidad al proceso.

El traslado de los datos desde Jardinería hacia Staging se realizó mediante consultas `INSERT ... SELECT` ejecutadas en un orden específico que respetó las dependencias entre tablas:

1. Oficina y categoría de producto, por ser entidades independientes.
2. Empleado, con control especial sobre la relación jerárquica del campo `ID_jefe`.
3. Cliente, validando que sus representantes de ventas existieran previamente en la tabla de empleados.
4. Pedido, condicionado a la existencia de clientes en staging.
5. Producto, vinculado a la categoría correspondiente.
6. Detalle de pedido, dependiente tanto de pedidos como de productos.

7. Pago, asociado directamente a clientes existentes.

Finalmente, se realizaron consultas de conteo comparativo entre origen y Staging, confirmando que los registros fueron trasladados correctamente. Así mismo, se verificó que todas las claves foráneas permanecieran habilitadas y confiables, lo que garantiza la integridad referencial de los datos.

En conclusión, el proceso de análisis y traslado aseguró que la información de Jardinería se almacenará en un entorno Staging estructurado, confiable y listo para servir de base en posteriores procesos analíticos y de toma de decisiones.

Conclusiones

- El diseño e implementación de la base de datos Staging para la empresa Jardinería constituye una solución eficaz para garantizar la calidad y consistencia de la información antes de ser utilizada en procesos analíticos o migrada hacia un data warehouse.
- Facilita el control de integridad referencial al replicar las claves primarias y foráneas del modelo transaccional, lo que asegura la coherencia entre clientes, pedidos, productos, empleados y pagos.
- Optimiza el proceso de migración de datos al establecer un entorno intermedio que permite detectar y corregir inconsistencias, evitando que registros huérfanos o duplicados lleguen a los sistemas analíticos.
- Aumenta la trazabilidad del proceso de carga gracias a la incorporación de campos técnicos como ETL_LoadDate y ETL_Source, que registran el origen y momento de inserción de los datos.
- Contribuye a la eficiencia en la gestión de la información, ya que simplifica la preparación de los datos y reduce la carga de trabajo sobre la base transaccional, liberándola para las operaciones diarias del negocio.
- En conclusión, la creación de la base de datos Staging no solo asegura la confiabilidad de la información que será utilizada en análisis estratégicos, sino que también fortalece la infraestructura tecnológica de la empresa Jardinería, garantizando una base sólida para la toma de decisiones fundamentadas en datos de calidad.

Bibliografia

- Coronel, C., & Morris, S. (2017). *Database systems: Design, implementation, & management* (12th ed.). Cengage
<https://www.cengage.com/c/database-systems-design-implementation-management-12e-coronel/9781337627900>
- Kimball, R., & Caserta, J. (2004). *The data warehouse ETL toolkit: Practical techniques for extracting, cleaning, conforming, and delivering data*. Wiley.
<https://www.wiley.com/en-us/The+Data+Warehouse+ETL+Toolkit-p-9780764567575>
- Microsoft. (2023). *CREATE DATABASE (Transact-SQL)*. Microsoft Learn.
<https://learn.microsoft.com/es-es/sql/t-sql/statements/create-database-transact-sql>
- Microsoft. (2023). *ALTER TABLE (Transact-SQL)*. Microsoft Learn.
<https://learn.microsoft.com/es-es/sql/t-sql/statements/alter-table-transact-sql>
- Microsoft. (2023). *INSERT (Transact-SQL)*. Microsoft Learn.
<https://learn.microsoft.com/es-es/sql/t-sql/statements/insert-transact-sql>
- Microsoft. (2023). *BACKUP DATABASE (Transact-SQL)*. Microsoft Learn.
<https://learn.microsoft.com/es-es/sql/t-sql/statements/backup-database-transact-sql>