

Creación de una base de datos Staging

Jefree Tamayo Herrera

Institución Universitaria Digital De Antioquia

PREICA2502B010064: Ingeniería De Software y Datos

Antonio Jesús Valderrama Jaramillo

16 de septiembre, 2025

ETL

Extracción, transformación y carga (ETL) es el proceso consistente de combinar datos de diferentes orígenes un gran repositorio central llamado almacenamiento de datos. ETL utiliza un conjunto de reglas comerciales para limpiar, y organizar datos en bruto y prepararlos para el almacenamiento, el análisis de datos y el machine learning (ML). Puede abordar necesidades de inteligencia empresarial, específicas mediante el análisis de datos, como la predicción del resultado de decisiones empresariales, la generación de informes y paneles.

Las organizaciones de hoy en día tienen datos estructurados y sistemas de administración de la relación con el cliente, pueden varias como datos de inventario, datos de marketing y redes sociales, datos de medición entre el tiempo como ventas de un año o en determinado tiempo, etc. Al aplicar el proceso de extracción, transformación y carga, los conjuntos de datos en bruto individuales se pueden preparar en un formato y una estructura que son mas consumibles para fines analíticos, lo que dan como resultado información mas significativa, por ejemplo, los minoristas en línea pueden analizar los datos de los puntos de venta para procanistar demanda y administrar el inventario.

ETL proporciona una vista consolidad de los datos para análisis e informes em profundidad. La administración de varios conjuntos de datos exige tiempo y coordinación, y puede provocar ineficacias y retrasos, pero el ETL combina las bases de datos y varias formas de datos en una sola vista unificada. El proceso de integración de los datos mejora la calidad de los datos y ahorra tiempo necesario para mover, categorizar o estandarizar datos, esto facilita el análisis, la visualización de grandes conjuntos de datos, el ETL brinda un contexto histórico profundo a los datos de las organizaciones. Una empresa puede combinar los datos heredados

con datos de nuevas plataformas y aplicaciones. Puede ver conjunto de datos más antiguos junto con información mas reciente, lo que le brinda una vista de largo plazo a los datos.

La utilización de un área de Staging permite asegurar la calidad y coherencia de los registros ya que posibilita incorporar claves de negocio (BK), atributos de control como fecha de carga, estado de registro, así como la identificación del origen de los datos. Posteriormente, esa información se transforma y se transfiere a un esquema OLAP (on-line Analytical Processing), diseñado con un enfoque dimensional mediante tablas de hechos y dimensiones, con el fin de optimizar el análisis histórico y apoyar la toma de decisiones estratégicas (Inmon, 2005)

Objetivos

Como objetivo tenemos la implementación de un proceso de integración de datos mediante la creación de una base datos Staging y su posterior carga en un esquema OLAP, con el fin de garantizar la calidad, validación, consistencia y disponibilidad de la información.

Objetivos específicos

- Revisar los datos almacenados en la base de datos jardinería para identificar cuales son relevantes y cuales se deben trasladar a la base de datos Staging
- Diseñar la estructura de tablas que estará en la base de datos Staging
- Construir las consultas que permitan traer los registros de jardinería a la base de datos Staging
- Construir el BK de ambas bases de datos

Surge una necesidad de establecer un proceso que permita centralizar y transformar los datos del base de datos jardinería de manera ordenada antes de llevarlo a un entorno analítico, pero la ausencia de una capa intermedia o puente genera riesgos de inconsistencias, duplicidad y perdida de trazabilidad de la información, lo que afecta directamente la confiabilidad de los resultados obtenidos.

Como sabemos las tablas Staging constituyen a una capa intermedia dentro del proceso ETL cuya finalidad es almacenar temporalmente los registros provenientes de la base de datos transaccional antes de ser integrados en un esquema analítico u OLAP, estas tablas replican la estructura de las tablas operativas para incorporan otros atributos adicionales que permiten mejorar la trazabilidad y el control de los datos. En la base de datos jardinería contábamos con diferentes tablas de dimensiones como cliente, empleado, pedido, pago, oficina, producto, etc, para el proceso de las tablas Staging no es necesario añadir todas las tablas base de la base de datos, sino solamente las necesarias para después hacer los respectivos análisis de datos, en este caso las tablas que seleccionamos para trasladar son las tablas cliente, empleado, pago, pedido y detalle de pedido.

```

1 • CREATE DATABASE stg_jardineria;
2 • USE stg_jardineria;
3
4 • CREATE TABLE stg_empleado (
5     id_stg INT AUTO_INCREMENT PRIMARY KEY,
6     bk_empleado INT,
7     nombre VARCHAR(50),
8     apellido1 VARCHAR(50),
9     apellido2 VARCHAR(50),
10    extension VARCHAR(10),
11    email VARCHAR(100),
12    ID_oficina INT,
13    ID_jefe INTEGER,
14    puesto VARCHAR(50),
15    fecha_carga DATETIME,
16    origen VARCHAR(100),
17    estado_registro VARCHAR(100)
18 );

```

Aquí podemos ver la creación de nuestra base de datos llamada stg_jardineria, el sufijo stg es para indicar que pertenece o estamos en el área de la base de datos Staging, este mismo sufijo se aplican en la diferentes tablas para diferenciar, como las tablas Staging son un espejo de o la idea es que los mismos registros de nuestra base de datos base se trasladen a las nuevas tablas de nuestra base de datos por ende llevaría los mismo atributos y al final agregamos tres atributos adicionales que nos ayudan a tener fecha de carga del registro, el origen o de donde proviene los registros por que se pueden trasladar desde diferentes fuentes. También creamos las Business Keys (BK) que son los identificadores únicos de cada empleado, cliente, oficina o empresa, en este caso como en la base de datos jardinería teníamos los identificadores únicos de cada tabla, así que lo que hice fue declarar o reutilizar esos mismos identificadores las BK de las nuevas tablas, de esta manera se asegura la trazabilidad de los registros entre la base de datos operativa y la capa intermedia.

```

83 • INSERT INTO stg_empleado (bk_empleado, nombre, apellido1, apellido2, extension, email, ID_oficina, ID_jefe, puesto, fecha_carga, origen,
84 SELECT
85 ID_empleado AS bk_empleado,
86 nombre,
87 apellido1,
88 apellido2,
89 extension,
90 email,
91 ID_oficina,
92 ID_jefe,
93 puesto,
94 NOW() AS fecha_carga,
95 "OLAP jardineria" AS origen,
96 "activo" AS estado_registro
97 FROM jardineria.empleado;

```

Lo siguiente es crear las consultas para trasladar los registros de las tablas de dimensiones de la base de datos base hacia nuestras tablas Staging, aquí hacemos la consulta para insertar o trasladar todos nuestros registros de la tabla empleado hacia la tabla stg_empleado y por último agregamos los nuevos atributos. La palabra reservada NOW () es para que al ejecutar o actualizar los registros, quede registrada a la fecha en tiempo real de cuando se realizó la consulta o la actualización.

```

159 • SELECT * FROM stg_empleado;
160 • SELECT * FROM stg_empleado;

```

	id_stg	bk_empleado	nombre	apellido1	apellido2	extension	email	ID_oficina	ID_jefe	puesto	fecha_carga	origen
▶	1	1	Marcos	Magaña	Perez	3897	marcos@jardineria.es	8	NULL	Director General	2025-09-15 23:54:39	OLAP jardineria
	2	2	Ruben	López	Martinez	2899	rlopez@jardineria.es	8	1	Subdirector Marketing	2025-09-15 23:54:39	OLAP jardineria
	3	3	Alberto	Soria	Carrasco	2837	asoria@jardineria.es	8	2	Subdirector Ventas	2025-09-15 23:54:39	OLAP jardineria
	4	4	Maria	Solís	Jerez	2847	msolis@jardineria.es	8	2	Secretaria	2025-09-15 23:54:39	OLAP jardineria
	5	5	Felipe	Rosas	Marquez	2844	frosas@jardineria.es	8	3	Representante Ventas	2025-09-15 23:54:39	OLAP jardineria
	6	6	Juan Carlos	Ortiz	Serrano	2845	cortiz@jardineria.es	8	3	Representante Ventas	2025-09-15 23:54:39	OLAP jardineria
	7	7	Carlos	Soria	Jimenez	2444	csoria@jardineria.es	4	3	Director Oficina	2025-09-15 23:54:39	OLAP jardineria
	8	8	Mariano	López	Murcia	2442	mlopez@jardineria.es	4	7	Representante Ventas	2025-09-15 23:54:39	OLAP jardineria

Por último consultamos o llamamos para ver la tabla de empleados y podemos ver que efectivamente los registros que teníamos en la tabla empleado de la base de datos operativa (jardinería) se trasladaron a nuestras nuevas tablas de manera organizada con su origen, fecha y estado.

Bibliografía

Amazon Web Services. (s.f.). *¿Qué es extracción, transformación y carga (ETL)?*

Recuperado el 16 de septiembre de 2025, de <https://aws.amazon.com/es/what-is/etl/>