

## Diseño de flujo de ingesta batch para consolidación de ventas regionales

### 1. Objetivo del flujo de ingesta

- Lograr consolidación diaria de las ventas provenientes de distintas sucursales del país.
- Responder a la necesidad del negocio de disponer de datos unificados, limpios y estandarizados para análisis y toma de decisiones estratégicas.

### 2. Estructura del flujo ETL

Extracción:

- Tipo de archivos: CSV provenientes de sucursales.
- Frecuencia: diaria (procesamiento nocturno).
- Origen: sistemas de ventas regionales.

Transformación:

- Limpieza de datos (eliminación de duplicados, tratamiento de valores nulos).
- Conversión de formatos (fechas, códigos de producto, divisas).
- Estandarización de campos para homogenizar la información.

Carga:

- Destino final: Data Warehouse centralizado (ejemplo: Amazon Redshift, Snowflake, o base de datos analítica en la nube).
- Opcional: exportación de copias de respaldo a almacenamiento (S3, HDFS).

### 3. Herramienta ETL seleccionada

Apache NiFi

- Justificación: permite ingesta desde múltiples fuentes (CSV, APIs, bases de datos), tiene interfaz visual intuitiva y soporta automatización de flujos batch recurrentes.
- Alternativa: Talend, por su potencia en integración de grandes volúmenes de datos y opciones avanzadas de calidad.

### 4. Buenas prácticas incluidas

- Validación de calidad de datos antes de la carga.
- Trazabilidad: registro de cada paso del flujo en logs.
- Programación fuera del horario de carga para optimizar recursos.
- Manejo de errores con reintentos automáticos y alertas en caso de fallos.

## 5. Representación del flujo (esquema técnico)

1. Extracción: NiFi recibe los archivos CSV diarios → se almacenan temporalmente.
2. Transformación: aplicación de limpieza → estandarización de formatos → consolidación de datos.
3. Carga: envío de datos procesados a Data Warehouse central.
4. Monitoreo: generación de reportes de ejecución y validación de calidad.