

Diagnóstico y propuesta de solución ante un problema de integración de datos

1. Análisis del caso

- Problemas de datos: bases de datos aisladas (silos), duplicidad de información, inconsistencias en registros.
- Riesgos: retrasos en reportes, decisiones erróneas por falta de una 'fuente única de la verdad', ineficiencia operativa.

2. Tipo de integración más adecuada

- Integración en batch + federación parcial.
 - * Batch para sincronizar grandes volúmenes de datos en horarios definidos (por ejemplo, fin de jornada).
 - * Federación para consultas rápidas entre sistemas sin necesidad de mover todos los datos.
- Justificación: el volumen es alto, pero no todas las áreas requieren tiempo real. Una estrategia híbrida balancea costo y agilidad.

3. Tareas del proceso de integración (ETL)

- Extracción: obtener datos de las distintas áreas (ventas, marketing, logística).
- Transformación: limpieza de duplicados, homologación de formatos (fechas, códigos postales, identificadores de cliente).
- Carga: centralizar en un Data Warehouse unificado o un lago de datos.
- Validaciones: reglas de integridad referencial, control de duplicados, chequeo de rangos válidos (ej. edad, fechas), monitoreo automático de calidad.

4. Herramientas recomendadas

- Talend o Pentaho: para flujos ETL tradicionales.
- Apache NiFi: para integraciones de datos en flujo continuo.
- HDFS o bases NoSQL (MongoDB, Cassandra): para almacenamiento escalable.

5. Propuesta de solución

Implementar un proceso de integración de datos basado en ETL con validaciones de calidad, centralizando la información en un repositorio común. Esto permitirá:

- Mejorar la agilidad analítica al contar con datos unificados.
- Tener una visión única del cliente entre todas las áreas.
- Reducir errores y duplicaciones que afectan la toma de decisiones.
- Aumentar la eficiencia operativa al eliminar silos de información.