

Ejercicio N° 1: Diagnóstico y propuesta de solución ante un problema de integración de datos

Diagnóstico

La organización presenta:

- Demoras en reportes por falta de integración.
- Duplicación de información entre sistemas aislados.
- Datos inconsistentes que generan decisiones poco confiables.
- Áreas críticas (ventas, marketing y logística) trabajan en silos de datos sin conexión.

Esto afecta la eficiencia, la confiabilidad de los reportes y la toma de decisiones estratégicas.

Propuesta de solución

Tipo de integración recomendada

- Integración mediante Data Warehouse (consolidación de datos) para unificar información de ventas, marketing y logística.
- Complementar con ETL/ELT para limpieza, normalización y carga de datos.
- Uso de herramientas de integración en tiempo real (ej. Kafka + Spark Streaming) si se requieren reportes instantáneos.

Tareas principales

1. Extracción: Captura de datos desde las bases de cada área.
2. Transformación:
 - Eliminación de duplicados.
 - Normalización de formatos (ej. fechas, monedas).
 - Validación de consistencia.
3. Carga: Consolidación en un Data Warehouse o Data Lake central.
4. Gobierno de datos: Definir políticas de calidad, trazabilidad y seguridad.
5. Monitoreo: Auditorías periódicas y alertas para detectar errores.

Herramientas sugeridas

- Apache NiFi / Talend / Airflow: Orquestación de ETL.
- Kafka + Spark Streaming: Integración en tiempo real.
- Data Warehouse en Snowflake, BigQuery o Redshift.
- Herramientas de BI (ej. Power BI, Tableau) para reportes confiables.

Beneficios esperados

- Unificación de datos entre áreas → fin de los silos de información.
- Reducción de duplicidad e inconsistencias → decisiones basadas en datos confiables.
- Reportes más rápidos y con una única “fuente de la verdad”.
- Visión 360° del negocio, alineando ventas, marketing y logística.
- Escalabilidad para integrar nuevas fuentes en el futuro.

Conclusión

La solución más adecuada es implementar un Data Warehouse centralizado con procesos ETL bien definidos, complementado con integración en streaming para casos de análisis en tiempo real. Esto permitirá mejorar la calidad, consistencia y disponibilidad de los datos, generando mayor agilidad y confiabilidad en la toma de decisiones.