

Tarea: Gestión y Transformación de Datos en Tiempo Real y su Integración con Procesos por Lotes

Objetivo:

Analizar y plantear las distintas necesidades, retos y soluciones al gestionar y transformar datos en tiempo real, así como explorar las opciones para unificar estos flujos con los procesos por lotes en una plataforma común.

Instrucciones:

1. Identificación de necesidades y retos:

- Define al menos **tres escenarios** donde se necesiten datos en tiempo real (por ejemplo: monitorización de dispositivos IoT, análisis de comportamiento de clientes, detección de fraudes).
- Enumera los principales retos técnicos y organizativos asociados a la gestión de datos en tiempo real (ejemplo: latencia, escalabilidad, consistencia de datos, integración de sistemas heredados).

2. Análisis de soluciones para datos en tiempo real:

- Investiga **tecnologías y herramientas** para procesar datos en tiempo real (por ejemplo: Apache Kafka, Flink, Spark Streaming).
- Explica cómo estas herramientas abordan los retos identificados.
- Evalúa los criterios de selección de estas herramientas (latencia, coste, integración con sistemas existentes, facilidad de uso).

3. Integración con procesos por lotes:

- Compara las características clave de los flujos de datos en tiempo real frente a los procesos por lotes (frecuencia de procesamiento, volúmenes de datos, casos de uso).
- Investiga **plataformas unificadas** que soporten tanto datos en tiempo real como por lotes (por ejemplo: Apache Beam, Databricks, Snowflake).
- Describe las ventajas y desventajas de adoptar una plataforma unificada.

4. Propuesta de solución:

- Diseña un esquema conceptual o diagrama que ilustre cómo los flujos de datos en tiempo real y los procesos por lotes podrían coexistir en una plataforma común.
- Explica cómo esta solución aborda:
 - La coherencia de datos entre flujos en tiempo real y lotes.
 - La escalabilidad de los sistemas.
 - La facilidad de mantenimiento.

5. Conclusión y evaluación:

- Escribe un breve informe (1-2 páginas) que resuma los hallazgos y proponga una solución adecuada para unificar flujos en tiempo real y procesos por lotes.
- Incluye recomendaciones sobre cuándo usar una plataforma unificada y cuándo mantener sistemas separados.

Entregables:

- **Informe escrito** que aborde las necesidades, soluciones y propuesta de integración.
- **Esquema o diagrama** que muestre la arquitectura de la solución.
- Lista de herramientas investigadas con sus pros y contras.

Criterios de evaluación:

- Profundidad y claridad del análisis de necesidades y soluciones. (2,5 puntos)
- Calidad del esquema conceptual o diagrama. (2,5 puntos)
- Relevancia y justificación de la propuesta de solución. (2,5 puntos)
- Uso de referencias técnicas y ejemplos concretos. (2,5 puntos)

Esta tarea está diseñada para fomentar el análisis crítico y la exploración práctica de herramientas y soluciones para gestionar datos en tiempo real y por lotes de manera eficiente.