

Tarea: Implementación de un Sistema de Procesamiento en Tiempo Real con Kafka y Dask

El objetivo de esta tarea es diseñar e implementar un sistema de procesamiento en tiempo real utilizando **Kafka y Dask**, integrando datos de una API pública. Los estudiantes deberán desarrollar un pipeline de ingesta, procesamiento y almacenamiento de datos para realizar análisis y reportes.

Descripción del Proyecto

Cada estudiante o grupo deberá seleccionar una **API de datos en tiempo real** y desarrollar una solución que:

1. Ingesta de Datos con Kafka

- Implementar un **Kafka Producer** que consuma datos de la API en tiempo real.
- Enviar los datos a un **Kafka Topic** adecuado.

2. Procesamiento en Paralelo con Dask

- Implementar un **Kafka Consumer** que reciba los datos del Topic.
- Utilizar **Dask para procesar los datos en paralelo**, aplicando transformaciones y cálculos relevantes.
- Realizar agregaciones, filtrados o cualquier procesamiento que tenga sentido para el tipo de datos seleccionado. Se debe generar INFORMACIÓN.

3. Almacenamiento y Visualización

- Guardar los datos procesados en un formato adecuado (**Parquet, CSV o base de datos**).
- Generar un reporte o visualización de los datos utilizando herramientas como **Pandas, Matplotlib o Dash**.

Sugerencias de APIs para utilizar

- **CoinGecko API** (precios de criptomonedas)
- **Open-Meteo API** (datos meteorológicos)
- **The Guardian API** (noticias en tiempo real)
- **Alpha Vantage API** (mercados financieros)