**Desarrollado por Jose Mauricio Calderon**

**Reto 4: Pipeline de datos end-to-end**

**Objetivo Principal**

Implementar un pipeline de datos end-to-end utilizando el ecosistema de Apache Kafka para procesar transacciones financieras en tiempo real y detectar posibles casos de fraude, este es un proceso, pero les aplica para cualquier proceso de modernización de Legados para procesos CDC.

**Tareas:**

* **Ingesta de Datos (50% del reto)**
  1. Configurar un Source Connector que lea un archivo CSV con transacciones financieras
  2. Publicar los datos en el tópico trx-fraud-transactions en formato JSON
  3. Configurar un Sink Connector para persistir los datos en PostgreSQL
* **Procesamiento de Streaming (50% del reto)**
  1. Crear streams en ksqlDB sobre el tópico de transacciones
  2. Implementar consultas continuas para filtrado, agregación y enriquecimiento
  3. Detectar transacciones sospechosas basadas en patrones de fraude

**Componentes del Flujo**

**1. Archivo CSV**: Contiene transacciones financieras con campos como account\_id, amount, timestamp, location, etc.

**2. Spooler CSV Source Connector**: Lee archivos CSV desde un directorio monitoreado

**3. Kafka Topic**: trx-fraud-transactions - almacena las transacciones en formato JSON

**4. ksqlDB**: Procesa las transacciones en tiempo real para detectar fraude

**5. PostgreSQL Sink Connector**: Almacena las transacciones procesadas en base de datos

**5. PostgreSQL**: Base de datos para almacenamiento persistente

Levantamiento de aplicaciones y servicios en Docker.

Interfaz de usuario gráfica, Aplicación, Correo electrónico

El contenido generado por IA puede ser incorrecto.

Creación del conector SpoolDir en KafkaConnect para tomar los archivos csv que se procesaran los cuales se identifican con el patron “transactions-\*.csv”

Texto

El contenido generado por IA puede ser incorrecto.

Creación del Conector Sink para registrar en postgres los datos procesados desde el archivo csv.

Captura de pantalla de un celular

El contenido generado por IA puede ser incorrecto.

Verificación de los Conectores creados:

Captura de pantalla de un celular

El contenido generado por IA puede ser incorrecto.

Captura de pantalla de un celular

El contenido generado por IA puede ser incorrecto.

Captura de pantalla de un celular

El contenido generado por IA puede ser incorrecto.

Verificación del Tópico creado en Kafka

Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.

Verificación del Consumidor creado en Kafka

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

El contenido generado por IA puede ser incorrecto.

Archivo CSV con 10000 transacciones fraudulentas



Se mueve el archivo a la ruta mapeada

Interfaz de usuario gráfica, Texto

El contenido generado por IA puede ser incorrecto.

El pipeline toma el archivo csv y lo procesa.

Interfaz de usuario gráfica, Texto

El contenido generado por IA puede ser incorrecto.

Los mensajes se ven en el tópico de Kafka

Interfaz de usuario gráfica, Tabla

El contenido generado por IA puede ser incorrecto.

Los mensajes en el tópico

Interfaz de usuario gráfica, Aplicación, Tabla, Excel

El contenido generado por IA puede ser incorrecto.

Se valida que el consumidor este arriba

Interfaz de usuario gráfica, Texto, Sitio web

El contenido generado por IA puede ser incorrecto.

Se validan los registros se persistan en la tabla tbl\_fraud\_transactions de postgres

Interfaz de usuario gráfica

El contenido generado por IA puede ser incorrecto.

**Pruebas de consultas desde Kafka SQL**

Se crea el stream

Texto

El contenido generado por IA puede ser incorrecto.

Se puede hacer consultas.

Texto

El contenido generado por IA puede ser incorrecto.

Se agrega EMIT CHANGES

Texto

El contenido generado por IA puede ser incorrecto.

Se cargan mas transacciones algunas cumplen y otras no la condición indicada.

Interfaz de usuario gráfica, Texto

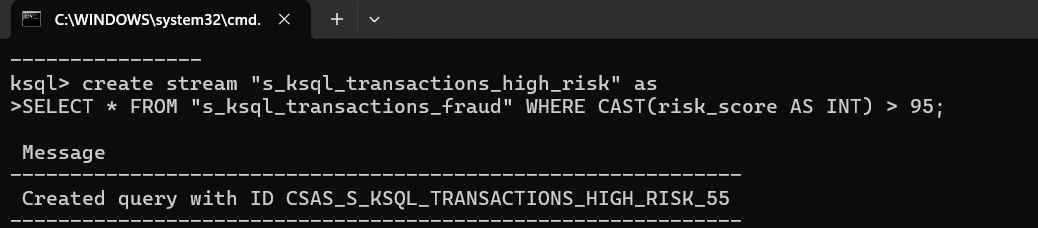
El contenido generado por IA puede ser incorrecto.

Se visualizan en la consola que llegan las transacciones

Texto

El contenido generado por IA puede ser incorrecto.

Se crear consulta para filtrar las transacciones con nivel de fraude mayor o igual a 95



Se crear consulta para agrupar transacciones de alto riesgo de fraude por tipo de fraude detectado.

Texto

El contenido generado por IA puede ser incorrecto.

Archivo de ksql utilizados:

