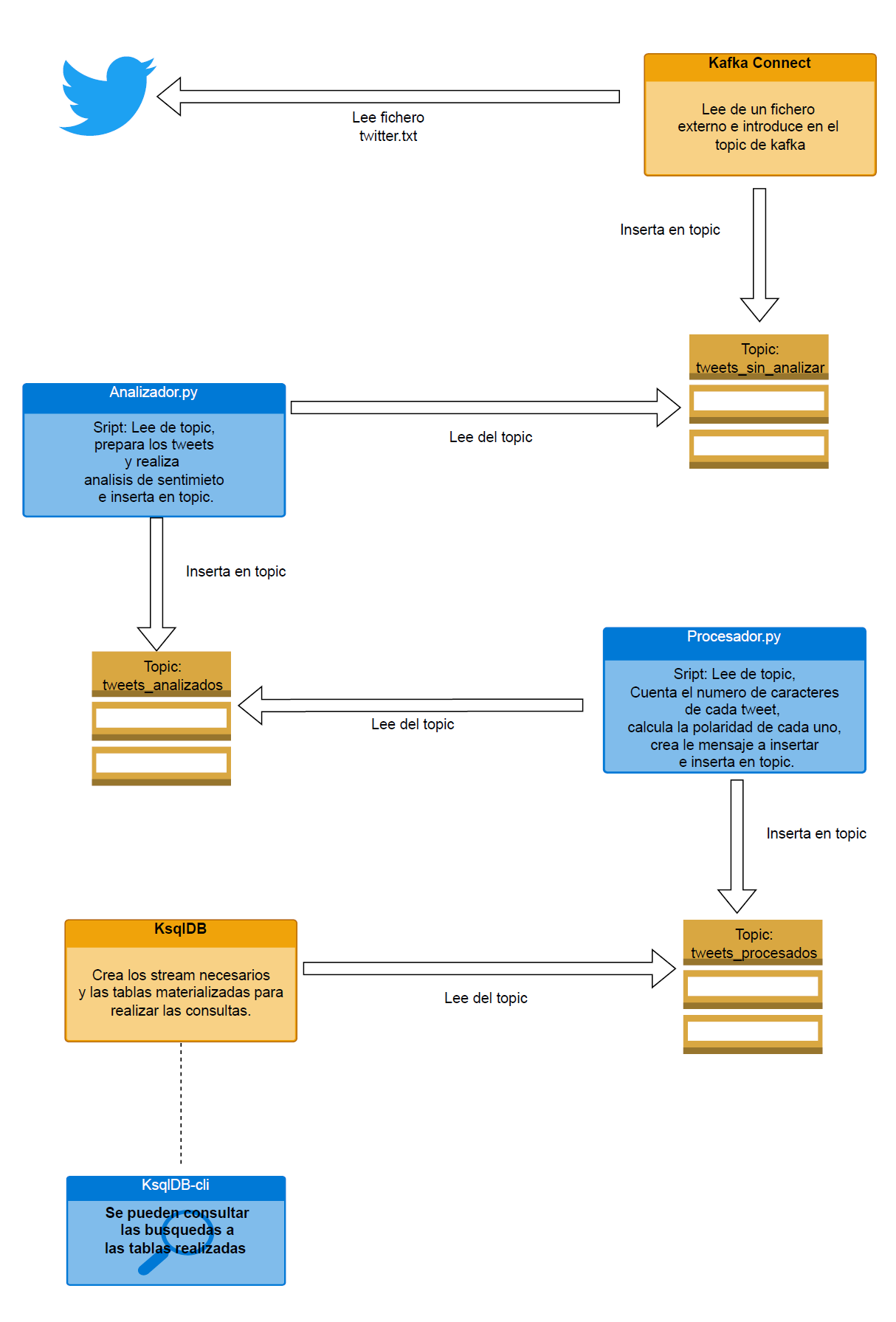
# Justificación de uso sistema “practicakafka”.

Diagrama de contendores



Justificación de uso

Para la practica he decidido usar las siguientes tecnologías:

* KafkaConnect

Se ha usado Kafka Connect para insertar tweets desde un TXT en cluster Kafka para luego realizar un análisis de sentimientos. Permite automatizar el flujo de datos desde el archivo hasta Kafka, facilitando un procesamiento continuo y en tiempo real a medida que se insertan los datos. Kafka Connect es una herramienta que sirve para integrar Kafka con diversas fuentes de datos de manera eficiente, desde ficheros en local (.txt, .csv, .json) como remotos. En este caso he usado un dataset descargado de kaggle con aproximadamente 190.000 tweets, por tener dificultades para lo conexión a twitter al ser de pago.

* Kafka producer- consumer

Se ha utilizado Kafka para comunicar los diferentes puntos de procesado del sistema, gracias al patrón de productor-consumidor con Kafka permite una comunicación eficiente y escalable entre los diferentes componentes de un sistema distribuido, facilitando la ingesta, procesamiento y distribución de datos en tiempo real.

* Kafka KsqlDB

Se ha utilizado Kafka KsqlDB por ser una herramienta eficaz para el procesamiento de datos en tiempo real y su capacidad de análisis en el ecosistema Kafka. Puede detectar patrones, tendencias o anomalías en los datos. Utiliza lenguaje SQL estándar para definir consultas sobre los flujos de datos, lo que le hace un entorno a KsqlDB familiar al desarrollador a la hora de realizar dichas consultas.

* Python

Desde mis scripts de Python ha sido relativamente cómodo configurar mis Productores y Consumidores, hasta crear los topics, se integra fácilmente con el ecosistema Kafka. Pudiendo crear un flujo de datos que deben de tener diferentes procesados. Todas las librerías de Python ayudan a tratar los datos de una manera estandarizada.