Agente Fulcrum,

El siguiente mensaje es altamente confidencial, léalo con detenimiento y asegúrese de que no caiga en las manos del Imperio.

Como bien sabe, existen tres sagas literarias que están gozando de un gran éxito en la capital imperial de Coruscant. Como técnica de propaganda, desde la Alianza Rebelde tenemos pensado diseñar y lanzar una campaña publicitaria basada en personajes de las sagas que inspiren a unirse a nuestra causa. Para ayudarnos con la tarea, nuestros espías han conseguido crear un sistema que analiza en tiempo real cuánta gente lee un texto relacionado con cualquiera de las tres.

Además, le proporcionaremos la siguiente información:

- Un listado con los personajes más importantes de dichas sagas.
- Un modelo que, dado un artículo, es capaz de decir sobre cuál de las tres sagas literarias habla.

Objetivo de la prueba

El objetivo de la prueba es diseñar un sistema que permita a nuestros analistas saber cuáles son los personajes que más interés despiertan en la gente en un periodo concreto. Para ello, deberá diseñar e implementar un sistema (junto con la arquitectura de ingesta, transformación y almacenamiento necesaria) que cumpla con las especificaciones que se detallan a continuación:

- El output del sistema debe ser el top N entidades sobre las que más se ha leído en una saga concreta, así como el número de personas que leyó sobre cada una de ellas.
- El sistema debe admitir como filtros opcionales un timestamp inicial y también final. El formato de estos timestamps queda a su elección.
- Los componentes del sistema quedan también a su elección, incluyendo la manera en la que se exponen los datos (API HTTP, base de datos, interfaz de consulta, etc.)
- Documente mínimamente cómo acceder a los datos. En caso de exponer los datos directamente a través de una base de datos, incluya al menos un par de consultas de ejemplo.

En resumen, el sistema debe admitir un número N de entidades, la etiqueta de una de las 3 sagas, y (opcionalmente) un timestamp inicial y otro final. El output esperado será cualquier estructura de datos que contenta las N entidades sobre las que más se leyó en ese periodo, junto con el volúmen total de lectores.

Detalle de los recursos

Como punto de partida para la prueba, dispone de los siguientes recursos:

`entities.txt`: listado de entidades/personajes más importantes de las tres sagas.

`docker-compose.yml:` docker-compose que levanta los siguientes servicios:

- **`classification-service`**: servicio de clasificación de artículos. Dispone de una ruta `/predict` en su puerto 5000 que espera una petición POST con la siguiente cuerpo: `{“text”: “texto a clasificar”}.` Las etiquetas emitas son: `"got", "lotr", "hp"`
- **`data-streaming-service`**: genera un stream de datos, que se publica en el topic `events` del contenedor de kafka. Cada uno de los eventos mide cuánta gente leyó un artículo concreto en el momento en el que se envía.
- **`zookeeper`**
- **`kafka`**

Adicionalmente, se levantará un contenedor llamado `listener-service` con el único fin de permitirle la exploración del stream de datos. Para poder acceder a él, simplemente debe entrar al contenedor con el siguiente comando:

```
docker-compose exec listener-service bash
```

y arrancar el siguiente proceso:

```
python3 kafka_consumer.py
```

Este proceso mostrará el stream de datos en tiempo real.

Instrucciones

El objetivo final de la prueba es ver cómo te desenvuelves al trabajar en este tipo de problemas. No te preocupes si el resultado no es 100% funcional. Intenta centrarte en hacer las cosas cómo las harías en un entorno real, documentando brevemente las decisiones de diseño, la arquitectura del sistema y next steps.

A la hora de entregar la prueba, recuerda incluir la documentación necesaria, instrucciones para ejecutar el resultado, y envíalo junto con el código y ficheros de tu prueba todo dentro de un único fichero comprimido.