# Practica de Kafka

## Requerimientos

* Docker
* Postman

## Configuración de Kafka

Dentro del directorio de Docker lanzamos lo siguientes comandos:

# Inicializamos los contenedores

docker-compose up

# Entramos al broker de kafka

docker exec -it broker bash

# Creamos los topics

kafka-topics --bootstrap-server localhost:29092 --topic tweets --create --partitions 3 --replication-factor 1

kafka-topics --bootstrap-server localhost:29092 --topic tweets-hashtags-lang --create --partitions 3 --replication-factor 1

kafka-topics --bootstrap-server localhost:29092 --topic tweets-trending-topics --create --partitions 3 --replication-factor 1

Lista de contenedores y puertos:

Kafka:          http://localhost:29092

Kafka Connect:  http://localhost:8083

ElasticSearch:  http://localhost:9200

Kibana:         http://localhost:5601

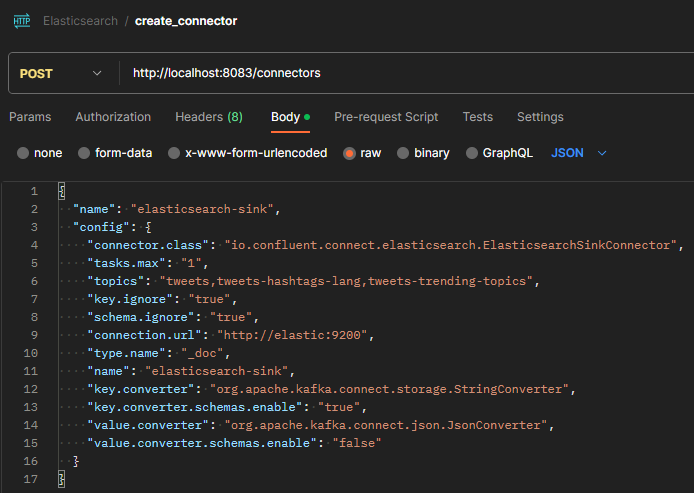
ksqldb-server:  http://localhost:8088

zookeeper:      http://localhost:2181

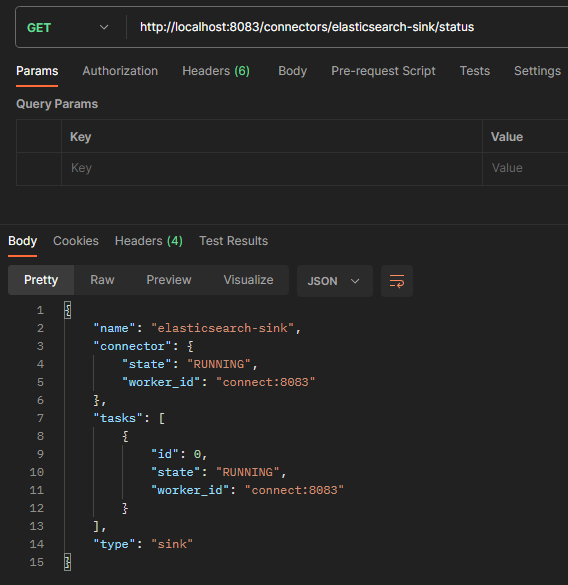
## Configuración de Elasticsearch

Elasticsearch es la base de datos NoSQL elegida para guardar nuestros topics. Para habilitar este flujo entre Elasticsearch y Kafka es necesario configurar un conector con Kafka Connect y los índices:

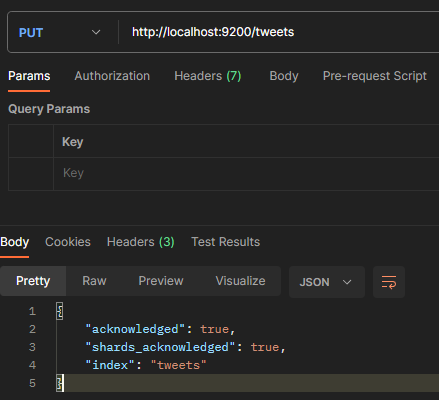
* Creamos el conector

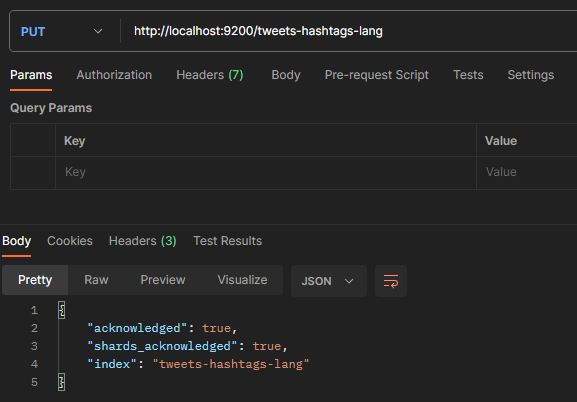


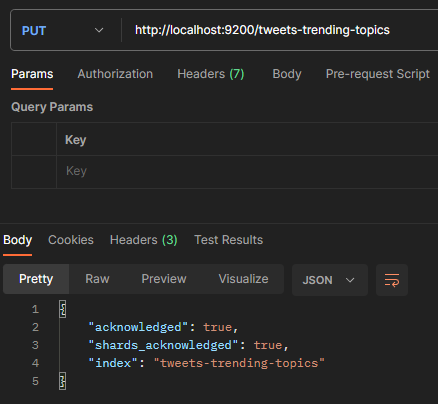
* Estado del conector



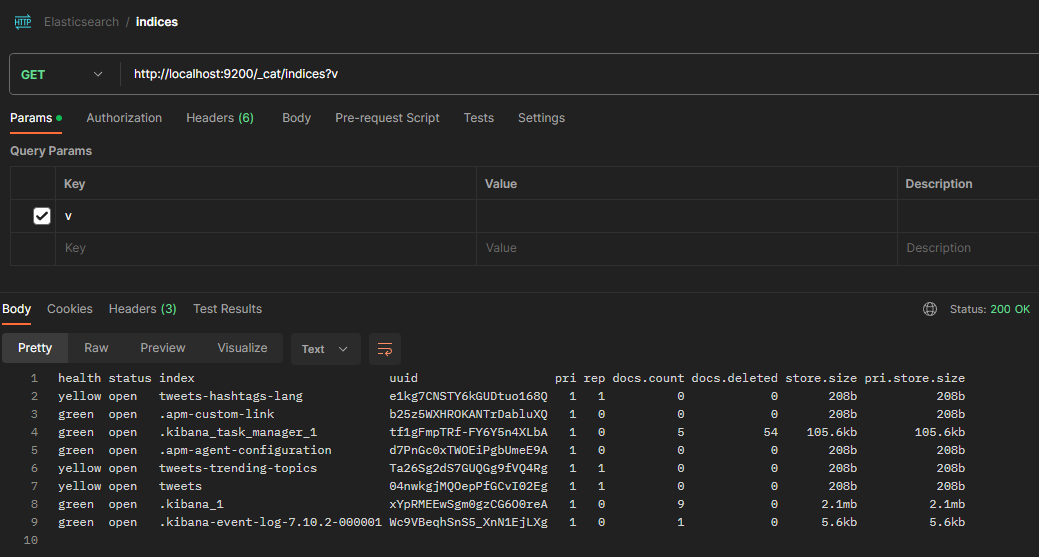
* Creación de índices para los topics





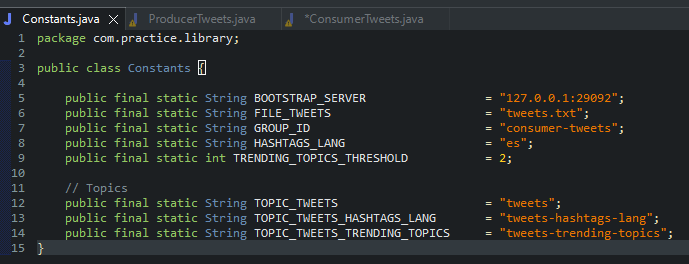


* Ver índices



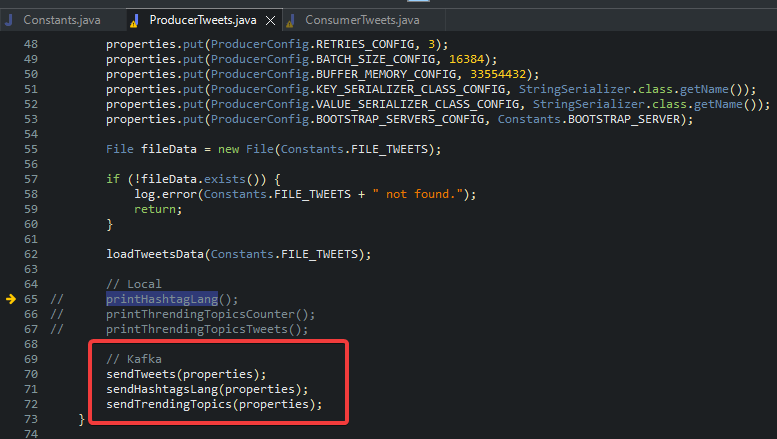
## Configuración de Java

Toda la configuración centraliza en el fichero Constants.java:



## Productor

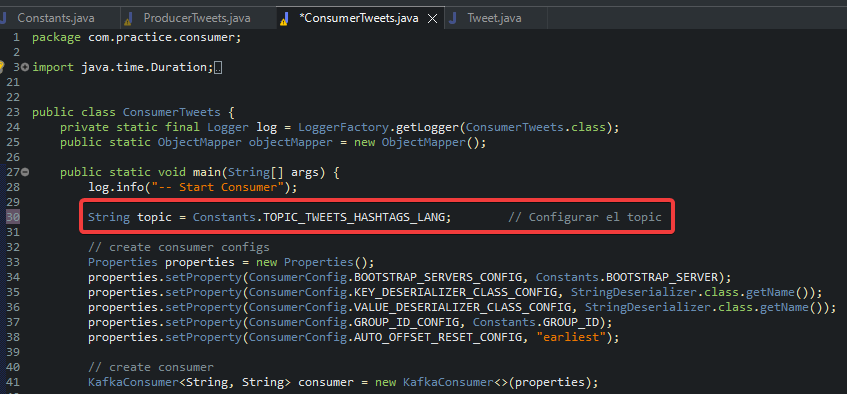
Por el lado del Productor, la carga de tweets se realiza por medio del fichero tweets.txt y el envío de tweets se realizará automáticamente a los topics definidos anteriormente:



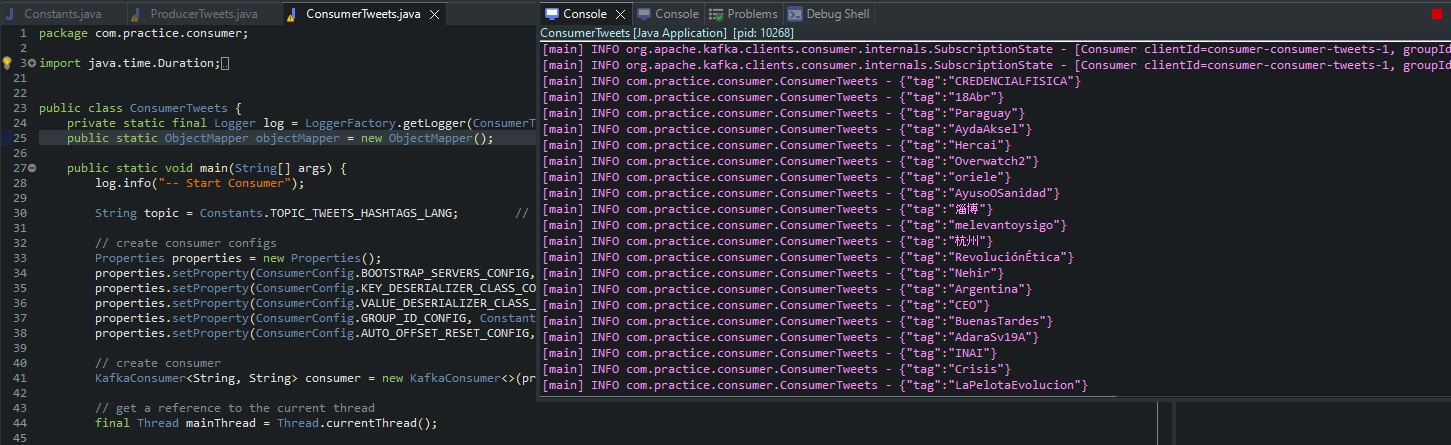
Procedemos a ejecutarlo, si todo ha ido bien los mensajes se llegarán a Kafka y se almacenarán en Elasticsearch.

## Consumidor

El consumidor está definido para leer un topic a la vez. Así pues, hay que configurar la línea marcada con la constante del topic que se quiera leer:

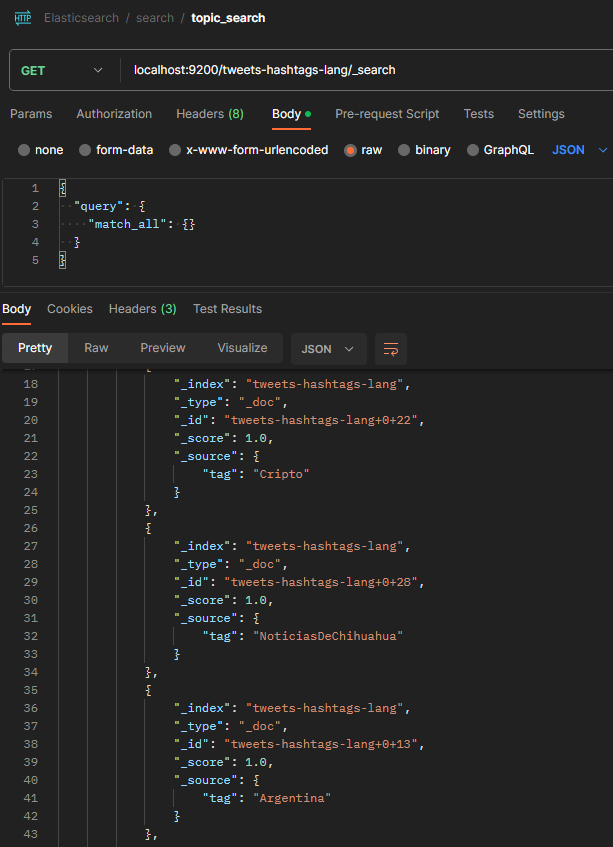


Procedemos a ejecutarlo:

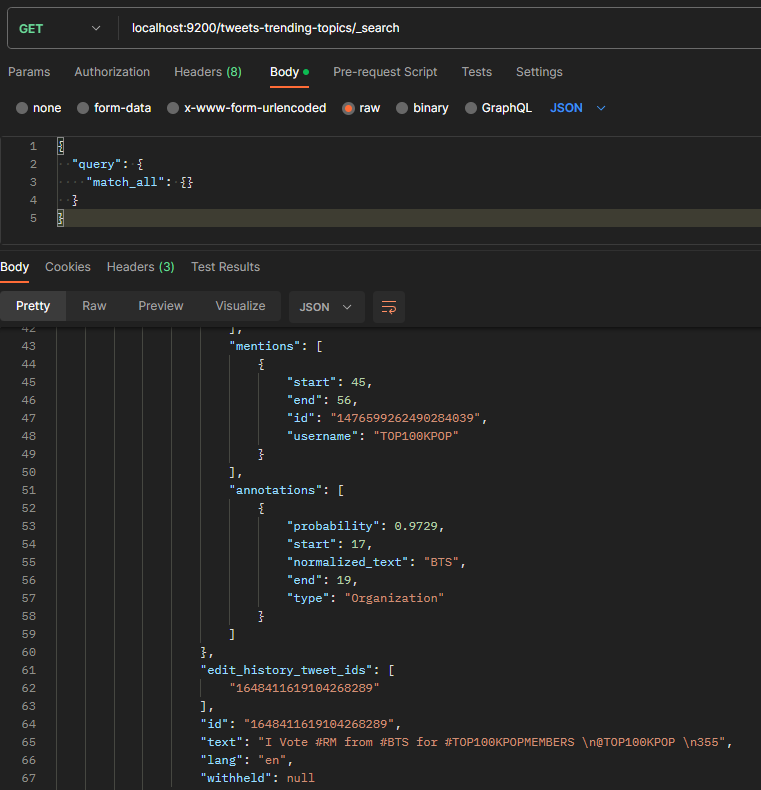


## Elasticsearch

Una vez configurado Elasticsearch y lanzado el Producer realizaremos una búsqueda de los topics:



Si todo ha ido bien veremos la lista de hashtags. El mismo proceso se realiza para leer los otros topics, hay que sustituir en la url el nombre del topic que queremos buscar:



La API de búsqueda tiene mucho juego a la hora de aplicar queries de búsqueda.

## Conclusión

He tenido ciertos problemas a la hora de realizar el filtrado de hashtags, si filtraba por prefijo de lenguaje “es” muchos de ellos estaban en tweets con leguaje muy diferente al español. No utilizo esta red social, pero deduzco que muchas de esas cuentas se crean con un lenguaje incorrecto. Tengo que confesar que no soy muy fan de las redes sociales y nunca he enviado un tweet en mi vida.

Por otro lado, yo soy programador .NET y al ser la práctica en Java me ha llevado más tiempo en realizarla por el aprendizaje del lenguaje y uso de Maven, pero ha estado bien aprender esta tecnología.

En definitiva, la práctica de Kafka ha sido realmente enriquecedora mí y mi carrera profesional, espero poder aplicar estos conocimientos en un proyecto real.