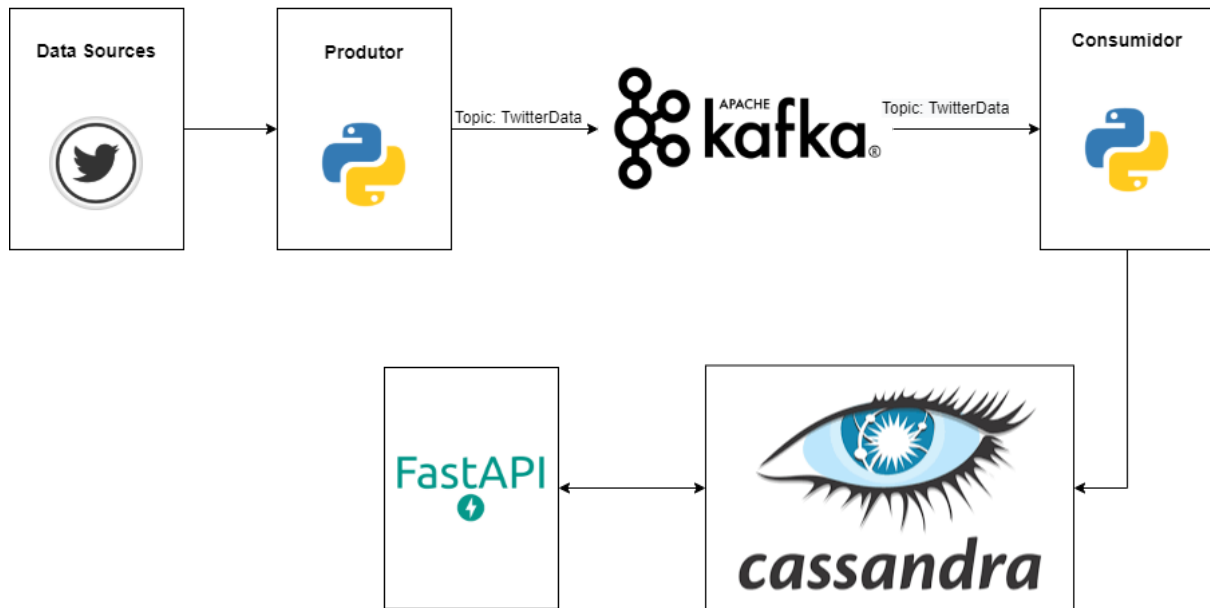


1. Diagrama da solução:



- **Produtor**
Utilizando a função `on_data` do Tweepy com o `KafkaProducer` para alimentar os dados brutos do Twitter no tópico do cluster Kafka.
- **Kafka**
Utilizado para mover os tweets capturados para dentro do data lake.
- **Consumer:**
Utilizado para consumir o tópico do cluster Kafka e inserir no Cassandra.
- **Cassandra:**
Utilizado como data lake, armazenamos os tweets na tabela tweets.
- **FastAPI:**
Interface web para consumo dos dados diretamente do data lake.

2. Benefícios e Tecnologias

- **Kafka:**
Apache Kafka é um dos mais populares modelos Publish-Subscribe que o mundo de BigData tem adotado em larga escala. Os eventos do Kafka podem ser resumidos em mensagens, sendo consumidas e produzidas através de tópicos. Tolerante a falhas.

- **Cassandra**

Banco de dados NoSql, distribuído baseado no Dynamo e a forma de organização dos dados é baseado no BigTable, altamente escalável, principalmente de forma horizontal. Muito utilizado hoje em dia para internet das coisas, logging, aplicações web/mobile ou qualquer outra aplicação que precise de muita escrita.

- **FastAPI**

Moderno e rápido (alta performance) framework web para construção de APIs com Python 3.6 ou superior.

Possui algumas dependências boas como:

Swagger UI:

Uma interface de documentação automática para API, que permite facilmente fazer requisições sem a necessidade de instalar um software.

ReDoc:

Uma outra opção Swagger que disponibiliza uma interface navegável da documentação.