Kafka Training

Extracting the patient data from a health file and loading it to KAFKA topics and then syncing it to DB.

Tech Stack - Spring boot or Python, Any relational DB, Confluent

Milestone 1: Read the .txt and push records to KAFKA topic (Auth-Topic)

1. Read the .txt file and parse the file with help of mapper.xlsx
2. Records starting from SUB to another SUB records belongs to a single subscriber, and this all record needs to be send as a single JSON
3. Create a output JSON and push to KAFKA topic.
4. This milestone will require data modelling.

Milestone 2: Split data to respective topics using KSQL

1. Read the JSON from Auth-Topic and split the JSON for (Subscriber, Patient, Case, Service) along with source identifier as “KSQL” using KSQL
2. Send the splitted objects to their respective topics

Milestone 3: Split data to respective topics using KSTREAMS

1. Read the JSON from Auth-Topic and split the JSON for (Subscriber, Patient, Case, Service) along with source identifier as “KSTREAMS” using KSTREAMS in spring boot.
2. Send the splitted objects to their resp. topics

Milestone 4: Sync data to DB using KCONNECTOR

1. Sync the data from resp. topics of (Subscriber, Patient, Case, Service) to their resp. DB tables. (It should be upsert instead of insert)

Milestone 5: Sync data to KSQL table

1. Sync data from KSQL streams to KSQL table.
2. Sync data present in Kafka topic (one behind the KSQL table) also to Postgres DB

Resources:

1. Confluent (Platform to run all Kafka related tech)
2. Kafka 101
3. Kafka Connect
4. KSQL
5. Spring Boot for Apache Kafka
6. Elephant SQL (In case if you’re not able to install any DB in your laptop locally)