SOFE 4630 Winter 202 - Cloud Computing

Project Milestone#2 : Data Ingestion Software

Course Group No: 12

Name : Ashwin Shanmugam

Student ID : 100700236

Date : 02/15/2022

Media Link : https://drive.google.com/drive/folders/1YluSp20cA2NS5i7qZsABX5yAv7kNyYMr?usp=sharing

What is EDA? What are its advantages and disadvantages?

- Event driven architecture is a type of software architecture and application design methodology.
- Events are used in event-driven architectures to trigger and establish connections between decoupled services.
- This architecture allows systems to share information in real-time.

Advantages :

- **loosely coupled :**
    - Components of EDA are loosely coupled.
    - This facilitates the logical separation of event creation and consumption.
    - Since it is loosely coupled, individual components can employ different technologies and could also be written in different programming languages.
    - It improves the scalability of applications.

- **Cost effective :**
    - In Event-driven architecture messages are published in the queue by the producer and received by the consumer.
    - Since it shares information in real-time, this eliminates the need for constant polling to check for events thereby reducing the bandwidth usage.

Disadvantages :

- **Error Handling :** It is difficult to debug and handle errors since modern applications will contain several messages from producers. They also contain several partitions and consumers.
- **Inconsistent :** Making changes to an existing event is challenging in EDA, which makes it hard for testing and debugging.

In Kafka, what's meant by cluster, broker, topic, replica, partition, zookeeper, controller, leader, consumer, producer, and consumer group?

Broker : It is a kafka server that runs in the Kafka cluster.

Topic : A Topic is a category/feed name to which records are stored and published.

Replica : Multiple copies of data spread across the kafka cluster.

Partition : Topics are divided into 1 or more partitions.

Producer : It sends produce requests to the kafka cluster.

Consumer : It subscribes to topics and consumes the messages sent by the producer.

A problem in the used YAML file to create the docker images is that the data inside Kafka clusters are not persistent which means if the docker images are down, all its messages are lost. Update the YAML file for persistent data (hint: it's related to the volume options in Kafka brokers and zookeeper). Describe how this update solves the problem.

```
volumes:
  - /home/srv/kafka:/kafka/kafka-logs-1
```