SOFE4630 Cloud Computing (Winter 2022 - Dr. M. El-darieby)

Lab 2: Kafka

Feb 15, 2022

Group 6

Cassidy Linhares	100615025

Part 1: Kafka Local & Volumes

Part 2: Kafka Cloud

What is EDA? What are its advantages and disadvantages?

EDA is event driven architecture. Unlike request/response architecture, EDA follows a more publish and subscribe method (Similar to that of subscribing to emails from your favourite shop).
The advantages is are not having to make the request and then wait, you just pull what you subscribed to, increases decoupling, more interrupt driven instead of poll driven. Making things loosely coupled allows for easy scalability. Another advantage is that failures don't affect the whole system, just the individual service. It is also good for high throughput and large data ingestion
The disadvantages include maintaining high availability since outages and stalling are inevitable. Also when replication begins to come into play then consistency can be hard to maintain and manage as well.

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

cluster	A group of brokers in kafka
broker	A broker is a middle-man that is responsible for communicating to publishers and subscribers since the publishers and subscribers don't know about each other. Think of a house realtor. You don't know who the buyer of your house is and you don't deal with them either. Your realtor will handle that for you. In kafka, each machine/server that has partitions is a broker. A broker can have multiple topics
topic	A topic is exactly like a topic in English class. It's used for organizing and grouping data within kafka. The topic lives within a broker. A topic has multiple partitions
replica	A replica, in kafka, is a duplicate of a partition from another cluster
partition	A partition is part of a topic and it is where an event message is stored. It is made for scalability and replication
zookeeper	The zookeeper is in charge of keeping track of the offsets for a consumer. An offset is similar to a time stamp so the consumer knows where they left off. The zookeeper is also responsible for monitoring the status of a kafka cluster, its partitions, and replicas.
controller	They are the broker that is in charge of administrative stuff like who will be leader if they die, assigning partitions, and checking if other brokers failed
leader	This is the broker where the consumer reads from. It will always have the most up-to-date data on each topic and partition in it. All the other brokers need to come and get its updates from it. The leader is there the producers make events to
consumer	These are subscribers and they read or use the data that is stored in kafka
producer	These are what make the events and publish them to the broker
Consumer group	A group of consumers that are subscribed to the same topics or have the same functions. Only one consumer in the group will pull the data from kafka but all will process the data. If one fails, then the other members will rebalance themselves

What are Docker Volumes?

Docker volumes are a way to persist data after the container shuts down. Every time you run docker-compose down, all the data it collected gets deleted with it so volumes are an easy way to save the container data to a location on the host machine.