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

Lab 2: Data Ingestion Software - Kafka Clusters

Feb 15, 2022

Group 6

Cassidy Linhares	100615025
Iliya Karac	100703933
Kaushik Ramani	100651855
Tanzir Hossain	100709113

GitHub: repo

VIDEO LINKS:

Part 1

Part 2

What is EDA? What are its advantages and disadvantages?

Event driven architecture or EDA is software architecture that follows events. There are many different types of architectures like MVC, Restful, microservice, service oriented or message driven. In the EDA there are 3 main parts. Event producers, event consumers, and the data stream. The producers and consumers treat the data stream like a broker. They can publish and subscribe to a topic. The stream ingests events and holds them in chronological order like cache but it does not delete them when they become outdated. Kafka is a hybrid between a controlled database and an event driven architecture with publishers and subscribers and a cluster that consists of brokers.

Pros of EDA :

- High throughput
- Disk based storage
- Scalable
- Fault tolerant
- Allows many producers and consumers
- Stores trillions of events depending on storage
- You can add storage

Cons EDA :

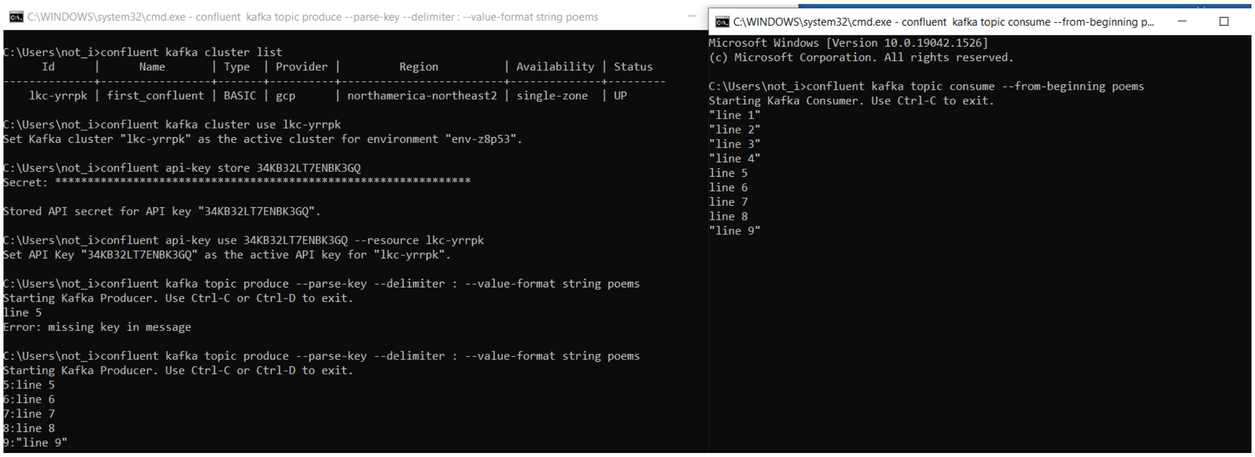
- CAP theorem if the system is large enough that it is distributed
- Not applicable in real time systems
- Latency issues in larger systems
- Complicated
- Usually not necessary
- Eventually will need do delete events

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

- **Cluster**: This is a group of brokers in Kafka
- **Broker**: this is a machine in a cluster that has the topics along with partitions and holds the events. There are usually multiple brokers in a cluster and one of the brokers is a leader.
- **Topic**: This is a way of categorizing events and they live inside of a broker. The same topic can be split across multiple brokers and each topic has partitions for further sorting/categorizing
- **Replica**: This is a duplicate of a partition
- **Partition**: This is a part of a topic that is used for further categorizing data. There can be multiple partitions across multiple brokers
- **Zookeeper**: Used for service synchronization and controller elections

- **Controller**: A broker that is responsible for managing partitions and replicas. Used mainly for admin tasks such as reassigning partitions. There can only be one controller at any given time
- **Leader**: The leader is the elected broker. They are responsible for all the main reads and writes. The other brokers need to update their topics/partitions from the leader
- **Consumer**: A service that reads data from a set of topic(s). Only messages that have been synchronized across all the brokers can be read.
- **Producer**: A service that creates messages/events to be stored in the kafka cluster. The producer specifies which topic to send the message to.
- **Consumer group:** Group of consumers subscribed to the same topics. 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

Installation of CLI:


C:\WINDOWS\system32\cmd.exe - confluent kafka topic produce --parse-key --delimiter : --value-format string poems
C:\WINDOWS\system32\cmd.exe - confluent kafka topic consume --from-beginning p...

```
C:\Users\not_i>confluent kafka cluster list
      Id      |       Name       | Type  | Provider |         Region          | Availability | Status
--------------+------------------+-------+----------+-------------------------+--------------+---------
    lkc-yrrpk | first_confluent | BASIC | gcp      | northamerica-northeast2 | single-zone  | UP

C:\Users\not_i>confluent kafka cluster use lkc-yrrpk
Set Kafka cluster "lkc-yrrpk" as the active cluster for environment "env-z8p53".

C:\Users\not_i>confluent api-key store 34KB32LT7ENBK3GQ
Secret: ****************************************************************

Stored API secret for API key "34KB32LT7ENBK3GQ".

C:\Users\not_i>confluent api-key use 34KB32LT7ENBK3GQ --resource lkc-yrrpk
Set API Key "34KB32LT7ENBK3GQ" as the active API key for "lkc-yrrpk".

C:\Users\not_i>confluent kafka topic produce --parse-key --delimiter : --value-format string poems
Starting Kafka Producer. Use Ctrl-C or Ctrl-D to exit.
line 5
Error: missing key in message

C:\Users\not_i>confluent kafka topic produce --parse-key --delimiter : --value-format string poems
Starting Kafka Producer. Use Ctrl-C or Ctrl-D to exit.
5:line 5
6:line 6
7:line 7
8:line 8
9:"line 9"
```

```
Microsoft Windows [Version 10.0.19042.1526]
(c) Microsoft Corporation. All rights reserved.

C:\Users\not_i>confluent kafka topic consume --from-beginning poems
Starting Kafka Consumer. Use Ctrl-C to exit.
"line 1"
"line 2"
"line 3"
"line 4"
line 5
line 6
line 7
line 8
"line 9"
```