Apache Kafka

Overview

Apache Kafka is a framework implementation of a software bus using stream-processing. It is an open-source software platform developed by the Apache Software Foundation written in Scala and Java. The project aims to provide a unified, high-throughput, low-latency platform for handling real-time data feeds.

Use cases of Kafka

- Service bus between micro Services
- Central Data hub for Realtime Stream processing and analytics
- Realtime Tracking
- Fradulence

Quick Start

Installation

- Refer Dev Setup documentation to know how to run docker.
- Make sure broker is up and running : `docker-compose ps`

Create Events in CLI

- Run command : `docker-compose exec broker sh` in the root of the project.
- Create a kafka topic

```
kafka-topics \
   --bootstrap-server http://broker:9092 \
   --topic test1 \
   --create \
   --replication-factor 1 \
   --partitions 6
```

- Run kafka-console-producer

```
kafka-console-producer \
   --topic test1 \
   --broker-list http://broker:9092 \
   --property parse.key=true \
   --property key.separator=,
```

- Pass the values like below

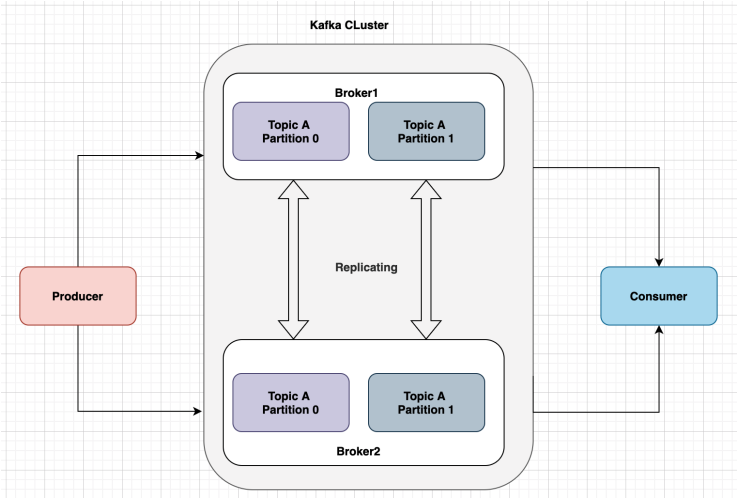
```
>
>
>alice,{"count":1}
>alice,{"count":2}
```

Consumer Events in CLI

- Parallely open another tab and point to root of the project and run : `docker-compose exec broker sh`
- Run below command to consume the event which we produced:

```
kafka-console-consumer \
    --topic test1 \
    --bootstrap-server http://broker:9092 \
    --property print.key=true \
    --from-beginning
alice {"count":0}
alice {"count":1}
alice {"count":2}
```

Architecture



Core Concepts

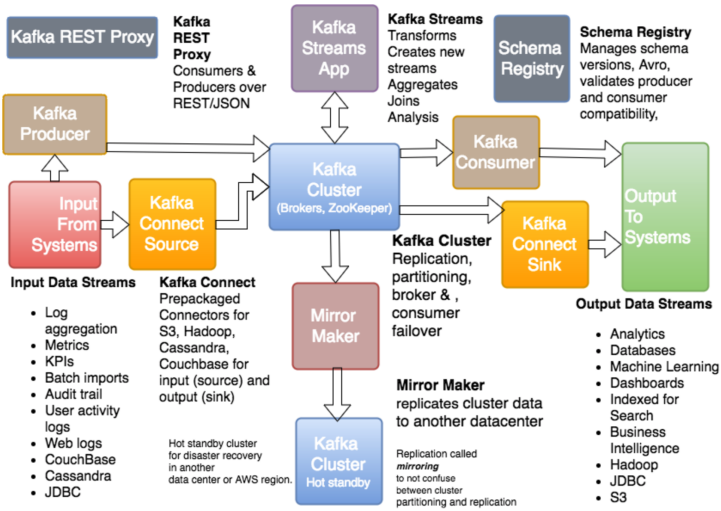
- Messages & Batches (Like records in a table)
- Schemas (like column names of a table)
- Topics & Partitions (like distributed table)
- Producers & Consumers
- Brokers & Clusters (nodes in a cluster)

Highlights of kafka

- Multiple producers and Consumers
- Disk Based Retention
- Scalable
- High Performance

Kafka Eco-System

Kafka Ecosystem: Diagram of Connect Source, Connect Sink, and Kafka Streams



References

https://kafka.apache.org/

https://kafka-tutorials.confluent.io/

https://www.oreilly.com/library/view/kafka-the-definitive/9781491936153/

https://docs.confluent.io/platform/current/connect/index.html#:~:text=Kafka Connect is a tool,into and out of Kafka.

https://docs.confluent.io/platform/current/overview.html