

# Handling Streaming Data with a Kafka Cluster

---

EXPERIENCING KAFKA AS A STREAMING PLATFORM



**Bogdan Sucaciu**

SOFTWARE ENGINEER

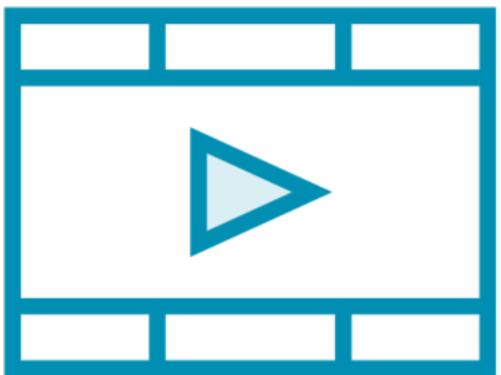
@BSucaciu    [bsucaciu.com](http://bsucaciu.com)



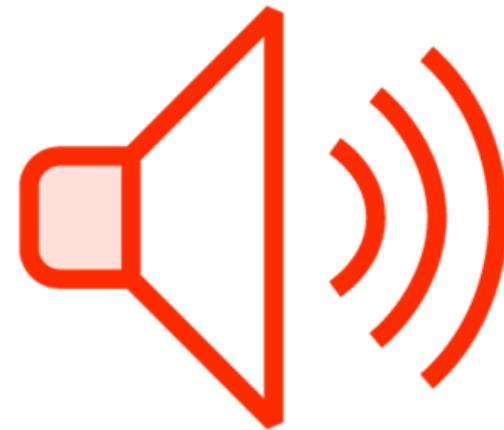
# What is Streaming?



# Streaming



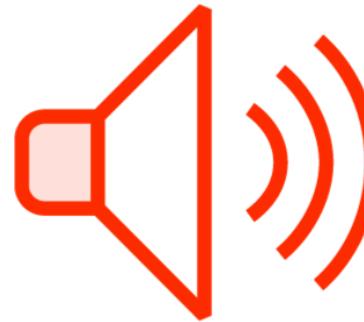
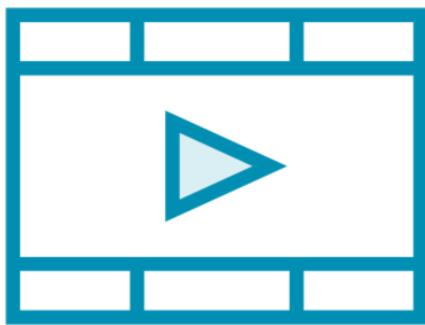
**Video Content**



**Audio Content**



# Streaming

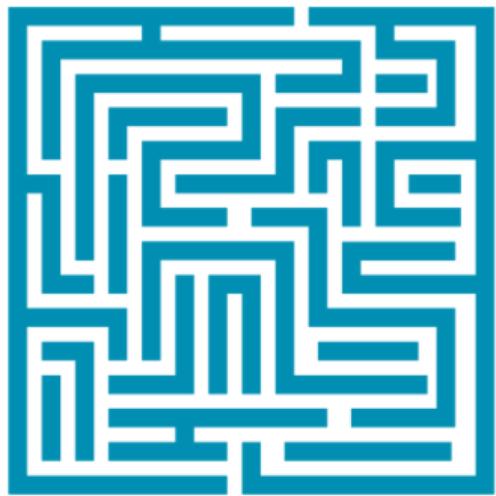


.....

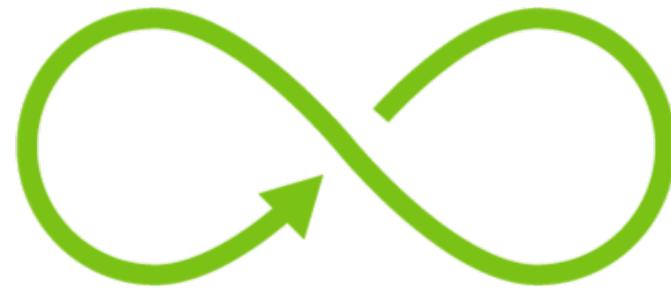
**Streaming Data**



# Streaming Data



**Unbounded**



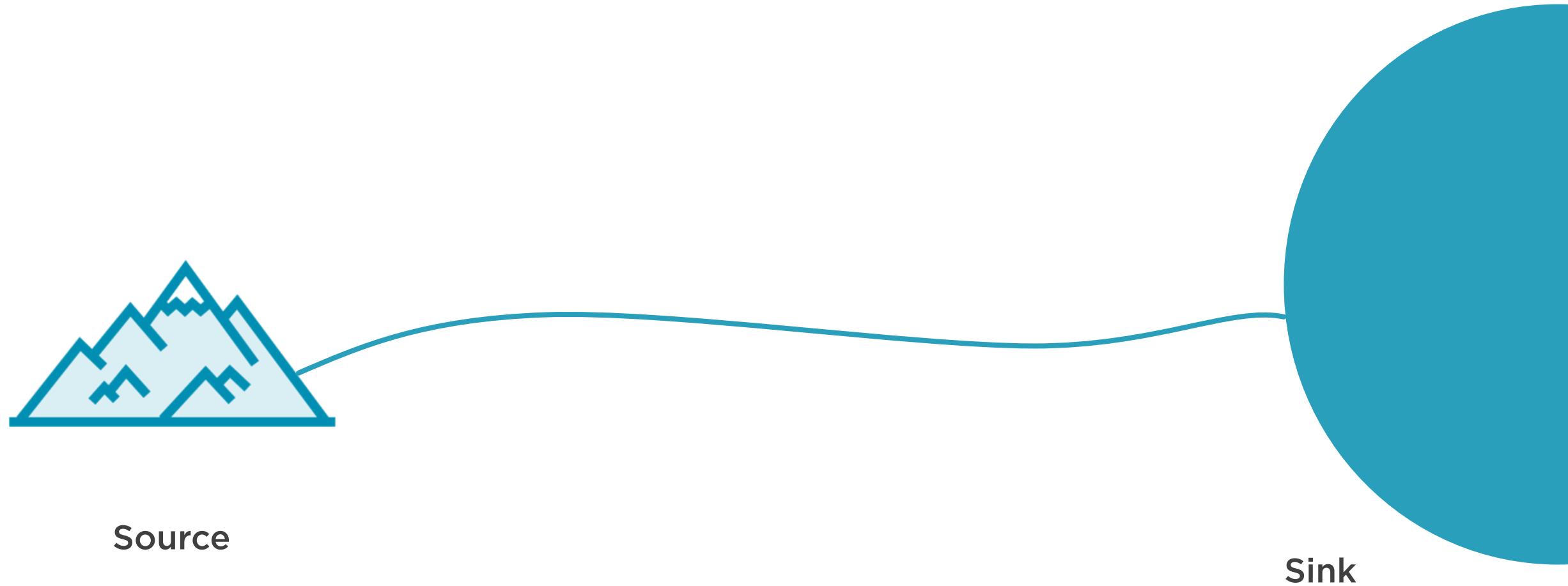
**Continuous**



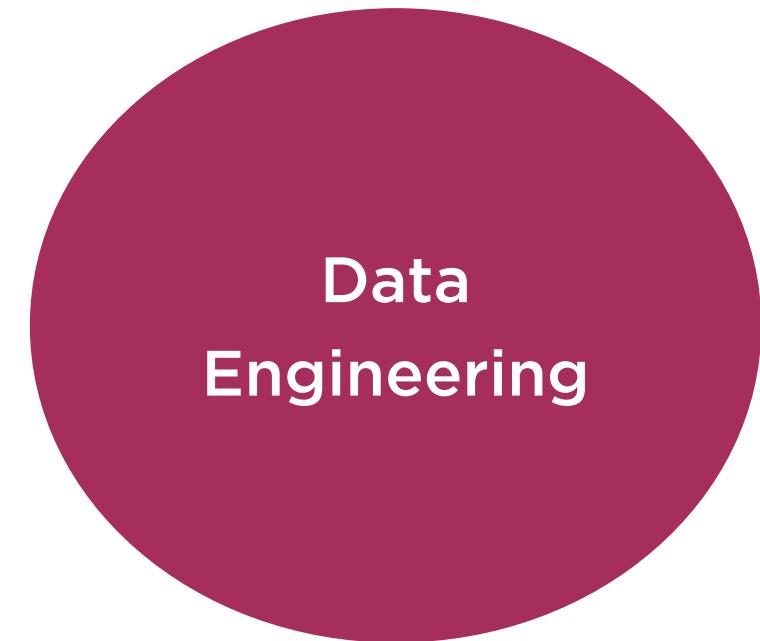
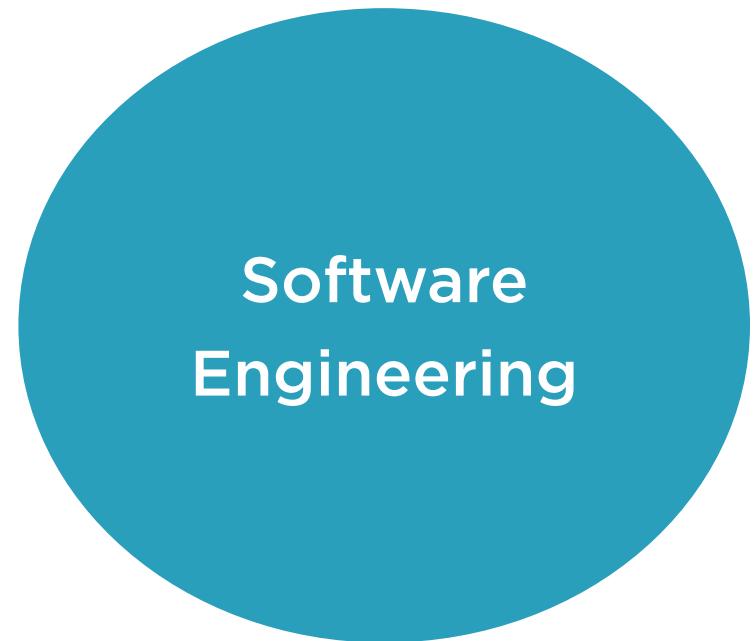
# Streaming Data



# Streaming Data



# Apache Kafka



Messaging  
System

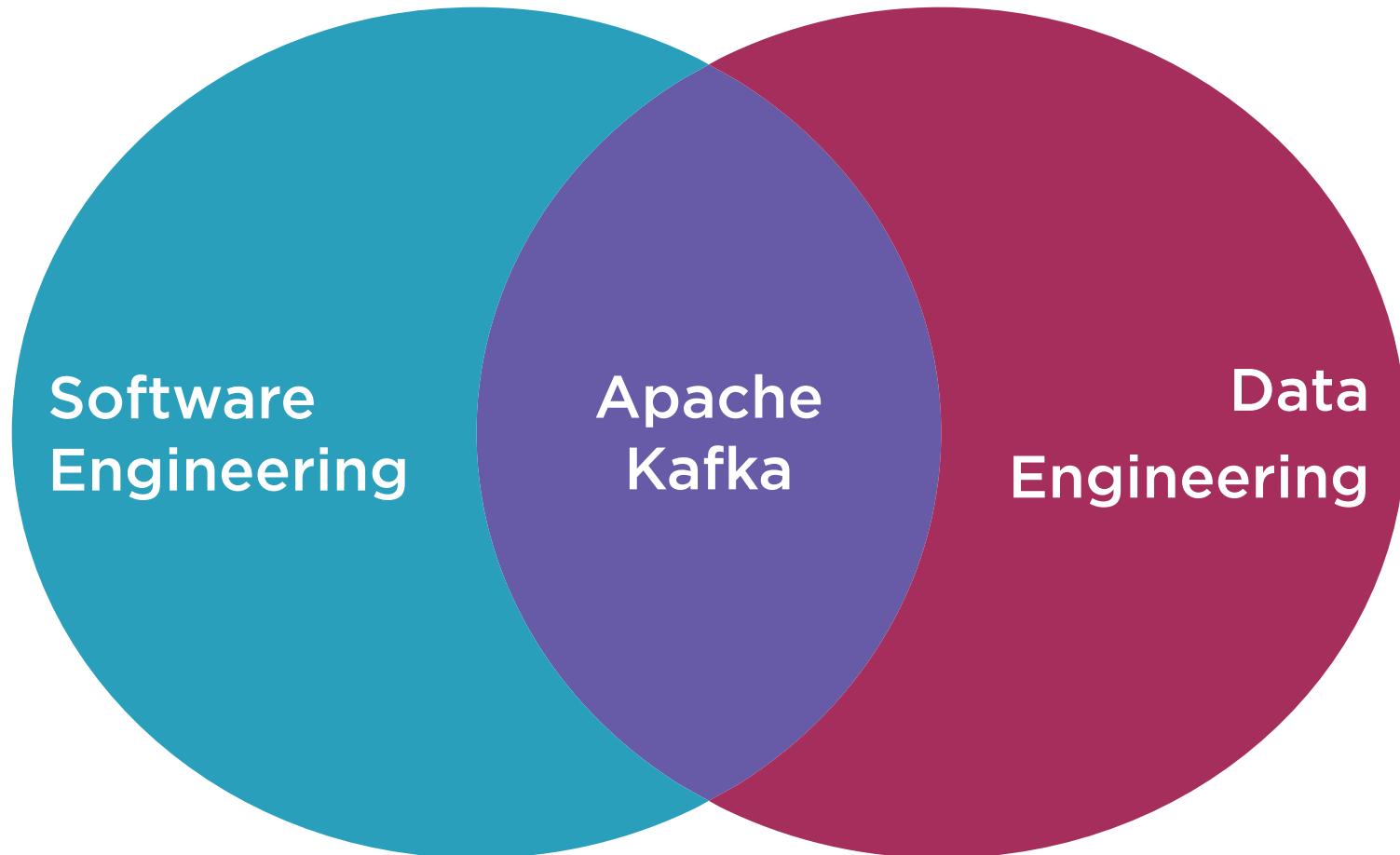
Event-Driven  
Architecture

Streaming  
Data

ETL



# Apache Kafka



# Pluralsight Courses

Getting Started with Apache Kafka

by Ryan Plant

This course will introduce you to Apache Kafka and provide a thorough tour of its architecture so you can start building your next enterprise system with it.

Resume Course    Bookmark    Add to Channel    Download Course

Streaming with Kafka

## Designing Event-driven Applications Using Apache Kafka Ecosystem

by Bogdan Sucaciuc

Apache Kafka has become one of the most popular tools available when it comes to enterprise messaging and streaming. In this course, you will learn the basics of creating an event-driven system using Apache Kafka and the ecosystem around it.

Resume Course    Bookmark    Add to Channel    Download Course



# Prerequisites



Docker



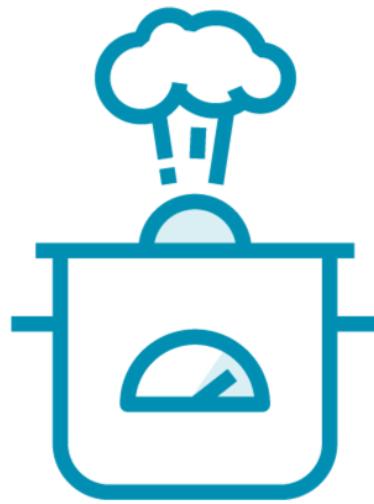
Java



# Online Food Delivery



Order



Prepare



Delivery



# Apache Kafka Architecture

---



# Apache Kafka History

2010

LinkedIn

Initial Development



# Apache Kafka History

2010

**LinkedIn**

Initial Development



2011

**Open Source**



# Apache Kafka History

2010  
**LinkedIn**  
Initial Development

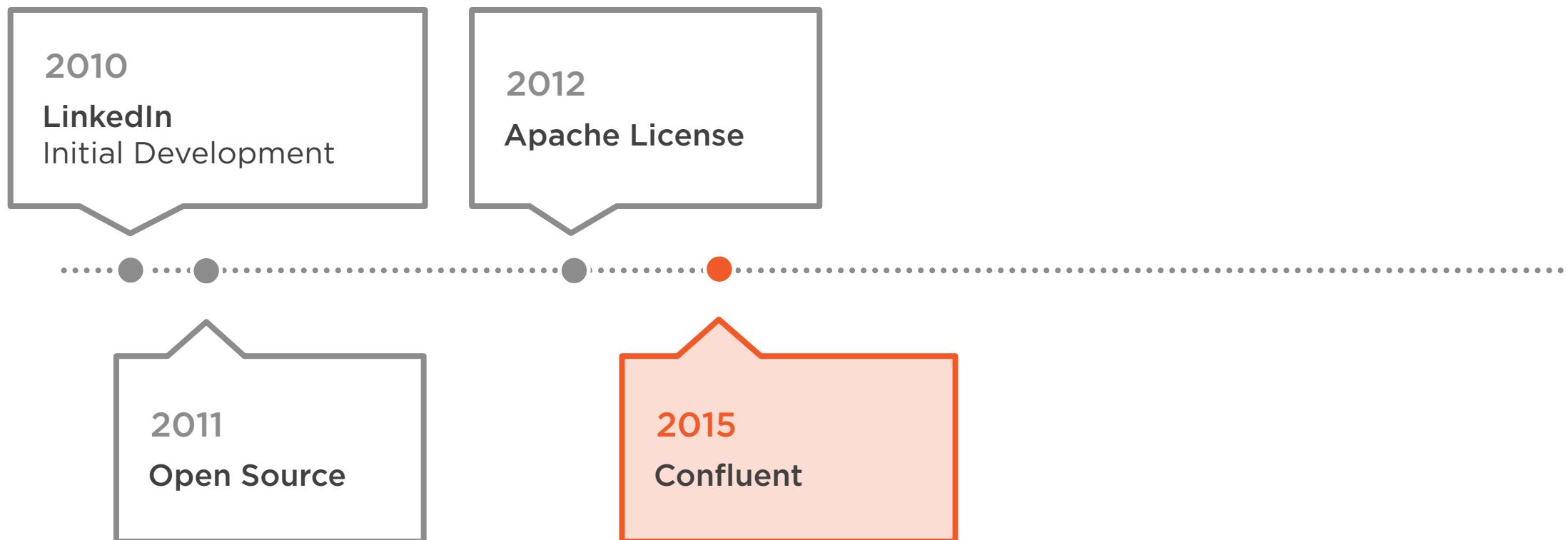
2012  
**Apache License**



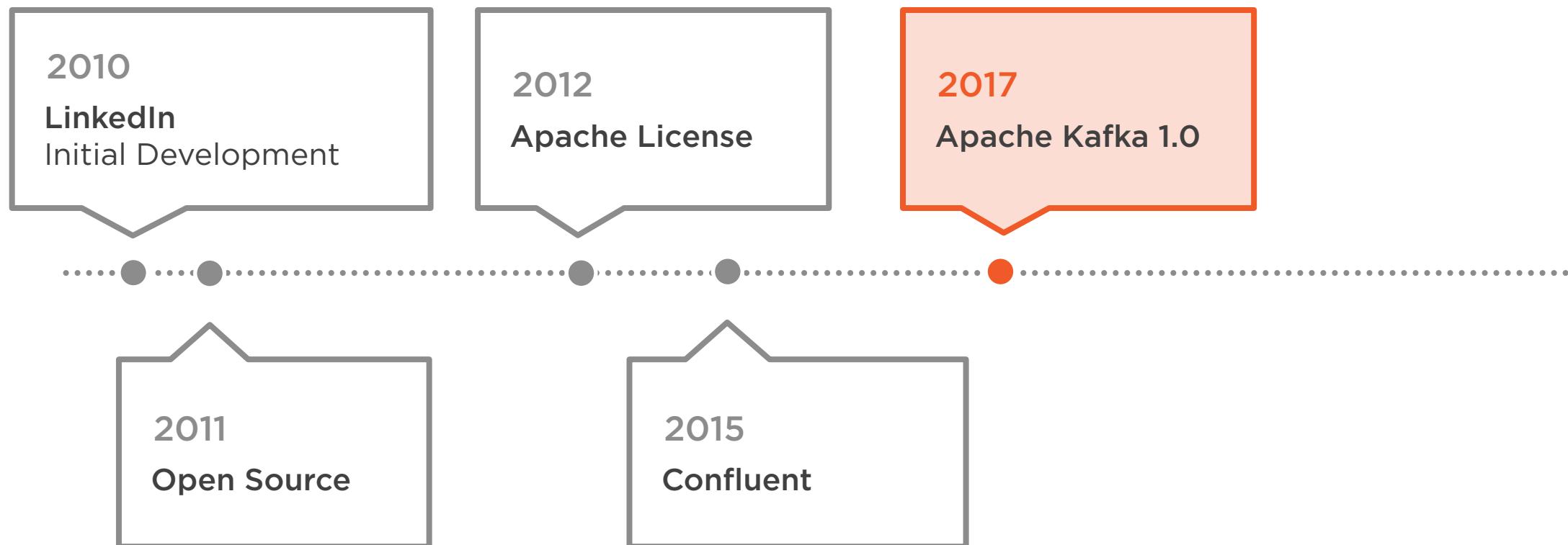
2011  
**Open Source**



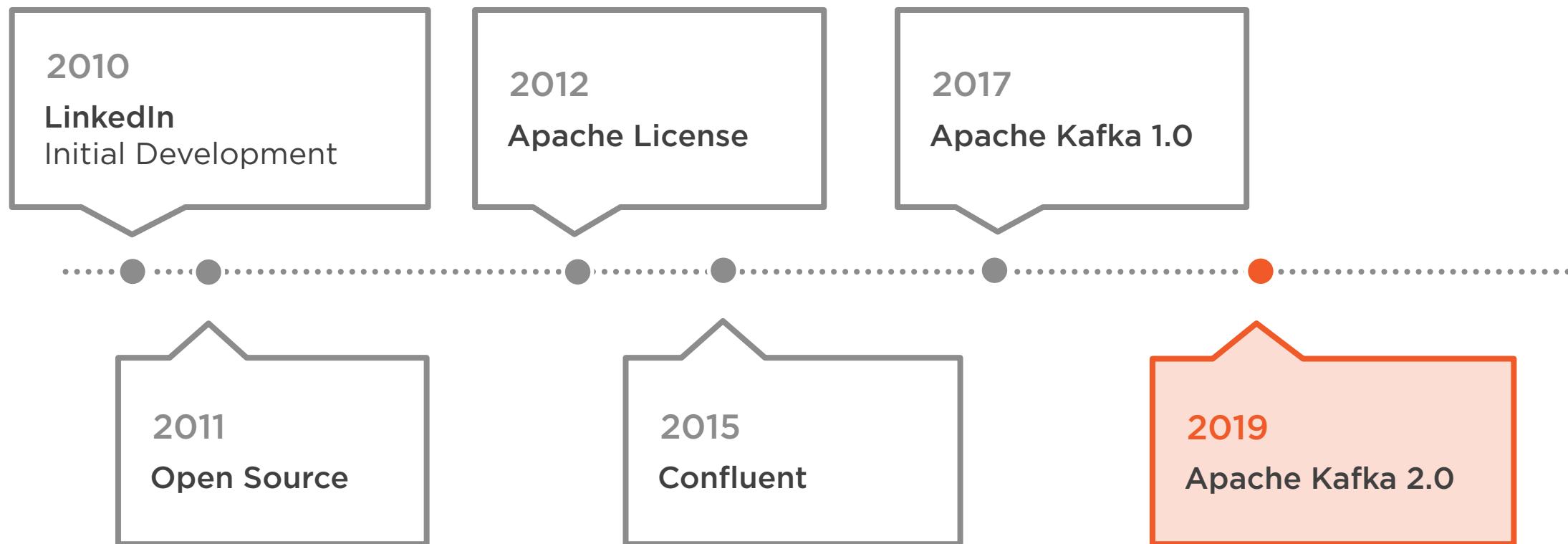
# Apache Kafka History



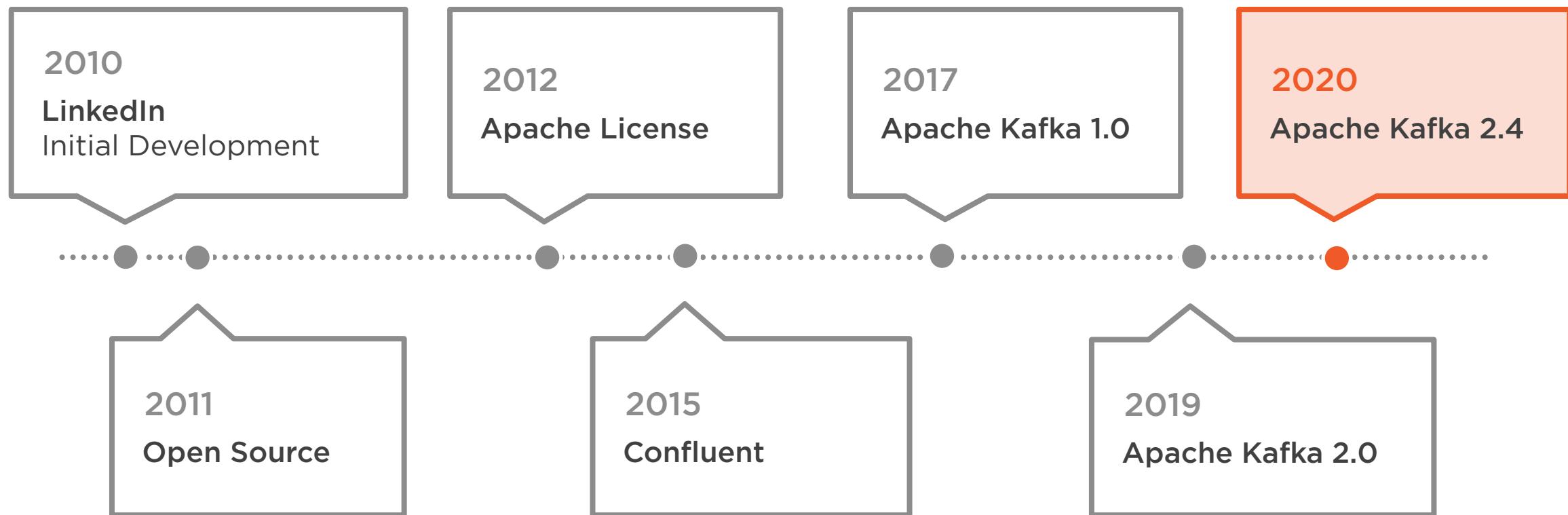
# Apache Kafka History

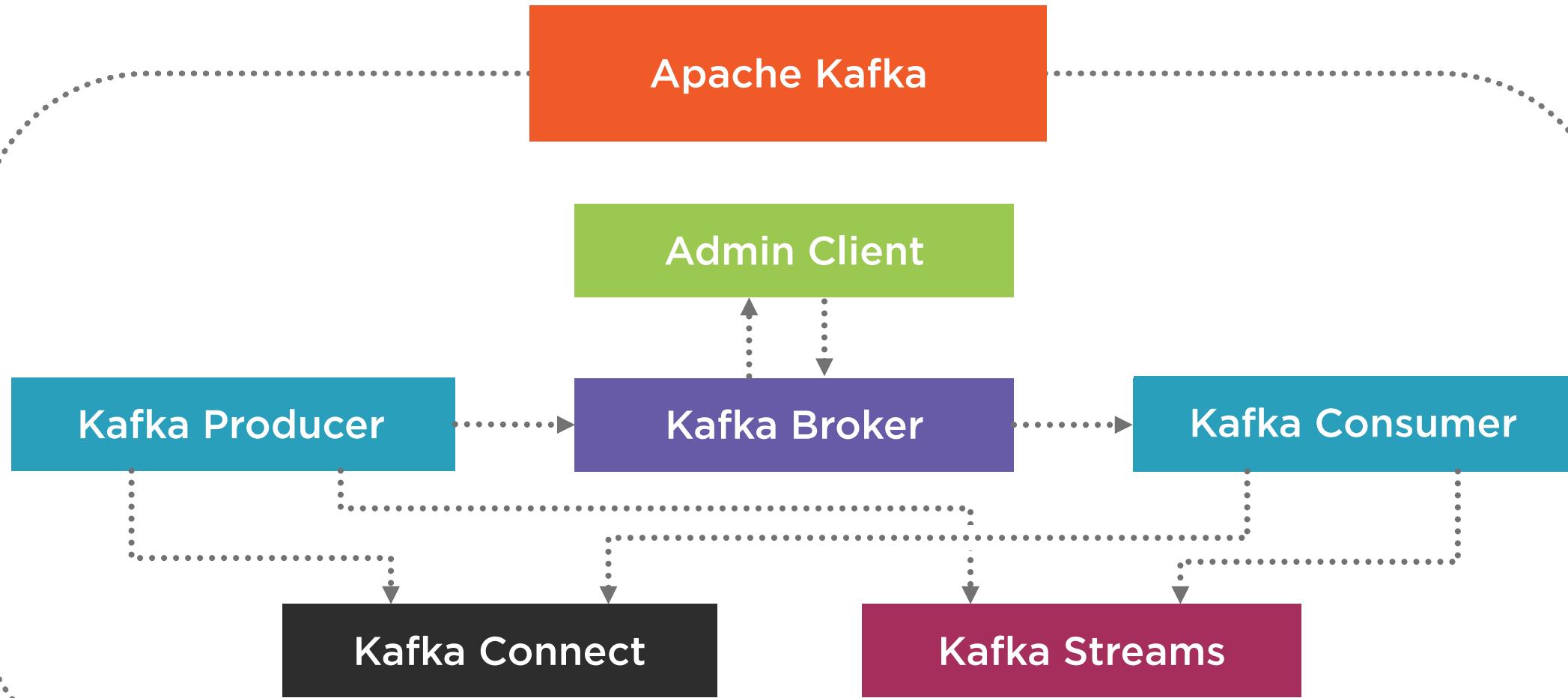


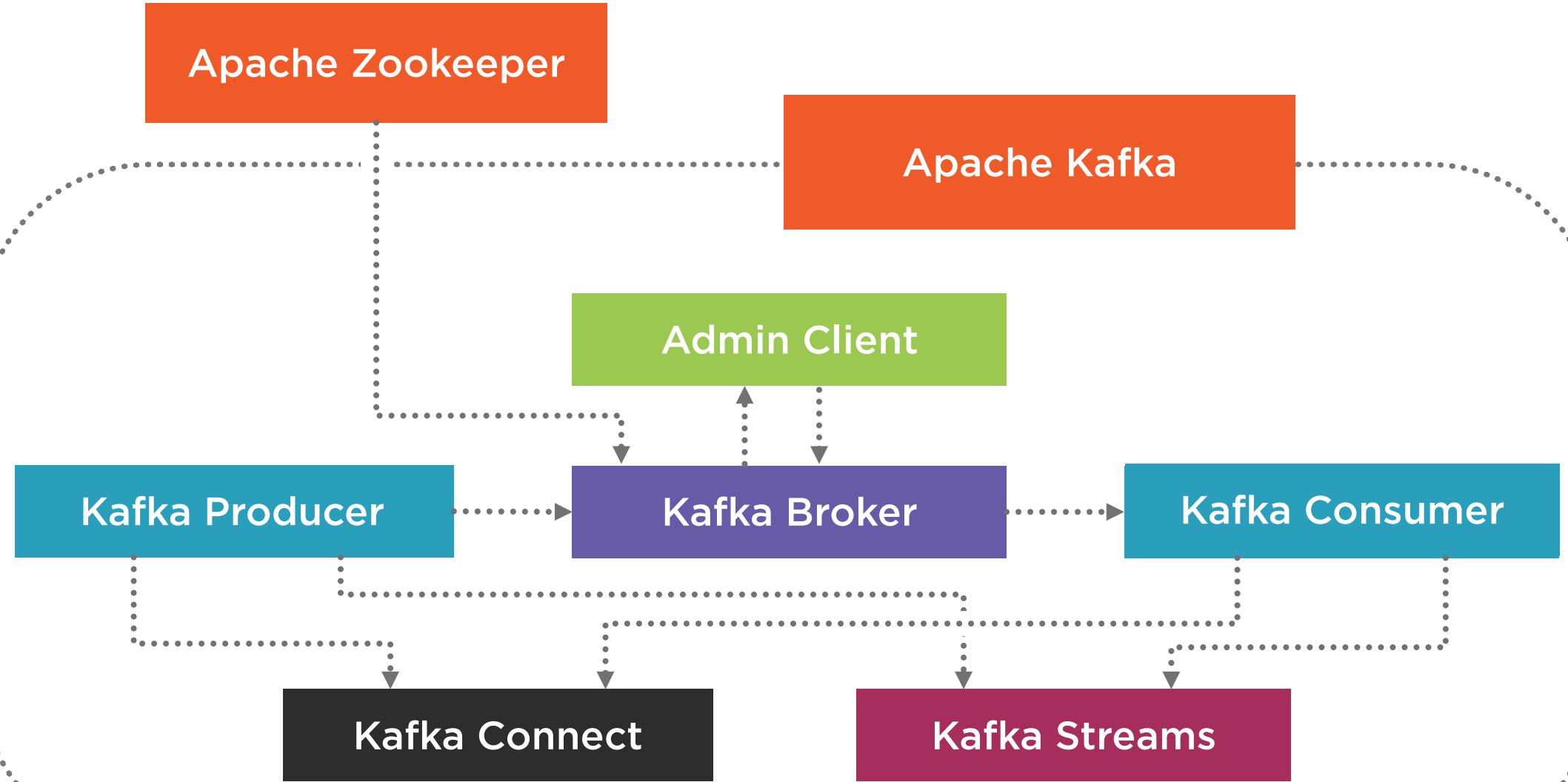
# Apache Kafka History



# Apache Kafka History

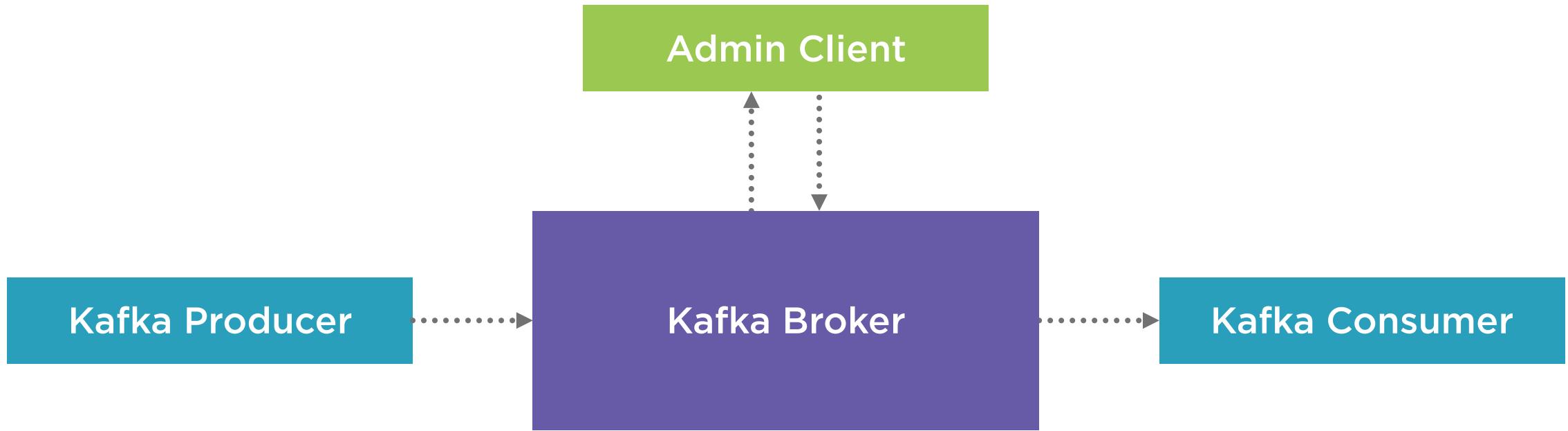






**KIP-500: Replace ZooKeeper with a Self-Managed Metadata Quorum**





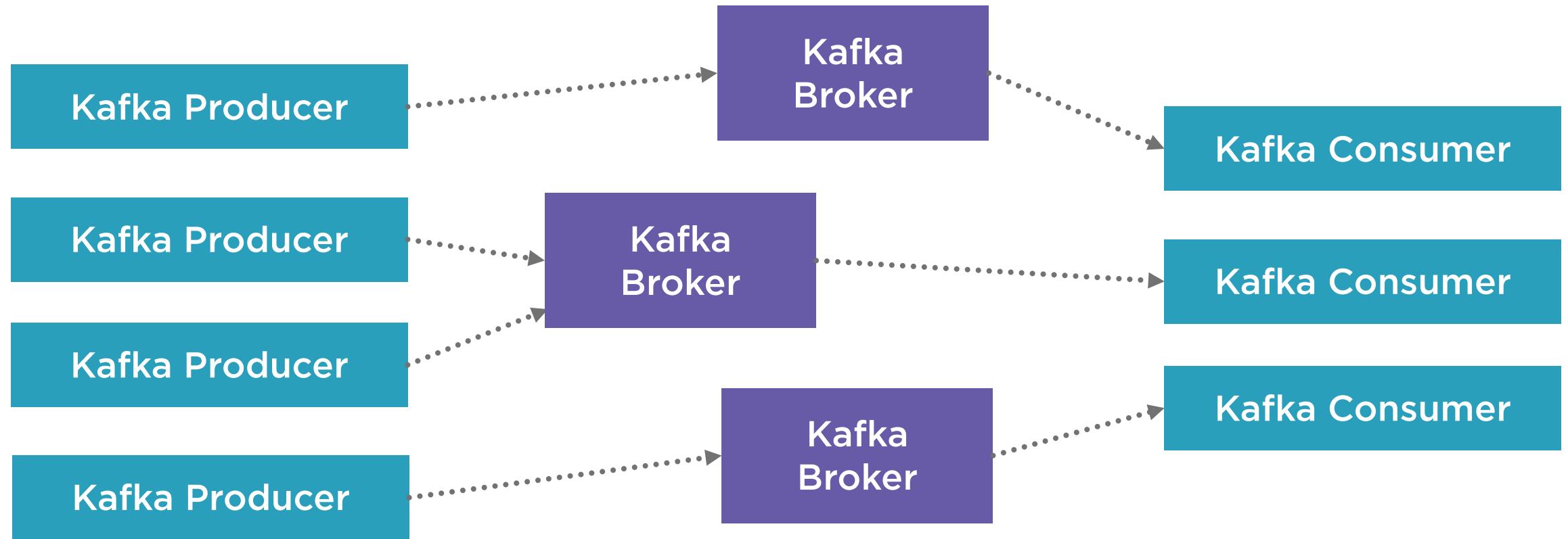




# Handling Streaming Data with a Kafka **Cluster**







# Events, Messages, Records

---



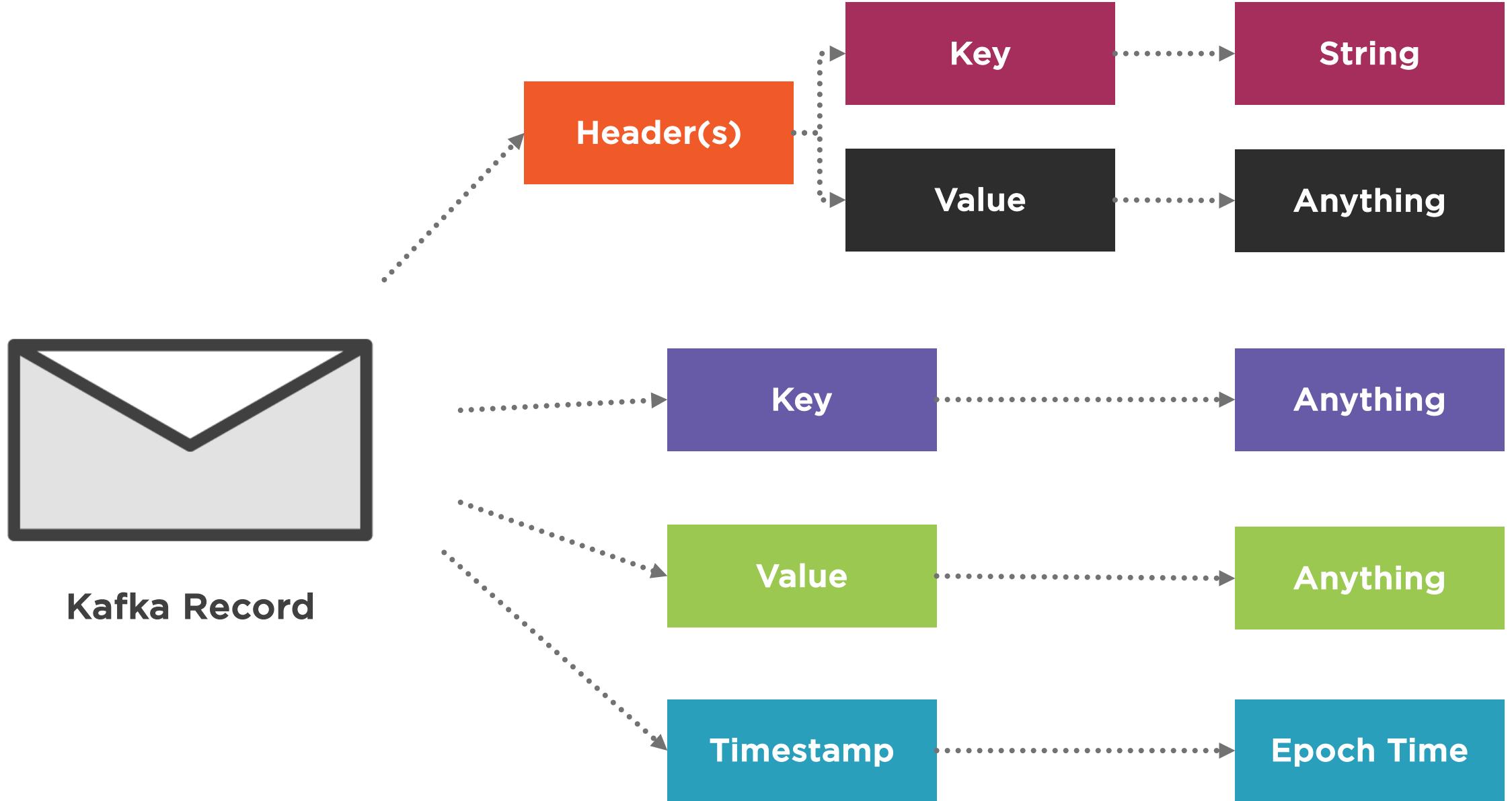
They all refer to  
the **same** thing



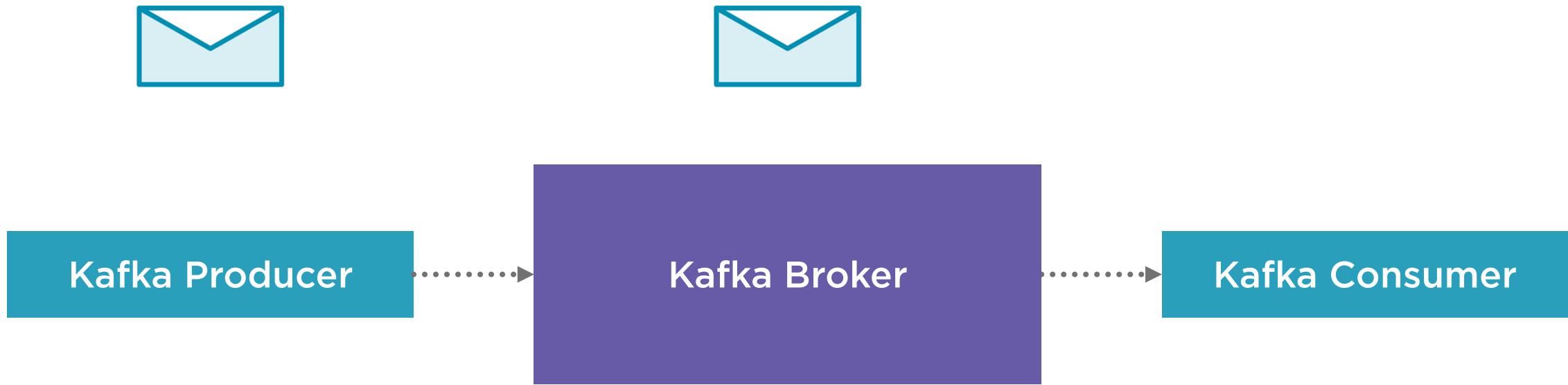
Kafka as an ...

Event Store  
Messaging System  
Streaming Platform





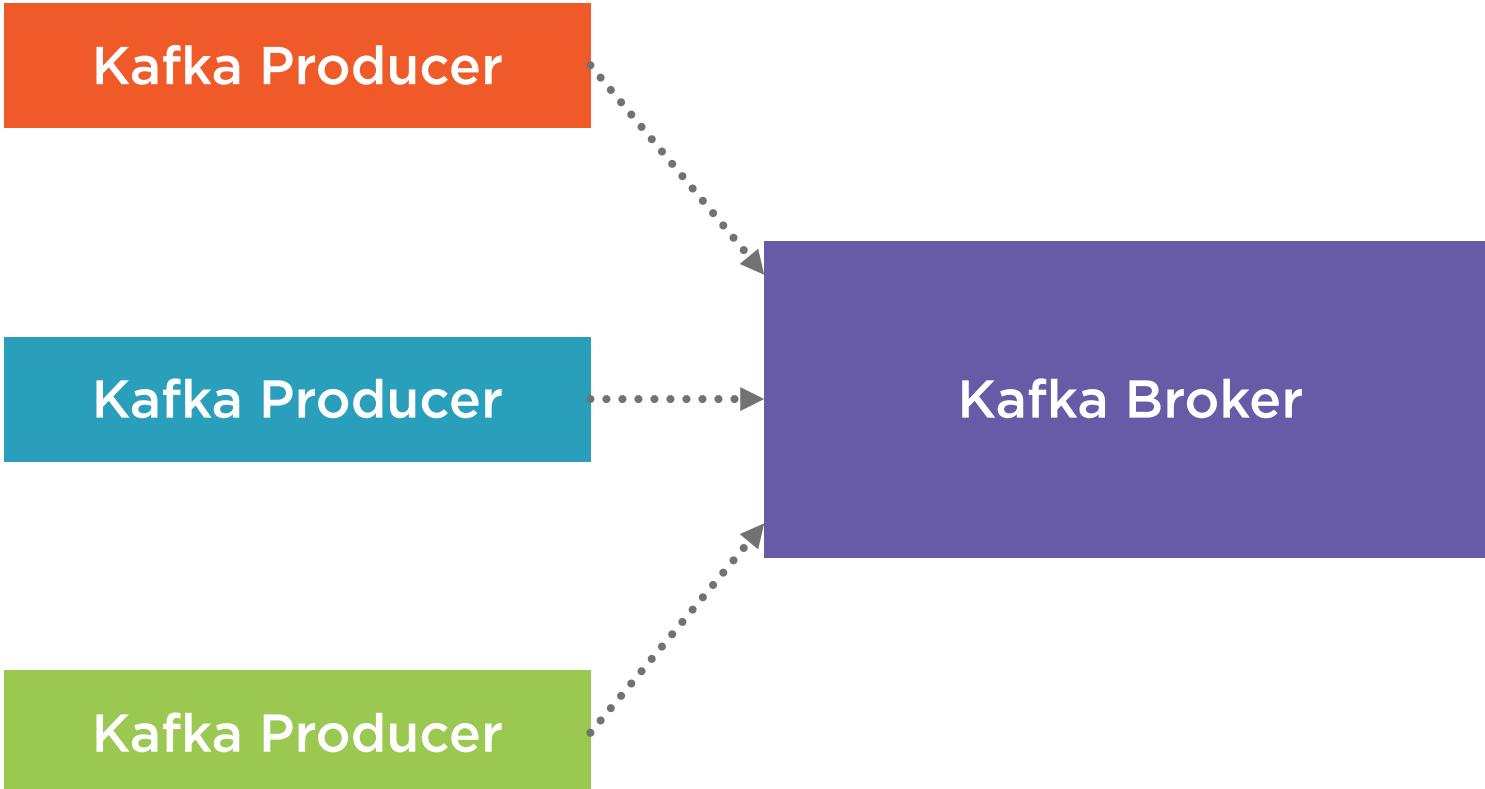
# Data Flow



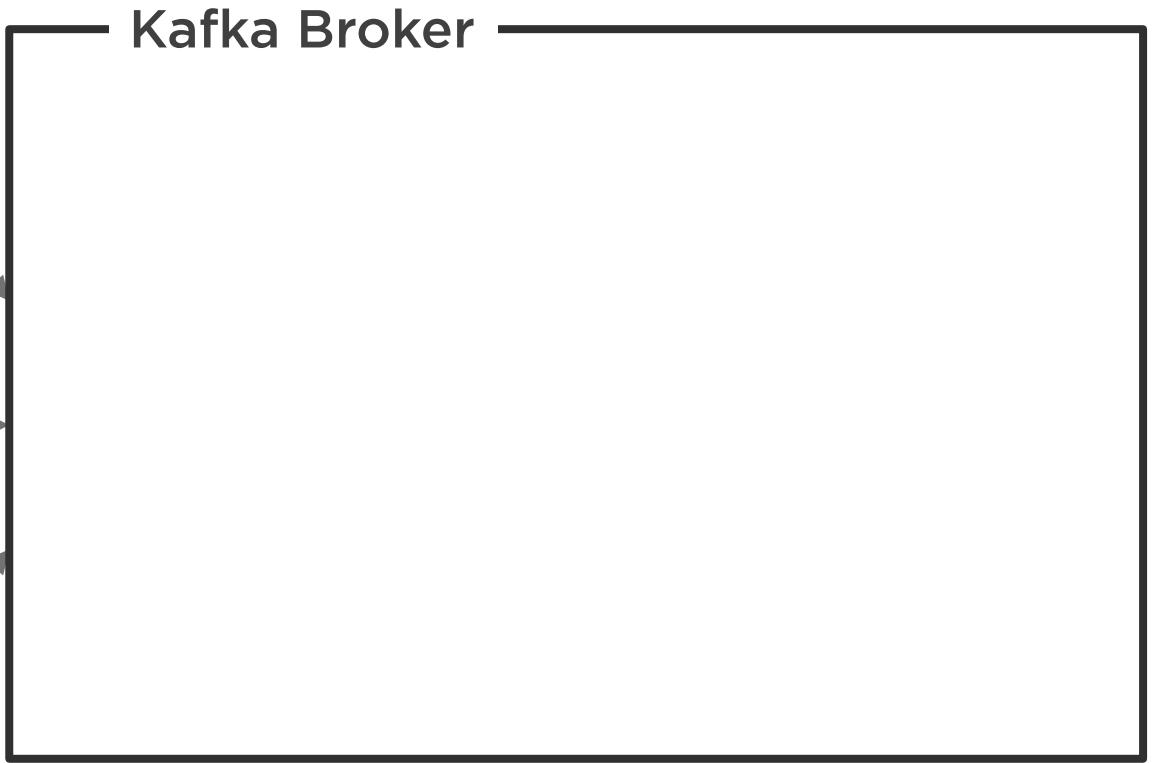
# Topics, Partitions, Logs

---

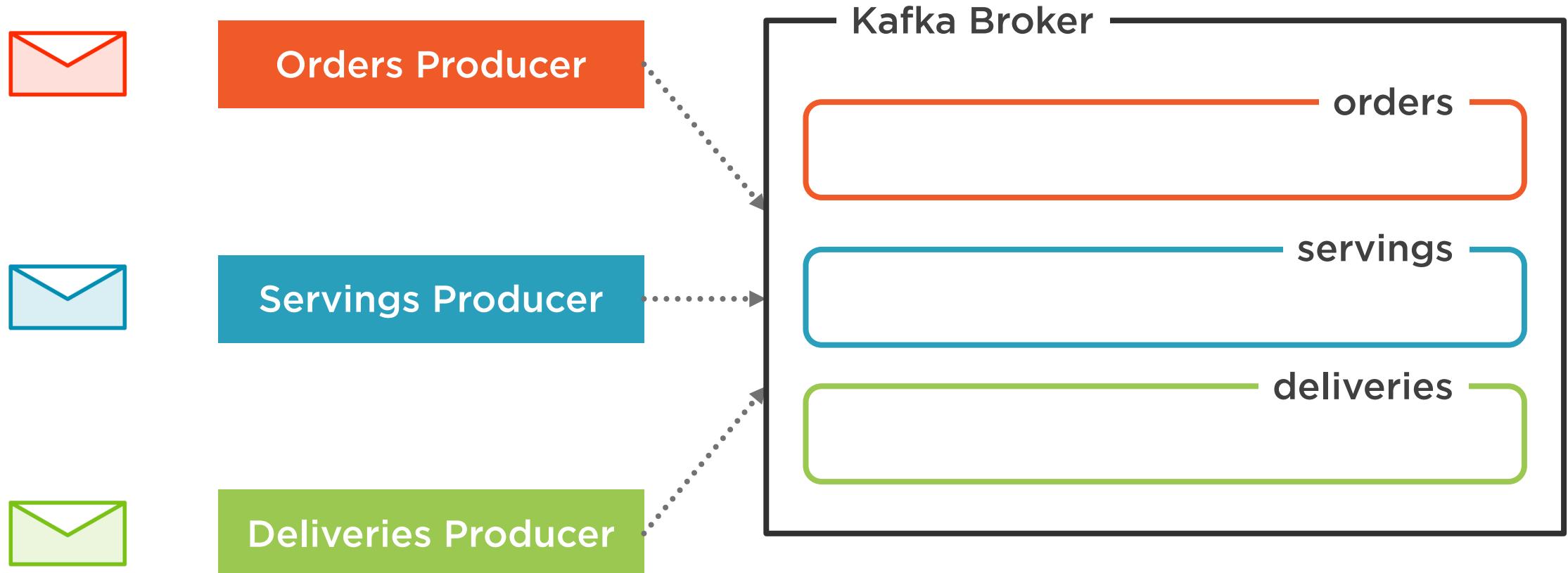




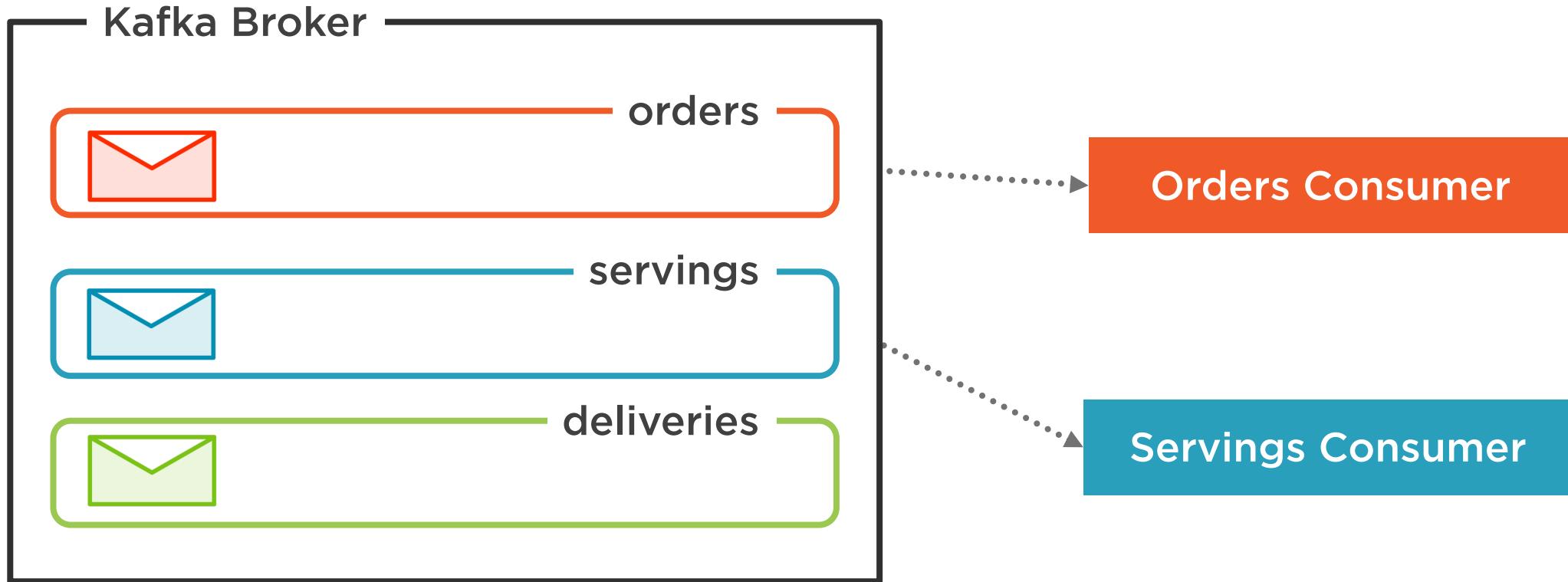
# Topics



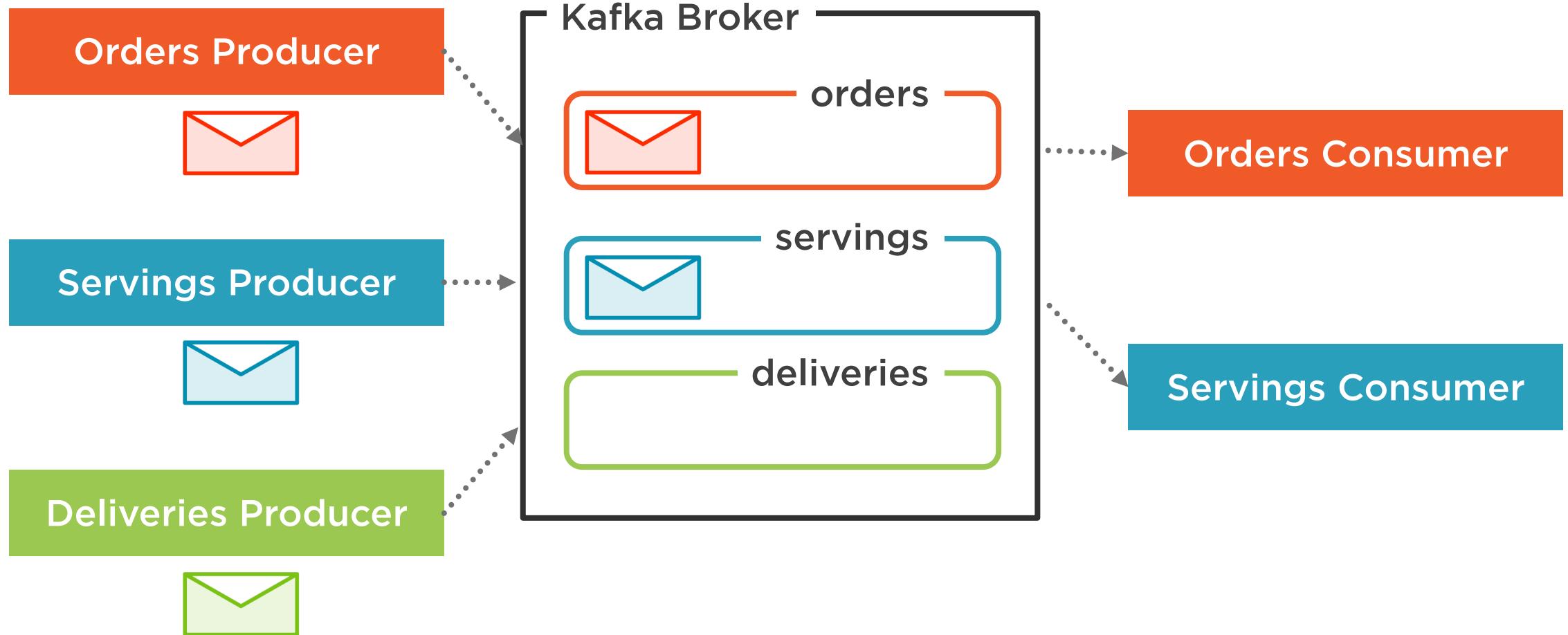
# Topics



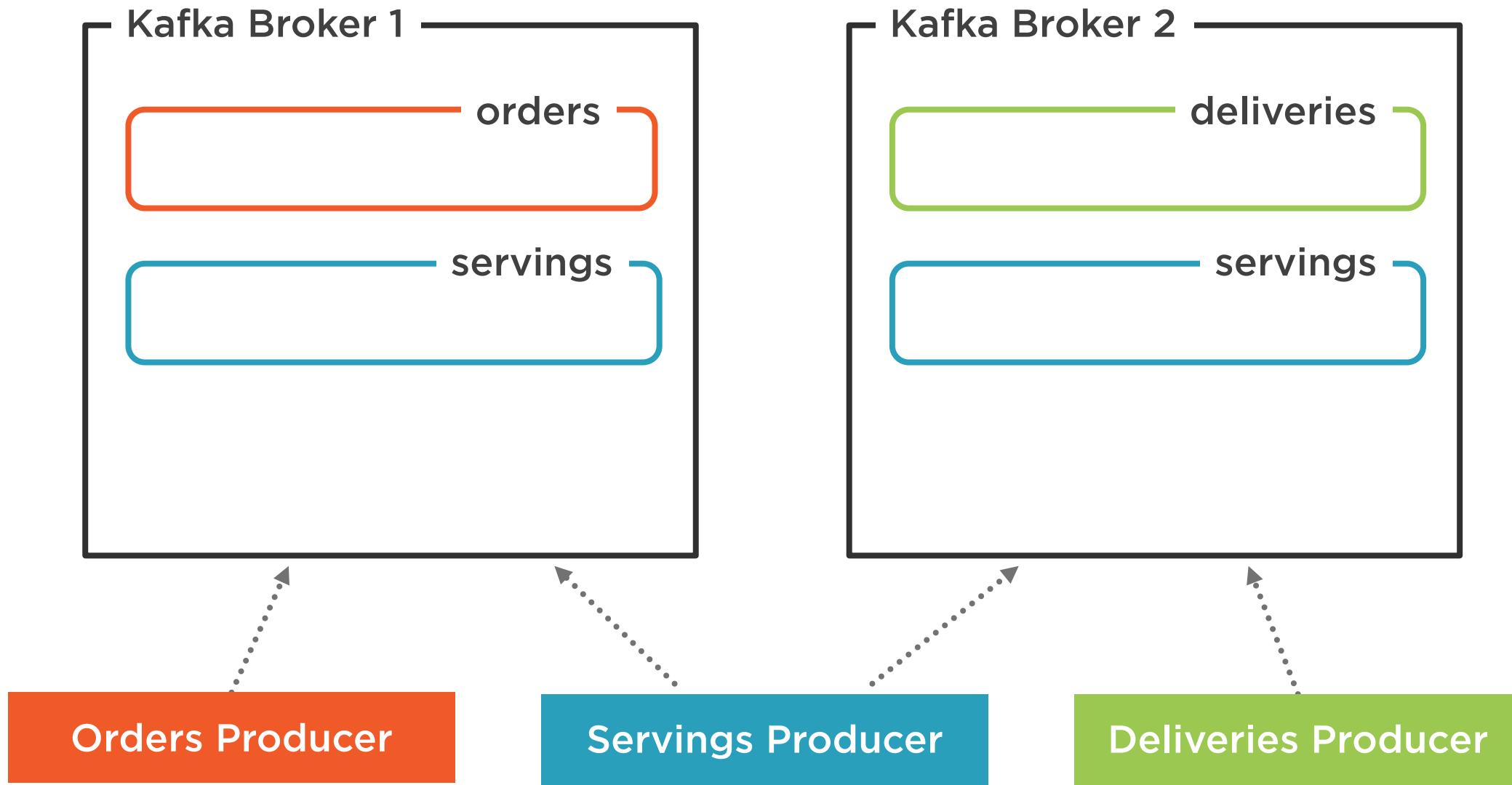
# Topics



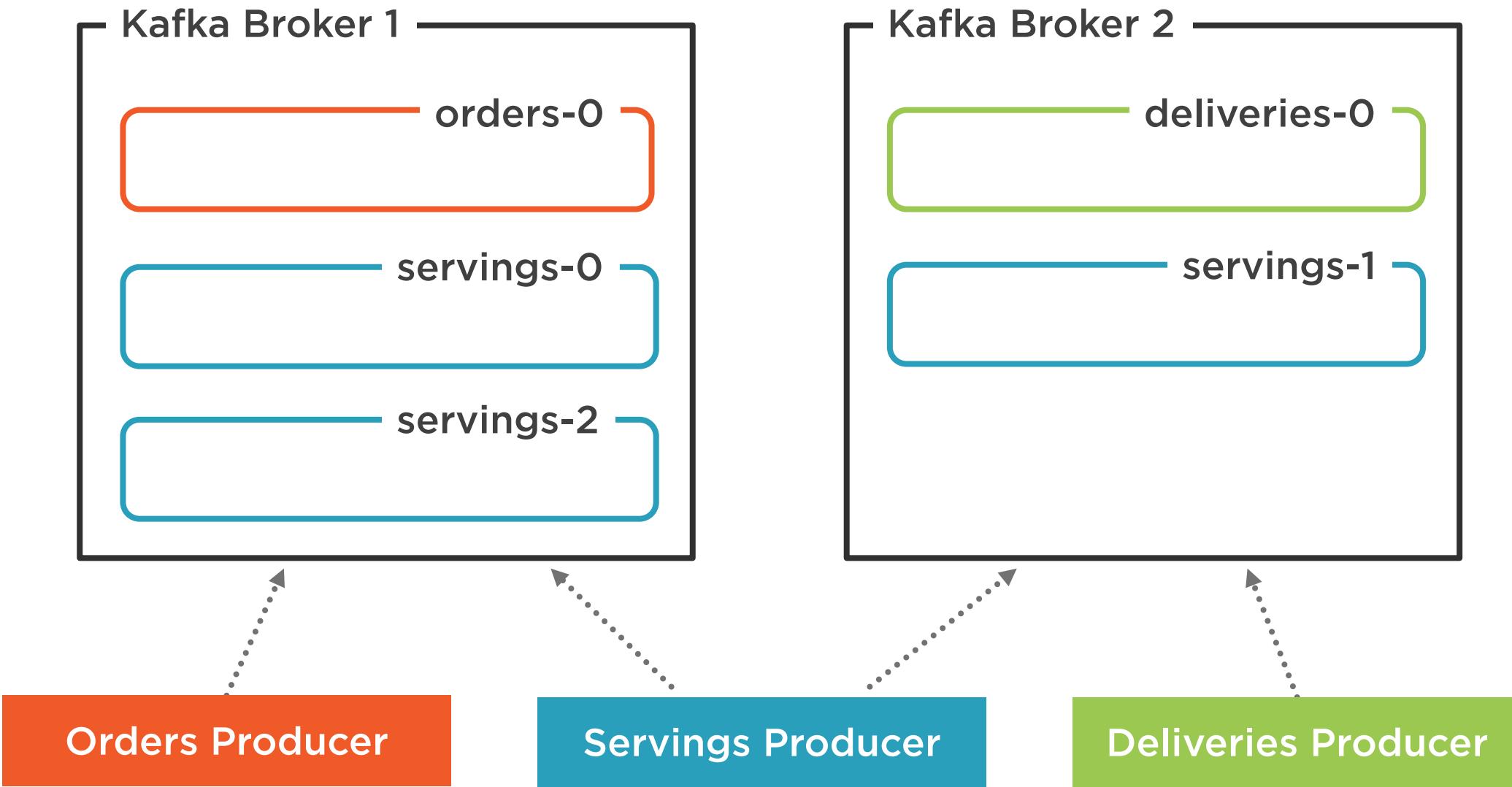
# Topics



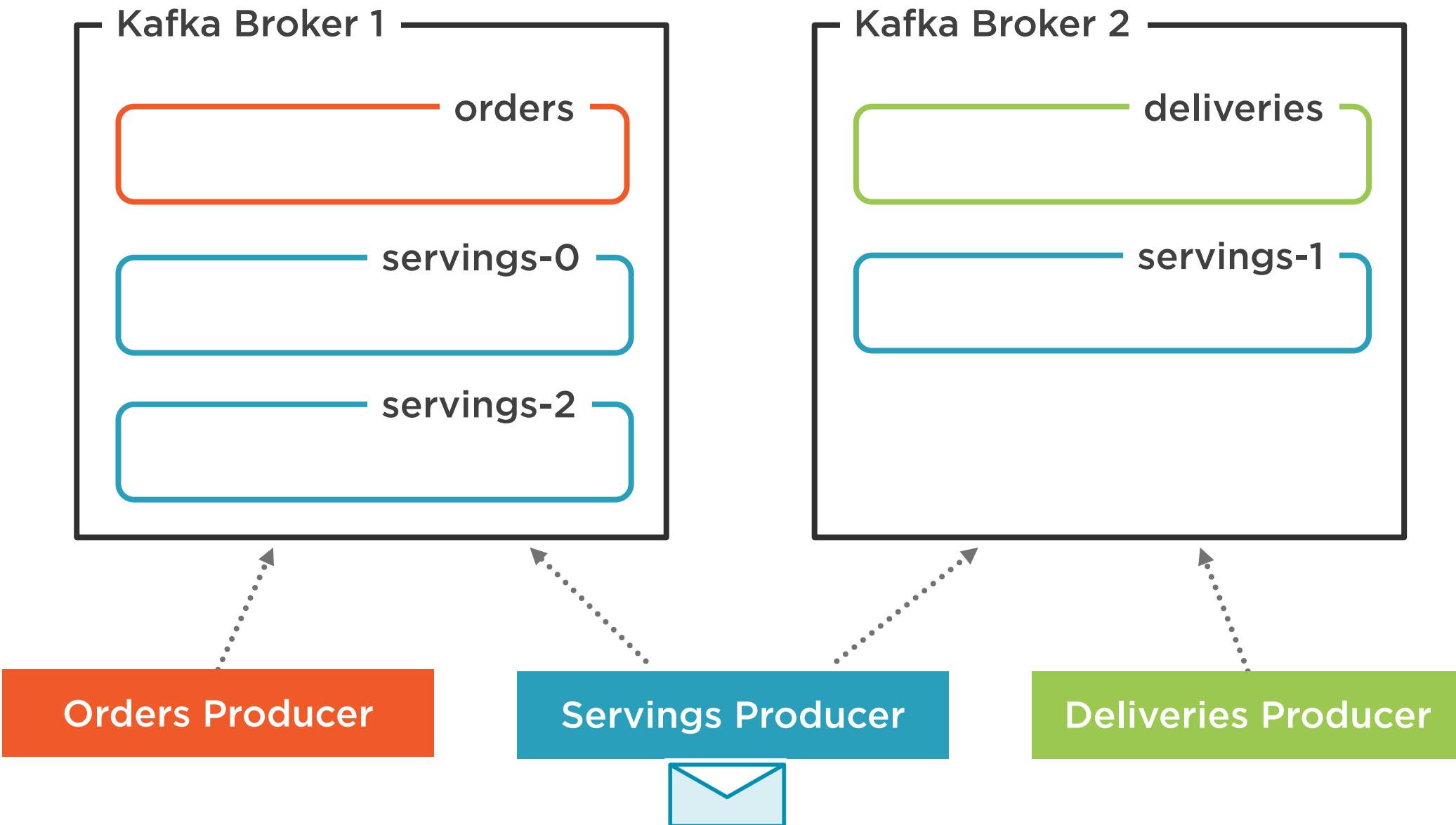
# Topics



# Partitions



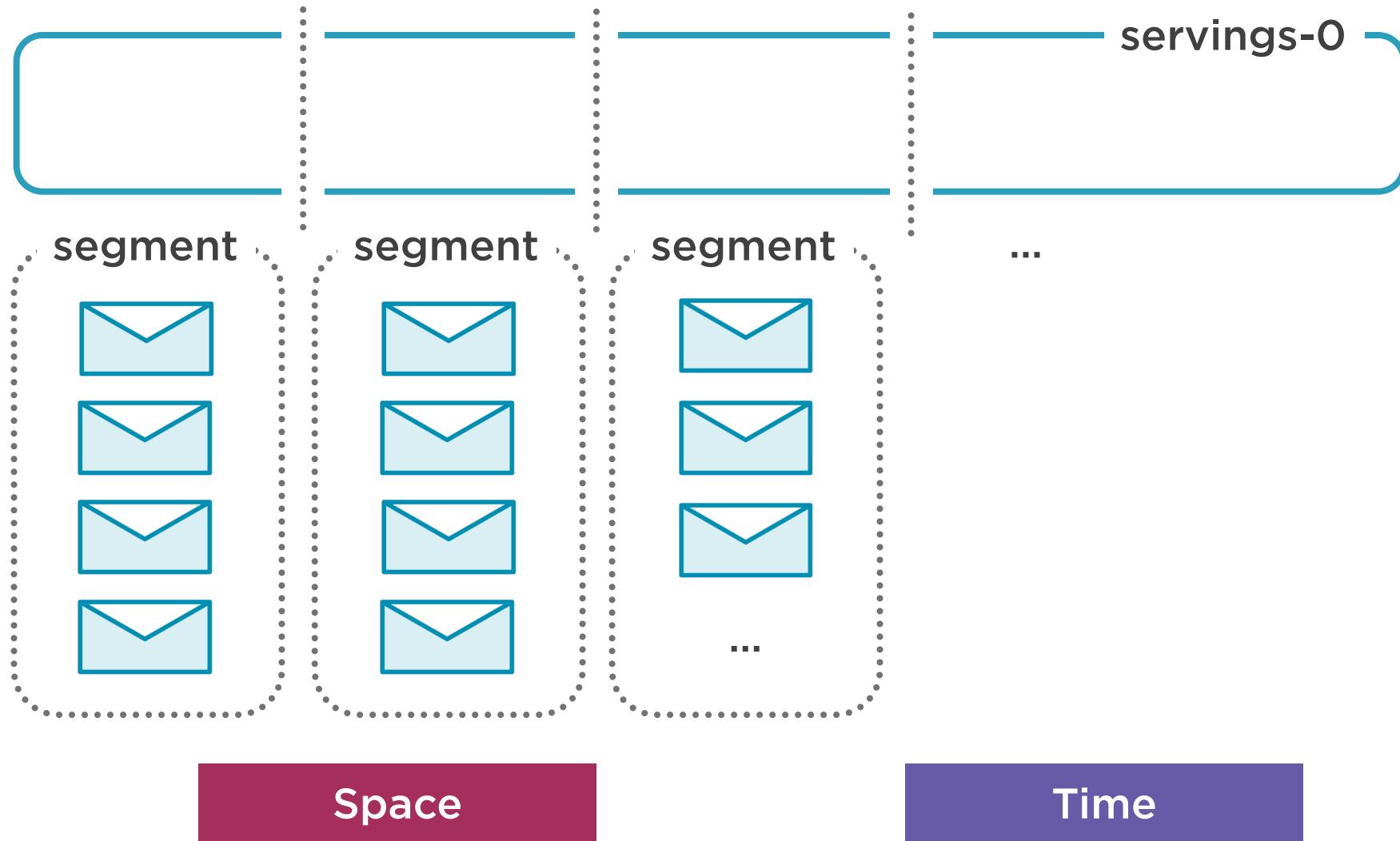
# Partitions



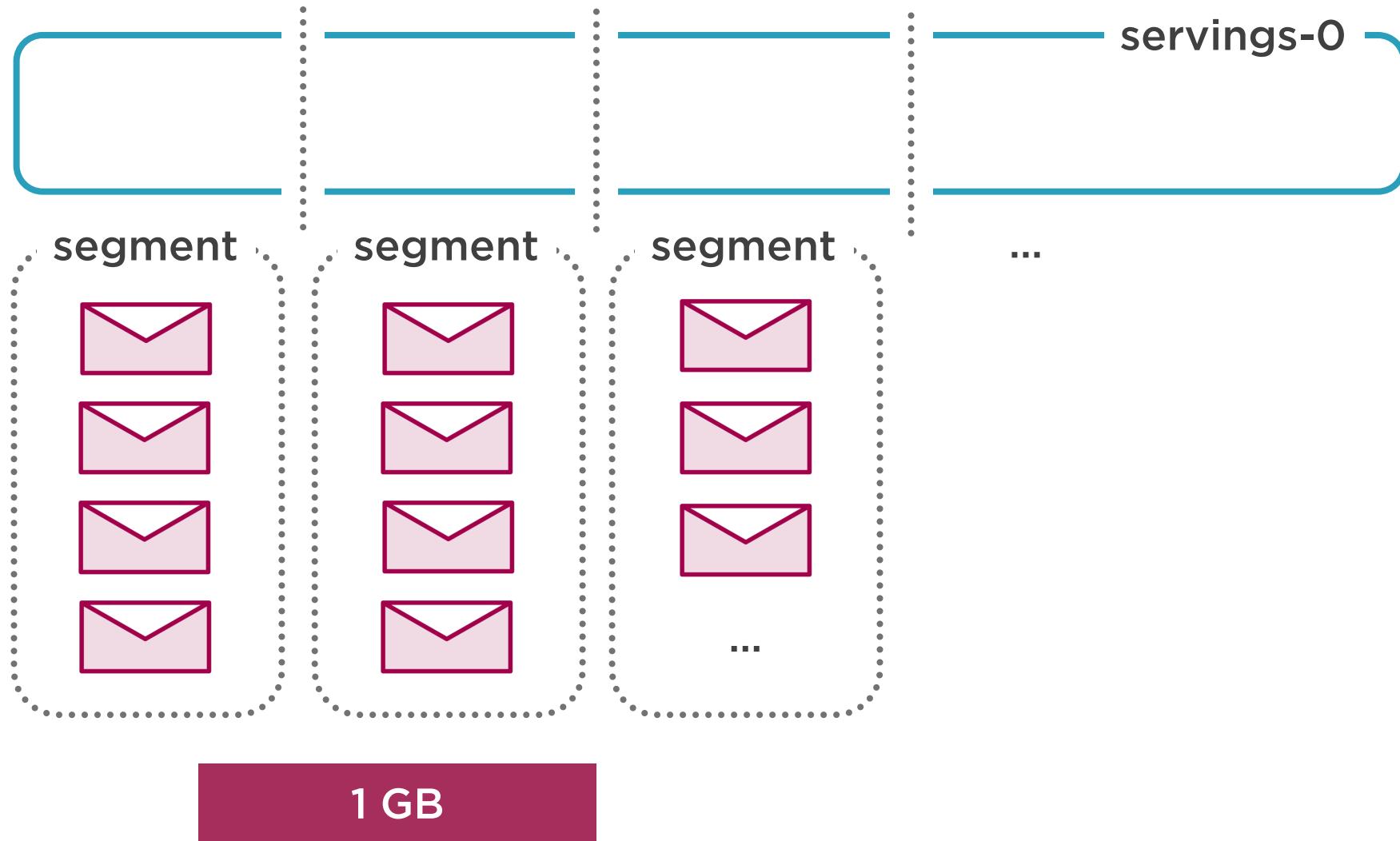
# Log



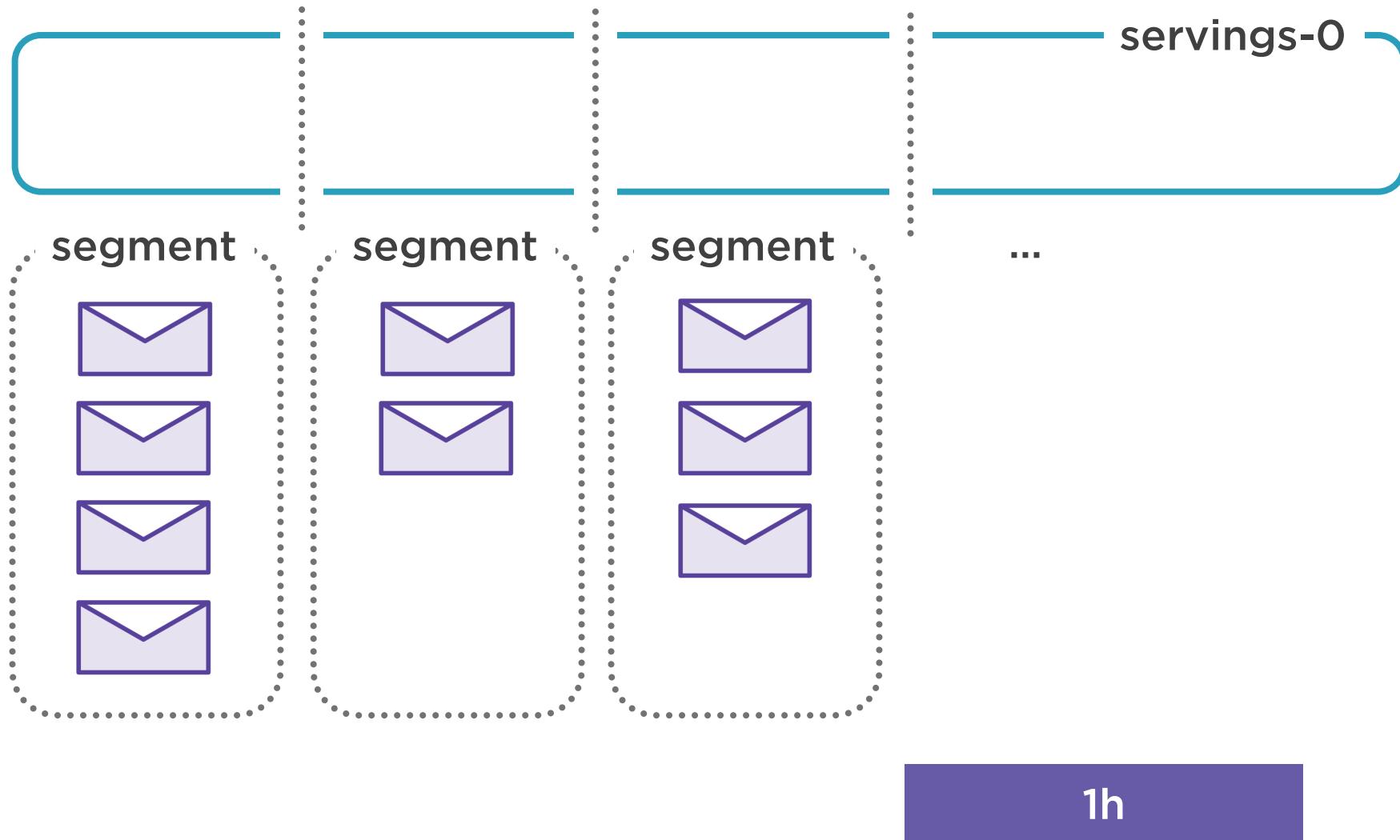
# Log



# Log



# Log



# Data Streams

---



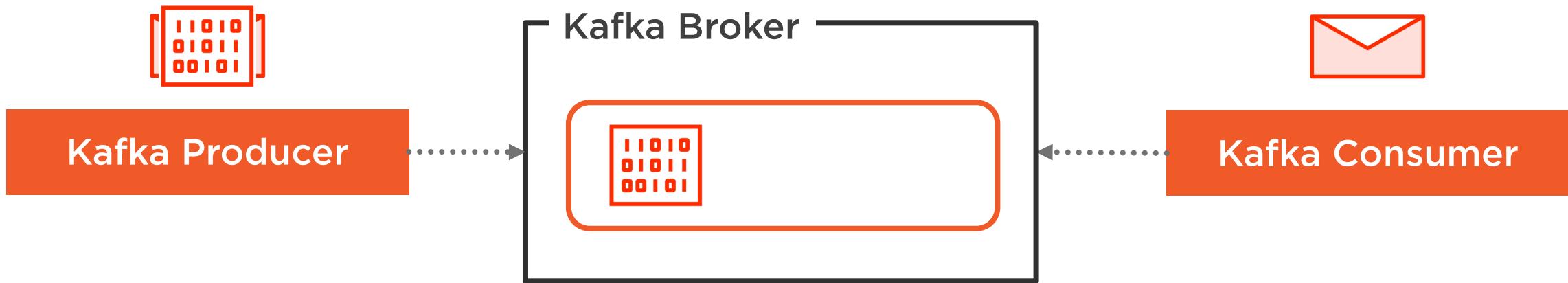
# Data Stream $\neq$ Kafka Topic



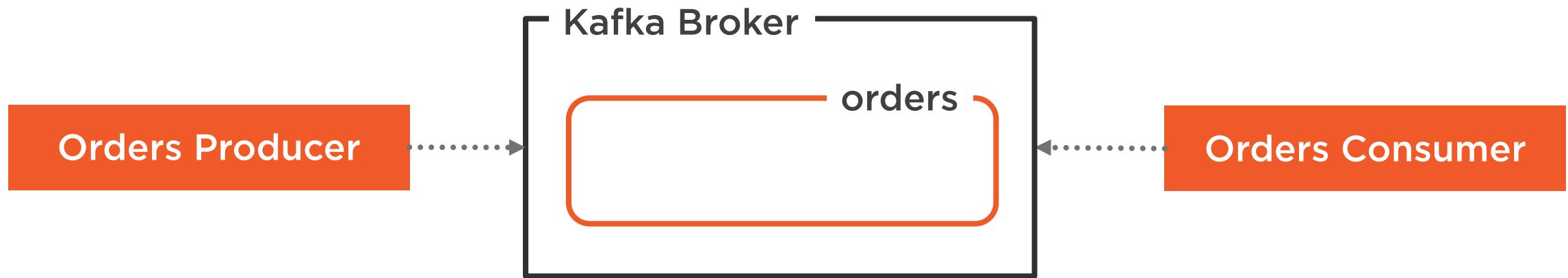
Data Stream =  
Kafka Topic + Data Governance



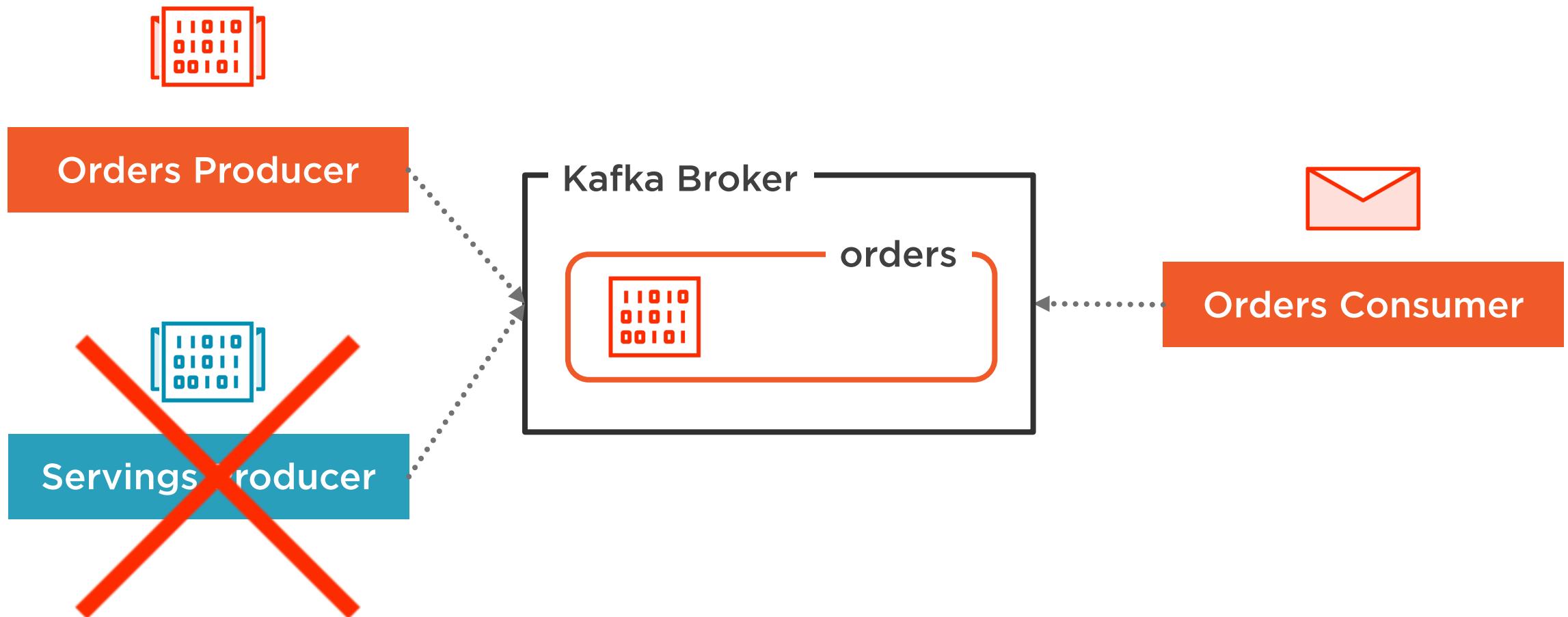
# Topics



# Topics



# Topics

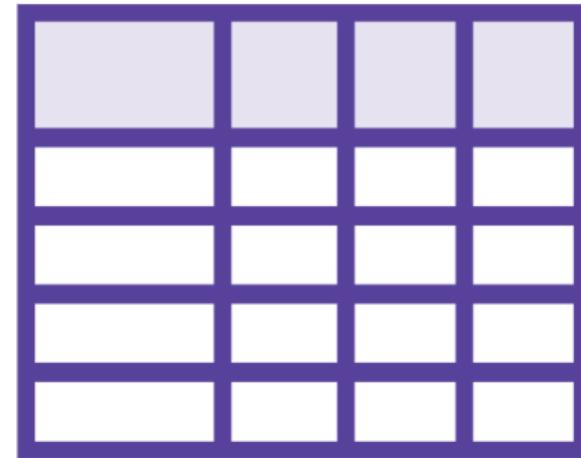


# Clean-up Policy



## Delete ( Stream )

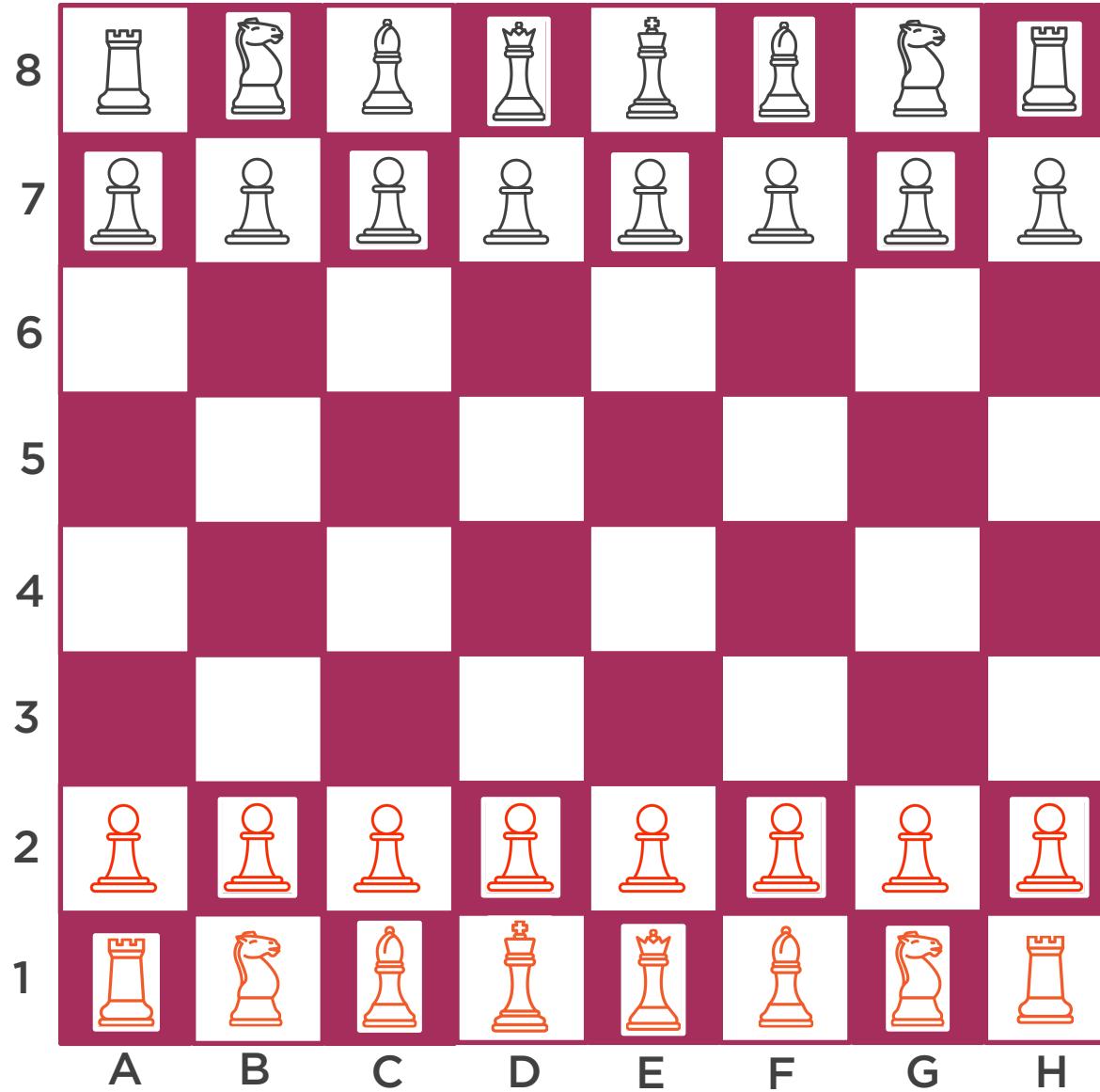
Delete records based on time or size factors



## Compact ( Table )

Similar to UPSERT functionality in a Database

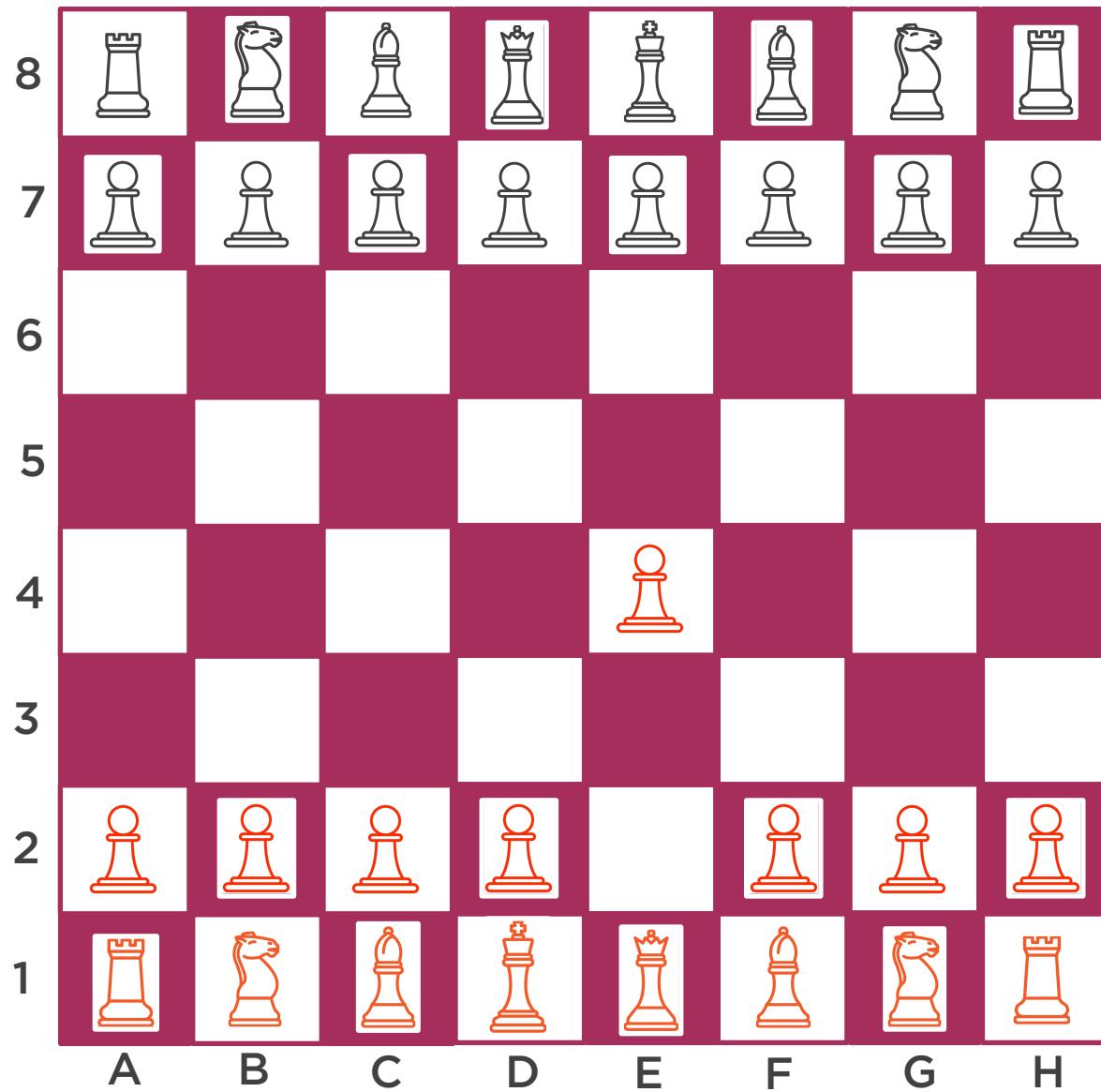




Stream ( delete )

Table ( compact )





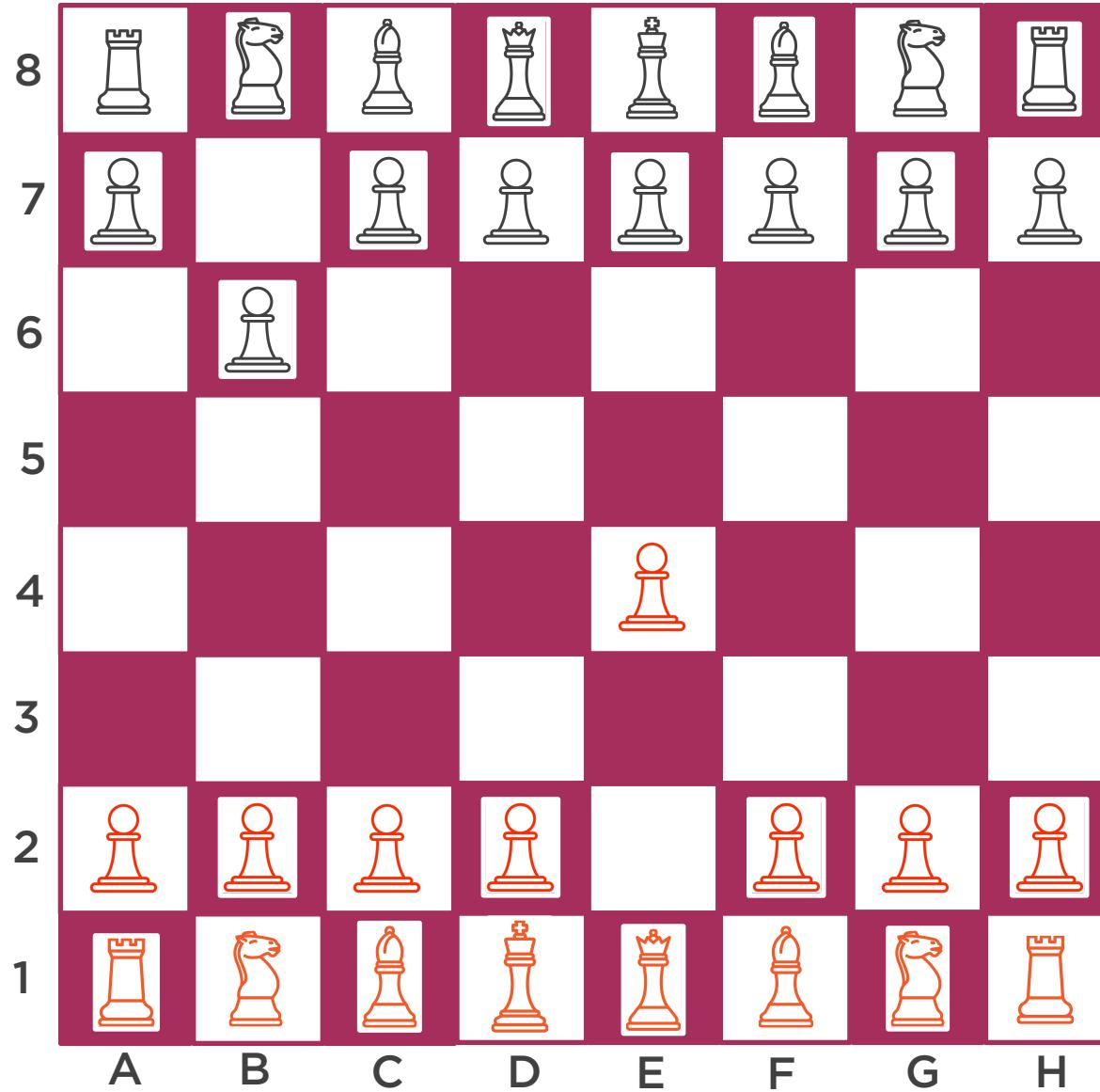
Stream ( delete )

e4

Table ( compact )

e4





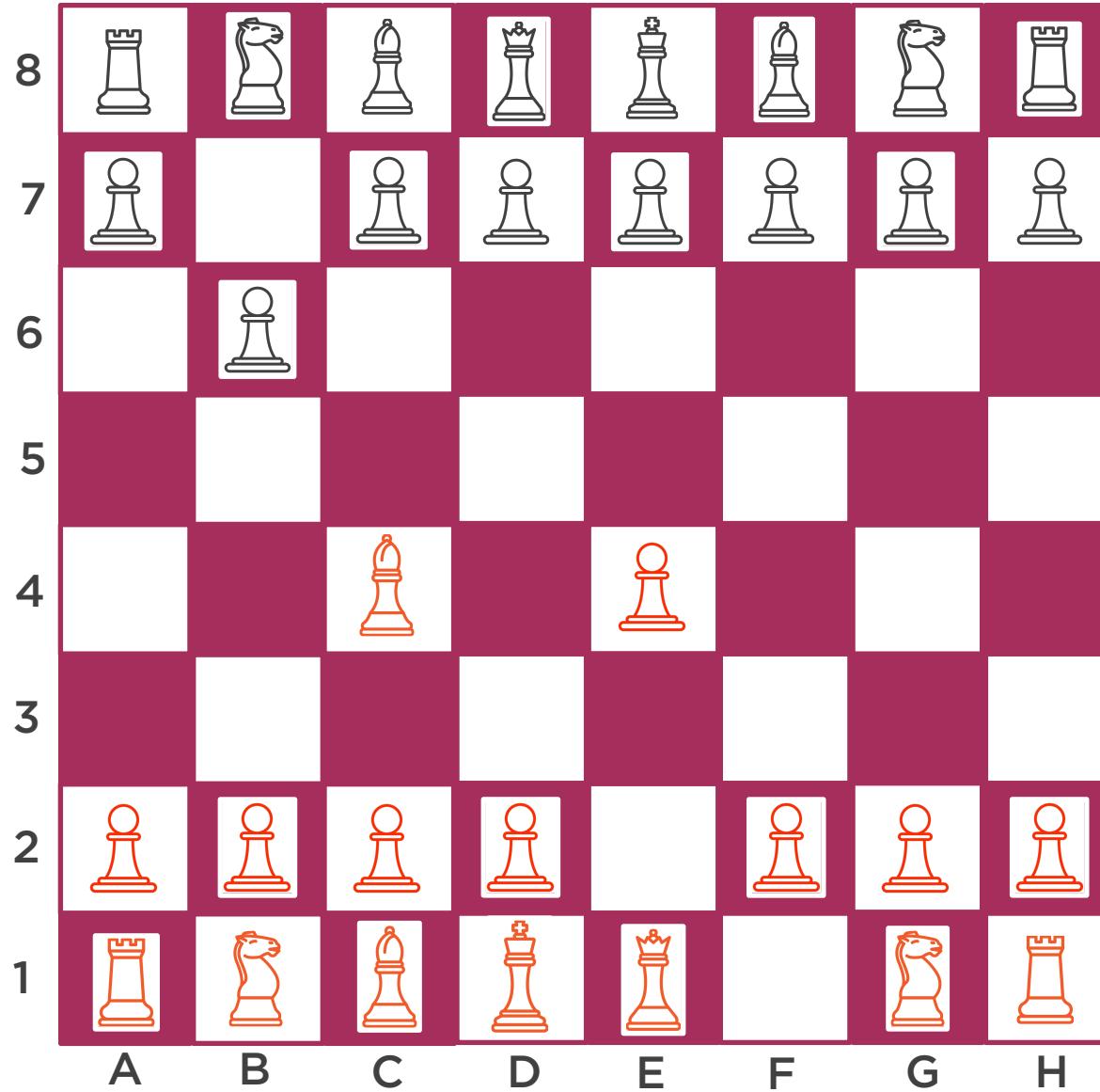
Stream ( delete )

e4 b6

Table ( compact )

e4 b6





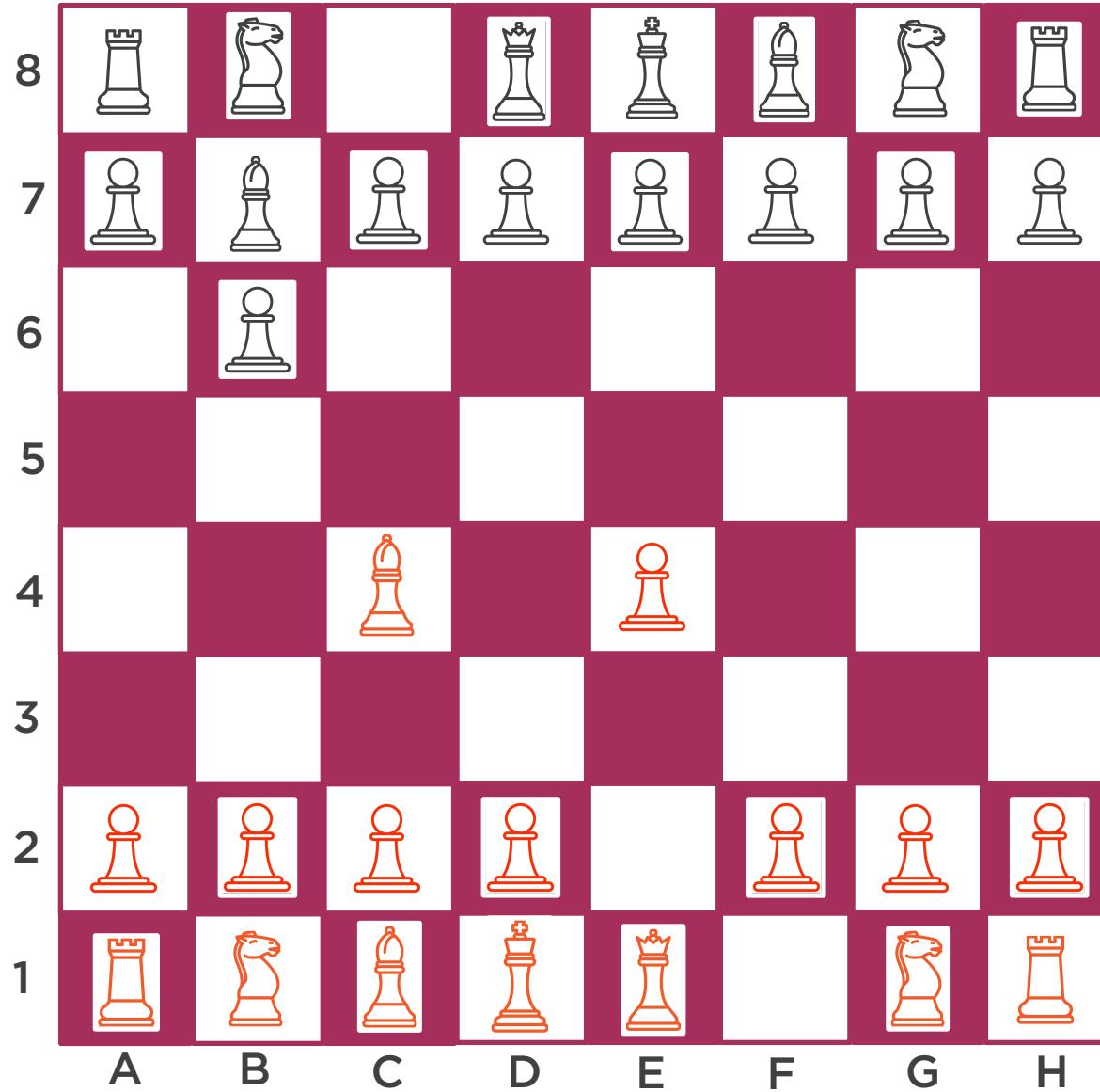
Stream ( delete )

e4 b6 Bc4

Table ( compact )

e4 b6 Bc4





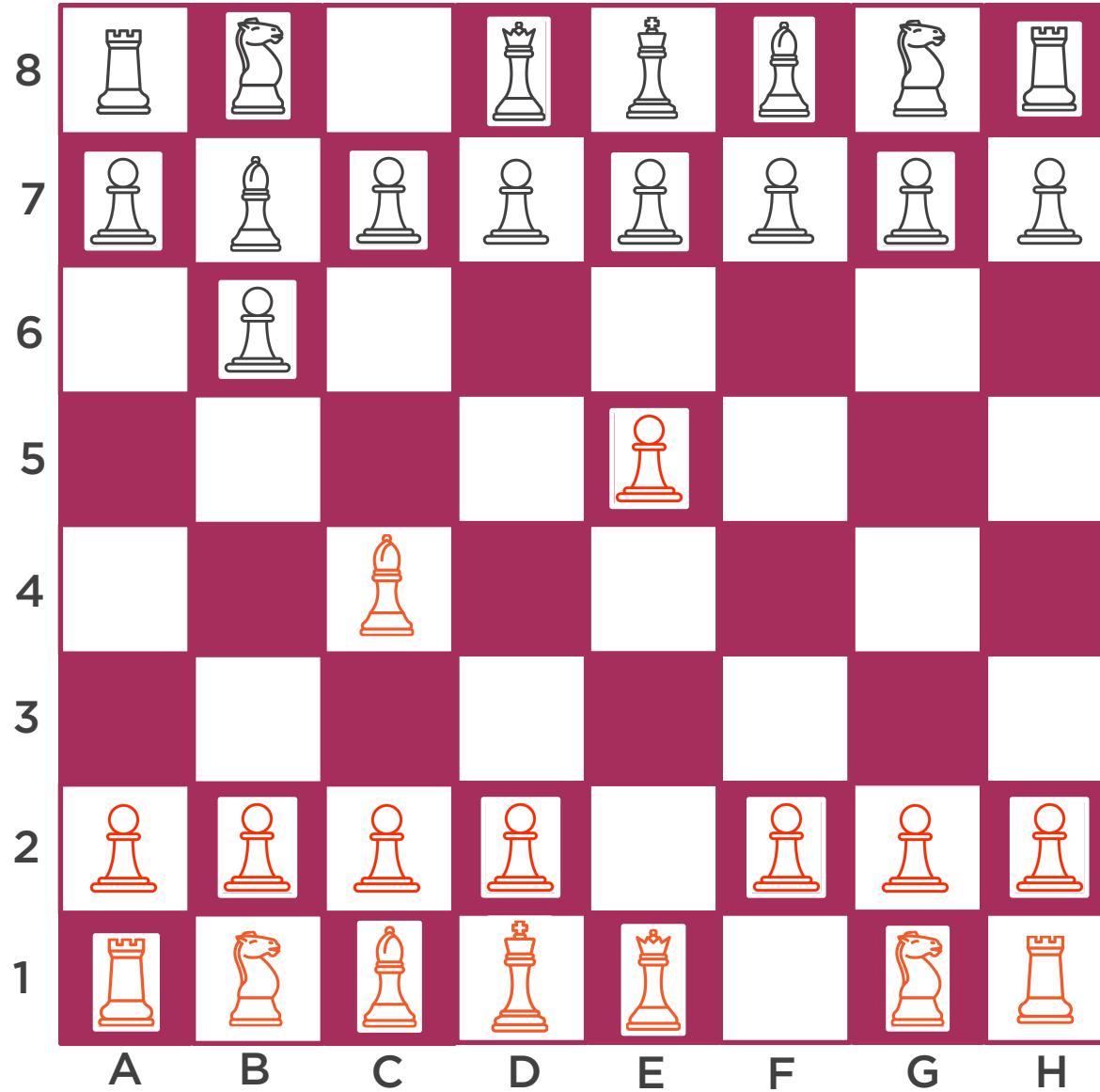
Stream ( delete )

e4 b6 Bc4 Bb7

Table ( compact )

e4 b6 Bc4 Bb7





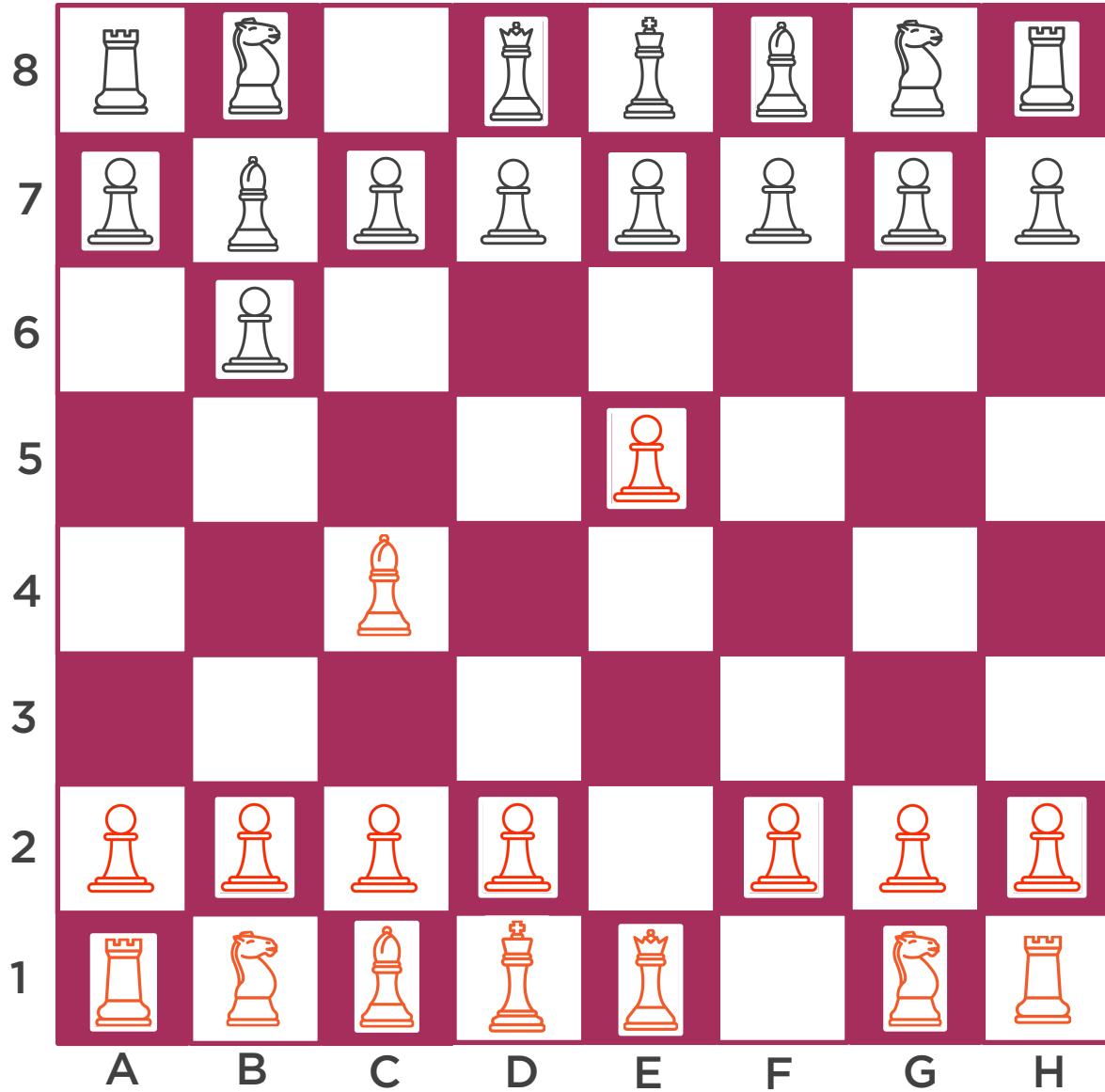
Stream ( delete )

e4 b6 Bc4 Bb7 e5

Table ( compact )

e4 b6 Bc4 Bb7 e5





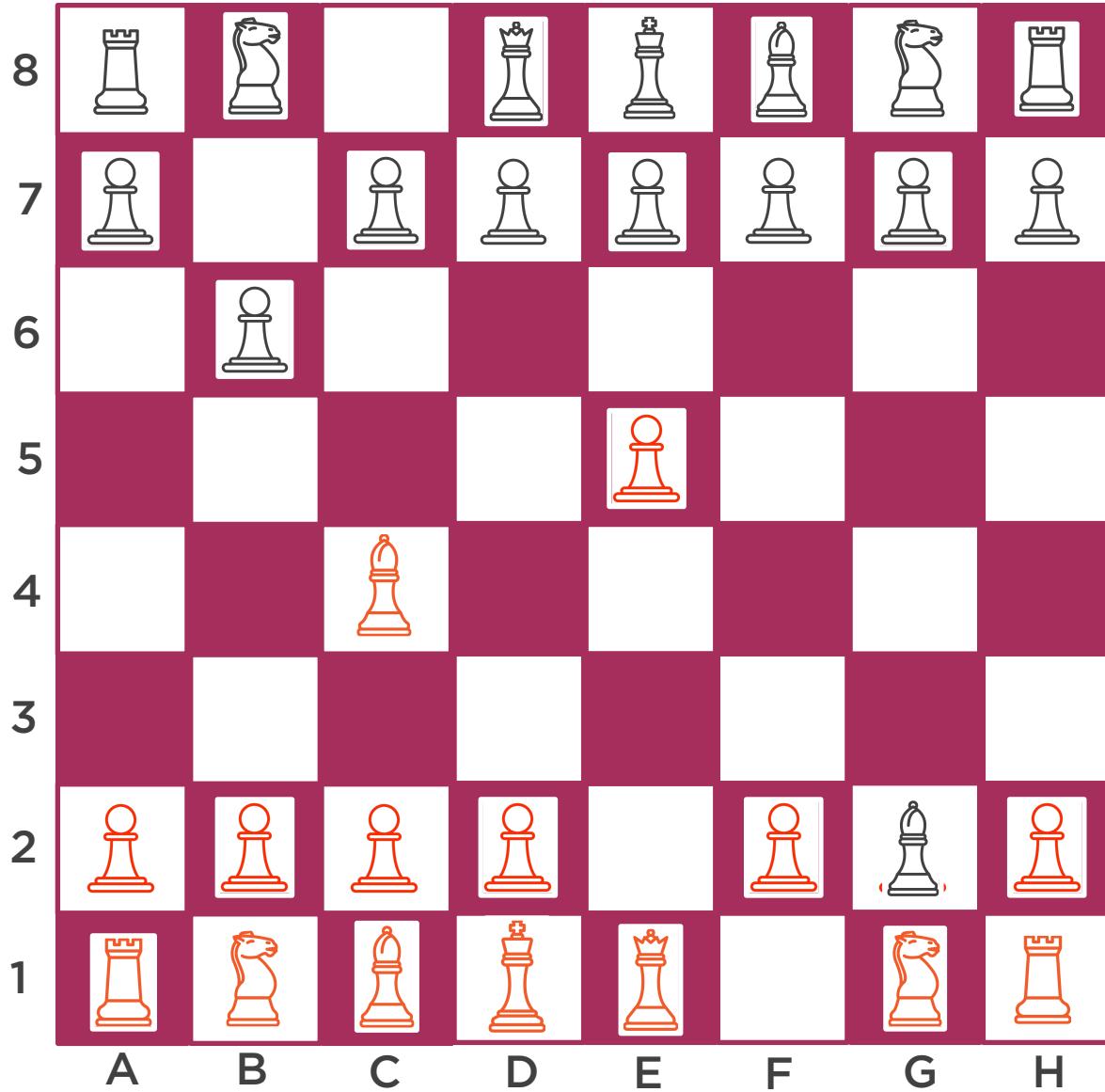
Stream ( delete )

e4 b6 Bc4 Bb7 e5

Table ( compact )

b6 Bc4 Bb7 e5





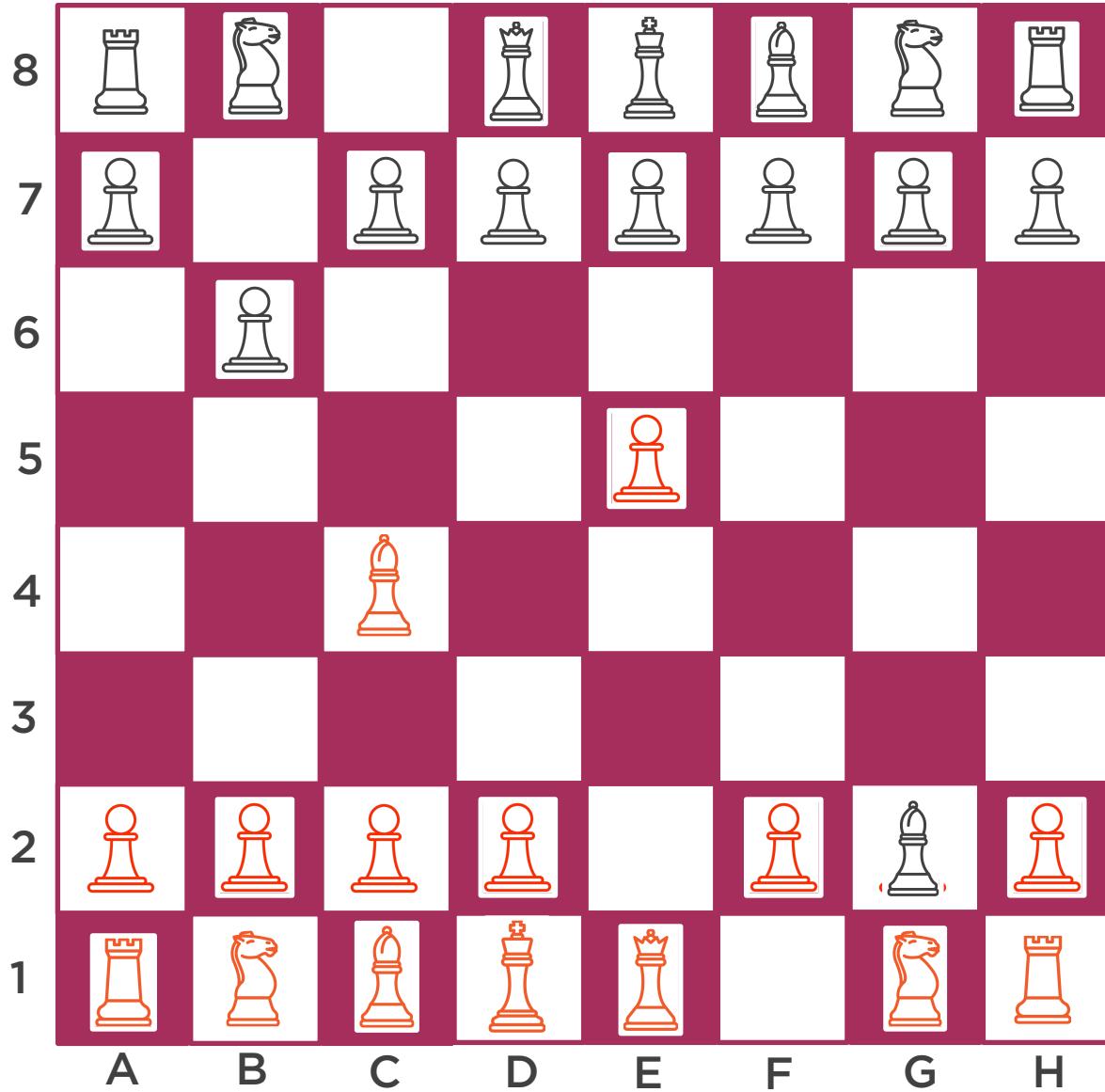
Stream ( delete )

e4 b6 Bc4 Bb7 e5 Bg2

Table ( compact )

b6 Bc4 Bb7 e5 Bg2





Stream ( delete )

e4 b6 Bc4 Bb7 e5 Bg2

Table ( compact )

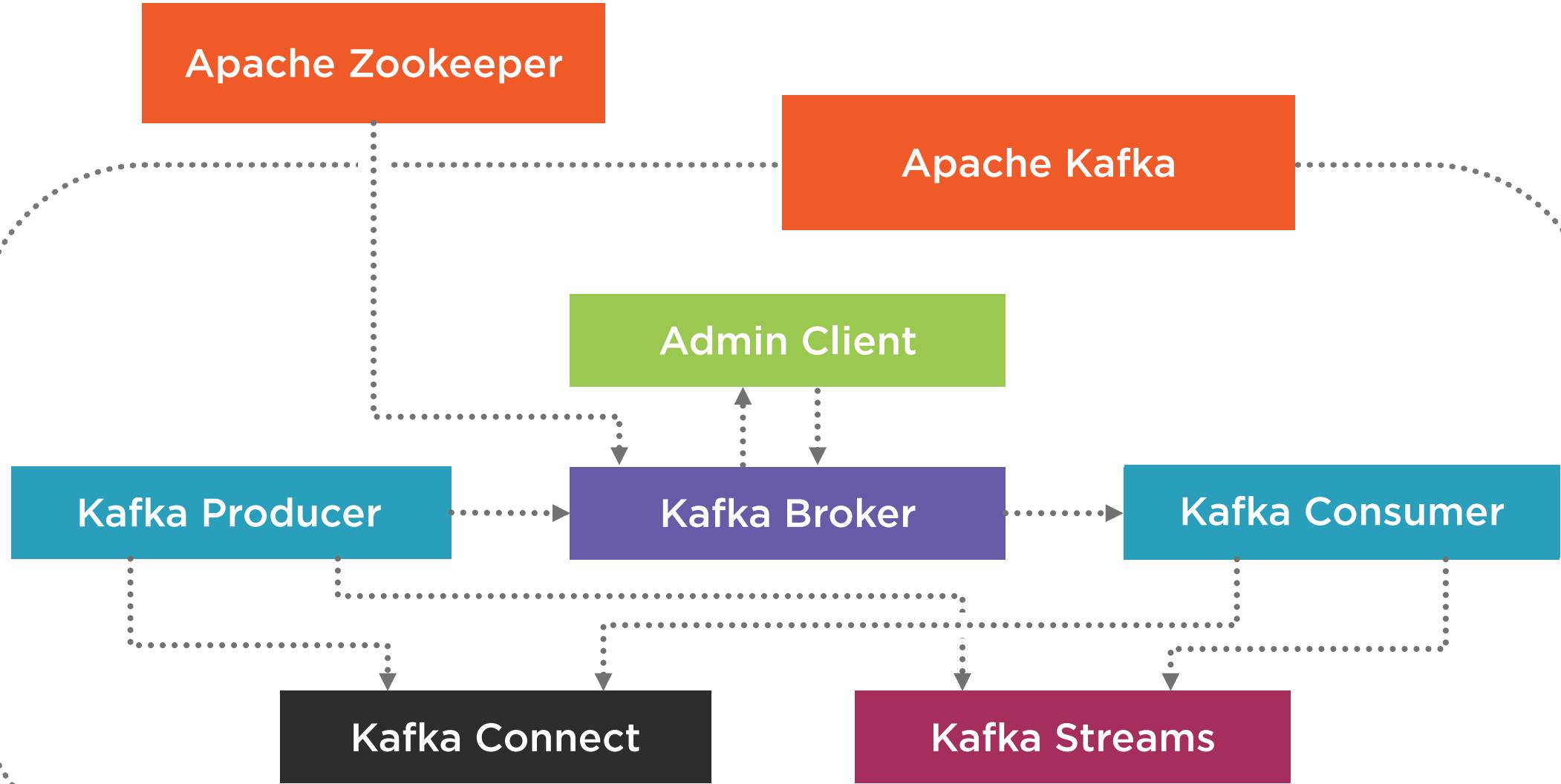
b6 Bc4 e5 Bg2



# Extending Apache Kafka

---





Apache Kafka is **NOT**  
enterprise-ready!



# Enterprise requirements ...

## Data Governance

- Creation and deletion of topics
- Enforcing data types
- Dealing with encrypted data

## Integration

## Testing

## Availability

## Security



# Extending Apache Kafka

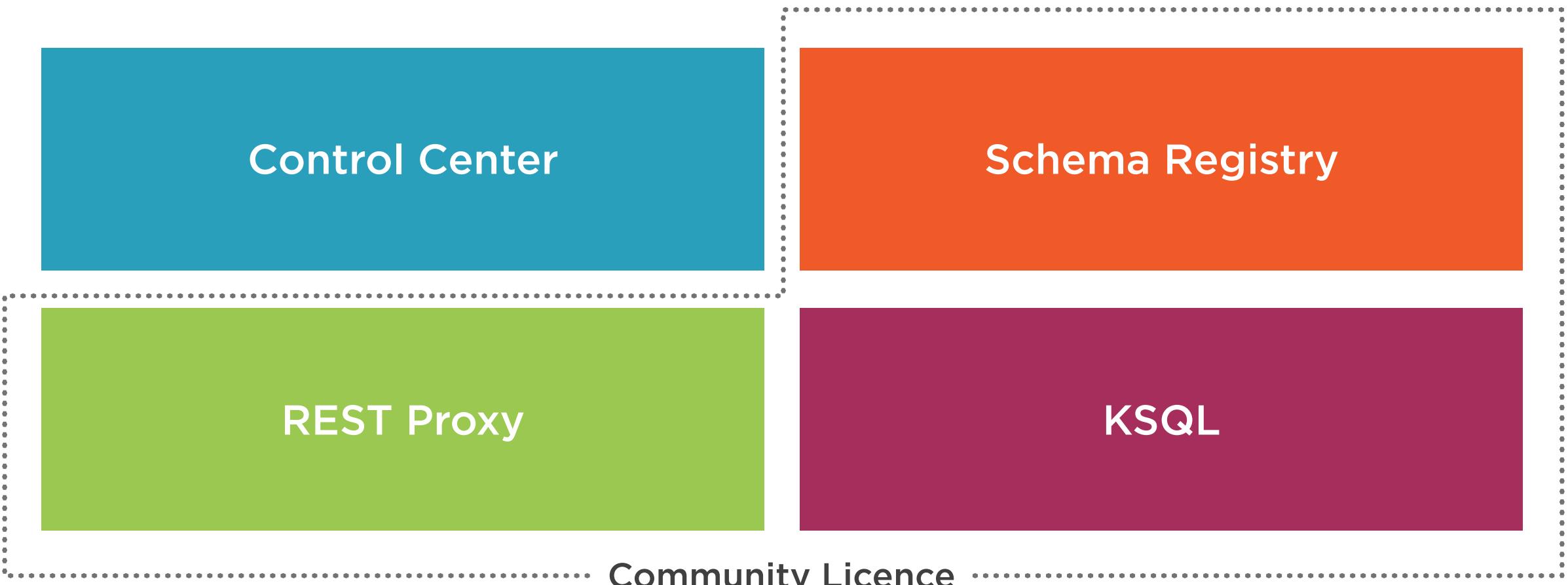
Open-Source /  
Community

Enterprise /  
Commercial

<https://cwiki.apache.org/confluence/display/KAFKA/Ecosystem>



# Confluent Platform



# Cloud Subscriptions



**Confluent Cloud**

**AWS MSK**

**IBM Event Streams**

**Aiven**

**Heroku**

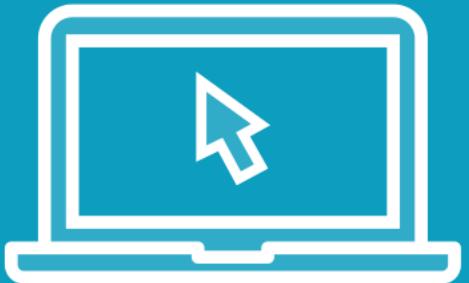
**CloudKarafka**

**Instaclustr**

**Etc.**



# Demo



**Install Apache Kafka**

**Install Confluent Platform**

**Prerequisites**

- Docker
- 8 GB RAM, 4x CPU



# Summary



**Apache Kafka Architecture**

**Records, Topics**

**Making Apache Kafka Enterprise Ready**

**Data Streams**

