애플리케이션 개발자 입장에서 살펴본 분산 이벤트 스트리밍 플랫폼

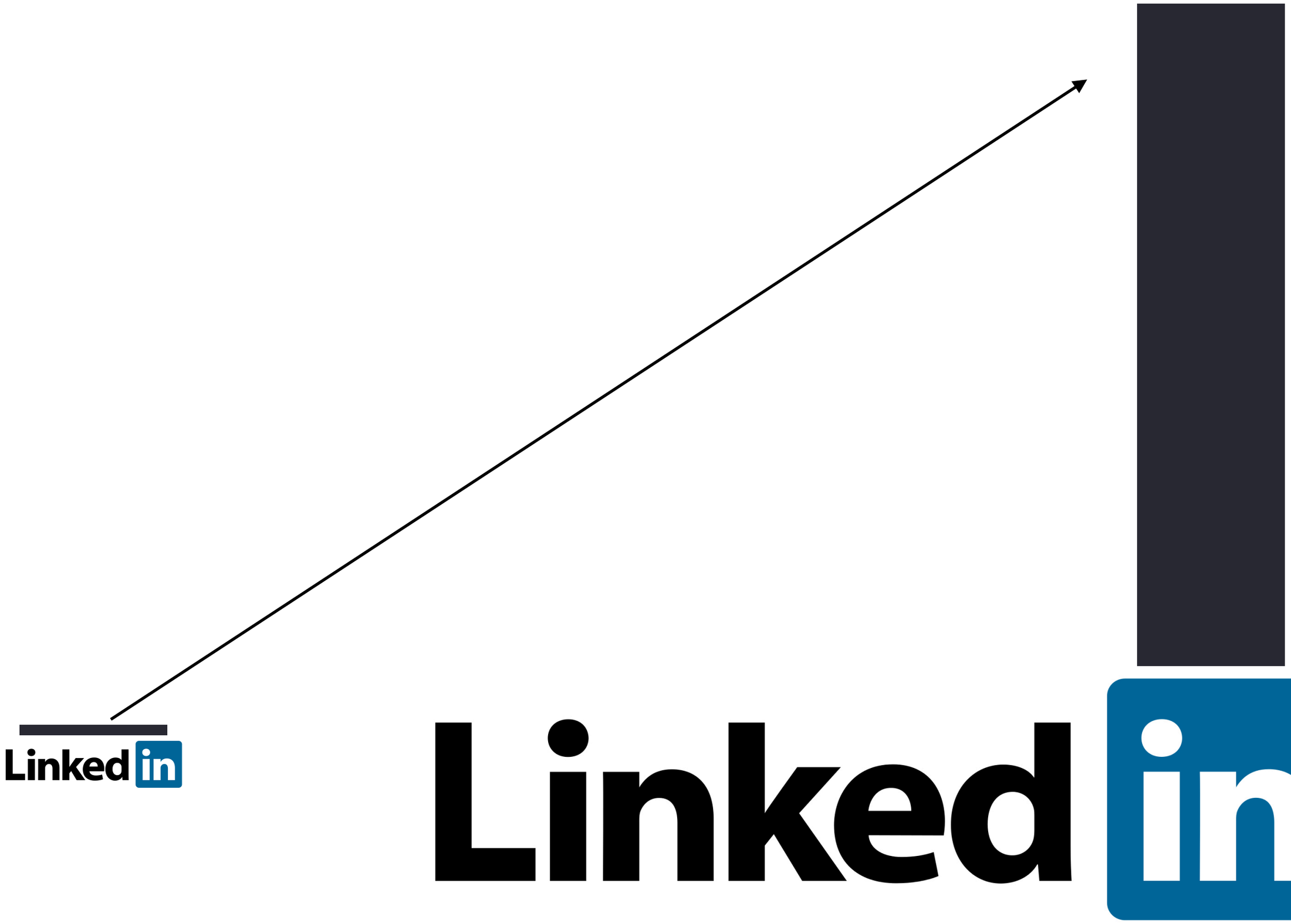
2019.04.03 @findstar

Kafka 란?

출생지

작은 규모

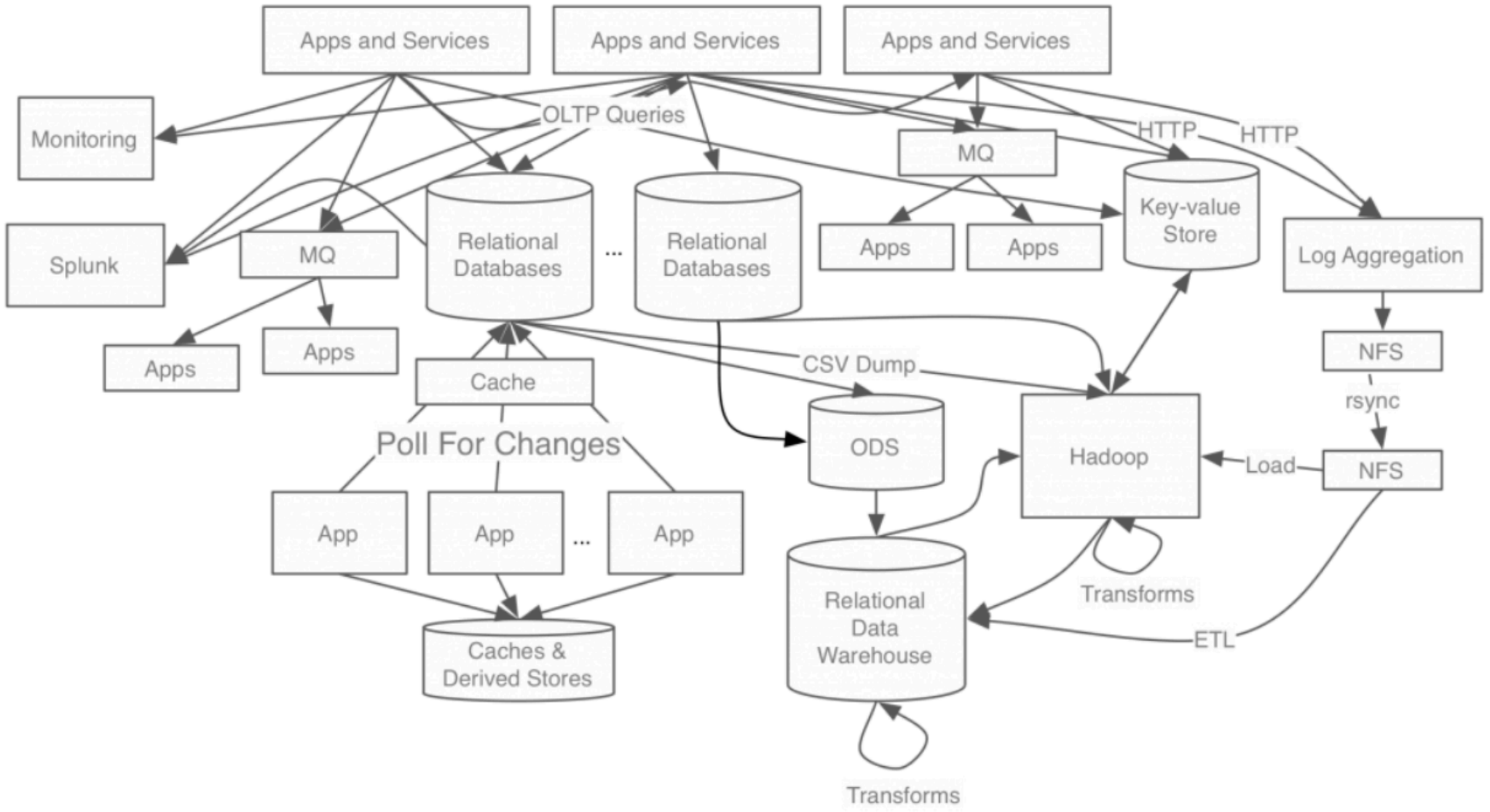
큰 규모



급격한 성장

부럽.. 우리도 이렇게 돈 벌었으면…

시스템 상황

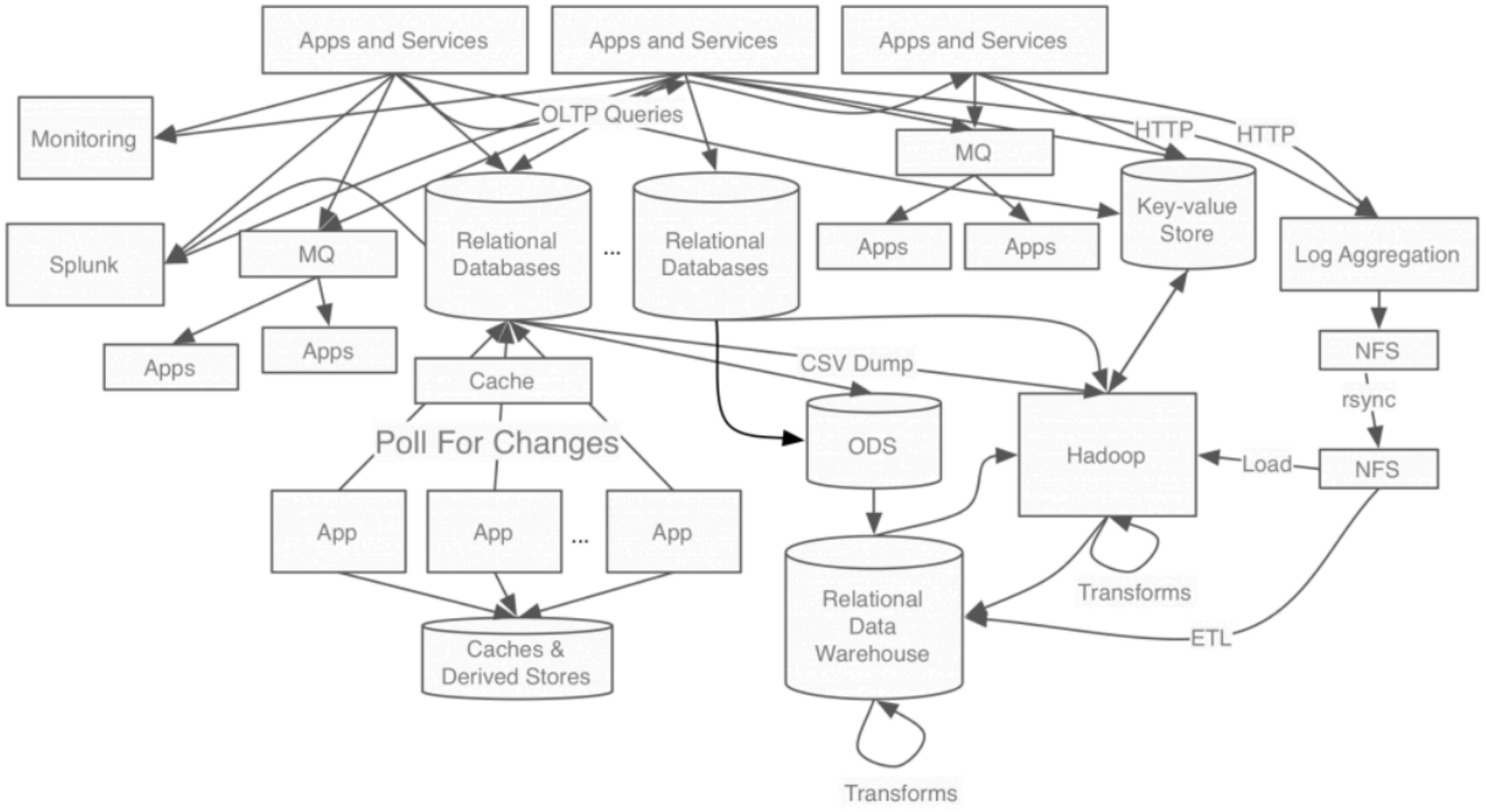


LinkedIn

시스템 상황

LinkedIn

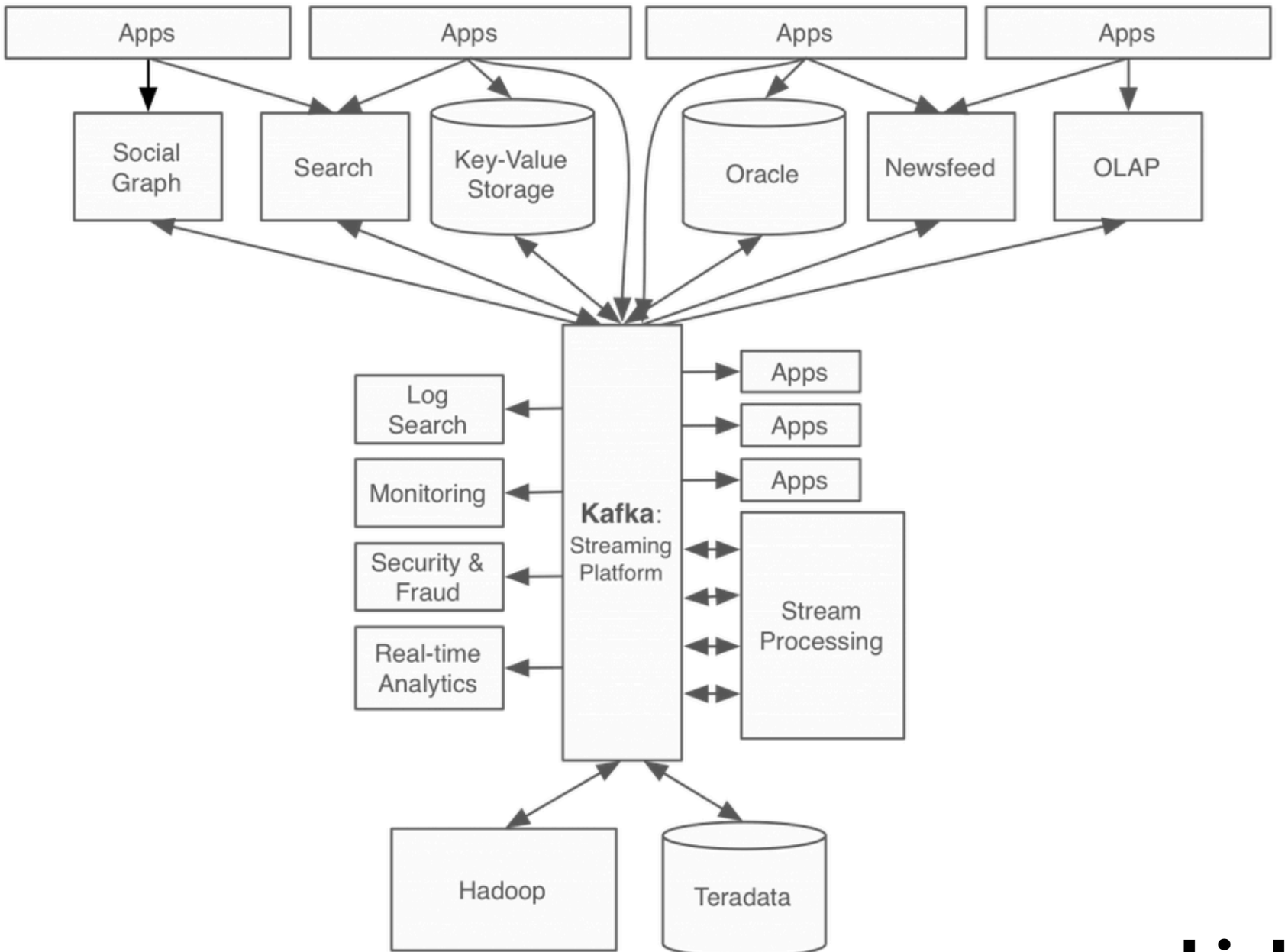
그래서..



From

LinkedIn

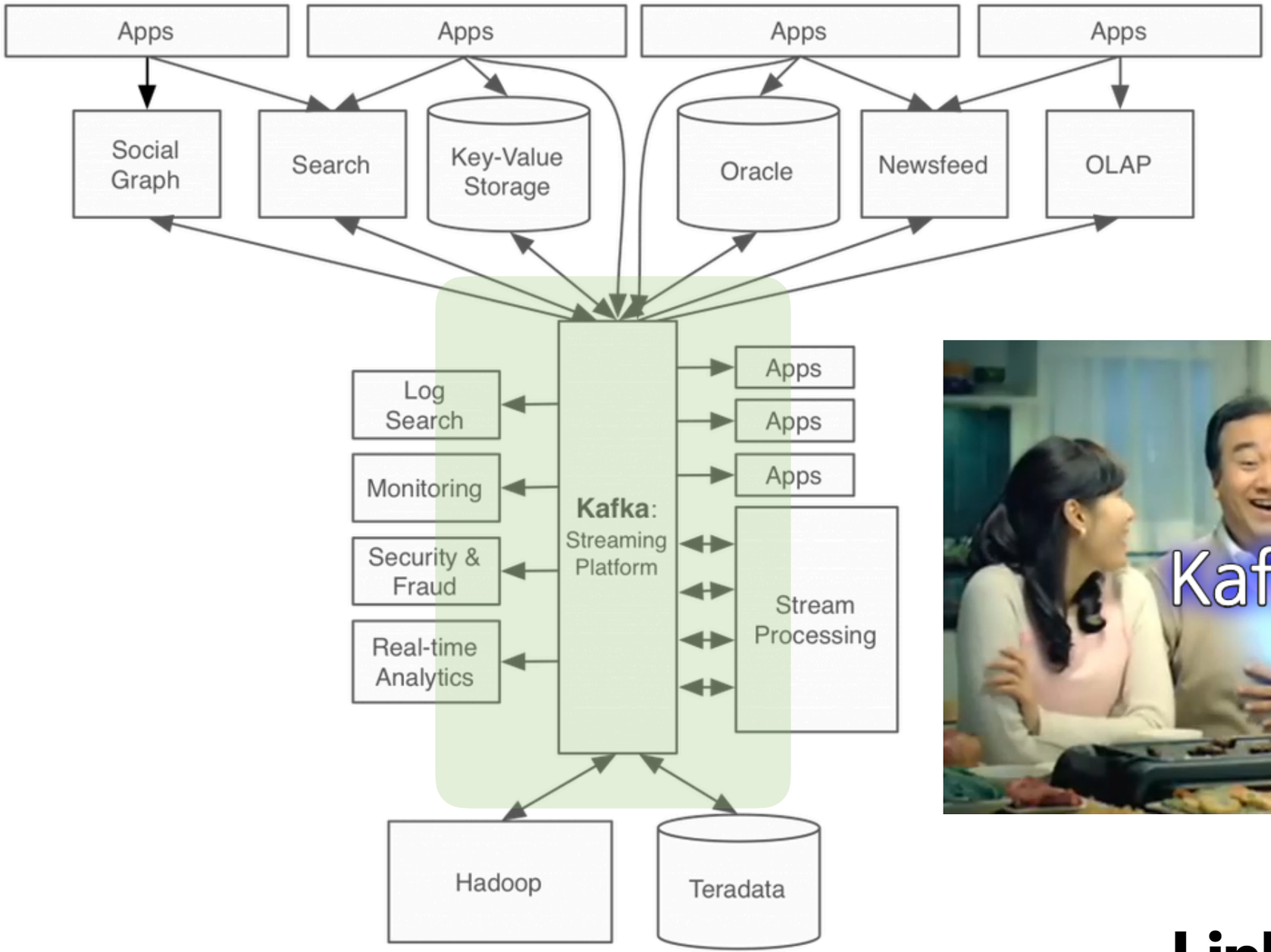
그래서..



To

LinkedIn

개선


Apps
Apps
Apps
Apps
Social Graph
Search
Key-Value Storage
Oracle
Newsfeed
OLAP
Log Search
Monitoring
Security & Fraud
Real-time Analytics
Kafka: Streaming Platform
Apps
Apps
Apps
Stream Processing
Hadoop
Teradata

Kafka

LinkedIn

Confluent

Jay Kreps
Co-founder & CEO

Neha Narkhede
Co-founder & CTO

Jun Rao
Co-founder

예전엔 linkedin 직원 지금은 C레벨..

https://www.confluent.io/about/#about_confluent

kafka

빠른

대용량

분산

메세지 큐

디스크에 저장

영속성

손쉬운 복제

kafka

1. 분산 클러스터 증설 용이

2. 디스크에 저장하므로 다운되도 메세지 남아 있음

3. 처리(소비)되는 메세지가 없이 쌓여 있어도 느려지지 않음

4. 대량의 메세지를 처리하면서도 속도도 빠름

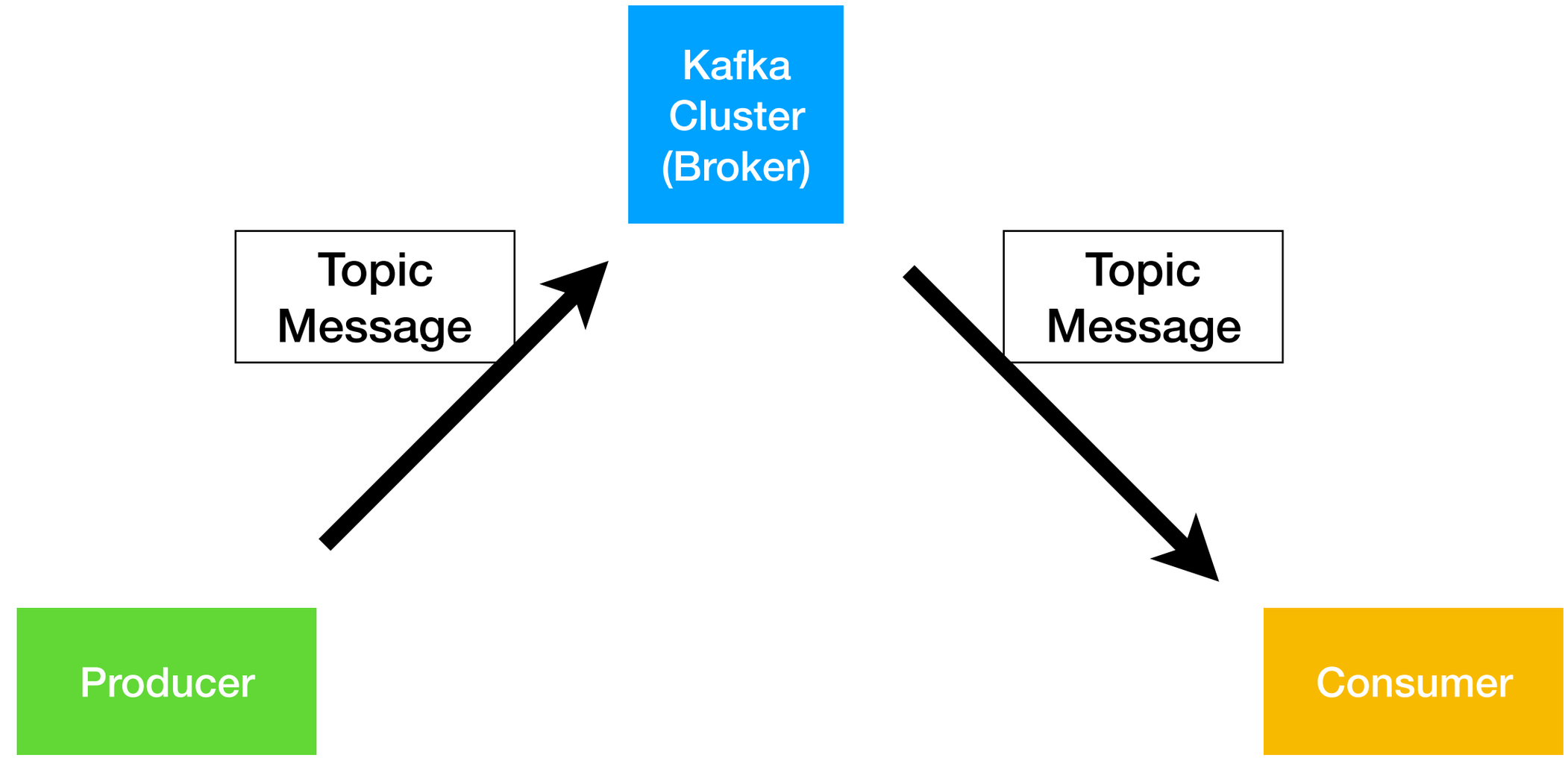
kafka

Topic 에 해당하는 Message를

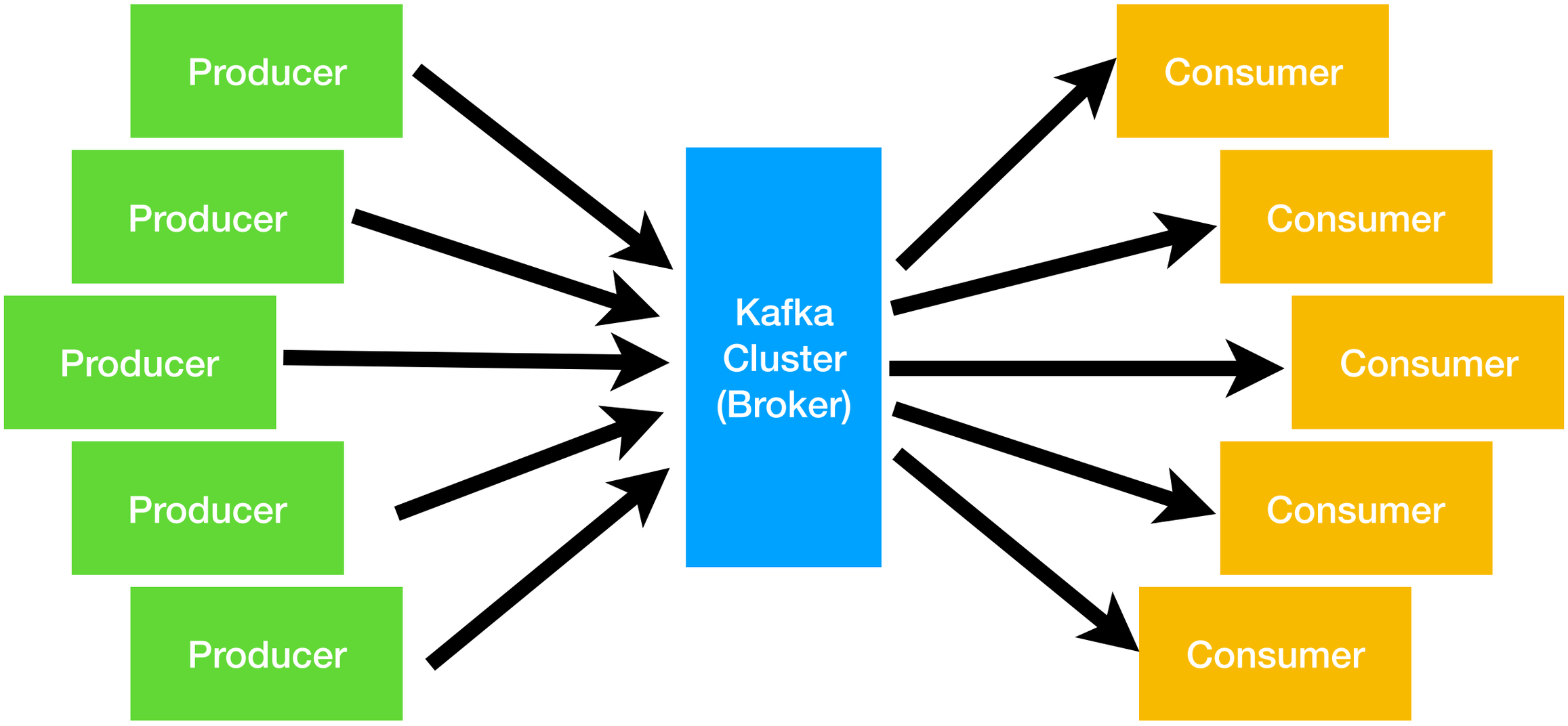
Producer 가 Producing(publish) 하면

Consumer 가 Consuming(subscribe) 한다.

kafka


Kafka
Cluster
(Broker)
Topic
Message
Topic
Message
Producer
Consumer

kafka

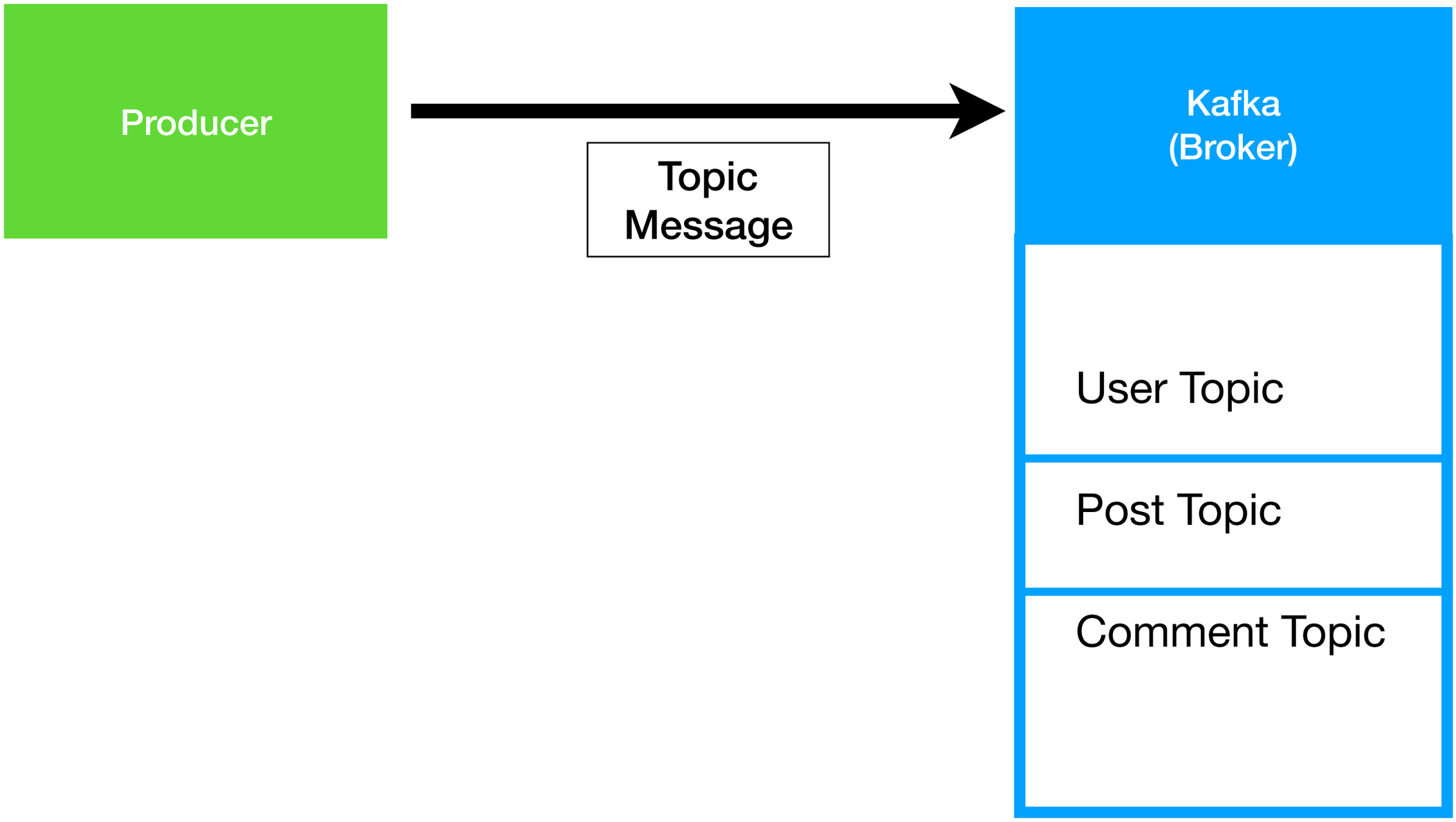


메세지 생산(pub)

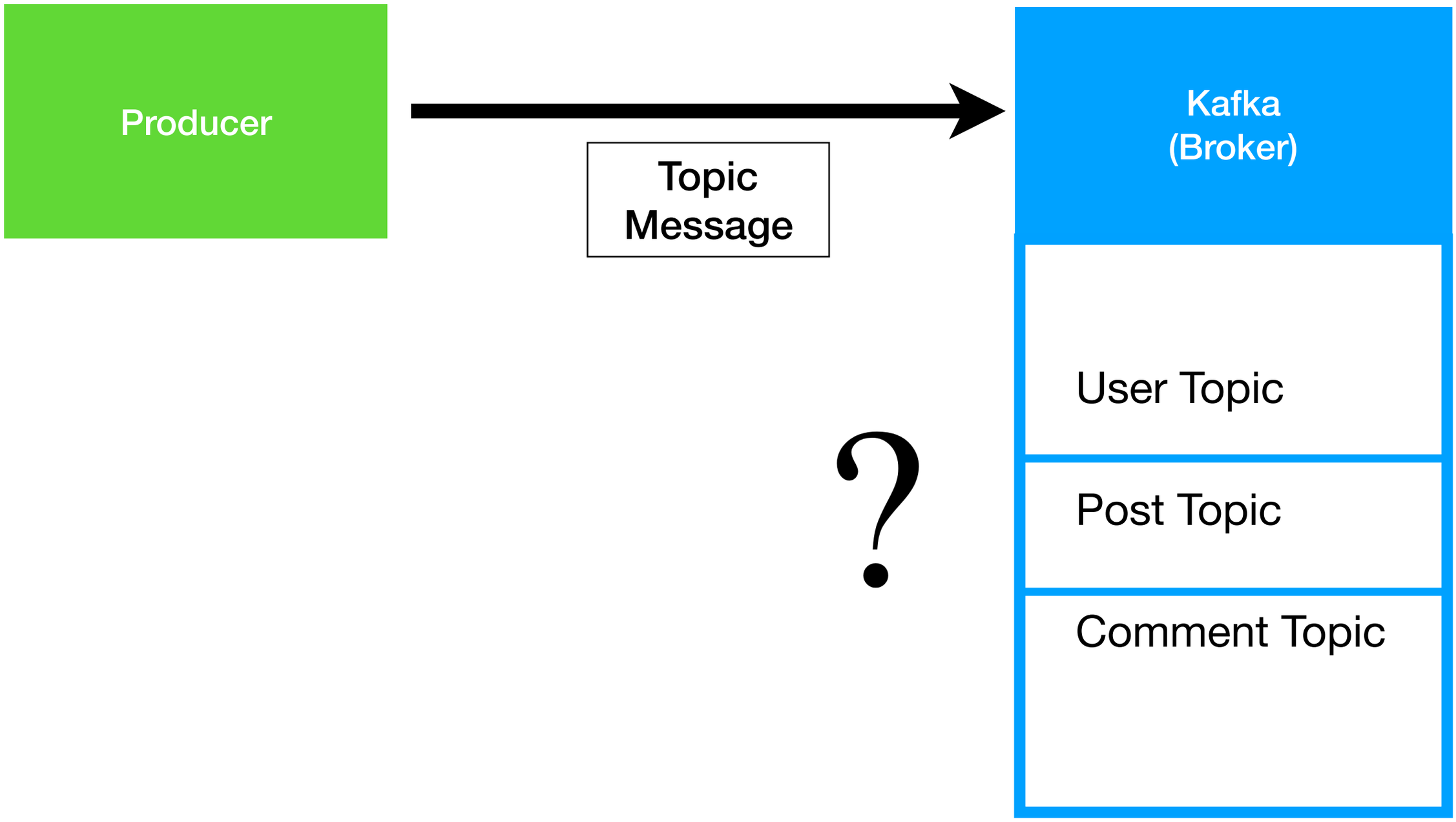
메세지 소비(sub)

Topic

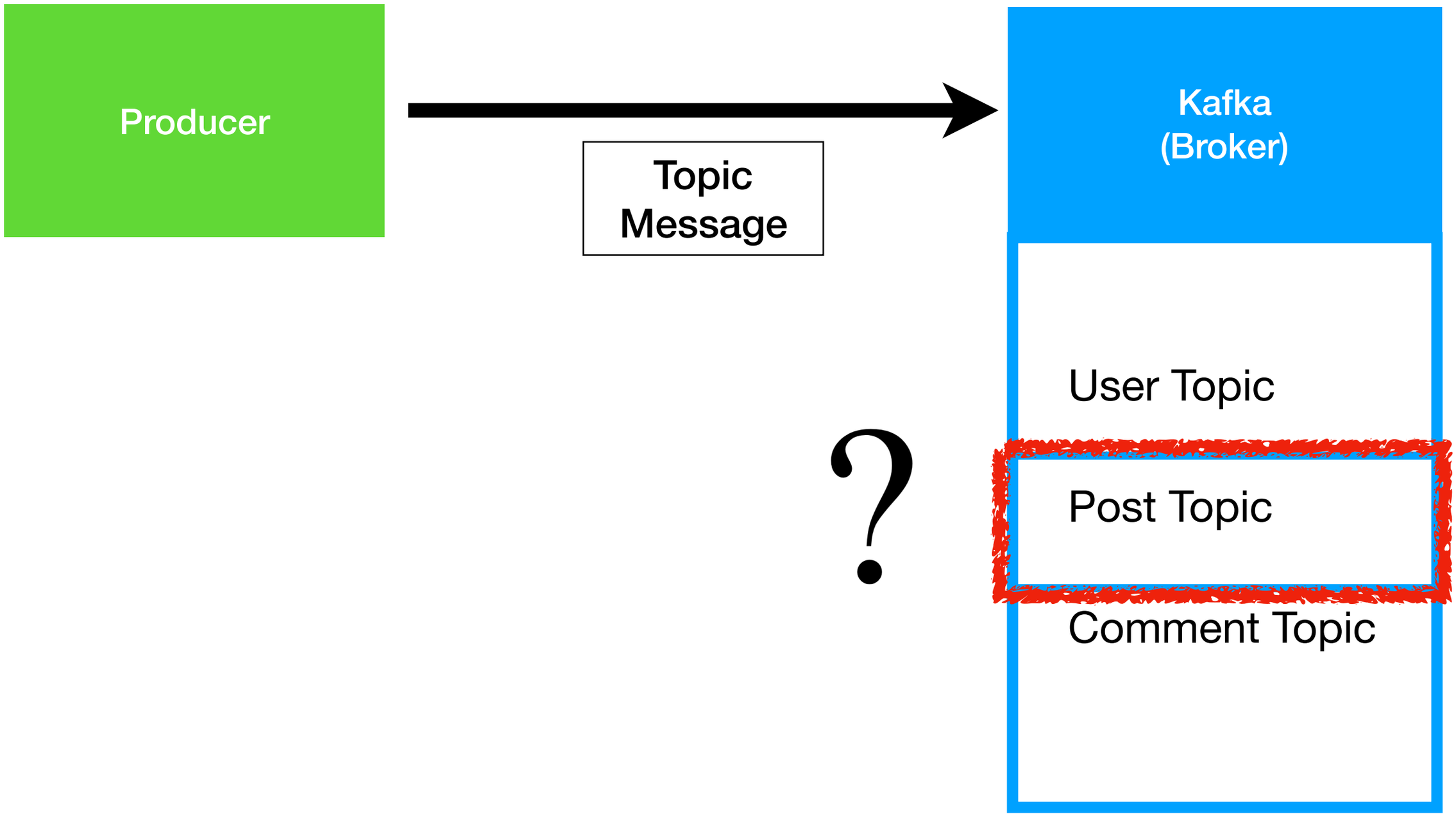
Topic


Producer
Topic
Message
Kafka
(Broker)
User Topic
Post Topic
Comment Topic

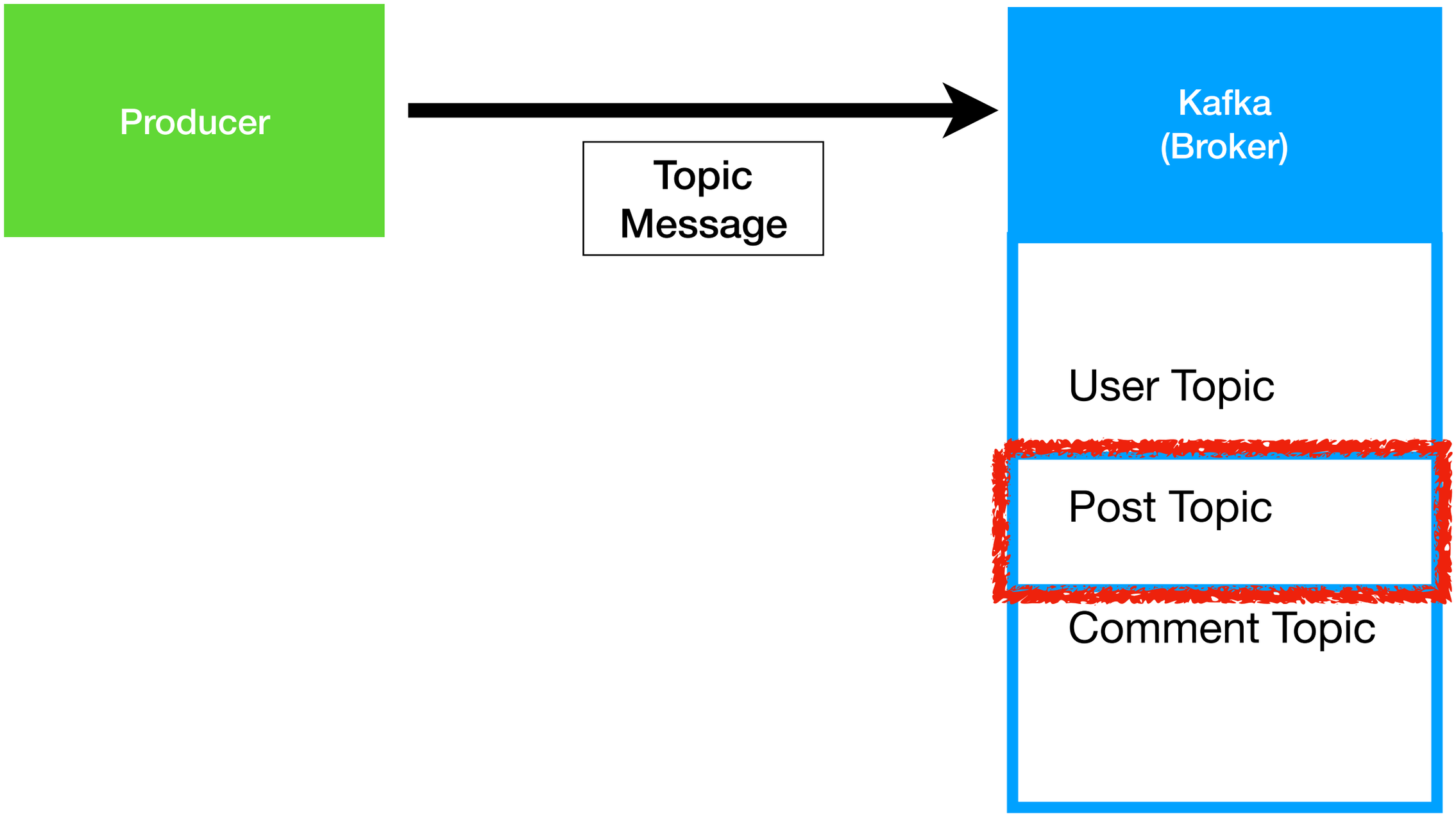
Topic


Producer
Topic
Message
?
Kafka
(Broker)
User Topic
Post Topic
Comment Topic

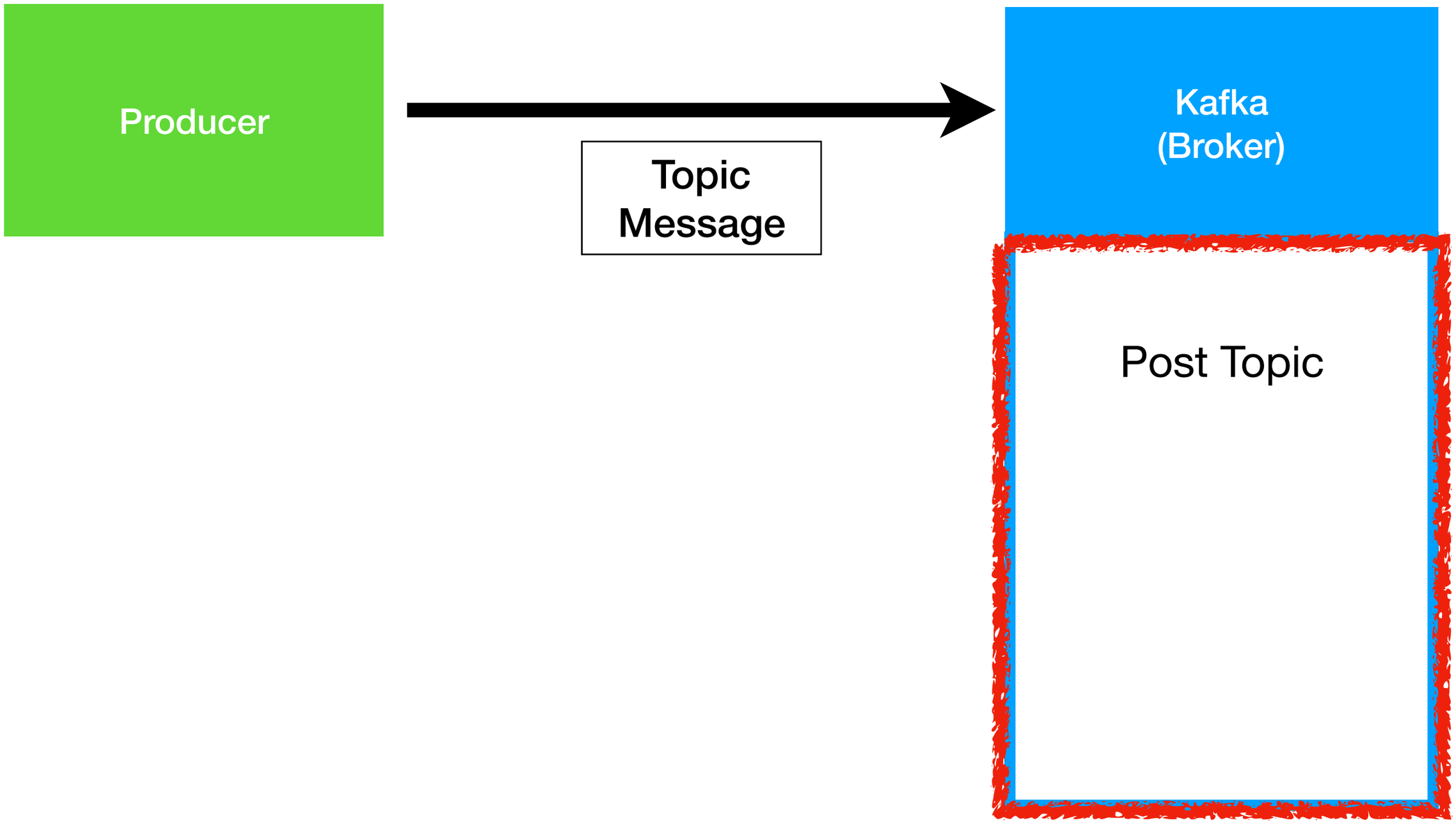
Topic


Producer
Topic
Message
?
Kafka
(Broker)
User Topic
Post Topic
Comment Topic

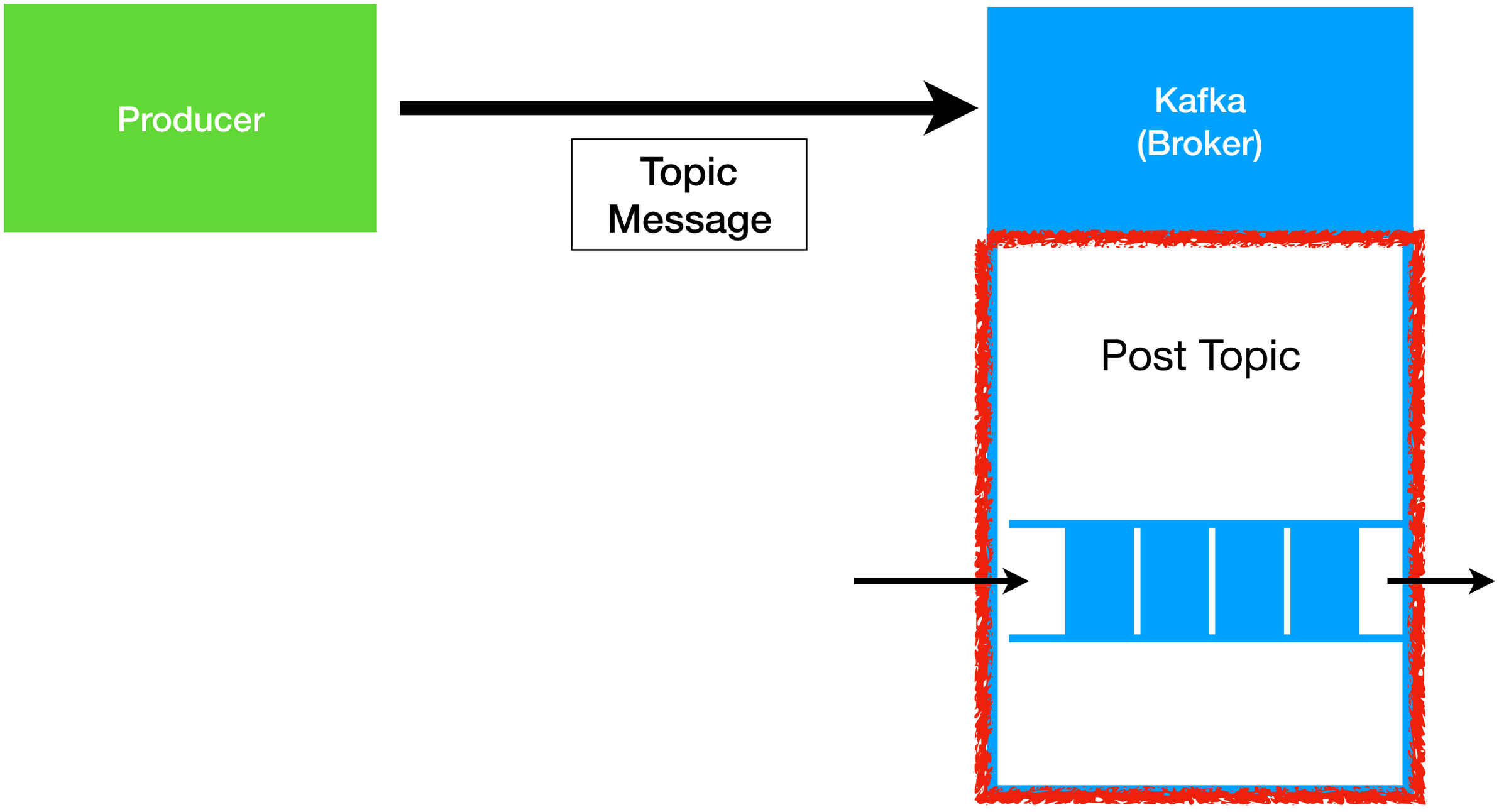
Partition


Producer
Topic
Message
Kafka
(Broker)
User Topic
Post Topic
Comment Topic

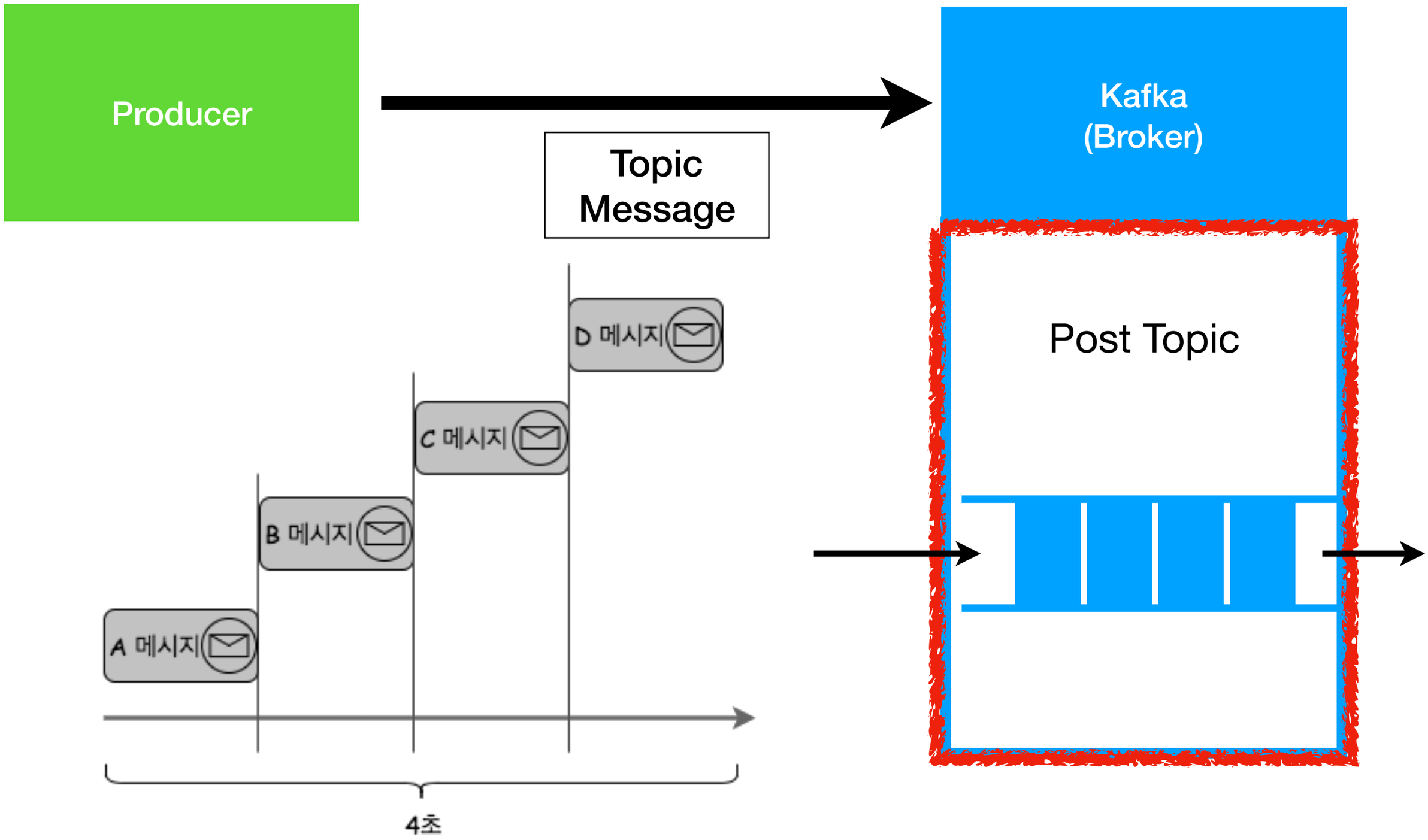
Partition



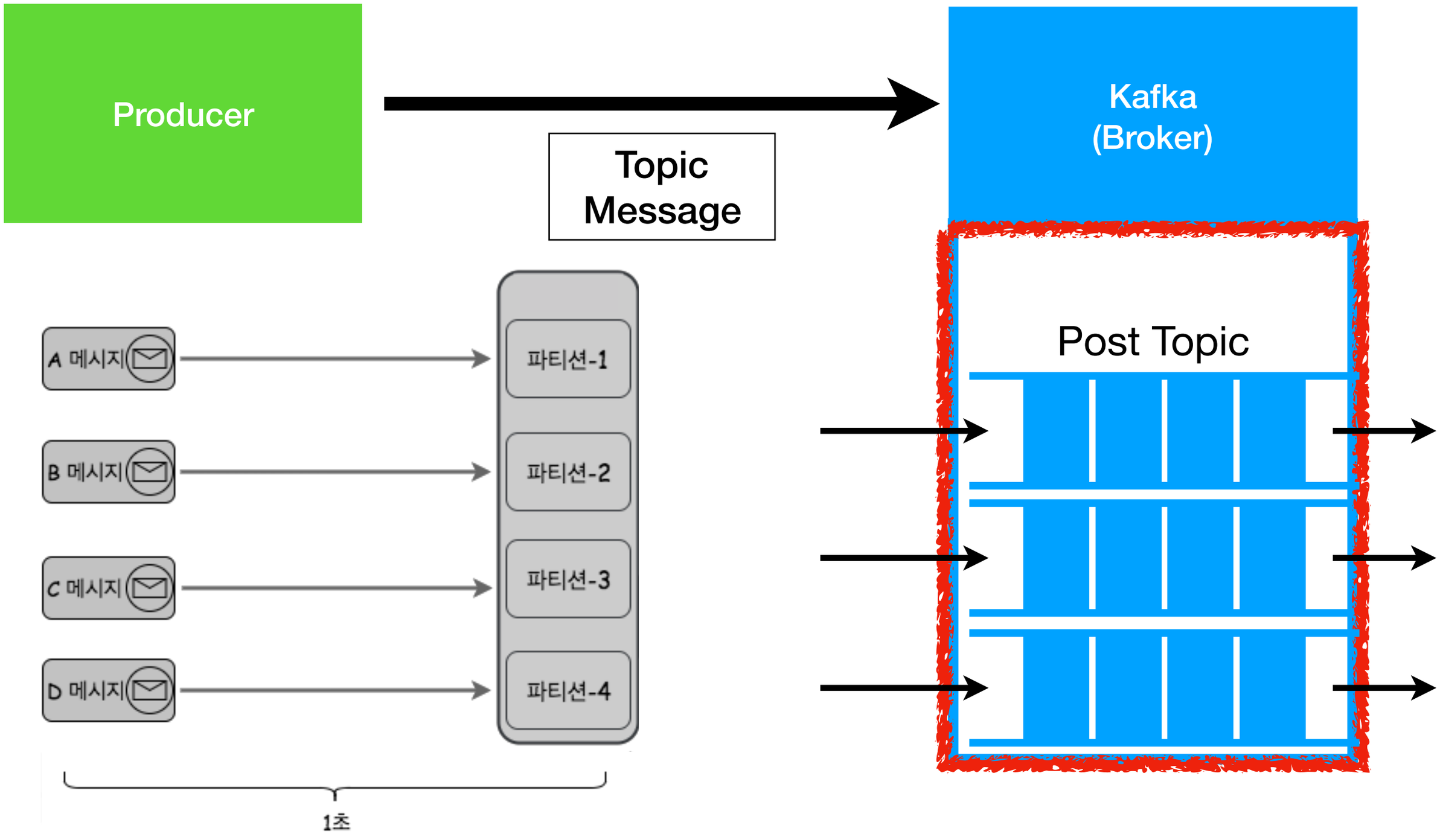
Partition



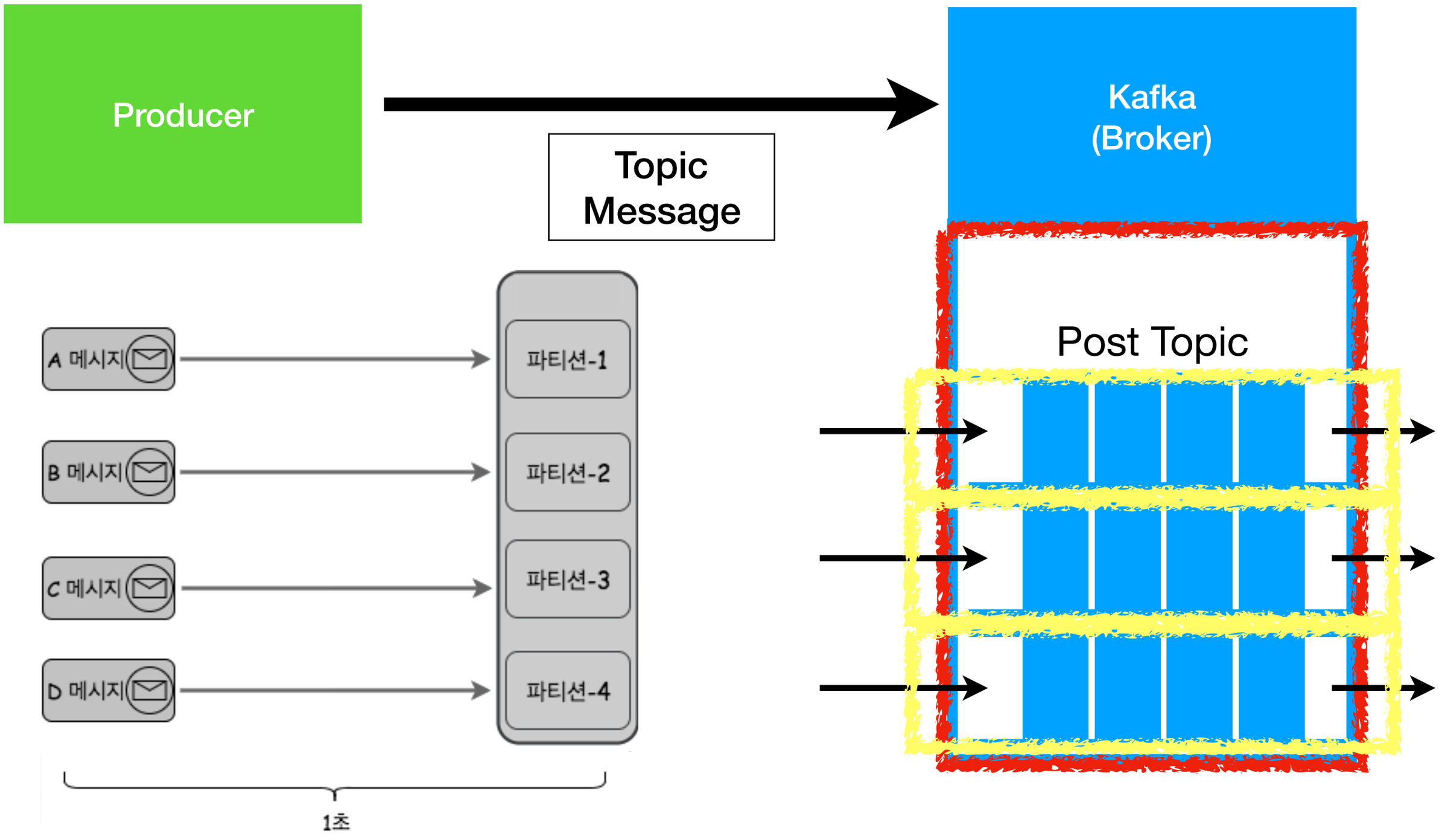
Partition


Producer
Topic
Message
Kafka
(Broker)
Post Topic
A 메시지
B 메시지
C 메시지
D 메시지
4초

Partition

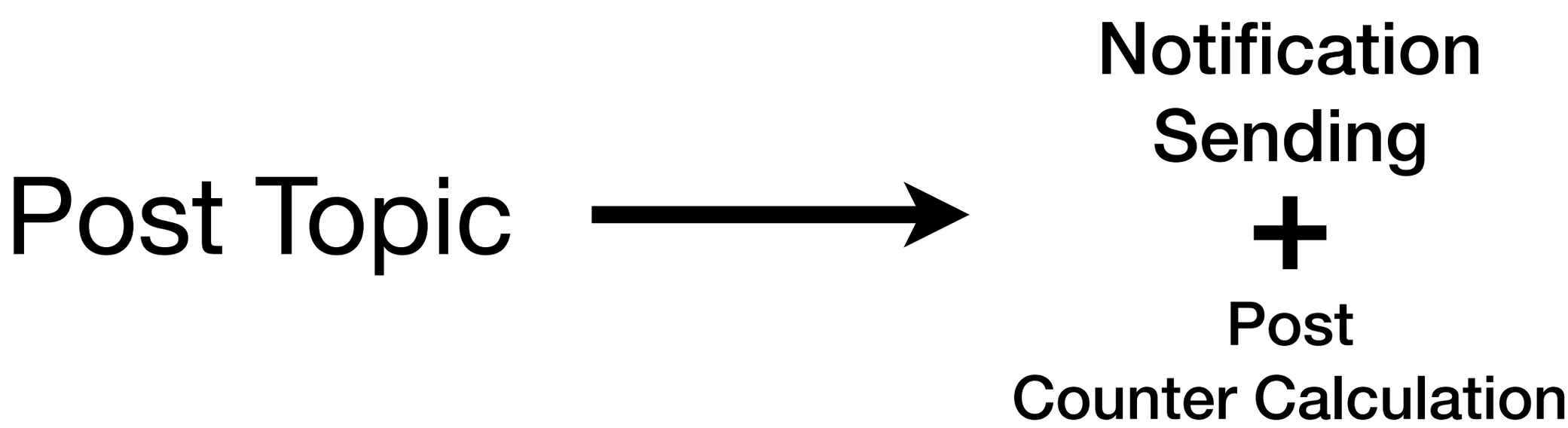

Producer
Topic
Message
Kafka
(Broker)
Post Topic
A 메시지
B 메시지
C 메시지
D 메시지
파티션-1
파티션-2
파티션-3
파티션-4
1초

Partition

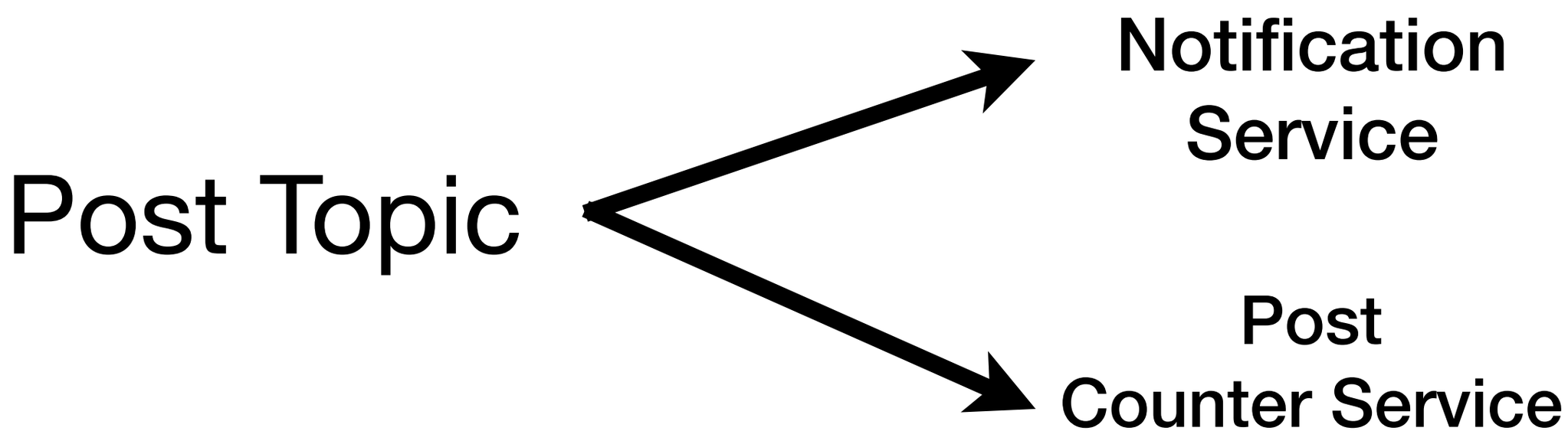

Producer
Topic
Message
Kafka
(Broker)
Post Topic
A 메시지
B 메시지
C 메시지
D 메시지
파티션-1
파티션-2
파티션-3
파티션-4
1초

Consumer Group

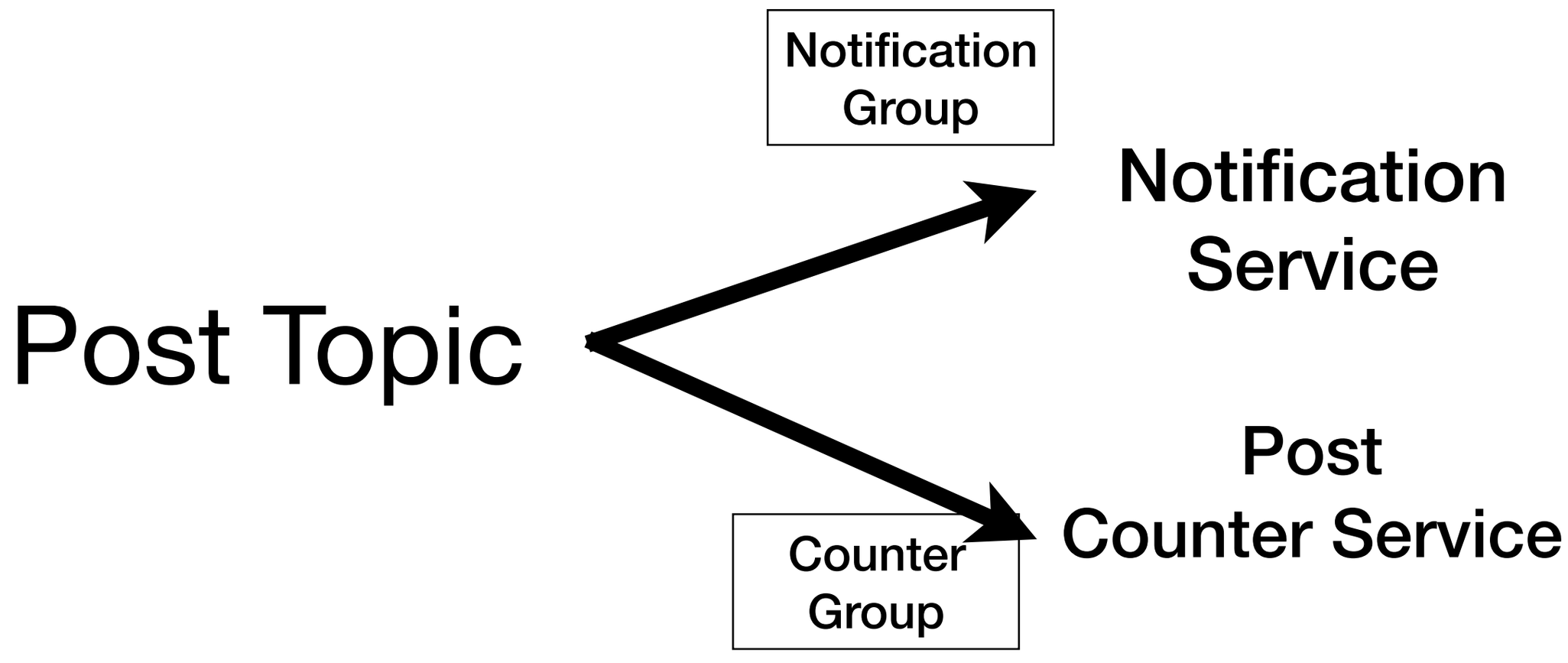
Consumer Group



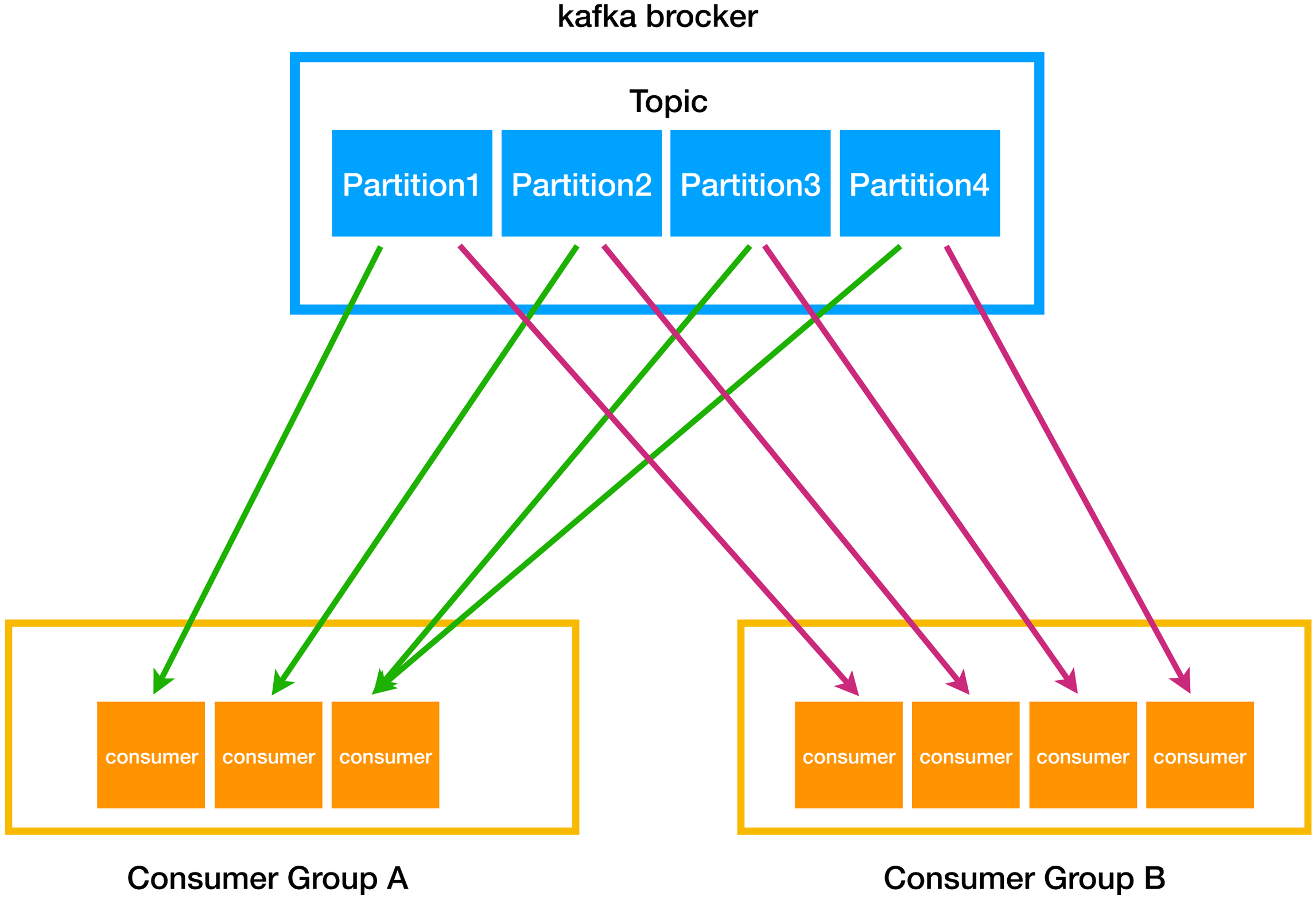
Consumer Group



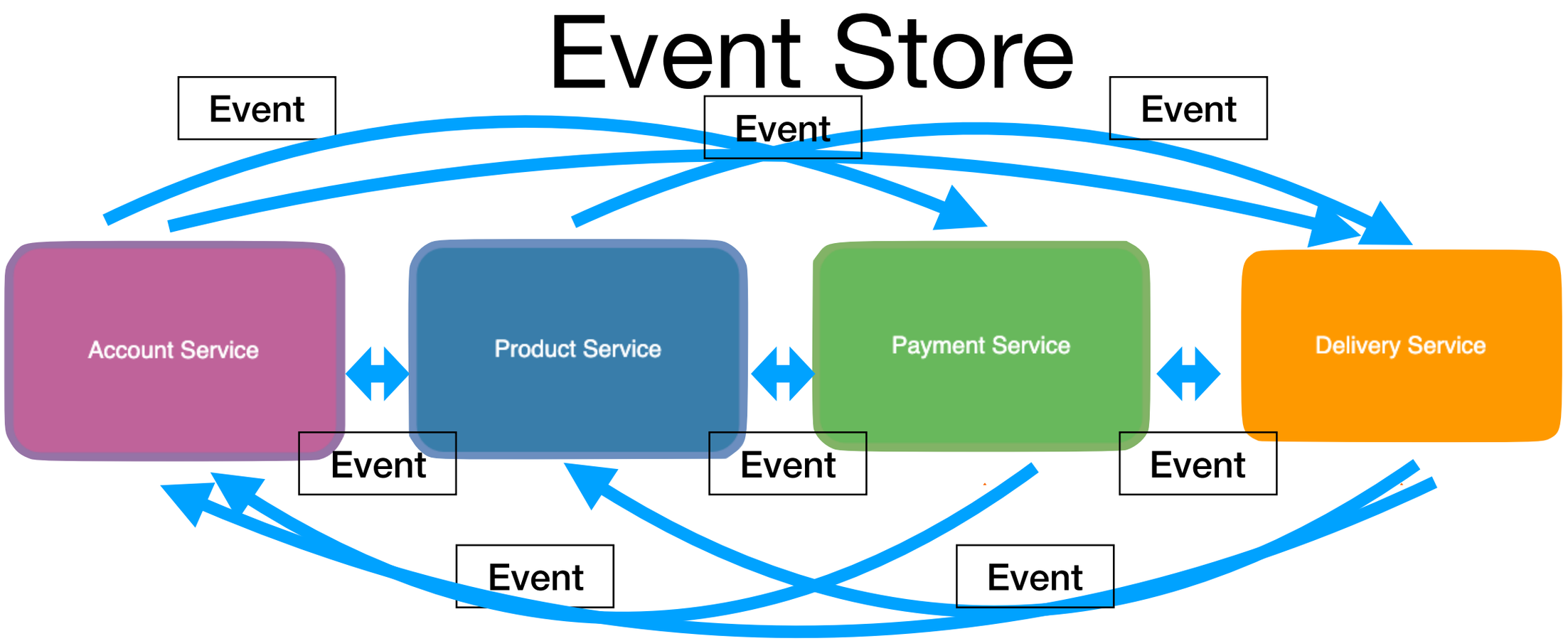
Consumer Group



Consumer Group

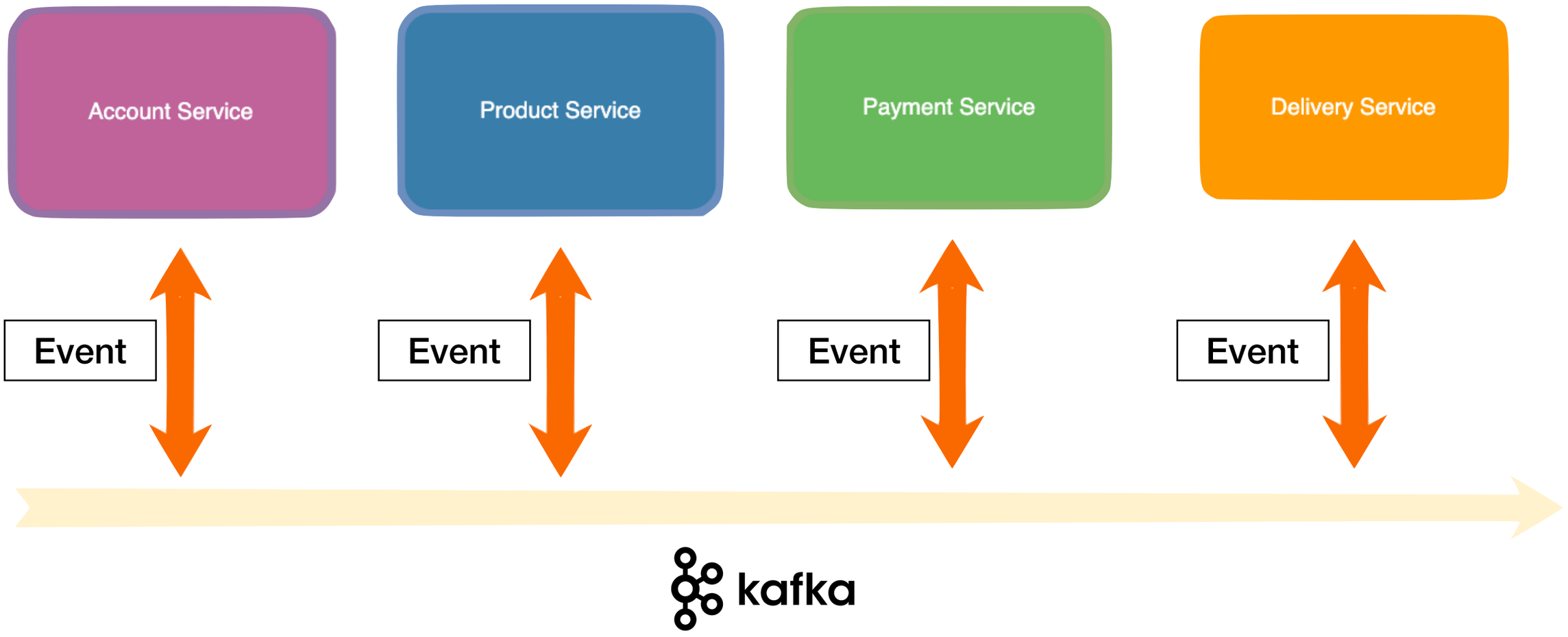


사용 Case 1


Event Store
Event
Event
Event
Account Service
Product Service
Payment Service
Delivery Service
Event
Event
Event
Event
Event

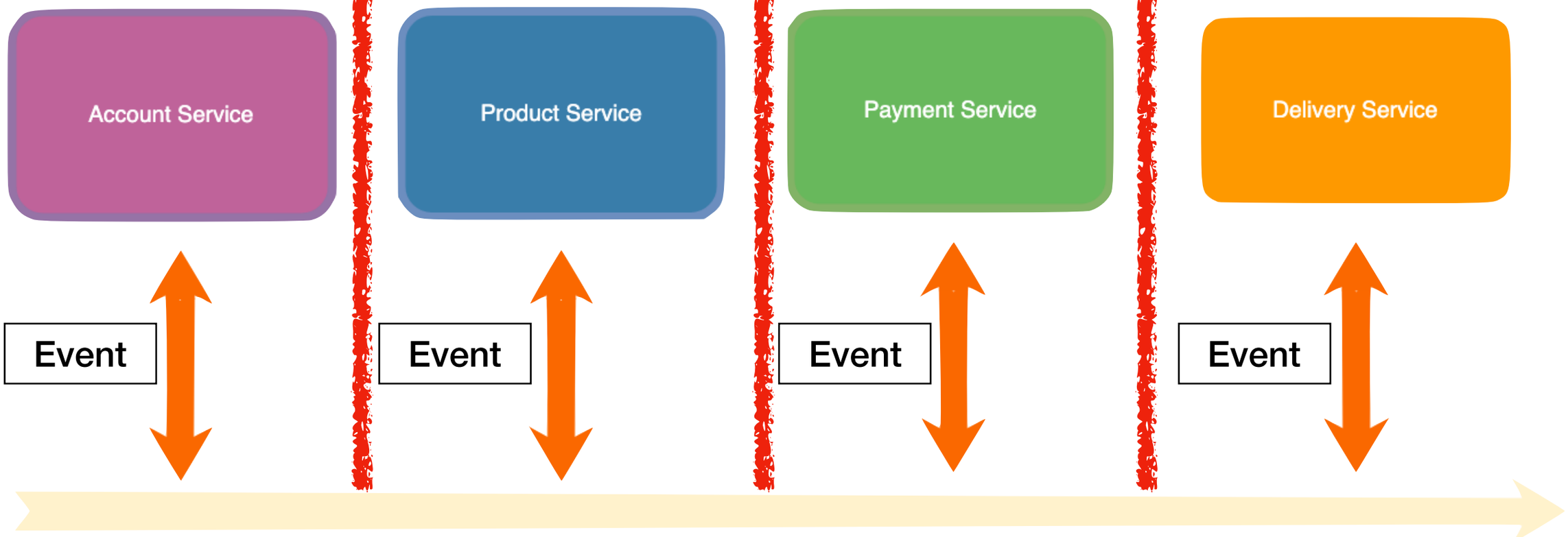
사용 Case 1

Event Store



사용 Case 1

Event Store

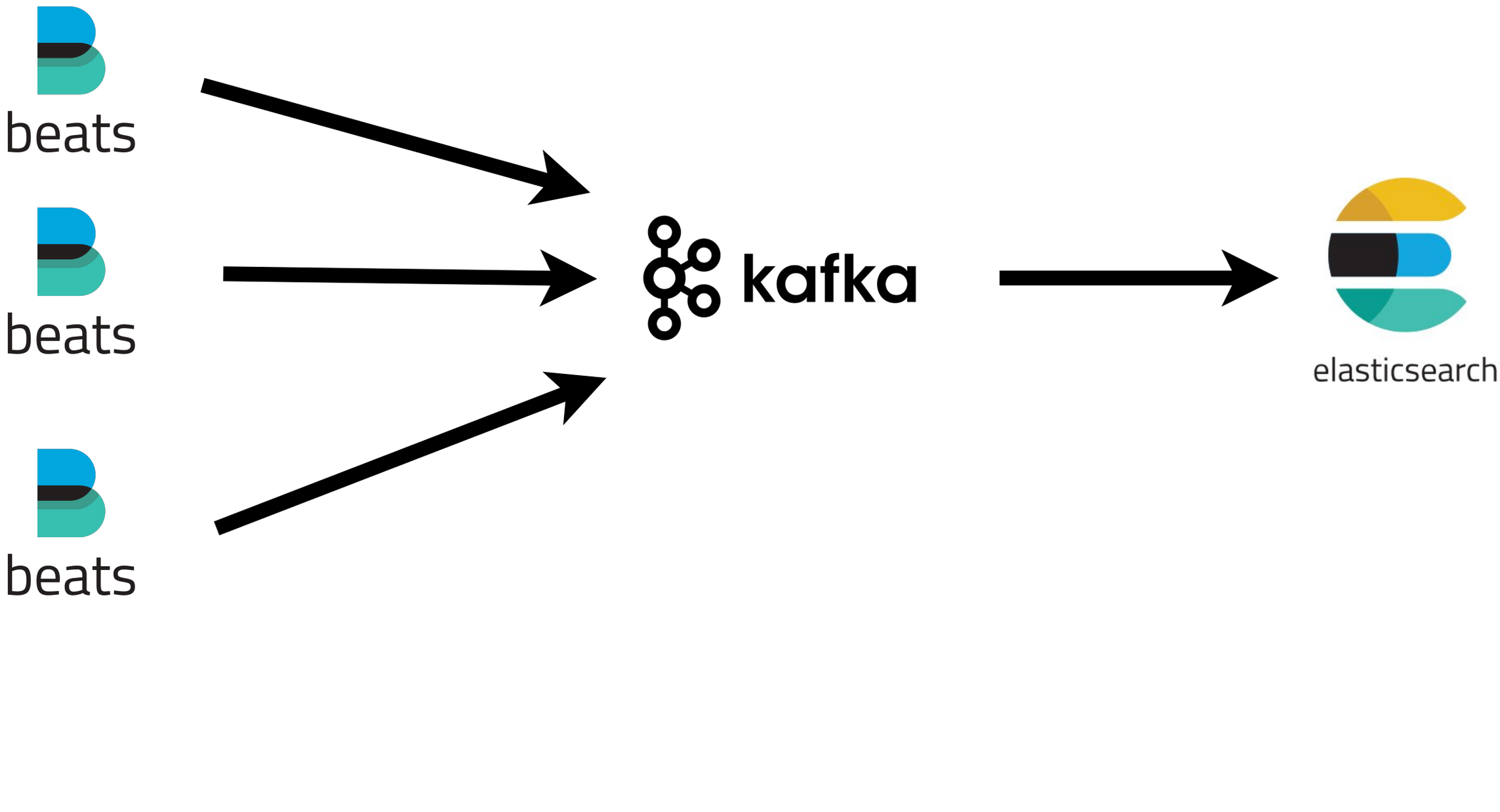


kafka

Decouple service

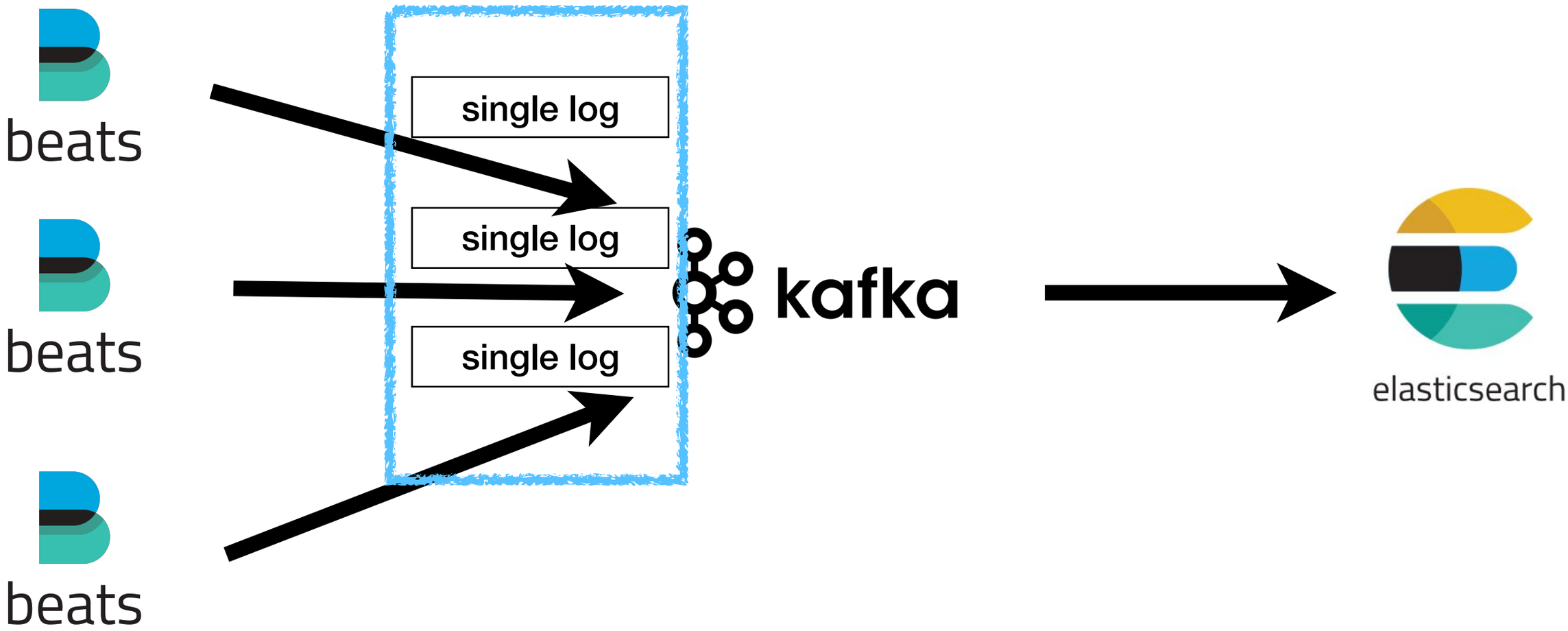
사용 Case 2

대용량 실시간 로그 처리



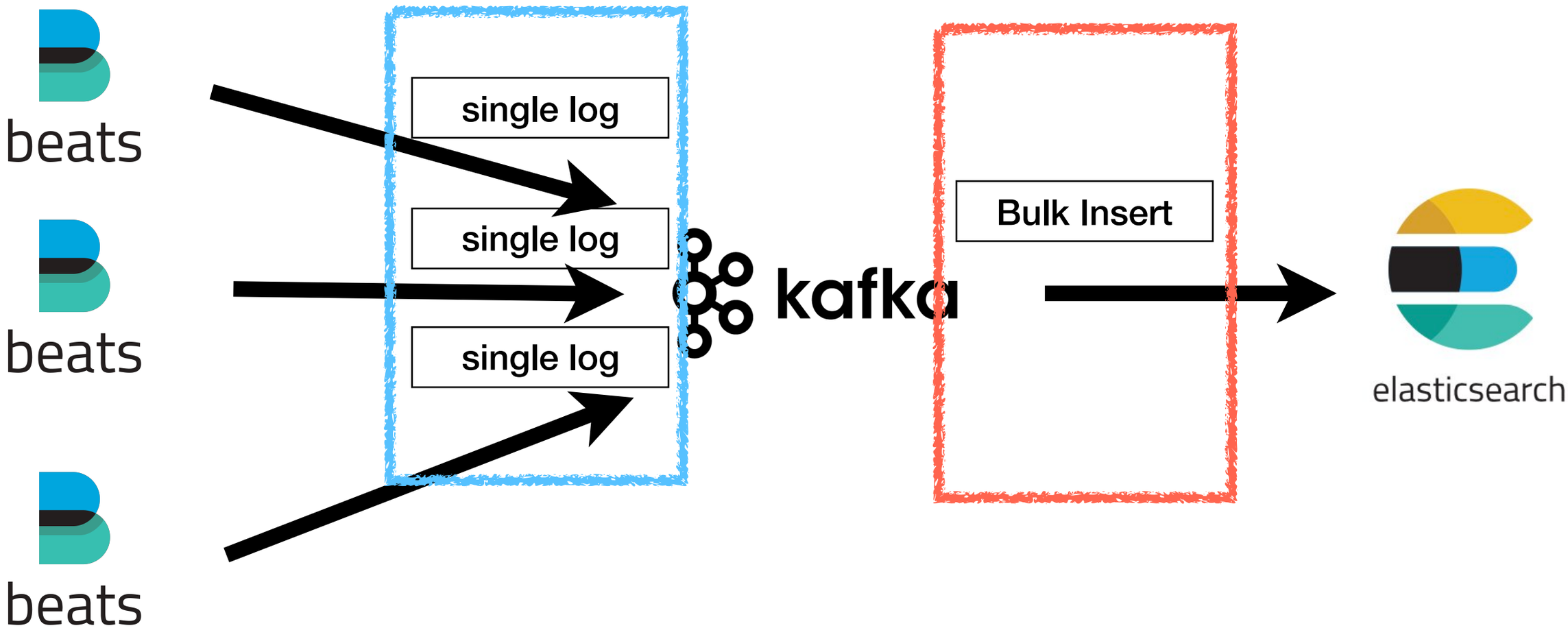
사용 Case 2

대용량 실시간 로그 처리



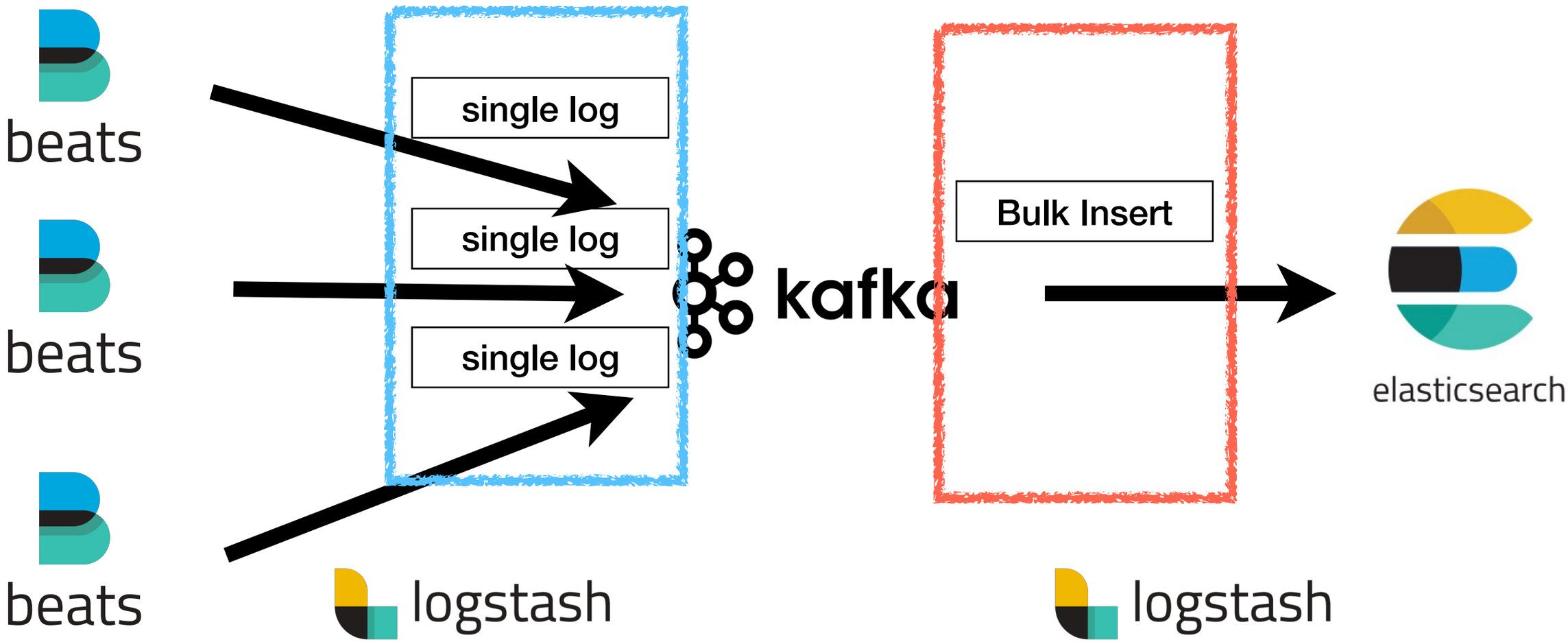
사용 Case 2

대용량 실시간 로그 처리



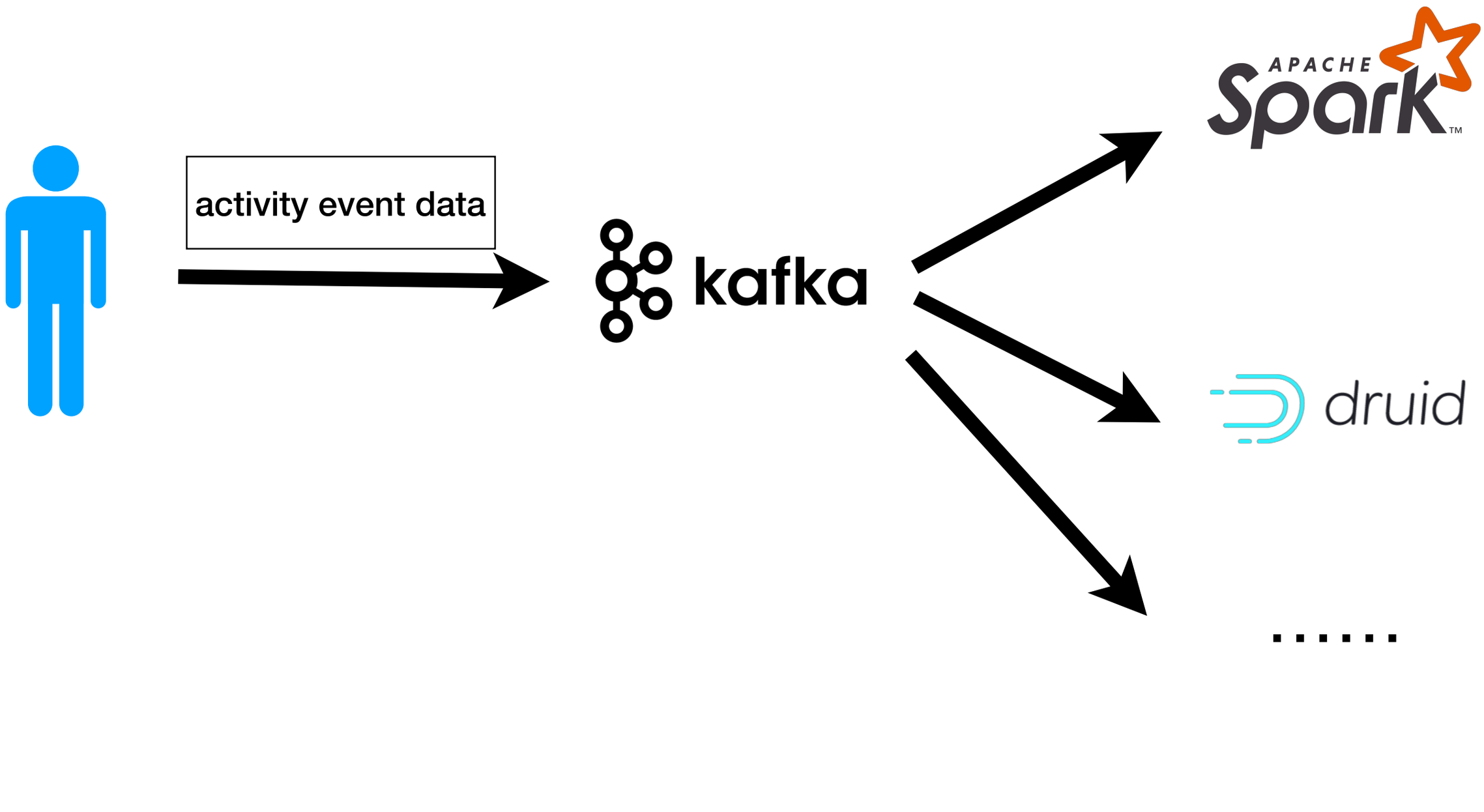
사용 Case 2

대용량 실시간 로그 처리



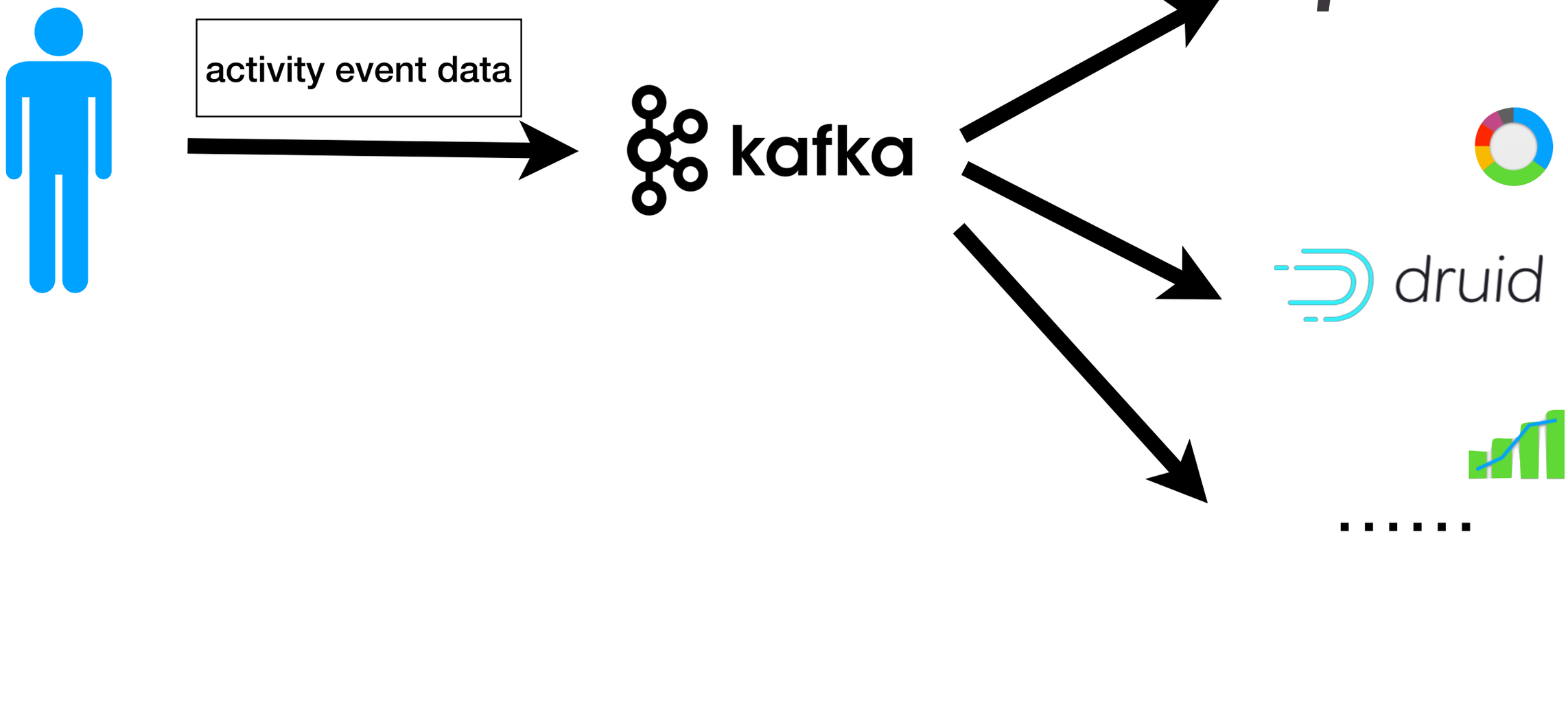
사용 Case 3

통계 / 분석



사용 Case 3

통계 / 분석



그래서 언제?

“그냥 바로 시스템에 끼얹으면 되나요?”

그래서 언제?


Festival del film Locarno
Concorso internazionale
tiff40
OFFICIAL SELECTION 2015
NYFF53
지금만나고
그때눈물리다
홍상수 감독작품
정재영 김민희 윤여정 기주봉 최화정 유준상 서영화 고아성
2015. 9. 24
청소년관람불가

그래서 언제?

시스템이 충분히 복잡해지고,

서비스별 연결을

개념적으로 따라가기

힘들어지면 도입을 추천합니다

그래서 어떻게?

1. 클러스터 구성

option1

Self Managed Software

Confluent Platform

option2

Fully Managed Service

Confluent Cloud

option3

그래서 어떻게?

2. 예제 example

confluentinc / **examples**

https://github.com/confluentinc/examples

Full list

Demo	Local	Docker	Category	Description
Avro	Y	N	Confluent Platform	Examples of client applications using Avro and Confluent Schema Registry
CDC with MySQL	N	Y	Data Pipelines	Self-paced steps to setup a change data capture (CDC) pipeline
CDC with Postgres	N	Y	Data Pipelines	Enrich event stream data with CDC data from Postgres and then stream into Elasticsearch
Clickstream	Y	Y	Stream Processing	Automated version of the KSQL Clickstream demo
Cloud clients	Y	N	Confluent Cloud	Examples of client applications in different programming languages connecting to Confluent Cloud
Connect and Kafka Streams	Y	N	Data Pipeline	Demonstrate various ways, with and without Kafka Connect, to get data into Kafka topics and then loaded for use by the Kafka Streams API
CP Demo	Y	Y	Confluent Platform	Confluent Platform Demo with a playbook for Kafka streaming ETL deployments
CP Quickstart	Y	Y	Confluent Platform	Confluent Platform Quickstart
GCP pipeline	N	Y	Data Pipeline	Work with Confluent Cloud to build cool pipelines into Google Cloud Platform (GCP)
Hybrid cloud	Y	Y	Confluent Cloud	End-to-end demo of a hybrid Kafka Cluster between Confluent Cloud and on-prem using Confluent Replicator
KSQL UDF	Y	N	Stream Processing	Advanced KSQL UDF use case for connected cars
KSQL workshop	N	Y	Stream Processing	showcases Kafka stream processing using KSQL and can run self-guided as a KSQL workshop

그래서 어떻게?

3. 진행 순서

1. Topic 설정
2. partition 설정
3. producing
4. consuming

그래서 어떻게?

4. coding ……

아파치 카프카

소프트웨어

아파치 카프카는 아파치 소프트웨어 재단이 스칼라로 개발한 오픈 소스 메시지 브로커 프로젝트이다. 이 프로젝트는 실시간 데이터 피드를 관리하기 위해 통일된, 높은 스루풋의 낮은 레이턴시를 지닌 플랫폼을 제공하는 것이 목표이다. 위키백과

최초 출시일: 2011년 1월

개발 상태: 지원 중

개발자: 아파치 소프트웨어 재단

종류: 메시지 브로커

라이선스: 아파치 라이선스 2.0

작성 언어: 스칼라, 자바

그래서 어떻게?

4. coding ……

아파치 카프카

소프트웨어

아파치 카프카는 아파치 소프트웨어 재단이 스칼라로 개발한 오픈 소스 메시지 브로커 프로젝트이다. 이 프로젝트는 실시간 데이터 피드를 관리하기 위해 통일된, 높은 스루풋의 낮은 레이턴시를 지닌 플랫폼을 제공하는 것이 목표이다. 위키백과

최초 출시일: 2011년 1월

개발 상태: 지원 중

개발자: 아파치 소프트웨어 재단

종류: 메시지 브로커

라이선스: 아파치 라이선스 2.0

작성 언어: 스칼라, 자바

그래서 어떻게?

JVM 언어 아니면 못쓰나요?

그래서 어떻게?

5. kafka-rest-proxy

confluentinc / kafka-rest

Kafka REST Proxy

kafka brocker 연결을

rest api 형태로

제공하는 proxy

rdkafka 보다 이게 좋더라…

왜?

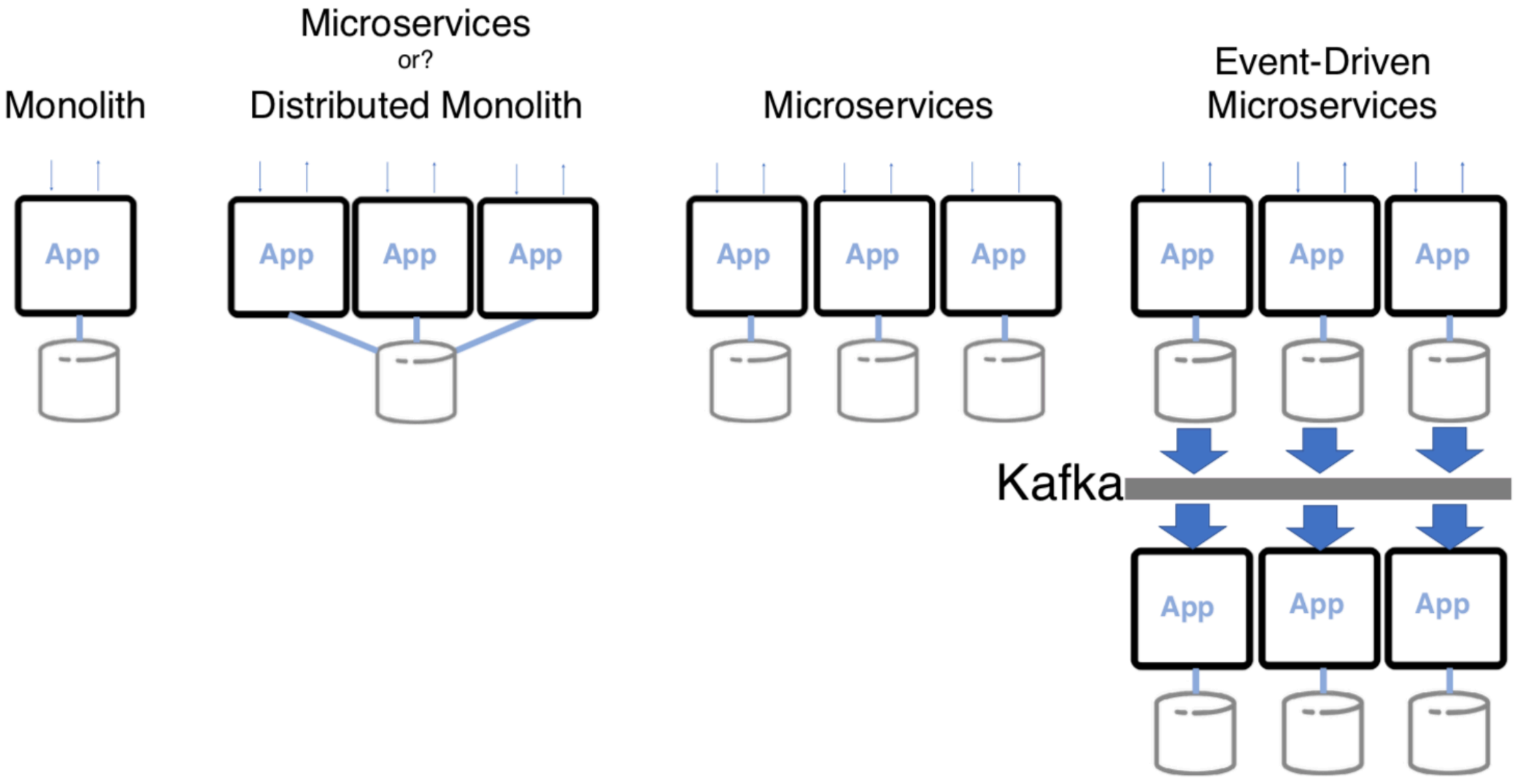
Data Pipeline

Event Driven System

Data Stream

MSA

생각해볼것


Monolith
Microservices
or?
Distributed Monolith
Microservices
Event-Driven
Microservices
App
App
App
App
App
App
App
App
App
Kafka
App
App
App

써보며 느낀점

- 시스템이 충분히 복잡해지면 도입한다
- 기반 인프라가 되어간다. DB 같은 필수요소 처럼..
- use case 가 다양하다..
- 데이터 구조를 잘 짜야한다.
- MSA 하려면 거의 필수라고 느껴진다.
- producer 1 : multi consumer 구조가 핵심

도입시 고민거리

- kafka message 는 어떻게 잘 만들어야 하는가?
- consumer 에서 business model의 상태는 어떻게 관리해야 하는가?
- fault tolerant 구성을 하기 위한 아키텍처

도입시 고민거리

결국, 우리에게 잘 맞는가?

안해보면 모른다… Just do it

요약

1. 1 producer : n consumer 좋아요
2. 메세지 모델링은 고민되요
3. 도입 활용은 적당한 필요를 기반으로
4. 언어는 jvm 이 편하고
5. 클러스터 구성은 managed 가 속편합니다.
6. MSA, Event Driven Architecture 에서는 필수라고 생각됩니다.

참고자료

https://kafka.apache.org/quickstart

https://github.com/confluentinc/kafka-rest

https://www.confluent.io/what-is-apache-kafka/

https://aws.amazon.com/ko/msk/getting-started/

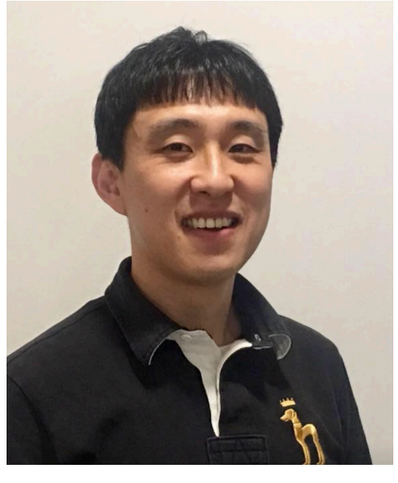
https://docs.confluent.io/current/kafka-rest/docs/index.html

https://content.pivotal.io/webinars/feb-27-microservices-events-and-breaking-the-data-monolith-webinar

https://www.confluent.io/kafka-summit-london18/keynote-the-death-and-rebirth-of-the-event-driven-architecture

추천도서

카프카, 데이터 플랫폼의 최강자



YES24

보다 자세한 내용은 이 책으로..

Q&A

궁금하신게 있다면.. 물어봐주세요?!?!