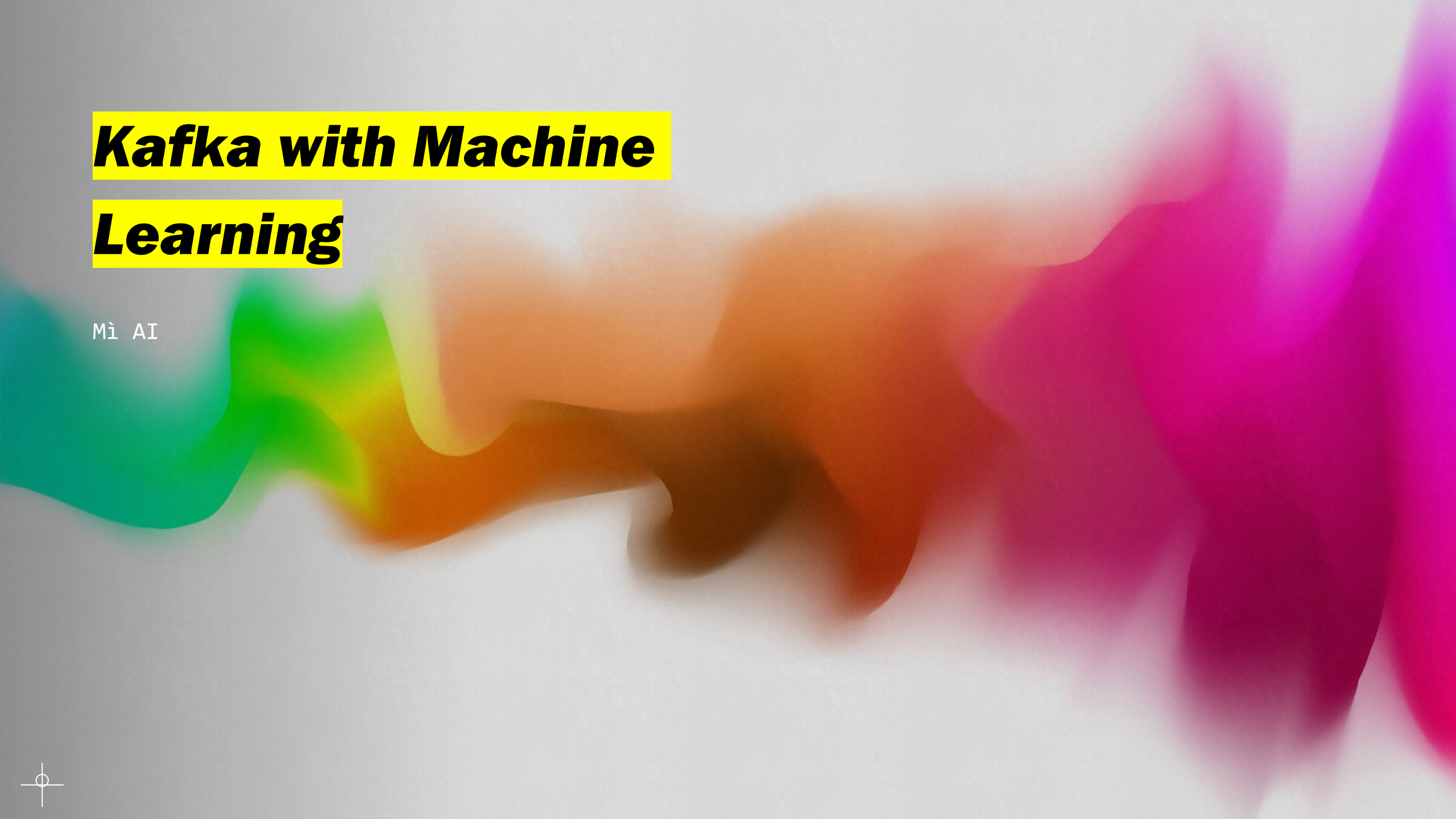
Kafka with Machine Learning

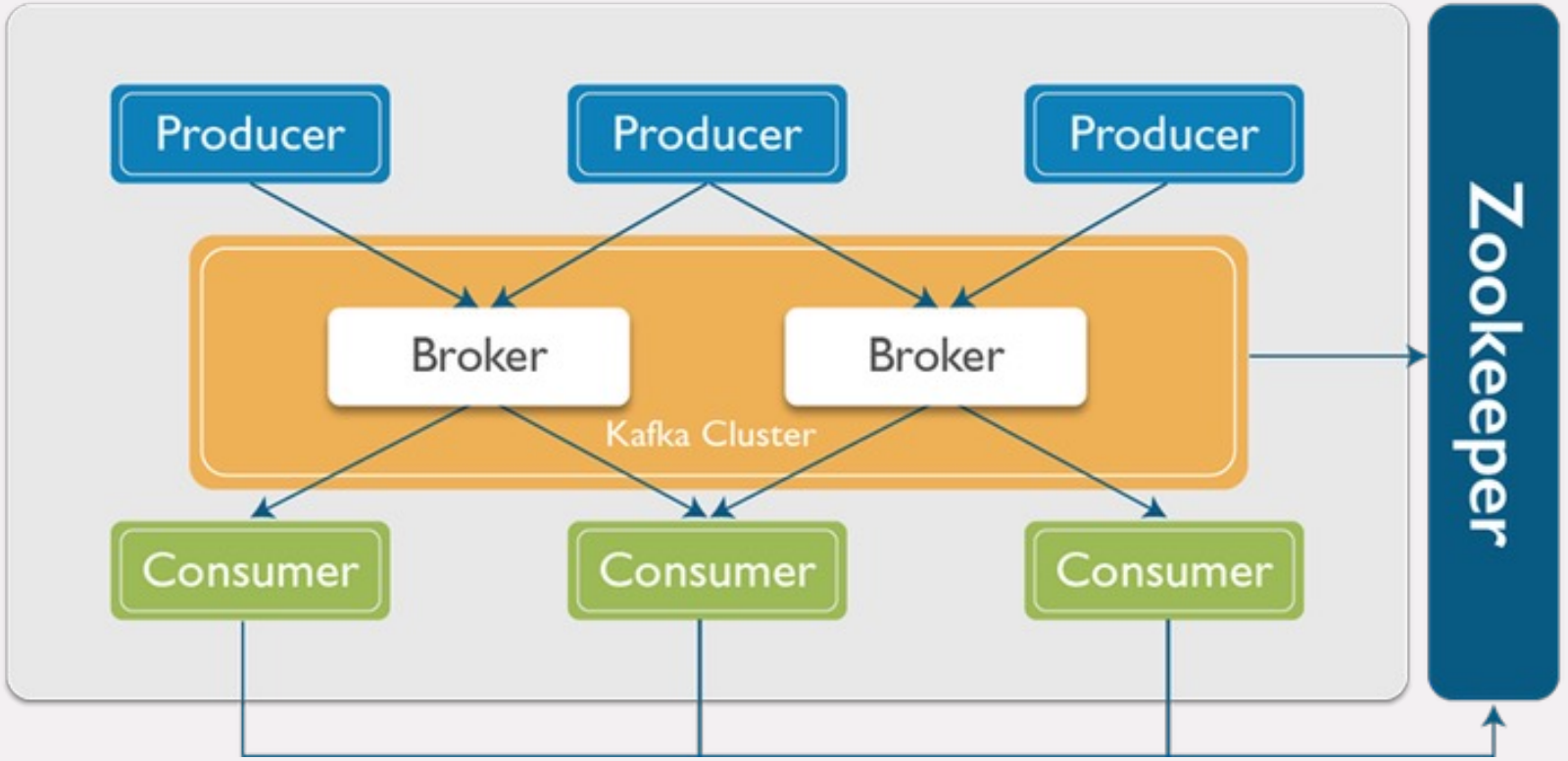
Mì AI



Today’s agenda

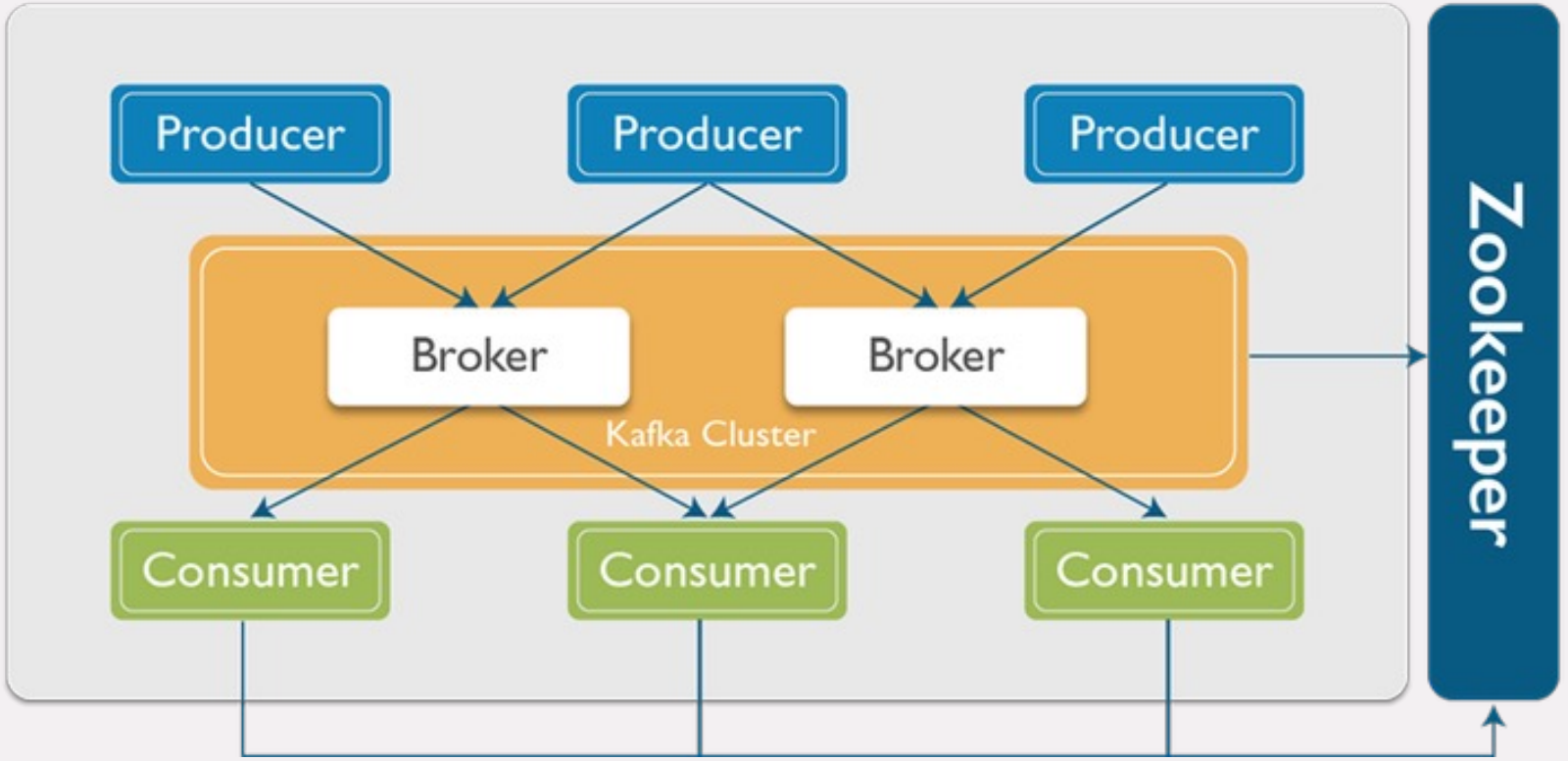
- What is Kafka?
- Why we need Kafka?
- System pipeline
- Handson

What is Kafka?



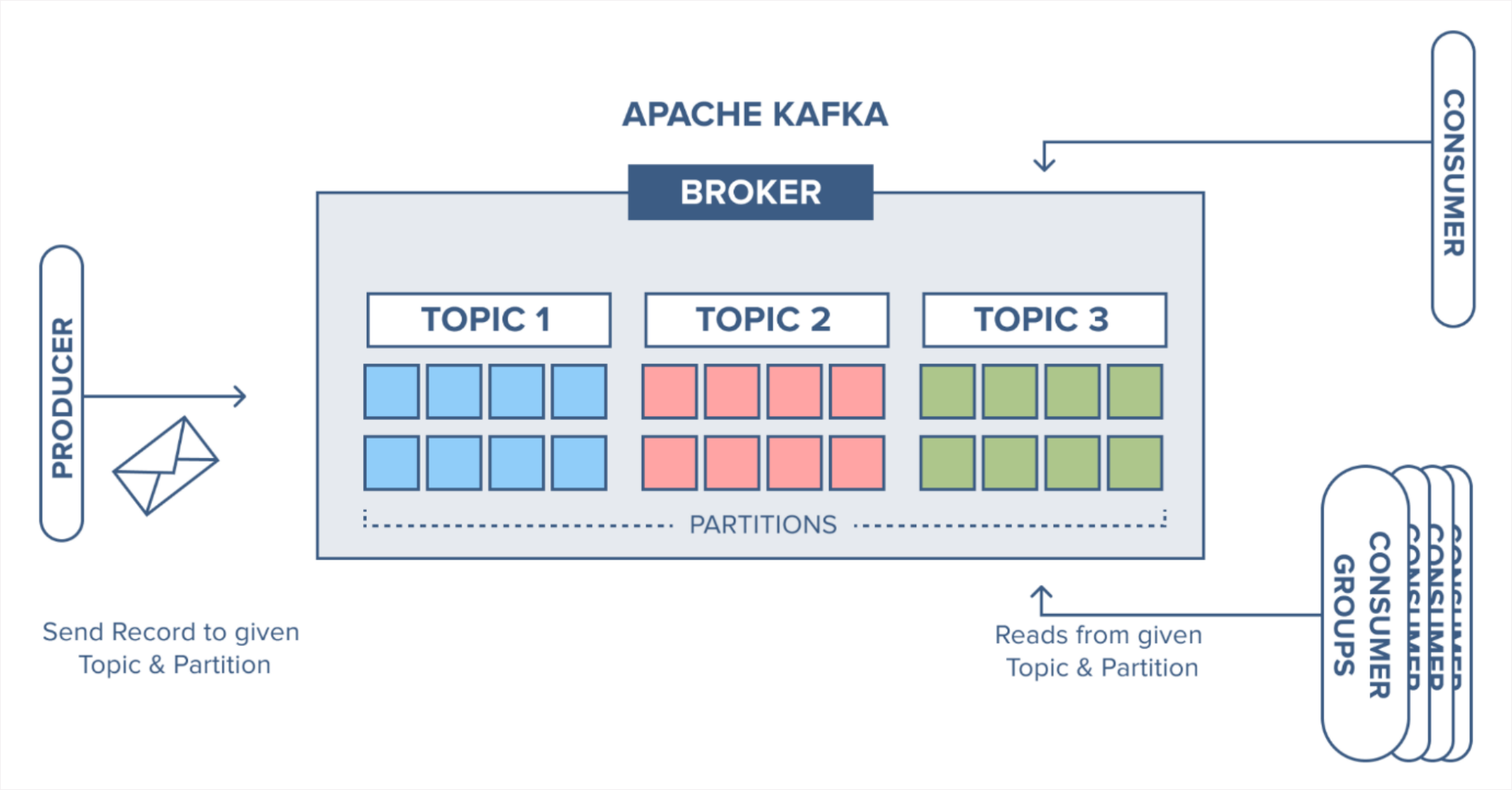
Apache Kafka là một nền tảng theo kiến trúc phân tán cho phép lưu trữ sự kiện và xử lý dữ liệu luồng mã nguồn mở được phát triển bởi Apache Software Foundation được viết bằng Java và Scala

What is Kafka?

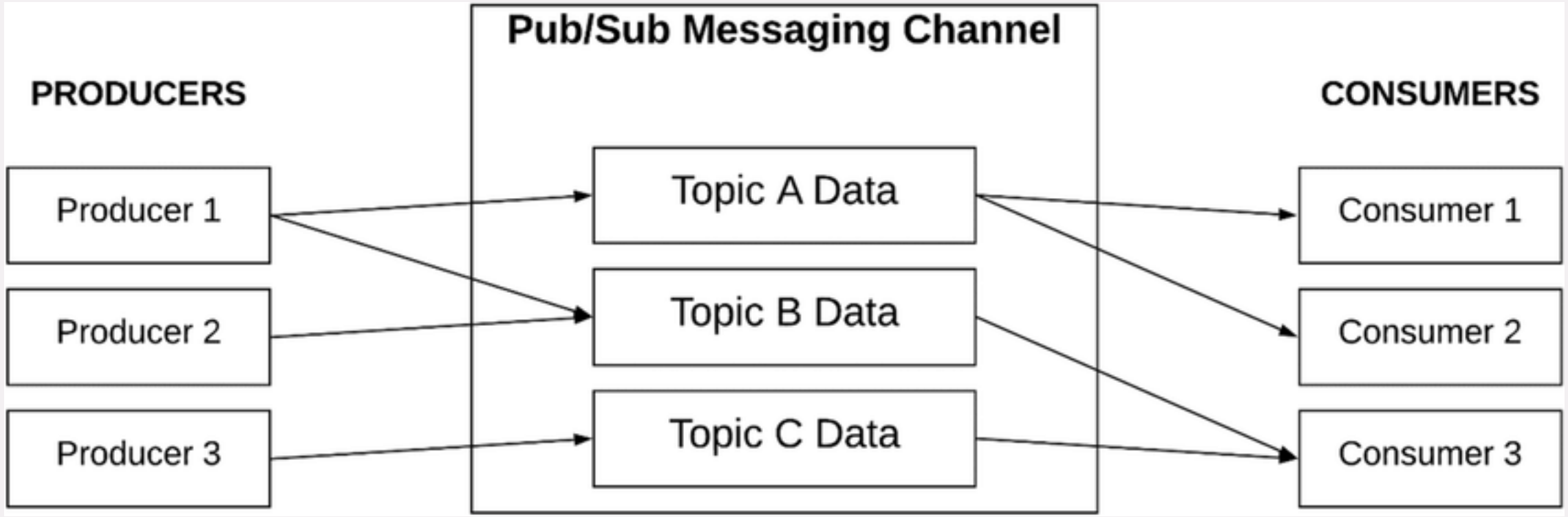


- Là hệ thống message pub/sub phân tán (distributed messaging system).
- Kafka là một trong những nền tảng message publish/ subscribe phân tán mã nguồn mở phổ biến nhất hiện nay và được xây dựng với mục đích để xử lý dữ liệu streaming real-tim e.

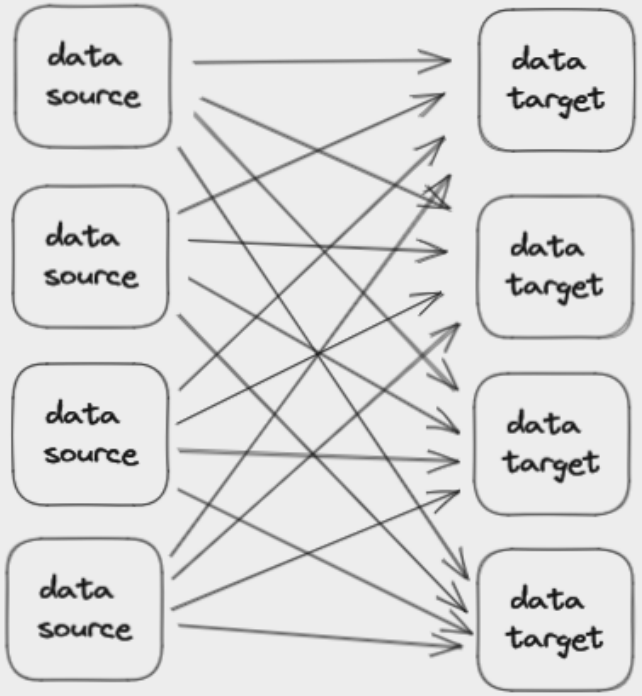
What is Kafka?

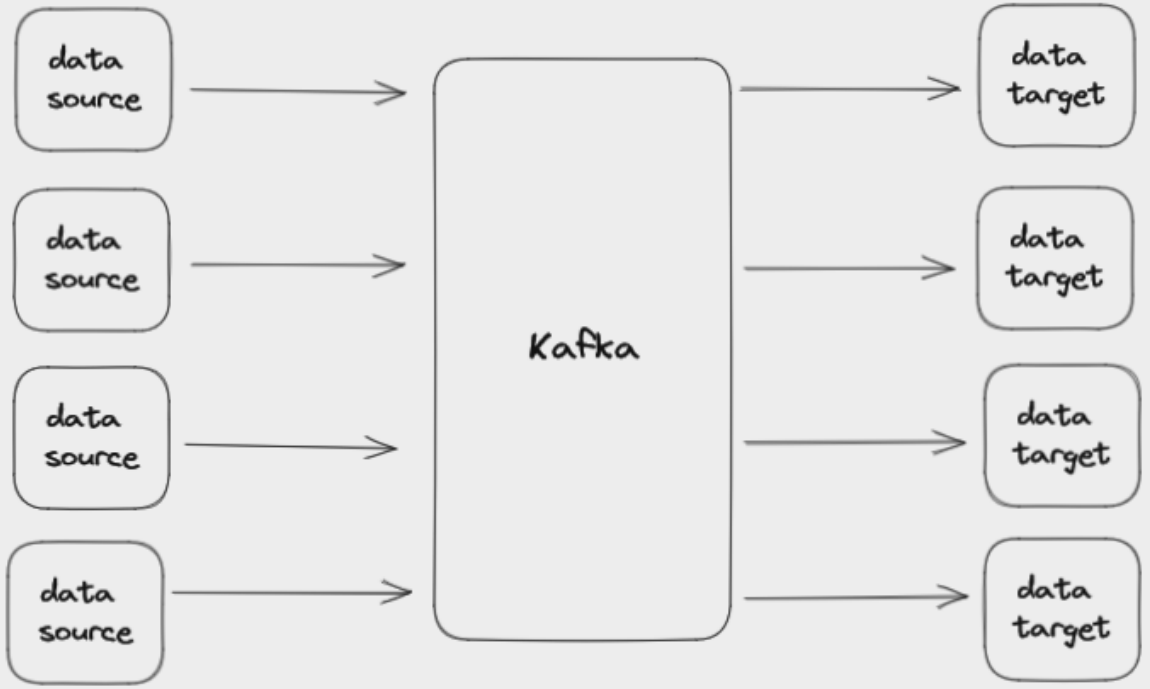


What is Kafka?



Why Kafka?


data source
data source
data source
data source
data target
data target
data target
data target

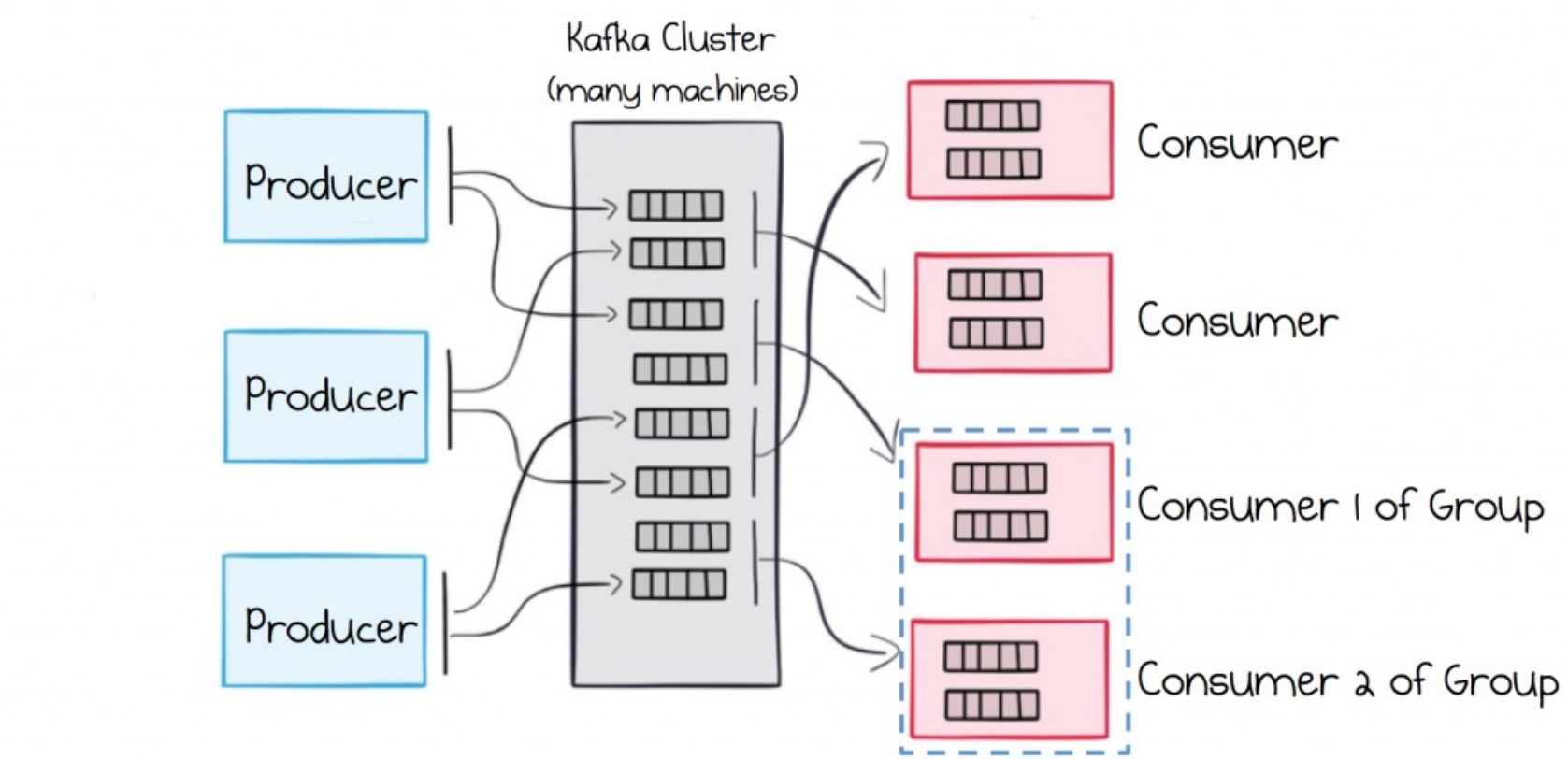

data source
data source
data source
data source
Kafka
data target
data target
data target
data target

Why Kafka?

Why Kafka?

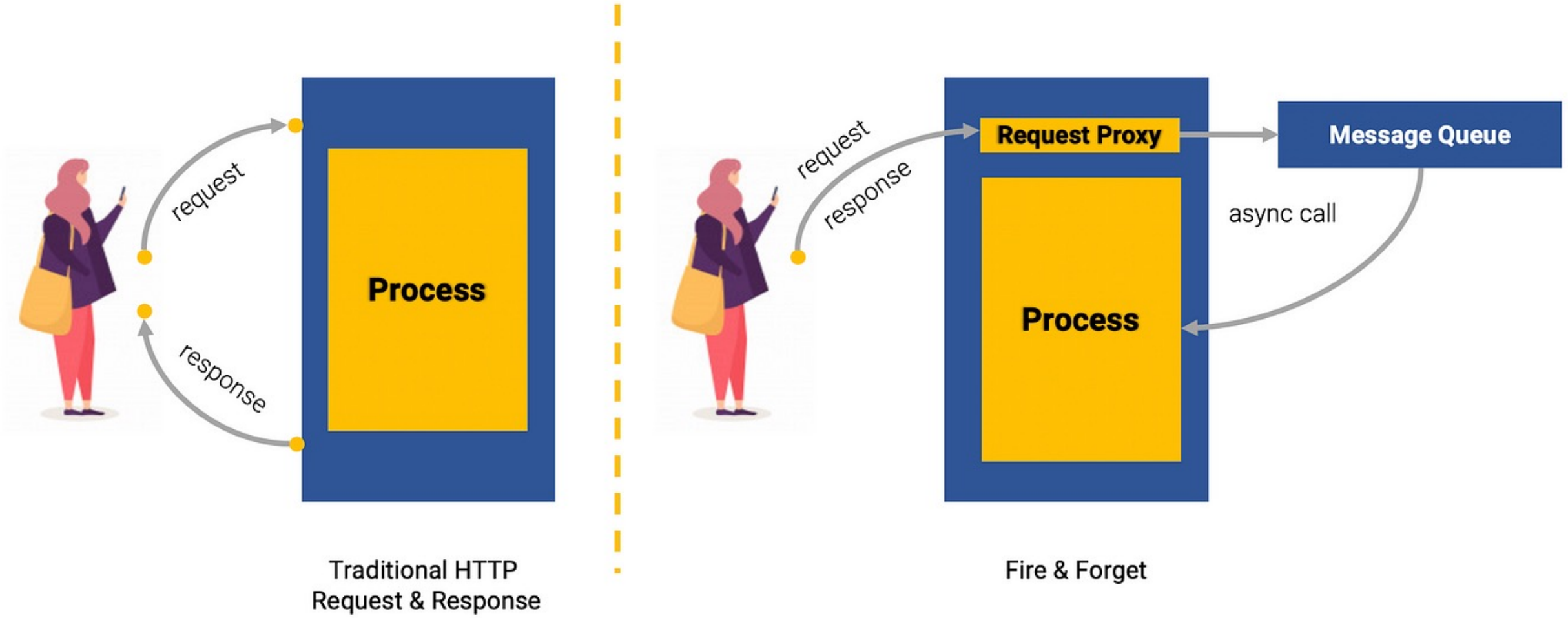

Top 10
Apache Kafka
Features
01
Performance
02
Replication
03
Durability
04
Zero
Downtime
05
Extensibility
06
Scalability
07
Fault
Tolerance
08
Reliability
09
High
Volume
10
Data
Transformations

Why Kafka?



Producers spread messages over many partitions, on many machines, where each partition is a little queue. Load balanced consumers (denoted a Consumer Group) share the partitions between them.

Why Kafka?


request
response
Process
Traditional HTTP
Request & Response
request
response
Request Proxy
Message Queue
async call
Process
Fire & Forget

Install Kafka on Ubuntu

1. Install Ubuntu: Ubuntu 20.04

2. Install JRE – Java Runtime

```
sudo apt update
sudo apt install default-jre
java --version
```

```
Output
openjdk version "11.0.11" 2021-04-20
OpenJDK Runtime Environment (build 11.0.11+9-Ubuntu-0ubuntu2.20.04)
OpenJDK 64-Bit Server VM (build 11.0.11+9-Ubuntu-0ubuntu2.20.04, mixed mode, sharing))
```

Install Kafka on Ubuntu

2. Install Kafka

- Download Kafka & Unzip
  - mkdir ~/Downloads
  - curl "https://downloads.apache.org/kafka/xxx/kafka_xxxx-xxx.tgz" -o ~/Downloads/kafka.tgz
  - mkdir ~/kafka && cd ~/kafka
  - tar -xvzf ~/Downloads/kafka.tgz --strip 1

- Change kafka config file
  - nano ~/kafka/config/server.properties
    - ✓ Add line “delete.topic.enable = true” at the EOF
    - ✓ Change log path: log.dirs=/home/thangnc/logs

Install Kafka on Ubuntu

2. Install Kafka

- Create systemd (Service) file for Zookeeper
  - Kafka uses Zookeeper to manage its cluster state and configurations
  - sudo nano /etc/systemd/system/zookeeper.service
  - Add file content

/etc/systemd/system/zookeeper.service

```
[Unit]
Requires=network.target remote-fs.target
After=network.target remote-fs.target

[Service]
Type=simple
User=kafka
ExecStart=/home/kafka/kafka/bin/zookeeper-server-start.sh /home/kafka/kafka/config/zookeep
ExecStop=/home/kafka/kafka/bin/zookeeper-server-stop.sh
Restart=on-abnormal

[Install]
WantedBy=multi-user.target
```

Install Kafka on Ubuntu

2. Install Kafka

- Create systemd (Service) file for Kafka
  - sudo nano /etc/systemd/system/kafka.service
  - Add file content

/etc/systemd/system/kafka.service

```
[Unit]
Requires=zookeeper.service
After=zookeeper.service

[Service]
Type=simple
User=kafka
ExecStart=/bin/sh -c '/home/kafka/kafka/bin/kafka-server-start.sh /home/kafka/kafka/config
ExecStop=/home/kafka/kafka/bin/kafka-server-stop.sh
Restart=on-abnormal

[Install]
WantedBy=multi-user.target
```

Install Kafka on Ubuntu

2. Install Kafka

- Start Kafka
  - sudo systemctl start kafka
  - sudo systemctl status kafka

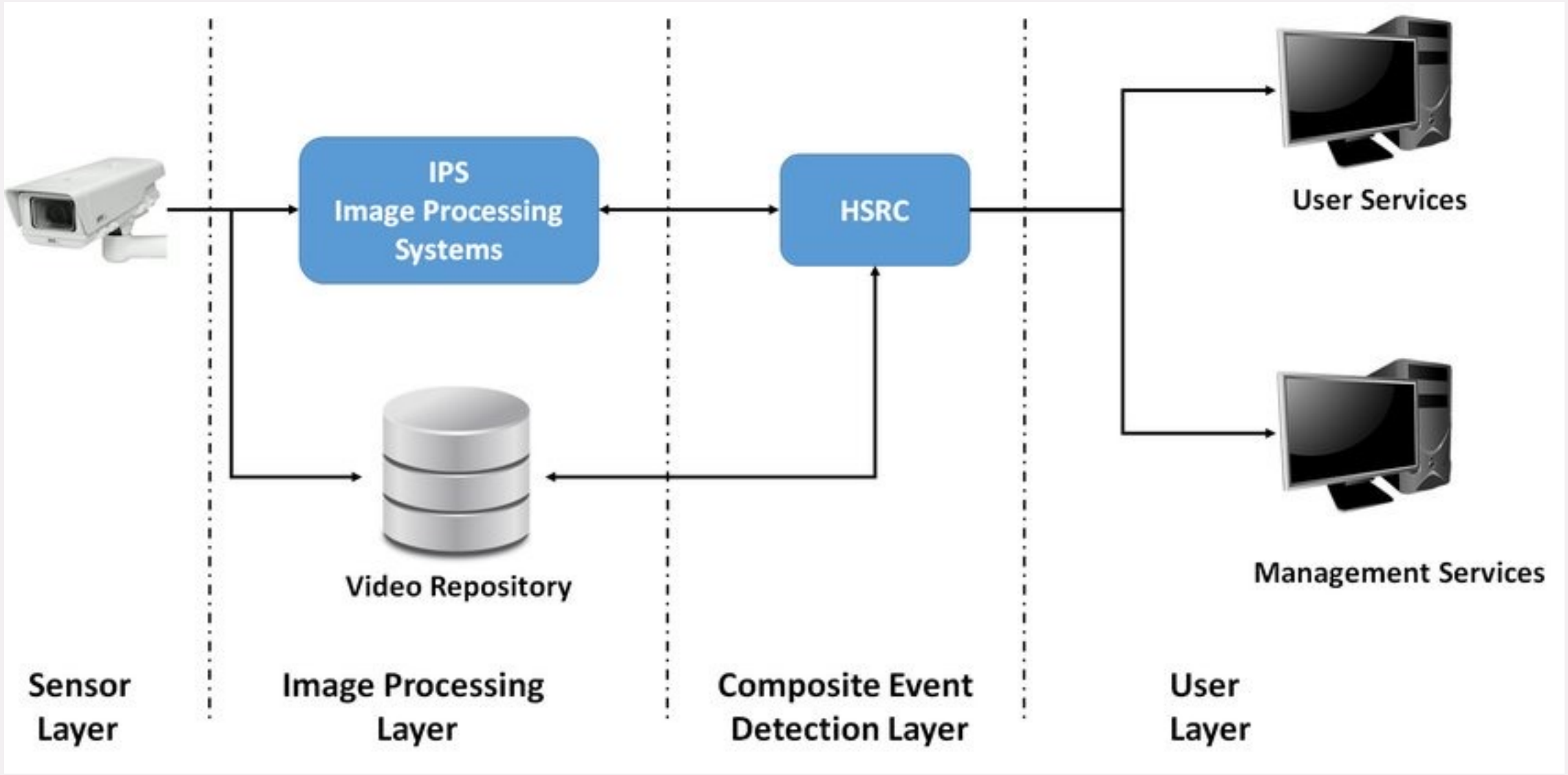
```
Output
● kafka.service
     Loaded: loaded (/etc/systemd/system/kafka.service; disabled; vendor preset>
     Active: active (running) since Wed 2023-02-01 23:44:12 UTC; 4s ago
   Main PID: 17770 (sh)
      Tasks: 69 (limit: 4677)
     Memory: 321.9M
     CGroup: /system.slice/kafka.service
             ├─17770 /bin/sh -c /home/kafka/kafka/bin/kafka-server-start.sh /ho>
             └─17793 java -Xmx1G -Xms1G -server -XX:+UseG1GC -XX:MaxGCPauseMill>
```

Install Kafka on Ubuntu

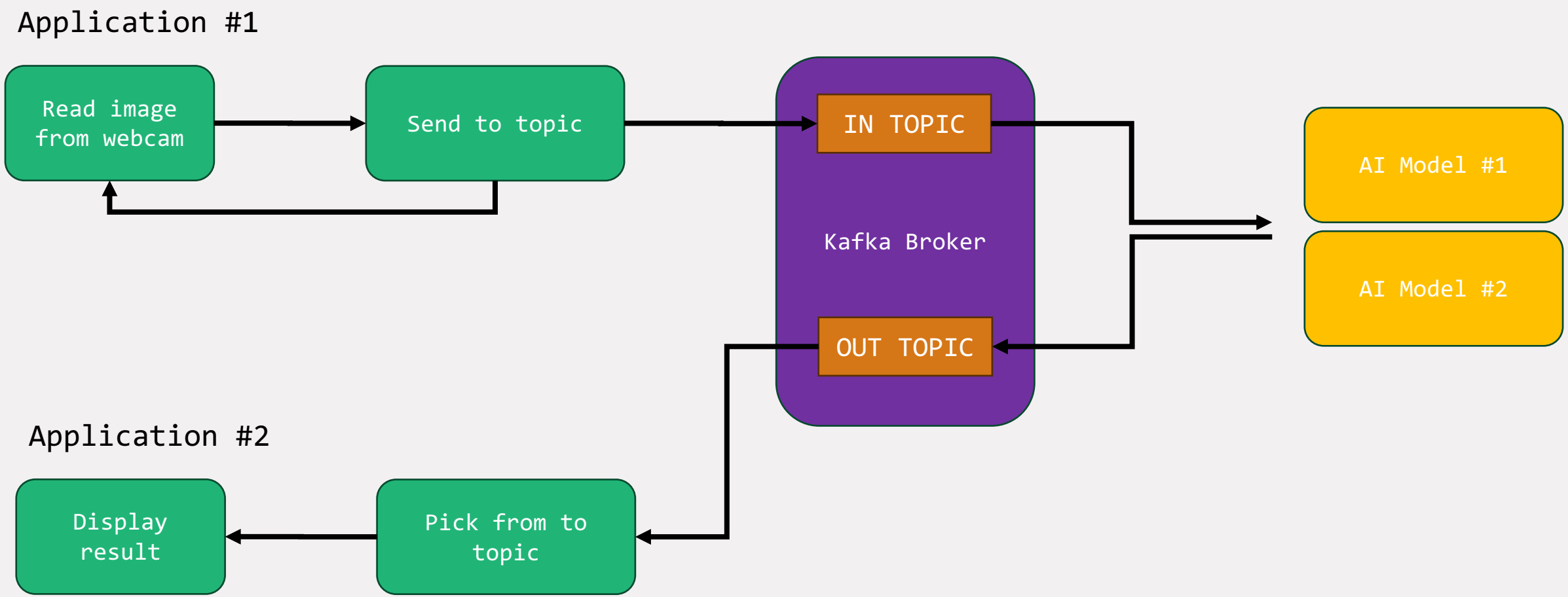
2. Install Kafka

- Enable kafka to run at boot
  - sudo systemctl enable zookeeper
  - sudo systemctl enable kafka
- Test Kafka
  - Create topic:
    - ~/kafka/bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic **miai2**
  - Publish test massage
    - echo "Hello, World" | ~/kafka/bin/kafka-console-producer.sh --broker-list localhost:9092 --topic **miai2** > /dev/null
  - Consume test massage:
    - ~/kafka/bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic **miai2** --from-beginning
- Turn on listener trong Kafka config (LAST BUT NOT LEAST)

Pipeline


IPS
Image Processing
Systems
HSRC
User Services
Video Repository
Management Services
Sensor
Layer
Image Processing
Layer
Composite Event
Detection Layer
User
Layer

Pipeline


Application #1
Read image from webcam
Send to topic
IN TOPIC
Kafka Broker
OUT TOPIC
AI Model #1
AI Model #2
Application #2
Display result
Pick from to topic

Handson