

Thu thập và phân tích hoạt động học tập của người học trên các Learning Management System

April , 2022
~~~~~

GVHD: TS. Nguyễn Quang Hùng

---

Sinh viên thực hiện:

Nguyễn Duy Khang - 2011364

Phan Phước Minh - 2010418

Nguyễn Đức An - 2010102

Đoàn Trần Cao Trí - 2010733



# Giới thiệu đề tài



CANVAS



Nền tảng hệ thống quản lý học tập như  
Moodle, Google Classroom, Canvas



Kafka, Spark

# Ý nghĩa đề tài



Code 1 lần, chạy nhiều nơi

Có thể phân tích và đưa ra kết quả từ các bộ dữ liệu thu thập được từ các LMS khác nhau theo thời gian thực.



Dẫn đầu trong thời đại Big Data

Tốc độ xử lý và phản hồi thông tin, cho ra kết quả nhanh.

# Yêu cầu đề tài

01

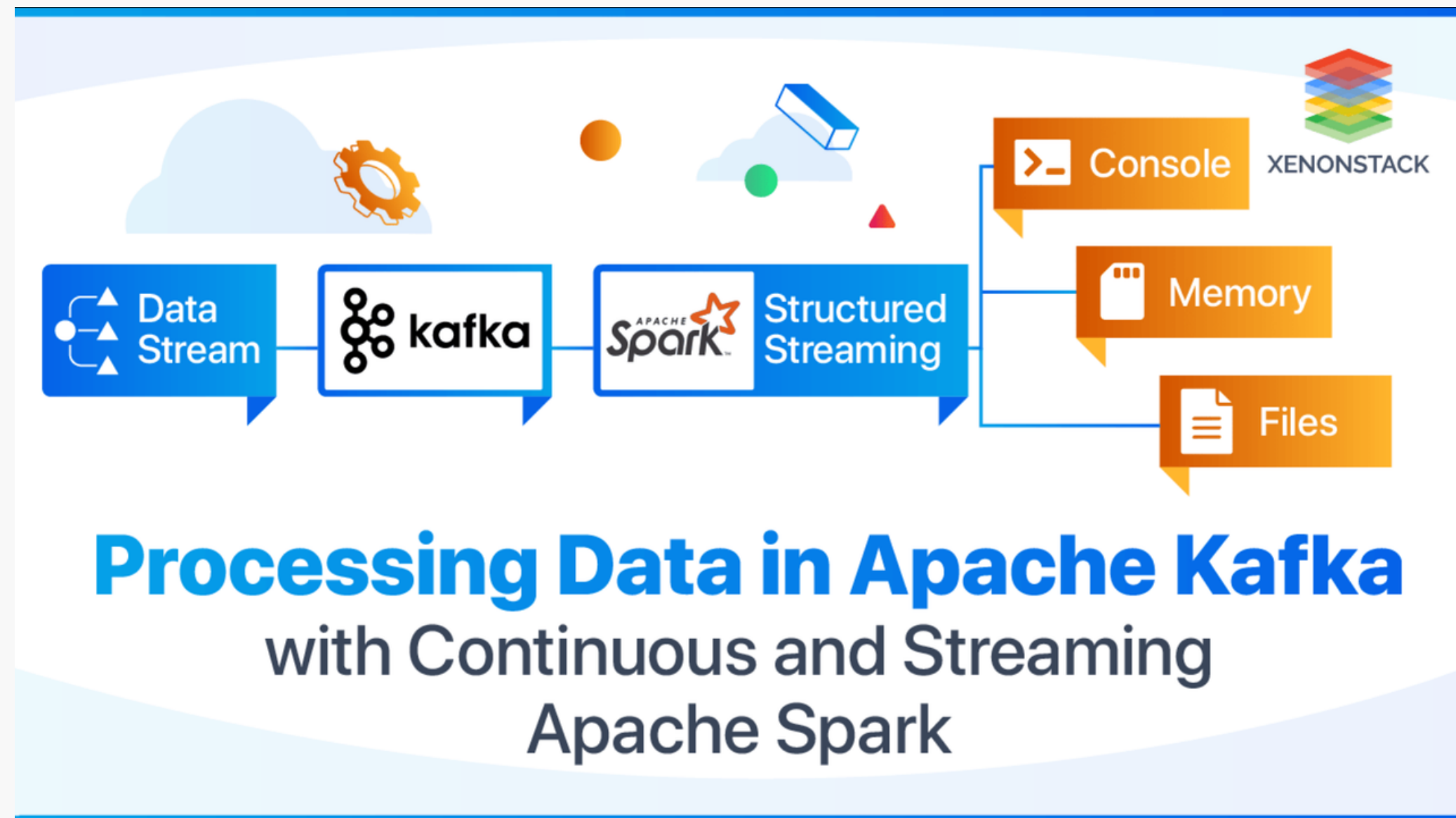
Big Data  
Machine Learning

02

Tìm hiểu về LMS như  
Moodle

03

Tìm hiểu về Kafka, Spark,  
Data Lake



- 1

Đưa ra những kết quả ý nghĩa dựa vào dữ liệu thu thập được

- 2

Thiết lập một hệ thống để thu thập dữ liệu từ Moodle và xử lý, đưa ra kết quả phân tích

# Kết quả đầu ra





# Learning Management Systems

giúp hỗ trợ việc quản lý, giám sát người học



LMS phổ biến nhất, được viết bằng PHP


# Đặc điểm

Mã nguồn mở


Phổ biến ở các trường ĐHQG TP HCM


Cho phép mở rộng thông qua các plugins





 ☒ Assignment

 ☐ Feedback

 ☐ Forum

 ☒ Quiz

 ☒ File

 ☒ URL

# Modules

# Những thông tin được thu thập

1

Sinh viên click xem trang khóa học, module khóa học

2

Sinh viên click làm quiz, điểm quiz

|    | A               | B              | C             | D                | E         | F                    | G               | H      | I            | J |
|----|-----------------|----------------|---------------|------------------|-----------|----------------------|-----------------|--------|--------------|---|
| 1  | Time            | User full name | Affected user | Event content    | Component | Event name           | Description     | Origin | IP address   |   |
| 2  | 17/10/21, 17:00 | -              | -             | Course: O System | Logs      | Log report viewed    | The user viewed | web    | 10.7.34.59   |   |
| 3  | 17/10/21, 17:00 | -              | -             | Course: O System | System    | Course viewed        | The user viewed | web    | 10.7.34.59   |   |
| 4  | 19/08/20, 19:00 | -              | -             | Course: O System | System    | Course viewed        | The user viewed | web    | 192.168.99.3 |   |
| 5  | 6/08/20, 20:00  | -              | -             | Course: O System | System    | Course viewed        | The user viewed | web    | 192.168.99.3 |   |
| 6  | 4/08/20, 12:00  | -              | -             | File: MIG        | File      | Course module viewed | The user viewed | web    | 192.168.99.3 |   |
| 7  | 4/08/20, 12:00  | -              | -             | Course: O System | System    | Course viewed        | The user viewed | web    | 192.168.99.3 |   |
| 8  | 6/07/20, 02:00  | -              | -             | File: Lab E      | File      | Course module viewed | The user viewed | web    | 192.168.99.3 |   |
| 9  | 6/07/20, 02:00  | -              | -             | File: Week       | File      | Course module viewed | The user viewed | web    | 192.168.99.3 |   |
| 10 | 6/07/20, 02:00  | -              | -             | File: Week       | File      | Course module viewed | The user viewed | web    | 192.168.99.3 |   |
| 11 | 6/07/20, 02:00  | -              | -             | Course: O System | System    | Course viewed        | The user viewed | web    | 192.168.99.3 |   |
| 12 | 5/07/20, 18:00  | -              | -             | File: Week       | File      | Course module viewed | The user viewed | web    | 192.168.99.3 |   |

# Một số ứng dụng

1

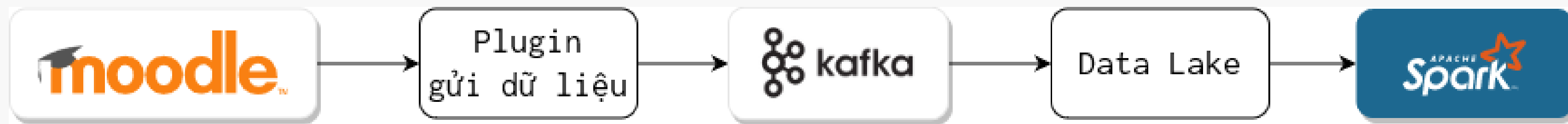
Cá nhân hóa trang  
Moodle của học viên

2

Dự đoán kết quả  
đậu/rớt của sinh viên

3

Dự đoán điểm  
của sinh viên

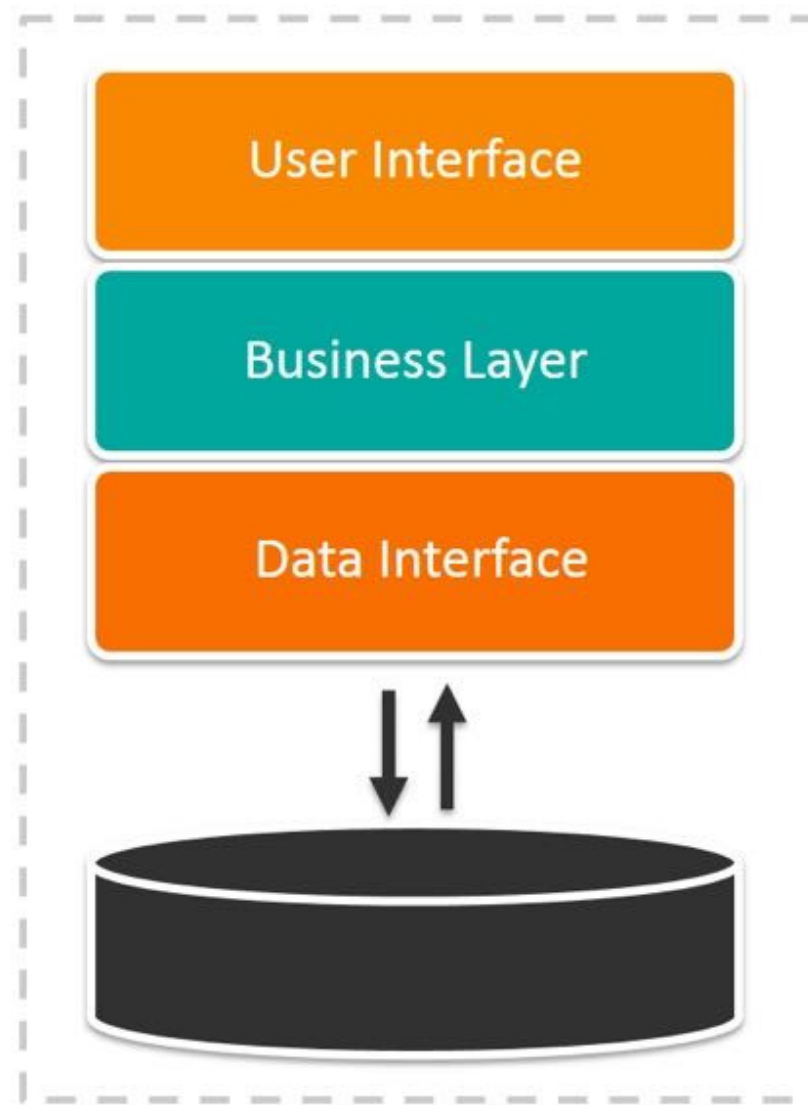


# Kafka

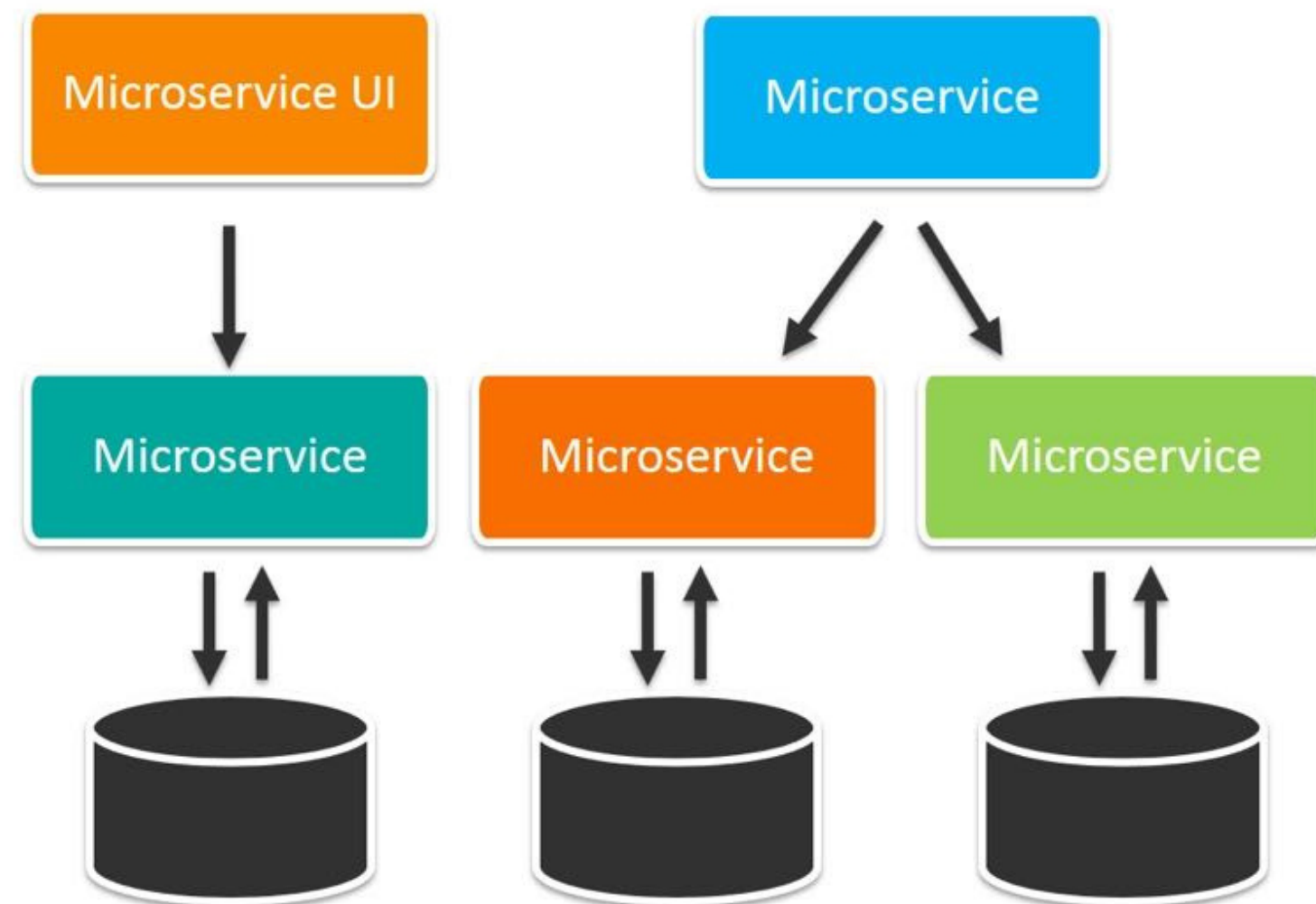


# Lịch sử ra đời

**Monolithic Architecture**



**Microservices Architecture**



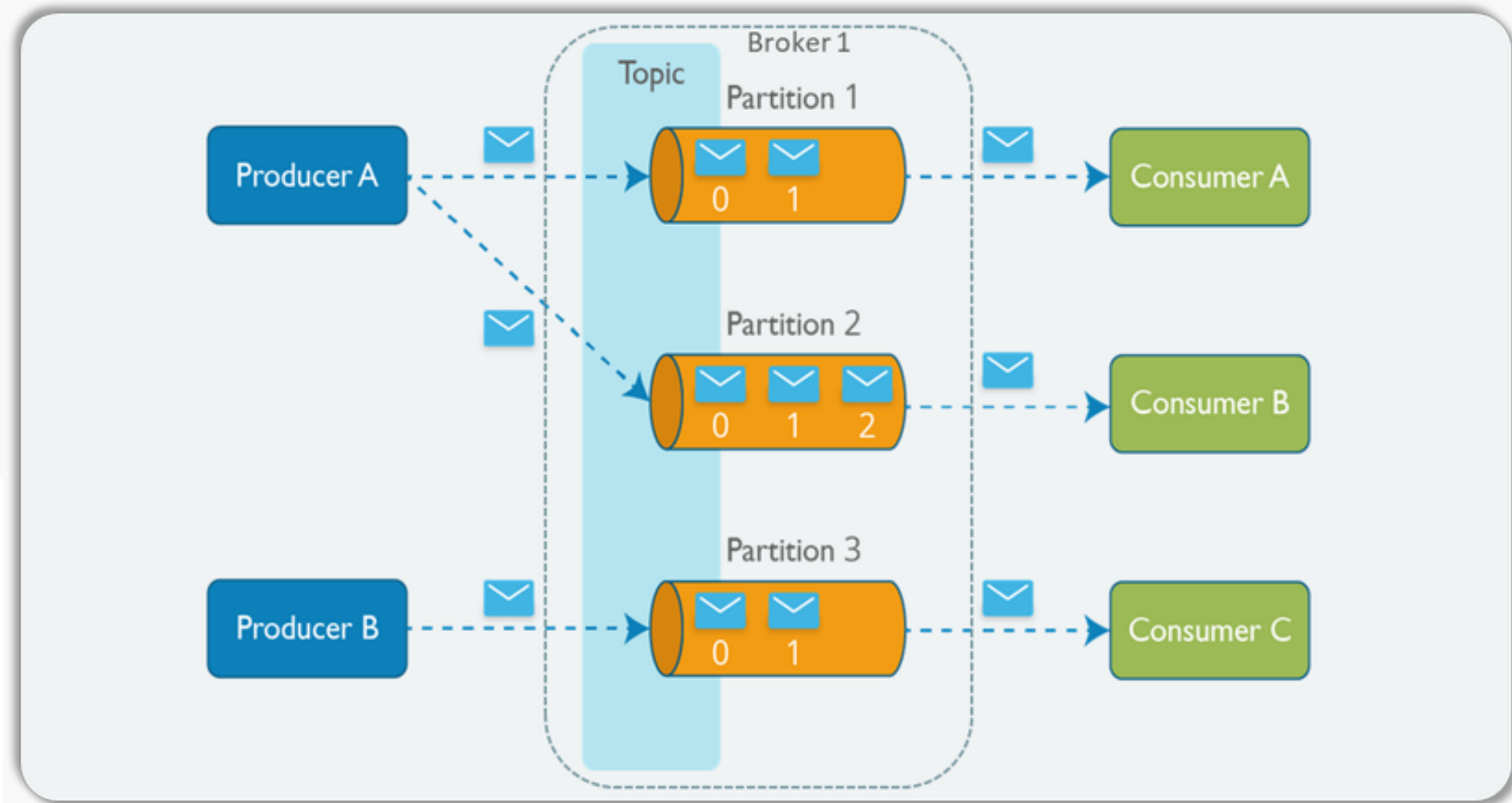
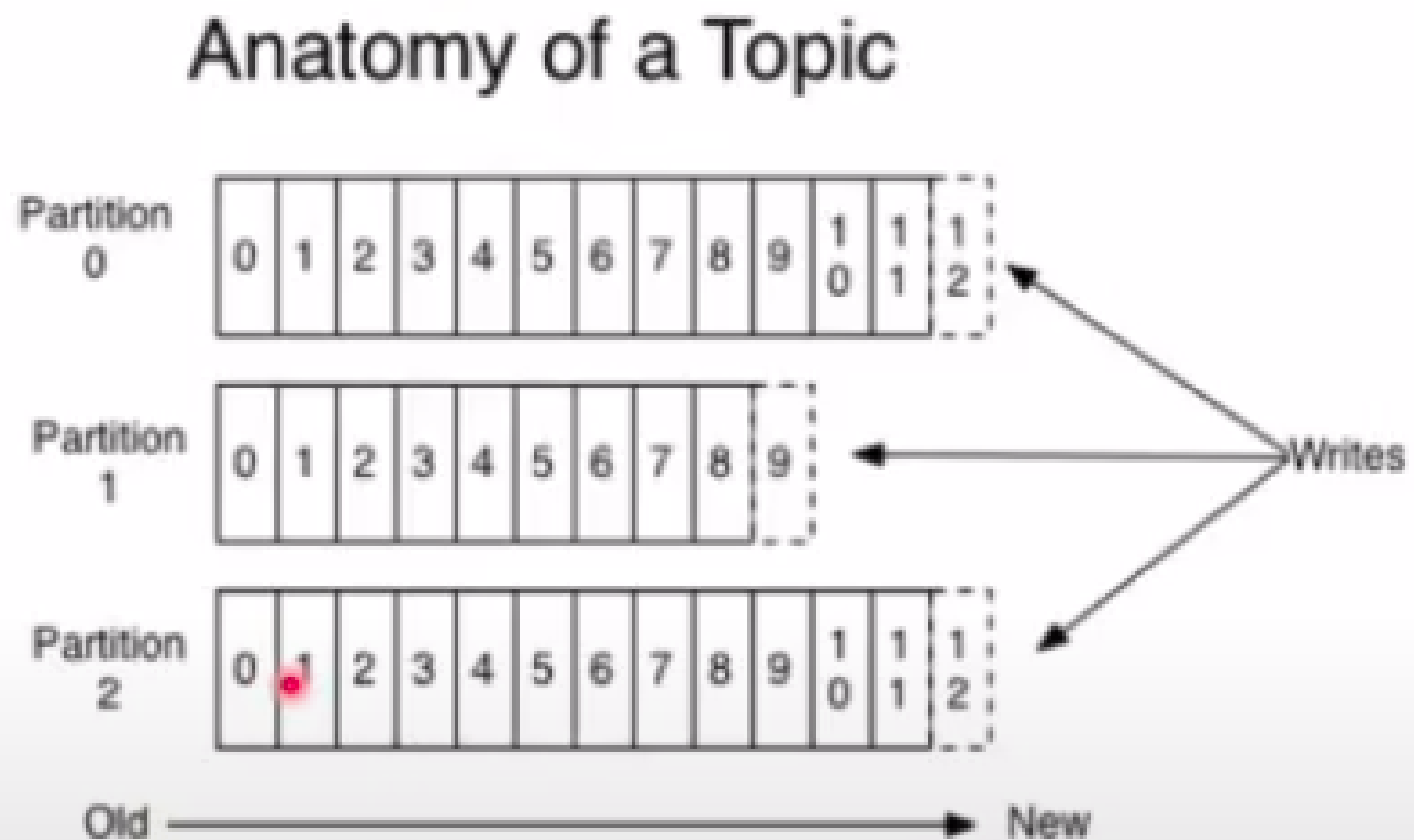
# Lịch sử ra đời

Được phát triển như một phương thức để giao tiếp

Được tạo ra bởi



# Một số thành phần của Kafka



Hiện thực theo cơ chế Pub-Sub



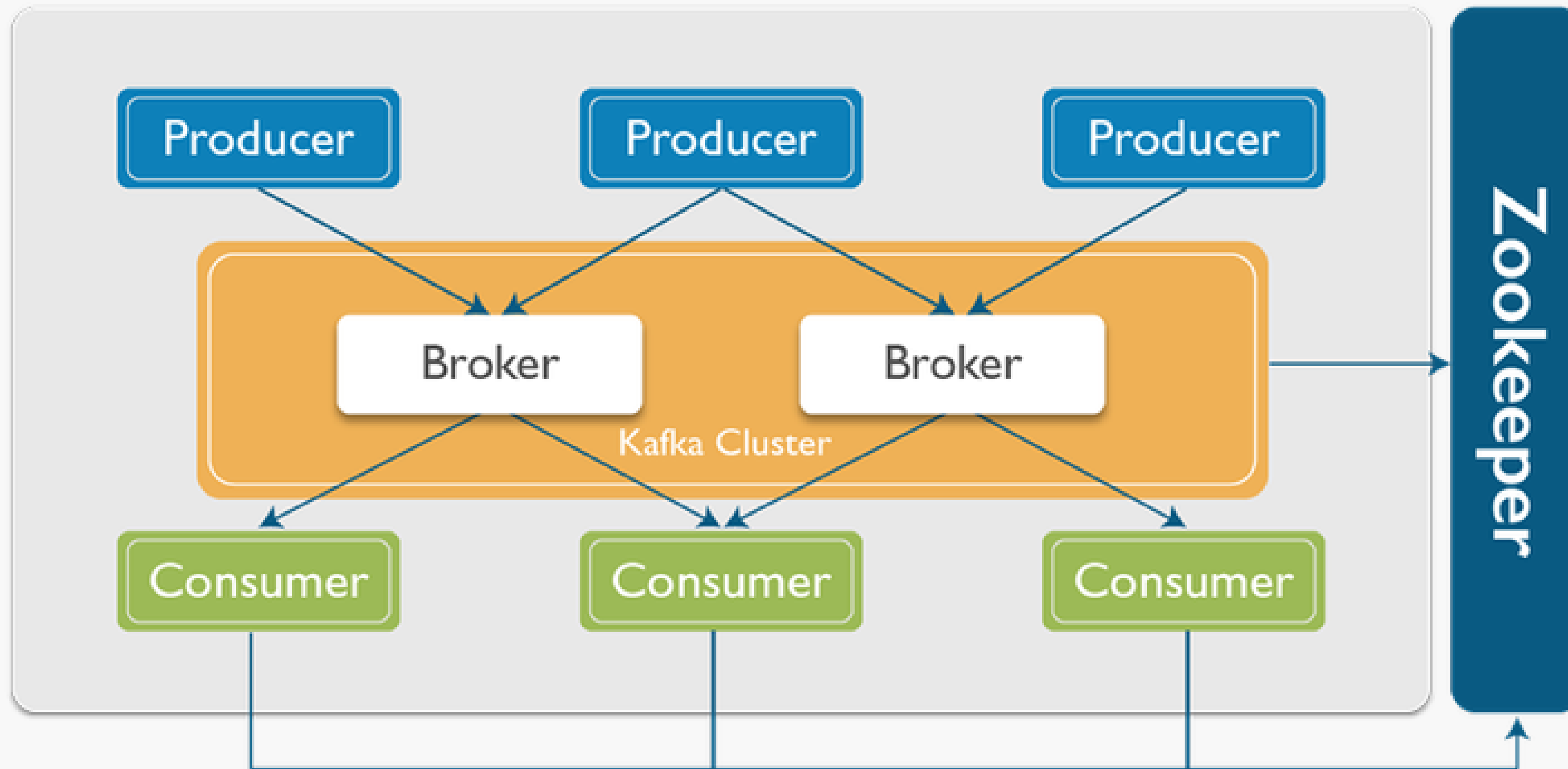
# Ưu điểm

- Xử lý thời gian thực
- Dữ liệu phân tán (distributed)
- Khả năng phục hồi (resilient)
- Khả năng chịu lỗi (fault tolerant)
- Khả năng mở rộng (scalable)

# Ứng dụng

- Vận hành hệ thống
- Phân tích dữ liệu
- Công nghệ phần mềm

# Zookeeper



# Zookeeper

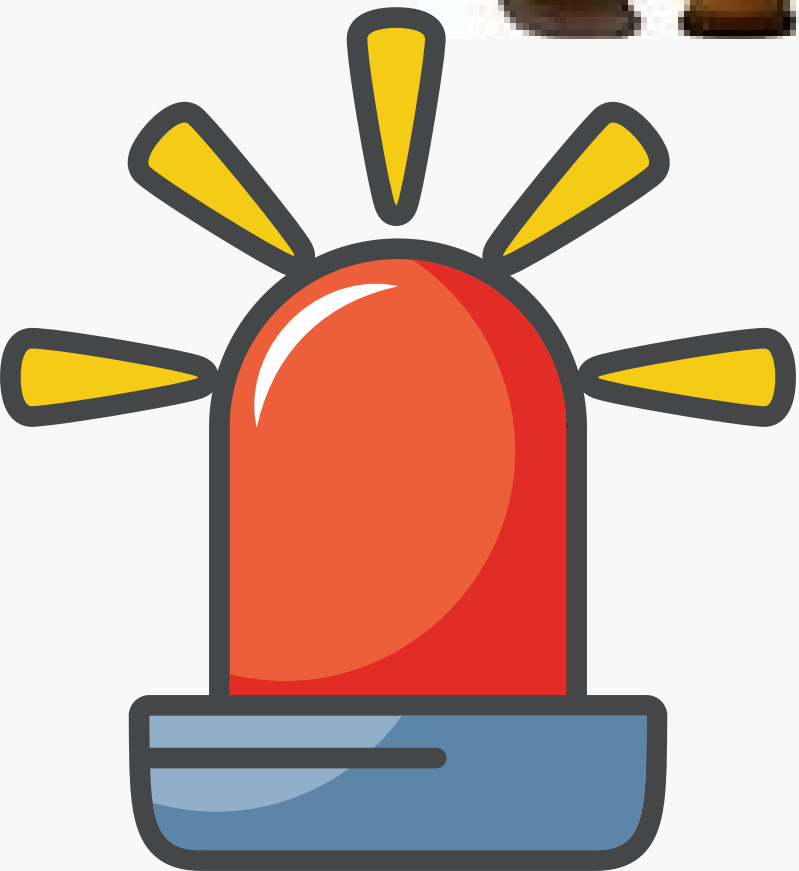
Zookeeper theo dõi, lưu trữ trạng thái của Kafka brokers, topics, partitions,...



# Zookeeper



Apache ZooKeeper



Kafka không thể sử dụng nếu không có Zookeeper.

# Kafka Connect

Thứ nhất, data pipeline với Apache Kafka là một trong hai endpoint. Ví dụ: lấy dữ liệu từ Kafka đẩy vào S3(Amazon Simple Storage Service) , hoặc lấy dữ liệu từ Database đẩy vào Kafka

Thứ hai, Sử dụng kafka để xây dựng một pipeline giữa 2 hệ thống riêng biệt. Ví dụ: lấy data từ Twitter đưa vào hệ thống Elasticsearch bao gồm việc lấy data từ Twitter đẩy vào Kafka sau đó từ kafka đưa vào Elasticsearch

Định nghĩa, ưu điểm, tính năng nổi bật

~~~~ Kafka

+

~~~~

# Timeliness.



## Producer

Producer có thể ghi dữ liệu vào Kafka định kỳ hoặc bất cứ khi nào cần.



## Consumer

Consumer có thể đọc dữ liệu mới nhất ngay khi nó được ghi vào, cũng có thể định kỳ đọc dữ liệu hàng giờ hoặc vài giờ một lần



# Reliability

Đối với các hệ thống lớn, việc mất mát dữ liệu, trùng lặp dữ liệu có thể gây ra hậu quả lớn. Kafka đảm bảo đảm bảo yêu cầu "exactly-once" cho mỗi sự kiện được đưa vào hệ thống

# Data Formats

Việc lựa chọn định dữ liệu tùy thuộc vào yêu cầu đặt ra của mỗi hệ thống, nghiệp vụ khác nhau. Kafka hỗ trợ rất nhiều: XML, text, JSON, Avro, CSV, ... Producer và Consumer có thể sử dụng bất kỳ bộ serializer nào để biểu diễn dữ liệu dưới bất kỳ định dạng nào bạn muốn



- High and Varying Throughput


Hoạt động như một bộ nhớ đệm cực lớn giữa producers và consumer, Kafka còn cung cấp một cơ chế backpressure giúp cho dữ liệu được giữ ở trong cụm cho đến khi consumer kịp xử lý.


- Security

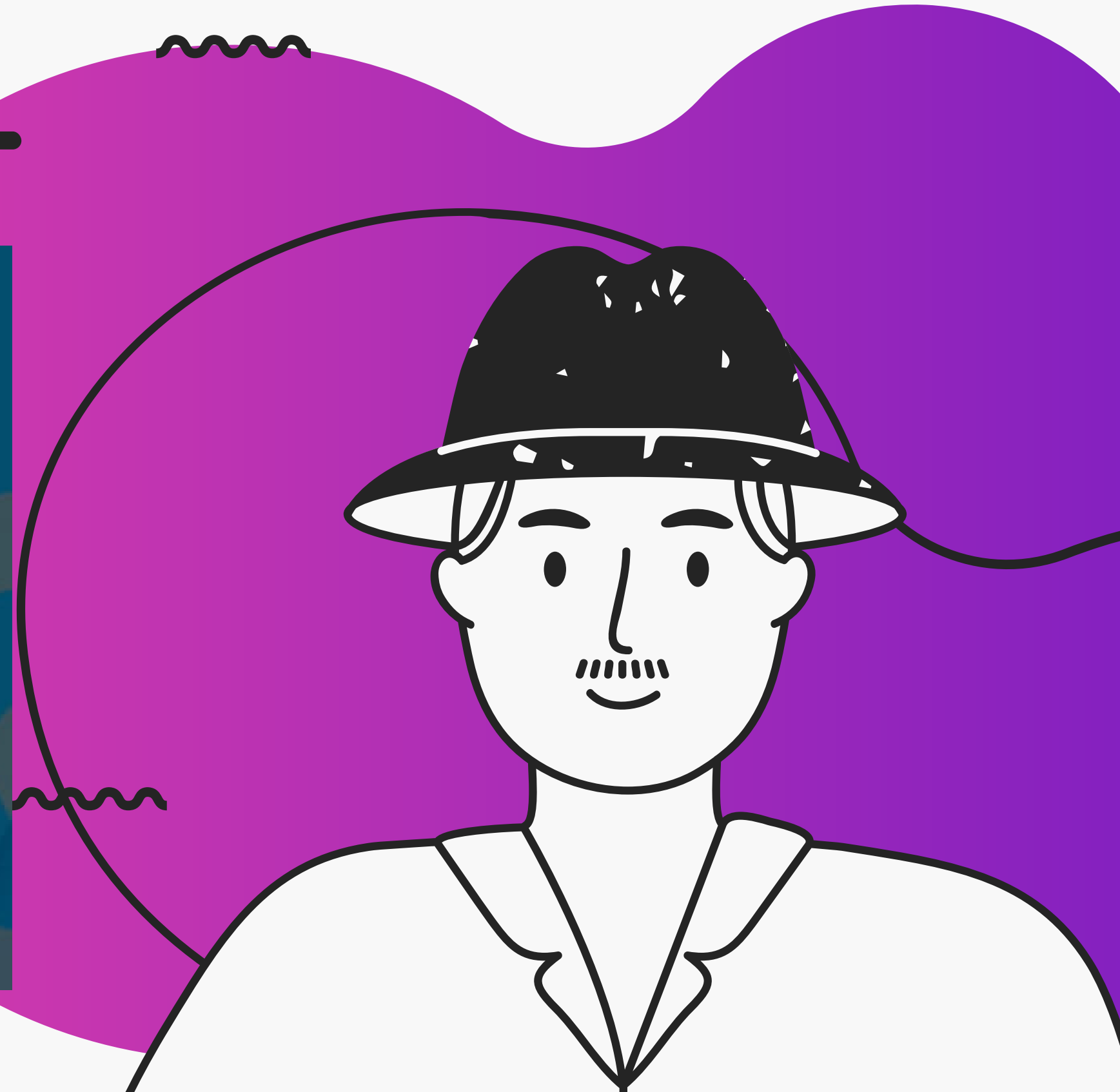
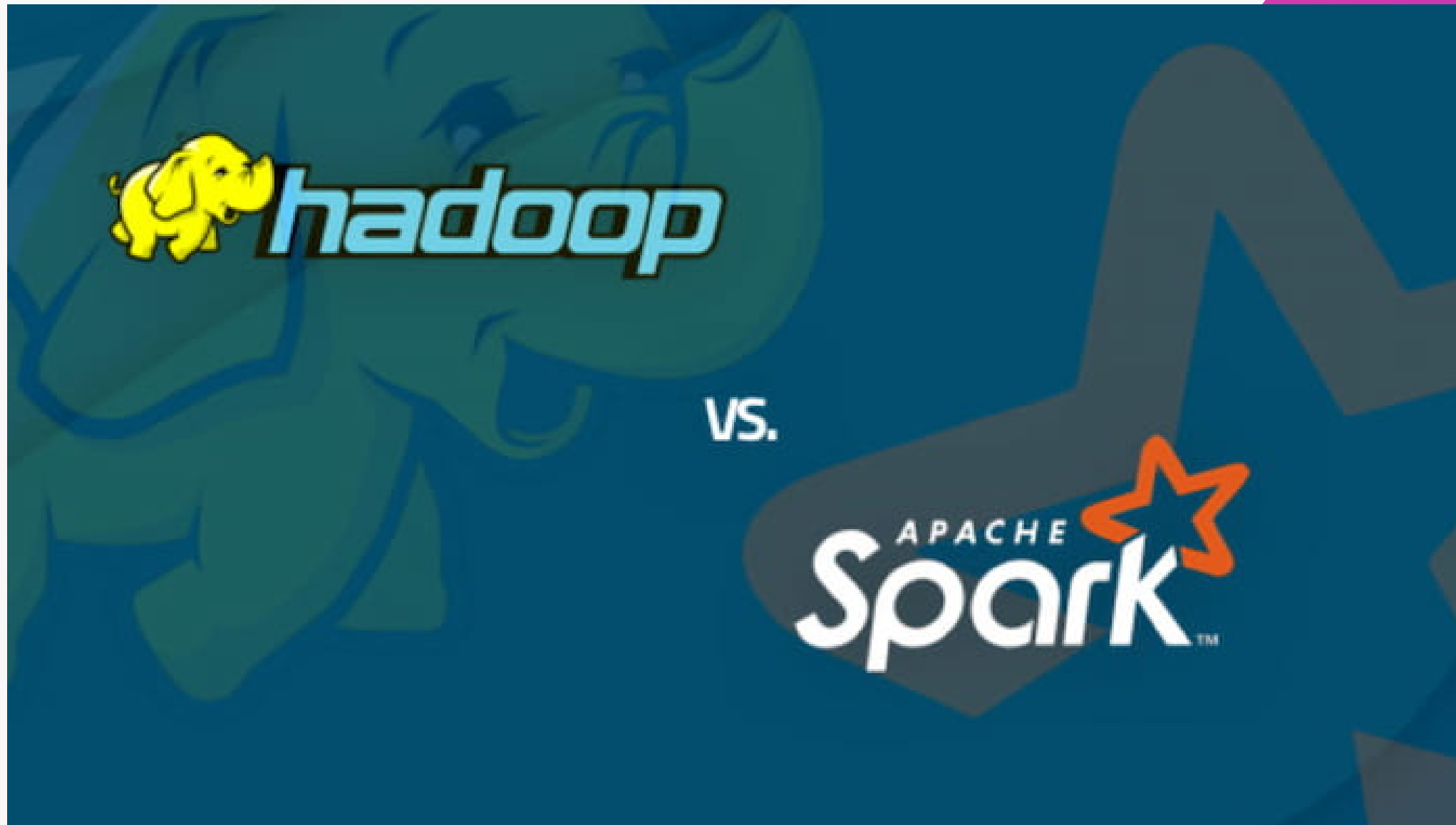
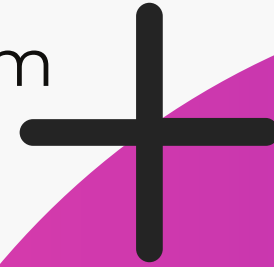
Bạn có thể chắc chắn rằng dữ liệu của bạn khi đi qua Kafka có thể được mã hóa nếu bạn muốn vì bạn chủ động trong ghi dữ liệu. Kafka cũng hỗ trợ cơ chế xác thực (SASL), do đó bạn có thể yên tâm rằng những người có quyền đọc mới được đọc topic của bạn.



# Apache Spark

Được tạo ra bởi Matei  
Zaharia tại UC Berkeley's  
AMPLab 

Được tạo vào năm 2009  
và mở mã nguồn vào năm  
2010 



# Đặc điểm của Spark



Tốc độ xử lý cao:

Xử lý mọi thứ trên RAM



Xử lý dữ liệu:

Spark xử lý dữ liệu theo lô và thời gian thực.



Tính tương thích:

Có thể tích hợp với tất cả các nguồn dữ liệu và định dạng tệp được hỗ trợ bởi cụm Hadoop.



Hỗ trợ ngôn ngữ:

Java, Scala, Python và R.

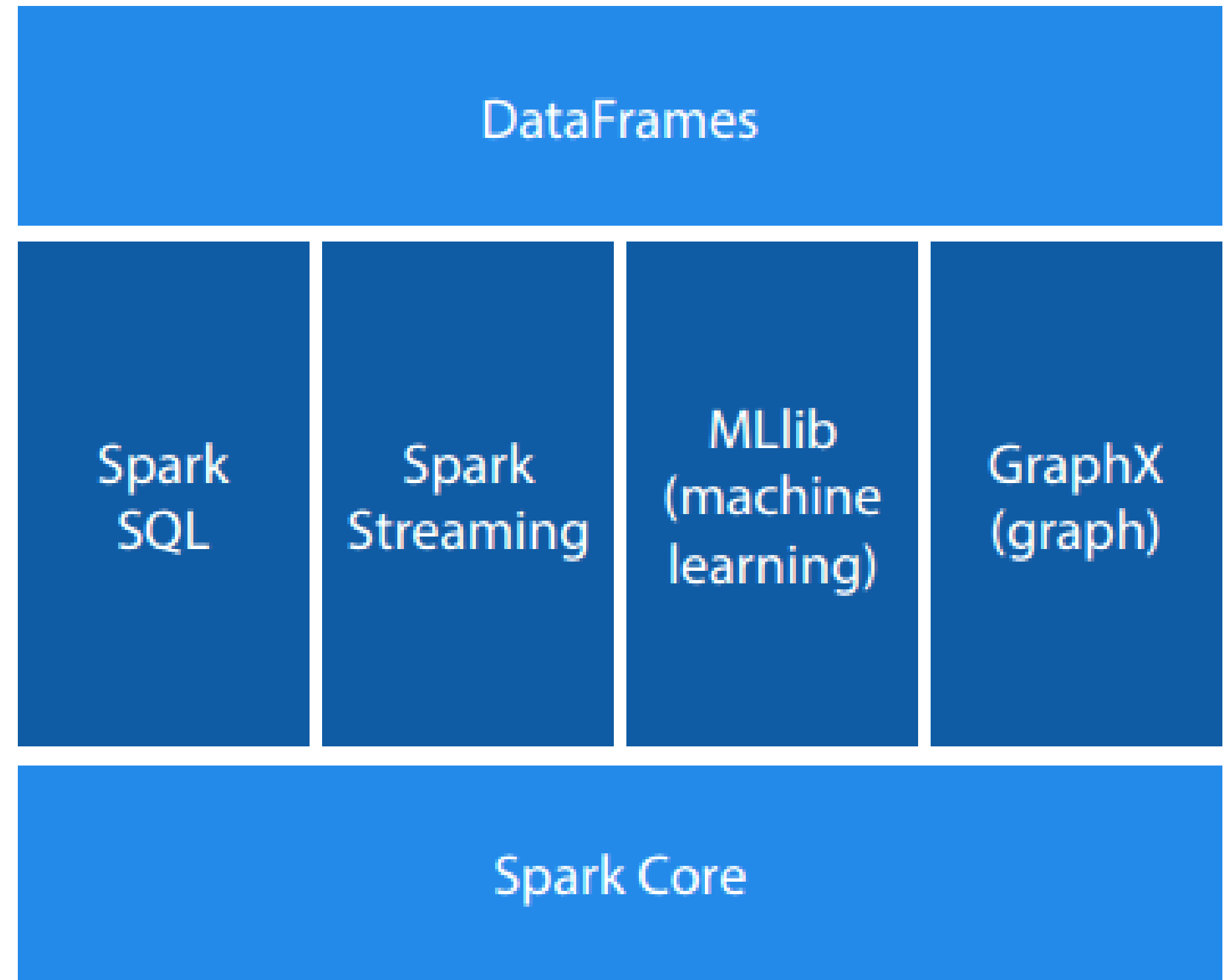


Xử lý thời gian thực:

Apache Spark có thể xử lý các luồng sự kiện dữ liệu thời gian thực với tốc độ hàng triệu sự kiện mỗi giây.

# PySpark

Apache Spark gồm có 5 thành phần chính : Spark Core, Spark Streaming, Spark SQL, MLlib và GraphX





# Các tính năng cơ bản

**SQL**

Truy xuất dữ liệu dataset

**Core**

API các ngôn ngữ hỗ trợ  
Java, Scala, Python

**GraphX**

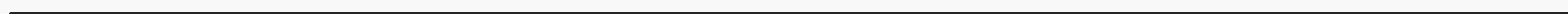
Hỗ trợ tính toán song song

**Streaming**

Nhận biết sự kiện và lấy dữ liệu trong thời gian thực.

**MLlib**

Hỗ trợ xây dựng model máy học



# Ý tưởng

- Huấn luyện model với dataset của các khóa học trước (supervised machine learning)
- Áp dụng các thuật toán clustering (phân cụm các nhóm học sinh dựa vào điểm số, ..)
- Áp dụng các thuật toán học máy như regression, tree decision để dự đoán kết quả học tập (hoàn thành / không hoàn thành) khóa học.
- Thông báo và hỗ trợ sinh viên những việc cần thiết để hoàn thành tốt khóa học.



# Tài liệu tham khảo

- Apache spark™ - unified engine for large-scale data analytics. <https://spark.apache.org/>
- Moodle - open-source learning platform. <https://moodle.org/>.



XIN CẢM ƠN  
MỌI NGƯỜI  
ĐÃ THEO DÕI

*thank  
you*

