Kafka

Hoạt động như 1 hệ thống ghi log phân tán đọc ghi theo chủ đề

Sử dụng với dòng dữ liệu lớn

Đảm bảo độ tin cậy, hiệu suất cao

Kiến trúc:

Producer: thành phần gửi dữ liệu tới kafka

Consumer: Thành phần nhận và xử lý dữ liệu

Broker: lưu trữ và quản lý dữ liệu

Topic: kênh phân loại message

Partition**:** Dữ liệu trong topic được chia nhỏ thành nhiều phân vùng

Zookeeper**:** theo dõi trạng thái của broker và topic

Tác dụng:

Tốc độ và hiệu suất cao.

Tính phân tán

dữ liệu có thể được sao chép

Sử dụng khi:

Xử lý dữ liệu theo thời gian thực

Ghi log

Truyền tải dữ liệu

Hệ thống giám sát

Nhược điểm:

Cấu hình phức tạp

Khó kiểm soát

Khái niệm: là mã nguồn mở cung cấp cơ chế xử lý dữ liệu dạng stream hàng đợi

Giao tiếp giữa các hệ thống phân tán nhưng nơi gửi và nơi tiêu thu không biết nhau(consumer, produce).

Consumer :nhận dữ liệu từ produce

Key word:

Broker: có thể có nh broker tạo thành cluster: trong cluster dữ liệu được chia sẻ với nhau. Trong broker có nh topic

Topic: là event lưu thông tin. Trong 1 topic có ít nhất 1 partition, nó không có cơ chế lấy lại dữ liệu đã đọc. khi có nh message được gửi đến nó sẽ chia cho các partition

Partition:mỗi partition có 1 offset riêng. mỗi bản ghi đại diện cho 1 offset

Offset:

Consummer group:

1 partition chỉ tương tác với 1 consumer nên số lượng consummer trong 1 topic phải nhỏ hơn số lượng partition;