Kafka là 1 hệ thống messaging queue phân tán, cho phép gửi và nhận (publish and subcribe) các bản ghi theo luồng.

Kafka có những tính năng nổi bật như:

* **Khả năng mở rộng**: Có nhiều khía cạnh trong khả năng mở rộng của Kafka, có thể kể đến như việc hỗ trợ mở rộng cụm Kafka theo chiều ngang (tương tự như cách mà Hdfs mở rộng), hay mở rộng các thành phần tương tác với cụm như producer, consumer… Kafka được thiết kế để mở rộng rất tốt.
* **Tính chịu tải và sự ổn định**: Kafka có thể làm việc với 1 lượng rất lớn dữ liệu, nhờ cơ chế mở rộng và partitioning, một hệ thống chỉ 7–8 nodes có thể xử lý tới hàng chục tỉ bản ghi mỗi ngày với tính ổn định cao.
* **Tin cậy dữ liệu**: Dữ liệu đẩy vào Kafka được lưu trữ phân tán, có replication, và được flush xuống ổ cứng trên từng node cùng với cơ chế quản lý từng bản ghi thông quan offset nên việc mất dữ liệu hay xử lý thiếu, sót, trùng lặp là rất khó xảy ra. Ngay cả khi có một hoặc vài node trên cụm bị dừng hoạt động, các ứng dụng tương tác với Kafka đều có thể hoạt động bình thường.
* **Zero-downtime**: Bất kì hệ thống nào kể cả phân tán hay không thì điều này đều cực kì quan trọng. Một hệ thống Kafka được cấu hình tốt có thể vận hành với thời gian chết (downtime) = 0. Mọi hoạt động từ việc thêm nguồn dữ liệu, mở rộng cụm hay thậm chí nâng cấp cấu hình đều có thể thực hiện mà không làm ảnh hưởng tới dịch vụ.

Real use-cases

Image for post



* Phát hiện gian lận hoặc bất thường trong giao dịch thẻ tín dụng: Thông tin giao dịch của khách hàng được xử lý thông qua các mô hình Machine — Learning để phát hiện bất thường hoặc gian lận, từ đó dừng các giao dịch nghi ngờ.
* Hệ thống xe tự lái: Các thông tin từ cảm biến được xử lý thời gian thực để đưa ra hành vi lái xe.
* Gợi ý sản phẩm dựa trên hành vi khách hàng: Các thao tác của người dùng trên hệ thống được ghi nhận thành các sự kiện và xử lý realtime, sau đó phân tích và đưa ra các sản phẩm, dịch vụ tương ứng với hành vi của khách hàng (Netflix đang sử dụng mô hình này để đưa ra recommend phim thời gian thực dựa trên tìm kiếm và hành vi xem nội dung).