

Apache Kafka

Damon Lopes e François Bruno

Sistema Distribuído

Um sistema distribuído é uma coletânea de softwares localizados em diferentes lugares trabalhando para atingir um mesmo objetivo. O sistema possui o foco na eliminação de gargalos e pontos centrais de falha geral.

Apache Kafka: o que é

O Apache Kafka é uma plataforma open-source de sistema distribuído que pode ser implementada tanto fisicamente como na nuvem. Baseia-se na comunicação entre servidores e clientes via uma comunicação TCP personalizada para alta performance.

- clusters de servidores/brokers
- plataforma de event streaming
- pub/sub: publish, subscribe, store além de processar (filtrar e agrupar, por exemplo) as streams em tempo real

O que é event streaming

- pub/sub com capacidades de armazenamento e processamento
- evento/mensagem/record: uma mudança de estado no sistema
- stream de eventos: fluxo dos publishers para os subscribers
- assegura o contínuo fluxo e interpretação dos dados para que as informações estejam nos lugares corretos da forma correta no momento correto

Produtores, Consumidores e Tópicos

- Tópicos: categorias onde dados são armazenados
- Produtores: aplicações que inserem dados em tópicos
- Consumidores: aplicações que lêem dados de tópicos

Um único consumidor pode não ser capaz de ler todos os eventos em streaming nos tópicos, por isso Kafka permite que grupos de consumidores se conectem a múltiplos tópicos ao mesmo tempo

Usos

- processar pagamentos e transações financeiras em tempo real
- telemetria logística de frotas
- acompanhamento e processamento de dados de dispositivos IoT
- monitoramento de pacientes em hospitais
- base fundamental para aplicações que usem arquiteturas movidas a eventos ou microsserviços