

# Arquitetura Genérica para Processamento de Eventos em Streaming utilizando Apache Kafka

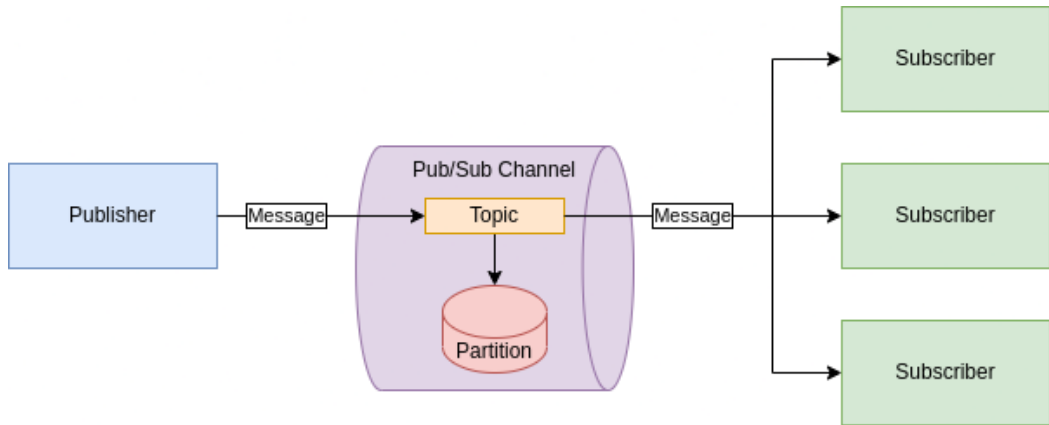
Jânio Prates Otoni

January 27, 2022

# Objetivos

- ▶ Avaliar a plataforma Kafka para *streaming* de dados de *IoT*
- ▶ **NÃO** é objetivo do trabalho processar esses dados

# Referencial



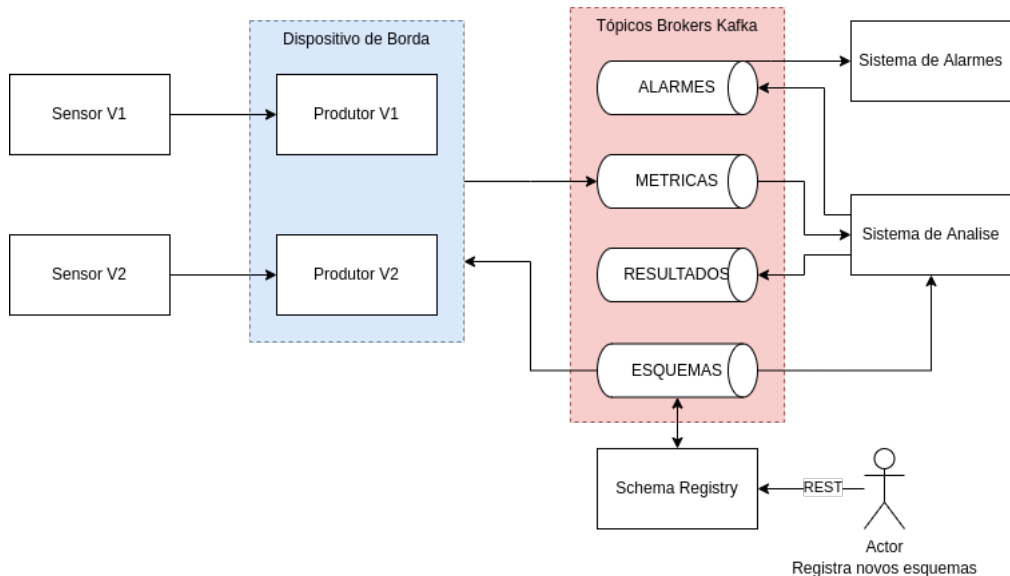
# Motivação

- ▶ Sistema de Internet das Coisas para Detecção de Incêndios Florestais
- ▶ Formato de Dados
  - ▶ Avro é amplamente utilizado em *big data*
  - ▶ Programas lêem e processam binários com mais facilidade
  - ▶ Formatos binários são mais compactos
  - ▶ Avro obriga a utilização de esquemas para a definição dos dados

## Trabalhos Relacionados

- ▶ Albuquerque, S., Fonseca, D., Lima, R. L., Milanés, A., Rocha, I., Santos, A. M., Silva, G., and Vieira, F. (2020). Sistema de internet das coisas para detecção de incêndios florestais. In Anais do XI Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais, pages 141150. SBC.
- ▶ Albuquerque, S., Silva, F., and Oliveira, D. (2021). Análise espacial da distribuição de sensores para detecção de incêndios florestais no parque estadual da serra do rola moça. In Anais do XII Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais, pages 1118, Porto Alegre, RS, Brasil. SBC.
- ▶ Mishra, B., Mishra, B., and Kertesz, A. (2021). Stress-testing mqtt brokers: A comparative analysis of performance measurements. *Energies*, 14(18):5817.

# Fluxo de Dados



# Experimentos

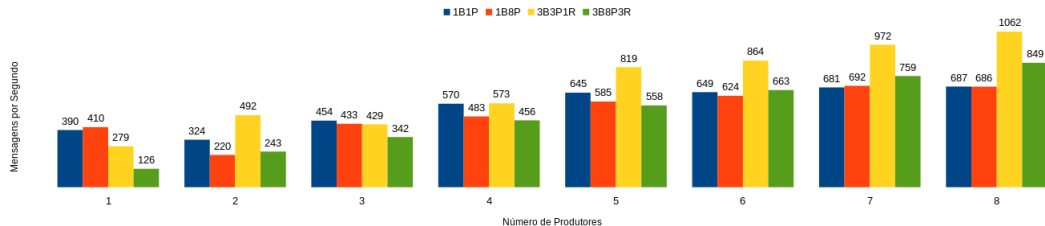
Quatro máquinas virtuais com:

- ▶ 1 vCPU ARM;
- ▶ 6 GB de memória RAM;
- ▶ 1 vNIC com 1 Gigabit Ethernet.

**Table:** Quantidade de mensagens por segundo e uso percentual de CPU relacionado ou número de produtores em execução

Produtores	1	2	3	4	5	6	7	8
msg/seg	808	1622	2260	2915	3624	4383	4876	5632
% CPU	96.35	99.80	99.68	100.00	100.00	100.00	100.00	100.00

# Número de Mensagens Produzidas em Cada Configuração do Cluster



1B1P 1 *broker* e 1 partição

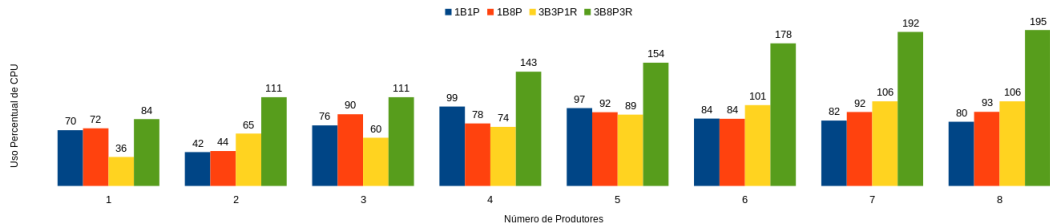
1B8P 1 *broker* e 8 partições

3B3P1R 3 *brokers*, 3 partições e 1 réplica

3B8P3R 3 *brokers*, 8 partições e 3 réplicas



# Uso de CPU em Cada Configuração do Cluster



## Conclusões e Trabalhos Futuros

- ▶ Oferece maior flexibilidade e uma quantidade maior de dados a serem trafegados
- ▶ Dependendo da escala, podem ser utilizados Raspberry Pi 4 Model B (4GB)
- ▶ Pode ser utilizado para análise de dados com baixa latência