

Sofia

TDC Floripa - 2019

globo.com

It's me Mauro



<https://www.linkedin.com/in/mauromurari/>



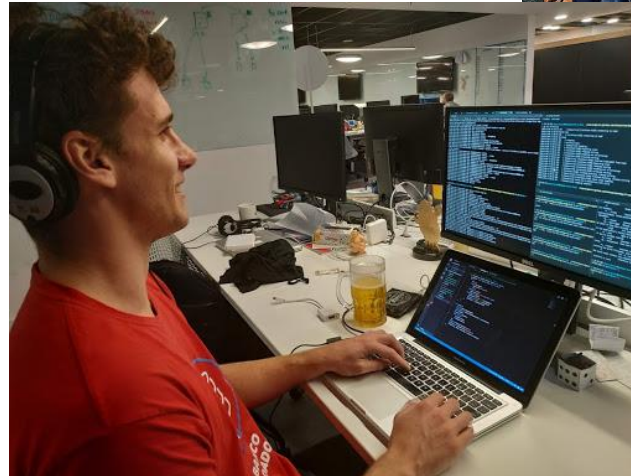
<https://github.com/otherpirate>



mauroamurari@gmail.com



@otherpirate



Problema

- Coletar dados para tomar decisão em tempo real
- Primeiro cliente: Serviço de roteamento e balanceamento de serviços
- Serviço de coleta de métricas: Statsd->Graphite

Requisitos

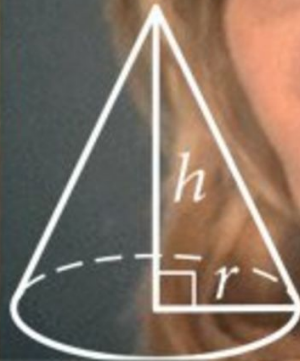
- Recebimento de eventos
- Dados disponíveis no máximo em um minuto
- Diversos formatos de entrada (JSON, Agent, UDP, Bin, Statsd)
- Pelo menos 15 milhões de eventos por dia (~2000/s)
- Granularidade de segundo em segundo
- Manter dados históricos



$$A = \pi r^2$$

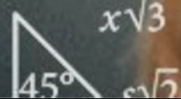
$$C = 2\pi r$$

$$V = \frac{1}{3} \pi r^2 h$$



$$V = \pi r^2 h$$

	30°	45°	60°
sin	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$
cos	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$
tan	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$



$$\int \sin x dx = -\cos x + C$$

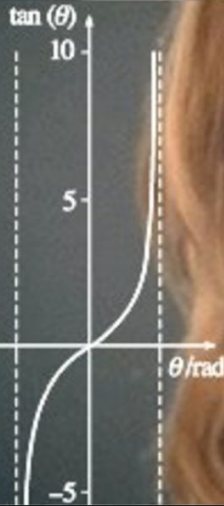
$$\int \frac{dx}{\cos^2 x} = \tan x + C$$

$$\int \tan x dx = -\ln|\cos x| + C$$

$$\int \frac{dx}{\sin x} = \ln \left| \tan \frac{x}{2} \right| + C$$

$$\int \frac{dx}{a^2 + x^2} = \frac{1}{a} \arctan \frac{x}{a} + C$$

$$\int \frac{dx}{x^2 - a^2} = \frac{1}{2a} \ln \left| \frac{x-a}{x+a} \right| + C$$



$$ax^2 + bx + c = 0$$

$$a\left(x^2 + \frac{b}{a}x + \frac{c}{a}\right) = 0$$

$$x^2 + 2\frac{b}{2a}x + \left(\frac{b}{2a}\right)^2 - \left(\frac{b}{2a}\right)^2 + \frac{c}{a} - \left(\frac{b}{2a}\right)^2 = 0$$

$$\left(x + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a^2} = 0$$

**IF YOUR DATA
HAS A TIMESTAMP**

**YOU NEED A TIME
SERIES DATABASE, HARRY**

Solução - Banco de dados

- InfluxDB
- TimescaleDB
- Prometheus
- Elasticsearch
- DruidDB
- Cassandra***

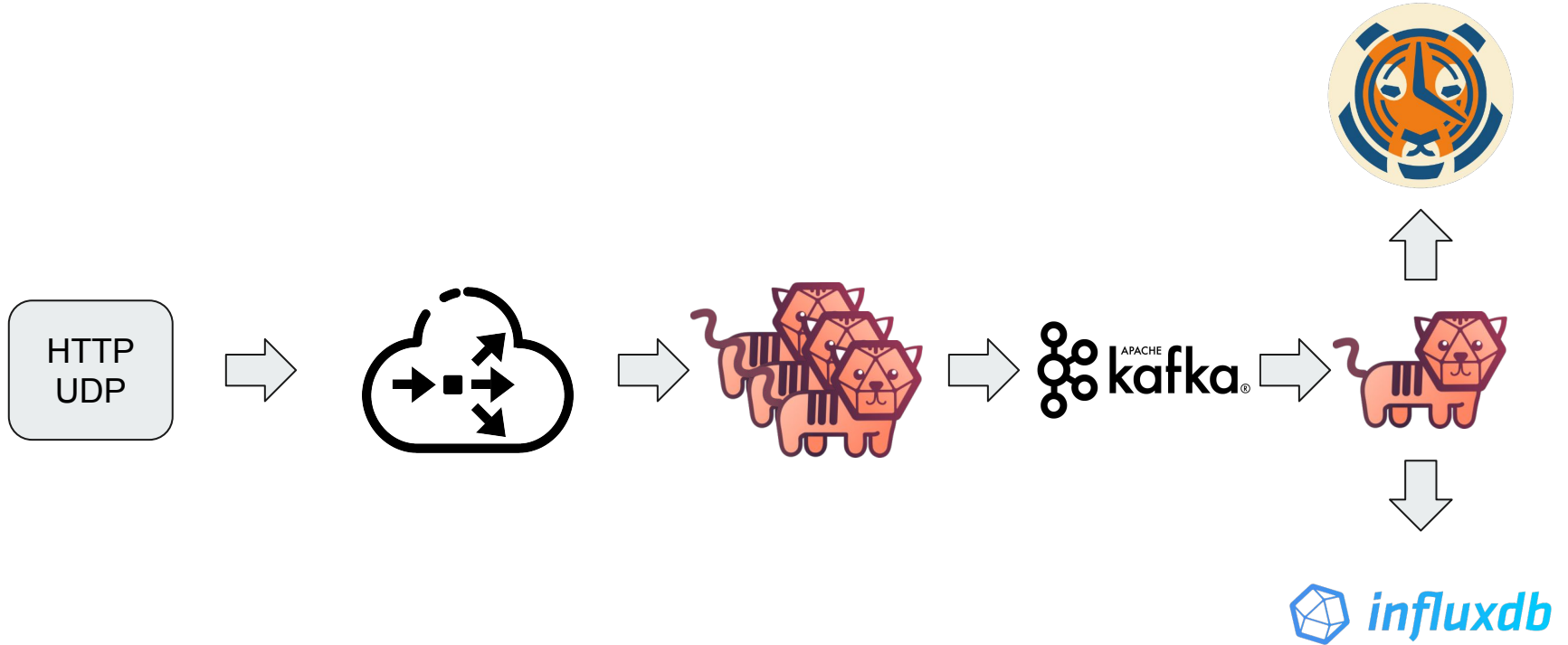
Solução - Benchmark

- TimescaleDB x InfluxDB
- 10 máquinas
- 80 cores
- Tamanho médio das mensagens 256kb
- 1B de mensagens
- Queries por: Time, fullscan, aggregate

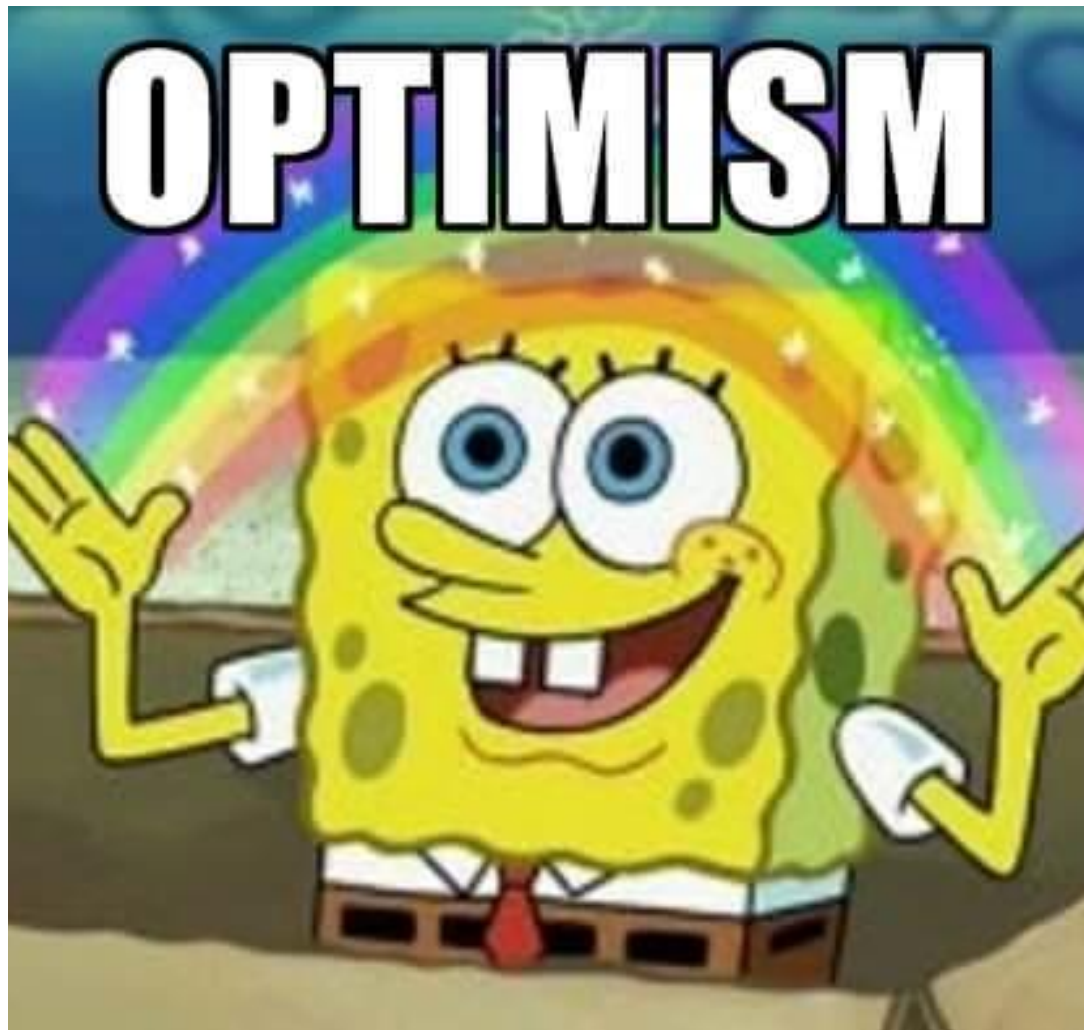
Solução - Input/Output

- KafkaBIN
- Telegraf
- Statsd

Virada 100%



OPTIMISM



Problemas

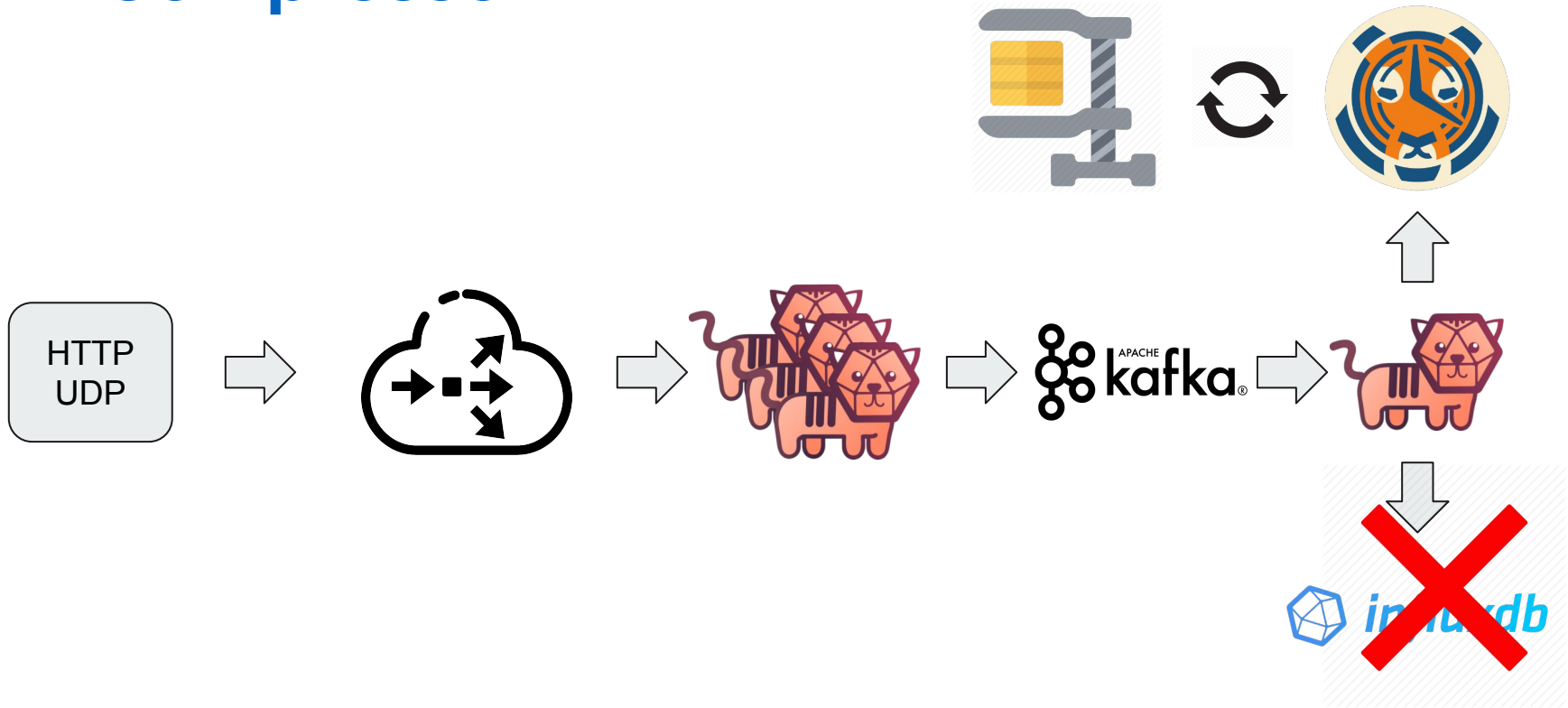
- Telegraf (Plugins, CPU, Memória)
- InfluxDB (Stop, Start)
- TimescaleDB (Disco, Backup)

TIME SERIES



(MOVING AVERAGE) SMOOTHER

Compressor



Resultados

- Segundo cliente: DBaaS
- 1B/dia (11k/s)
- Dado acessível em 2 segundos
- Dados armazenados:
 - 1/1s durante 7 dias
 - 1/1m durante 30 dias
 - 1/1hr durante 5 anos

Stack

- Telegraf (Golang)
 - Input: UDP, HTTP -> Kafka
 - Output: Kafka -> TimescaleDB, Cassandra***
- Filas: Kafka
- Compressor: Python, MongoDB, Flask
- Visualização: Grafana

Infra

- Telegraf2Kafka
 - 20 VMs
 - 76 CPUs
 - 200 GB
- Kafka
 - 3VM
 - 24 CPUs
 - 64GB
 - 200GB Dados
- Telegraf2TimeScale
 - 1 VM
 - 8 CPUs
 - 6 GB
- TimeScaleDB
 - 2VM
 - 48 CPUs
 - 256GB
 - 2TB dados

FACT



MONEY TALKS

Futuro

- Detecção de anomalia
- Automatizar
- Preparar DataLake
- ML/AI

Quer resolver esse tipo de problema?

<https://talentos.globo.com/#/oportunidades>

- Rio de Janeiro
- São Paulo
- Porto Alegre

Thanks!



 <https://www.linkedin.com/in/mauromurari/>

 <https://github.com/otherpirate>

 mauro.murari@corp.globo.com

 @otherpirate