



KAFKA + AIRFLOW



uma combinação que
dá certo



Juliany Raiol

- Manaus/Amazonas
- Desenvolvedora de Software na Loggi
- PyLadies Manaus & PyData Manaus
- Sistemas Distribuídos e Cloud
- Ficção Científica, Dança e Patins <3



@julianyraiol

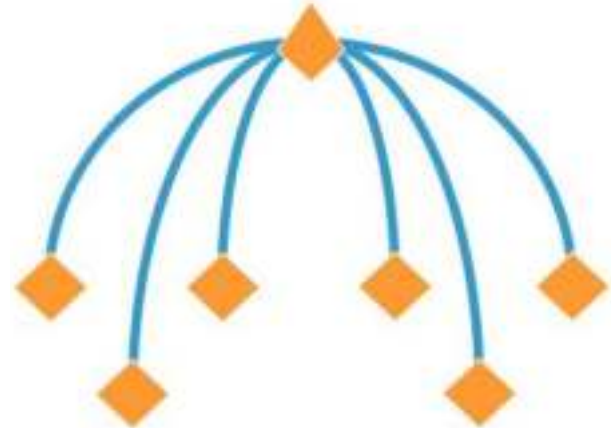


Comunidade Python em Manaus



pyladies
Manaus

@pyladiesmanaus



@pydatamanaus



Pessoas > Tecnologia

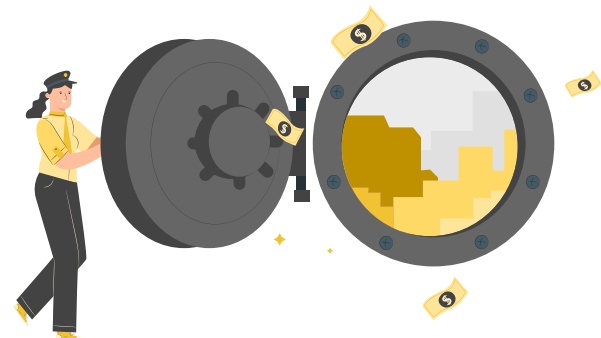


PyLadies Manaus

Giovana de Lucca – DataViz:
Detalhes para aprimorar
sua apresentação



Dados são o novo petróleo



**Toda organização é
alimentada por dados**





Era do Big Data

- Ganhou mais força a partir de 2010
 - Internet das coisas
 - Aprendizado de Máquina
- Diversas fontes de dados



Não dá para usar o **len()**





**Existe uma grande vantagem em
ver o que está acontecendo os
seus dados em **tempo real****

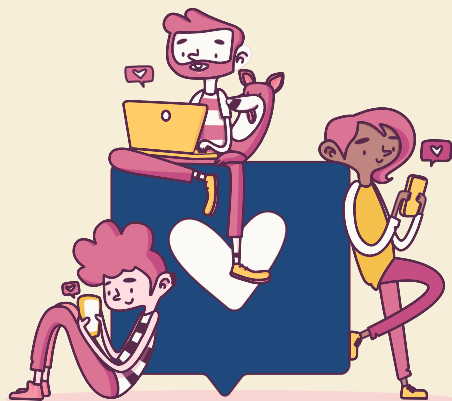
Hilary Mason



**Data Streaming ==
Fluxo de Dados Contínuo**



Pessoas ao redor do mundo



envia

recebe



recebe

envia

Pessoas ao redor do mundo

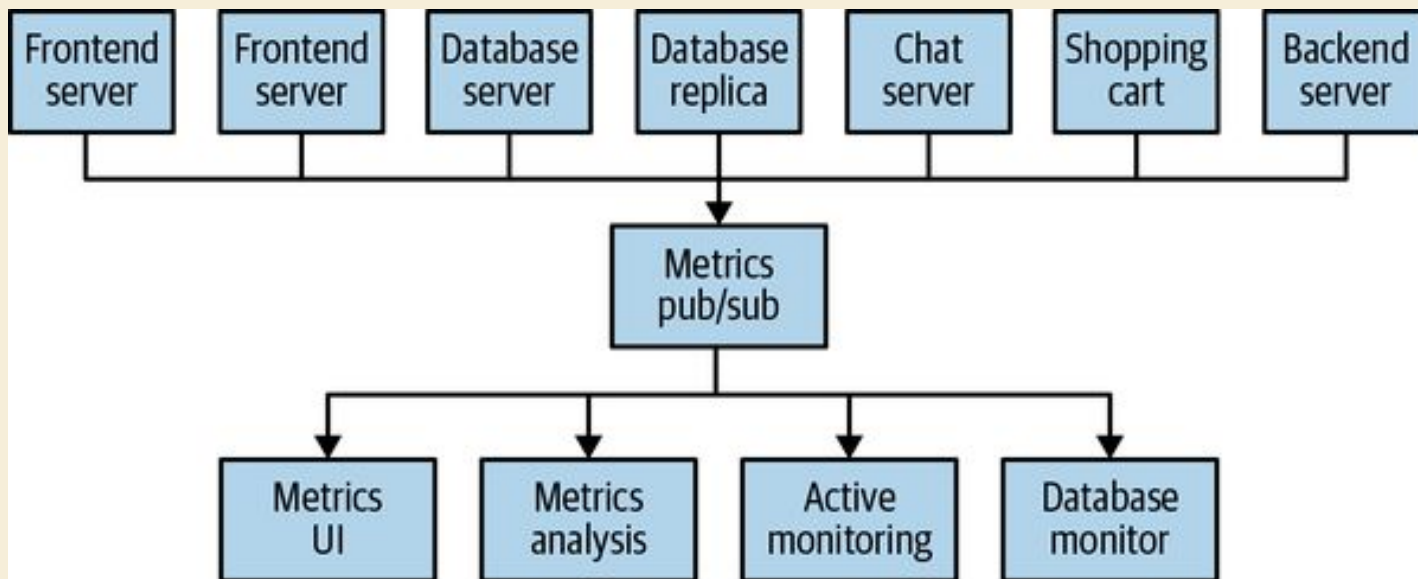


Arquitetura de sistemas escalável



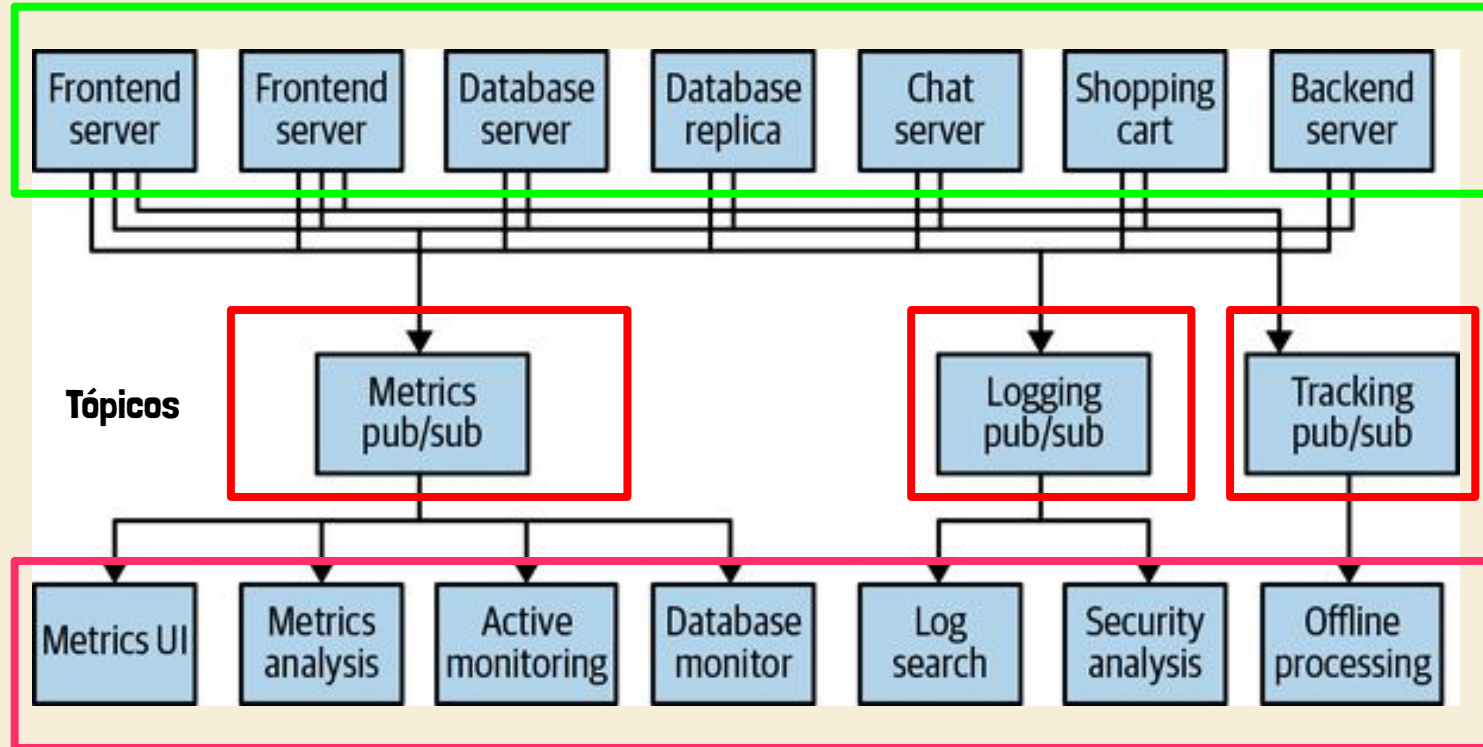
Arquitetura Orientada a Eventos

- Utiliza eventos para a tomada de decisão entre sistemas.
- **Evento, Produtor, Consumidor**



Aplicando pub/sub

Produtores



Consumidores

**Como colocar essa
arquitetura em prática?**

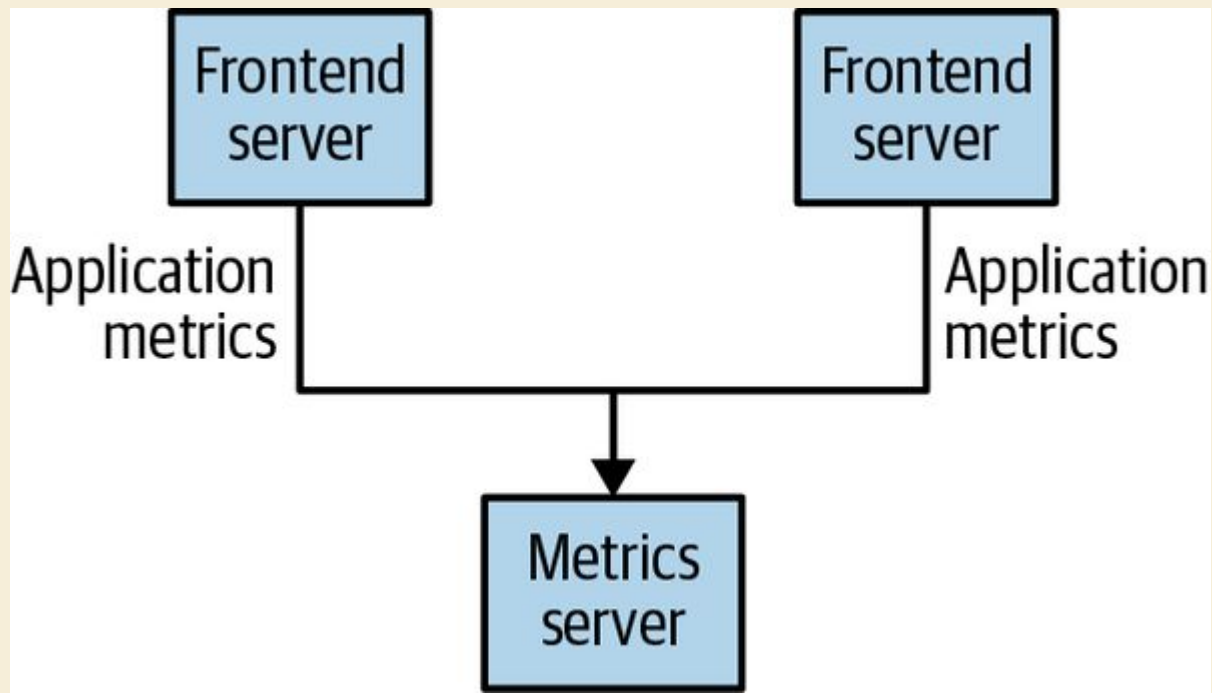


Apache Kafka

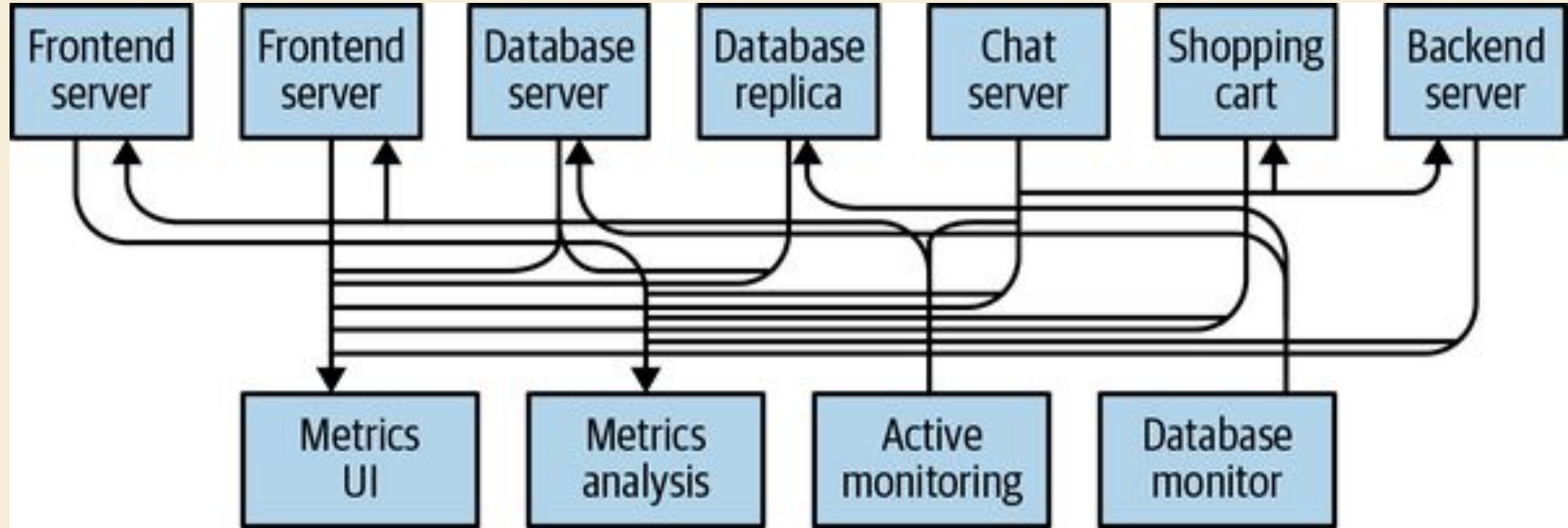
- Surgiu para resolver um problema no LinkedIn e virou open source em 2010;
- Ganhou popularidade e agora faz parte do cotidiano de diversas empresas.
- Camada de transporte;
- Não trabalha com filas;



Em linhas gerais, começamos com:

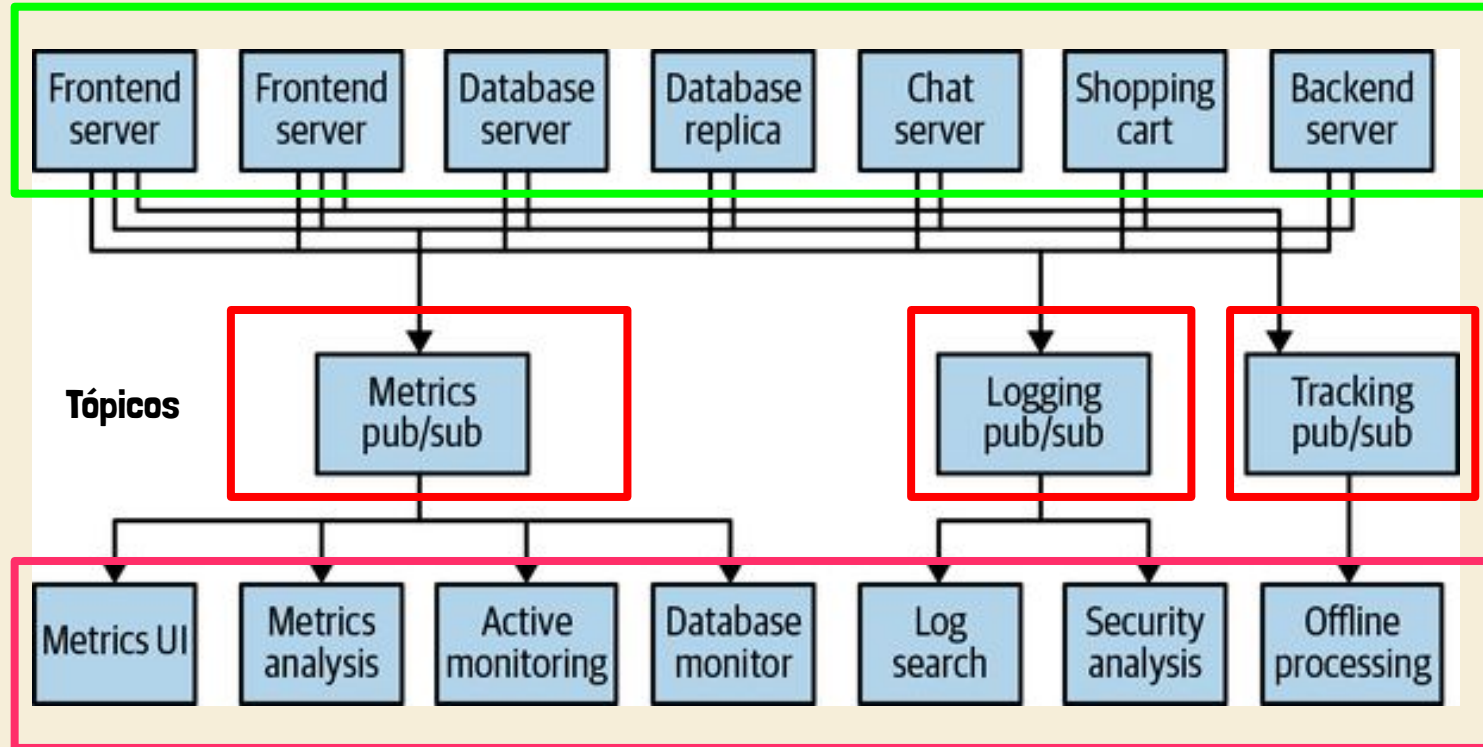


Depois de um tempo, as coisas podem perder o controle



Aplicando pub/sub

Produtores



Consumidores

Vejamos um exemplo básico



producer.py

```
import json

from const import config
from data import data_source
from confluent_kafka import Producer

producer = Producer(config)
topic = "gorn_example"

def confirmacao_entrega(err, msg):
    if err is not None:
        print('Entrega de mensagem com falha: {}'.format(err))
    else:
        print('Mensagem entregue no tópico {} e na partição [{}]'.format(msg.topic(), msg.partition()))

for data in data_source:
    producer.poll(0)
    producer.produce(topic, json.dumps(data), callback=confirmacao_entrega)

producer.flush()
```


consumer.py

```
import json
from const import config
from confluent_kafka import Consumer

topic = "gorn_example"

consumer = Consumer(config)
consumer.subscribe([topic])

while True:
    try:
        message = consumer.poll(1.0)
        if message is None:
            continue
        converted_message = json.loads(message.value().decode("utf-8"))
        print(converted_message)

    except KeyboardInterrupt:
        break

consumer.close()
```

**Como transformar isso em
uma tarefa repetitiva?**



Apache Airflow

- Open Source
- Orquestrador de pipelines
- Usa Python para escrever as DAGS



**O Airflow
funciona como
um despertador**



dag_kafka.py

```
from datetime import datetime
from airflow import DAG
from airflow.operators.python_operator import PythonOperator
from app.kafka_airflow import KafkaAirflow

def get_kafka_messages():
    kafka_messages = KafkaAirflow()
    return kafka_messages.consume_messages()

default_args = {
    "owner": "grypyrn",
}

dag = DAG(
    dag_id="dag_exemplo_gorn",
    start_date=datetime(2023, 11, 3),
    tags=["kafka", "airflow"],
    catchup=False,
    default_args=default_args,
    schedule_interval="@ 15,19,23 * * *",
)

get_kafka_messages_task = PythonOperator(
    task_id="get_kafka_messages",
    python_callable=get_kafka_messages,
    dag=dag
)

get_kafka_messages_task
```

“

**Pra martelo tudo é
prego!**



”

Outras alternativas

- Kafka

- RedPanda
- Amazon Kinesis Data Streams
- Azure Event Hubs
- Google Cloud Dataflow



- Airflow

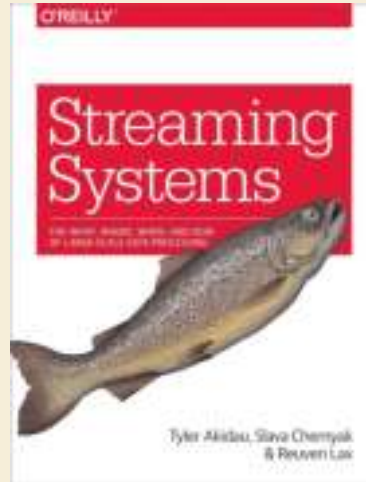
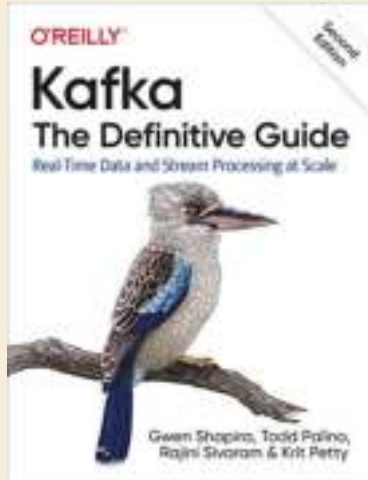
- Cronjob Kubernetes
- Celery
- Github Actions



Onde aprender mais (Documentação)

- Documentação do Airflow:
<https://airflow.apache.org/>
- Documentação do Kafka:
<https://kafka.apache.org/documentation/>
- Live de Python:
https://www.youtube.com/watch?v=LMk78Dmk0qU&ab_channel=EduardoMendes

Onde aprender mais (Livros)





Obrigada!



@julianyraiol