



# Do BI ao Stream

grupo **ZAP**

# André Ronquetti

- Gerente de Engenharia no GrupoZAP
- 4 anos fiz uma transição de carreira para Gestão de Pessoas
- Faço código nos fins de semana
- [andre.ronquetti@gruposzap.com](mailto:andre.ronquetti@gruposzap.com)
- [LinkedIn](#)



Quem é o grupo **ZAP** ?

 +  VivaReal





**BI clássico**  
**0% Real time**  
**Centralizado**  
**Limite da**  
**capacidade**



**Redshift**  
**Centralizado**  
**1% Real time**  
**Estrutura**  
**incompleta**

The logo for 'sua house' features the word 'sua' in a light green font and 'house' in a grey font.

O melhor e mais eficiente sistema de gestão, com CRM, plataforma de BI, controle de leads e muito mais.

The logo for 'data ZAP' has 'data' in grey and 'ZAP' in a stylized green font with a blue accent.

Empresa de inteligência imobiliária que auxilia no processo de decisão para construção, investimento, financiamento e negociação.

The logo for 'fipec ZAP' shows 'fipec' in grey and 'ZAP' in green with a red accent.

Principal índice de preço de imóveis, de abrangência nacional. Parceria entre Fipe e o Grupo ZAP.

The logo for 'zap' consists of a blue stylized 'z' followed by 'ap' in orange.

O portal especialista em imóveis.

The logo for 'VivaReal' features a blue house icon with a white 'V' inside, followed by the text 'VivaReal' in blue.

O maior portal de imóveis do Brasil.

The logo for 'conecta imobili' has 'conecta' in grey and 'imobili' in a lighter grey font.

A plataforma que coloca o mercado imobiliário no centro de tudo.

The logo for 'GEOIMOVEL' has 'GEO' in orange and 'IMOVEL' in grey.

Plataforma de soluções e pesquisas imobiliárias com as mais diferenciadas ferramentas e instrumentos de análise mercadológica para as empresas que atuam no setor imobiliário

# Desafio pós fusão

- Criar uma stack de Data robusta e escalável.
- Que sustente todo o pool de empresas do GrupoZAP.
- Estrutura fléxivel
- Stack Near Real time

**Um time de Data  
com 100 pessoas**

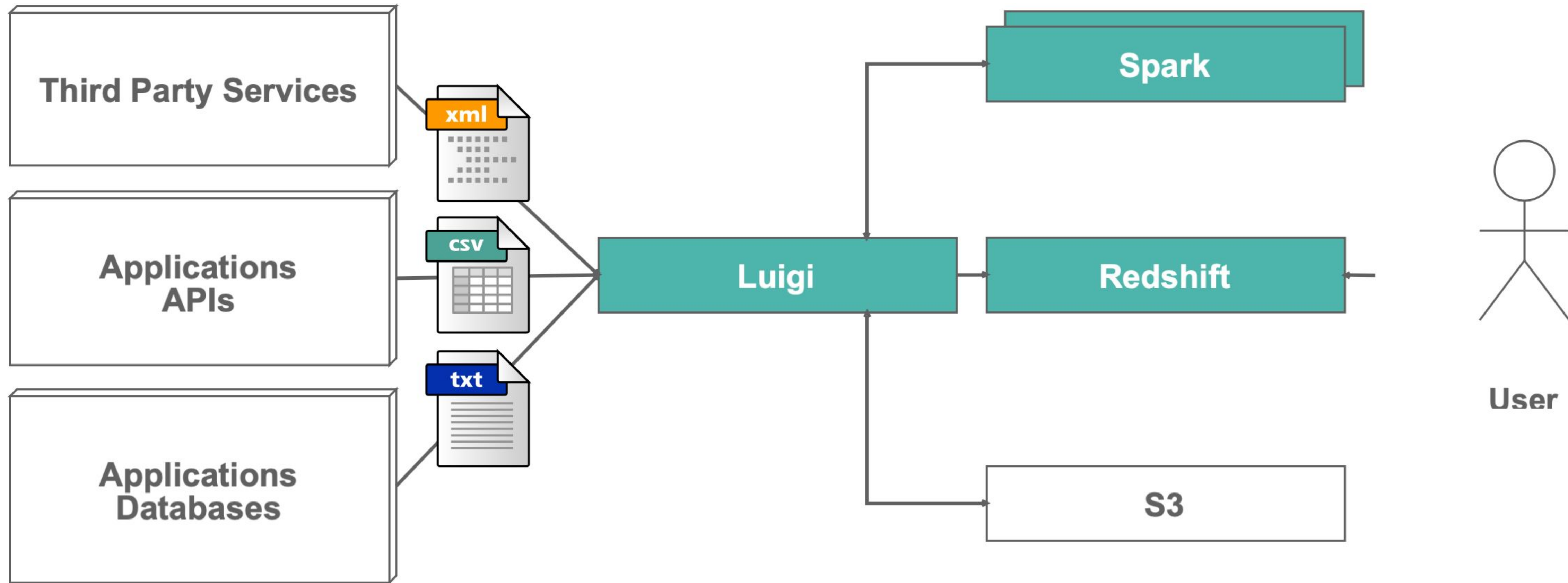


# Descentralização de Dados

**“Qualquer pessoa  
que precise de uma  
resposta deve ter  
acesso aos dados  
de forma fácil e  
simples.”**

# Consumer Analytics



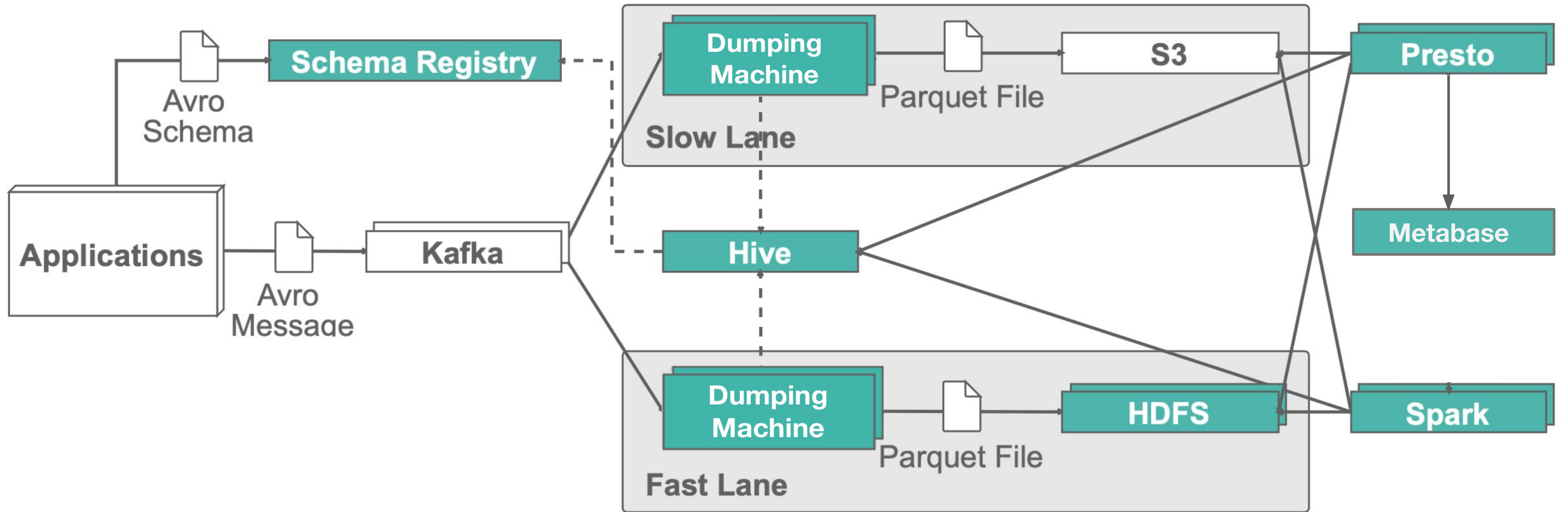


# Consumer Analytics

- Prós
  - Arquitetura simples
- Contras
  - Não **escalável**
  - Dados em **D-1**
  - Alto **custo**
  - Muita **lógica de negócio** concentrada



**kubernetes**



# Metabase

DATA PLATFORM

Esta questão está escrita em SQL.

ABRIR EDITOR



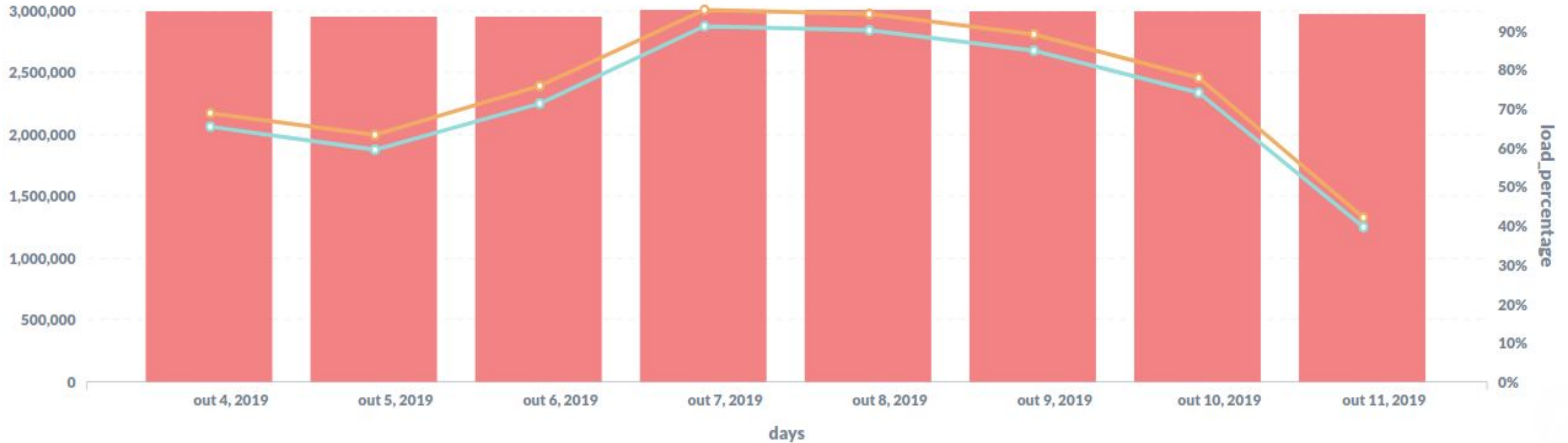
DISPLAY

Line

Atualizar



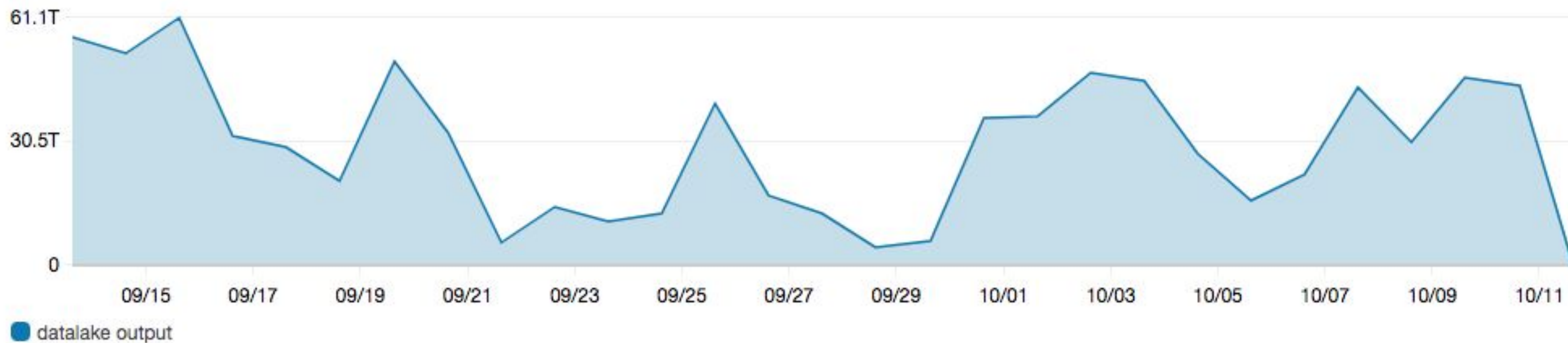
count\_all\_event\_table count\_event load\_percentage





## datalake-download-day

No unit



## datalake-download-day

# 854<sub>T</sub>

datalake output

### Qual o total de usuários no Metabase?

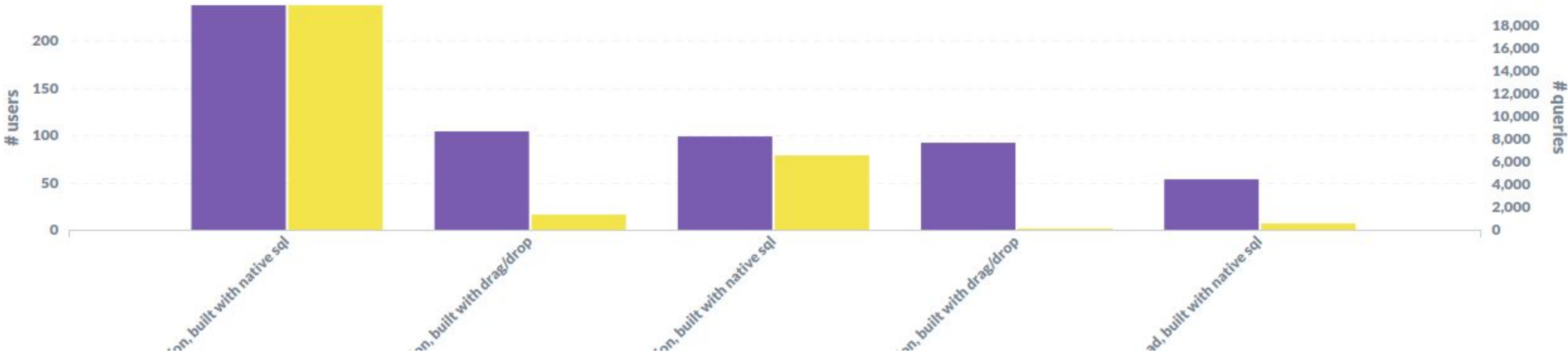


### Há quantos dias nossos usuários estão sem logar no Metabase?



### De que maneira as queries foram executadas no Metabase nos últimos 30 dias?

● # users ● # queries



# Consumer Analytics

- **Prós**

- Escalabilidade
- **Real-time**
- **Baixo custo**
- Totalmente **agnóstica**
- Atende múltiplas **necessidades**

- **Contras**

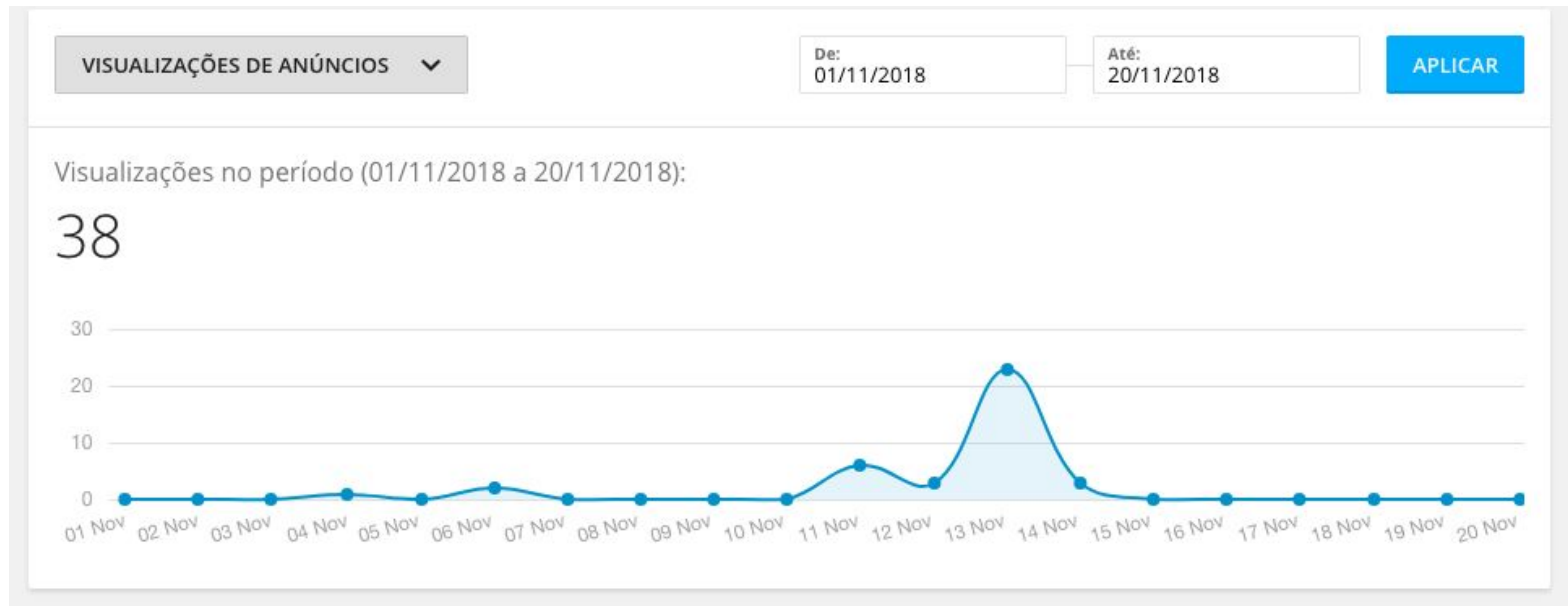
- Arquitetura complexa
- Muita configuração

# Application Analytics

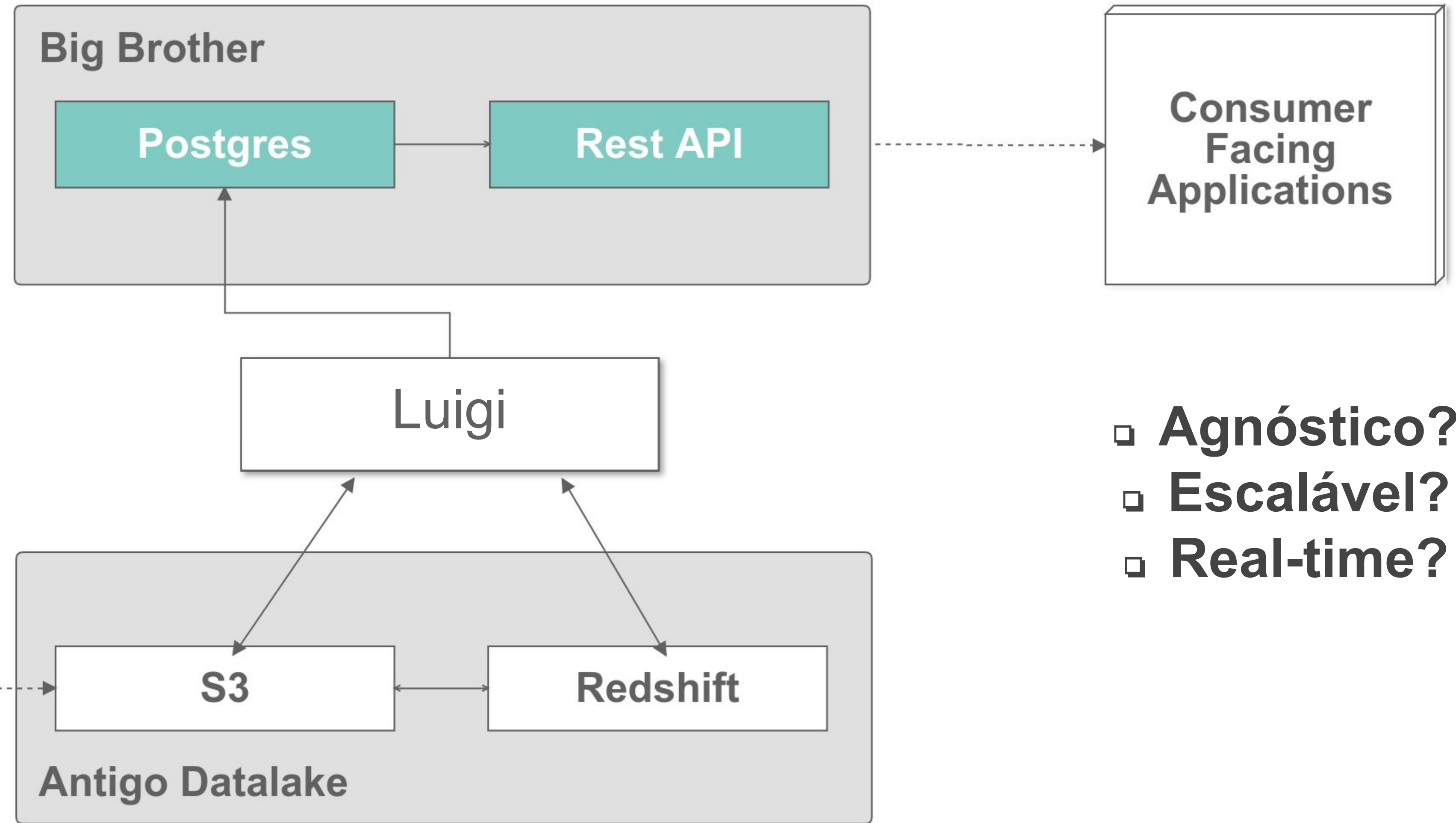


# Cenário Data Product

- **Mostrar métricas para o usuário em um dashboard**



# Antiga stack: Big Brother



# Druid

- ▣ **OLAP / Time Series**
- ▣ **Dados históricos e realtime**
- ▣ **Ingestão: Kafka, HDFS, S3**
- ▣ **Armazenamento Colunar**

# Druid - Segmentos

**Time**

**Dimensions**

**Metrics**

Timestamp	Id Anunciante	Id Anuncio	Plataforma	Count
2018-11-20T14:00:00	123123	84782	ZAP	1
2018-11-20T14:00:01	765232	36287	VIVA	1
2018-11-20T14:00:02	28362	37487	ZAP	1
2018-11-20T14:00:03	1726372	23434	VIVA	1



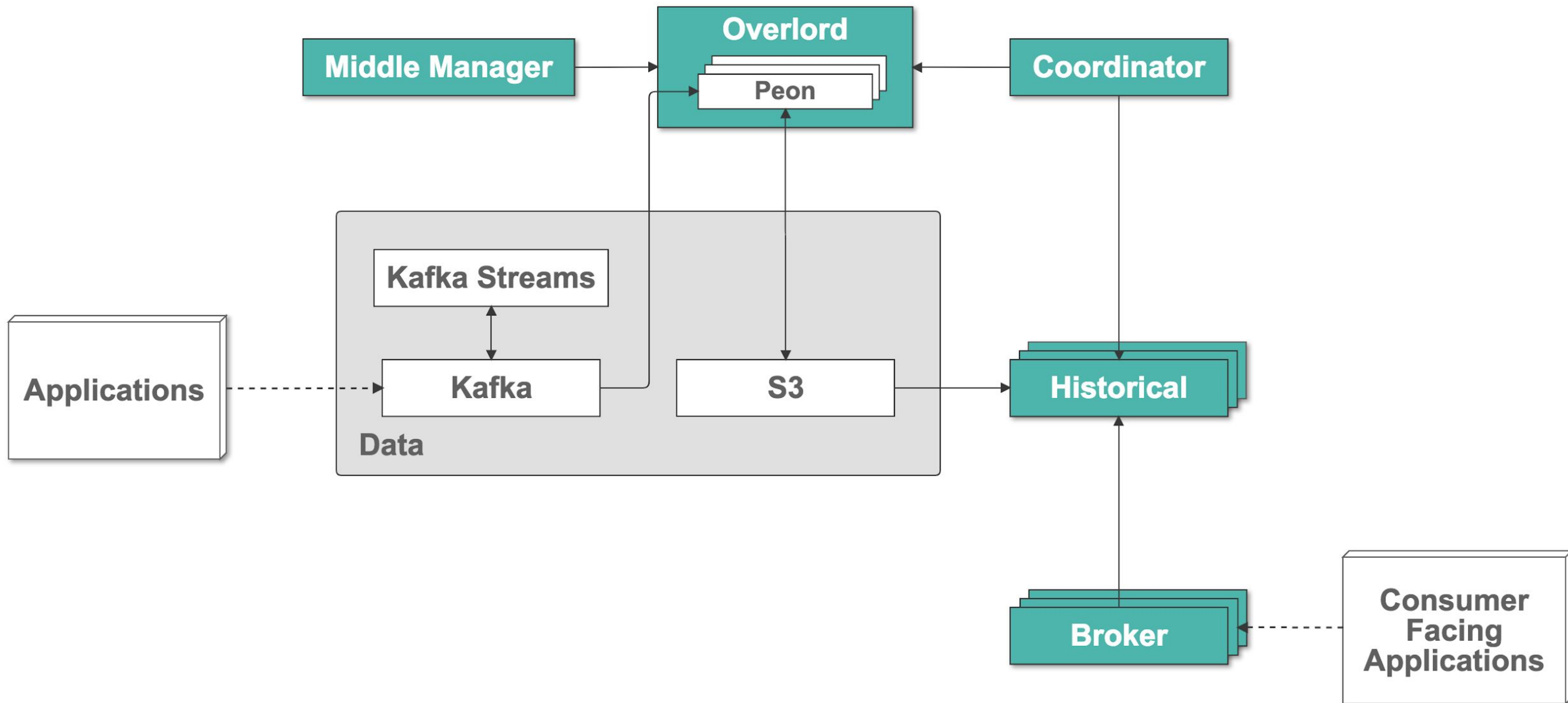
# Druid

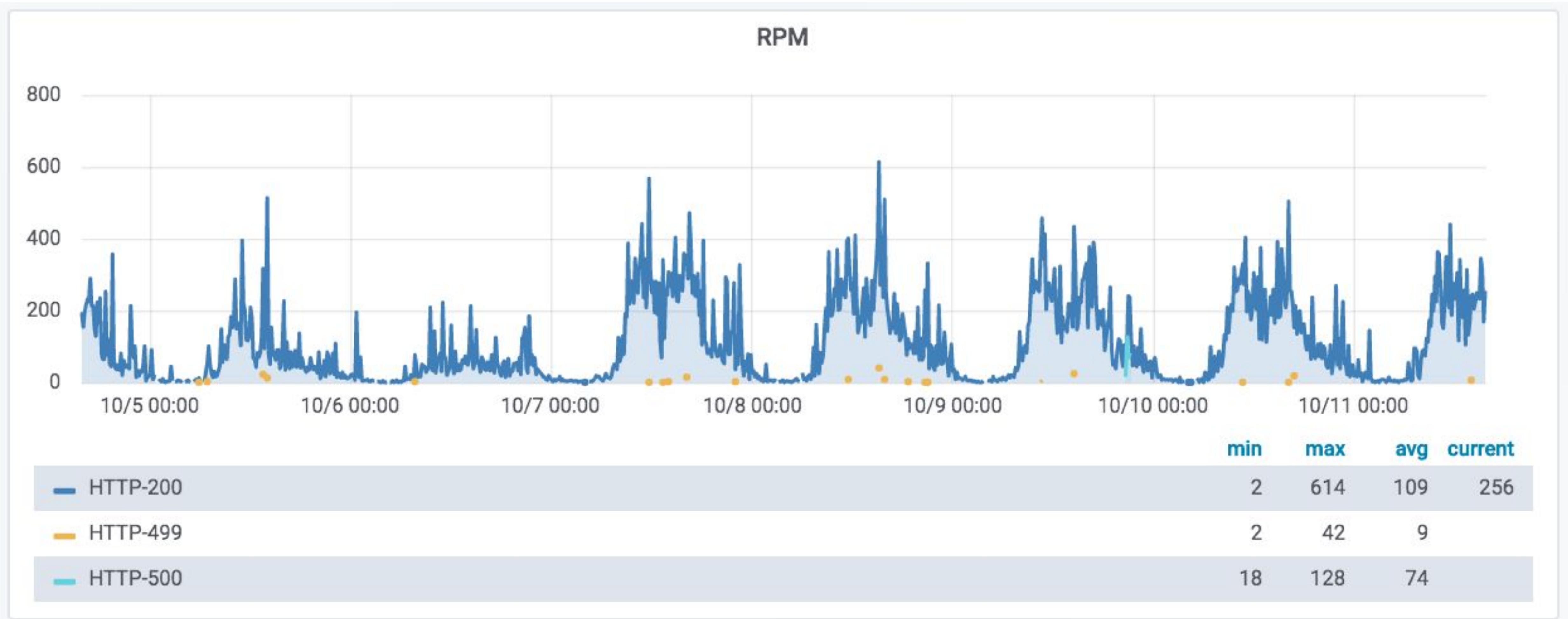
```
"parseSpec": {  
  "format": "timeAndDims",  
  "timestampSpec": {  
    "column": "date",  
    "format": "yyyy-MM-dd"  
  },  
  "dimensionsSpec": {  
    "dimensions": [  
      "date",  
      "publisher_id",  
      "listing_id",  
      "application",  
      "platform",  
      "name"  
    ]  
  }  
}
```

# Druid

```
"metricsSpec": [  
  {  
    "name": "count",  
    "type": "count"  
  }  
],  
"granularitySpec": {  
  "type": "uniform",  
  "segmentGranularity": "DAY",  
  "queryGranularity": "HOUR"  
}
```

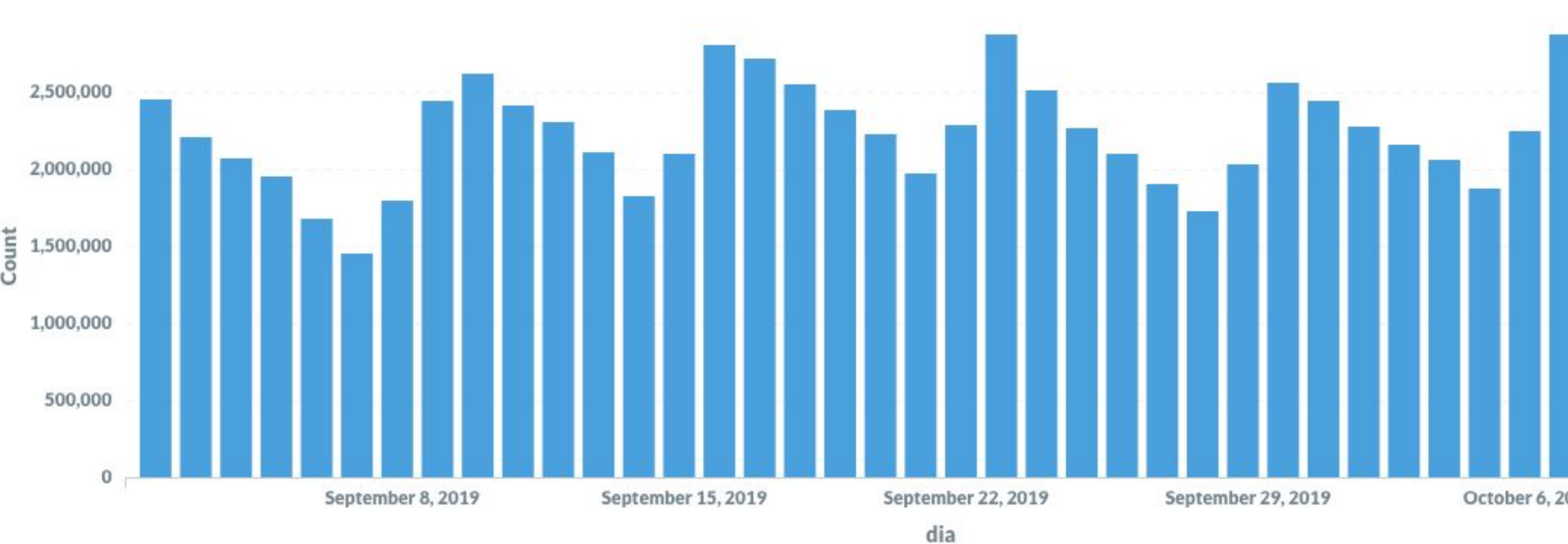
# Druid - Arquitetura





**256 RPM**

**109 ms avg**



**2.5MM de mensagens novas  
por dia**

**1.3T compactados em 20GB no período de 1 ano em 4 meses**

**3.5G de que são compactados em 40 MB por dia**

# Druid

- **Prós**

- **Escalabilidade**
- **Distribuído**
- **Real-time**
- **Queries rápidas**
- **Agnóstico**

- **Contras**

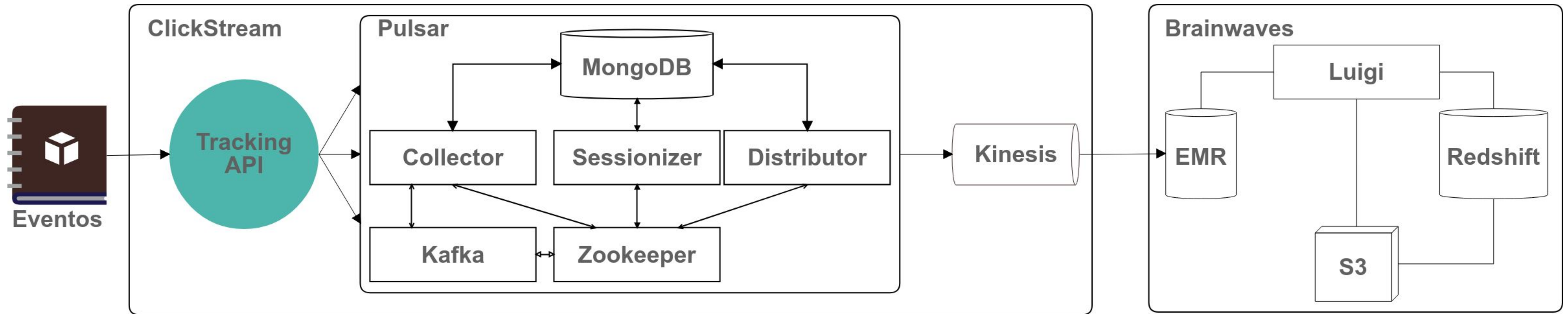
- **Arquitetura complexa**
- **Poucas fontes de informações**
- **Muuuuuitas configurações**



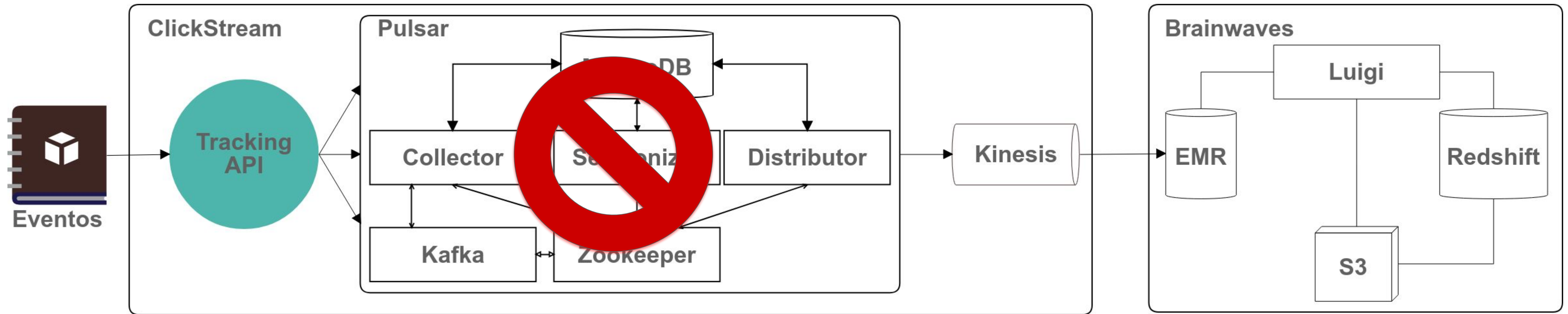




# Antiga - Stack Cerebro

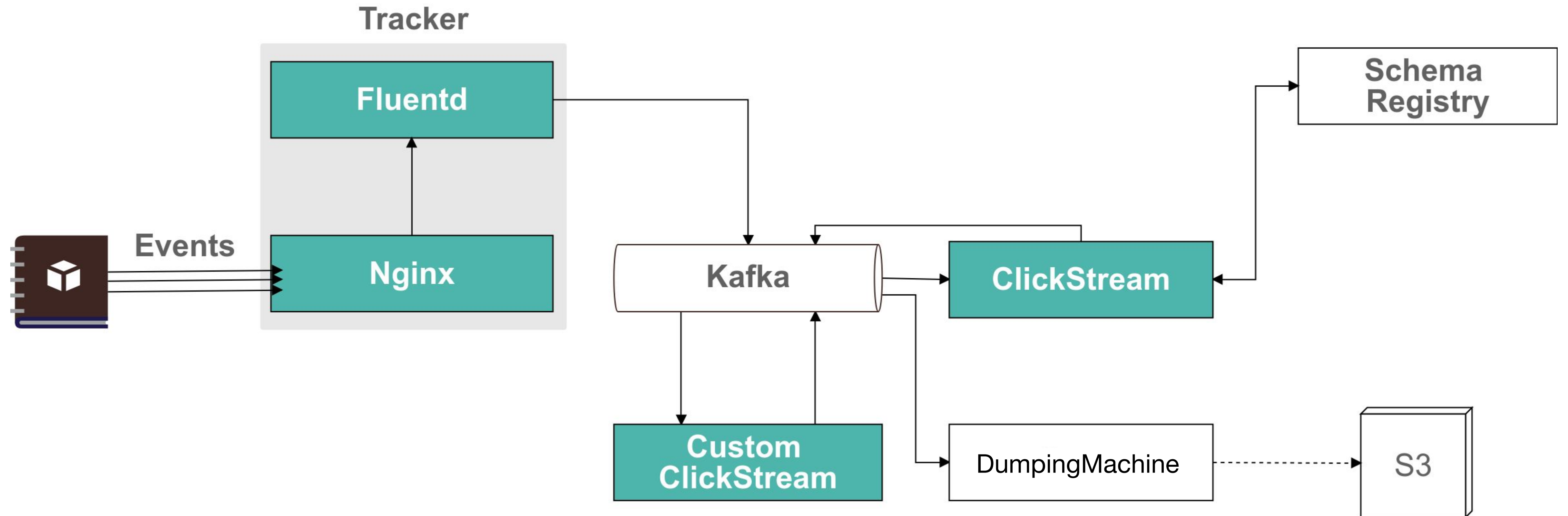


# Antiga - Stack Cerebro





# Fluxo ClickStream



- ❑ Dados segmentados por data
- ❑ Dados no formato **Parquet** (Columns Data)
- ❑ Dados para consulta no Hive/Presto

# Fluxo ClickStream

{json}

```
{  
  "application": "ZAP",  
  "platform": "DESKTOP",  
  "anonymous_id": "cjfvn",  
  "name": "pageview",  
  "type": "PAGEVIEW"  
}
```



# Fluxo ClickStream



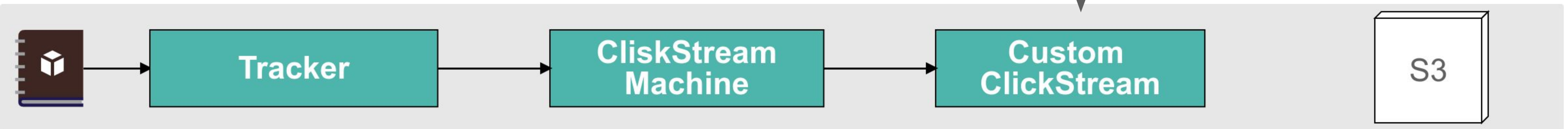
```
{  
  "application": "ZAP",  
  "platform": "DESKTOP",  
  "anonymous_id": "cjfvn",  
  "name": "pageview",  
  "type": "PAGEVIEW",  
  "session_id": "541d",  
  "location": "sp",  
  "device": "iphone"  
}
```



# Fluxo ClickStream



```
{  
  {  
    "application": "ZAP",  
    "name": "pageview",  
  },  
  {  
    "application": "ZAP",  
    "name": "click PDP",  
  }  
}
```

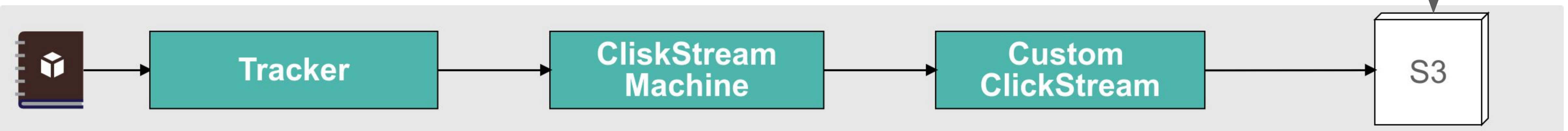




# Fluxo ClickStream



```
{  
  "application": "ZAP",  
  "platform": "DESKTOP",  
  "anonymous_id": "cjfvn",  
  "name": "pageview",  
  "type": "PAGEVIEW",  
  "session_id": "541d",  
  "location": "sp",  
  "device": "iphone"  
}
```



# Fluxo ClickStream

{json}

```
{  
  "application": "ZAP",  
  "platform": "DESKTOP",  
  "anonymous_id": "cjfvnxz6",  
  "name": "pageview home",  
  "type": "PAGEVIEW"  
}
```



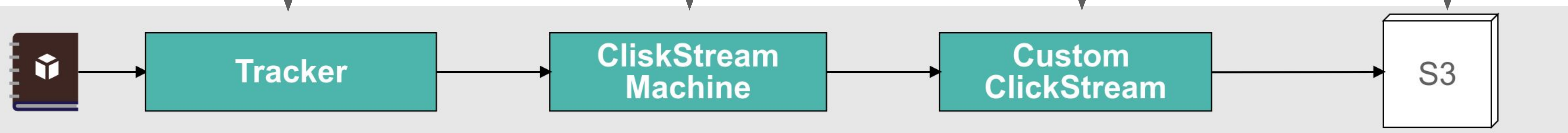
```
{  
  "application": "ZAP",  
  "platform": "DESKTOP",  
  "anonymous_id": "cjfvnxz6",  
  "name": "pageview home",  
  "type": "PAGEVIEW",  
  "session_id": "541564d5d",  
  "location": "são paulo",  
  "device_family": "iphone"  
}
```

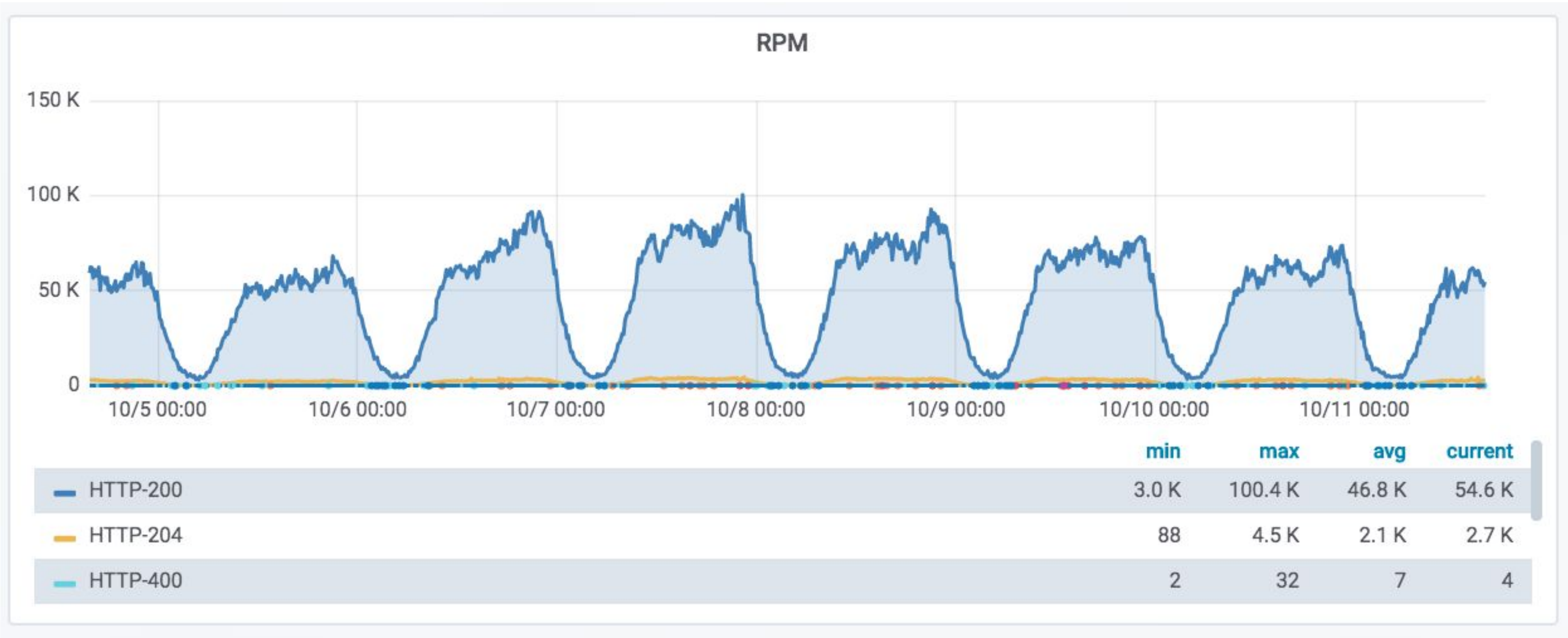


```
{  
  {  
    "application": "ZAP",  
    "name": "pageview home",  
  },  
  {  
    "application": "ZAP",  
    "name": "click PDP",  
  }  
}
```

Parquet

```
{  
  "application": "ZAP",  
  "platform": "DESKTOP",  
  "anonymous_id": "cjfvnxz6",  
  "name": "pageview home",  
  "type": "PAGEVIEW",  
  "session_id": "541564d5d",  
  "location": "são paulo",  
  "device_family": "iphone"  
}
```



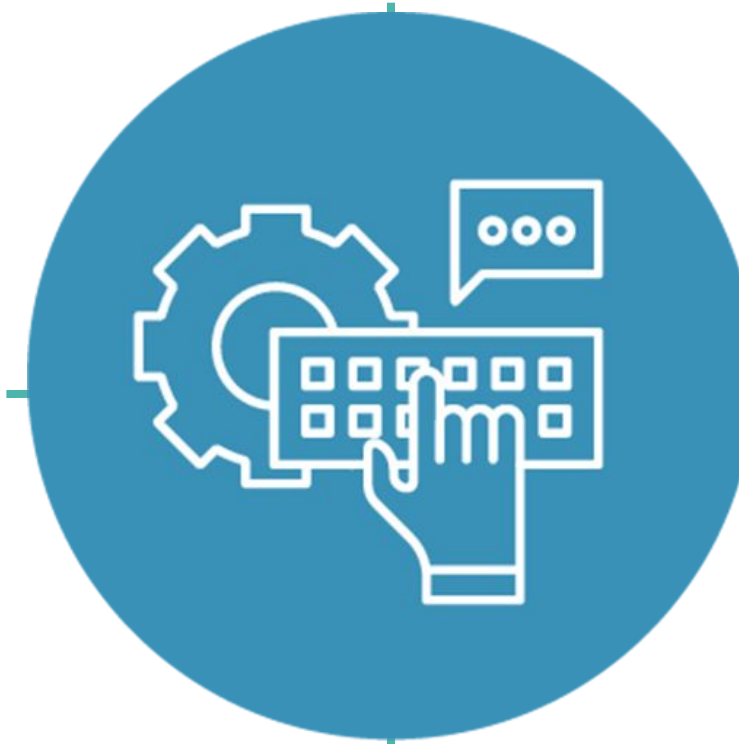
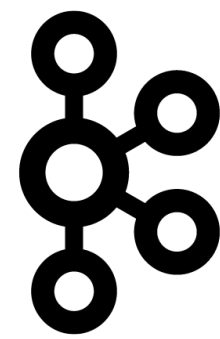
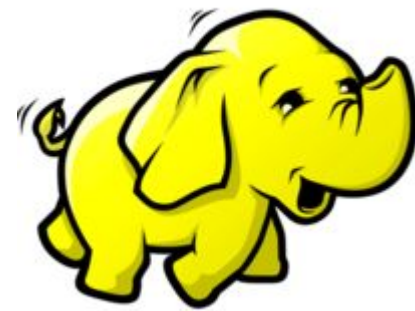


**46.8K RPM**

**60MM Eventos/dia**

# Data Map

## Big Data Tools



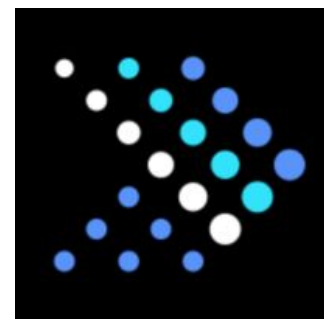
## Visualization and Format Files



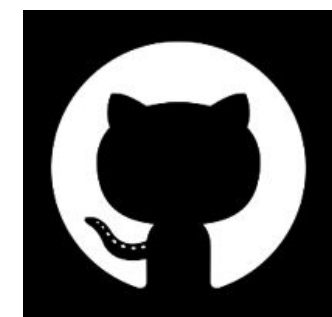
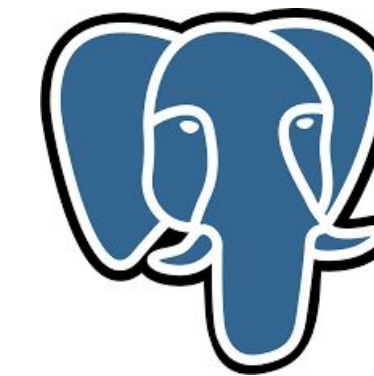
Parquet



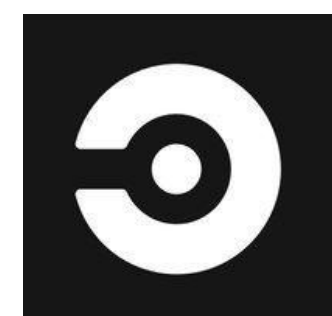
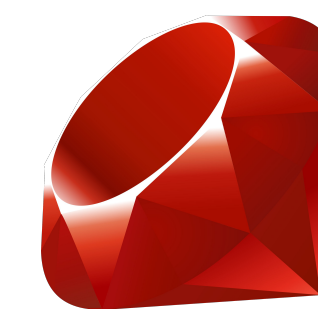
fluentd



## Infra Data Tools



## Dev Tools





grupo **ZAP**



[gruposzap.com](http://gruposzap.com)

**Obrigado !!!**