

Como a RuPaul entrega insights para os nossos anunciantes

Maria Eduarda Nubié @ Grupo ZAP
Robinson Nakamura @ Grupo ZAP



grupo ZAP

GEOIMOVEL

fipe ZAP

data ZAP

conecta imobili

suahouse

 zap

 VivaReal



Maria Eduarda Nubié

madu@grupozap.com

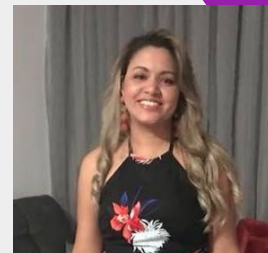
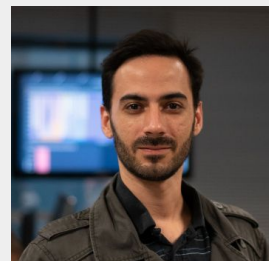
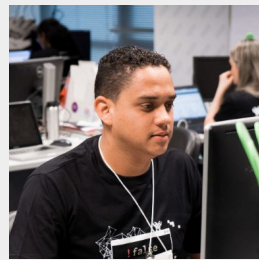
Data Analyst

Robinson Nakamura

naka@grupozap.com

Data Engineer







Quem é a **RuPaul**?



Quem é?

Drag Queen mais bem sucedida dos EUA. Transforma o comum em algo

fantástico



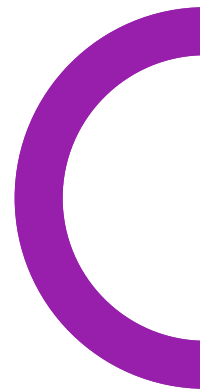


RuPaul no **Grupo ZAP**



Imobiliária Feliz

O Naka deixou para trás a carreira de engenheiro de dados e foi empreender no mercado imobiliário de Recife, vendendo apartamentos em Boa Viagem



Informações no Canal Pro

Entre outras features, o Canal Pro traz o total de visualizações e de mensagens que cada anúncio recebeu.

Desempenho

Nos últimos 30 dias 10/09/2019 A 10/10/2019

 1678

Visualização de anúncios

 73

Mensagens de contato

 42

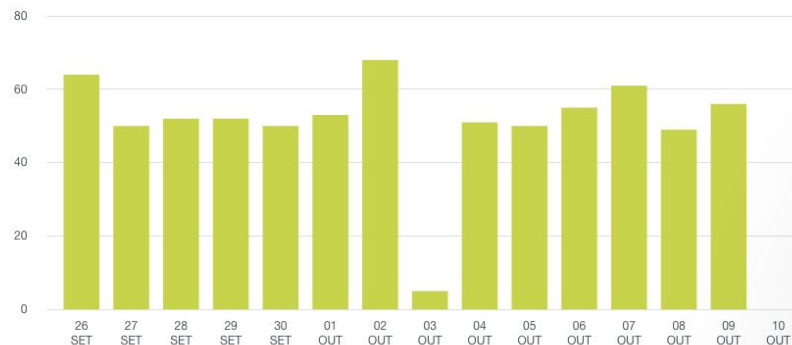
Exibições de telefone

Por período 26/09/2019 A 10/10/2019

15 dias 

716 Visualizações de anúncios



Não seria **fantástico** se
a **Imobiliária Feliz**
soubesse como está
posicionada no
mercado de venda de
apartamentos em Boa
Viagem?

Através de um índice de performance, poderíamos trazer o ***insight*** sobre onde a **Imobiliária Feliz** está pior, igual ou melhor que a concorrência.



FaceApp



~35 mil Nakinhas

A **RuPaul** é a
plataforma que
gera o **Relatório
de Desempenho**
do Grupo ZAP



RuPaul & **Arquitetura**





~240 milhões

Dados de Anúncios por Mês





~390 GB

Dados de Anúncios por Mês



Precisávamos de uma arquitetura que conseguisse realizar **transformações** nos nossos dados, trazendo análises ***fantásticas*** para nossos anunciantes.



Alaska

Sasha

Violet

Processa
os dados
do Data
Lake

Gera os
PDFs
com as
análises

Envia os
PDFs por
e-mail

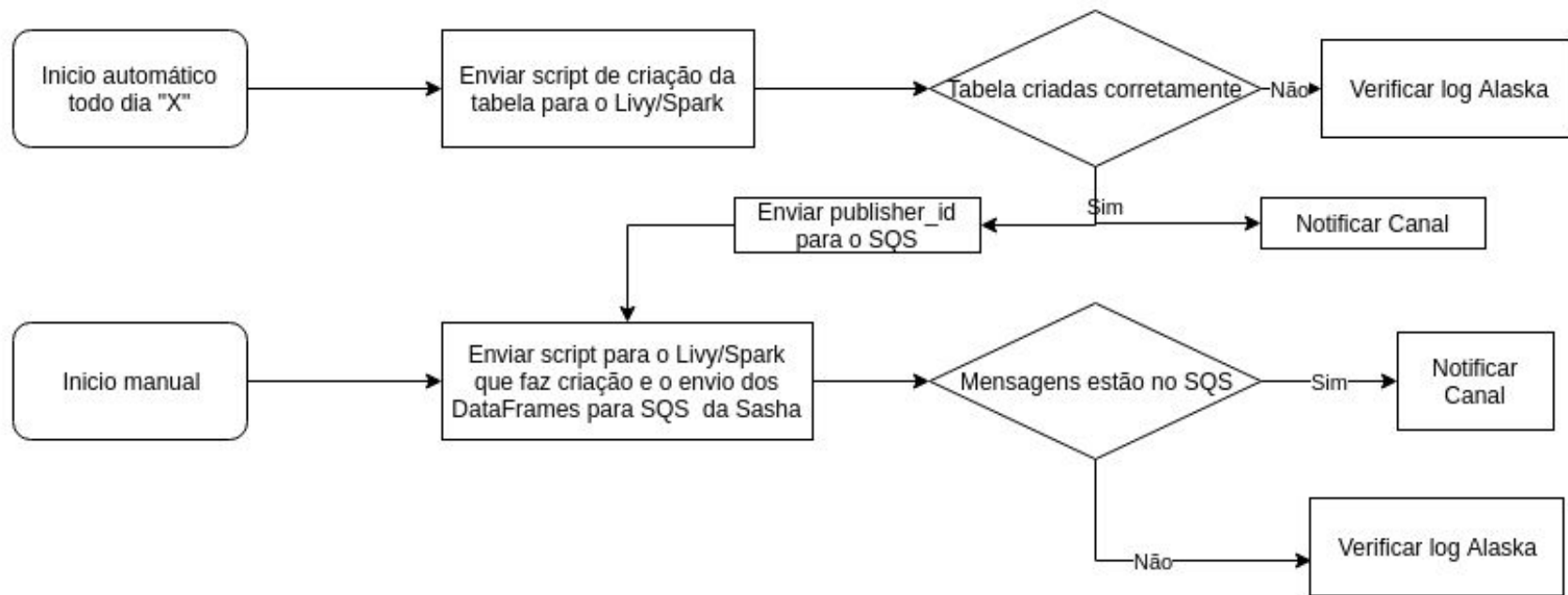


RuPaul & **Alaska**



Alaska

Criação da tabela agregada de listings com os leads do periodo



Ofertas e Contatos nos Últimos 6 Meses

ESTE GRÁFICO CONTÉM:

Informações dos resultados gerados nos últimos meses.

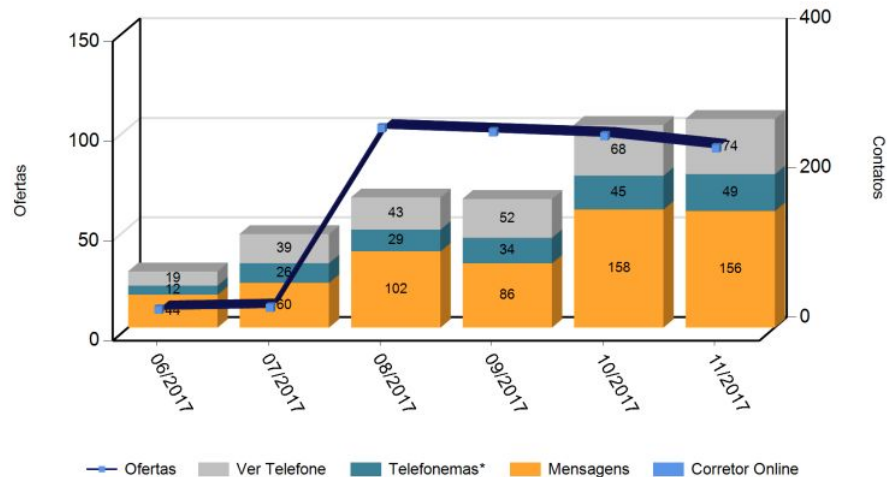
COMO ANALISAR ESTE GRÁFICO:

A coluna/barra corresponde aos contatos gerados pelo internauta em suas ofertas.

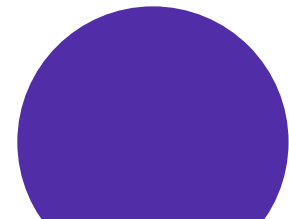
A linha corresponde à média de ofertas veiculadas pela imobiliária no portal.

QUAL CONCLUSÃO PODEMOS TIRAR DESTE GRÁFICO:

Acompanhar a média de oferta inserida em cada mês, com o número de contatos recebidos no mesmo período.



* Estimativa de ligações recebidas com base em estatísticas do serviço ZAP Fone



Quais eram os requisitos?

Escalável

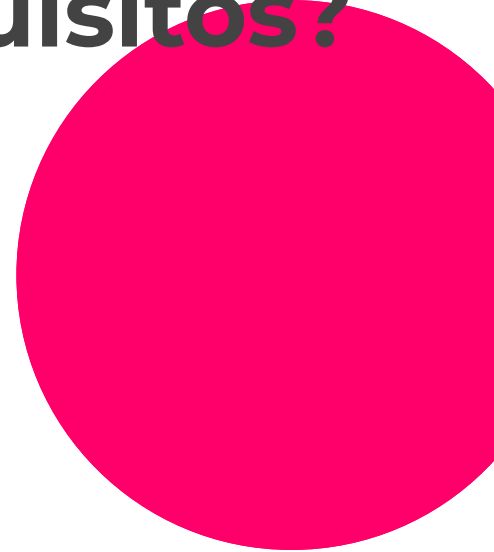
01

Fácil de alterar

02

Suporte de Data
Platform

03

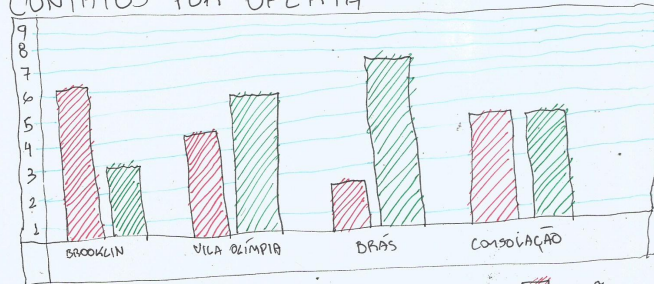


ANÁLISE DE DESEMPENHO

DE 12/03/18 A 12/04/18

CIDADE: SÃO PAULO

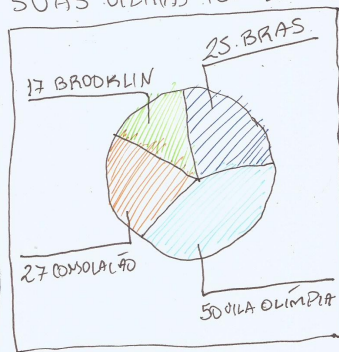
CONTATOS POR OFERTA



BAIRROS PRÓXIMOS



SUAS OFERTAS POR BAIRRO



Enxugamos a versão antiga do Relatório de Desempenho

O resultado foi esse belo desenho

POC1.ipynb

menubie Fix filename references and store data using %store 1f96ac6 on Jun 24, 2018

1 contributor

1149 lines (1148 sloc) 336 KB

```
In [1]: import matplotlib
import numpy as np
import pandas as pd

pd.set_option('display.max_columns', 500)
pd.set_option('float_format', '{:f}'.format)

get_ipython().magic('matplotlib inline')
```

Objetivo

Temos por objetivo entender se temos informação suficiente e habilidade para implementar o desenho do relatório de Análise de Desempenho proposto pelo Robin [neste arquivo](#).

Qual a ideia do relatório?

- Os anúncios ativos no dia x do anunciante tal tiveram esse tanto de leads para o período y;
- Para cada agrupamento (grupo 1, 2, 3) temos um trio de informações:
 - Distribuição da Carteira;
 - Análise da Concorrência;
 - Parking lot: Perfil médio da concorrência;
 - Top 15 bairros da cidade de atuação.

Grupos

(Nível 1, Nível 2, Nível 3)

- Nível 1: Venda, locação, ambos
- Nível 2: Comercial, residencial
- Nível 3: Verticais (casa) e horizontais (resto)

Estamos pensando em escolher apenas uma combinação para grupo foco.

Localização

(Estado, Cidade, Bairro)

Relação oferta e demanda

Basicamente, fazemos para cada localização a somatória da quantidade de leads recebida no período dividido pelo total de dias ativos

```
In [3]: %store -r df
df.head()
```

```
Out[3]:
```

	codimobiliaria	cliente_codcliente	cliente_nomefantasia	anunciantetipo	codimovel	datacriacao	nomeagrupamentotip
0	57750.000000	3092630	Manoel Santos Pereira	b2c	17953191	2018-03-28 15:27:00.690	Residencial
1	57750.000000	3092630	Manoel Santos Pereira	b2c	17743996	2018-03-12 19:31:11.760	Residencial

Pandas para validar amostras dos nossos dados no Data Lake



Paradigma Descentralizado dos Dados



```
In [3]: import findspark
findspark.init()

import sys

from pyspark.sql import functions as fn
from pyspark.sql.types import LongType, BooleanType

START_DATE = "2018-09-01"
END_DATE = "2018-09-30"
TABLE_NAME = "advertiser_advisor.rd_{}_{}" \
              .format(START_DATE, END_DATE) \
              .replace("-", "_")

print(TABLE_NAME)

##### start - spark.py
from pyspark.sql import SparkSession

def create_spark_context(name, cores="8", memory="16G", timeout="10000",
                        hive="", spark=""):

    return SparkSession.builder \
        .appName(name) \
        .master(spark) \
        .config("spark.cores.max", cores) \
        .config("spark.executor.memory", memory) \
        .config("spark.sql.broadcastTimeout", timeout) \
        .config("hive.metastore.uris", hive) \
        .config('spark.hadoop.mapreduce.fileoutputcommitter.algorithm.version',
                '2') \
        .config('spark.hadoop.mapreduce.fileoutputcommitter.cleanup-failures.ignored',
```

Adaptação do script da análise para PySpark no Jupyter Hub



**Jupyter
Hub**


**Spark
Submit**

**Livy
Spark**



Spark Submit local

- O que é o Spark Submit?
- Por que optamos em usá-lo ao invés do meu scriptoso?
- Dados trafegando do cluster para a minha máquina.



Spark Submit no k8s

- Colocar o spark-submit no K8s significaria perder latência de rede.
- Dockerização estava impossível.
- Mas precisávamos muito para conseguir fazer um teste de stress de todos os nossos dados.



**olá, livy
spark**

- Livy-Spark vs Spark Submit
- API rest para subir os pythinho pro Spark
- Não precisaríamos nos preocupar mais com a configuração de deploy pra usar spark-submit e afins, apenas chamamos a API deles

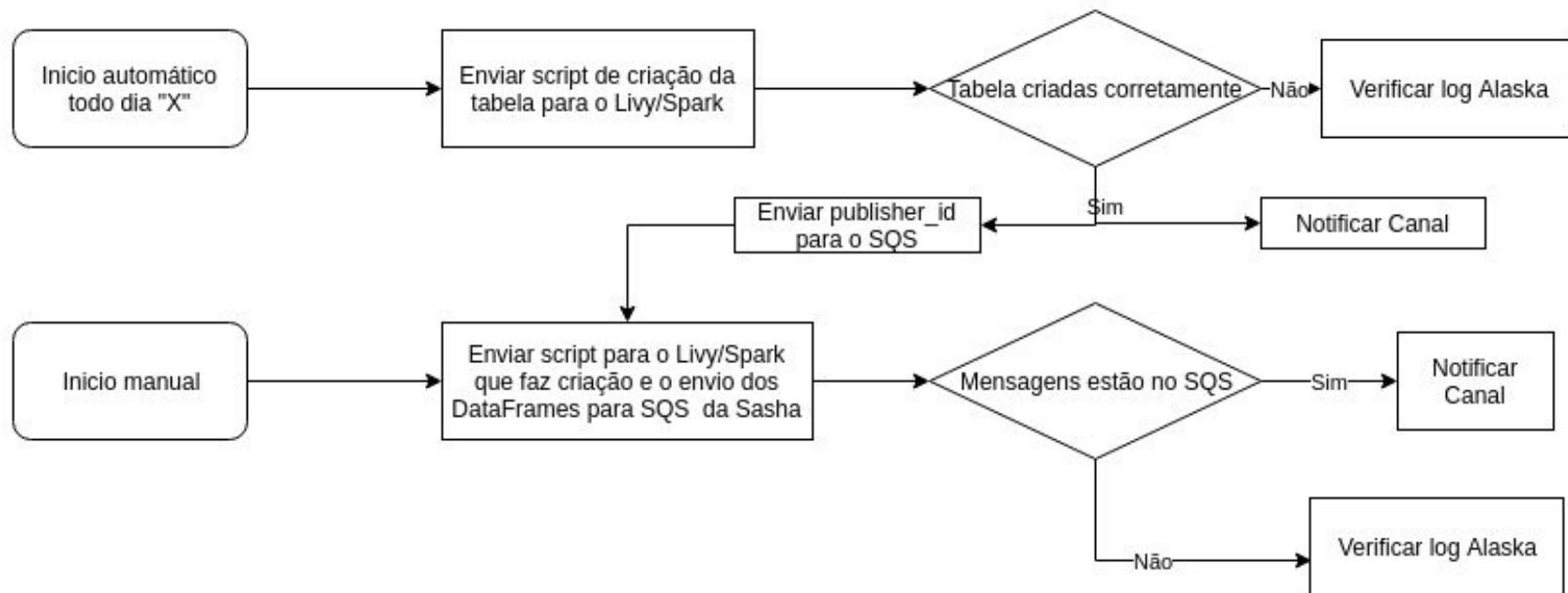
**Jupyter
Hub**



**Livy
Spark**

Alaska

Criação da tabela agregada de listings com os leads do periodo





O que o RuPaul **entrega**?



canal **pro**

Olá, Imobiliária Feliz

Você está recebendo o seu **relatório de desempenho**
referente a Fevereiro de 2019

Nele você encontrará informações precisas para analisar sua
performance, saber quais são seus pontos fortes e as
oportunidades de crescimento.

Veja o relatório

Abraços,
Equipe Grupo ZAP

E-mail para o anunciante com o seu relatório





Relatório de Desempenho

JUL 2019

O Relatório de desempenho é um documento que reúne informações importantes que ajudarão na análise da performance dos seus anúncios. Irá trazer mais clareza sobre seus pontos fortes e, assim, apresentar novas oportunidades de crescimento. Reserve um tempo para analisá-lo, e, se surgirem dúvidas, ou feedbacks, [fale com a gente](#).

Análise de desempenho

Como interpretar?

Podemos interpretar os gráficos da sessão de desempenho de duas maneiras diferentes. A primeira interpretação dá uma perspectiva sobre a sua concorrência: No portal o seu desempenho está melhor, pior ou igual ao desempenho geral?

A segunda interpretação está voltada para a comparação entre regiões, tipologias e faixas de preço: Qual bairro está performando melhor? E quanto ao total de dormitórios? E assim por diante.

Como é definida a sua concorrência?

A sua concorrência é composta por todos os anunciantes do portal, com ofertas nos bairros que você atende, no mesmo negócio (venda e/ou aluguel) e com os mesmos tipos de imóveis (apartamento, casa, terreno) que você trabalha.

Como é calculado o desempenho?

O desempenho é um índice criado para permitir que você avalie e analise a performance de suas ofertas no portal. Para calcular esse indicador, são consideradas três informações:

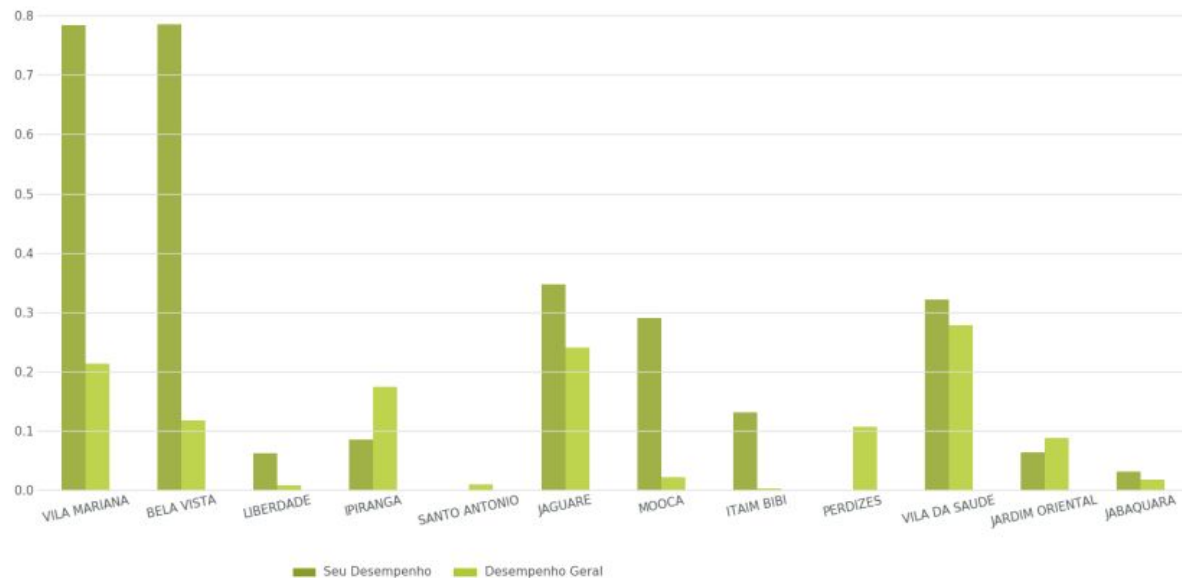
1. Total de contatos recebidos (mensagens, telefonemas, exibições de telefone)
2. Quantidade de ofertas publicadas
3. Quanto tempo essas ofertas ficaram ativas

Com base nesses fatores, geramos um índice que indica o retorno que você obteve com os anúncios publicados no período. A partir destes *insights* uma análise mais detalhada deverá ser feita para uma maior compreensão do comportamento apresentado.

Bairros em que você atua

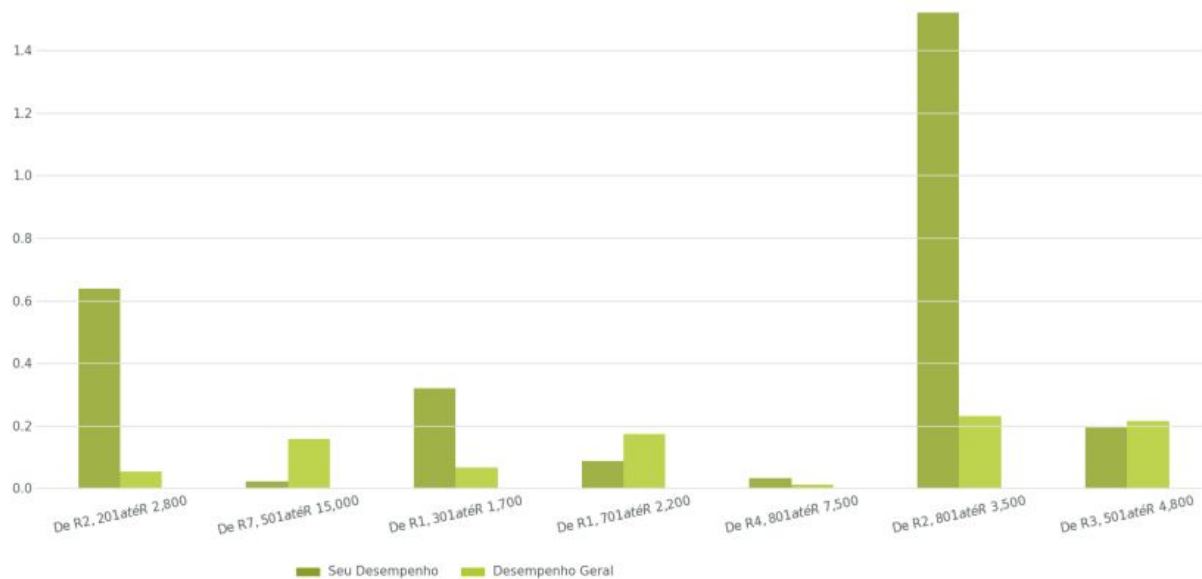
Veja em quais bairros você está indo bem e em quais há potencial para crescimento.

A análise está limitada a, no máximo, 15 bairros de você tem maior quantidade de imóveis e melhor desempenho.



Valor da locação

Descubra quais são as faixas de valor de aluguel que seus anúncios têm o melhor desempenho e identifique oportunidades para ajustar suas campanhas.





~1 hora

Processamento da Alaska



Qual a relação de PDFs gerados pelo Relatório de Desempenho?



This question is written in SQL.

ID DO ANUNCIANTE

VISUALIZATION

Table

Refresh

advertiser_id	business_name	start_date	portal	_col4
9aa47444-e7b3-92be-b2e8-9413ce70a69c	SHIRLEY ROSANA FERREIRA	2019-08-01	VIVAREAL	https://s3.amazonaws.com/qa-advisor-sauron/rupaul/08-19/9aa47444-e7b3-92be-b2e8-9413ce70a69c.pdf
9cb511dd-913c-4a0d-3805-99e3375f5e8a	MORATA NEGÓCIOS IMOBILIÁRIOS LTDA	2019-08-01	VIVAREAL	https://s3.amazonaws.com/qa-advisor-sauron/rupaul/08-19/9cb511dd-913c-4a0d-3805-99e3375f5e8a.pdf
4e2c1290-05df-509e-f97a-8b51e1ea0c0a	BENS DE RAIZ IMOVEIS	2019-08-01	ZAP	https://s3.amazonaws.com/qa-advisor-sauron/rupaul/08-19/4e2c1290-05df-509e-f97a-8b51e1ea0c0a.pdf
75381ff5-5e6a-3962-1422-533250adfd9	MASUDA CORRETORA DE IMÓVEIS	2019-08-01	VIVAREAL	https://s3.amazonaws.com/qa-advisor-sauron/rupaul/08-19/75381ff5-5e6a-3962-1422-533250adfd9.pdf
a0b7bbca-7503-23e0-11b2-e11a70b3a121	ANDERSON VICENTE DA SILVA	2019-08-01	VIVAREAL	https://s3.amazonaws.com/qa-advisor-sauron/rupaul/08-19/a0b7bbca-7503-23e0-11b2-e11a70b3a121.pdf
5d52f862-ba33-20a8-8329-cd86070e3f7e	MARIO TAKEO HIRABAE	2019-08-01	VIVAREAL	https://s3.amazonaws.com/qa-advisor-sauron/rupaul/08-19/5d52f862-ba33-20a8-8329-cd86070e3f7e.pdf
dad4edbc-9ae4-6a87-92d1-652b2acab683	FLORISVALDO DE SOUZA SALVIONE	2019-08-01	ZAP	https://s3.amazonaws.com/qa-advisor-sauron/rupaul/08-19/dad4edbc-9ae4-6a87-92d1-652b2acab683.pdf
93680c01-f456-dbbe-cc17-c776d2e825a5	JOSE VALDINEIS LOPES PINHEIRO	2019-08-01	VIVAREAL	https://s3.amazonaws.com/qa-advisor-sauron/rupaul/08-19/93680c01-f456-dbbe-cc17-c776d2e825a5.pdf
2cbdb801-b147-733e-f709-0cf39821e628	MURILLO NIERO DAS NEVES	2019-08-01	VIVAREAL	https://s3.amazonaws.com/qa-advisor-sauron/rupaul/08-19/2cbdb801-b147-733e-f709-0cf39821e628.pdf
656ef420-736f-13c7-94c8-d09bec293b69	LUIZ FERNANDO DO PRADO BRANDÃO	2019-08-01	VIVAREAL	https://s3.amazonaws.com/qa-advisor-sauron/rupaul/08-19/656ef420-736f-13c7-94c8-d09bec293b69.pdf

PDFs para os
nossos usuários de
Relacionamento



+90%

Usuários cadastrados no Metabase aumentaram de 138 para 263





Search...

Ask a question



New question

SAVE

DATA: **Dados para o Relatório de Desempenho** FILTERED BY: + Add filters to narrow your answer VIEW: **Raw data** GROUPED BY: + Add a grouping ...

VISUALIZATION

Table

Refresh

Showing first 2,000 rows

ID do Anunciante	Razão Social	Tipo da Transação	Cidade	Número de Quartos	Tipo do Imóvel	CNPJ/CPF	E-mail	Total de Leads	Leads de Ligação	Leads de Mensagem
006aefaa-6616-6ac6-ae31-90aaef4eaede	TOMMY CARLOS DE ALMEIDA VASCONCELOS	Venda	GUARULHOS	2 quartos	Apartamento	21692531816	boscoconsultorimobiliario@gmail.com			
006aefaa-6616-6ac6-ae31-90aaef4eaede	TOMMY CARLOS DE ALMEIDA VASCONCELOS	Venda	GUARULHOS	2 quartos	Apartamento	21692531816	boscoconsultorimobiliario@gmail.com	2	0	
006aefaa-6616-6ac6-ae31-90aaef4eaede	TOMMY CARLOS DE ALMEIDA VASCONCELOS	Venda	SANTA ISABEL	Nenhum quarto	Lote Terreno Residencial	21692531816	boscoconsultorimobiliario@gmail.com	1	0	
006aefaa-6616-6ac6-ae31-90aaef4eaede	TOMMY CARLOS DE ALMEIDA VASCONCELOS	Venda	GUARULHOS	3 quartos	Casa	21692531816	boscoconsultorimobiliario@gmail.com			
006aefaa-6616-6ac6-ae31-90aaef4eaede	TOMMY CARLOS DE ALMEIDA VASCONCELOS	Venda	GUARULHOS	2 quartos	Apartamento	21692531816	boscoconsultorimobiliario@gmail.com	3	0	
006aefaa-6616-6ac6-ae31-90aaef4eaede	TOMMY CARLOS DE ALMEIDA VASCONCELOS	Venda	GUARULHOS	2 quartos	Casa	21692531816	boscoconsultorimobiliario@gmail.com			
006aefaa-6616-6ac6-ae31-90aaef4eaede	TOMMY CARLOS DE ALMEIDA VASCONCELOS	Venda	GUARULHOS	2 quartos	Apartamento	21692531816	boscoconsultorimobiliario@gmail.com			
006aefaa-6616-6ac6-ae31-90aaef4eaede	TOMMY CARLOS DE ALMEIDA VASCONCELOS	Venda	GUARULHOS	2 quartos	Apartamento	21692531816	boscoconsultorimobiliario@gmail.com	1	0	
006aefaa-6616-6ac6-ae31-90aaef4eaede	TOMMY CARLOS DE ALMEIDA VASCONCELOS	Venda	GUARULHOS	2 quartos	Apartamento	21692531816	boscoconsultorimobiliario@gmail.com	9	0	
006aefaa-6616-6ac6-ae31-90aaef4eaede	TOMMY CARLOS DE ALMEIDA VASCONCELOS	Venda	GUARULHOS	2 quartos	Apartamento	21692531816	boscoconsultorimobiliario@gmail.com	1	0	
006aefaa-6616-6ac6-ae31-90aaef4eaede	TOMMY CARLOS DE ALMEIDA VASCONCELOS	Venda	GUARULHOS	2 quartos	Condomínio Residencial	21692531816	boscoconsultorimobiliario@gmail.com	2	0	
006aefaa-6616-6ac6-ae31-90aaef4eaede	TOMMY CARLOS DE ALMEIDA VASCONCELOS	Venda	GUARULHOS	3 quartos	Apartamento	21692531816	boscoconsultorimobiliario@gmail.com	1	0	
006aefaa-6616-6ac6-ae31-90aaef4eaede	TOMMY CARLOS DE ALMEIDA VASCONCELOS	Venda	GUARULHOS	2 quartos	Apartamento	21692531816	boscoconsultorimobiliario@gmail.com	1	0	
006aefaa-6616-6ac6-ae31-90aaef4eaede	TOMMY CARLOS DE ALMEIDA VASCONCELOS	Venda	GUARULHOS	3 quartos	Apartamento	21692531816	boscoconsultorimobiliario@gmail.com	2	0	
006aefaa-6616-6ac6-ae31-90aaef4eaede	TOMMY CARLOS DE ALMEIDA VASCONCELOS	Venda	GUARULHOS	2 quartos	Apartamento	21692531816	boscoconsultorimobiliario@gmail.com			
03912fe5-c077-2ea2-a6d0-9b29475fac...	PAMELA ANDRADE LOURENCO PEREDA	Venda	SAO PAULO	2 quartos	Apartamento	29033259877	pamela@imoveimoveis.com.br			

Qual o total de leads agrupado por mês e tipo de imóvel filtrado por anunciante?

LEADS [NEW]



This question is written in SQL.

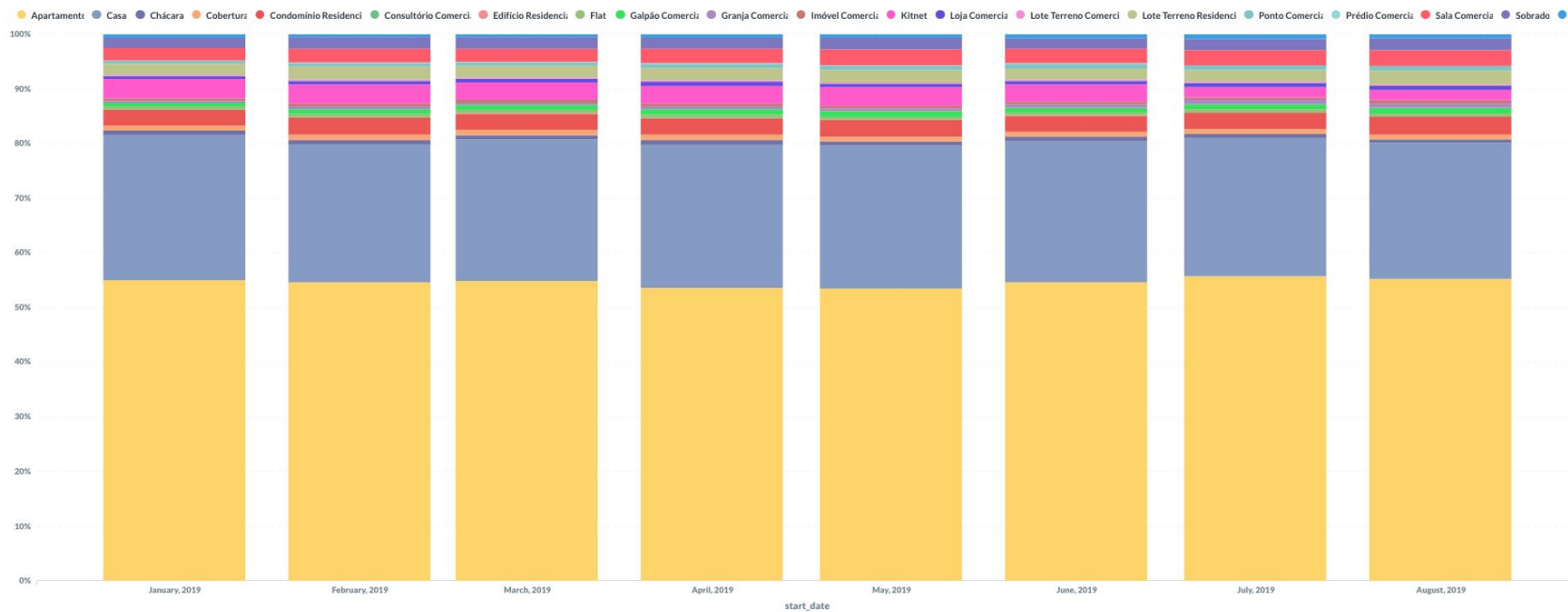
ID do Anunciante

OPEN EDITOR

VISUALIZATION

Bar

Refresh



Imobiliária Feliz... Feliz!



grupo **ZAP**

Relatório de Desempenho

JUL 2019

O Relatório de desempenho é um documento que reúne informações importantes que ajudarão na análise da performance dos seus anúncios. Irá trazer mais clareza sobre seus pontos fortes e, assim, apresentar novas oportunidades de crescimento. Reserve um tempo para analisá-lo, e, se surgirem dúvidas, ou feedbacks, [fale com a gente](#).



RuPaul + **Data Platform**





RuPaul



AVM



Lich King



Scores do ranking



Fipe ZAP



Análises



Snapshots



Extrações



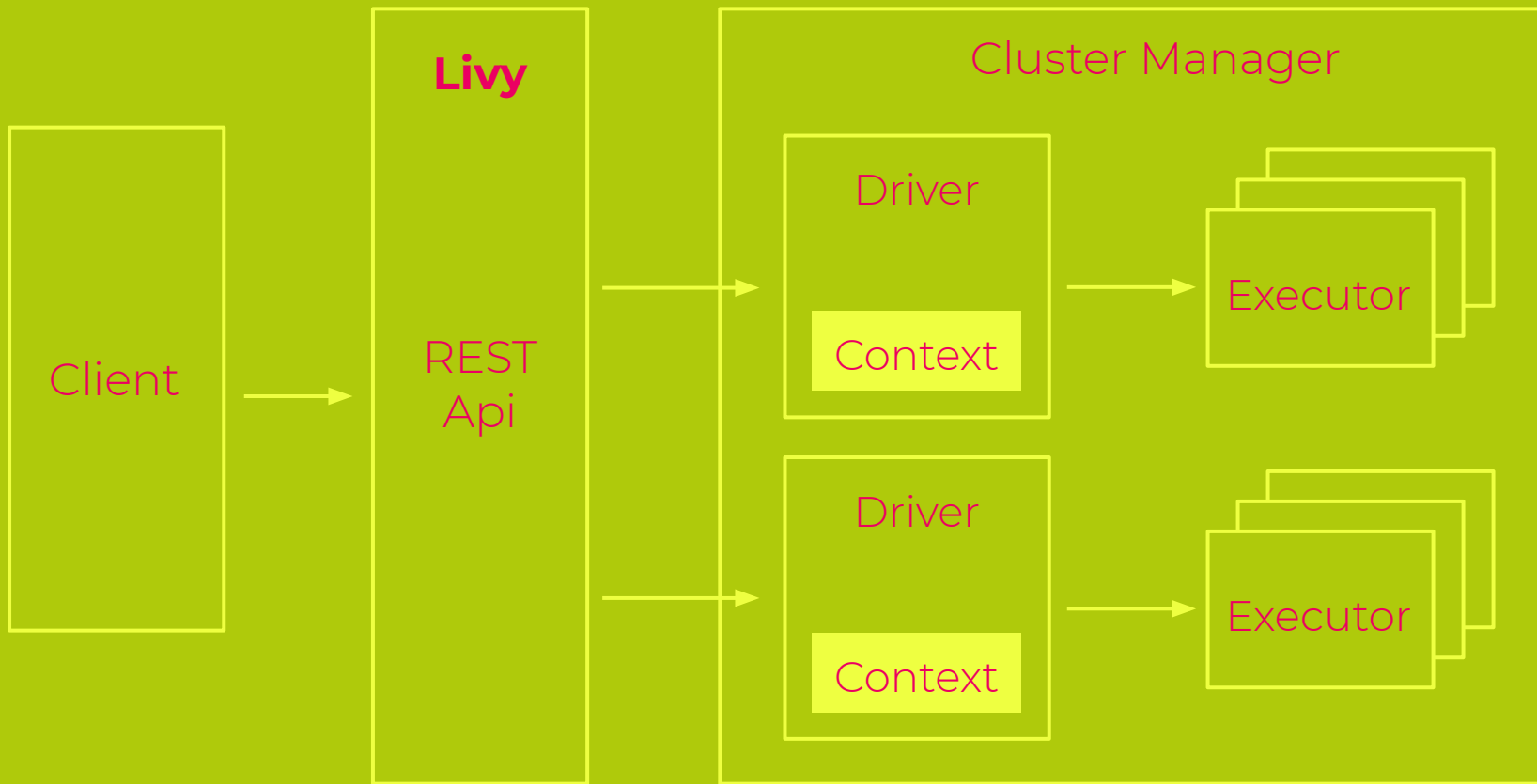
~800T/mês

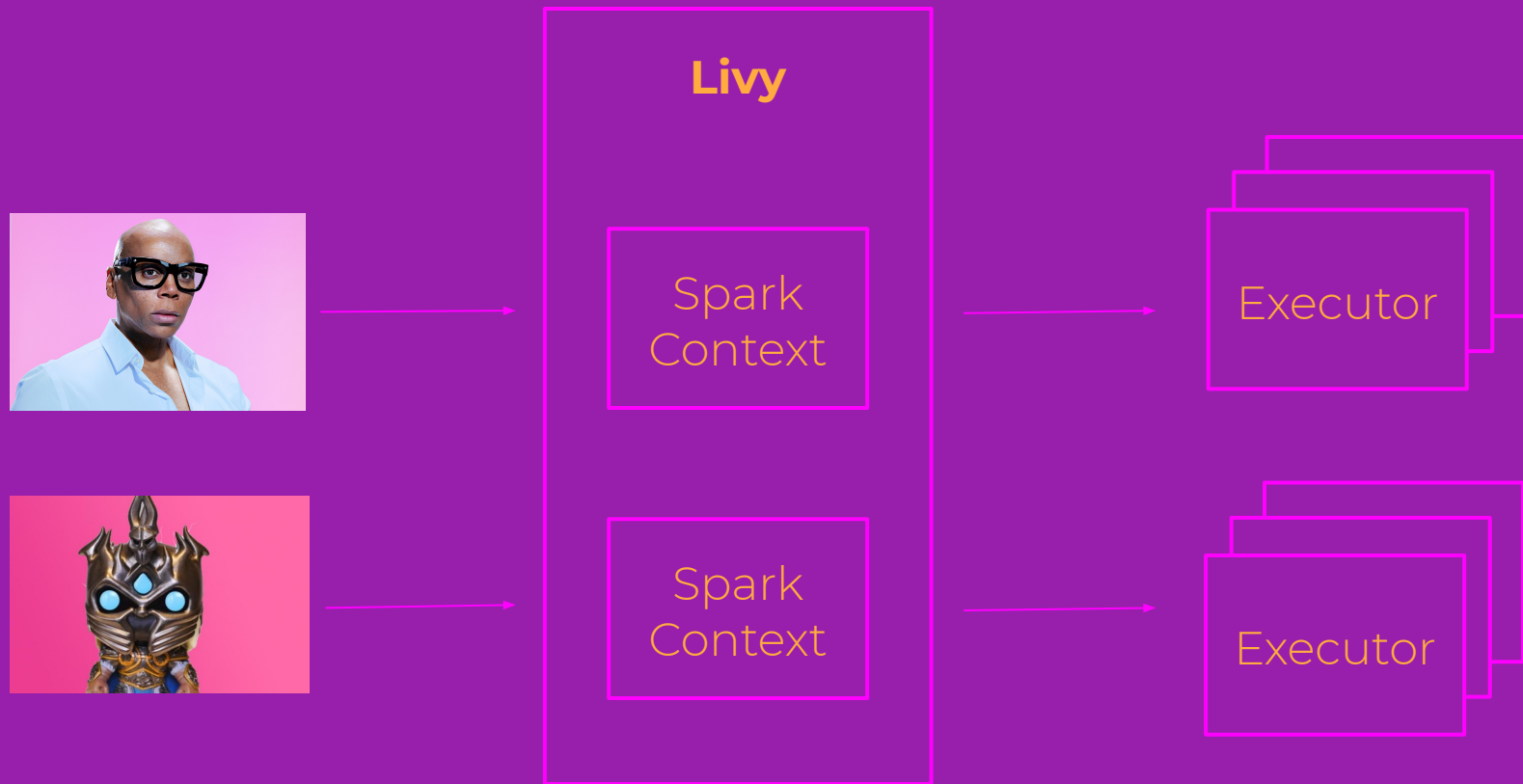




Arquitetura Efêmera de Spark com Livy no Kubernetes







```
{  
  "kind": "pyspark",  
  "name": "rupaul",  
  "driverMemory": "1G",  
  "executorCores": 7,  
  "executorMemory": "50G",  
  "numExecutors": 5  
}
```



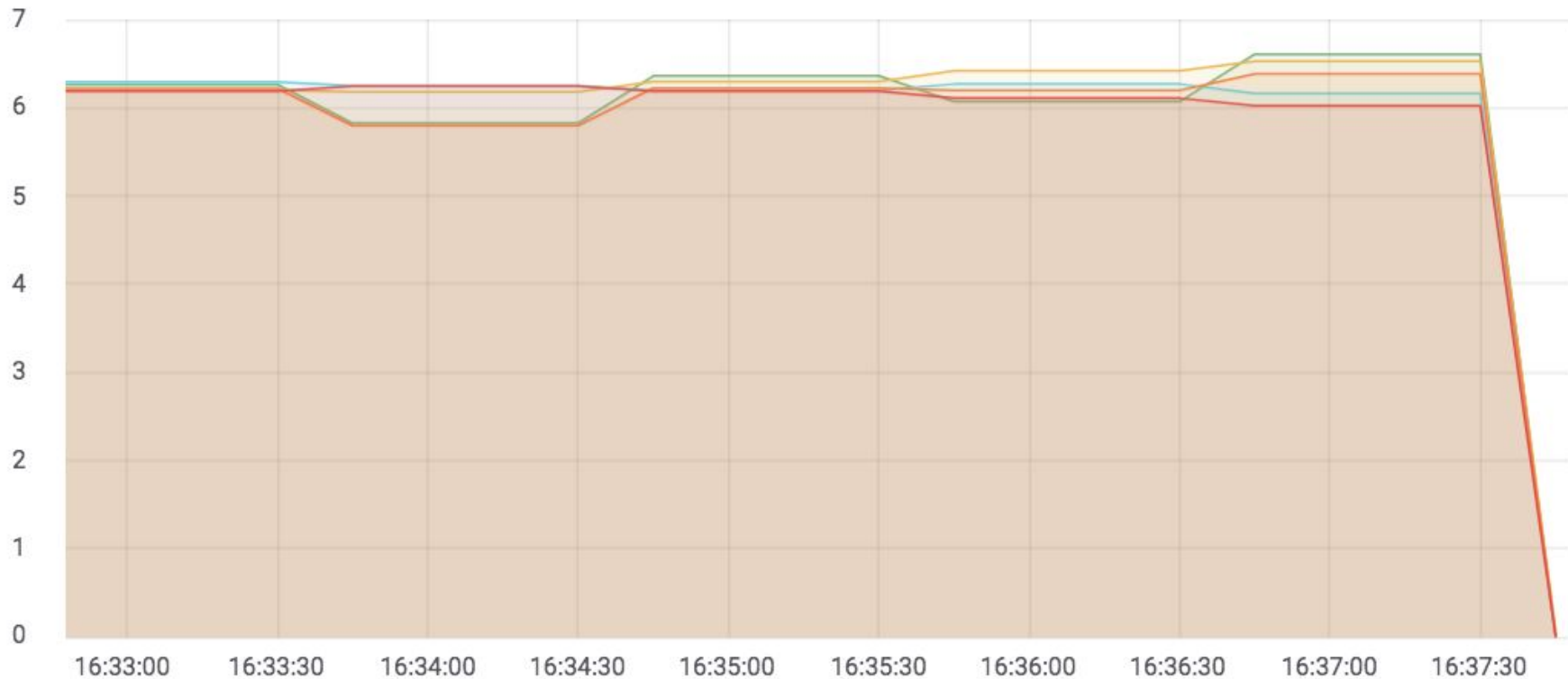
```
{  
  "code": "df = spark.sql('select * from listings')  
          df.write('s3a://datalake/test')"  
}
```



Métricas Para o Usuário



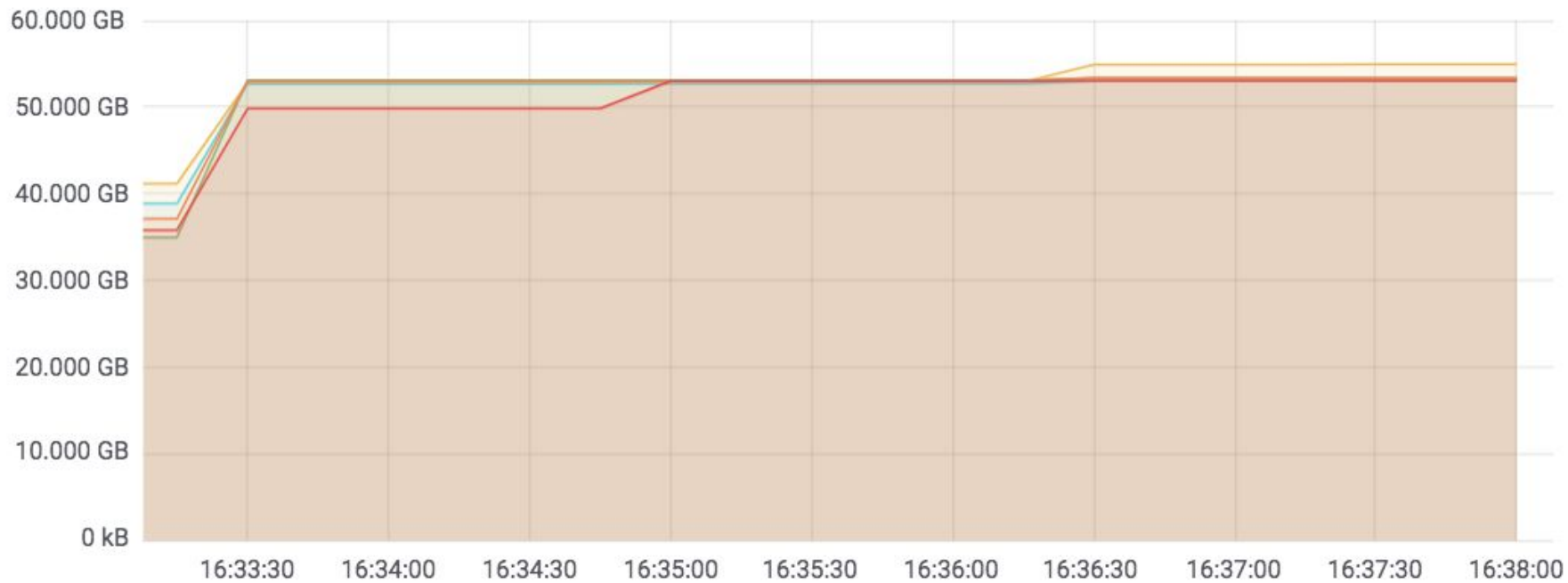
CPU



Legend:

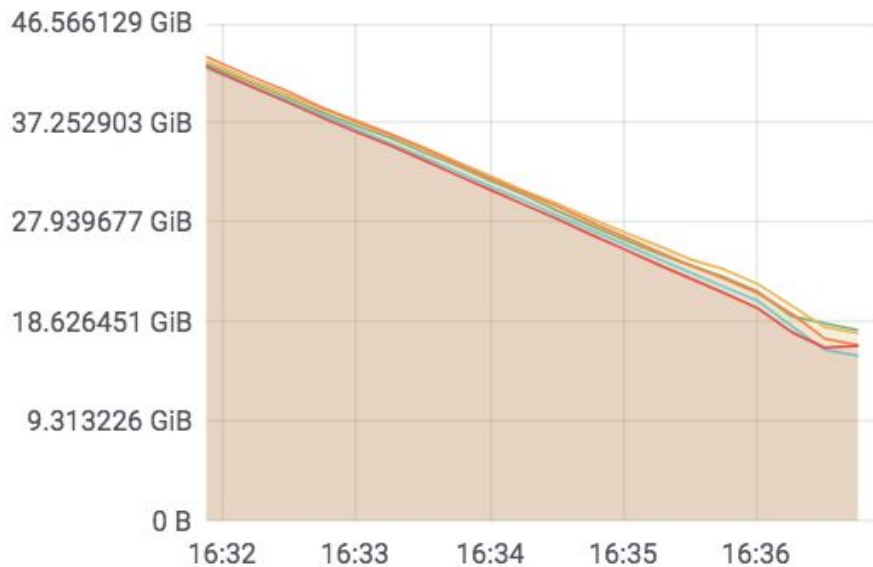
- rupaul-1570735600135-exec-1
- rupaul-1570735600383-exec-2
- rupaul-1570735600450-exec-3
- rupaul-1570735600514-exec-4
- rupaul-1570735600565-exec-5

Memory



rupaul-1570735600135-exec-1 rupaul-1570735600383-exec-2 rupaul-1570735600450-exec-3 rupaul-1570735600514-exec-4
rupaul-1570735600565-exec-5

Node Disk Space AVAILABLE



Node Disk Space AVAILABLE

Metric ▾	Total
rupaul-1570735600565-exec-5	16.37 GiB
rupaul-1570735600514-exec-4	16.60 GiB
rupaul-1570735600450-exec-3	15.47 GiB
rupaul-1570735600383-exec-2	17.16 GiB
rupaul-1570735600135-exec-1	17.97 GiB



Vantagens



- **Remover concorrência entre jobs**
- **Não ter um cluster de pé 24/7**
- **Clusters em namespaces diferentes**
- **Spot Instances**
- **Gerenciar job spark via api REST**



Desvantagens



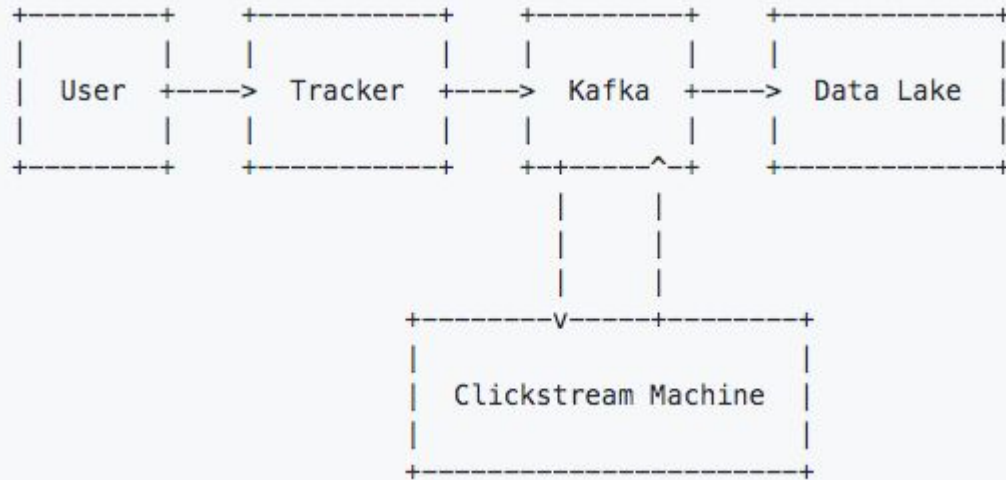
- **Suporta apenas 10 drivers concorrentes**
- **Não controla duplicação de contextos**



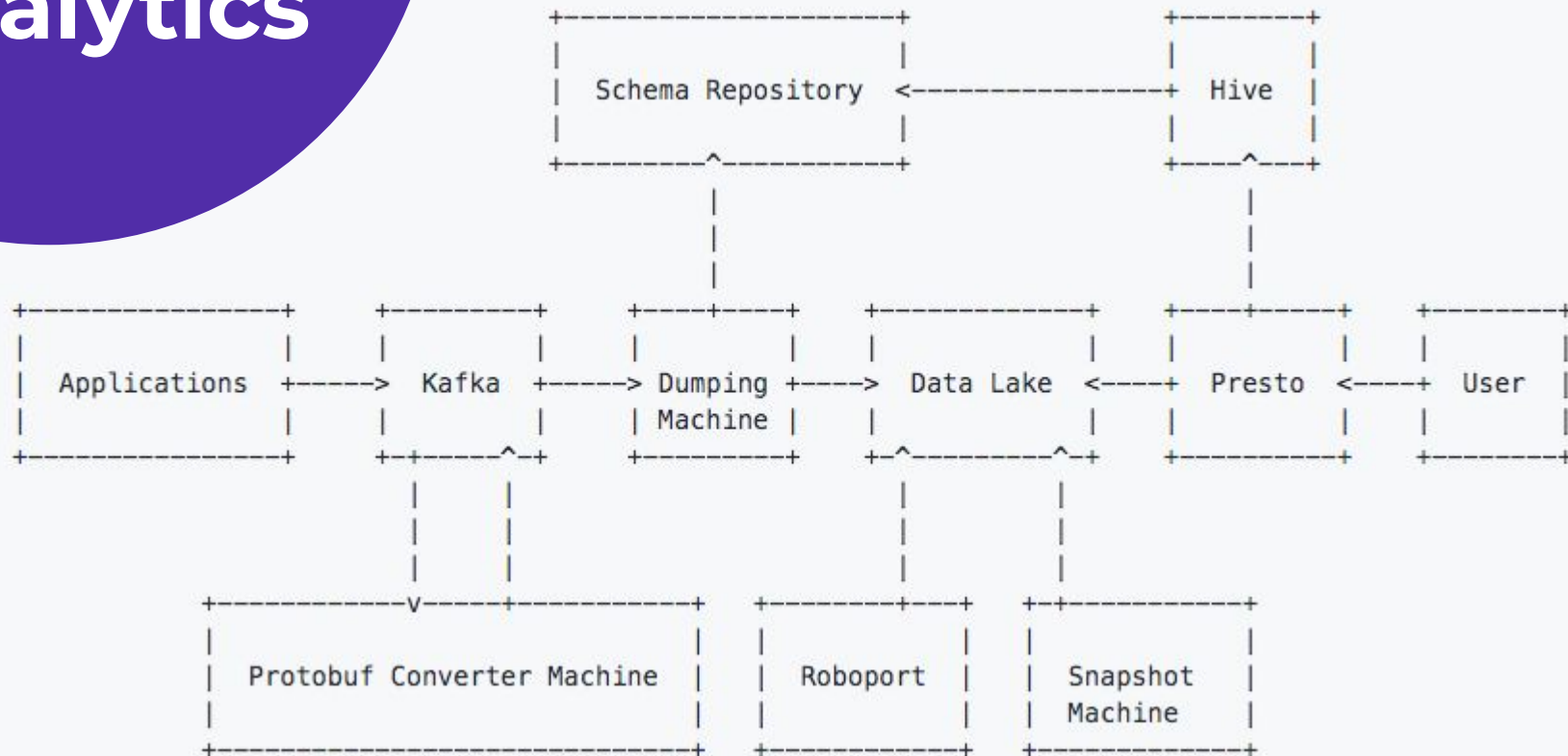
Stack Data Platform

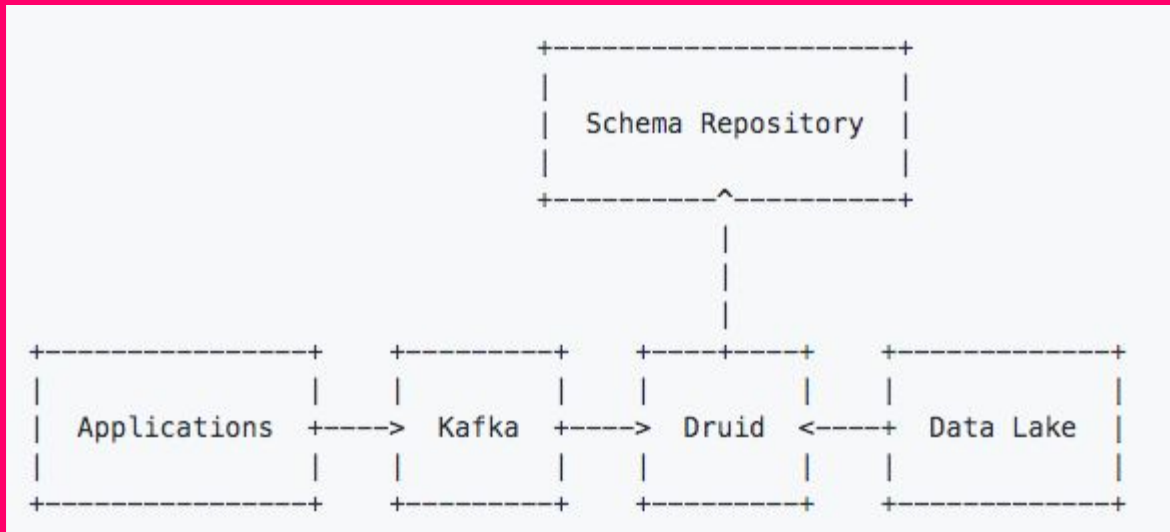


Clickstream



Consumer Analytics





Application Analytics





Thanks!





madu@gruposzap.com
naka@gruposzap.com

vagas.gruposzap.com
medium.com/tech-gruposzap

