



Arquitetura de dados na Birdie





Considerações iniciais

Quem sou eu ?

- Marcos Camargo
- BCC 015
- Engenheiro de dados na Birdie

Sobre a Birdie?

Nossa missão é melhorar a experiência de escolha de produtos para consumidores

- Descobrir novos produtos
- Validar produtos de acordo com a necessidade
- Tomar decisões com mais confiança
- <https://birdie.ai/>



Sumário

- E aí, o que o Birdie faz?
- Um pouco sobre nossa arquitetura e processo de dados
- ETL na coleta de ofertas de *e-commerces*
- Desafios encontrados
- Considerações finais

Usuários passam
por centenas de abas

**para conseguir fazer
uma escolha de compra.**



Segundo o Google

Usuários passam em média **55 dias** em uma jornada de compra

Passam por **185 touchpoint** diferentes

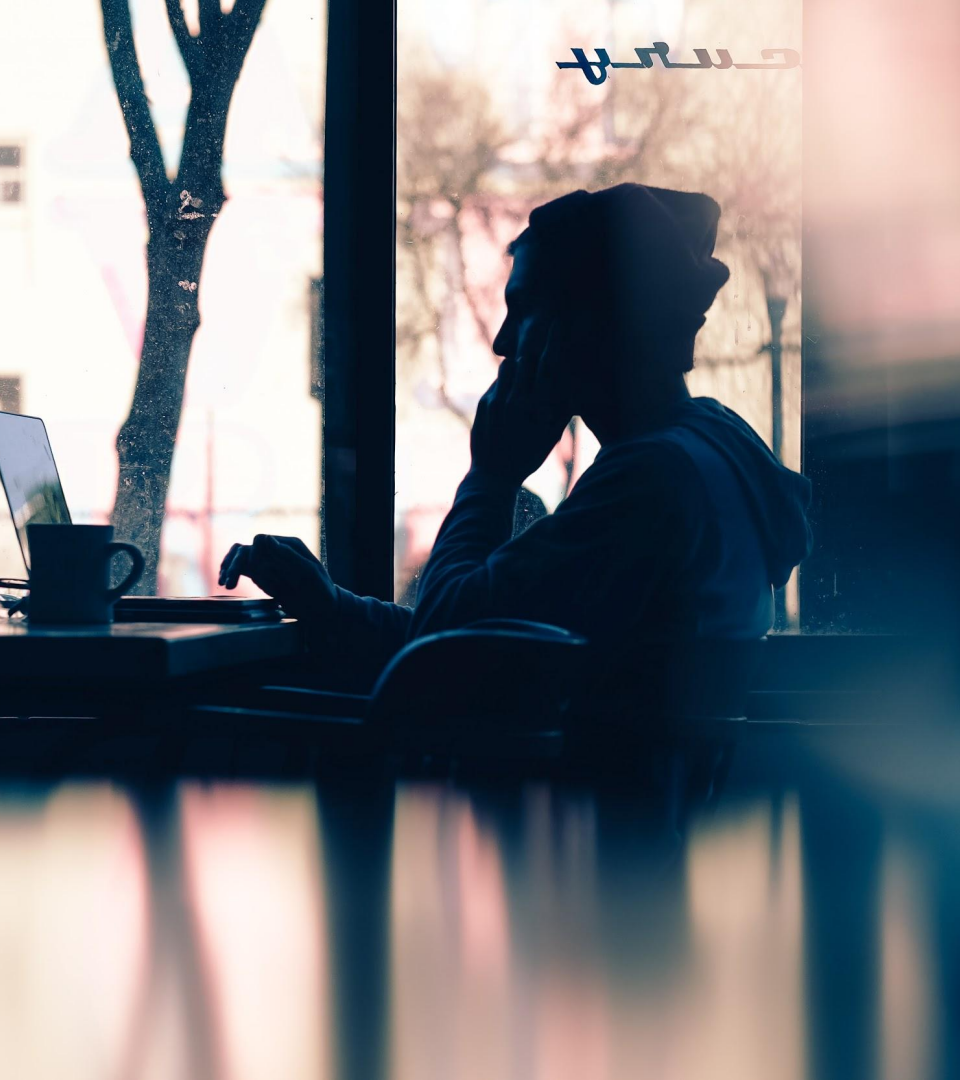
E ainda assim **34%** abandona o carrinho de compra por
não se sentir seguros o suficiente para comprar

As pessoas se frustram durante
o processo de pesquisa e comparação de produtos

**por não terem clareza
dos critérios necessários
para se fazer uma boa escolha.**

E o que os consumidores querem?

Ver informações que façam sentido pra ele e para o uso que ele vai fazer com o produto.



BRUNO

O que a Birdie faz?

Como podemos reduzir a complexidade, o tempo e o esforço nas compras online?

Construímos uma base de conhecimento de produtos.

Para aprender o que os consumidores falam e entregar uma boa experiência de compra, mais clara e mais relevante.



**E como
fazemos isso?**



Principais produtos

Busca semântica de produtos

- Tradução de buscas utilizando linguagem natural para aspectos extraídos de produtos

Rankeamento de produtos

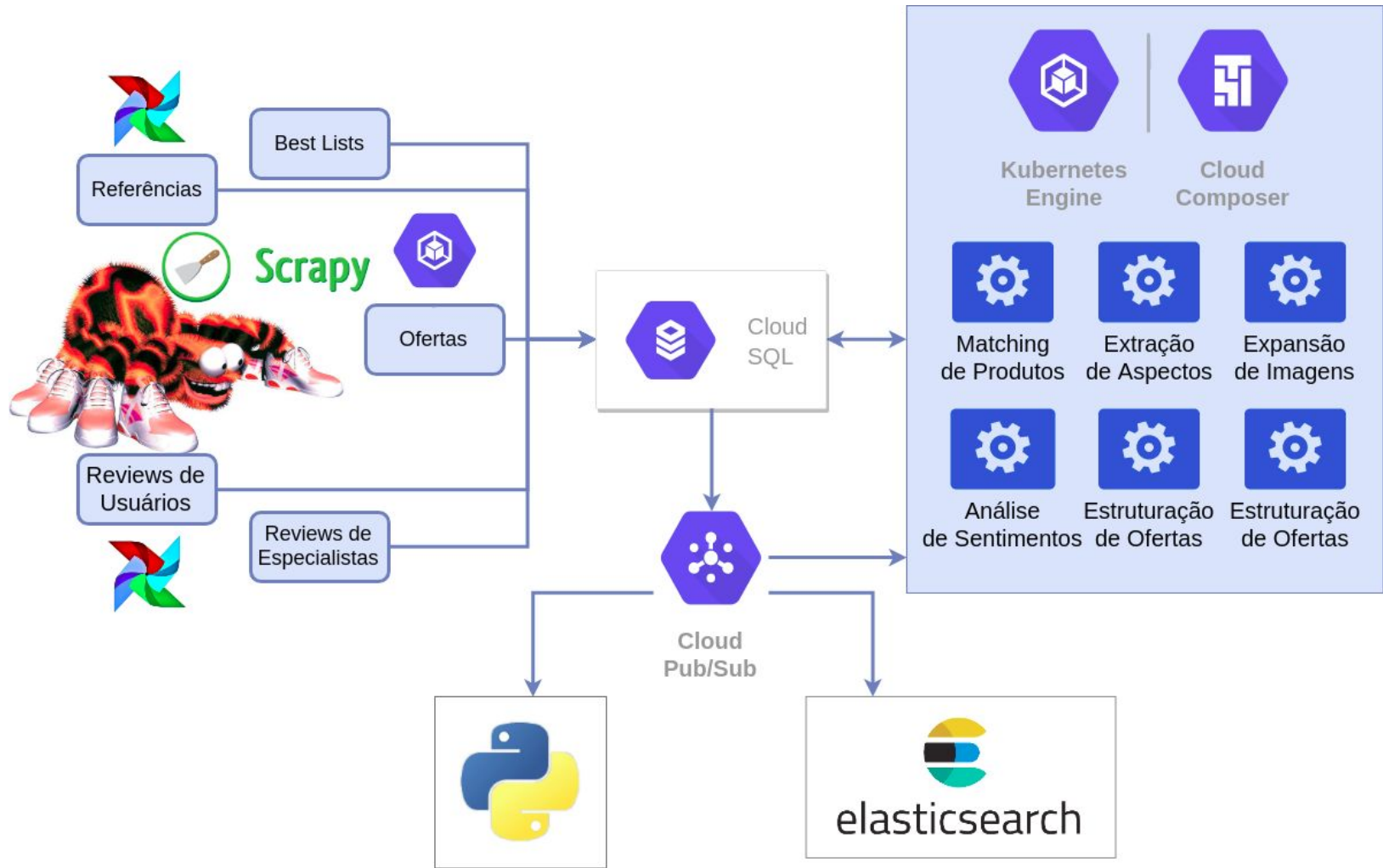
- Geração de notas com base nas informações coletadas sobre os produtos



Dados coletados

- Ofertas
 - Informações básicas
 - Preços
 - Imagens
 - Perguntas e respostas
- Especificações técnicas
- Reviews de especialistas
- Best Lists
- Reviews de usuários

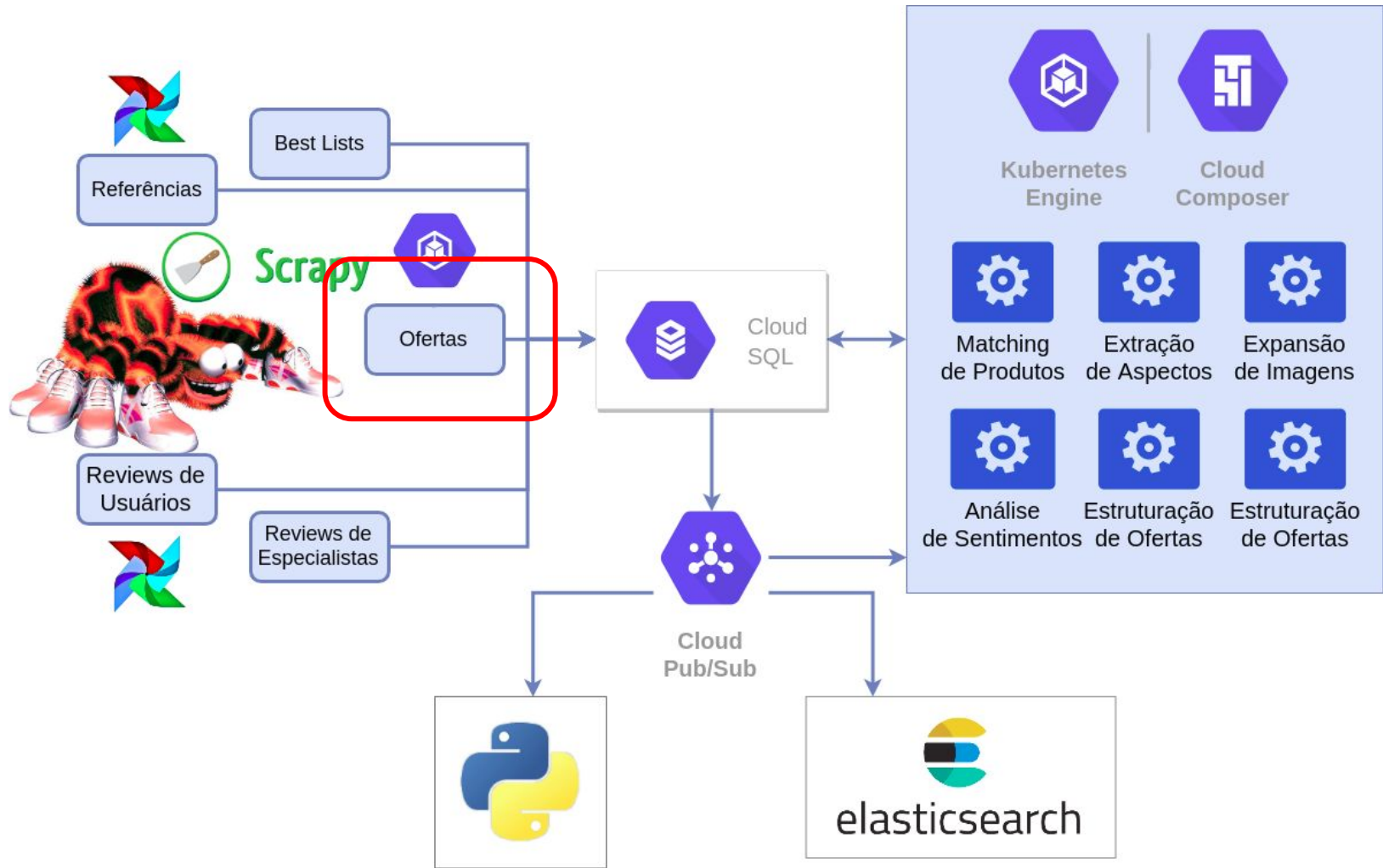
Nossa arquitetura de dados





Tecnologias utilizadas

- PostgreSQL
- Elasticsearch
- Airflow
- Kubernetes
- GitLab CI/CD
- Google Cloud Platform
- Python
- Flask



Sony - 65" Class - LED - X900F Series - 2160p - Smart - 4K Ultra HD TV with HDR

Model: XBR65X900F SKU: 6179726

★★★★★ 4.8 (1,893 Customers) | 4.0 (4 Experts) | 130 Answered Questions



[Price Match Guarantee](#)

\$1,299.99 or **\$54.17/mo.***

Save \$300

Was \$1,599.99

suggested payments with
24-Month Financing

[Show me how >](#)

Save \$30 on Sling TV and STARZ (new subscribers only)

Hot offer Save \$50 on TV Connect and Setup

Open-Box: from \$1,117.99

Protect your product

☆☆☆☆☆ Not yet reviewed

2 Years

\$169.99

5 Years

\$329.99

No plan

selected

[Learn about Standard Protection Plans](#)

Delivery: [Unavailable for 00820 >](#)

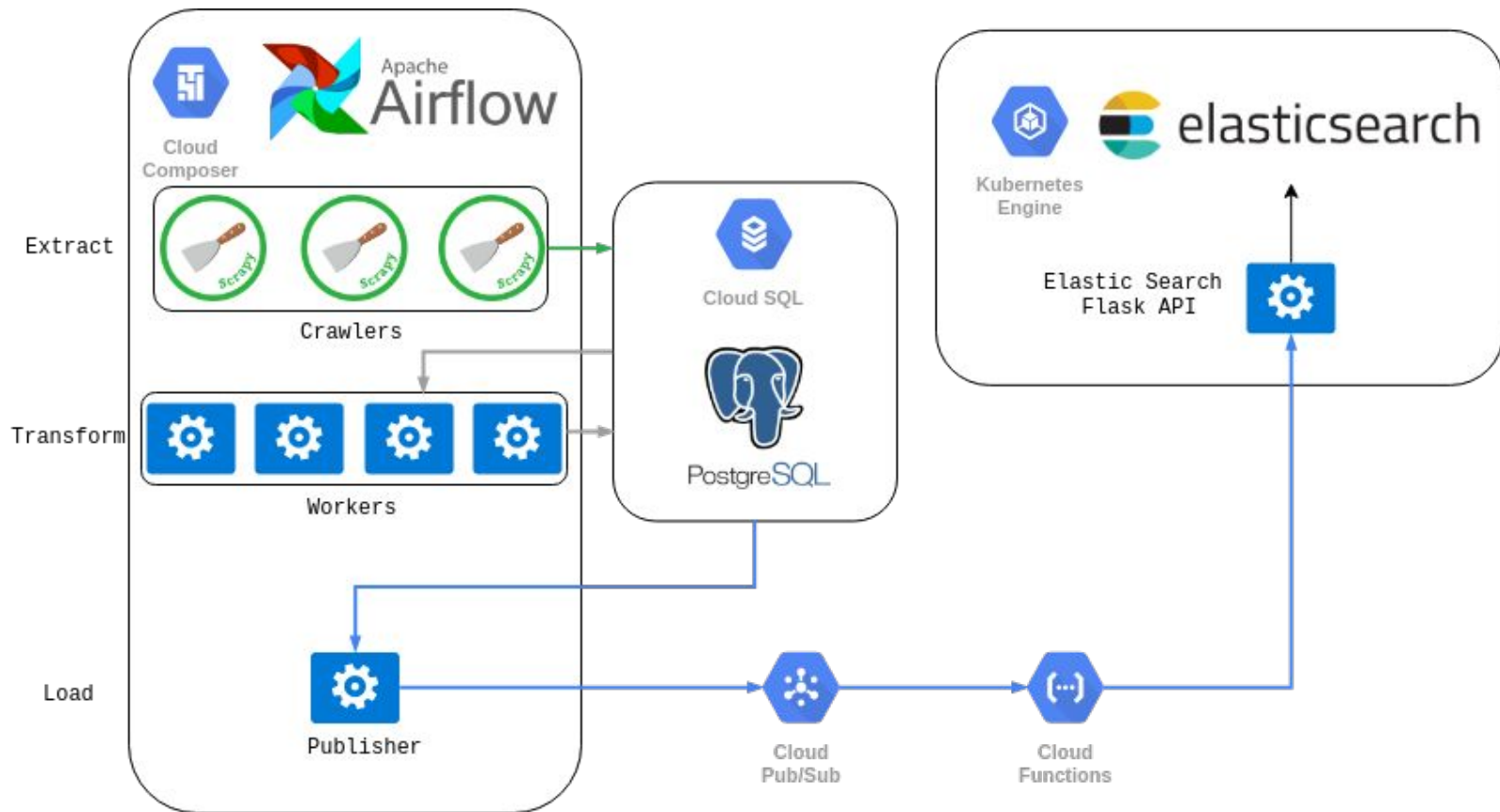
Store Pickup: [Unavailable within 250 miles of Plaza Carolina Mall Update Location >](#)

[On Display at Plaza Carolina Mall](#) [Change Store](#)

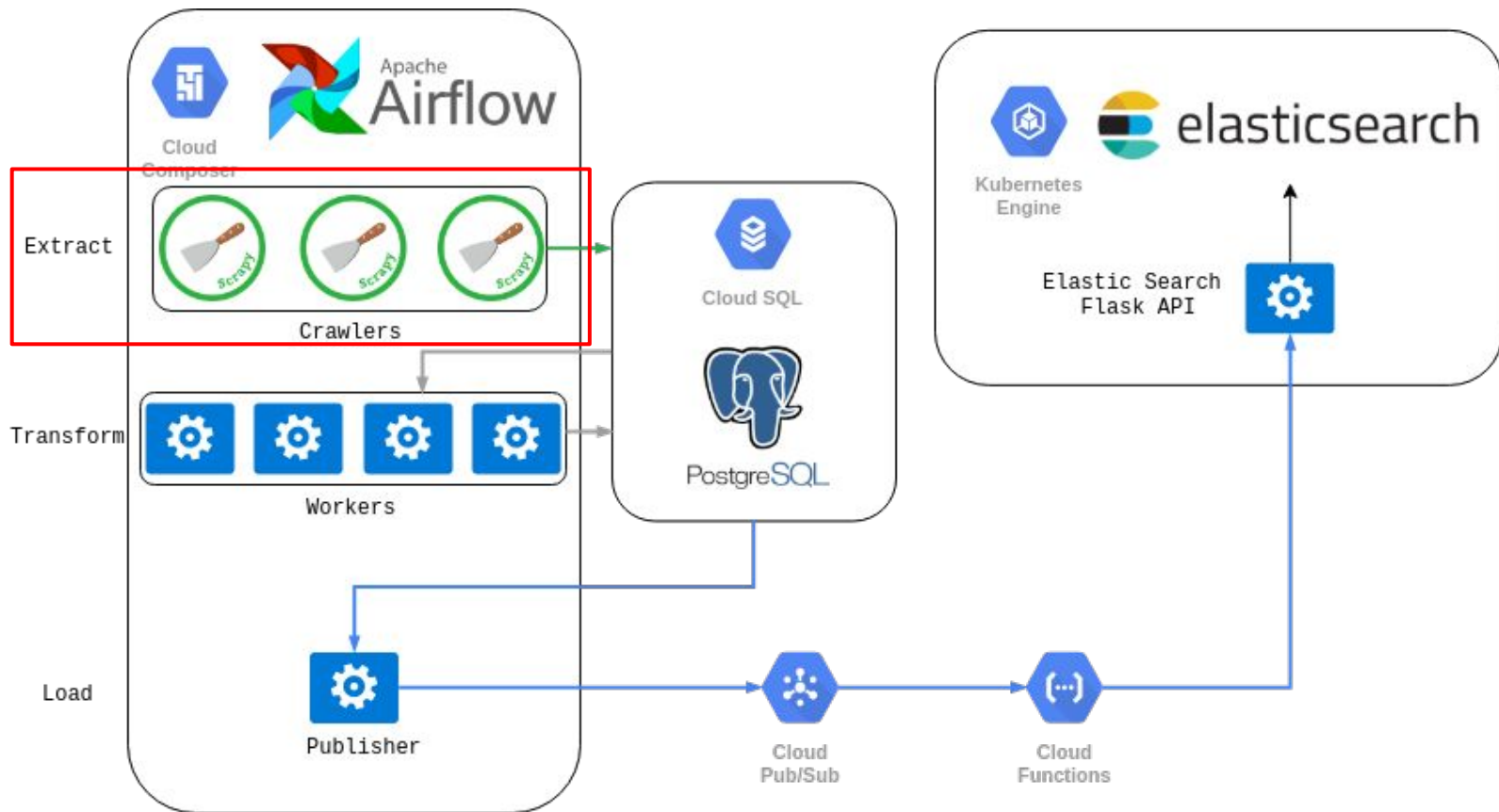
Sold Out

ETL na coleta e processamento de ofertas de e-commerces

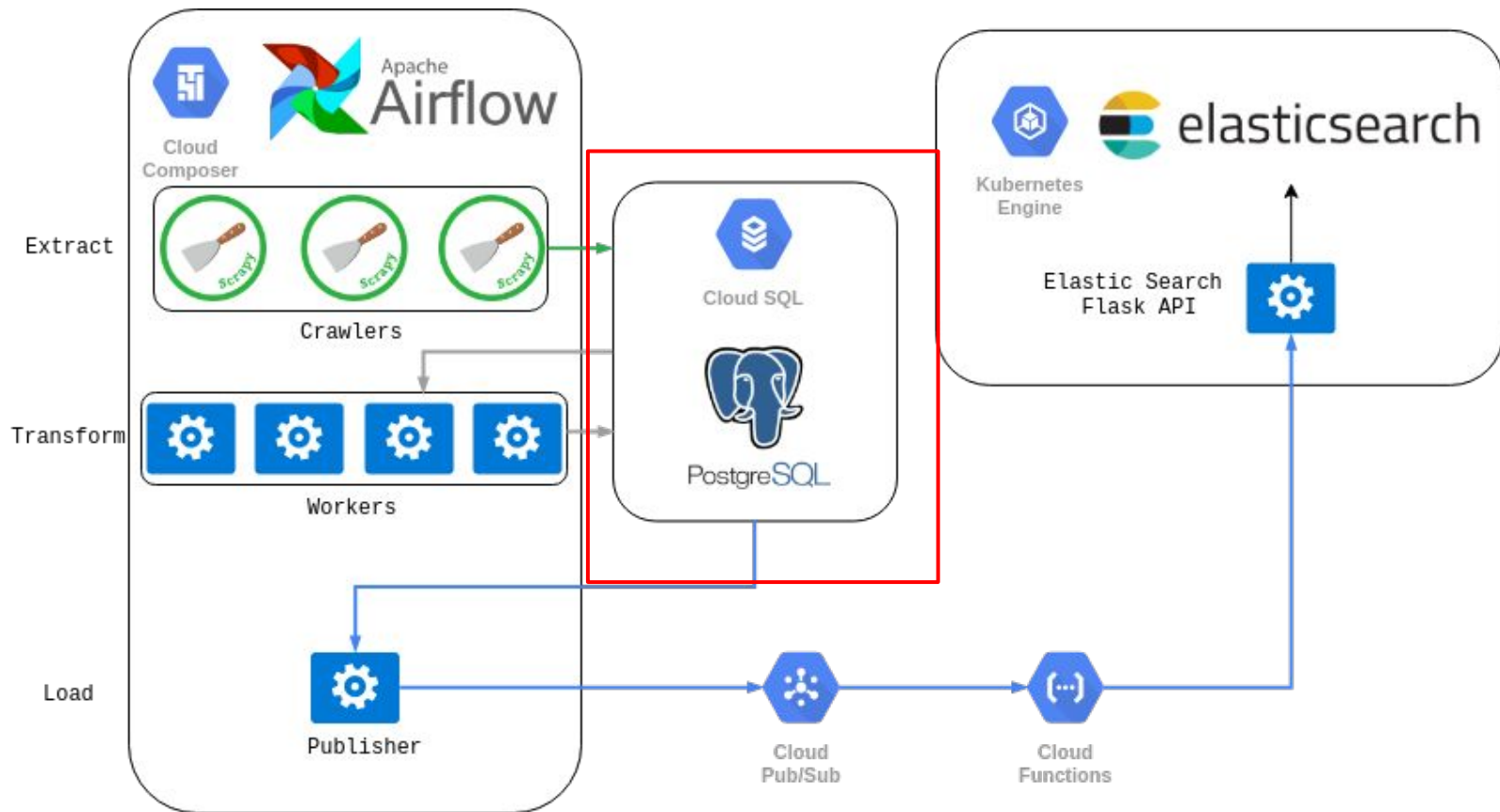




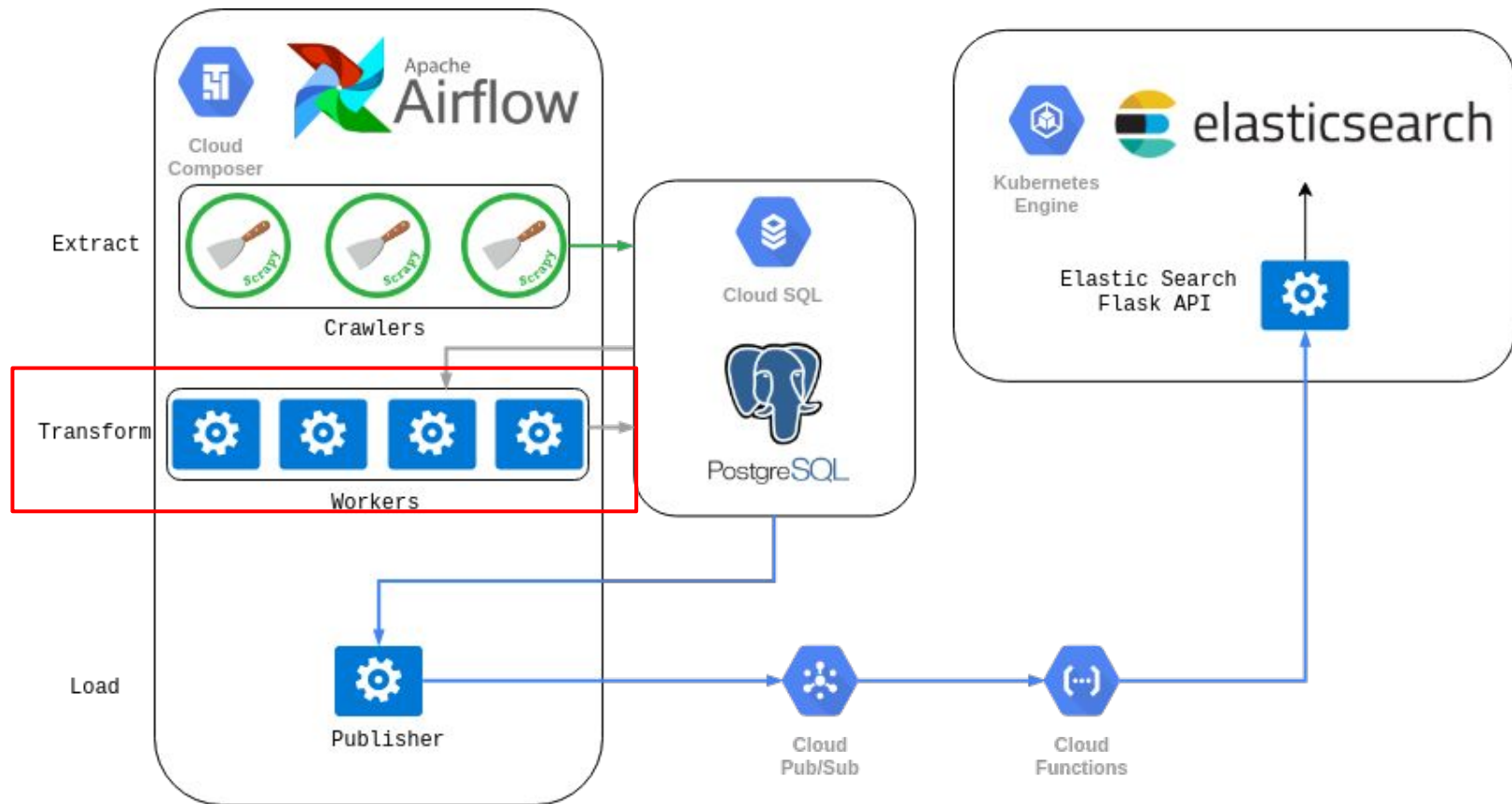
Arquitetura de transformação de dados agendados no Airflow



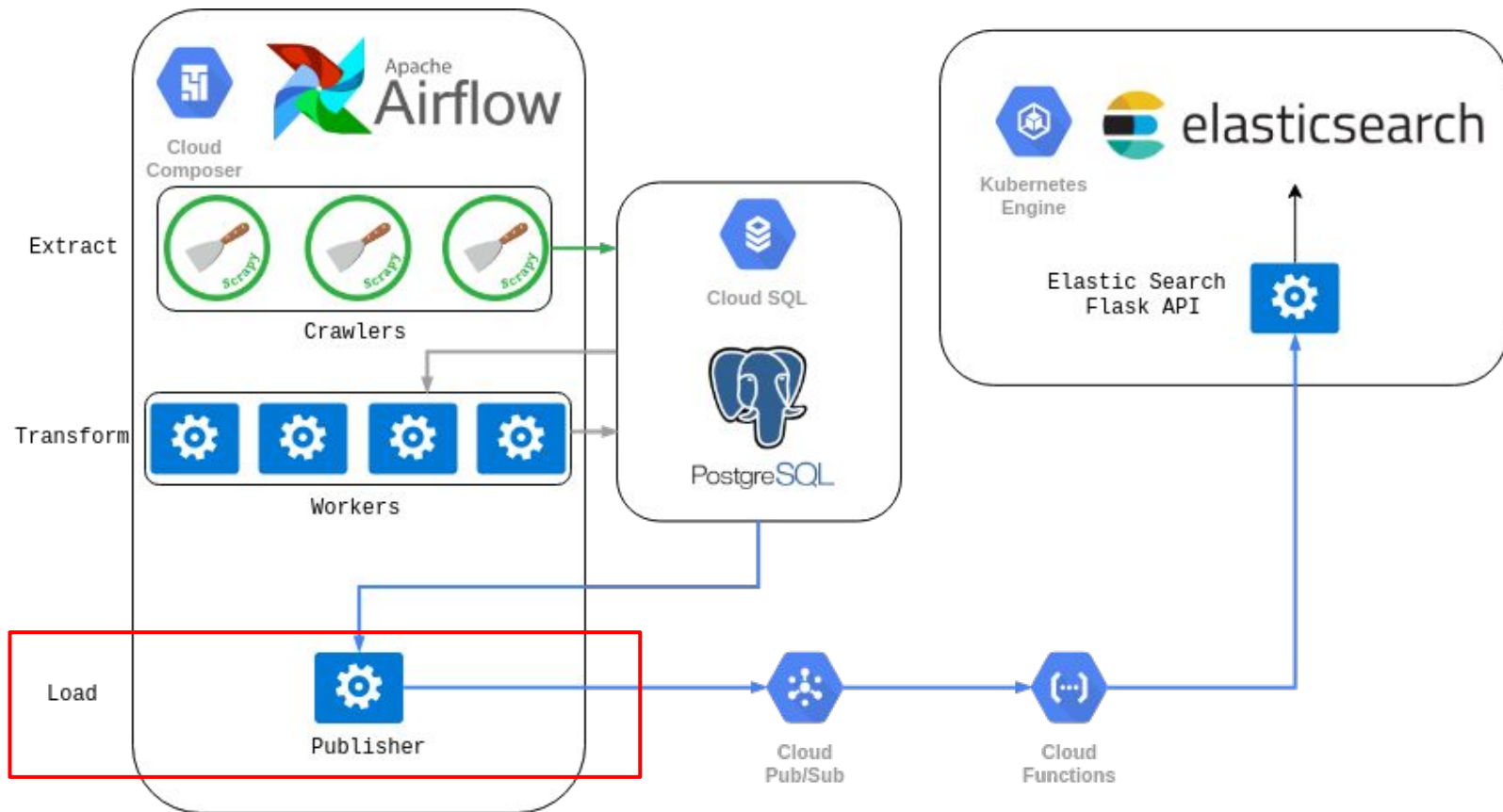
Arquitectura de transformação de dados agendados no Airflow



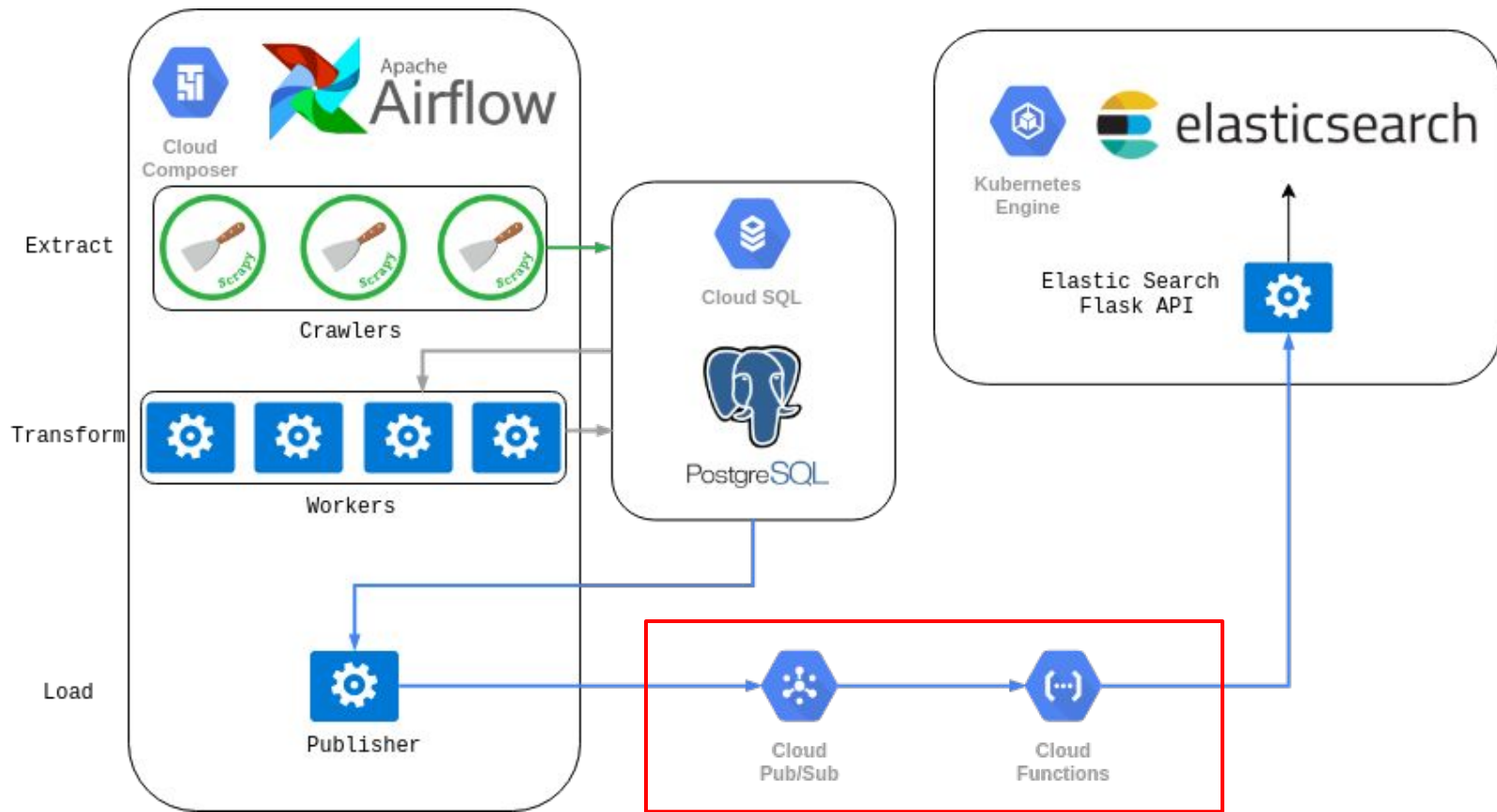
Arquitectura de transformação de dados agendados no Airflow



Arquitectura de transformação de dados agendados no Airflow



Arquitetura de transformação de dados agendados no Airflow



Arquitetura de transformação de dados agendados no Airflow



Crawlear a amazon.com

- 12 milhões de produtos*, sem incluir livros
- Hoje temos 220 mil ofertas coletadas das nossas categorias de interesse
- Nossos preços são atualizados de hora em hora
- Coletamos novas ofertas semanalmente

* Fonte: <https://www.retailtouchpoints.com/resources/type/infographics/how-many-products-does-amazon-carry>



Problema: Explorar todas as ofertas

- O processamento acontece em um único worker
 - O worker não conseguia continuar o processamento de onde parou
 - Não é tolerante a falhas
 - Todo deploy gerava problemas



Problema: Explorar todas as ofertas

- Processo distribuído
 - Procura de urls distribuídas por categorias
 - Crawlers de ofertas orientados a eventos



Microservices e Event-Driven

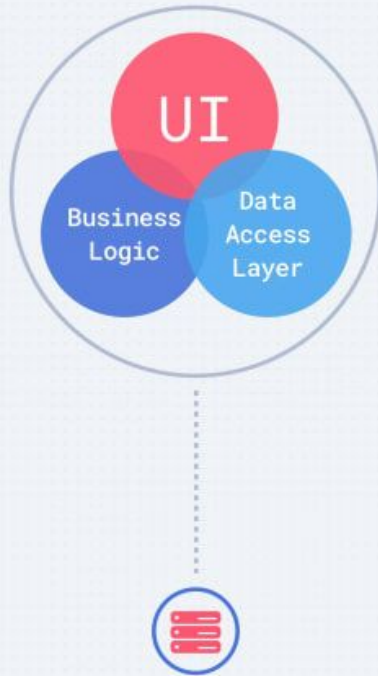
- Prós

- Cada serviço com seu banco
- Escalabilidade
- Fornecer cada parte da arquitetura como serviço (API)
- Degradação graciosa

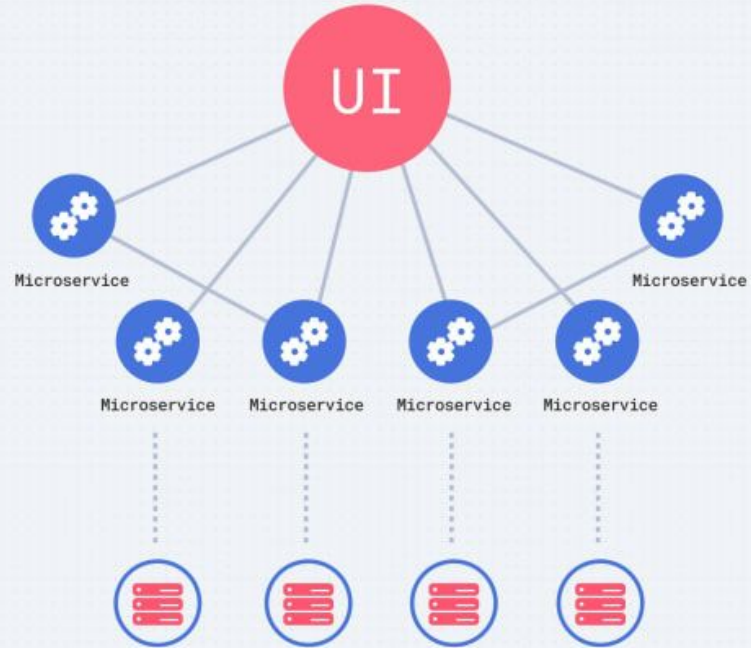
- Contras

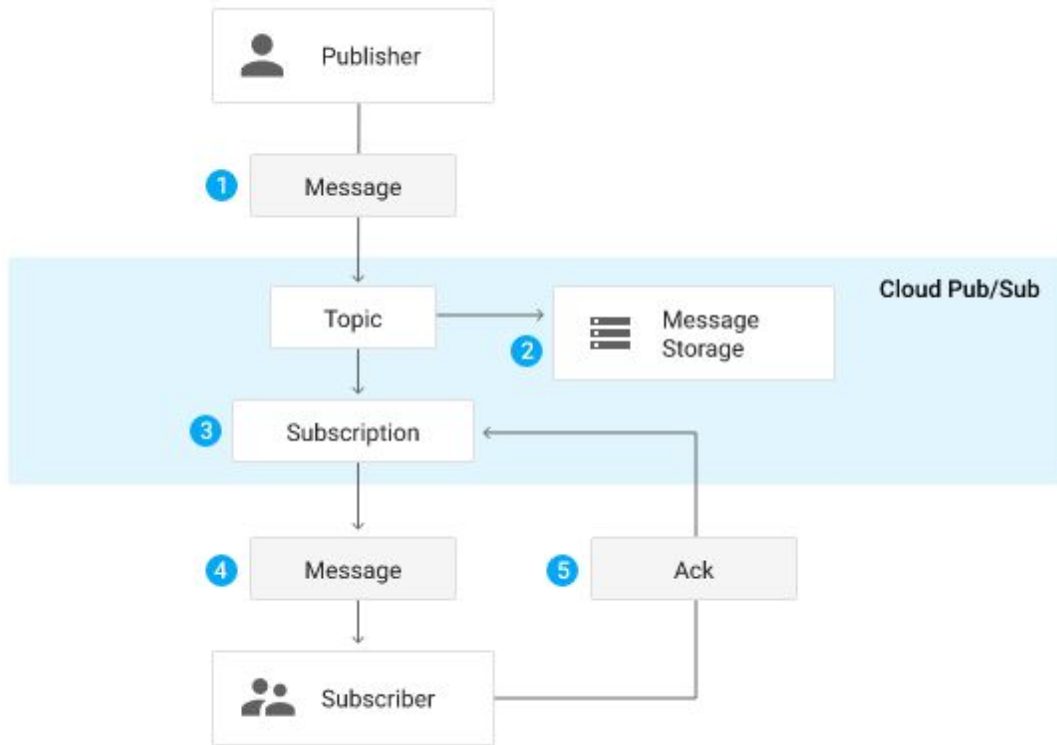
- Overhead de comunicação
- Replicação de dados
- Problemas de sincronização

Monolithic Architecture

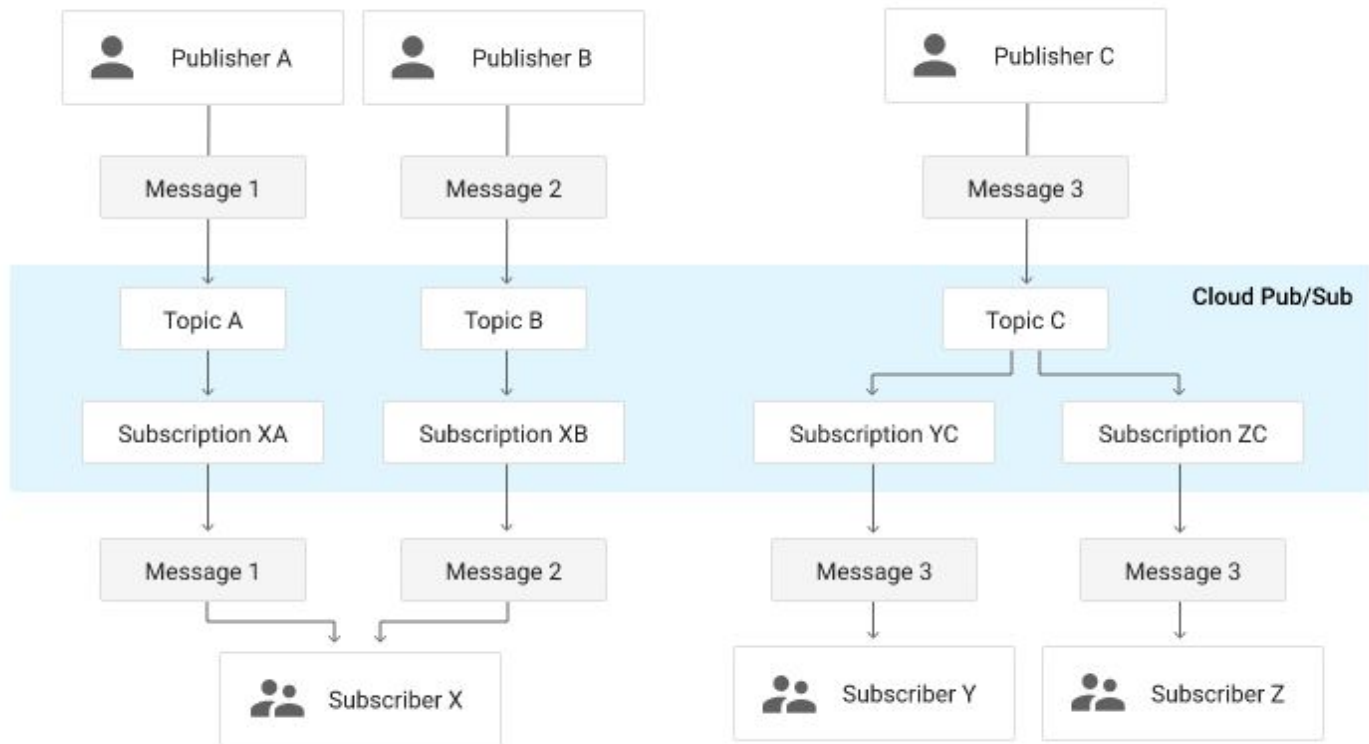


Microservices Architecture





Fluxo de mensagens no Pub/Sub. Fonte: <https://cloud.google.com/pubsub/docs/overview?hl=pt-br>



Relacionamentos Publisher e Subscriber . Fonte: <https://cloud.google.com/pubsub/docs/overview?hl=pt-br>

Considerações finais



250,861

Total of offers

↑ 0.32% • was 250,072 last day

Number of offers per retailer

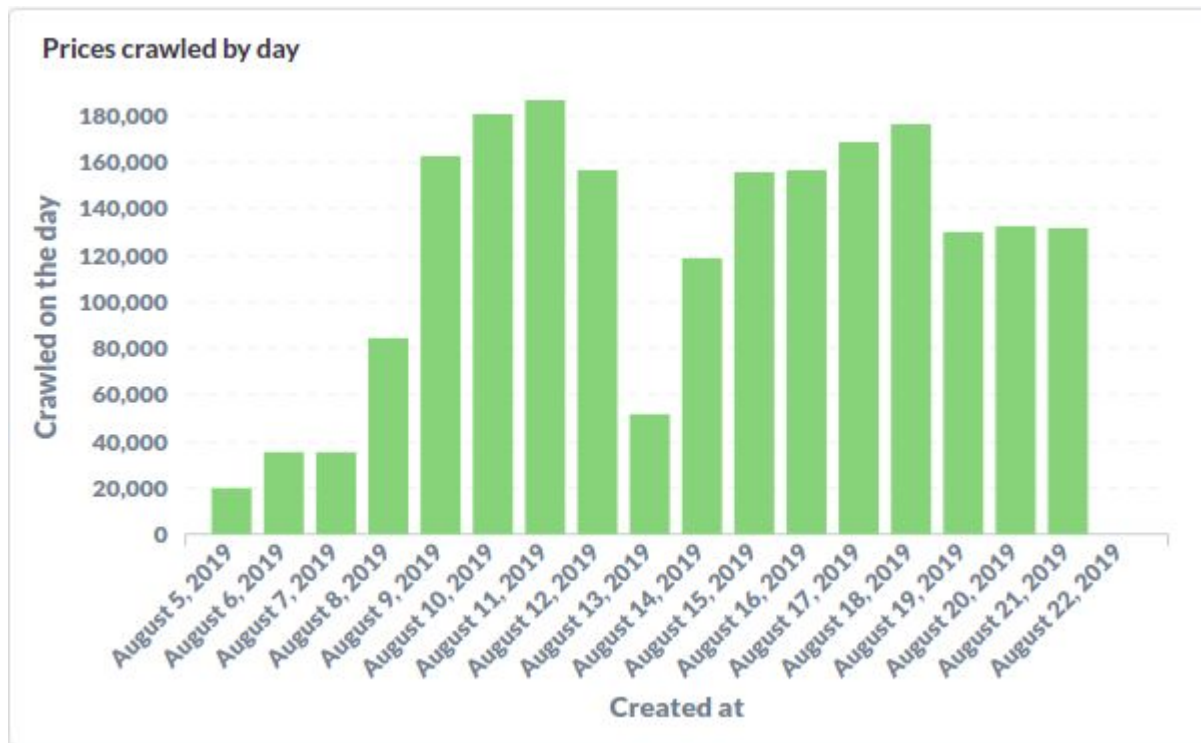
Retailer	Per retailer	Percentage
amazon_us	218,457	87%
ebay_us	8,795	3.5%
walmart_us	8,086	3.2%
bestbuy_us	7,946	3.2%
bh_us	5,737	2.3%
target	987	0.39%
costco_us	291	0.12%

Number of offer per category

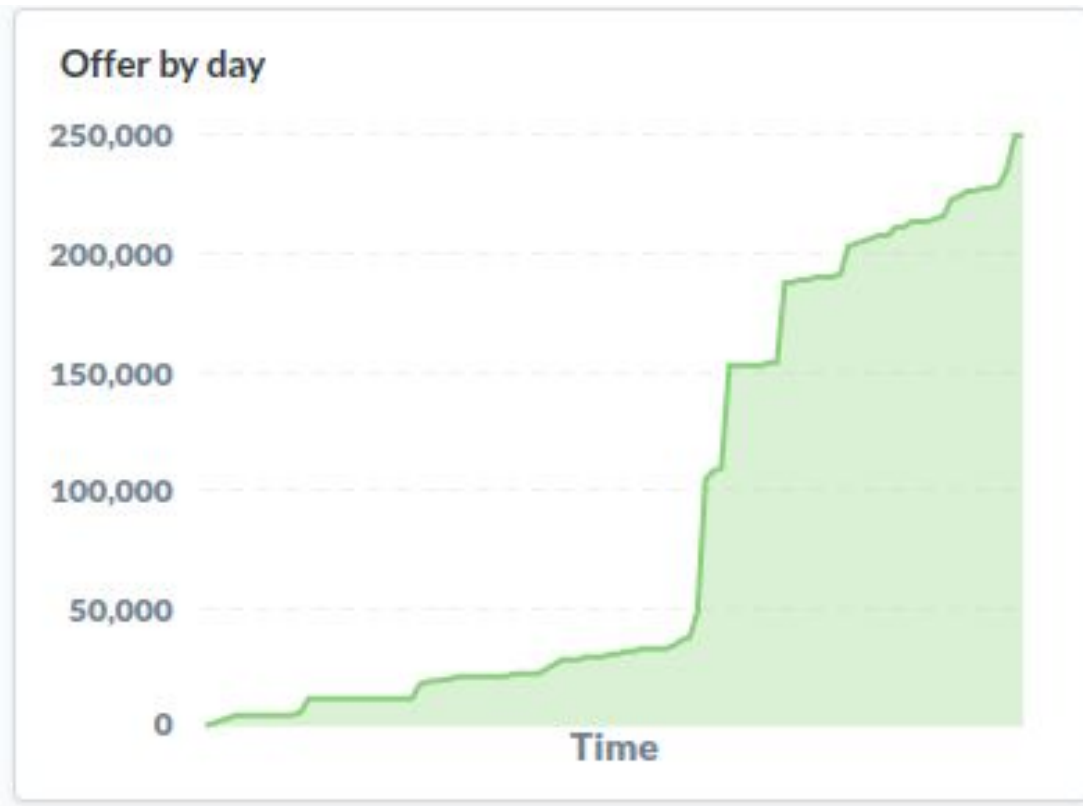
Category Name	Count
Smartphones	50,593
Headphones	39,124
Smartwatches	27,728
Laptops	21,281
Digital Cameras	19,646
Tablets	15,138
Mattresses	11,286

Rows 1-7 of 16 ◀ ▶

Total de ofertas coletadas



Números de preços coletados por dia



Número de ofertas coletadas de acordo com o tempo

Obrigado!

Dúvidas, perguntas?

Contato

Email: marcos.camargo@birdie.ai

LinkedIn: [/in/marcoscrcamargo](https://www.linkedin.com/in/marcoscrcamargo)

Telegram: [@marcosrcamargo](https://www.telegram.com/@marcosrcamargo)

Vagas

Estágio em Ciência de Dados

Estágio Front/Back

Link da apresentação:

<http://tiny.cc/h5vjbz>