



Como o iFood usa o FTS do PostgreSQL

PGConf.Brasil 2019

**1º A 3 DE AGOSTO
SÃO PAULO - SP**

**PALESTRAS | TUTORIAIS | LIGHTNING TALKS
+ 1 DIA DE WORKSHOPS**

PGCONF.COM.BR





E aí, já pediu um iFood?

Um pouco sobre o iFood

Infraestrutura

Alguns números:

+900

Instâncias no pico

+90

Instâncias de DB

99%

Automatizado



Por que não o RDS?

- Sem acesso a todos os parâmetros do servidor, OS e database.
- Sem acesso remoto do host via ssh.
- Sem acesso ao super user.
- Não se pode instalar outras aplicações no host do banco, Ex: pgbouncer, pgpool.
- Falta de visibilidade do que está acontecendo com o seu banco de dados.





Então como fizemos nosso próprio RDS?

- Com configuration manager.
- Com infrastructure as code.
- Com shell script.
- Com scripts em python.



CHEF



HashiCorp

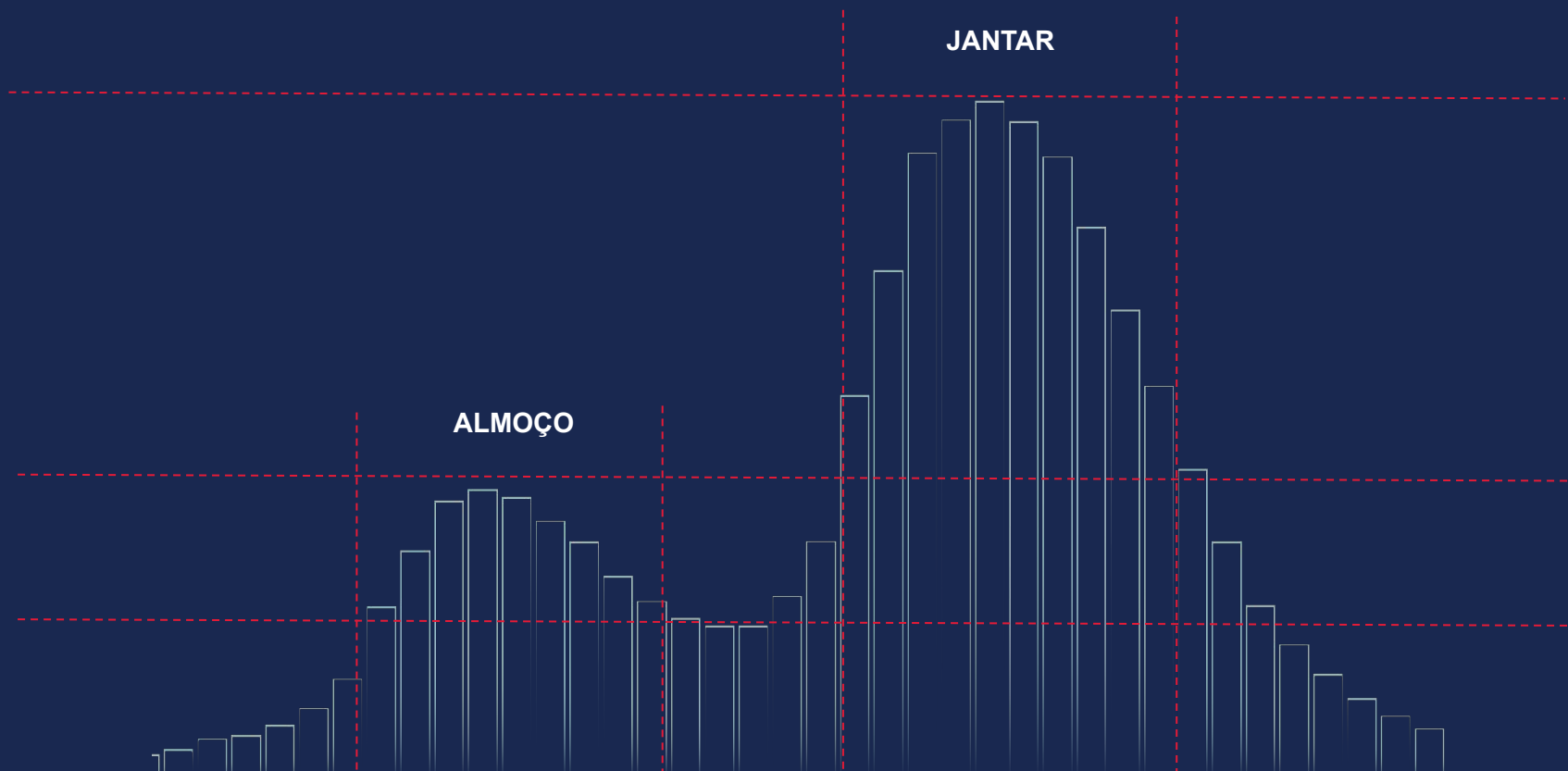
Terraform

```
#!/bin/bash
```



python™

Infraestrutura



Crescimento...



BRAZILIAN FOOD DELIVERY EXPANSION

GLOBAL MARKET NUMBERS

Current Revenue in Online Food Delivery
US\$82B in 2018 | +17.1% yoy

Future Revenue Expected
Growth rate (CAGR 2018-2023) of 10.7% → resulting in a market volume of **US\$137B** by 2023.

*according to Statista 2018

iFood orders per day in Brazil



ORDERS IN MARCH 2019

17.4 million

Present in 500 cities in every state, in Brazil

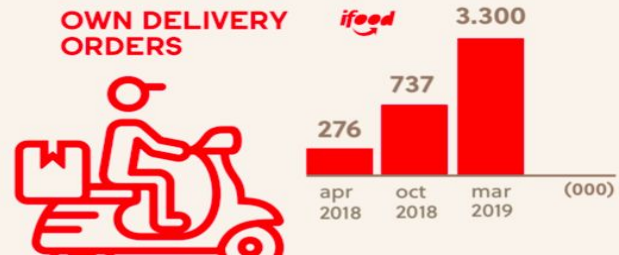
17x the nearest competitor * in terms of daily active users

*SimilarWeb data on Android App daily active users for January 2019

RESTAURANT GROWTH



OWN DELIVERY ORDERS



Plataforma de busca

A importância de uma busca de qualidade

Uma busca de qualidade, deve retornar os pratos que o usuário espera, corrigindo erros de digitação e sugerindo termos similares. Tudo isso pra aumentar a conversão

13:46

Yakissoba

Restaurantes (0) Pratos (51)

Yakisoba misto (leve 2 pague 1)
Macarrão de yakisoba, frango, carne, brócolis, cenoura, moyashi, acelga e molho shoyu
R\$ 41,00
Teste Cheff Solutions 2 ★ - • 45-55 min

Yakisoba misto(fome)+1 rolinho primavera
Macarrão de yakisoba, frango, carne, brócolis, cenoura, moyashi, acelga e mo...
R\$ 37,50
Teste Cheff Solutions 2 ★ - • 45-55 min

Combo: Yakisoba misto+2 rolinhos+bebida
Um yakisoba fome + dois rolinhos ou doce ou salgado e mais um bebida de sua escolha
R\$ 21,74 ~~R\$ 54,37~~
Teste Cheff Solutions 2 ★ - • 45-55 min

Combo: Yakisoba misto+2 rolinhos primavera+bebida
Escolha o sabor do rolinho, doce ou salgado e escolha a bebida

13:48

Esfira

Restaurantes (0) Pratos (237)

Esfíha premium de quatro queijos com Philadelphia
Deliciosa esfíha de 4 queijos com queijo Philadelphia.
A partir de R\$ 3,40
Teste Voxline - Hbb ★ 5,0 • 38-48 min

O problema: dados de vários serviços integrados na mesma tela

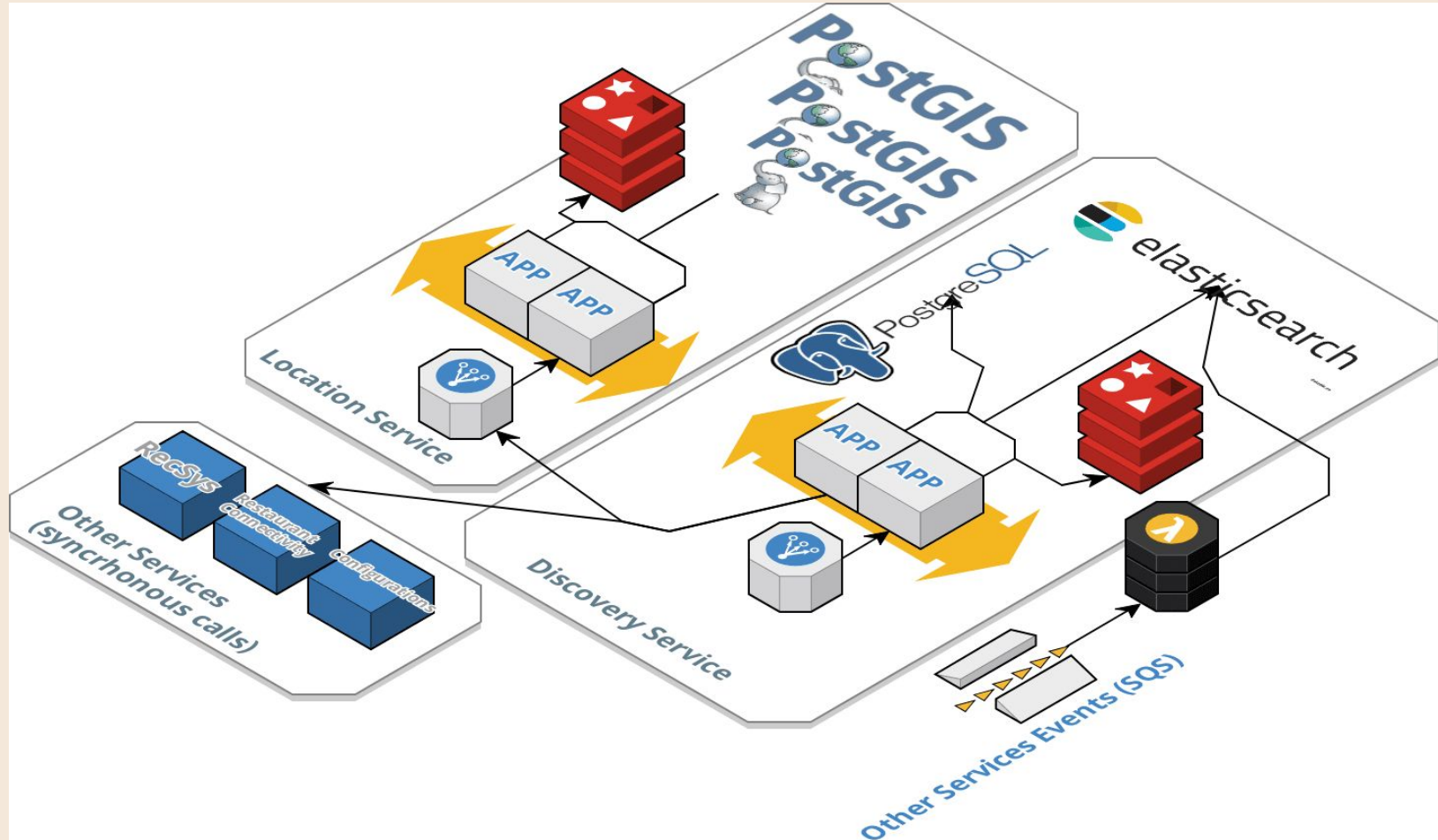
The image shows a mobile application interface for food delivery. The top status bar displays the time 20:25, signal strength, Wi-Fi, and battery level at 89%. The app header includes the word "ENTREGAR" and a location dropdown menu set to "Agora em R Teste, 100". A search bar contains the text "Prato ou restaurante" and a "Filtros" button. Below the search bar, a section titled "Restaurantes" lists three items:

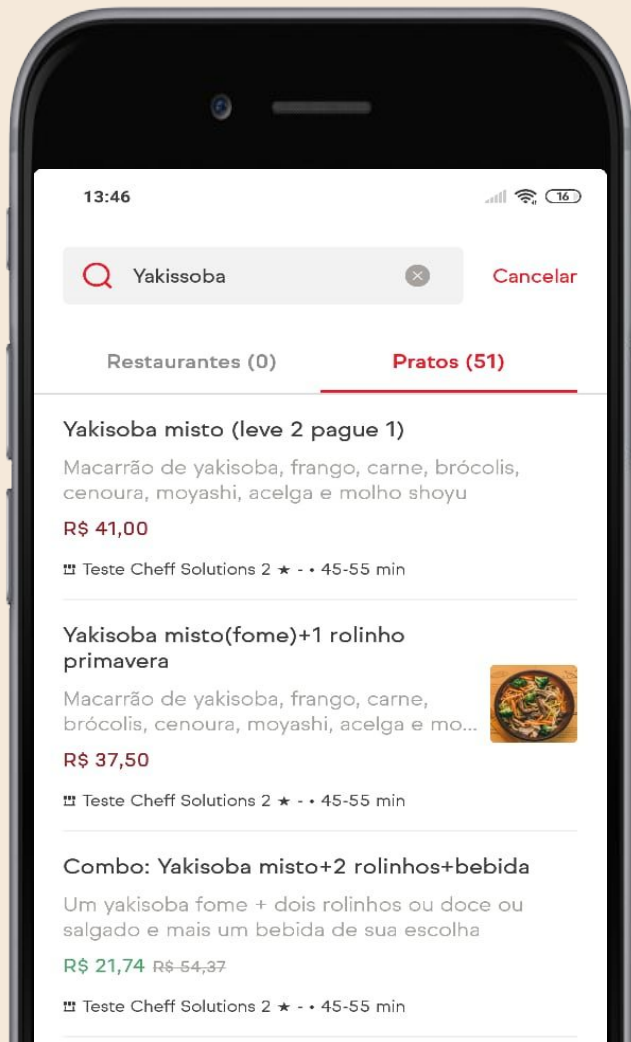
- Modelo 9 Teste Pos2**: Pizza • 0-10 min • 0,3 km • \$\$, with a 4.0 star rating and "ENTREGA VARIÁVEL" delivery status.
- Teste Datamaxi**: Brasileira • 1-11 min • 0 km • \$, with no rating and "ENTREGA R\$ 7,00" delivery status.
- Restaurante lfood**: Alemã • 30-40 min • 0 km • \$, with no rating and "ENTREGA GRÁTIS" delivery status.

Annotations with arrows point to various elements:

- Lat/Lon**: Points to the location dropdown menu.
- Termo**: Points to the search bar.
- Restaurantes e pratos**: Points to the list of restaurant entries.
- Avaliação**: Points to the star rating (4,0) for the first restaurant.
- Entrega**: Points to the delivery status label "ENTREGA VARIÁVEL".
- Tempo**: Points to the estimated delivery time "0-10 min".
- Distância**: Points to the distance "0,3 km".

Solução antiga





Uma busca deve ter:

- Dados de qualidade
- Stemming (entregar resultados baseados na fonética da palavra)
- Ranqueamento (Atribuição de peso)
- Pesquisa Fuzzy para erros de ortografia
- Acentuação

Demo

Steps

- Análise do documento
- Stemming
 - Remover sufixos
 - Converter para lower case
 - Remover stop words
- Realizando a busca
- Ranking

Full text search no PostgreSQL

- Integração total
- Suporte de configurações definidas pelo usuário
- Dicionários (ispell, snowball, thesaurus), parsers
- Ranking por relevância
- Índices: GIN e GiST
- Fácil de implementar
- Integridade dos dados

Stemming

- Processo de reduzir palavras flexionadas (ou às vezes derivadas) ao seu tronco

```
# select to_tsvector('portuguese', 'pizzaria');
```

```
to_tsvector
```

```
-----
```

```
'pizz':1
```

```
# select to_tsvector('portuguese', 'pizza');
```

```
to_tsvector
```

```
-----
```

```
'pizz':1
```

Stop Words

- A função `tsvector` remove palavras irrelevantes. Isso é importante para encontrar resultados mais relevantes

```
# select to_tsvector('portuguese', 'pizza de queijo com bacon') fts;
```

```
fts
```

```
-----
```

```
'bacon':5 'pizz':1 'queij':3
```

Prioridade

- Atribui peso a cada elemento que compõe o documento

```
# # SELECT setweight(to_tsvector('portuguese', 'pizza de queijo'), 'A') fts;  
fts
```

```
-----  
'pizz':1A 'queij':3A
```

Stemming e prioridade

```
# SELECT (  
    setweight(to_tsvector('portuguese', 'pizza'), 'A') ||  
    setweight(to_tsvector('portuguese', 'com queijo e bacon'), 'B')  
);  
?column?
```

'bacon':5B 'pizz':1A 'queij':3B

Pesquisa de termos semelhantes

- Temos um dicionário que encontra termos similares e utiliza os mais relevantes na consulta, com isso agregamos outros resultados ao termo buscado

```
# .. WHERE search_platform.unaccent(s.word) % search_platform.unaccent('pastel')
```

word	similarity
-----+-----	
pastel	1
pastell	0.666666687
pastelão	0.600000024
pastel09	0.600000024
pastelao	0.600000024
pastello	0.600000024

Gerando o tsquery

- A partir desses termos similares, geramos um tsquery, considerando na busca as opções mais relevantes

```
# select ifood_plainto_tsquery('pastel');
```

```
pastel | pastell | pastelao | pastelzinho
```

Busca

```
# ... ifood_plainto_tsquery('pizza') AS q(tsquery) ON catalog_item.fts @@ q.tsquery
```

- Função `ifood_plainto_tsquery` gera o `tsquery`
- `catalog_item.fts` é o campo do Full Text Search persistido

GIN

- Estratégias customizadas para tipos de dados específicos
- Mais Lento para escrita
- Mais rápido para buscas

GiST

- Mais Rápido para escrita
- Mais lento em leitura

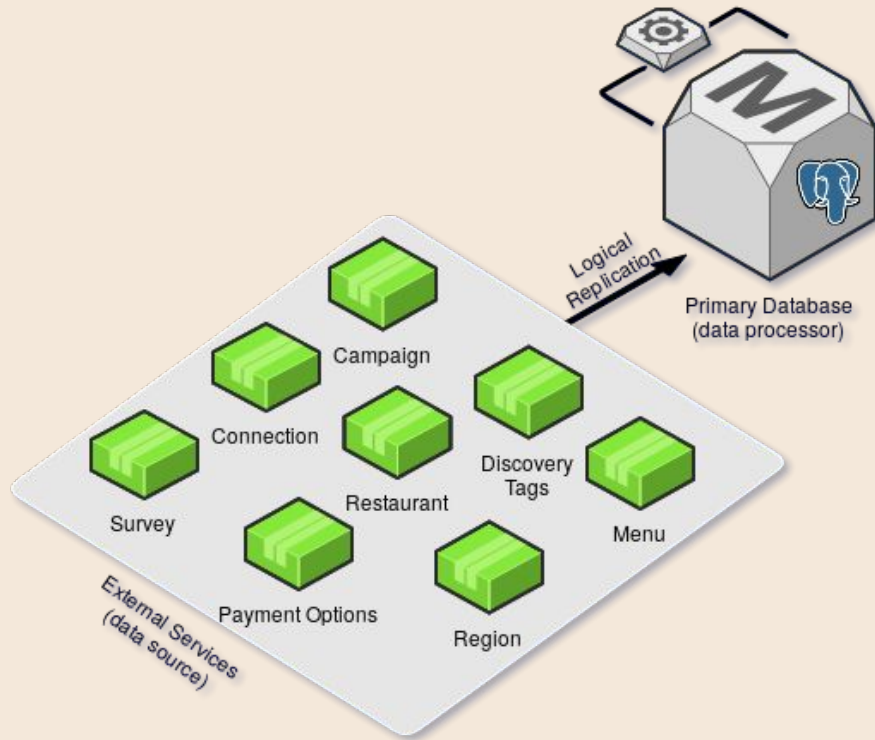
Pesquisa Fuzzy para erros de ortografia

- Ordenamos pelo score, para não retornar tudo para o RecSys

ts_rank(fts, q.tsquery) AS term_score

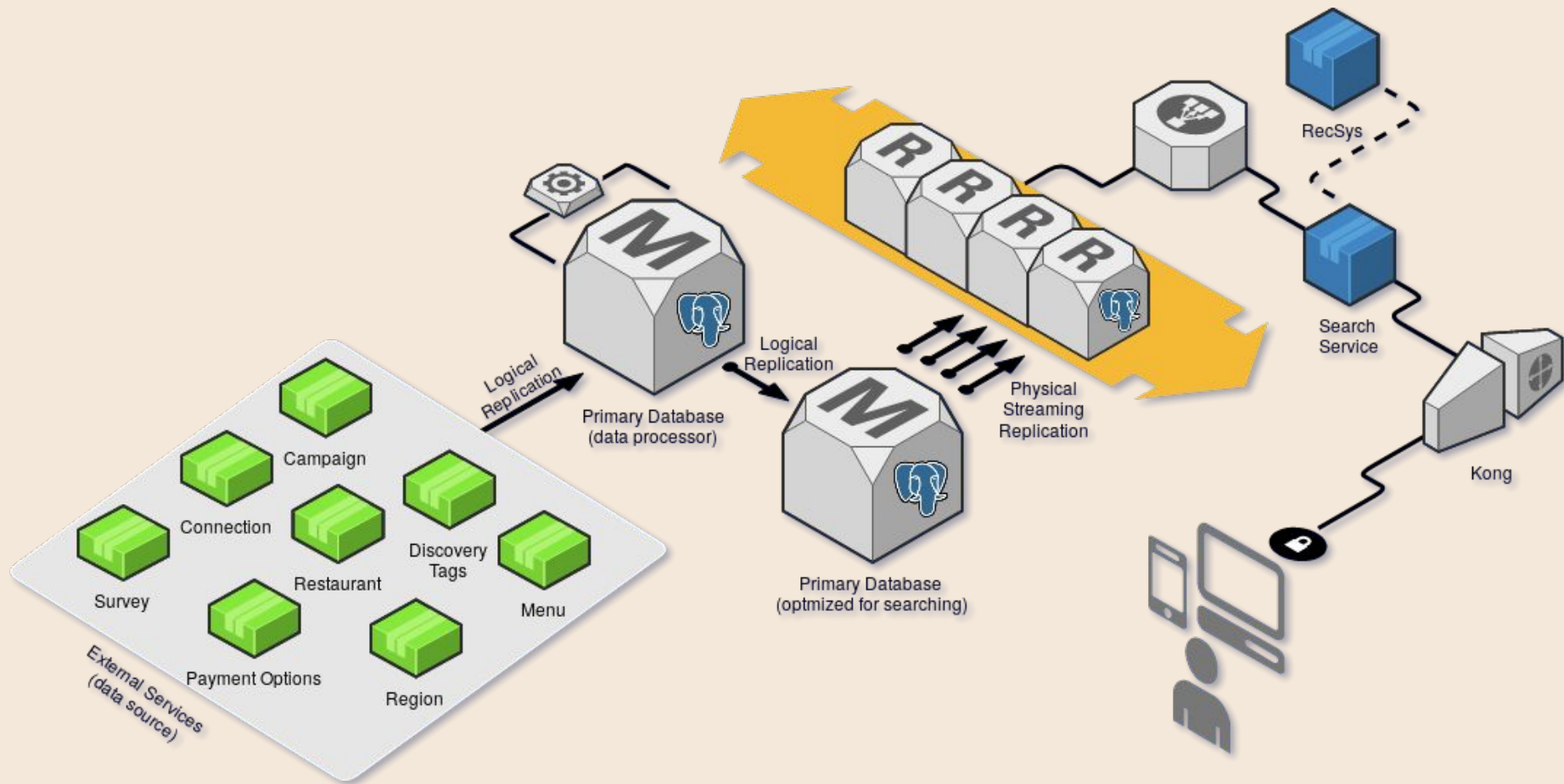
Processamento dos dados

Cada tabela replicada tem um trigger, que processa os dados alimentando um novo schema otimizado para buscas



- Essas triggers não mudam o comportamento das aplicações

Arquitetura atual



Hoje temos em média ~13 milhões de buscas
por mês





Obrigado!

Lucas Viecelli

lucas.viecelli@ifood.com.br

telefone 49-99194-4888