



M&C SOLUTIONS



Integration

Sprint 3



AGENDA

Fluxo de entendimento do que será abordado na apresentação.

01

Atualizações
da última sprint

02

Objetivo da
sprint

03

Nova
arquitetura

04

Estrutura do
Data
Warehouse

05

Status
report

06

Próximos
passos

ATUALIZAÇÕES DA SPRINT 2

Ajustes no pacote

Finalização e refinamento do código da biblioteca criada.

Carregamento dos dados

Foi realizada a inserção de novas fontes de dados no DataLake.

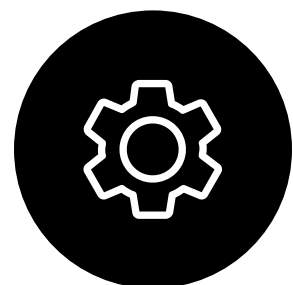
Ambiente AWS

Migração do ambiente para usuário Root (antes era o Labs)

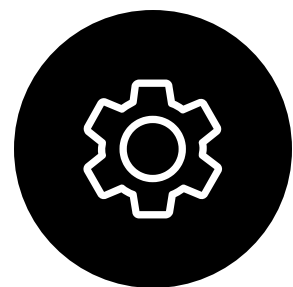
Ajustes na ingestão de dados

Ajustes na arquitetura de ingestão de dados

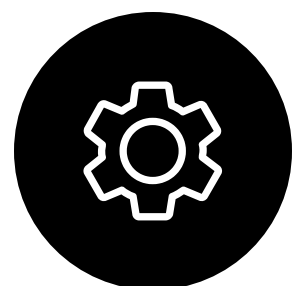
OBJETIVO DA SPRINT



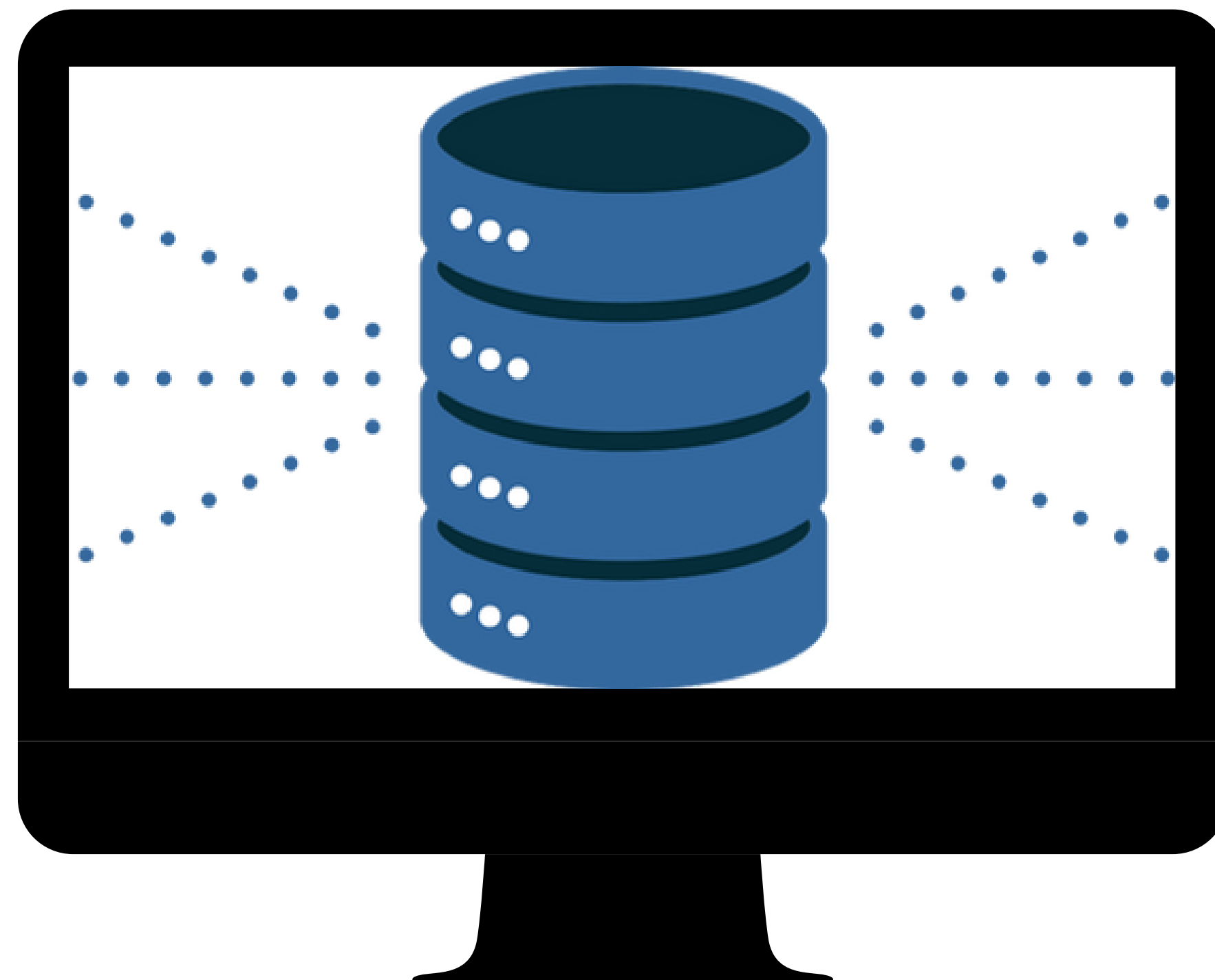
Criação do data warehouse



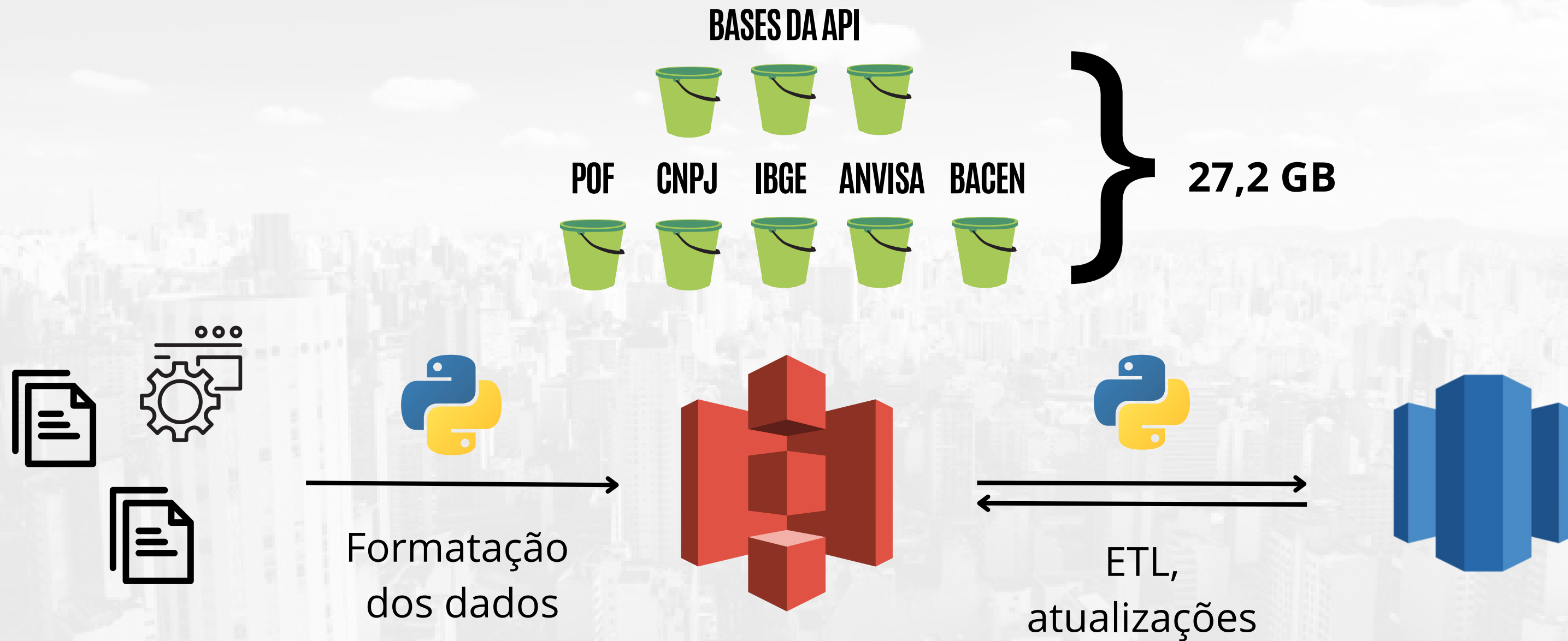
Inserção dos dados no datalake/datawarehouse



Transformação dos dados



NOVA ARQUITETURA DE INGESTÃO



Sistemas de processamento de dados



OLAP

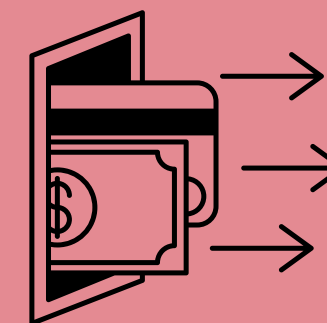
Inferências analíticas

Processamento de dados complexos

Orientada a objetos

Utilizada para serviços de suporte a decisão e dados históricos

Facilita o processamento de queries complexas



OLTP

Operações diárias e recorrentes

Orientada a aplicações

Utilizada para serviços de transações simples

Boa para dados atualizados em tempo real

Organização do Datawarehouse

Posicionamento dos dados de forma relacional e organizando-os num sistema de arquivos distribuídos:

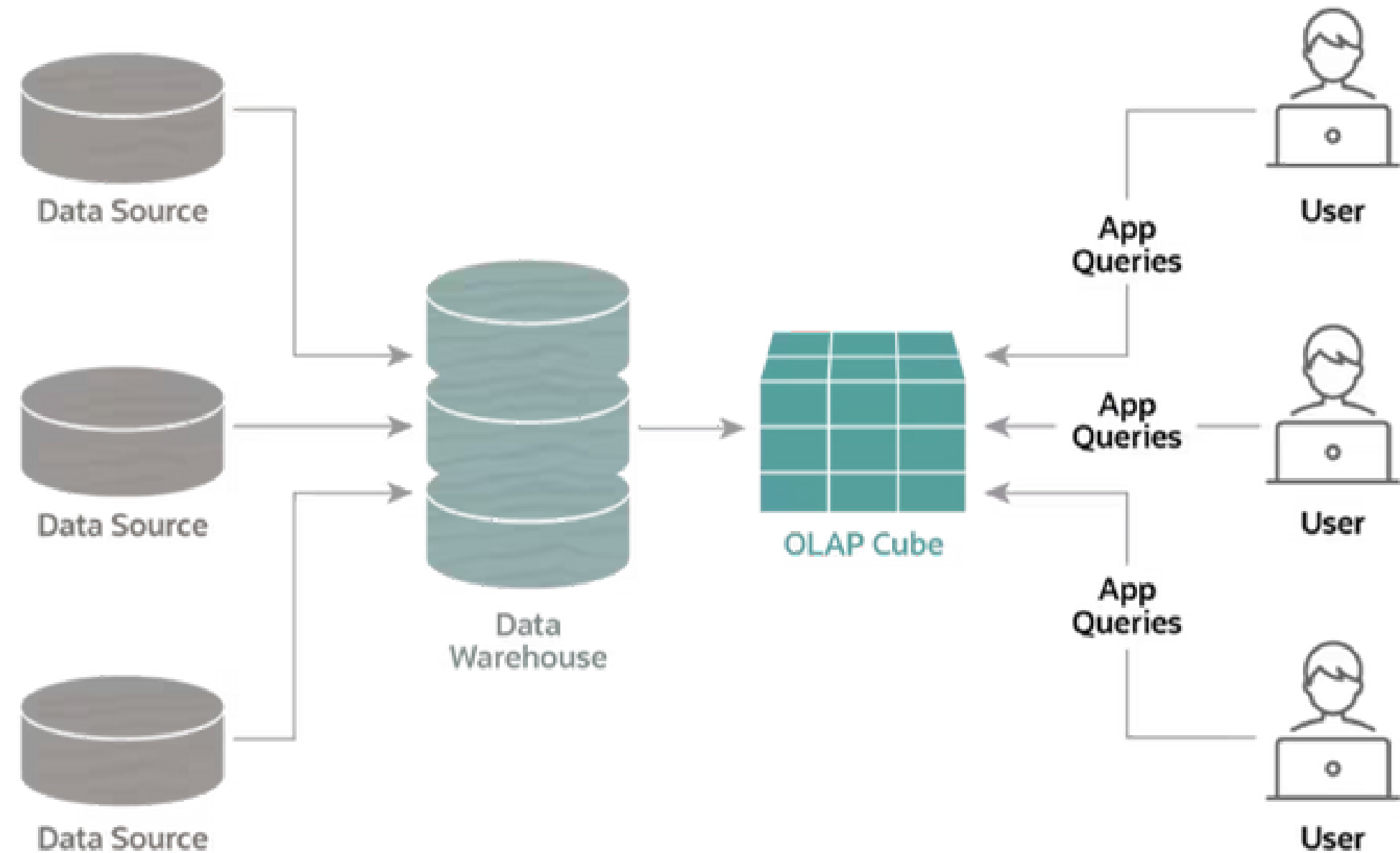


O Redshift é um serviço OLAP e tem processamento distribuído

REDSHIFT

O Redshift proporciona:

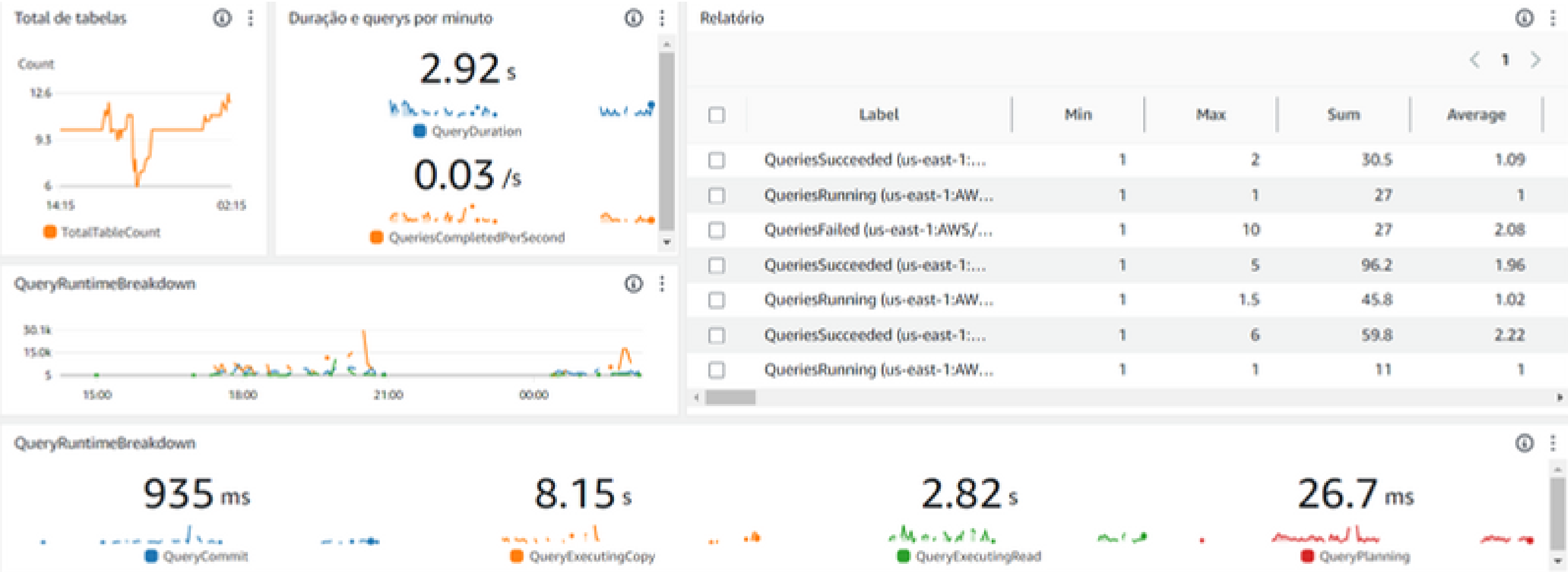
- Um processamento rápido de grandes quantidades de dados;
- Uma visão mais abrangente e detalhada;
- Facilidade na navegação e na compreensão dos dados.



The image features four decorative green network diagrams, each consisting of a central node connected to several peripheral nodes, located in the corners of the page. Two horizontal black lines with diagonal steps and circular endpoints frame the central text.

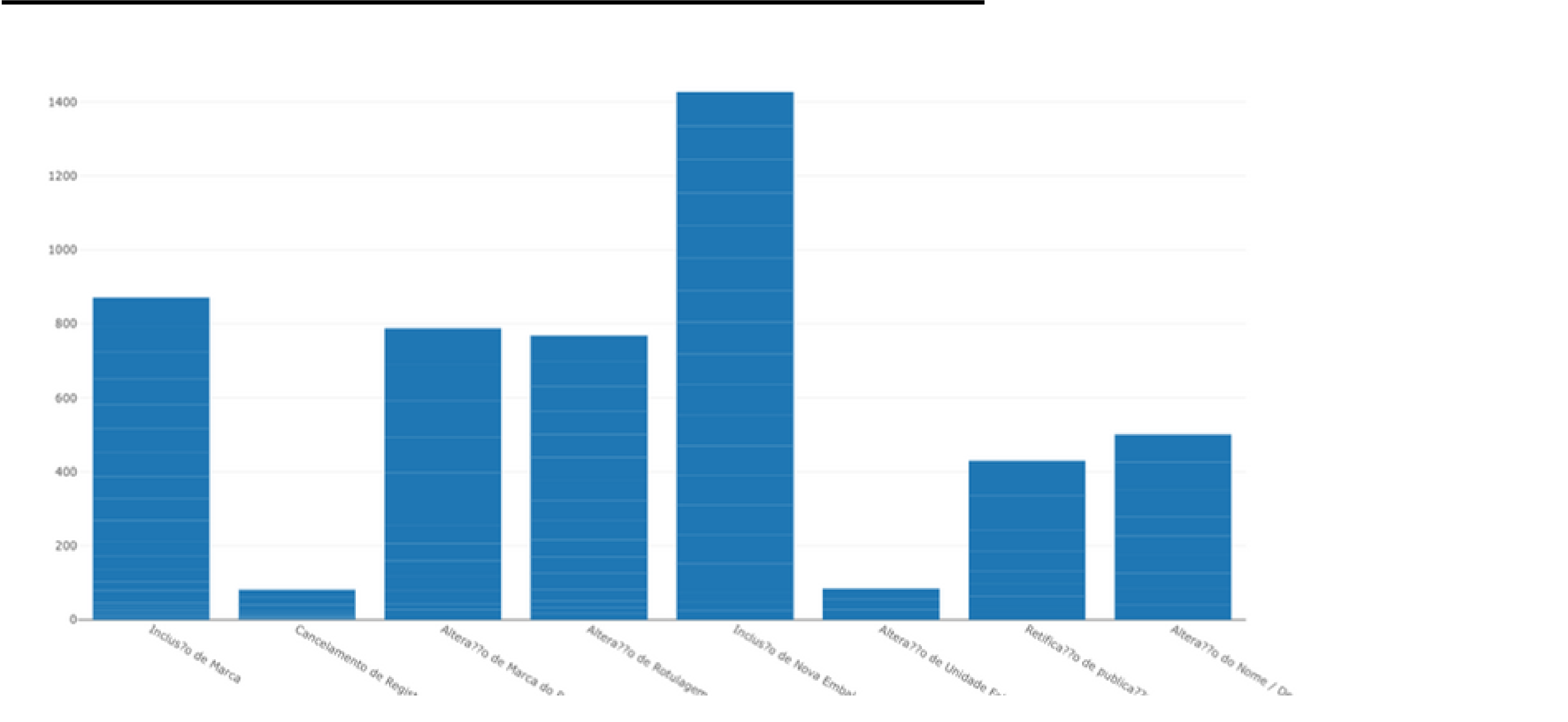
SQL VIEW REDSHIFT

MONITORAMENTO DE SERVIÇOS



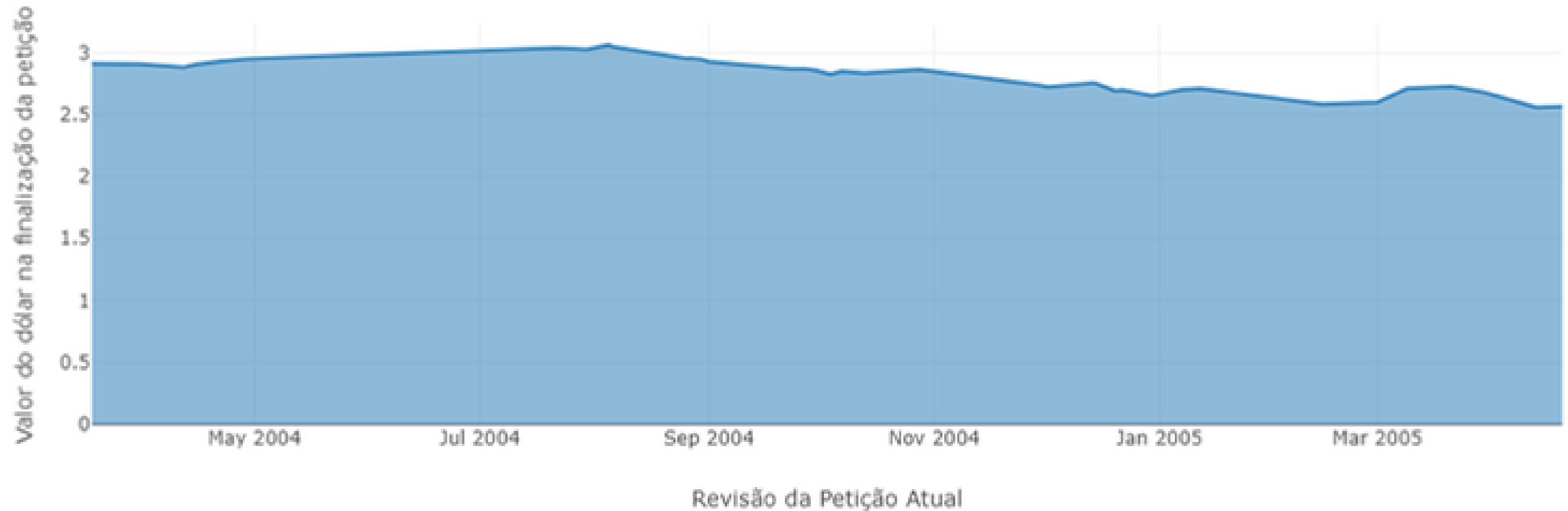
SOMA DE PETIÇÕES

Dados de Petição por descrição (SUM)



COTAÇÃO A PARTIR DA FINALIZAÇÃO DE PETIÇÃO

Dados de Petição de Alimentos vs Dados do Bacen



INSIGHTS EXTRAÍDOS DAS VIEWS

E IMPLEMENTAÇÕES AINDA PENDENTES

01

Avaliação da
Competição e Decisões
de precificação

02

Custo de Produção e
Impacto Econômico

03

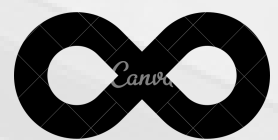
Maior entendimento da
projeção de
consumidores

04

Expansão de mercado e
posicionamento por
região

STATUS REPORT

Etapa	Status	Data
Desenvolvimento dos scripts em python	Feito	Sprint 1
Identificação das fontes de dados e suas características	Feito	Sprint 3
Seleção dos serviços da AWS	Feito	Sprint 3
Descrição da estrutura do Data Lake/Data Warehouse	Feito	Sprint 3
Aspectos de monitoração e segurança do fluxo de dados	Feito	Sprint 3
Documentação da Estrutura do Datawarehouse	Em andamento	Sprint 3
Abastecimento completo do Datawarehouse	Em andamento	Sprint 3 e 4



PRÓXIMOS PASSOS

O que será feito nas próximas sprints

01



Entendimento do problema

02



Prototipação e estrutura de Ingestão de dados

03



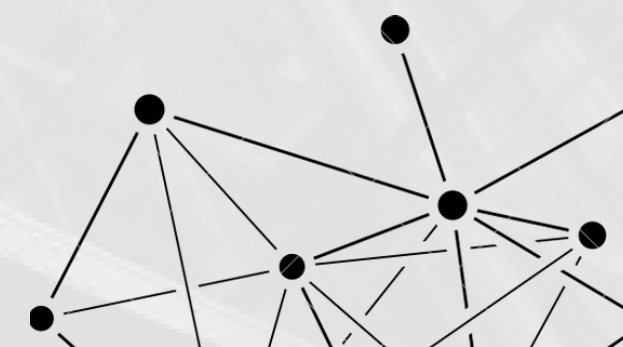
Análise de impacto ético e Data Lake/Data Warehouse

04

Infográfico e Ensemble

05

Pipeline de BigData na AWS e entrega do projeto



TIME



Rodrigo Campos



Eric Tachdjian



Rodrigo Martins



Vinicius Fernandes



Michel Mansur



Lucas Conti



Estamos disponíveis para mais dúvidas!

OBRIGADO PELA ATENÇÃO



Integration