



84° Python Floripa

# Apache Superset

Marília Melo Favalesso

# Marília Melo Favalesso

- 🧠 Desenvolvedora de IA | MSc | PhD
- 🐍 Python & Comunidades
- 🐈 Gatos, pizza e bicicleta nas horas vagas

✉️ [marilia.melo.favalesso@gmail.com](mailto:marilia.melo.favalesso@gmail.com)

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

🔗 github: [/mmfava](https://github.com/mmfava)

🔗 site: [www.mariliafavalesso.com](http://www.mariliafavalesso.com)



***"Superset by heart"***

---

Entusiasta de ferramentas Open-Source ❤️

# Agenda

1. Overview
2. Arquitetura
3. Funcionalidades
4. Instalação
5. Código
6. SQL Lab
7. Datasets
8. Charts
9. Dashboards e Filtros
10. Mão na Massa

# Overview

---

O **Apache Superset** é uma **Ferramenta para Business Intelligence (BI) open-source** que permite a exploração e visualização de dados de forma iterativa e escalável.

# Business Intelligence (BI)

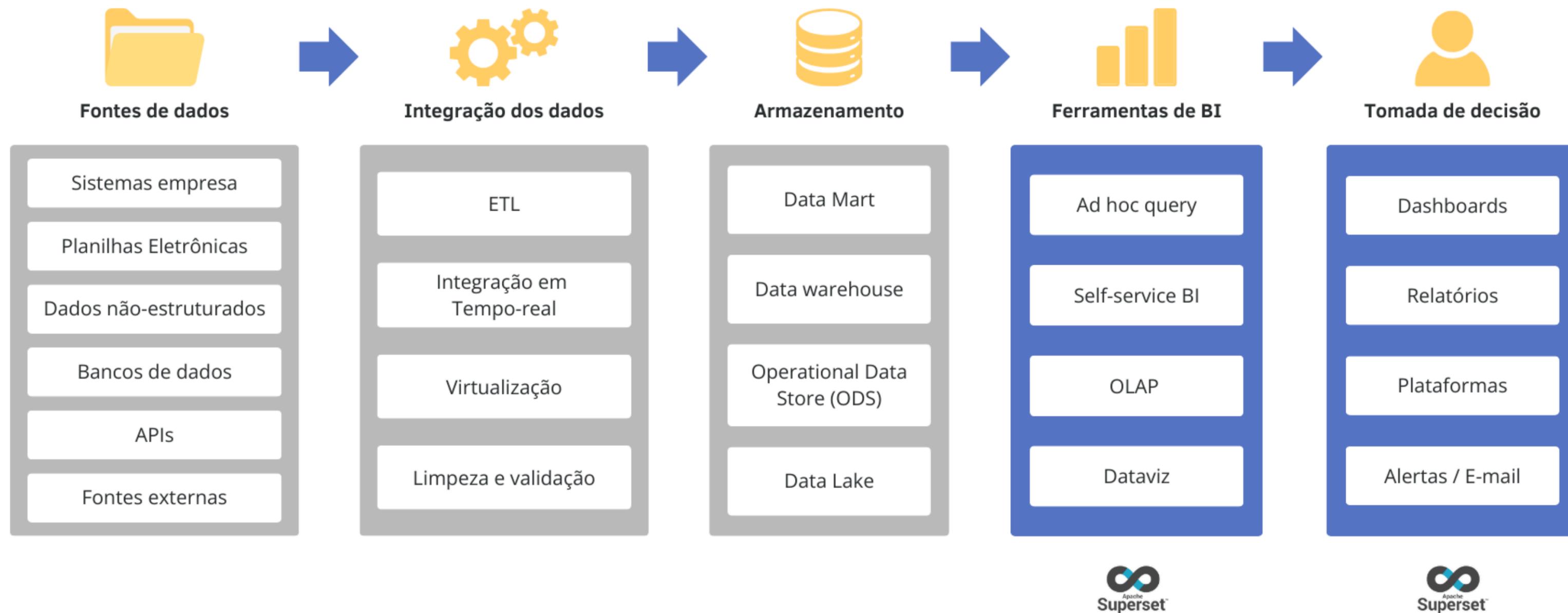
## Inteligência de Negócios

- Conjunto de processos, tecnologias e ferramentas que **coletam, organizam, analisam e visualizam dados**.
- Transforma dados dispersos em **informações úteis** e insights claros.
- Facilita decisões estratégicas, operacionais e táticas, permitindo ações rápidas e informadas.
- Ajuda gestores a entenderem rapidamente **o que está acontecendo e o porquê**.

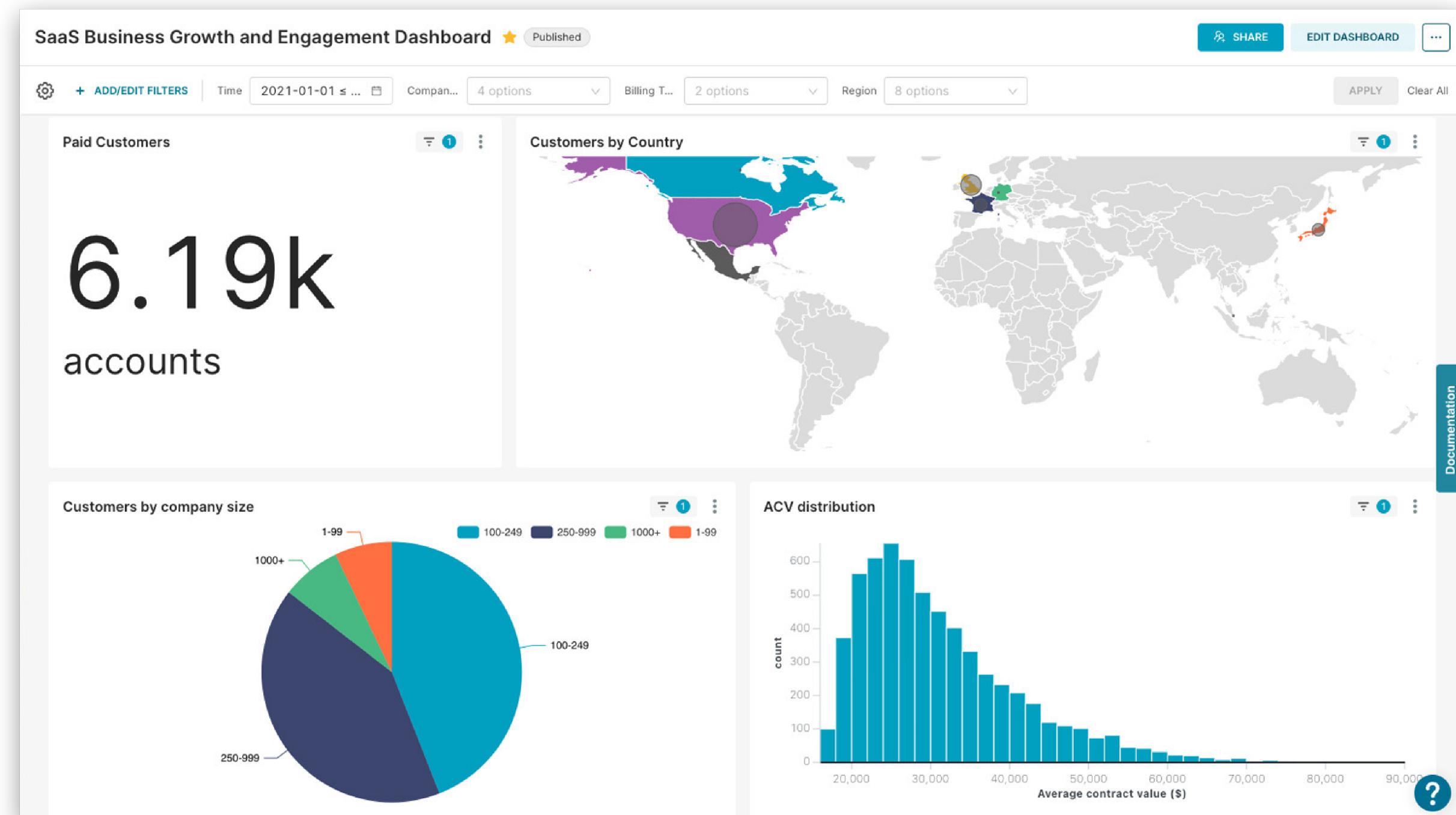
❖ O objetivo é o de apoiar decisões inteligentes e baseadas em fatos  
❖ ~Data-Driven(DWx 2025)!



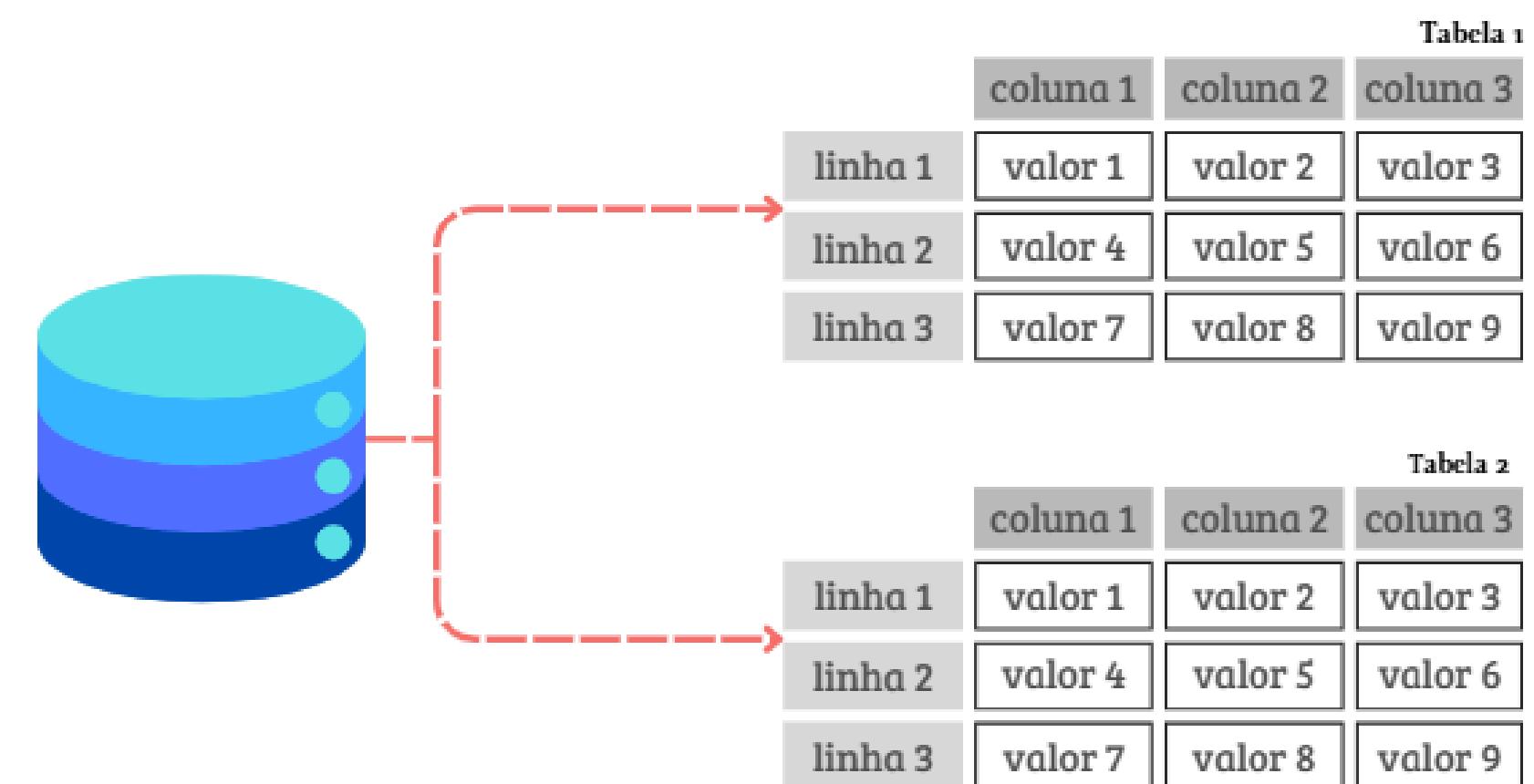
# Apache Superset



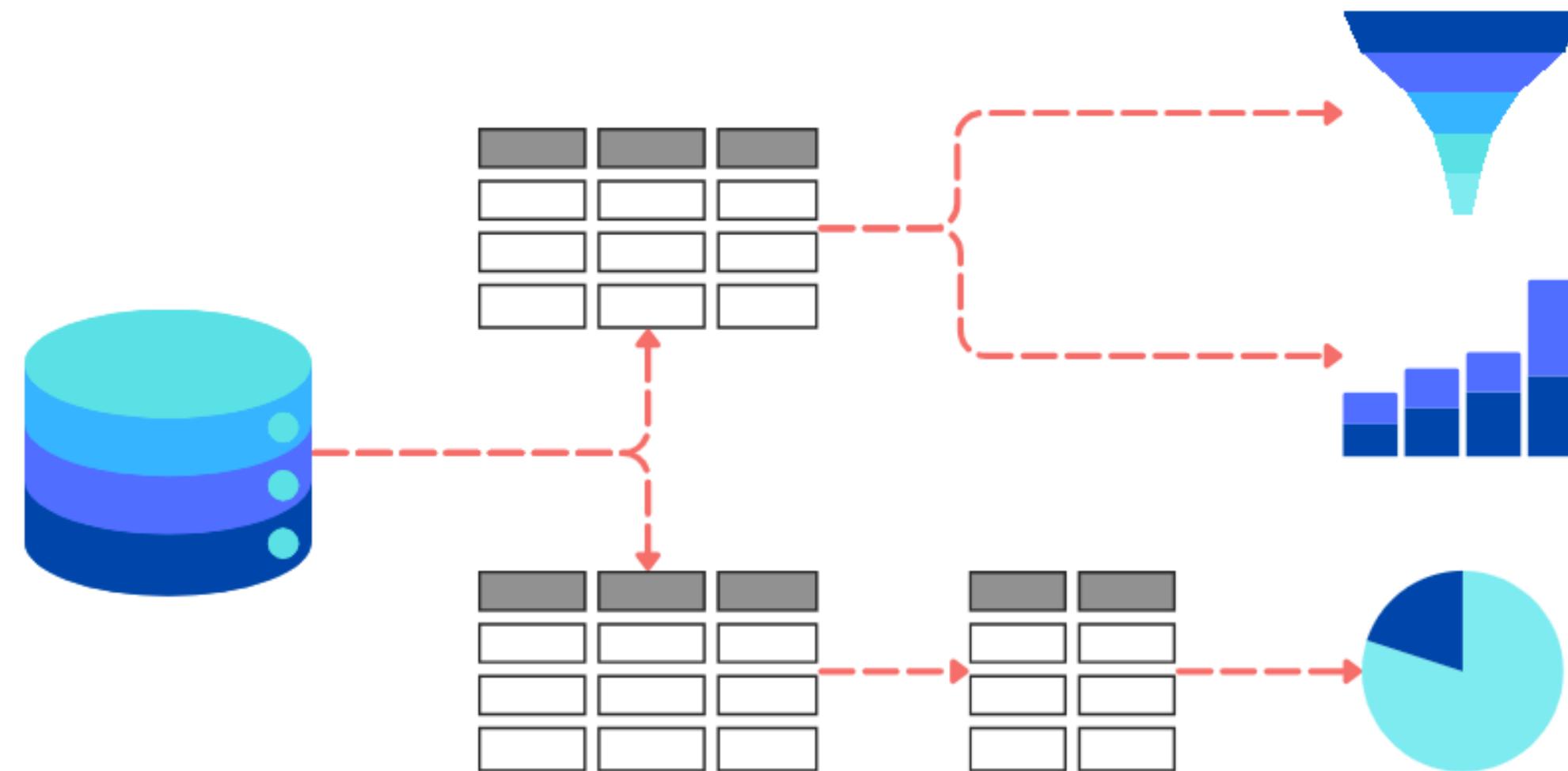
Adaptado de Yasar & Pratt 2024 (techtaget).



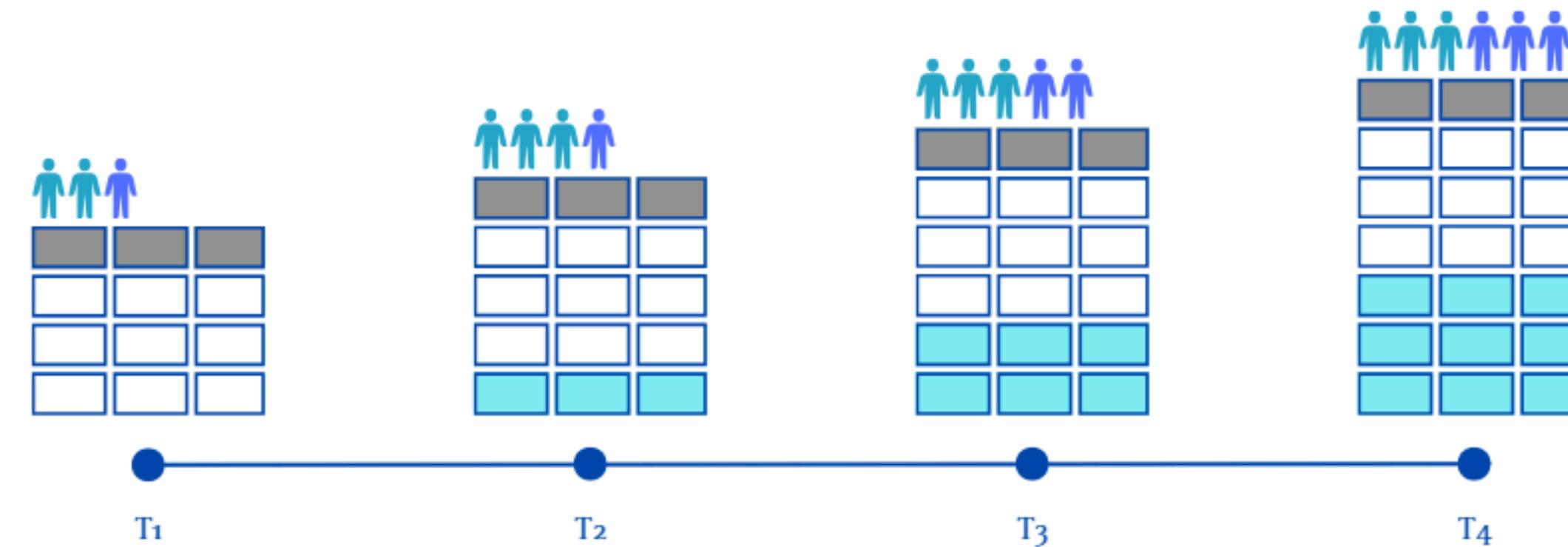
No **Apache Superset**, a análise e visualização de dados ocorrem a partir de **conjuntos de dados estruturados** – tabelas organizadas em **linhas e colunas** – extraídas de bancos de dados compatíveis com **SQL**.



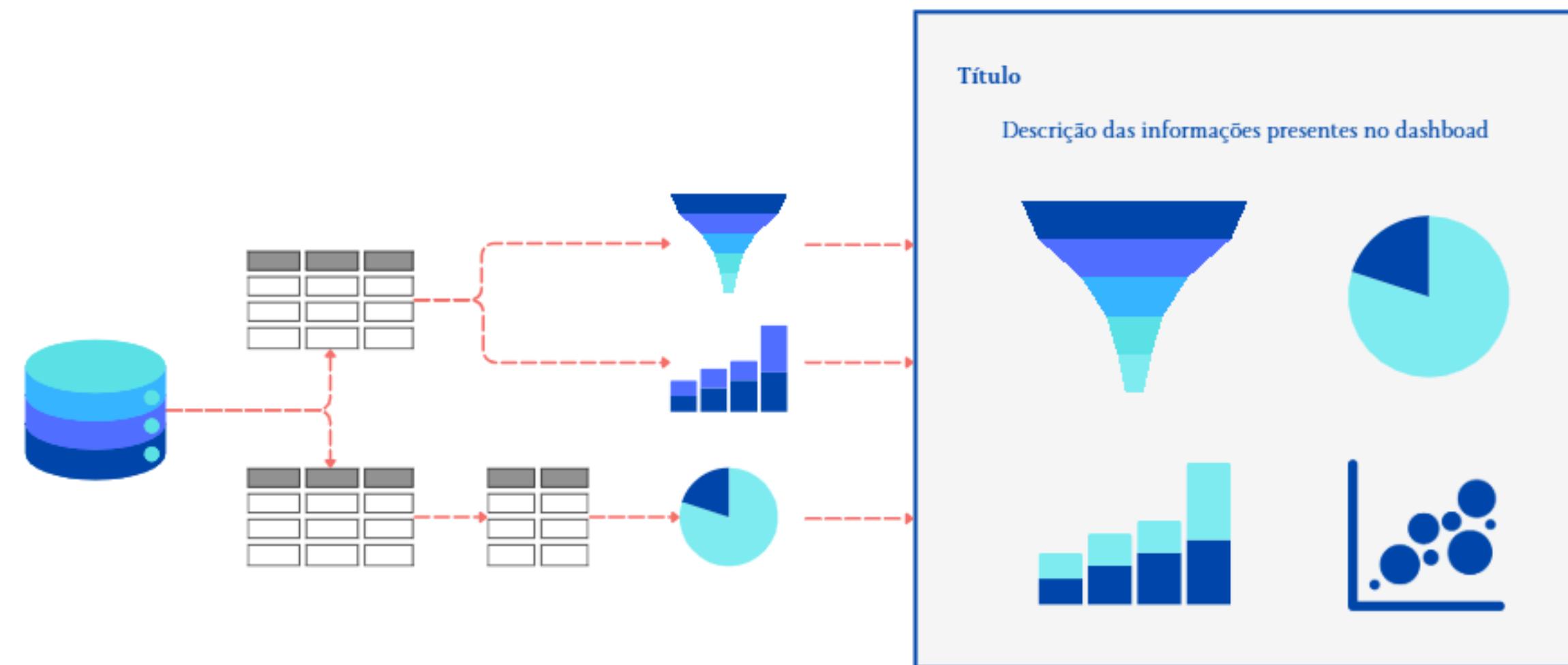
Os dados são manipulados e summarizados para compor **visualizações interativas e personalizadas**.  
O **Apache Superset** oferece **mais de 40 tipos de gráficos e visualizações**.



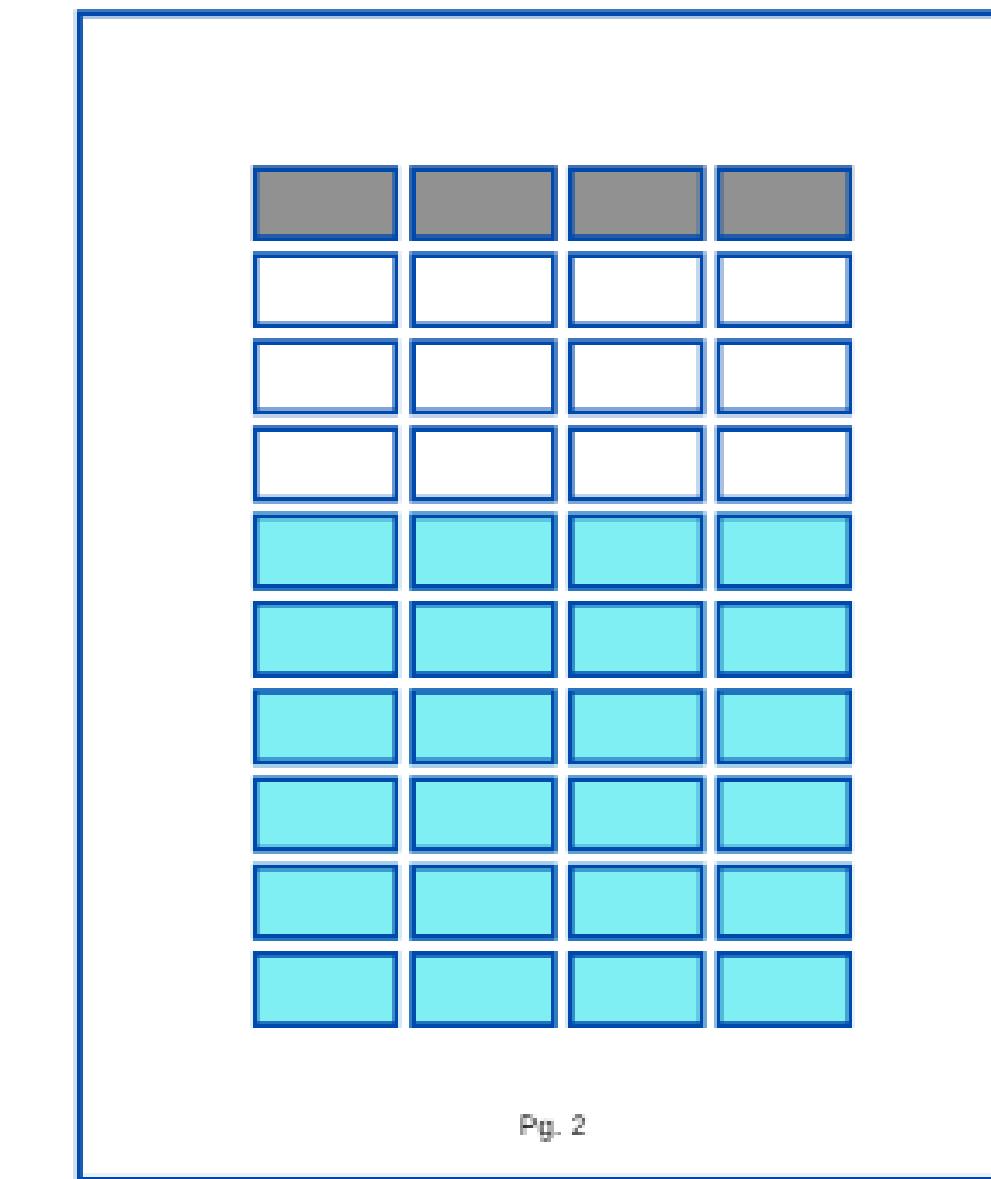
**Quando os datasets são atualizados na fonte original, o Apache Superset reflete essas mudanças de forma automática, garantindo que as visualizações de dados estejam sempre atualizadas.**



As visualizações geradas podem ser organizadas em **dashboards interativos**, fornecendo uma visão abrangente dos dados com filtros, indicadores e relatórios dinâmicos.



Os **dashboards** podem incluir tabelas e gráficos personalizados, ser exportados como **PDFs, imagens ou relatórios automatizados** e compartilhados com equipes para acompanhamento de métricas.





Performance summary

Incoming payments

Outgoing payments

Beneficiaries

Refunds

Users

Activity log

Help & Support

Resources



Multiple portals  
Flywire University, Flywire University, Flywire ...



Peter Gibbons



## Performance summary

Date 2023-05-09T17...

APPLY

Clear All



Note: Payment charts display values in the currency specific to each portal. When viewing multiple portals with different currencies, amounts remain in their respective portal currencies and are not converted into a single currency.

INCOMING PAYMENTS REFUNDS

Weekly Disbursement Amount

43.2k

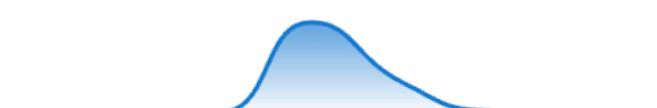
-68.1% Week Over Week



Monthly Disbursement Amount

115k

-83.1% Month Over Month



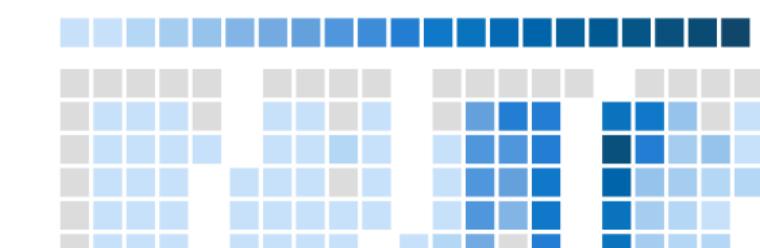
Yearly Disbursement Amount

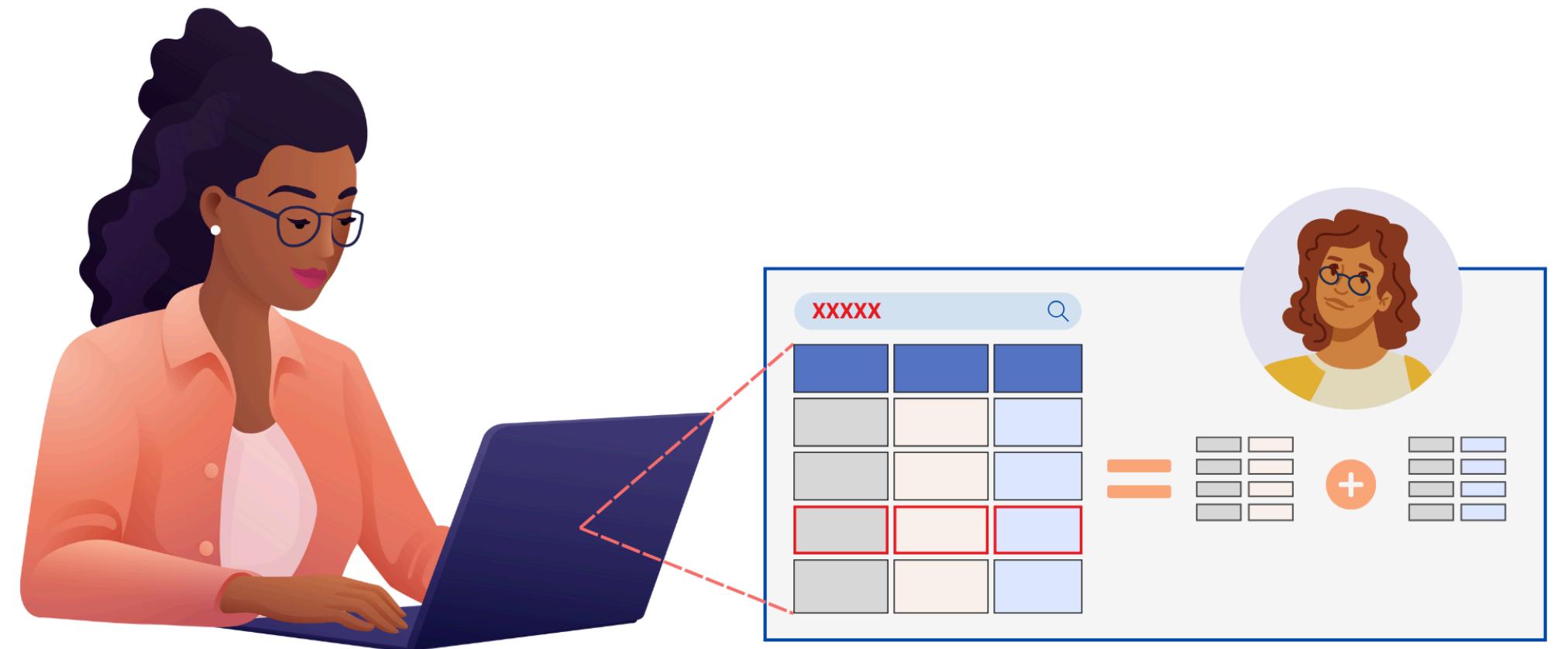
103M

-0.1% Year Over Year



Daily Payment Amount Last 365 Days





*Self-serve analytics para todos os níveis de usuários!*  
Oferece um construtor de visualizações e customizações **sem código**.



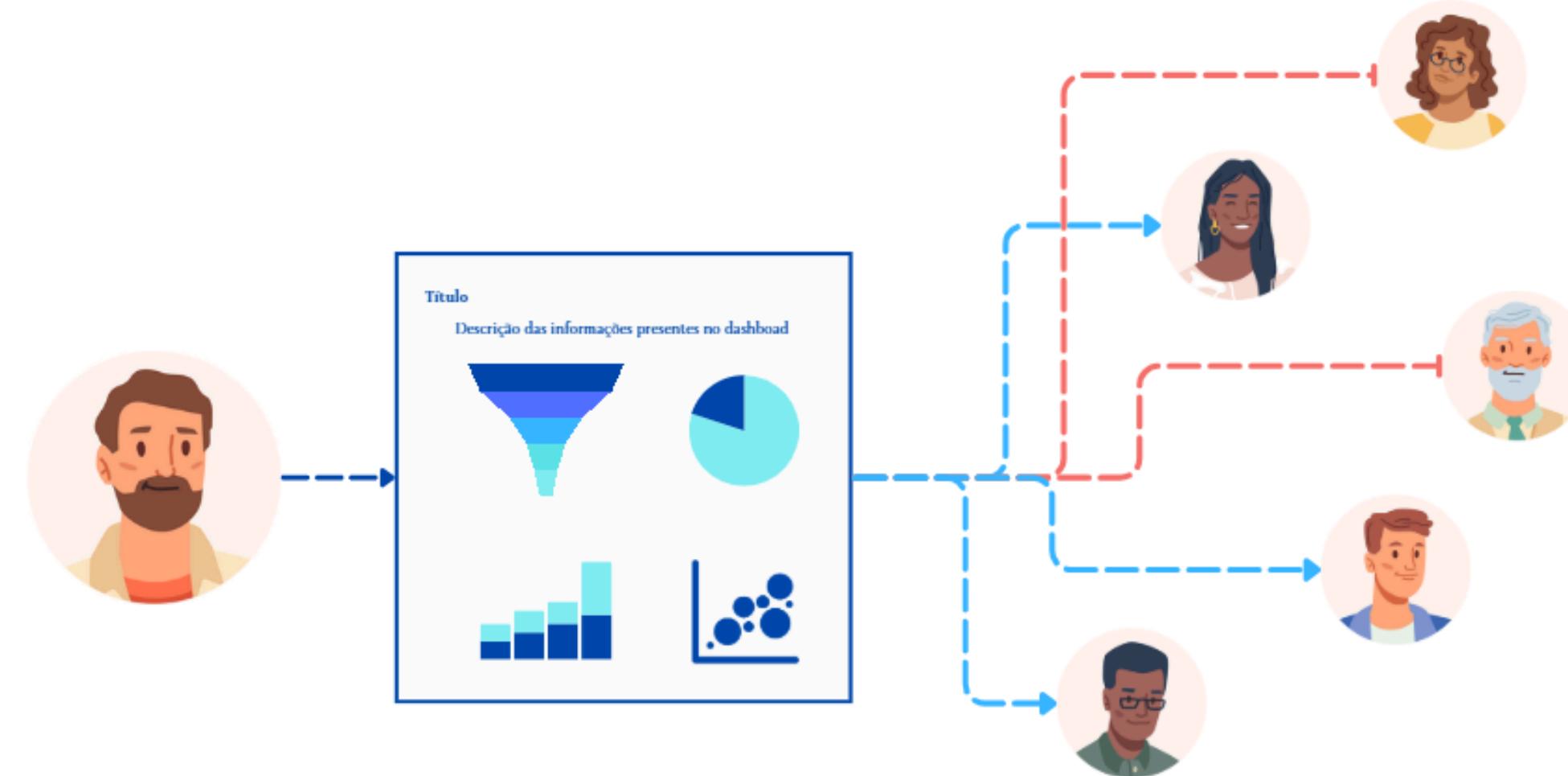
**SELECT** coluna1, coluna2 **FROM** schema.tabela1

Selezione

coluna1 e coluna 2

da

tabela 1

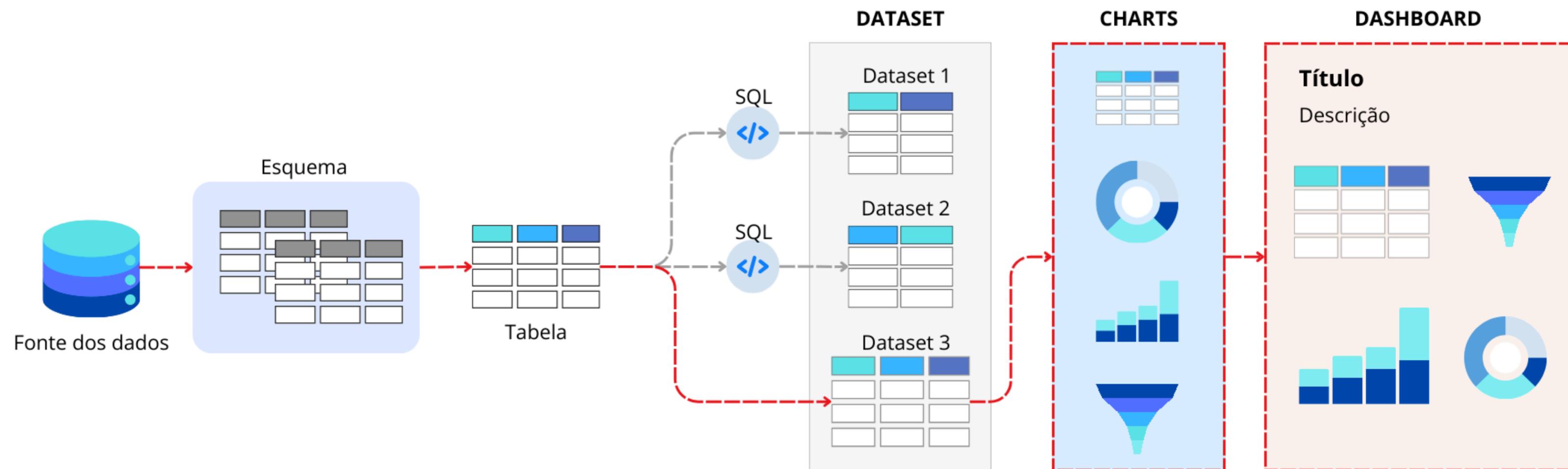


O **Apache Superset** oferece um sistema de permissões flexível, permitindo a definição de **diferentes perfis de usuários**, cada um com níveis de acesso e funcionalidades personalizáveis, garantindo controle granular sobre a plataforma.

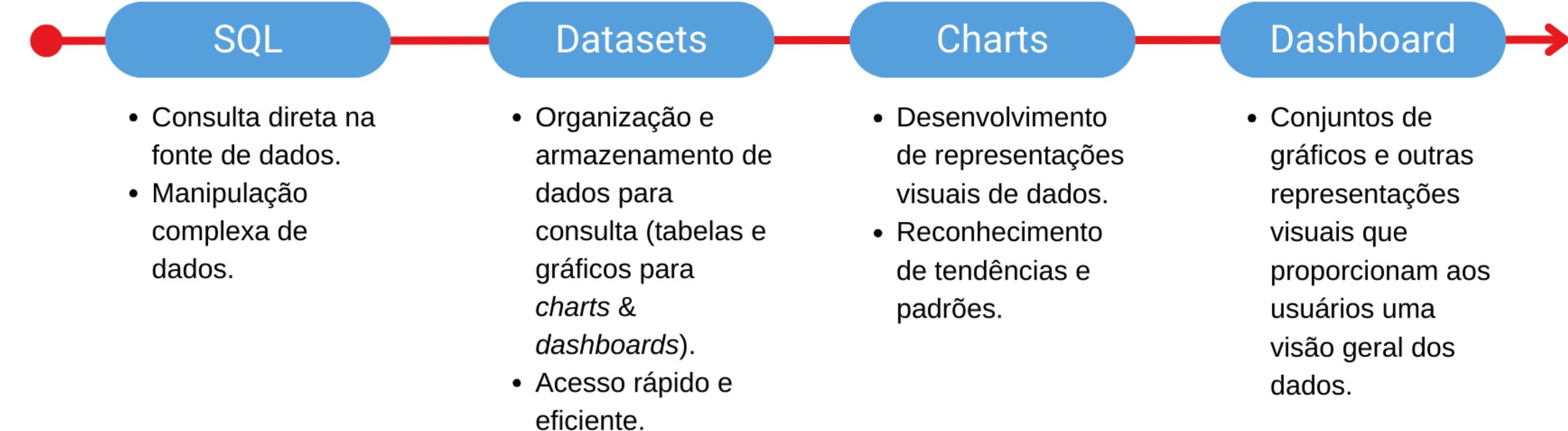
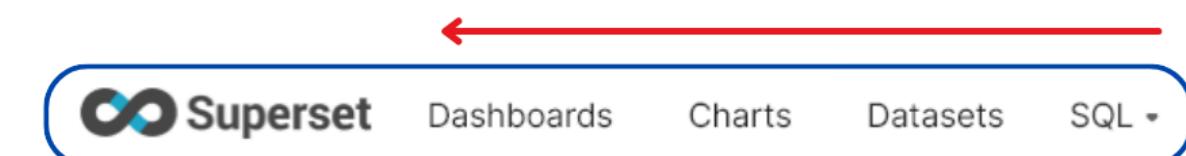
A ferramenta é um projeto de **código aberto (*open-source*)** e se beneficia de uma grande comunidade de desenvolvedores e usuários que contribuem para a sua melhoria contínua.



# Workflow



# Menu

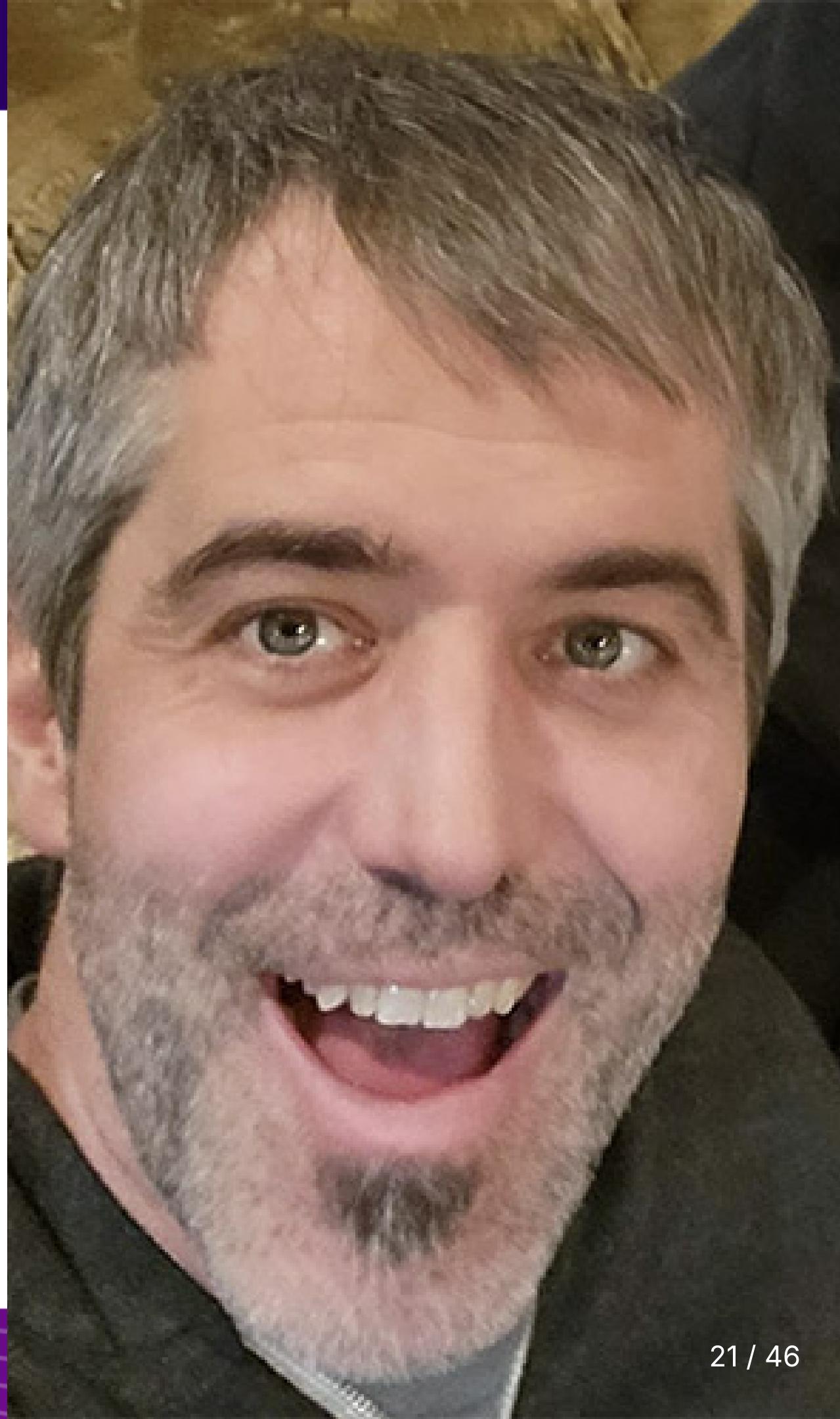


# História

## Max Beauchemin

- Projeto de hackathon no Airbnb (2015) cujo objetivo era a criação de uma ferramenta para a visualização de dados em código aberto.
- Cresceu rapidamente e superou o Tableau como principal solução de visualização de dados do Airbnb.
- Em 2016, Superset tornou-se um projeto de código aberto completo, incubado pela Apache Software Foundation.
- Empresas como Airbnb, Lyft e Twitter (X) são usuários corporativos do Superset.

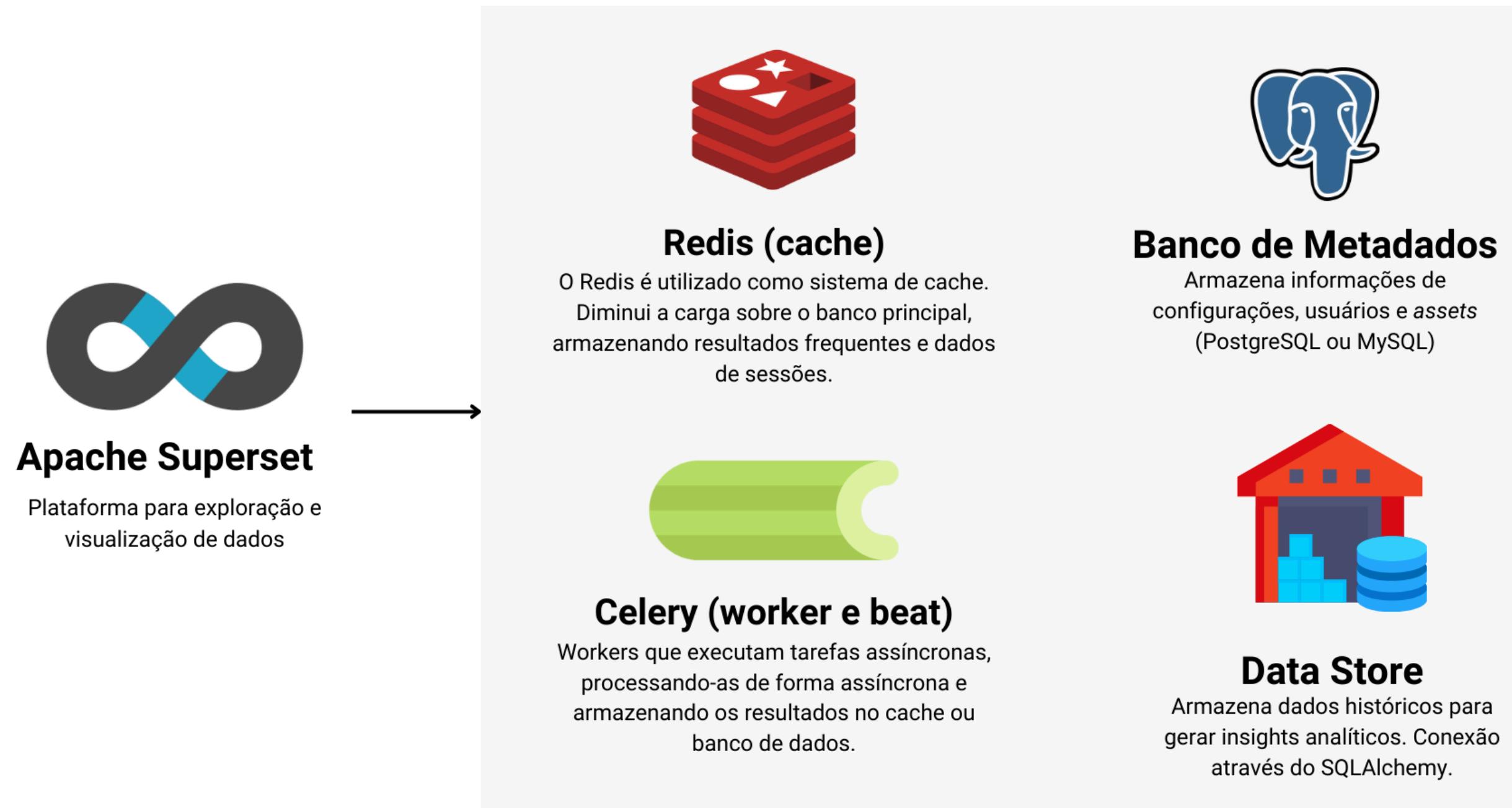
Maxime criou o **Apache Superset** para superar as limitações do **Tableau** no Airbnb, que não suportava **Presto** e **Druid**, além de ser caro e pouco escalável. O objetivo era desenvolver uma solução **open-source**, flexível e acessível para análise e visualização de dados. [🔗](#)



# Arquitetura

---

O Apache Superset foi projetado para **escalar de forma flexível junto ao seu negócio**. E quando você se sentir confiante no básico, há diversas maneiras de explorar ainda mais todo o potencial da ferramenta.



Uma instalação típica do Superset é composta pelos componentes ilustrados.

# Docker Compose

Container	Função
superset_app	Serviço principal do Superset (interface web e backend Flask)
superset_worker	Executor de tarefas assíncronas com Celery (ex: geração de gráficos, e-mails)
superset_worker_beat	Agenda tarefas periódicas no Celery (ex: atualizações agendadas)
superset_nginx	Proxy reverso que expõe o Superset na porta 80
superset_db	Banco de dados PostgreSQL que armazena metadados do Superset
superset_cache	Redis usado como cache e backend do Celery
superset_init	Executa a configuração inicial (migrar DB, criar admin, etc.) — finaliza após rodar
superset_node	Servidor do frontend em modo dev (Webpack dev server com hot reload)
superset_websocket	Suporte a comunicação WebSocket (ex: dashboards em tempo real)

# Vantagens

- ▶ **Escalável:** fácil de crescer, só adicionar mais servidores
- ▶ **Rápido:** usa **Redis (caching)** para acelerar consultas
- ▶ **Robusto:** tarefas pesadas rodam separadas, sem travar o sistema
- ▶ **Conectável:** funciona com vários bancos de dados via SQLAlchemy

Ideal para empresas que precisam de **desempenho** e **crescimento fácil**

# Fontes de dados

A ferramenta se conecta a diversas fontes de dados compatíveis com **SQL**, incluindo **Presto**, **Trino**, **Athena** e muitos outros.

-  Suporte a **bancos relacionais e data warehouses**
-  Utiliza **SQLAlchemy** para abstração e integração com diferentes bancos de dados.



# Funcionalidades

---

*"Open-source é uma abordagem superior para criar e distribuir software porque fornece garantias importantes sobre como o software pode ser descoberto, testado, operado, colaborado e empacotado."*

Maxime Beauchemin (2021) 

# RBAC (Role-Based Access Control)

*Controle de acesso baseado em funções*

Define **o que o usuário pode ver e fazer**, de acordo com seus **papéis (roles)**.

**Exemplos de permissões:**

- ▶ Ver dashboards
- ▶ Editar datasets
- ▶ Executar SQL
- ▶ Gerenciar usuários

**Como funciona**

- Roles padrão: *Admin, Alpha, Gamma*
- Possível criar roles personalizados
- Usuários podem ter múltiplos roles

### Add Data Access Role

Name \*  The name of the data access role. Create a nice understandable role name like: Sales Data Access

Users

A list of users to associate this role to. These users will have data access to the list of permissions contained on this role

Permissions

- All database access
- All dataset access** 
- All query access
- Database access on BigQuery - Financial
- Database access on BigQuery - Financial.(id:None)
- Schema access on BigQuery - Financial.Cryptoanalysis1
- Dataset access on BigQuery - Financial.BTC-USD
- Database access on examples

**SAVE**

# RLS (Row-Level Security)

*Segurança em nível de linha*

Restringe **quais dados o usuário pode ver**, aplicando **filtros por linha**.

**Exemplo:**

João vê só dados da região **Sudeste**,  
Maria vê apenas da **Região Sul**.

```
WHERE regiao IN ('sul')
```

**Como funciona**

- Filtros criados em *Security > Row Level Security*
- Associados a datasets e roles
- Aplicados automaticamente nas queries

# Row Level Security (RLS)

## Segurança em nível de linha (RLS)

Edit Row level security filter

Filter Type	Regular
Regular filters add where clauses to queries if a user belongs to a role referenced in the filter. Base filters apply filters to all queries except the roles defined in the filter, and can be used to define what users can see if no RLS filters within a filter group apply to them.	
Tables	* public.Video Game Sales
These are the tables this filter will be applied to.	
Roles	* Wii Platform
Create roles in Data Access Roles. For regular filters, these are the roles this filter will be applied to. For base filters, these are the roles that the filter DOES NOT apply to, e.g. Admin if admin should see all data.	
Group Key	platform
Filters with the same group key will be ORed together within the group, while different filter groups will be ANDed together. Undefined group keys are treated as unique groups, i.e. are not grouped together. For example, if a table has three filters, of which two are for departments Finance and Marketing (group key = 'department'), and one refers to the region Europe (group key = 'region'), the filter clause would apply the filter (department = 'Finance' OR department = 'Marketing') AND (region = 'Europe').	
Clause *	platform = 'Wii'
This is the condition that will be added to the WHERE clause. For example, to only return rows for a particular client, you might define a regular filter with the clause 'client_id = 9'. To display no rows unless a user belongs to a RLS filter role, a base filter can be created with the clause '1 = 0' (always false).	
<button>SAVE</button> <a href="#">Cancel</a>	

# SSO (Single Sign-On)

*Acesso único para múltiplos sistemas*

Permite que o usuário **faça login uma única vez** para acessar várias ferramentas (ex: Superset, Slack, Drive).

## Como funciona:

- ▶ Integração com um provedor de identidade (ex: Google, Azure AD, Okta)
- ▶ Autenticação centralizada e segura

## Vantagens:

Menos senhas, mais segurança e melhor experiênciass

# Alertas e Relatórios

- Cria relatórios de forma automática
- Notificações via email ou Slack
- Gatilhos com consultas SQL

## Alert condition

DATABASE \*

examples

SQL QUERY \*

```
1 SELECT COUNT("user_action_id") as "nrfails" from user_trac  
2 WHERE  
3   "user_action_id" = '1003'  
4 AND  
5   "dtm" > '2020-12-23 00:00:01'
```

TRIGGER ALERT IF...

> (Larger than)

VALUE \*

0

# Jinja

*Template para consultas eficientes*

- Variáveis para **parametrizar** filtros
- **Reutilize** e simplifique consultas SQL
- Para análises **iterativas** e **complexas**



```
SELECT *
FROM tbl
WHERE (
    {% if from_dttm is not none %}
        dttm_col > '{{ from_dttm }}' AND
    {% endif %}
    {% if to_dttm is not none %}
        dttm_col < '{{ to_dttm }}' AND
    {% endif %}
    1 = 1
)
```

# Grupos

Metrics <span style="border: 1px solid #0070C0; border-radius: 50%; padding: 2px;">3</span>	Columns <span style="border: 1px solid #0070C0; border-radius: 50%; padding: 2px;">11</span>	Calculated Columns <span style="border: 1px solid #0070C0; border-radius: 50%; padding: 2px;">1</span>	Settings
Metric	Label	SQL Expression	
▶ count	COUNT(*)	COUNT (*)	
▶ Cases Latin America	New Cases Latin America	<pre>SUM( CASE WHEN country IN ( 'Mexico','Guatemala', 'Honduras', 'El Salvador', 'Nicaragua',  'Costa Rica','Panama' , 'Colombia', 'Venezuela', 'Ecuador','Peru', 'Bolivia', 'Chile', 'French Guiana', 'Paraguay', 'Brazil', 'Argentina','Uruguay','Cuba','Dominican Republic', 'Haiti' ) THEN new_cases ELSE 0 END )</pre>	
▶ Cases_Europe	Cases Europe	<pre>sum(case when country in ('Albania','Andorra','Armenia','Austria','Azerk Herzegovina','Bulgaria','Croatia','Cyprus','Czechia','Denmark','Estonia' 'Macedonia','Norway','Poland','Portugal','Romania','Russia','San Marino',</pre>	

# Importar / Exportar

*Migração e versionamento de dados no Superset*

- ▶ Compartilhe dashboards e datasets
- ▶ Suporte a CSV, Excel, JSON, YAML
- ▶ Migração entre instâncias
- ▶ Versionamento em código

## **Usos comuns:**

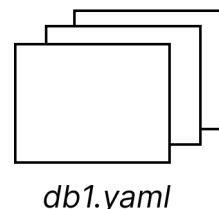
Backup, migração, compartilhamento e CI/CD

# Github

# Preset

## Files in Git

Database Assets



db1.yaml

Charts Assets

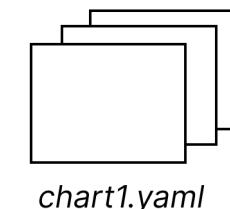
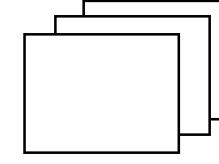


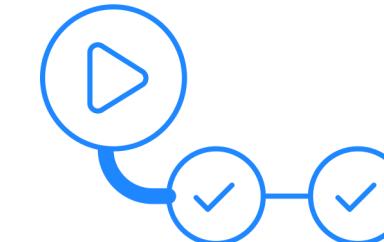
chart1.yaml

Dashboard Assets

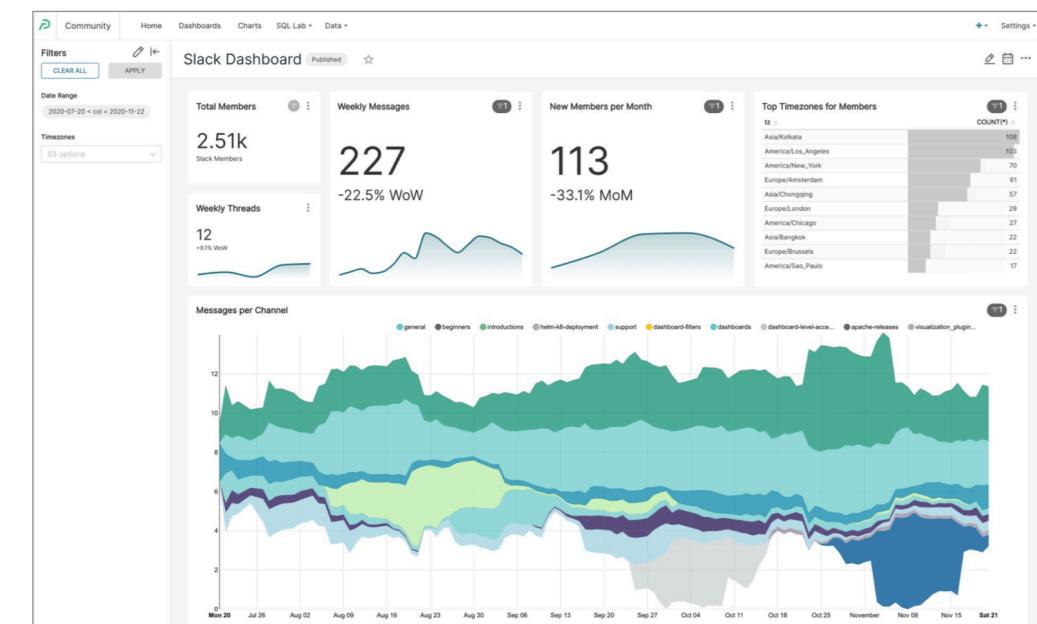


dashboard1.yaml

## Github Actions



## Preset Workspace



# Embed

*Incorpore em outras aplicações*

- ▶ Geração de **links embed**
- ▶ Portais, intranets e aplicações web

## Exemplos de uso:

Portais corporativos, sistemas internos, relatórios interativos

### Embed

This dashboard is ready to embed. In your application, use the following values to configure the SDK:

- Embedded Dashboard ID: `132d5a9f-cce6-40f7-a850-f881a3ee5741`
- Superset Domain: <https://011e6a5e.us2a.app.preset.io>
- Workspace ID: `011e6a5e`
- Team ID: `e7d91b7d`

For further instructions, consult the [Preset Frontend SDK documentation](#).

#### Settings

Allowed Domains (comma separated)  ⓘ

[DEACTIVATE](#) [SAVE CHANGES](#)

# Outros

- ▶ Maptools
- ▶ Handlebars
- ▶ Thumbnails
- ▶ ...

Consulte a documentação oficial, em configurações!

# Funcionalidades opcionais

Algumas funcionalidades adicionais do Apache Superset dependem da ativação de componentes auxiliares, como a **camada de cache**.

Por exemplo, os **alertas e relatórios automatizados** e as **miniaturas (thumbnails) dos dashboards**.

 Já habilitado com o docker-compose!

# Instalação

---

*The fastest way to try Superset locally is using Docker Compose on a Linux or Mac OSX computer. Superset does not have official support for Windows (documentação oficial [🔗](#))*

# Docker

A comunidade Apache Superset usa extensivamente o **Docker** para desenvolvimento, lançamento e produção do Superset.

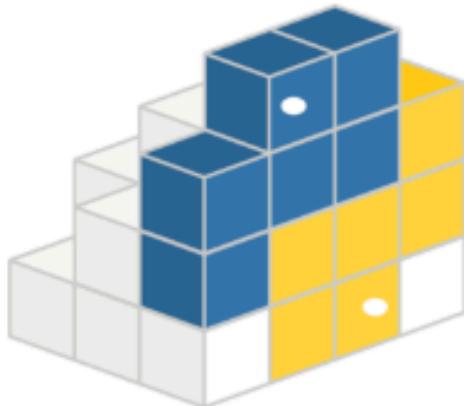
## \*\*Docker-compose\*\*:

Template para versão de desenvolvimento e para produção.

```
2 # Licensed to the Apache Software Foundation (ASF) under one or more
3 # contributor license agreements. See the NOTICE file distributed with
4 # this work for additional information regarding copyright ownership.
5 # The ASF licenses this file to You under the Apache License, Version 2.0
6 # (the "License"); you may not use this file except in compliance with
7 # the License. You may obtain a copy of the License at
8 #
9 #     http://www.apache.org/licenses/LICENSE-2.0
10 #
11 # Unless required by applicable law or agreed to in writing, software
12 # distributed under the License is distributed on an "AS IS" BASIS,
13 # WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
14 # See the License for the specific language governing permissions and
15 # limitations under the License.
16 #
17 #
18 # -----
19 # We don't support docker compose for production environments.
20 # If you choose to use this type of deployment make sure to
21 # create your own docker environment file (docker/.env) with your own
22 # unique random secure passwords and SECRET_KEY.
23 #
24 x-superset-image: &superset-image apachesuperset.docker.scarf.sh/apache/superset:${TAG:-latest-dev}
25 x-superset-volumes:
26   &superset-volumes # /app/pythonpath_docker will be appended to the PYTHONPATH in the final container
27   - ./docker:/app/docker
28   - superset_home:/app/superset_home
29
30 >Run All Services
31 services:
32   >Run Service
33   redis:
34     image: redis:7
35     container_name: superset_cache
36     restart: unless-stopped
37     volumes:
38       - redis:/data
39
38 >Run Service
39 db:
40   env_file:
41     - path: docker/.env # default
42       required: true
43     - path: docker/.env-local # optional override
44       required: false
45     image: postgres:16
46     container_name: superset_db
47     restart: unless-stopped
48     volumes:
49       - db_home:/var/lib/postgresql/data
50       - ./docker/docker-entrypoint-initdb.d:/docker-entrypoint-initdb.d
```



# Installing Superset from PyPI



This page describes how to install Superset using the [apache-superset](#) package published on PyPI.

## OS Dependencies

Superset stores database connection information in its metadata database. For that purpose, we use the [cryptography](#) Python library to encrypt connection passwords. Unfortunately, this library has OS level dependencies.

### Debian and Ubuntu

In Ubuntu **20.04 and 22.04** the following command will ensure that the required dependencies are installed:

```
sudo apt-get install build-essential libssl-dev libffi-dev python3-dev python3-pip libsasl2-dev lib
```

OS Dependencies

Python Virtual Environ

Installing and Initia

# Código e Ferramenta

---

*"The Future of Business Intelligence is Open Source"*

Maxime Beauchemin (2021) 

# Referências

- **Preset (ferramenta baseada em Superset)**

<https://docs.preset.io>

- **Documentação oficial do Apache Superset**

<https://superset.apache.org/docs/>

- **Repositório no GitHub (Apache Superset)**

<https://github.com/apache/superset>

- **Docker e Docker Compose**

<https://docs.docker.com/engine/install/ubuntu/>

- **Artigo do criador do Superset (Max Beauchemin)**

["The Future of Business Intelligence is Open Source" – Medium](#)

# Obrigada!

---

**Marília Melo Favalesso - PhD, Cientista de Dados**

 LinkedIn: /mariliafavalesso  Email: marilia.melo.favalesso@gmail.com