



84° Python Floripa

Apache Superset

Marília Melo Favalesso

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

 marilia.melo.favalesso@gmail.com

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

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

 site: 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.

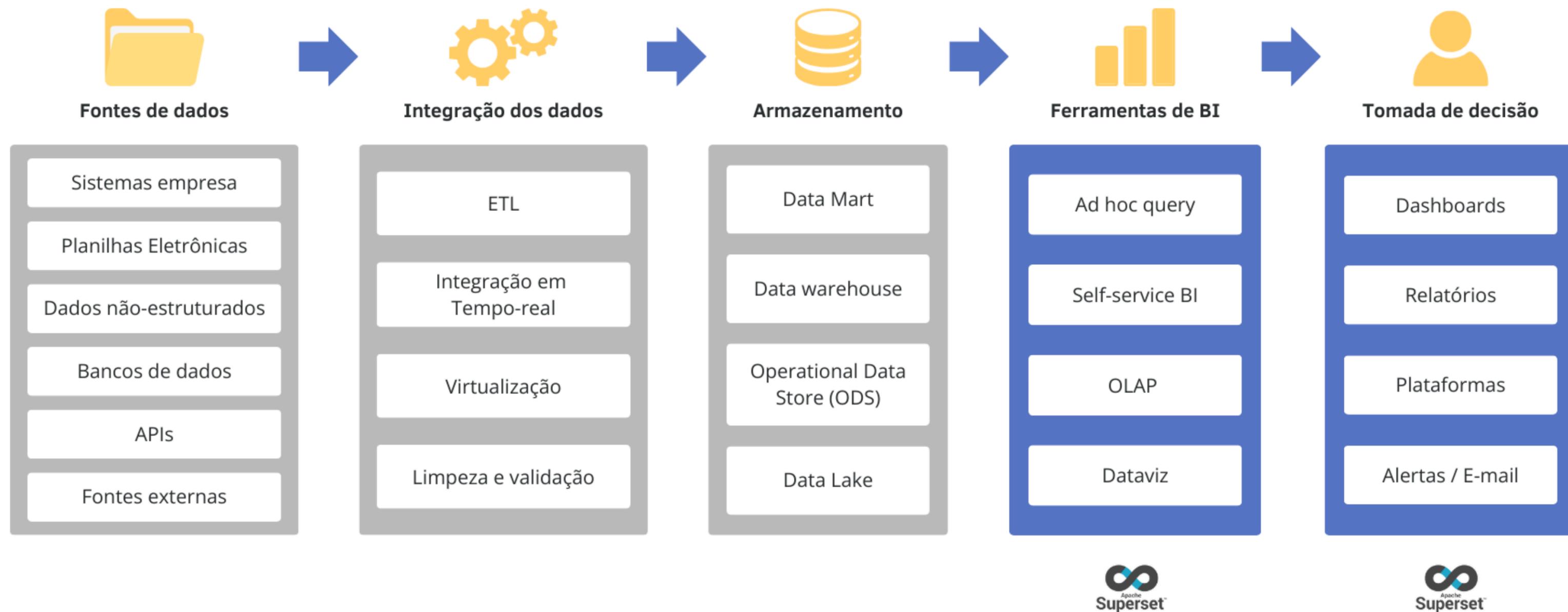
Inteligência de Negócios

- 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**.

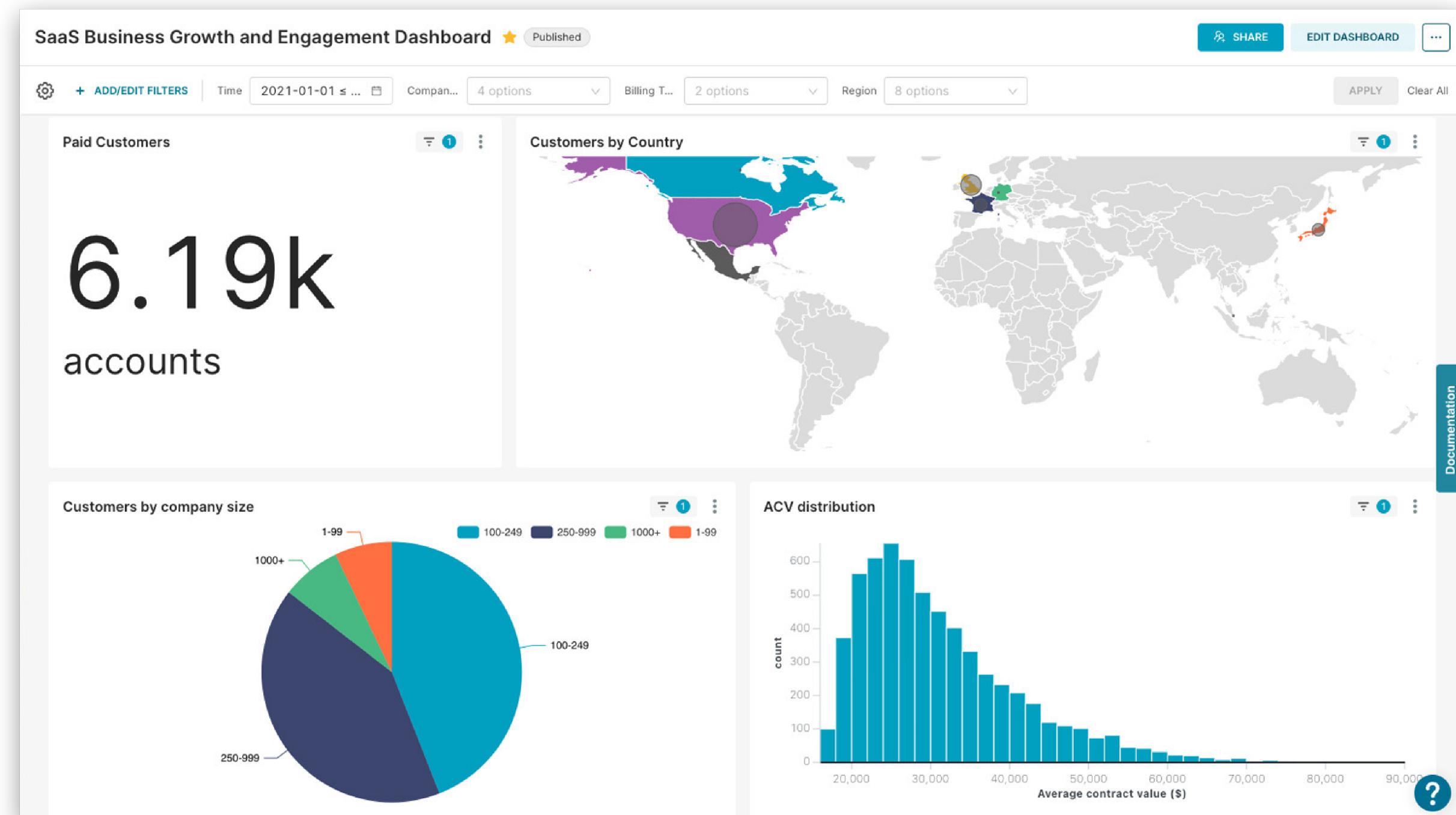
- ❖ O objetivo é o de apoiar decisões inteligentes e baseadas em fatos
- ❖ ~Data-Driven(DWx 2025)!



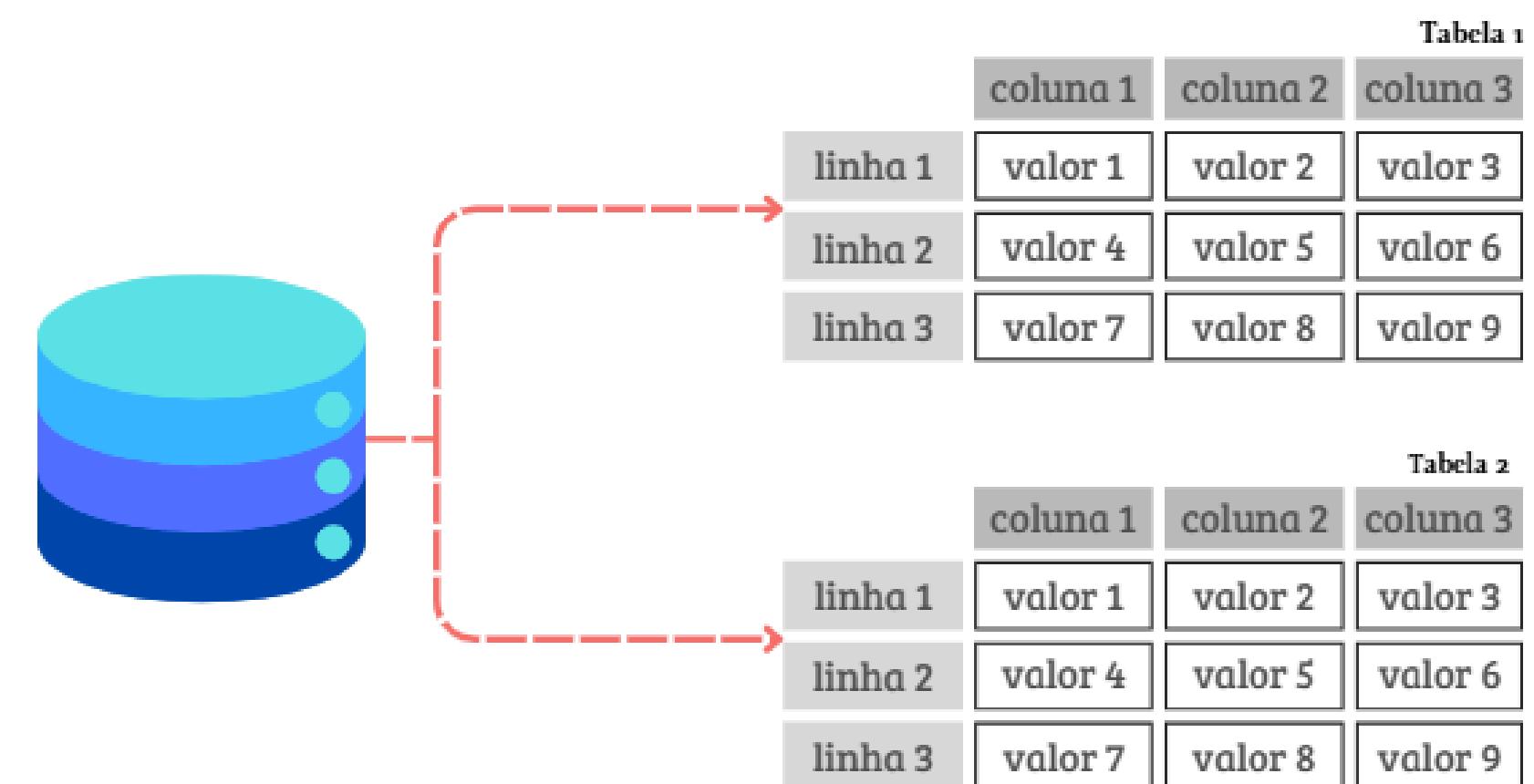
Apache Superset



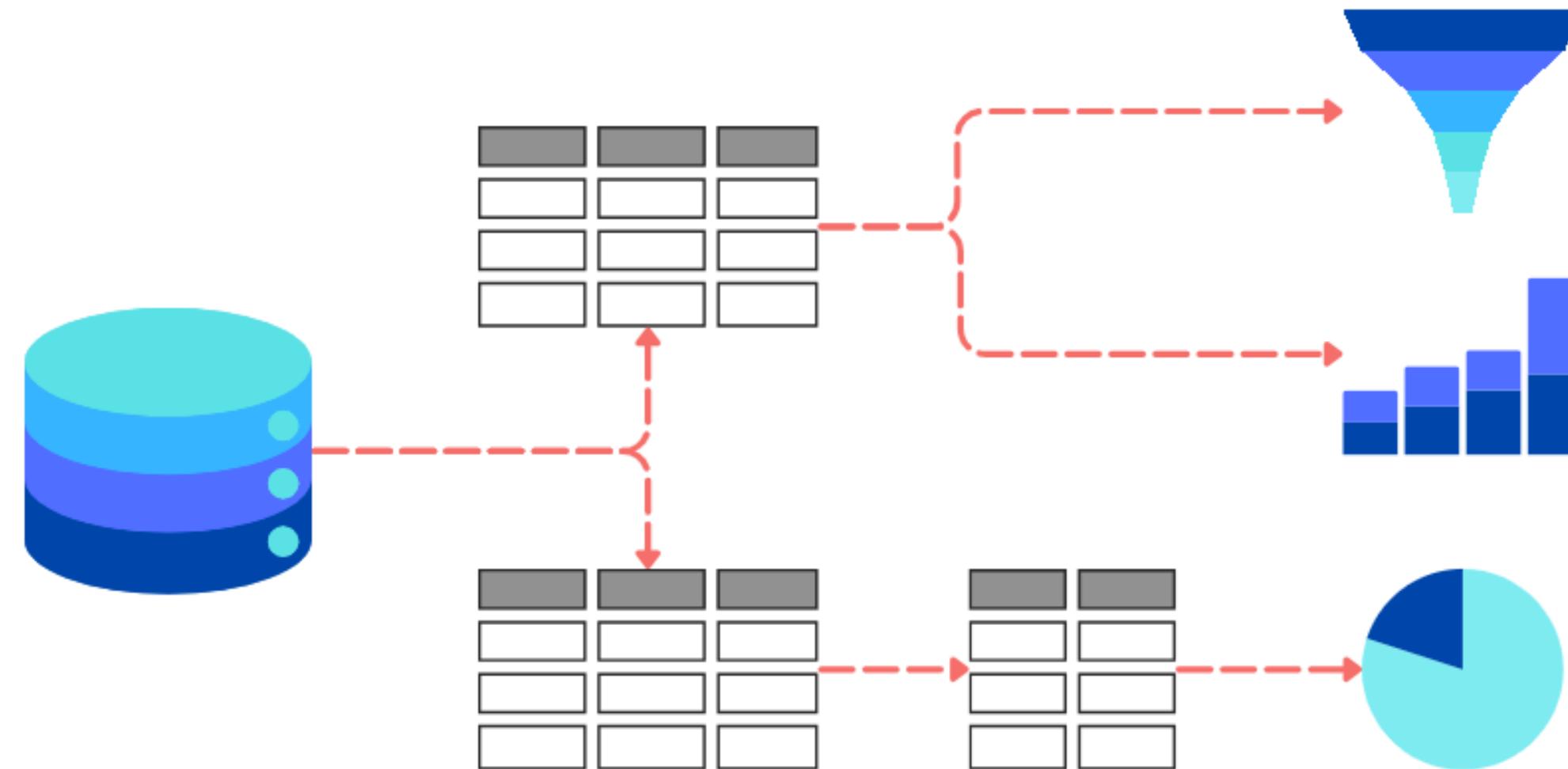
Adaptado de Yasar & Pratt 2024 (techtaget).



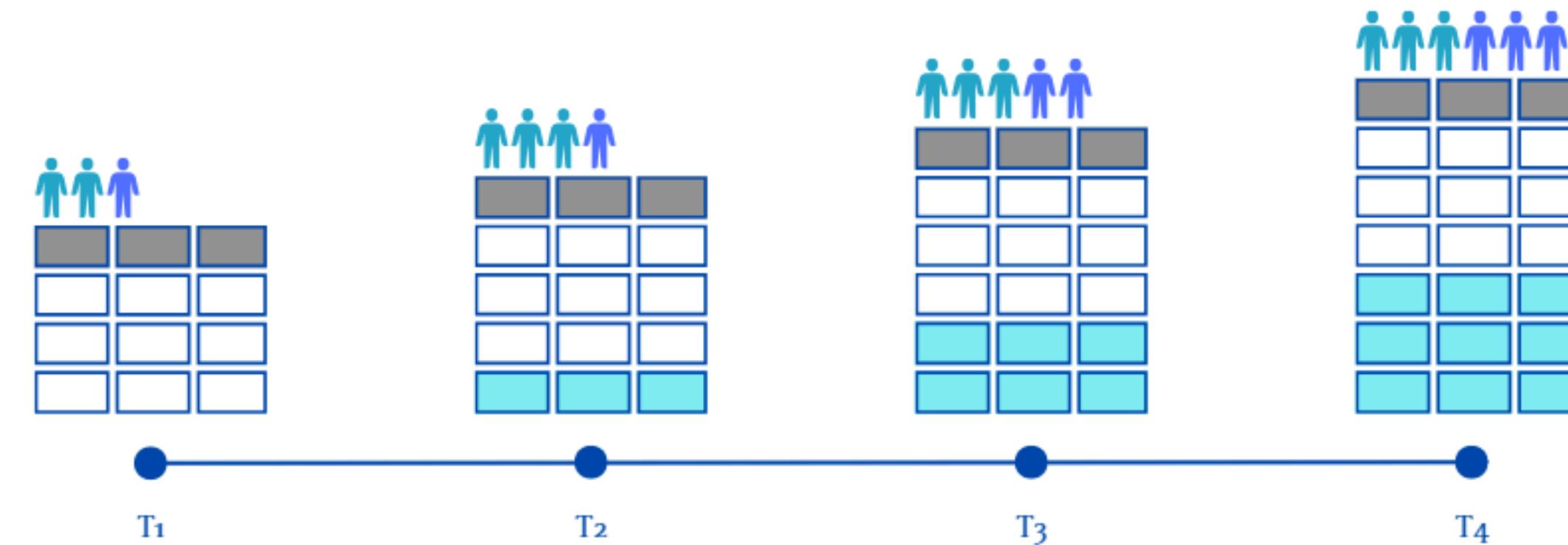
No **Apache Superset**, a análise e visualização de dados ocorrem a partir de **conjuntos de dados estruturados** – tabelas organizadas em **linhas e colunas** – extraídas de bancos de dados compatíveis com **SQL**.



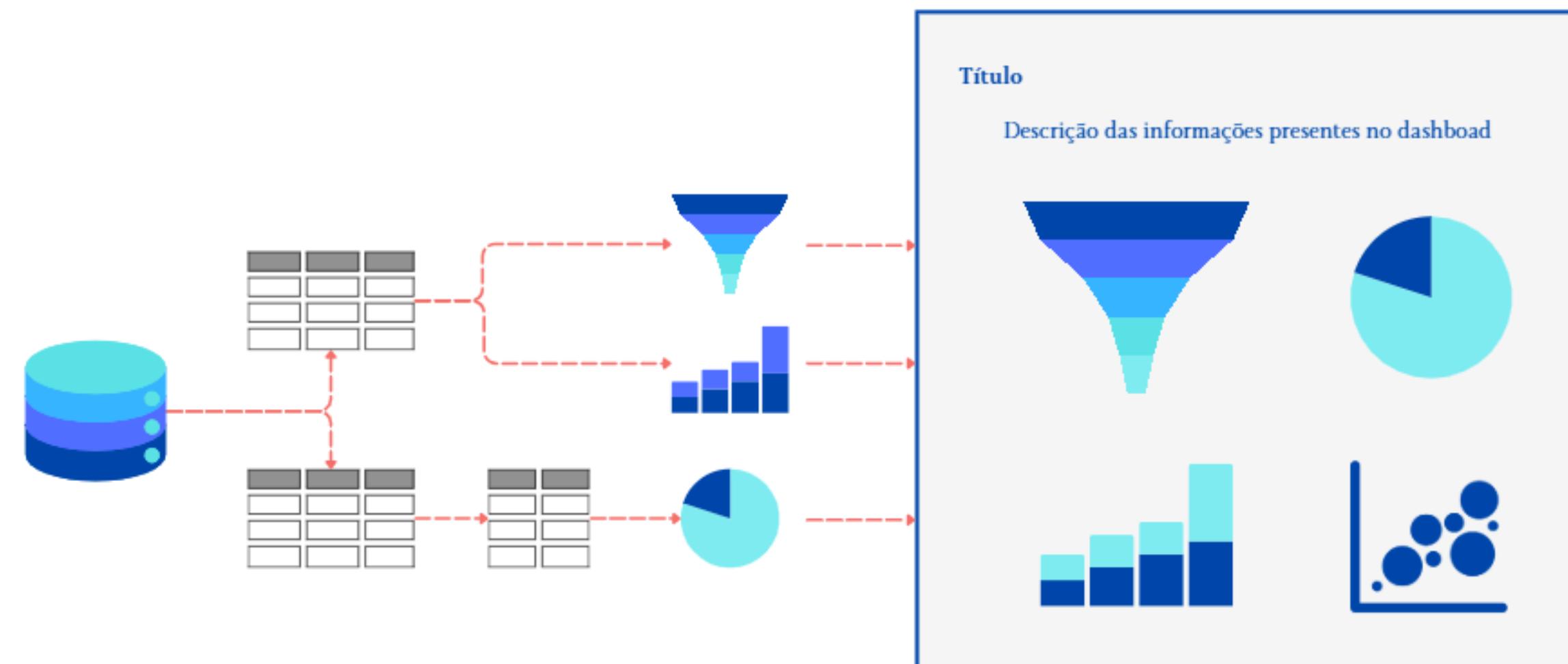
Os dados são manipulados e summarizados para compor **visualizações interativas e personalizadas**.
O **Apache Superset** oferece **mais de 40 tipos de gráficos e visualizações**.



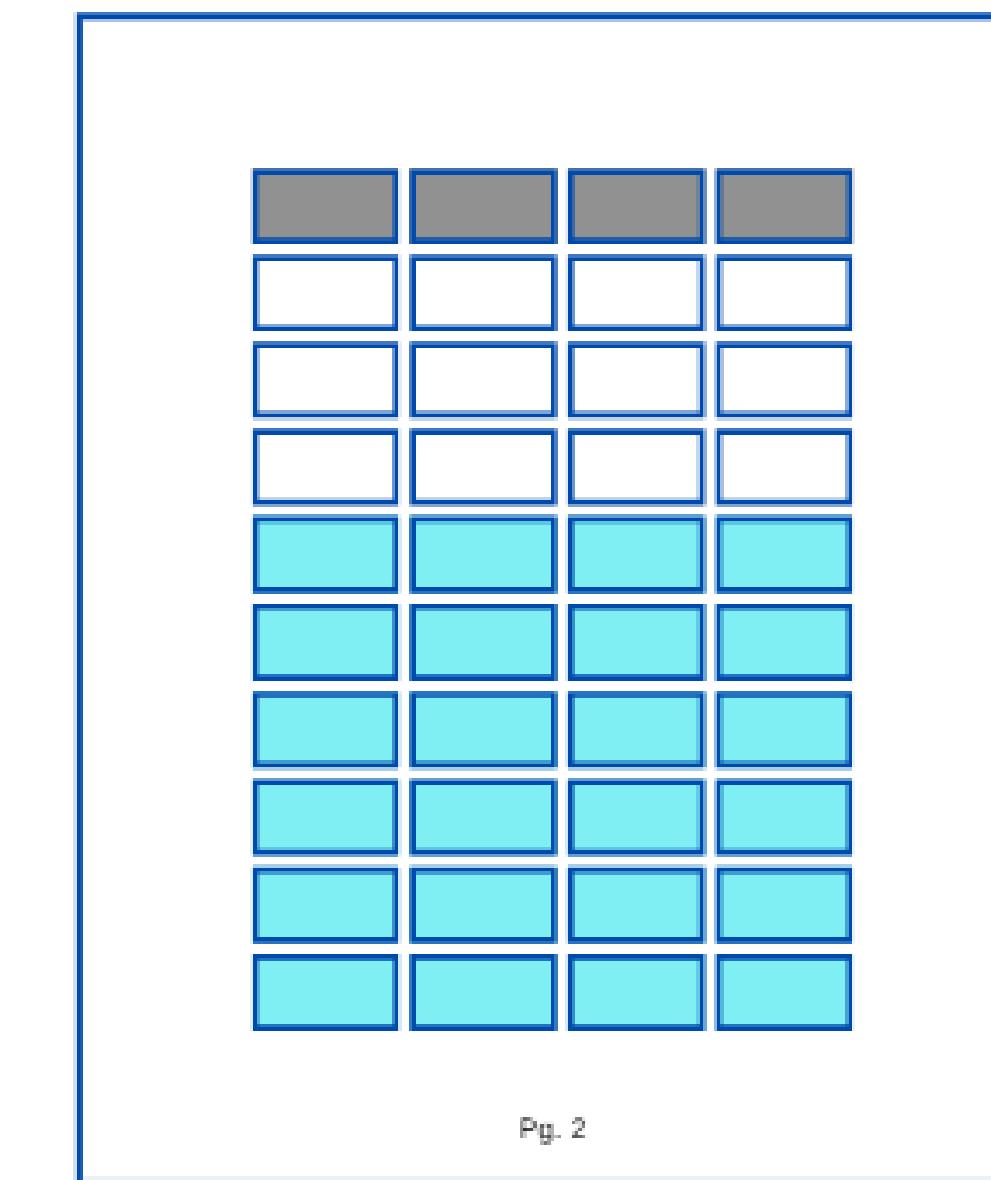
Quando os datasets são atualizados na fonte original, o Apache Superset reflete essas mudanças de forma automática, garantindo que as visualizações de dados estejam sempre atualizadas.

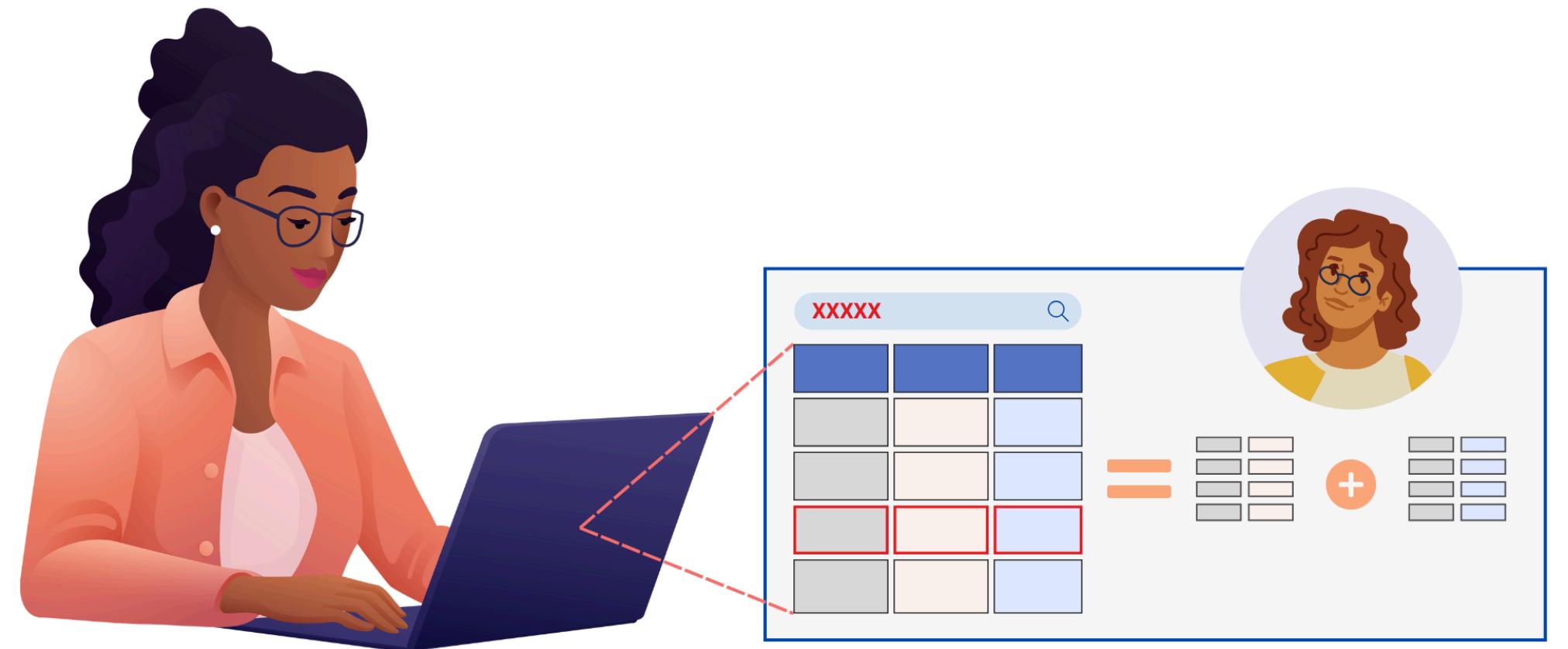


As visualizações geradas podem ser organizadas em **dashboards interativos**, fornecendo uma visão abrangente dos dados com filtros, indicadores e relatórios dinâmicos.

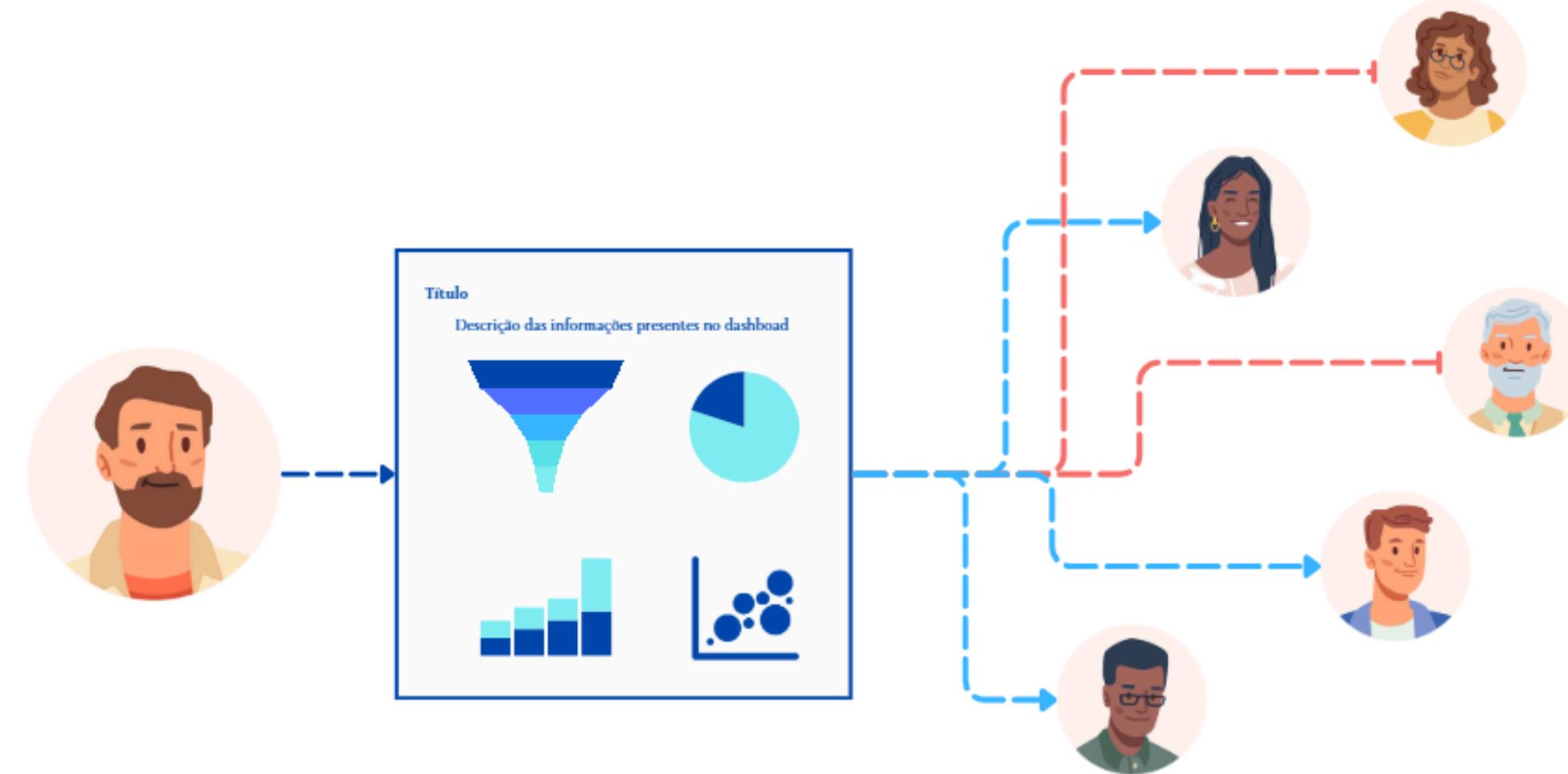


Os **dashboards** podem incluir tabelas e gráficos personalizados, ser exportados como **PDFs, imagens ou relatórios automatizados** e compartilhados com equipes para acompanhamento de métricas.





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

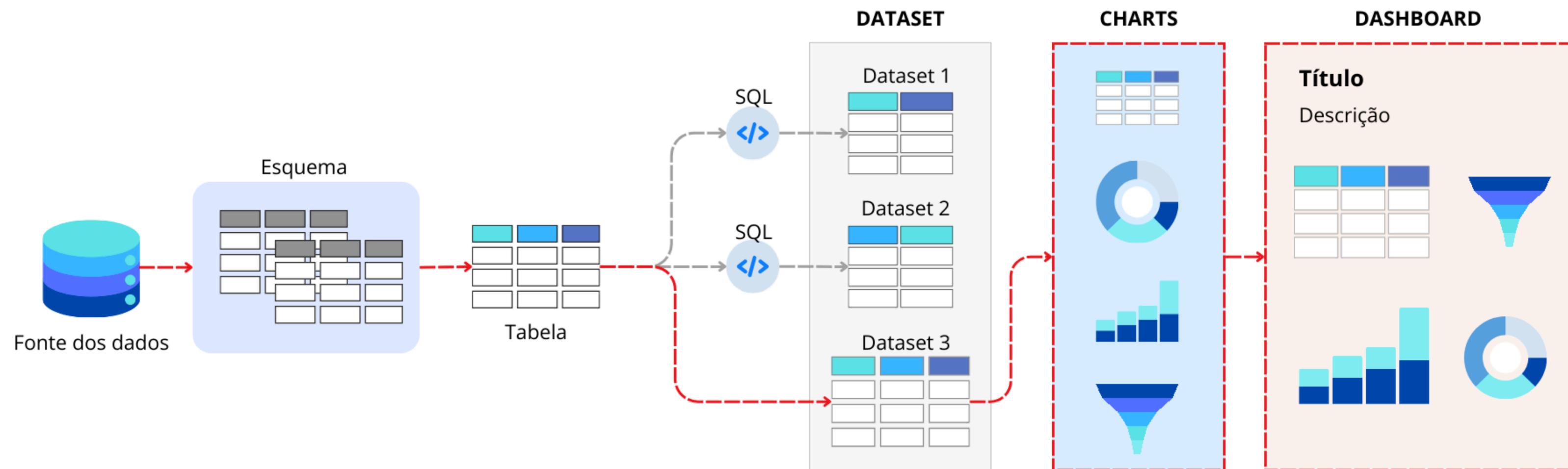


O **Apache Superset** oferece um sistema de permissões flexível, permitindo a definição de **diferentes perfis de usuários**, cada um com níveis de acesso e funcionalidades personalizáveis, garantindo controle granular sobre a plataforma.

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



História

- 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.

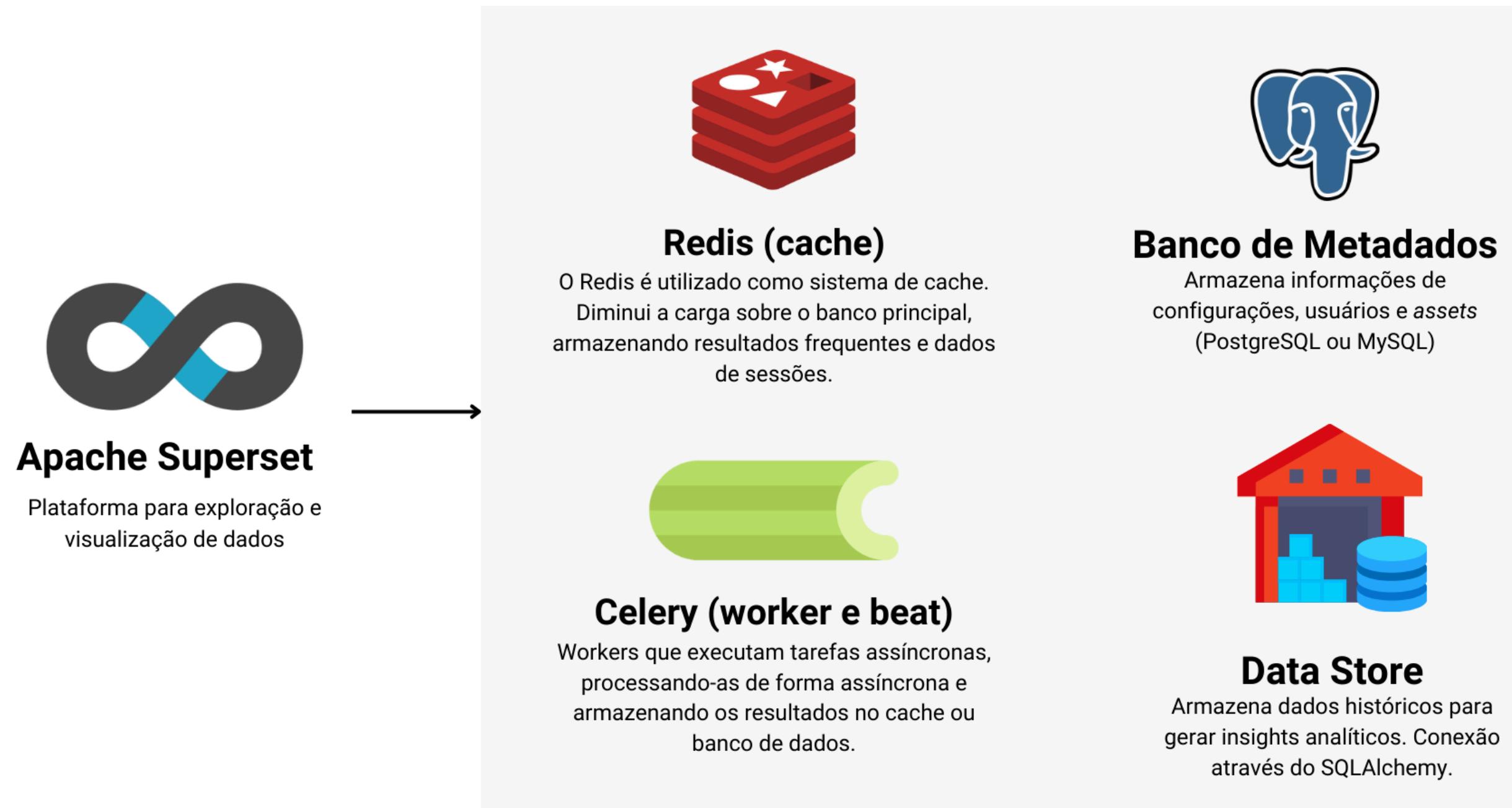
Maxime criou o **Apache Superset** para superar as limitações do **Tableau** no Airbnb, que não suportava **Presto** e **Druid**, além de ser caro e pouco escalável. O objetivo era desenvolver uma solução **open-source**, flexível e acessível para análise e visualização de dados. [🔗](#)



Arquitetura

"Como o Superset é de código aberto e o Airbnb contribuiativamente para o projeto, eles conseguiram turbinar o Superset com componentes internos com relativa facilidade."

Maxime Beauchemin (2021) 



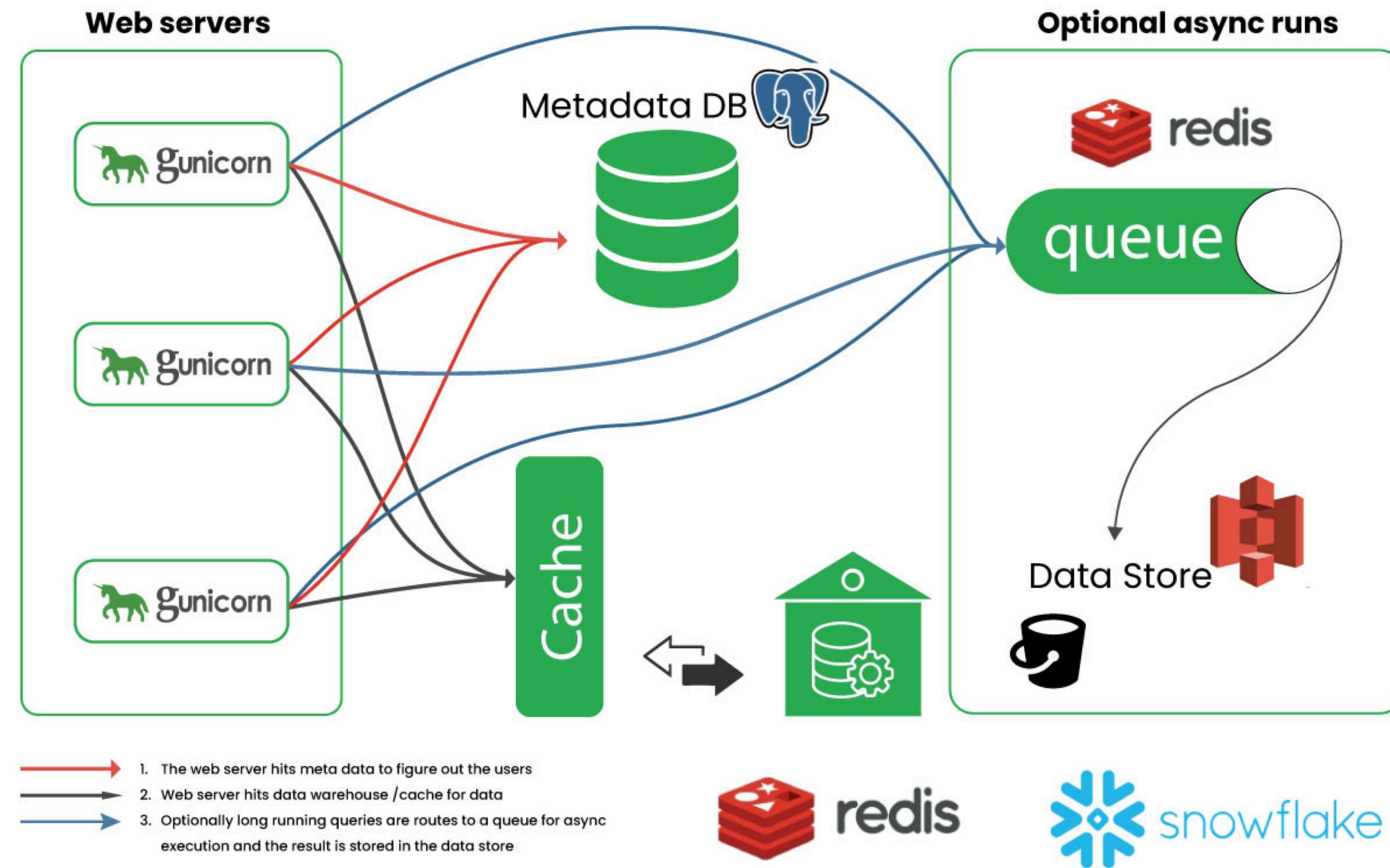
Uma instalação típica do Superset é composta pelos componentes ilustrados.

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**
-  Compatibilidade depende de **driver Python DB-API**
-  Utiliza **SQLAlchemy** para abstração e integração





Autora: Abhilasha Sinha (2023) - WalkingTree

Vantagens

- **Escalabilidade horizontal:** permite aumentar a capacidade facilmente adicionando servidores Gunicorn e Workers.
- **Desempenho:** o uso de cache Redis reduz significativamente o tempo de resposta das consultas mais comuns.
- **Resiliência:** tarefas longas não sobrecarregam servidores web, pois são processadas por workers independentes.
- **Integração flexível:** conecta-se facilmente a múltiplas fontes de dados usando SQLAlchemy.

Essa arquitetura torna o Superset uma opção interessante para ambientes empresariais, especialmente onde **performance e escalabilidade** são necessárias.

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) 

Controle de Acesso - RBAC

Controle de Acesso Baseado em Funções (RBAC)

Mecanismo de segurança que restringe o acesso a recursos com base nas funções dos usuários.

- ▶ Gerencia permissões de usuários com funções personalizáveis.
- ▶ Permissões específicas (ex.: visualizar, executar query).
- ▶ Usuários são atribuídos a funções para controlar níveis de acesso.

Funções Predefinidas:

Admin (acesso total), Alpha (uso avançado), Gamma (acesso básico).

Single Sign-On (SSO)

Processo de autenticação que permite acesso a múltiplas aplicações com um único login.

- ▶ Integração simplificada com provedores OAuth (Google, Okta, Azure, etc.)
- ▶ Autenticação única para maior conveniência dos usuários

Benefícios:

Redução dos esforços administrativos, simplifica o login e aumenta a segurança corporativa.

Segurança

- ▶ Criptografia avançada para dados sensíveis
- ▶ Auditoria detalhada e logs de acessos
- ▶ Suporte para mascaramento de dados
- ▶ Criptografia em trânsito (HTTPS) e em repouso para dados sensíveis.
- ▶ Suporta múltiplos métodos (banco de dados, LDAP, OAuth).
- ▶ RBAC garante controle de acesso granular.
- ▶ Configurável para regulamentações - política de acesso de dados

Alertas e Relatórios

- ▶ Criação automatizada de relatórios periódicos
- ▶ Notificações via email ou plataformas de mensagem (Slack, Teams)
- ▶ Configuração flexível de gatilhos baseados em consultas SQL

Caching

Armazena resultados de consultas para acelerar o carregamento de dashboards.

- ▶ Armazena em cache resultados de consultas, metadados e dados de visualização.
- ▶ Tempos de expiração e intervalos de atualização configuráveis para balancear desempenho e atualidade.

Benefício:

Melhora a velocidade para dashboards complexos.

Consultas Assíncronas

Executa consultas longas em segundo plano sem bloquear a interface do usuário.

- ▶ Usa Celery, uma fila de tarefas distribuída, para processamento assíncrono.
- ▶ Suporte a execução de consultas pesadas em segundo plano
- ▶ Gerenciamento eficiente de recursos computacionais
- ▶ Maior estabilidade e performance do sistema
- ▶ Configurável via configurações de backend do Superset.

Benefício:

Melhora a experiência do usuário para grandes conjuntos de dados.

Templates SQL



Consultas SQL dinâmicas usando modelagem jinja!

- ▶ Parametrize consultas com variáveis (ex.: intervalos de datas, filtros).

Exemplo: **SELECT * FROM vendas WHERE data >= {{ data_inicio }}**

- ▶ Reutilização de consultas SQL através de modelos parametrizados
- ▶ Facilidade para criar consultas dinâmicas e flexíveis
- ▶ Maior produtividade na criação e manutenção de dashboards

Caso de Uso Comum:

Iteração / Análises complexas

Rede e Segurança

- ▶ Compatibilidade com VPN e ambientes seguros
- ▶ Suporte a HTTPS e SSL para comunicação segura
- ▶ Facilidade para implantação em redes corporativas restritas

Propósito:

Garante implantação segura em ambientes restritos.

Importar/Exportar

Importe e exporte configurações do superset e de *assets*.

- ▶ Facilidade na migração e compartilhamento de dashboards e datasets.
- ▶ Compatibilidade com diversos formatos populares (CSV, Excel, JSON, YALM).
- ▶ Gerenciamento simplificado e ágil do ciclo de vida dos dados.
- ▶ Importe de uma instância para outra!
- ▶ Versionamento em código.

Caso de Uso:

Backup, migração, compartilhamento de conexões de banco de dados e versionamento.

Outros

- ▶ Maptools
- ▶ Handlebars
- ▶ Embed
- ▶ ...

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

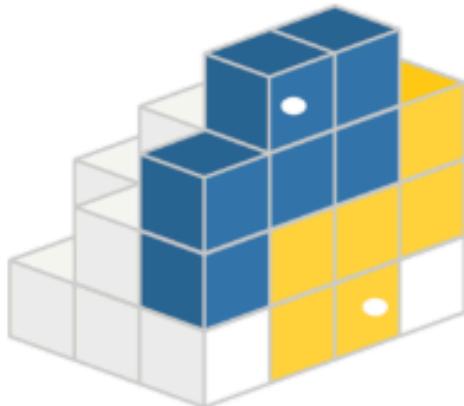
A comunidade Apache Superset usa extensivamente o **Docker** para desenvolvimento, lançamento e produção do Superset.

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
39 >Run Service
40 db:
41   env_file:
42     - path: docker/.env # default
43       required: true
44     - path: docker/.env-local # optional override
45       required: false
46   image: postgres:16
47   container_name: superset_db
48   restart: unless-stopped
49   volumes:
50     - db_home:/var/lib/postgresql/data
51     - ./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) 

Obrigada!

Marília Melo Favalesso - PhD, Cientista de Dados

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