

intro to Apache Superset



Evan Rusackas &
Maxime Beauchemin
10/21/2024

Well, hello there!



Evan Rusackas

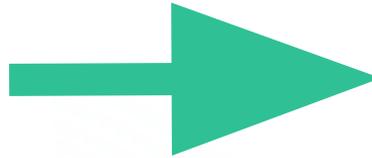
Community @ Preset
Apache Superset PMC Member
Github: @rusackas



Maxime Beauchemin

CEO @ Preset
Apache Superset VP
Github: @mistercrunch

Global Switch Day



What we'll cover

What is Apache Superset?

A Peek Under The Hood

Customizing Superset

How to Try It

What *is* Apache Superset?

Well, it's...

- Open-Source
- Enterprise-Ready
- Business Intelligence
- *(with all the bells and whistles)*



64,054



1,523



14,337

What *IS* the ASF?

The Apache Software Foundation (ASF) is a non-profit organization that supports the development of open-source software projects.

Apache provides:

- **Governance**, a.k.a. "The Apache Way"
- **Oversight** via security/infrastructure teams, and the Apache board
- **Legal support**: IP protection, software licensing support, and more
- **Infrastructure**: tooling, repositories, mailing lists, etc.



Origin Story



Maxime Beauchemin

Original creator of Apache Superset & Airflow



Built out of **Airbnb**, Superset is the most popular open source business intelligence platform with a large and growing community of data professionals around the world



Broad ecosystem of supported databases



Used by everyone*



TESLA



INTERCOM



CAPITAL SERVICE S.A.



amino

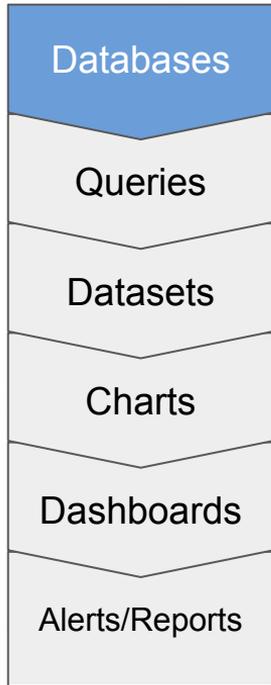


Shopee



+ tens of thousands of organization!

Basic Components: Database Connections



Connect a database [Close]

STEP 1 OF 3

Select a database to connect

Google BigQuery

PostgreSQL

Snowflake

MySQL

Databricks

Google Sheets

Or choose from a list of other databases we support:

SUPPORTED DATABASES

Choose a database... [Search]

- Airtable
- Amazon Athena
- Amazon Redshift
- Apache Doris
- Apache Drill
- Apache Druid
- Apache Hive

Edit database [Close]

BASIC | ADVANCED

DISPLAY NAME *

Superset Database

Pick a name to help you identify this database.

SQLALCHEMY URI *

postgresql://examples:XXXXXXXXX@db:5432/examples

Refer to the [SQLAlchemy docs](#) for more information on how to structure your URI.

TEST CONNECTION

CLOSE | FINISH

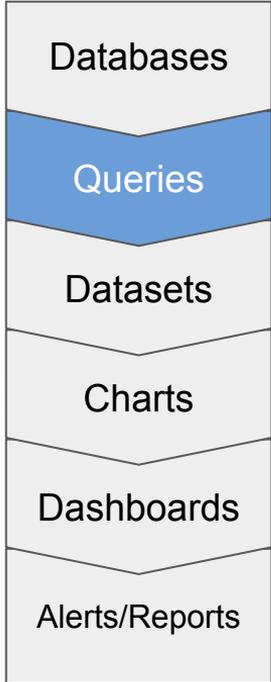
Databases

Expose in SQL Lab | AQE

Type a value | Select or type a value | Select or type a value

- man's Repos on Peaka
- Peaka Hubspot CRM
- Examples (powered by MotherDuck/DuckDB)
- Competition
- AIN - BigQuery Superset READONLY
- MySQL test
- Smarie Playground 2 [test label change]
- etadb test
- Factor DB
- Google Sheets - CS
- Google Sheet Examples
- e Uploads
- semarie-playground don't use
- s Metrics Testing
- ies Impacts Dec 22
- gQuery Open Datasets

Basic Components: **Advanced SQL IDE**



Superset Dashboards Charts Datasets SQL

Untitled Query 1

DATABASE: postgresql examples

SCHEMA: public

SEE TABLE SCHEMA: cleaned_sales_data

cleaned_sales_data

order_number	BIGINT
quantity_ordered	BIGINT
price_each	DOUBLE_PRECISION
order_line_number	BIGINT
sales	DOUBLE_PRECISION
order_date	TIMESTAMP
status	TEXT
quarter	BIGINT
month	BIGINT
year	BIGINT
product_line	TEXT
msrp	BIGINT
product_code	TEXT
customer_name	TEXT
phone	TEXT
address_line1	TEXT
address_line2	TEXT
city	TEXT
state	TEXT
postal_code	TEXT
country	TEXT

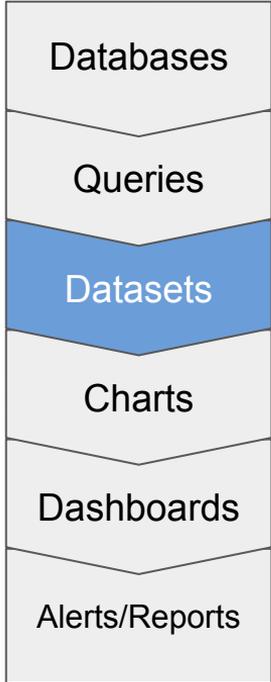
```
1 SELECT
2   "sales",
3   "order_date",
4   "status",
5   "quarter",
6   "month",
7   "year",
8   "product_line"
9 FROM cleaned_sales_data
10 LIMIT 10000
```

RUN LIMIT: 1000 00:00:00.70 SAVE COPY LINK

RESULTS QUERY HISTORY PREVIEW: 'CLEANED_SALES_DATA'

State	Started	Duration	Progress	Rows	SQL	Results	Actions
✓	17:09:34	00:00:01.52	<div style="width: 100%;"></div>	1000	SELECT "sales", "order_date", "status", "quarter", { ... }	View	✎ +
✓	16:52:44	00:00:00.14	<div style="width: 100%;"></div>	55	SELECT DATE_TRUNC('week', year) AS __timestamp, SUM("NY_GNP_PCAP_CD") AS "(foo)" FROM wb_health_population GROUP BY DATE_TRUNC('week', year) { ... }	View	✎ + 🗑
✗	16:52:29	00:00:00.10	0%		SELECT DATE_TRUNC('week', year) AS __timestamp, SUM("NY_GNP_PCAP_CD") AS "(foo)" FROM wb_health_population GROUP BY DATE_TRUNC('week', year) { ... }		✎ + 🗑
✗	16:53:31	00:00:00.08	0%		wb_health_population		✎ + 🗑

Basic Components: **Datasets**



Datasets

Bulk select + Dataset

Name: Q. Type a value | Type: Select or type a value | Database: Select or type a value | Schema: Select or type a value | Owner: Select or type a value | Certified: Select or type a value | Modified by: Select or type a value

Name	Type	Database	Schema	Owners	Last modified	Actions
marketing_blog	Physical	MAIN - BigQuery Superset READONLY	core		8 days ago	
manager_team_history	Physical	MAIN - BigQuery Superset READONLY	core_history		8 days ago	
Query core_recurly.account_subscription_data	Virtual	MAIN - BigQuery				
Peaka Hubspot Deals	Virtual	Peaka Hubspot				
viewers & creators by team 08/28/2024 10:11:16	Virtual	MAIN - BigQuery				
viewers & creators by team 08/27/2024 18:16:51	Virtual	MAIN - BigQuery				
viewers & creators by team	Virtual	MAIN - BigQuery				
Query core_history.hubspot_deal_history	Virtual	MAIN - BigQuery				
hubspot_deal_history	Physical	MAIN - BigQuery				
World Bank Health Data	Physical	examples (to po				
Vehicle Sales	Physical	examples (to po				
Query birth_names 03/28/2024 14:23:40	Virtual	examples (to po				
Query COVID Vaccines 03/01/2022 10:33:14	Virtual	examples (to po				
Query Slack Members and Channels 02/24/2022 18:55:47	Virtual	examples (to po				

Edit Dataset marketing_blog

Be careful. Changing these settings will affect all charts using this dataset, including charts owned by other people.

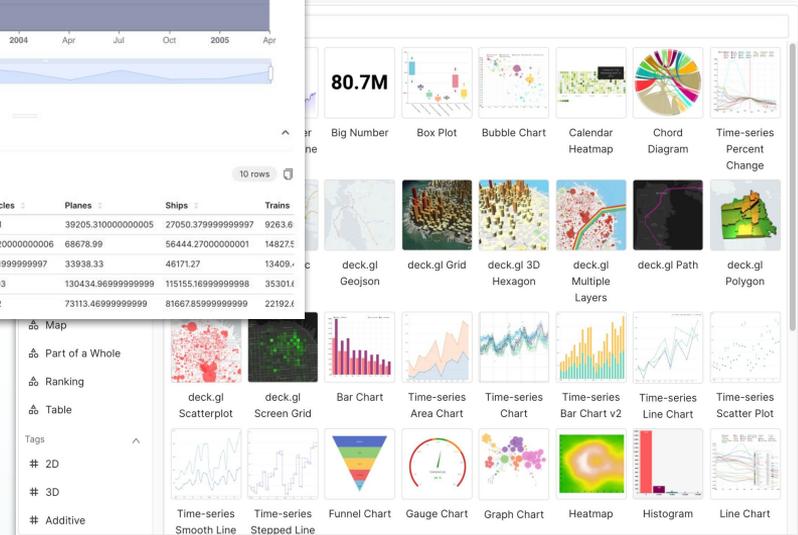
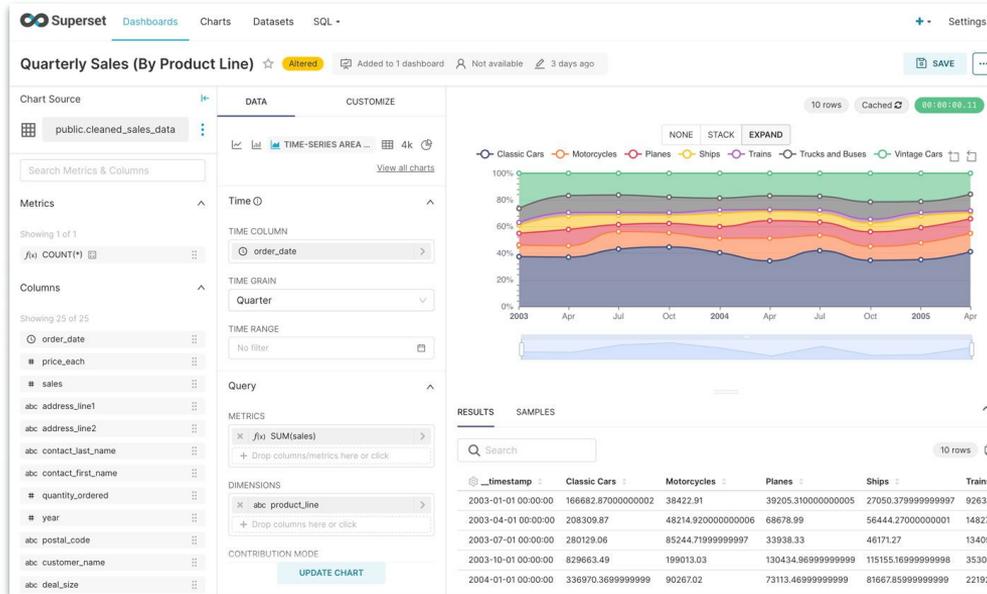
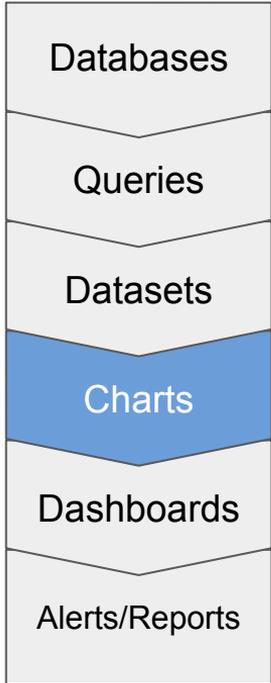
Source Metrics (6) Columns (28) Calculated columns (0) Settings

Sync columns from source

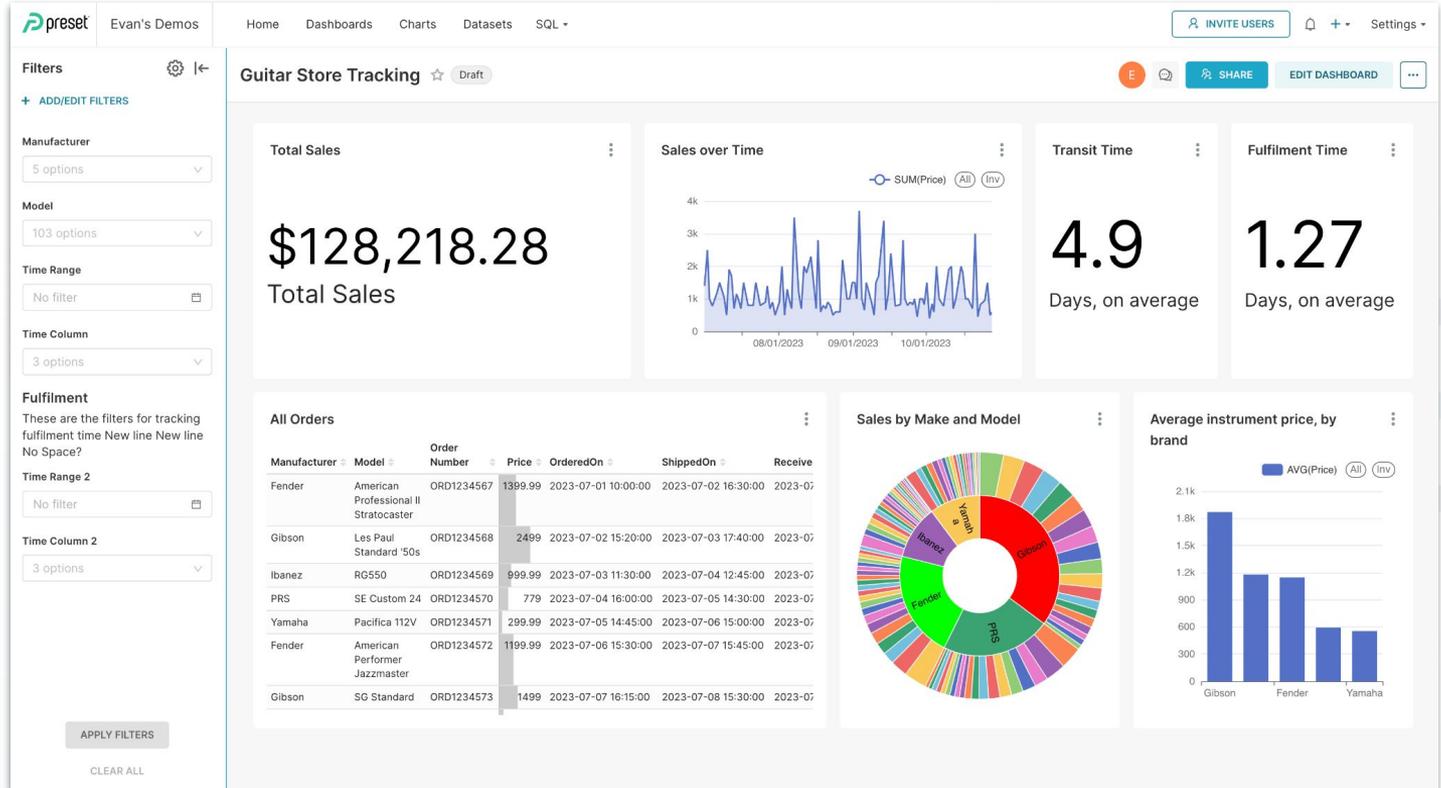
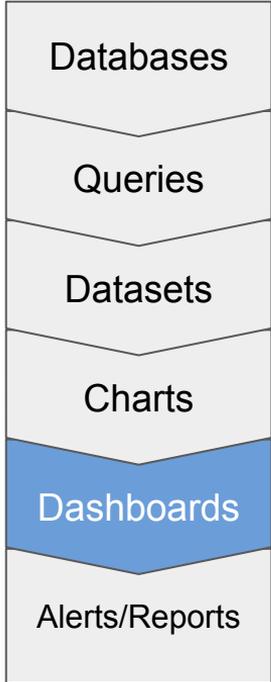
Column	Data type	Is temporal	Default datetime	Is filterable	Is dimension
full_referrer	STRING	<input type="checkbox"/>	<input type="radio"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
session_referrer	STRING	<input type="checkbox"/>	<input type="radio"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
session_referrer_domain	STRING	<input type="checkbox"/>	<input type="radio"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
dt	DATE	<input checked="" type="checkbox"/>	<input checked="" type="radio"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
dtm	TIMESTAMP	<input checked="" type="checkbox"/>	<input type="radio"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
blog_slug	STRING	<input type="checkbox"/>	<input type="radio"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Cancel Save

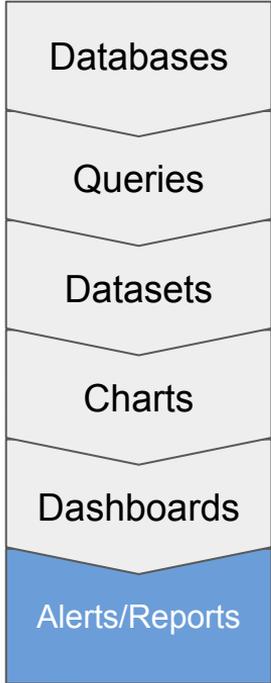
Basic Components: No-code Chart Creation



Basic Components: Interactive Dashboards



Basic Components: Alerts/Reports



Add Alert

General information *
Set up basic details, such as name and description.

Alert condition *
Define the database, SQL query, and triggering conditions for alert.

Alert contents *
Customize data source, filters, and layout.

Schedule ✓
Define delivery schedule, timezone, and frequency settings.

SCHEDULE TYPE *
Recurring (every)

SCHEDULE *
Every week

on MON

at 9 : 0

TIMEZONE *
GMT -06:00 (Mountain Daylight Time)

LOG RETENTION *
90 days

GRACE PERIOD
Time in seconds

CANCEL ADD

Edit Alert

General information ✓
Set up basic details, such as name and description.

Alert condition ✓
Define the database, SQL query, and triggering conditions for alert.

Database *
Preset Cloud Analytics

SQL Query *

```
54 )
55 )
56 )
57 )
58 )
59 )
60 )
61 )
62 )
63 select sum(qtd) as issue_count from (
64   select unknown_org_count as qtd
65   from unknown_orgs
66   UNION ALL
67   select unknown_support_tier_count as qtd
68   from wrong_license_info
69 )
```

Trigger Alert If... * Value *
> (Larger than) 0

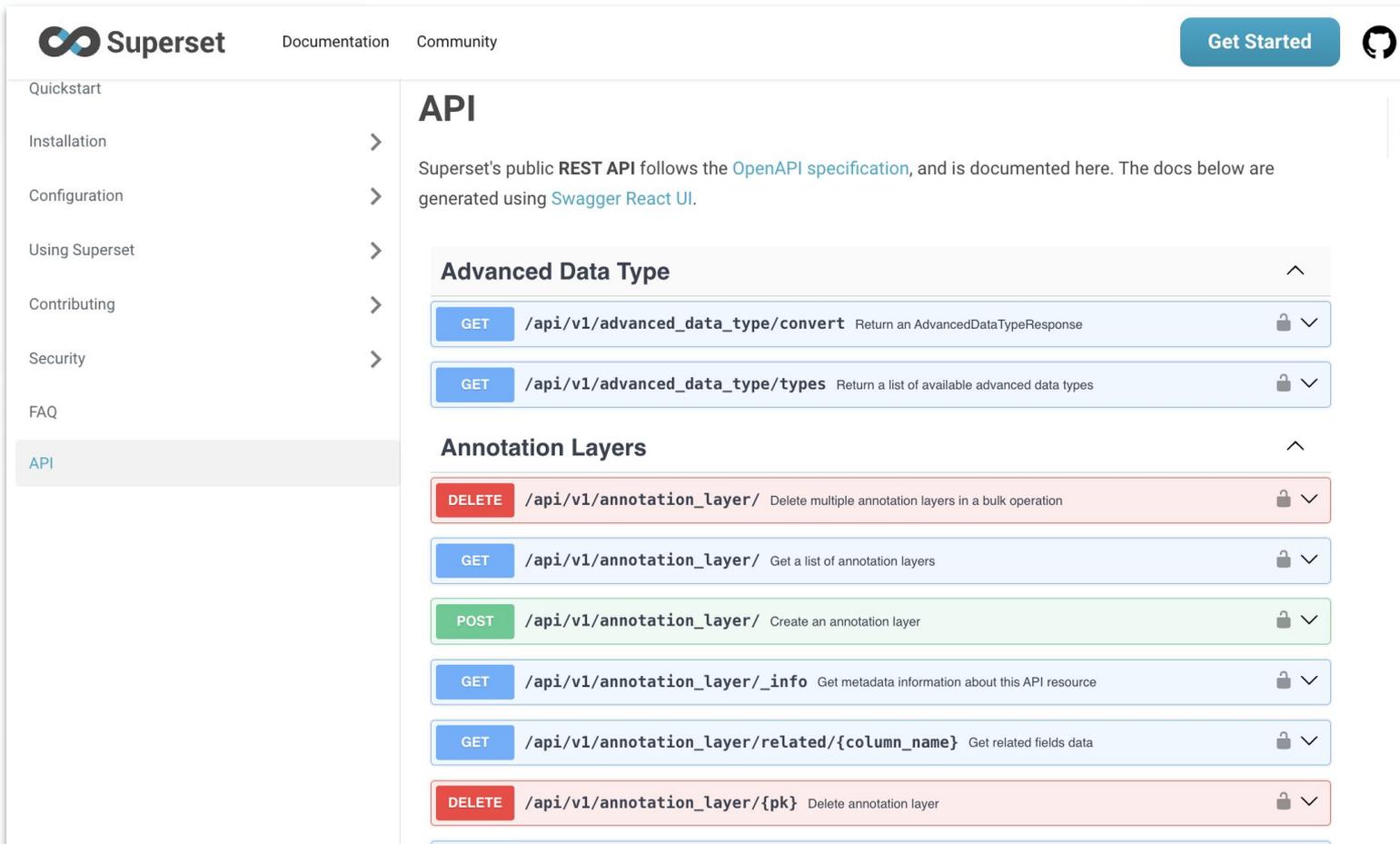
Alert contents ✓
Customize data source, filters, and layout.

Schedule ✓
Define delivery schedule, timezone, and frequency settings.

Notification method ✓
Choose notification method and recipients.

Cancel Save

API/CLI (TODO)



Superset Documentation Community [Get Started](#)

- Quickstart
- Installation >
- Configuration >
- Using Superset >
- Contributing >
- Security >
- FAQ
- API**

API

Superset's public **REST API** follows the [OpenAPI specification](#), and is documented here. The docs below are generated using [Swagger React UI](#).

Advanced Data Type

- GET** `/api/v1/advanced_data_type/convert` Return an AdvancedDataTypeResponse
- GET** `/api/v1/advanced_data_type/types` Return a list of available advanced data types

Annotation Layers

- DELETE** `/api/v1/annotation_layer/` Delete multiple annotation layers in a bulk operation
- GET** `/api/v1/annotation_layer/` Get a list of annotation layers
- POST** `/api/v1/annotation_layer/` Create an annotation layer
- GET** `/api/v1/annotation_layer/_info` Get metadata information about this API resource
- GET** `/api/v1/annotation_layer/related/{column_name}` Get related fields data
- DELETE** `/api/v1/annotation_layer/{pk}` Delete annotation layer

What is Preset?



Open-source BI is here!

-  Competitive, modern BI
-  No vendor lock-in
-  Massive open-source community
-  Fully customizable & integrable

60k+

Github 

1k+

Contributors

150+

Monthly
downloads



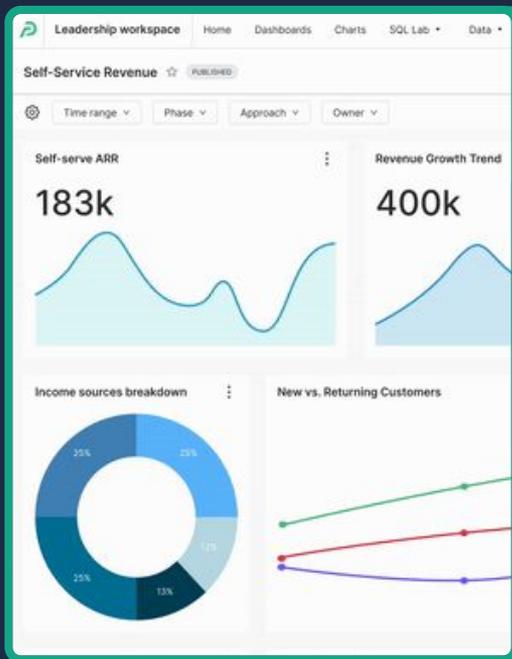
Open-source BI, turbocharged!

-  Open-source BI, turbocharged
-  Hosted, hassle-free Superset
-  AI Assist, collaboration & more
-  Enterprise-ready: SSO, compliance, support
-  Embedded & OEM support
-  Automation: assets as code, APIs



Preset: **Open, cutting-edge BI for dynamic teams**

Preset supercharges open-source Apache Superset™ with a scalable, enterprise-grade hosted solution, extending it with key features and offering a broad platform that allows you to deliver modern Business Intelligence, embedded analytics, and tailored data experiences.



Preset-Only Enhancements

Workspaces | Embedded | Integrations | APIs | Headless BI | [Collaboration & Communication](#) | [AI features](#) | UX/UI Improvements

Governance & Security

RBAC | SSO | SOC 2 | RLS

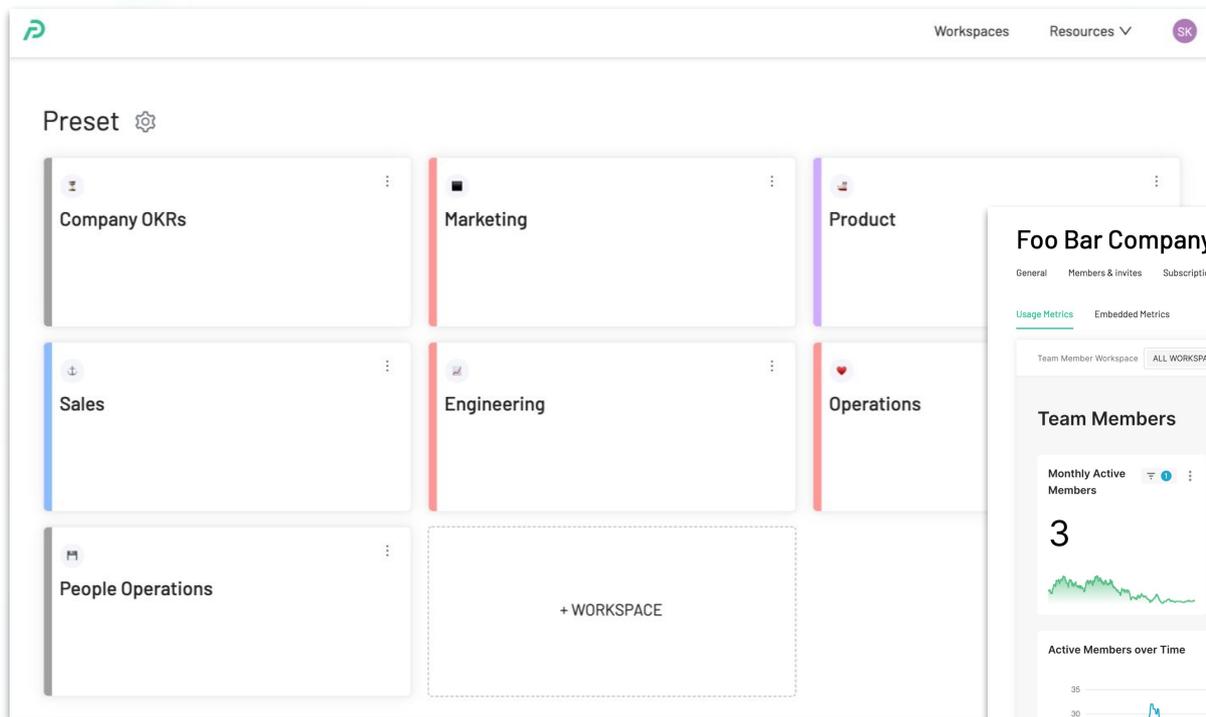
Cloud & Performance

Auto Deployment | Auto Scale | Auto Upgrades | Caching | Private Cloud

Apache Superset Core

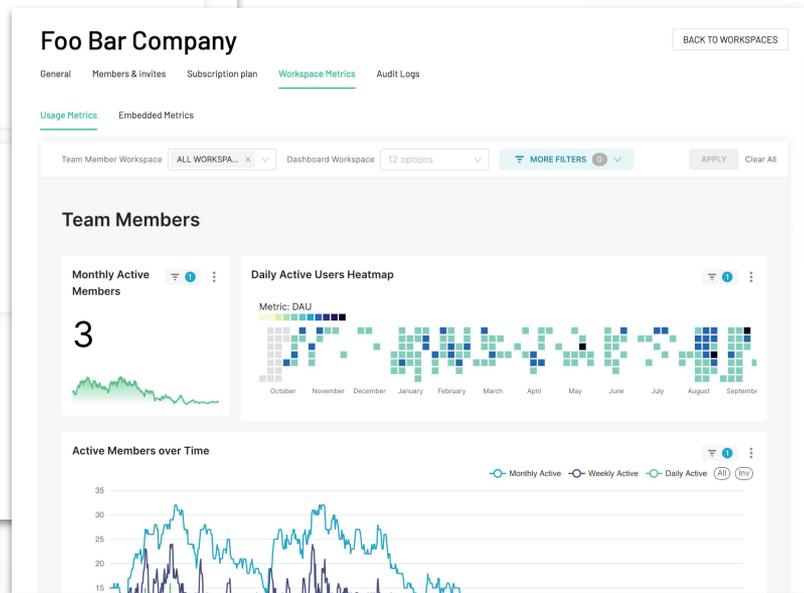
Charts & Dashboards | Visualizations | Semantic Layer | SQL IDE | Drag & Drop Exploration

Preset Manager: Why settle for *just one* Superset instance?



The screenshot shows the 'Preset Manager' interface in Superset. At the top, there are tabs for 'Workspaces' and 'Resources' with a user profile icon 'SK'. Below the 'Preset' header, a grid of workspace presets is displayed:

- Company OKRs
- Marketing
- Product
- Sales
- Engineering
- Operations
- People Operations
- + WORKSPACE

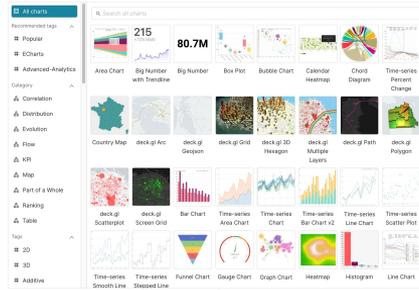


The screenshot shows a dashboard for 'Foo Bar Company' with the following components:

- Navigation: BACK TO WORKSPACES, General, Members & invites, Subscription plan, **Workspace Metrics**, Audit Logs
- Sub-navigation: Usage Metrics, Embedded Metrics
- Filters: Team Member Workspace, ALL WORKSPA... x, Dashboard Workspace, 12 options, MORE FILTERS 0, APPLY, Clear All
- Section: Team Members
- Card: Monthly Active Members (3)
- Card: Daily Active Users Heatmap (Metric: DAU) with a heatmap and line chart showing activity from October to September.
- Card: Active Members over Time with a line chart showing Monthly Active, Weekly Active, and Daily Active members over time.

A Peek Under the Hood

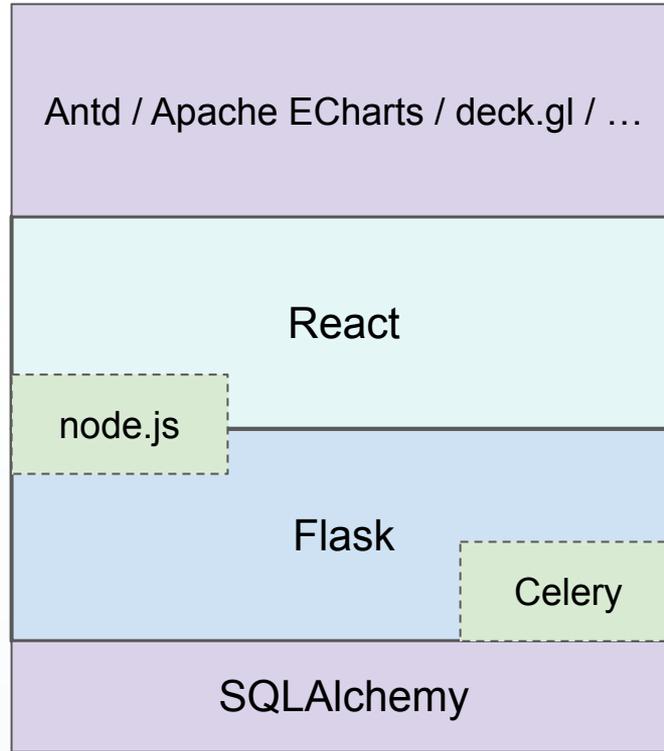
Architecture



SQL



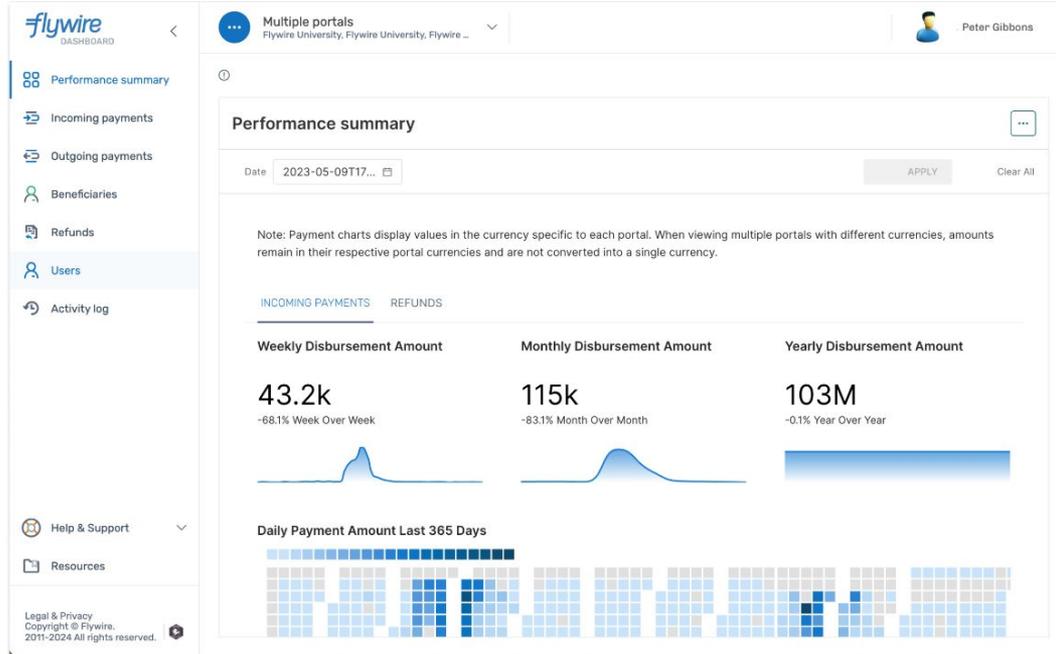
The stack



- Async queries
- Alerts & reports
- Cache warmup
- Thumbnails

Customizing Superset

Embedding / Co-branding



Custom CSS Templating

Live CSS editor

```
1 .navbar {  
2   transition: opacity 0.5s ease;  
3   opacity: 0.05;  
4 }  
5 .navbar:hover {  
6   opacity: 1;  
7 }  
8 .chart-header .header{  
9   font-weight: @font-weight-normal;  
10  font-size: 12px;  
11 }  
12 /*
```

LOAD A CSS TEMPLATE

Flat

Courier Black

The dashboard is titled "Sample dashboard - Piermarco" and includes a "SHARE" and "EDIT" button. It features a sidebar with "Filters" and "APPLY FILTERS" / "CLEAR ALL" buttons. A modal window titled "EDIT RANGE TIME" is open, showing options for "Range type" (Last) and "Configure Time Range" (Last year). The main content area displays two summary cards: "30 Registrations, last 30 days" and "5.5k Registrations, last 90 days". Below these is a line chart titled "Active user trends" showing data from Jan to Dec.

Month	Active Users
Jan	500
Feb	1000
Mar	500
Apr	500
May	500
Jun	500
Jul	500
Aug	1500
Sep	1200
Oct	2000
Nov	1000
Dec	500

WIP: Comprehensive Theming



Ant Design

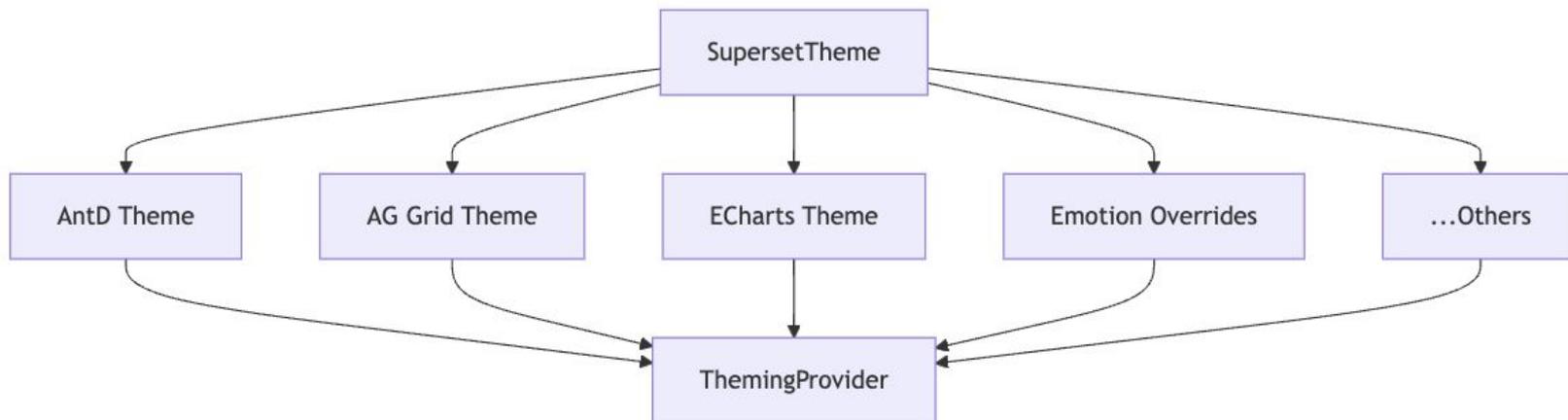


ECHARTS

AG Grid



Emotion



Handlebars Chart

HandlebarTutorial

SAVE ...

Chart Source

public.Vehicle Sales

Search Metrics & Columns

Metrics

Showing 1 of 1

f(x) COUNT(*)

Columns

Showing 28 of 28

order_date

price_each

sales

address_line1

deal_size

product_code

product_line

state

status

city

country

DATA

CUSTOMIZE

```
4 <th>Product Type</th>
5 <th>Total Sales</th>
6 </tr>
7 </thead>
8 <tbody>
9   {{#each data}}
10  <tr>
11    <td>{{this.Product_Type}}</td>
12    <td>{{this.Total_Sales}}</td>
13  </tr>
14  {{/each}}
15 </tbody>
16 </table>
```

CSS Styles

```
1 .data-table {
2   border-collapse: collapse;
3   width: 100%;
4 }
5
6 .data-table th, .data-table td {
7   border: 1px solid black;
8   padding: 8px;
9   text-align: left;
10 }
11
12 .data-table th {
13   background-color: #f2f2f2;
14 }
```

UPDATE CHART

7 rows 00:00:01.38

Product Type	Total Sales
Classic Cars	3919615.6599999997
Vintage Cars	1903150.8399999992
Motorcycles	1166388.3400000003
Trucks and Buses	1127789.8399999996
Planes	975003.5700000001
Ships	71443713
Trains	226243.46999999997

RESULTS SAMPLES

Search

7 rows

Product_Type	Total_Sales
Classic Cars	3919615.6599999997
Vintage Cars	1903150.8399999992
Motorcycles	1166388.3400000003
Trucks and Buses	1127789.8399999996

BYO Viz (new SDK coming soon!)

The screenshot displays a data visualization tool interface. On the left is a configuration sidebar with sections: DATA, CUSTOMIZE, Chart type, VISUALIZATION TYPE (Liquid), Time, TIME COLUMN (ds), TIME RANGE (No filter), Query, METRIC (f(x) ratio), and FILTERS. The main area shows a liquid gauge chart with a blue fill and a wavy top edge, displaying the value 59.66%. The top of the main area includes a title '- untitled', '1 rows', a green timer '00:00:00:17', and various utility icons. At the bottom of the main area is a '> Data' link.

⚡ RUN 💾 SAVE

DATA CUSTOMIZE

Chart type ^

VISUALIZATION TYPE
Liquid

Time ⓘ ^

TIME COLUMN
ds >

TIME RANGE
No filter

Query ^

METRIC
f(x) ratio >

FILTERS
+ Drop columns/metrics here or click

- untitled 1 rows 00:00:00:17 📅 🔗 ✉ <> 📄 .JSON 📄 .CSV ☰

59.66%

> Data

Superset Improvement Proposals (SIPs)

- When a change is potentially controversial, or large in scope, we open a SIP
- Discussion on official and unofficial channels
- Consensus
- Tracking/execution

The screenshot displays a board of Superset Improvement Proposals (SIPs) organized into three columns: VOTING (in the mailing list) with 5 items, APPROVED (in mailing list) with 24 items, and IMPLEMENTED / DONE with 32 items. Each item includes a Superset issue number, a title, and status tags.

Column	Item	Title	Status
VOTING (in the mailing list) 5	superset #21385	[SIP-93] Proposal for non-blocking SQL Lab persistence	sip
	superset #11167	[SIP-54] Proposal to improve export/import of objects	enhancement:request sip
	superset #6814	[SIP-16] Proposal for loading additional polygon encodings	sip
	superset #15362	[SIP-70] Secret keys rotations and stop using single SECRET_KEY	inactive sip
APPROVED (in mailing list) 24	superset #21789	[SIP-91] - Enable SSH Tunneling on Database Connections	2022.7 sip
	superset #17187	[SIP-75] Proposal for Embeddable Dashboards	preset-io sip
	superset #10408	[SIP-51] Dashboard Level Access	enhancement:request sip
	superset #10254	[SIP-48] Using Ant Design as our primary component library	!deprecated-label:org:preset enhancement:request sip
IMPLEMENTED / DONE 32	superset #8318	[SIP-24] Proposal to introduce Flask app factory pattern	!deprecated-label:org:preset enhancement:request inactive sip
	superset #5565	[SIP-2] Deprecate support for python <3.6	sip
	superset #17852	[SIP-78] Proposal for Advanced Types	sip
	superset #5528	[SIP-1] Split the Table visualization into two visualizations	sip

How we adapt and work: the “Superset Operational Model”

- *Everyone* has a part in Superset’s software lifecycle
- Anyone can join Town Hall, make proposals, and discuss issues
- Working groups delegate or volunteer for projects
- Project boards, and additional syncs to track discussions
- Results/updates/issues cycle back to Slack and Town Hall



Why Believe In Open-Source BI?

The Future Of Business Intelligence Is Open Source

Maxime Beauchemin March 05, 2021



Subscribe

While *“software is [still actively] eating the world”*, it’s also clear that **open source is taking over software**.

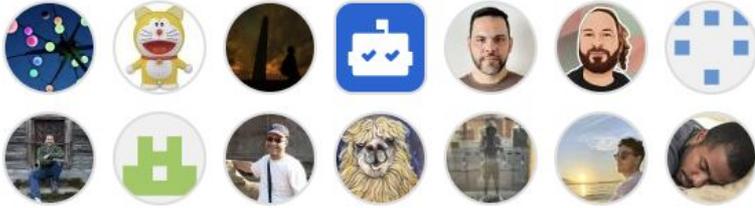
Simply put, **open source is a superior approach at building and distributing software** because it provides important guarantees around how software can be discovered, tried, operated, collaborated on and packaged. For those reasons, it is not surprising that it has taken over most of the modern data stack: infrastructure, data warehouses, databases, orchestration, data processing, AI/ML and beyond.

Adoption



Community / Velocity

Contributors 1,182



[+ 1,168 contributors](#)

Last 28 days Stats

NEW Compare with the previous period

☆ Stars

545



🔗 Pull Requests

Opened Merged

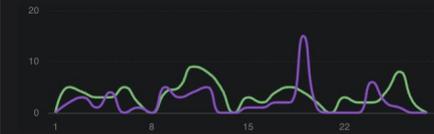
176 132



🔍 Issues

Opened Closed

94 59

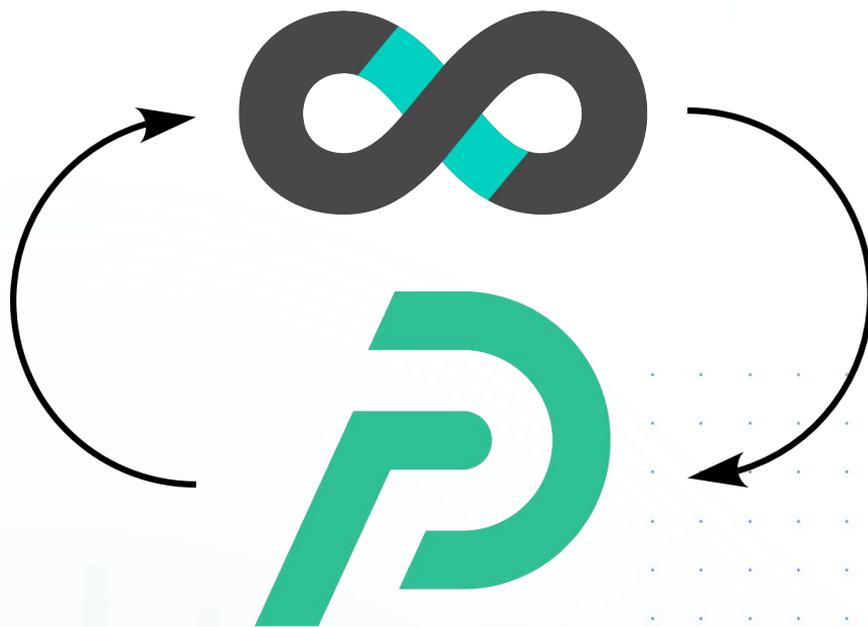


📄 Commits

827



No Vendor Lock-In



AI Bots Welcome



Slack Bots



Issue Bots



SupersetBot™

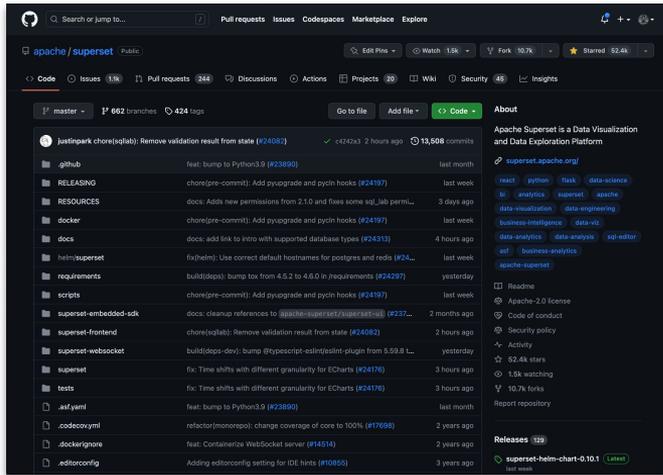


PR Review Bots

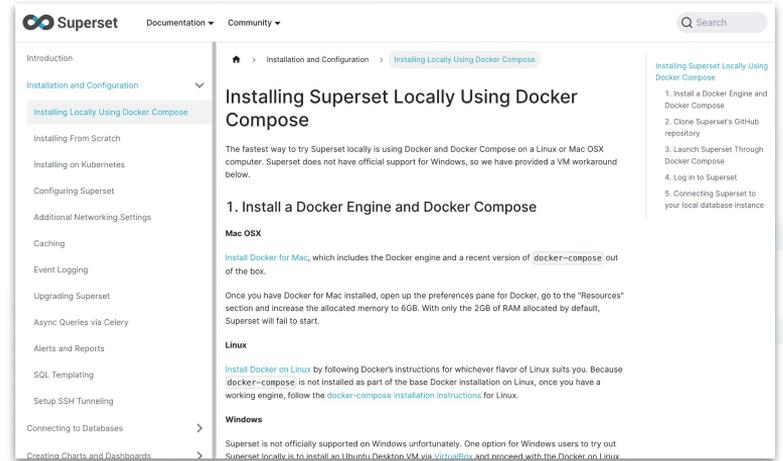


How to Try It

Try out open-source Superset!



Check out the repo!
github.com/apache/superset



Installation/Docs
superset.apache.org

Try out Preset!

Starter

Hassle free Superset in the cloud, best for small teams.

Free

Forever, for up to 5 users.

Get started

Features

- ✓ Unlimited dashboards and charts
- ✓ No-code chart builder
- ✓ Collaborative SQL editor
- ✓ Over 40 visualization types
- ✓ Chart and dashboard cache
- ✓ Semantic layer
- ✓ Up to 5 users
- ✓ 1 workspace
- ✓ In depth docs and tutorials

Professional

Collaborate securely across your entire organization. Try it free for 14 days.

\$20/month

Per user, billed annually.
\$25 per user, per month billed monthly.

TRY FOR FREE

Everything in Starter, plus

- ✓ 2 additional workspaces
- ✓ Role-based access controls (RBAC)
- ✓ Scheduled Slack or email reports
- ✓ Scheduled Slack or email alerts
- ✓ Multi-region support
- ✓ Standard customer support

Add-on:

- ✓ Embedded dashboards*
- ✓ Embedded dashboard viewer licenses (\$500/mo for 50 Viewers**)

Enterprise

Enterprise deployment at scale, best for large organizations.

Custom

Get a customized price estimate for your team.

Contact us

Everything in Professional, plus

- ✓ Additional workspaces
- ✓ API access
- ✓ dbt integration
- ✓ Managed Private Cloud deployment
- ✓ Database connection via SSH Tunnels
- ✓ Single sign-on (SSO)
- ✓ SCIM integration
- ✓ Audit logs & usage metrics
- ✓ Enterprise customer support
- ✓ Dedicated support via Slack
- ✓ Enterprise SLA

Ask us on Slack

<https://bit.ly/join-superset-slack>

Membership

18,537

Total ↑ 274 (+2%) ▾

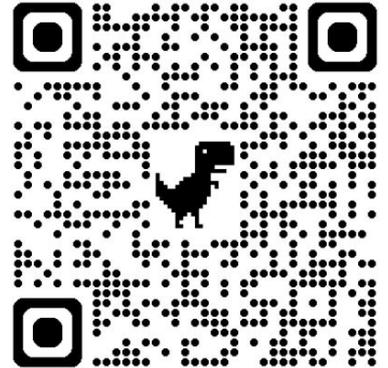


Questions/Demos

Join us at the OSA Com table

Tomorrow/Sun from 12:30-14:30

OSA Com →





That's all Folks!