

A decorative graphic spanning the width of the slide, positioned between the logo and the title. It consists of a teal-colored area chart at the bottom, a light blue line chart with circular markers in the middle, and a lime green line chart with circular markers at the top. The lines are jagged, representing data trends over time.

# The **history** and **anatomy** of Apache **Superset**

# Maxime Beauchemin

- **Open source leader & community builder**
  - Creator of Apache **Superset**
  - Creator of Apache **Airflow**
- Digital artist
- Influencer in the **data engineering** space
- 15+ years of experience in **data & analytics**
- Entrepreneur



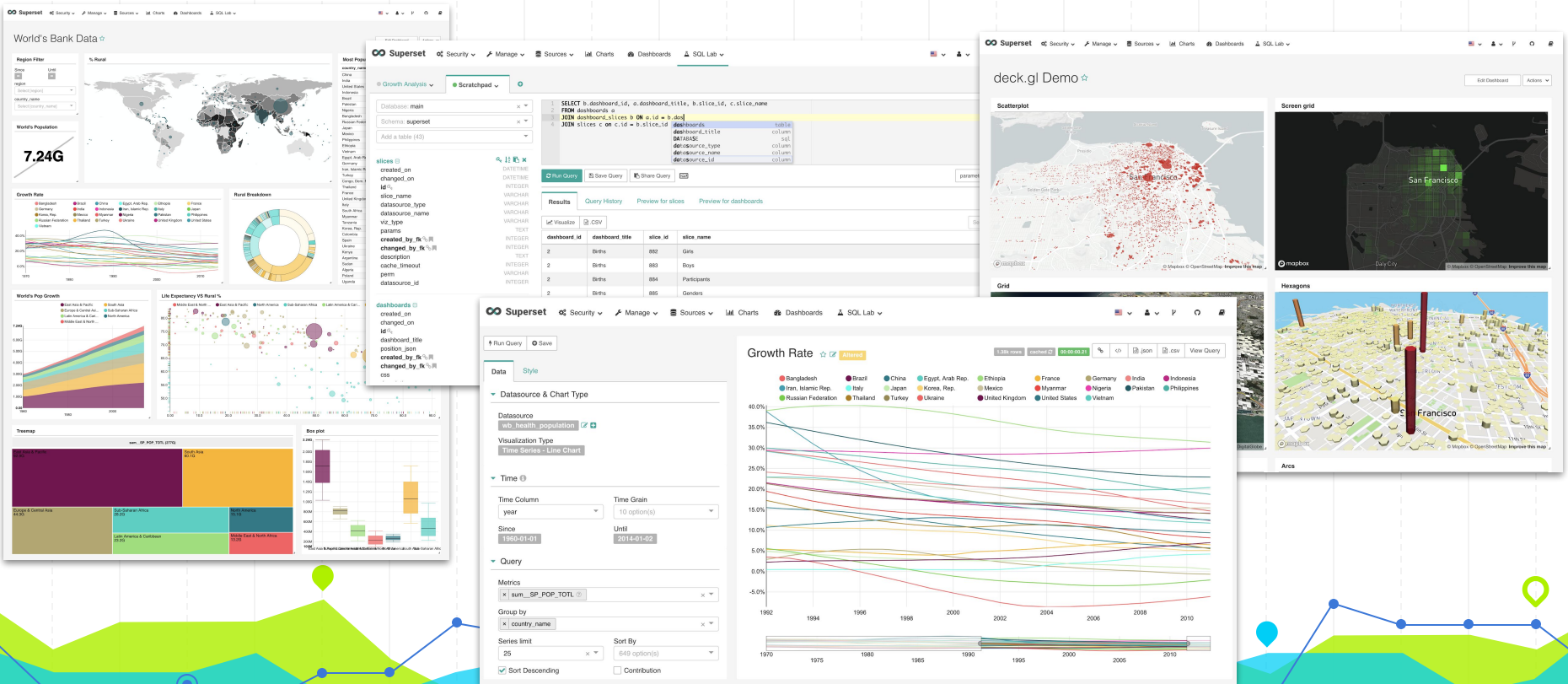
1,202 contributions in the last year

Contribution settings ▾



# Apache Superset

## A data visualization and exploration platform



## World's Bank Data

## Region Filter

Since

region

Select [region]

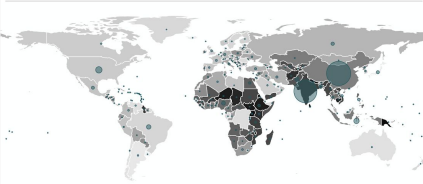
country\_name

Select [country\_name]

## World's Population

7.24G

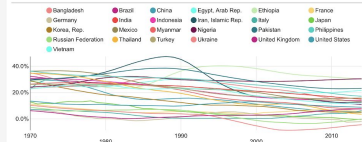
## % Rural



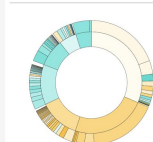
## Most Populated Countries

country_name	sum_SP_POP_TOTL
China	1,360,000,000
India	1,200,000,000
United States	310,000,000
Indonesia	254,000,000
Brazil	209,000,000
France	199,000,000
Nigeria	177,000,000
Bangladesh	159,000,000
Russian Federation	144,000,000
Japan	127,000,000
Mexico	125,000,000
Philippines	96,100,000
Ethiopia	97,000,000
Vietnam	80,700,000
Egypt, Arab Rep.	89,600,000
Germany	82,300,000
Iran, Islamic Rep.	76,100,000
Turkey	73,900,000
Comp. Dem. Rep.	74,900,000
Thailand	67,700,000
France	66,200,000
United Kingdom	64,500,000
Italy	61,300,000
South Africa	54,000,000
Myanmar	53,400,000
Tanzania	51,800,000
Korea, Rep.	50,400,000
Colombia	47,300,000
Spain	46,400,000
Ukraine	46,400,000
Kenya	44,500,000
Argentina	43,000,000
Sudan	39,400,000
Algeria	38,900,000
Poland	38,000,000
Uganda	27,800,000

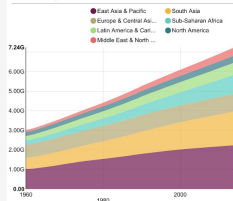
## Growth Rate



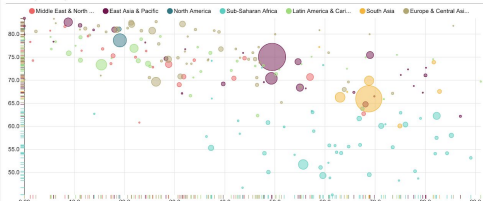
## Rural Breakdown



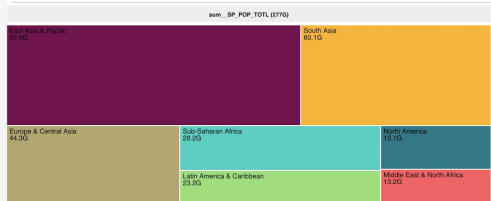
## World's Pop Growth



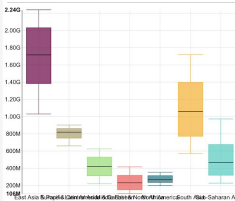
## Life Expectancy VS Rural %



## Treemap

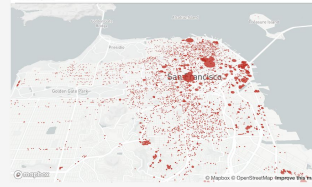


## Box plot

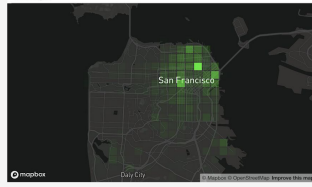


## deck.gl Demo

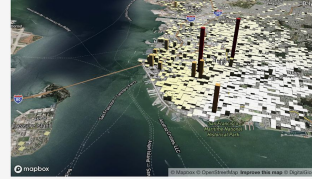
## Scatterplot



## Screen grid



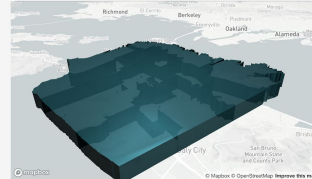
## Grid



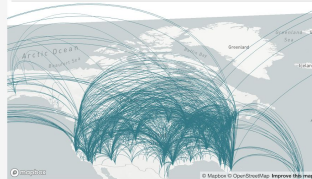
## Hexagons



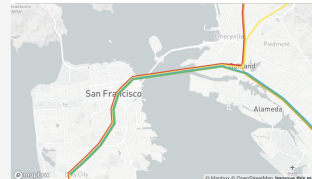
## Polygons



## Arcs



## Path



Run Query Save

Data Style

### Datasource & Chart Type

Datasource

wb\_health\_population

Visualization Type

Time Series - Line Chart

### Time

Time Column

year

Time Grain

10 option(s)

Since

1960-01-01

Until

2014-01-02

### Query

Metrics

x sum\_SP\_POP\_TOTL

Group by

x country\_name

Series limit

25

Sort By

649 option(s)

☒ Sort Descending

☐ Contribution

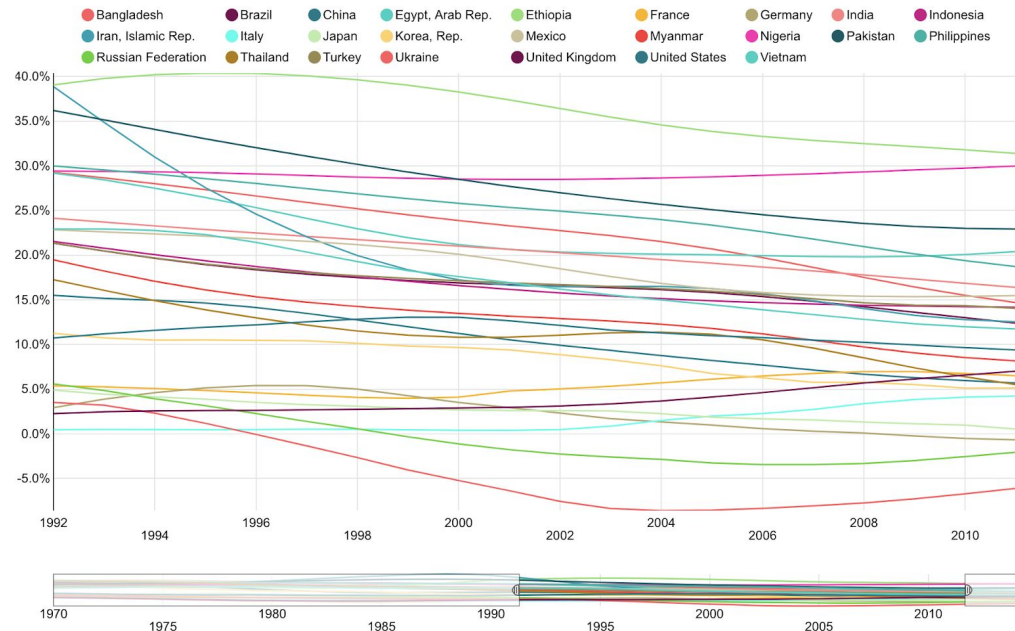
## Growth Rate

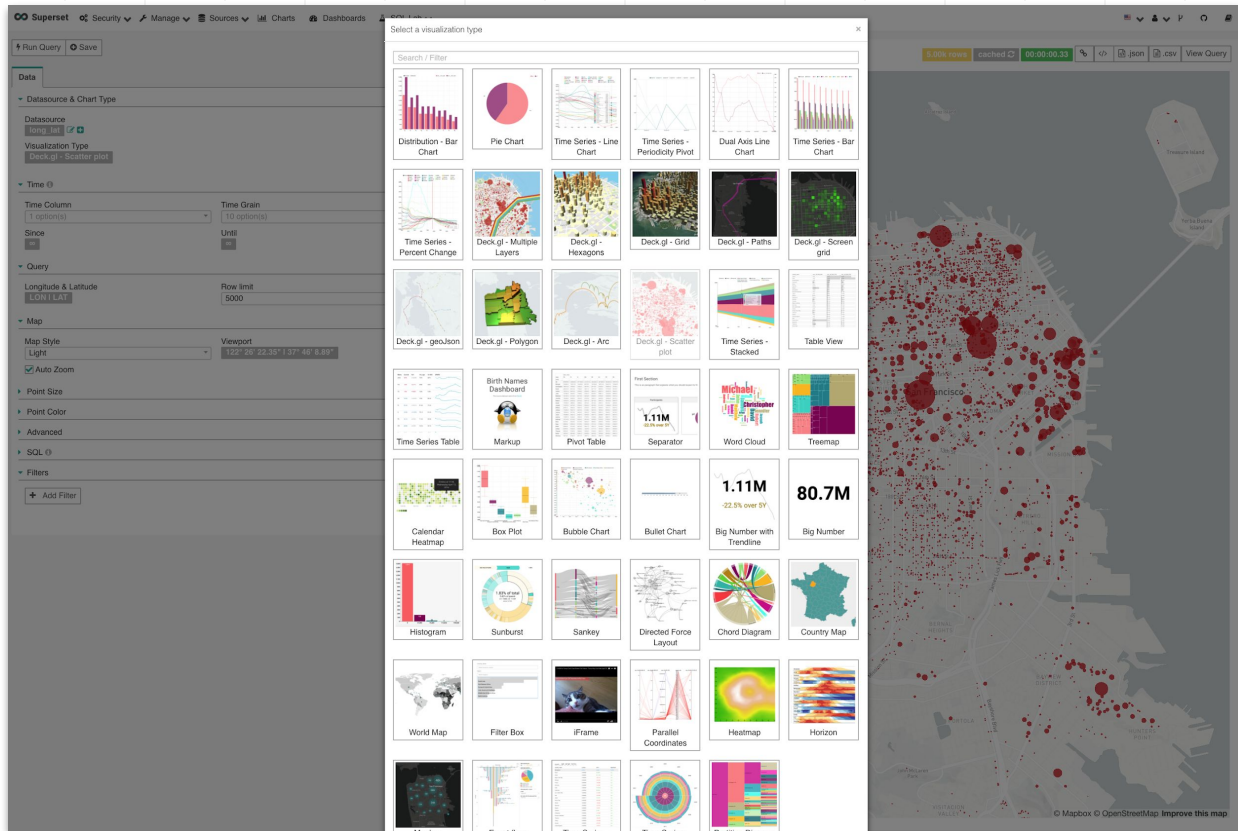
1.38k rows

cached

00:00:00.21

JSON CSV View Query







Growth Analysis

Scratchpad

Database: main

Schema: superset

Add a table (43)

## slices

created\_on  
changed\_on  
id  
slice\_name  
datasource\_type  
datasource\_name  
viz\_type  
params  
created\_by\_fk  
changed\_by\_fk  
description  
cache\_timeout  
perm  
datasource\_id



DATETIME  
DATETIME  
INTEGER  
VARCHAR  
VARCHAR  
VARCHAR  
VARCHAR  
TEXT  
INTEGER  
INTEGER  
TEXT  
INTEGER  
VARCHAR  
INTEGER

## dashboards

created\_on  
changed\_on  
id  
dashboard\_title  
position\_json  
created\_by\_fk  
changed\_by\_fk  
css



DATETIME  
DATETIME  
INTEGER  
VARCHAR  
TEXT  
INTEGER  
INTEGER  
TEXT  
TEXT

```
1 SELECT b.dashboard_id, a.dashboard_title, b.slice_id, c.slice_name
2 FROM dashboards a
3 JOIN dashboard_slices b ON a.id = b.das
4 JOIN slices c on c.id = b.slice_id
```

dashboards	table
dashboard_title	column
DATABASE	sql
datasource_type	column
datasource_name	column
datasource_id	column

Run Query

Save Query

Share Query

SQL

parameters

00:00:00.15

## Results

Query History

Preview for slices

Preview for dashboards

Visualize

CSV

Search Results

dashboard_id	dashboard_title	slice_id	slice_name
2	Births	882	Girls
2	Births	883	Boys
2	Births	884	Participants
2	Births	885	Genders
2	Births	886	Genders by State
2	Births	887	Trends
2	Births	888	Average and Sum Trends
2	Births	889	Title
2	Births	890	Name Cloud





# Apache Superset

A data visualization and exploration platform

- Easy-to-use & **fast “time-to-dashboard”**
- **Enterprise-ready** (RBAC) & **cloud-native**
- Richest set of **visualizations (50+)**
  - Solid geospatial visualization
- Lightweight **semantic layer**
- Works with a wide array of databases
- Deep integration with **Druid**
- A thriving and growing **community**

# The early days

Panoramix



Caravel





THE  
APACHE™  
SOFTWARE FOUNDATION

# The Superset Project

👁 Unwatch ▾

1,178

★ Unstar

23,626

🍴 Fork

4,379

📄 3,960 commits

🌿 75 branches

📦 142 releases

🚀 1 environment

👤 350 contributors

📄 Apache-2.0

- Thriving & accelerating open source community
- Most promising open source BI solution
- 1500 WAU at Airbnb (replaced Tableau), 400 WAU at Lyft
- 12 + committed engineers at 3 leading tech companies



# Stack

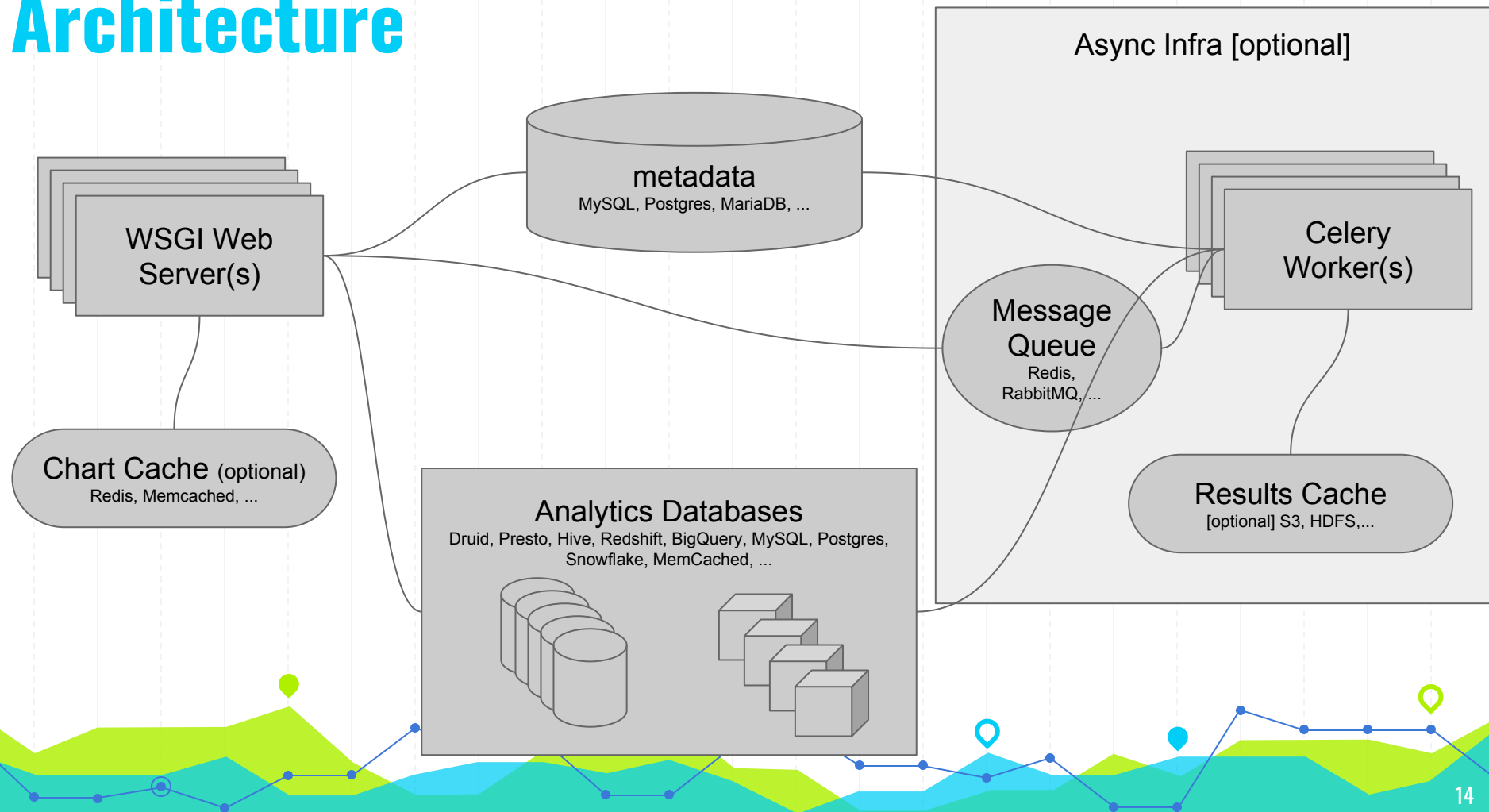
## ES6 **Javascript** Frontend

- React / Redux
- webpack / eslint / jest
- Broken down as many packages @superset-ui/\*
- nvd3, data-ui (VX), blocks, ...

## **Python** Backend

- Flask.\* + Flask App Builder
- Pandas
- SQLAlchemy (ORM + SQL Toolkit)
- Many utility libs (sqlparse, dateutils, ...)

# Architecture



# Challenges





# Challenge: a fast pace repo

February 25, 2019 – March 25, 2019

Period: 1 month ▾

## Overview



112 Active Pull Requests



95 Active Issues

 74

Merged Pull Requests

 38

Proposed Pull Requests

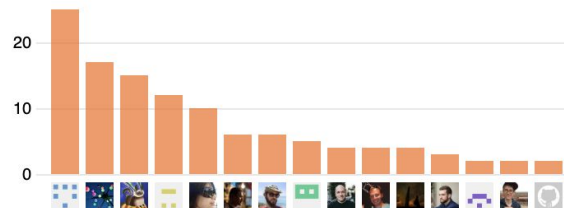
 49

Closed Issues

 46

New Issues

Excluding merges, **29 authors** have pushed **68 commits** to master and **131 commits** to all branches. On master, **129 files** have changed and there have been **4,581 additions** and **4,758 deletions**.



 74 Pull requests merged by 27 people

# Challenge: a huge dependency tree

```
superset@0.31.0-rc20 /Users/maximebeauchemin/code/superset/superset/assets
├── @babel/cli@7.2.3
├── chalk@2.4.1
├── anymatch@2.0.0
├── micromatch@3.1.10 count all dependencies
├── arr-diff@4.0.0
├── array-unique@0.3.2
├── braces@2.3.2
├── arr-flatten@1.1.0 deduped
├── array-unique@0.3.2 deduped
├── extend-shallow@2.0.1 deduped (0.64 seconds)
├── is-extendable@0.1.1 deduped
├── fill-range@4.0.0 deduped number of npm modules installed
├── extend-shallow@2.0.1 deduped how-to-get-number-of-npm-modules
├── is-extendable@0.1.1 deduped
├── is-number@3.0.0
├── kind-of@3.2.2 deduped You can use: npm ls --only=dev npm ls --only=prod
├── is-buffer@1.1.6 deduped -prod Pipe it out to grep to count.
├── repeat-string@1.6.1 deduped install and npm audit counts?
├── to-regex-range@2.1.1 deduped
├── isobject@3.0.1 deduped command to list the number of dependencies
├── repeat-element@1.1.3 deduped ar of installed packages, including ...
├── snapdragon@0.8.2 deduped ser-installed packages?
├── snapdragon-node@2.1.1 deduped verflow.com
├── split-string@3.1.0 deduped
├── to-regex@3.0.2 deduped
├── define-property@2.0.2 deduped documentation
├── extend-shallow@3.0.2
├── assign-symbols@1.0.0 deduped is on my dila.html *
├── is-extendable@1.0.1 deduped how to print out about all the versions of packages
├── is-plain-object@2.0.4 deduped nature. Positional arguments are ...
├── extglob@2.0.4
├── array-unique@0.3.2 deduped
├── define-property@1.0.0 deduped for robot/npm-dependencies: Count
├── is-descriptor@1.0.2 deduped listing npm-dependencies *
├── is-accessor-descriptor@1.0.0 deduped the results. npm-dependencies yargs
├── kind-of@6.0.2 deduped for: yargs@latest. - anali-regex@2.0.0
├── is-data-descriptor@1.0.0 deduped
├── kind-of@6.0.2 deduped
├── kind-of@6.0.2 deduped
├── expand-brackets@2.1.4 deduped way to check all your npm depend
├── debug@2.6.9 deduped the-easiest-way-to-check-all-your-dep
├── define-property@0.2.5 deduped the-easiest-way-to-check-all-your-dep
├── is-descriptor@0.1.6 deduped the-easiest-way-to-check-all-your-dep
├── is-accessor-descriptor@0.1.6 deduped the wc (word count) command
├── kind-of@3.2.2 deduped
├── is-buffer@1.1.6 deduped
├── is-data-descriptor@0.1.4 deduped
├── kind-of@3.2.2 deduped
├── is-buffer@1.1.6 deduped
├── kind-of@5.1.0 deduped know if NPM packages are installed?
├── extend-shallow@2.0.1 deduped
├── is-extendable@0.1.1 deduped
├── posix-character-classes@0.1.1 deduped
├── regex-not@1.0.2 deduped how to update and edit the
```

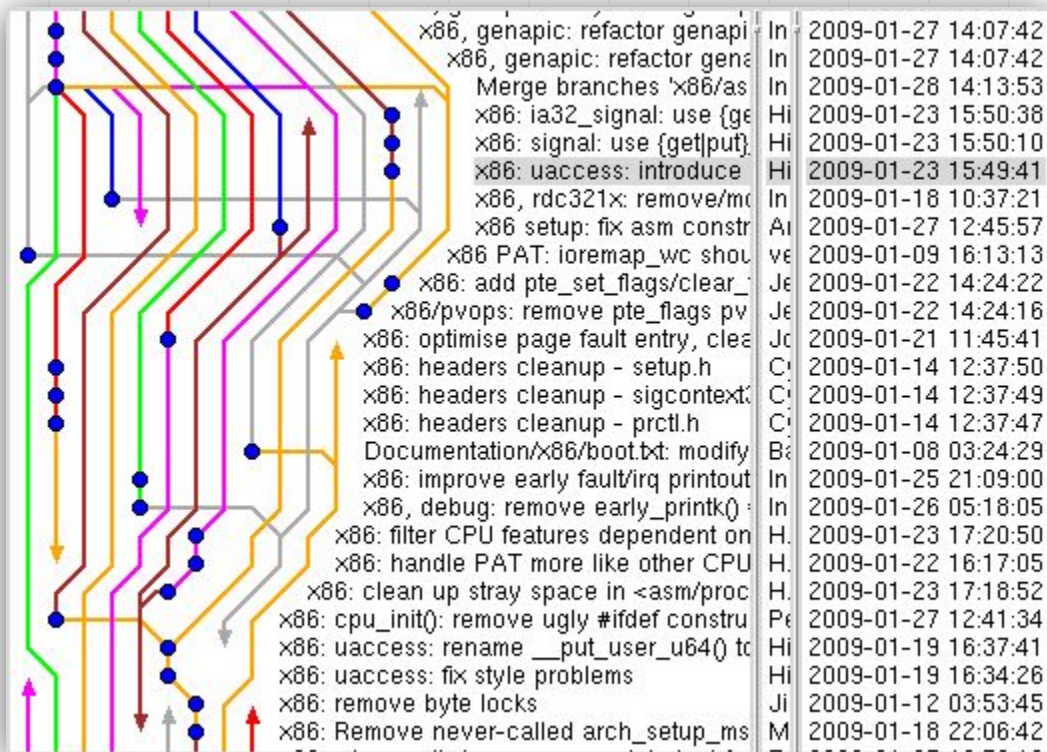
## Javascript:

- 88 production packages
- 68 dev packages
- `ls node_modules/ | wc -l == 1242`

## Python

- 35 direct dependencies
- ~66 leaves in the dep tree

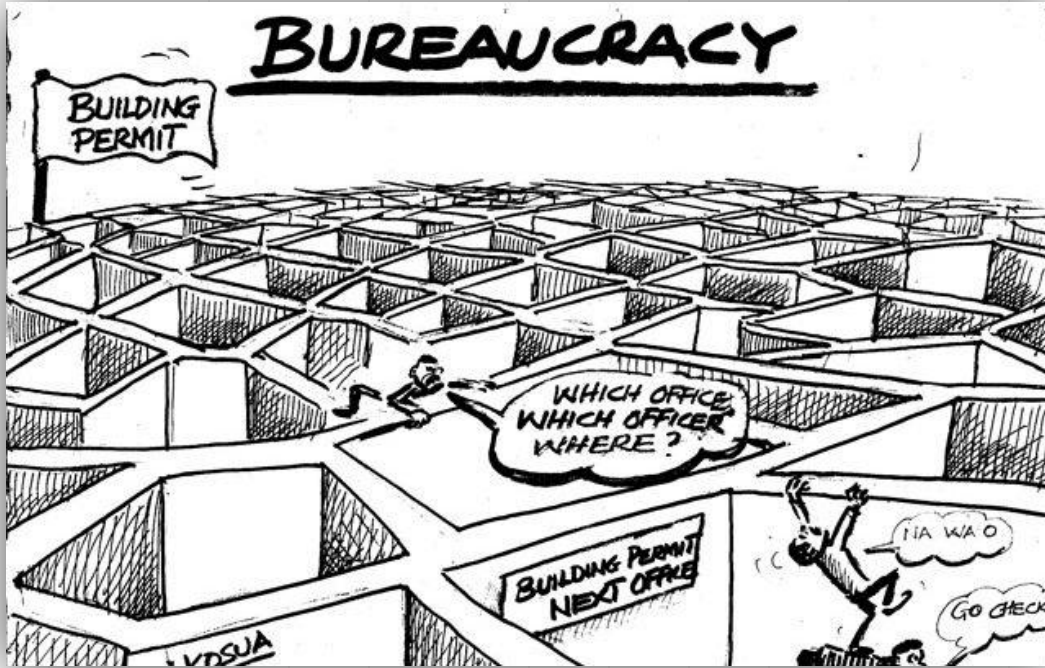
# Challenge: Release Management



# Challenge: Coordination



# Challenge: ASF bureaucracy



# Roadmap

- Steady Apache-approved releases
- Quality & polish ++
- Thumbnails + cards!
- A formal data access layer API
- Embeddable components
- Schedule simple data pipelines



# What's next!?

- Automated root cause analysis & anomaly detection
- Assisted dashboard generation
- Collaborative workspaces & social features
- Mobile!
- Data governance & auditing
- Integrated notebooks
- Storytelling
- Specialized visualization packages
- ML models introspection
- Alerts, notifications, email/mobile delivery



# Conclusion

- I'm looking to help companies onboard!
- Interested in working on Superset!?

[max@preset.io](mailto:max@preset.io)

[github.com/apache/incubator-superset](https://github.com/apache/incubator-superset)

[apache-superset.slack.com](https://apache-superset.slack.com)