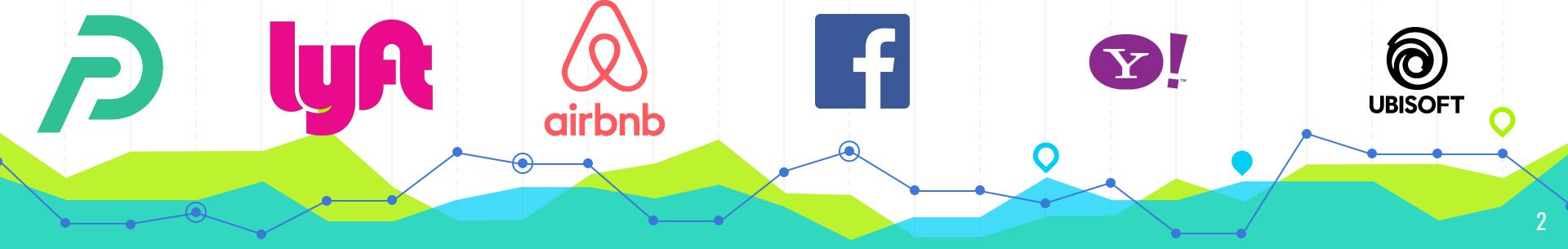




# The **history** and **anatomy** of Apache **Superset**

# 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



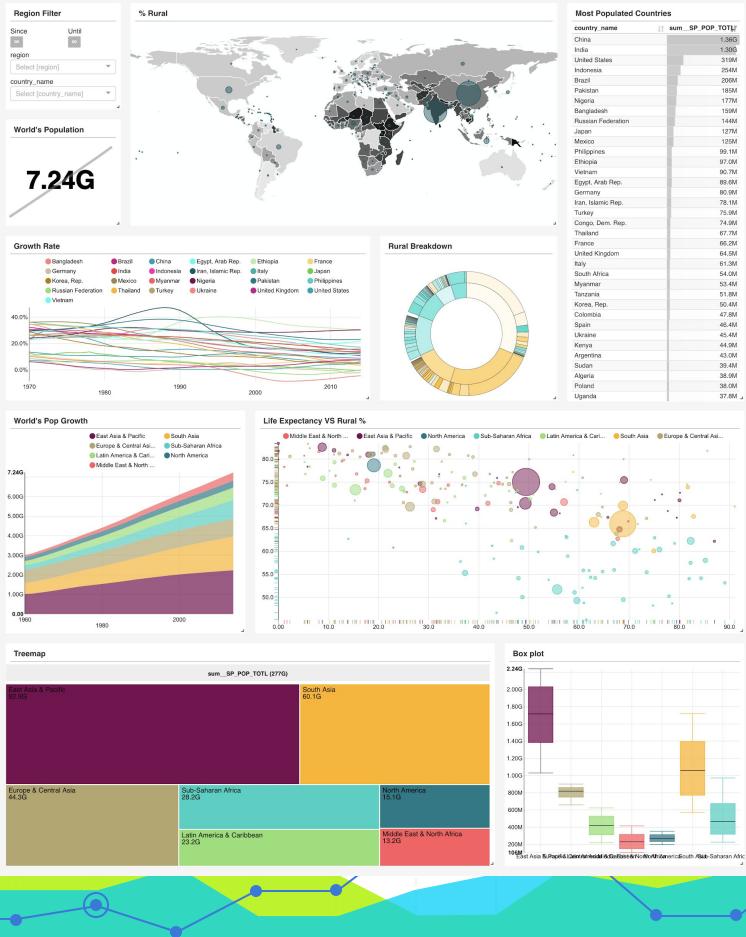
# Apache Superset

A data visualization and exploration platform

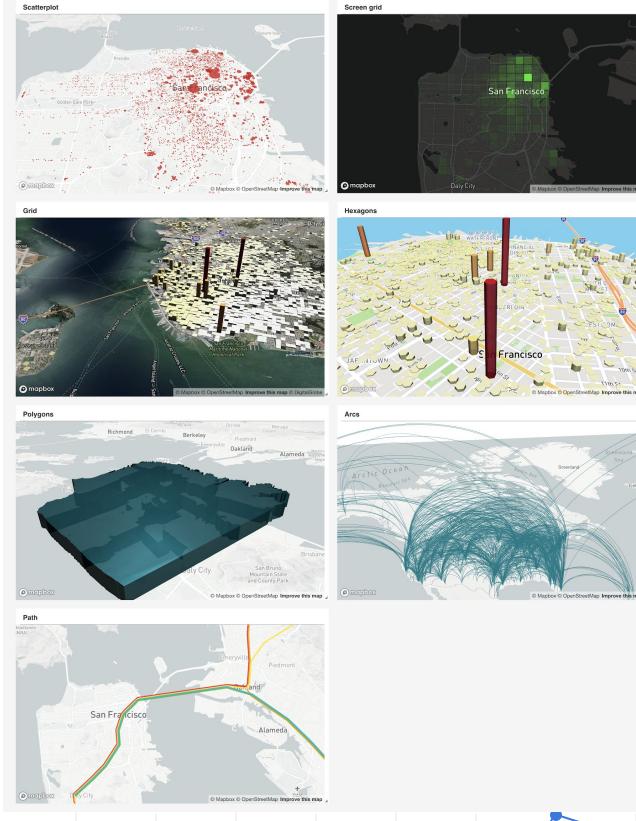
The image displays five distinct Apache Superset dashboards, illustrating its versatility across different data types and visualization techniques:

- World's Bank Data:** This dashboard features a world map showing rural populations, a donut chart of rural breakdowns, a line chart of world population growth from 1950 to 2010, and a treemap visualization of global health populations.
- Growth Analysis:** This screenshot shows the Superset SQL Lab interface. A query is being run against the 'main' database schema to select data from the 'dashboards' and 'dashboard\_slices' tables. The results show columns like dashboard\_id, dashboard\_title, slice\_id, and slice\_name.
- deck.gl Demo:** This dashboard demonstrates the use of the deck.gl library for 3D data visualization. It includes a scatterplot of red dots over a map, a grid overlay, and a screen grid view of San Francisco.
- Growth Rate:** This time-series chart tracks the growth rate of various countries from 1990 to 2010. The Y-axis ranges from -5.0% to 40.0%, and the X-axis shows years from 1992 to 2010. The legend lists countries including Bangladesh, Brazil, China, Egypt, Arab Rep., Ethiopia, France, Germany, India, Indonesia, Italy, Japan, Korea, Mexico, Myanmar, Nigeria, Pakistan, Philippines, Russia, Thailand, Turkey, Ukraine, United Kingdom, United States, and Vietnam.
- Complex Dashboard:** This dashboard is a collage of various visualizations:
  - A box plot showing data distribution for different regions.
  - A line chart with a blue line and green shaded area.
  - A 3D map of San Francisco with hexagonal overlays and colored arcs.
  - A scatterplot of red dots over a map.
  - A screen grid view of San Francisco.

## World's Bank Data ⚡



## deck.gl Demo ⚡



Run Query
 Save

---

Data
Style

---

▼ Datasource & Chart Type

Datasource

wb\_health\_population

Visualization Type

Time Series - Line Chart

---

▼ Time

Time Column
Time Grain

year

10 option(s)

Since
Until

1960-01-01

2014-01-02

---

▼ Query

Metrics

x sum\_SP\_POP\_TOTL

Group by

x country\_name

Series limit
Sort By

25



649 option(s)

Sort Descending

Contribution



1.38k rows cached

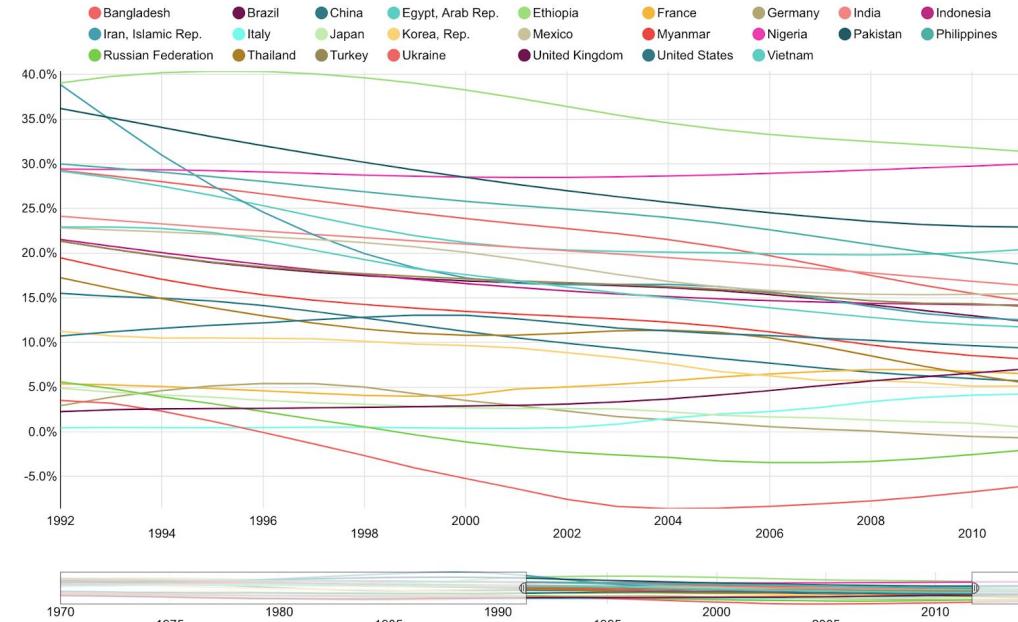
00:00:00

21

.json

.CSV

[View Query](#)



The screenshot shows a complex data visualization interface, likely Superset, with a sidebar on the left and a main content area on the right.

**Left Sidebar:**

- Run Query** and **Save** buttons.
- Data** section:
  - Datasource & Chart Type**: Long table.
  - Visualization Type**: Deck.gl - Scatter plot.
- Time** section:
  - Time Column**: 1 option(s).
  - Time Grain**: 10 option(s).
  - Since** and **Until** date pickers.
- Query** section:
  - Longitude & Latitude**: LON | LAT.
  - Row limit**: 5000.
- Map** section:
  - Map Style**: Light.
  - Viewport**: 122° 26' 22.35" | 37° 46' 8.89".
  - Auto Zoom.
- Point Size**, **Point Color**, **Advanced**, and **SQL** sections.
- Filters** section with a "+ Add Filter" button.

**Main Content Area:**

- Select a visualization type** dropdown.
- Search / Filter** input field.
- Visualization Grid** (4x6 grid):
  - Distribution - Bar Chart, Pie Chart, Time Series - Line Chart, Time Series - Periodicity Pivot, Dual Axis Line Chart, Time Series - Bar Chart.
  - Time Series - Percent Change, Deck.gl - Multiple Layers, Deck.gl - Hexagons, Deck.gl - Grid, Deck.gl - Paths, Deck.gl - Screen grid.
  - Deck.gl - geoJson, Deck.gl - Polygon, Deck.gl - Arc, Deck.gl - Scatter plot, Time Series - Stacked, Table View.
  - Time Series Table, Birth Names Dashboard, Markup, Pivot Table, First Section, 1.11M, Separator, Michael, Word Cloud, Treemap.
  - Calendar Heatmap, Box Plot, Bubble Chart, Bullet Chart, Big Number with Trendline, 1.11M, 22.5% over SY, 80.7M, Big Number.
  - Histogram, Sunburst, Sankey, Directed Force Layout, Chord Diagram, Country Map.
  - World Map, Filter Box, iFrame, Parallel Coordinates, Heatmap, Horizon.
  - Mapbox, Event Flow, Time Grids, Time Series, Plastic Pollution.
- Map View** on the right side showing a scatter plot of data points (red circles) over a geographical area.





Time Table    localhost:8081/superset/explore/?form...

Datasource Editor for birth\_names

Settings    Columns 7    Calculated Columns 1    Metrics 2

**Basic**

**Physical Table**

Database: mysql main  
Select a schema (6)

birth\_names

The pointer to a physical table. Keep in mind that the chart is associated to this Superset logical table, and this logical table points the physical table referenced here.

**Description**

1

**Default URL**

Default URL to redirect to when accessing from the datasource list page

**Autocomplete filters**

Whether to populate autocomplete filters options

**Autocomplete Query Predicate**

Use Legacy Datasource Editor

**Advanced**

**SQL**

1

When specifying SQL, the datasource acts as a view. Superset will use this statement as a subquery while grouping and filtering on the generated parent queries.

**Cache Timeout**

The duration of time in seconds before the cache is invalidated

**Hours offset**

0

**Template parameters**

A set of parameters that become available in the query using Jinja templating syntax

Save    Cancel

Time Table    localhost:8081/superset/explore/?form...

Datasource Editor for birth\_names

Settings    Columns 7    Calculated Columns 1    Metrics 2

**Basic**

**Physical Table**

Database: mysql main  
Select a schema (6)

birth\_names

The pointer to a physical table. Keep in mind that the chart is associated to this Superset logical table, and this logical table points the physical table referenced here.

**Description**

1

**Default URL**

Default URL to redirect to when accessing from the datasource list page

**Autocomplete filters**

Whether to populate autocomplete filters options

**Autocomplete Query Predicate**

Use Legacy Datasource Editor

**Advanced**

**SQL**

1

When specifying SQL, the datasource acts as a view. Superset will use this statement as a subquery while grouping and filtering on the generated parent queries.

**Cache Timeout**

The duration of time in seconds before the cache is invalidated

**Hours offset**

0

**Template parameters**

A set of parameters that become available in the query using Jinja templating syntax

Save    Cancel

# Apache Superset

A data visualization and exploration platform

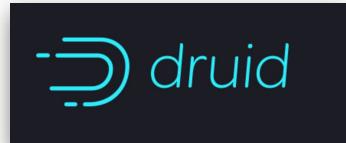
- Easy-to-use & fast “time-to-dashboard”
- Enterprise-ready (RBAC) & cloud-native
- Richest set of visualizations (50+)
  - Solid geospatial visualization
- Lightweight semantic layer
- Works with a wide array of databases
- Deep integration with Druid
- A thriving and growing community

# The early days

Panoramix



Caravel





THE

APACHE<sup>TM</sup>

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

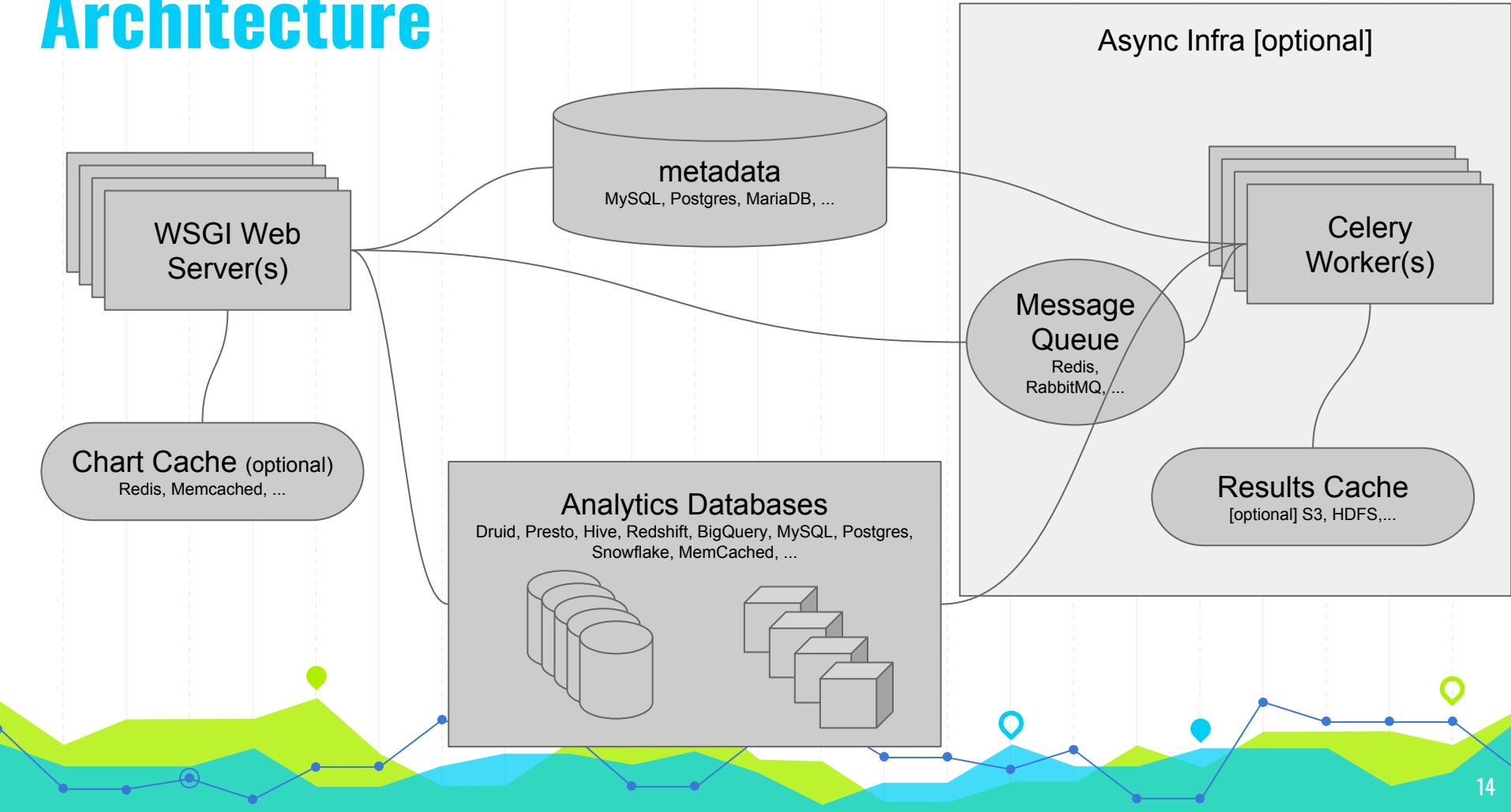
## ES6 Javascript Frontend

- 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

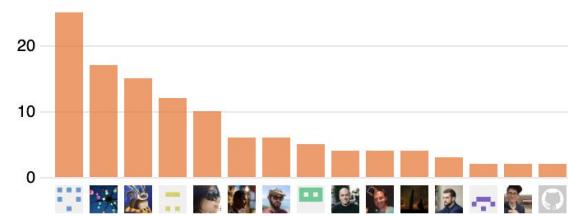


49  
Closed Issues



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



16

# Challenge: a huge dependency tree

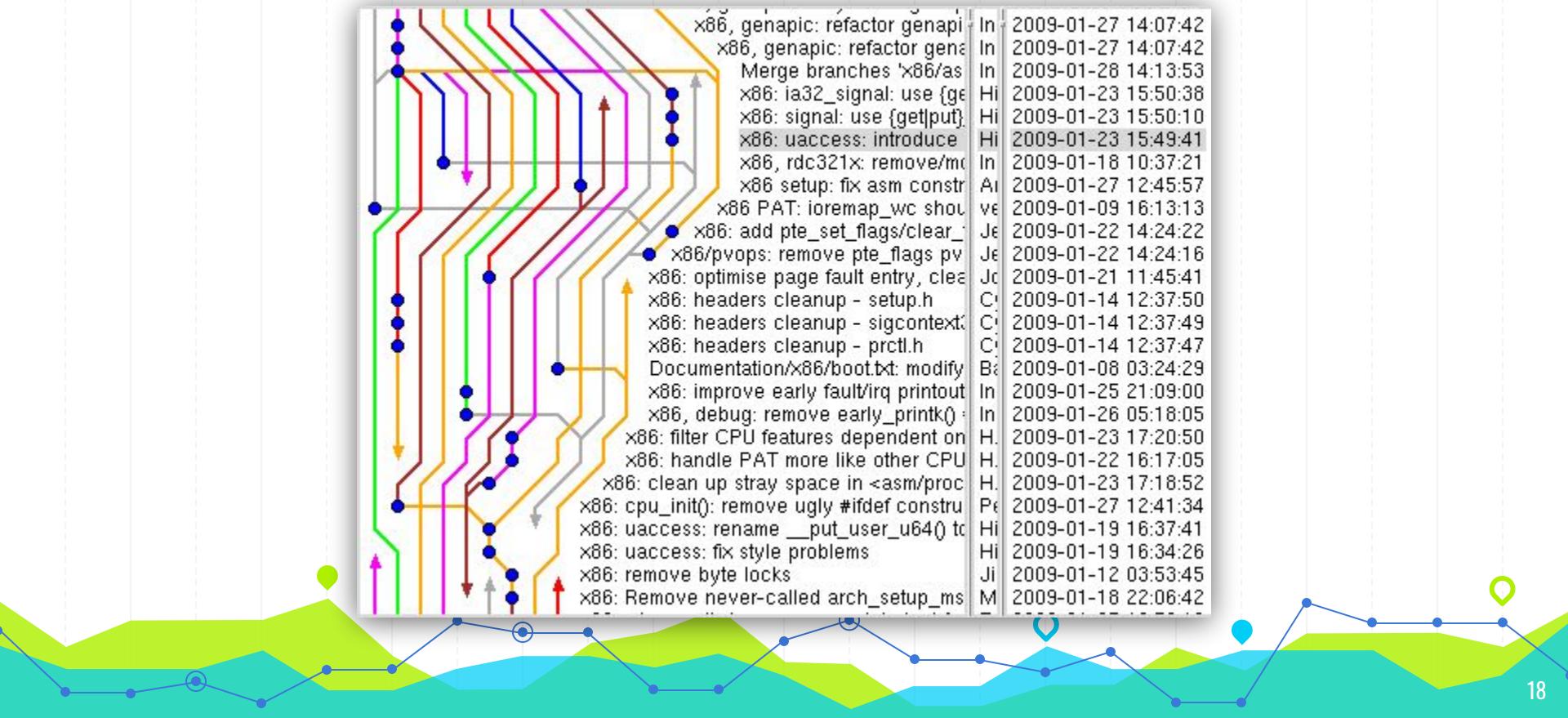
# Javascript:

- 88 production packages
  - 68 dev packages
  - ls node\_modules/ | wc -l == 1242

# Python

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

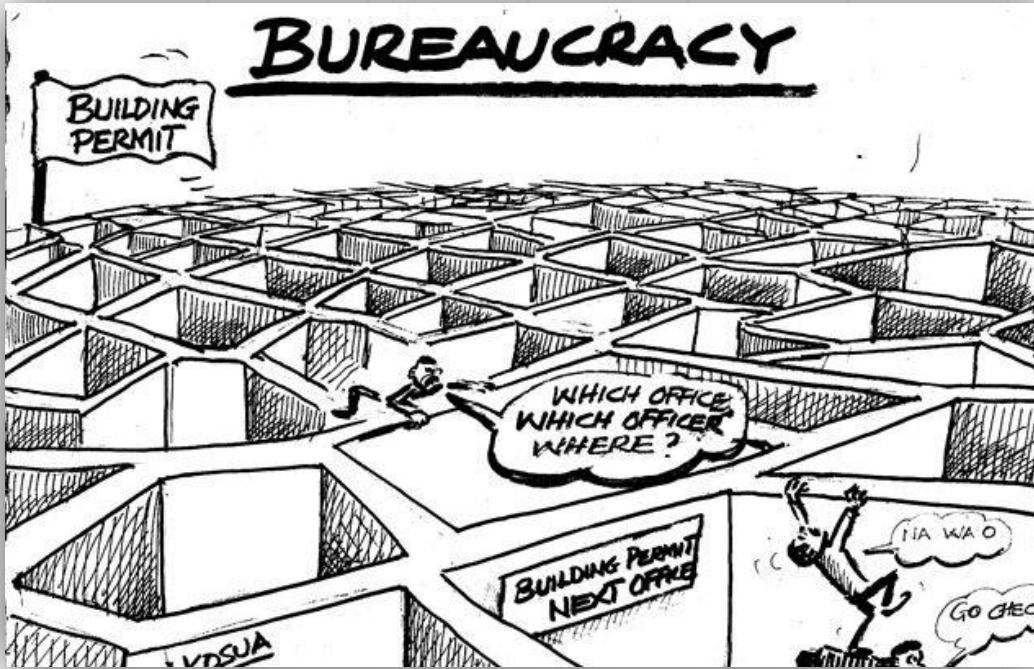
# Challenge: Release Management



# 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)