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
```sql
SELECT b.dashboard_id, a.dashboard_title, b.slice_id, c.slice_name
FROM dashboards a
JOIN dashboard_slices b ON a.id = b.dash
JOIN slices c ON c.id = b.slice_id
```

### Results

<table>
<thead>
<tr>
<th>dashboard_id</th>
<th>dashboard_title</th>
<th>slice_id</th>
<th>slice_name</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Births</td>
<td>882</td>
<td>Girls</td>
</tr>
<tr>
<td>2</td>
<td>Births</td>
<td>883</td>
<td>Boys</td>
</tr>
<tr>
<td>2</td>
<td>Births</td>
<td>884</td>
<td>Participants</td>
</tr>
<tr>
<td>2</td>
<td>Births</td>
<td>885</td>
<td>Genders</td>
</tr>
<tr>
<td>2</td>
<td>Births</td>
<td>886</td>
<td>Genders by State</td>
</tr>
<tr>
<td>2</td>
<td>Births</td>
<td>887</td>
<td>Trends</td>
</tr>
<tr>
<td>2</td>
<td>Births</td>
<td>888</td>
<td>Average and Sum Trends</td>
</tr>
<tr>
<td>2</td>
<td>Births</td>
<td>889</td>
<td>Title</td>
</tr>
<tr>
<td>2</td>
<td>Births</td>
<td>890</td>
<td>Name Cloud</td>
</tr>
</tbody>
</table>

---

**Datatypes:**
- `datetime`: DATETIME
- `integer`: INTEGER
- `varchar`: VARCHAR
- `text`: TEXT
- `column`: column
- `table`: table
- `sol`: sol
- `sql`: sql
- `database`: database
- `dataresource_type`: dataresource_type
- `dataresource_name`: dataresource_name
- `dataresource_id`: dataresource_id
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 Superset Project

- 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

WSGI Web Server(s)

Chart Cache (optional)
Redis, Memcached, ...

metadata
MySQL, Postgres, MariaDB, ...

Analytics Databases
Druid, Presto, Hive, Redshift, BigQuery, MySQL, Postgres, Snowflake, MemCached, ...

Async Infra [optional]

Celery Worker(s)

Message Queue
Redis, RabbitMQ, ...

Results Cache
[optional] S3, HDFS,...
Challenges
Challenge: a fast pace repo

February 25, 2019 – March 25, 2019

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

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

<table>
<thead>
<tr>
<th>Task</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refactor genapic: x66</td>
<td>2009-01-27</td>
</tr>
<tr>
<td>Refactor genapic: x66</td>
<td>2009-01-27</td>
</tr>
<tr>
<td>Merge branches x66/asm</td>
<td>2009-01-26</td>
</tr>
<tr>
<td>Move all code to x64</td>
<td>2009-01-26</td>
</tr>
<tr>
<td>x86: ia32_signal: use {get</td>
<td>put}</td>
</tr>
<tr>
<td>x86: signal: use {get</td>
<td>put}</td>
</tr>
<tr>
<td>x86: uaccess: introduce</td>
<td>2009-01-23</td>
</tr>
<tr>
<td>x86: rdc321x: remove/move</td>
<td>2009-01-18</td>
</tr>
<tr>
<td>x86 setup: fix asm constructs</td>
<td>2009-01-27</td>
</tr>
<tr>
<td>x86 PAT: ioremap_wc show</td>
<td>2009-01-09</td>
</tr>
<tr>
<td>x86: add pte_set_flags/clear</td>
<td>2009-01-22</td>
</tr>
<tr>
<td>x86/pvops: remove pte_flags pv</td>
<td>2009-01-22</td>
</tr>
<tr>
<td>x86: optimise page fault entry, clear</td>
<td>2009-01-21</td>
</tr>
<tr>
<td>x86: headers cleanup - setup.h</td>
<td>2009-01-14</td>
</tr>
<tr>
<td>x86: headers cleanup - sigcontext.h</td>
<td>2009-01-14</td>
</tr>
<tr>
<td>x86: headers cleanup - prctl.h</td>
<td>2009-01-14</td>
</tr>
<tr>
<td>Documentation/x66/boots.txt: modify</td>
<td>2009-01-08</td>
</tr>
<tr>
<td>x86: improve early fault/irq printout</td>
<td>2009-01-25</td>
</tr>
<tr>
<td>x86, debug: remove early_print()</td>
<td>2009-01-26</td>
</tr>
<tr>
<td>x86: filter CPU features dependent on</td>
<td>2009-01-23</td>
</tr>
<tr>
<td>x86: handle PAT more like other CPU</td>
<td>2009-01-22</td>
</tr>
<tr>
<td>x86: clean up stray space in asm/proc</td>
<td>2009-01-23</td>
</tr>
<tr>
<td>x86: cpu_init(): remove ugly ifdef constructs</td>
<td>2009-01-27</td>
</tr>
<tr>
<td>x86: uaccess: rename __put_user_u64() to</td>
<td>2009-01-26</td>
</tr>
<tr>
<td>x86: uaccess: fix style problems</td>
<td>2009-01-19</td>
</tr>
<tr>
<td>x86: remove byte locks</td>
<td>2009-01-12</td>
</tr>
<tr>
<td>x86: Remove never-called arch_setup_ms</td>
<td>2009-01-18</td>
</tr>
</tbody>
</table>
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

[Links]
github.com/apache/incubator-superset
apache-superset.slack.com