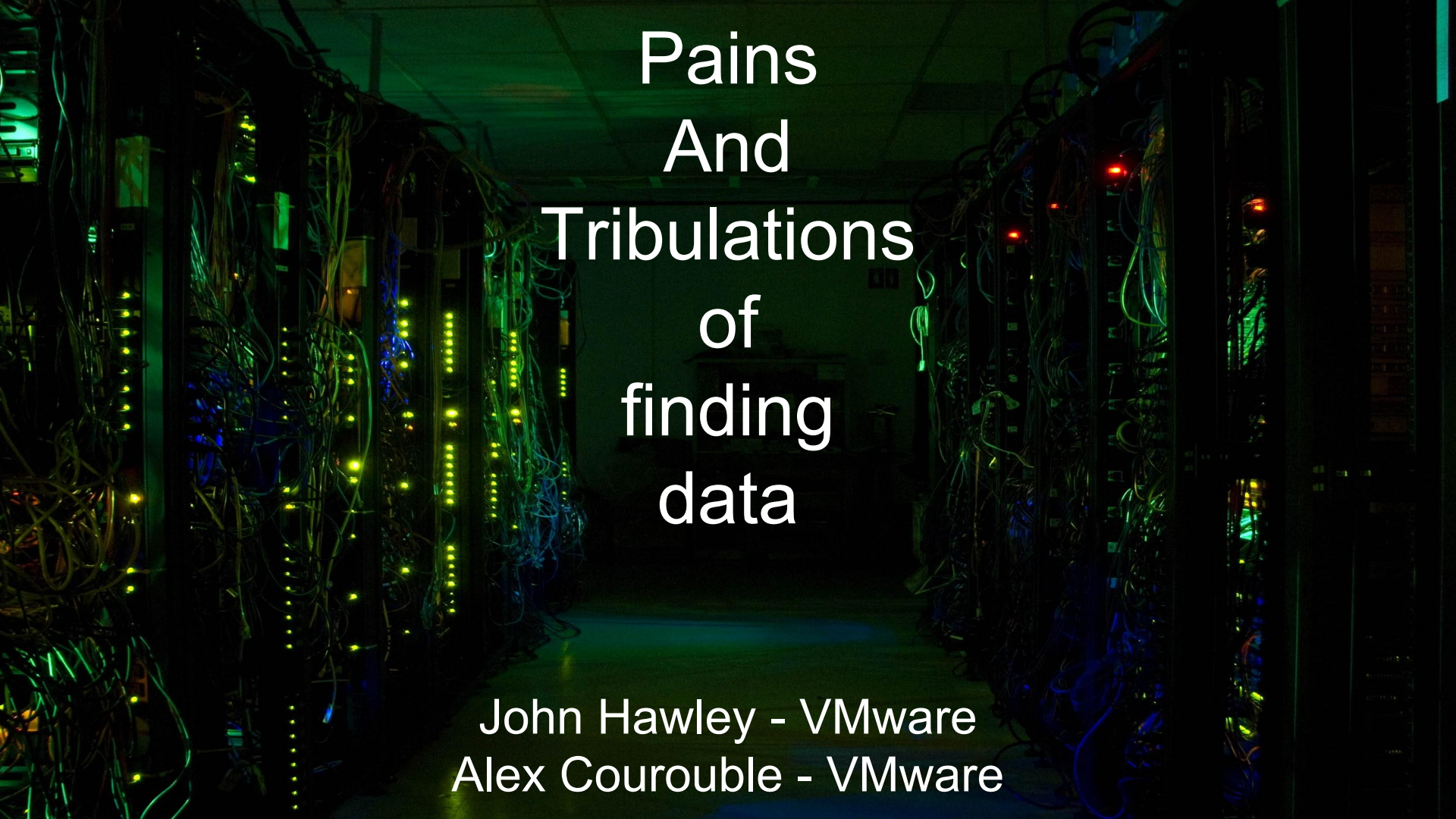
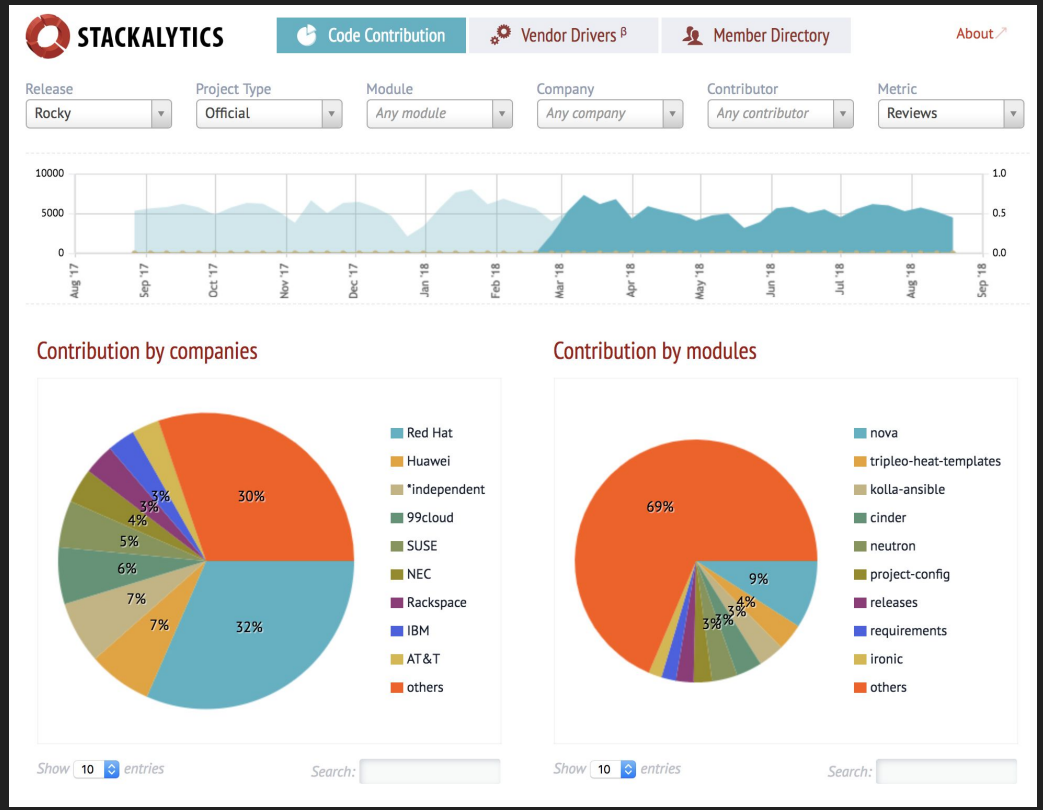


John Hawley - VMware
Alex Courouble - VMware

What’s gotten us this far...

Stackalytics for Openstack



Devstats for CNCF

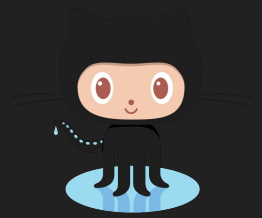
Grimoire Lab

Impressive list of data source:

Slack

Jenkins

Git



Github

Gitlab

Twitter

Telegram

Bugzilla

Dockerhub

Stackexchange

(And more)

The questions we are asking are fundamentally hard

But access to the data makes it harder to even start!

GitHub’s API

- Is painfully slow
- Seemingly misses data when you ask for it
- Is painfully slow
- Values are inconsistently returned, and may randomly change.
- Is painfully slow
- Event streams sometimes never deliver some data
- Did I mention, it’s painfully slow?

But access to the data makes it harder to even start!

GH Archive will save us!

- Based on the activity stream
  - That’s known to sometimes never deliver some events
- To make any real sense of it, you’d need to pull in and parse the entire stream
- There’s still a good chance you’ll be missing data
  - (real question is, does that matter for you and your queries?)
- E-mail addresses obfuscated
  - "author":{"email":"00bd4d57bfb0456e4d4147a0954ac944447a5fcb@gmail.com", [...]
  - "author":{"email":"dac17826b12bfd084b5f0d5793214561 39c4f746@b8457f37-d9ea-0310-8a92-e5e31aec5664", [...]

But access to the data makes it harder to even start!

Google BigQuery

To the rescue! (?)

I don’t have an answer here that’s “good”

Query the data yourself

- It’ll take forever, and now you are on the hook

Convince GH Archive , and Google, to stop obfuscating the data, and to help folks understand what projects are/aren’t included

- Highly doubtful to happen

Pie in the sky dream would be a distributed effort to grab the data and collectively share it

Or, Maybe GitHub’s new owner would be willing to discuss some of these issues more?

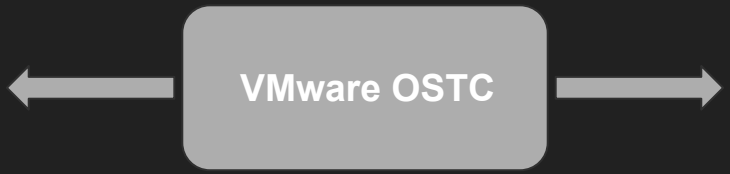
Measuring VMware OSTC’s impact in OSS

- Actively contributing to an undefined number of OSS projects
- Contributions include:
  - Pull Requests Merged
  - Issues Opened
  - Commits Merged (non-github)
  - Code Reviews
  - ...
- Github api works great for low scale user-centric use cases.

Tracking OSTC’s contributions

Github hosted projects

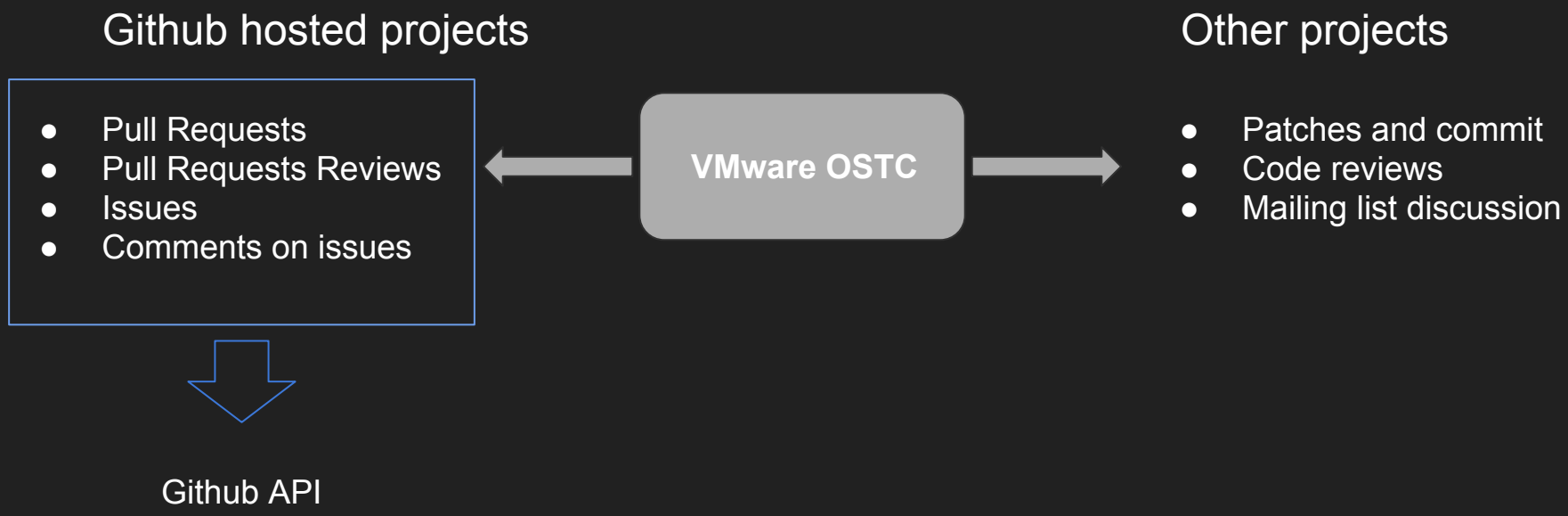
- Pull Requests
- Pull Requests Reviews
- Issues
- Comments on issues



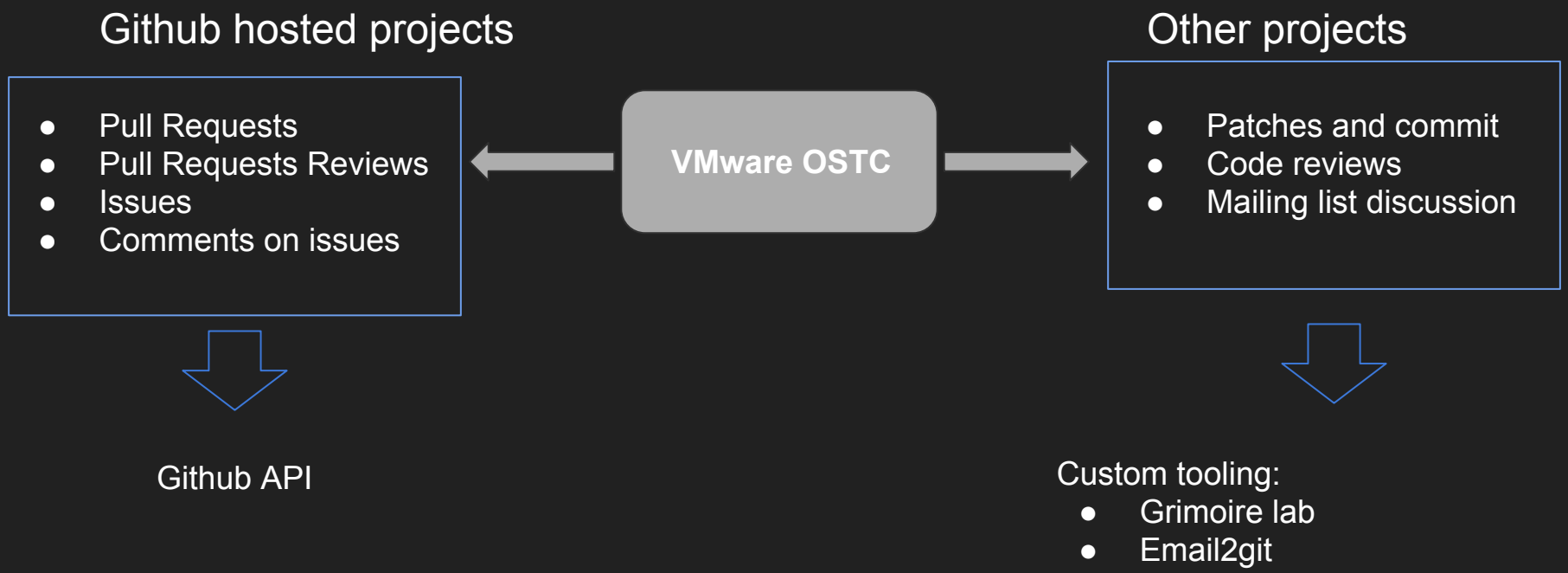
Other projects

- Patches and commit
- Code reviews
- Mailing list discussion

Tracking OSTC's contributions



Tracking OSTC's contributions



Questions, Comments, Concerns?

John Hawley - VMware - @warty9
warthog9@eaglescrag.net

Alex Courouble - VMware - @alexcourouble -

acourouble@vmware.com