

Mistakes to Avoid when Open Sourcing Proprietary Tech

Open Source Summit North America 2018

@IndeedEng | opensource.indeedeng.io



Hi, I'm Jack.

Jack Humphrey, VP Engineering
@IndeedEng

@youknowjack | <http://jackhumphrey.me>



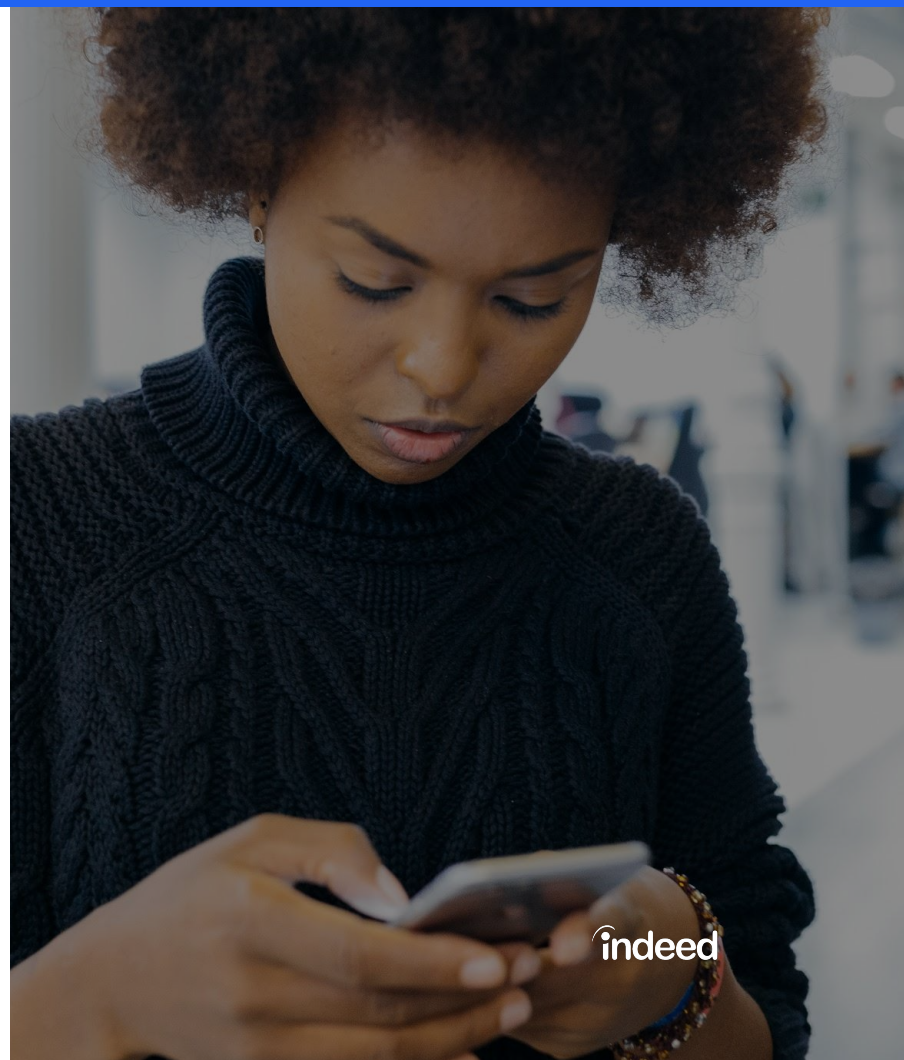
**We help
people
get
jobs.**

60 countries

30 languages

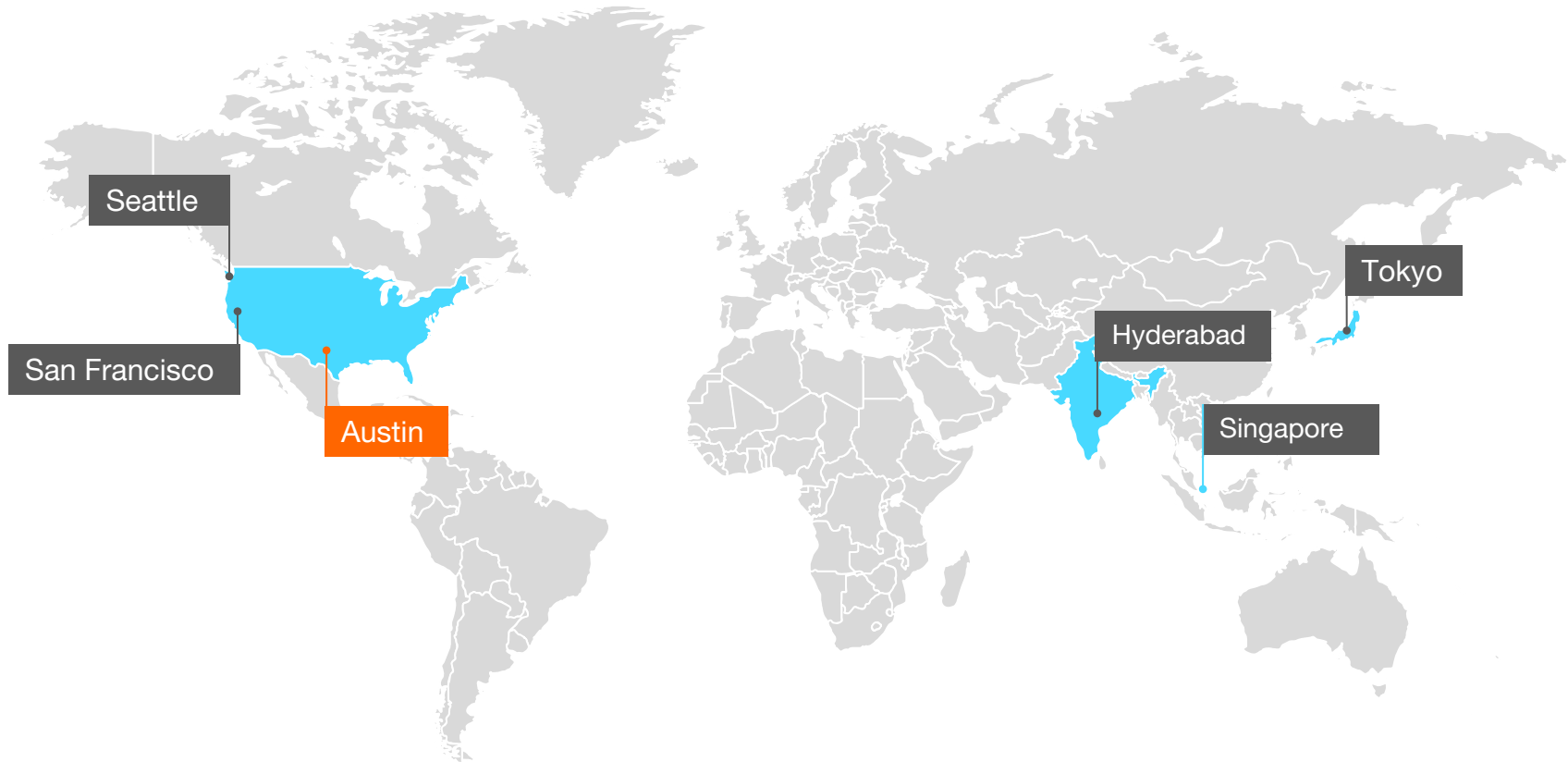
200M unique visitors

20M jobs



indeed

Engineering offices



And now: Vancouver!

Indeed Announced Today That It Will Acquire Resume.com to Enable Job Seekers to Build and Download Personalized Resumes for Free

Austin, TX (May 25, 2018) — Indeed announced today that it acquired resume.com, an online service for creating personalized resumes. Over 3.5 million job seekers currently use the site, creating more than 40,000

Today: learn from our failures

Prologue: Questions to ask before open sourcing

Chapter 1: External forks and rotten bits

Chapter 2: Failure to connect

Chapter 3: Hiding the work

Chapter 4: Time for an OSPO

Epilogue: Reviewing the mistakes and lessons learned

“Maybe we should open source
[COOL CODENAME].”

SET COURSE FOR OPEN SOURCE



MAKE IT SO

Belay that order!

Questions to ask

1

Why do you want to open source?

2

Can you commit the necessary resources?

3

Do you know the common pitfalls?

1. Why do we want to open source?

Build awareness in the community and attract talent to Indeed

Create opportunities for Indeed's software engineers

Get help maintaining and enhancing our software

Give back to Open Source

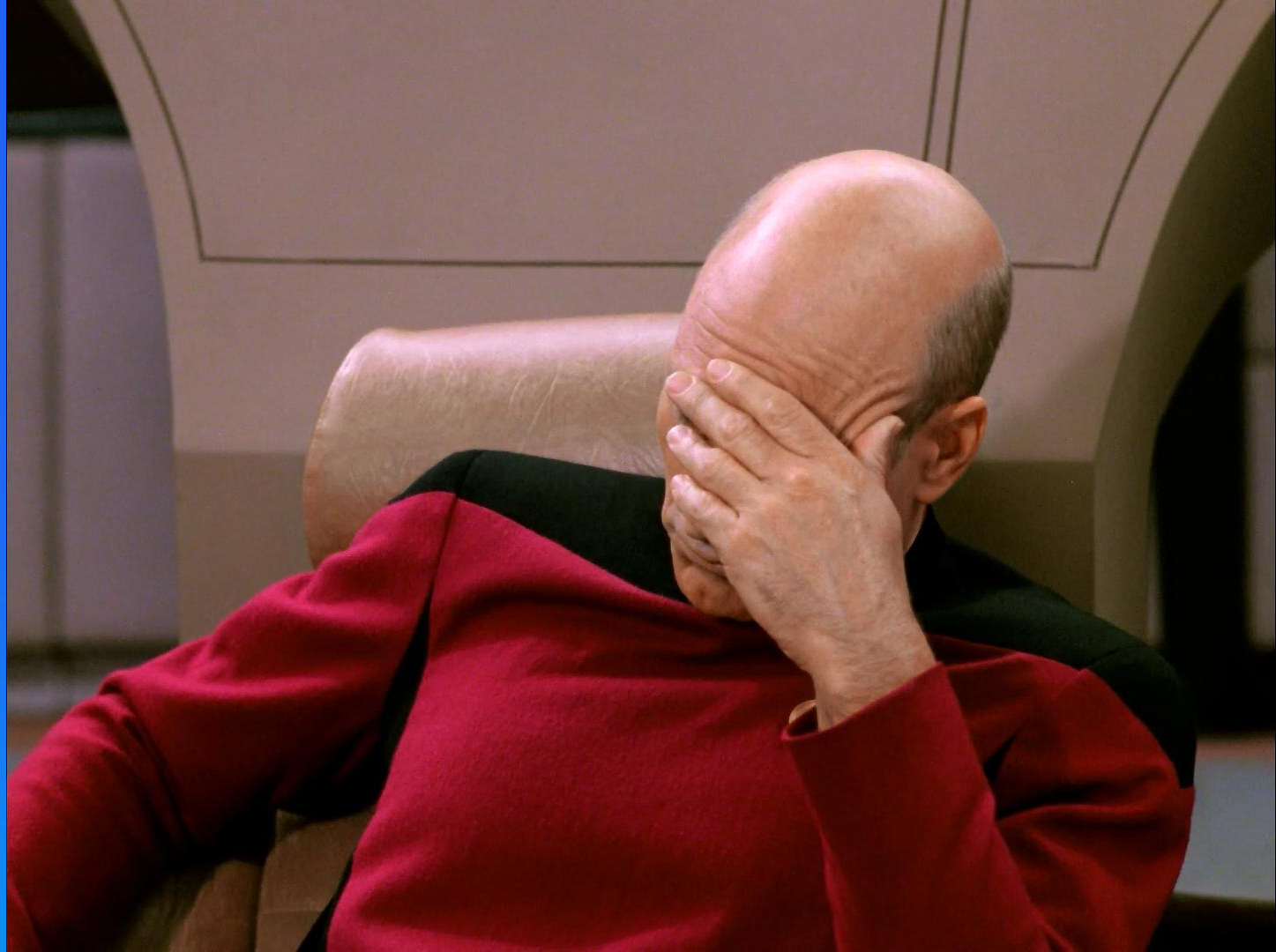
2. Have we committed enough resources?

Not yet.

3. Did we avoid all the pitfalls?

Nope!

Passionately Incompetent



Chapter 1

External forks and rotten bits

2009: We built something cool

Monolithic App \Rightarrow Service-Oriented Architecture

How could we distribute load to services without a load balancer?

Solution: A peer-to-peer scheme for service load distribution

Boxcar, a distributed services framework



2010: Painstaking analysis took place

2:41 PM **Doug**: I'm thinking of
proposing that we open source boxcar
as well.... Thoughts?

2:45 PM **me**: i think that would be cool

Make it so

1. An engineer volunteered to open source it.
2. Cleaned up the internal dependencies.
3. Picked a license.
4. Wrote a project overview (324 words).
5. August 2010: Published to Google Code, as an **external fork**:
"The plan is to keep our internal versions, and propagate changes on a semi-regular schedule."
6. January 2011: Presented Boxcar to some other tech companies in Austin

We blogged about it

Boxcar: A self-balancing distributed services protocol

Posted on December 17, 2012 by [Ketan Gangatirkar](#) and [Courtney Jeffries](#)

At Indeed, we have an unrelenting obsession with speed. We believe site performance is critical to providing the best experience for job seekers. Since Indeed launched in 2004, traffic growth has been swift and steady. Today, we have over 85 million unique visitors and over 2 billion searches per month. Traffic to Indeed continues to rapidly increase, so ensuring our systems scale with this traffic really matters. In order for our site to be fast and reliable, we developed Boxcar, an RPC framework and load distribution strategy for interacting with remote services using a generic layer to transport raw bytes.

go.indeed.com/boxcar



Will any of this become open source?



Ian Ragsdale

@iragsdale

Follow



Replying to @IndeedEng

[@IndeedEng](#) very interesting article. Any info forthcoming on how clients discover servers? Will any of this become open source?

12:58 AM - 18 Dec 2012



The bits rotted

We eventually added a deprecation notice.

<https://code.google.com/archive/p/boxcar/>



MISTAKE

Release OSS project
as an external fork of
internal version

LESSON

Use OSS release
internally

MISTAKE

Neglected external
validation and
promotional activities

LESSON

Validate the need for
your project, work
closely with early
adopters, and talk
about it to anyone
who will listen

Oh and don't wait 2 years to blog about it:

Boxcar: A self-balancing distributed services protocol

Posted on December 17, 2012 by [Ketan Gangatirkar](#) and [Courtney Jeffries](#)

At Indeed, we have an unrelenting obsession with speed. We believe site performance is critical to providing the

Posted on December 17, 2012

growth has been swift and steady.

es per month. Traffic to Indeed continues

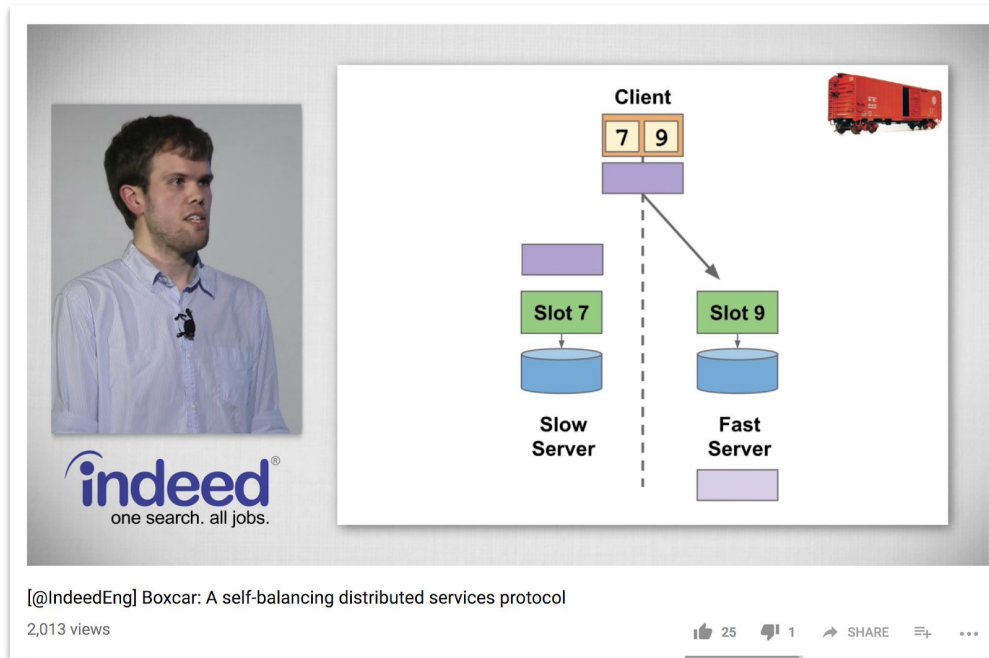
matters. In order for our site to be fast

and reliable, we developed Boxcar, an RPC framework and load distribution strategy for interacting with remote services using a generic layer to transport raw bytes.

go.indeed.com/boxcar



...or 3 years to present about it



The video player shows a man in a light blue shirt on the left. To his right is a diagram of the Boxcar protocol. The diagram features a 'Client' at the top with two boxes labeled '7' and '9'. A dashed vertical line separates the 'Slow Server' on the left from the 'Fast Server' on the right. The 'Slow Server' has a 'Slot 7' (green box) above a database icon. The 'Fast Server' has a 'Slot 9' (green box) above a database icon. An arrow points from the 'Client' to the 'Fast Server' Slot 9. A red toy train is in the top right corner of the diagram. Below the video, the text reads: '@IndeedEng] Boxcar: A self-balancing distributed services protocol', '2,013 views', '25' likes, '1' comment, 'SHARE', and a menu icon.

indeed[®]
one search. all jobs.

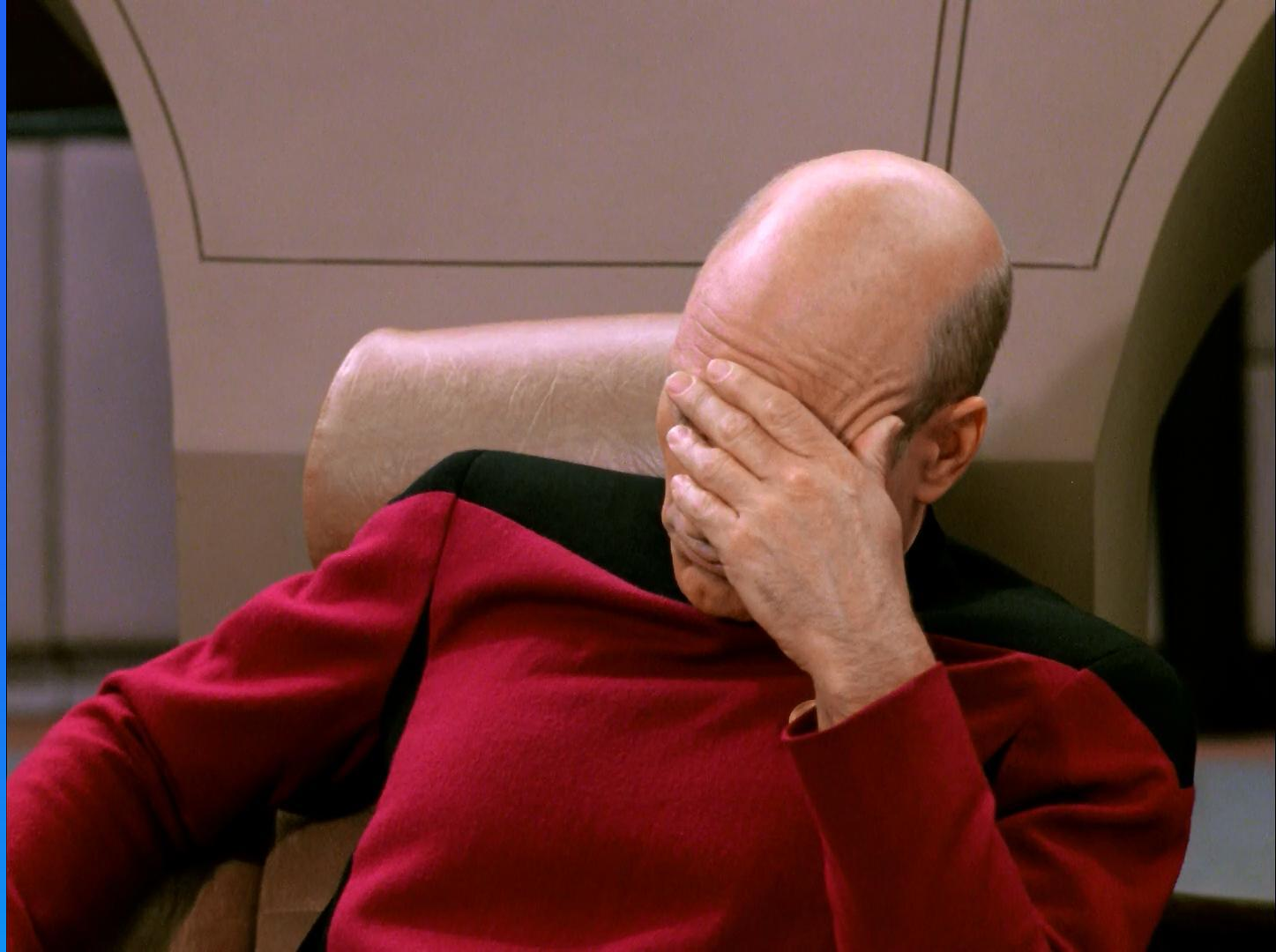
[@IndeedEng] Boxcar: A self-balancing distributed services protocol

2,013 views

25 1 SHARE ...

go.indeed.com/boxcar-talk





Chapter 2

Failure to connect

We built something cool

We loved A/B testing, but adjusting test allocations required code changes/deploys and took too long.

Solution: **Proctor** A/B Testing Framework
Allows dynamic test definition without code changes




opensource.indeedeng.io/proctor

Make it so

1. Recruited engineers to work on it
2. Cleaned up the internal dependencies
3. Figured out how to build, publish, and consume internally
4. Published some of our useful common libraries on Github
5. Wrote great documentation (using Github Pages)
6. October 2013: Released the core components --
github.com/indeedeng/proctor -- management webapp was
delivered later
7. And...

We talked about it



The thumbnail features a man in a striped shirt on the left and a presentation slide on the right. The slide has the title 'Proctor' in bold, followed by the text 'Indeed's **Open Source** Java Framework for Managing A/B Tests and More', and the URL github.com/indeedeng/proctor. The Indeed logo is at the bottom left of the slide. Below the video frame, the text '@IndeedEng: Managing Experiments and Behavior Dynamically with Proctor' is displayed, along with '1,893 views' and engagement icons for likes (15), comments (0), and shares.

Proctor

Indeed's **Open Source** Java Framework for
Managing A/B Tests and More

github.com/indeedeng/proctor

indeed[®]
one search. all jobs.

indeed

@IndeedEng: Managing Experiments and Behavior Dynamically with Proctor

1,893 views

15 0 SHARE ...

go.indeed.com/proctor-talk



We got an early adopter

Homeaway started using it...
and contributed some code!

Git support added

[Browse files](#)

 master (#1)  proctor-parent-1.1.55 ... proctor-parent-1.0.16



jdcheng authored and **chriscolon** committed on Jun 19, 2014

1 parent [bdcc098](#)

commit [f708df515f00931849610dabf3eceb02a80e2179](#)



HomeAway®



We got an unsolicited PR from a stranger

Expose raw JSON matrix for consumption by other servers #2

New issue

Merged chriscolon merged 1 commit into `indeedeng:master` from `benjamincochran:master` on Dec 1, 2014

Conversation 1

Commits 1

Checks 0

Files changed 1

+8 -0



benjamincochran commented on Nov 24, 2014

Contributor

...

Reviewers

No reviews

This is meant to be used with clients implementing UrlProctorLoaders.



indeed

We got Google traffic



java a/b testing



All

Shopping

News

Videos

Images

More

Settings

Tools

About 59,500,000 results (0.24 seconds)

GitHub - indeedeng/proctor: Proctor is a Java-based A/B testing ...

<https://github.com/indeedeng/proctor> ▼

Proctor. Proctor is a **A/B testing** framework written in **Java** that enables data-driven product design at Indeed. Proctor consists of data-model, client specification, ...



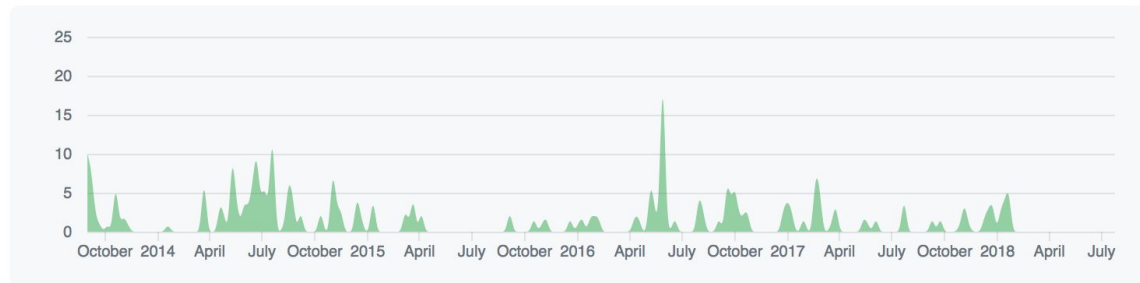


Not so fast, Captain

Sep 29, 2013 – Aug 22, 2018

Contributions: **Commits** ▼

Contributions to master, excluding merge commits



2018: 1 contributor

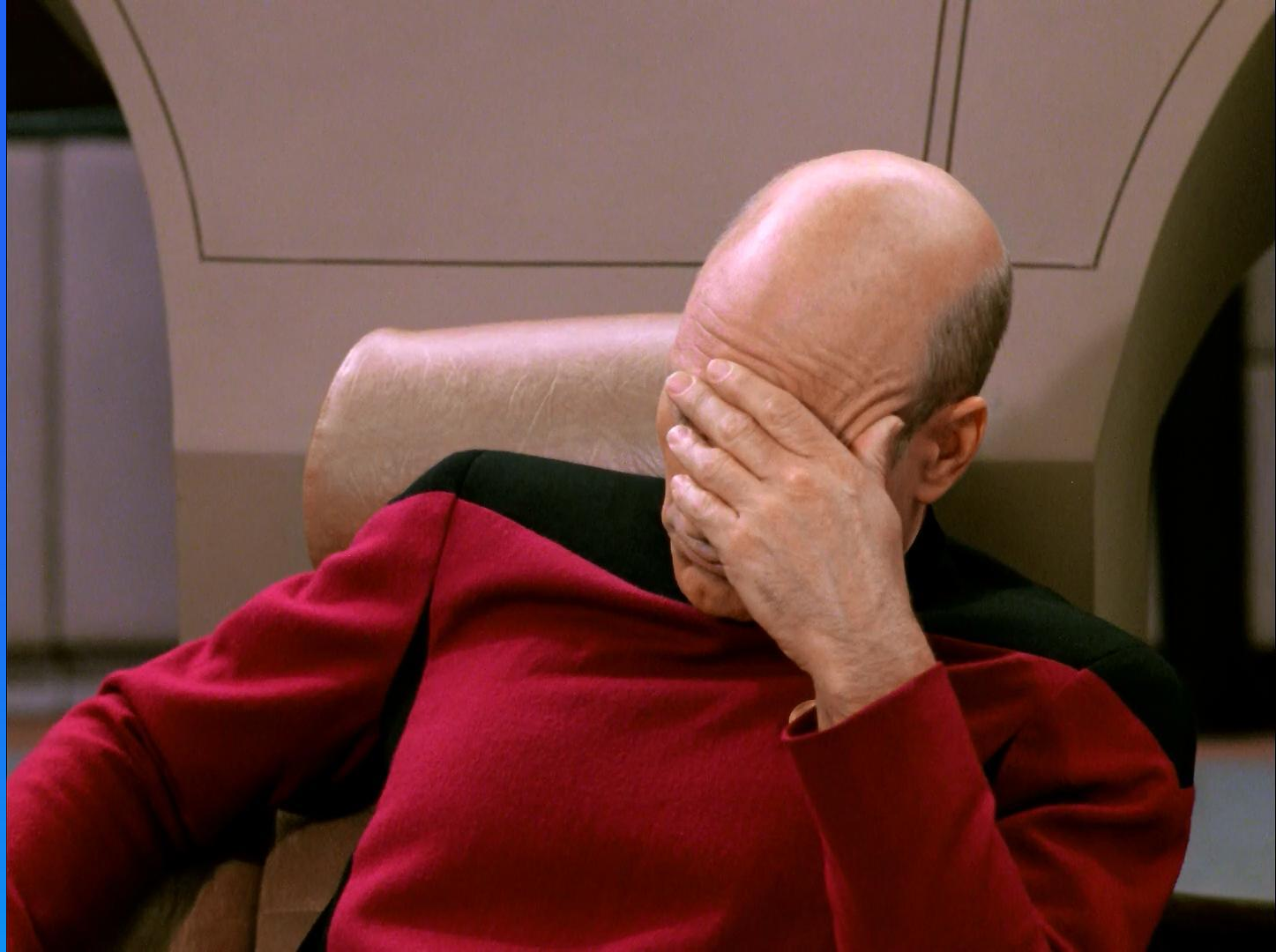
All time: 11 contributors, 63 forks

MISTAKE

Under-invested in
promotion and
community building

LESSON

Invest in building a
community of users
and contributors



Chapter 3

Hiding the work

We built something cool

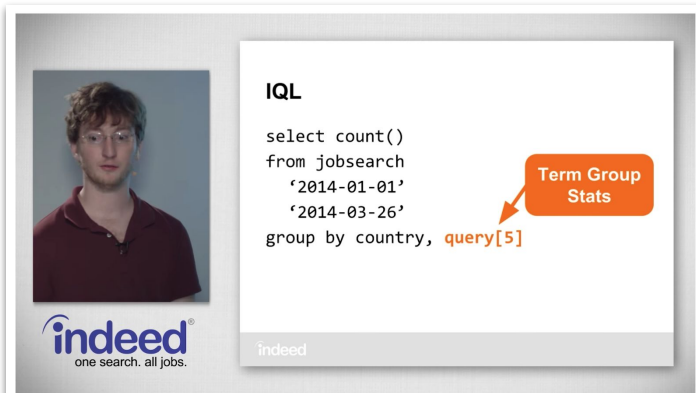
We love to make data-driven product decisions, but analyzing data can take too long and require specialized skill sets

Solution: **Imhotep**, a data analytics platform fast, interactive, ad hoc queries and aggregate results for large time-series datasets.



opensource.indeedeng.io/imhotep

We talked about it



IQL

```
select count()  
from jobsearch  
  '2014-01-01'  
  '2014-03-26'  
group by country, query[5]
```

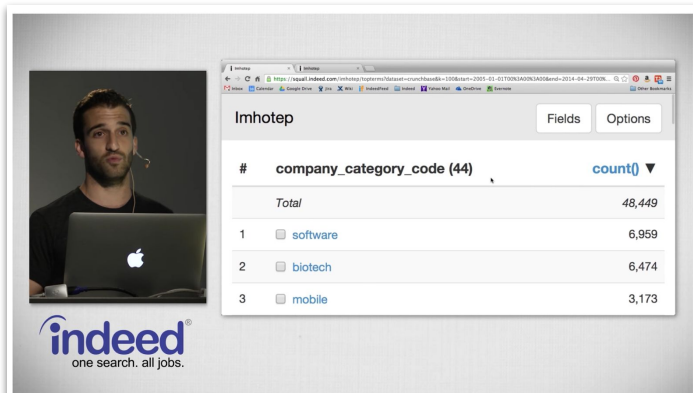
Term Group Stats

indeed[®]
one search. all jobs.

[@IndeedEng] Imhotep: Large Scale Analytics and Machine Learning at Indeed

1,674 views

9 0 SHARE ...



Imhotep

Fields Options

company_category_code (44) count() ▼

Total	48,449
1 software	6,959
2 biotech	6,474
3 mobile	3,173

indeed[®]
one search. all jobs.

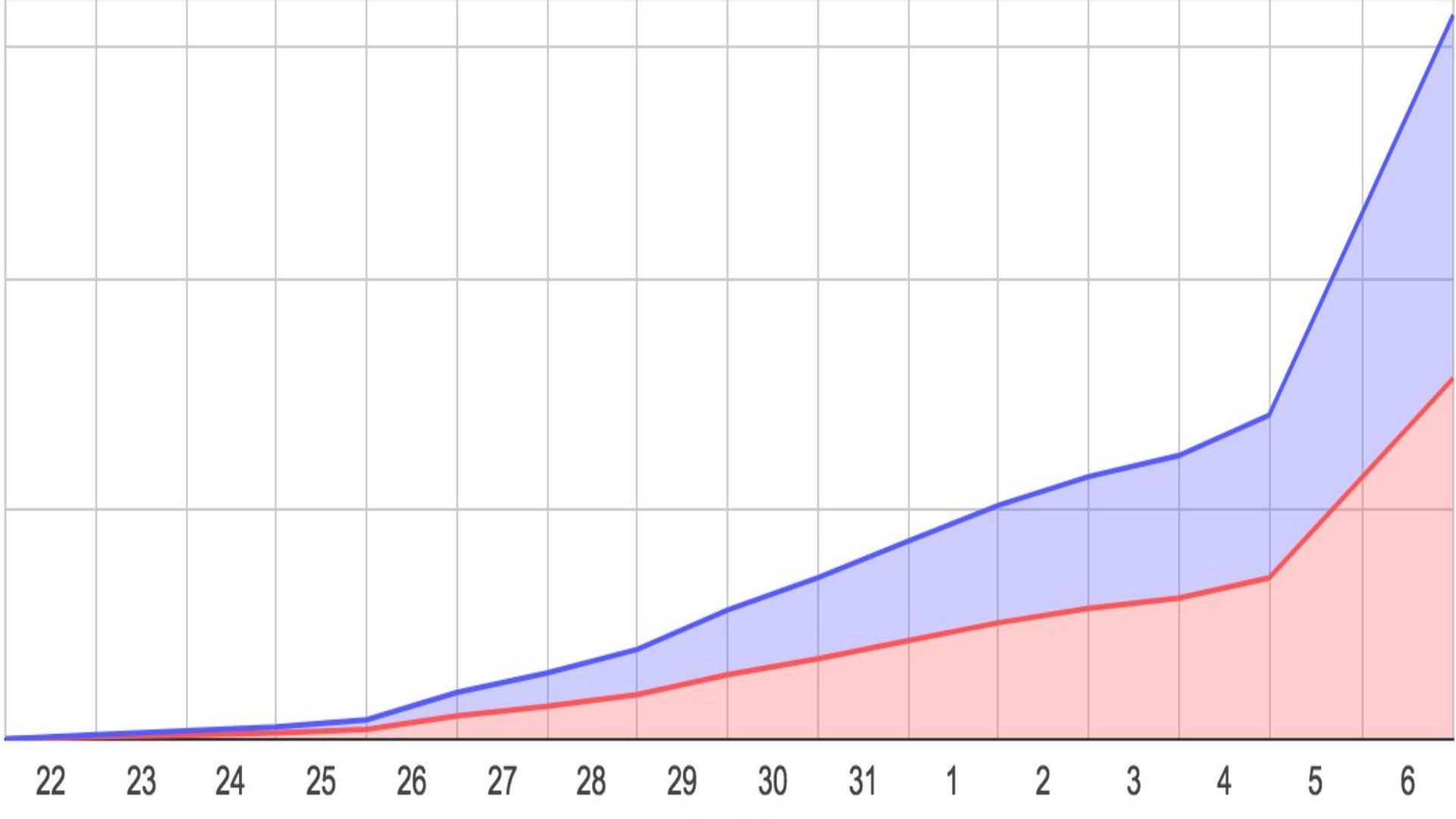
[@IndeedEng]: Large scale interactive analytics with Imhotep

1,950 views

1 1 1 SHARE ...

talks.indeedeng.io






Make it so

1. Recruited engineers to work on it *part-time*.
2. Cleaned up the internal dependencies.
3. Refactored on an open source fork.
4. Added some AWS CloudFormation scripts to make deployment easy.
5. Wrote lots of documentation.
6. October 2014: Declared it ready, flipped the “Public” switch on the repos.
7. And...

We held a local workshop



The screenshot shows the Imhotep interactive analytics interface. The URL bar displays a complex query: `demo.imhotep.works/iq/[#q]=from+wikipedia+2014-09-18+2014-09-19+group+by+title[1000+by+numRequests]+select+numRequests&view=ta...`

The interface includes a query builder with the following fields:

- from:** wikipedia
- timerange:** 2014-09-18 to 2014-09-19
- where:** (empty)
- group by:** title[1000 by numRequests]
- select:** numRequests

Buttons for **Run**, **TSV**, **CSV**, **Bash script**, **Help**, and **Settings** are visible.

On the left, a list of fields is shown under "Hide Fields": allbit, bytesServed, categories, linksOut, numRequests, title, titleWords, unixtime.

The results table displays the following data:


#	title[1000 by numRequests]	numRequests
1	numRequests	212,353,063
2	Undefined	1,738,274
3	undefined	1,558,187
4	Less_(Unix)	737,617
5	Scottish_independence_referendum,_2014	248,039
6	Carrie_Buck	236,148
7	English_alphabet	217,292
8	Alphabet	215,495
9	Schizophrenia_and_smoking	100,806

The Indeed logo and tagline "one search. all jobs." are visible at the bottom left of the video frame.


@IndeedEng: Interactive Analytics with Imhotep
2,814 views

10 2 SHARE ...

We talked at a conference (GASP!)




O'REILLY
OSCON
OPEN SOURCE CONVENTION



MAY 8-11, 2017
AUSTIN, TX
oscon.com
#oscon


Imhotep: scalable, **efficient**, fast



Most popular dataset: 39B events, 384 fields

- 5.7TB on disk (146 bytes/event)

No need to sample — use all the data



We got our first *known* external user

"Imhotep is a fast and powerful tool for analyzing extremely large datasets. I've found no other open source or paid alternatives that match Imhotep's speed, flexibility, and ease of collaboration. I'm glad I can bring the same kind of deep understanding and data-driven decision making to Hometogo."

- Tom Bergman, VP of Product at Hometogo



We got our first *known* external user

"Imhotep is a fast and powerful tool for analyzing extremely large datasets. I've found no other open source or paid alternatives that match Imhotep's speed, flexibility, and ease of collaboration. I'm glad I can bring the same kind of deep understanding and data-driven decision making to Hometogo."

**- Tom Bergman, VP of Product at Hometogo;
former Product Manager at Indeed**

A screenshot of a tweet from @IndeedEng. The tweet content is a video player showing Tom Bergman using a laptop. The laptop screen displays the Imhotep interface with a table of data. The table has columns for 'id', 'category', 'numRequests', 'title', and 'url'. The data rows show various categories and their corresponding request counts.

#	category	numRequests	title	url
1	undefined	1,738,274	undefined	undefined
2	undefined	1,558,187	undefined	undefined
3	Less_10%	737,617	Scandinavia_independence_referendum_2014	Scandinavia_independence_referendum_2014
4	Career_Buzz	246,039	English_alphabet	English_alphabet
5	Alphabet	215,495	Alphabet	Alphabet

@IndeedEng: Interactive Analytics with Imhotep
2,814 views

MISTAKE

Failed to offer easy
ways to evaluate and
adopt

LESSON

Invest in ease of
evaluation and
adoption,
as a feature

MISTAKE

Under-invested in
promotion and
community building

LESSON

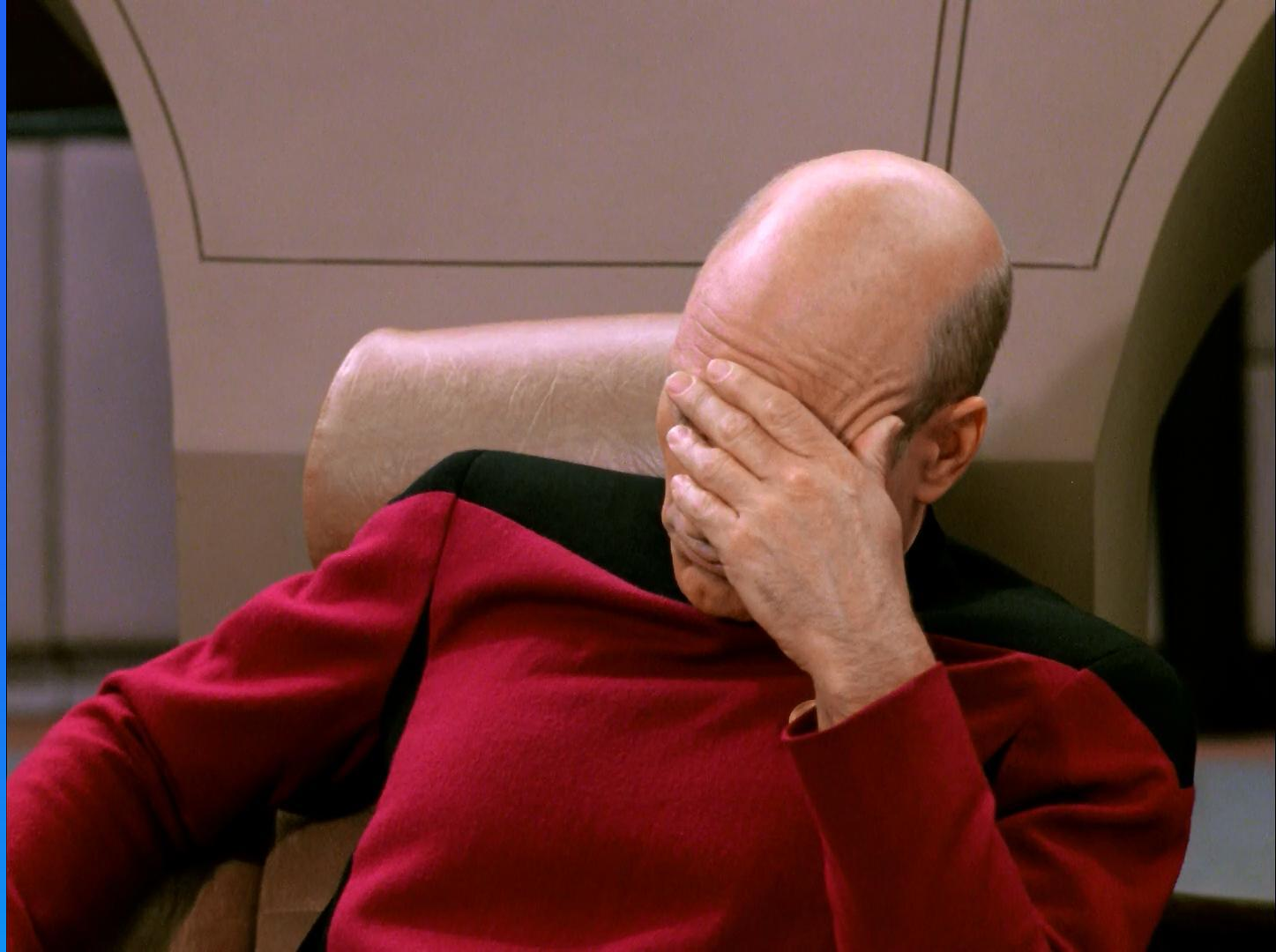
Invest in building a
community of users
and contributors

MISTAKE

Developed OSS
release behind
closed doors
(and again, an
external fork)

LESSON

Work in the open
(and use your own
OSS release)



Chapter 4

Time for an OSPO

Jack's "duh" moment

Me:

"Hmm. Maybe I need people who understand open source better than I do working on our open source program full-time.

Hey, boss -- can I hire someone to build a staffed open source program?"

Boss:

"Yeah, sure, sounds good."

Me:

"Cool. I'll get right on that."

Initial job description

We are seeking an engineering leader to develop and lead our open source program... As the Open Source Engineering Lead, you will:

- Coordinate teams across Indeed Engineering working on open source projects
- Promote our projects to the appropriate communities
- Cultivate an open source mindset at Indeed
- Establish best practices and process for releases
- Guide and review documentation
- Manage individual project priorities (feature requests and fixes)
- Review and triage public issues, questions, and pull requests
- Develop and integrate software to support the open source release process
- Survey Indeed's internal technologies and recommend new candidate projects

Initial job description

We are seeking an engineering leader to develop and lead our open source program... As the Open Source Engineering Lead, you will:

- Coordinate teams across Indeed Engineering working on **open source projects**
- Promote **our projects** to the appropriate communities
- Cultivate an open source mindset at Indeed
- Establish best practices and process for **releases**
- Guide and review **documentation**
- Manage individual **project priorities** (feature requests and fixes)
- **Review and triage** public issues, questions, and pull requests
- Develop and integrate software to support the open source **release process**
- Survey Indeed's internal technologies and recommend new **candidate projects**

Got advice from the community

Insight:

We could achieve our goals for Open Source in other ways.
Releasing our own tech was just one of multiple strategies.

"INDEED, YOU NEED AN OSPO"



duaneobrien.github.io/Talks/OSPOPitch



Duane O'Brien

Head of Open Source at Indeed.com



0 to 4
and
growing!

Software Engineer - Open Source

Indeed ★★★★★ 306 reviews

San Francisco, CA

As a Software Engineer on the Open Source team, you will.... As



CLOUD NATIVE
COMPUTING FOUNDATION

APACHE[®]
SOFTWARE FOUNDATION



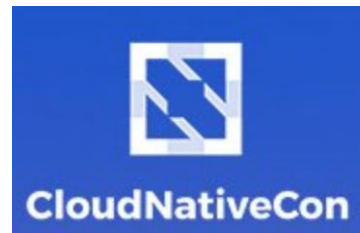
python SOFTWARE
FOUNDATION



webpack

foss
asia

O'REILLY[®]
OSCON



MISTAKE

Neglected other
strategies for
achieving Open
Source goals

LESSON

An Open Source
Program can run
multiple strategies to
achieve your goals

Epilogue

Let's Review

Lessons

Work in the open

Use OSS release internally

Validate need, support early adopters,
and **promote**

Invest in building a **community**
of users and contributors

Invest in ease of
evaluation and adoption

Pursue other strategies
to achieve your goals

Lessons

Work in the open

Use OSS release internally

Validate need, support early adopters,
and **promote**

Invest in building a **community**
of users and contributors

Invest in ease of
evaluation and adoption

Pursue other strategies
to achieve your goals

Lessons

Work in the open

Use OSS release internally

Validate need, support early adopters,
and **promote**

Invest in building a **community**
of users and contributors

Invest in ease of
evaluation and adoption

Pursue other strategies
to achieve your goals

Lessons

Work in the open

Use OSS release internally

**Validate need, support early adopters,
and promote**

Invest in building a **community**
of users and contributors

Invest in ease of
evaluation and adoption

Pursue other strategies
to achieve your goals

Lessons

Work in the open

Use OSS release internally

Validate need, support early adopters,
and **promote**

**Invest in building a community
of users and contributors**

Invest in ease of
evaluation and adoption

Pursue other strategies
to achieve your goals

Lessons

Work in the open

Use OSS release internally

Validate need, support early adopters,
and **promote**

Invest in building a **community**
of users and contributors

**Invest in ease of
evaluation and adoption**

Pursue other strategies
to achieve your goals

Lessons

Work in the open

Use OSS release internally

Validate need, support early adopters,
and **promote**

Invest in building a **community**
of users and contributors

Invest in ease of
evaluation and adoption

**Pursue other strategies
to achieve your goals**

Lessons

Work in the open

Use OSS release internally

Validate need, support early adopters,
and **promote**

Invest in building a **community**
of users and contributors

Invest in ease of
evaluation and adoption

Pursue other strategies
to achieve your goals

Any questions?

@youknowjack | jackhumphrey.me

@IndeedEng | opensource.indeedeng.io