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
Engineering offices

- Seattle
- San Francisco
- Austin
- Hyderabad
- Tokyo
- Singapore
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!
<table>
<thead>
<tr>
<th>Questions to ask</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Why do you want to open source?</td>
</tr>
<tr>
<td>2. Can you commit the necessary resources?</td>
</tr>
<tr>
<td>3. Do you know the common pitfalls?</td>
</tr>
</tbody>
</table>
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 ⇒ 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?

@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 best user experience and growth has been swift and steady. Traffic to Indeed continues to increase by millions per month. 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.

[link]

go.indeed.com/boxcar
...or 3 years to present about it

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

go.indeed.com/proctor-talk
We got an early adopter

Homeaway started using it... and contributed some code!
We got an unsolicited PR from a stranger

Exposé raw JSON matrix for consumption by other servers

#2

chriscolon merged 1 commit into [indeedeng:master](https://example.com) from [benjamingochran:master](https://example.com) on Dec 1, 2014

Conversation 1Commits 1Checks 0Files changed 1

benjamingochran commented on Nov 24, 2014

This is meant to be used with clients implementing UrlProctorLoaders.
We got Google traffic

Google search results for "java a/b testing"

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

Proctor is an 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 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
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
We talked at a conference (GASP!)

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
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”
Software Engineer - Open Source

Indeed ★★★★★ 306 reviews
San Francisco, CA
As a Software Engineer on the Open Source team, you will.... As
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
<table>
<thead>
<tr>
<th>Lessons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work in the open</td>
</tr>
<tr>
<td>Use OSS release internally</td>
</tr>
<tr>
<td>Validate need, support early adopters, and <strong>promote</strong></td>
</tr>
<tr>
<td>Invest in building a <strong>community</strong> of users and contributors</td>
</tr>
<tr>
<td>Invest in ease of evaluation and adoption</td>
</tr>
<tr>
<td>Pursue other strategies to achieve your goals</td>
</tr>
<tr>
<td>Lessons</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td><strong>Work in the open</strong></td>
</tr>
<tr>
<td>Validate need, support early adopters, and <strong>promote</strong></td>
</tr>
<tr>
<td>Invest in ease of evaluation and adoption</td>
</tr>
<tr>
<td>Use OSS release internally</td>
</tr>
<tr>
<td>Invest in building a <strong>community</strong> of users and contributors</td>
</tr>
<tr>
<td>Pursue other strategies to achieve your goals</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Work in the open</th>
<th>Use OSS release internally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validate need, support early adopters, and <strong>promote</strong></td>
<td><strong>Invest in building a community of users and contributors</strong></td>
</tr>
<tr>
<td>Invest in ease of evaluation and adoption</td>
<td>Pursue other strategies to achieve your goals</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Work in the open</th>
<th>Use OSS release internally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validate need, support early adopters, and <strong>promote</strong></td>
<td>Invest in building a <strong>community</strong> of users and contributors</td>
</tr>
<tr>
<td>Invest in ease of evaluation and adoption</td>
<td>Pursue other strategies to achieve your goals</td>
</tr>
</tbody>
</table>
Any questions?

@youknowjack | jackhumphrey.me

@IndeedEng | opensource.indeedeng.io