Case Study: Adopting SRE Principles at Stack Overflow

Tom Limoncelli, SRE
Stack Exchange, Inc. @YesThatTom
Tom’s SRE Creds

- Google SRE from 2006-2012

- Worked mostly on internal infrastructure and virtualization services (Ganeti) as well as external services like Blog Search.

- Joined Stack Exchange, Inc. in June 2012.
Who is Tom Limoncelli?

Google SRE 2006-2013.

Author, blogger, speaker.

Sysadmin since 1988. Worked at Bell Labs and many smaller companies

SRE at Stack Exchange, Inc

serverfault.com / stackoverflow.com
WHAT IS STACK EXCHANGE, INC?

- 140+ Question and Answer Web sites:
  - StackOverflow (coders)
  - Serverfault (sysadmins)
  - Over 130 more: Grammar, Math, Physics, Poker, Pets, Aviation, each major religion

- Stack Overflow Careers
- Helping the best developers and sysadmins find jobs that don’t suck.
Q: What can I feed a vegetarian at my Pesach seder?
Wait, wait, it's not a question for Seasoned Advice. :-) I have a vegetarian (not vegan) coming to my seder this year. Most of the vegetarian staples (as far as I know) are chametz or kitniyot ... in the context of the seder -- e.g. made ahead and held, not prepared immediately before serving (because we'll all be busy with the seder). What kinds of vegetarian main dishes satisfy all the Pesach ...

Q: Eating in a non-kosher household [duplicate]
to be able to go eat at my parent's for dinner, but I know they do not keep Kosher, be it keeping separate dishes/sinks or even checking labels. Can I eat at their house if my mother promises to cook a vegetarian, kosher-style meal? ...

Q: Eating dairy after pareve on meat equipment?
Oreos are made that's different from, say, my pot of vegetarian chili. So if I've got this vegetarian chili made in a meat pot, can I have ice cream for dessert? (This isn't a p'sak question, even if I get an answer before lunch.) ...

Q: Can a person convert if they are living with a non-Jewish person?
compromises would the Beth Din allow the person to convert? The flatmate is vegetarian so that's one less thing to worry about. We will however have separate dishes - my sets for meat, dairy and pareve, and theirs.
Snake not interested in food

So I have a female Californian king snake and I'm starting to get worried about her not eating.

First of all she has not eaten in 4 months. It's winter at the moment and I know they do not eat at winter as it happened last year with her not eating for an extended period but it was more like 3 months. I keep her vivarium temperature is between 25-32C.

She spends most of the time sleeping, when she isn't she is very, very active feeling around her enclosure.

She handles extremely well, however I'm worried this could be part of the main problem she is completely non-aggressive even when I attempt to feed her, she just seems to want to get out of her feeding container and has no interest in the mouse what so ever.

I handled her today and she feels lighter and thinner.

If anyone can help I'd really appreciate it.
Keywords | Location
--- | ---
python | 94041

116 jobs for “python” near Mountain View, CA

🌟 Software Engineer (Python/Django)
Cowayl • South San Francisco, CA • OFFERS RELOCATION
Our operations are powered by our own custom-built infrastructure on a Python/Django/Postgres...
python, django, postgresql, computer-science, robotics

🌟 Django/Python Developer
Perkville • Oakland, CA
We’re looking for a full-stack engineer fluent in Python, experienced with Django, to join our...
django, mysql, python, javascript, ini

🌟 Python / PyQt Software Engineer
ToyTalk, Inc. • San Francisco, CA
This authoring system, called PullString, is written in Python and PyQt, binds to several...
python, pyqt, c++, swig, osx

🌟 Python Developer
YouGov • Palo Alto, CA • WORK REMOTELY
YouGov seeks a full-time, full stack Python developer to join the team building our survey...
python, postgresql, django, rabbitmq, mongodb

🌟 Senior Platform Engineer @ Scribd
Scribd • San Francisco, CA
An automated classification system for automatically optimizing across numerous machine...
ruby-on-rails, javascript, query, python, mysql
Stack Exchange is...

- Smaller than you’d think:
  - Dozens of machines, not hundreds

- Mixed Windows and Linux
- Rendering and DB: C#, IIS, SQLServer
- Linux: Cache, Load Bal, Sec, Search, +
ASSESSMENT SO FAR
(i.e. How do we stack up?)
SITE RELIABILITY PRACTICES (BEN TREYNOR’S SREcon14 Keynote)

1. Hire only coders.
2. Have an SLA for your service.
3. Measure and report performance against the SLA.
4. Use Error Budgets and gate launches on them.
5. Have a common staffing pool for SRE and Developers.
6. Have excess Ops work overflow to the Dev team.
7. Cap SRE operational load at 50 percent.
8. Share 5 percent of Ops work with the Dev team.
9. Oncall teams should have at least eight people at one location, or six people at each of multiple locations.
10. Aim for a maximum of two events per oncall shift.
11. Do a postmortem for every event.
12. Postmortems are blameless and focus on process and technology, not people.
1. Hire only coders.
2. Have an SLA for your service.
3. Measure and report performance against the SLA.
4. Use Error Budgets and gate launches on them.
5. Have a common staffing pool for SRE and Developers.
6. Have excess Ops work overflow to the Dev team.
7. Cap SRE operational load at 50 percent.
8. Share 5 percent of Ops work with the Dev team.
9. Oncall teams should have at least eight people at one location, or six people at each of multiple locations.
10. Aim for a maximum of two events per oncall shift.
11. Do a postmortem for every event.
12. Postmortems are blameless and focus on process and technology, not people.
2015

1. Hire only coders.
2. Have an SLA for your service.
3. Measure and report performance against the SLA.
4. Use Error Budgets and gate launches on them.
5. Have a common staffing pool for SRE and Developers.
6. Have excess Ops work overflow to the Dev team.
7. Cap SRE operational load at 50 percent.
8. Share 5 percent of Ops work with the Dev team.
9. Oncall teams should have at least eight people at one location, or six people at each of multiple locations.
10. Aim for a maximum of two events per oncall shift.
11. Do a postmortem for every event.
12. Postmortems are blameless and focus on process and technology, not people.
Management/budget policies that put value on operations:
5. Have a common staffing pool for SRE and Developers.

Handle Outages/Incidents Awesomely
10. Aim for a maximum of two events per oncall shift.
11. Do a postmortem for every event.
12. Postmortems are blameless and focus on process and technology, not people.

SLA-Driven Operations & Monitoring
2. Have an SLA for your service.
3. Measure and report performance against the SLA.
4. Use Error Budgets and gate launches on them.

Control Operations Overload
1. Hire only coders.
6. Have excess Ops work overflow to the Dev team.
7. Cap SRE operational load at 50 percent.
8. Share 5 percent of Ops work with the Dev team.
9. Oncall teams should have at least eight people at one location, or six people at each of multiple locations.
...THE EASY ONES...
**Management/Budget Dynamics:**

5. Have a common staffing pool for SRE and Developers.
Handle Outages/Incidents Awesomely

10. Aim for a maximum of two events per oncall shift.

11. Do a postmortem for every event.

12. Postmortems are blameless and focus on process and technology, not people.
WHY IMPORTANT?

• Postmortems are how we learn from outages.
• Each outage educates the team.
• The team then educates the entire company.
• The net result: The organization gets smarter over time.

• We must create an environment where people have time to PM every event.
I dunno about anybody else, but I really like getting these post-mortem reports. Not only is it **nice to know what happened**, but it’s also great to see **how you guys handled it in the moment** and **how you plan to prevent these events** going forward. Really neato. Thanks for the great work :)

—-Anna
• Template in Appendix D

• Web search [John Allspaw Postmortem]
...THE HARDER ONES...
SLA-Driven Operations & Monitoring

2. Have an SLA for your service.

3. Measure and report performance against the SLA.

4. Use Error Budgets and gate launches on them.
Defining SLAs is difficult because...

- they are complex
- it means talking with product management

- Important because SLAs are fundamental to managing a service.
Monitoring

- Collect a time-series of measurements.
- NOT “up/down” monitoring.
- Example:
  03:04:12:00 server1.cpu.loadave 3.1
  03:04:17:00 server1.cpu.loadave 3.0
  03:04:22:00 server1.cpu.loadave 2.9
  03:04:27:00 server1.cpu.loadave 0.1
• Up/down monitoring == bad
  • You get alerted when the system is down.
  • The outage is “built in”

• Time-series == good:
  • Alerting when “sick”, not “down”.
  • Prevents outages
  • Long-term history needed for planning.
STATUS

• Dissatisfied with existing monitoring tools.

• So we built our own: [www.bosun.org](http://www.bosun.org)
  • The first monitoring system with an IDE for designing complex alerts!
  • Uses OpenTSDB for storage
    • Rewrote agent in Go, works w/Linux+Win
  • Can use Graphite as a back-end (NEW!)
STATUS (2)

• We made it easy for Devs to specify what is collected and define alerts.
• They think it's fun!
• Found+fixed many previously unknown performance issues.
• Other departments have made more interesting rules than we have.
• SLAs: In process of defining.
• Time series rules
• www.bosun.org
SRE IS CONCERNED WITH...

UNRESPONSIVE speed and performance
UNSTABLE server availability
OVERLOADING server upgrades
UNAVAILABLE hardware upgrades
availability
SLA

speed and performance
security
capacity planning
software upgrades
hardware upgrades
availability

You don’t know when to stop optimizing and/or spending.

You know the gaps.
You can prioritize.
You know when to stop!
It isn't a service if it isn't monitored. If there is no monitoring then you're just running software.
It isn't a service if it isn't monitored against an SLA. If there is no SLA-based monitoring then you're just running software.
SRE is SLA-driven system administration
EXAMPLE: CDN DEPLOYMENT

• Enable proxy-style CDN for all content. Better TCP handshake, compression, improvement over HTTP.
• Ooops! Performance got worse!
• Bug fix involved vendor code overhaul
• Why didn’t other customers notice this?
• Isn’t anyone doing before&after comparison?
• Measure twice, cut once!
CONTROL OPERATIONS OVERLOAD

1. Hire only coders.

6. Have excess Ops work overflow to the Dev team.

7. Cap SRE operational load at 50 percent.

8. Share 5% Ops work w/ the Dev team.

9. Oncall teams.. 6-8 ppl/location
ENDING

TRANSACTIONAL SYSTEM ADMINISTRATION
TRANSACTIONAL SYSTEM ADMINISTRATION

• Users make requests
• Sysadmins fulfill requests

• Bad because:
  • Wrong relationship dynamic
  • Scales by hiring more people (yuck!)
You don’t process transactions.
You run the system that processes transactions.

Better because:
- Scales
- SLAs
- Better service for customers
- Less boring for the sysadmins
Example 1

• We didn’t create monitoring rules for Devs.
• We created and run a system that lets Devs be self-sufficient

• Better because:
  • Devs have better knowledge of internals
  • “Workforce multiplier”
Example 2 (Very “Enterprise-Y”)

- SREs shouldn’t install and deliver laptops to users.
- SREs should maintain the system that:
  - Knows when a new laptop is needed:
    - Tracks HR database for new employees
    - Emails users when eligible for refresh
    - A portal for special ordering machines
  - Automates OS installation for use by a “laptop delivery crew” in each building. (Non-technical clerks and IT coordinator)
  - Does capacity planning, generates purchase orders, etc. for the purchasing/finance group.
- Better because:
  - OMG do I have to explain?
STATUS

• Stack Exchange SRE:
  • Got rid of the SRE “ticket system”
  • Dedicated SRE “point person(s)” on each development team
  • SRE-related work is scheduled and assigned to SRE in a cooperative manner (agile-style).

• SREs have more time for service development, optimization, projects, etc.
IN CONCLUSION

• SLAs SLAs SLAs SLAs
• End transactional system administration
• Spread this philosophy to the enterprise
• Try www.Bosun.org

• Q&A
ADDITIONAL RESOURCES

The Practice of Cloud System Administration
www.informit.com/TPOSA
Discount code TPOSA35

- Ch 7: Defines SRE
- Ch 14: Postmortems
- Ch 15: Drills/DiRT/Game Days
- Ch 16-17: Monitoring
- Ch 18: Capacity Planning
- Ch 19+20+App A: SLAs/KPIs

Our SRE team has a blog: blog.serverfault.com
Follow me on Twitter: @YesThatTom
We’re hiring in NYC! stackexchange.com/work-here

I’ll be the speaker at these Meetups this week:
Tuesday, 6pm — Large-Scale Production Eng Meetup (at URL Cafe, Yahoo! Sunnyvale)
Thursday, 7:30pm sharp! — BayLISA (at LinkedIn office, Sunnyvale)
IN CONCLUSION

• SLAs SLAs SLAs SLAs
• End transactional system administration
• Spread this philosophy to the enterprise
• Try www.Bosun.org

• Q&A