Performance optimization in Linux

Tales from the trenches

Alex Chistyakov, Principal Engineer, Git in Sky

Linux Piter 2015

Git in Sky
TUNING YOUR INFRASTRUCTURE

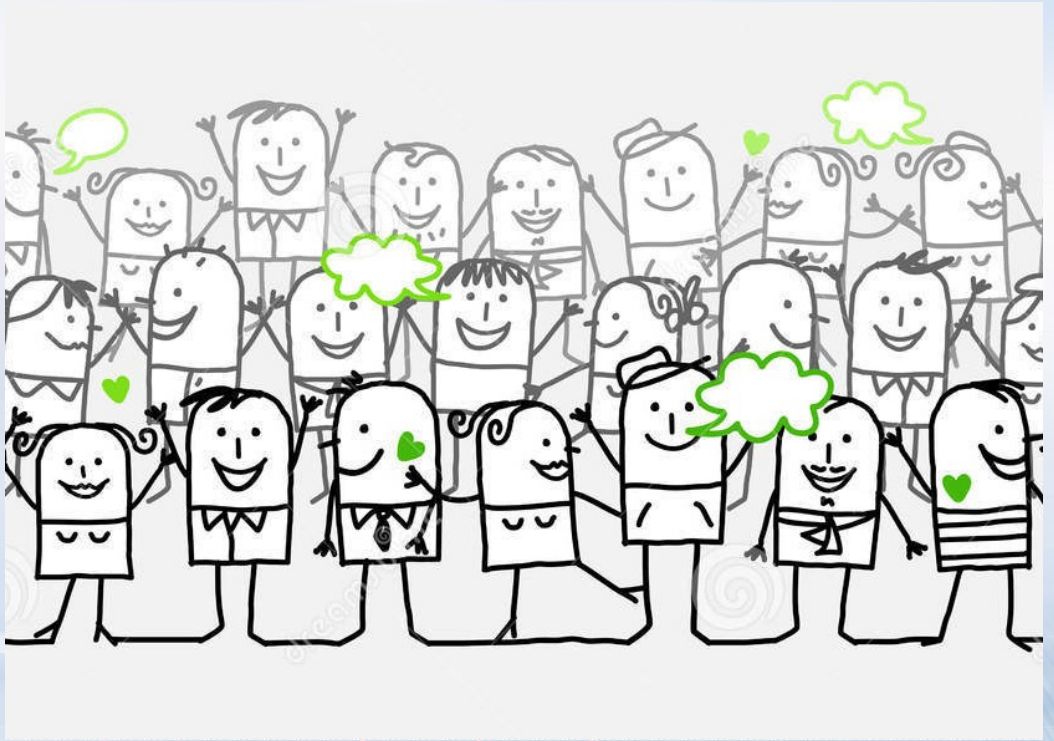
Who are we?

- A small consulting company based in SPb.
- Web operations
- Automation
- Performance tuning
- Sponsors of local DevOps meetup

Git in Sky
TUNING YOUR INFRASTRUCTURE

Who are you?

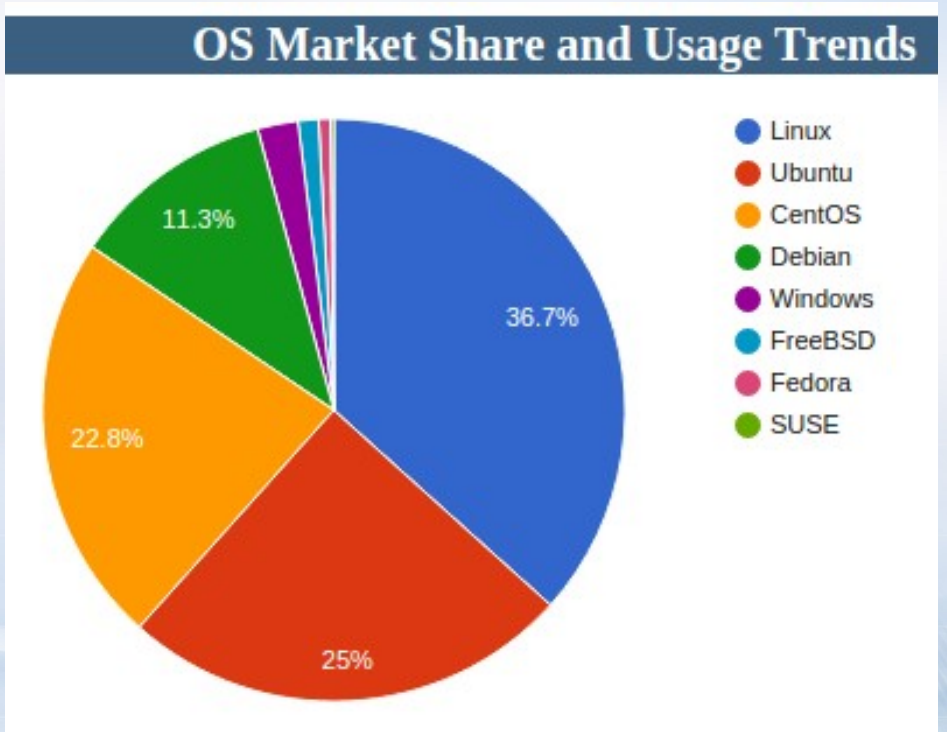
- Linux fans?
- Developers?
- Web developers, maybe?
- System architects?
- Performance engineers?



Git in Sky
TUNING YOUR INFRASTRUCTURE

Okay, why Linux?

- Is there anything else?
- According to W3Cook stats, Linux serves 95.8% of public web sites
- And it's on the desktop too!
- (At least on my desktop)



Git in Sky
TUNING YOUR INFRASTRUCTURE

Linux in the perfect world

It's just you. **http://gitinsky.com** is up.

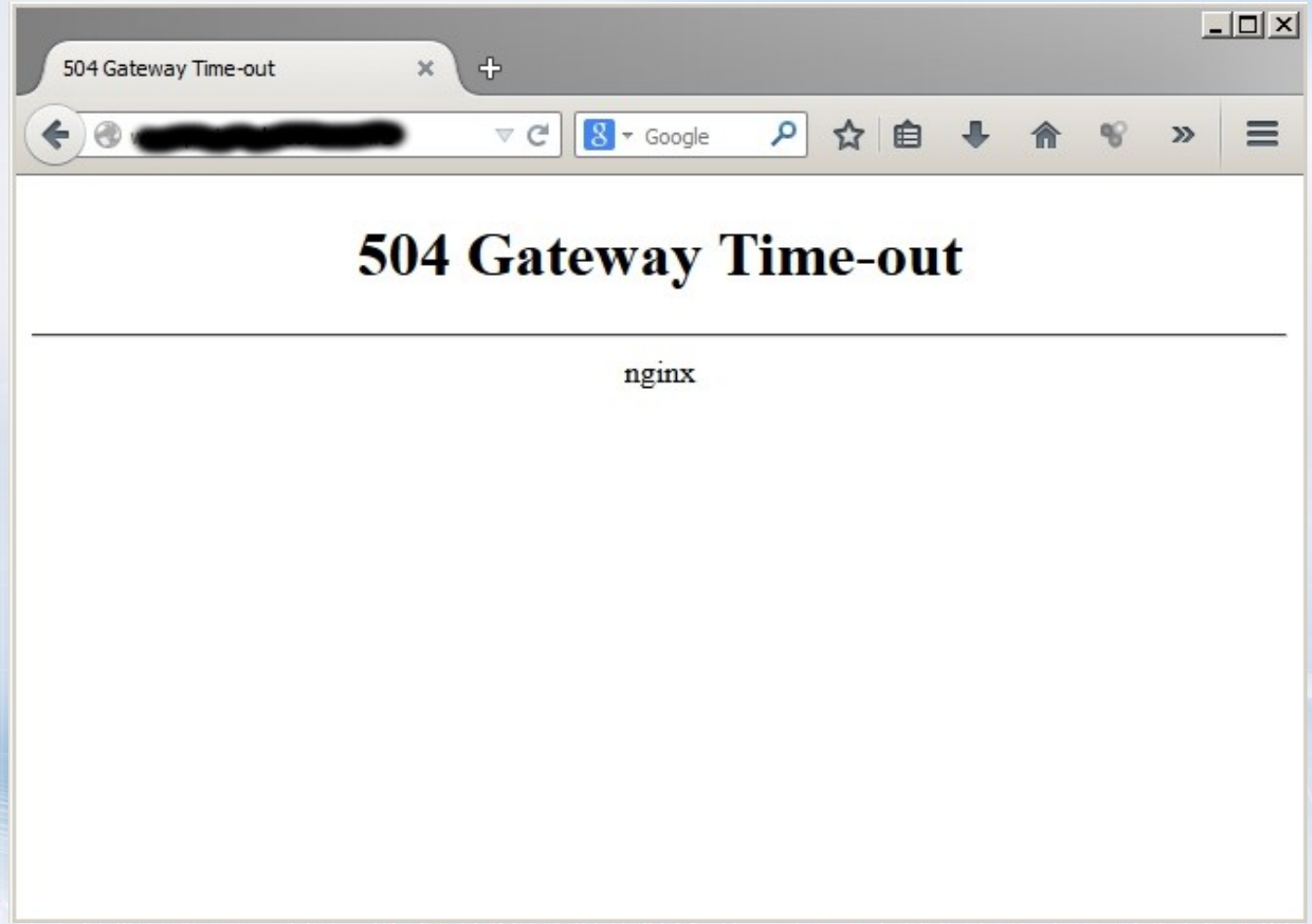
Check another site?

Looking for great web hosting?
Move to SiteGround and get the best!

Short URL at isup.me

Git in Sky
TUNING YOUR INFRASTRUCTURE

Linux in the real world



Git in Sky
TUNING YOUR INFRASTRUCTURE

504 on the main page!

- A customer is stressed extremely
- Reaction should be quick and effective
- The obvious plan does not work
- We should be prepared!

Git in Sky
TUNING YOUR INFRASTRUCTURE

The obvious plan

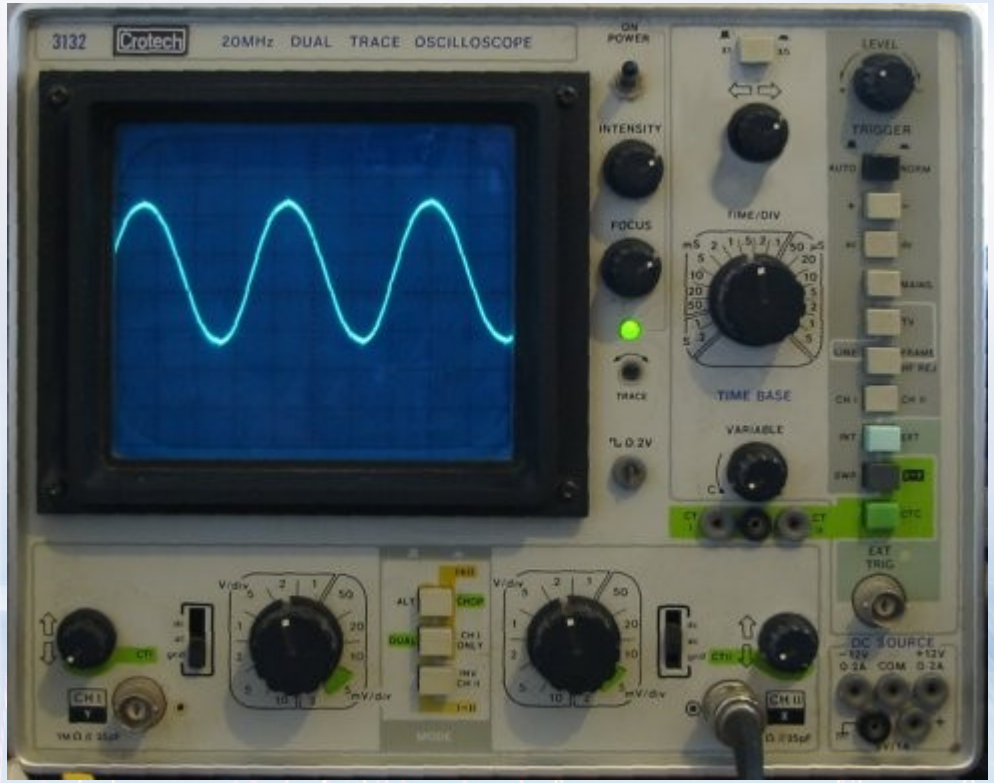
- 1) Change something
- 2) Expect the situation to become better
- 3) Wait anxiously
- 4) ????
- 5) PROFIT!!!
- This plan is quite popular in fact

  for some reason (simplicity?)

Git in Sky
TUNING YOUR INFRASTRUCTURE

The proper plan (top secret!)

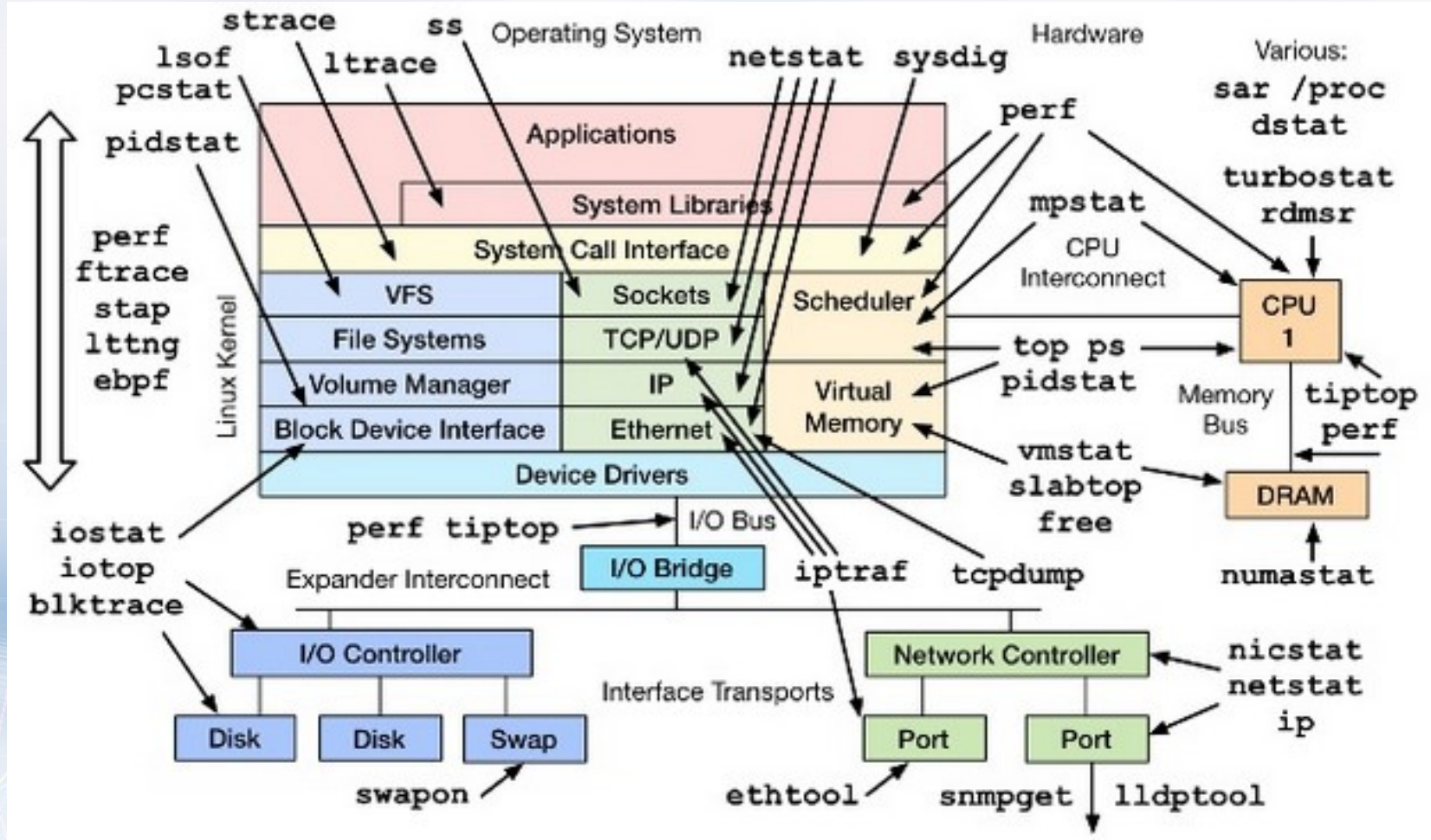
- 1) Gather metrics (you have them already, don't you?)
- 2) Analyze metrics
- 3) Elaborate a hypothesis
- 4) Plan and make a single change
- 5) Repeat until success
- If you were proficient in physics at high school, this plan should sound extremely familiar

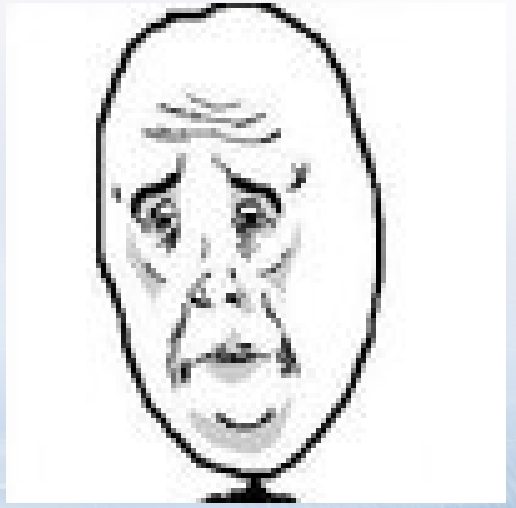


Git in Sky
TUNING YOUR INFRASTRUCTURE

How to collect/analyze metrics?

- Brendan Gregg's observability tools diagram





Git in Sky
TUNING YOUR INFRASTRUCTURE

How do we collect/analyze?

- `atop` (30 sec intervals)

Git in Sky
TUNING YOUR INFRASTRUCTURE

How do we collect/analyze?

- `atop` (30 sec intervals)
- Ex-graphite stack (Grafana, InfluxDB or OpenTSDB or Cyanite/Cassandra, **but NOT Whisper**, collectd)

Git in Sky
TUNING YOUR INFRASTRUCTURE

How do we collect/analyze?

- `atop` (30 sec intervals)
- Ex-graphite stack (Grafana, InfluxDB or OpenTSDB or Cyanite/Cassandra, **but NOT Whisper**, collectd)
- NewRelic

Git in Sky
TUNING YOUR INFRASTRUCTURE

How do we collect/analyze?

- `atop` (30 sec intervals)
- Ex-graphite stack (Grafana, InfluxDB or OpenTSDB or Cyanite/Cassandra, **but NOT Whisper**, collectd)
- NewRelic
- `pidstat` (not `iotop`)

Git in Sky
TUNING YOUR INFRASTRUCTURE

How do we collect/analyze?

- `atop` (30 sec intervals)
- Ex-graphite stack (Grafana, InfluxDB or OpenTSDB or Cyanite/Cassandra, **but NOT Whisper**, collectd)
- NewRelic
- `pidstat` (not `iotop`)
- `perf top` and `perf record`

Git in Sky
TUNING YOUR INFRASTRUCTURE

How do we collect/analyze?

- `atop` (30 sec intervals)
- Ex-graphite stack (Grafana, InfluxDB or OpenTSDB or Cyanite/Cassandra, **but NOT Whisper**, collectd)
- NewRelic
- `pidstat` (not `iotop`)
- `perf top` and `perf record`
- `sysdig`

Git in Sky
TUNING YOUR INFRASTRUCTURE

How do we collect/analyze?

- `atop` (30 sec intervals)
- Ex-graphite stack (Grafana, InfluxDB or OpenTSDB or Cyanite/Cassandra, **but NOT Whisper**, collectd)
- NewRelic
- `pidstat` (not `iotop`)
- `perf top` and `perf record`
- `sysdig`
- `iostat -x 1`

Git in Sky
TUNING YOUR INFRASTRUCTURE

Most common case (a no-brainer)

- CPU time is too high (a lucky customer got an SSD)
- Disk saturation is above 60% (a not-so-lucky customer)

Git in Sky
TUNING YOUR INFRASTRUCTURE

Most common case (a no-brainer)

- CPU time is too high (a lucky customer got an SSD)
- Disk saturation is above 60% (a not-so-lucky customer)
- PHP and, of course, MySQL

Git in Sky
TUNING YOUR INFRASTRUCTURE

Most common case (a no-brainer)

- CPU time is too high (a lucky customer got an SSD)
- Disk saturation is above 60% (a not-so-lucky customer)
- PHP and, of course, MySQL
- InnoDB buffers are too low
- Synchronous commit is 'on'

Git in Sky
TUNING YOUR INFRASTRUCTURE

Most common case (a no-brainer)

- CPU time is too high (a lucky customer got an SSD)
- Disk saturation is above 60% (a not-so-lucky customer)
- PHP and, of course, MySQL
- InnoDB buffers are too low
- Synchronous commit is 'on'
- Too many slow queries
- Queries with 'filesort' in execution plan

Git in Sky
TUNING YOUR INFRASTRUCTURE

How to solve

- Install Anemometer, turn on slow queries log
- Range queries based on their cumulative exec time
- Read and understand execution plans
- Blame developers
- Cry in vain

Git in Sky
TUNING YOUR INFRASTRUCTURE

Story time!

- Case #1: measuring the measurer
- It seems that everybody still uses Graphite/Whisper

Git in Sky
TUNING YOUR INFRASTRUCTURE

Story time!

- Case #1: measuring the measurer
- It seems that everybody still uses Graphite/Whisper
- Even the big guys like Mail.Ru

Git in Sky
TUNING YOUR INFRASTRUCTURE

Story time!

- Case #1: measuring the measurer
- It seems that everybody still uses Graphite/Whisper
- Even the big guys like Mail.Ru
- Well, because SAS HDDs are cheap...

Git in Sky
TUNING YOUR INFRASTRUCTURE

Story time!

- Case #1: measuring the measurer
- It seems that everybody still uses Graphite/Whisper
- Even the big guys like Mail.Ru
- Well, because SAS HDDs are cheap…
- But...

Git in Sky
TUNING YOUR INFRASTRUCTURE

A challenger appears!

- InfluxDB vs. Whisper, July 2015
- The same set of metrics (carbon-relay-ng in the middle)
- And the winner is...

Git in Sky
TUNING YOUR INFRASTRUCTURE

Okay, we hate magic

- Whisper is just a set of RRD-like files on a plain old FS
- 20000 metrics lead you to 20000 files
- Accessing 20000 files every 10 secs is a major pain
- InfluxDB is a time series database based on an LSM-tree
- InfluxDB is much more write-optimized than 20000 separate files on your ext4/XFS/you-name-it
- But, of course, SAS drives are quite cheap...

Git in Sky
TUNING YOUR INFRASTRUCTURE

In case you are not scared

- Case #2: the site got a SUDS (sudden unexpected death syndrome)

Git in Sky
TUNING YOUR INFRASTRUCTURE

In case you are not scared

- Case #2: the site got a SUDS (sudden unexpected death syndrome)
- Symptoms: everything slows down to a crawl (sounds familiar)

Git in Sky
TUNING YOUR INFRASTRUCTURE

In case you are not scared

- Case #2: the site got a SUDS (sudden unexpected death syndrome)
- Symptoms: everything slows down to a crawl (sounds familiar)
- NewRelic shows nothing unusual

Git in Sky
TUNING YOUR INFRASTRUCTURE

In case you are not scared

- Case #2: the site got a SUDS (sudden unexpected death syndrome)
- Symptoms: everything slows down to a crawl (sounds familiar)
- NewRelic shows nothing unusual
- Well, if parameters more suitable for a busy site that for a very low traffic one can be called “nothing unusual”
- And this site is not busy at all

Git in Sky
TUNING YOUR INFRASTRUCTURE

Diagnostic card

- PHP is OK
- MySQL does not sort anything
- Top queries in MySQL sorted by total exec time are all indexed
- Every MySQL query runs very slow when there is even moderate load

Git in Sky
TUNING YOUR INFRASTRUCTURE

But how did we solve it?

- Even a modern rig w/decent Xeons and SATA HDDs can be turned into a slug
- As simple as disabling AHCI in BIOS and staying on plain IDE
- Well this one was not that hard but was quite unusual
- Rented servers do not suffer from problems like this because they are configured uniformly
- I can't easily explain how I came upon this solution, pure intuition seemed to be involved

Git in Sky
TUNING YOUR INFRASTRUCTURE

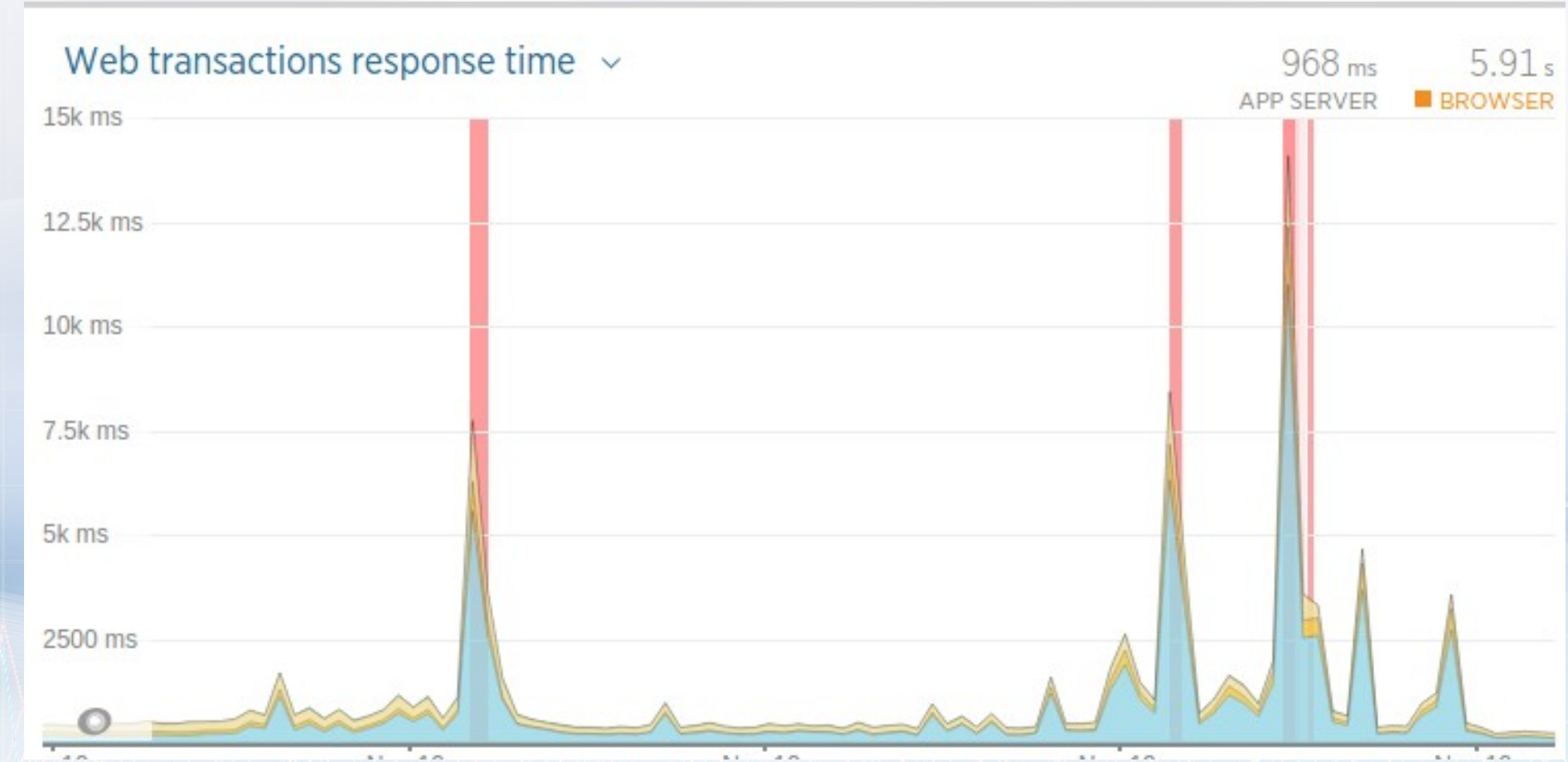
Not scared enough yet?

- Then, case #3: another site got SUDS

Git in Sky
TUNING YOUR INFRASTRUCTURE

Not scared enough yet?

- Then, case #3: another site got SUDS



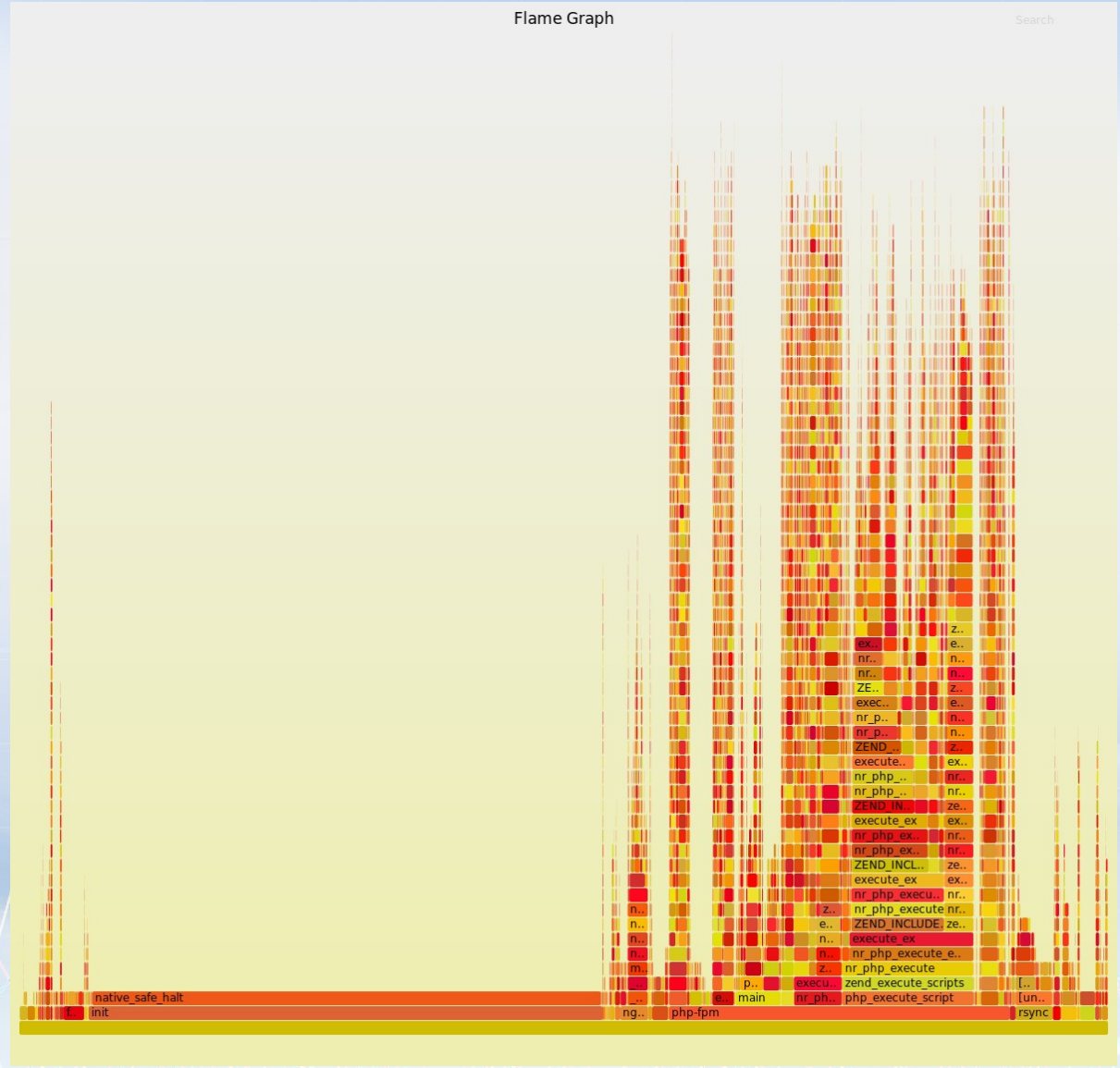
Git in Sky
TUNING YOUR INFRASTRUCTURE

Diagnostic card

- NewRelic blames PHP code
- Even the SSH console is slow
- Nothing unusual or unexpected in daily CPU load graphs
- CPU flamegraph shows nothing

Git in Sky
TUNING YOUR INFRASTRUCTURE

What is a «CPU flamegraph»?



Git in Sky
TUNING YOUR INFRASTRUCTURE

How did we solve it

- Analyzed atop recorded stats for outage periods
- atop is quite smart in fact and color suspected values in red or blue
- IRQ % is over 50%

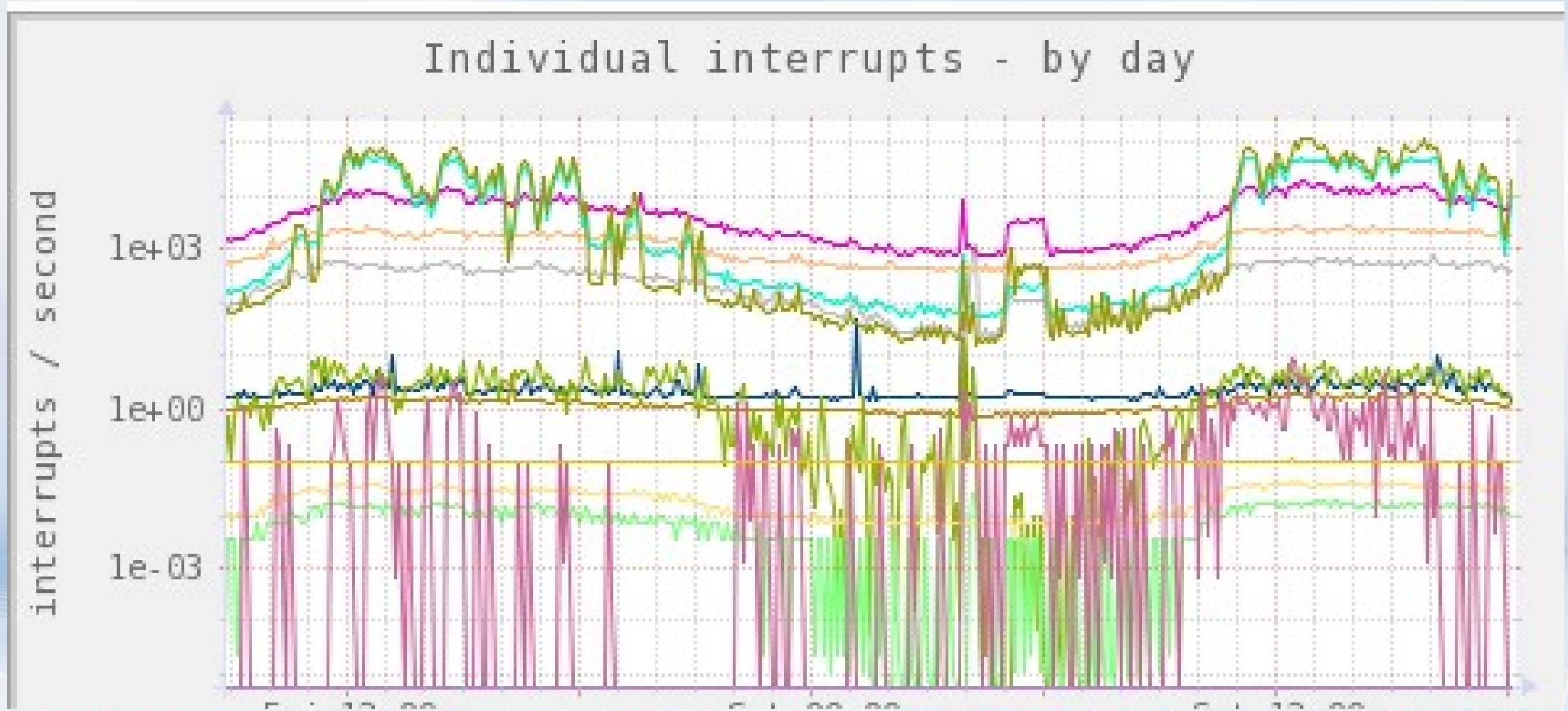
Git in Sky
TUNING YOUR INFRASTRUCTURE

How did we solve it

- Analyzed atop recorded stats for outage periods
- atop is quite smart in fact and color suspected values in red or blue
- IRQ % is over 50%
- But what is “IRQ %” anyway?
- Oh, who cares, let's install Munin and get per-interrupt graphs

Git in Sky
TUNING YOUR INFRASTRUCTURE

A blast from the past



Git in Sky
TUNING YOUR INFRASTRUCTURE

How did we solve it

- Well we have not solved it yet
- The graph from previous slide is for past two days
- But at least we have a plan!
- https://help.ubuntu.com/community/ReschedulingInterrupts

Git in Sky
TUNING YOUR INFRASTRUCTURE

Summary

- Linux is cool
- Performance engineering is hard
- Don't panic!

Git in Sky
TUNING YOUR INFRASTRUCTURE

Thank you!

- Questions?
- Oh, BTW you can hire us!
- http://gitinsky.com
- alex@gitinsky.com
- Please do not forget to attend our meetups:
- http://meetup.com/Docker-Spb, http://meetup.com/Ansible-Spb, http://meetup.com/DevOps-40