History of Prometheus

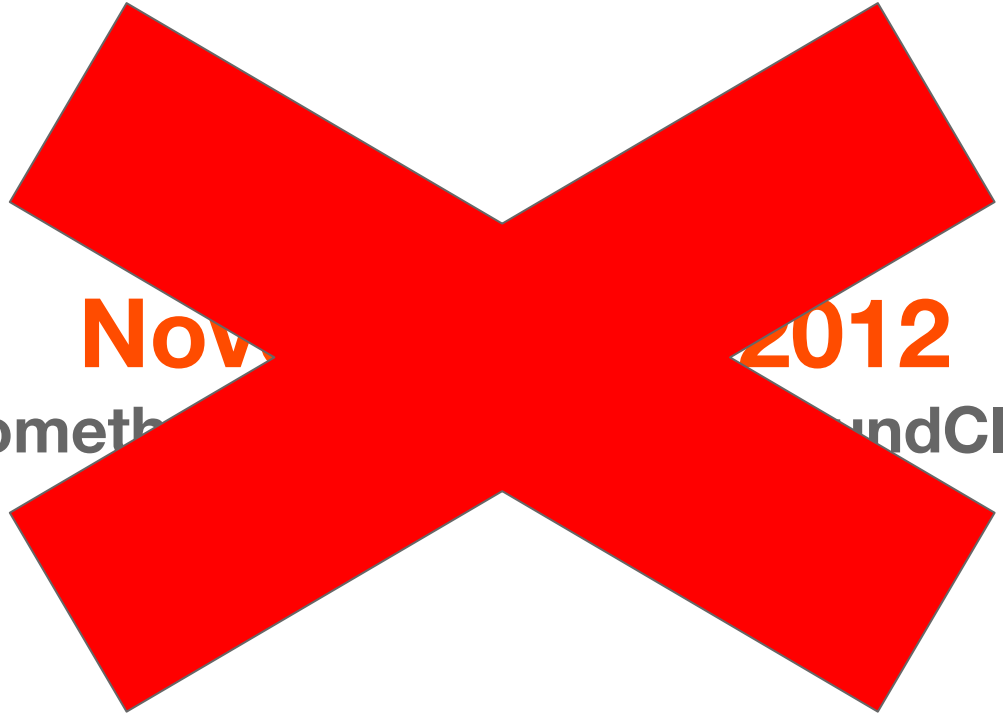
at SoundCloud

Tobias Schmidt - PromCon August 26, 2016

github.com/grobie - @dagrobie

November 24, 2012

Prometheus was started at SoundCloud



Nov 2012

Prometh ndCloud

Beginnings

- Prometheus was started by Matt T. Proud before he joined SoundCloud
- Huge need for a great open-source time-series monitoring system
  - Not limited to infrastructure or even software monitoring
  - Econometrics, bio-chemical, environmental, all kind of sensors, …
  - Only OpenTSDB (complex) and RRDtool based (insufficient)
- Research started in February 2012
  - C, C++, Java, Go
  - Cassandra, LevelDB
- Oldest repository is client_golang from March 2012
- Server in Go started in August 2012, public repo November 2012
- Matt joined SoundCloud September 2012
- Julius joined SoundCloud October 2012

Goals

- Faceted data / Metrics with label dimensions
- Powerful and flexible query language
- Operationally very simple
  - No complex orchestration
  - No dependencies
- Strong interfaces
  - Clearly defined ingestion formats
  - Typed query language
  - Rules, storage, query API
- Proven, mature technologies
  - Protocol buffers, LevelDB
  - Go was actually a gamble

Julius' reaction

"You're crazy!
But this is pretty cool, too."

Julius Volz, October 2012

2012

SoundCloud before Prometheus

Infrastructure

- Very advanced: container orchestration system
  - Bazooka
  - Fast deployment of new services
  - Highly dynamic deploys
- Huge monolithic Rails application
- Most other services were not in critical path
  - But there were already over 200 repositories
- Mostly common web service architecture
  - Load balancers, application servers, caches, databases

Monitoring

- Most of the available open-source tools
  - Nagios, Graphite, StatsD, Ganglia, Cacti, Munin
- Typical Rails SaaS tools
  - NewRelic, Airbrake
- On-call
  - One ops team had the pager
  - First infrastructure teams going on-call
  - Almost exclusively blackbox alerting
- Problems
  - Very little instrumentation
  - No per-instance application metrics at all
  - StatsD and Graphite scalability issues

Prometheus

- First prototypes exist
- Matt and Julius begin working on different parts in isolation
- Matt
  - Works on server and client_golang
- Julius
  - Implements configuration, query language and web interface
- They put it together - And it works!

2013

Introduction of Prometheus and service discovery

Early 2013

- January
  - Still a side project
  - Created prometheus-developers@googlegroups.com
  - Acquired prometheus.io
  - A test instance is already continuously running
  - Bazooka instrumentation with Prometheus - first metrics!
- February
  - Internal tech-talk
  - Start allocating work time for Prometheus
  - client_java is born
- Definition of client exposition formats (JSON, protobuf)

Service Discovery

- So far
  - Statically configured DNS names
  - Driven by and coupled to configuration management (Chef)
  - Service-to-service communication over load-balancers
  - No consistency in naming
  - Undescriptive: `bazooka-lb.r.internal.soundcloud.com:8043`
- DNS service discovery
  - One system to discover Chef and Bazooka services
  - One naming: `<service>.<job>.<env>.<product>.<zone>`
  - Prometheus support! No more static configs
- Lesson learned:
  - Don’t put team or product dimensions into your routing scheme

The sad moment

Matt leaves SoundCloud October 2013

Highlights 2013

- Service discovery
- SoundCloud
  - Transitioned to “You build it, you own it, you’re on-call for it”
  - First teams start writing runbooks for their services
  - Björn joined in October! But worked on other projects first.
- Prometheus
  - Clients for Go, Java and Ruby exist
  - Alertmanager was born in July
  - Work on PromDash starts in August (Grafana doesn’t exist yet)
  - 8 teams use Prometheus already, 10 servers
  - No point releases yet (git sha based continuous deploys)

2014

Maturation

Pushgateway

- Pull vs. Push - A religious debate.
- Need to monitor ephemeral jobs (Map/Reduce on Hadoop, chef-client, ...)
- Work on Pushgateway starts in January

“It is explicitly not an aggregator, but rather a metrics cache [...]”

Bernerd Schaefer, January 2013

Prometheus usage outside of SoundCloud

- #prometheus on freenode gets created on February 25th
- First external users:
  - Johannes Ziemke works now for Docker and introduces Prometheus
  - Brian Brazil leaves Google and joins Boxever, learns about Prometheus from Björn

“If it didn’t exist I would have created it.”

Brian Brazil

Prometheus brought *too much* fire

- The first (and only *knock on wood*) outage caused by Prometheus
- Pull makes it easy to run additional test Prometheus servers all the time
- `/metrics.json` gets renamed to `/metrics`

```
    resp, _ := t.httpClient.Do(req)
+   if resp.StatusCode != http.StatusOK {
+       return fmt.Errorf("server returned HTTP status %s", resp.Status)
+   }
    defer resp.Body.Close()
```

- Connection doesn't get closed, targets run out of file descriptors
- Bazooka managers can't receive heartbeats from containers anymore
  - Marks all containers as lost

Lessons learned

- Availability checking is hard - be conservative
- Instrument everything
  - Missing whitebox metrics and alerting on critical components
  - Process collector is born, providing `process_open_fds` among others
- Prevent alerting fatigue
  - Page was ignored as it was similar to constant alert noise
- “Measure twice, cut once” during outages.

Highlights 2014

- SoundCloud
  - Full transition to micro-service architecture undergo
  - Causes many instabilities, lack of metrics becomes apparent
  - All new services have Prometheus instrumentation
- Prometheus
  - New text exposition format (deprecated JSON format)
  - Huge storage rewrite (from LevelDB to chunks on filesystem)
  - First exploration of long-term storage (OpenTSDB)
  - *Lot's* of performance improvements
  - Server releases 0.1.0 - 0.8.0

2015

Prometheus grows big

Public release

January 26, 2015

“Prometheus is now a standalone open-source project and maintained independently of any company.”

Highlights 2015

- SoundCloud
  - Fabian Reinartz joins SoundCloud to work full-time on Prometheus
  - All services have Prometheus instrumentation and alerts
  - Prometheus indispensable for operations (still many instabilities)
- Prometheus
  - Unexpected growth and adoption by other companies
  - Histograms
  - New service discoveries (file, consul, marathon, kubernetes, ec2, …)
  - Double-Delta encoding v1
  - Server releases 0.9.0 - 0.16.1 (no semver yet)

2016

Current state

Latest developments

- SoundCloud
  - Migrated to Grafana (dashboard templating!)
  - New services come with monitoring batteries included (graphs+alerts)
  - Noise free alerting thanks to Alertmanager
  - Very stable - reached goal of 99.9% availability every month
- Prometheus
  - Variable bit-wide encoding v2
  - Prometheus joins the CNCF on May 9th
  - Prometheus reaches 1.0.0 and promises API stability on July 18th
  - First Prometheus Conference in Berlin on August 25th - 26th

Incoming Request Rate (HTTP & Thrift)

19K rps

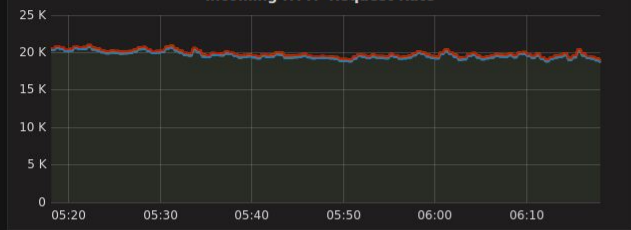
99th Percentile Latency

162.6 ms

Error Rate

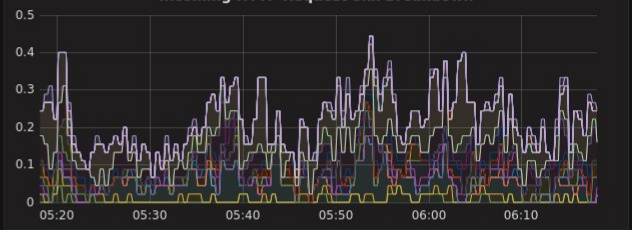
0.24 rps

Incoming HTTP Request Rate

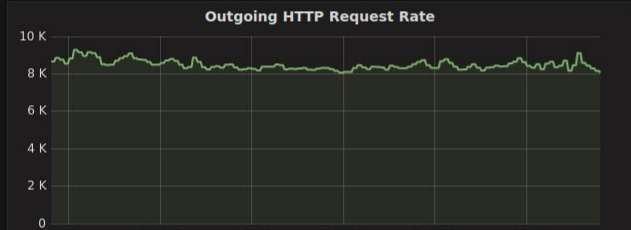


Incoming HTTP Request Latency

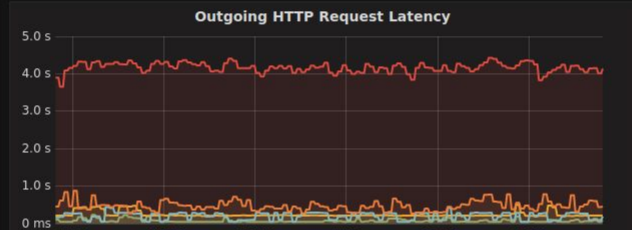
Incoming HTTP Request 5xx Breakdown



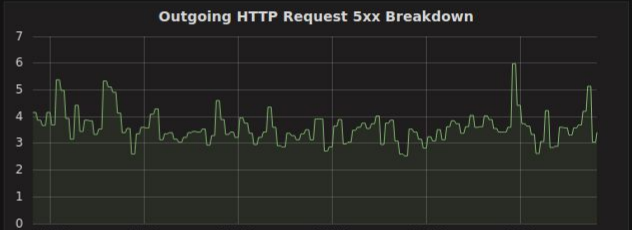
Outgoing HTTP Request Rate



Outgoing HTTP Request Latency

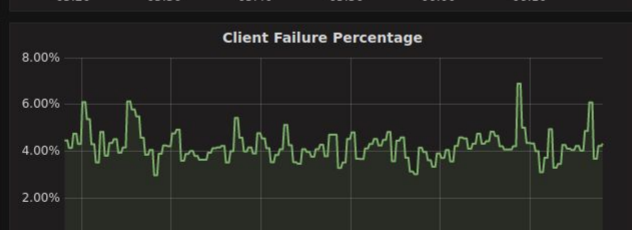


Outgoing HTTP Request 5xx Breakdown

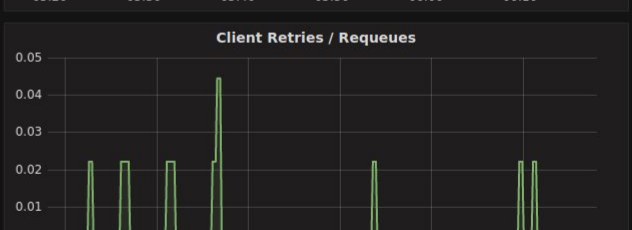


Client Failure Accruals

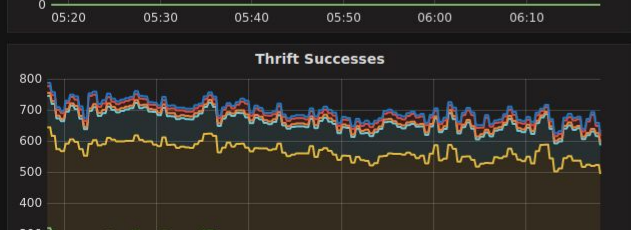
Client Failure Percentage



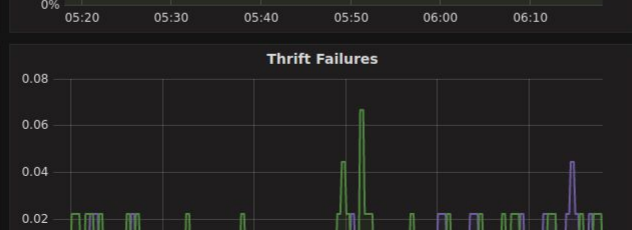
Client Retries / Requeues



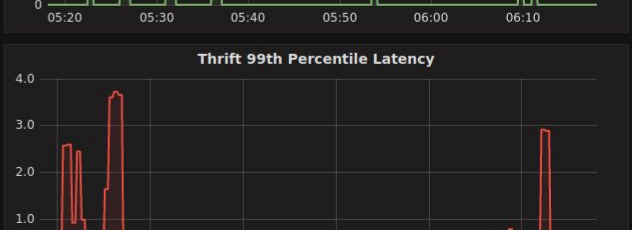
Thrift Successes



Thrift Failures



Thrift 99th Percentile Latency



Some things we have learned

- Hard work pays off!
  - Even if you don’t think you can make it.
- Go was a perfect choice
  - Huge community
  - Many natural integrations (Bazooka, Kubernetes, Etcd)
  - Great language features (concurrency, compilation time, single binary)
- Don’t deliver features at the compromise of your goals!
- Provide clear documentation around your vision and goals
  - We might have needed to think about a governance structure earlier

CHRONOLOGICAL
NON-EXHAUSTIVE

Acknowledgements

- Matt T. Proud - The father of Prometheus.
- Julius “jrv” Volz - `ETooManyThings`, The mother of Prometheus.
- Bernerd Schaefer - server, Pushgateway
- Johannes “fish” Ziemke - service discovery, operation
- Stuart “stn” Nelson - PromDash
- Björn “beorn7” Rabenstein - storage, client_golang
- Brian “bbrazil” Brazil - `ETooManyThings`, keeping us on track
- Fabian “fabxc” Reinartz - The coding machine.
- Ben Kochie, Matthias Rampke, Richi Hartmann, Jimmy Dyson, Steve Durrheimer, *and so many, many more!*

Thank you

Tobias Schmidt - PromCon August 26, 2016

github.com/grobie - @dagrobie

SOUNDCLOUD