Prometheus

What the hype is about

Richard Hartmann,
RichiH@{freenode,OFTC,IRCnet},
richih@{debian,fosdem,richih}.org,
@TwitchiH

2019-04-09

Lucky me...

This was supposed to happen in a room, not on the show floor, sorry about the audio.

...at least these are text slides...

On the other hand, I always try to have some time for Q&A; this might not work here...

whoami

- Richard ”RichiH” Hartmann
- Swiss army chainsaw at SpaceNet
- Project lead for building one of the most modern datacenters in Europe
- Debian Developer
- FOSDEM, DebConf, DENOGx, PromCon staff
- Prometheus team member

Show of hands

- Who has heard of Prometheus?
- Who is considering to use Prometheus?
- Who is POCing Prometheus?
- Who uses Prometheus in production?

Prometheus 101

- Inspired by Google's Borgmon
- Time series database
- unit64 millisecond timestamp, float64 value
- Instrumentation & exporters
- Not for event logging
- Dashboarding via Grafana

Main selling points

- Highly dynamic, built-in service discovery
- No hierarchical model, n-dimensional label set
- PromQL: for processing, graphing, alerting, and export
- Simple operation
- Highly efficient

Working assumptions & concepts

- Prometheus is a pull-based system
- Black-box monitoring: Looking at a service from the outside (Does the server answer to HTTP requests?)
- White-box monitoring: Instrumention code from the inside (How much time does this subroutine take?)
- Every service should have its own metrics endpoint
- Hard API commitments within major versions
- No built-in TLS yet, use reverse proxies for now

Time series

- Time series are recorded values which change over time
- Individual events are usually merged into counters and/or histograms
- Changing values are recorded as gauges
- Typical examples
  - Access rates to a webserver (counter)
  - Temperatures in a datacenter (gauge)

Efficiency

- 1,000,000+ samples/second no problem on currect hardware
- 200,000 samples/second/core
- 16 bytes/sample compressed to 1.36 bytes/sample
- Cheap ingestion & storage means more data for you

Exposition format

```
http_requests_total{env="prod",method="post",code="200"} 1027
http_requests_total{env="prod",method="post",code="400"} 3
http_requests_total{env="prod",method="post",code="500"} 12
http_requests_total{env="prod",method="get",code="200"} 20
http_requests_total{env="test",method="post",code="200"} 372
http_requests_total{env="test",method="post",code="400"} 75
```

PromQL vs SQL

```
avg by(city) (temperature_celsius{country="germany"})

SELECT city, AVG(value) FROM temperature_celsius WHERE \
 country="germany" GROUP BY city

rate(errors{job="foo"}[5m]) / rate(total{job="foo"}[5m])

SELECT errors.job, errors.instance, [...more labels...], \
 rate(errors.value, 5m) / rate(total.value, 5m) \
 FROM errors JOIN total ON [...all label equalities...] \
 WHERE errors.job="foo" AND total.job="foo"
```

Grafana

- Supports dozens of data sources
- Modern UI
- Allows for complex data manipulation and visualization
- Native Prometheus support
- New feature: Interactive exploration of Prometheus data

Borg

- Kubernetes is Borg
- Prometheus is Borgmon
- Google couldn't have run Borg without Borgmon
- (We will ignore Omega & Monarch in the context of this talk)
- Kubernetes & Prometheus are designed and written with each other in mind

Cloud

- Metrics are a sweet spot in visibility and scalability
  - Observability also means being able to handle the amounts of data
- n-dimensional label sets are what enable cloud scale metrics collection
  - Alternative: Have fun extracting knowledge from a monitoring data lake
- Prometheus handles churn better than a lot of other monitoring software
  - A nameless company has a pod lifetime of 15m; if they want to run analysis over the last two weeks...

Anthos ;)

- Running Prometheus is simple, but its HA story is very opinionated
- Two friendly forks tackle this on different levels:
  - Cortex: Horizontally scalable index
  - Thanos: Horizontally scalable storage
  - Ongoing to merge the two together: Corthanos
  - Might become a build-time flag in Prometheus proper at some point?
- Running Cortex and Thanos is a bit like running K8s; consider aaS

Scale

Modern workloads are impossible to handle without modern observability tooling

Scale

And it's likely not your core business to handle metrics

Thanks!

Thanks for listening!

Questions?

See slide footer for contact info.