zalando

Zalandos Quest for Operating 10K Micro Services

DevOpsCon Berlin - June 2022

Heinrich Hartmann - Head of SRE - Zalando SE

about me

Head of SRE @ zalando

DataScientist @ CIRCONUS

Mathematician @ The University of Bonn
Doctor of Philosophy (PhD), Matchematik
2008 - 2011

Recent Talks / Publications

- How to measure Latency (P99 Conf 21)
- State of the Histogram (SLOConf 2021)
- Statistics for Engineers (2014..2019)
- Latency SLOs Done Right (FOSDEM 2019)
- Circllhist - A Histogram Data Structure of IT Operations (arxiv)

zalando

Zalando Business

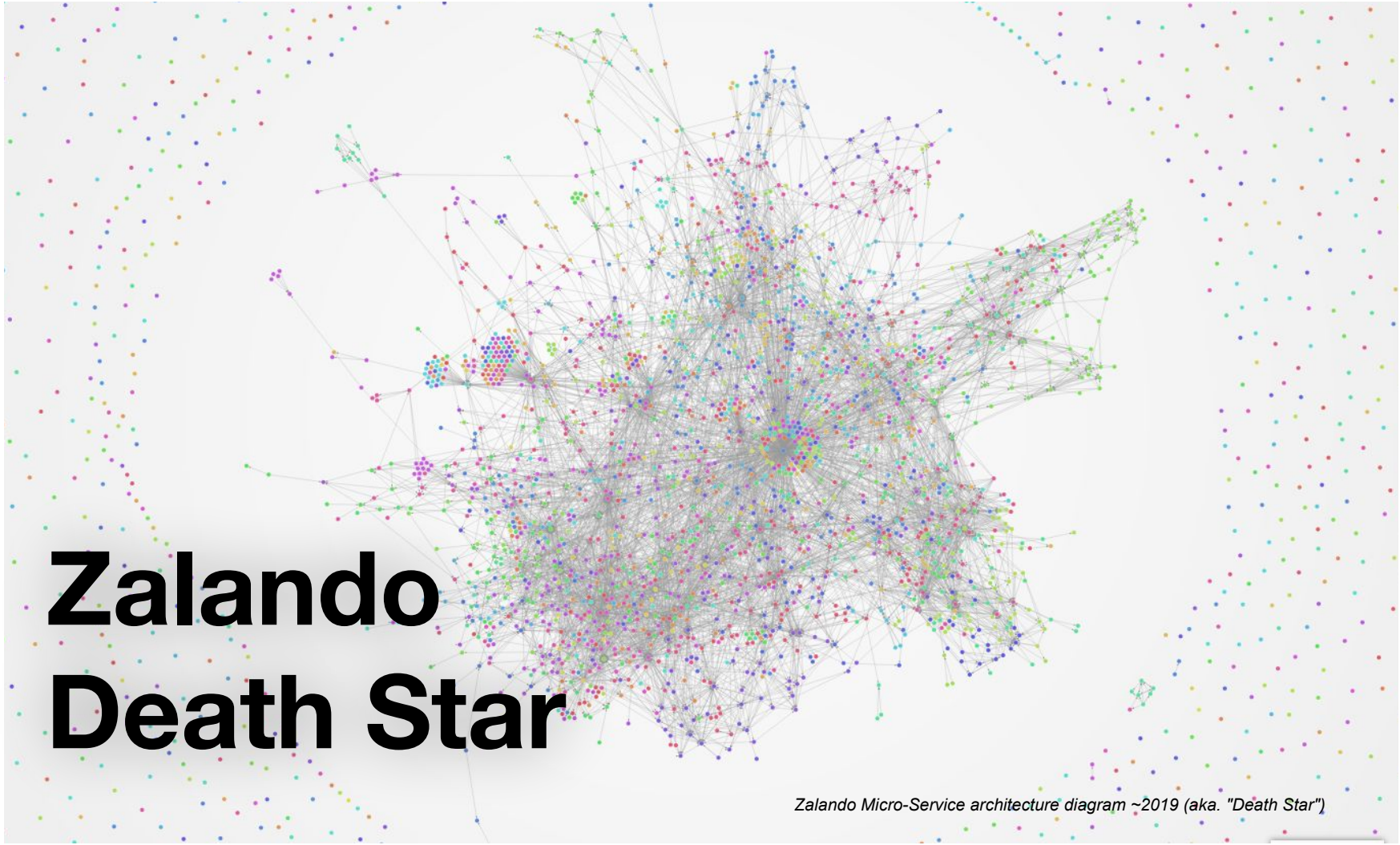
- Largest Fashion Retailer in EU
- 10B+ Annual Revenue
- 50M+ active Customers
- 23+ Countries
- 17K Employees

zalando

Zalando Tech

- 2.500+ SWE on Staff
- 200+ teams
- In AWS Frankfurt
- Up to 10K EC2 nodes
- 200+ k8s clusters
- 5K+ Micro Services

- Internal Platform providing
  - Managed k8s
  - Managed Postgres
  - Managed Kafka
  - Managed ML Infrastructure
  - ...

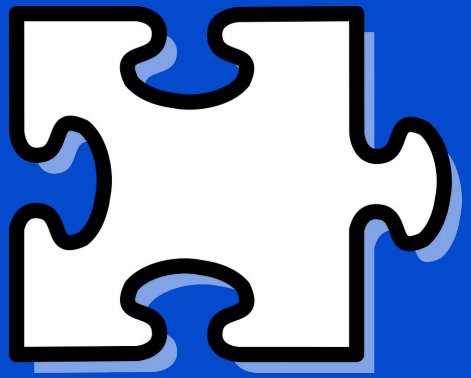


Zalando Micro-Service architecture diagram ~2019 (aka. "Death Star")

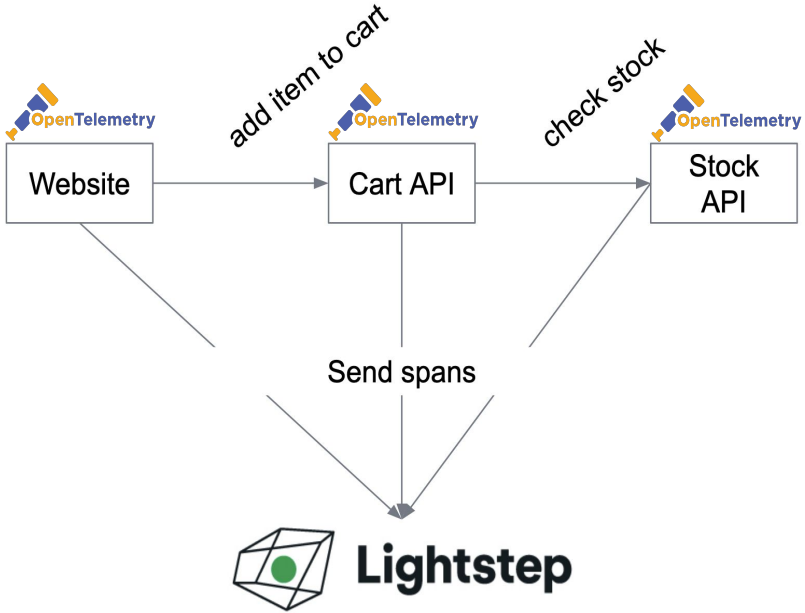
Service Diagramming Exercise @ Zalando ~2017

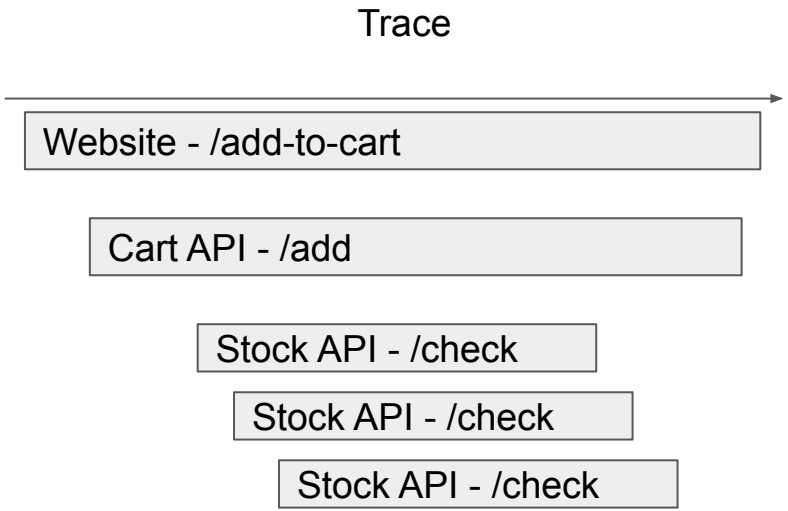
zalando

Distributed Tracing

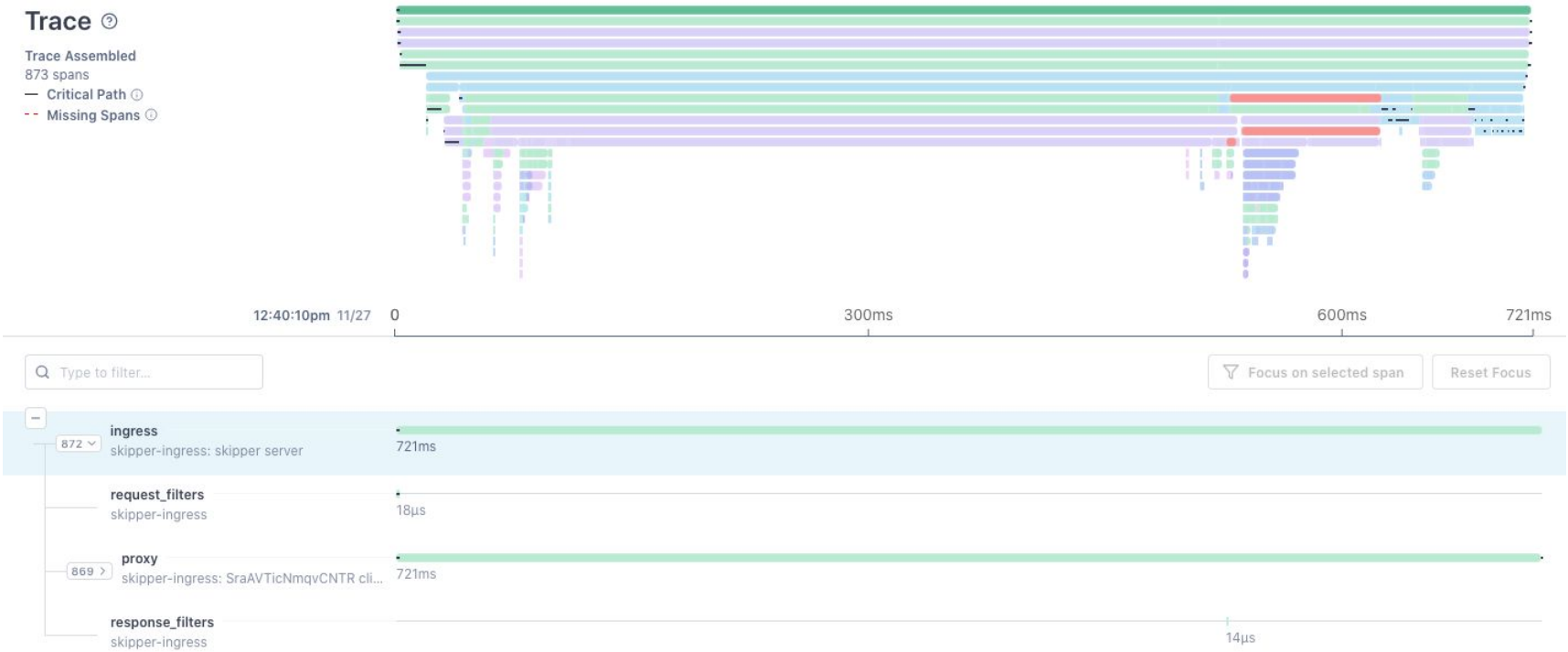


Heinrich Hartmann - DevOpsCon Berlin / June 2022


OpenTelemetry
add item to cart
OpenTelemetry
check stock
OpenTelemetry
Website
Cart API
Stock API
Send spans
Lightstep


Trace
Website - /add-to-cart
Cart API - /add
Stock API - /check
Stock API - /check
Stock API - /check

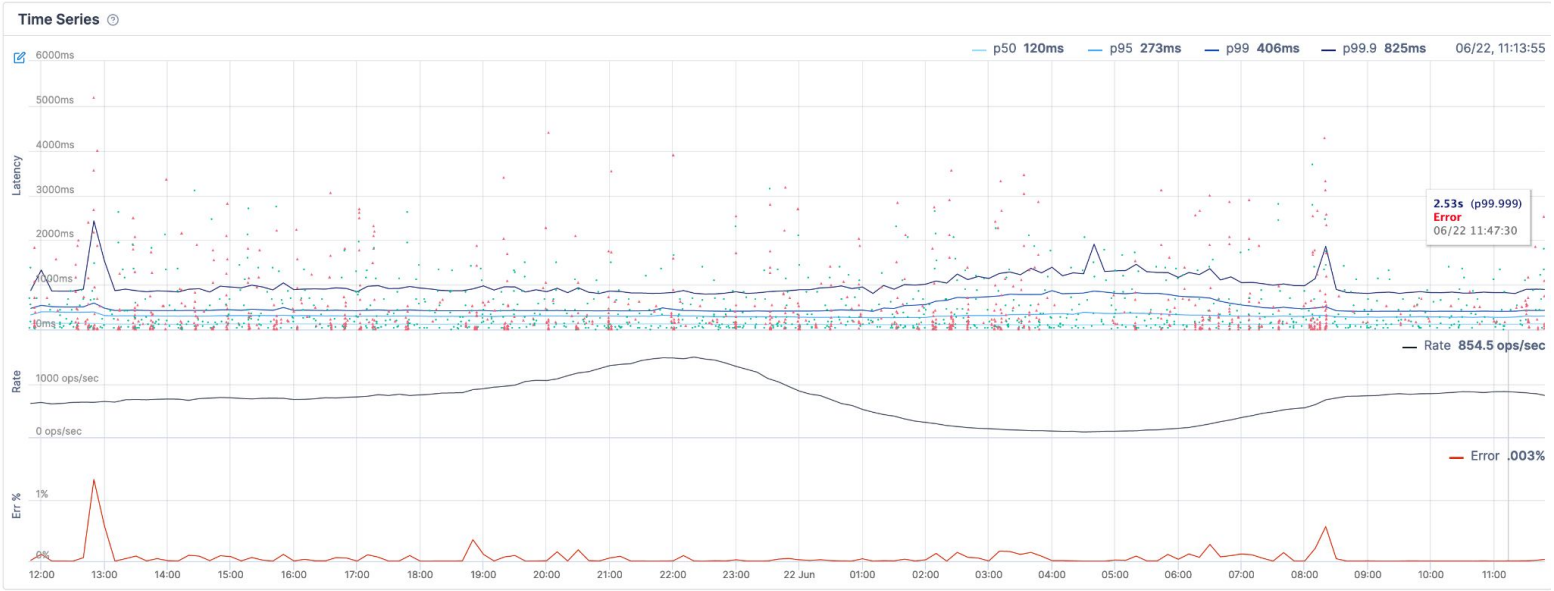
zalando


Trace
Trace Assembled
873 spans
Critical Path
Missing Spans
12:40:10pm 11/27
0
300ms
600ms
721ms
Type to filter...
Focus on selected span
Reset Focus
872
ingress
skipper-ingress: skipper server
721ms
request_filters
skipper-ingress
18µs
869
proxy
skipper-ingress: SraAVTicNmqvCNTR cli...
721ms
response_filters
skipper-ingress
14µs

R

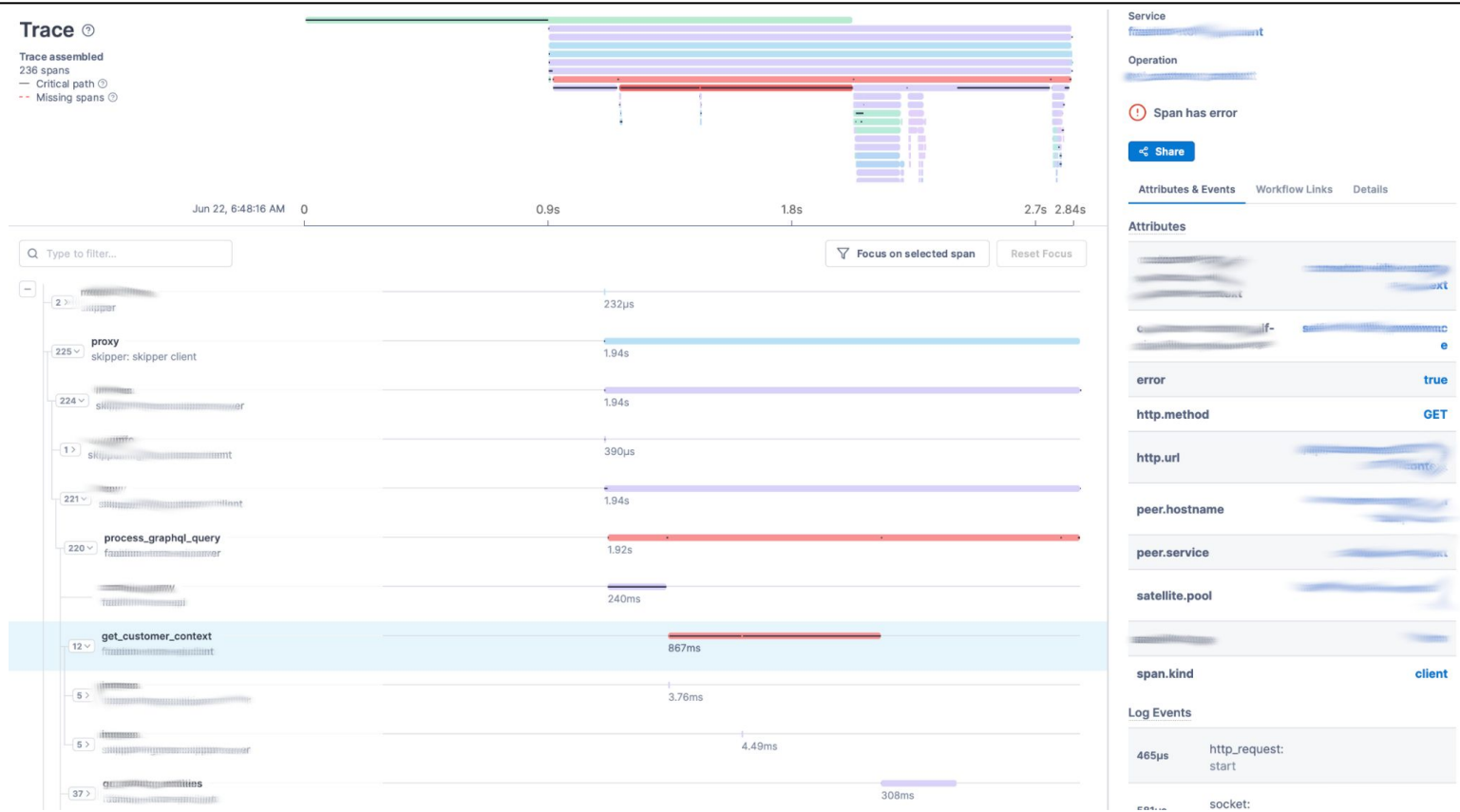
E

D



Lightstep

zalando



Lightstep

Tracing at Zalando

- First introduced in 2019
- >3K Applications Instrumented with Tracing (OpenTracing, OpenTelemetry)
- 10M traced operations/second peak
- 3d raw data retention
- 50% sampling applied before ingestion

HeinrichHartmann.com

Sampling Error Calculator

View Source — Stemwede, 2022-05-29

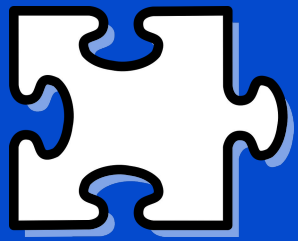
# Sampling			
Sampling Rate	50 %		

# Request Rates			
Request Rate	10 rps		
Time window	1 min (60 sec)		
Population	Total 600 requests contained in 60 sec time-window.		
## Samping Effects on Request Rate Estimates			929 iterations
Sample	We expect to retrain 300 requests, after sampling with 50% probability.		
	Value	Standard Error	Realtive Error
Estimate Req. Count	600.0 req	± 24.49 req	4.08%
Simulate Req. Count	599.9 req	± 24.80 req	4.13%
Estimate Req. Rate	10.0 rps	± 0.41 rps	4.08%
Simulate Req. Rate	10.0 rps	± 0.41 rps	4.13%

Sampling Error Calculator available on HeinrichHartmann.com

- Observability SDKs
  Productize Observability for common Engineering patterns:
  - Language Runtimes
  - HTTP/REST APIs
  - DB clients
  - other libraries

- Standardized Dashboards
  For supported technologies like:
  k8s, Redis, Kafka, Postgres, ...

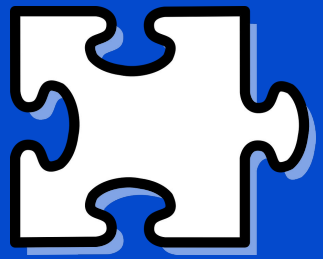


Productize Operational Know-How

OpenTelemetry

Grafana

zalando

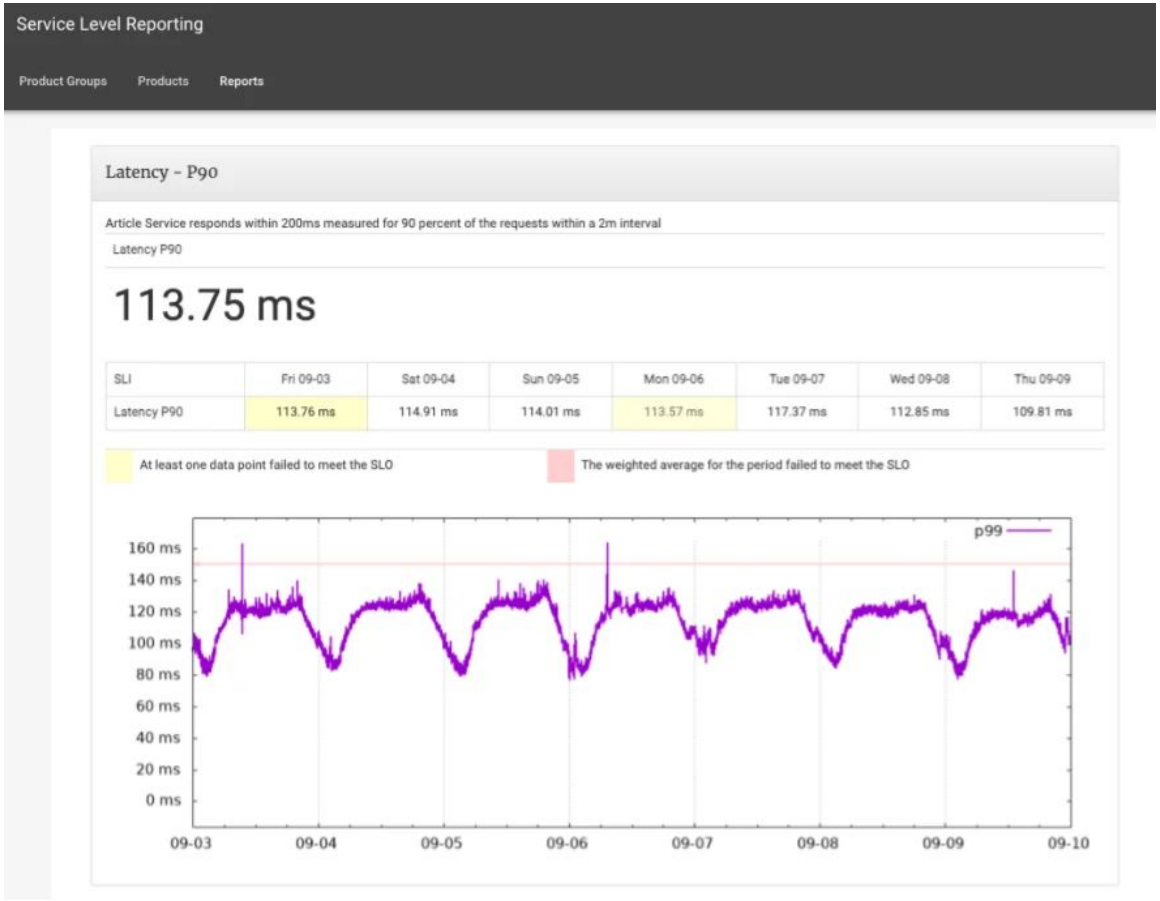


Operation based SLOs

P. Alves - Operation Based SLOs (engineering.zalando.com)

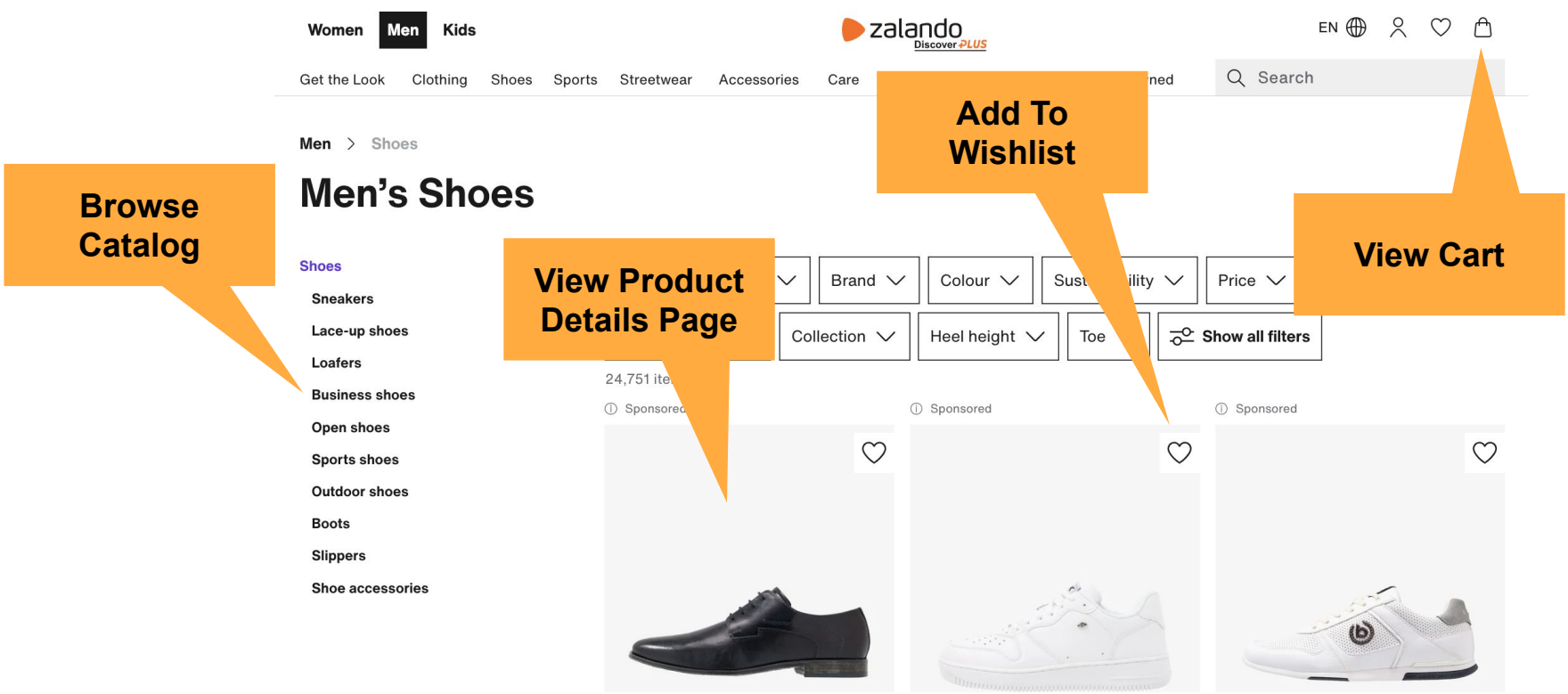
Service List

- Proxy Web
- Rendering Engine
- Checkout Service
- Payment Gateway
- Payment Service
- Risk Service
- Accounting Service
- Stock Service
- Customer Service
- Order Service
- Random BI Service
- Coupon Service
- Typical Payment Blackbox
- Logistics Service
- Mail Notification Service
- Authentication Service
- Another Shady Service
- Machine Learning Shenanigans
- Article Service
- …



zalando

Critical Business Operations

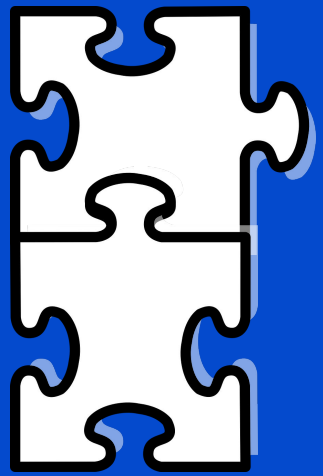


Availability

Operation	Organization	Target	Current 27 Mar - 25 Apr	Previous 27 Feb - 27 Mar
Fashion Store			99.99% ▲0.01%	99.98%
Add to cart	DX	99.00%	99.98% ▲0.02%	99.96%
View customer profile	DX	98.00%	99.77% ▲0.02%	99.75%
Place order	DX	99.00%	99.65% ▲0.14%	99.51%
Remove from cart	DX	99.00%	99.89% ▲0.19%	99.70%
View catalog	DX	99.00%	99.89%	99.88%
Show order history	DX	99.00%	99.97%	99.97%
View cart	DX	99.00%	99.98% ▲0.04%	99.94%

zalando

SLO Based Alerting



Heinrich Hartmann - DevOpsCon Berlin / June 2022

- SLOs quantify customer experience
- SLO-alerting avoids false-positives and implements Symptom Based Alerting.
- Zalando Alerting Strategy

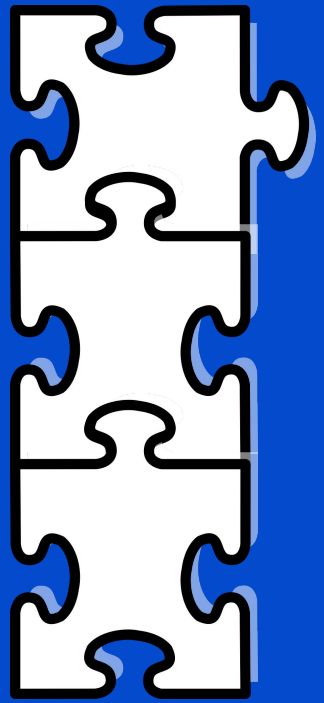
Page only if an SLO is at risk of being breached.

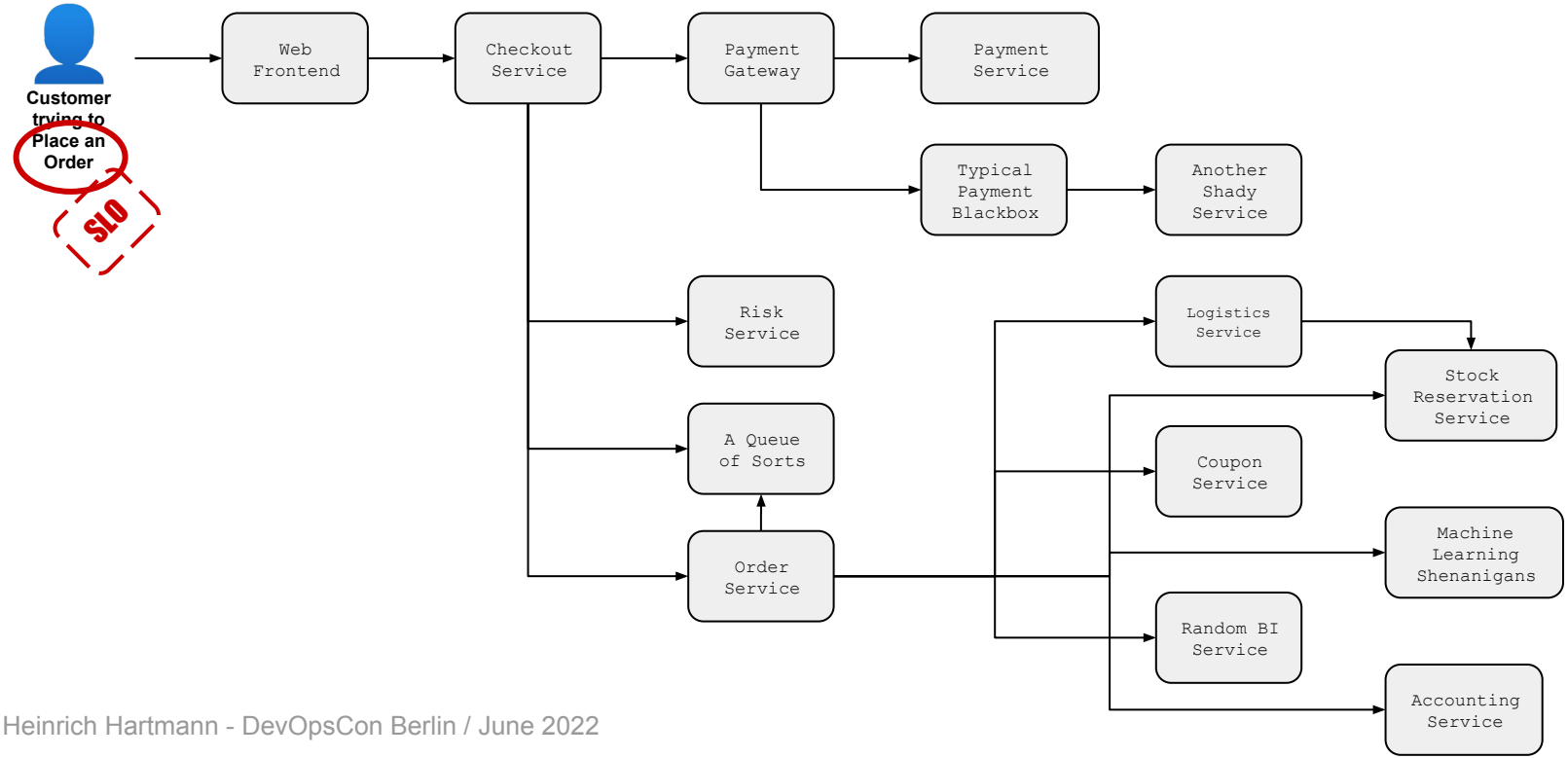
- Allow for cause-based non-paging alerts. But only wake people up, if there is an actual user-facing problem.

zalando

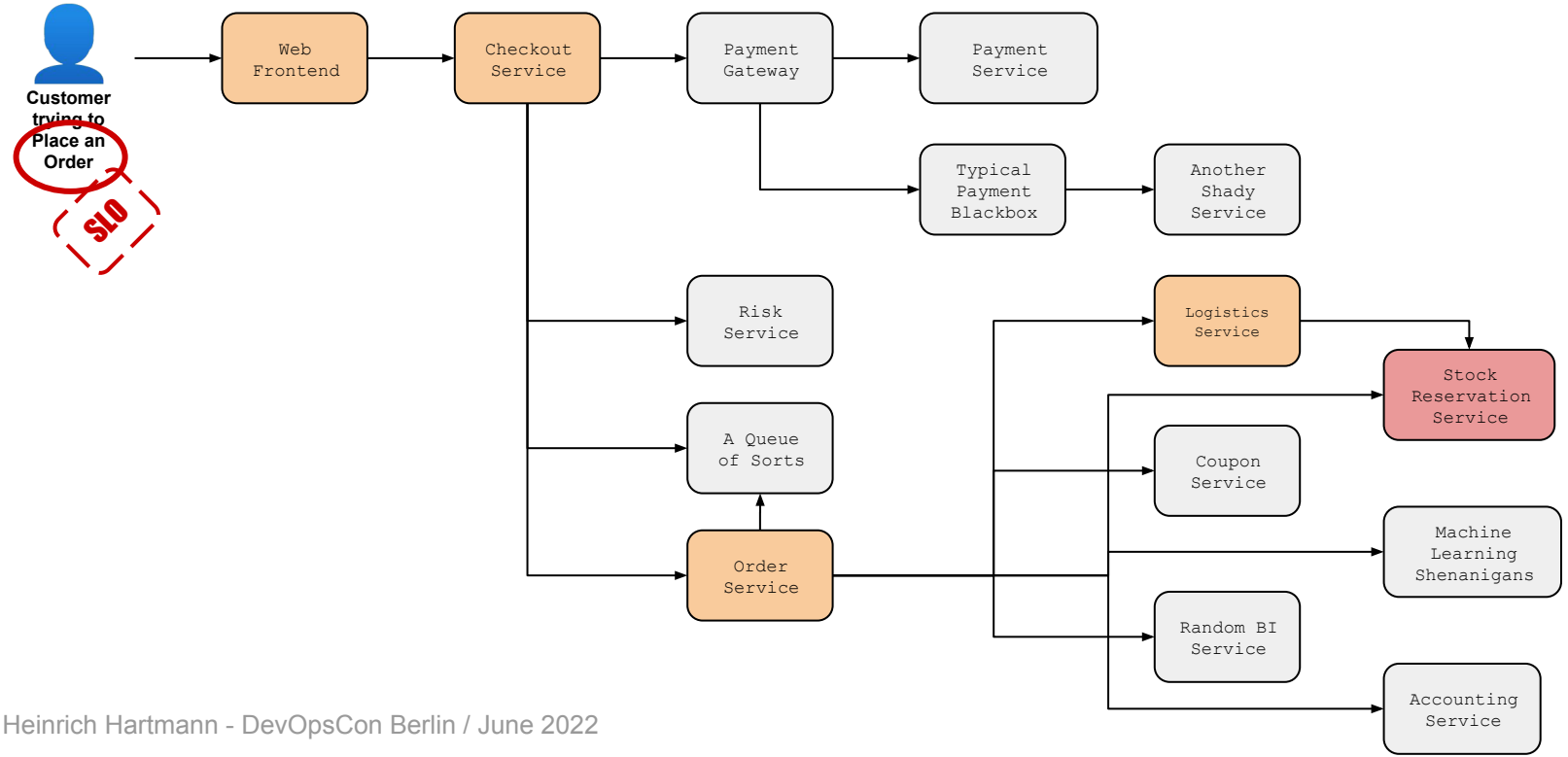
Adaptive Paging

L. Mineiro - Are we on the same page? SRECon 2019 / Login;



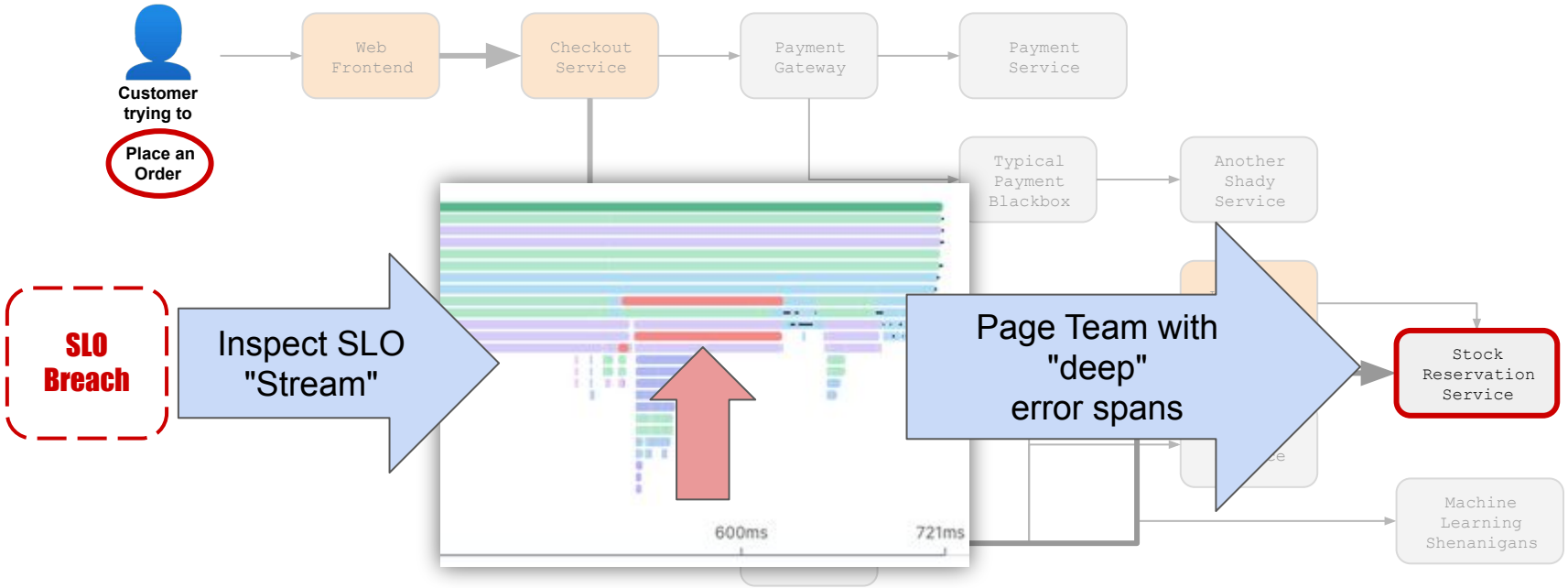

Customer trying to Place an Order
SLO
Web Frontend
Checkout Service
Payment Gateway
Payment Service
Typical Payment Blackbox
Another Shady Service
Risk Service
A Queue of Sorts
Order Service
Logistics Service
Stock Reservation Service
Coupon Service
Machine Learning Shenanigans
Random BI Service
Accounting Service

The Christmas Tree Problem



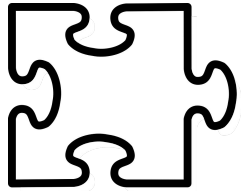
zalando

Adaptive Paging

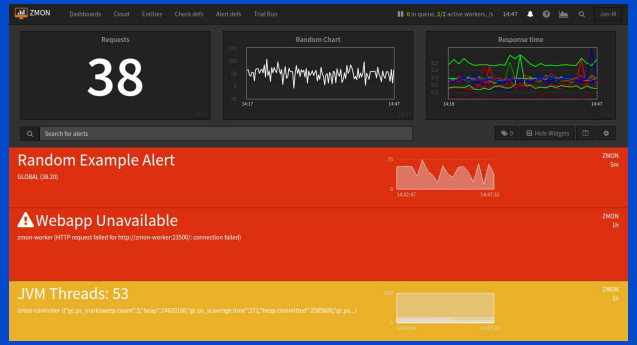


L. Mineiro - Are we on the same page? SRECon 2019 / Login;

zalando



Outsource Metrics Storage



- In-House Monitoring System ZMON
- Metrics Storage Operational pain-point
- Outsource Metrics Storage in 2021

Metrics POC Requirements

Highly Scalable, Reliable, "Straight Forward" metrics solution.

Load Profile

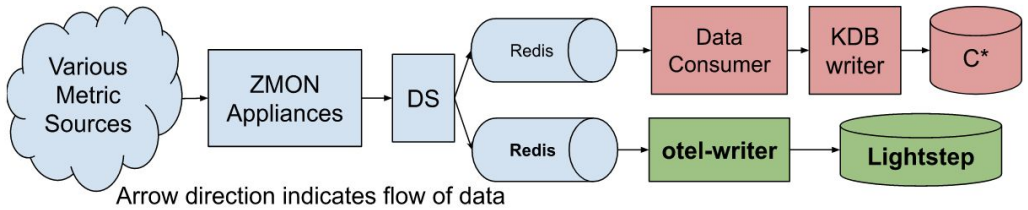
- 120M metrics ingestion peak
- 300 rps read load

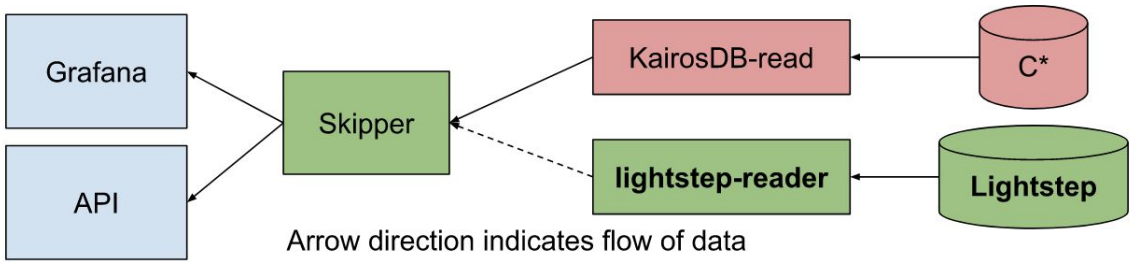
- No fancy analytics / query patterns
- 30 day data retention
- Cost efficient. Competitive with self-hosted solution (inc. staff costs)

Our pick

Runner-up

Metrics Transition Architecture





Next Up

- More Distributed Tracing
- More Standardisation
- More SLOs
- More Load-Testing

@HeinrichHartmann