

Zalando's 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 \(FOSDEM 2019\)](#)
- [CircIhist - A Histogram Data Structure of IT Operations \(arxiv\)](#)

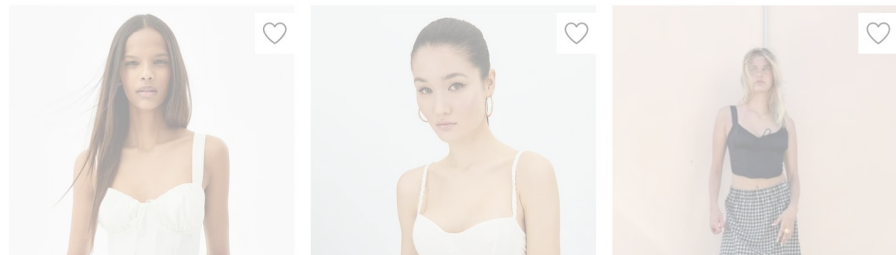
Sponsored Sponsored Sponsored



Zalando Business

Urban Classics
Basic T-shirt - pale green
From 11,99 €
Originally: 14,99 € up to 23%

8 Offener
MET KANT - Top - black
40 €
23%
PLUS Premium Delivery



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



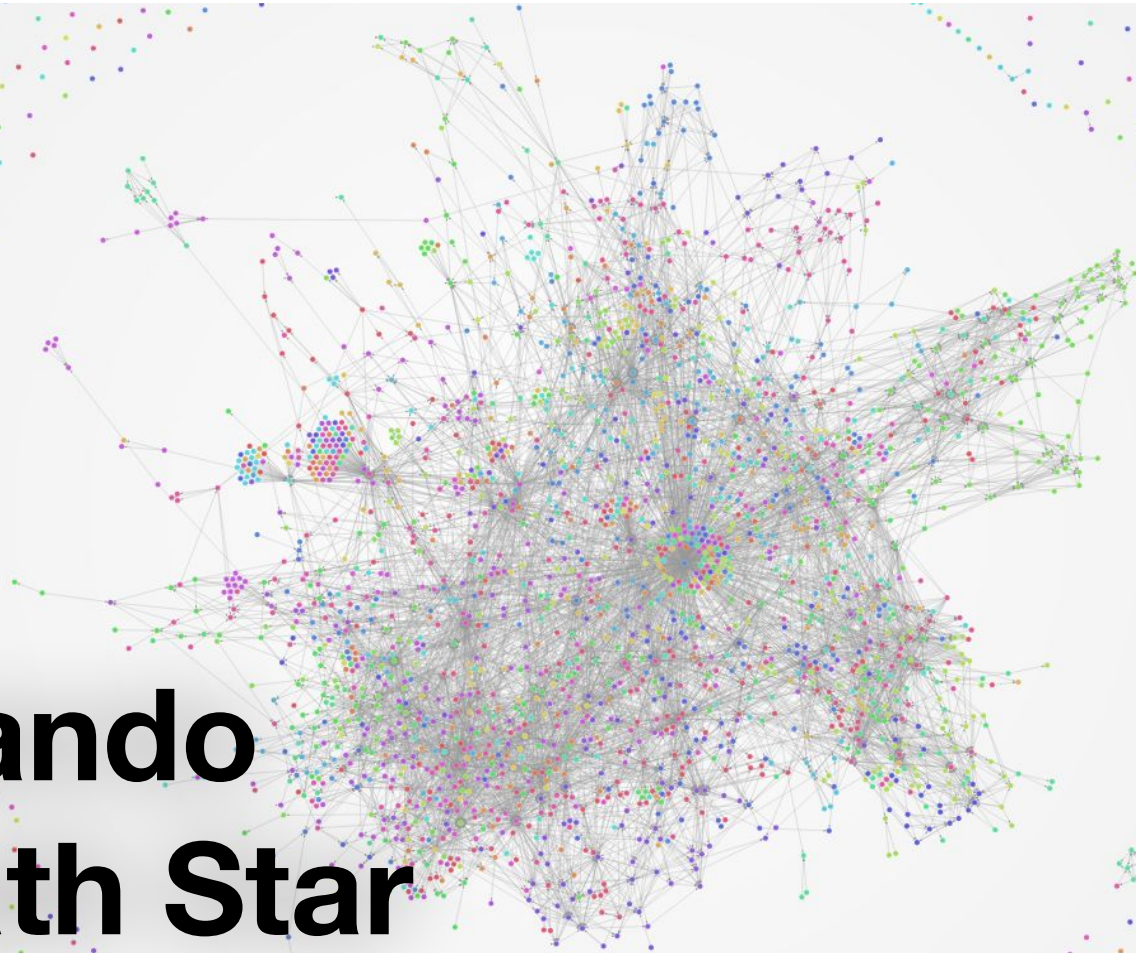
Zalando Tech

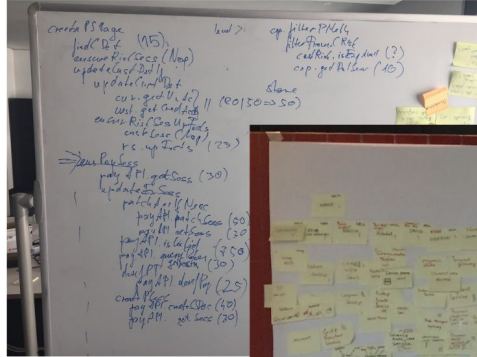
- 2.500+ SWE on Staff
- 200+ teams
- In AWS Frankfurt
- Up to 10K EC2 nodes
- 200+ k8s clusters
- 5K+ Micro Services

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

Zalando Death Star

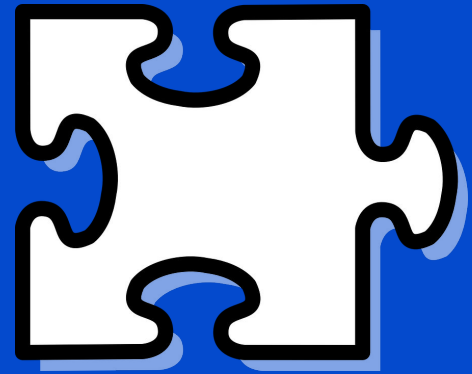
Zalando Micro-Service architecture diagram ~2019 (aka. "Death Star")

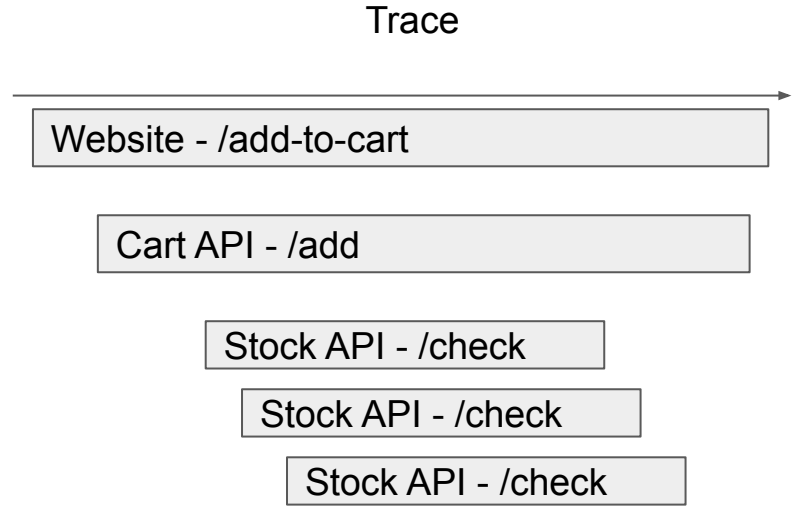
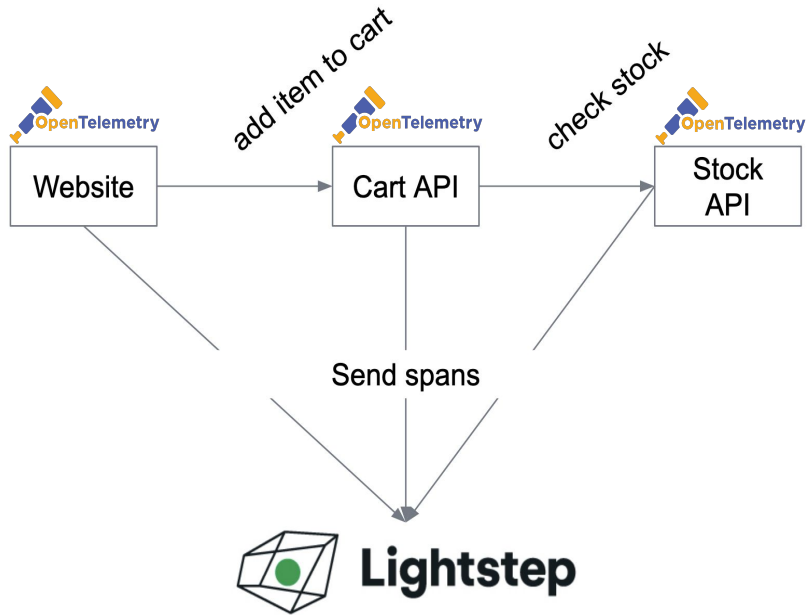




Service Diagramming Exercise @ Zalando ~2017

Distributed Tracing





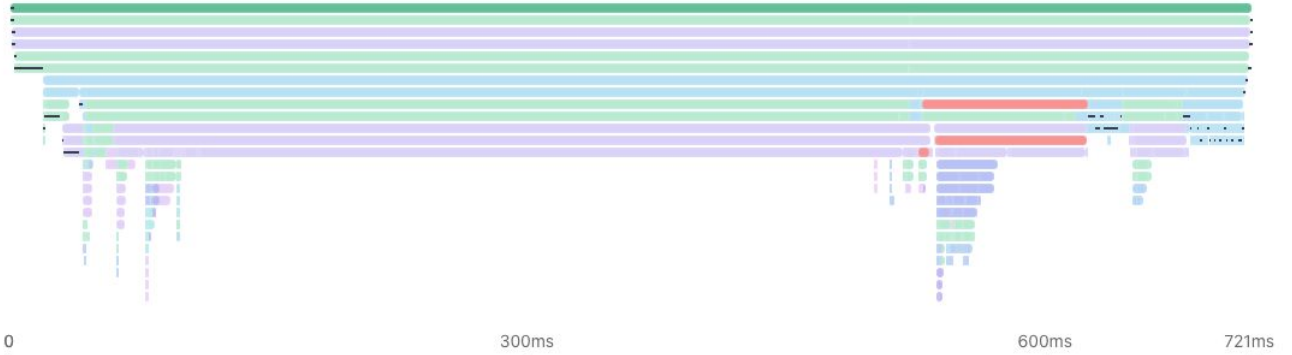
Trace ⓘ

Trace Assembled

873 spans

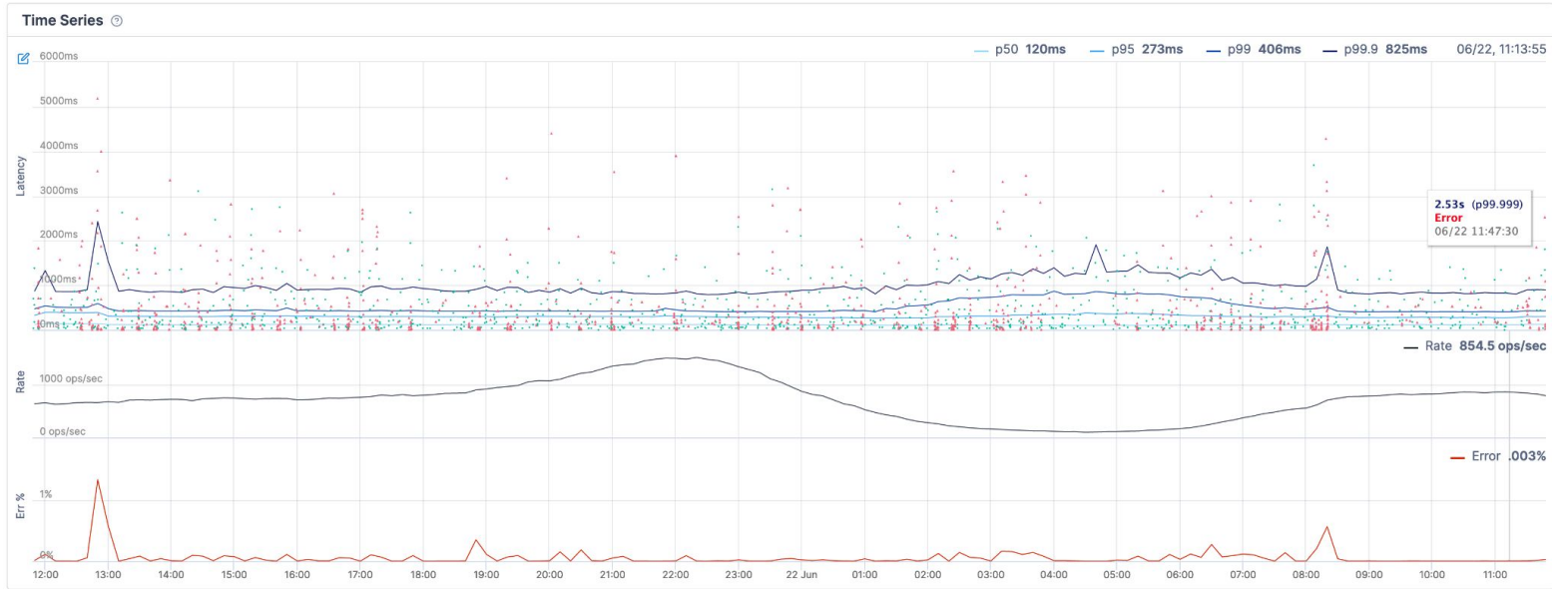
— Critical Path ⓘ

-- Missing Spans ⓘ



-	ingress	skipper-ingress: skipper server	721ms
	request_filters	skipper-ingress	18μs
872 ▾	proxy	skipper-ingress: SraAVTicNmqvCNTR cli...	721ms
	response_filters	skipper-ingress	14μs

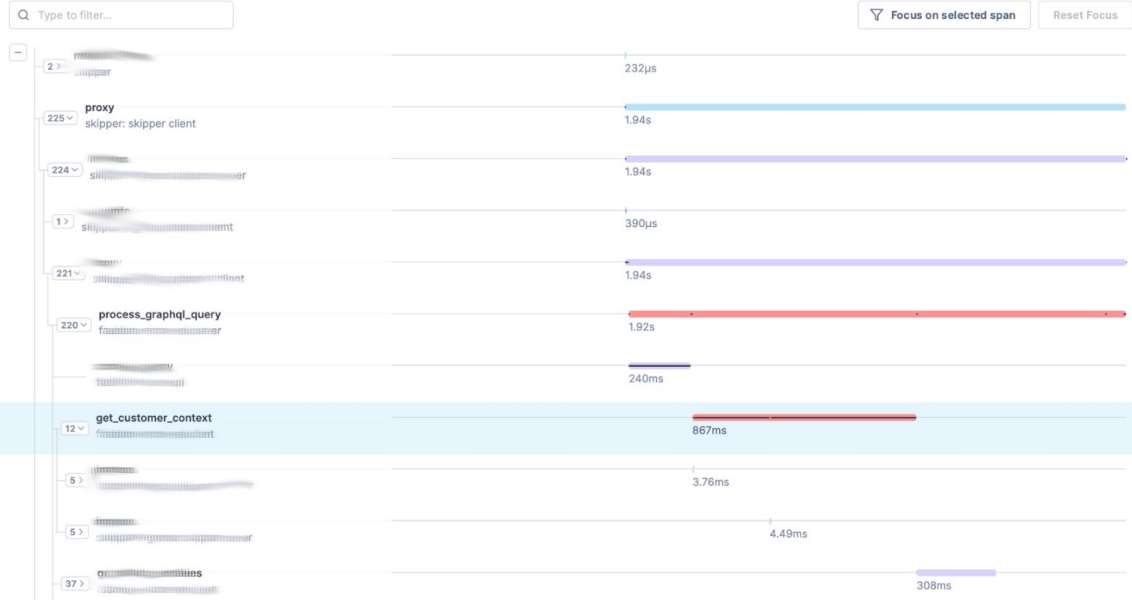
R
E
D



Trace ⊙

Trace assembled
236 spans
— Critical path ⊙
- - Missing spans ⊙

Jun 22, 6:48:16 AM 0 0.9s 1.8s 2.7s 2.84s



Service

Operation

! Span has error

[Share](#)

[Attributes & Events](#) [Workflow Links](#) [Details](#)

Attributes

error	true
http.method	GET
http.url	
peer.hostname	
peer.service	
satellite.pool	
span.kind	client

Log Events

465µs	http_request: start
...	socket:

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



Sampling Error Calculator

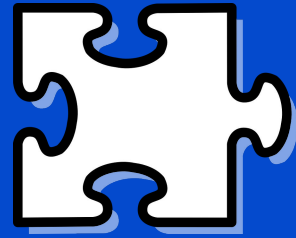
[View Source](#)

Stemwede, 2022-05-29

# Sampling			
Sampling Rate	<input type="text" value="50"/> %		
# Request Rates			
Request Rate	<input type="text" value="10"/> rps		
Time window	<input type="text" value="1"/> min	(60 sec)	
Population	Total 600 requests contained in 60 sec time-window.		
## Sampling Effects on Request Rate Estimates			929 iterations
Sample	We expect to retrain 300 requests, after sampling with 50% probability.		
	Value	Standard Error	Relative 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, ...

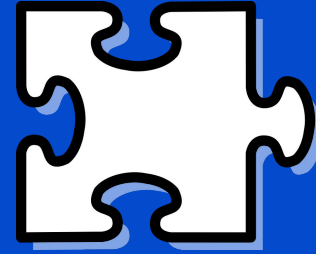


Productize Operational Know-How

Operation based SLOs

P. Alves - [Operation Based SLOs](https://engineering.zalando.com) (engineering.zalando.com)

Heinrich Hartmann - DevOpsCon Berlin / June 2022



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
- ...

Service Level Reporting

Product Groups Products Reports

Latency - P90

Article Service responds within 200ms measured for 90 percent of the requests within a 2m interval

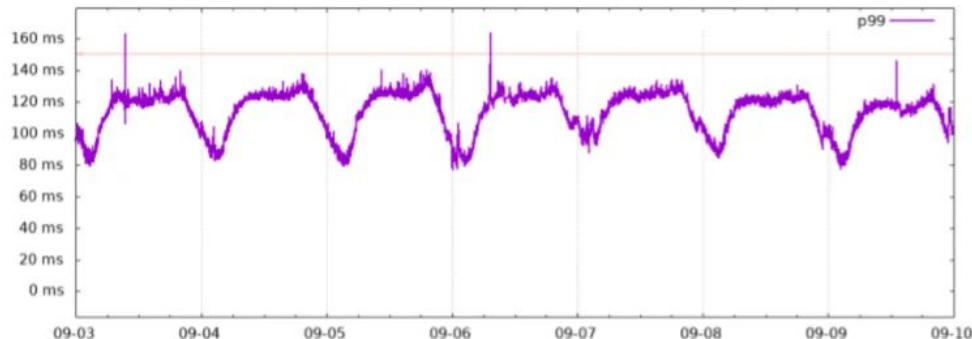
Latency P90

113.75 ms

SLI	Fri 09-03	Sat 09-04	Sun 09-05	Mon 09-06	Tue 09-07	Wed 09-08	Thu 09-09
Latency P90	113.76 ms	114.91 ms	114.01 ms	113.57 ms	117.37 ms	112.85 ms	109.81 ms

At least one data point failed to meet the SLO

The weighted average for the period failed to meet the SLO



The image shows a screenshot of the Zalando website's 'Men's Shoes' category page. The page layout includes a top navigation bar with 'Women', 'Men', and 'Kids' tabs, and a search bar. Below the navigation, there are category filters for 'Sneakers', 'Lace-up shoes', 'Loafers', 'Business shoes', 'Open shoes', 'Sports shoes', 'Outdoor shoes', 'Boots', 'Slippers', and 'Shoe accessories'. A filter bar contains dropdown menus for 'Brand', 'Colour', 'Sustainability', 'Price', 'Collection', 'Heel height', and 'Toe', along with a 'Show all filters' button. Three shoe products are displayed in a grid, each with a heart icon for adding to a wishlist. Four orange callout boxes highlight key business operations: 'Browse Catalog' points to the category filters, 'View Product Details Page' points to the first shoe product, 'Add To Wishlist' points to the heart icon on the second shoe, and 'View Cart' points to the shopping cart icon in the top right corner.

Browse Catalog

View Product Details Page

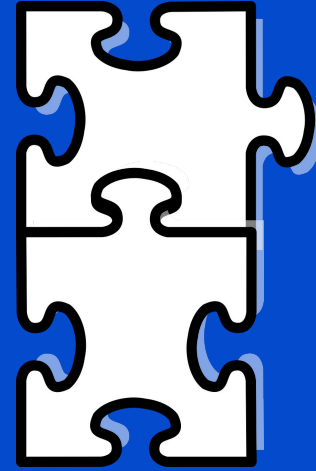
Add To Wishlist

View Cart

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%

SLO Based Alerting



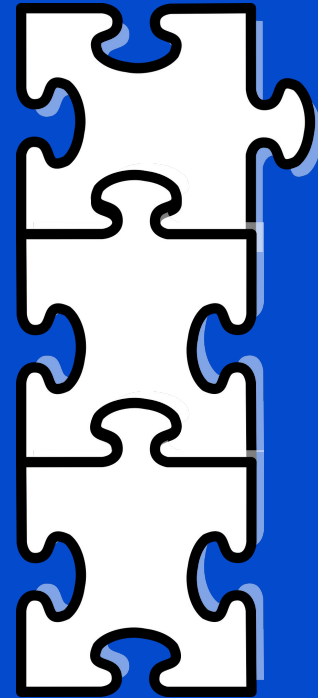
- SLOs quantify customer experience
- SLO-alerting avoids false-positives and implements [Symptom Based Alerting](#).
- [Zalando Alerting Strategy](#)

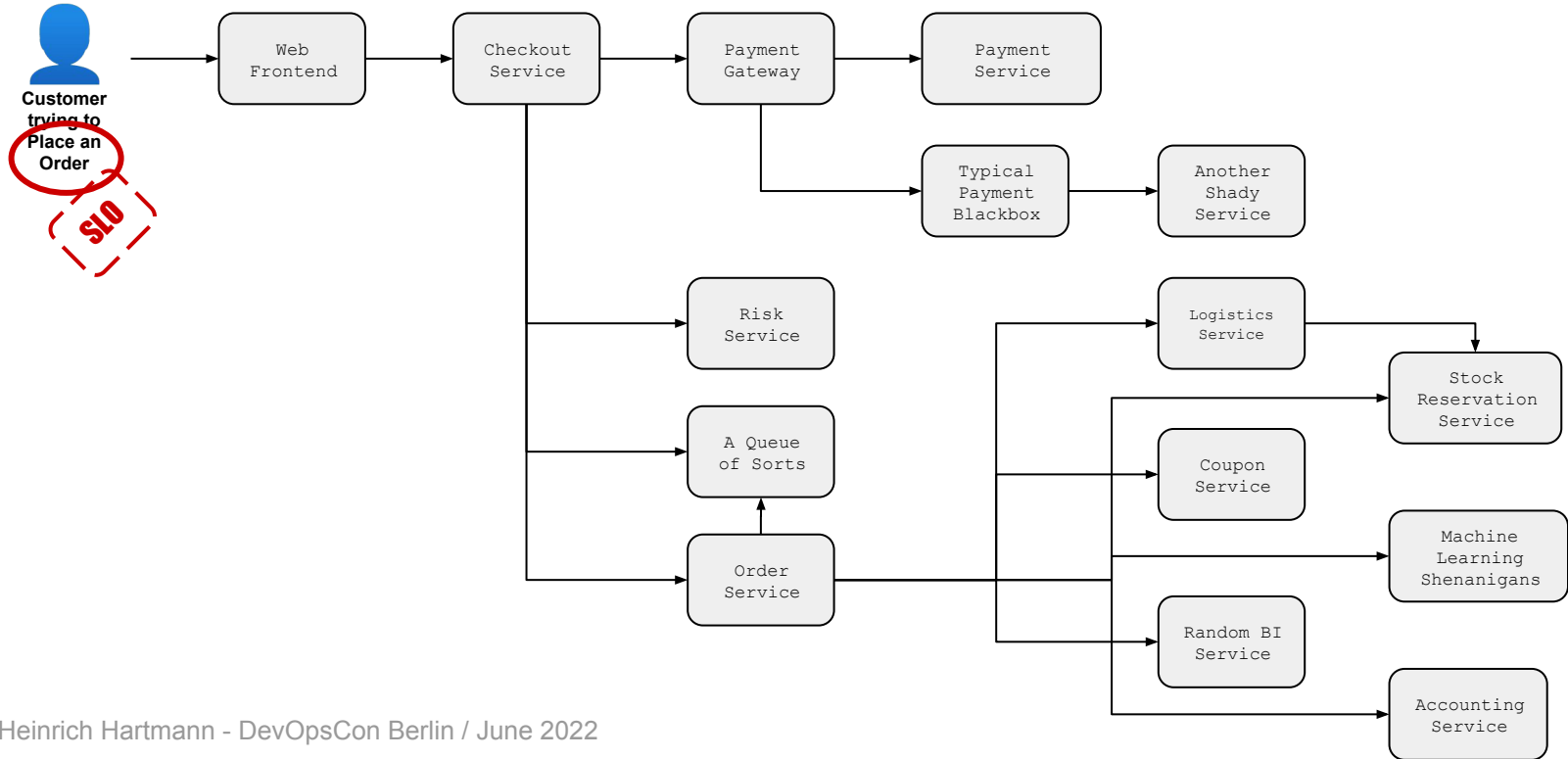
Page only if an SLO is at risk of being breached.

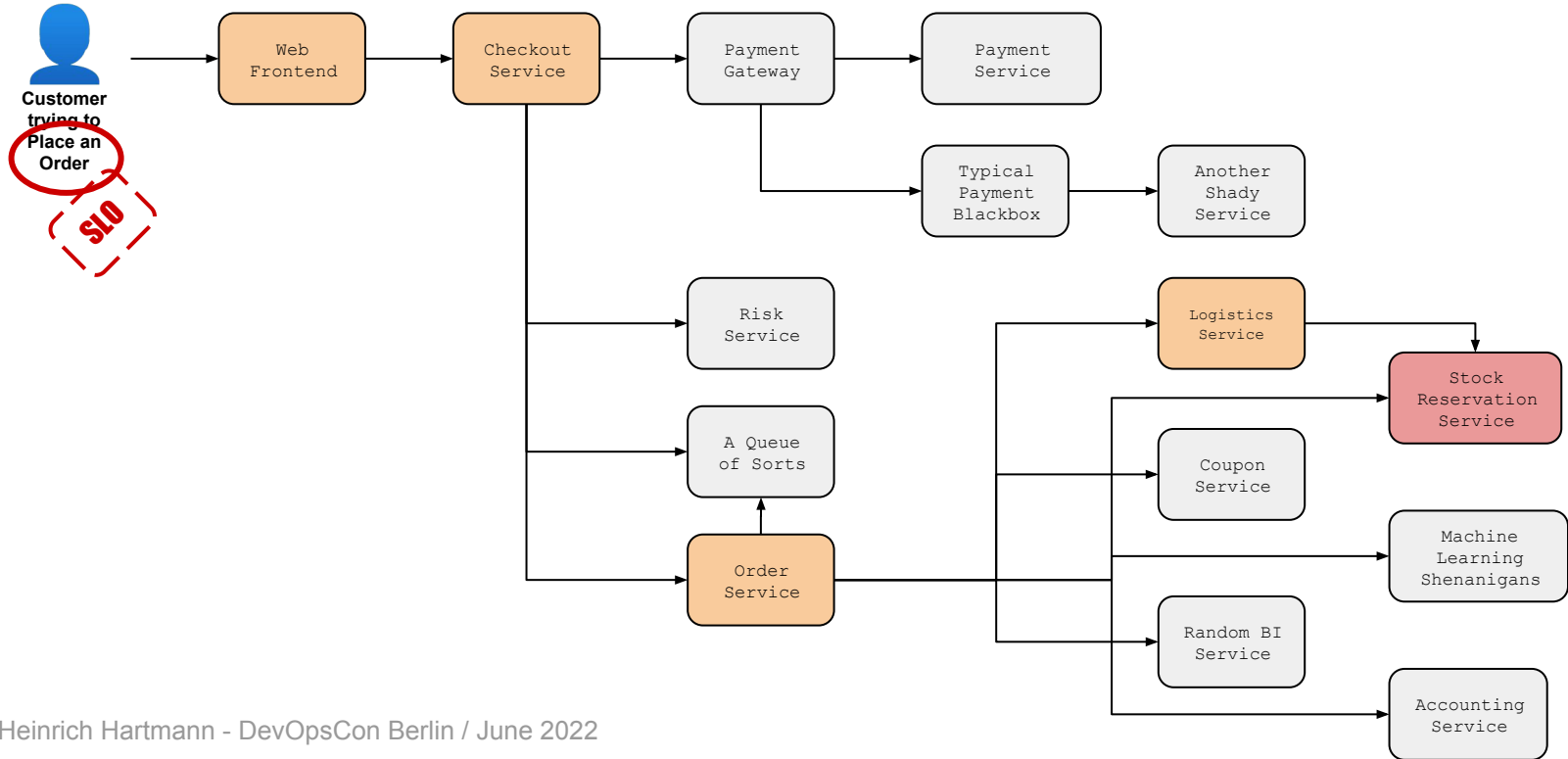
- Allow for cause-based non-paging alerts. But only wake people up, if there is an actual user-facing problem.

Adaptive Paging

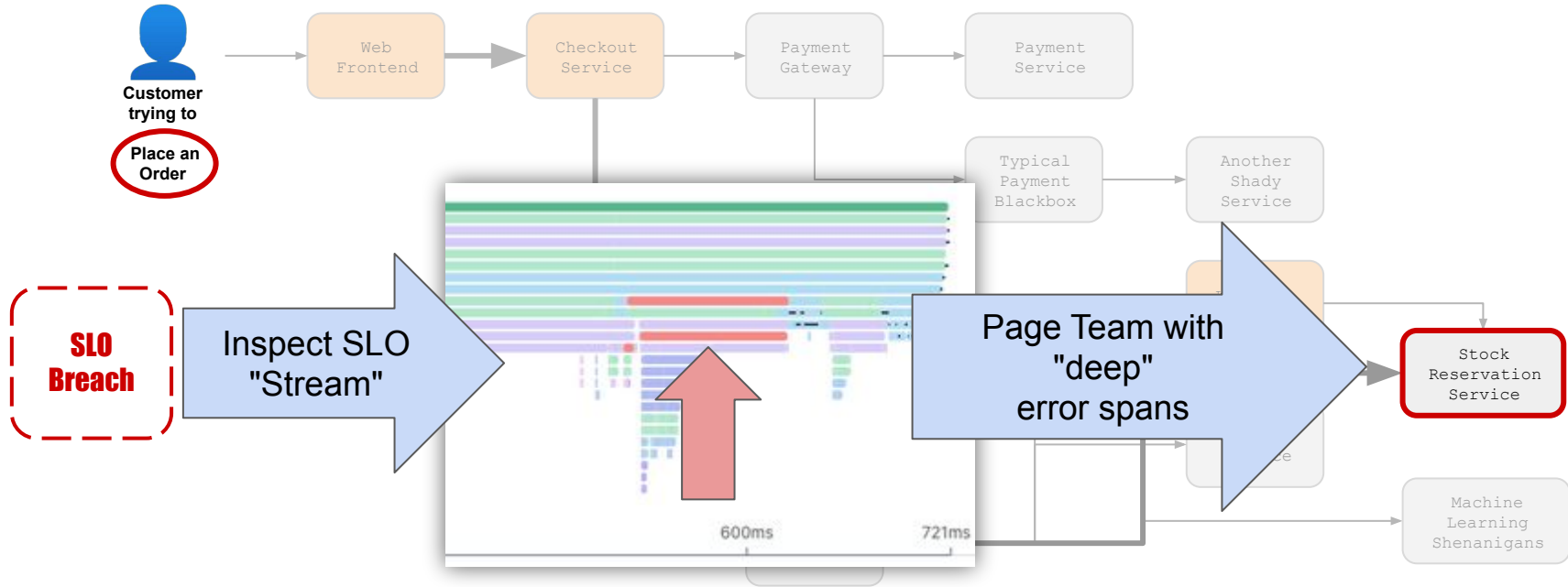
L. Mineiro - Are we on the same page? [SRECon 2019](#) / [Login](#):

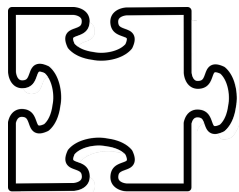




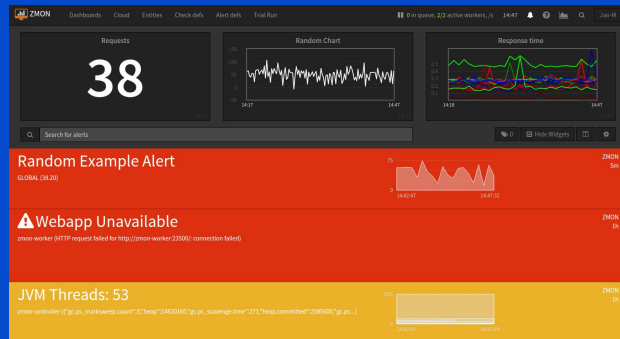


Adaptive Paging





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



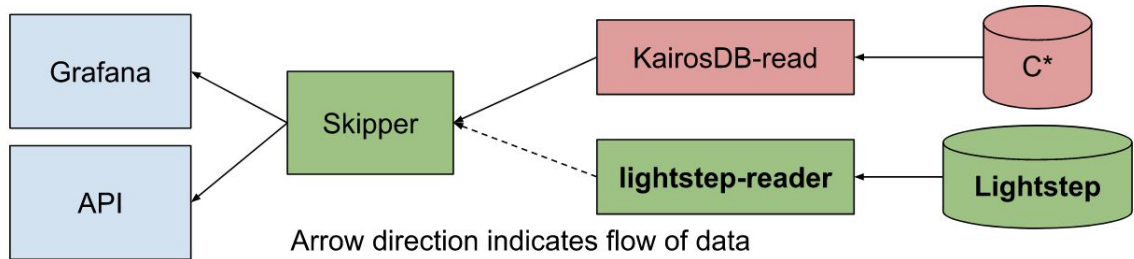
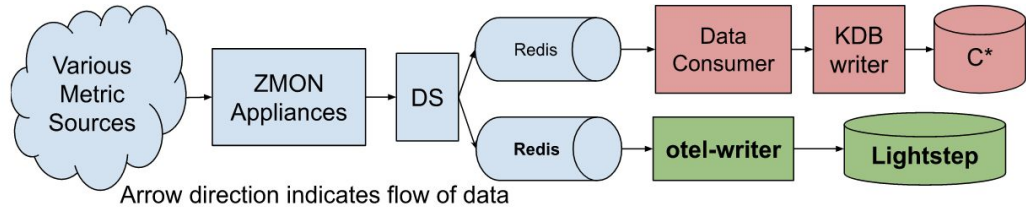
Lightstep Metrics

Runner-up



CIRCONUS

Metrics Transition Architecture





Next Up

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

@HeinrichHartmann