

Infrastruktur

- Repo vorbereiten

```
git clone --recurse-submodules https://github.com/frzifus/ContainerConf-Workshop-2024.git
```

- Jaeger / Datenbank / „other“-app
 - cd ContainerConf-Workshop-2024/source
 - podman-compose up



OpenTelemetry Workshop

Instrumentieren von Applikationen

Heiko Rupp & Benedikt Bongartz



Agenda

- Erster Teil:
 - Ein wenig Theorie
 - Präparation Infrastruktur
 - Quarkus - REST - App als Java-Beispiel
 - Python - REST - Server
 - ggfls: Asynchrone Kommunikation via Kafka
- Zweiter Teil:
 - Kubernetes / Weitere Themen



Ein wenig Theorie

Theorie: Warum Tracing / OpenTelemetry

- Verfolgen von Aufrufen innerhalb einer Applikation
 - Finden von Unregelmäßigkeiten
 - Performance Optimierung
- Micro-Services
 - Viele Komponenten
- Aber auch Monolithen
 - Es ist ja nie nur ein Binary alleine

Theorie: Warum Tracing / OpenTelemetry

- In Container-Umgebungen (z.B. Kubernetes / OpenShift)
 - Immutable Container
 - Ich kann nicht einfach eine Shell aufmachen
 - Ich kann oft nicht einfach einen Debugger / Profiler anhängen
 - Hohe Parallelität
 - An welche Instanz müsste ich den Debugger anhängen?



Theorie: Warum Tracing / OpenTelemetry

- Für Monolithen
 - Normal Cluster von 3+ Instanzen wegen HA
 - Datenbank
 - Anbindung an ERP / Payment / ...
- => Also auch hier ein verteiltes System

Theorie: Warum Tracing / OpenTelemetry

- „Passives“ Verfolgen von Aufrufen
- Weiteres Tool neben Metriken und Logs
- Metriken:
 - Zustand des Servers (CPU-Last, Speicher, ...)
 - Keine Info zu einem Request
- Logs:
 - Geschwätzig. Schwer mit Logs von anderer App zu korrelieren

Theorie: Warum Tracing / OpenTelemetry

OpenTelemetry



- Herstellerunabhängige Telemetriedatenerfassung
- Sammlung von
 - Spezifikationen
 - APIs
 - Werkzeugen
- zum Sammeln von **Traces**, Logs und Metriken
- Sprachen- und Tool- unabhängig
- <https://opentelemetry.io/>

Tools die wir nutzen



- Open-Source Distributed Tracing System
- Entwickelt von Uber, jetzt Teil der CNCF (Cloud Native Computing Foundation)
- **OTLP Native:** Jaeger 2.0 basiert auf dem OpenTelemetry Collector!
- **End-to-End Distributed Tracing:** Verfolgt Anfragen über Microservices hinweg
- **Performance Monitoring:** Identifiziert Latenzen und Flaschenhälse
- **Abhängigkeitsanalyse:** Veranschaulicht Service-Abhängigkeiten
- **Root Cause Analysis:** Unterstützt bei der Fehlersuche und Analyse von Performance-Problemen

Tools die wir nutzen



- Open-Source Monitoring- und Alarmsystem
- Ursprünglich von SoundCloud entwickelt, jetzt Teil der CNCF (Cloud Native Computing Foundation)
- **OTLP Native:** Seit Prometheus 3.0
- **Zeitreihen-Datenbank:** Speichert Metriken als Zeitreihen-Daten (z.B. CPU-Auslastung, Anfragen pro Sekunde)
- **Pull-basierte Datenerfassung:** Abfrage von Metriken durch Abholen (Pull) von definierten Endpunkten
- **Flexible Abfragesprache (PromQL):** Leistungsfähige Sprache zur Abfrage und Analyse von Metriken

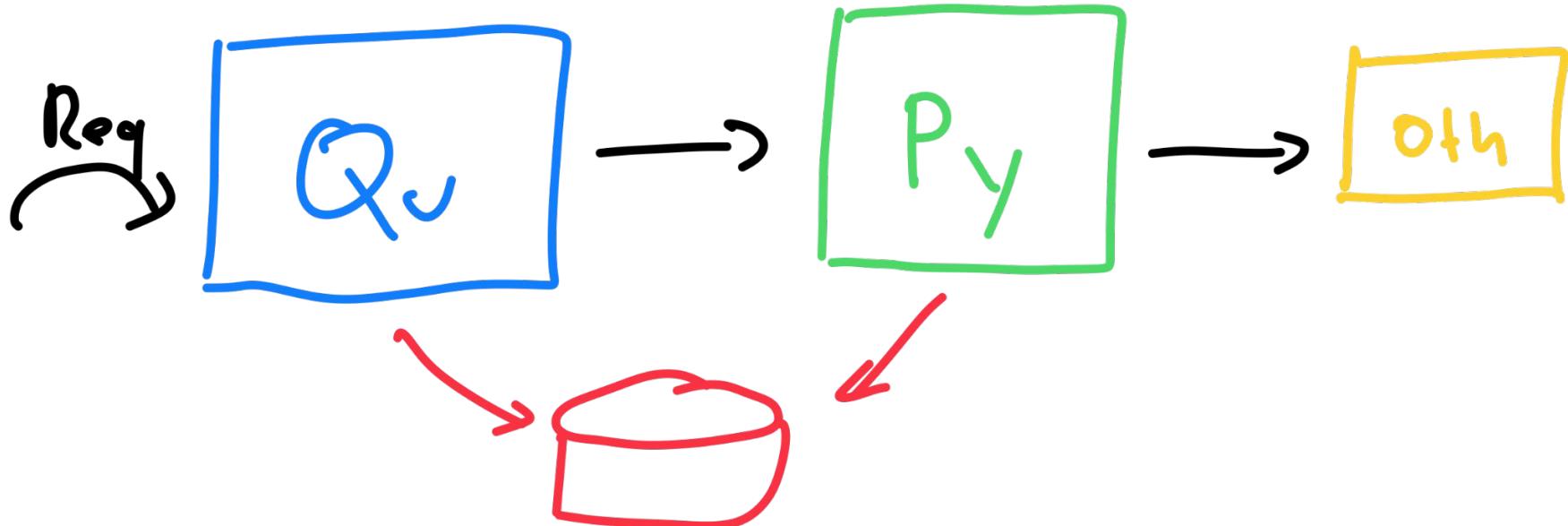
OBSERVABILITY: THE



Red Hat

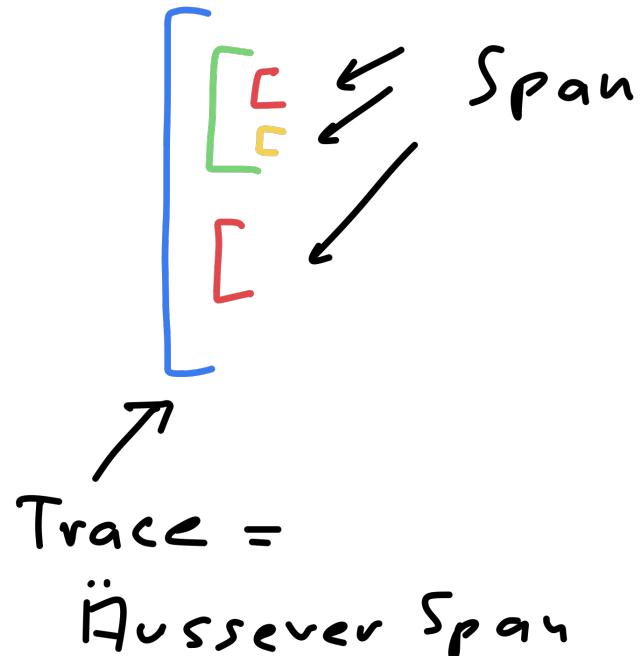
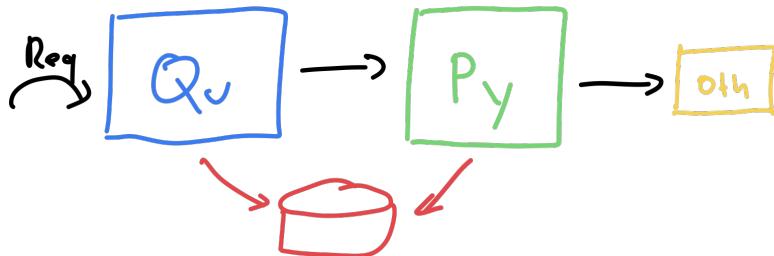
Theorie: Trace / Span - was ist das?

Die Replikator-Applikation



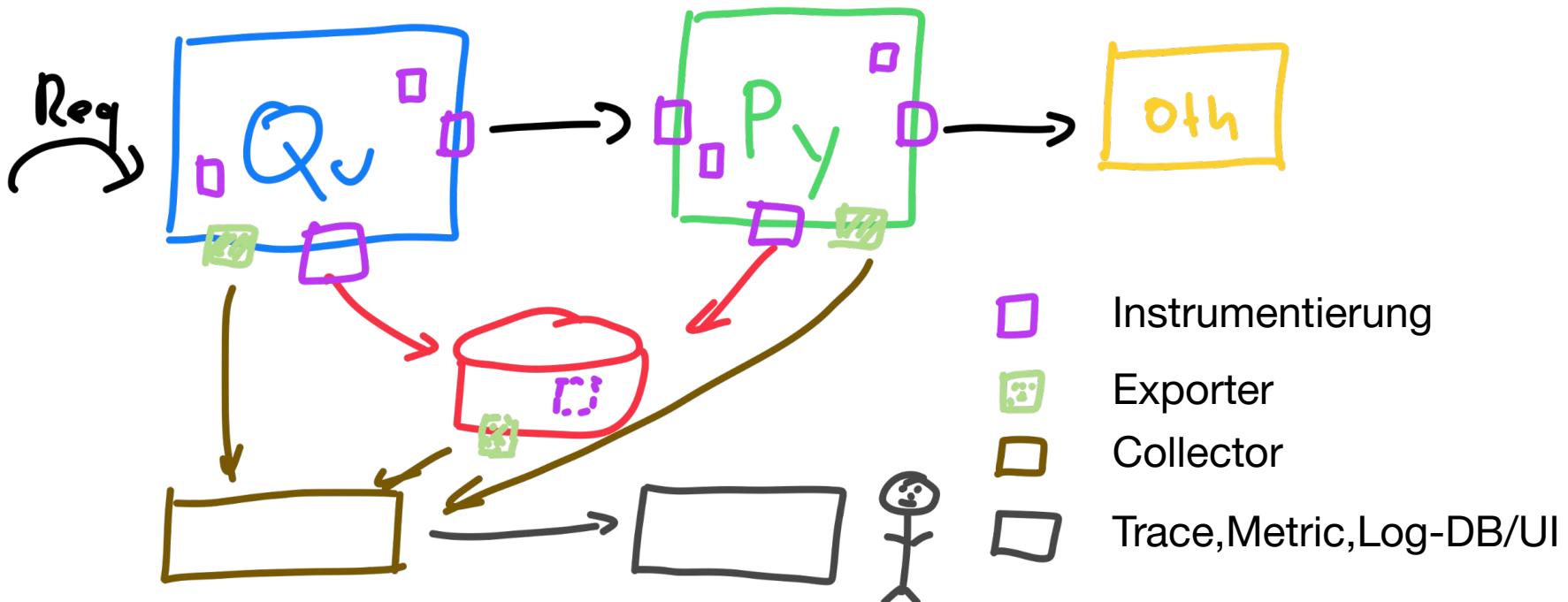
Theorie: Trace / Span - was ist das?

Aufrufe verfolgen...



Red Hat

Theorie: Tracing-Architektur



Präparation Infrastruktur

Infrastruktur

- Repo vorbereiten

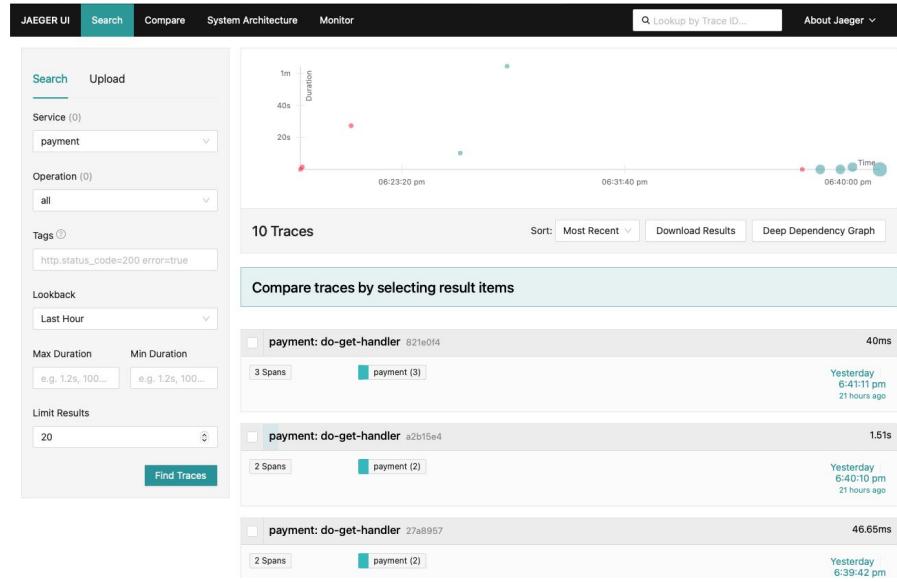
```
git clone --recurse-submodules https://github.com/frzifus/ContainerConf-Workshop-2024.git
```

- Jaeger / Datenbank / „other“-app
 - cd ContainerConf-Workshop-2024/source
 - podman-compose up



Jaeger

<http://localhost:16686/search>



Red Hat

Quarkus-REST-Applikation

Setup

- cd source/quarkus-replicator
- ./mvnw quarkus:dev

Simpler Server

replicator/TeaResource.java

```
@Path("/tea")
public class TeaResource {
    @GET
    @Produces(MediaType.TEXT_PLAIN)
    public String makeTea(@QueryParam("kind") String kind)
        throws Exception {
        return "Here is your " + kind + " tea - enjoy!";
    }
}
```

<http://localhost:8080/>



Sehen Sie, Sie sehen nichts

Wir müssen die Instrumentierung anschalten

pom.xml

```
<dependency>
    <groupId>io.quarkus</groupId>
    <artifactId>quarkus-opentelemetry</artifactId>
</dependency>
```

application.properties

```
quarkus.application.name=replicator
quarkus.otel.exporter.otlp.endpoint=http://${
    otel.host:localhost}:4317
```



JAEGER UI Search Compare System Architecture Monitor Lookup by Trace ID... About Jaeger ▾

Search Upload

Service (2)
replicator

jaeger-all-in-one
replicator

Tags ⓘ
http.status_code=200 error=true

Lookback
Last Hour

Max Duration Min Duration
e.g. 1.2s, 100... e.g. 1.2s, 100...

Limit Results
20

Find Traces

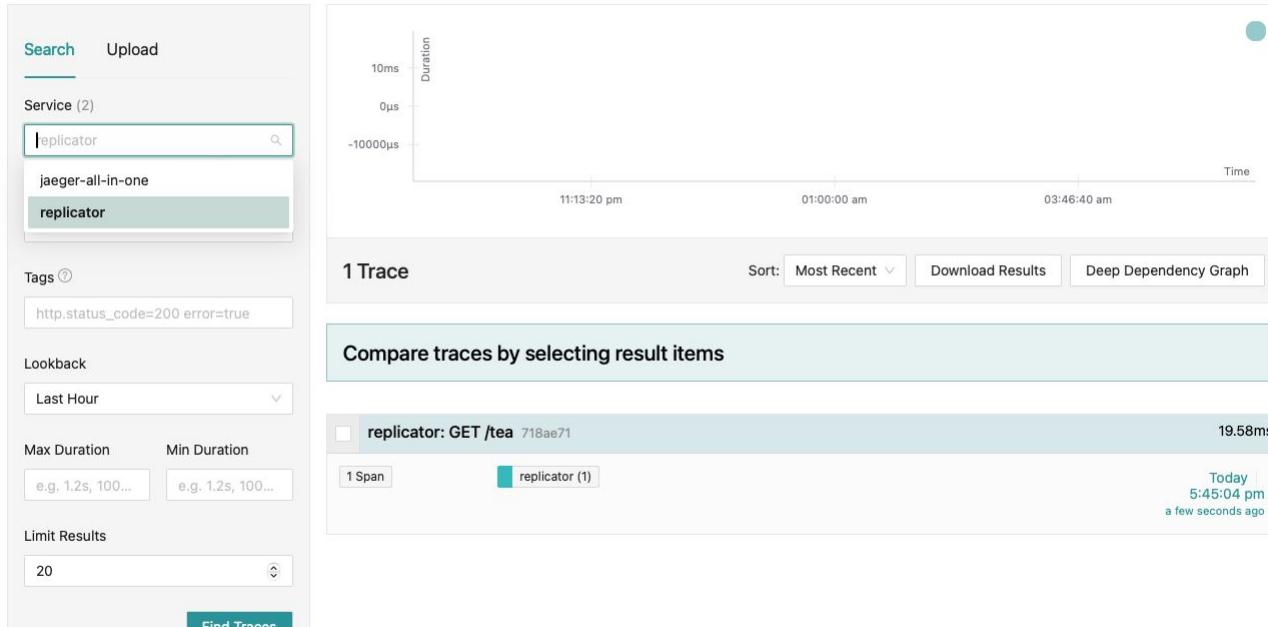
Duration
10ms
0µs
-10000µs

Time
11:13:20 pm 01:00:00 am 03:46:40 am

1 Trace Sort: Most Recent Download Results Deep Dependency Graph

Compare traces by selecting result items

replicator: GET /tea 718ae71 19.58ms
1 Span replicator (1)
Today | 5:45:04 pm
a few seconds ago



<http://localhost:16686/search>



Tees via Datenbank validieren

TeaResource.java

```
String name = kind.toLowerCase(Locale.ROOT) ;  
  
Tea tea = Tea.findByName(name) ;  
if (tea == null) {  
    throw new NotFoundException("No such tea " + kind) ;  
}
```



JDBC-Instrumentierung

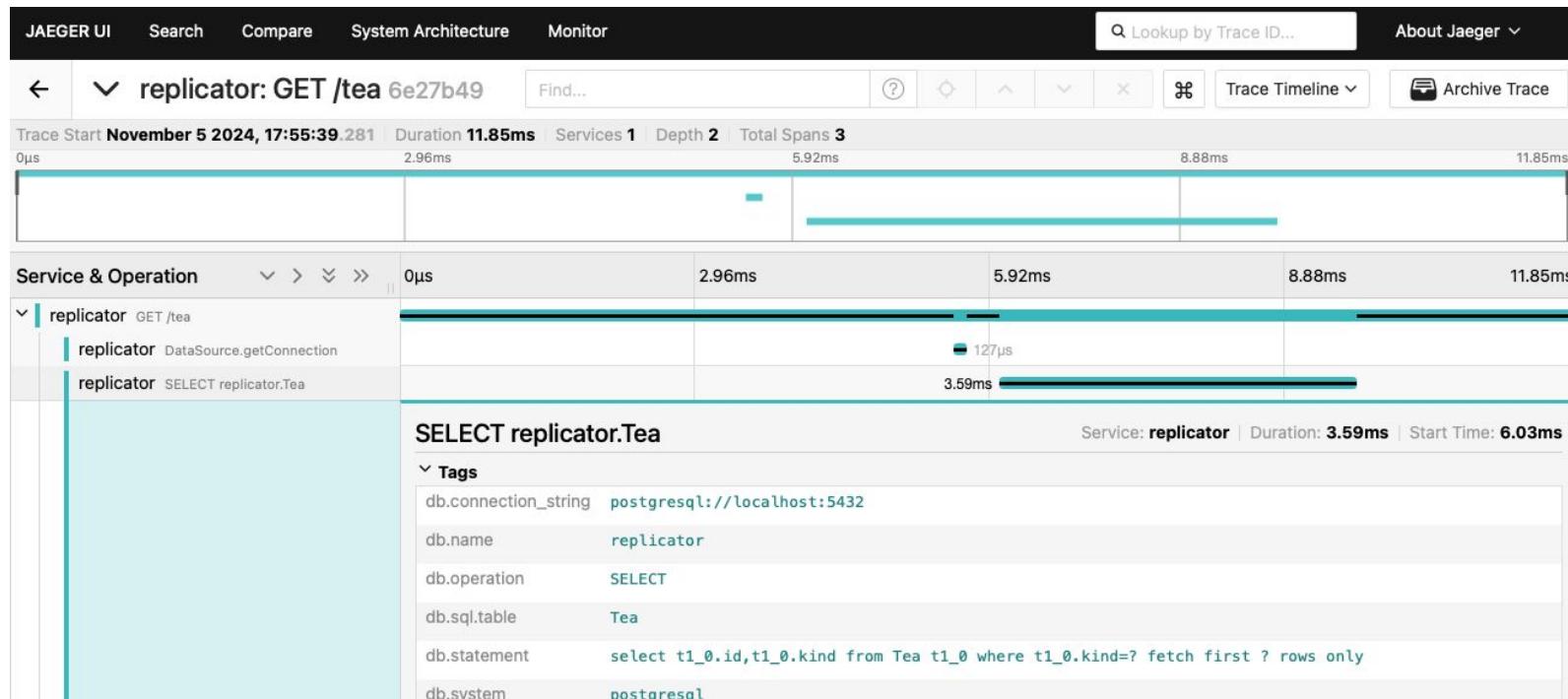
pom.xml

```
<dependency>
    <groupId>io.opentelemetry.instrumentation</groupId>
    <artifactId>opentelemetry-jdbc</artifactId>
</dependency>
```

application.properties

```
quarkus.datasource.jdbc.telemetry=true
```





<http://localhost:16686/search>



Tee aufbrühen

TeaResource.java

```
@Inject  
Brewery brewery;  
  
brewery.brewTea(kind);
```



Tee aufbrühen - eigener Span

Brewery.java

```
@ApplicationScoped  
public class Brewery {  
  
    @WithSpan()  
    void brewTea(@SpanAttribute value = "kind
```



[JAEGER UI](#)[Search](#)[Compare](#)[System Architecture](#)[Monitor](#)

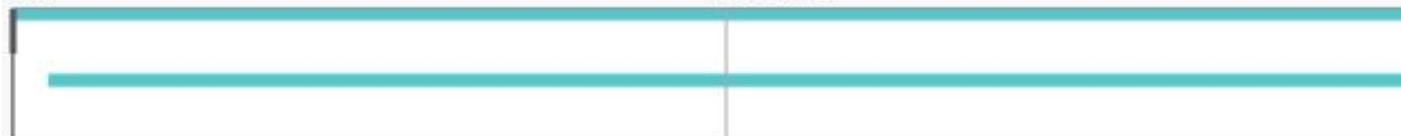
replicator: GET /tea dc410ee

 Find...Trace Start **November 5 2024, 18:09:23.231**Duration **374.28ms**Services **1**Depth **2**

Tot

0µs

93.57ms



Service & Operation

[▼](#) [▶](#) [▽](#) [»](#)

0µs

93.57ms

replicator GET /tea

replicator Brewery.brewTea

**Red Hat**

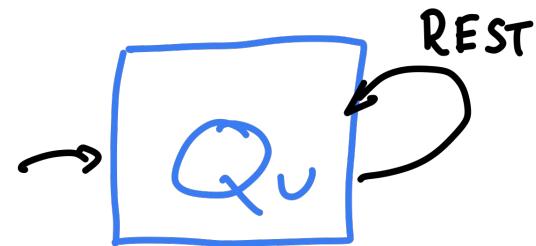
„Externen“ PaymentService aufrufen

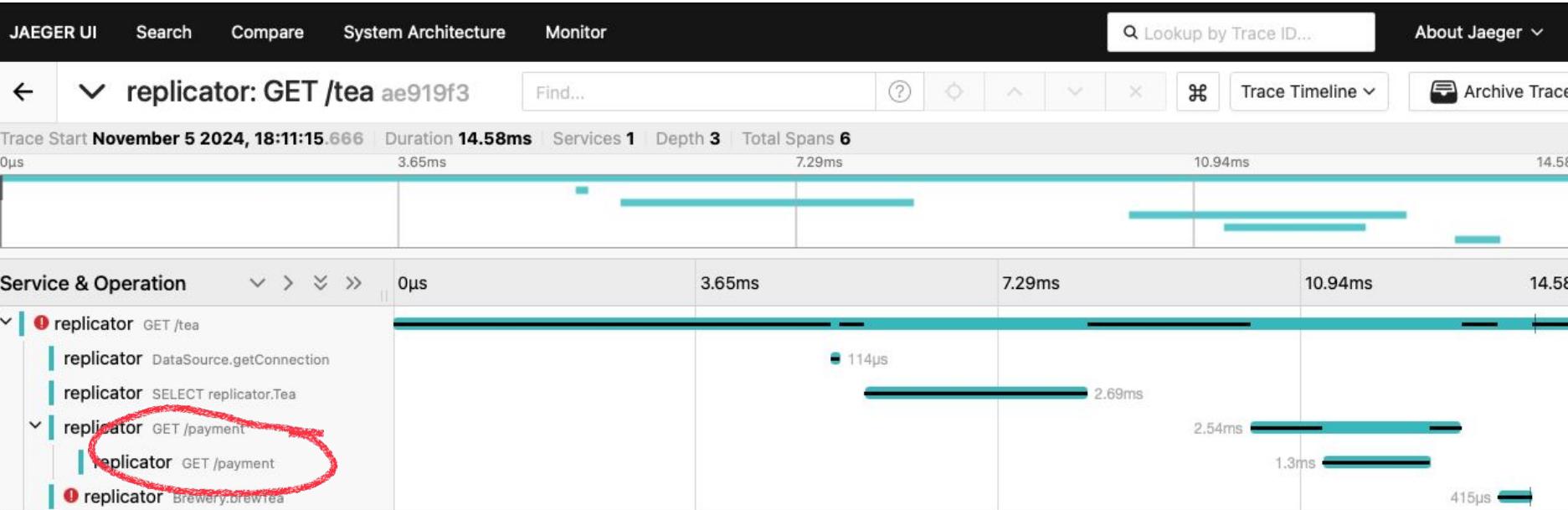
TeaResource.java

```
@RestClient  
PaymentService paymentService;
```

```
boolean paid = checkPayment(kind);  
if (!paid) {  
    throw new NotPaidException(kind);  
}
```

```
private boolean checkPayment(String kind) {  
    return Boolean.parseBoolean(paymentService.isPaid(kind));  
}
```





Geschafft für den Moment...



Python-REST-Applikation

Setup

- Neues Terminal-Fenster
- \$ cd source/python-pay
- Virtual env anlegen: \$ python3 -m venv ./venv
- Aktivieren: \$. venv/bin/activate
- Pakete installieren: \$ pip install -r requirements.txt



Start mit minimalem Server

server-minimal.py

› python3 server-minimal.py

Server started on port 8787

› curl localhost:8787

true%

127.0.0.1 - - [05/Nov/2024 22:17:06] "GET / HTTP/1.1" 200 -



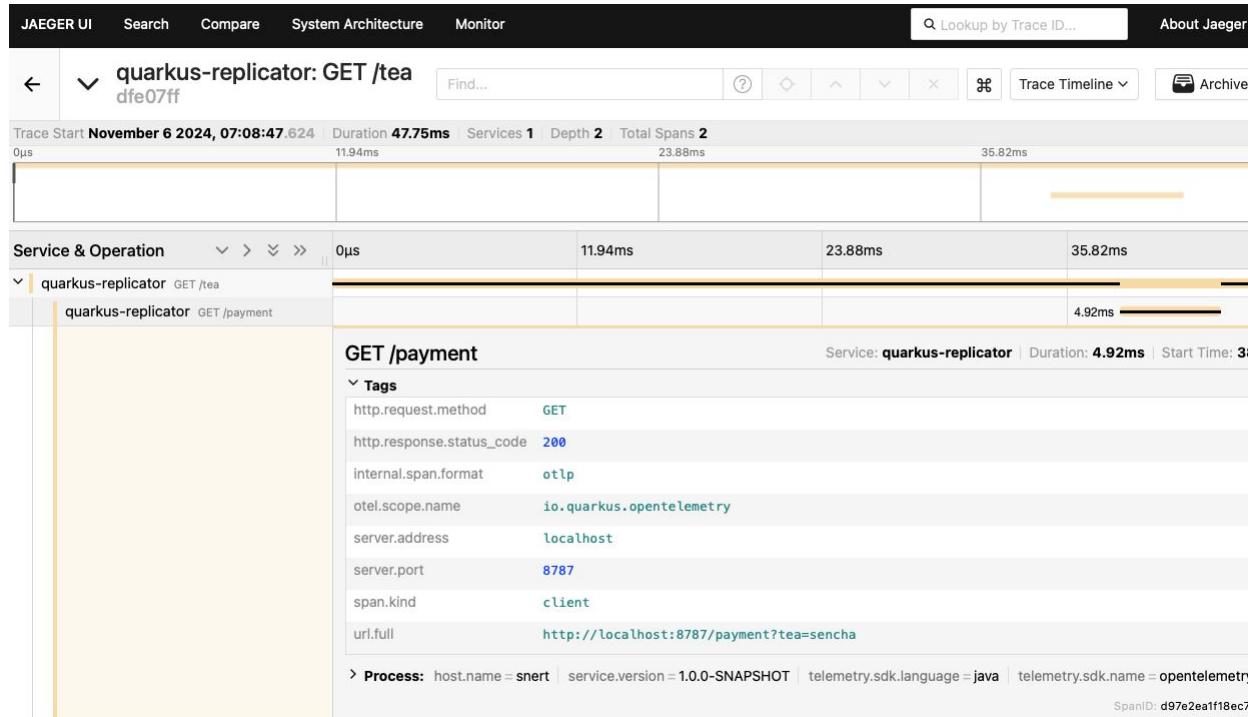
Replicator mit externen Payment

```
# The remote payment service we consult
quarkus.rest-client."de.bsd.replicator.PaymentService".url= \
http://${payment.hostport}:localhost:8080}
```

```
$ mvn quarkus:dev -Dpayment.hostport=localhost:8787
```



Wir sehen noch nix in Jaeger



Trace-Exporter

server1.py

```
# Set up exporting
resource = Resource(attributes={
    SERVICE_NAME: "payment"
})

# Configure the provider with the service name
provider = TracerProvider(resource=resource)
# We need to provide the /v1/traces part when we use the http-exporter on port 4318
# For the grpc endpoint on port 4317, this is not needed.
processor = BatchSpanProcessor(OTLPSpanExporter(endpoint="http://localhost:4317"))
provider.add_span_processor(processor)
trace.set_tracer_provider(provider)

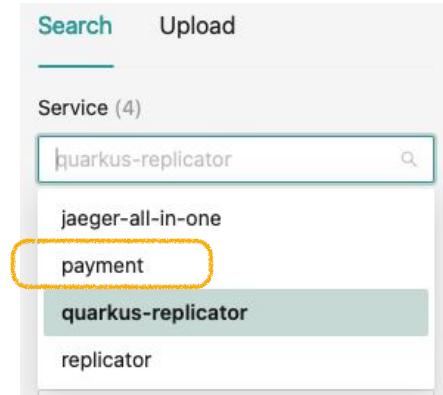
tracer = trace.get_tracer(__name__)
```



Instrumentierung von Hand

server1-1.py

```
def do_GET(self):
    ctx = {}
    with tracer.start_as_current_span("do-get-handler",
        context=ctx) as span:
        self.send_response(200)
    ...
    ...
```

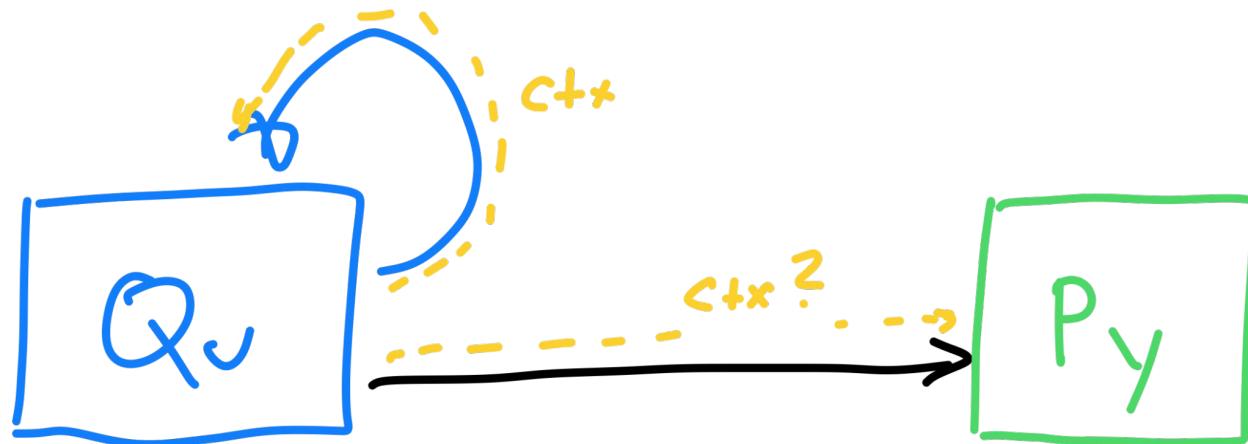


Red Hat

Warum tut das nicht?

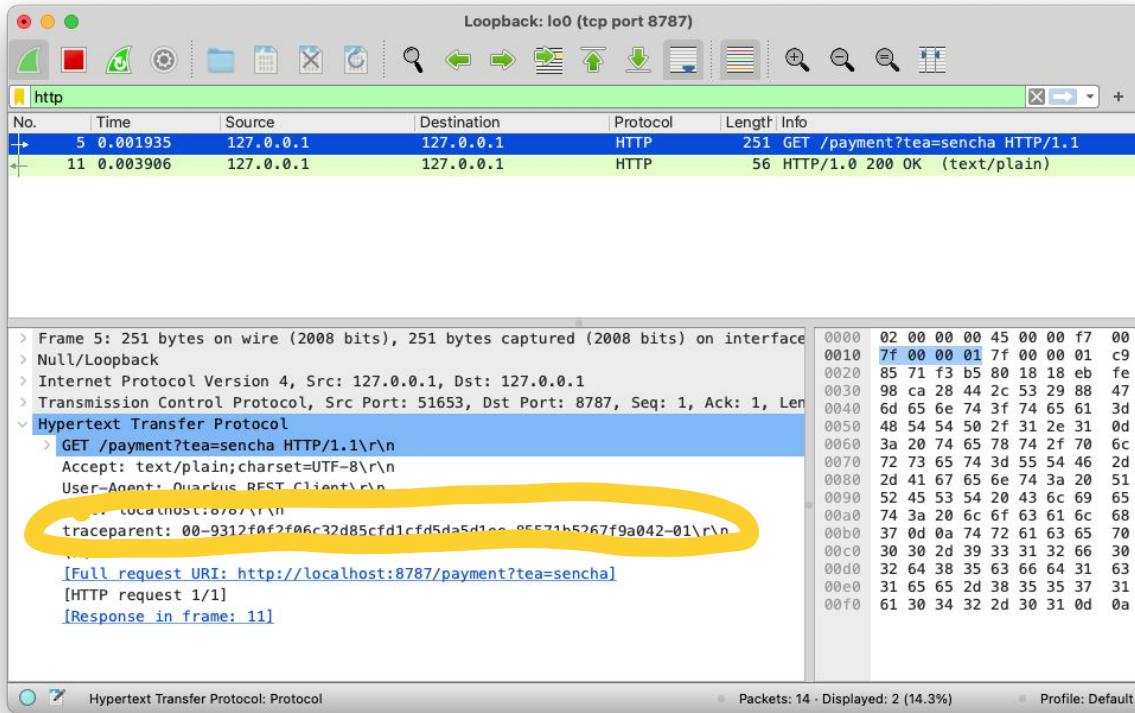
Beim Payment in Quarkus hat es doch auch getan?

- Quarkus hat sowohl REST-Sever als auch -Client instrumentiert
- Wir müssen den Kontext weiter geben



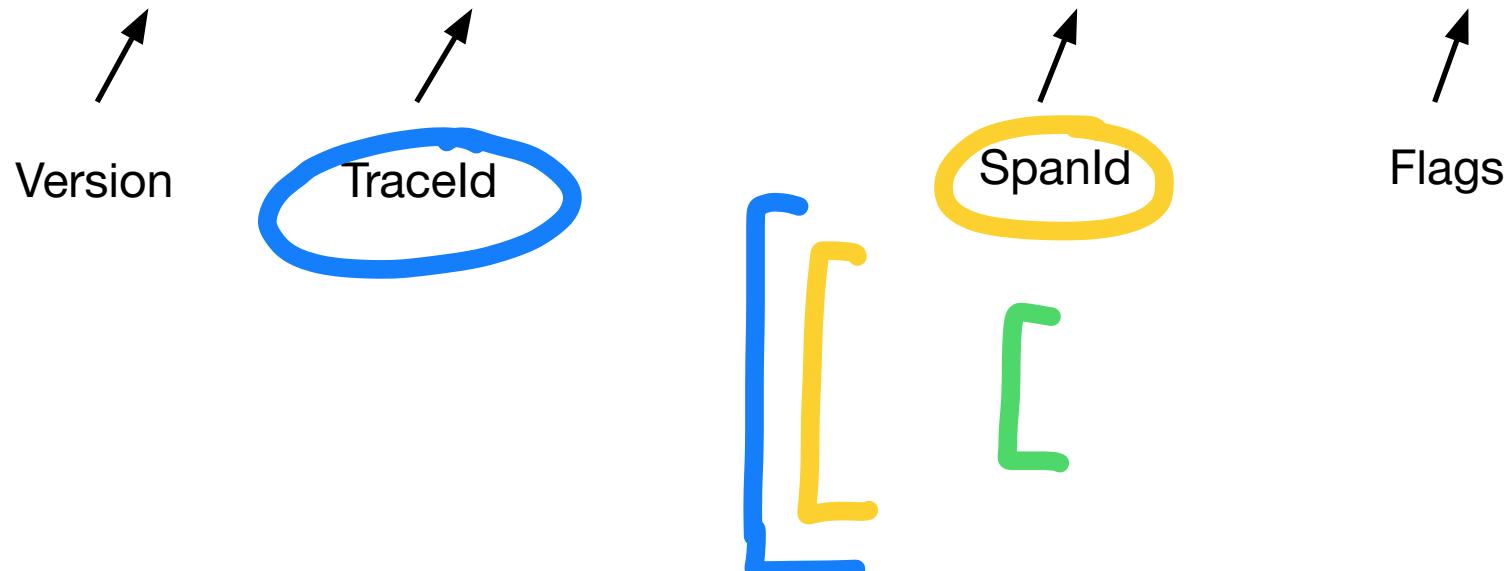
Einschub: Kontext weiterleiten

Header anschauen in Wireshark



traceparent-http—header

traceparent: 00-9312f0f2f06c32d85cf1cf5da5d1ee-85571b5267f9a042-01



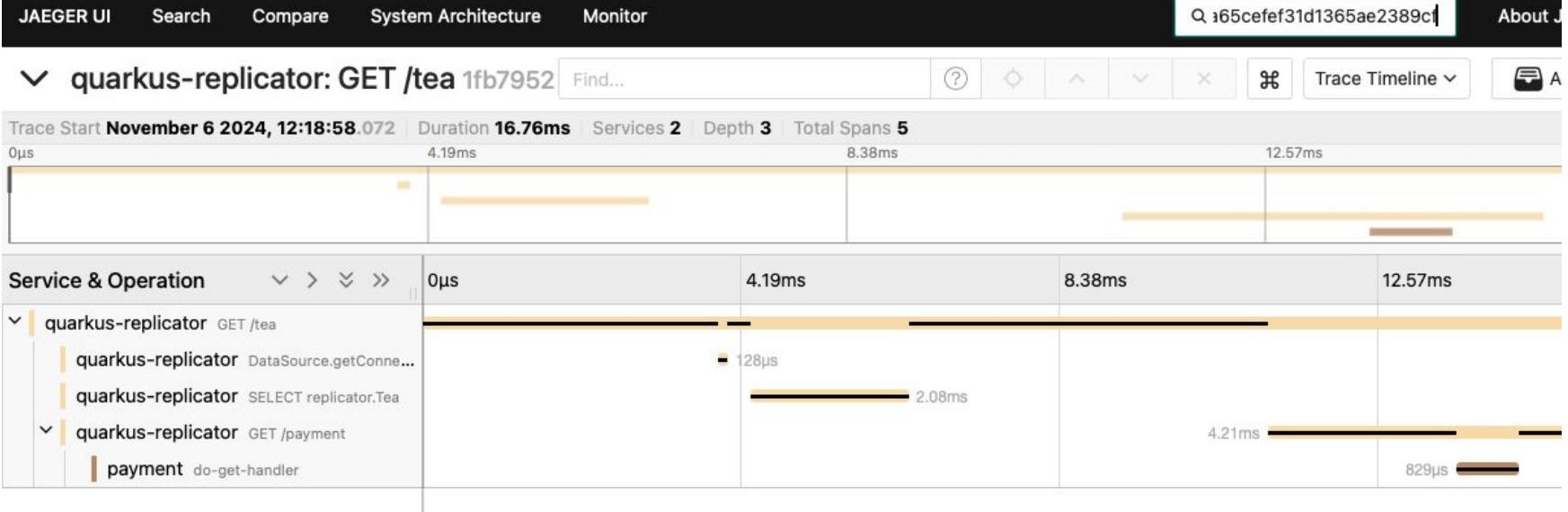
Traceparent in Code

server2.py

```
inc_trace = self.headers["traceparent"]
ctx = {}
span_context = None
if inc_trace is not None:
    print(inc_trace)
    span_context = extract_trace_data(inc_trace)
    ctx = trace
        .set_span_in_context(NonRecordingSpan(span_context))

with tracer.start_as_current_span("do-get-handler",
    context=ctx) as span:
```



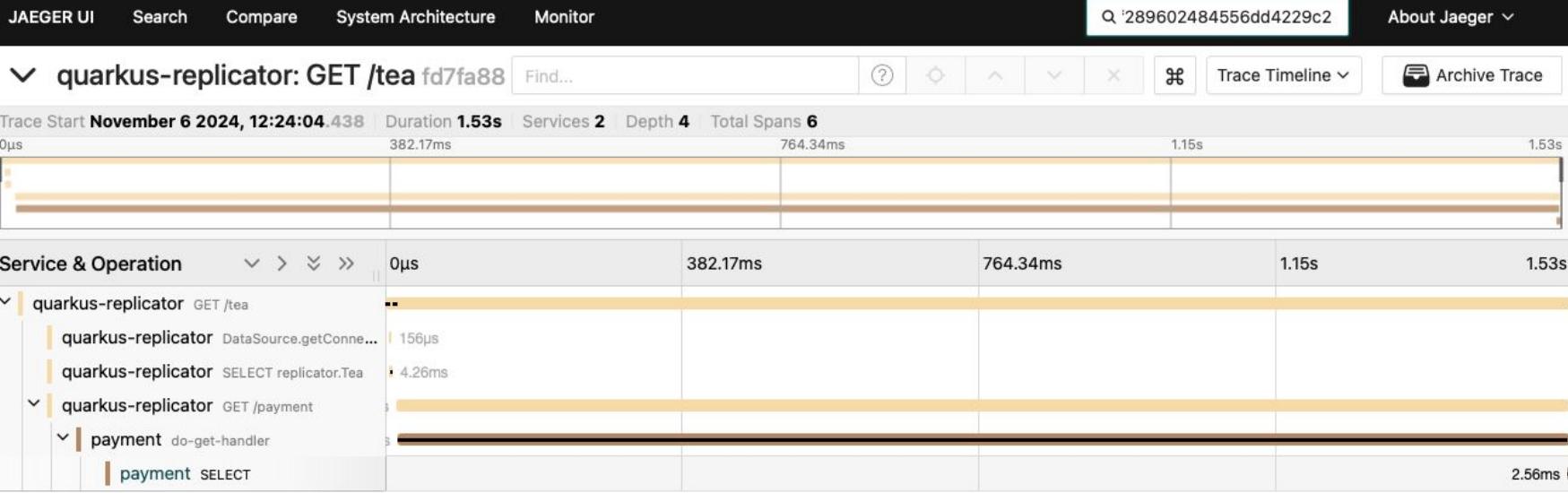


Ist der Tee bezahlt?

server3.py

```
Psycopg2Instrumentor().instrument(enable_commenter=True,  
commenter_options={})  
  
connection = psycopg2.connect(database="replicator",  
                             user="demo",  
                             password="lala7",  
                             host="localhost",  
                             port=5432)  
  
cursor = connection.cursor()  
  
cursor.execute("SELECT * FROM tea WHERE kind=%s;" , (tea, ) )
```

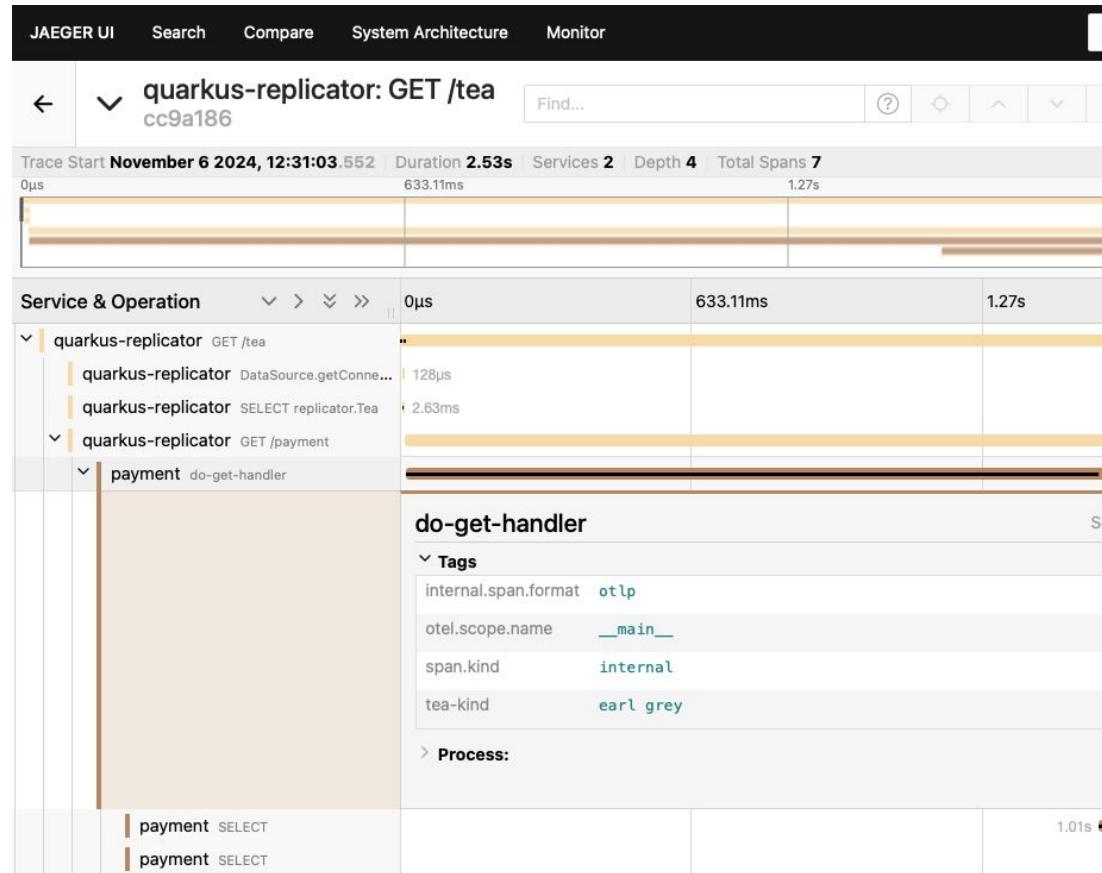




Tee-Sorte im Span festhalten

server4.py

```
span.set_attribute("tea-kind", the_tea)
```



Ein wenig Abrunden, Fehler verhindern..

server5.py

```
# We get a request like 'GET /payment?tea=sencha HTTP/1.1',
#       let's parse it down
req = self.requestline
if not '?' in req:
    span.set_attribute('request_line', req)
    self.send_response(400)
    self.end_headers()
    # This must come after sending the headers
    span.set_status(HttpStatus.ERROR, 'Bad request line')
return
```



... weiteren Server aufrufen

server5.py

```
r = requests.get('http://' + args.other_hostport)  
print(r.status_code)
```

```
RequestsInstrumentor().instrument()
```

› curl -i localhost:8787





payment: do-get-handler 91f333e

Find...

Trace Start **November 6 2024, 13:00:11.590**Duration **1.52s**Services **1**Depth **2**Total Spans **2**

0μs

379.04ms

758.07ms

Service & Operation

> < >>

0μs

379.04ms

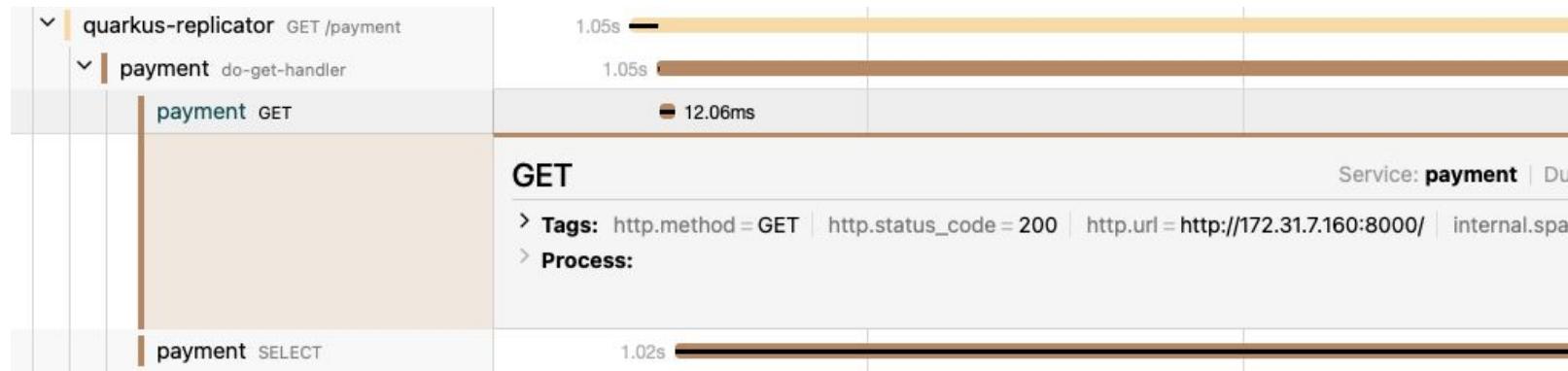
payment do-get-handler

do-get-handler

Tags

error	true
internal.span.format	otlp
otel.scope.name	__main__
otel.status_code	ERROR
otel.status_description	Bad request line

**Red Hat**



Umgebungs“variablen“

12-Faktor-App

```
python server5.py \
    -pg.host=localhost \
    -otel_host=localhost \
    -other_hostport=localhost:8000
```

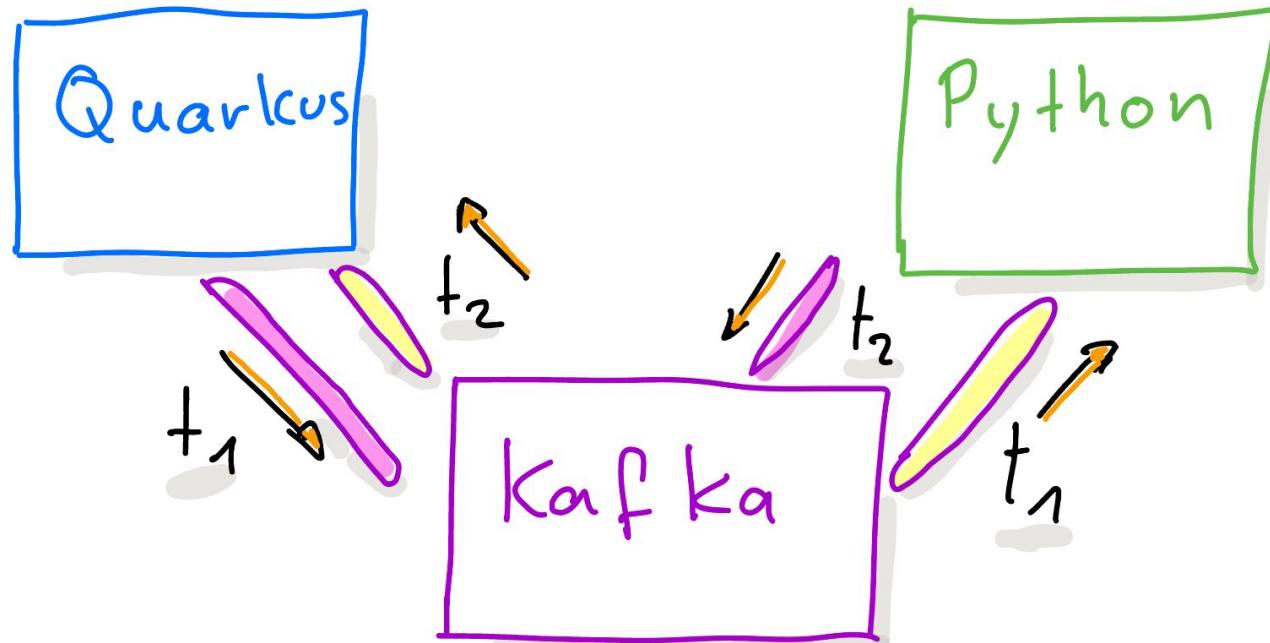


tl;dr



Wie geht das asynchron?

zB via Apache Kafka?



Quarkus

TeaResource.java

```
@Inject  
LoggerService los;  
  
los.sendLog(kind, paid);
```

```
public class LoggerService {  
  
    @Inject  
    @Channel("topic1")  
    Emitter<String> emitter;  
  
    public void sendLog(String tea,  
        boolean paid) {  
        String message = tea + ':'  
        + paid;  
        emitter.send(message);  
    }  
}
```



Kafkacat

```
$ kcat -C -t topic1 -b localhost:9092 -J
{
  "topic": "topic1",
  "Headers": [
    "traceparent", "00-8533576e564e74dca5015e3e0391f11b-131d95f9275598ad-01"
  ],
  "Key": null,
  "payload": "matcha:true"
}
```



Python

kaf-relay.py

```
consumer = KafkaConsumer('topic1')
producer = KafkaProducer()

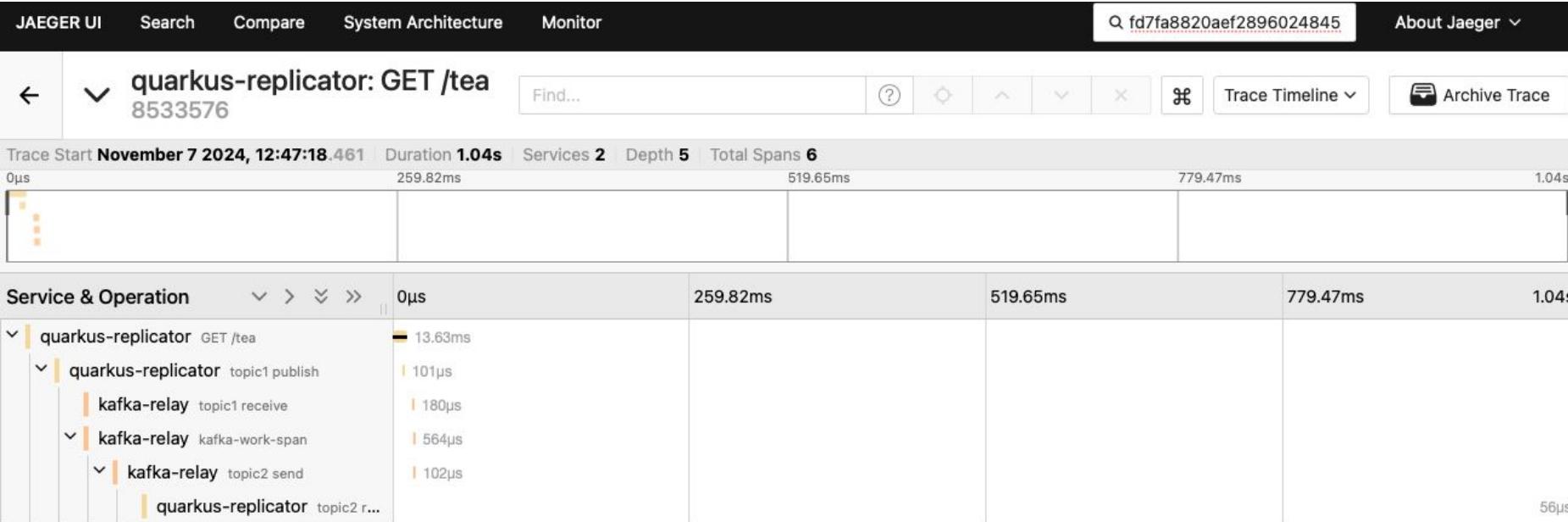
# Loop over incoming messages, process then and forward to topic2
for msg in consumer:
    trace_parent = get_trace_parent_header(msg)

    # Create a SpanContext object from the header value
    span_context = extract_trace_data(trace_parent)

    # Use this SpanContext as parent
    ctx = trace.set_span_in_context(NonRecordingSpan(span_context))

    with tracer.start_as_current_span("kafka-work-span", context=ctx)
        as span:
        # do the work
        body = msg.value.decode('utf-8')
        body = body + ' from Python'
        # and send it off to topic2
        producer.send('topic2', body.encode('utf-8'))
```





Geschafft für den Moment...

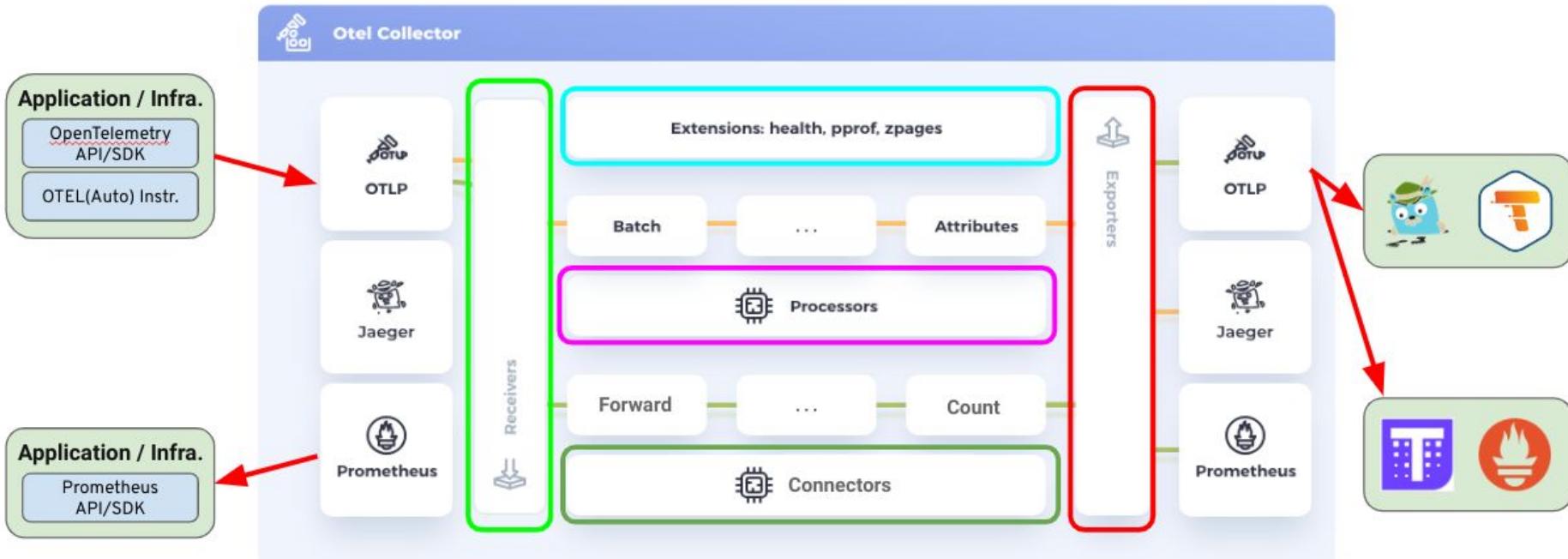


Datenerfassung mit OpenTelemetry in der Kubernetes/OpenShift Umgebung

- Überblick über das OpenTelemetry-Projekt
 - Architektur und Einsatz des OpenTelemetry Collectors
 - Konfiguration und Deployment des Collectors
- Datenerfassung in Kubernetes und OpenShift
 - Einführung des Operators: Konfiguration und Deployment
 - Integration mit Service- und PodMonitors

OpenTelemetry Collector

Architektur



OpenTelemetry Collector

Konfiguration

```
1 ---
2 receivers:
3   otlp:
4     protocols:
5       grpc:
6       http:
7
8 exporters:
9   prometheus:
10    endpoint: "0.0.0.0:8889"
11
12 otlp:
13   endpoint: jaeger:4317
14   tls:
15     insecure: true
16
17 connectors:
18   spanmetrics:
19
20 processors:
21   batch:
22
23 service:
24   pipelines:
25     traces:
26       receivers: [otlp]
27       processors: [batch]
28       exporters: [spanmetrics, otlp]
29     metrics/spanmetrics:
30       receivers: [spanmetrics]
31       exporters: [prometheus]
```

receivers:
prometheus: **(Optional)**
config:
scrape_configs:
- job_name: 'otel-collector'
scrape_interval: 5s
static_configs:
- targets: ['0.0.0.0:8888']



OpenTelemetry Collector

Simple - Konfiguration

```
---
```

```
receivers:
  otlp:
    protocols:
      http:
        endpoint: 0.0.0.0:4318
processors:
  batch:

exporters:
  debug:

service:
  pipelines:
    traces:
      receivers: [otlp]
      processors: [batch]
      exporters: [debug]
```

<https://opentelemetry.io/docs/collector/configuration/>



OpenTelemetry Collector

Konfiguration - Lokal testen

```
podman run --rm -it --network=host  
ghcr.io/open-telemetry/opentelemetry-collector-contrib/telemetrygen traces --otlp-insecure  
--otlp-http --duration 45s --rate 1 --otlp-endpoint="127.0.0.1:4318"
```

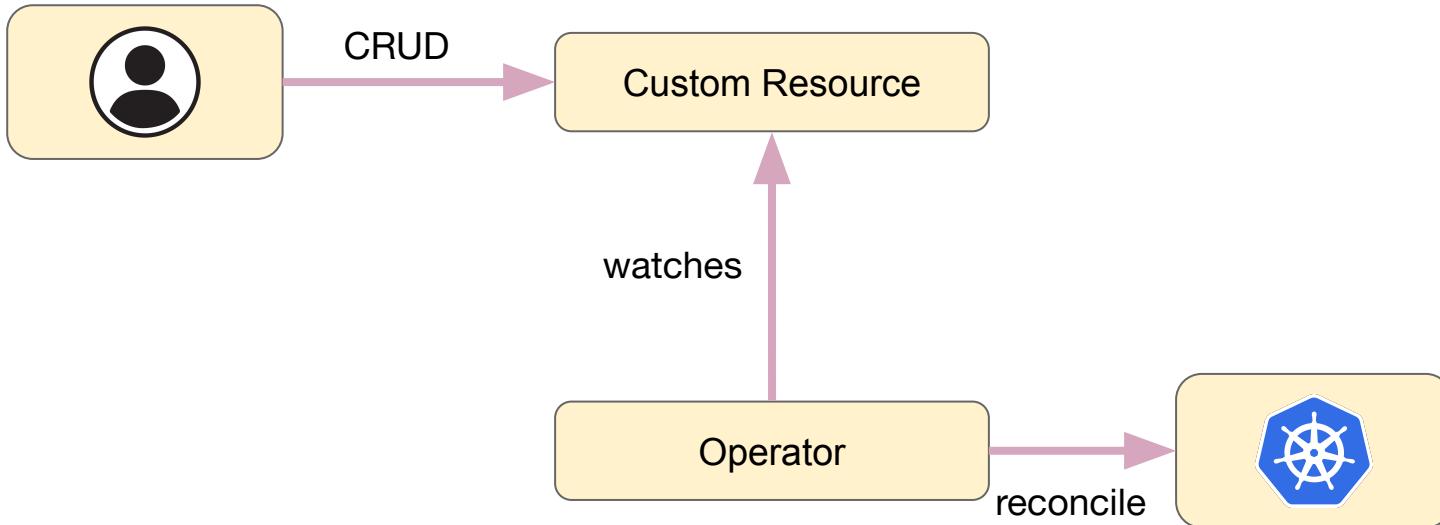
```
podman run --rm -it --network=host -v $(pwd)/config.yaml:/etc/otelcol-contrib/config.yaml:z  
ghcr.io/open-telemetry/opentelemetry-collector-releases/opentelemetry-collector-contrib:0.113.0
```

```
...  
2024-11-10T22:06:51.127Z info service@v0.113.0/service.go:261 Everything is ready. Begin running and processing data.  
2024-11-10T22:07:02.752Z info Traces {"kind": "exporter", "data_type": "traces", "name": "debug", "resource spans": 1, "spans": 2}  
2024-11-10T22:07:03.755Z info Traces {"kind": "exporter", "data_type": "traces", "name": "debug", "resource spans": 1, "spans": 2}  
2024-11-10T22:07:05.759Z info Traces {"kind": "exporter", "data_type": "traces", "name": "debug", "resource spans": 1, "spans": 2}  
...
```



OpenTelemetry Collector

Operator



OpenTelemetry Operator

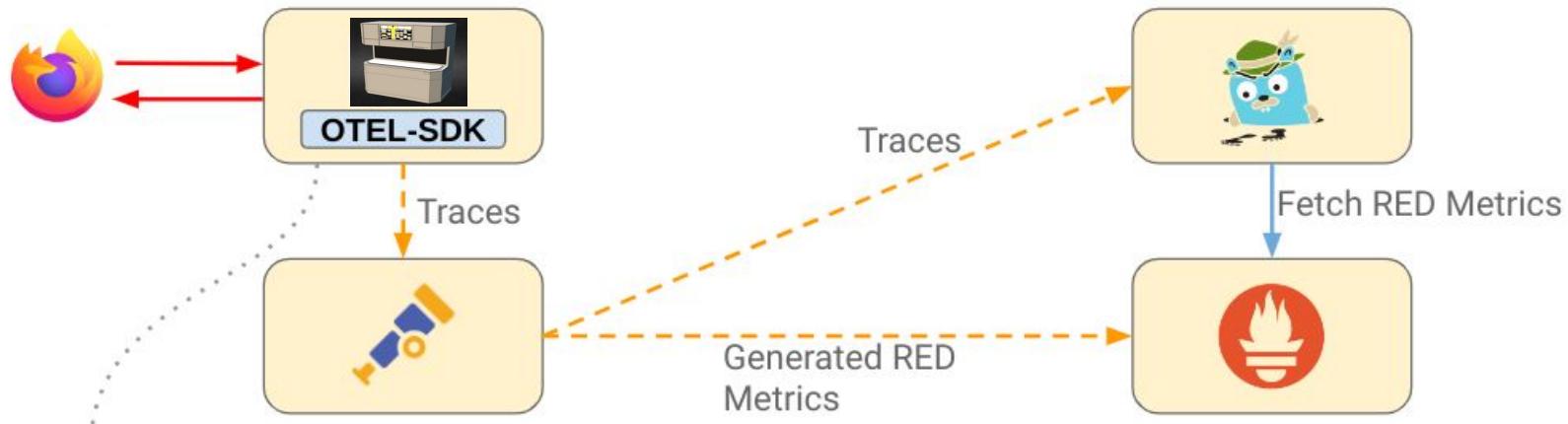
- Stellt Schnittstellen bereit:
 - opentelemetry.io/v1beta1
 - OpenTelemetryCollector
 - opentelemetry.io/v1alpha1
 - OpenTelemetryCollector
 - Instrumentation
 - OpAMPBridge
 - Sampling (in development)

OpenTelemetry Operator

- Wechsel zu Repo

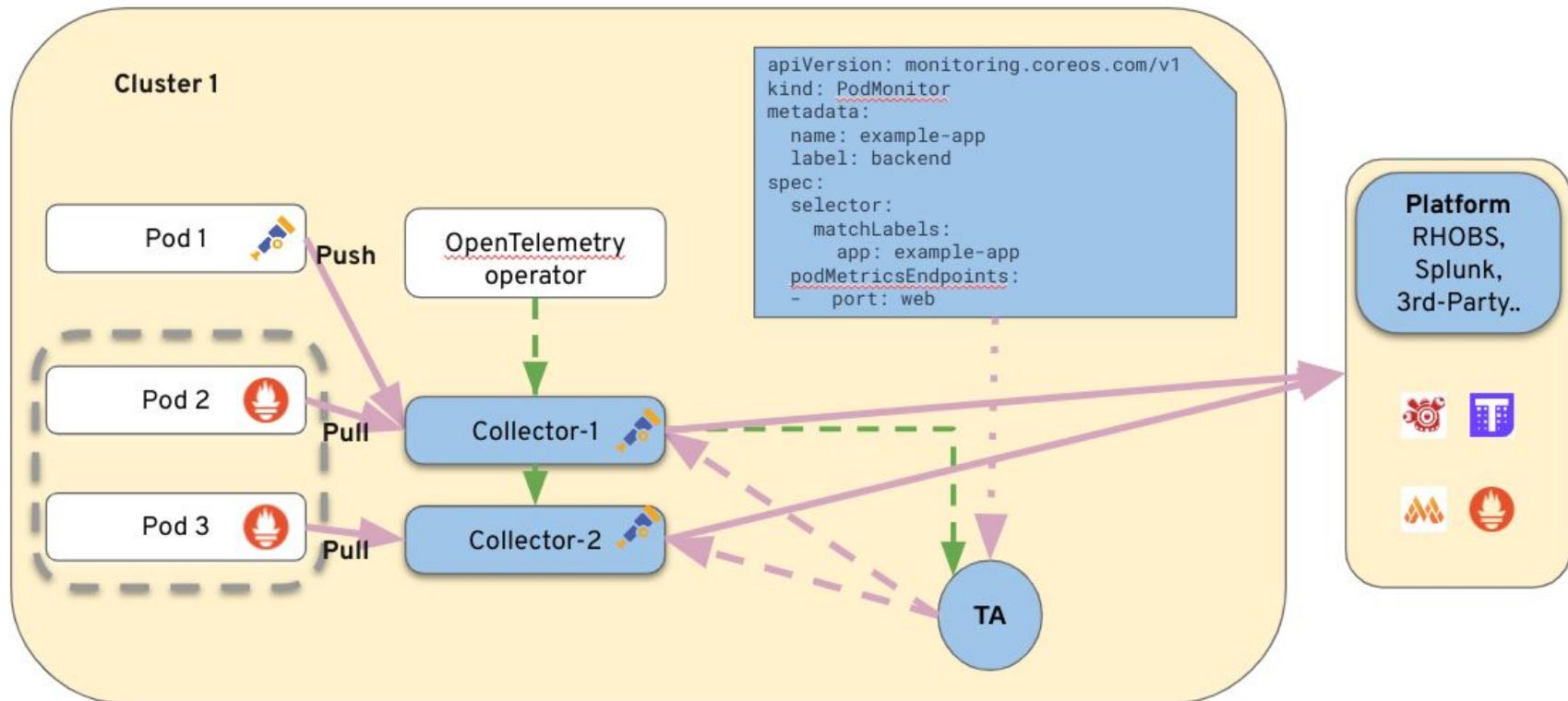
<https://github.com/frzifus/ContainerConf-Workshop-2024/blob/main/30-otel-k8s.md>

OpenTelemetry Collector



```
with tracer.start_as_current_span("do-get-handler",  
    context=ctx) as span:
```

OpenTelemetry Collector



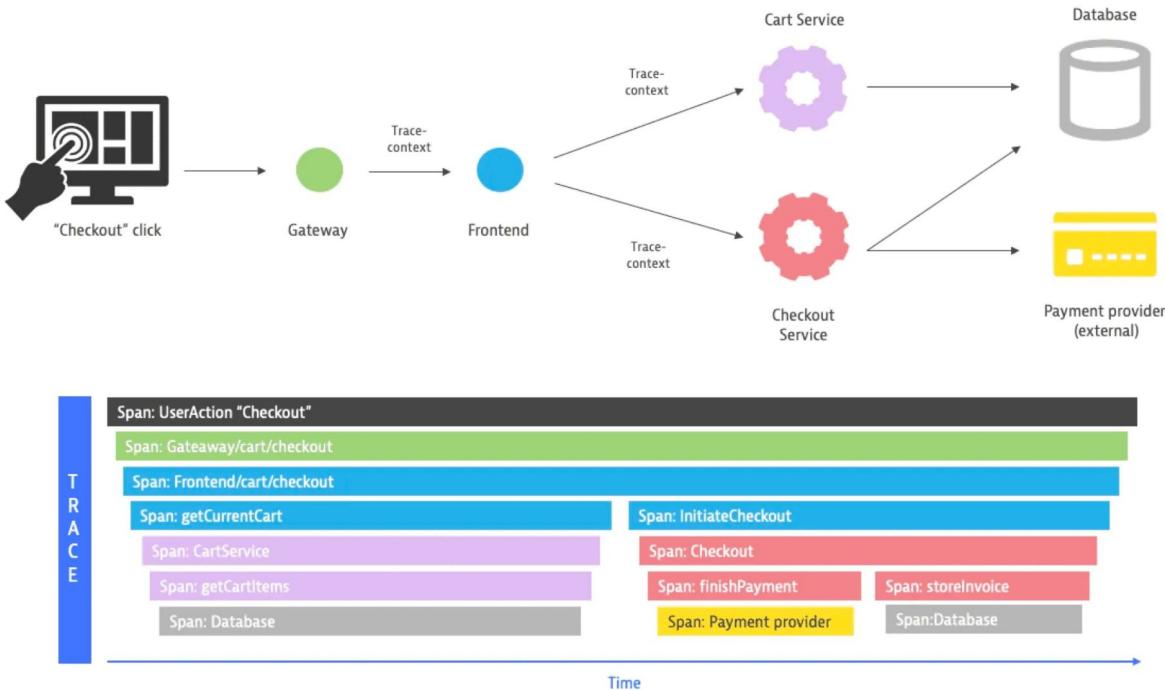
Geschafft für den Moment...



Sampling und Filtering: Optimierung der Observability in verteilten Systemen

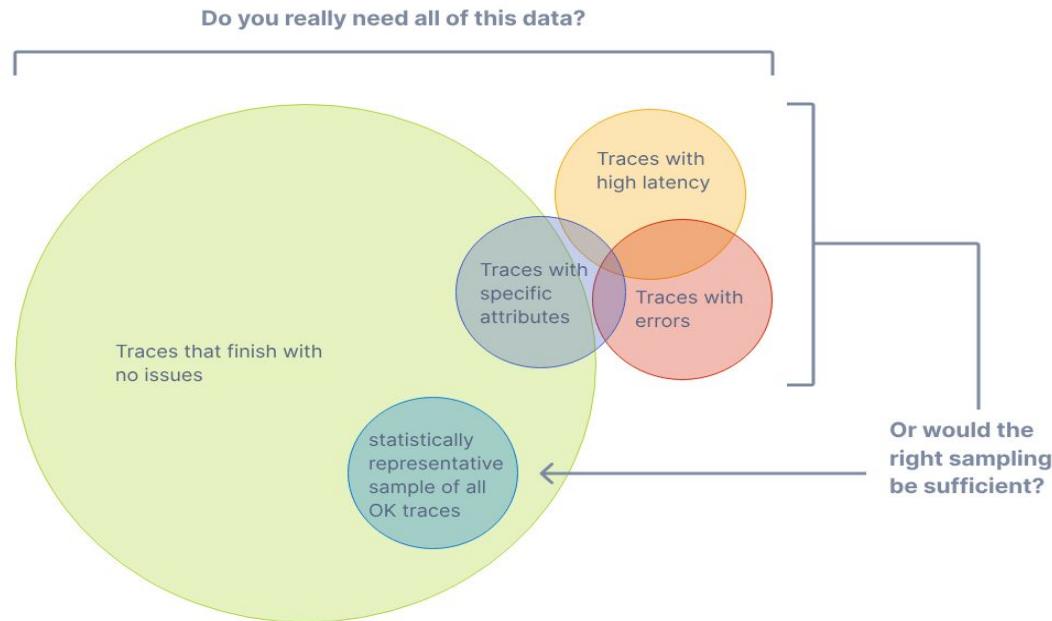
- Finale Theorierunde
 - Was ist sampling und warum sollte es uns interessieren?
 - Konfiguration des Tailbased sampling processors
 - Probleme und Lösungen
 - Demo setup

Sampling Theorie



<https://henesgokdag.medium.com/distributed-tracing-9300d55e7245>

Sampling Theorie



<https://opentelemetry.io/docs/concepts/sampling/>

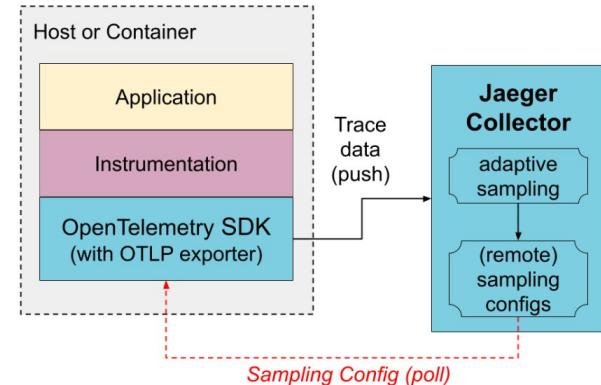
Sampling Theorie

- Setup (production environment, telco workloads)
 - ~30 Microservices, ~110 Nodes, ~2350 Pods, ~940 CPU
- Kostenberechnung bei der Speicherung aller Telemetriedaten
 - $1.100.000 \text{ traces/min} \rightarrow 1.100.000 * 60 * 24 * 30 * 100\% = 47.500.200.000$
 - = 237.600 \$ (27.01.2024, region=eu-west-1)
- Wie lassen sich die Unverhältnismäßig hohe Kosten reduzieren?
Wünschenswert wäre eine Reduktion auf die **~0.1%** relevanten Daten.

Sampling Theorie

Head-based sampling

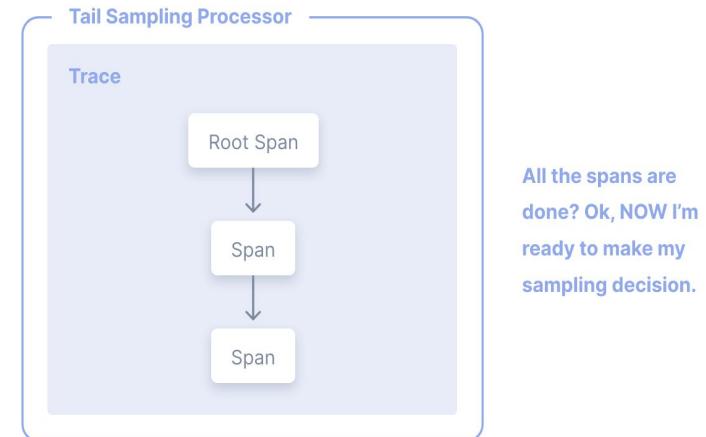
- Sampling entscheidung wird bei **beginn** des Traces getroffen und über die services hinweg propagiert.
- Effizient, einfach zu verstehen und zu konfigurieren
- Konfigurationsoptionen sind in SDKs verfügbar.
 - Manuell z.B. über Umgebungsvariablen
 - Jaeger Remote Sampling-Erweiterung ([docs](#))
- Alternativer Probabilistischer Probenehmer-Prozessor ([docs](#))



Sampling Theorie

Tail-based sampling

- Die Entscheidung über Sample wird am **Ende** eines traces getroffen
- Eine Informierte entscheidung kann nur getroffen werden, wenn alle spans zu einem trace vorliegen
- Ermöglicht komplexe Richtlinien
- **Verbraucht zusätzliche Ressourcen**
- Verwende tail-basiertes Sampling, wenn seltene oder extreme Fälle untersucht werden sollen, welche erhebliche Auswirkungen haben könnten oder besondere Aufmerksamkeit erfordern



<https://opentelemetry.io/docs/concepts/sampling#tail-sampling>

Sampling Theorie

Kostenreduktion

- Calculated cost telo/xray setup - 1.1 Mio traces per Minute
 - **100%** sampling => 237.600 \$
 - **0.1%** sampling => 237 \$ + (sampling cost)
- Zusätzliche Kosten für Tailbased Sampling
 - (Cluster resource limits 15000Mi + 6500m CPU)
 - Instanz optionen
 - Type: t2.xlarge per H: 0,1856 USD Cores: 4 Mem: 16 GiB
 - Xray(237 \$) + t2.xlarge(133 \$) = **370 \$**
 - Type: t2.2xlarge per H: 0,3712 USD Cores: 8 Mem: 32 GiB
 - Xray(237 \$) + t2.2xlarge(267 \$) = **504 \$**

Sampling Theorie

Konfiguration

```
apiVersion: opentelemetry.io/v1alpha1
kind: OpenTelemetryCollector
metadata:
  name: sampling-col
spec:
  mode: deployment
  replicas: 5
  resources: ...
  config: |
    receivers:
      otlp:
    processors:
      memory_limiter: ...
      batch/traces: ...
      tail_sampling: ...
    exporters:
      otlp/tempo: ...
  service:
    pipelines:
      traces:
        receivers: [otlp]
        processors: [memory_limiter, batch/traces, tail_sampling]
        exporters: [otlp/tempo]
```

```
resources
requests:
  memory: "3000Mi"
  cpu: "1300m"
limits:
  memory: "3000Mi"
  cpu: "1300m"

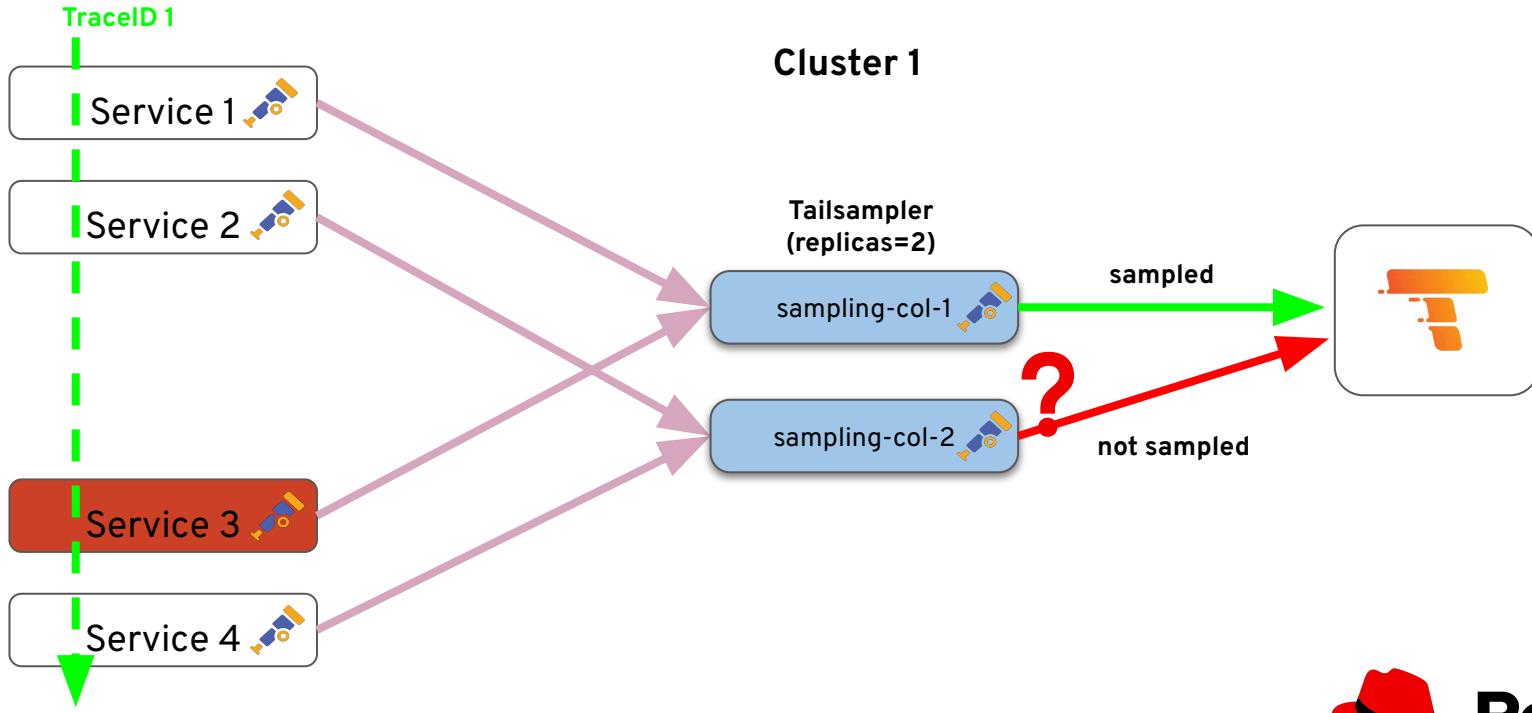
tail_sampling:
  decision_wait: 10s
  num_traces: 100000
  policies
    - name: policy-errors-retain
      type: status_code
      status_code: {status_codes: [ERROR]}
    - name: policy-probabilistic
      type: probabilistic
      probabilistic:
        sampling_percentage: 10
```



Red Hat

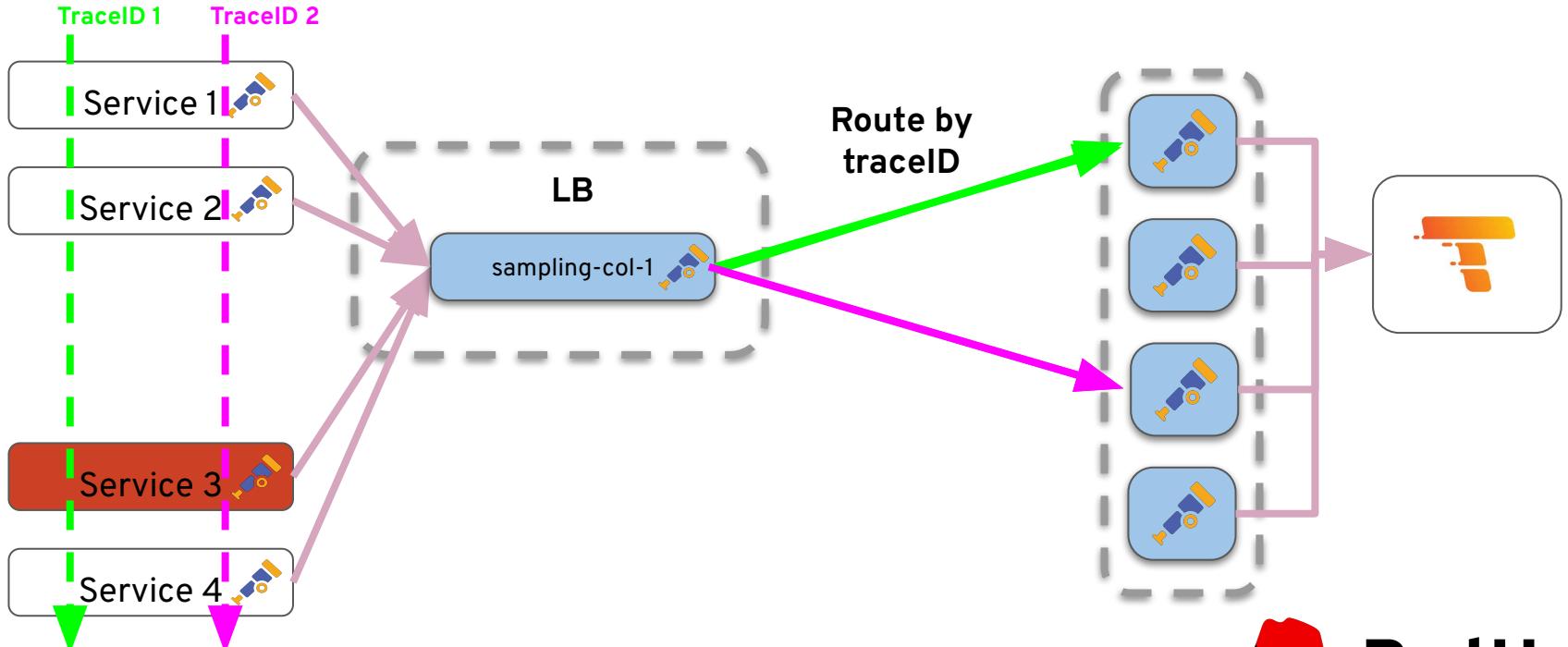
Sampling Theorie

Probleme



Sampling Theorie

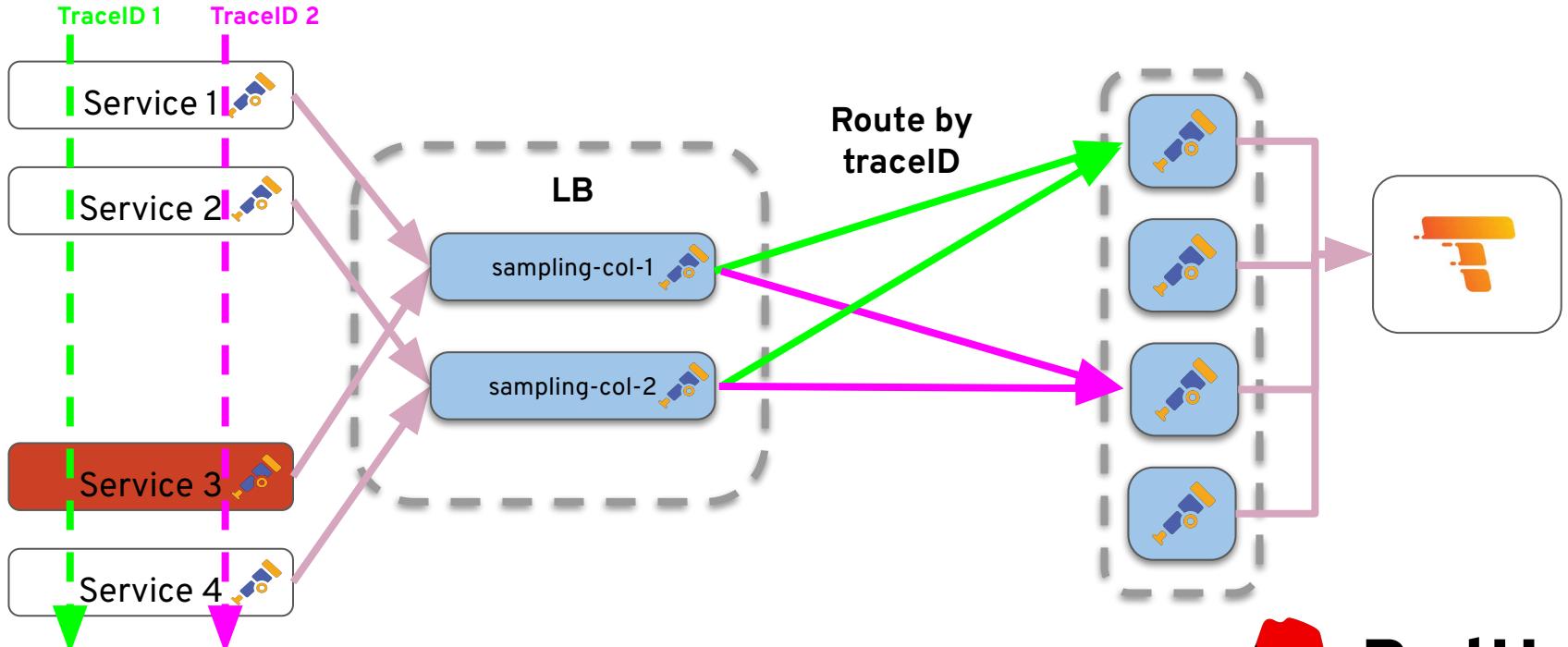
Scaling



Red Hat

Sampling Theorie

Scaling



Red Hat

Sampling Theorie

Load-balancing exporter

```
apiVersion: opentelemetry.io/v1alpha1
kind: OpenTelemetryCollector
metadata:
  name: telemetry-lb
spec:
  mode: deployment
  replicas: 2
  config: |
    exporters:
      loadbalancing:
        routing_key: traceID
        protocol:
          otlp:
            sending_queue
            queue_size: 4000
      resolver:
        dns:
          hostname: sampling-collector-headless.observability-backend.svc.cluster.local
          port: 4317
  service:
    pipelines:
      traces:
        exporters: [loadbalancing]
```

```
resolver:
  k8s:
    service: lb-svc.kube-public
    ports:
      - 15317
      - 16317
  resolver
    static:
      hostnames:
        - backend-1:4317
        - backend-2:4317
        - backend-3:4317
        - backend-4:4317
```



Geschafft für den Moment...



Teilnehmer Projekt

- TBD

OpenTelemetry Collector

Edge

