



# Distributed tracing in action at MAX IV

Anton Joubert

# Overview



Intro

(Too many!) Examples

Performance impact

Compute resources for backend

Issues / ideas for improvement

Outro

# Introduction

# Overview

## *Observability*

Understand a complex system from outside  
Traces, metrics, logs

## *OpenTelemetry*

Observability framework  
Vendor agnostic  
Many programming languages

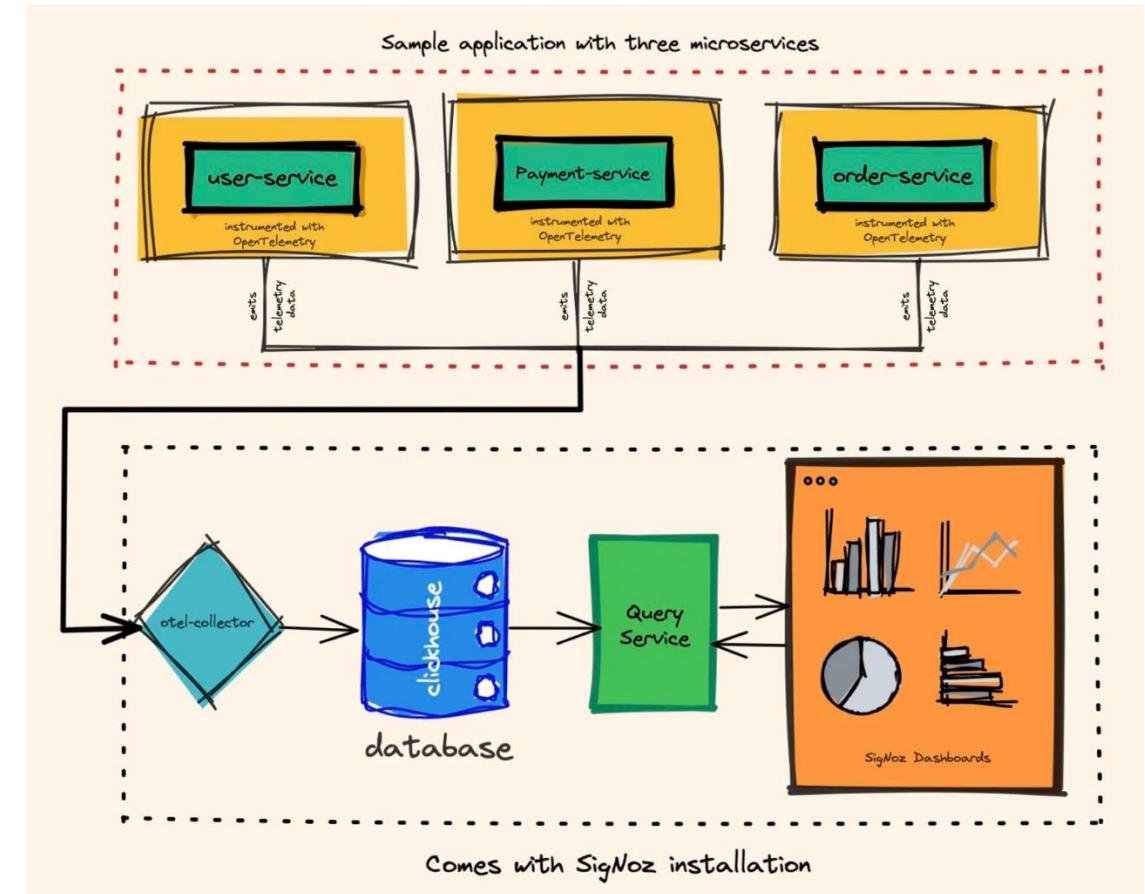


Image credit: <https://signoz.io/blog/opentelemetry-backend/>

# Grafana Tempo and Loki

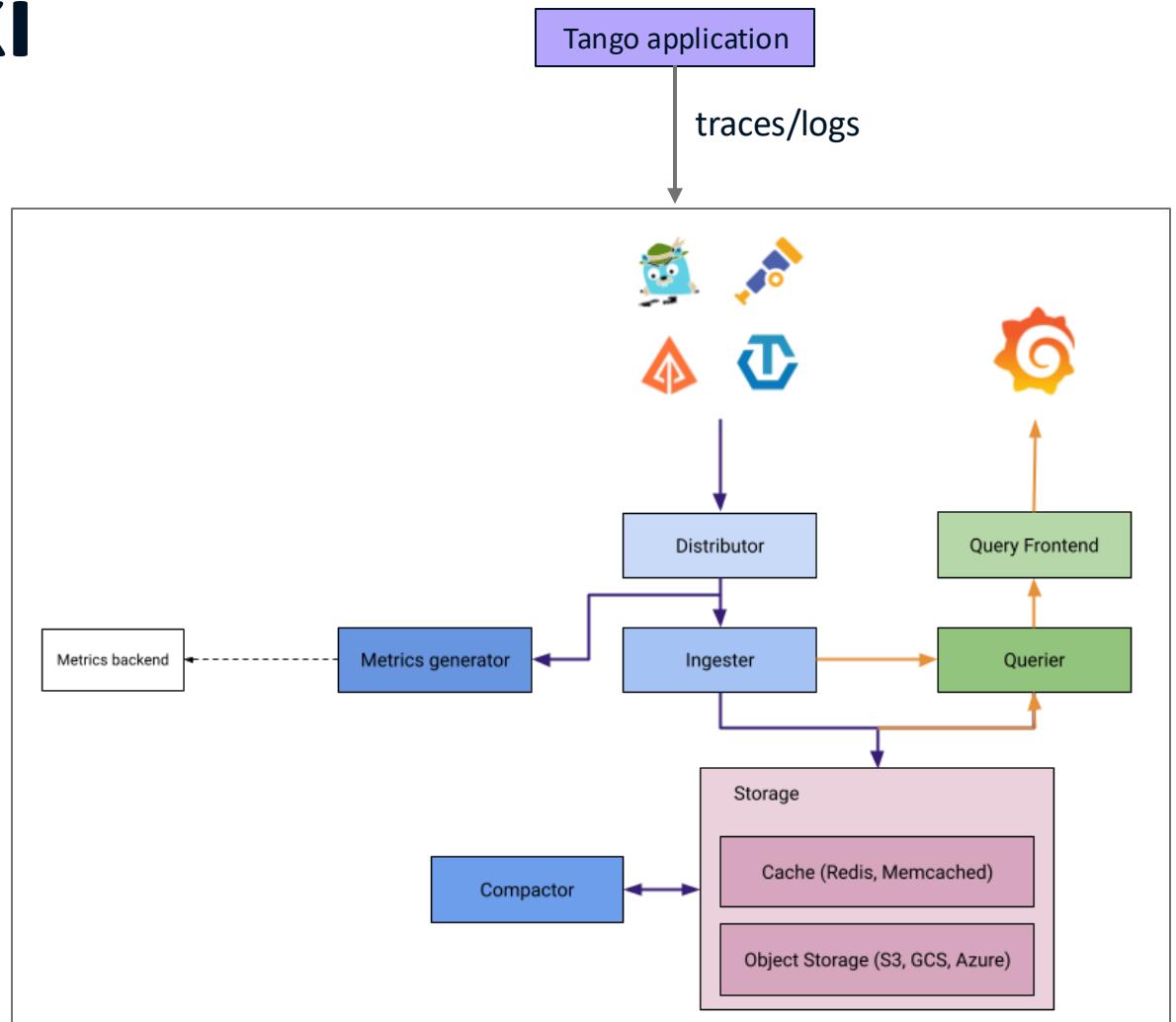
*Tempo*

Backend for traces and metrics

*Loki*

Backend for logs

Image credit: <https://grafana.com/docs/tempo/latest/operations/architecture/>



# Enabling in Tango

Available since Tango 10.0.0

Set environment variables, for example:

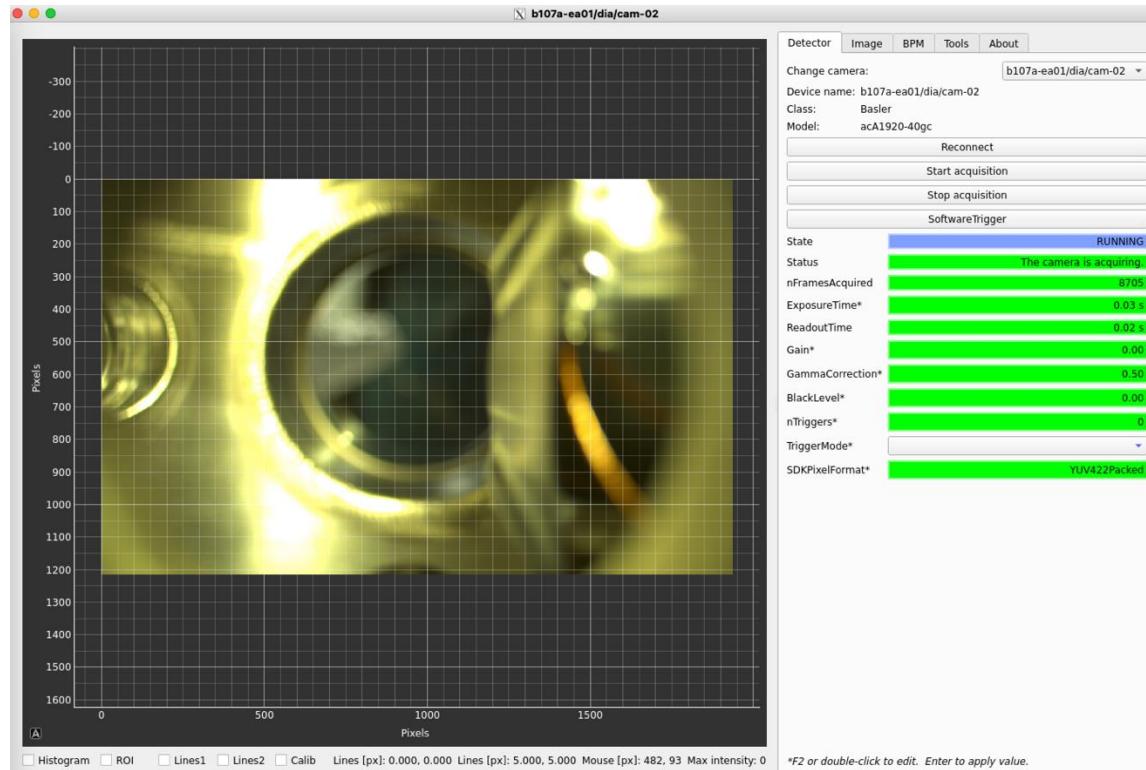
```
TANGO_TELEMETRY_ENABLE=on
TANGO_TELEMETRY_TRACES_EXPORTER=http
TANGO_TELEMETRY_TRACES_ENDPOINT=https://traces.institute.org:443/v1/traces
TANGO_TELEMETRY_LOGS_EXPORTER=http
TANGO_TELEMETRY_LOGS_ENDPOINT=https://logs.institute.org:443/otlp/v1/logs
```

Can also use HTTP, or GRPC for transport

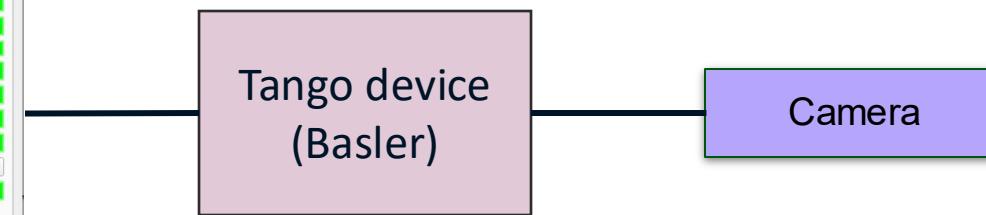
Can set environment variables in /etc/tangorc

# Examples

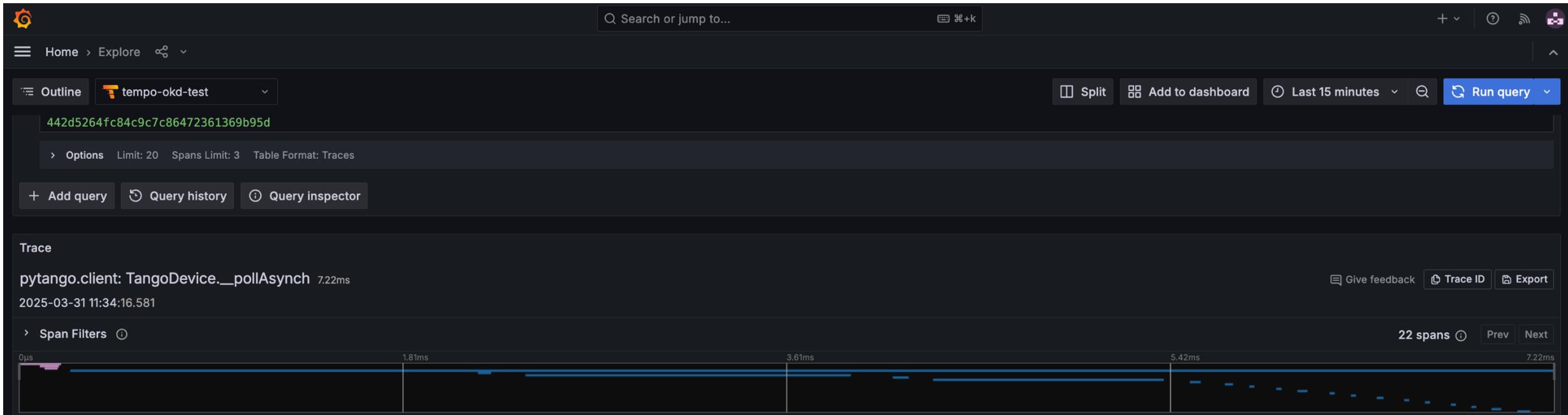
# GUI and Tango device for camera



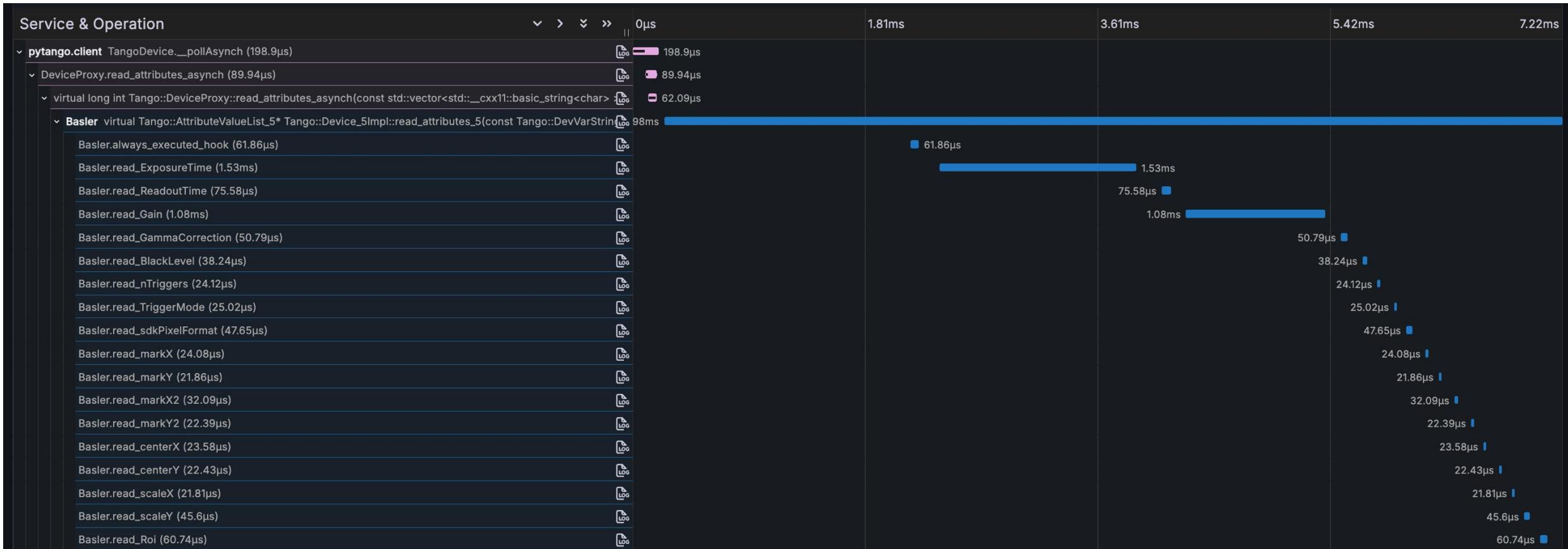
Taurus GUI (Lux Viewer)



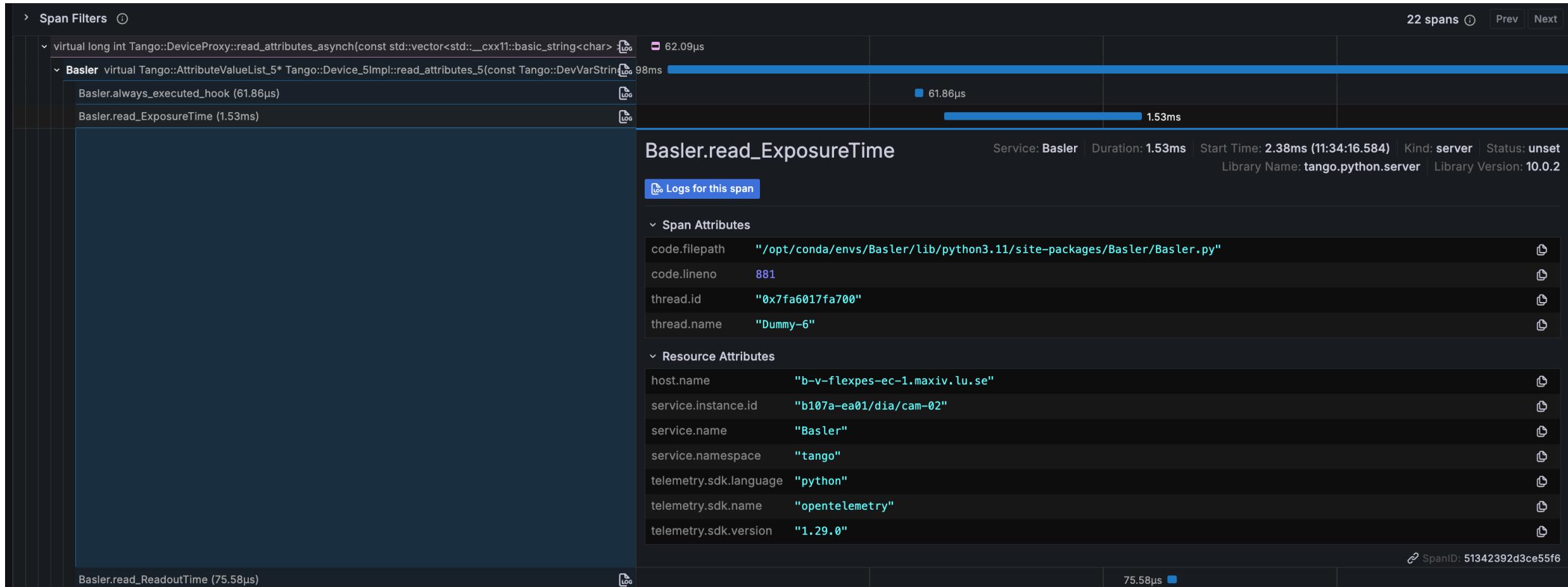
# Trace overview



# Trace detail



# Span detail



# Simple search

The screenshot shows the Tempo UI interface for searching traces. The search bar at the top contains the text "tempo-okd-test". The search panel below has the following filters applied:

- Service Name: Basler
- Span Name: Select value
- Status: Select value
- Duration: span > 10ms
- Tags: span Select tag = Select value

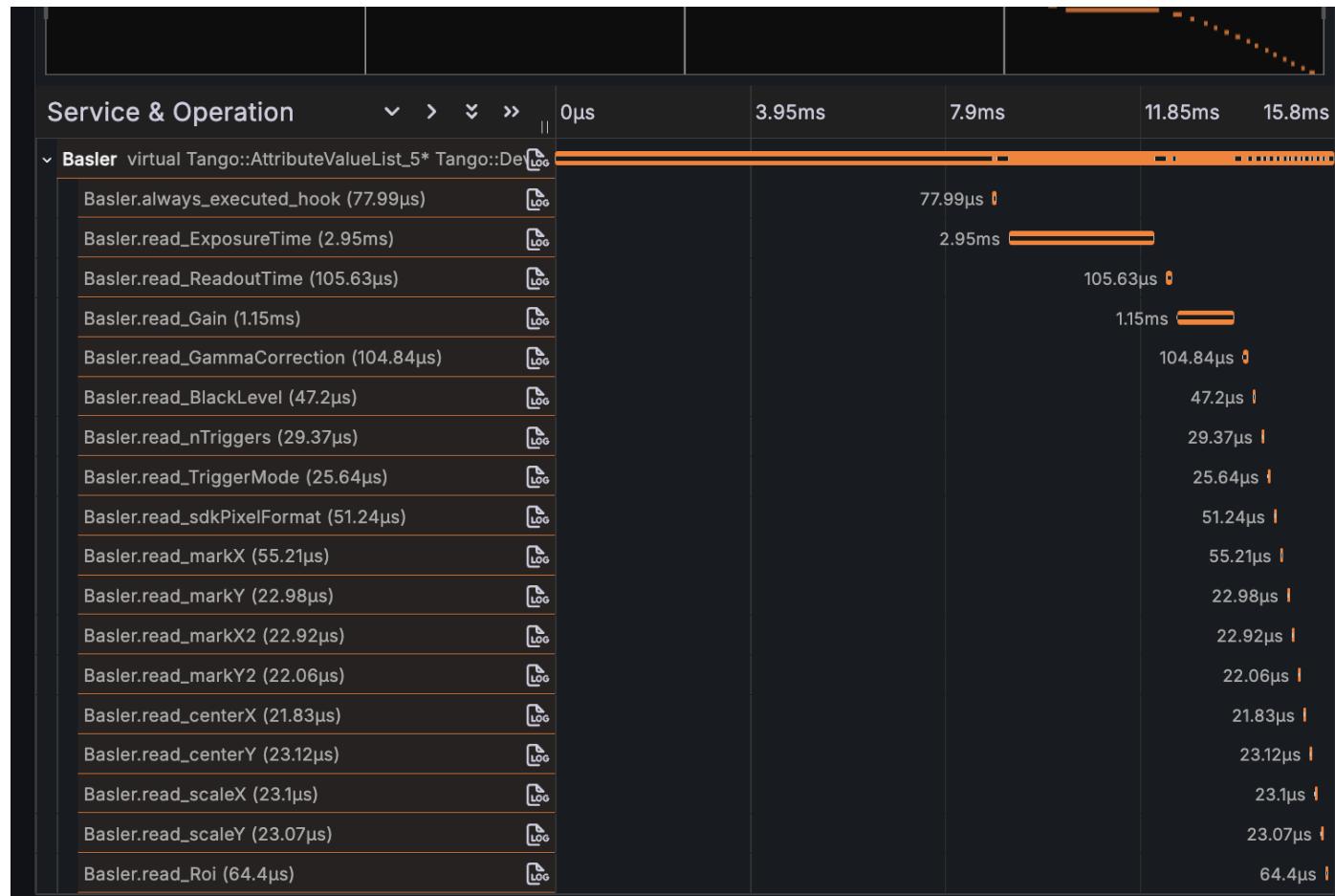
The TraceQL query generated by these filters is: `{resource.service.name="Basler" && duration>10ms}`. Below the search panel, there are options for "Options", "Limit: 20", "Spans Limit: 3", and "Table Format: Traces".

At the bottom, there are buttons for "+ Add query", "Query history", and "Query inspector".

The "Table - Traces" section displays the results:

Trace ID	Start time	Service	Name	Duration
516ae0ad947812d0520...	2025-03-31 11:47:32	Basler	virtual Tango::Attribute	10 ms
3b92629b059e004e4e...	2025-03-31 11:47:23	Basler	virtual Tango::Attribute	15 ms
1b15aaed97336d67dde...	2025-03-31 11:42:14	Basler	virtual Tango::Attribute	11 ms

# Slow traces



# Slow traces

Service & Operation	0µs	6.96ms	13.92ms	20.88ms	27.84ms
Basler virtual Tango::AttributeValueList_5* Tango::De	0µs	6.96ms	13.92ms	20.88ms	27.84ms
Basler.always_executed_hook (59.91µs)	59.91µs				
Basler.read_ExposureTime (1.14ms)	1.14ms				
Basler.read_ReadoutTime (56.97µs)	56.97µs				
Basler.read_Gain (1.11ms)	1.11ms				
Basler.read_GammaCorrection (44.79µs)	44.79µs				
Basler.read_BlackLevel (35.87µs)	35.87µs				
Basler.read_nTriggers (155.33µs)		155.33µs			
Basler.read_TriggerMode (36.56µs)		36.56µs			
Basler.read_sdkPixelFormat (78.57µs)		78.57µs			
Basler.read_markX (29.85µs)		29.85µs			
Basler.read_markY (26.71µs)		26.71µs			
Basler.read_markX2 (26.31µs)		26.31µs			
Basler.read_markY2 (27.81µs)		27.81µs			
Basler.read_centerX (26.96µs)		26.96µs			
Basler.read_centerY (26.41µs)		26.41µs			
Basler.read_scaleX (26.43µs)		26.43µs			
Basler.read_scaleY (25.49µs)		25.49µs			
Basler.read_Roi (72.91µs)		72.91µs			

# Logs

The image shows two side-by-side log analysis interfaces. Both have their respective titles highlighted with orange boxes.

**Left Panel (Tempo):**

- Title:** tempo-rancher-obs-test
- Trace View:** Shows a trace for "pytango.client: TaurusCommandButton.\_onClicked" with a duration of 15.56ms. It displays 7 spans with timing details: 0μs, 3.89ms, 7.78ms, 11.67ms, and 15.56ms. A "Logs" icon is highlighted with an orange box at the bottom of the trace view.
- Service & Operation View:** Shows a tree structure of operations. One node under "virtual Tango::DeviceData Tango::Connection" is expanded, showing three sub-nodes: "Basler", "Basler.always\_executed\_hook (57.34μs)", "Basler.is\_Arm\_allowed (25.74μs)", and "Basler.Arm (14.43ms)".

**Right Panel (Loki):**

- Title:** loki-rancher-obs-test
- Logs View:** Displays log entries. One entry is highlighted with an orange box:

```
Basler.py 1 cppTango unknown b107a-ea01/dia/cam-02 Basler tango 5a8b75d14205c39e
b-v-flexpes-csdb-0:18000 882665 server Basler/B107A-EA01-02 cpp opentelemetry 1.18.0
labels:
26125483f3cd54b02bdaf1ca9be5fa8d
```
- Time Range:** The time range is set from 12:00:06 to 12:00:06.

# Log fields

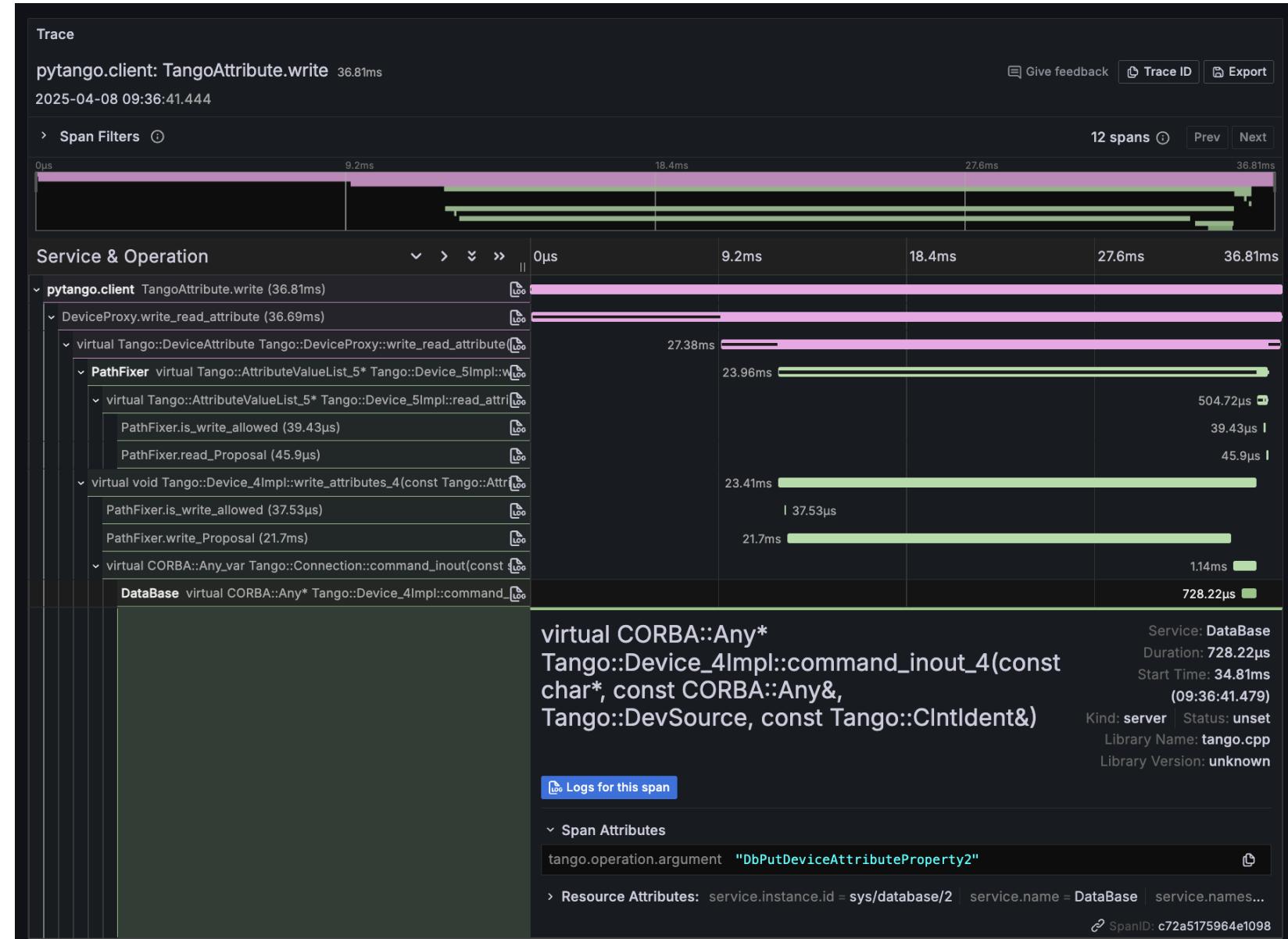
```
Common Basler.py 1830 debug 1 1744020003562754196 cppTango unknown b107a-ea01/dia/cam-02 Basler tango 5 DEBUG 0122fe970c13e55d
labels: b-v-flexpes-csdb-0:10000 802665 server Basler/B107A-EA01-02 cpp opentelemetry 1.18.0 1c80da6dd55d5e11e2b140ea1b73735e
Line limit: 1000 (1 returned) Total bytes processed: 628 kB

▼ 2025-04-07 12:00:03.562 Stopping the acquisition.

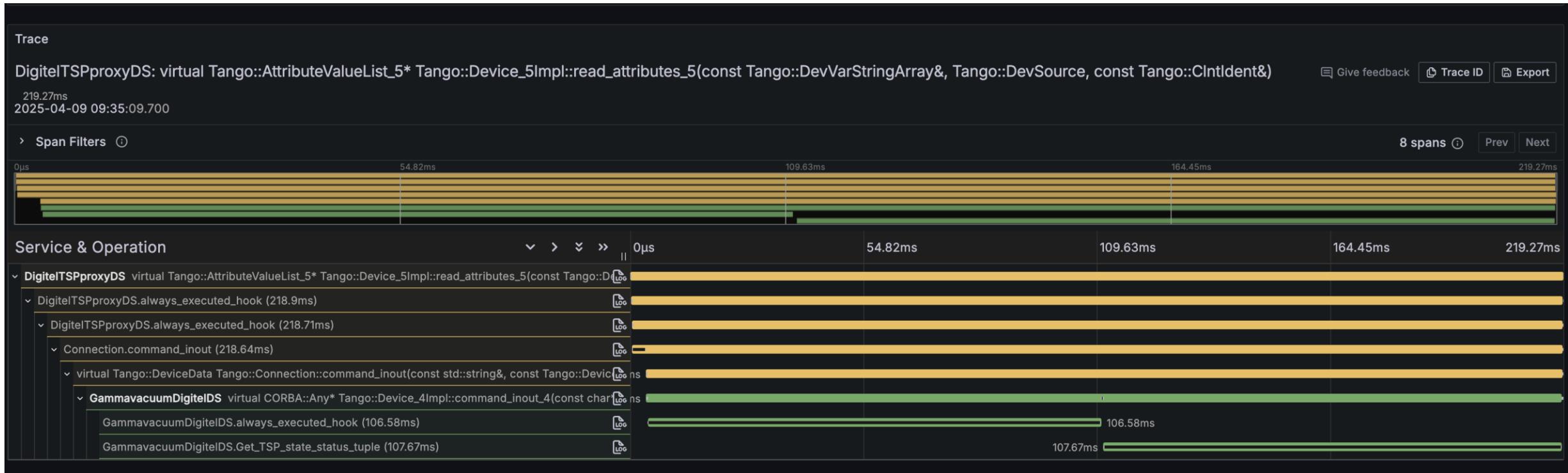
Fields
⊕ Q ⚡ code_filepath Basler.py
⊕ Q ⚡ code_lineno 1830
⊕ Q ⚡ detected_level debug
⊕ Q ⚡ flags 1
⊕ Q ⚡ observed_timestamp 1744020003562754196
⊕ Q ⚡ scope_name cppTango
⊕ Q ⚡ scope_version unknown
⊕ Q ⚡ service_instance_id b107a-ea01/dia/cam-02
⊕ Q ⚡ service_name Basler
⊕ Q ⚡ service_namespace tango
⊕ Q ⚡ severity_number 5
⊕ Q ⚡ severity_text DEBUG
⊕ Q ⚡ span_id 0122fe970c13e55d
⊕ Q ⚡ tango_host b-v-flexpes-csdb-0:10000
⊕ Q ⚡ tango_process_id 802665
⊕ Q ⚡ tango_process_kind server
⊕ Q ⚡ tango_server_name Basler/B107A-EA01-02
⊕ Q ⚡ telemetry_sdk_language cpp
⊕ Q ⚡ telemetry_sdk_name opentelemetry
⊕ Q ⚡ telemetry_sdk_version 1.18.0
⊕ Q ⚡ trace_id 1c80da6dd55d5e11e2b140ea1b73735e

Download ▾
Older logs
Start of range
^
12:00:03
—
12:00:03
^
```

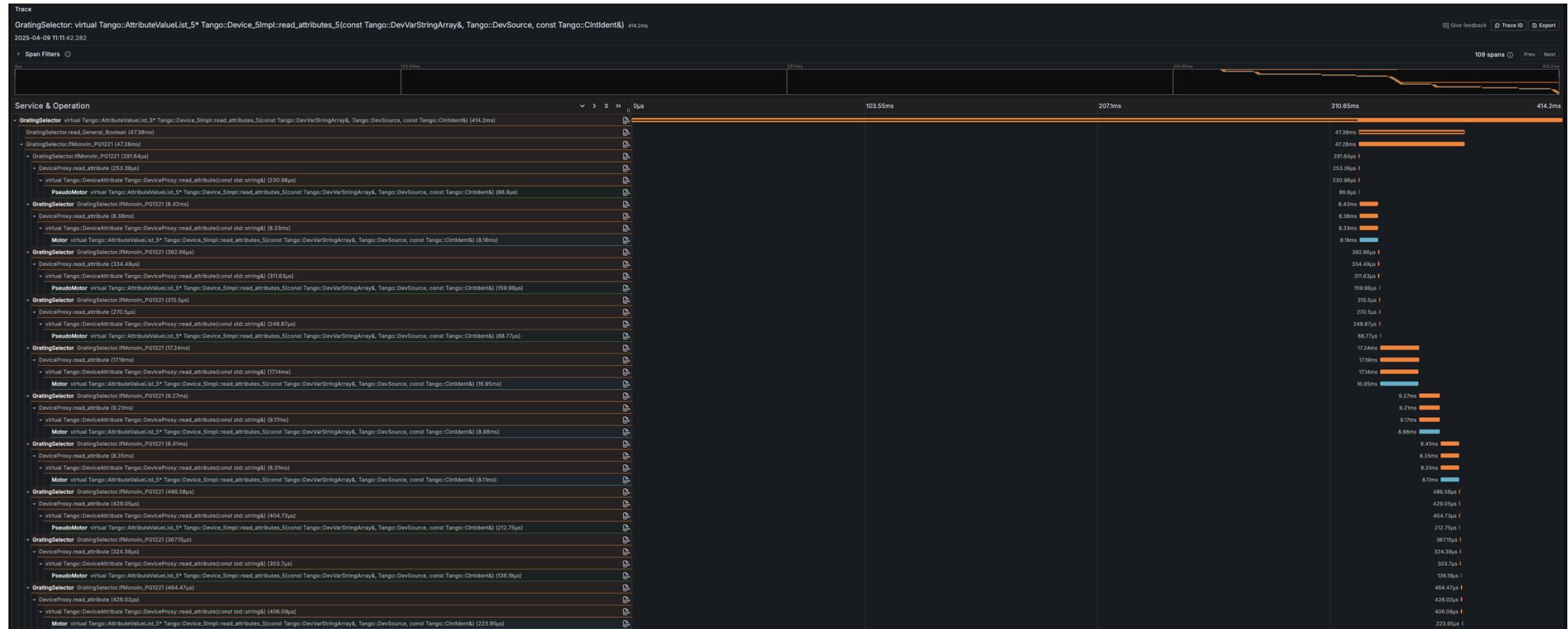
# Multi-device



# Two Python devices



# Deep traces



# Process info

```
$ export OTEL_EXPERIMENTAL_RESOURCE_DETECTORS=process  
$ python
```

Logs for this span	
> Span Attributes: code.filepath = <stdin>   code.lineno = 2   thread.id = 0x7fa71538c140   thread.name = MainThread	
Resource Attributes	
host.name	"b-v-flexpes-cc-0.maxiv.lu.se"
process.command	""
	[ ... ]
process.command_args	"]"
process.command_line	""
process.executable.name	"/opt/conda/envs/sardana/bin/python"
process.executable.path	"/opt/conda/envs/sardana/bin"
process.owner	"antjou"
process.parent_pid	331770
process.pid	333402
process.runtime.description	"3.11.11   packaged by conda-forge   (main, Dec 5 2024, 14:17:24) [GCC 13.3.0]"
process.runtime.name	"cpython"
process.runtime.version	"3.11.11"
service.name	"pytango.client"
service.namespace	"tango"

# TraceQL

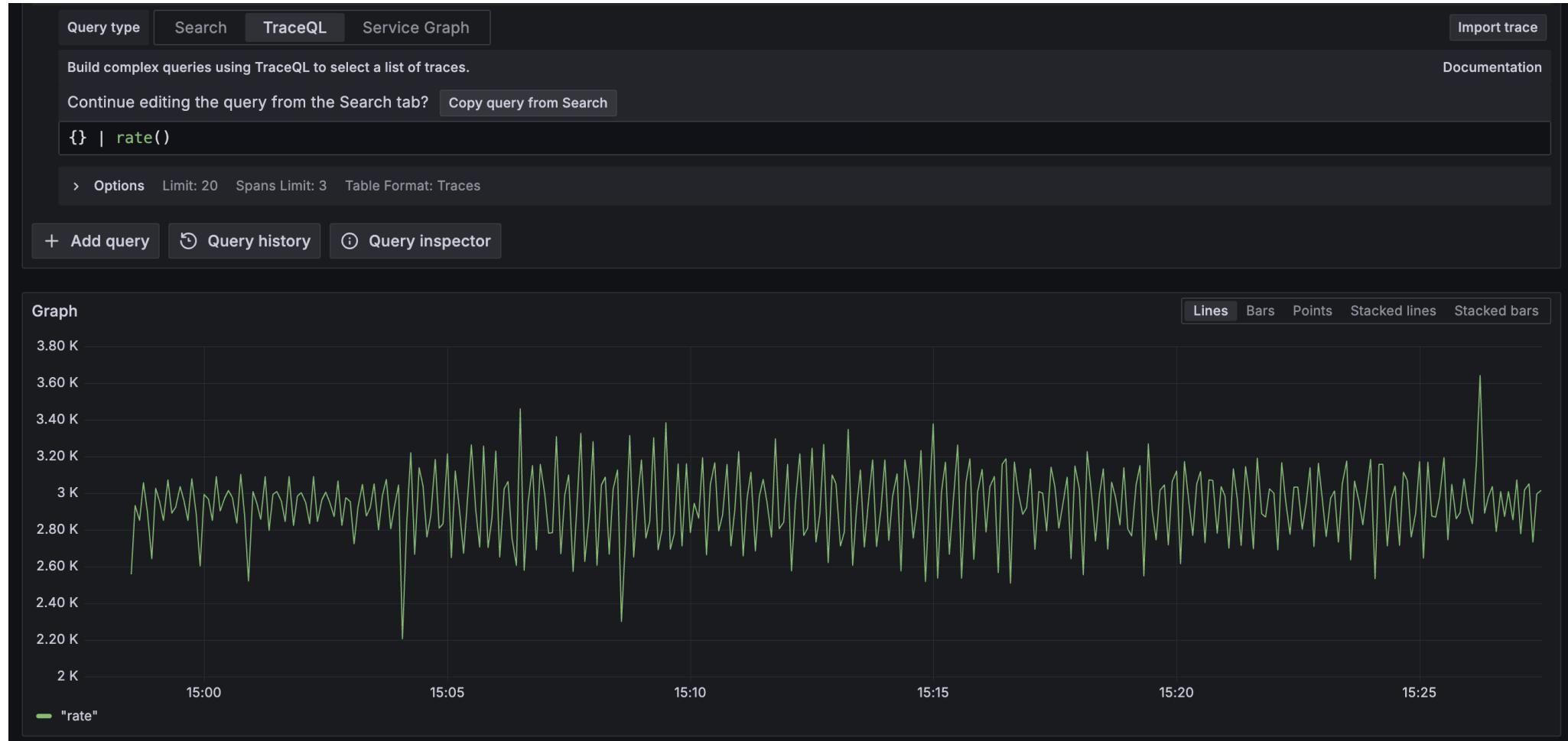
The screenshot shows the TraceQL interface with the following components:

- Header:** Includes tabs for "Query type" (Search, TraceQL, Service Graph), "Import trace", and "Documentation".
- Query Editor:** A text input field containing the TraceQL query: `[{"service": "pytango.client", "name": "_WorkItem.run"}]`. Below it are buttons for "Options", "Limit: 20", "Spans Limit: 3", and "Table Format: Traces".
- Buttons:** "+ Add query", "Query history", and "Query inspector".
- Table - Traces:** A table with columns: Trace ID, Start time, Service, Name, and Duration. The data is as follows:

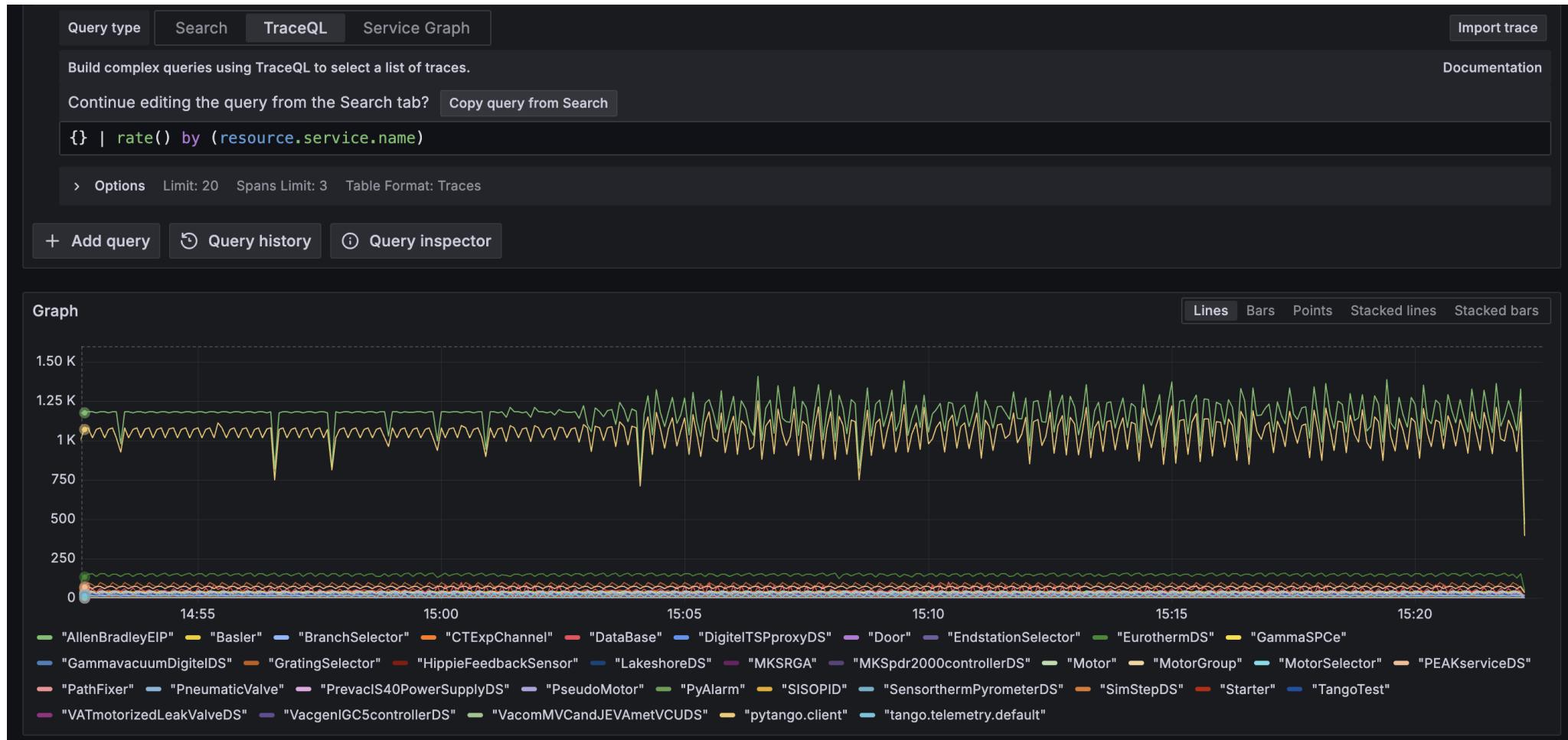
Trace ID	Start time	Service	Name	Duration
> 692c0818ca9e409894...	2025-05-19 15:25:55.589	pytango.client	_WorkItem.run	3 ms
> 99b7396b570bf518fc5...	2025-05-19 15:25:55.340	pytango.client	_WorkItem.run	5 ms
> 1d6fbe14f0aee64ec9a3...	2025-05-19 15:25:55.200	pytango.client	_WorkItem.run	7 ms
> 42e895482fae50e6141...	2025-05-19 15:25:54.991	pytango.client	_WorkItem.run	3 ms
> 85e26e133ab41cac06c...	2025-05-19 15:25:54.795	pytango.client	SimStepTangoMotorController.ReadOne	104 ms
> 395440b9082eeb5cbb...	2025-05-19 15:25:52.388	DataBase	virtual CORBA::Any* Tango::Device_4Impl::command_inou	<1ms
> 9fd8606089f8a6f4831...	2025-05-19 15:25:52.372	MKSRGA	virtual Tango::AttributeValueList_5* Tango::Device_5Impl::	4 ms

# TraceQL

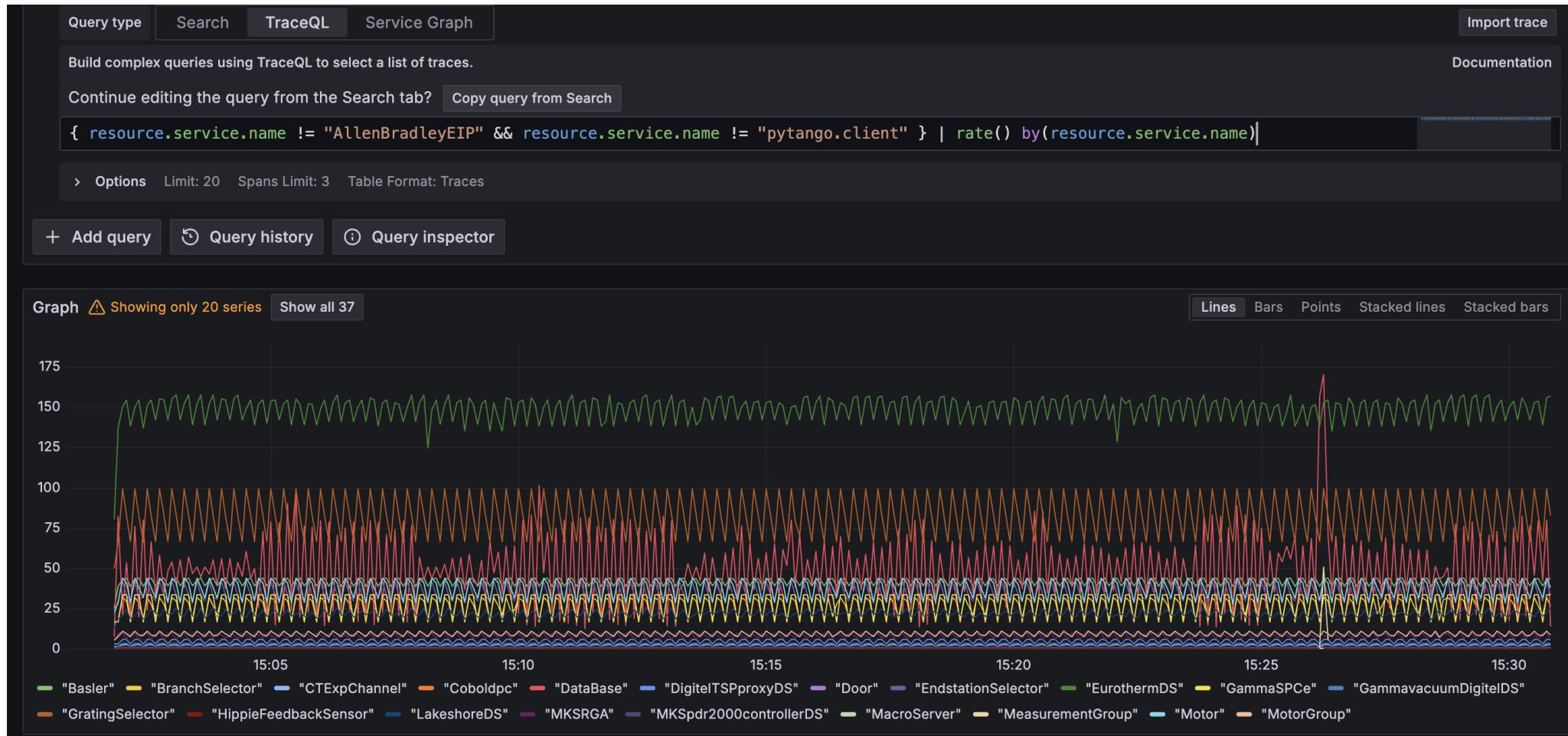
Note: max time range for `rate()` is 3h



# TraceQL



# TraceQL



# TraceQL



# TraceQL examples

```
{ .service.name = "BrooksMFC0254"  
  && duration > 500ms  
  && resource.service.instance.id = "b108a-d100830cab11/gas/vfcc-02" }  
  
{ resource.host.name = "host.institute.org" } >> { .service.name = "BrooksMFC0254" }  
  
{ span.tango.operation.argument = "double_scalar" }  
  
{ span.tango.operation.argument = "ZmqEventSubscriptionChange" }  
  
{ name = "DeviceProxy.subscribe_event" }  
  
{ name = "PandaPGMPCAPCoTiCtrl.ReadOne" }  
  
{ } | rate()  
  
{ } | rate() by(resource.host.hostname)  
  
{ } | quantile_over_time(duration, 0.99, 0.9, 0.5)
```

# Real-life usage

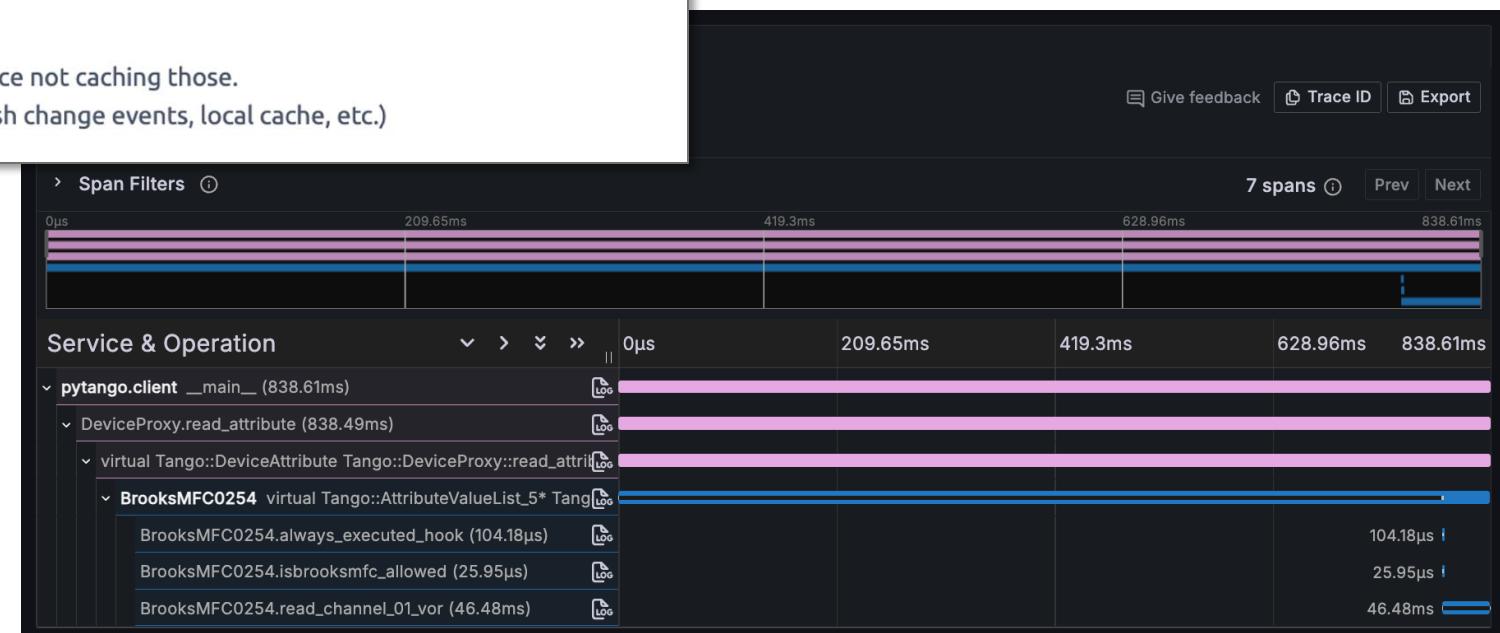
## #7752 SPECIES: Investigate slow communication with MFC and MOXA

### USER STORY

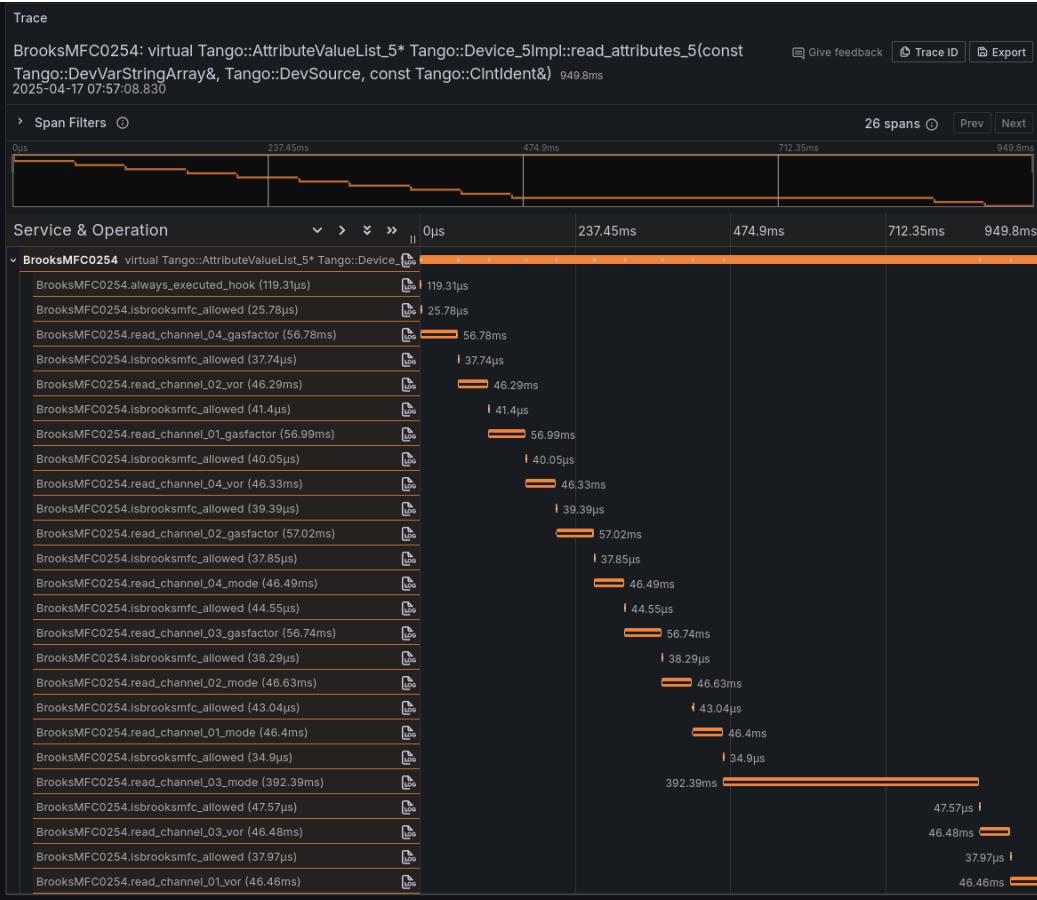
When multiple GUIs that are reading the MFC attributes are open, we notice that the device takes a considerable amount of time to respond.

This is likely due to taurus reading the attributes constantly and the tango device not caching those.

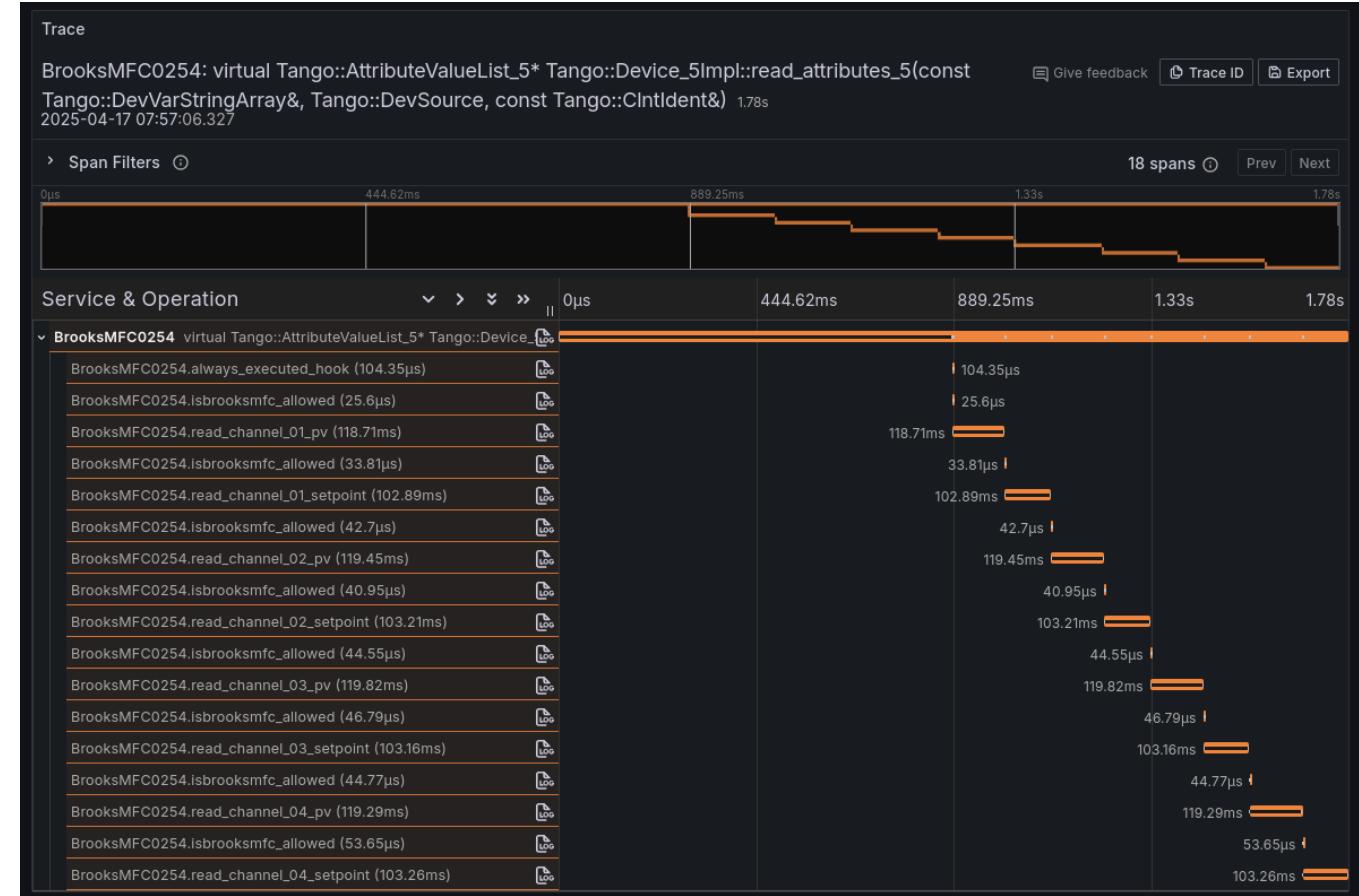
Investigate if that is the actual cause and if it is, implement better solution (push change events, local cache, etc.)



## Request from Taurus GUI (12 attributes)



## Request from Tango internal polling (8 other attributes)



Approx. 50 ms to 400 ms to read from hardware

The screenshot shows the Tempo interface with the title bar "tempo-rancher-obs-test". The main area has a header "A (tempo-rancher-obs-test)" with tabs for "Query type", "Search", "TraceQL", and "Service Graph". Below the tabs is a text input for TraceQL queries, containing the following code:

```
{ resource.service.instance.id = "b108a-d100830cab11/gas/vfcc-02" && duration > 500ms && status != error }
```

Below the TraceQL input are buttons for "Options", "Limit: 20", "Spans Limit: 3", and "Table Format: Traces". At the bottom of the interface are buttons for "+ Add query", "Query history", and "Query inspector".

The main content area is titled "Table - Traces" and displays a table with the following columns: Trace ID, Start time, Service, Name, and Duration. The table contains 15 rows of trace data, all from the service "BrooksMFC0254" and name "virtual Tango::AttributeValu". The durations range from 0.944 ms to 1.26 s.

Trace ID	Start time	Service	Name	Duration
c975d860a8f6128787b...	2025-04-14 13:53:33.327	BrooksMFC0254	virtual Tango::AttributeValu	1.26 s
6a415e03d810aafffb3e...	2025-04-14 13:53:30.328	BrooksMFC0254	virtual Tango::AttributeValu	1.23 s
62104301ca0899e9d8...	2025-04-14 13:53:27.327	BrooksMFC0254	virtual Tango::AttributeValu	1.23 s
96bdcc56ca87a38b23...	2025-04-14 13:53:24.327	BrooksMFC0254	virtual Tango::AttributeValu	1.20 s
e5cc07c53ad4abcee9d...	2025-04-14 13:53:21.327	BrooksMFC0254	virtual Tango::AttributeValu	1.22 s
75c8a6a24417c3c7874...	2025-04-14 13:53:18.327	BrooksMFC0254	virtual Tango::AttributeValu	1.11 s
1dd2c490fe7f957de7b...	2025-04-14 13:53:15.327	BrooksMFC0254	virtual Tango::AttributeValu	1.23 s
c8590a140e6f1583075...	2025-04-14 13:53:12.327	BrooksMFC0254	virtual Tango::AttributeValu	1.02 s
b9257ac0df61ecbf0bd...	2025-04-14 13:53:09.327	BrooksMFC0254	virtual Tango::AttributeValu	1.23 s
b945db2a8d1cc0e852...	2025-04-14 13:53:06.327	BrooksMFC0254	virtual Tango::AttributeValu	946 ms
f7accf9ad631f245029a...	2025-04-14 13:53:03.328	BrooksMFC0254	virtual Tango::AttributeValu	1.22 s
19855d03034e33a08a...	2025-04-14 13:53:00.327	BrooksMFC0254	virtual Tango::AttributeValu	954 ms
6c4870888ab762d54e...	2025-04-14 13:52:57.327	BrooksMFC0254	virtual Tango::AttributeValu	1.23 s
b56957db6c526ffdd6b...	2025-04-14 13:52:54.328	BrooksMFC0254	virtual Tango::AttributeValu	944 ms

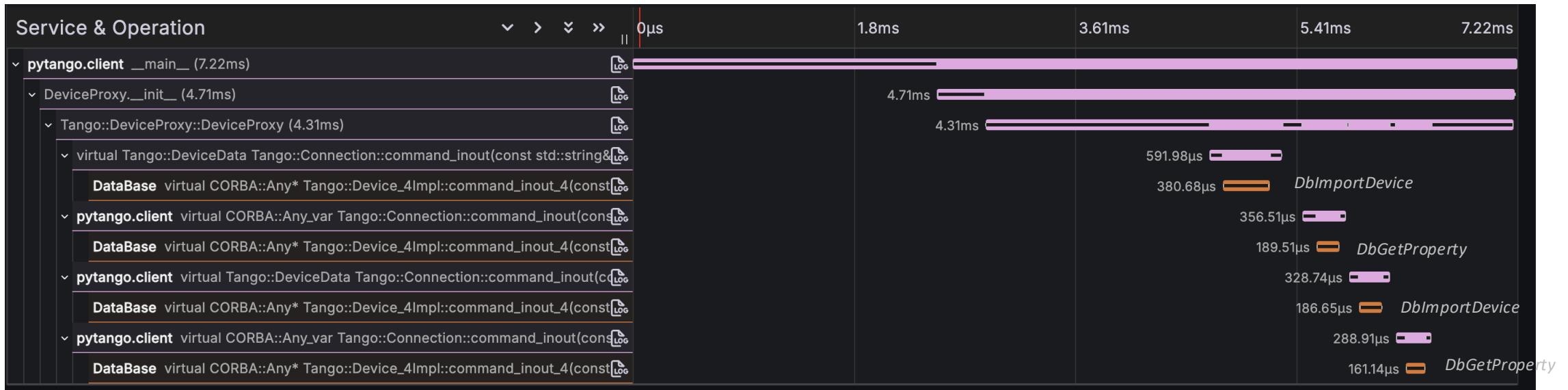


# Event subscriptions

```
>>> import tango
>>> dp = tango.DeviceProxy("b107a-ea01/dia/cam-02")
>>> eid = dp.subscribe_event("lastImage", tango.EventType.CHANGE_EVENT, tango.utils.EventCallback())
2025-05-21 19:20:15.559263 B107A-EA01/DIA/CAM-02 LASTIMAGE CHANGE [ATTR_VALID] [[0 0 0 ... 0 0 0]
[0 0 0 ... 0 0 0]
[0 0 0 ... 0 0 0]
...
[0 0 0 ... 0 0 0]
[0 0 0 ... 0 0 0]
[0 0 0 ... 0 0 0]]
>>> dp.unsubscribe_event("lastImage")
Traceback (most recent call last):
  File "/opt/conda/envs/sardana/lib/python3.11/site-packages/tango/device_proxy.py", line 1794, in
__DeviceProxy__unsubscribe_event
    evt_info = se[event_id]
    ~~^~~~~~^~~~~~^
...
KeyError: 'This device proxy does not own this subscription lastImage'
>>> dp.unsubscribe_event(eid)
>>>
```

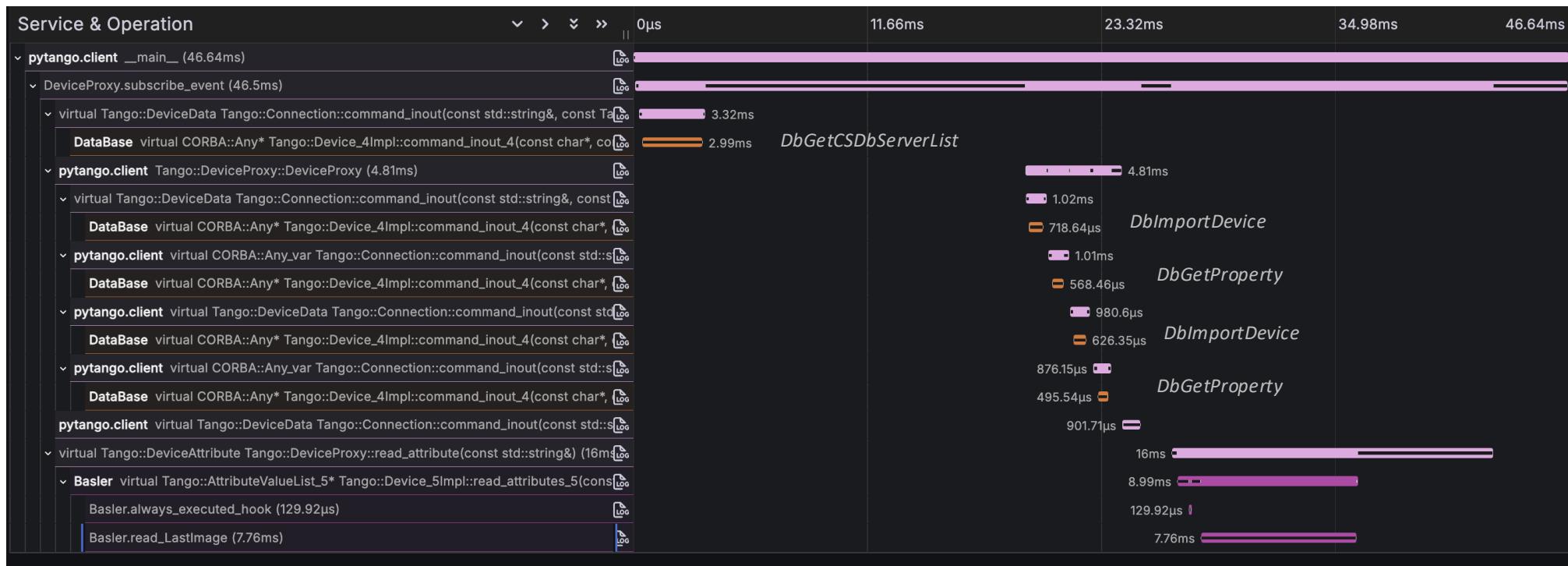
# Event subscriptions

```
>>> dp = tango.DeviceProxy("b107a-ea01/dia/cam-02")
```



# Event subscriptions

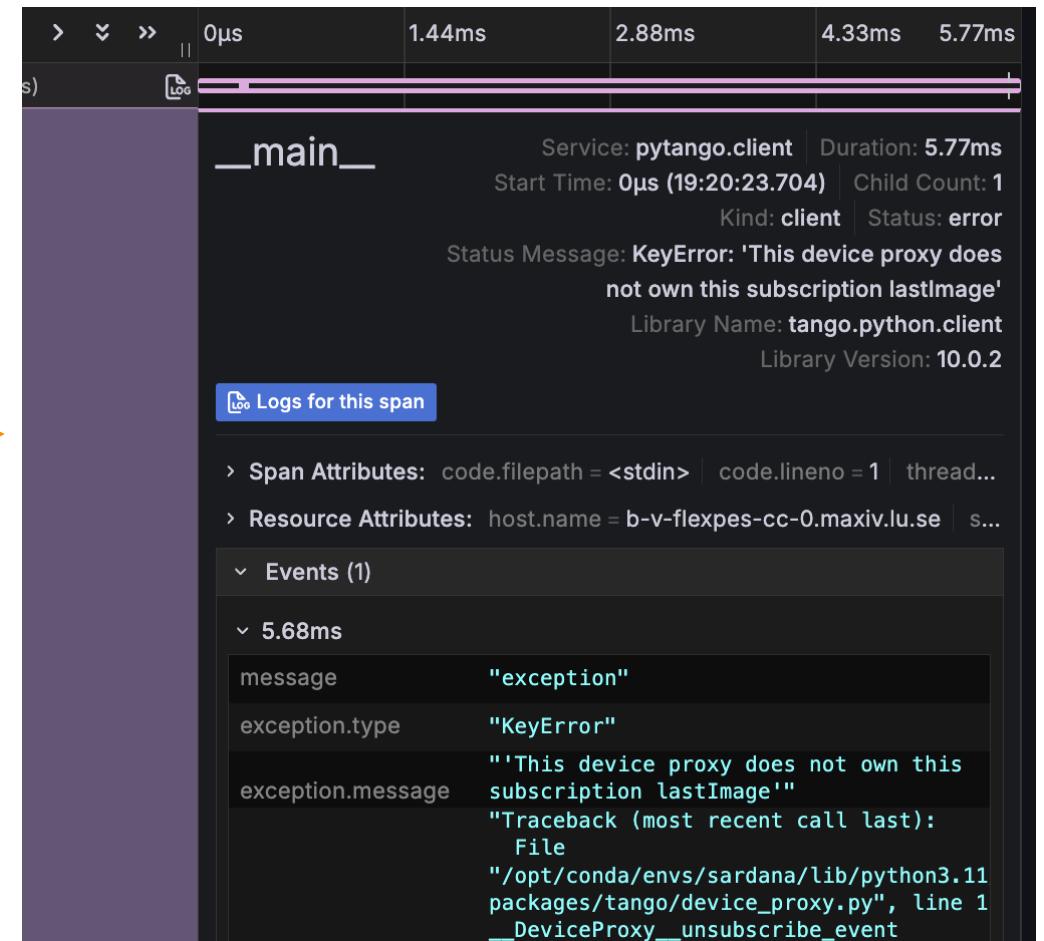
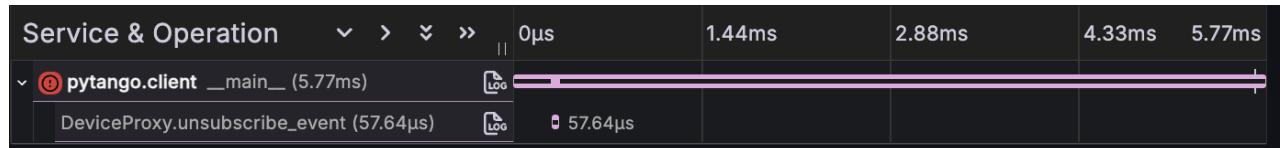
```
>>> eid = dp.subscribe_event("lastImage", tango.EventType.CHANGE_EVENT, tango.utils.EventCallback())
```



# Event subscriptions

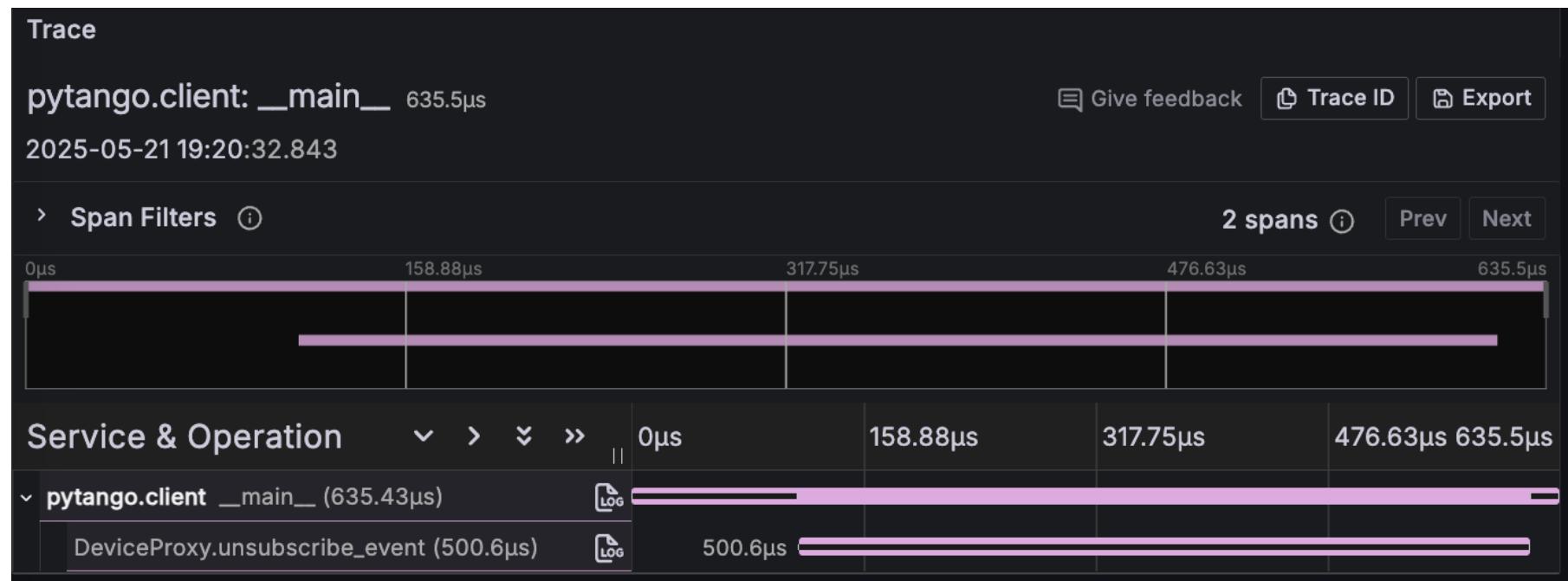
```
>>> dp.unsubscribe_event("lastImage")
Traceback (most recent call last):
...

```



# Event subscriptions

```
>>> dp.unsubscribe_event(eid)
```



# Event subscriptions

The screenshot displays two side-by-side views of the Tempo UI interface, both titled "tempo-rancher-obs-test".

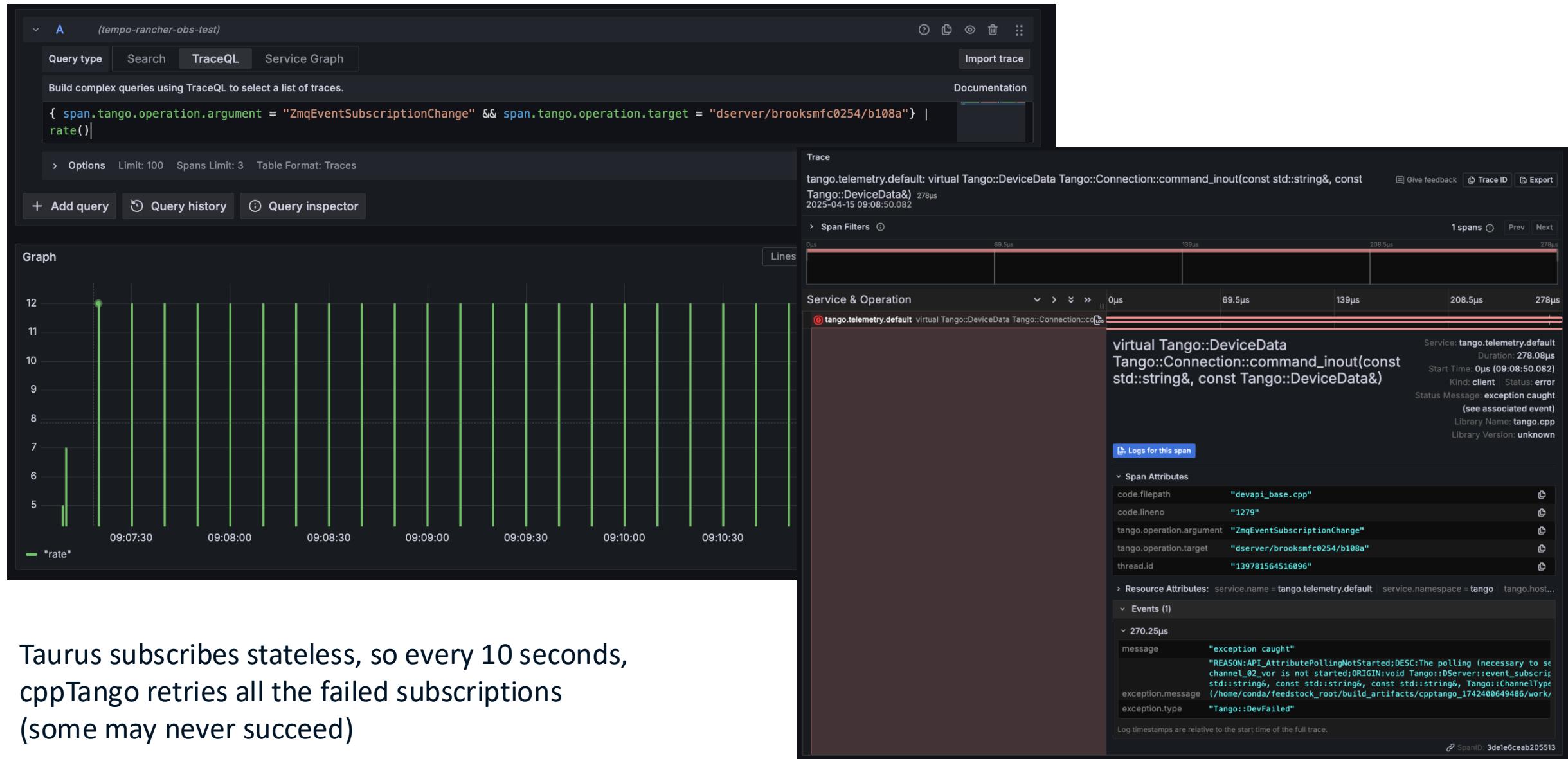
**Left Panel (Query Results):**

- Query Type:** TraceQL
- Search Query:** { name =~ ".\*subscribe.\*" }
- Options:** Limit: 20, Spans Limit: 3, Table Format: Traces
- Table - Traces:**

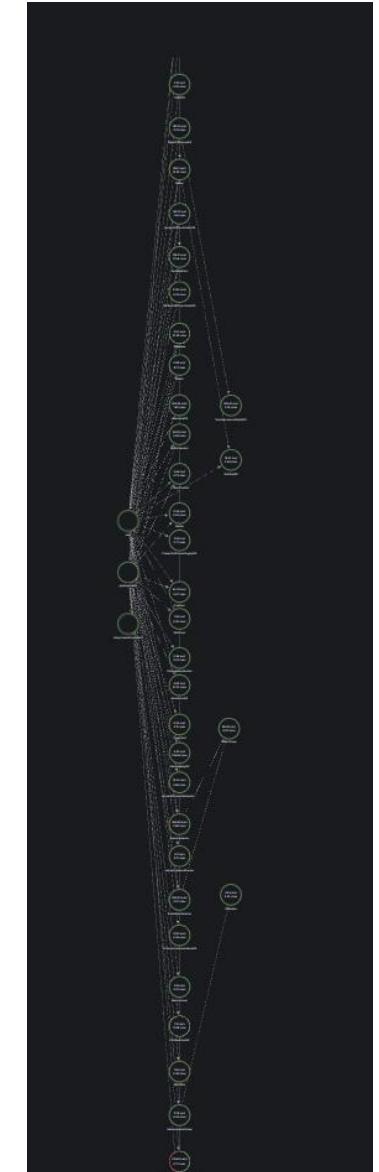
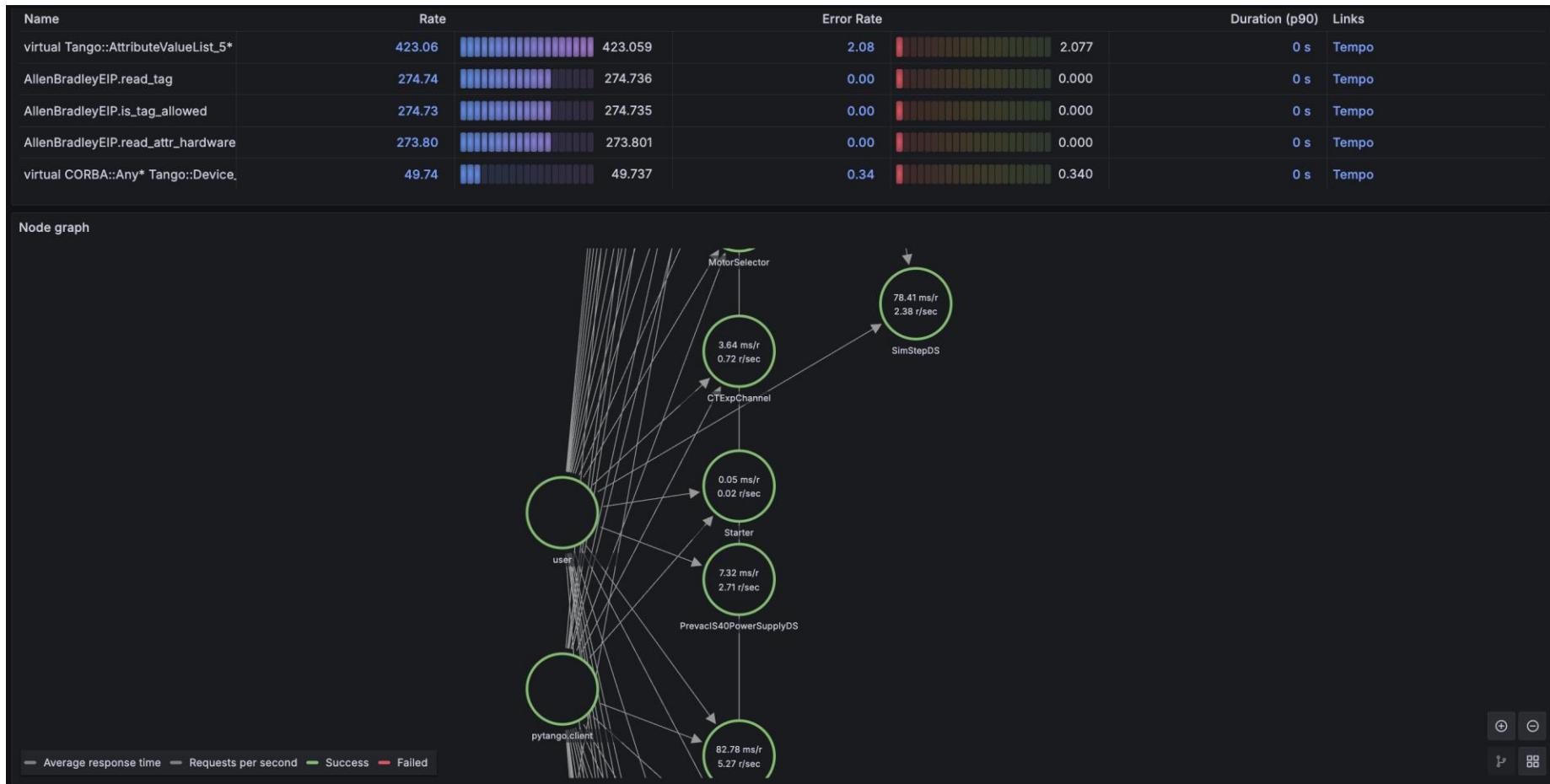
Trace ID	Start time	Service	Name	Duration
9c4b8dbf885aecae62...	2025-04-04 14:08:05	<root span not yet received		<1ms
86724b00380b08687e...	2025-04-04 14:08:05	<root span not yet received		<1ms
9e40a9576bc8402520...	2025-04-04 14:08:05	<root span not yet received		<1ms
9b2f5c1ecb612eb4bbf7...	2025-04-04 14:08:05	<root span not yet received		<1ms
9621321c167777e7fea...	2025-04-04 13:52:48	pytango.client	TangoAttribute._call_dev_h	<1ms
84a6b696c454396508...	2025-04-04 13:52:48	pytango.client	TangoAttribute._call_dev_h	1 ms
c34fb6a19cd712c04b7...	2025-04-04 13:52:48	pytango.client	TangoAttribute._subscribeC	2 ms
9d428b12832aa9fbec5...	2025-04-04 13:52:48	pytango.client	TangoAttribute._call_dev_h	<1ms
553fcae12b461f5db9f7...	2025-04-04 13:52:48	pytango.client	TangoAttribute._call_dev_h	2 ms
8ac3dd53eac2ac20e8...	2025-04-04 13:52:48	pytango.client	TangoAttribute._subscribeC	30 ms
8eef993b0653727bd72...	2025-04-04 13:52:23	pytango.client	TangoAttribute._call_dev_h	<1ms
f9522d69abc3ece8d3e...	2025-04-04 13:52:23	pytango.client	TangoAttribute._call_dev_h	<1ms
2d69d5ce8ed1d165423...	2025-04-04 13:52:23	pytango.client	TangoAttribute._subscribeC	2 ms

**Right Panel (Trace Detail):**

- Trace ID:** 8ac3dd53eac2ac20e82305c3d84d0c93
- Service & Operation:** pytango.client: TangoAttribute.\_subscribeConfEvents (30.08ms)
- Span Filters:** 10 spans
- Timeline:** 0μs, 7.52ms, 15.04ms, 22.56ms, 30.08ms
- Operations:** DeviceProxy.subscribe\_event (30.01ms), virtual Tango::DeviceData Tango::Connection::comm, virtual CORBA::Any\_var Tango::Connection::comm, Tango::DeviceProxy::DeviceProxy (5.17ms), virtual Tango::DeviceData Tango::Connection::comm, virtual CORBA::Any\_var Tango::Connection::comm, virtual Tango::DeviceData Tango::Connection::comm, virtual CORBA::Any\_var Tango::Connection::comm, virtual Tango::DeviceData Tango::Connection::comm, virtual Tango::DeviceData Tango::Connection::comm



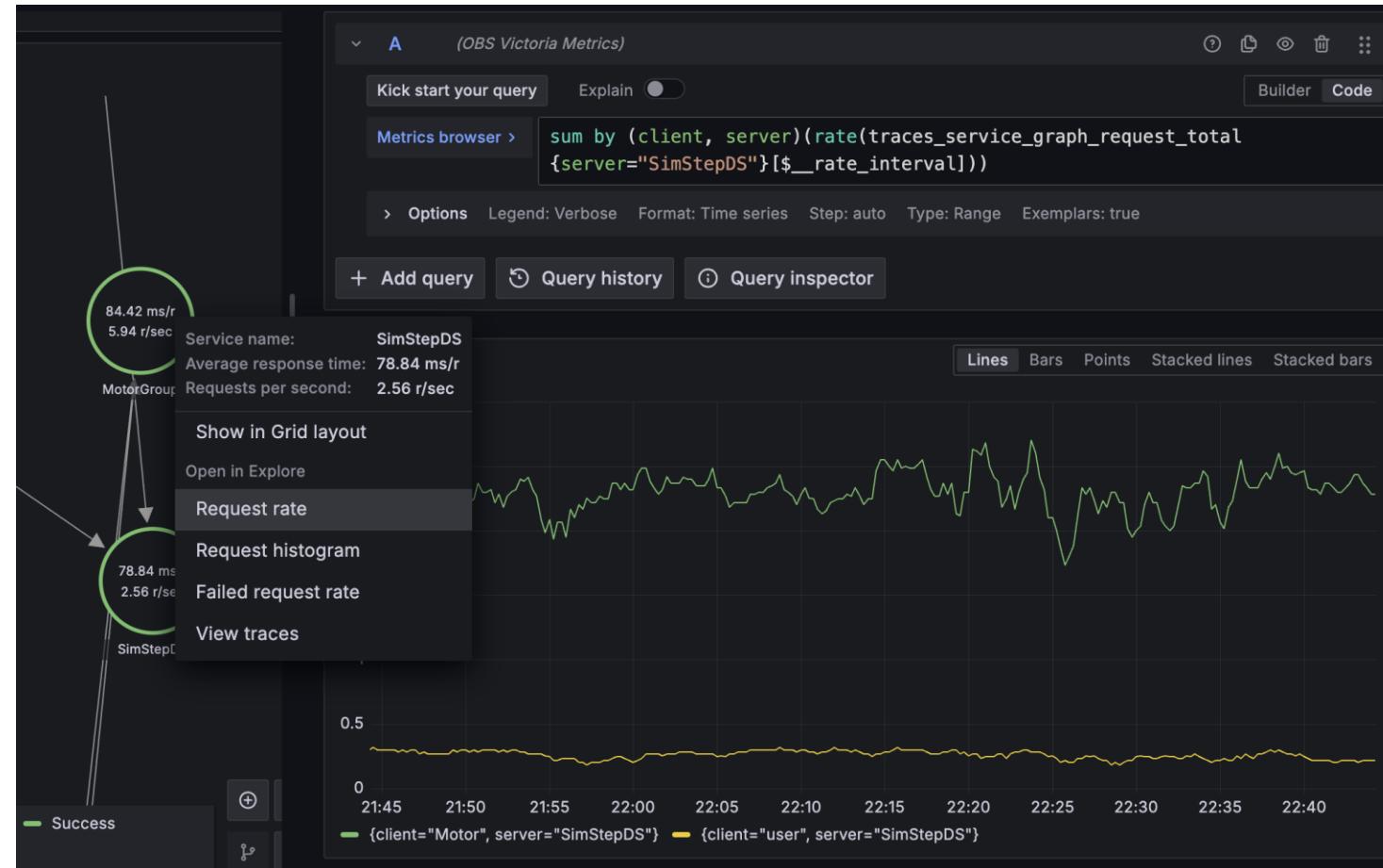
# Service graph



# Service graph



# Service graph



# Performance impact

# CPU usage



# Compute resources for backend

# Infrastructure

Resources (2 replicas)

8 cores @ 2GHz each

16 GB RAM (250 MB / component: distributor, ingestor, compactor, querier, etc.)

Storage

50 GB/day (only keeping 150 GB)

S3 storage

250 read/writes per second

Ceph using 4 CPU cores, 20 GB RAM (minio would be better)

10k rpm SAS drives

Network

3k spans/second => ~230 HTTP requests/sec

# Tango applications

Telemetry enabled for only 1 out of 16 beamlines

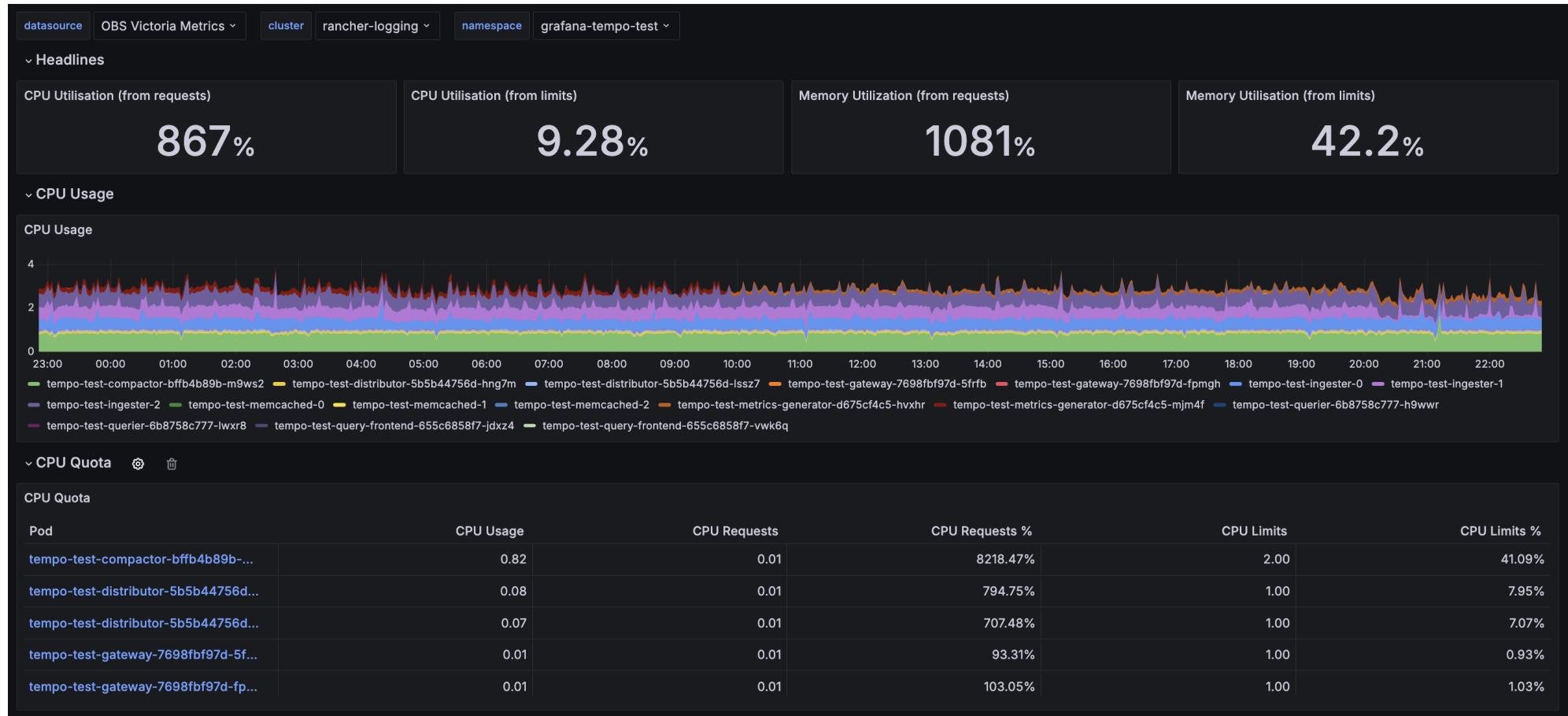
900 devices with 11k commands, 17 k attributes

1 control room workstation (client applications)

Logging only enabled on 1 device

(MAX IV has 24k devices, in total)

# K8s compute resources



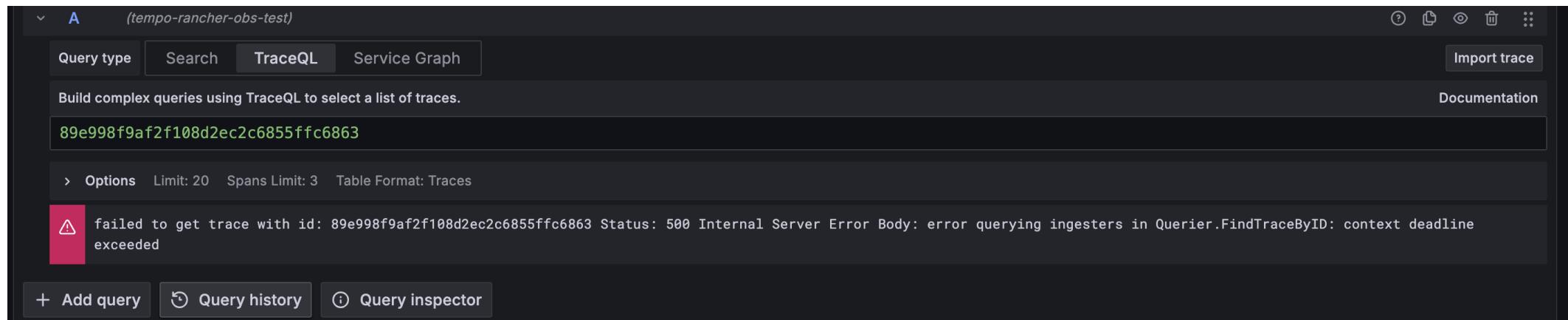
# K8s compute resources



# Tempo monitoring



# Timeouts when querying backend

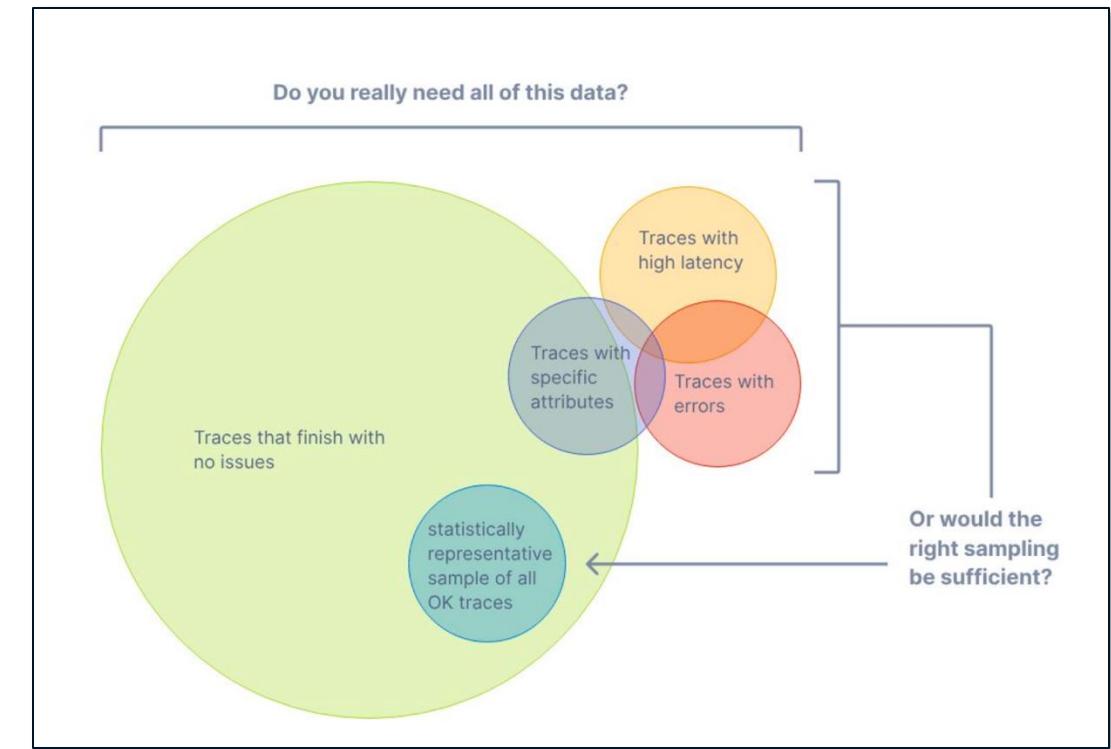


Resources are probably insufficient.  
Also see some errors pushing spans from clients

# Sampling to reduce trace volume

Plan to use Grafana Alloy to implement tail sampling

Keep traces if  
duration > 5ms  
status is error  
5% of "normal" traces



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

# Issues / ideas for improvement

# Warning logs

Sardana spock output:

```
File Edit View Search Terminal Help
Operation will be saved in /data/visitors/flexpes/20250094/2025040808/raw/EXAFS_Na_compounds.h5 (HDF5::NXscan from NXscanH5_FileRecorder)
Scan #3507 started at Tue Apr  8 18:30:05 2025. It will take at least 0:13:26.700000
#Pt No beamline_energy_mlcorr a_slit2_v_real b107a_em_03_ti b107a_em_03_ch1 b107a_em_03_ch2 b107a_em_03_ch3 b107a_em_03_ch4 pcap_energy_av pcap_energy_min pcap_energy_max pcap_cff_av dt
 0      1050      50.015     0.3    9.38795e-11   6.61168e-11   6.95358e-11   4.78591e-09   1050      1050      1050      2.25001  2.30754
 1      1050.45     50.02     0.3    9.43228e-11   6.61131e-11   5.72815e-11   4.78362e-09   1050.48     1050.48     1050.48   2.25001  5.39012
 2      1050.99     50.0155    0.3    9.44075e-11   6.65705e-11   6.72559e-11   4.81808e-09   1050.97     1050.97     1050.99   2.24999  10.7521
 3      1051.42     50.011     0.3    9.42124e-11   6.67454e-11   7.32813e-11   4.82544e-09   1051.46     1051.46     1051.46   2.25    13.4973
 4      1051.92     50.021     0.3    9.38026e-11   6.6851e-11    7.05943e-11   4.82565e-09   1051.92     1051.92     1051.91   2.25002  18.6754
,Body[connection error: desc = "transport: Error while dialing: dial tcp 10.42.225.243:9095: connect: connection refused" _client.cc:115 [OTLP HTTP Client] Export failed, Status:503, Header: content-length: 119
[Error] File: /home/conda/feedstock_root/build_artifacts/opentelemetry-sdk_1735643008820/work/exporters/otlp/src/otlp_http_exporter.cc:144 [OTLP TRACE HTTP Exporter] ERROR: Export 1 trace span(s) error: 1
 5      1052.42     50.014     0.3    9.32053e-11   6.68975e-11   6.3714e-11   4.82167e-09   1052.4     1052.41     1052.41   2.24999  24.3119
 6      1052.86     50.015     0.3    9.25303e-11   6.67555e-11   5.91543e-11   4.81324e-09   1052.86     1052.86     1052.86   2.25    29.4297
 7      1053.33     50.0105    0.3    9.18877e-11   6.66973e-11   7.35663e-11   4.80865e-09   1053.33     1053.33     1053.35   2.25    35.0107
 8      1053.77     50.019     0.3    9.15922e-11   6.67282e-11   5.61416e-11   4.809e-09    1053.77     1053.77     1053.77   2.25001  39.765
```

Exiting Taurus application:

```
...
MainThread      WARNING 2025-04-04 14:08:02,003 opentelemetry.sdk.trace.export: Already shutdown,
dropping span.
MainThread      WARNING 2025-04-04 14:08:02,004 opentelemetry.sdk.trace.export: Already shutdown,
dropping span.
```

# Improvement ideas

Shorter trace name for cppTango? E.g. change:

```
virtual CORBA::Any* Tango::Device_4Impl::command_inout_4(const char*, const CORBA::Any&,
Tango::DevSource, const Tango::ClntIdent&)
```

to:

`Tango::Device_4Impl::command_inout_4` and put full signature in a span attribute

Include data from read/write/command/get/put property in span attribute (limit size?)

Device must record client info (e.g., if client hasn't opted in to providing telemetry info, won't know where request came from)

Fix crash when doing a short client session with PyTango

Enable/disable/configure at runtime, similar to logging. [cppTango #1398](#)

# Other clients

Instrument clients like Taurus GUI, Sardana, TangoGQL

# Conclusion

# Summary

Very powerful feature

Negligible impact on Tango device server performance

Backend needs “a lot” of compute resources

Trace sampling will be needed to reduce trace volume

Room for improvement

# Useful links

PyTango documentation

<https://tango-controls.readthedocs.io/projects/pytango/en/latest/how-to/telemetry.html>

Observability and OpenTelemetry

<https://opentelemetry.io/docs/concepts/observability-primer/>

<https://opentelemetry.io/docs/what-is-opentelemetry/>

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

MAX IV