Karabo Data Logging: InfluxDB Backend and Grafana UI

Dr. Gero Flucke *et al.*
European XFEL GmbH

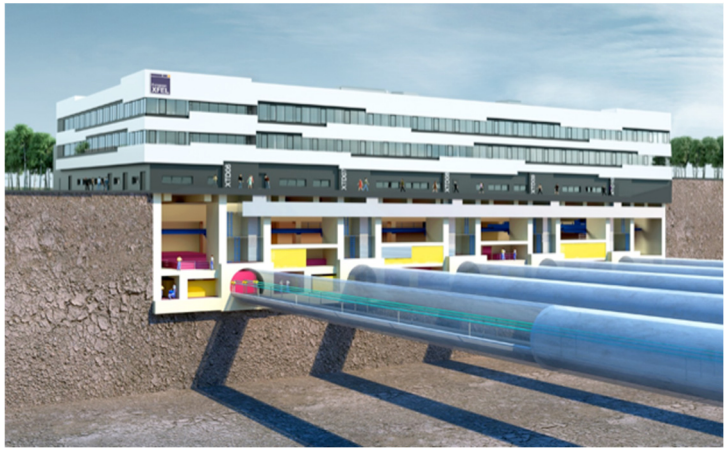
18th International Conference on Accelerator and Large Experimental Physics Control Systems
(ICALEPCS 2021)
Virtual Conference

October 18-22, 2021

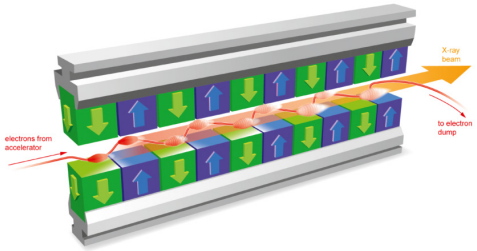
Outline

- The **Karabo** Control System at the **European XFEL**
- Karabo Data Logging
  - From custom file-based backend to **InfluxDB**
- Boosting visualisation of historic data using **Grafana**

The Home of Karabo:
European X-ray Free Electron Laser (XFEL):



- Linear electron accelerator
  - 10 Hz of "trains" of up to 2700 pulses
  - run by DESY
- Undulators
  creating X-ray laser photons



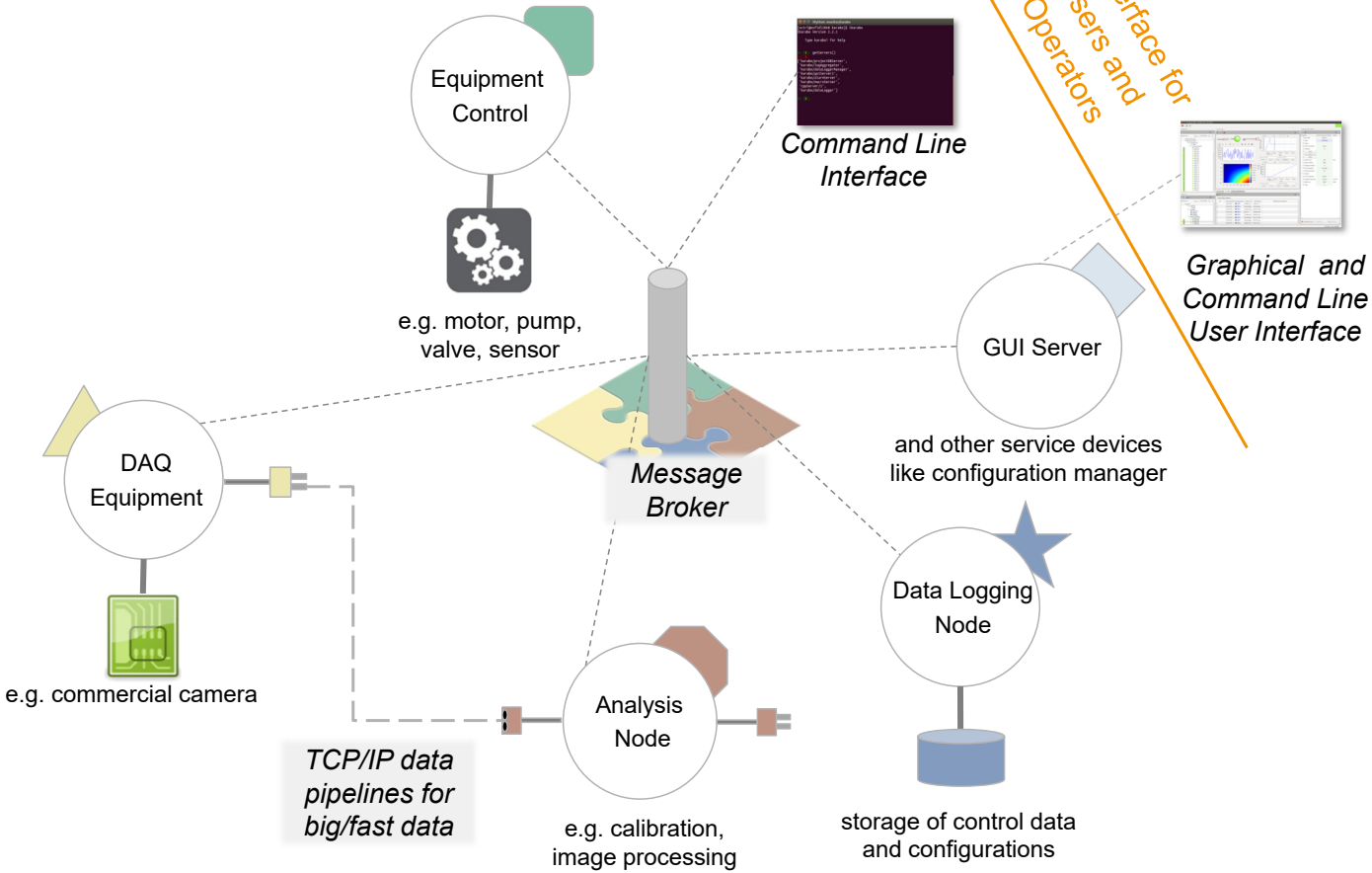
- Photon beam steered
  - through 3 tunnels
  - to 6 instruments

Karabo:
Designed and developed
for control, data acquisition, analysis

Karabo: Device Based Communication via a Message Broker

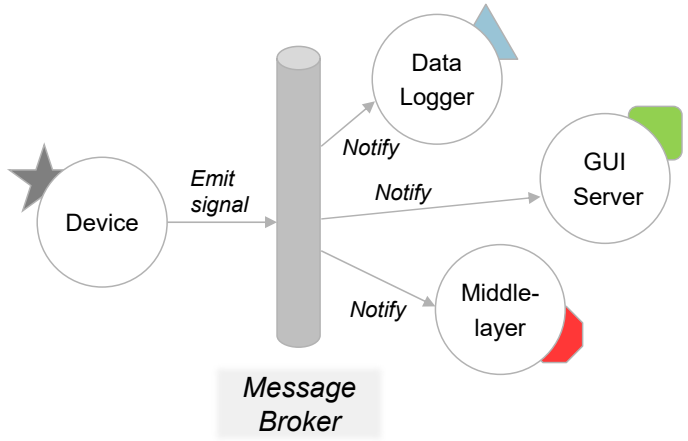
Self-describing
Karabo Devices

- Equipment control, e.g. motors, valves,…
- Detectors
  - e.g. cameras
- Online data analysis
- Data Logging
- Other system services
  - GUI entry point
  - DAQ for big data (not shown)



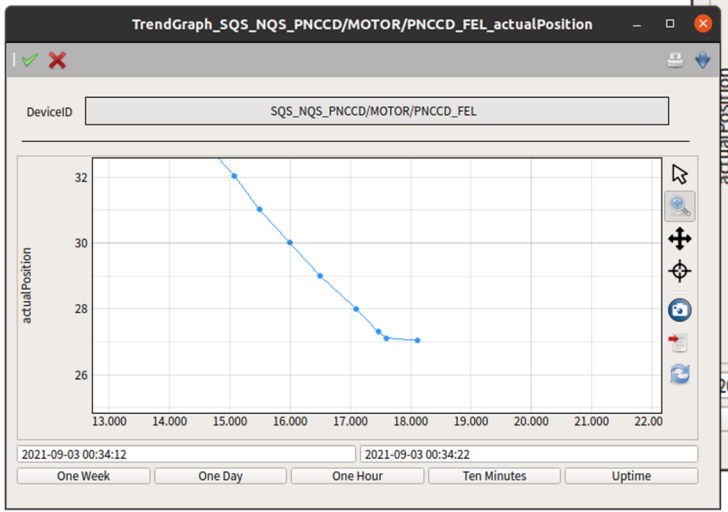
Karabo: Event-Driven Broker Communication

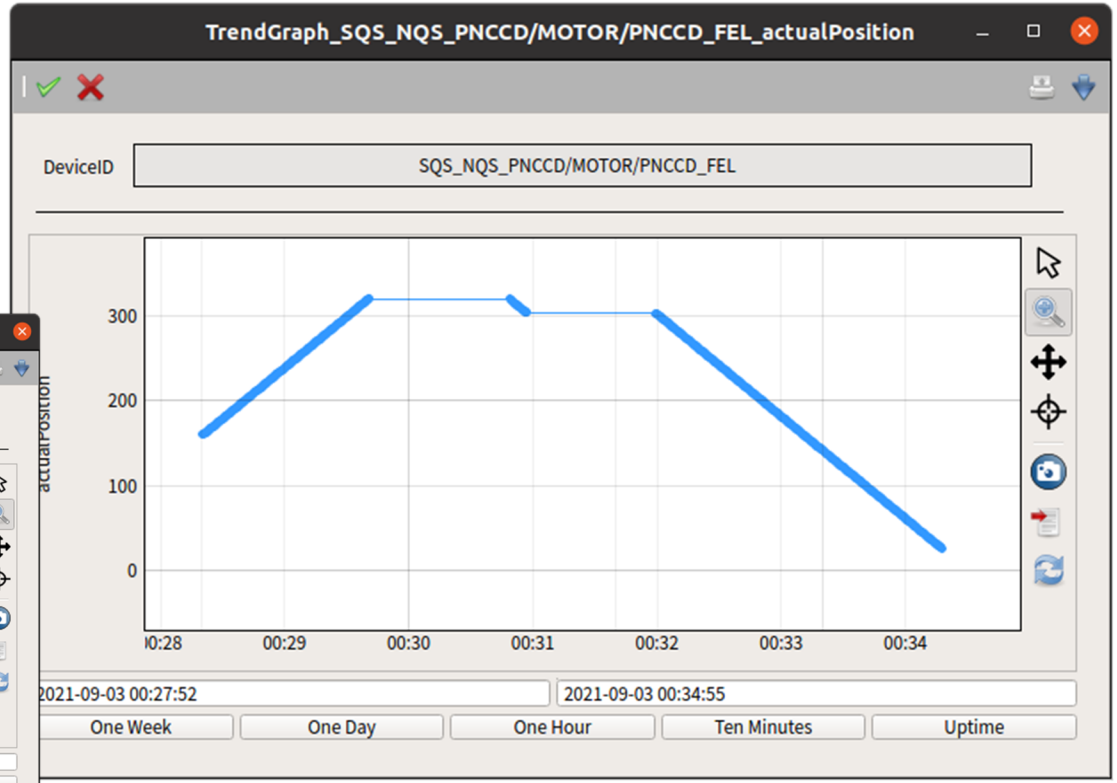
- Cross device signal/slot subscription
  - Devices subscribe slots to a remote “signal”.
- When signal is “emitted”, all subscribed slots are called.
  - Single message to the broker
    - ► Avoids publishing overhead for “popular” devices
  - Regular **polling obsolete**.
- Used by Data Logger Devices
  - Logging task distributed among a handful of “loggers”
  - Each logger subscribes to the “property update” signals of its share of devices



Data Logging and Retrieval – What for?

- Main control use cases:
  - View historic data to spot trends and analyse incidents
  - Restore configurations known to have worked "the day before"





Karabo Data Logging: How?

- Original, temporary solution from 2014:
  - Custom-built ascii file backend (one directory per logged device)

```
20200719T184128.573059Z|1595184088.573059|804227972|actualPosition|FLOAT|35.91043||VALID
```

  - Drawbacks
    - ▶ Human readable data does not scale well: routine access only for 3 months
    - ▶ Indexing (for fast retrieval) needed custom implementation
    - ▶ No support for statistical treatment
      (data reduction by simple down sampling: e.g. every 2nd data point only)
    - ▶ Reading back data can be slow

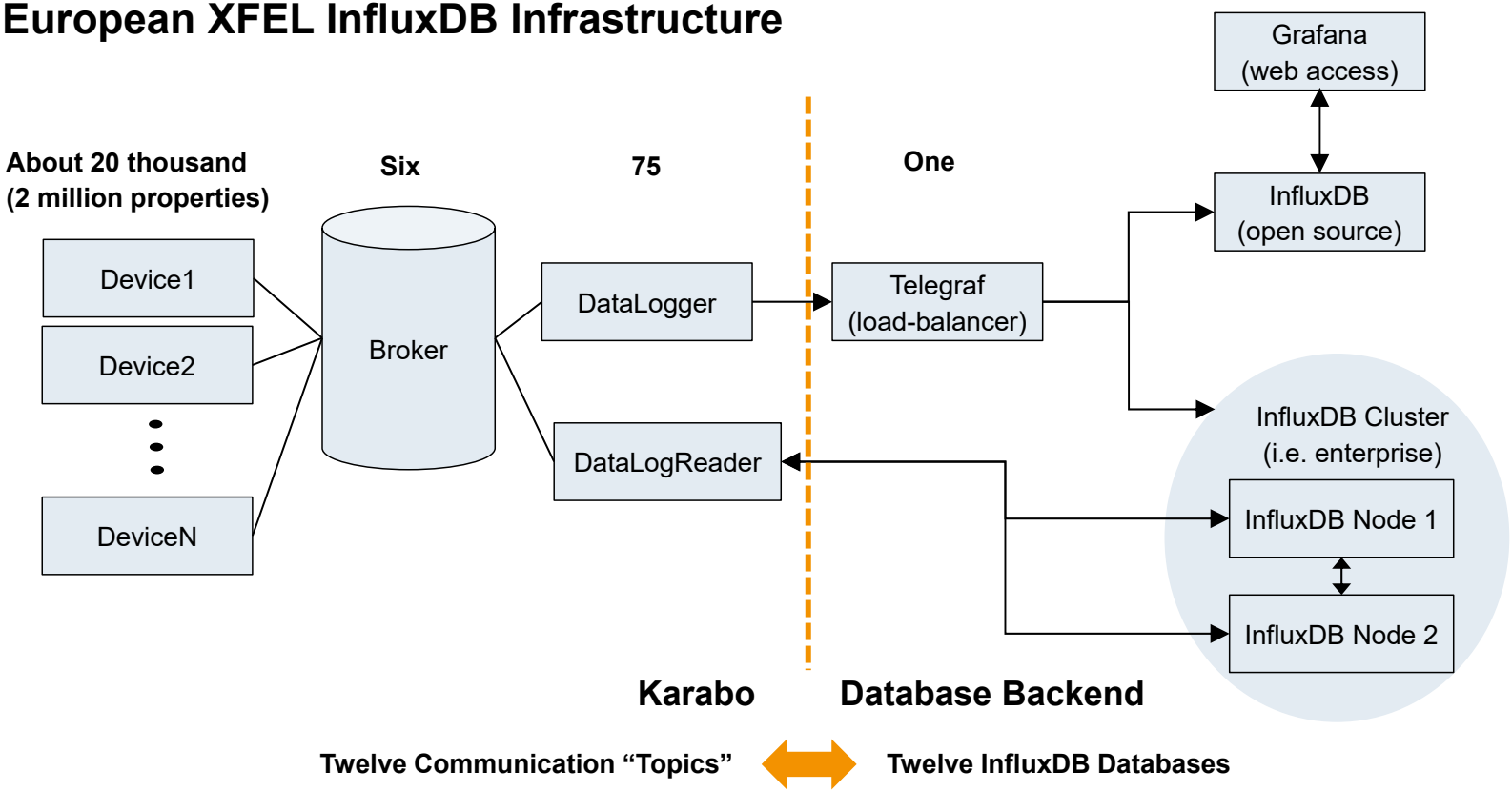
- Time series database suits well:
  - Logging to InfluxDB:
    - ▶ Karabo proto-type in 2018
    - ▶ Serious development 2019 and 2020
    - ▶ In production since summer 2020
    - ▶ Migrated all data since January 2020

Types: Karabo	InfluxDB
BOOL	Boolean
FLOAT, DOUBLE	Float (special NaN treatment)
[U]INT8, … [U]INT32, INT64	Integer
UINT64	Integer re-interpreted as int64 ⚠
STRING	String (mangling escape characters) ⚠
VECTOR_[U]INT8, …	Comma separated string
VECTOR_STRING	Base64 encoded JSON string
CHAR, VECTOR_CHAR (raw data), VECTOR_HASH (table), SCHEMA (self-description)	Base64 encoded (Karabo) binary

Karabo-less interpretation

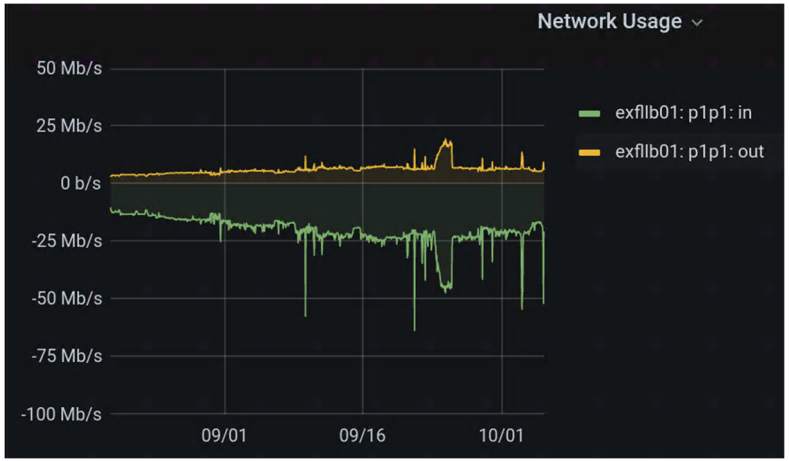
Karabo-specific

European XFEL InfluxDB Infrastructure



InfluxDB @ EuXFEL

- About 20 thousands of Karabo devices (each mapped to InfluxDB "measurement")
  - In total about 2 million properties to log (InfluxDB "fields")
- 75 Karabo data logger devices writing to database backend
  - Plus data log reader devices to request historic data
- Single load-balancer ("Telegraf") receiving all data
  - Input data rate about 20 Mb/s (Sept. 2021, see ➔)
  - Duplicated (but zipped) output rate about 8 Mb/s
- Stored in InfluxDB 1.8 cluster (2 nodes)
  - Enterprise edition, 5*24 h support by InfluxData
  - Additionally: single open source node for Grafana
- > 240 billion property updates stored in InfluxDB
  - 14 TB per InfluxDB node
  - Increase per month: ~ 10 Billion



Operational Experience

- Enhanced Karabo experience:
  - Can afford much longer data availability than previously (was 3 months)
    - ► 3 years retention are planned
  - Can afford (& implemented) averaging instead of "random" down sampling if data reduction needed
    - ► Spikes short in time can be reliably spotted
    - ► More uniform view of data if requested for two slightly different intervals
  - Much faster retrieval of historic trendline data
    - ► About 1 second for data of 1 week that updates every 5 seconds (incl. averaging)
- Beware! Influx is a "sharded" time series database:
  - Data with future (wrong) timestamp compromises performance
  - Had a case with timestamps months in the future:
    - ► Caused delayed data availability: > 1 hour instead of 30 seconds

InfluxDB beyond the Karabo Control System: Grafana

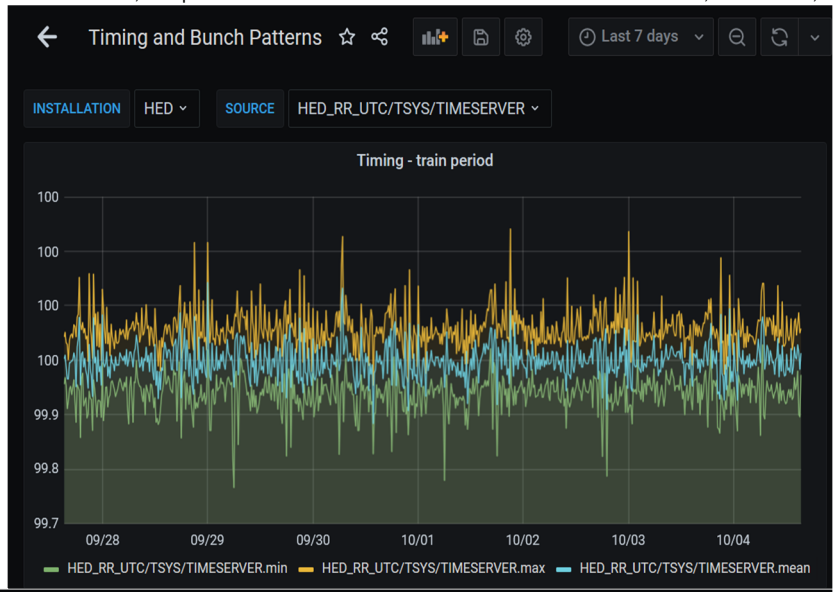
- Using non-custom data storage backend opens doors
  - Decouple analysis of historic data from control message flow (i.e. broker)
  - Allows visualisation with external tools

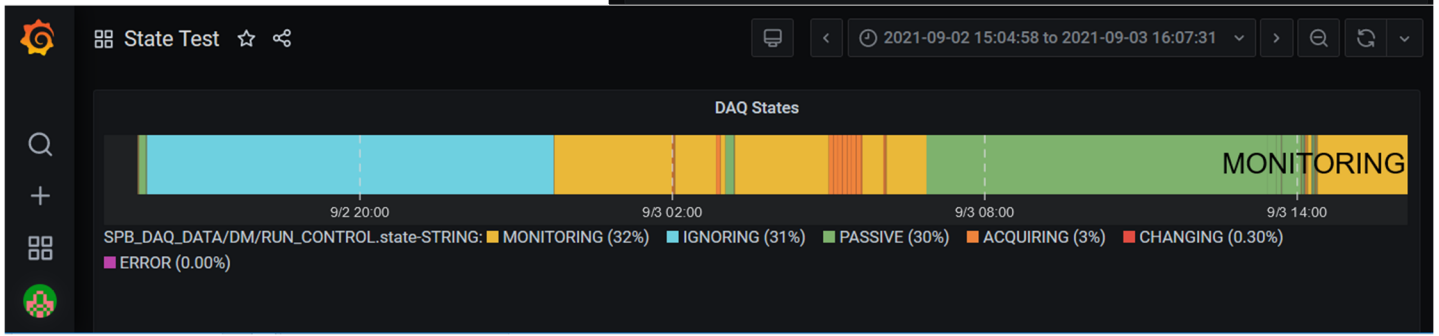
- Grafana web service
  - Feature rich
  - Powerful query language: *Flux*
  - Online community with examples
    - ► Low entry threshold for non-developers
  - Access control to view or edit "dashboards"
  - More difficult for non-numeric data

- Example: Correlate data of different Karabo devices
  - Here: ratio of 2 beam intensity measurements

Grafana Examples

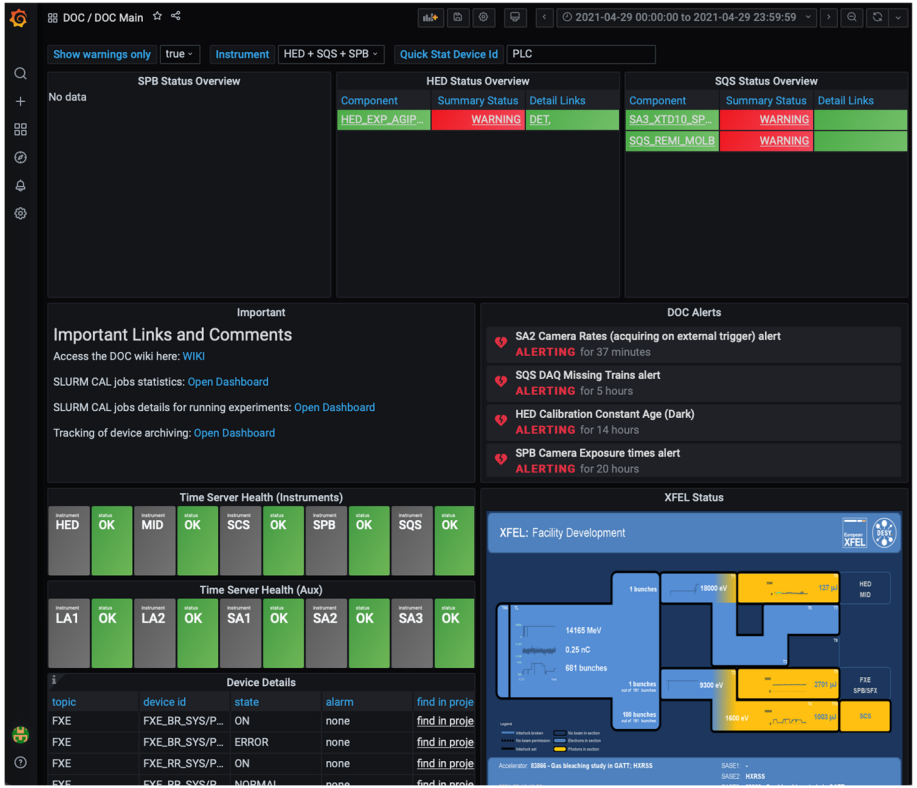
- Different options to down sample data
  - Here: Overlay Min, Mean and Max
- Visualise string changes
  - Grafana plugin "Discrete"
  - Here: State of the **D**ata **A**c**Q**uisition





Enabling Centralised Monitoring

- EuXFEL Data Operation Centre: **DOC**
- Unified monitoring and support as a co-effort
  - Controls, Electronics, Data Analysis, IT & Data Management, Detector Operation
- Monitors overall system health via many Grafana dashboards
- Non-software staff actively creates monitoring dashboards



Summary

- **Karabo** controls photon beam-lines and instruments at the European XFEL
  - Broker-based and event-driven
- Its custom-built text-file based data logging system lacks scalability
- Replaced in 2020 by a time series data base: **InfluxDB**
  - Transparent transition for Karabo users
  - Increased data reading speed
  - Can afford 3 years data retention
- InfluxDB backend enables data visualisation with external tools: **Grafana**
- Rich feature set of Grafana enables the new **Data Operation Centre** to quickly spot operational problems
  - Often before users and operators at the scientific instruments notice.

Authors

- Members (and former members) of the European XFEL Groups for Controls and IT/Data Management:
  - G. Flucke, V. Bondar, R. Costa, W. Ehsan, S. G. Esenov, R. Fabbri, G. Giovanetti, D. Goeries, S. Hauf, D. G. Hickin, A. Klimovskaia, A. Lein, L. Maia, D. Mamchyk, A. Parenti, G. Previtali, A. Silenzi, D. P. Spruce, J. Szuba, M. Teichmann, K. Wrona, C. Youngman