PAUL SCHERRER INSTITUT

PSI

Dominik Werder :: Paul Scherrer Institut

Modular, maintainable and scalable archiving of EPICS Channel Access and general timestamped data

ICALEPCS Cape Town :: October 11, 2023

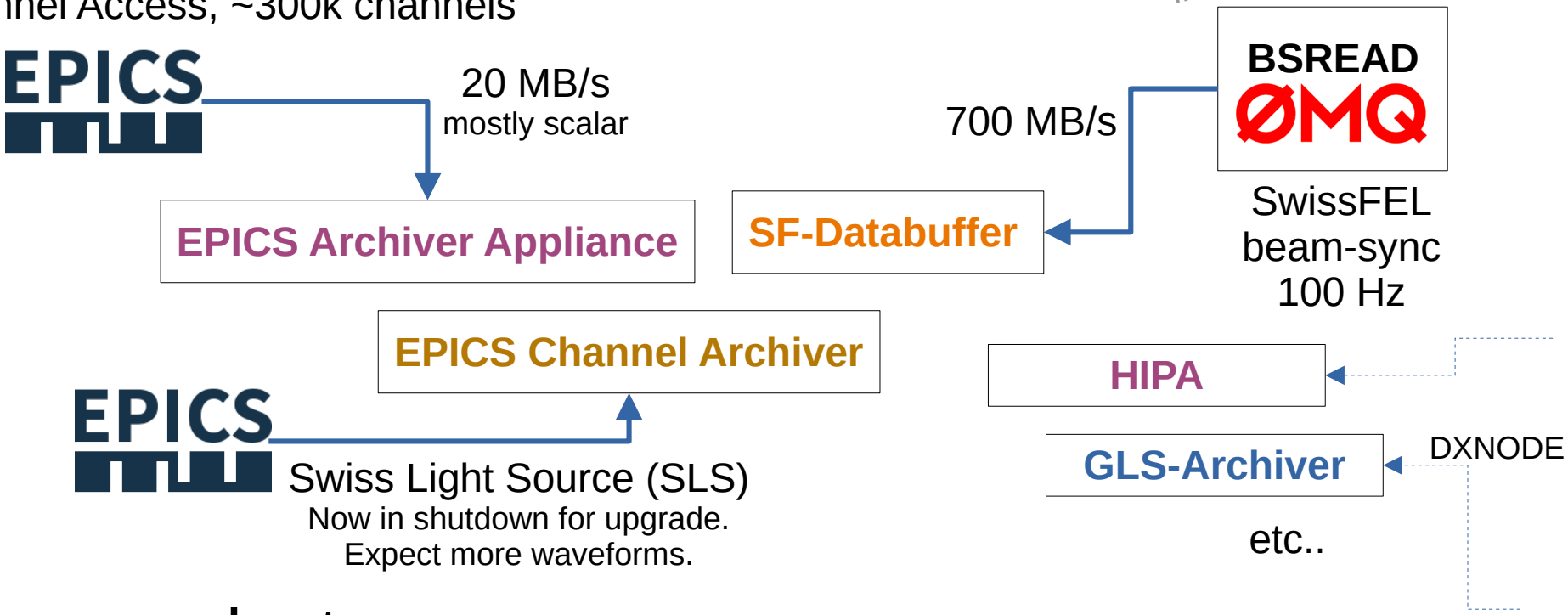
Archiving and Buffering at PSI Accelerators

Archiving and Buffering at PSI Accelerators

Archiving and Buffering at PSI Accelerators

SwissFEL:
Channel Access, ~300k channels

Simplified, wild, but still incomplete picture..



Many products.

Some are deprecated.

Maintain?

De-centralize?

Scale?

Availability?

New features?

Goals for future archiving

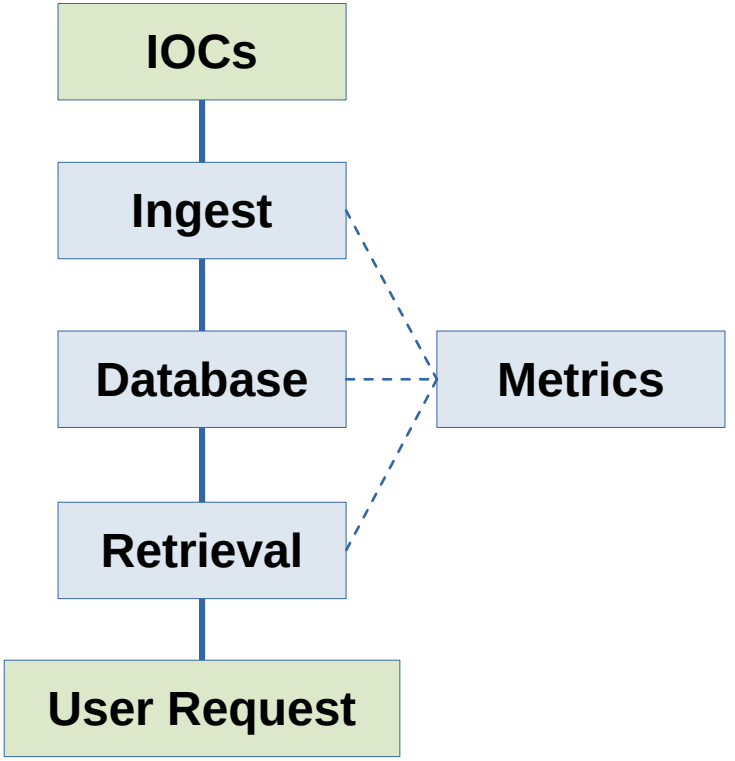
Unify Archiving

- Reduce number of products.
- Same setup across facilities.

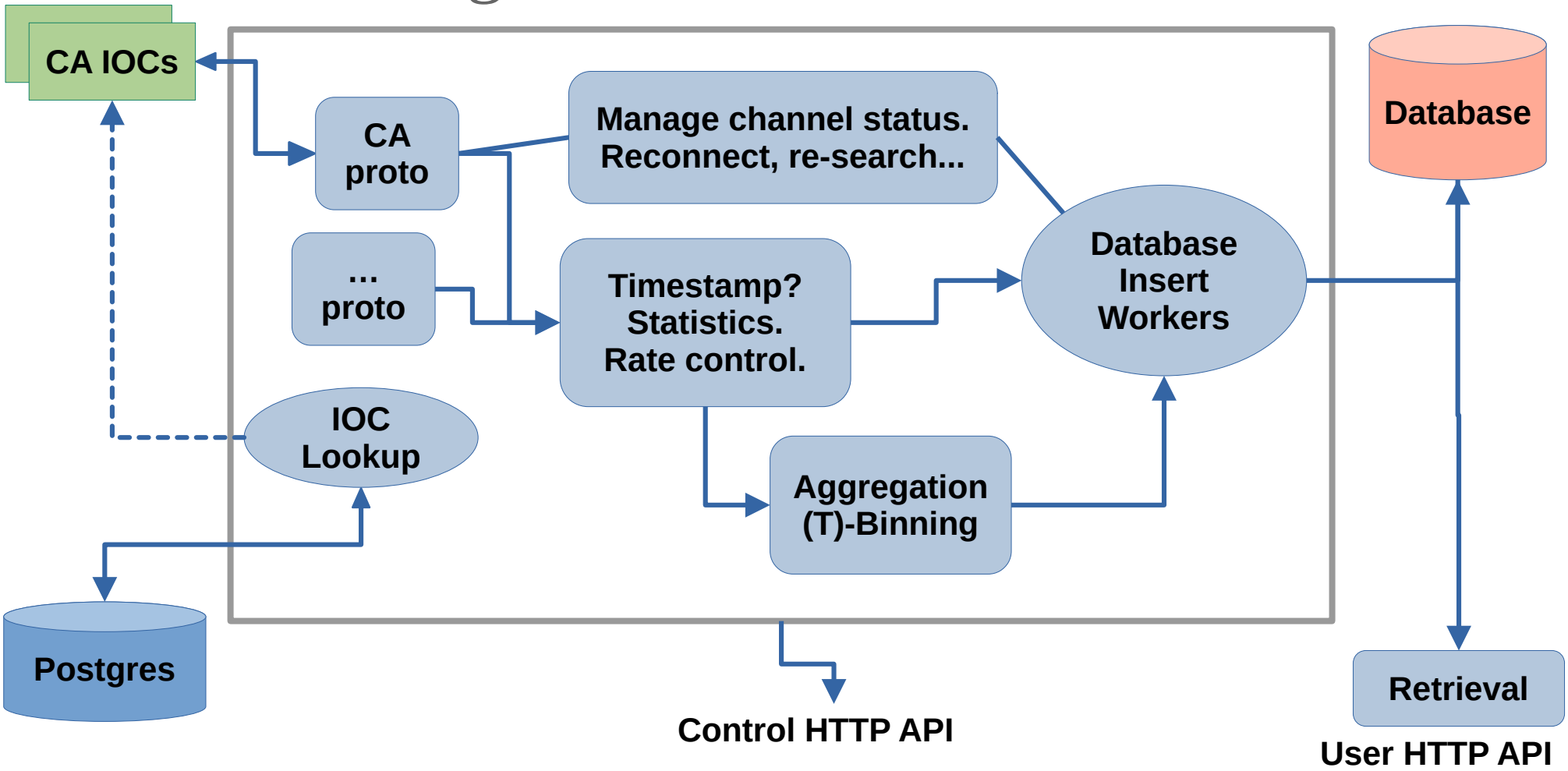
Simplify Operation

- Scale by adding nodes.
- Avoid manual sharding.
- Redundancy and availability.
- Configurable at runtime (REST).
- Inspection, monitoring.

Increase Modularity



Ingest Service Data Flow



Ingest Service Architecture

Rust:

- Memory safety.
- No garbage collector.
- Ownership and borrow model.
- Codegen via LLVM.

Async execution: 100% futures, using async/await, Tokio executor.

No manual locking, no manual threads, no shared memory.
Channels to pass data.

Shared library dependencies: only **basic** OS libs.

Test under **Valgrind** with reduced random production load.

ScyllaDB as Archiver Database

ScyllaDB

- Open Source, but with enterprise support.
- Rewrite (C++) of Cassandra architecture.
  (typed key-cluster, with cluster an ordered list of values)
- Hot Scalable: add/remove nodes at runtime.
  No node is special.
- Replication (can depend on channel).

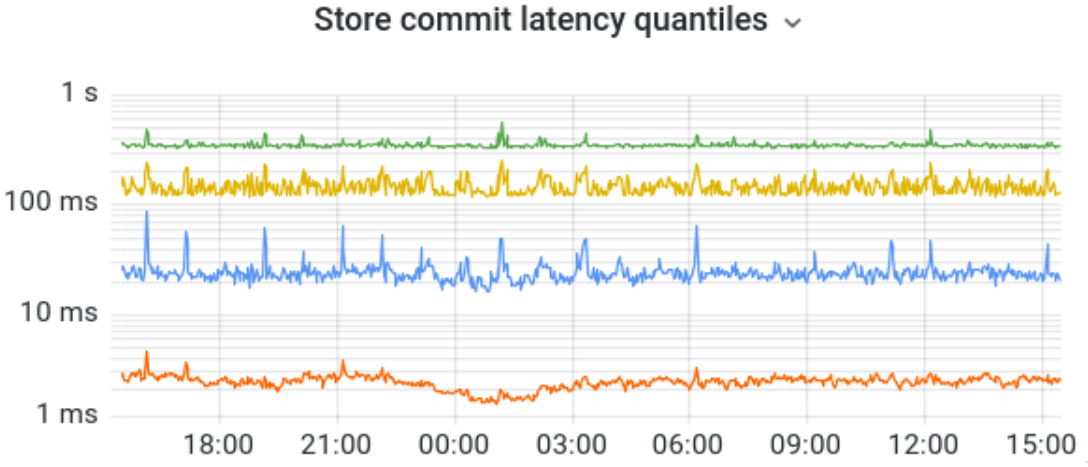
Metrics and Monitoring

Instrumented to monitor operation.

Prometheus

Grafana

Evalute our new hardware.



Quantiles shown:
0.999, 0.99, 0.90, 0.50.

Ingest SwissFEL Channel Access

Development cluster, 4 nodes ScyllaDB:

- Xeon E5-2680 v2 “Ivy Bridge”, year 2013, spinning disks…
- Ingest service with 310k SwissFEL channels, from 1290 IOCs:
  - 270 k/s
  - 21 MB/s

Test on more modern, 2 nodes ScyllaDB, SSD:

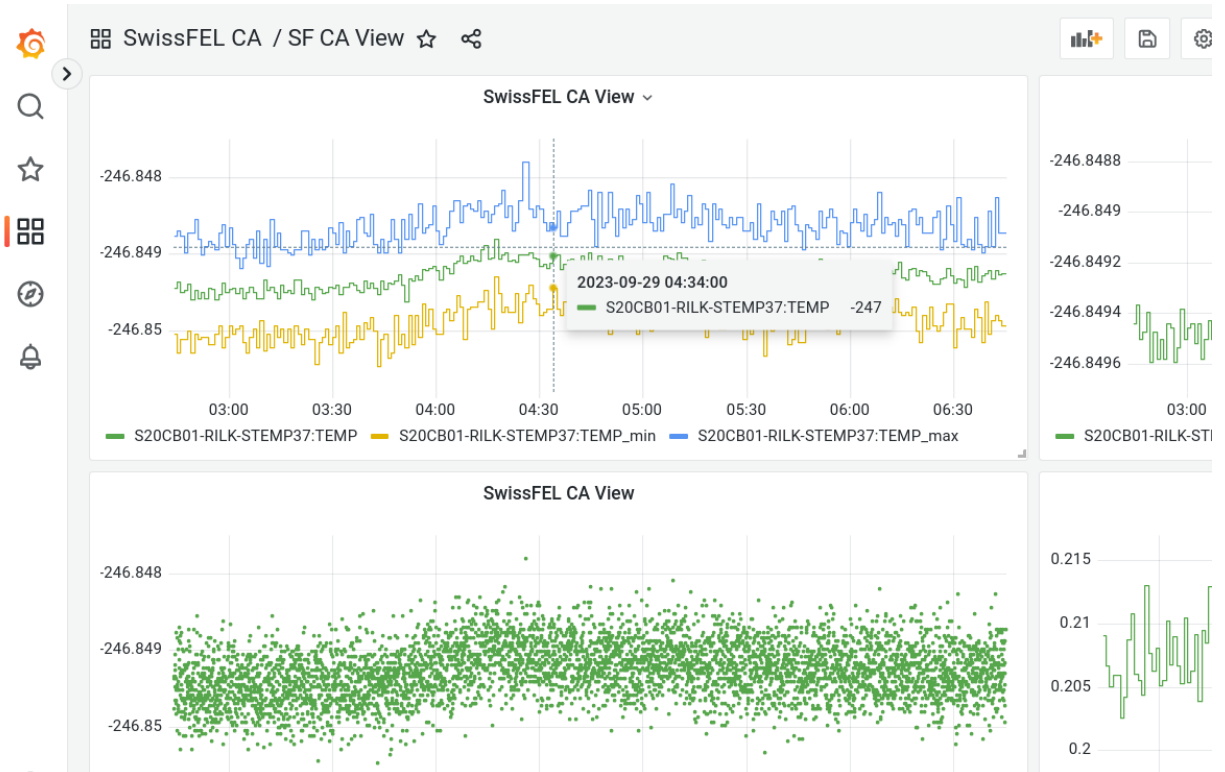
- 60k channels, run each test 30 min.
- 100% scalar float64, 1000 k/s.
- 70% scalar int32, 10% wave 8kB, 10% 64kB, 10% 512kB: 360 MB/s.
  P99.9 insert latency: 46 ms. Max: 85 ms. Replication factor: 2.

Looking forward to receive the new hardware

Retrieve Data from Database

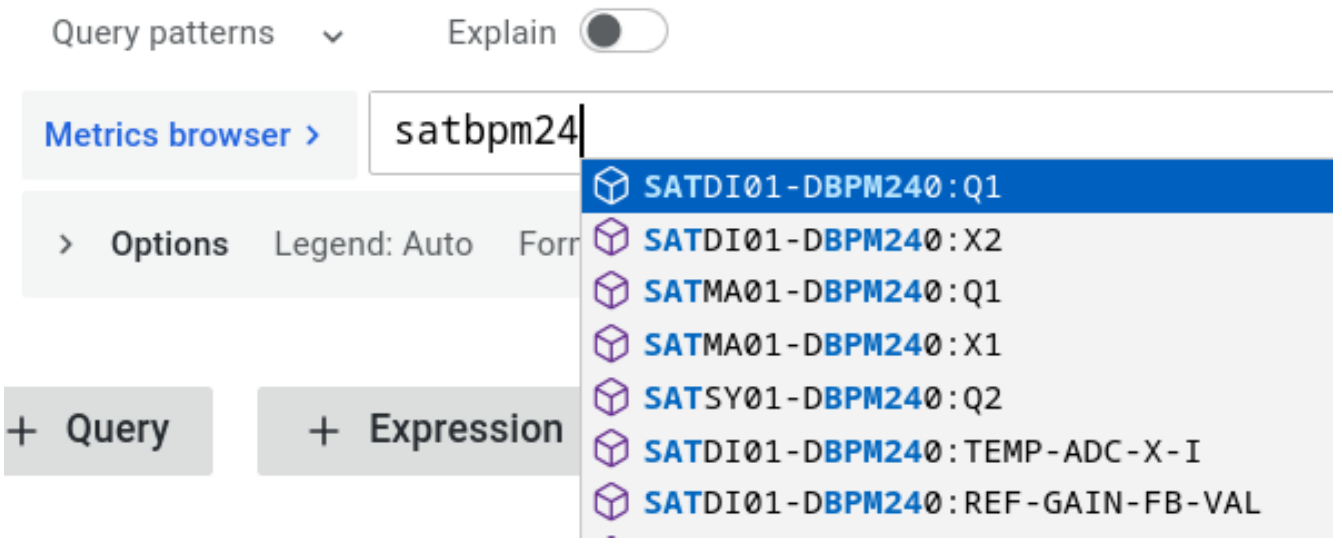
Retrieval service, REST, various formats:

Interface with Grafana.



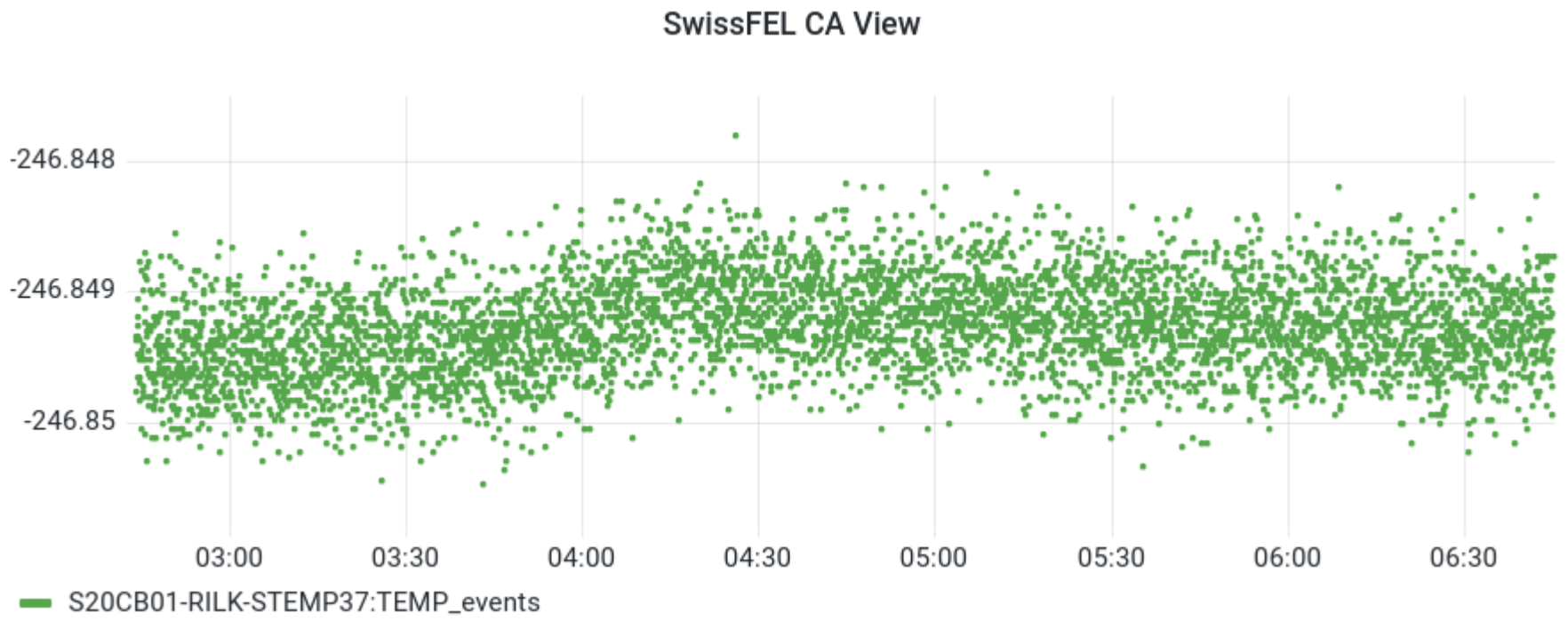
Basic Grafana Viewer

Auto-completion:



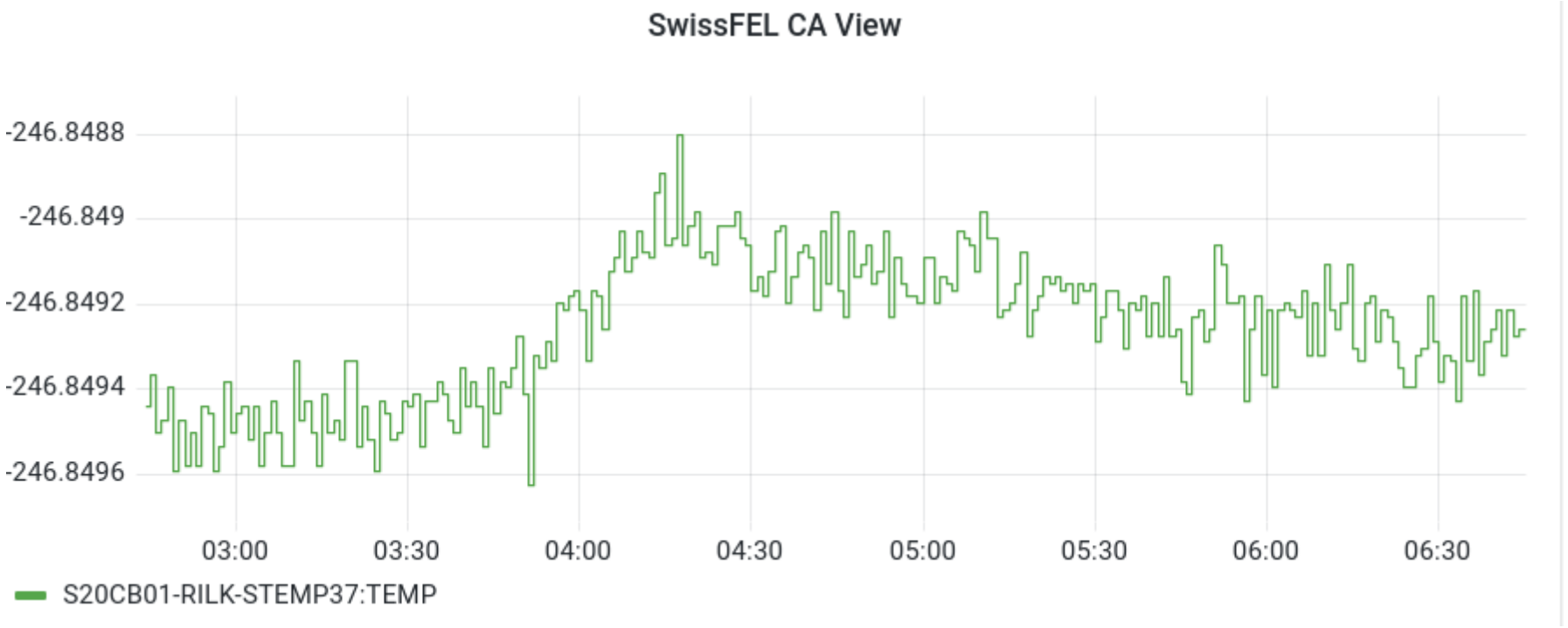
Basic Grafana Viewer

Plot individual channel changes:



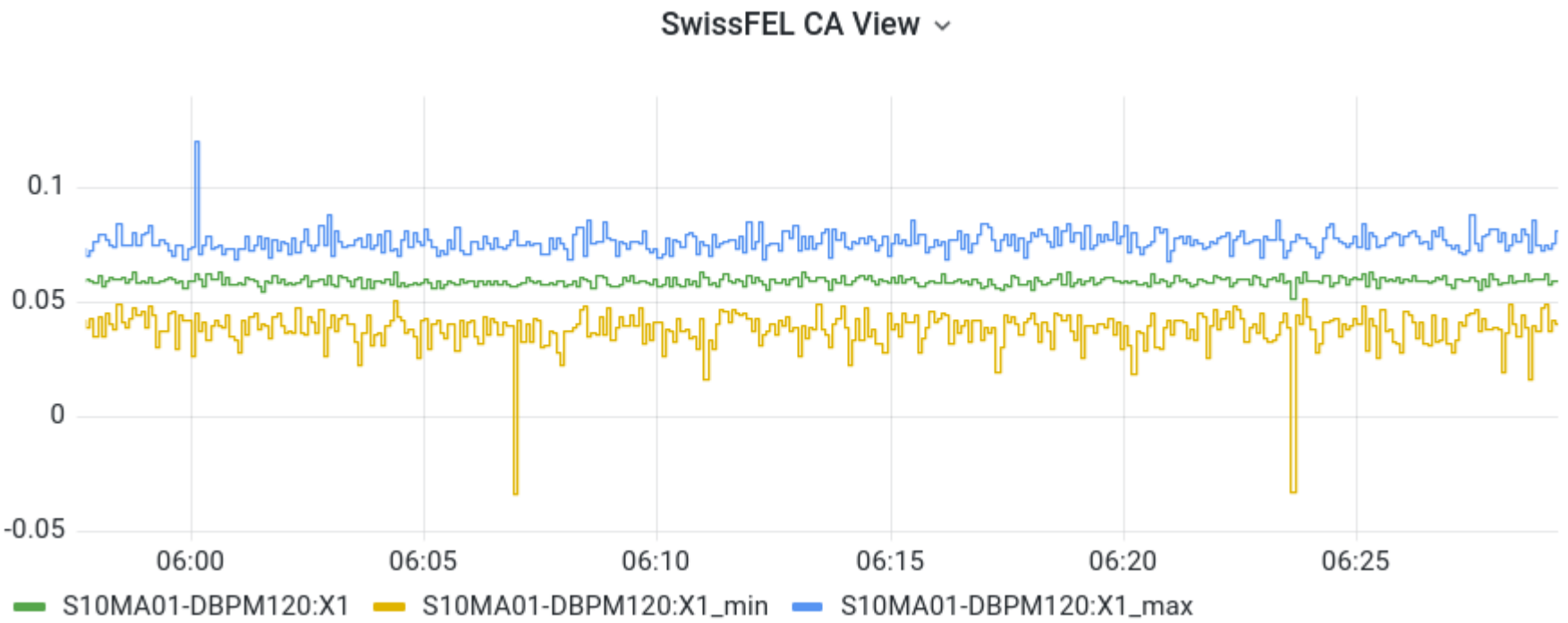
Basic Grafana Viewer

Time-weighted binned average:



Basic Grafana Viewer

Time-weighted binned average, min, max:



Outlook

Prepare for beta test in Q1 2024:

- Plan: run beside existing, compare.
- Refactor and polish.

Receive and evaluate new hardware:

- 4 nodes Scylla cluster.

Plans:

- Compare with other database.
- Support more inputs (bsread, …).

Thanks:
T. Humar, controls colleagues.