WinCC Open Architecture

Next Generation Archiver (NGA)

a progress update

Rafal Kulaga, Antonin Kveton
BE-ICS in collaboration with Siemens/ETM

CERN Openlab Technical Workshop
27th March 2024

What is WinCC OA?

- WinCC OA is a SCADA (Supervisory Control and Data Acquisition) system widely adopted by CERN
- Used for monitoring and control of industrial processes and equipment

Examples of WinCC OA systems at CERN

>800 instances deployed at CERN

gas distribution

electric grid

vacuum

environment & radiation

detector controls

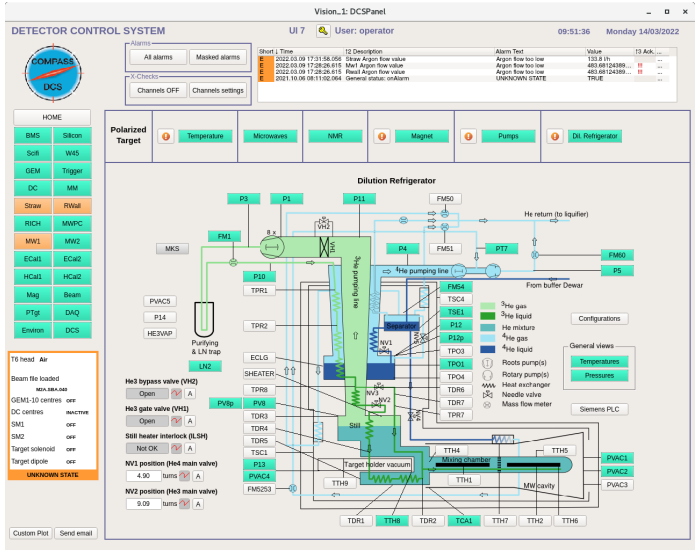
interlocks & safety

cryogenics

cooling & ventilation

Examples of WinCC OA systems at CERN

COMPASS experiment Detector Control System UI



Examples of WinCC OA systems at CERN

gas distribution

electric grid

~ 300,000 signals
~ 10 million events per day

vacuum

environment & radiation

detector controls

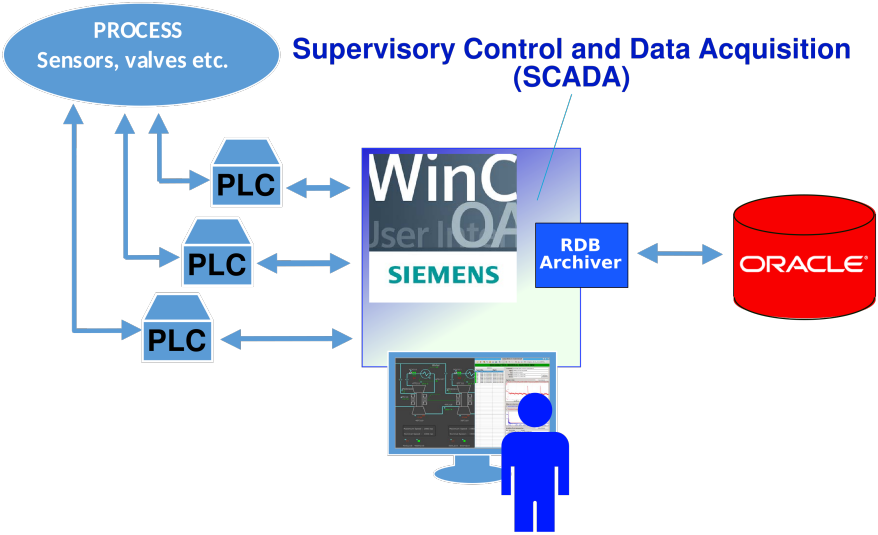
interlocks & safety

Quench Protection System
~ 130,000 signals
~ 17 billion events per day

cryogenics

cooling & ventilation

Archiving in WinCC OA – old architecture

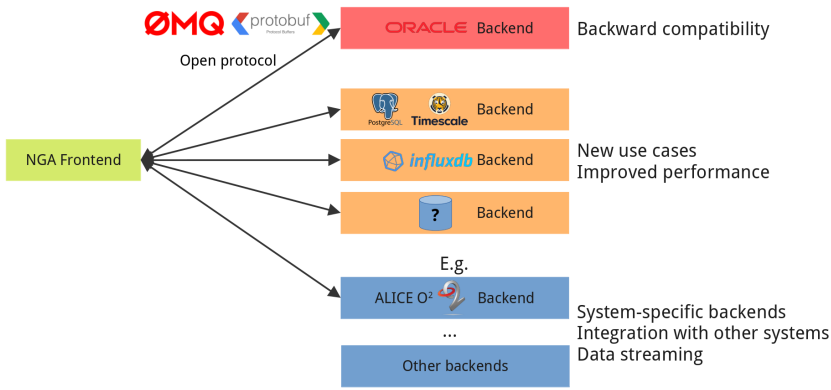


Old architecture worked well for more than 15 years, but...

- Technical debt and closed architecture prevented us from enabling new use cases
- There is a need to support alternative databases to reduce vendor-lock in
- Many database systems exist in today's market which are open-source and free
- Some of them are specialized to store time-series data, which is exactly our use case
- Next Generation Archiver enables archiving to those database systems

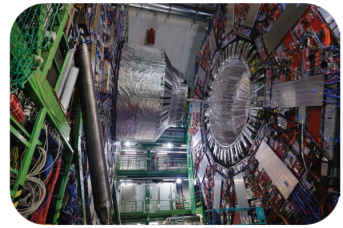
Next Generation Archiver (NGA) – architecture

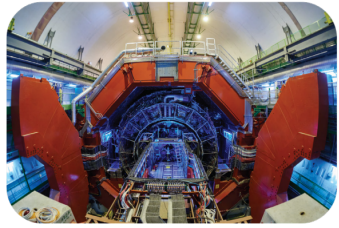
- One generic frontend which forwards queries to pluggable DB-specific backends
- Possibility to use multiple backends in parallel
- Custom backends can be written by the user → generic solution for streaming and custom processing of data coming from WinCC OA systems



Next Generation Archiver – 2023 highlights

- All the big 4 LHC experiments were migrated to the NGA with the Oracle backend during the last year-end technical stop – 3 years earlier than originally planned
- Smooth migration of more than 500 systems was ensured by automating all the steps
- Significant progress on the TimescaleDB support





TimescaleDB

- A PostgreSQL extension specialized in handling time-series data while remaining relational
- PostgreSQL is already supported by the DB On Demand service at CERN
- Fully compatible with SQL
- Data is organised in so-called *hypertables*, which is a way to partition the data in time (and optionally other dimensions)
- Comes with many interesting features, such as data retention policies, compression and continuous aggregates (cached results of aggregate queries – e.g., daily averages of your data)

Progress on TimescaleDB support in 2023

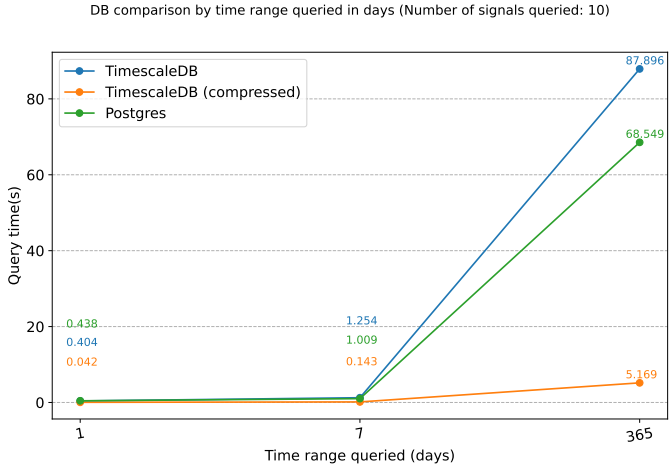
- Initiated talks with IT-DA regarding central CERN support for TimescaleDB
- Continued benchmarking efforts of TimescaleDB (more on that soon)
- Continued development of the TimescaleDB backend (considerable progress)
- Wrote a major test suite for the backend

Benchmarking databases

- We have implemented ArchBench – a toolkit that allows us to imitate write and read workloads of large distributed WinCC OA systems
- Initially, the benchmarking focus was on finding an optimal TimescaleDB schema and evaluating compression performance
- We evaluated performance of different kinds of indexes, types of partitioning, ways of storing arrays, ways of writing into the database (INSERT vs COPY) and much more...
- Recently, we ran distributed benchmarks at a scale very close to what would be a real workload at CERN

Key findings from PostgreSQL and TimescaleDB benchmarks

- Measured write performance of around 100,000 rows/sec for a single session is more than sufficient for our use case
- Using TimescaleDB compression results in $>$ 90% space savings
- Compression also significantly improves the performance of typical queries:



Plans for 2024 and beyond

- Provide support to more than 500 production systems at CERN that are using the NGA
- Continue the development of the TimescaleDB backend and release its preview version for testing at CERN
- Continue the discussions with the database group regarding central CERN support for TimescaleDB
- Evaluate different ways of using continuous aggregates in our use cases
- Finish missing features in the Oracle backend

Thank you!