Next Generation Archiver (NGA) for WinCC OA

Martin Zemko,
Rafal Kulaga

BE-ICS-STF in collaboration
with Siemens / ETM

CERN openlab

Outline

WinCC OA

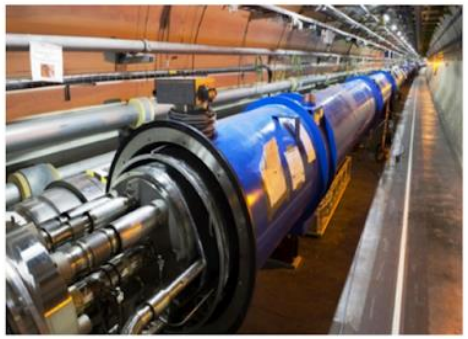
- WinCC OA is a SCADA (Supervisory Control and Data Acquisition) system widely adopted by CERN
- Used for monitoring and control of industrial processes and equipment

WinCC OA at CERN

- WinCC OA is used in more than **800 systems** across CERN


Gas distribution


Electrical grid


Vacuum


Environment


Detector controls

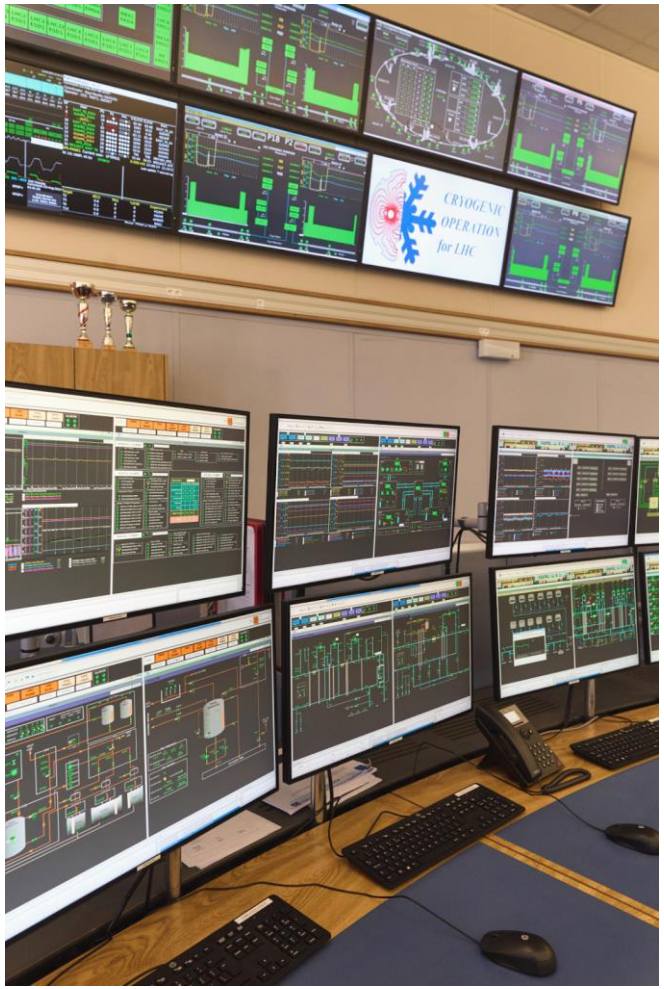

Interlocks, safety


Cryogenics


Cooling, ventilation

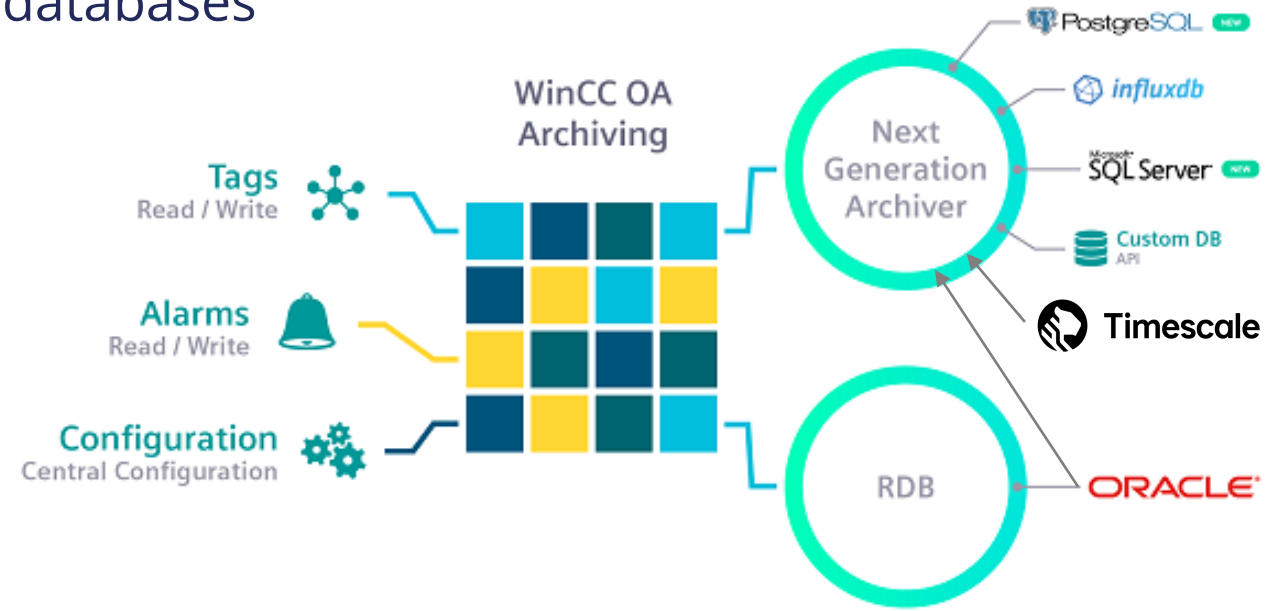
WinCC OA Archiving at CERN

- Archiving is the essential feature of WinCC OA
- It archives time series data (events & alarms)
- Supports both writing and querying of historical data



Next Generation Archiver

- Original RDB Archiver supports Oracle database only
- Next Generation Archiver supports multiple databases through pluggable backends
- Some of them are specialized for storing time-series data (e.g., TimescaleDB, Influx)



Source: Siemens
https://support.industry.siemens.com/cs/document/109796197/delivery-release-–-wincc-open-architecture-v3-18

CERN strategy:

- Oracle backend – backward compatibility with RDB Archiver
- **TimescaleDB backend** – new solution, to be deployed in parallel to Oracle

Progress in 2024

- Finished the migration of all ALICE, ATLAS, CMS and LHCb systems to the NGA (with the Oracle backend)
- Finalized initial TimescaleDB backend implementation → already in evaluation by ATLAS
- Multiple improvements and bugfixes in the TimescaleDB backend
- Extensive benchmarking of the TimescaleDB database and backend
- Integration of WinCC OA software into the Archive Benchmark suite (benchmarks can include WinCC OA)

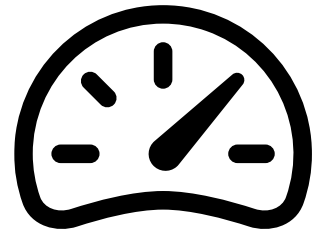
Database benchmarking

We needed to:

- Evaluate different databases and schemas in various scenarios
- Perform tests with different read and write workloads
- Ensure reproducibility of benchmarks

Archive Benchmark toolkit

- Python-based database benchmarking toolkit developed by our team at BE-ICS-STF
- Specialized in time-series data



ArchBench

TimescaleDB

PostgreSQL extension for efficient handling of large-scale time-series data

Main features:

- Simple deployment on the Database-on-demand service
- Reliable, scalable, and highly performant solution
- Automatic time-based data partitioning of hypertables into chunks
- Data compression
- Continuous aggregation

Example test data sets

Medium data set

1-year data: 1.3B rows, 209.7 GB
4-year data: 5.3B rows, 826.2 GB

Update frequency	Changes per day	Series count	Rows per year
Very low	1	20k	7.3M
Low	10	10k	36.5M
Medium	100	5k	182.5M
High	1k	2k	730.0M
Very high	10k	100	365.0M
		Total rows	**1,321.3M**

Extra-large data set

1-year data: 9.3B rows, 1.4 TB

Update frequency	Changes per day	Series count	Rows per year
Very low	10	500k	1,825M
Low	100	50k	1,825M
Medium	1k	5k	1,825M
High	10k	500	1,825M
Very high	100k	50	1,825M
		Total rows	**9,125M**

Performance measurements

- We focused mainly on **PostgreSQL** and **Timescale**

PostgreSQL	**Timescale**	**Timescale**
(DB part only)	(DB part only)	(DB & client part)
67k	**77k**	**40k**
rows per second	rows per second	rows per second

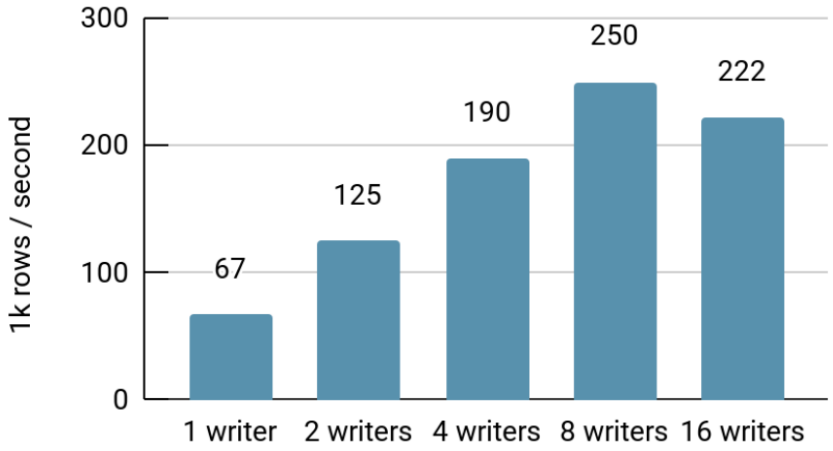
Reference: 20k rows/s total write throughput for 250 connections for an example production Oracle DB

Timescale performance remains very good even with:

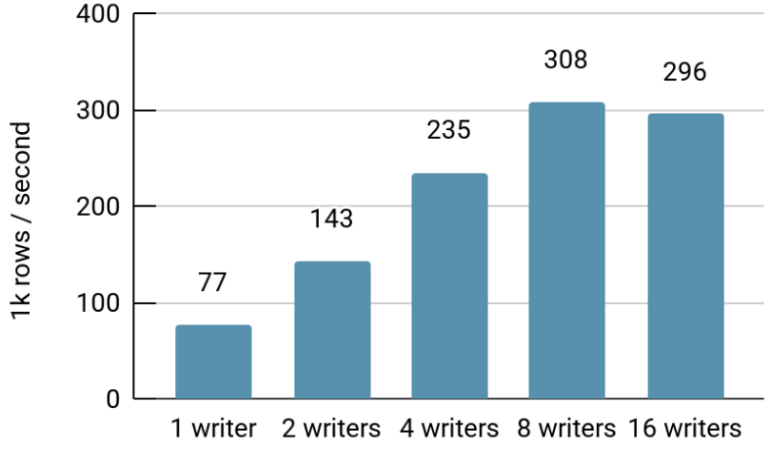
- large data sets
- high number of time series
- multiple parallel writer

Parallel writing

PostgreSQL



Timescale

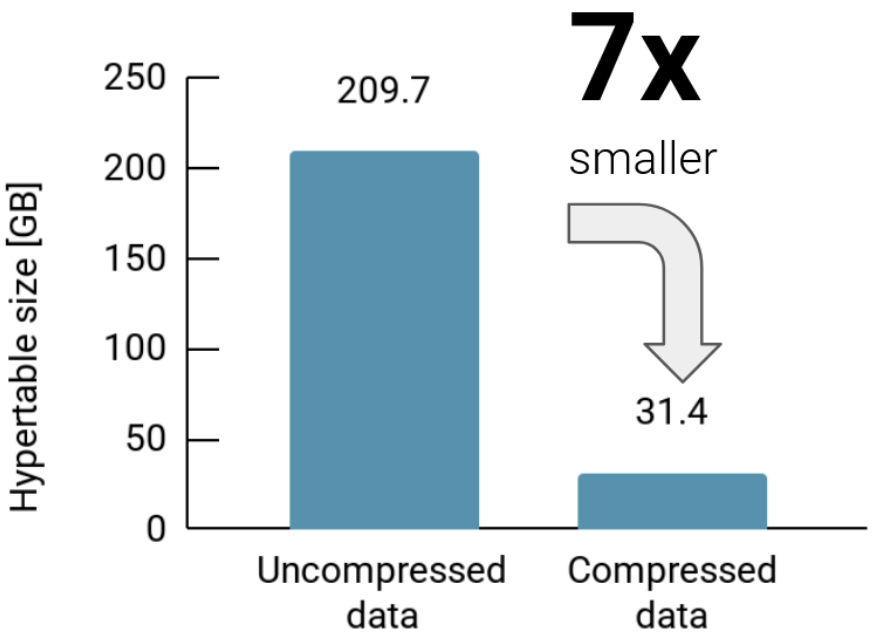


- Our benchmarks show high scalability with the number of writers
- We plan to explore writing performance with more than 16 writers in the near future

Data compression

Significant storage space savings

- between 78% and 87% for synthetic data
- between 90% and 95% for production data



Compression is a CPU-intensive operation

- 2 hours to compress the medium data set
- 2.5 hours to decompress the medium data set

Plans for 2025 and beyond

- Further tests and development of the TimescaleDB backend
- Pilot deployments in production systems (in parallel to the Oracle backend)
- Scaling up the benchmarks to the full size of our production systems
- Improving the integration of new use cases supported by the NGA and TimescaleDB
  - Parallel archiving
  - Data downsampling
- CERN-wide production deployment of the TimescaleDB backend from Q2 2027

Summary

- NGA is now the primary archiving solution at CERN, supporting over 500 systems
- TimescaleDB backend will complement the Oracle backend as one of CERN's two main archiving solutions
- TimescaleDB has proven to be the most suitable technology for further evaluation and development
- ArchiveBenchmark toolkit has demonstrated exceptional flexibility, and we will continue using it for development and testing

THANK YOU!

CERN openlab