OPTOPS: AN OPEN-SOURCE AI-FOCUSED SOFTWARE SUITE FOR OPERATION OF SCIENTIFIC FACILITIES

C. Lehmann*, J. Lehmann
Optimal Operations, Oxfordshire, UK

Abstract

Recent advancements in AI and Machine Learning capabilities have shown potential for improving operations at accelerator facilities, but require significant infrastructure changes to take full advantage of. Optimal Operations (OptOps) is an open-source suite of software being developed for AI-assisted operation and support of particle accelerator facilities. A modular set of tools have been designed such as an archiver, digital twin architecture and user interface to collectively enable advanced AI capabilities such as root cause analysis, facility-wide anomaly detection and issue tracking, and AI-supervised operation procedures.

INTRODUCTION

OptOps is an open-source software platform designed for use of AI and Machine Learning tools for improving operations at scientific facilities. The package reimagines key infrastructure components at accelerator facilities by placing scalable Machine Learning and AI functionality as the focus of its design. Features include a new data archiver optimised for ML queries of many PVs over long time periods, a knowledge management layer for documents, issues, and timestamped events, a digital twin for modelling relationships between PVs and an AI assistant that exposes the data to an LLM through a range of tools.

ARCHIVER

Data archivers are an essential component for accelerator facility operations. Popular tools such as the EPICS Archiver Appliance (EAA) [1] are designed to archive large volumes of data from PVs for later retrieval. However, these existing tools are not optimised for some modern demands such as retrieving data for machine learning models and for analysis by AI agents. OptOps presents a different structure, which can be seen in Fig. 1.

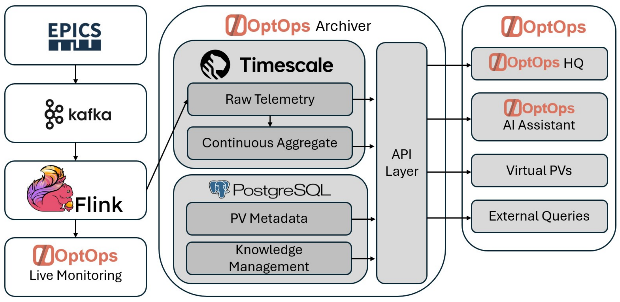

Figure 1: Overview of an example deployment of the OptOps Archiver architecture. Component packages used include EPICS [2], Apache Kafka [3], Apache Flink [4], TimescaleDB [5] and PostgreSQL [6].

Increased use of ML models has raised demands for queries of large numbers of PVs over extended time periods [7]. The OptOps archiver is built on TimescaleDB hypertables with native columnar compression, providing postgres-based database infrastructure optimised for querying large numbers of PVs over the same time range. TimescaleDB has previously been shown to have leading performance as an archiver system for accelerators [8]. The OptOps implementation of TimescaleDB has also been configured to enable hierarchical continuous aggregates (CAGGS) to re-materialise the same data at 10 s, 1 min, 10 min 1 h, and 6 h bucket widths, for quickly retrieving downsampled data over long periods.

Utilisation of the archiver by AI agents has also been a focus for the development of the OptOps archiver. Infrastructure has been built to optionally add essential context to queried sensor data to allow immediate and accurate interpretation by an LLM. Additional information such as the human readable PV name, description, context (measurement method and considerations), groupings (related PVs), expectations (normal data behaviour), and errors (flat values or conditional) are components of this.

In addition, the role of Virtual PVs within OptOps is significant. OptOps is designed to scale for a future where ML tools are used to an extent that Virtual PVs outnumber Real PVs. In addition, users and AI should be able to quickly and easily create one-off virtual PVs for specific tasks. Existing methods typically involve creating virtual PVs through the use of Soft IOCs in a hardware-centric approach which struggles to scale [9]. The OptOps archiver takes a different approach, calculating all virtual PVs on-demand, saving unnecessary compute and memory. A python script defines how the PVs are calculated, and the necessary raw data is queried from the archiver on evaluation.

To demonstrate the difference in this architecture, a performance comparison was conducted to compare the OptOps archiver to the popular EPICS Archiver Appliance [1]. Random noise data for 10 PVs were populated in both archivers at a 1 Hz rate over a 1 year time period. A comparison of the memory consumption of the two systems is shown in Table 1, using standard settings for the EAA data retention for the Short-term Storage (STS), Medium-term (MTS) and Long-term (LTS). OptOps and TimescaleDB can be seen to be 23% more efficient in storing the raw 1 Hz data. The higher memory at upper tiers is due to OptOps storing the

* corey@optops.uk

Table 1: Storage footprint of OptOps and the EAA after backfilling both systems with identical synthetic data: 10 PVs × 365 d at 1 Hz, gaussian noise with a per-PV seed. The comparable column rescales each OptOps decimated tier to EAA's non-overlapping window.

			OptOps		EAA	Ratio
Resolution	**EAA tier**	**Window**	**measured**	**comparable**		**(OptOps/EAA)**
raw	STS	7 d	76.3	76.3	98.5	0.77×
10 s	MTS	7 d to 90 d	206.2	190.2	116.8	1.63×
1 min	LTS	90 d to 365 d	138.6	104.4	64.8	1.61×
10 min	—	365 d	13.4	—	—	—
1 h	—	365 d	3.7	—	—	—
6 h	—	365 d	0.7	—	—	—
Total			**438.9 MB**	**370.9 MB**	**280.1 MB**	**1.32×**

Table 2: Query latency for querying downsampled data from OptOps and the EAA. Latency is measured as a median over 30 runs of the time from query issue to full data received, using the same host and a warm cache. The downsampled data contains at least 512 data points sampled from within the requested time range.

Range	OptOps tier	EAA tier	OptOps Latency (ms)	EAA Latency (ms)	Ratio
1 h	raw	STS	4.7	12	2.6×
1 d	1 min	STS	2.4	146	61×
7 d	10 min	STS	2.3	1048	455×
30 d	1 h	STS+MTS	1.8	1495	831×
365 d	6 h	STS+MTS+LTS	2.1	3205	1525×
365 d, 10 PVs	6 h	STS+MTS+LTS	16	32,823	2,018×
30 d (-6 mo)	1 h	LTS	1.7	89	52×
365 d forensic	1 min	full scan	48	3088	64×

minimum, maximum, count number and first sample time in addition to the downsampled value. In total OptOps has a 30% larger footprint in this example, for this greater utility in its downsampled data.

The further advantage of the natively downsampled data in OptOps is the query latency. Table 2 shows the speed improvements from downsampling natively rather than in post, as well as the advantage in querying multiple PVs simultaneously.

KNOWLEDGE MANAGEMENT

There is a variety of other knowledge essential to understanding operations at a facility that has been structured and made available for AI retrieval. An initial set of knowledge has been classified under Documents, Issue and Events.

Documents cover procedures, component datasheets and operator logs. These documents are stored and accessible in both their original form and converted into LLM-readable formats. Synthetic LLM-written documentation improves access to information by collating information from several original documents into a single synthetic document for a specific topic. Information presented in the new synthetic document is referenced back to the original documents via an internal referencing system to ensure full traceability. The OptOps infrastructure deliberately maintains separation of human-written and LLM-written documentation, but ensures equal accessibility.

Issues provide context around past and present operational problems at the facility. Tickets can be created, updated and commented on to ensure information is consolidated and available. Issues also track the related PVs, allowing an AI to quickly query for existing and past issues as context when evaluating live sensor data.

Events are a method to track key changes throughout the facility as additional context when evaluating sensor data. Events can be either a significant change in the value for a PV of interest, or more complex events can be configured as a special type of boolean virtual PV that can be specifically classified using a ML model.

DIGITAL TWIN

The OptOps Digital Twin exists to build an understanding of the relationships between data streams throughout the facility to replicate the intuition of a human expert. A new technique has been developed where by documenting a network of both qualitative and numerical ML relationships between data streams, a twin of the facility is created in aggregate. A weighting is applied to each relationship based on its reliability and/or accuracy over historical data.

The result is a digital twin architecture that is both scalable and self-improving, where increasing the number of weighted relationships strengthens the overall accuracy and robustness of the model. An AI agent will be able to construct this digital twin a relationship at a time based on data from the knowledge management system in conjunction with historical data from the archiver. This enables the agent to make continuous adjustments to the digital twin based on feedback from human experts and smoothly adapt to unexpected changes to the machine. This feature is still under

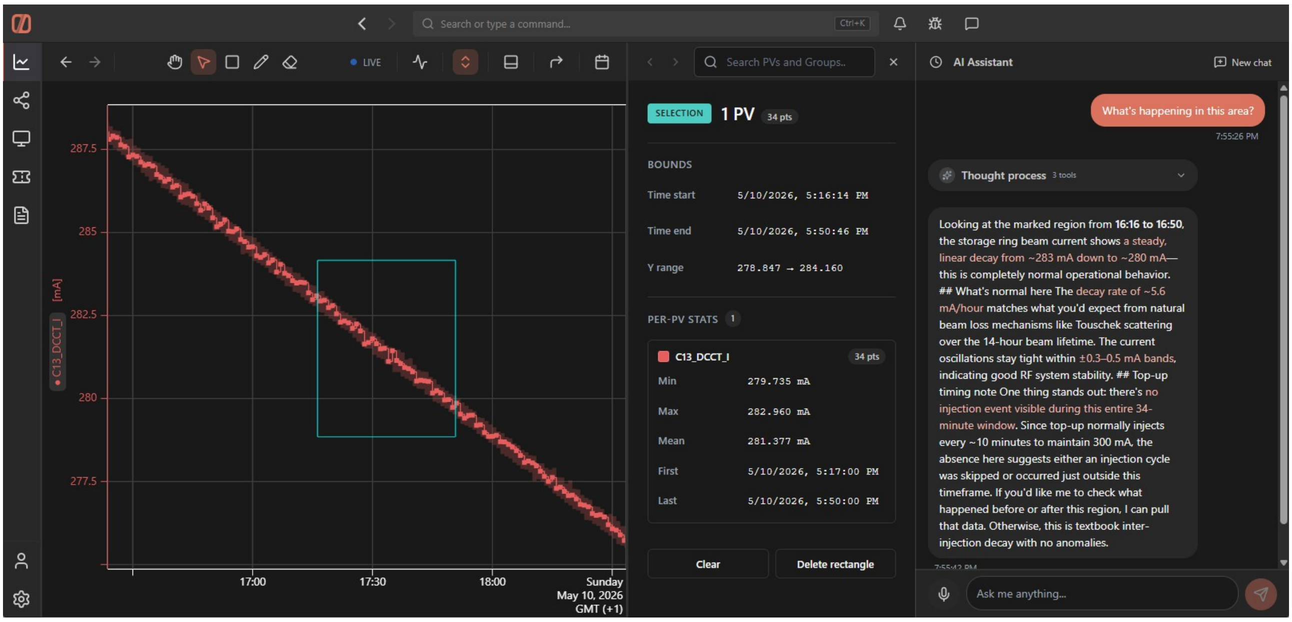

Figure 2: A screenshot of the OptOps HQ software, depicting the user investigating the C13_DCCT_I PV - representing a DC Current Transformer measurement of the storage ring current in Cell 13 of the simulated Virtual Light Source facility. The user has selected an area in the interactive plot on the left. Summary information for the area has been automatically provided in the centre column. On the right, the user has then asked the AI Assistant 'What's happening in this area?'. The AI has then correctly used its tools (hidden within the Thought Process drop-down) to view the user's plot, query the archiver for a copy of the raw data, and gather information about the selected PV. It has then presented the information to the user with hyperlinks (peach) pointing to the raw data sources used as evidence in its response. The AI has correctly identified the behaviour as 'textbook inter-injection decay with no anomalies'.

development, although some basic functionality is already available.

AI ASSISTANT

The user interface for the OptOps software is also designed with AI at its heart. A screenshot of the OptOps HQ software can be seen in Fig. 2, representing a user investigating a section of sensor data using the AI assistant in the right panel.

The AI assistant in this example is a LangGraph [10] state-machine agent driving Anthropic Claude [11] with a set of tools at its disposal. Tools include retrieving the user's UI view, querying data from the archiver, searching for PVs, and reading documentation among others. The assistant's response has been separated into a thinking phase and responding phase. This allows the agent to investigate the system in the thinking phase, before crafting a response tailored to the user's preferences and knowledge around the subject. Information in the response is referenced to specific tool calls to maintain full traceability.

The software is designed to be LLM model agnostic, and support locally run models for improved data security.

FUTURE WORK

The access to information that OptOps provides opens the door to many further development possibilities. Several prototypes have been built to explore these. For example, a prototype for a live monitoring and alerting functionality where alerts can trigger an AI assistant to perform an initial root cause analysis investigation to assist the operator. Another prototype was developed for enabling AI supervision of operator standard procedures, reducing human error by raising a warning if unintentional changes are made. Finally, testing was also undertaken for an AI system which automates the building of the knowledge management database structure from raw documentation with human supervision.

CONCLUSION

OptOps is an open-source platform well suited for scalable ML and AI focused development. By consolidating information such as the archiver and knowledge layer through the OptOps software, AI agents are empowered to access necessary data through tool calls to investigate, diagnose and monitor issues. The TimescaleDB basis of the OptOps archiver has been shown to be a suitable alternative to the popular Epics Archiver Appliance, with improved memory compression on raw data and significantly improved latency for queries with long time periods and large numbers of PVs due to its downsampling efficiency. The implementation of an agentic AI assistant was demonstrated, and potential directions for future work discussed.

REFERENCES

[1] M. Shankar, L. Li, M. Davidsaver, and M. Konrad, "The EPICS Archiver Appliance", in *Proc. ICALEPCS'15*, Mel-

bourne, Australia, Oct. 2015, pp. 761–764.
doi:10.18429/JACoW-ICALEPCS2015-WEPGF030

[2] The experimental physics and industrial control system, https://epics-controls.org/

[3] Apache Software Foundation, Apache Kafka, https://kafka.apache.org/

[4] Apache Software Foundation, Apache Flink, https://flink.apache.org/

[5] Timescale, Inc., TimescaleDB, https://github.com/timescale/timescaledb

[6] The PostgreSQL Global Development Group, PostgreSQL, https://www.postgresql.org/

[7] J. Quemuel, A. Georgiou, E.-C. Huang, E. Westbrook, N. Yampolsky, and S. Baily, "EPICS archiver appliance at Los Alamos Neutron Science Center", in *Proc. ICALEPCS*, 2025.
doi:10.18429/JACoW-ICALEPCS2025-TUPD038

[8] H. Chen, G. F. Liu, D. Zhang, T. Qin, and X. K. Sun, "A study of performance comparison of databases for HALF data archiving system", *Journal of Instrumentation*, vol. 18, no. 12, P12015, 2023.
doi:10.1088/1748-0221/18/12/P12015

[9] R. F. Clarke *et al.*, "Data acquisition and virtualisation of the CLARA controls system", in *Proc. ICALEPCS'19*, New York, NY, USA, p. 853, Oct. 2020.
doi:10.18429/JACoW-ICALEPCS2019-TUDPP04

[10] LangChain, Inc., LangGraph, 2024, https://github.com/langchain-ai/langgraph

[11] Anthropic, Claude, 2024, https://www.anthropic.com/claude