

JLab LDRD2410 Quarterly Report FY24Q4

Proposal Name:

Streaming Readout Real-Time Development and Testing Platform

Report Date:

Sep. 30, 2024

Principal Investigator:

David Lawrence PhD

1 Work-scope Highlights

Work continues with progress towards multiple milestones. Some issues encountered that slowed progress in some areas. This results in some schedule lag. The milestone schedule for the 2nd year proposal were adjusted slightly to account for that. We will continue to monitor this closely and will communicate with the program managers if we foresee any reduction in scope is necessary.

The major highlights are:

- Postdoc Ayan Roy left for faculty position at CNU
- Nathan Brei and Jeng Tsai added to project
- R&D on process monitoring using Prometheus with *process_exporter* and *node_exporter*
- DPUX development
- R&D on Data Lake technologies (Kafka, Redia, xMsg, Chronicle Queues)

2 Data Capture

Software that was developed earlier in the project to quickly verify the packet capture data format and generate basic plots was used as a basis for a new, JANA-based configuration. This implements a trivial event identifier and event filter. The algorithms are placeholders that can be used as examples for deeper analysis.

Aggregation of the packets from multiple streams is handled via a dedicated program merges EVIO blocks representing single time frames from each VTP into a correctly formatted EVIO bank where all VTPs are represented. Figure 1 illustrates the data processing chain described. A draft of this has been completed at the end of FY24Q4, but testing and the actual event identifier and filter algorithms are still required. This work will extend into FY25Q1.

3 Milestone Progress

Table 1 shows the status of the FY24 milestones.

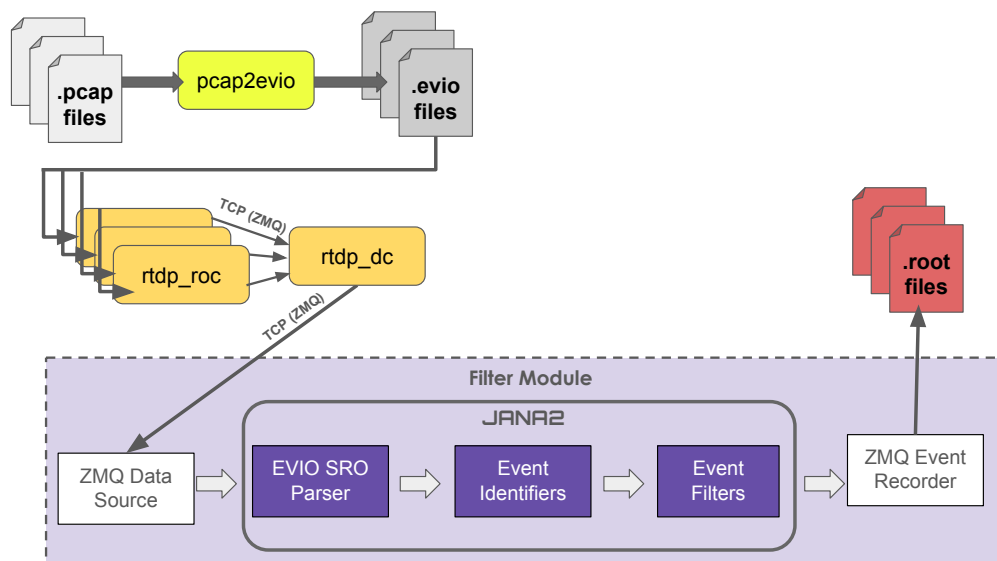


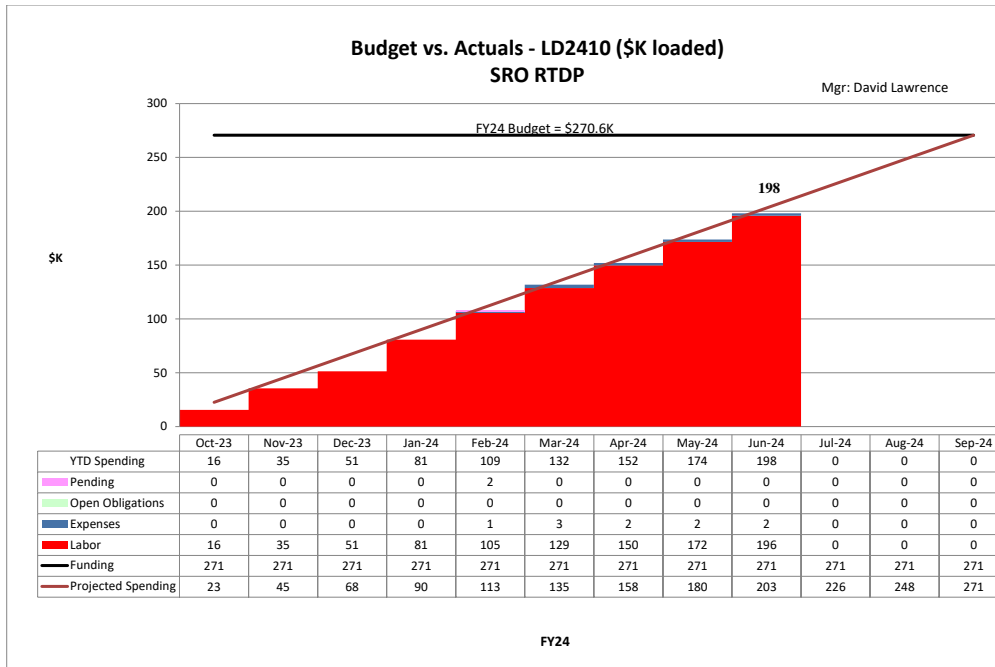
Figure 1: Data processing chain for data captured from CLAS12 during packet capture exercise in May.

ID	Task	status	Comments
M01	Create prototype ERSAP configurations for INDRA and CLAS12 test systems	■	A CLAS12 example and "Hello World" example have been placed in Github. INDRA has not been done yet.
M02	Identify or capture SRO formatted data from CLAS12 and INDRA test systems with data tag/filtering capability (output data ready for further offline processing)	✓	Data was captured at various beam currents from CLAS12 on Dec. 17. INDRA data capture done using pulser inputs to SAMPA setup.
M03	Evaluate existing solutions for configuring and launching remote distributed processes	✓	see evaluations in document on EPSCI wiki.
M04	Establish code repository(s), project site, and method of documentation	✓	This has been done here: https://github.com/JeffersonLab/SRO-RTDP
M05	Create stream splitter program for EVIO or HIPO data formatted files	✓	Created for GlueX. (See text for details on HIPO)
M06	Create stream splitter program for simulated data in PODIO for ePIC	✓	Prototype tested using FABRIC testbed. Simulated ePIC data sent from CERN to 8 different US sites.
M07	Create VTP emulator using files produced by stream splitter	✓	Done for GlueX raw data. Addressed using tcpreplay for CLAS12 packet captured data.
M08	Create controller program to synchronize multiple VTP emulators	✓	Satisfied through alternate design using synchronized system clocks.
M09	Determine appropriate schema for all aspects of monitoring	■	Monitoring info. extracted as JSON records from both docker and /proc sources on Linux. Display in Grafana prototyped, but not yet complete.
M10	Establish databases for monitoring system using existing JLab servers.	✓	This will be better accomplished by saving the time-series DB to files specific to a workflow and serving them with a temporary container.
M11	Integrate Hydra as monitoring component.	■	Work done to containerize GlueX online monitoring in order to allow full test with Hydra. The Hydra is nearly complete containerizing Hydra (for off project purposes) which we will use.
M12	Integrate off-line data analysis framework into platform for CLAS12 data	■	Stand alone example exists, but general framework not ready to implement.
M13	Integrate off-line data analysis framework into platform for ePIC or GlueX simulated data	■	Stand alone example exists, but general framework not ready to implement.
M14	Integrate example JANA2 analysis into platform	■	Stand alone example exists, but general framework not ready to implement.

Table 1: FY24 Milestones

4 Budget

Figure 2 shows the project spending as of the end of FY24Q4. This is considered on track with the project expectations.



M:\budget\FY2024\FY24_Level 2\CST\Monthly Reports\FY24 WBS 1.10 CST Spending Report Master Worksheet - TS, 7/3/2024

Figure 2: LD2410 Project Spending through FY24Q2. Values are in \$K.

5 Concerns

Personnel changes and some unanticipated R&D roadblocks slowed development. Proposal presented for year 2 renewal adjusted milestones slightly to accommodate positioning us to develop more momentum by end of FY25Q1.

Acknowledgements

The research described in this report was conducted under the Laboratory Directed Research and Development Program at Thomas Jefferson National Accelerator Facility for the U.S. Department of Energy.

Appendix: Full Project Milestones (revised for year 2)

- **Y1Q1**

- M01: Create prototype ERSAP configurations for INDRA and CLAS12 test systems
- M02: Identify or capture SRO formatted data from CLAS12 and INDRA test systems with data tag/filtering capability (output data ready for further offline processing)
- M03: Evaluate existing solutions for configuring and launching remote distributed processes
- M04: Establish code repository(s), project site, and method of documentation

- **Y1Q2**

- M05: Create stream splitter program for EVIO or HIPO data formatted files
- M06: Create stream splitter program for simulated data in PODIO for ePIC
- M07: Create VTP emulator using files produced by stream splitter
- M08: Create controller program to synchronize multiple VTP emulators

- **Y1Q3**

- M09: Determine appropriate schema for all aspects of monitoring system.
- M10: Establish databases for monitoring system using existing JLab servers.
- M11: Integrate Hydra as monitoring component.

- **Y1Q4**

- M12: Integrate off-line data analysis framework into platform for CLAS12 data
- M13: Integrate off-line data analysis framework into platform for ePIC or GlueX simulated data
- M14: Integrate example JANA2 analysis into platform

- **Y2Q1**

- M15: Establish general framework for RTDP simulation
- M16: Create configurable CPU proxy component
- M17: Create configurable GPU proxy component (hardware and software)
- M18: Create configurable FPGA proxy component (hardware and software)

- **Y2Q2**

- M19: Create functioning hardware GPU component (e.g., CLAS12 L3)
- M20: Create functioning hardware FPGA component (e.g., ML4FPGA)
- M21: Configure simulation of full SRO system using existing JLab hardware resources

- **Y2Q3**

- M22: Establish working test of system that transfers ≥ 100 Gbps from CH to compute center
- M23: Establish working test of system that includes GPU component for portion of the stream
- M24: Establish working test of system that includes FPGA component for portion of the stream
- M25: Test system with remote compute facility (e.g., BNL or NERSC) at limits of available resources

- **Y2Q4**

- M26: Configure system that results in stream(s) being received by JLab from external source
- M27: Collaborate with HPDF group to evaluate processing SRO data at JLab for external experiments
- M28: Complete documentation for platform to be used by non-experts