RTDP
SRO Real-Time Development Platform

EPSCI
Experimental Physics Software and Computing Infrastructure

Network Serialization for EDM4eic

or

Streaming PODIO ROOT files to EICrecon

David Lawrence

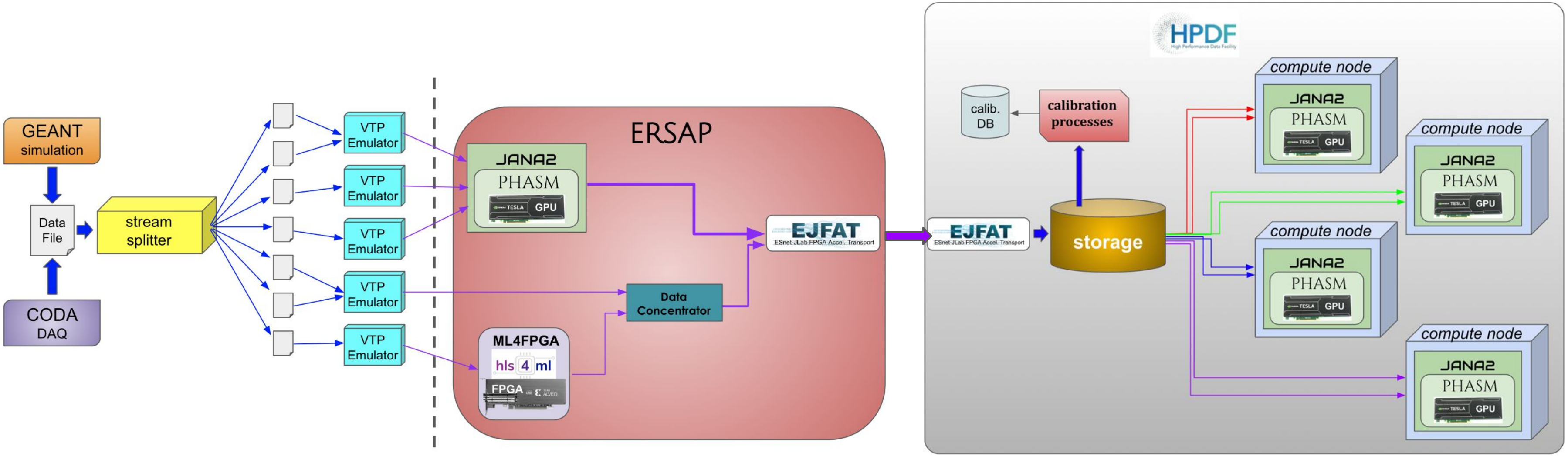
ePIC Reconstruction WG Meeting
June 18, 2024

Project funded by JLab LDRD (LD2410)

RTDP
SRO Real-Time Development Platform

Simple Example of a Streaming Readout (SRO) System

EPSCI
Experimental Physics Software and Computing Infrastructure



Highly configurable multi-stream source allows realistic streaming simulations

Onsite components will implement first stages of data filtering/reduction

Offsite processing must incorporate built-in calibration latencies and storage. This will also help inform HPDF design

HOSS!

EPJ Web of Conferences **251**, 04005 (2021)
https://doi.org/10.1051/epjconf/202125104005

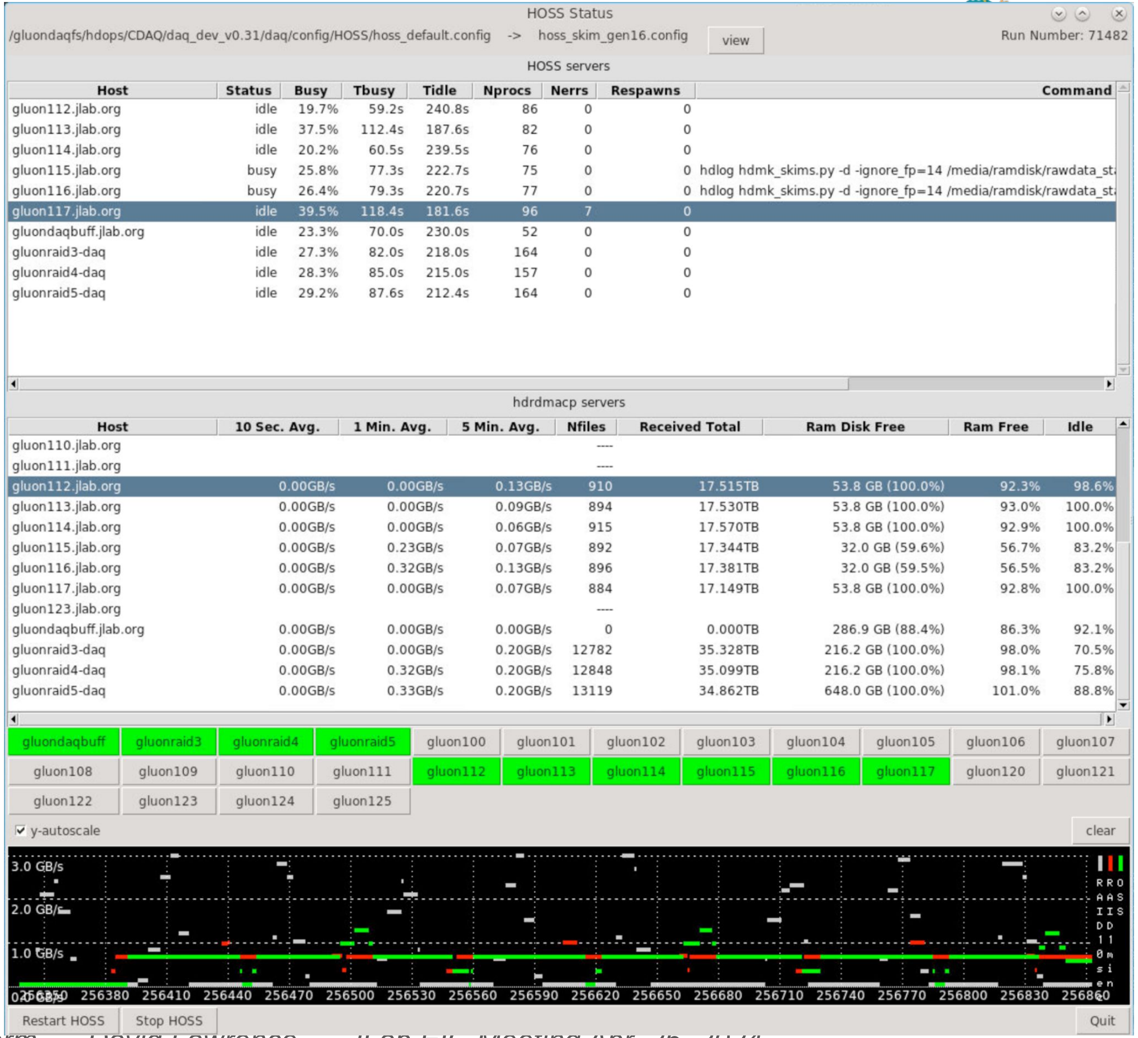
HOSS = Hall-D Online Skim System

This is responsible for distributing raw data to multiple RAID disks

DAQ system writes to files in RAM disk.

Files transferred using RDMA only after file is closed

Fine time structure naturally introduced wherever buffering is implemented



Host	Status	Busy	Tbusy	Tidle	Nprocs	Nerrs	Respawns	Command
gluon112.jlab.org	idle	19.7%	59.2s	240.8s	86	0	0	
gluon113.jlab.org	idle	37.5%	112.4s	187.6s	82	0	0	
gluon114.jlab.org	idle	20.2%	60.5s	239.5s	76	0	0	
gluon115.jlab.org	busy	25.8%	77.3s	222.7s	75	0	0	hdlog hdmk_skims.py -d -ignore_fp=14 /media/ramdisk/rawdata_st
gluon116.jlab.org	busy	26.4%	79.3s	220.7s	77	0	0	hdlog hdmk_skims.py -d -ignore_fp=14 /media/ramdisk/rawdata_st
gluon117.jlab.org	idle	39.5%	118.4s	181.6s	96	7	0	
gluondaqbuff.jlab.org	idle	23.3%	70.0s	230.0s	52	0	0	
gluonraid3-daq	idle	27.3%	82.0s	218.0s	164	0	0	
gluonraid4-daq	idle	28.3%	85.0s	215.0s	157	0	0	
gluonraid5-daq	idle	29.2%	87.6s	212.4s	164	0	0	

Host	10 Sec. Avg.	1 Min. Avg.	5 Min. Avg.	Nfiles	Received Total	Ram Disk Free	Ram Free	Idle
gluon110.jlab.org				----				
gluon111.jlab.org				----				
gluon112.jlab.org	0.00GB/s	0.00GB/s	0.13GB/s	910	17.515TB	53.8 GB (100.0%)	92.3%	98.6%
gluon113.jlab.org	0.00GB/s	0.00GB/s	0.09GB/s	894	17.530TB	53.8 GB (100.0%)	93.0%	100.0%
gluon114.jlab.org	0.00GB/s	0.00GB/s	0.06GB/s	915	17.570TB	53.8 GB (100.0%)	92.9%	100.0%
gluon115.jlab.org	0.00GB/s	0.23GB/s	0.07GB/s	892	17.344TB	32.0 GB (59.6%)	56.7%	83.2%
gluon116.jlab.org	0.00GB/s	0.32GB/s	0.13GB/s	896	17.381TB	32.0 GB (59.5%)	56.5%	83.2%
gluon117.jlab.org	0.00GB/s	0.00GB/s	0.07GB/s	884	17.149TB	53.8 GB (100.0%)	92.8%	100.0%
gluon123.jlab.org				----				
gluondaqbuff.jlab.org	0.00GB/s	0.00GB/s	0.00GB/s	0	0.000TB	286.9 GB (88.4%)	86.3%	92.1%
gluonraid3-daq	0.00GB/s	0.00GB/s	0.20GB/s	12782	35.328TB	216.2 GB (100.0%)	98.0%	70.5%
gluonraid4-daq	0.00GB/s	0.32GB/s	0.20GB/s	12848	35.099TB	216.2 GB (100.0%)	98.1%	75.8%
gluonraid5-daq	0.00GB/s	0.33GB/s	0.20GB/s	13119	34.862TB	648.0 GB (100.0%)	101.0%	88.8%

Streaming Data

Perception

Reality

RTDP

SRO Real-Time Development Platform

LDRD2410 milestones

EPSCI

Experimental Physics Software and Computing Infrastructure

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**: Create configurable CPU proxy component
- **M16**: Create configurable GPU proxy component (hardware and software)
- **M17**: Create configurable FPGA proxy component (hardware and software)
- **M18**: Create functioning hardware GPU component (e.g. CLAS12 L3)
- **M19**: Create functioning hardware FPGA component (e.g ML4FPGA)

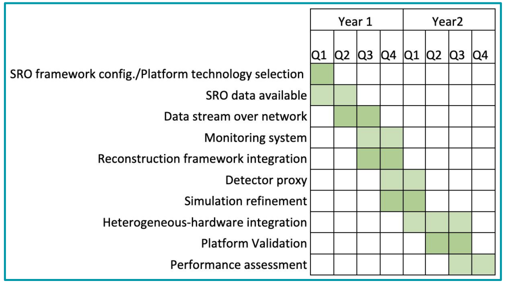
Y2Q2

- **M20**: Impose artificial time structure on stream sources to mimic beam-like conditions
- **M21**: Configure simulation of full SRO system using existing JLab hardware resources

Y2Q3

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

Milestones and Schedule



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

RTDP
SRO Real-Time Development Platform

PODIO ROOT Streaming

EPSCI
Experimental Physics Software and Computing Infrastructure

*Disclaimer: This is not an efficient way to stream data and will almost certainly **not** be part of the final ePIC standard streaming configuration. It **does** provide a continuous streaming source that can be consumed by the current ePIC reconstruction software allowing other components of the streaming system to be sketched out.*

Multiple pieces

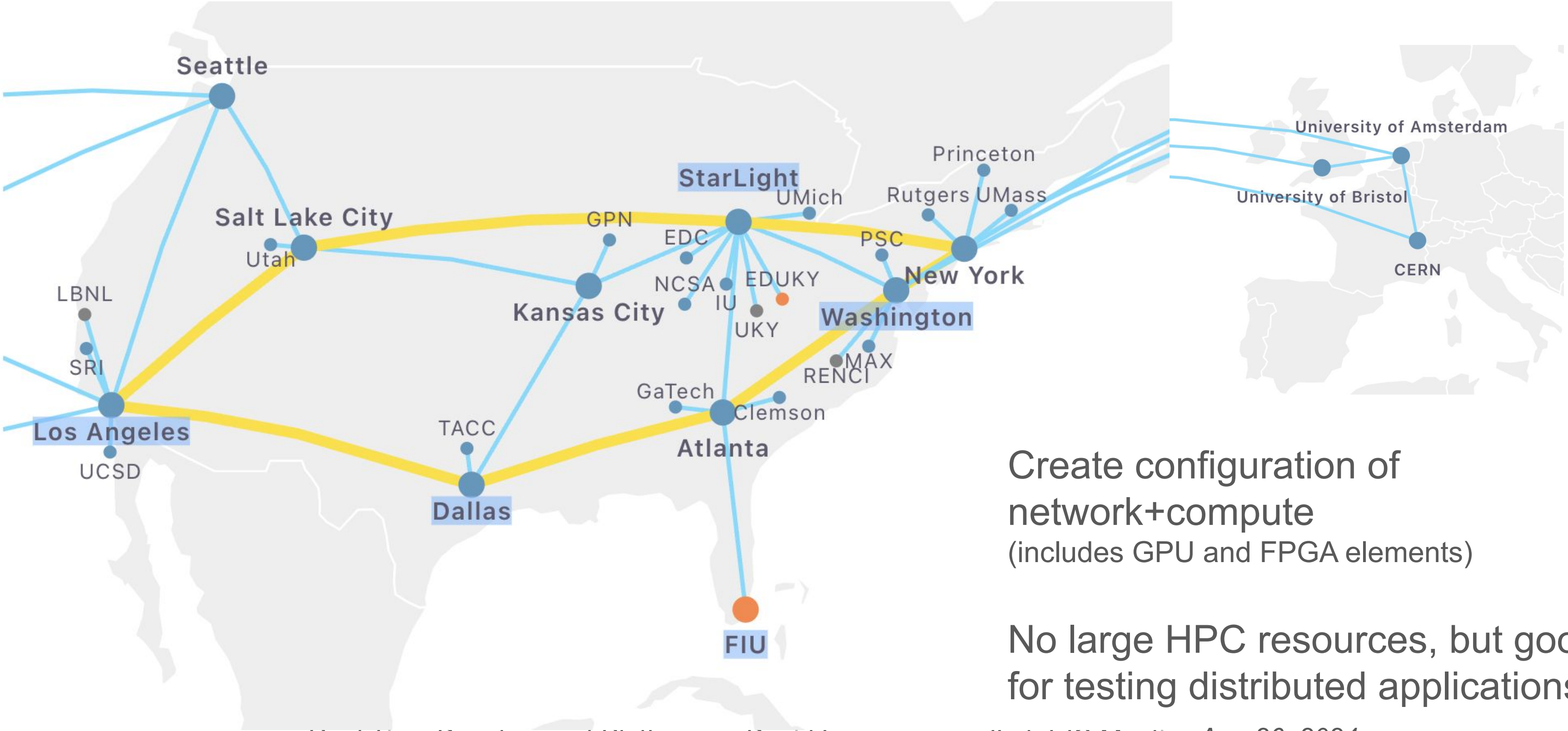
1. Modified PODIO *(allows data to come in form other than ROOT file on disk)*
   Added `openTDirectory()` to ROOTReader as alternative to `openFiles()`
   https://github.com/AIDASoft/podio/issues/565
   https://github.com/AIDASoft/podio/pull/579

2. podio2tcp utility *(read from ROOT file and send data over network)*
   Use zmq PUSH-PULL to do automatic load balancing
   https://github.com/JeffersonLab/SRO-RTDP/tree/main/src/utilities/cpp/podio2tcp

3. podiostream JANA2 event source plugin
   Reads events from network and provides them to EICrecon

FABRIC Testbed

https://portal.fabric-testbed.net/



Create configuration of network+compute
(includes GPU and FPGA elements)

No large HPC resources, but good for testing distributed applications

RTDP

SRO Real-Time Development Platform

FABRIC test

Tested sending data from CERN to 8 locations in US.

Four EICrecon processes running at each location

32 consumers total

(event processing rate is few seconds per event)

Input file: SIDIS 10x100

/work/eic2/EPIC/EVGEN/SIDIS/pythia6-eic/1.0.0/10x100/q2_0to1/pythia_ep_noradcor_10x100_q2_0.000000001_1.0_run48.ab.hepmc3.tree.root

Name	EICreconTCPmulti
Lease Expiration (UTC)	2024-04-15 11:54:13 +0000
Lease Start (UTC)	2024-04-14 11:54:14 +0000
Project ID	a7818636-1fa1-4e77-bb03-d171598b0862
State	StableOK

ID	Name	Cores	RAM	Disk	Image	Image Type	Host	Site
Nodes								
19bdb4d0-619a-4abe-8d98-6ad35a986b43	server_CERN	4	16	100	docker_rocky_8	qcow2	cern-w1.fabric-testbed.net	CERN
abc1026b-fc73-4ed7-9759-e37c29ded40b	worker_ATLA	4	16	100	docker_rocky_8	qcow2	atla-w1.fabric-testbed.net	ATLA
eeb8ae21-e7c7-4cc1-8940-6cbe0fe9a3ba	worker_DALL	4	16	100	docker_rocky_8	qcow2	dall-w1.fabric-testbed.net	DALL
ac0936ca-b44e-4388-aa58-0a4c775234b9	worker_KANS	4	16	100	docker_rocky_8	qcow2	kans-w1.fabric-testbed.net	KANS
710cdd4d-1c2a-4c6d-9126-3aadf1ecc076	worker_LOSA	4	16	100	docker_rocky_8	qcow2	losa-w2.fabric-testbed.net	LOSA
1b3e6ea2-1f39-467e-9bf9-cd756356a0f7	worker_NEWY	4	16	100	docker_rocky_8	qcow2	newy-w1.fabric-testbed.net	NEWY
709bf2f5-e684-4671-816e-042b50c94751	worker_SALT	4	16	100	docker_rocky_8	qcow2	salt-w2.fabric-testbed.net	SALT
ff915fef-a6ee-466e-9917-8e53a158b6e8	worker_SEAT	4	16	100	docker_rocky_8	qcow2	seat-w2.fabric-testbed.net	SEAT
99cf0a22-0da4-4c9b-a6b3-dc0ede261950	worker_WASH	4	16	100	docker_rocky_8	qcow2	wash-w1.fabric-testbed.net	WASH

EPSCI

Experimental Physics Software and Computing Infrastructure

RTDP
SRO Real-Time Development Platform

Summary and Outlook

EPSCI
Experimental Physics Software and Computing Infrastructure

- The RTDP (Real Time Development Platform) project seeks to create a software tool for developing and testing streaming readout (SRO) systems
- Components will be configured and connected to emulate systems, allowing for real components to be modularly swapped with emulators
- Code developed and tested for sending PODIO (EDM4eic) over the network to EICrecon. (PR still pending)
- Additional streaming configurations being developed that will serve as test cases for RTDP development. LDRD project completion scheduled for Sept. 2026

Backup Slides

RTDP
SRO Real-Time Development Platform

EPSCI
Experimental Physics Software and Computing Infrastructure

RTDP poster from ACAT2024

RTDP: Streaming Readout Real-Time Development and Testing Platform

Authors: Ayan Roy, David Lawrence, Jeng-Yuan Tsai, Marco Battaglieri, Markus Diefenthaler, Vardan Gyurjyan, Xinxin (Cissie) Mei

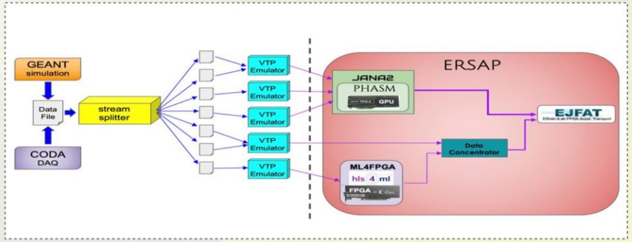
MOTIVATION

Experimental Nuclear Physics is moving towards a Streaming Readout (SRO) paradigm

Complex pipelines integrating heterogeneous hardware and varied software may have interference effects

Simulation and testing of complex SRO systems is needed to assist in their design and validation

Testing of complete, integrated SRO systems at scale for future experiments requires new tooling

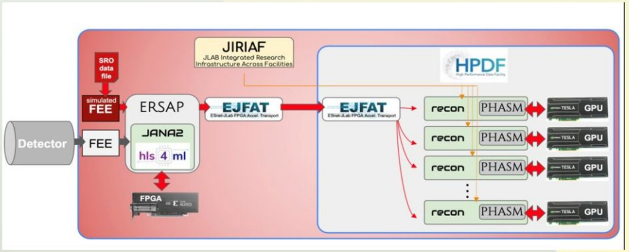


OBJECTIVE

- Deployment of a distributed (quasi) real-time SRO data processing model includes data calibration and full traditional off-line reconstruction.
- Framework optimization using GEANT-generated and archived beam-on data.
- Optimized framework validation with beam-on tests.
- Assessment of needed network and computing resources.
- Assessment of the performance for different hardware platforms.
- Identify potential issues relevant to a future HPDF in receiving and processing SRO data.

APPLICATION

- SRO Experiments requiring intricate configurations can be defined with user-friendly YAML
- Individual components such as calibration or data transport can be represented by software simulation modules
- Full simulation can include mixture of real and simulated components
- Scale from fully simulated on single PC to full use of hardware in distributed system



MEASURE OF SUCCESS

Specific milestones and objectives of the project include:

- Ability to launch synchronized processes across multiple nodes
- Integrated monitoring of all components in the system
- Ability to configure and simulate an experiment similar in size to the planned SoLID experiment at JLab
- Test with 400Gbps transfer speed, at least one FPGA and at least 1 GPU component

GOAL

- Create a platform to seamlessly process data from SRO to analysis on compute centers in various configurations
- Fully developed software platform that is capable of monitoring the components in a fully developed streaming system.
- Tools for fully simulating a real-time SRO data processing network from Front End Electronics to large compute.

ACKNOWLEDGEMENT

This project is funded through the Thomas Jefferson National Accelerator Facility LDRD program. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics under contract DE-AC05-06OR23177.

PROGRESS

RTDP is at the early stages of development. Here are some out of many things we have worked on:

- Captured CLAS12 data, streamed across the Jlab campus using a 100Gbps high-speed NIC featuring hardware timestamps.
- Captured data using synchronized streams from multiple network sources.

FUTURE WORKS

- Create stream splitter program for EVIO or HIPO data formatted files
- Create stream splitter program for simulated data in PODIO for ePIC
- Create VTP emulator using files produced by stream splitter
- Integrate Hydra as monitoring component.

Jefferson Lab   EPSCI Experimental Physics Software and Computing Infrastructure   U.S. DEPARTMENT OF ENERGY   JSA

RTDP

SRO Real-Time Development Platform

LDRD2410 milestones

EPSCI

Experimental Physics Software and Computing Infrastructure

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**: Create configurable CPU proxy component
- **M16**: Create configurable GPU proxy component (hardware and software)
- **M17**: Create configurable FPGA proxy component (hardware and software)
- **M18**: Create functioning hardware GPU component (e.g. CLAS12 L3)
- **M19**: Create functioning hardware FPGA component (e.g ML4FPGA)

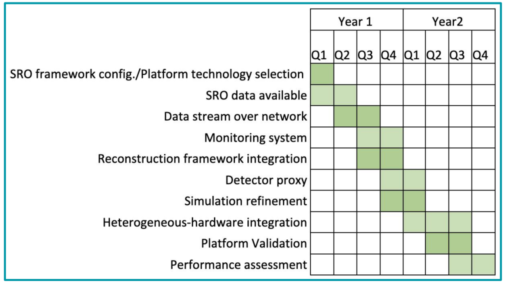
Y2Q2

- **M20**: Impose artificial time structure on stream sources to mimic beam-like conditions
- **M21**: Configure simulation of full SRO system using existing JLab hardware resources

Y2Q3

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

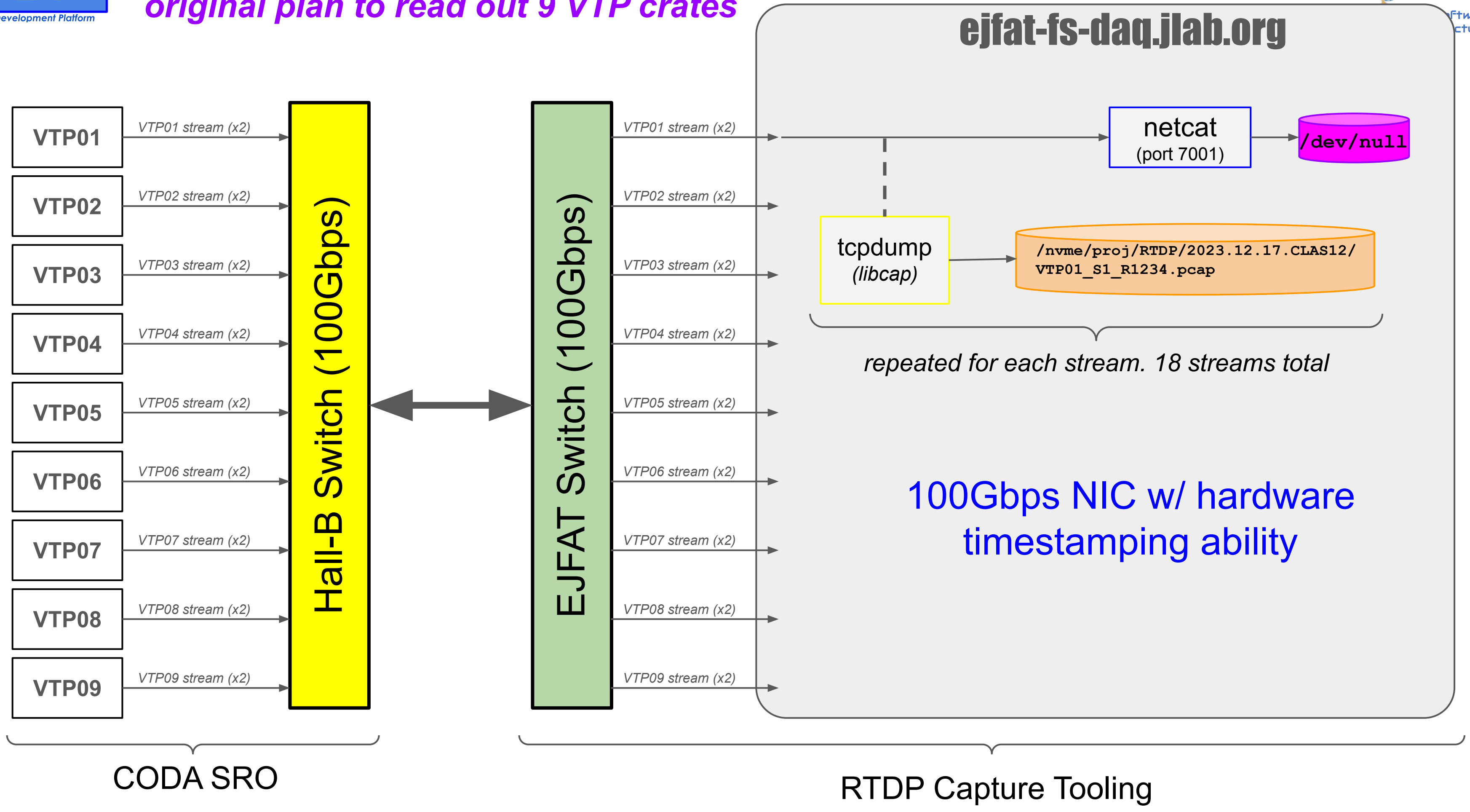
Milestones and Schedule



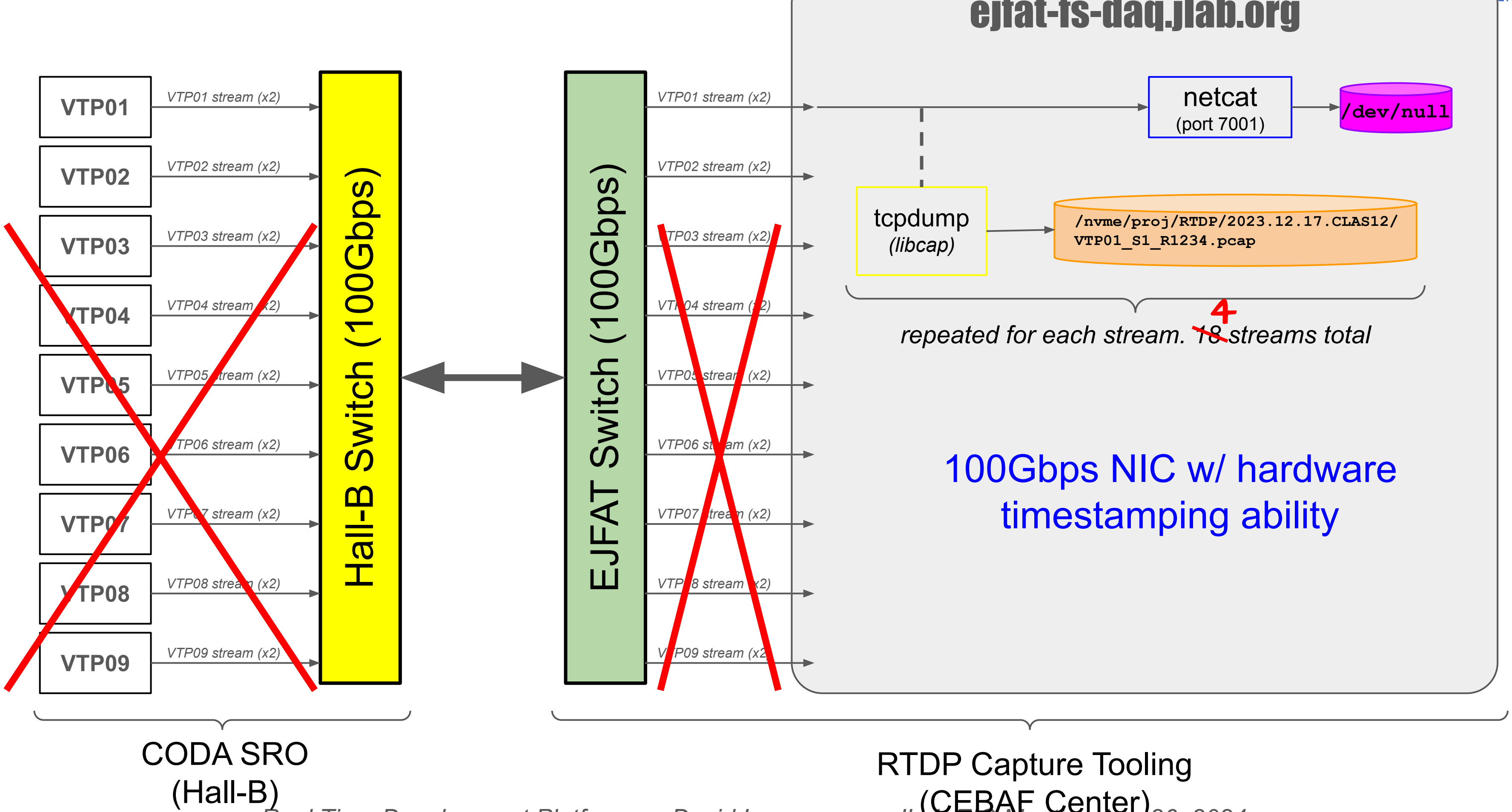
Y2Q4

- **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

original plan to read out 9 VTP crates



revised plan read out 2 VTP crates



Multiple beam currents captured in 10 minute runs

Run	Start time	Beam Current	filename	File Size
176	2023-12-17 8:40:49	90nA	CLAS12_ECAL_PCAL_S2_2023-12-17_08-37-33.pcap	11GB
177	2023-12-17 8:54:25	10nA	CLAS12_ECAL_PCAL_S2_2023-12-17_08-51-31.pcap	6.4GB
178	2023-12-17 9:12:23	100nA	CLAS12_ECAL_PCAL_S2_2023-12-17_09-08-04.pcap	16.2GB
179	2023-12-17 9:32:13	50nA	CLAS12_ECAL_PCAL_S2_2023-12-17_09-29-15.pcap	9.5GB
180	2023-12-17 9:48:41	150nA	CLAS12_ECAL_PCAL_S2_2023-12-17_09-44-43.pcap	14.1GB
181	2023-12-17 10:03:07	75nA	CLAS12_ECAL_PCAL_S2_2023-12-17_10-00-51.pcap	12.0GB
182	2023-12-17 10:20:42	25nA	CLAS12_ECAL_PCAL_S2_2023-12-17_10-18-11.pcap	7.2GB

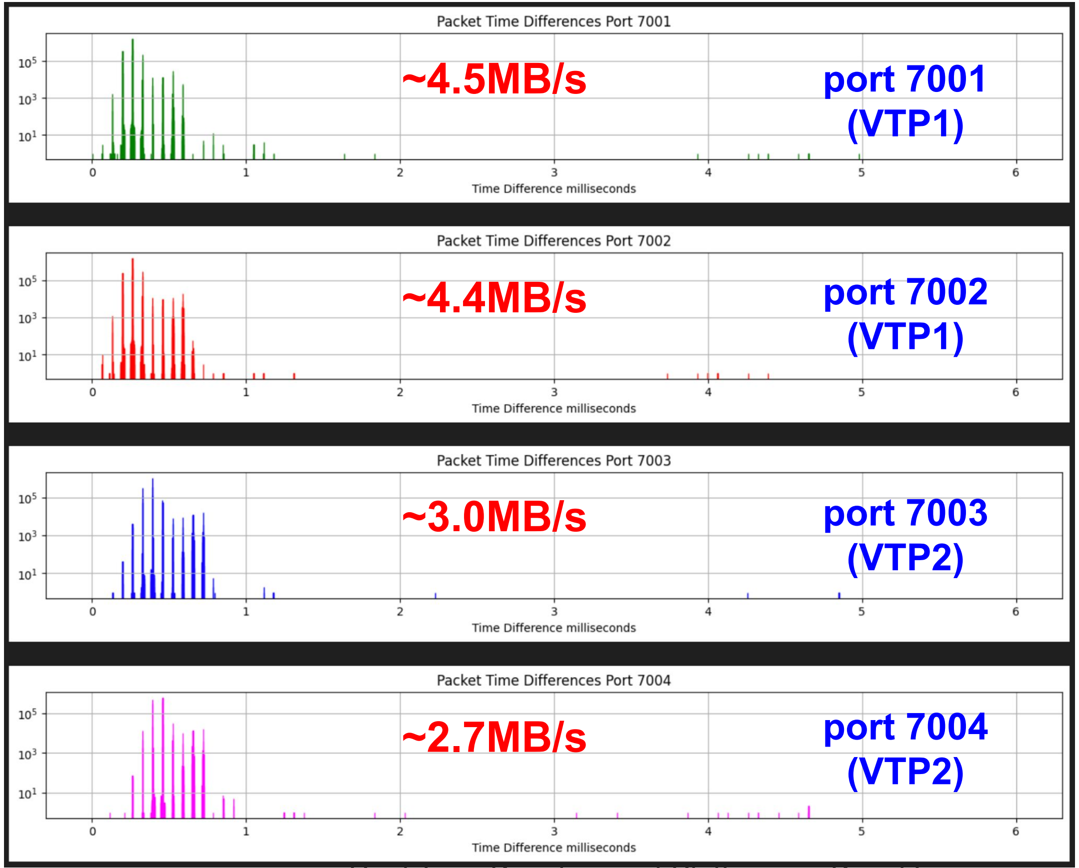
Copied to tape: **`/mss/epsci/RTDP/2023.12.17.CLAS12`**

Time difference between packets by port

RTDP
SRO Real-Time Development Platform

EPSCI
Experimental Physics Software and Computing Infrastructure

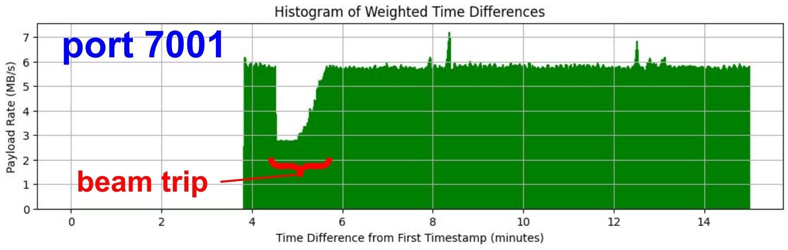
100nA

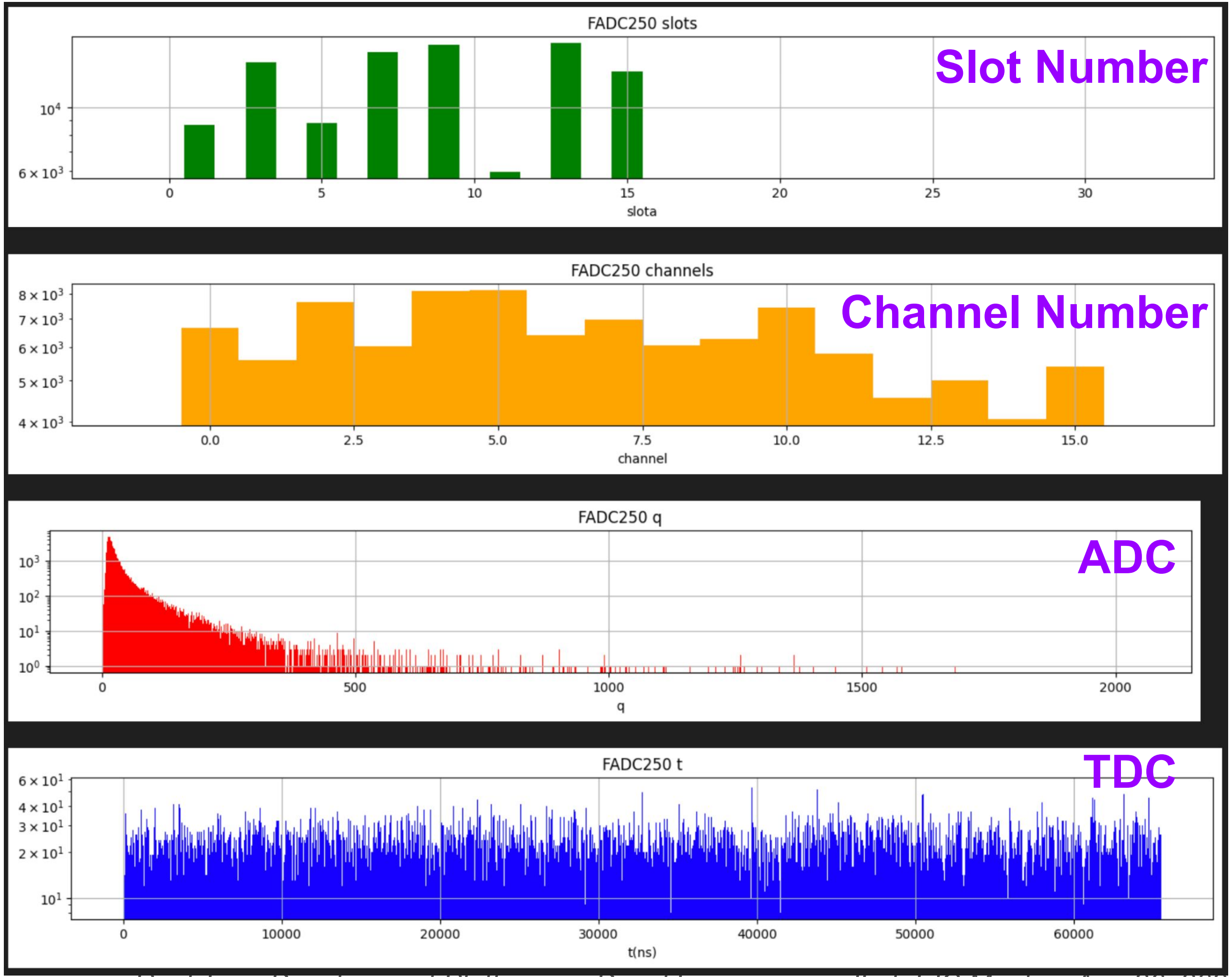


Standard 1.5kB MTU used for TCP packets

Packets w/ time structure metadata captured into industry standard .pcap files

Crates report every 64μs, even if no hits present for that crate





TDC timing is relative to frame start time

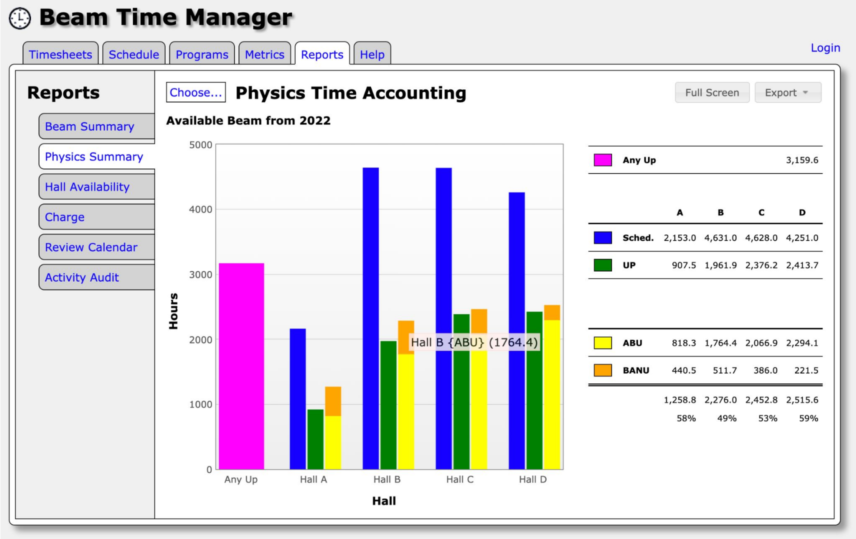
SRO Benefit to Future Experiments

RTDP SRO Real-Time Development Platform

EPSCI Experimental Physics Software and Computing Infrastructure

Hall-B has ~1640 ABU/year
10% increase in recorded beam -> additional 164 hours/year
$10k/hr * 164 = $1.64M/year *n.b. not accounting for multi-hall operation*

1764 hours in CY22



1517 hours in CY23