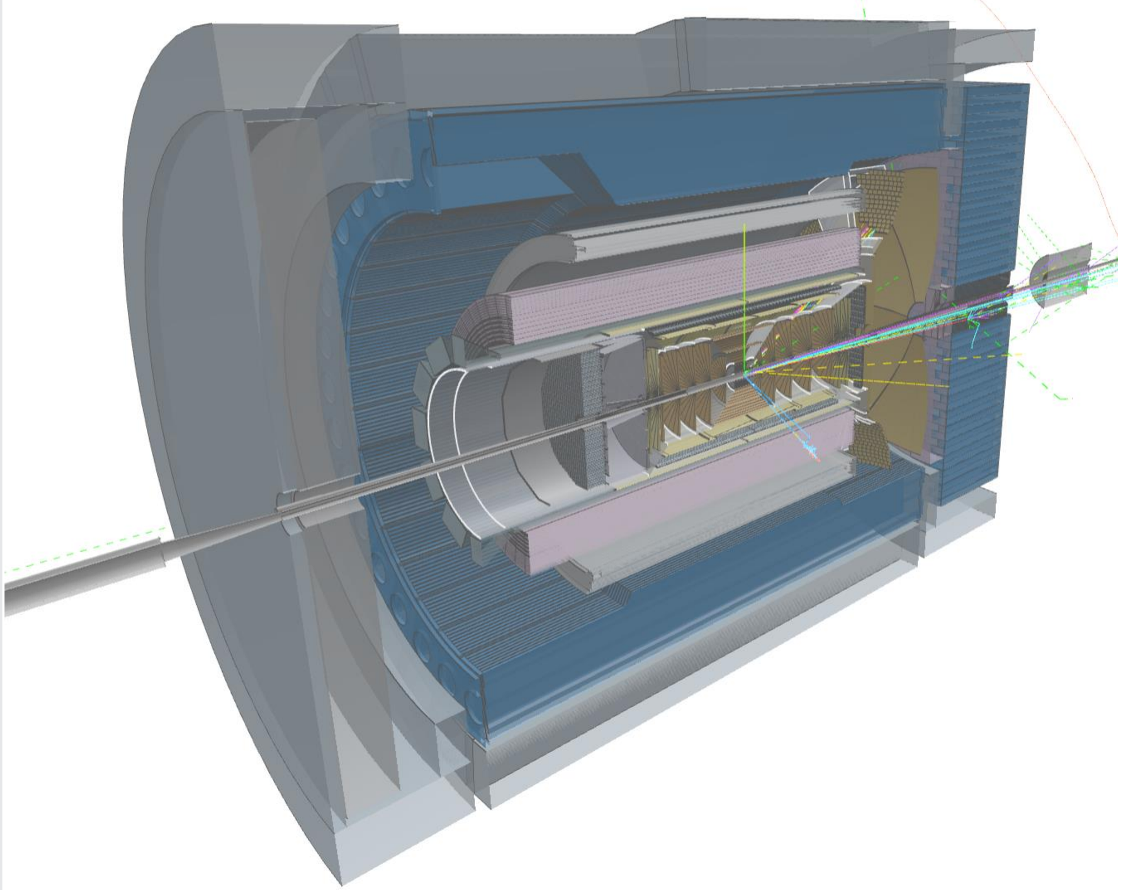
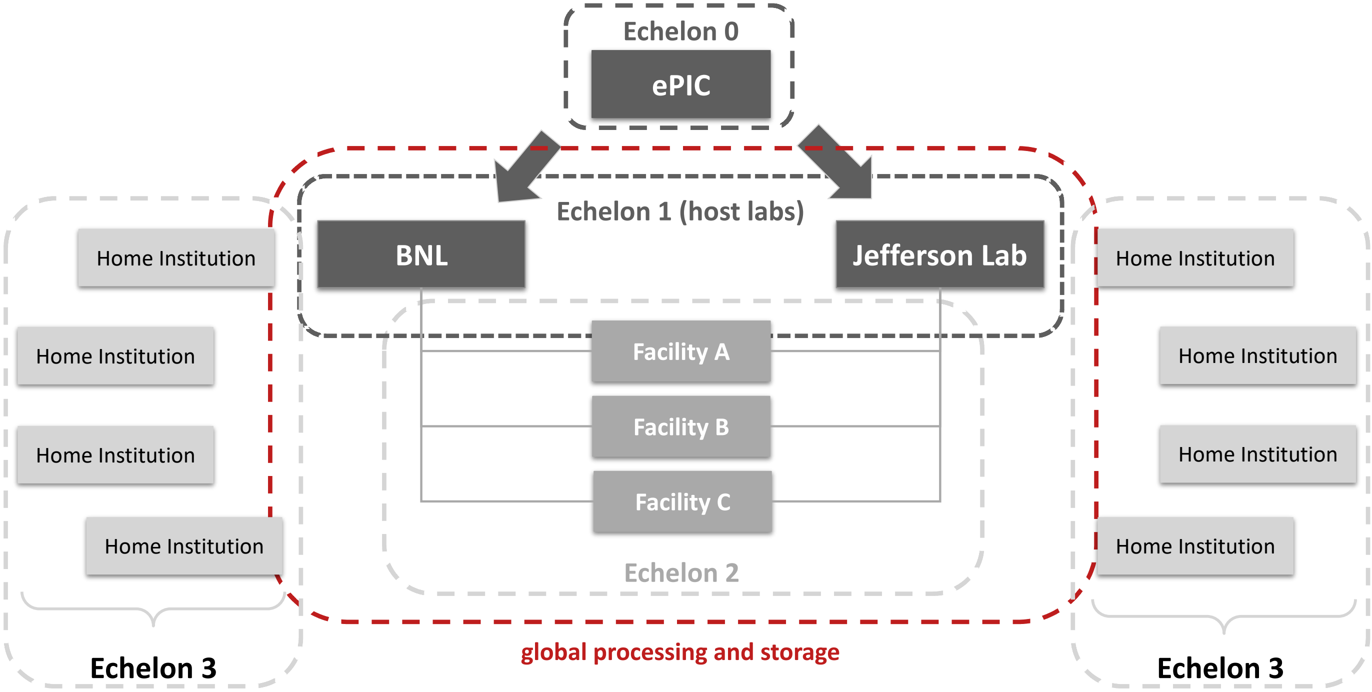
ePIC Software & Computing Report





Markus Diefenthaler (Jefferson Lab)

ePIC Software & Computing Organization

Coordinators and WG Conveners

Development

Dmitry Kalinkin

Simulation WG: Chao Peng, Sakib Rahman. **Reconstruction WG**: Derek Anderson, Chandra Chatterjee (nominated), Shujie Li.

Infrastructure

Torre Wenaus

Streaming Computing WG: Marco Battaglieri, Taku Gunji, Jeff Landgraf.

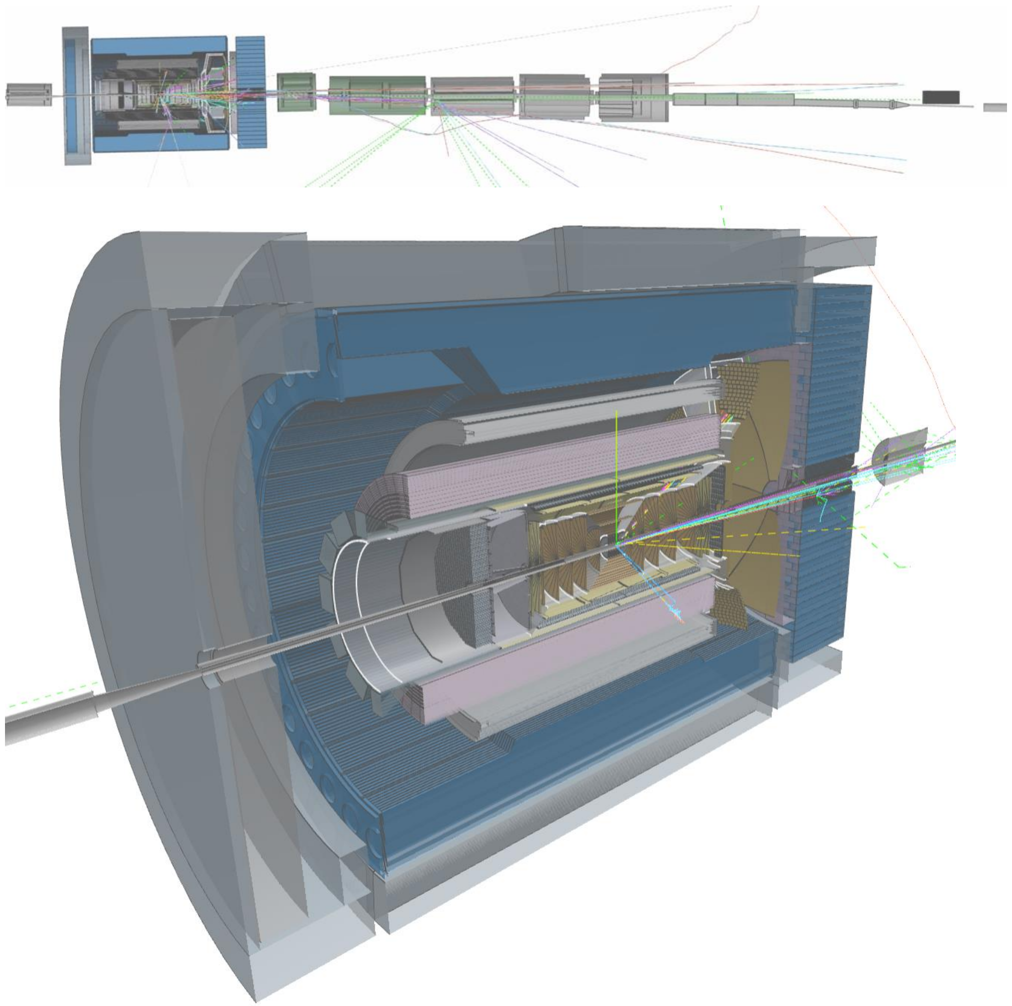
Operations

Wouter Deconinck

Production WG: Thomas Britton, Sakib Rahman. **User Learning WG**: Stephen Kay, Holly Szumila-Vance.

“Software is the Soul of the Detector”



Captured from the ePIC Event Display

Great Software for Great Science:

- **Design and Construction:** Integrated and validated simulations are essential for evaluating detector performance and determining physics reach.
- **Operation:** Rapid processing of streamed data using streaming readout, AI, and distributed computing. Autonomous experimentation and control.
- **Physics Analysis:** Software and data enable discovery.

- We **work together**, on a global scale and with other fields, on great software for great science.
- We focus on **modern scientific software & computing practices** to ensure the **long-term success** of the **EIC scientific program**.

ECSAC Review and Current Priorities

ECSAC Review

Computing Model and Testbeds

Software and Simulations for the preTDR

EIC Computing & Software Advisory Committee (ECSAC) Review

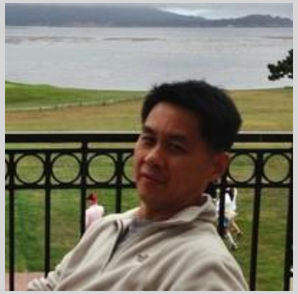
Ilya Baldin
(JLab)

Simone Campana
(CERN, chair)

Mario Cromaz
(LBNL)

Pere Mato
(PIC)

Verena Martinez
Outschoorn
(UMass Amherst)



Wei Yang
(SLAC)

Reviews occur annually, with a charge reflective of the EIC schedule, the stage of the ePIC experiment, and impending deadlines. The 2025 review was held at Jefferson Lab on October 6–7.

Charge of 2025 ECSAC Review:

- Progress toward TDR readiness and implementation of the computing model
- Short- and long-term resource planning
- ECSJI resource support and laboratory collaboration (see ECSJI Report later today)
- Engagement with the broader software and computing community
- ECSJI-ePIC collaboration effectiveness

ECSAC Review (Oct. 6–7): Highlights and Recommendations

From the ECSAC Closeout Report

"We commend the ePIC collaboration for the progress in Software & Computing. It is very positive to see the active engagement of additional partners from U.S. and international institutes. There has been substantial progress in software and computing development, notably in streaming data processing, simulation campaigns, and software readiness for the pre-TDR."

ECSAC commended ePIC's progress and made five recommendations:

1. We **recommend** the ePIC collaboration to establish a multi-year plan for SW&C, focusing on the deliverables for the next three years. Such a plan should be presented at the next ECSAC meeting in Fall 2026.
2. We **recommend** the ePIC collaboration and the host labs to start developing a data management and lifecycle plan, agreed with the Funding Agencies.
3. We **recommend** the ePIC collaboration to continue the commissioning plan through the ongoing demonstrators and testbeds. Such a commissioning process should include assessing the E0-E1 interfaces and workflows.
4. We **recommend** the ePIC to continue engaging the institutes (including the host labs) to identify possible additional contributions for SW&C developments and operations. We expect to hear about the level of engagement of the institutes and the success in addressing resource gaps at the next ECSAC meeting.
5. We **recommend** implementing the EICO as described during the review. We suggest a staged approach focusing initially on the aspects more related to the collaboration (Collaboration/Overview boards) and the links with WLCG.

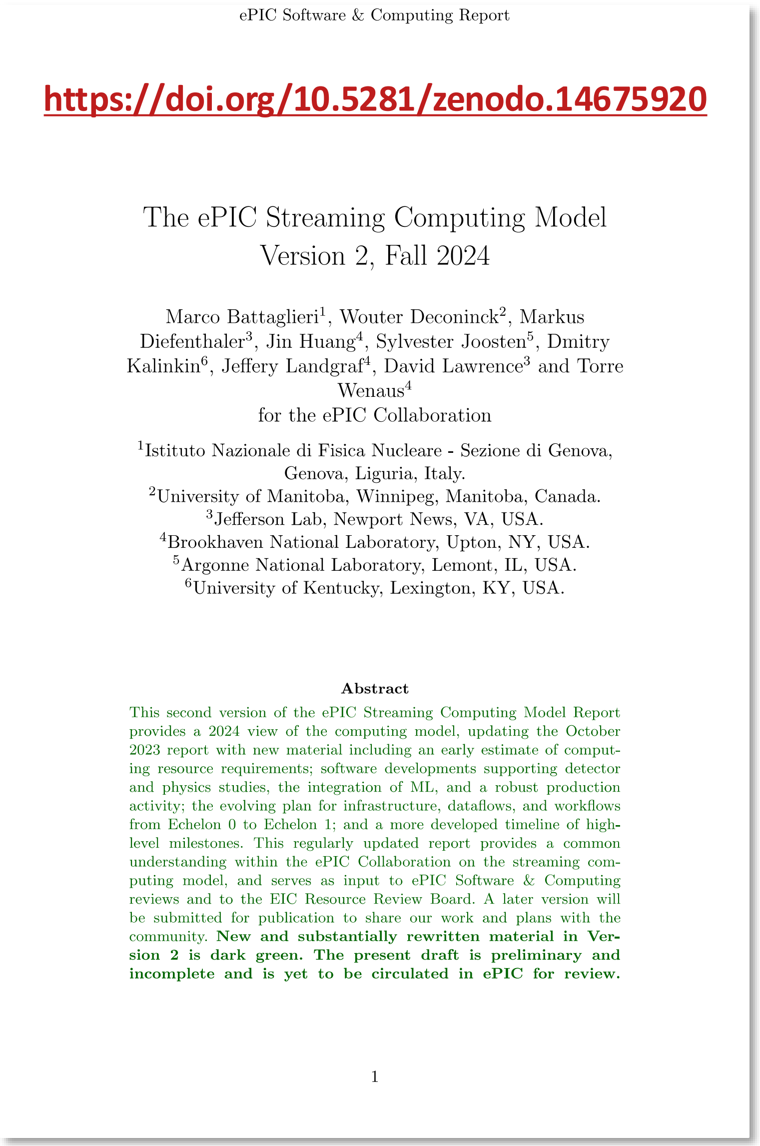
ECSAC Review and Current Priorities

ECSAC Review

Computing Model and Testbeds

Software and Simulations for the preTDR

The ePIC Streaming Computing Model


ePIC Software & Computing Report

https://doi.org/10.5281/zenodo.14675920

The ePIC Streaming Computing Model
Version 2, Fall 2024

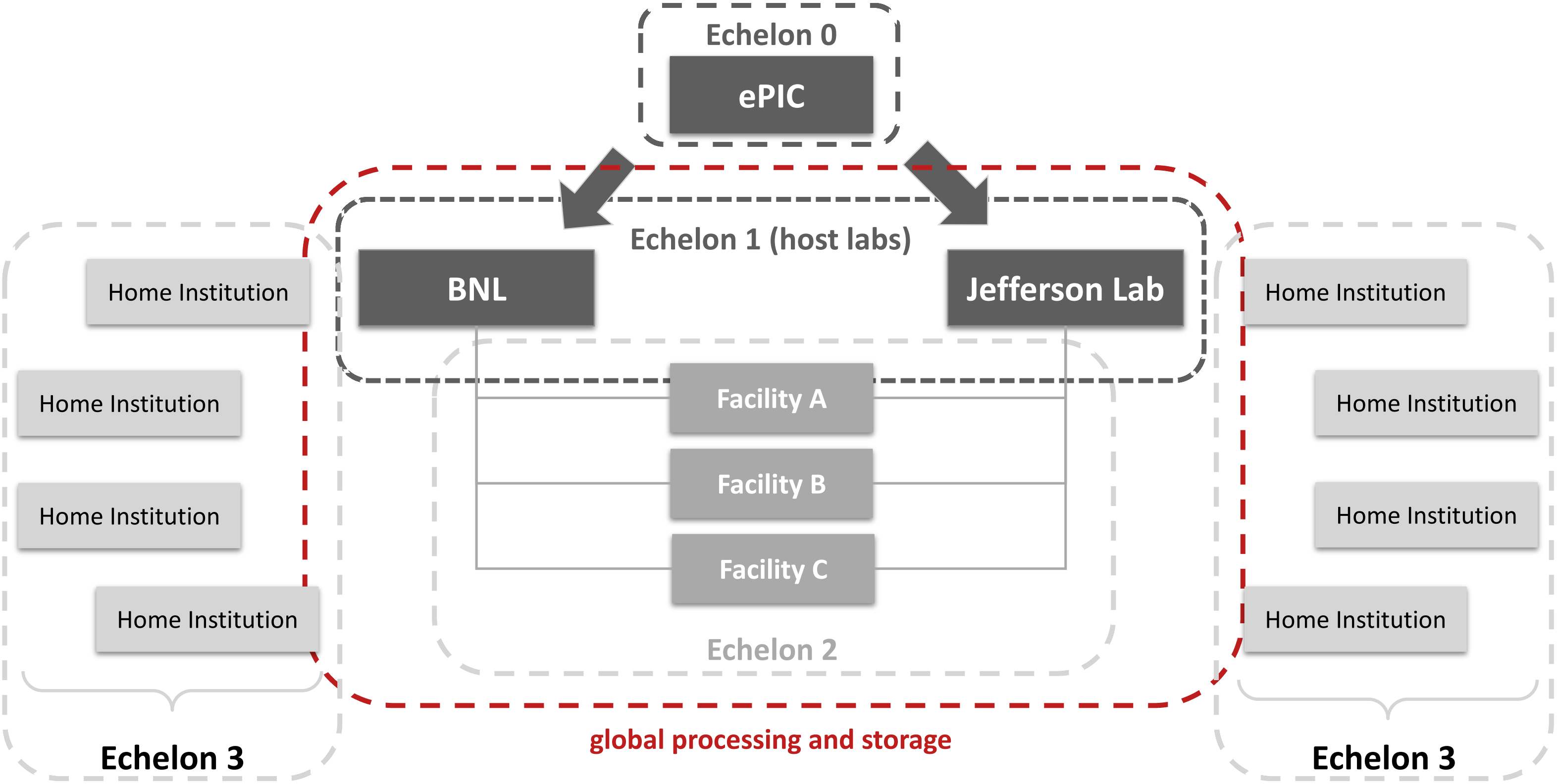
Marco Battaglieri[1], Wouter Deconinck[2], Markus Diefenthaler[3], Jin Huang[4], Sylvester Joosten[5], Dmitry Kalinkin[6], Jeffery Landgraf[4], David Lawrence[3] and Torre Wenaus[4]
for the ePIC Collaboration

[1]Istituto Nazionale di Fisica Nucleare - Sezione di Genova, Genova, Liguria, Italy.
[2]University of Manitoba, Winnipeg, Manitoba, Canada.
[3]Jefferson Lab, Newport News, VA, USA.
[4]Brookhaven National Laboratory, Upton, NY, USA.
[5]Argonne National Laboratory, Lemont, IL, USA.
[6]University of Kentucky, Lexington, KY, USA.

Abstract

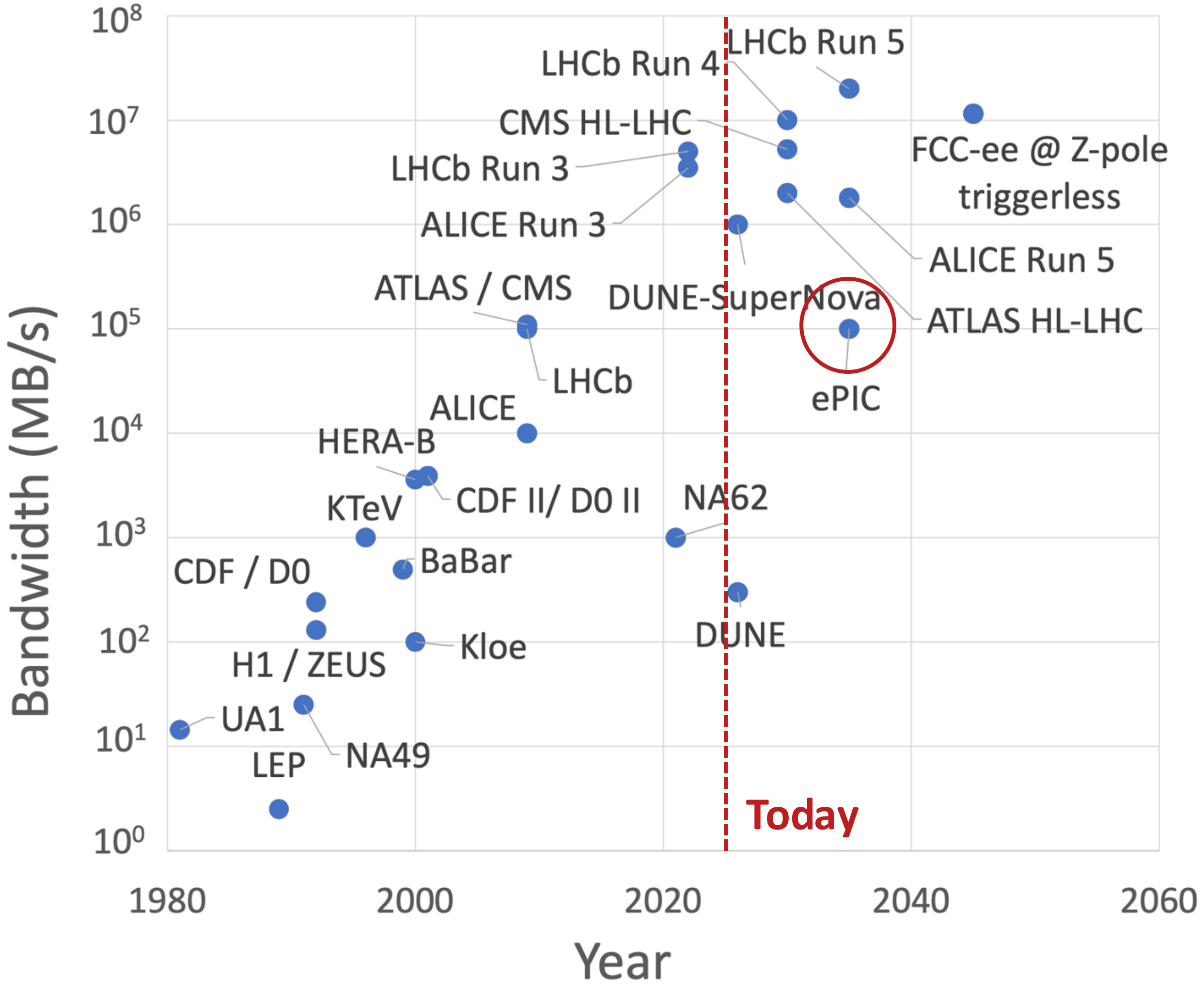
This second version of the ePIC Streaming Computing Model Report provides a 2024 view of the computing model, updating the October 2023 report with new material including an early estimate of computing resource requirements; software developments supporting detector and physics studies, the integration of ML, and a robust production activity; the evolving plan for infrastructure, dataflows, and workflows from Echelon 0 to Echelon 1; and a more developed timeline of high-level milestones. This regularly updated report provides a common understanding within the ePIC Collaboration on the streaming computing model, and serves as input to ePIC Software & Computing reviews and to the EIC Resource Review Board. A later version will be submitted for publication to share our work and plans with the community. **New and substantially rewritten material in Version 2 is dark green. The present draft is preliminary and incomplete and is yet to be circulated in ePIC for review.**

1



We developed the ePIC Streaming Computing Model to accelerate the pace of discovery and enhance scientific precision through improved management of systematic uncertainties. The model is documented in a detailed report and was reviewed during the 2023 and 2024 ECSAC reviews.

ePIC Within the Global Particle Physics Experiments Landscape



Streaming Readout

Data rate of up to 100 Gbit/s

after low-level data reduction in the Streaming DAQ

Aarrestad, Thea, and Dorothea vom Bruch. *Trigger and Data Acquisition: Challenges and Perspectives*. Presentation at the Open Symposium on the European Strategy for Particle Physics, Venice, Italy, June 23, 2025. https://agenda.infn.it/event/44943/contributions/265988/

Prototyping Ideas and Tools in Testbeds

With active testbeds and functional prototypes now in place, the effort is moving from design to implementation. These developments aim to define and test the interface between DAQ and computing, and to mitigate risks in the integrated DAQ-computing system.

- **Streaming orchestration**, i.e., a workflow and workload management system for streaming data—is essential for system testing. A requirements document has been developed and is now guiding testbed and prototype development.
  - **Testbed plans** are taking concrete shape:

Streaming reconstruction: Raw data stream to event identification, reconstruction, and analysis.

Streaming orchestration: Developing E0-E2 streaming workflows in the testbed, utilizing Rucio and PanDA.

Streaming processing: Developing E0-E2 streaming workflows using EJFAT.

  - Not covered in this presentation but starting efforts:

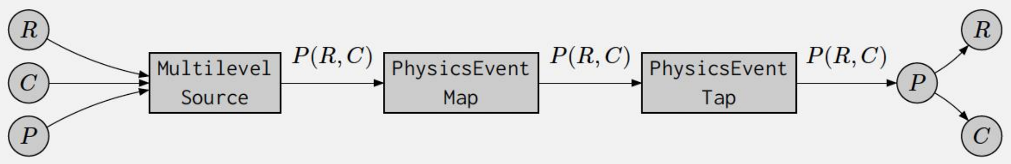
Streaming analysis: Demonstrate simulation data production streaming to E2 site.

Rapid data processing: Autonomous calibration workflow for one detector system.

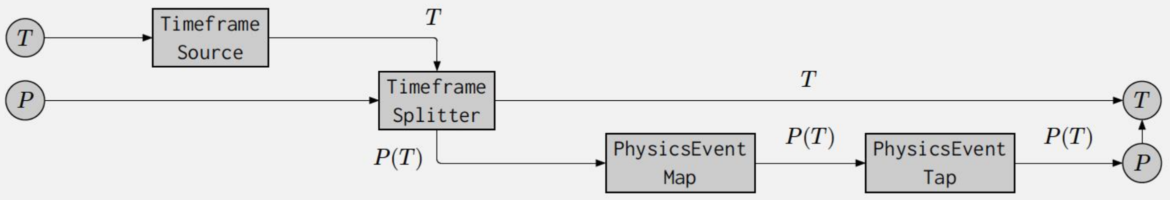
JANA2 for Streaming Processing

- Multithreaded JANA2 framework provides a component-level hierarchical decomposition of data boundaries into **Run**, **Timeframe**, **PhysicsEvent**, and **Subevent** levels. This is essential for streaming processing.
- The **Folder** and **Unfolder** component interfaces enable traversal of this hierarchy by supporting operations such as splitting and merging data streams. This functionality has been tested and validated within **EICrecon**.

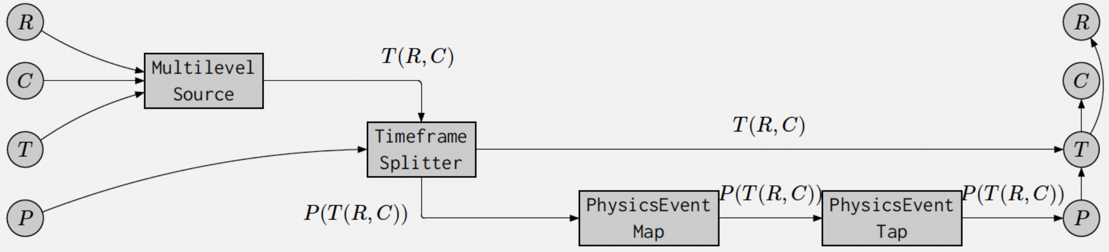
Introducing multilevel sources



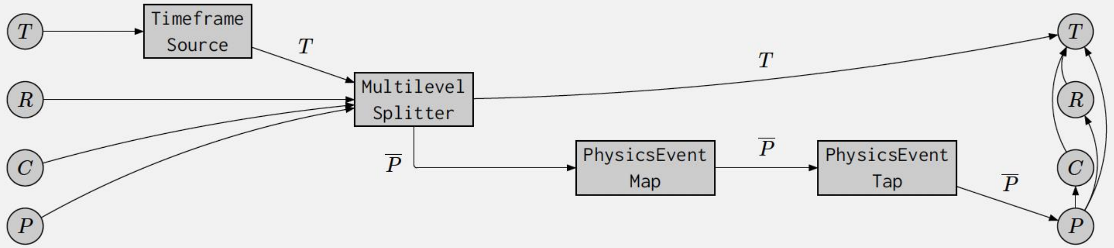
EICrecon timeframe splitting



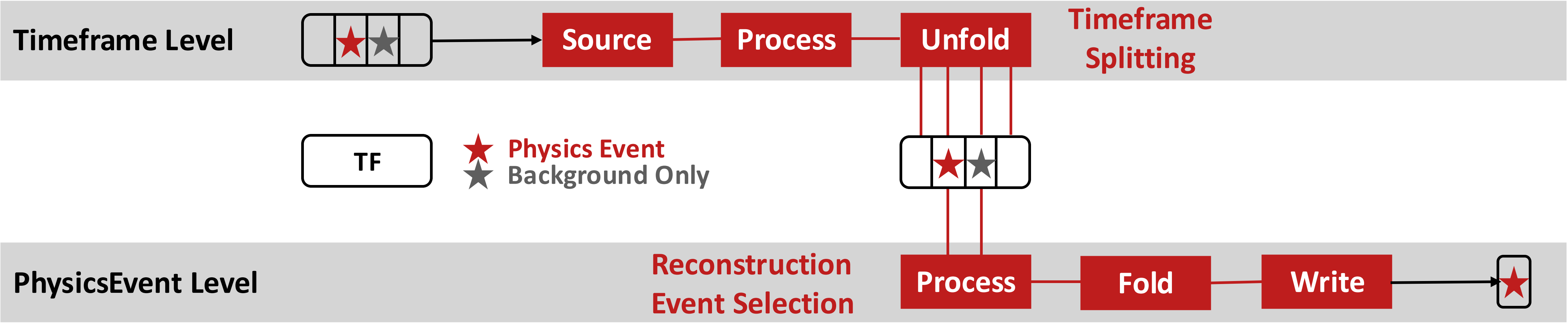
Multilevel sources with timeframe splitting

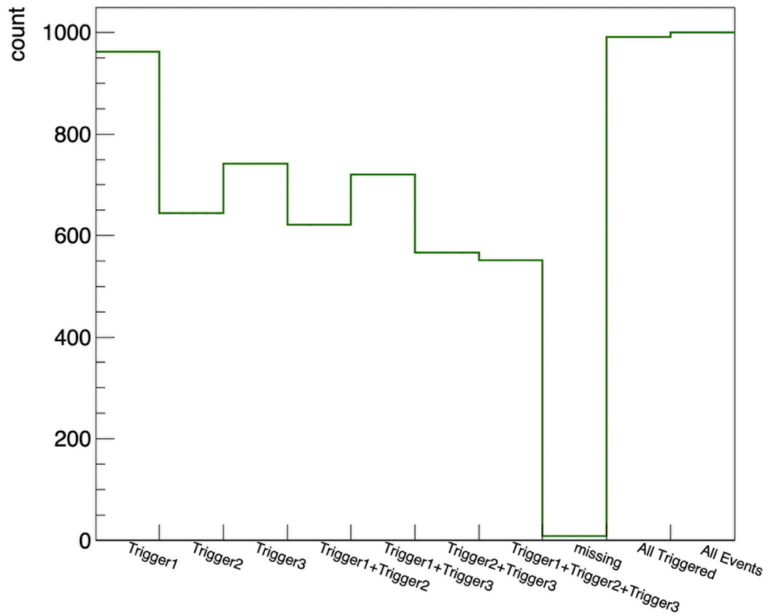


Timeframe sources with multilevel splitting



Streaming Reconstruction Prototype: Event-Building in JANA2





Trigger 1: Coincidence hits in the SVT Endcap and Forward MPGD Endcap.
Trigger 2: Coincidence hits in the TOF Barrel and MPGD Barrel.
Trigger 3: Coincidence hits in the SVT Endcap and Backward MPGD Endcap.

A straightforward event selection based on raw simulated hits in the SVT, MPGD, and TOF detector systems achieves greater than 99 % efficiency and less than 1 % background in identifying physics events in TFs.

Streaming Orchestration Testbed

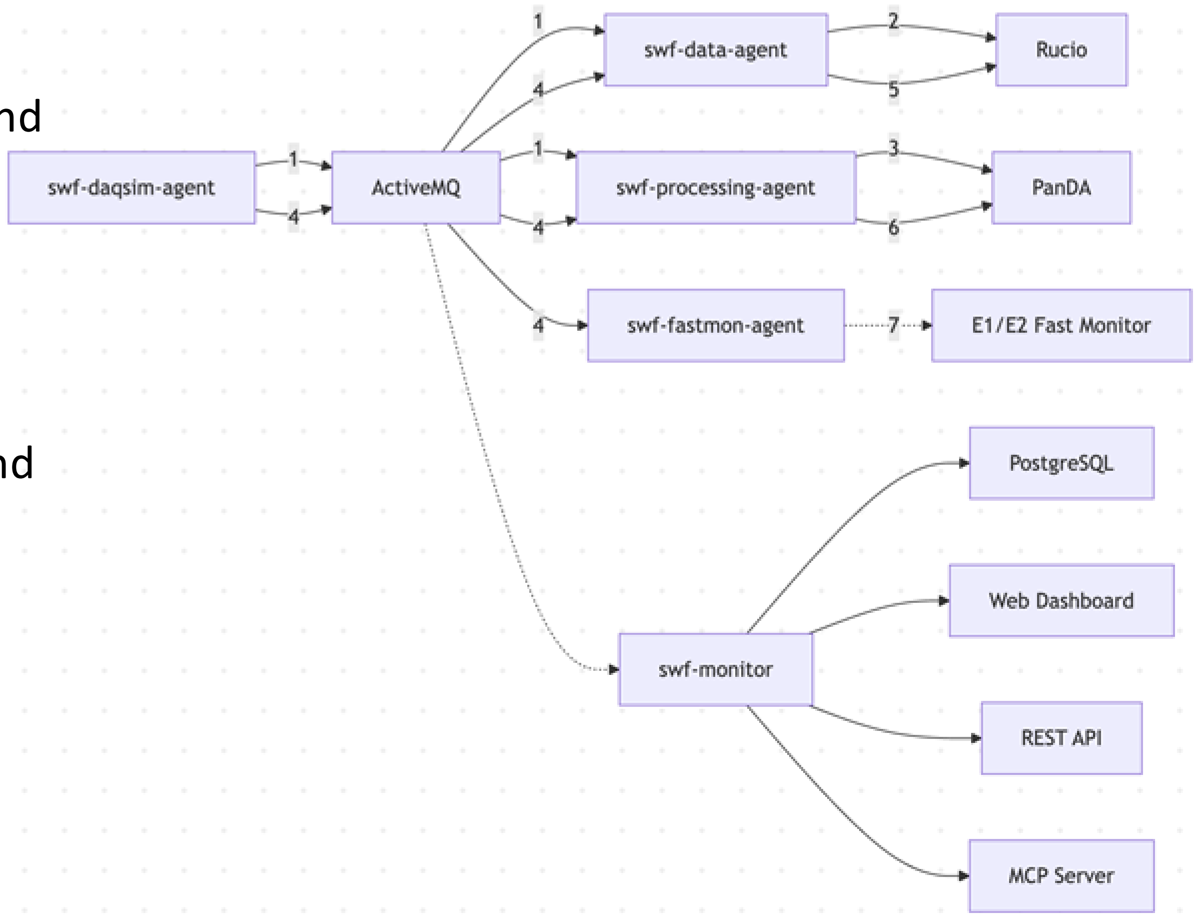
Motivation:
- Evaluate how well existing distributed computing tools support streaming orchestration.
- Focus on practical deployment and performance in realistic environments.

- **Design Precepts**:
  - Robust geographical distribution across real-world networks.
  - Full automation of data processing workflows.
  - Complete exposure of system status and operational analytics.
- **Approach**:
  - PanDA and Rucio align with the stated design precepts and are deployed in live testbed instances at BNL.
  - Assume that data is delivered in STFs, each consisting of 1000 aggregated TFs, with a size of ~2 GB at a rate of ~1 Hz.
- **System Architecture:** Simulated DAQ with Distributed Agents and Monitoring
  - A **set of collaborating agents** communicate via ActiveMQ and form the core of the system:
    - **Data handling agent**: executes Rucio actions based on data and conditions.
    - **Processing agent**: executes PanDA actions based on available data and conditions.
    - **Monitoring agent**: skims STF data for fast feedback.
  - The **driver simulates the DAQ** and other external components to evaluate their impact.
  - A **monitoring backend** with a database aggregates and exposes system state.

End-to-End Workflow: From Run Start to Monitoring

Workflow driven by DAQ simulator and 3 agents communicating via ActiveMQ.
All system activity and state recorded in the database via REST and displayed in the monitor.

1. **Run Start** DAQ simulator broadcasts a message signaling the start of a new data-taking run (**working**).
2. **Dataset Creation** The data agent receives the message and instructs Rucio to create a dataset for the run (**working**).
3. **Processing Task** Processing agent sets up a PanDA task based on the run start message (**working**).
4. **STF Available** DAQ simulator broadcasts availability of a new STF file; this continues while the run is active.
5. **STF Transfer** The data agent triggers Rucio registration and transfers the STF to E1 storage (**not yet integrated**).
6. **STF Processing** PanDA detects the new STF at E1, transferred via Rucio, and launches jobs to process it (**working**).
7. **Fast Monitoring** The fastmon agent sees the STF broadcast, performs a partial read, and injects a data sample into the E1/E2 monitoring stream (**fast monitor emulation working based on STF metadata**).



State Machine: Stream-Oriented Workflow States and Substates

States

- no_beam
  - Collider not operating
- beam
  - Collider operating
- run
  - Physics running
- calib
  - Dedicated calibration period
- test
  - Testing, debugging
  - Any substates can be present during test

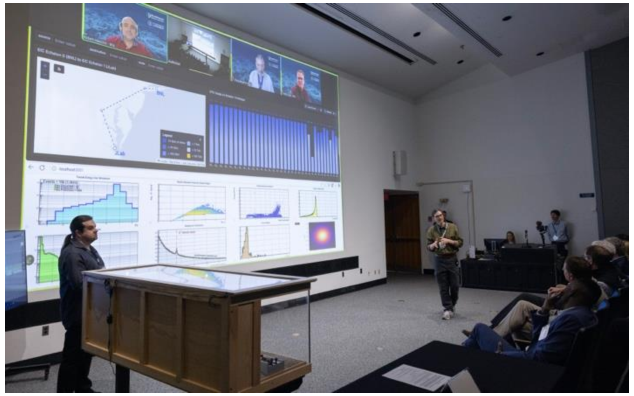
Substates

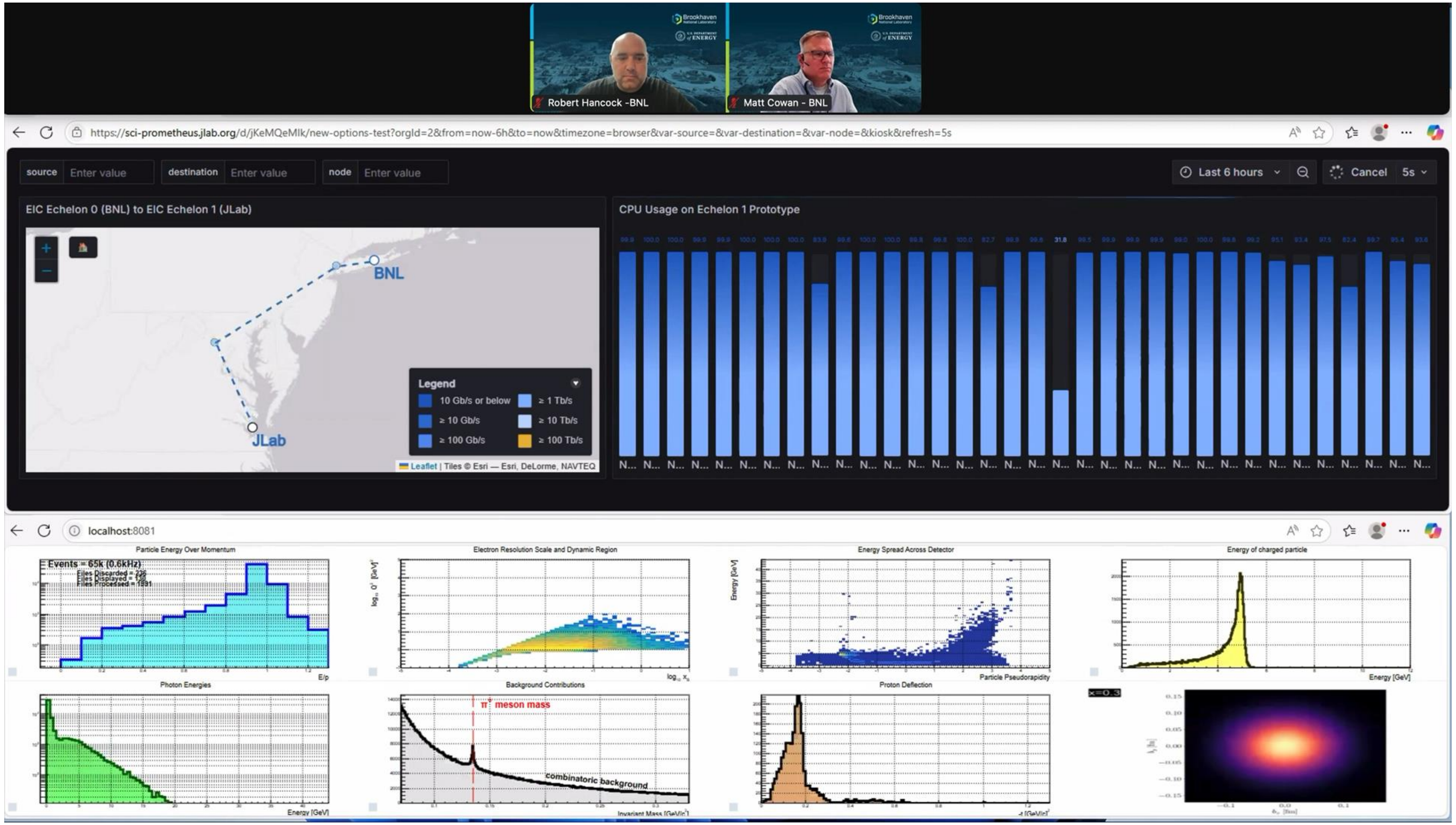
- not_ready
  - detector not ready for physics datataking
  - occurs during states: no_beam, beam, calib
- ready
  - collider and detector ready for physics, but not declared as good for physics
  - when declared good for physics, transitions from beam/ready to run/physics
  - occurs during states: beam
- physics
  - collider and detector declared good for physics
  - if collider or detector drop out of good for physics, state transitions out of 'run' to 'beam' or 'off'
  - occurs during states: run
- standby
  - collider and detector still good for physics, but standing by
  - occurs during states: run
- lumi
  - detector, machine data that is input to luminosity calculations
  - occurs during states: beam, run
- eic
  - machine data, machine configuration
  - occurs during states: all
- epic
  - detector configuration, data
  - occurs during states: all
- daq
  - info, config transmitted from DAQ
  - occurs during states: all
- calib
  - a catch-all for a great many calib data types, we can start small
  - occurs during states:

The baseline workflow exercises nearly all states, with the exception of detector and machine configurations.

Streaming Data Processing Demonstration

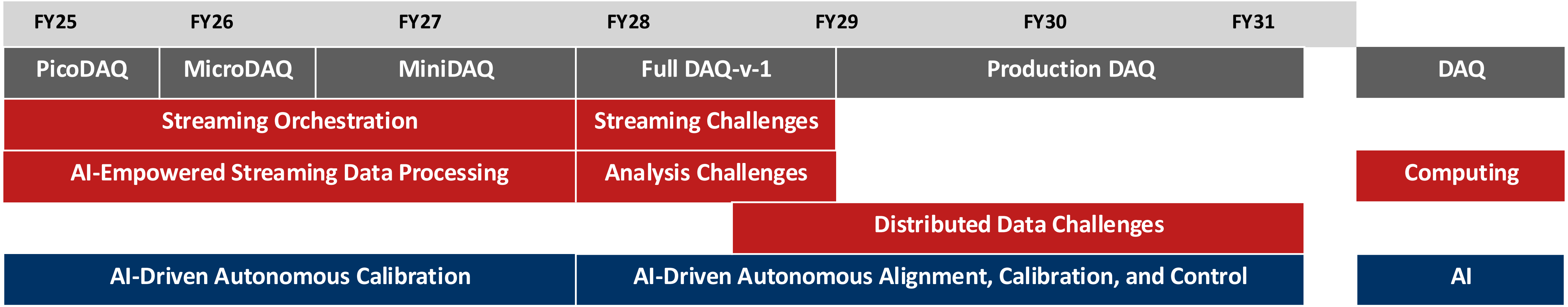
Secretary Wright's Visit to Jefferson Lab on August 21





We successfully showcased **real-time data transfer from BNL to JLab** using the ESnet-JLab FPGA Accelerated Transport (EJFAT) Load Balancer and its **processing at JLab**. It also prompted collaboration with BNL to test current tools and establish a clear network path.

Streaming DAQ and Computing Milestones



- **Compute-Detector Integration**:
  - Joint deliverables between **DAQ** and **computing** to develop integrated systems for detector readout, data processing, and ultimately physics analysis.
  - **Key role of AI(/ML)**: Empowering data processing and enabling autonomous experimentation and control.

- **FY28Q1 deliverables**:
  - fully functional testbed for streaming orchestration,
  - autonomous calibration workflow for one detector system,
  - AI/ML-empowered streaming reconstruction.

Distributed Computing for ePIC

- **Echelon 1** sites uniquely perform the **low-latency streaming workflows**:
  - Archiving and monitoring of the streaming data, prompt reconstruction and rapid diagnostics.
- Apart from low-latency, **Echelon 2** sites fully participate in use cases and **accelerate** them.
- **Priority**: Establishing EIC International Computing Organization (EICO):

ECSAC recommended a staged implementation starting with the Collaboration and Overview Boards. See ECSJI presentation for details on EICO.

Use Case	Echelon 0	Echelon 1	Echelon 2	Echelon 3
Streaming Data Storage and Monitoring	✓	✓		
Alignment and Calibration		✓	✓	
Prompt Reconstruction		✓		
First Full Reconstruction		✓	✓	
Reprocessing		✓	✓	
Simulation		✓	✓	
Physics Analysis		✓	✓	✓
AI Modeling and Digital Twin		✓	✓	

Substantial role for Echelon 2 in preliminary resource requirements model

Assumed Fraction of Use Case Done Outside Echelon 1	
Alignment and Calibration	50%
First Full Reconstruction	40%
Reprocessing	60%
Simulation	75%

Distributed Computing Model

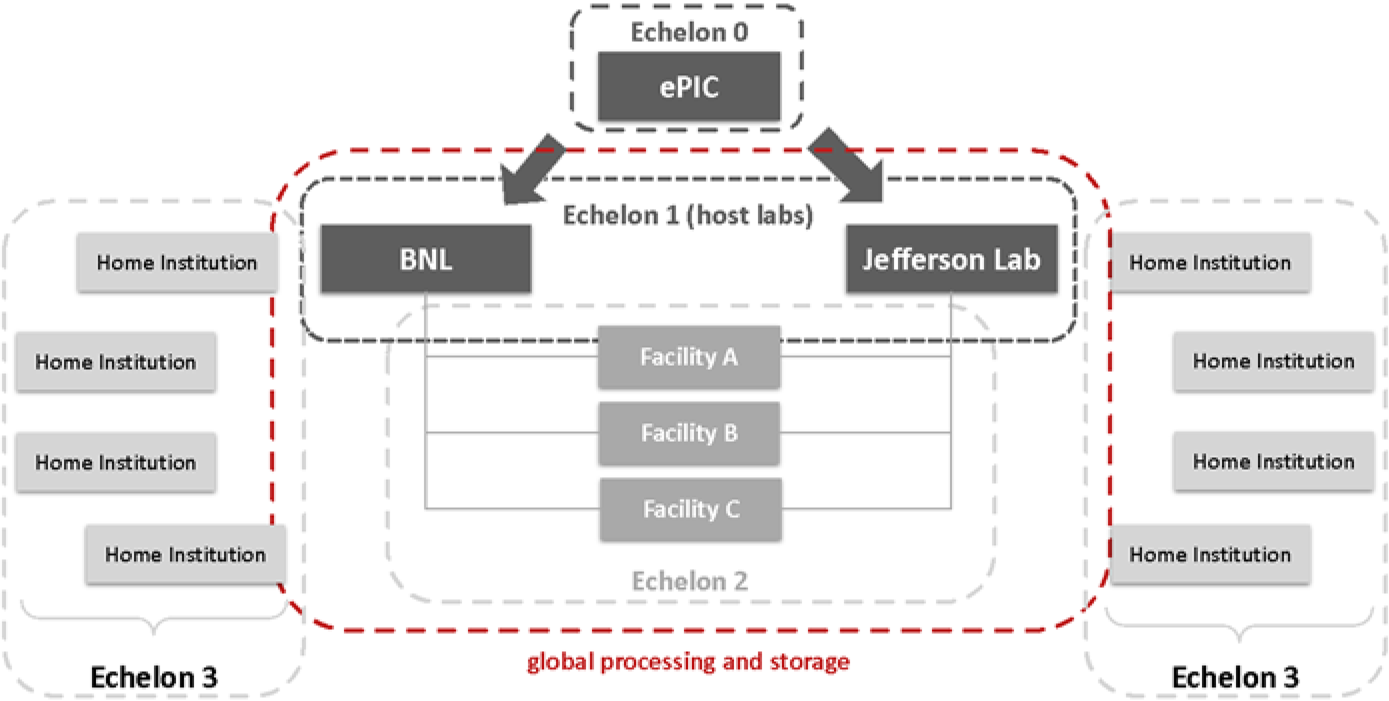
Building Echelon 2 capacity through early integration in simulation campaigns.

- Central infrastructure at **JLab**:
  - HTCondor submit node
  - Rucio main server
  - Rucio storage element
- Central infrastructure at **BNL**:
  - HTCondor submit node
  - Rucio storage element

- Active integration of **Canada** and **Italy** as compute providers integrated through OSG.

- Commissioning of **Canada** as Rucio storage element provider.
- Planning in progress for **Japan** and **Taiwan** as storage providers.

- Discussions on potential compute and storage contributions are underway with **France** and the **United Kingdom**.



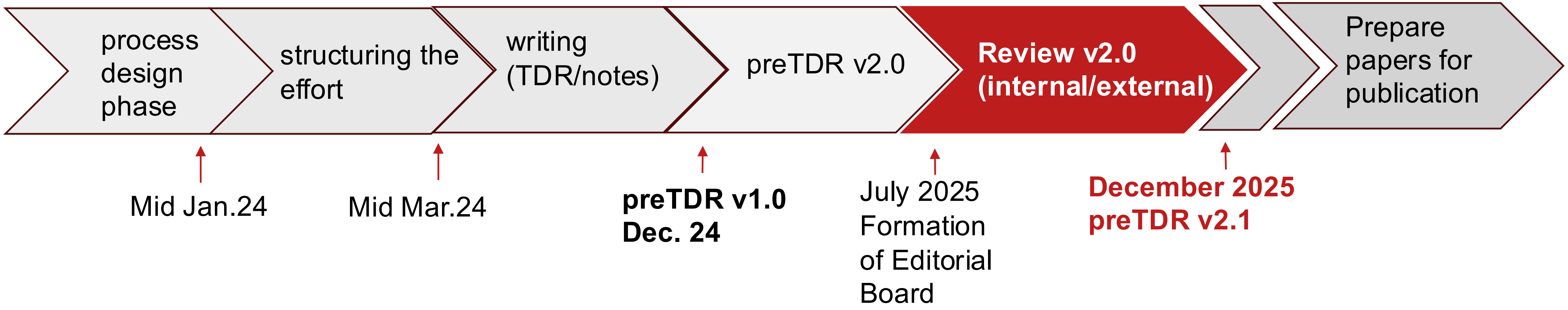
ECSAC Review and Current Priorities

ECSAC Review

Computing Model
and Testbeds

Software and
Simulations for
the preTDR

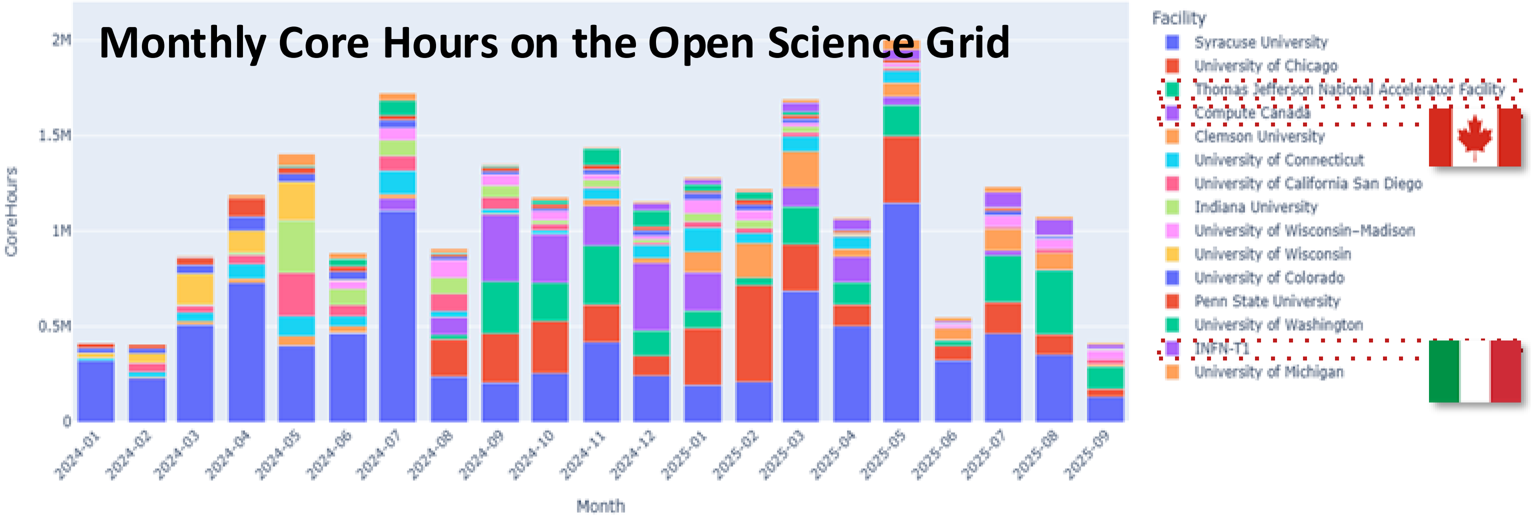
Software and Simulation Readiness for the preTDR

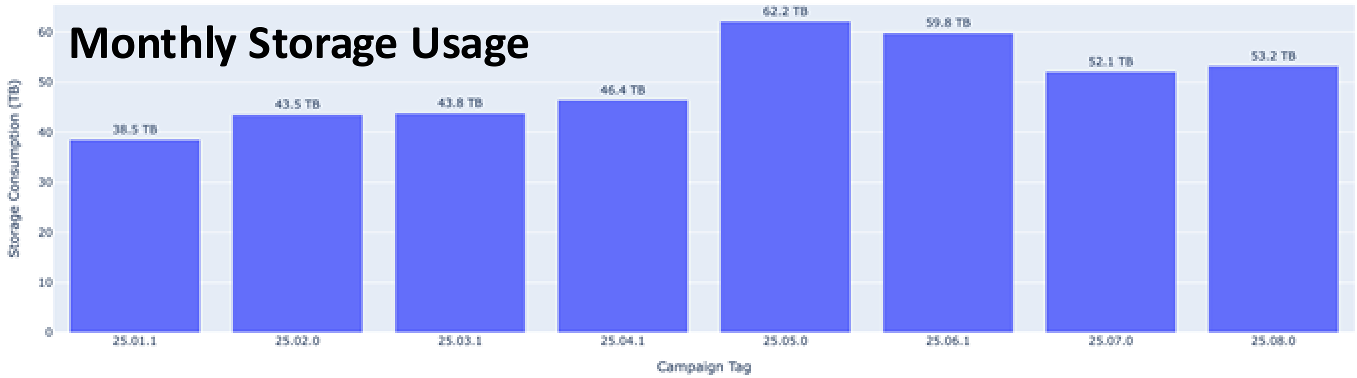


- **In 2024, we defined preTDR readiness for software and simulations with the collaboration—and successfully met those goals. In 2025, we renewed our priorities at the Frascati Collaboration Meeting, with many topics under our charge already reaching an advanced stage and positioning us well for the next phase.**
- Our advanced software and large-scale simulations serve as the backbone of both the pre-TDR efforts and the development of the Early Science Program
- Strong emphasis is placed on **coordination with Physics WGs** on simulation targets, reconstruction
- Surveys of DSCs provided insight into both ongoing activities and future plans.

Simulation Campaigns

- We provide simulation productions tailored to the needs of the collaboration, as defined by the DSCs and PWGs.
- Simulation campaigns are conducted monthly, based on the software release for the corresponding month.
- These simulations serve as the standard for detector and physics studies for the preTDR and also the Early Science Program.
- In the past year, monthly simulation campaigns consumed approximately 15 million core hours on the OSG, generating over 500 TB of simulation data. We anticipate a ~2x storage increase, and are in talks with institutions in Japan, Taiwan, Italy and UK for additional resources.





We are capable of integrating new detector geometry and algorithms within a month, processing millions of events needed to assess scientific impact.

Working Together Worldwide for Great Software and Great Science

ePIC Software & Computing is engaging the full collaboration in development

- Successful **landing page** for onboarding new collaboration members.
- **Tutorials inside the collaboration** built to onboard detector experts and analyzers into the software framework, organized by User Learning working group.
- **Continuous integration** and **deployment** of software efforts to reduce time from contributions by experts to output in nightly environment and monthly campaigns.

ePIC Software & Computing is engaging with other EIC communities

- **Tutorials on general applicability of EIC software** at CFNS/EIC/CTEQ summer schools
- **Community expansion** through workshops, including recent **HSF-India/ePIC meeting**.
- Engagement with AI4EIC workshops series, with bidirectional exchange of expertise.

ePIC Software & Computing is engaging external software projects and compute providers

- **Active engagement in community software dependencies:** ACTS, Key4HEP, etc.
- **Active engagement in compute providers in the US and abroad:** OSG, Italy, Canada, etc.
- **Participation and leadership in global NHEP software ecosystem:** HEP Software Foundation; representation at CHEP and other computing conferences and workshops.

Summary

ECSAC Review

- **ECSAC Review, commended progress** notably in streaming data processing, simulation campaigns, and software readiness for the pre-TDR.
- ECSAC recommended **multi-year planning**, a **data lifecycle strategy**, **continued testbed development**, **institute engagement to address workforce gaps**, and a **staged implementation of EICO**.

Computing Model and Testbeds

- **Streaming Computing Model** is defined, reviewed, and actively guiding ongoing developments. Its integration of computing and detector systems of ePIC experiment **maximizes scientific output** and **accelerates scientific discovery**.
- Effort **transitioning from design to implementation**, with active and functional prototypes.
- **Deliverables are aligned with EIC Project / Streaming DAQ milestones**.
- ePIC will be **compute-intensive experiment**; **substantial role for Echelon 2** foreseen.

Software and Simulations for the preTDR

- **Advanced software and large-scale simulations** serve as the backbone of the preTDR and Early Science Program efforts.
- We are **coordinating effectively** with the DSCs and PWGs.
- These efforts **position us well for the upcoming phases** of the (pre)TDR.

Backup

Science Drivers for Streaming Readout at ePIC

Broad ePIC Science Program:

- Plethora of observables, with less distinct topologies where every event is significant.

Moderate Signal Rate:

	EIC	RHIC	LHC → HL-LHC
Collision species	$\vec{e}+\vec{p}, \vec{e}+A$	$\vec{p}+\vec{p}/A, A+A$	$p+p/A, A+A$
Maximum x-N C.M. energy	140 GeV	510 GeV	13 TeV
Peak x-N luminosity	10^{34} cm^{-2} s^{-1}	10^{32} cm^{-2} s^{-1}	$10^{34} \rightarrow 10^{35}$ cm^{-2} s^{-1}
x-N cross section	**50 µb**	**40 mb**	**80 mb**
Maximum collision rate	500 kHz	10 MHz	1-6 GHz
$dN_{ch}/d\eta$	0.1-Few	~3	~6
Charged particle rate	**4×10^6 N_{ch}/s**	**6×10^7 N_{ch}/s**	**3×10^{10}+ N_{ch}/s**

Enabling Next-Generation Compute-Detector Integration

- **Maximize Science**: **Capture every collision signal, including background.**
  - **High-precision measurements**: Control of systematic uncertainties is critical.
  - Event selection using all available detector data for **holistic reconstruction**:
    - **Eliminate trigger bias** and provide accurate estimation of uncertainties during event selection.
  - Streaming background estimates ideal to **reduce background** and related systematic uncertainties.
- **Accelerate Science**: **Rapid turnaround of two weeks for data for physics analyses**.
  - Timeline driven by alignment and calibration.
  - Subsystem experts indicate a two-week turnaround is feasible.
- **Technologies**: Compute-detector integration using:

Streaming Readout	Artificial Intelligence	Heterogeneous Computing
for continuous data flow of the full detector information.	for rapid processing (autonomous alignment, calibration, and validation).	for acceleration (CPU, GPU).

Computing Resource Needs (2034) and Their Implications

Processing by Use Case [cores]	Echelon 1	Echelon 2
Streaming Data Storage and Monitoring	-	-
Alignment and Calibration	6,004	6,004
Prompt Reconstruction	60,037	-
First Full Reconstruction	72,045	48,030
Reprocessing	144,089	216,134
Simulation	123,326	369,979
Total estimate processing	**405,501**	**640,147**

O(1M) core-years to process a year of data:

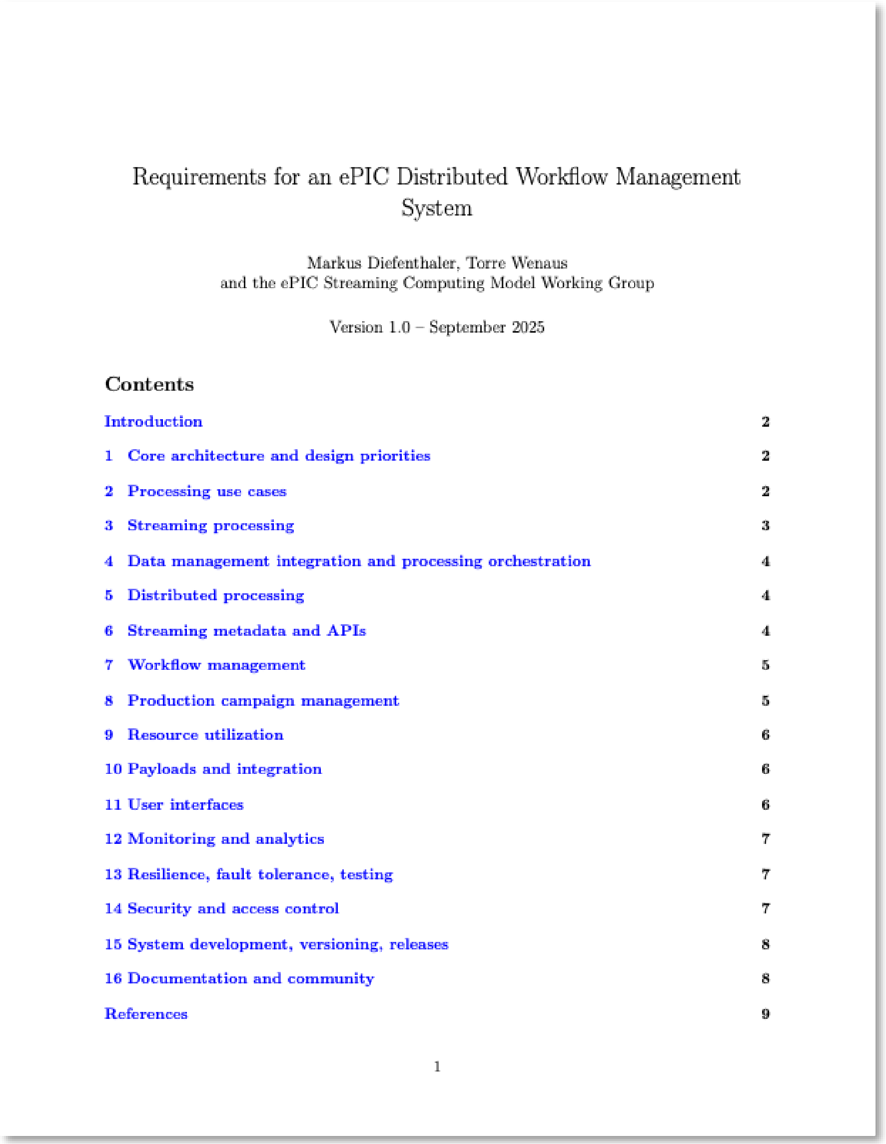
- Even with performance gains over the years, the required processing scale remains substantial.
- Highlights the need to leverage distributed and opportunistic resources from the outset.

Storage Estimates by Use Case [PB]	Echelon 1	Echelon 2
Streaming Data Storage and Monitoring	71	35
Alignment and Calibration	1.8	1.8
Prompt Reconstruction	4.4	-
First Full Reconstruction	8.9	3.0
Reprocessing	9	9
Simulation	107	107
Total estimate storage	**201**	**156**

~350 PB to store data of one year.

ePIC is a compute-intensive experiment. Its science must not be limited by computing constraints.

Requirements for Streaming Orchestration



Requirements for an ePIC Distributed Workflow Management System

Markus Diefenthaler, Torre Wenaus
and the ePIC Streaming Computing Model Working Group

Version 1.0 – September 2025

Contents

1

Requirements Document

- Builds upon and guides further development of the **ePIC Streaming Computing Model.**
- Developed collaboratively by the **ePIC Streaming Computing Model WG**.
- Informed by **lessons learned from other experiments** and streaming systems.

Key Themes

- **Scalable and Automated Workflows**: Low overhead, automated orchestration, and real-time processing across E1–2.
- **Streaming-First Design**: Native support for near real-time processing.
- **Integrated Data Management**: Tight coupling with Rucio-based DDM for data-driven workflows and provenance tracking.
- **Flexible & User-Centric Interfaces**: CLI, REST, and web interfaces with support for custom dashboards and diagnostics.
- **Robust Monitoring & Resilience**: Real-time analytics, fault tolerance, and automated recovery mechanisms.
- **Community and Documentation Focus**: Open development, transparent processes, and collaborative design.

Streaming DAQ and Computing Milestones

	FY25	FY26	FY27	FY28	FY29	FY30	FY31
DAQ	PicoDAQ	MicroDAQ	MiniDAQ	Full DAQ-v-1	Production DAQ		
Computing	Streaming Orchestration			Streaming Challenges			
Computing	AI-Empowered Streaming Data Processing			Analysis Challenges			
Computing				Distributed Data Challenges			
AI	AI-Driven Autonomous Calibration			AI-Driven Autonomous Alignment, Calibration, and Control			

- **Compute-Detector Integration**:
  - Joint deliverables between DAQ and **computing** to develop integrated systems for detector readout, data processing, and ultimately physics analysis.
  - **Key role of AI(/ML)**: Empowering data processing and enabling autonomous experimentation and control.

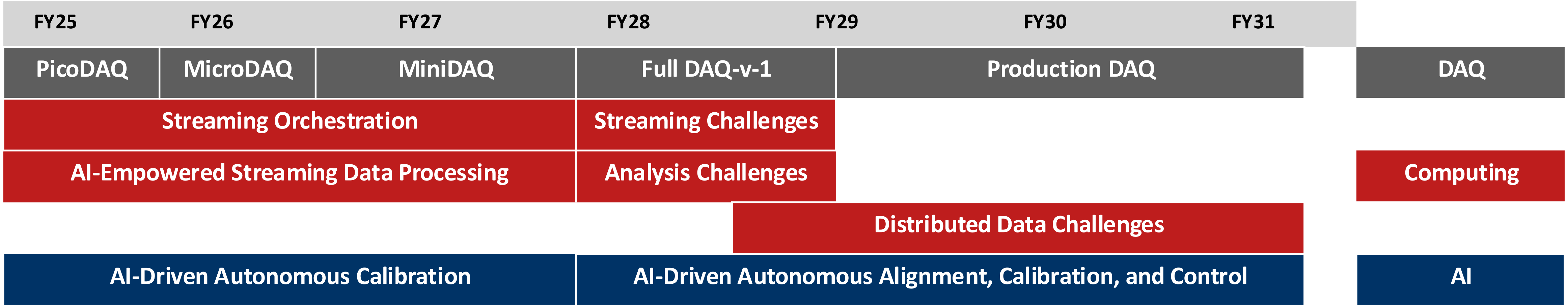
Streaming DAQ and Computing Milestones

FY25	FY26	FY27	FY28	FY29	FY30	FY31	
PicoDAQ	MicroDAQ	MiniDAQ	Full DAQ-v-1	Production DAQ			DAQ
Streaming Orchestration			Streaming Challenges				
AI-Empowered Streaming Data Processing			Analysis Challenges				Computing
				Distributed Data Challenges			
AI-Driven Autonomous Calibration			AI-Driven Autonomous Alignment, Calibration, and Control				AI

Streaming DAQ Milestones and Deliverables

FY26Q1: PicoDAQ: Readout test setups
FY26Q4: MicroDAQ: Readout detector data in test stand using engineering articles
FY28Q1: MiniDAQ: Readout detector data using full hardware and timing chain
FY29Q2: Full DAQ-v1: Full functionality DAQ ready for full system integration & testing
FY31Q3: Production DAQ: Ready for cosmics

Streaming DAQ and Computing Milestones



Streaming Orchestration Milestones and Deliverables

- ✓ **Requirement documents** for streaming orchestration developed.
- **FY28 Q1 Goal**: Deliver a functional testbed for calibrating one detector system using simulated streaming data.
- Progress is ongoing in testbed development:
  - We are evaluating streaming orchestration using **PanDA + Rucio** (slides 15–20).
  - We have demonstrated streaming data processing using **EJFAT** (slides 21–22).
  - Additional prototypes under consideration: LHCb Allen, SPADI Alliance.

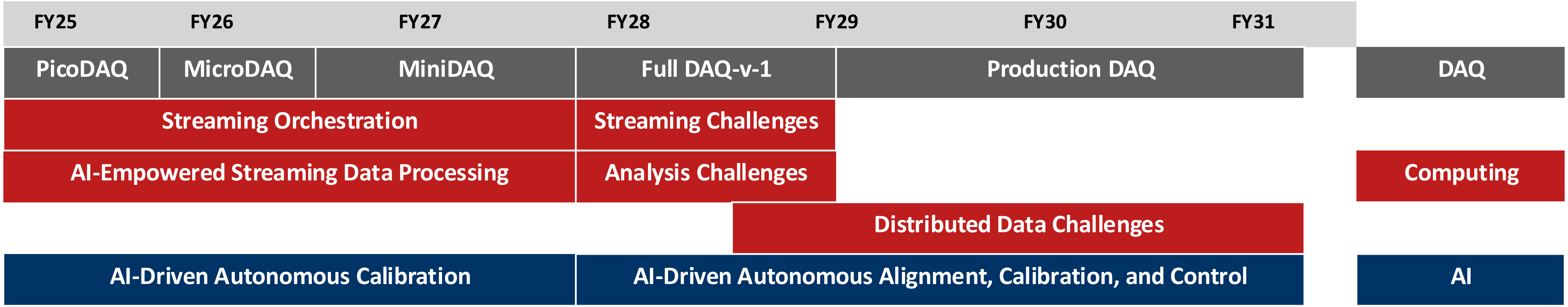
Streaming DAQ and Computing Milestones

FY25	FY26	FY27	FY28	FY29	FY30	FY31	
PicoDAQ	MicroDAQ	MiniDAQ	Full DAQ-v-1	Production DAQ			DAQ
Streaming Orchestration			Streaming Challenges				
AI-Empowered Streaming Data Processing			Analysis Challenges				Computing
				Distributed Data Challenges			
AI-Driven Autonomous Calibration			AI-Driven Autonomous Alignment, Calibration, and Control				AI

Streaming Data Processing Milestones and Deliverables

- ✓ **JANA2 enables data processing at the timeframe, event, and sub-event levels.**
- **FY28 Q1 Goal**: Achieve streaming data reconstruction with high efficiency in identifying physics collision events in simulations, including varying levels of background. This includes an AI/ML challenge focused on developing algorithms for distinguishing physics events from background.
- Progress is ongoing in streaming data reconstruction (slides 13–14).

Streaming DAQ and Computing Milestones

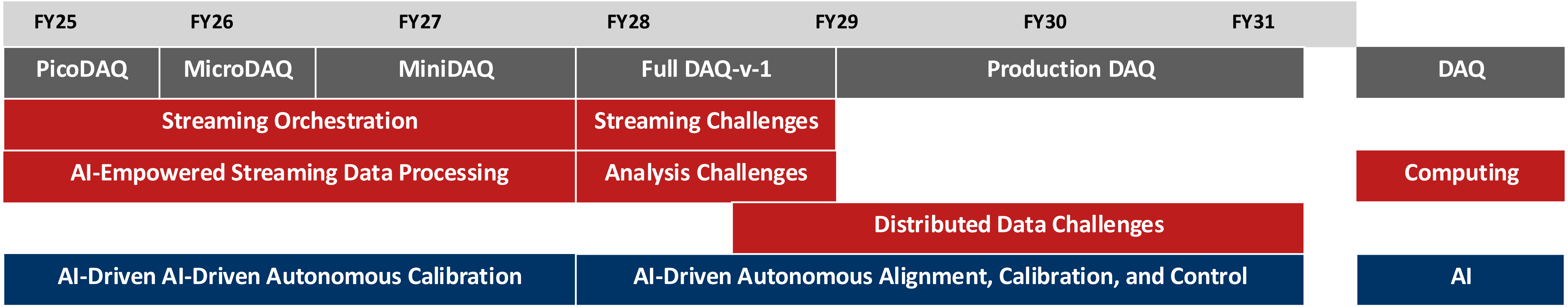


Streaming, Analysis, and Distributed Data Challenges

Streaming Data Challenge	E0+E1	Archiving and monitoring of TFs / STFs. Streaming reconstruction based on TFs / STFs.
Analysis Challenge	E1+E2	Autonomous alignment and calibration based on STFs Exercising end-to-end workflows from STFs to physics analysis.
Distributed Data Challenge	E1+E2	Exercising scaling and capability tests.

- Preparations focus on streaming orchestration, data processing, and monthly simulation campaigns that support detector and physics analysis for the preTDR and Early Science Program, while also driving near-term software development goals.

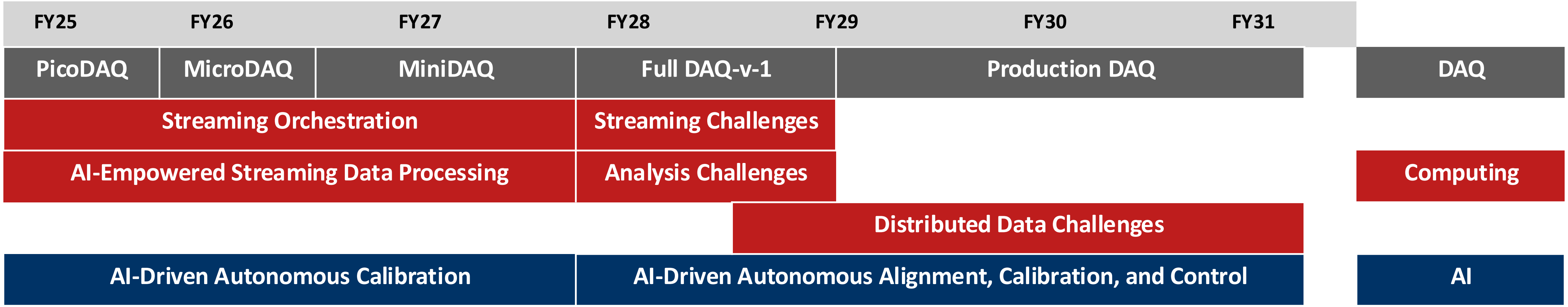
Streaming DAQ and Computing Milestones



AI-Driven Autonomous Calibration

- Progress continues on understanding alignment and calibration workflows in collaboration with subsystem experts, with a focus on identifying timelines and interdependencies.
- The strategy for autonomy involves algorithms for change detection and agentic workflows.
- **FY28 Q1 Goal**: Autonomous calibration of one detector system using simulated streaming data.

Streaming DAQ and Computing Milestones



AI-Driven Autonomous Alignment, Calibration, and Control

- Initial efforts focus on enabling fast access to detector data in a format suitable for real-time AI processing.
- Develop tooling to couple data stream processing with control systems with an emphasis on safety and reliability.

Development Priorities

Chao Peng (ANL), Sakib Rahman (BNL)

Physics and Detector Simulation

- **Charge**:
  - Development of accurate MC simulations using a suite of physics and background generators and detector simulation based on Geant4 and DD4hep
- **Priorities for 2025**:
  - Continue to support the **detector design** and integration with services.
  - Collaborate with the EIC Project to evaluate the **differences between the engineering and simulation designs**, and lead discussions with the DSCs on how to address these differences.
  - Continue to support the development of **background modeling** and implement its timing structure in physics and detector simulations, together with the Background TF.
  - Enable **simulation of streaming readout** by providing the option to switch between streaming data and event data modes.
  - Coordinate the **development of digitization and noise models** with the DSCs and the Electronics and DAQ WG.

Development Priorities

Derek Anderson JLab), Chandra Chatterjee (INFN Trieste,), Shujie Li (LBNL)

Reconstruction Framework and Algorithms

- **Charge**:
  - Development of a holistic and modular reconstruction for the integrated ePIC detector.
- **Priorities for 2025**:
  - Drive the **development of the reconstruction framework to meet ePIC needs**, e.g., on modularity or streaming data processing.
  - Host collaboration-wide discussions on all aspects of reconstruction, driving the **work toward holistic reconstruction**.
  - Enable reconstruction algorithms to **handle physics events with background**.
  - Collaborate with **PWGs** on **shared reconstruction priorities**, which currently include:
    - Secondary vertexing
    - Hadron identification
    - Particle flow algorithms for jet reconstruction
    - Event kinematics
  - Integrate continued development of **web-based event display** in reconstruction efforts.

Infrastructure Priorities

Marco Battaglieri (INFN Genoa 🇮🇹), Taku Gunji (Tokyo 🇯🇵), Jeff Landgraf (BNL)

Streaming Computing Model

- **Charge**:
  - Development of the computing model for the compute-detector integration using streaming readout, AI/ML, and heterogeneous computing, in collaboration with the Electronics and DAQ WG.
- **Priorities for 2025**:
  - Define **requirements for streaming orchestration** and **set up corresponding testbeds**:
    - Develop a testbed for event reconstruction from streamed data in EICrecon, separating signal from background events and demonstrating how we will reconstruct physics events.
    - Establish an initial testbed for super time frame building and processing, and deliver a corresponding requirements document.
  - Document alignment and calibration workflows jointly with the DSCs and identify **requirements for autonomous alignment and calibration**.
  - **Publish the ePIC Streaming Computing Model report**, and the related section in the (pre)TDR.

Operation Priorities

Thomas Britton (JLab), Sakib Rahman (BNL)

Production

- **Charge**:
  - Responsible for the coordination and production of simulation campaigns based on priorities from the Technical and Analysis Coordinators.
  - Develop automated production workflows that scale with the needs of the collaboration.
- **Priorities for 2025**:
  - **Automation Priorities**:
    - Improve the exposure and organization of monitoring so that no one needs to be an OSG expert to track progress, thereby enabling more individuals to participate in operating the monitoring.
    - Explore workflow and workload management tools.
  - **Simulation Campaign Priorities**:
    - Roll out Rucio to the collaboration as the default method for finding and accessing simulation productions.
    - Establish liaisons with DSCs and PWGs to actively participate in the simulation campaigns.

Operation Priorities

Stephen Kay (York 🇬🇧), Holly Szumila-Vance (FIU)

User Learning

- **Charge**:
  - Responsible for onboarding via a landing page for new collaboration members and additional appropriate mechanisms.
  - Responsible for support via documentation, help desk, and training.
  - Ensure that software is discoverable (easy to use with only minimal instructions) and simulated data and metadata is findable.
- **Priorities for 2025**:
  - New initiative: **Roadmap towards discoverable software**.
  - **Revised and frequently updated FAQs**.
  - **Rolling schedule of software tutorials** that incorporates updated versions of existing tutorials, new material, and relevant resources from the HSF Training WG.

Tutorials, Helpdesk and User Surveys

Successful Landing Page for Onboarding

- Many new collaborators successfully onboarded themselves using only the landing page
- A survey of early career scientists in ePIC received this feedback: *“Having this talk/form [survey] already sets the EIC above lots of other places I've worked with”*

Any member of the collaboration can **directly contribute to any of the websites by submitting Pull Requests**.

Tutorial Series

- Understanding the Simulation Output (Shujie Li)
- Analysis and Working with the Simulation Output (Stephen Kay)
- Getting Started with a Physics Analysis (Alex Jentsch)
- Inclusive Kinematics Reconstruction (Stephen Maple)

Helpdesk on Mattermost

- A place to ask questions in a respectful and open environment. Well appreciated.
- Experts are readily available, and questions are addressed efficiently.