ePIC Streaming Computing Model Intro

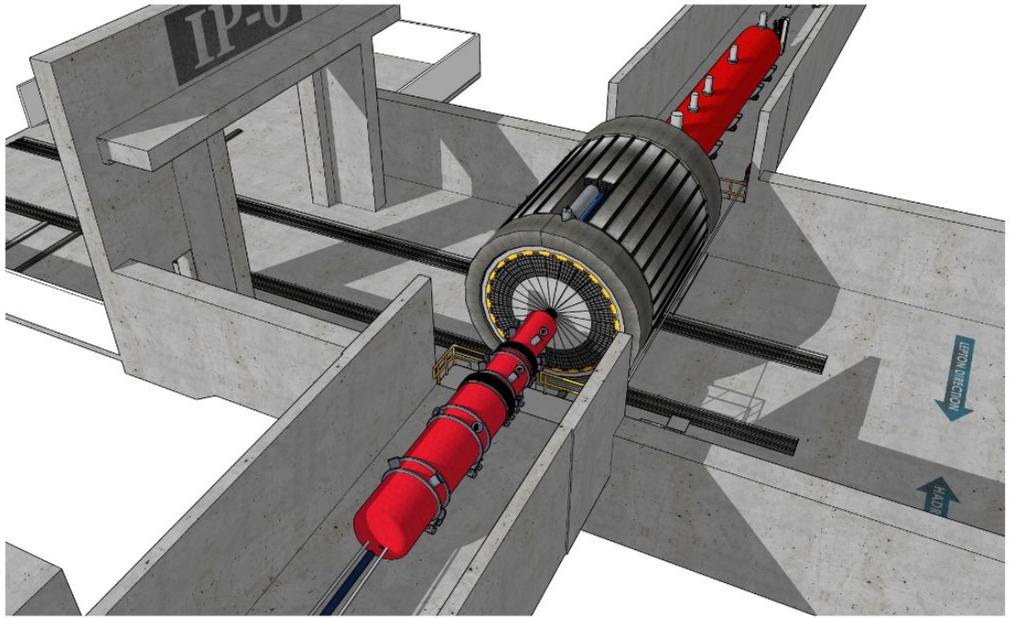
Torre Wenaus (BNL) for the ePIC Streaming Computing Model Working Group

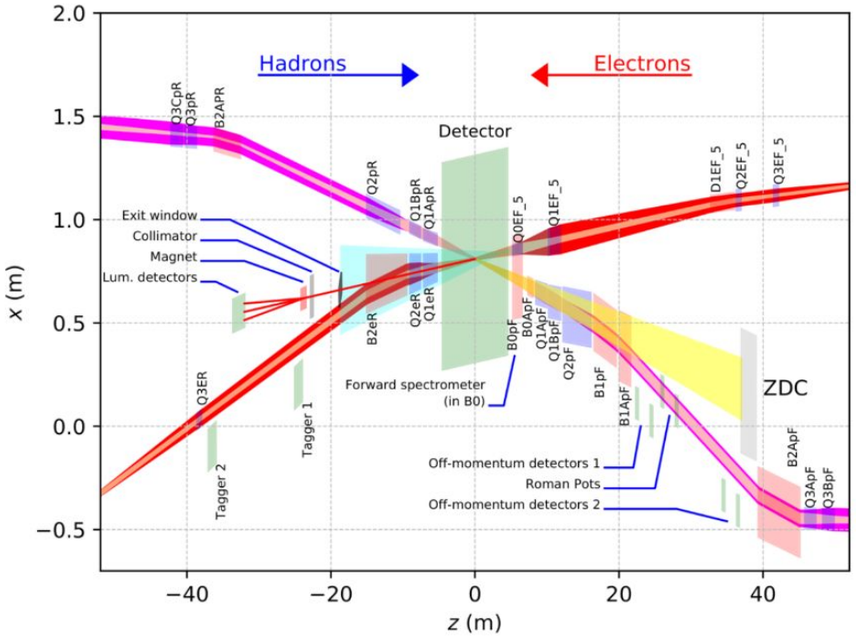
ePIC S&C meeting
Apr 25 2024
CERN

EIC streaming motivation: Maximizing physics reach

Integrated interaction and detector region (90 m)

Get ~100% acceptance for all final state particles, and measure them with good resolution. All particles count!





Compute-Detector Integration

Extend integrated interaction and detector region into detector readout (electronics), data acquisition, data processing and reconstruction, and physics analysis.

Streaming Readout Capability Due to Moderate Signal Rate:

- **Capture every collision signal**, including background.
- Event selection using all available detector data for **holistic reconstruction**:
  - **Eliminate trigger bias** and provide accurate estimation of uncertainties during event selection.

The overall model, streaming DAQ to global processing

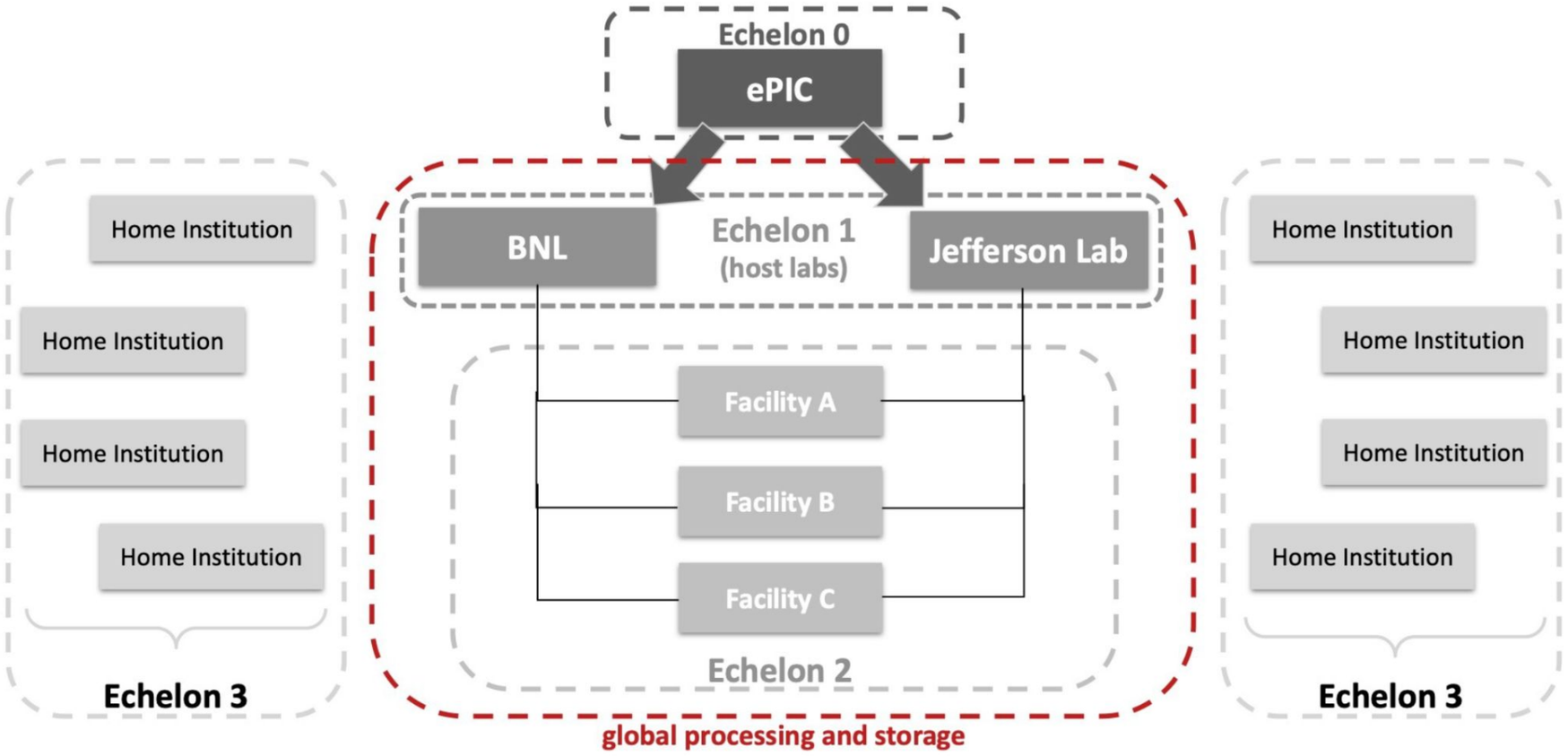
Four tiers:

Echelon 0:
ePIC experiment

Echelon 1:
Two host lab facilities

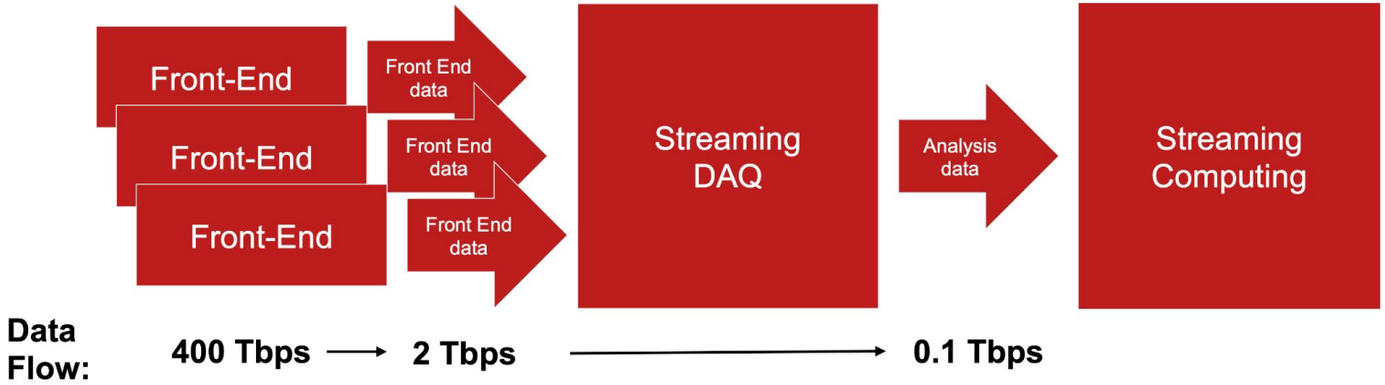
Echelon 2:
Global processing and data management

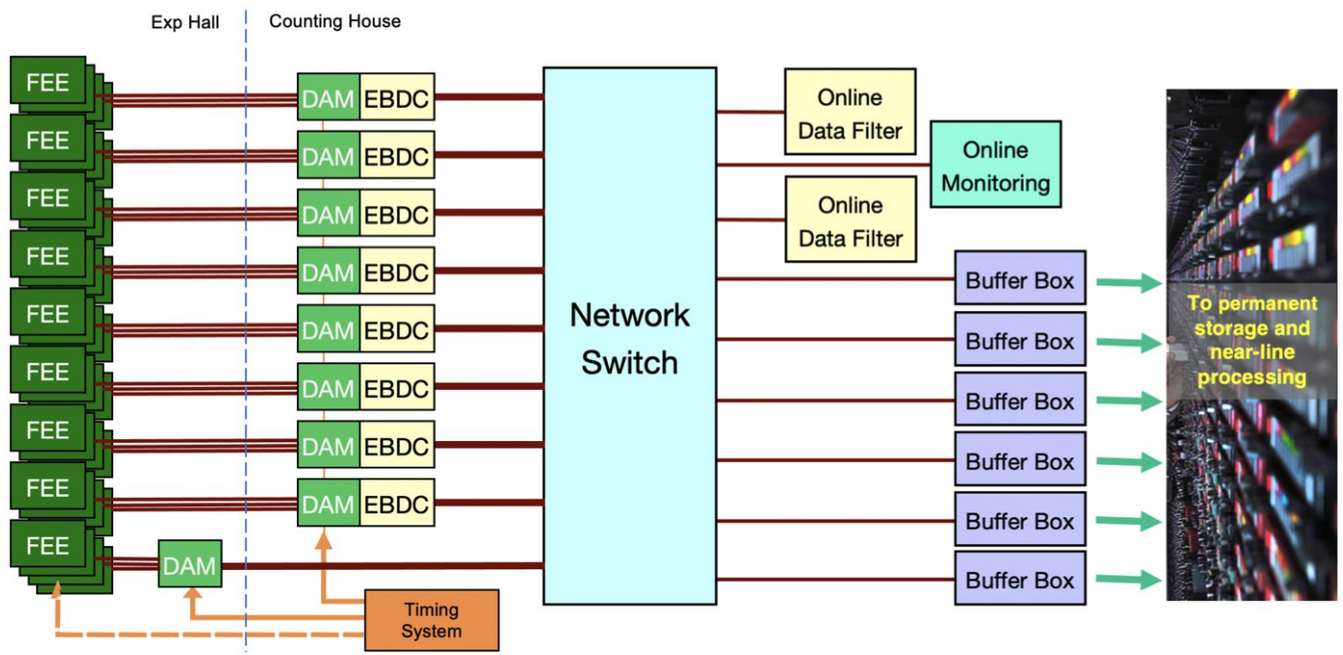
Echelon 3:
Home institutes: where the analyzers are



ePIC streaming: DAQ to storage

- Each detector element determines whether it has been “hit”
- Passes time stamped data up the DAQ chain
- Full detector data is available in the counting room for detector-wide decisions
- A light touch on data reduction and filtering: noise reduction, zero suppression, compression
- Send a truly minbias sample to storage
  - Eliminating trigger bias and associated systematics
- The stream to storage is the working definition of DAQ - offline divide
- Data emerges from DAQ organized in ~microsecond-wide ‘time frames’ containing all subdetector data for hundreds of events
- time frames => events in the prompt processing taking place at Echelon 1



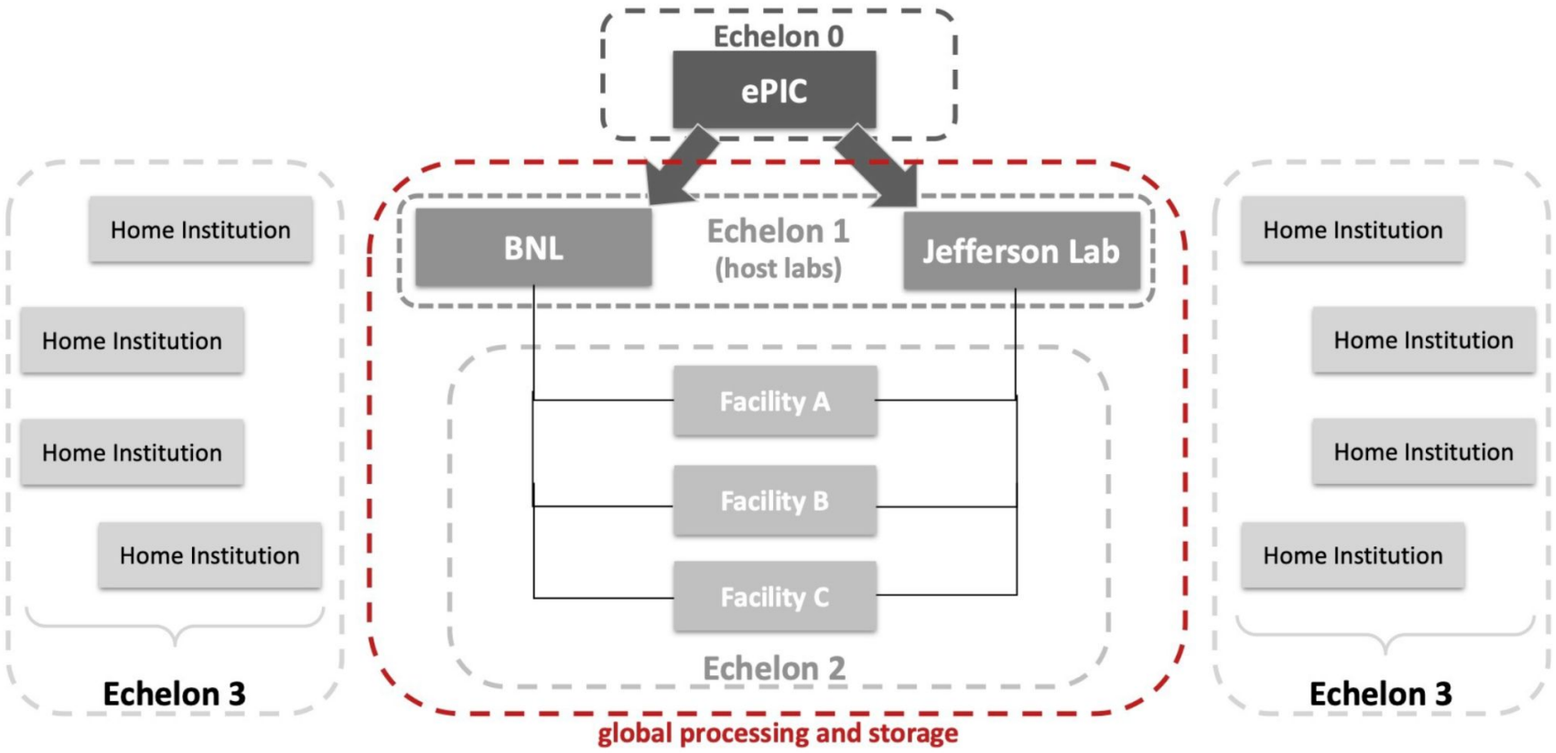


I. Vvenaus

ePIC Streaming Computing Model

[ePIC streaming computing model report V1, Oct 2023]

- Echelon 0
  - ePIC experiment, counting house
    - and SDCC enclave?
  - symmetric delivery of raw data stream to two host labs, two complete copies where downstream processing can proceed
  - exactly where and how the data stream is sent to the two host labs is TBD
- Echelon 1
  - Host lab computing facilities
  - Both with a complete archival data sample, and equal opportunity to process the data
- Echelon 2
  - Global processing and data storage facilities with a committed contribution to ePIC (managed by a Resource Review Board)
- Echelon 3
  - Home institute computing
  - No formal agreements but must be served well, this is where most of the physicists doing analysis are
- E1 + E2 + E3 seen as a 'web', as for LHC (after some evolution), not a hierarchy
- Opportunistic resources (e.g. OSG) will be (already is) a very productive source of computing
- Processing is a mix of
  - quasi-continuous processing of fine grained data (prompt streaming processing)
  - batch style processing (offline)



Brookhaven National Laboratory

Mapping of processing use cases to the Echelons

Echelon 0: ePIC Experiment
Echelon 1: Host Labs
Echelon 2: Global processing and data facilities
Echelon 3: Home institute computing

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

*** Opportunistically**

That’s it, keeping it short!

Enough context from the ePIC side to now hear about and discuss LHCb