Ideas for Real-Time Analysis for HL-LHC using the CMS DAQ System

Remigius K Mommsen
Fermilab

Fermilab

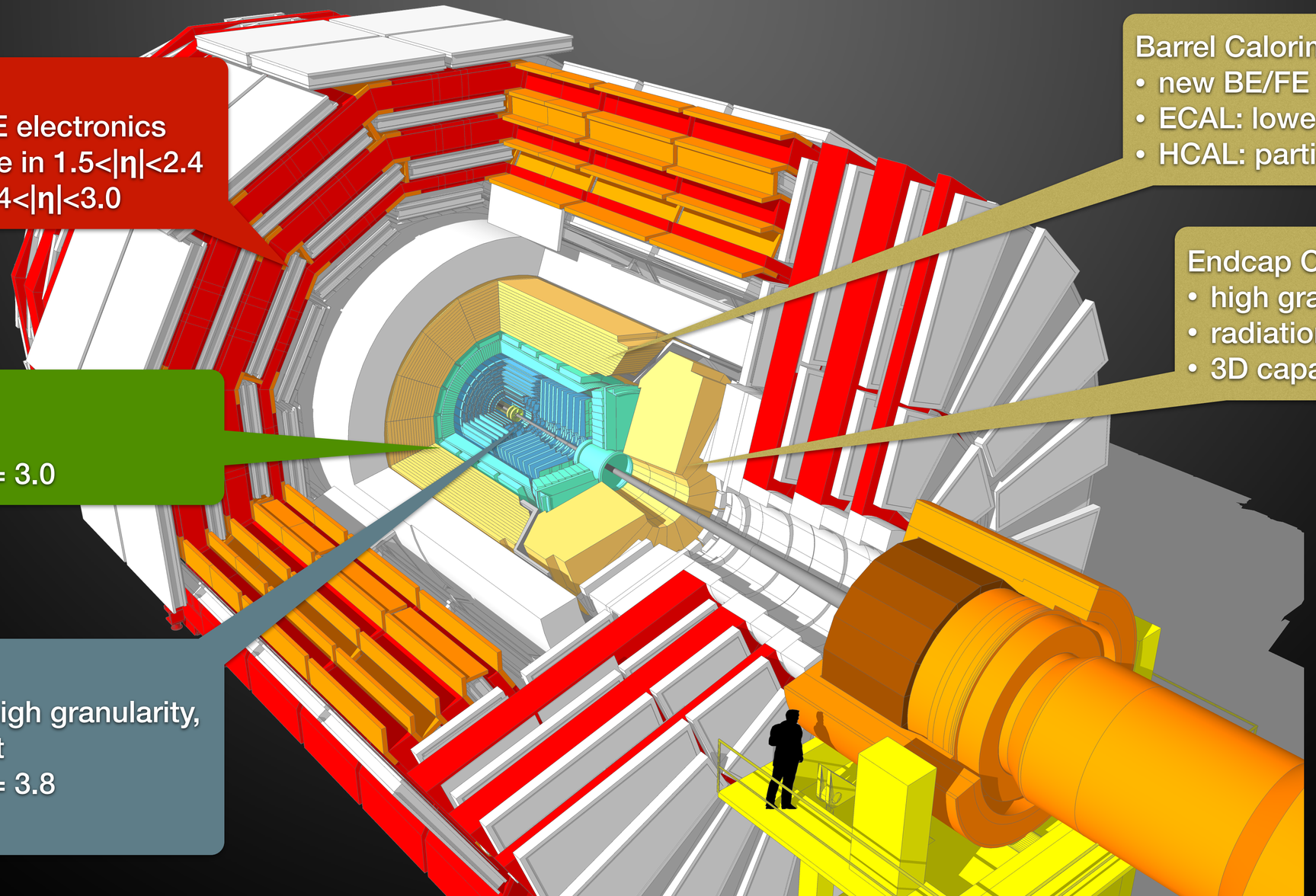
Disclaimer

The idea of the L1 scouting originates from Emilio Meschi (CERN)

This talk is based to a large extend on material presented by Hannes Sakulin (CERN)
at CHEP 2019, Adelaide, Australia

Any mistakes or misinterpretations are mine

All-new CMS for HL-LHC (2027 onwards)



Muon Systems
- new DT/CSC BE/FE electronics
- GEM/RPC coverage in 1.5<|η|<2.4
- Muon-tagging in 2.4<|η|<3.0

Barrel Calorimeters
- new BE/FE electronics
- ECAL: lower temperature
- HCAL: partially new scintillator

Endcap Calorimeters
- high granularity calorimeter
- radiation tolerant scinitllator
- 3D capability and timing

MIP Timing Detector
- 30-60 ps resolution
- coverage up to |η| = 3.0

Tracker
- radiation tolerant, high granularity, low material budget
- coverage up to |η| = 3.8
- track trigger at L1

L1 rate:
750 kHz

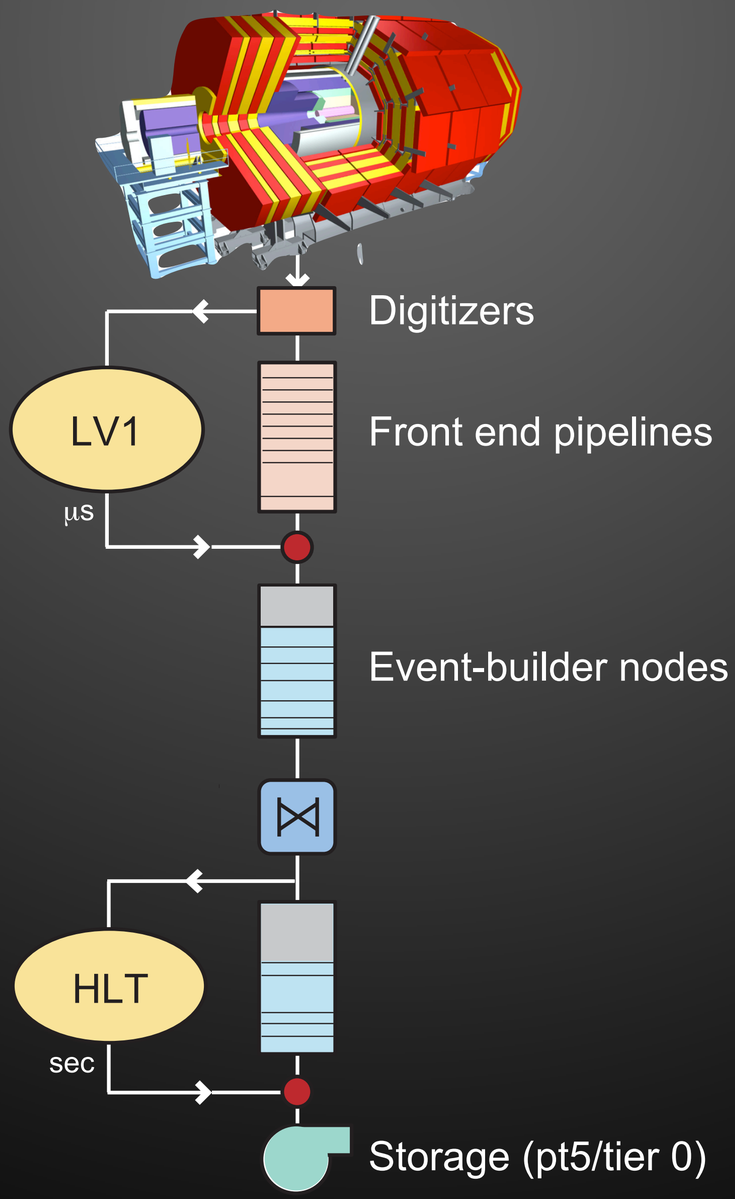
HLT rate:
~7.5 kHz

Event size:
7.5 MB

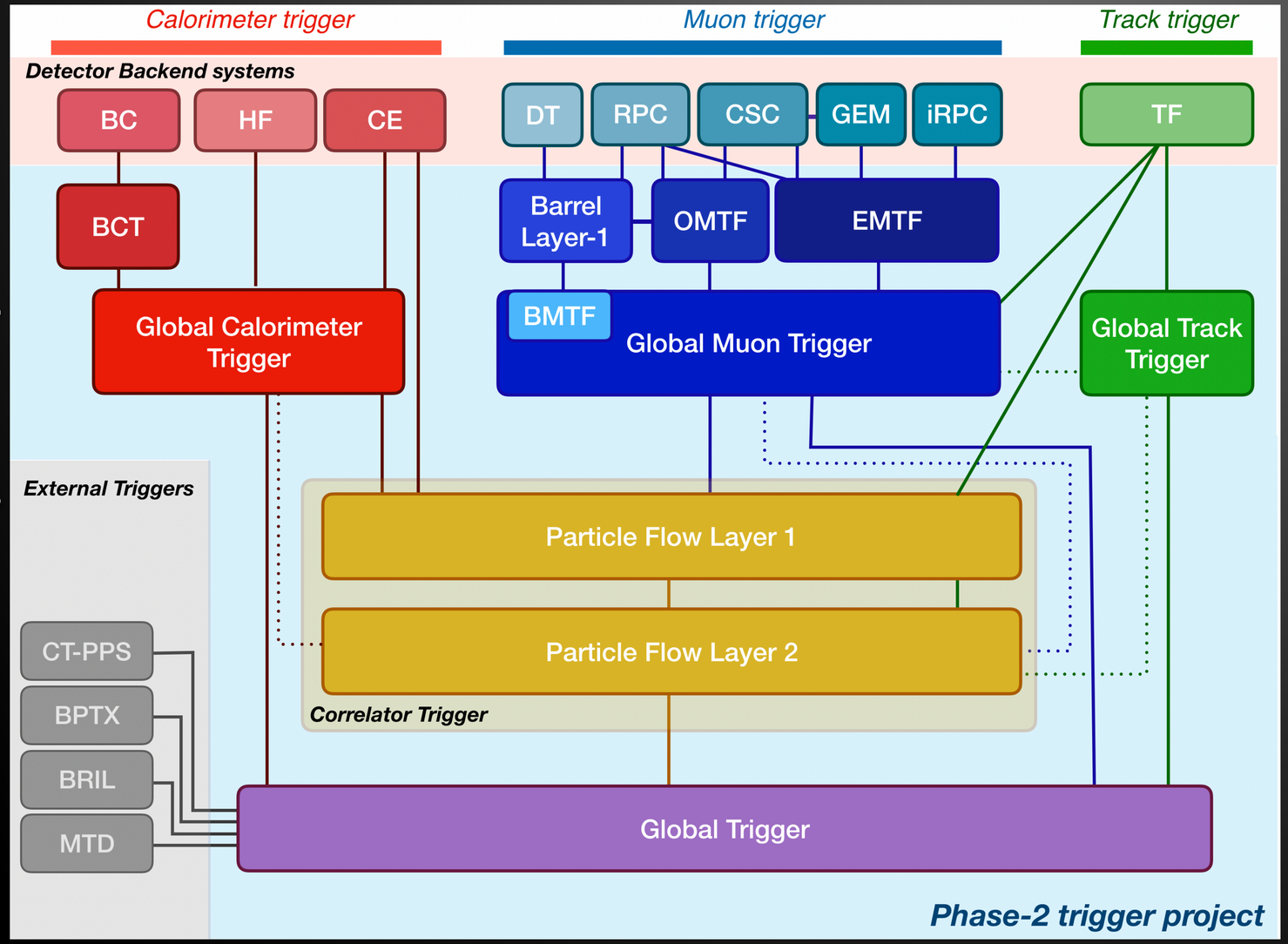
300 TB/s
@ 40 MHz

CMS Trigger & DAQ — 2 Trigger Levels Only

Phase 0 & 1 — 2008-24		Phase 2 — 2027-
40 MHz		40 MHz
100 kHz		750 kHz
1.5 MB event size		7.5 MB event size
0.15 TB/s		5.5 TB/s
2 kHz		7.5 kHz



L1 Trigger for HL-LHC



High resolution objects

- Tracker track reconstruction in firmware
- Vertex finding
- Kalman filter muon reconstruction
- Displaced muons
- High precision calorimetry
- Particle flow reconstruction

Topological algorithms including invariant/transverse mass cuts

Machine learning algorithms

Inter-bx algorithms (limited to +/- 3 bx)

What is Real-Time Analysis?

Analyze events while the data is being taken

- Partial events with limited resolution
- Full events with sub-optimal calibrations
- Much higher rate than possible with offline analysis
- Stringent time constrains

Store summary results for certain topologies at higher rate

- E.g. low-mass di-jets, three-jet resonances, di-muons

LHCb will does most analysis in “real-time”

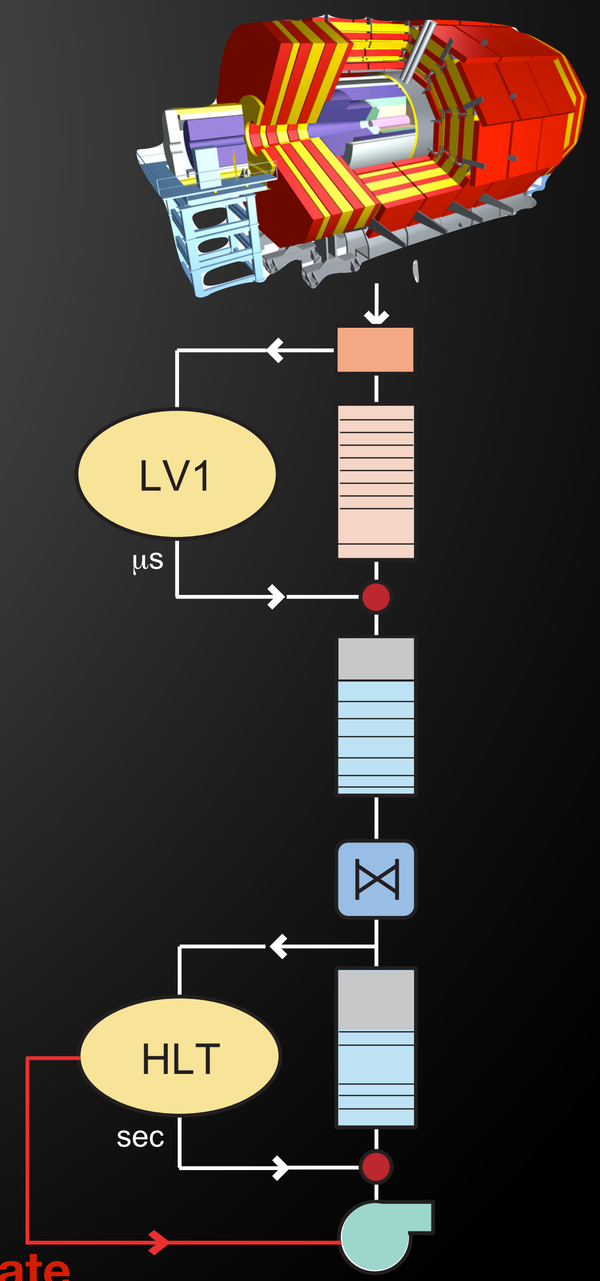
- 2-step HLT selection
- 2nd step is run after calibrations have been done
- Same physics quality as offline for most objects

HLT Real-time Analysis

Data Scouting at HLT used successfully in CMS since 2011

- Save HLT physics objects to disk
- Perform offline analysis on these objects rather than on offline reconstructed entities
- No raw data is saved and no further reconstruction is performed for these events
- Typically 1-5 kHz of scouting data O(100 MB/s)

Run2 dataset	Event content	Event size	Rate	L1 seeds/HLT Cuts	Comment
ScoutingPFMuon	[Muons],electrons, photons,pfMet, pfJets	13 kB	0.6kHz	Many L1 seeds, pT>3 GeV, M>10 GeV	pfCandidates of jet constit., not in 2017, reworked in 2018
ScoutingPFHT	[Muons],electrons, photons,pfMet, pfJets	13 kB	0.6kHz	Many L1 seeds, HT=410 GeV	pfCandidates of jet constit.
ScoutingCaloMuon	muons(displ-) vertices, (btag) calojets,calomet	3 kB (7kB `18)	~3kHz	Many L1 seeds, pT>3 GeV,	Added tracks around muons in 2018, only since 2017
ScoutingCaloHT	muons,(displ-) vertices, (btag) calojets,calomet	3 kB (7kB`18)	~3kHz	Many L1 seeds, HT=250 GeV	Added tracks around muons in 2018



L1 Trigger Scouting

Acquire L1 trigger data at full bunch crossing rate

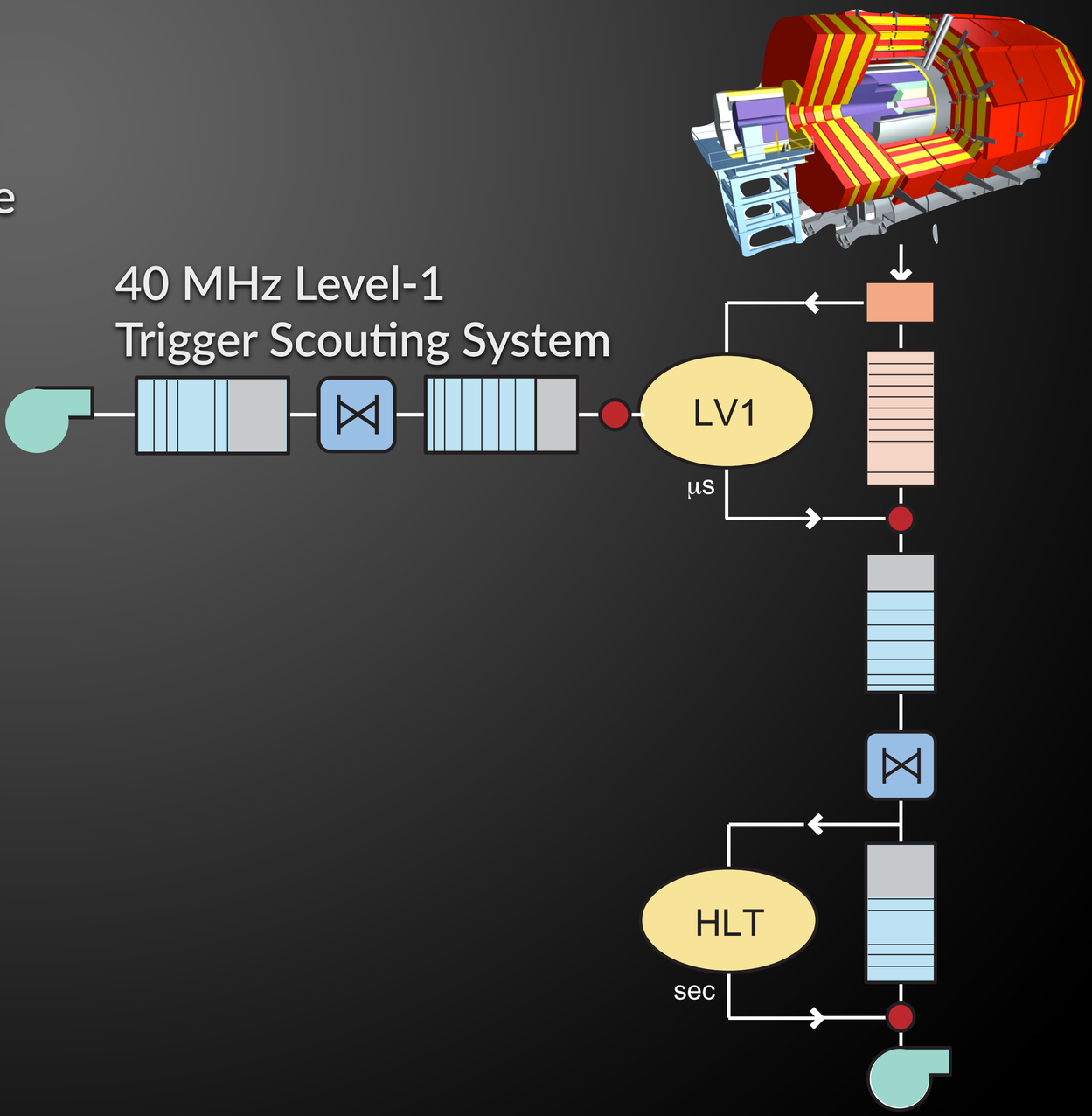
- No back pressure
- Drop data if system cannot keep up with rate

Analyze certain topologies at full rate

- Real-time analysis
- Store tiny event record

Planned for HL-LHC

- Prototyping now
- Testing during run 3



Physics to Look at with L1 scouting (non-exhaustive)

Physics use case

- Rare process
- Difficult to select at Level-1 trigger - despite upgraded L1 trigger
  (Available cuts give low efficiency at attributed rate budget)
- Analysis is possible with resolution available at Level-1
- Scouting for new signal -> then point L1 trigger to it

Several Physics channels identified where L1 scouting could potentially make a difference

- Dark Photon
  - $D^{*0} \to D^0 A, A \to ee/A \to \mu\mu, A \to ee, A \to jj$
- Flavor anomalies:
  - $W' \to \tau\nu$ @ m(W')=300 GeV
- Classic B-physics:
  - ($B \to \tau\tau$ or $\mu\mu$?)
- Hidden Hadronic physics (dark sector?)
  - Soft bombs, RPV SUSY
- Kaon anomalies
  - $K_s \to \mu\mu$
- Standard Model
  - QCD measurement, Other hadronic objects?

A list of models we have created by consulting with theorists

Its not exhaustive but covers several classes of topics

Phil Harris(MIT) + Many others
(Nhan Tran, Jeff Berryhill, David Sperka, Amit Lath,Marc Osherson, Abhijith Gandrakota,Dylan Rankin,Javier Duarte,Nick Smith,Matt Low, Paddy Fox,Nikita Blinov,Marat Freytsis, Jure Zupan)

Other uses for Level-1 Trigger scouting

Scouting provides invaluable diagnostic and monitoring opportunities as well

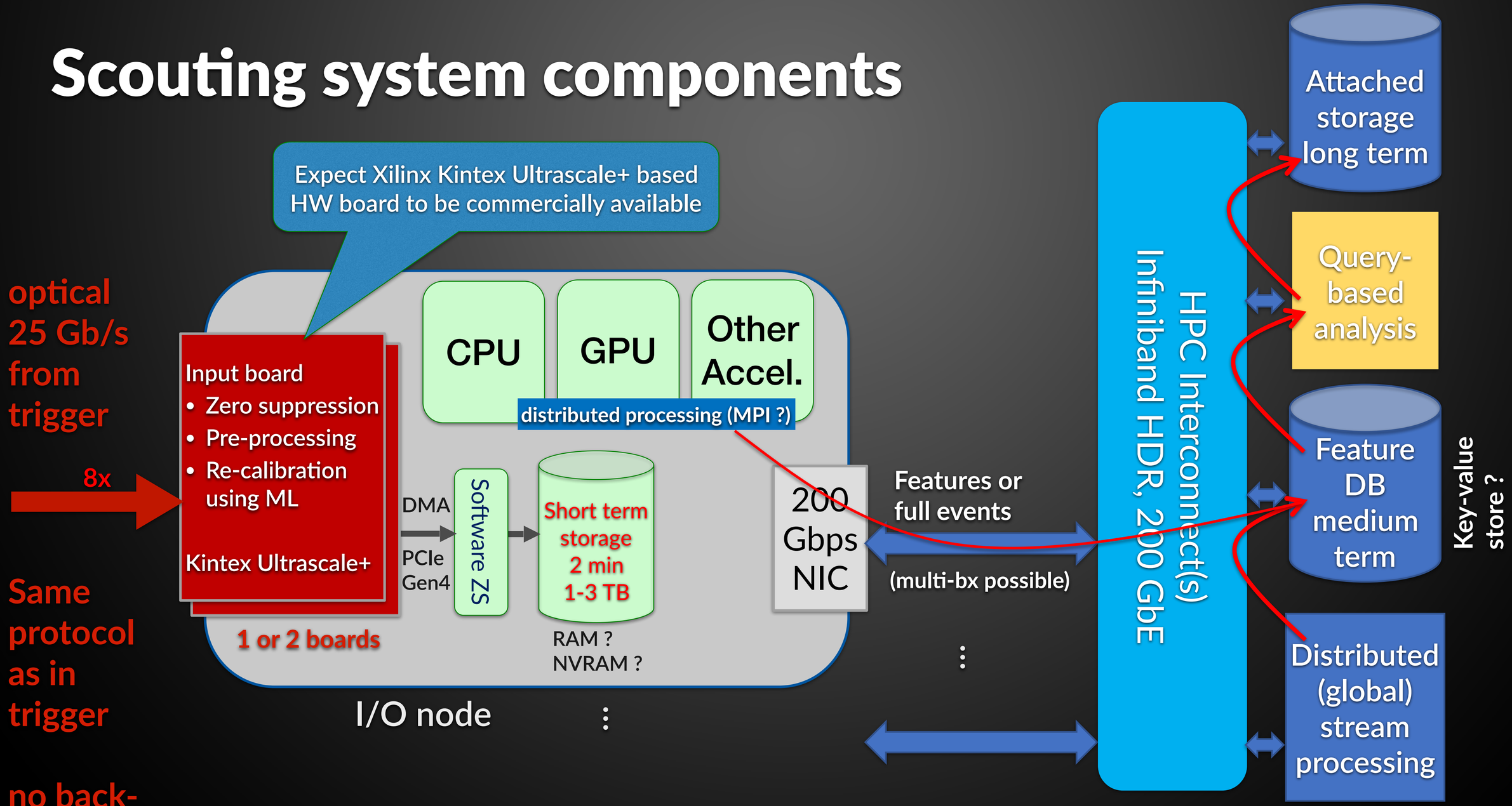
- BX-to-BX correlations available at all times (cosmics, pre/post firing, etc.)
- Real-time heat maps to immediately spot problematic channels
- High-stat cross-check of algorithms (e.g. GT inputs/outputs)

Per-bunch luminosity measurement using physics channels with high statistics

Anomaly detection with deep-learning algorithms

HL-LHC 40 MHz L1 Scouting Stageable Architecture

Scouting system components



Ingredients

Trigger data captured directly from the Level-1 using spare outputs of the processing boards

- Assuming same 16/25 Gbps serial optical links used for the Level-1 interconnects and using the same protocol

Input hardware: PCIe boards with (modest) FPGA in 1U PC (I/O node) – (uGMT scouting uses KCU1500 [limited to 16 Gbps])

- Zero-suppression, local pre-processing (e.g. re-calibration using ML) in FPGA
- DMA to host memory for short-term buffering (~2 min)
- Baseline: eight optical inputs per board (PCIe Gen4 ~ 200Gbps over 16 lanes), one or two input boards per PC

I/O nodes (CPU, GPU, other accelerators) use distributed algorithms to extract features while data are buffered in memory

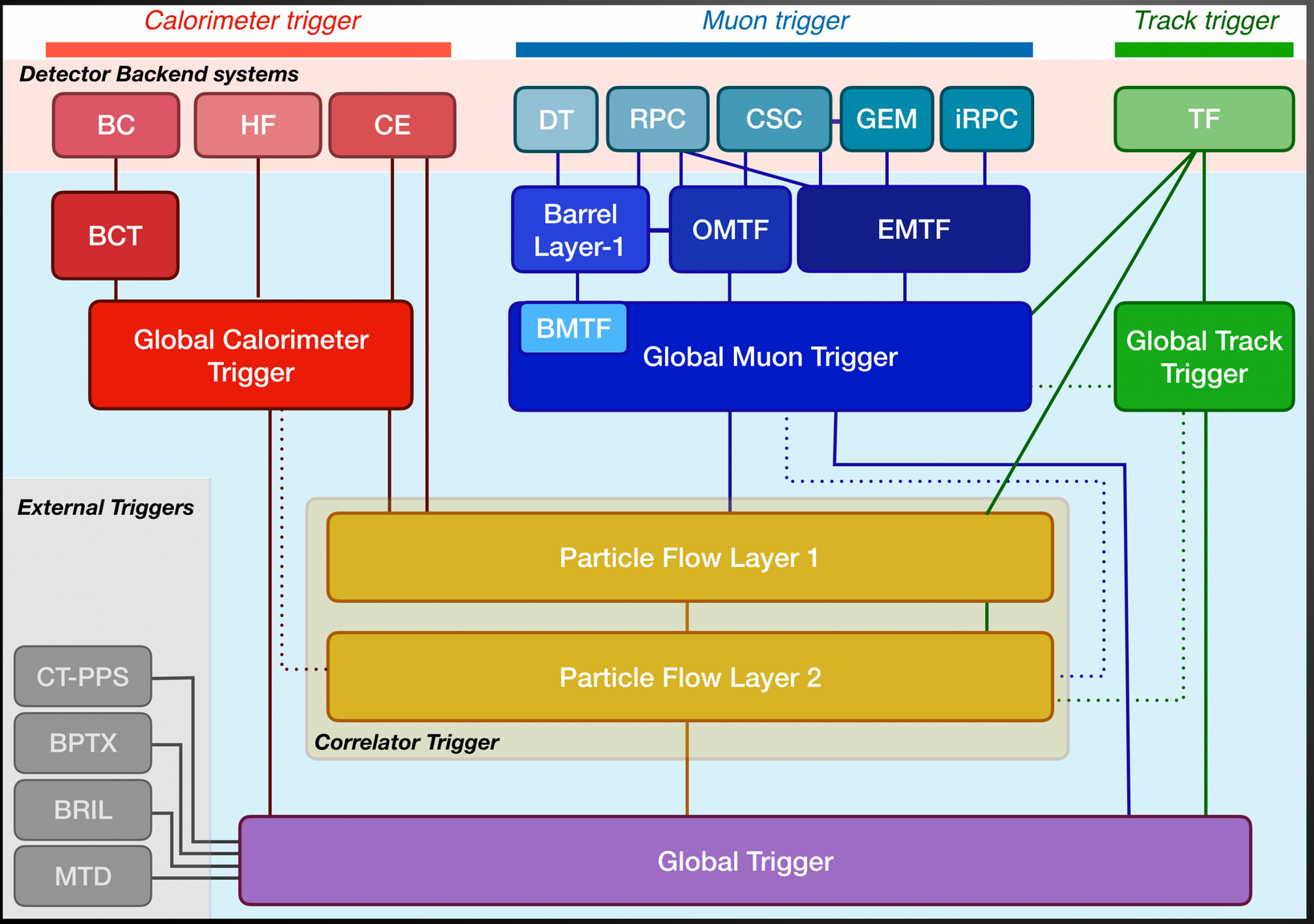
- 1-3 TB short-term buffer (e.g. NVRAM, could be cheaper with acceptable latency)
- 200 Gbps low-latency interconnect (e.g. InfiniBand HDR or 200 GbE)
- Interesting features and/or full “events” (multi-bx possible) streamed over interconnect to global processing “farm”

Distributed global stream processing and storage into “feature DB”

- Organizes features in “searchable” data structures
- Search-engine-like system optimized for numerical data, medium term storage (e.g. key-value store)

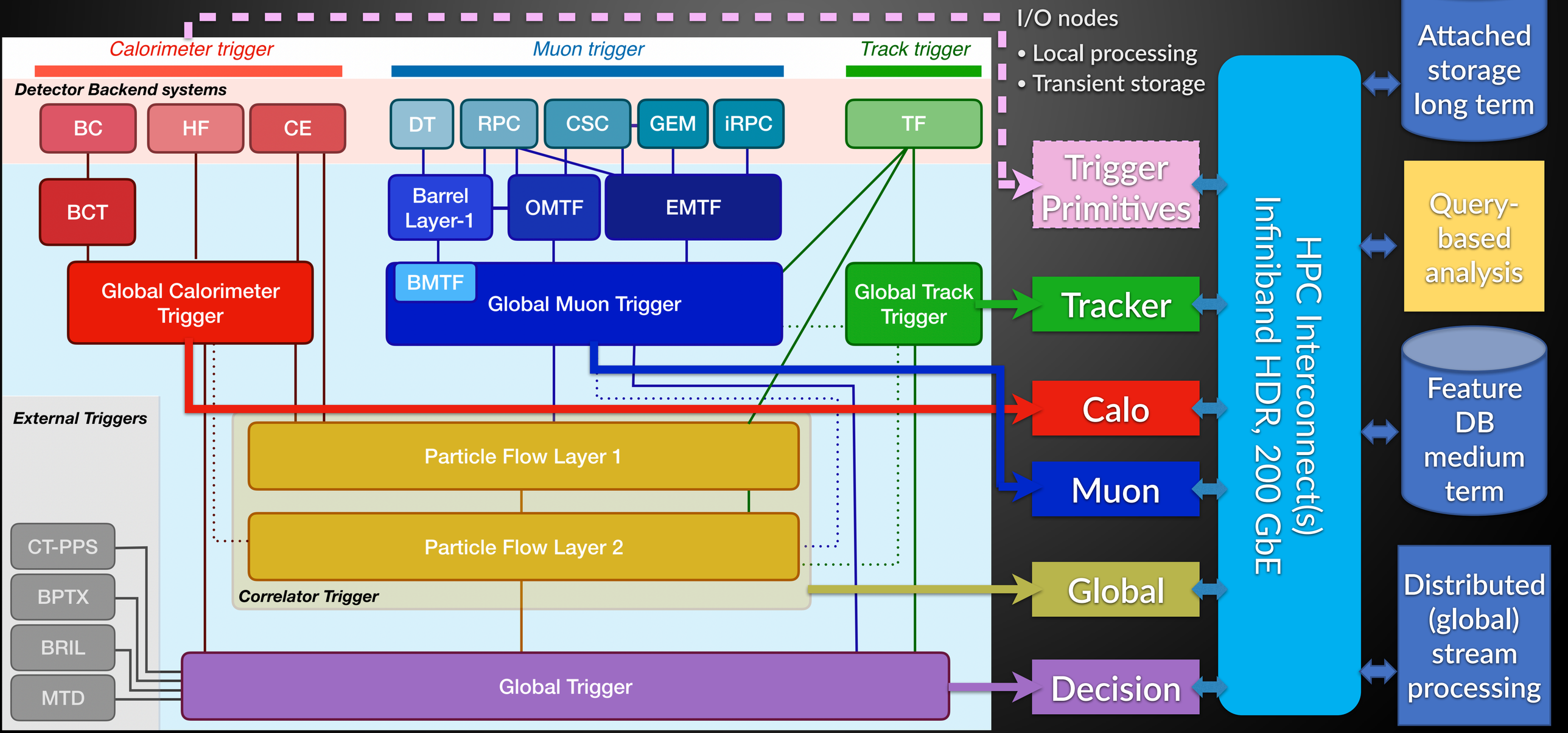
Analysis by query, analysis results to permanent storage

L1 Trigger System



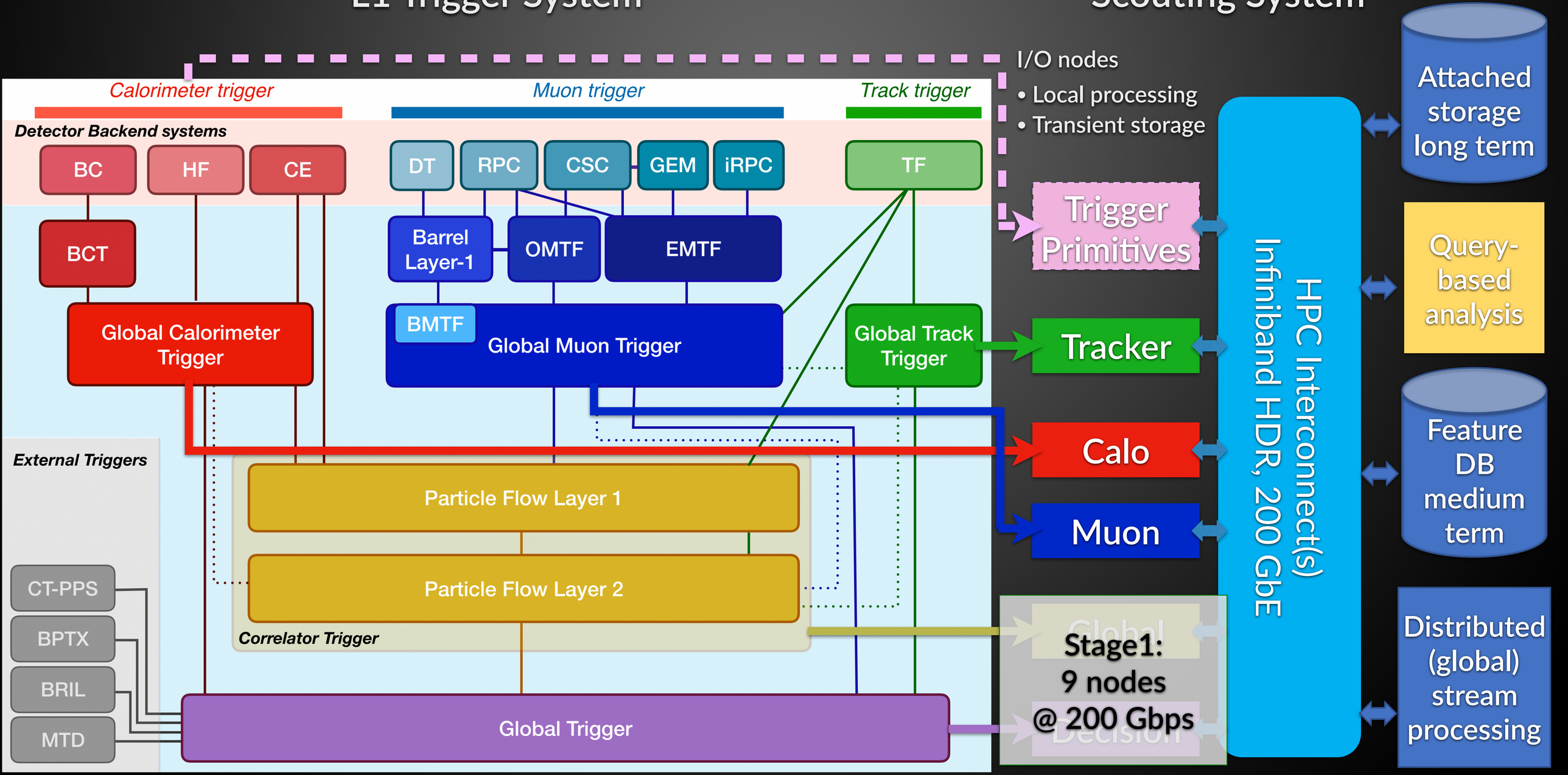
L1 Trigger System

Scouting System



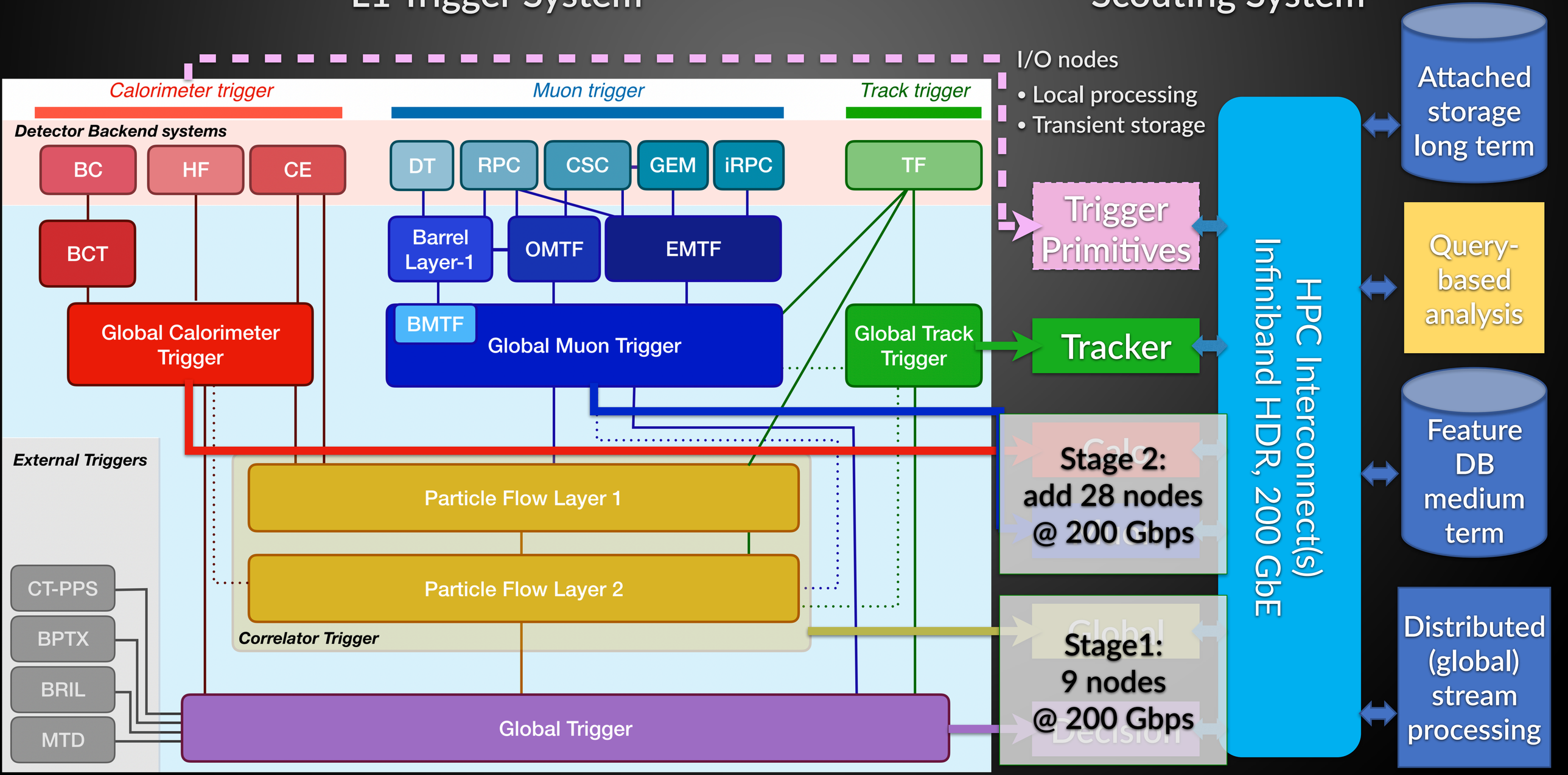
L1 Trigger System

Scouting System



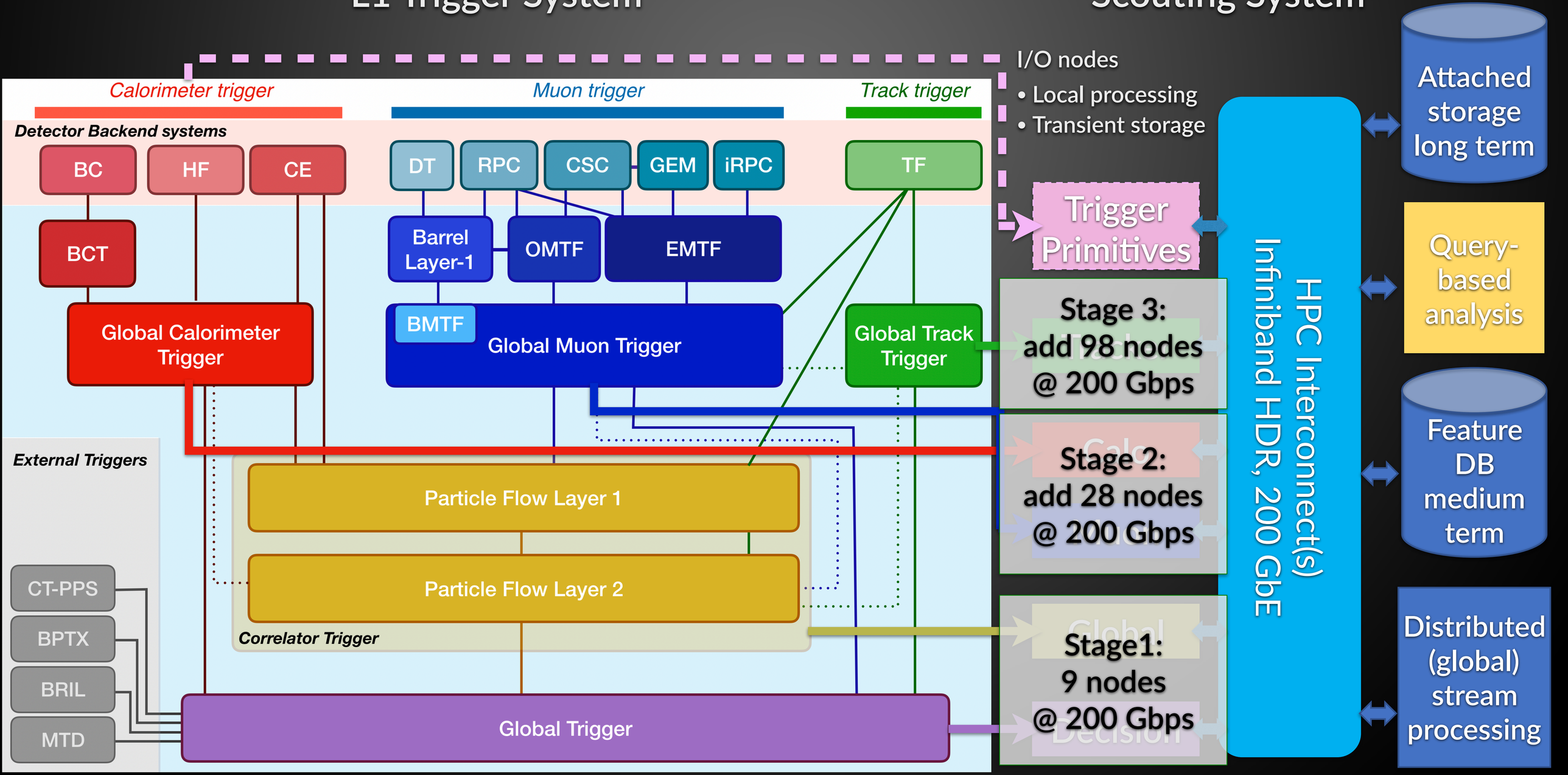
L1 Trigger System

Scouting System



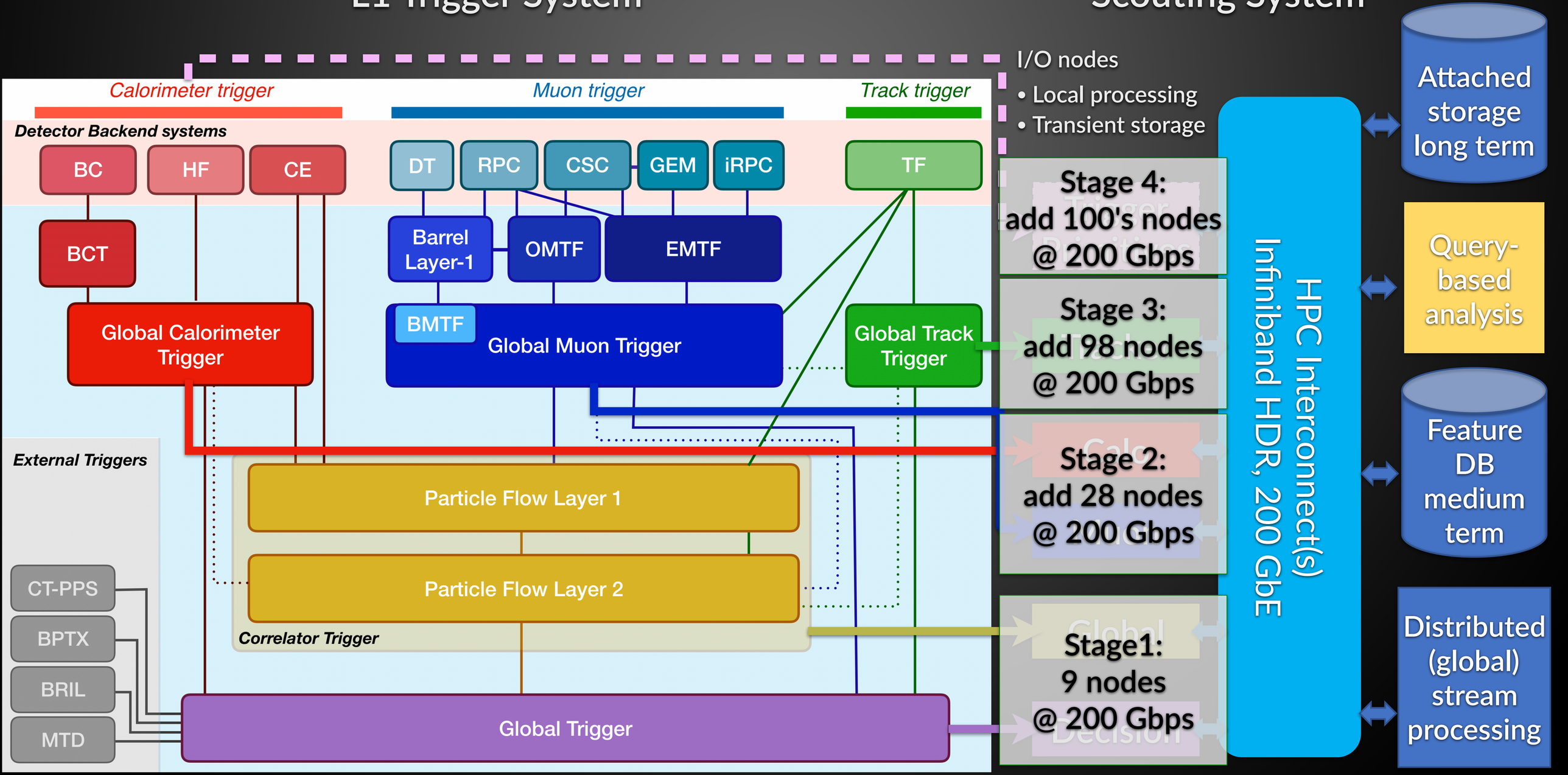
L1 Trigger System

Scouting System



L1 Trigger System

Scouting System



GMT scouting prototype in Run 2

Global Muon Trigger Scouting in Run 2

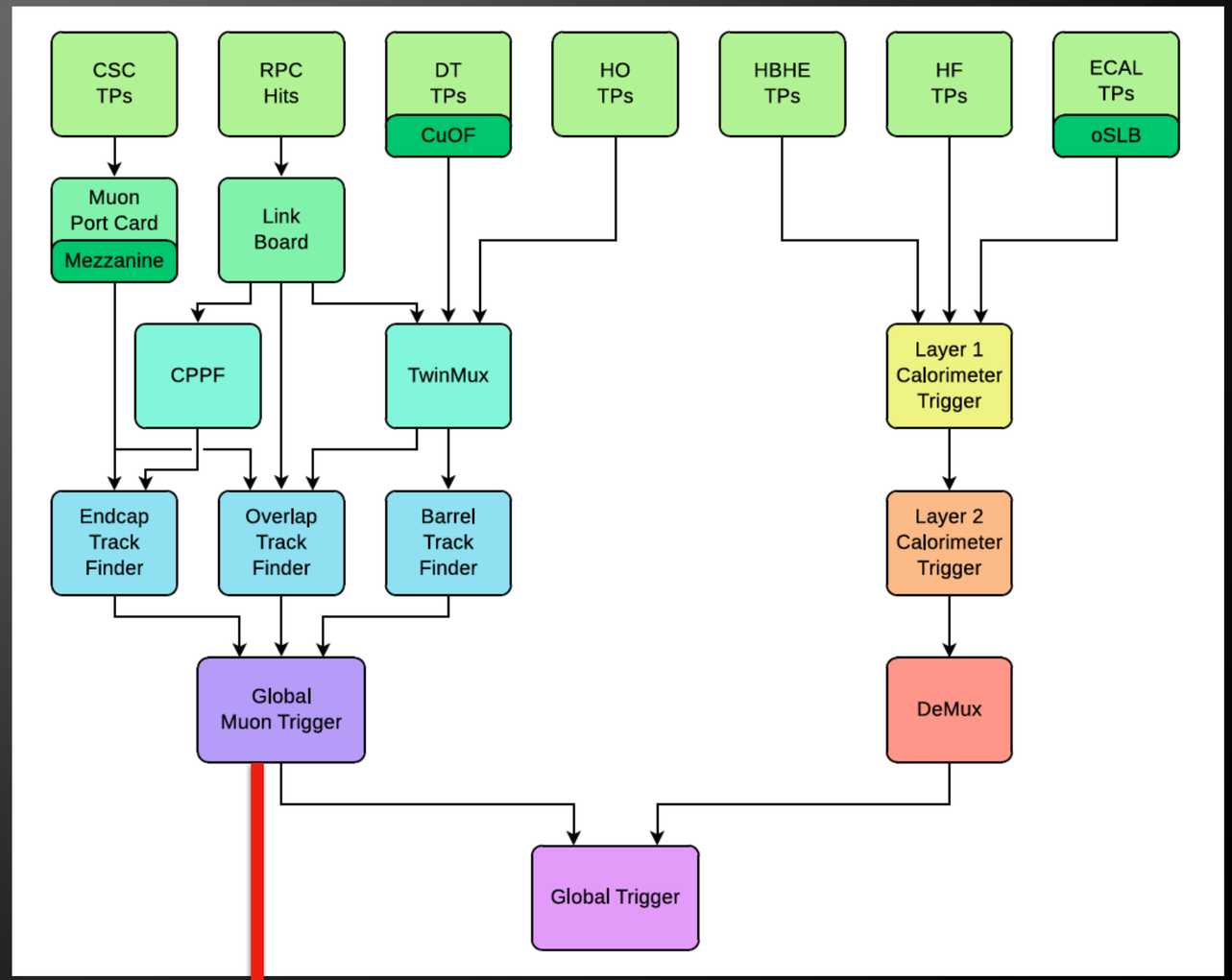
When: Oct / Nov 2018

Types of runs:

- 1 week of pp run
- Large part of HI run

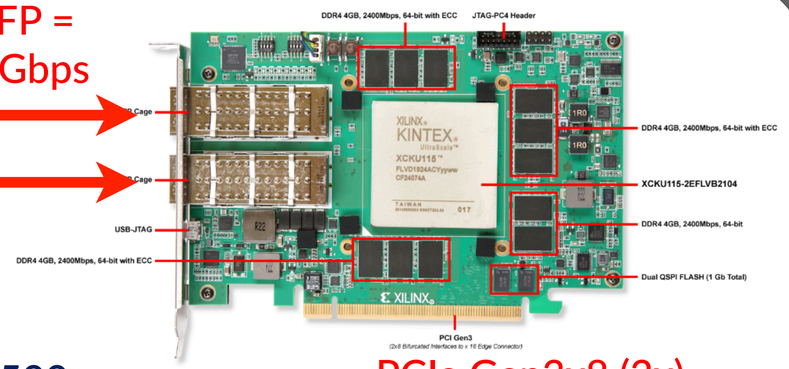
Capture @ 40 MHz

- Up to 8 final muon candidates
- Up to 8 intermediate muon candidates from barrel region
- GMT adds bunch and orbit counters



Global Muon Trigger (GMT) Scouting Prototype

2x QSFP =
8x 10 Gbps



KCU 1500
Xilinx Kintex Ultrascale 115

PCIe Gen3x8 (2x)

optical
10 Gb/s
from
GMT

8x

8 GB/s

KCU1500
firmware ZS
(1/20)

DMA

PCIe
Gen3

Software ZS
(1/8)

RAMdisk
mount

10
Gbps
NIC

10/40
Gbps
switch

40
Gbps
NIC

RAM
disk

BZIP (1/2)

RAID
8 TB

Infini
band
NIC

Lustre

max 800 MB/s

max 100 MB/s

Dell R720

max 50 MB/s

Dell R720

Controller PC
Firmware update & monitoring

1.1 TB/24 hour beam day
after compression in pp @2E34

uGMT scouting in action

Data collected in the last week of pp running

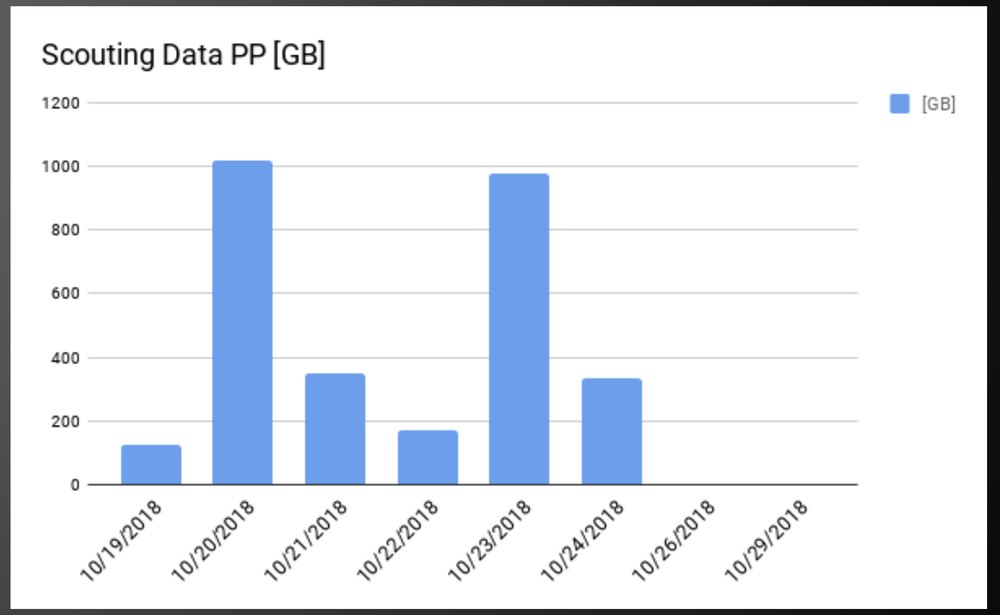
- Online zero suppression to variable-size block (x8 compression)
- Bzip2 to disk (~x2 compression)
  - About 2.1 GB per 1/pb
- Experimental setup, captured ~50% of data

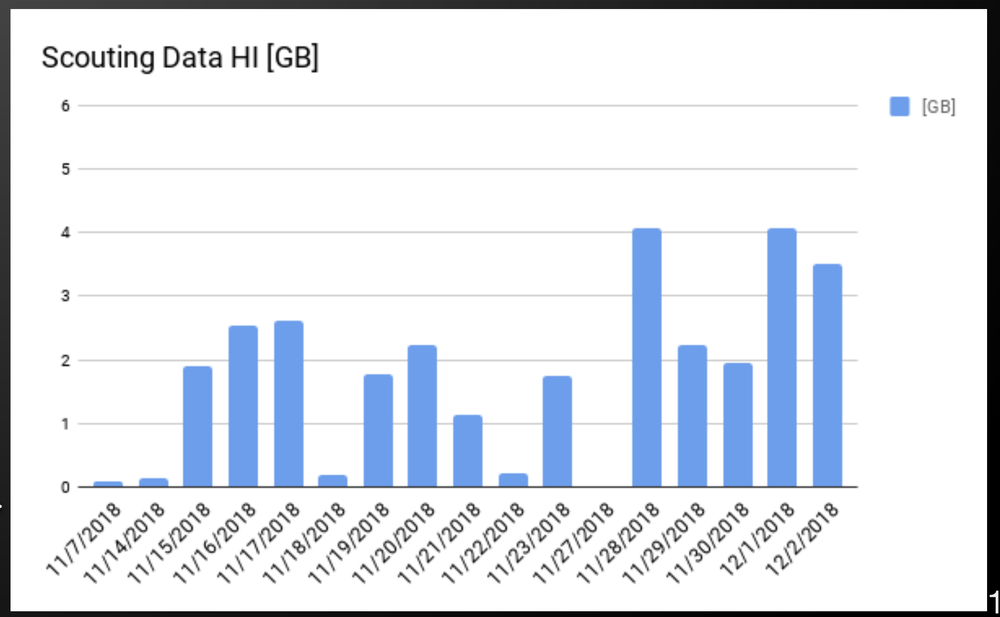
…and for the entire HI run

- About 28 MB per 1/ub
- Large contribution from cosmics

About one trillion non-empty BXs collected

- About 1 in 20 non-empty BX in pp





<31…eta extrapolated…23><22…quality…19><18…transverse momentum…10>
<9…phi extrapolated…0><31…reserved…30><29…eta…21><20…phi…11>
<10…index bits…4><3charge valid><2charge><1…iso…0>

LHC Emittance scan analysis

Emittance scan = method to determine beam overlap

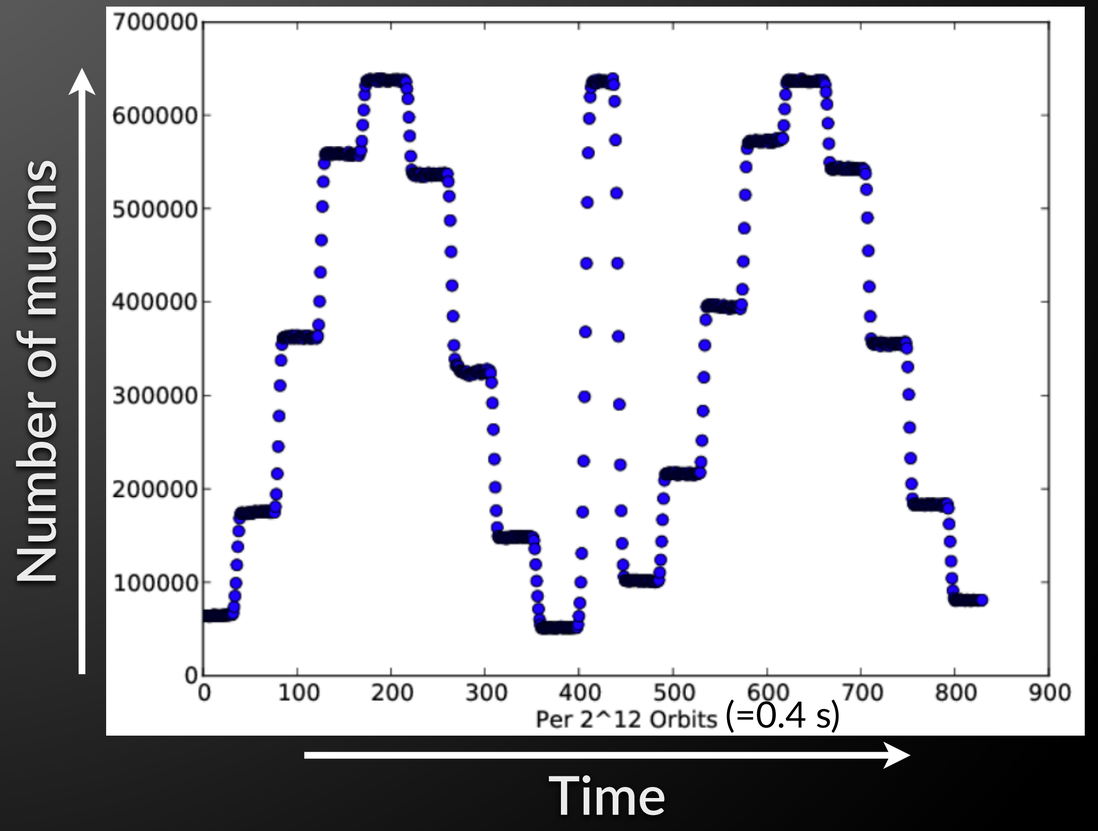
Beams moved in x (or y) w.r.t. each other

Measure interaction rate by counting muons from GMT 40 MHz scouting

- High statistic needed for per-bunch crossing analysis

Results from GMT scouting compatible with other luminometers

Fill 7333 “late” emittance scan

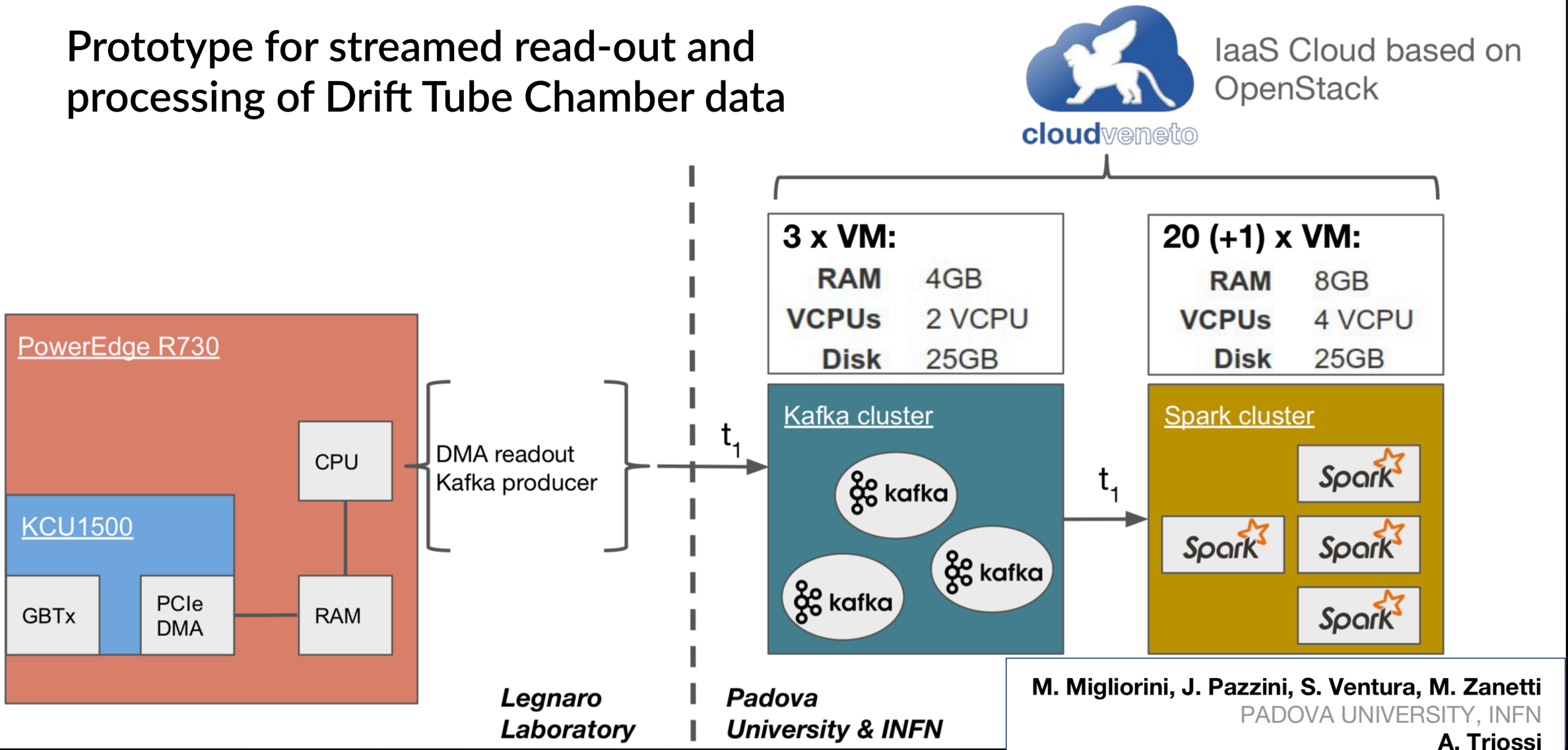


Tim Brueckler et al. (BRIL)

Stream processing prototype (Legnaro / Padova)

Stream processing: Apache Kafka & Spark

Prototype for streamed read-out and processing of Drift Tube Chamber data



M. Migliorini, J. Pazzini, S. Ventura, M. Zanetti
PADOVA UNIVERSITY, INFN
A. Triossi
IPHC STRASBOURG

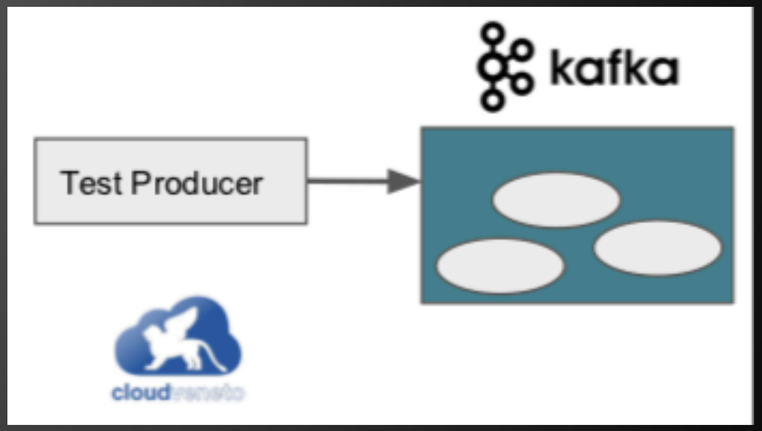
Measuring the Throughput with Kafka

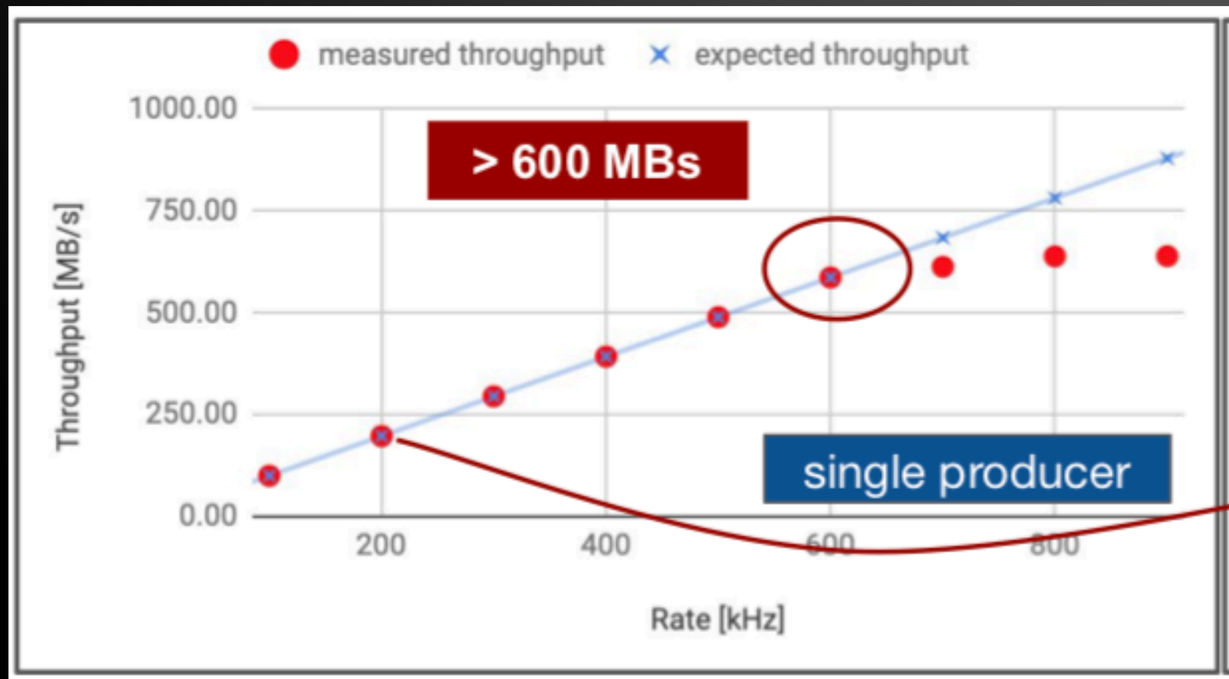
Using Kafka Java producers from a cloud node

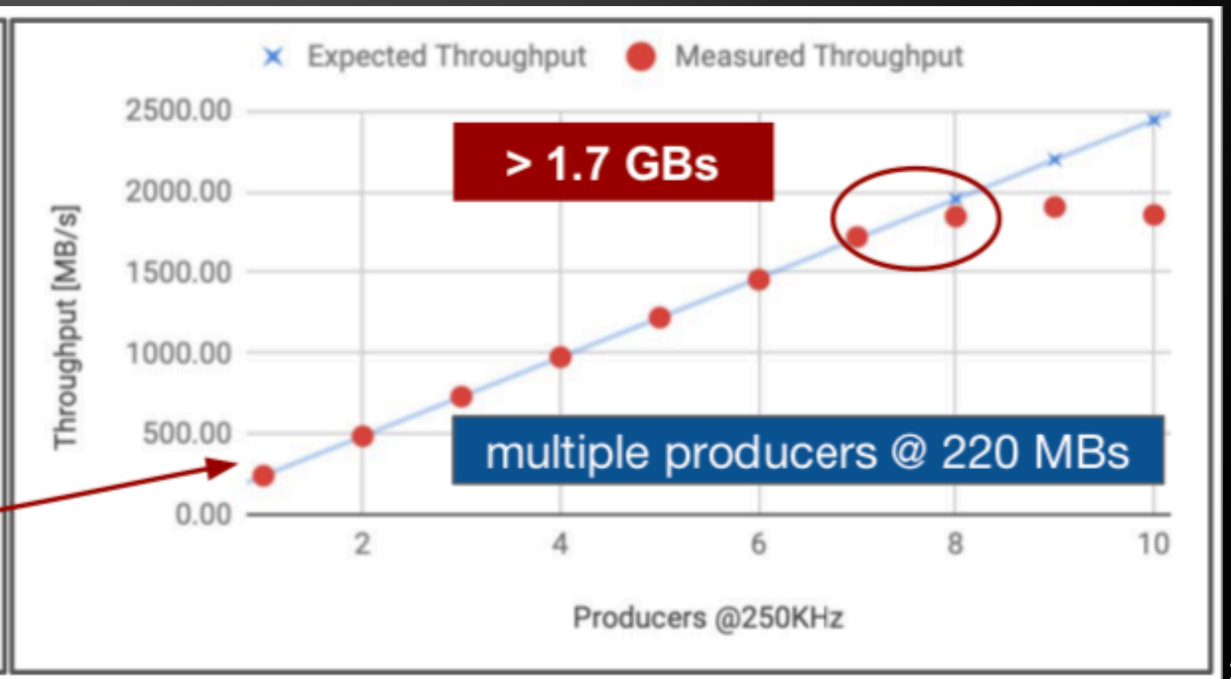
- 1 single producer (equivalent to 1 KCU in current setup)
- Multiple producers

Topic with 80 partitions on 3 brokers

Stream processing being optimized







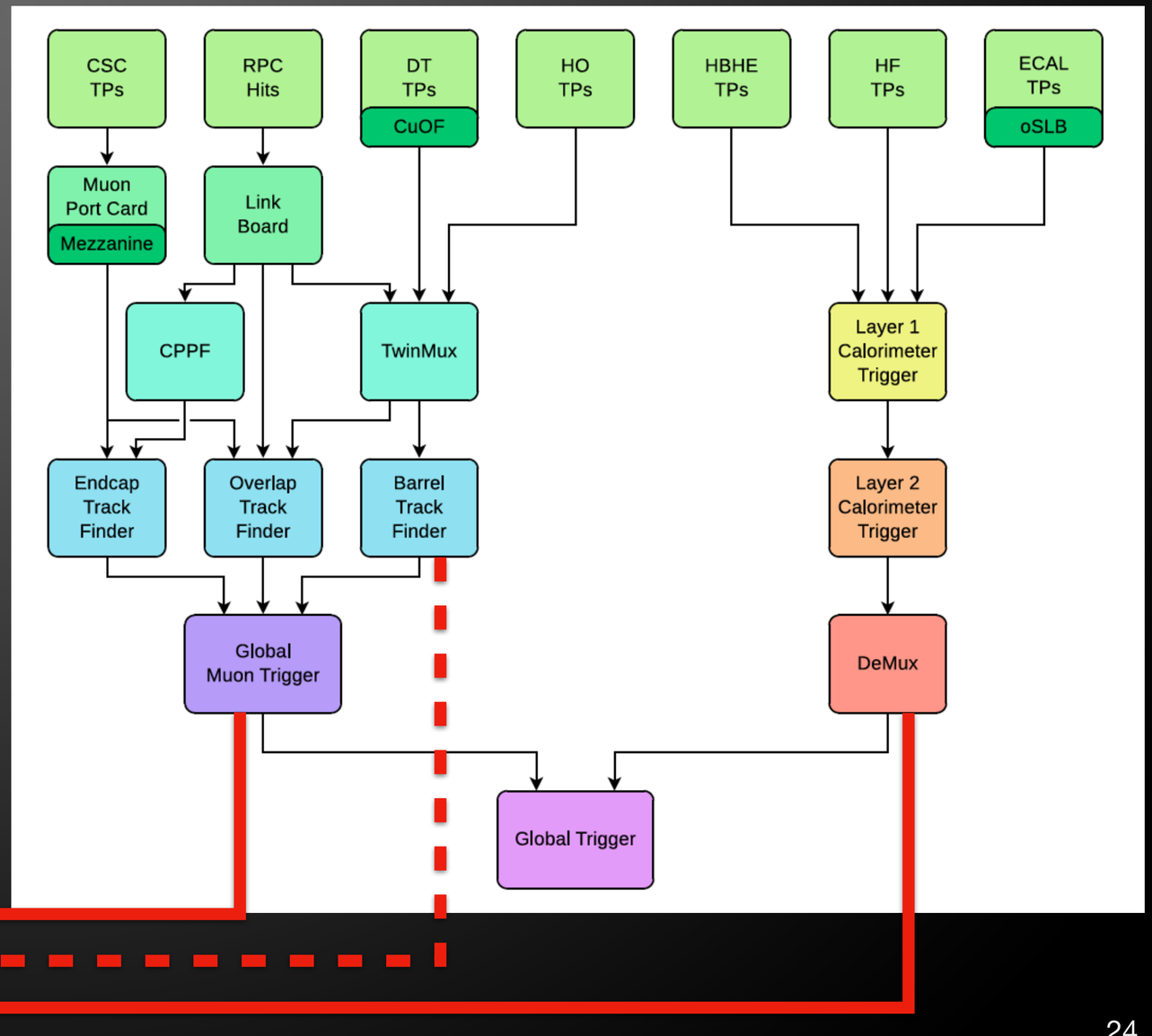
GMT + Calo scouting prototype for Run 3

Plan for Run 3 (2021): Muon + Calo Scouting

When: starting 2021

Capture @ 40 MHz

- Up to 8 final muon candidates
- Barrel Muon Kalman Filter muons (displaced Muons)
  - Through GMT or directly
- Calorimeter objects: jets, e/g, sums

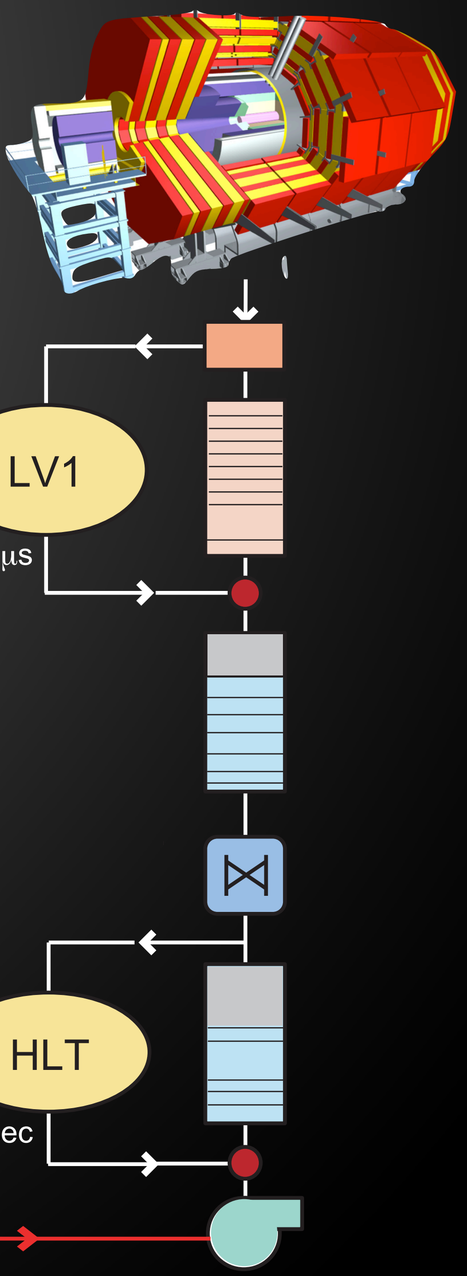


High-Level Trigger

Scouting at HLT for Run 3

CMS will continue to use scouting technique in run 3

- Constant luminosity during most of the fill
  - No longer have spare bandwidth & CPU as luminosity goes down
- GPUs available on all HLT nodes
  - Allows for more objects to be reconstructed on HLT
  - Full pixel tracking for all events
  - Enables more particle-flow algorithms to be run online
  - Opens the door for deep-learning applications on HLT
- Detailed plan is being worked out



Other Possibilities for HLT Scouting

Plenty of disk space on local HLT machine

- Could store some pre-selected events on local disk as long as there's space
- Run analysis on these events during interfills or technical stops when CPU is available
- Bookkeeping of number of events/recorded luminosity is challenging
- Needs analysis topics which are insensitive to delivered vs. recorded luminosity

Large buffer space in event-builder would allow to delay HLT selection

- Idea of a large key-value store (DAQDB) for event building pursued as openlab project
  - Based on relatively low-price large 3D XPoint memory pool
    - Store events for a few hours before final HLT selection
    - Could allow to run prompt calibration
    - More precision for 2nd stage selection and real-time analysis on HLT
- Could complement L1 trigger scouting by making full event available for selected L1 triggers

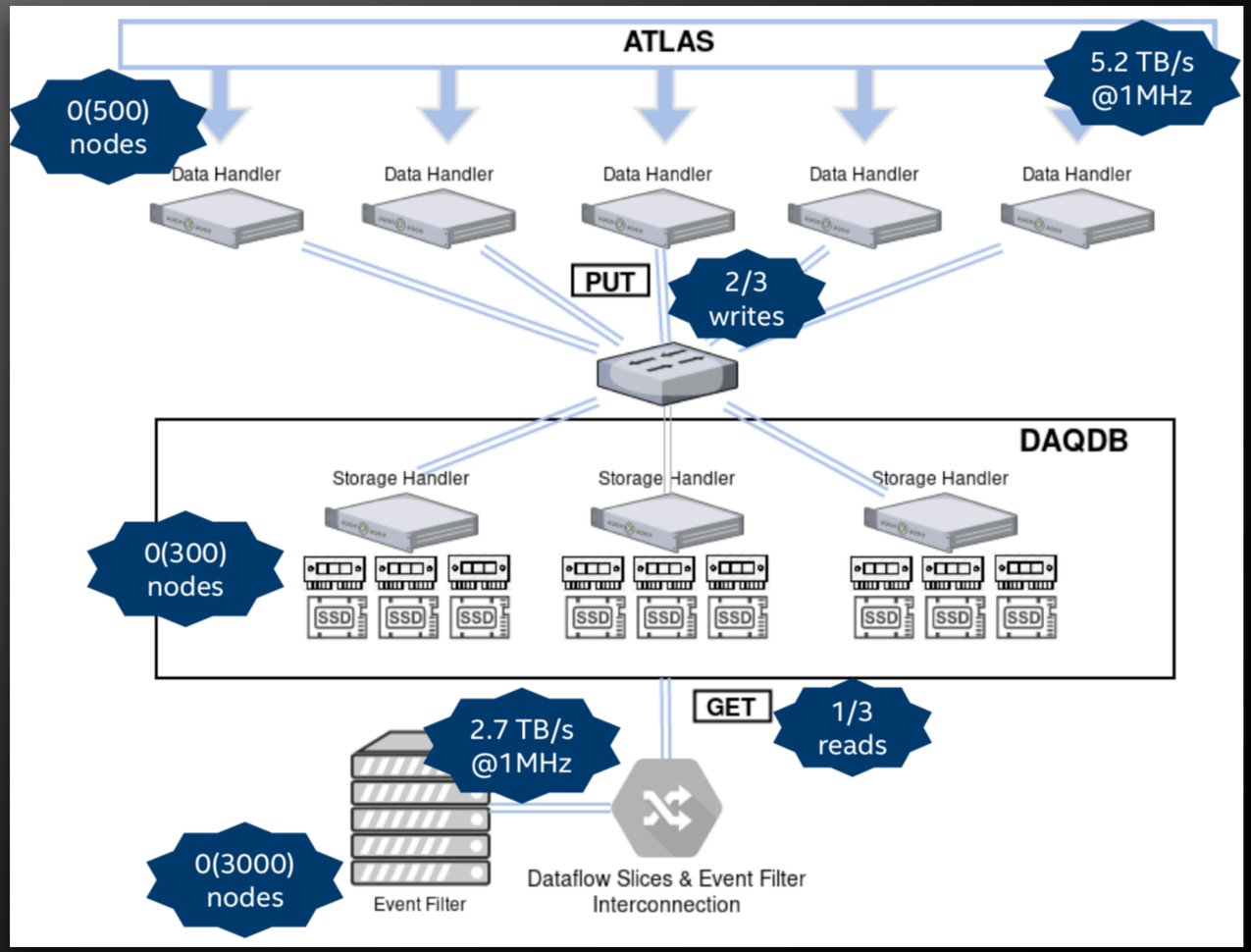
DAQDB being Integrated with ATLAS TDAQ

DAQDB

- Designed for Intel Optane Persistent Memory
- Data persistence with strong performance and affordable capacity
- Second-line NVMe-based storage to further extend the capacity
- Data structure based on Adaptive Radix Trie (ARTree) for efficient range queries
- DAQ-specific API featuring compound keys, range queries, and next event retrieval

Complete dataflow simulation

- Writer application with embedded DAQDB
- Client applications for getting fragments



Summary

CMS is planning a 40 MHz L1 Trigger scouting for system for HL-LHC

- Promising for physics (high resolution objects available at L1)
- Invaluable diagnostic and monitoring tool for the trigger
- Additional per-bx luminometer

Prototyped Global Muon Trigger Scouting in run 2

- Planning to capture all Global trigger inputs in run 3

HLT scouting technique will be expanded for run 3

R&D ongoing on various fronts

- HW inference engines
- Stream processing: e.g. Kafka / Spark prototype
- Distributed algorithms (MPI)
- NVRAM latency
- Searchable Feature DB
- Key-value store to assemble and buffer event fragments before HLT selection