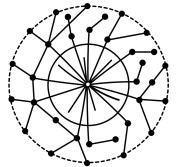
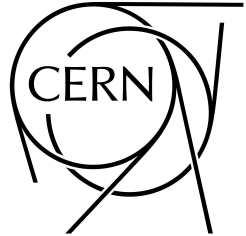
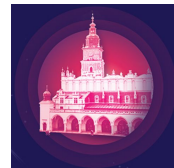




# An online data processing system for the CMS Level-1 Trigger data scouting demonstrator

**Matteo Migliorini**  
on behalf of the CMS Collaboration

**CHEP2024**  
**Krakow, Poland**  
October 19-25, 2024



**NextGen**

# Trigger system of the CMS experiment

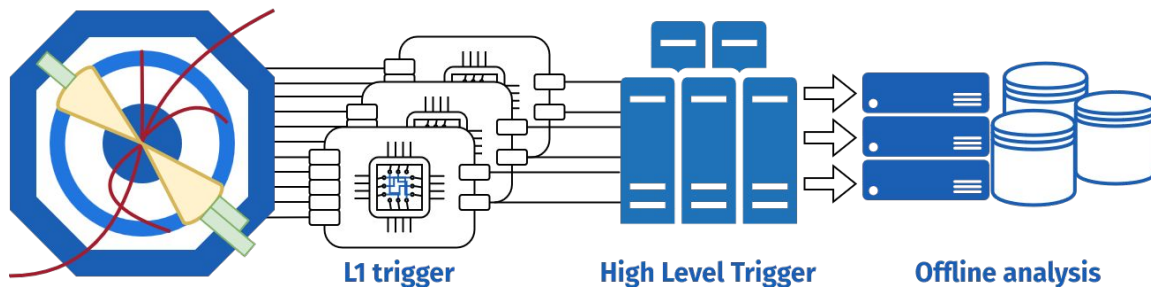
Compact Muon Solenoid: general purpose detector at the LHC

Offline analysis based on the collision events selected online

- Impossible to store everything!
- Trigger system selecting a small fraction of interesting events

⇒ Allowed to successfully probe many aspects of the Standard Models

... but is there a chance we are missing something?



## L1 Trigger

Hardware reconstruction of events based on reduced set of information and granularity  
⇒ **selecting 100kHz** of events

## High Level Trigger

Software based reconstruction using full detector granularity  
⇒ **~1kHz** of events stored for offline analysis

Working at LHC bunch-crossing rate of **40MHz**

## Goal of L1 Scouting

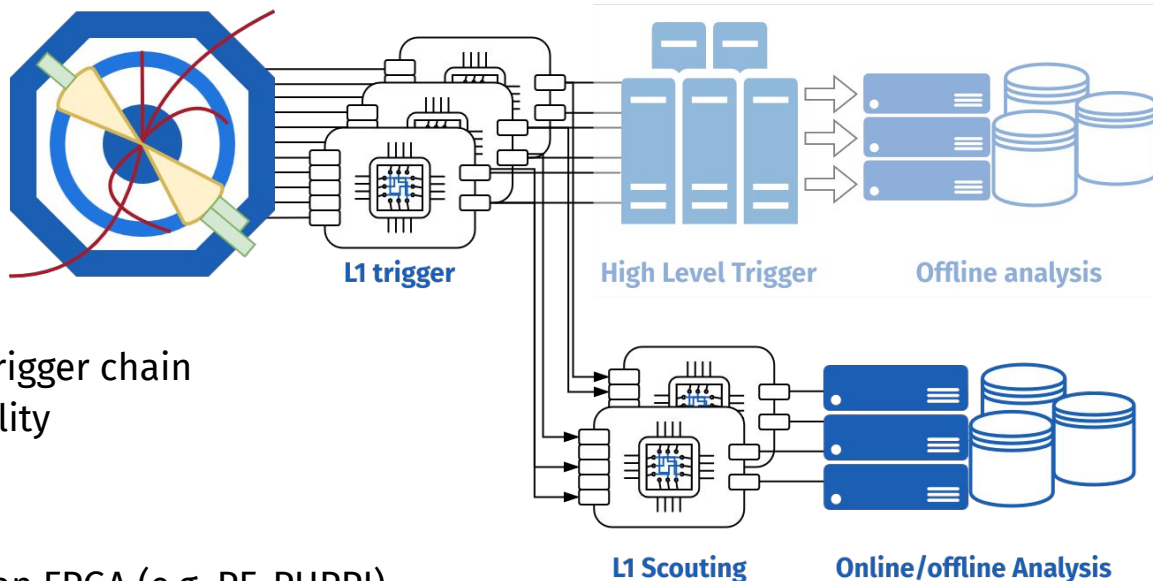
Capture objects reconstructed in the L1 trigger at the bunch crossing rate of **40MHz**

- ⇒ Capture signatures evading the trigger chain
- ⇒ Tradeoff with reconstruction quality

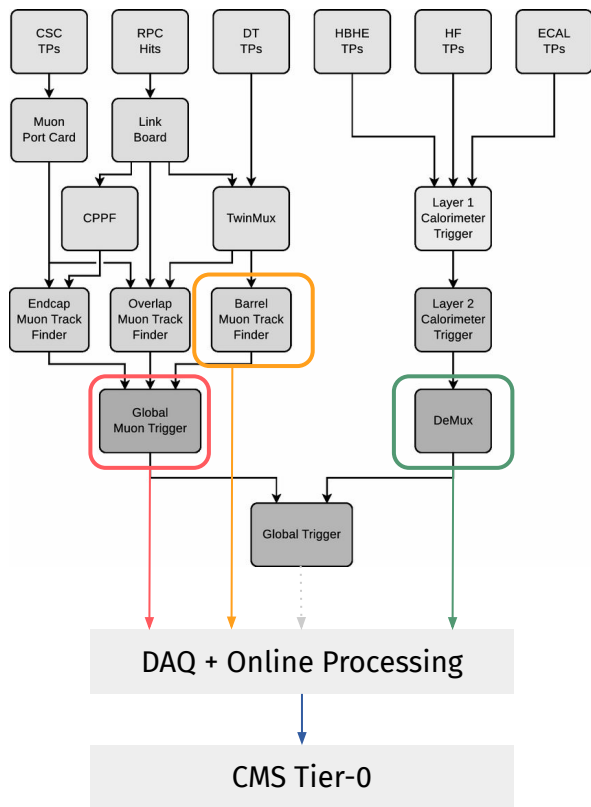
**Target:** Phase-2 upgrade L1-trigger

- Advanced object reconstruction on FPGA (e.g. PF, PUPPI)
- Dedicated poster: [CMS L1 Data Scouting at HL-LHC](#)

**Currently:** Developed a demonstrator system during Run 3



# L1 Scouting Run 3 demonstrator



Developed demonstrator during LHC Run 3

- Gain experience in both implementation and operation

Collect Trigger Primitives from the current system:

- Muon Stubs from Barrel Muon Track Finder
- Muons from the Global Muon Trigger
- Jets,  $e/\gamma$ ,  $\tau$  and energy sums from the Calorimeter Trigger

## Goals of the demonstrator

- ⇒ Collect **all** primitives @ 40MHz
- ⇒ Online processing and analysis streams creation
- ⇒ Transfer results to Tier-0 for data distribution

# Overview of the demonstrator

## FPGA boards concentrating L1 Trigger links

- Zero suppression and transmission to processing farm via 100Gb TCP/IP

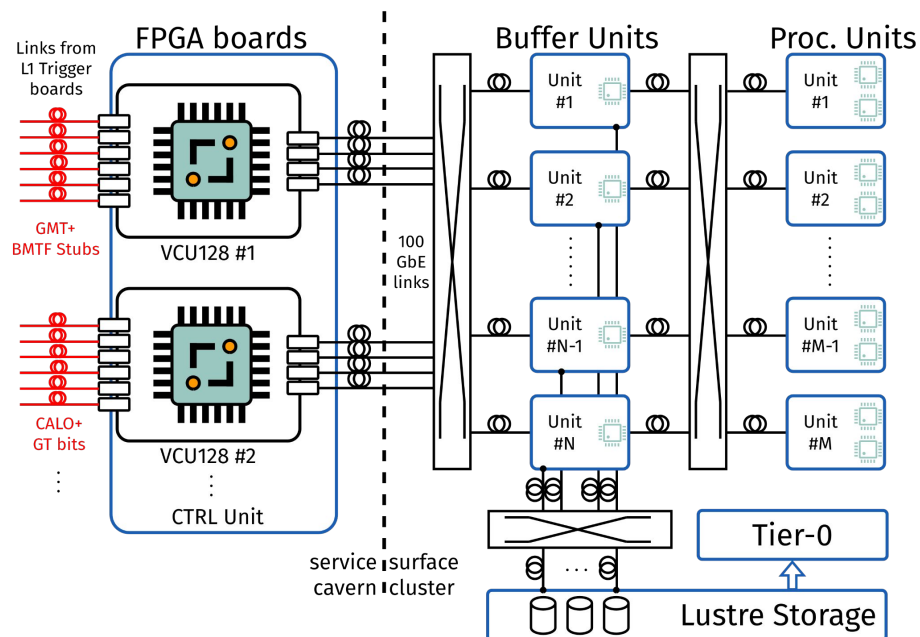
## Buffer Units receives TCP streams

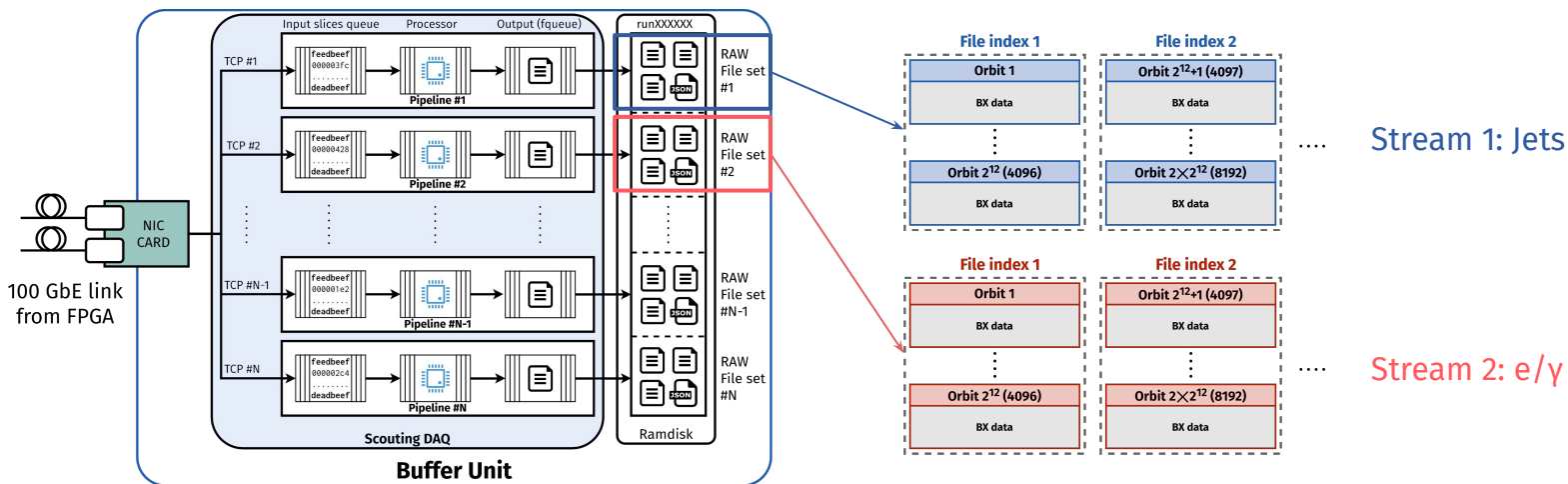
- Temporarily store raw data in ramdisk

## Processing Units

- Merge stream fragments from sources
- Online processing and stream creation

Results transferred to Tier-0, available for offline analysis





## Intel TBB-based DAQ software

- Running on the Buffer Units
- Every TCP stream processed by a pipeline
- Reformatting of input packets and writing raw data file in the ramdisk

## Each stream is split based on a common counter

- LHC Orbit counter (+1 every 3564 BX, ~90μs)
  - **Identical for all subsystems!**
- Data from multiple streams will be merged based on the orbit counter

A **unique file index** is assigned to each process

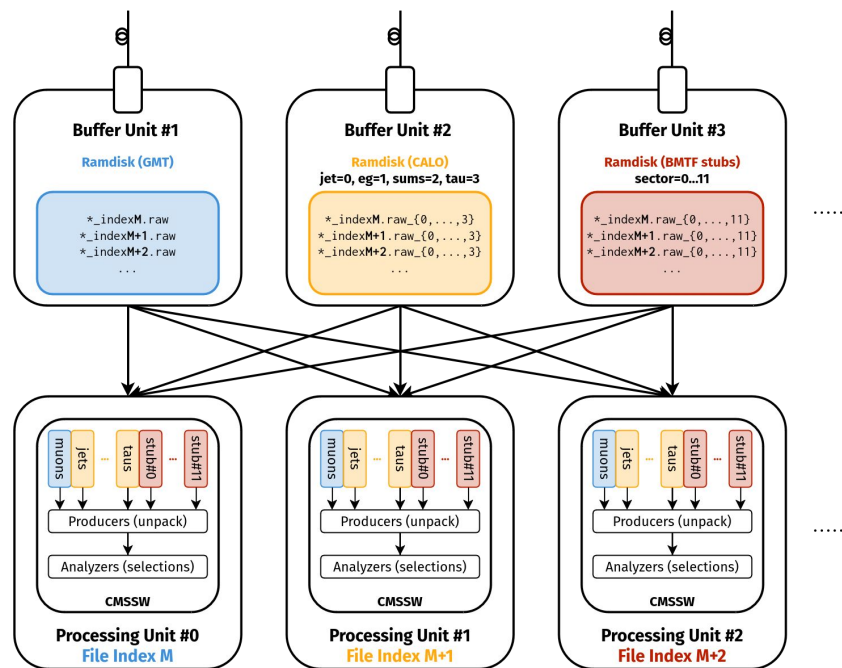
- Read raw data with the same file index from all the buffer units
- Guaranteed to contain the same set of orbits!

Orbit building: merge raw data corresponding to the same orbit from all data stream

- **Orbit based processing**  $\Leftrightarrow$  **single event (BX)**
- Performed at fixed rate of 11kHz

Implemented using standard reconstruction framework (CMSSW)

- Similar to HLT  $\rightarrow$  allows to reuse existing components and infrastructure



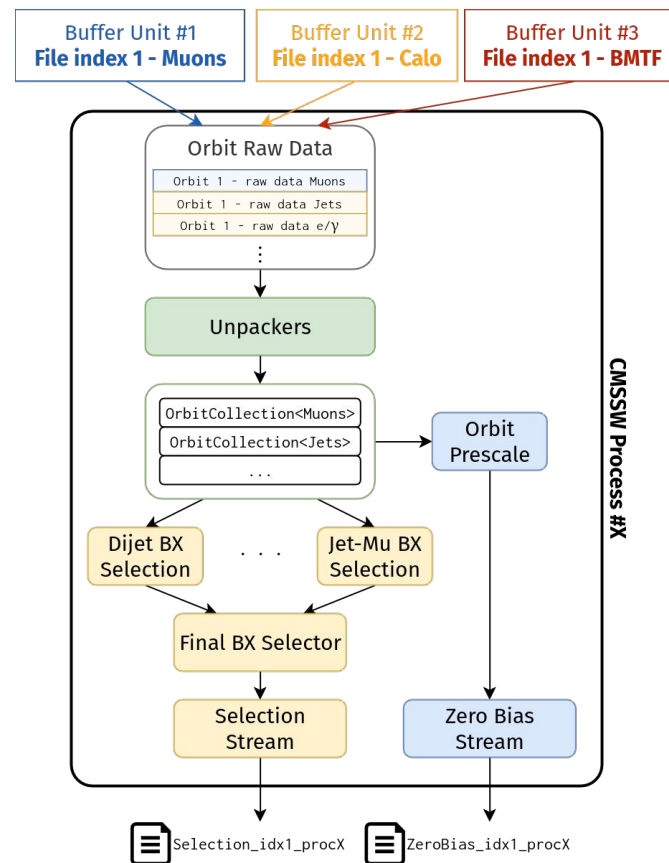
Starting from OrbitRawData created during the orbit building

- **Unpackers** interpret raw data and create OrbitCollections for each object, i.e. data from all 3564 BXs

Two distinct data-paths

- **Zero Bias**: store all collections for a fraction of the orbits
- **Selection**: no prescale on the orbit, but store only selected BXs and objects
  - Online analysis targeting signatures such as dijets, ecc.
  - FinalBxSelector removing duplicates

Results for each input file index stored in custom raw data format in the Processing Units local disk





Data from all the Processing units merged before transfer to Tier-0

- One file per “Lumisection” (LS)
- $2^{18}$  orbits  $\sim$  23.3s of data taking

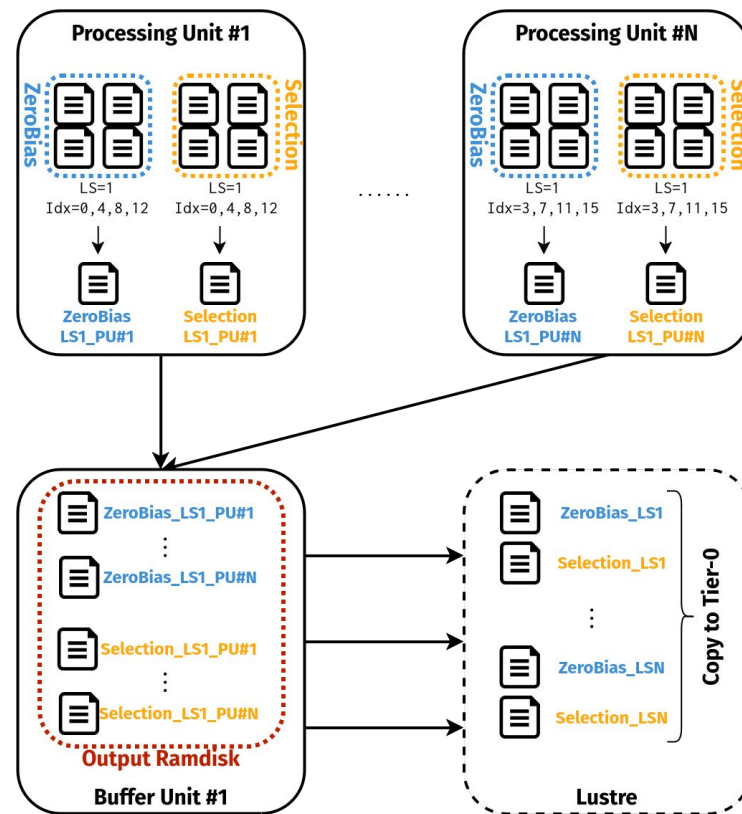
Same architecture as CMS central DAQ:

⇒ Local merging in each PU of files per LS

⇒ Moved to Buffer Unit output ramdisk, merge results from all PUs and moved to Lustre FS

⇒ Copied to Tier-0

- Perform “repacking” of the raw files
- Convert to ROOT-based format for persistence



System integrated into CMS Run Control system, successfully operated in parallel since the beginning of 2024 LHC pp run

- Final **Selection stream** deployed in June, **over 60fb<sup>-1</sup>** of unprescaled data collected with it!
- Prescale of **ZeroBias stream** gradually increased from 1 to 15, collected **≈10fb<sup>-1</sup>**

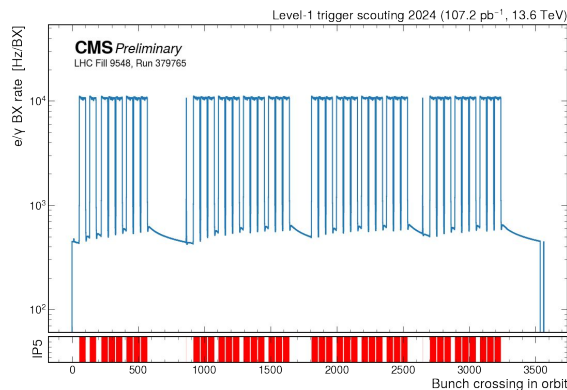
During a typical 2024 LHC fill at peak lumi  $L_{inst} = 2 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$ , PU~63

- DAQ software receiving 6 GB/s and writing around 4 GB/s to Buffer Units ramdisks
- Files consumed by 5 Processing units (x2 AMD EPYC "Milan" 7763 CPUs)
- Throughput of ~ 275 MB/s repacked data (x1.5 smaller than streamer files)

⇒ Stored approximately 1.2 PB available for offline analysis

- 53% Zero Bias and 47% Selection stream

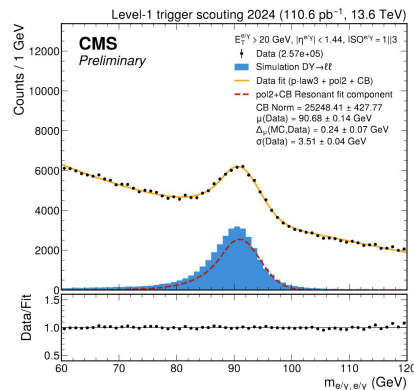
# Preliminary results



**Zero Bias:** access to all bunch crossings

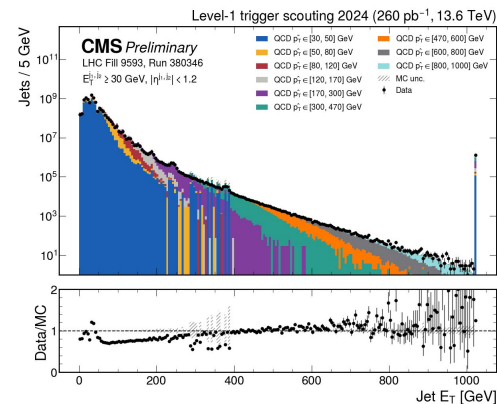
Rate of BX with with e/γ (1970 colliding BXs)

- ~11kHz for colliding BX ⇒ every BX contains at least a TP!
- Monitoring at the full BX rate



**Selection stream:** individual bunch crossings

- Data well modeled by the L1T simulation
- SM well known resonances visible (Z→ee)
- Preliminary studies of dijet events
  - Low thresholds on jets, low mass dijet search!



Implemented and operated a demonstrator of the CMS L1 Trigger scouting system

- Successfully collected data from the Run 3 trigger during to 2024
- Exploring new technologies: [Real-time Level-1 Trigger Data Scouting at CMS using CXL Memory Lake](#)

Beginning the analysis of this uncharted data ⇒ **a measurement will complete the demonstrator!**

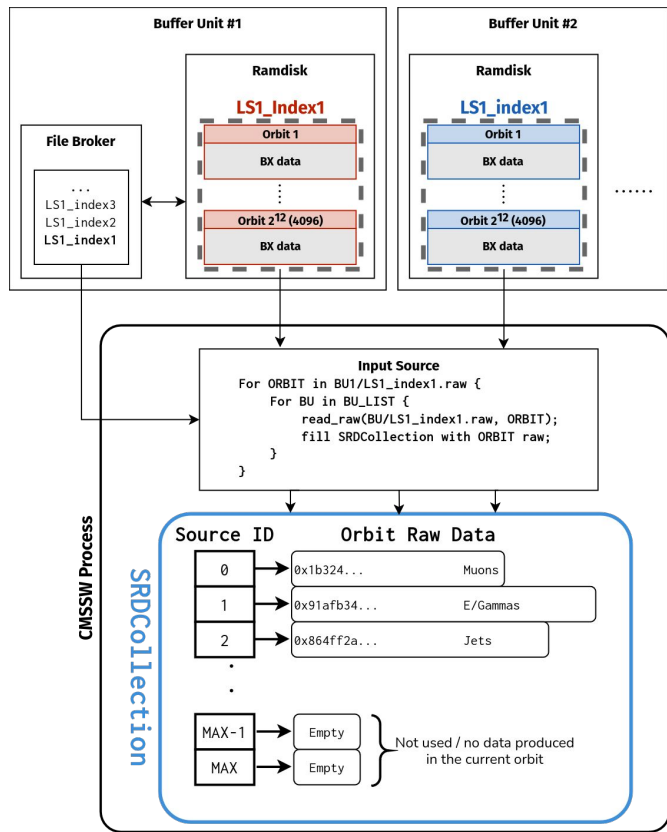
- Low mass dijet search, HSCP over multiple BX and other exocitic signatures
- ⇒ Stay tuned!

Building a demonstrator for Phase-2 system

- Based on the experience gained with the Run 3 demonstrator
- [Checkout the dedicated poster!](#)
  - From design to expected analysis performance!

# Backup

# Orbit Building steps



Ramdisk monitored by the File Broker (on one Buffer Unit)

- Checking when a new file appear in the ramdisk

When a CMSSW is IDLE it can request a new file

- E.g. File Broker provides LS1\_index1

InputSource will then read the content of files with name LS1\_index1 from all the RUBUs' ramdisks

- For every orbit (event header in the files) create a new CMSSW event
- Create a RawDataCollection and fill it with raw data from each source
- SourceID indicates the origin of the fragment's raw data

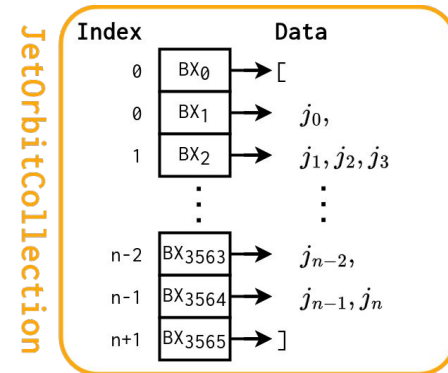
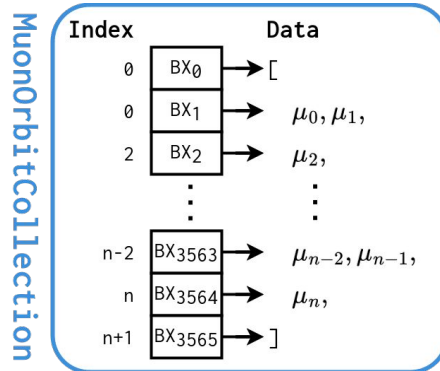
# Orbit collections

OrbitCollection<T> implemented using 2 “flat vectors”, Index and Data

- Conceptually equivalent to a collection with 2 nested arrays
- Outermost = BX, inner objects

Data contains a sequential list of objects for the entire orbit (i.e. all BX)

Index contains one entry per BX, where the value is the starting position index of the objects related to that BX stored in the data vector

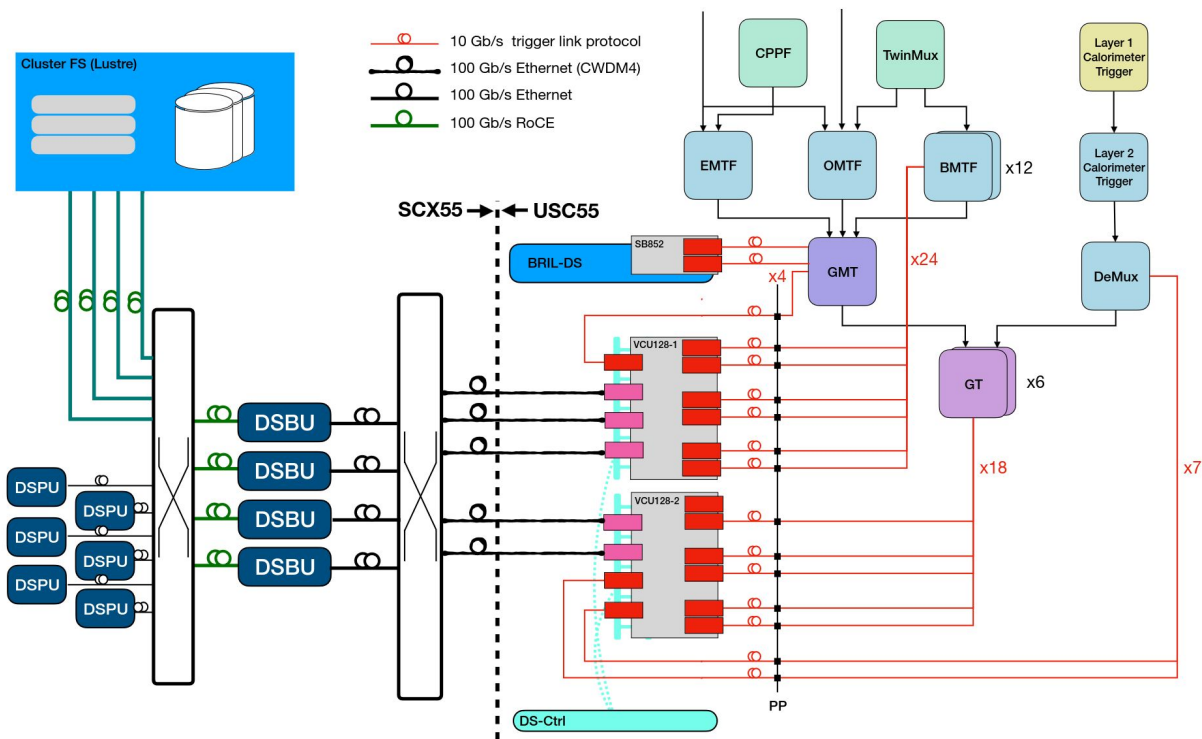


In the muons collection:

- Index at BX<sub>1</sub> is 0, 2 at BX<sub>2</sub>
- This means that there are two muons in BX<sub>1</sub>
  - Data[0], Data[1]

⇒ Content of BX<sub>N</sub> stored in data vector ranging from Index[BX<sub>N</sub>] to Index[BX<sub>N+1</sub>]

# L1 Scouting Run 3 Demonstrator



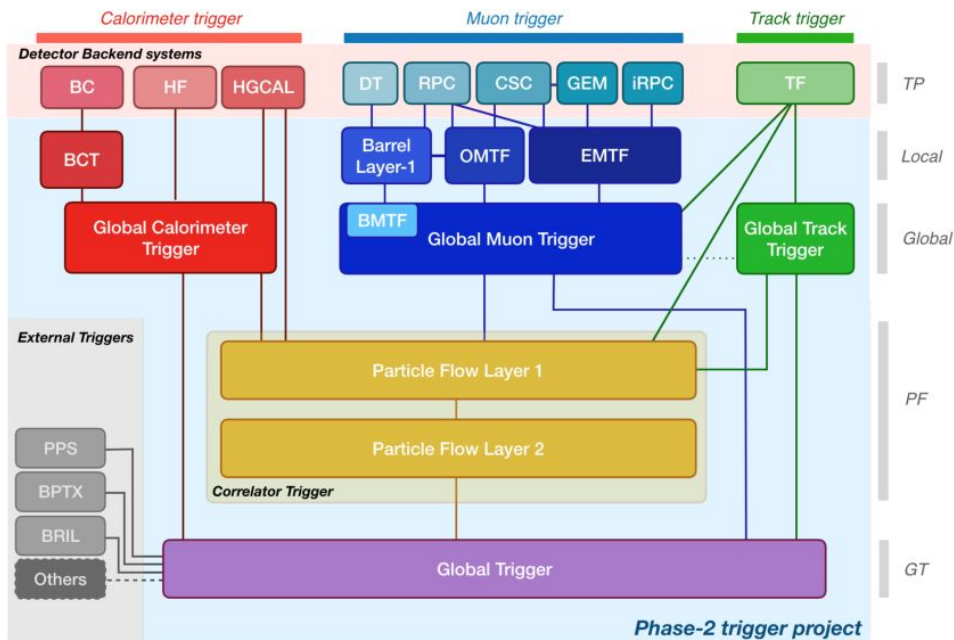
## Detailed view of the demonstrator components

- Two VCU128 boards used to concentrated trigger links
- Each hosts 4 on-board QSFPs and 6 mounted in n FMC mezzanine
- Similar characteristic as the phase-2 boards

Boards controlled via DMA



# Phase-2 L1 Trigger upgrade



⇒ Advanced object reconstruction on FPGA

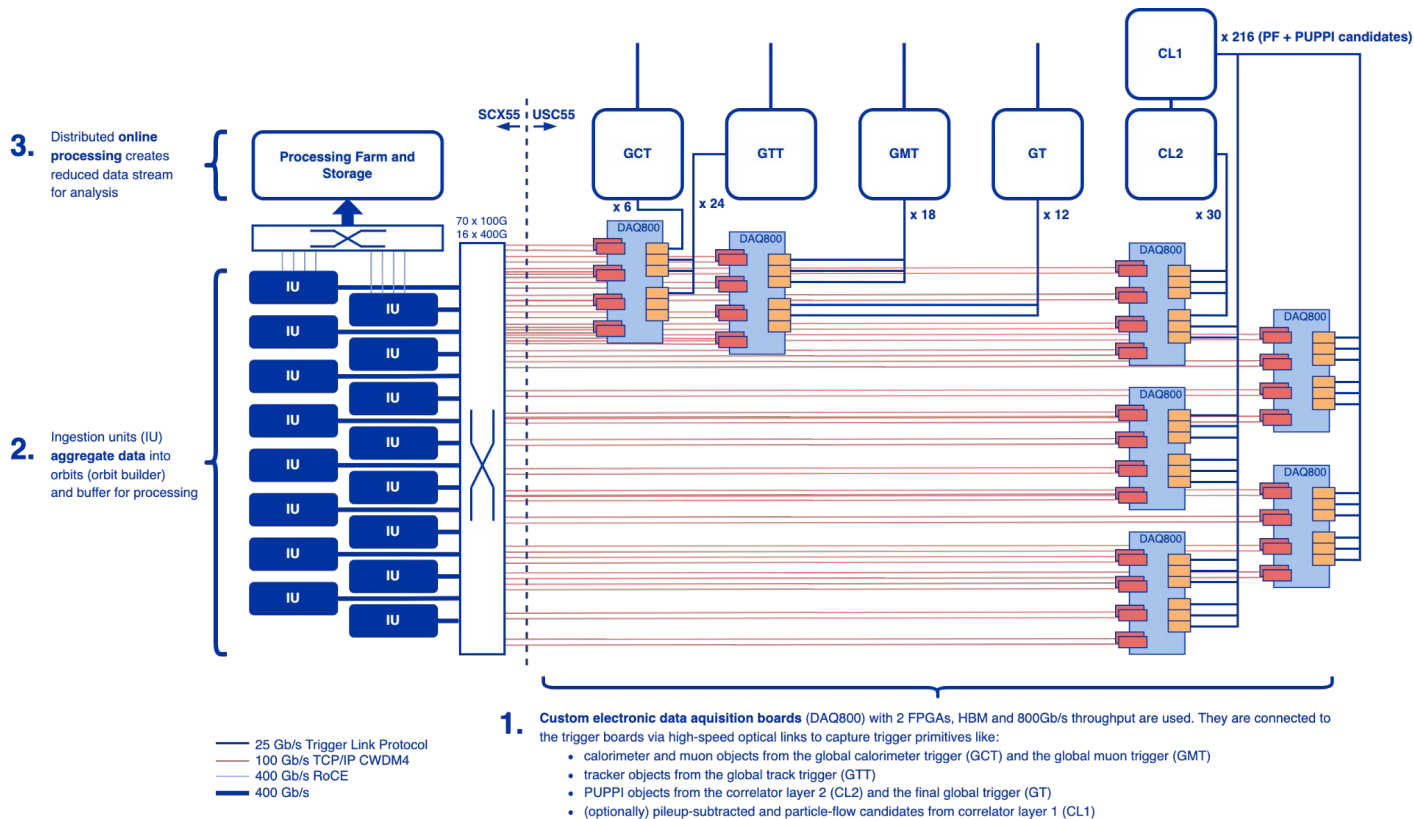
⇒ Global Calorimeter and Muon Trigger: higher granularity objects

⇒ Global Track Trigger: tracker tracks, vertex finding

⇒ Correlator Trigger: Particle Flow reconstruction

⇒ **Resolution often comparable to offline level**

# Phase-2 L1 Scouting system



Bunch crossing data selected and tagged if any of the following selection is valid:

“**Dijet30**”: At least 2 jets with  $E_T \geq 30$  GeV

“**SingleMu0 BMTF**”: At least a muon in the barrel

“**DoubleMu0Qual8**”: At least two muons with  $hwQual \geq 8$

“**High-Multiplicity Jet**”: At least 4 jets with  $E_T \geq 20$  GeV

“**Jet-Mu**”: targeting Heavy Flavour jets

- A jet with  $E_T \geq 30$  GeV
- A muon inside jet cone ( $\Delta R < 0.4$ )

⇒ **Covering large number of possible analysis and studies!**

Collection name	Average Size (kBytes/Orbit)	
	Uncompressed	Compressed
TauOrbitCollection	422	60
EGammaOrbitCollection	238	35
BxSumsOrbitCollection	226	30
JetOrbitCollection	103	15
MuonOrbitCollection	38	4
BMTFStubOrbitCollection	22	2

Collection Name	Average Size (kBytes/Orbit)	
	Uncompressed	Compressed
EGammaOrbitCollection	43	5
BxSumsOrbitCollection	35	4.1
JetOrbitCollection	31	3.3
MuonOrbitCollection	20	1.2
BMTFStubOrbitCollection	18	0.7
FinalBxSelector_SelBx	0.9	0.2
DijetEt30_SelBx	0.6	0.1
HMJetMult4Et20_SelBx	0.4	0.1
SingleMuPt0BMTF_SelBx	0.1	0.04
DoubleMuPt0Qual8_SelBx	0.1	0.04
MuTagJetEt30Dr0p4_SelBx	0.02	0.007

**Zero Bias:** 146 kBytes per Orbit

- Typical prescale ~15
- Throughput  $146\text{kB} \cdot 11\text{kHz} / 15 = 109\text{ MB/s}$

**Selection:** 15 kBytes per orbit

- No prescale & no taus
- Throughput  $15\text{kB} \cdot 11\text{kHz} = 165\text{ MB/s}$