The new ALICE DAQ system for LHC Run 3

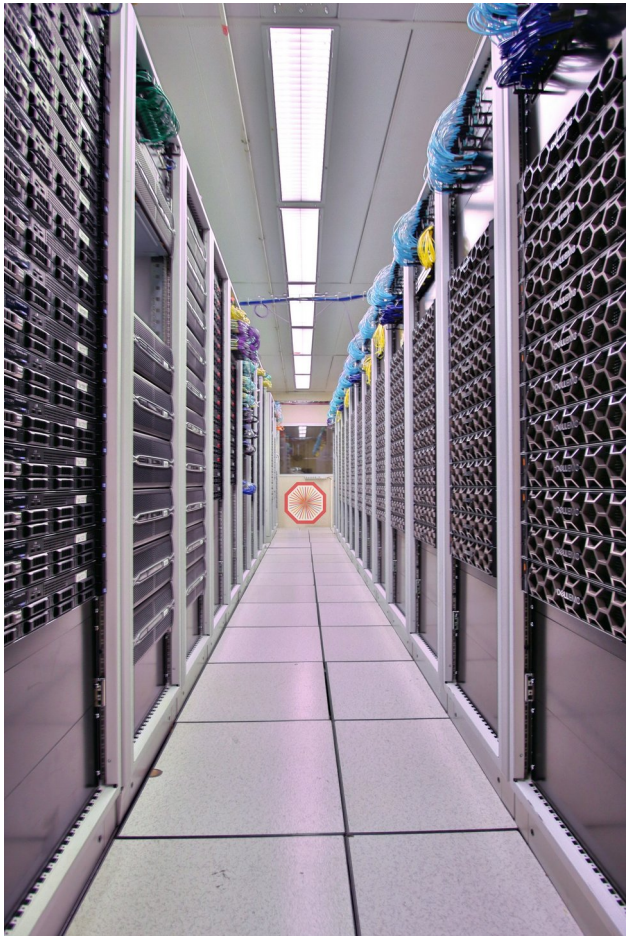
Sylvain Chapeland

for the ALICE collaboration

The 2022 International Workshop on the High Energy Circular Electron Positron Collider

Outline

- The ALICE detector
- Data Acquisition system (DAQ)
  - Architecture and implementation
  - Insights on the readout
- Commissioning

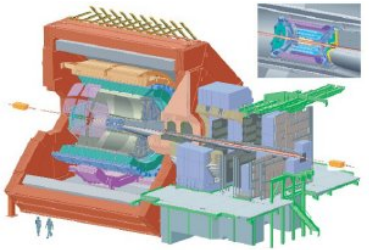


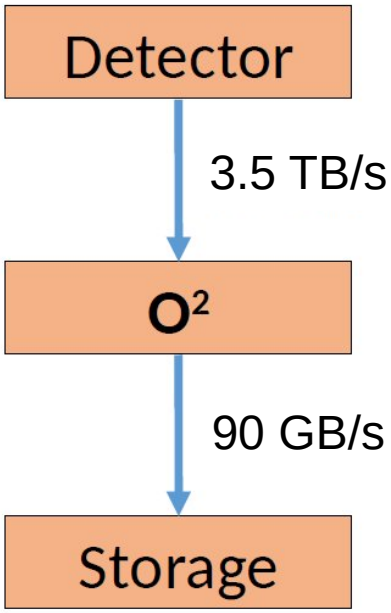
ALICE

- Upgrade highlights
  - Time Projection Chamber (TPC)
    36 GEM readout chambers, 500k channels, continuous readout
  - Inner Tracking System (ITS)
    CMOS chips in 7 layers, 12 Gpixels
  - Muon Forward Tracker (MFT), Fast Interaction Trigger (FIT)
  - 15 subdetectors in total

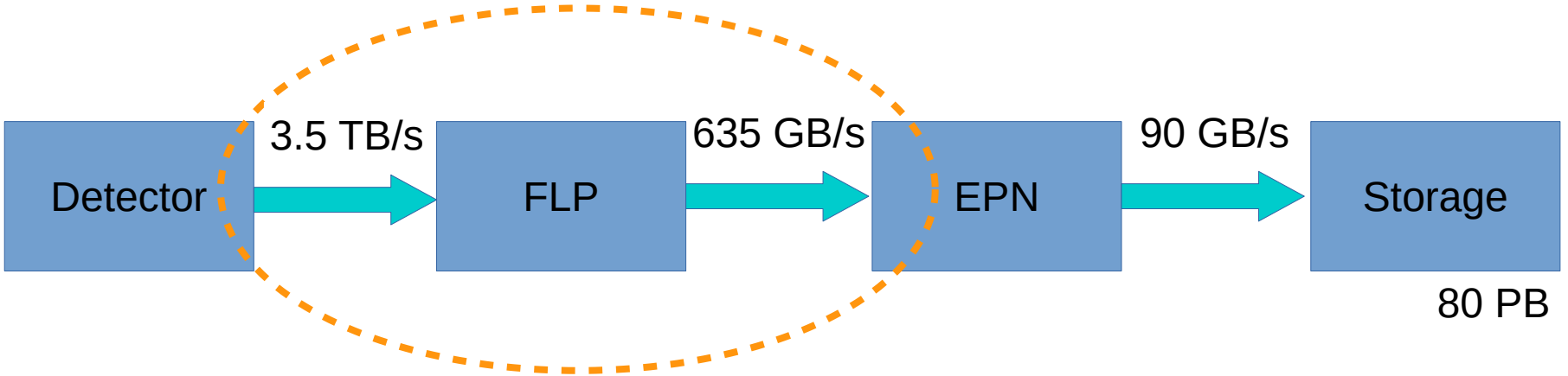
- Increased data throughput: *x100 vs LHC Run 2*
  - 3.5 TB/s from the detector
  - 8000 optical links

- Demanding online processing and compression
  - O2: online – offline system
  - Synchronous and Asynchronous processing





Data Acquisition (DAQ) - but not only, fuzzy limit with processing



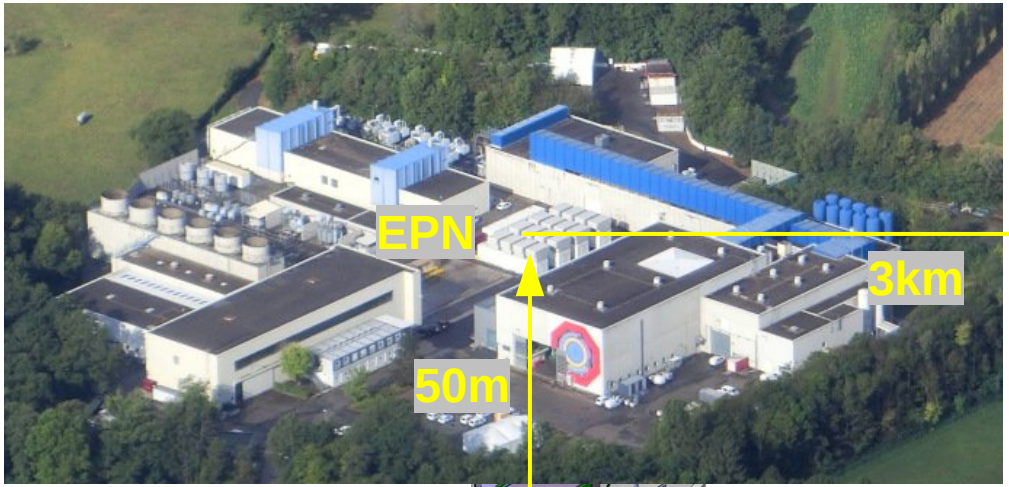
- Readout: First Level Processors (FLP)
- Reconstruction: Event Processing Nodes (EPN)
- Data reduction in 2 steps

Systems layout

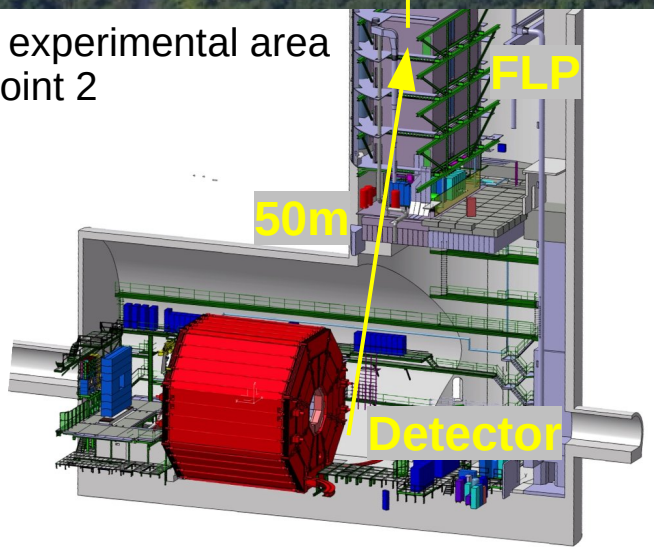
- Many technical choices driven by existing facilities constraints
  - Distance
  - Space
  - Power
  - Cooling

- Following slides to illustrate some of them
  - General to close-up views of CERN / ALICE / DAQ

Aerial view of LHC, CERN, and ALICE

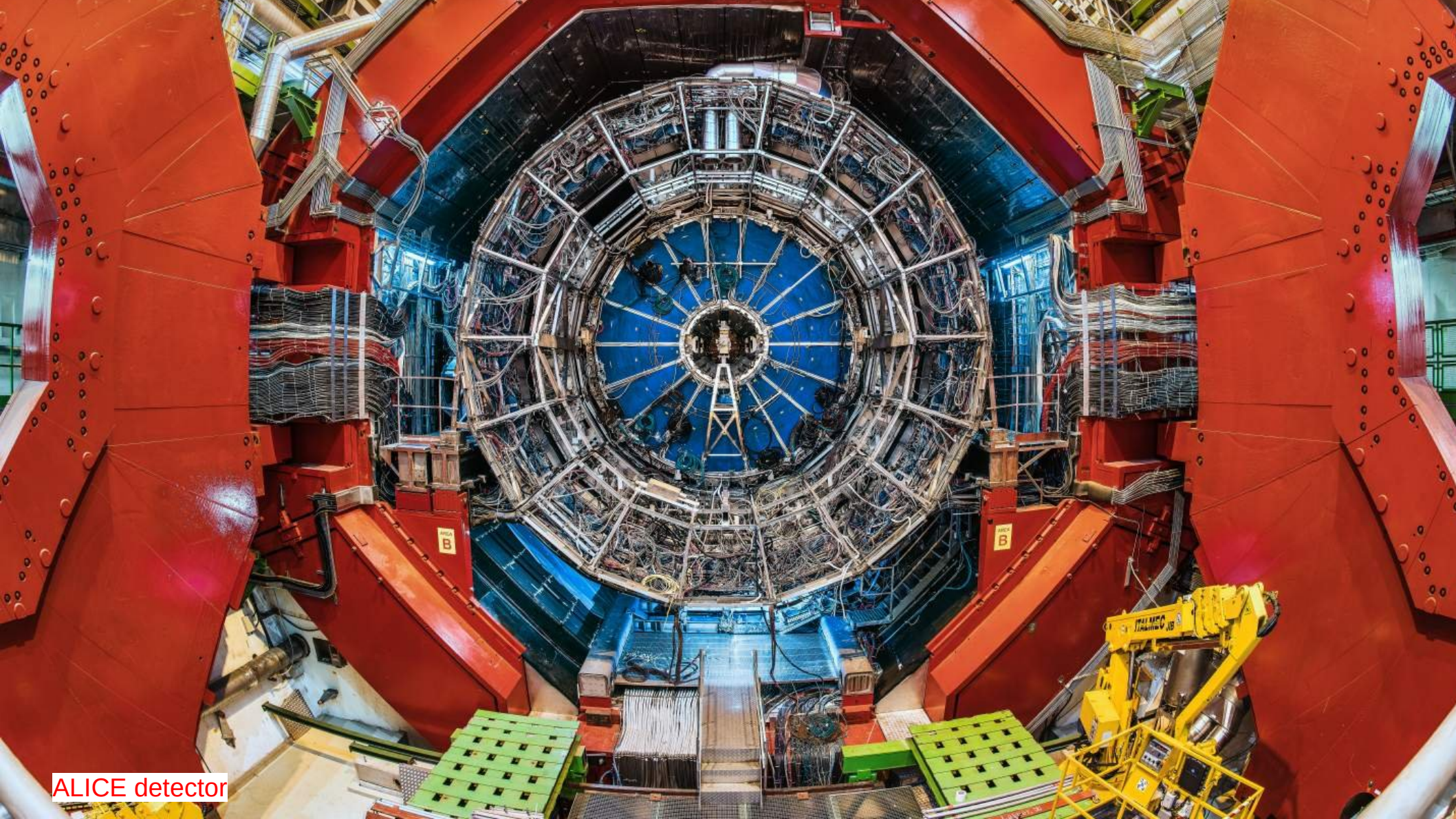


ALICE experimental area
LHC Point 2



CERN Computing Center

Distance between components
(actual links length are longer)


ALICE detector

FLP farm

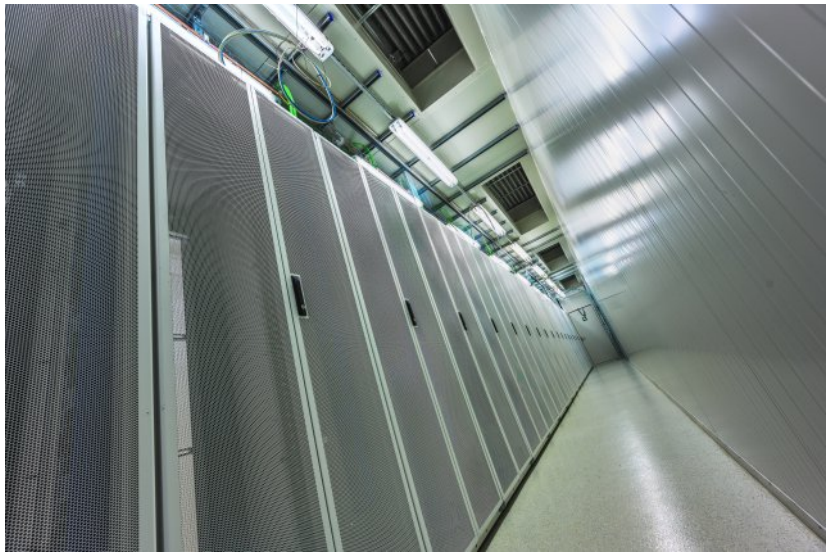
- Cold water cooling (rack doors with forced air flow)
- 200 servers, 500 readout cards (up to 3 per host)
  - Mostly 2x 10cores CPUs Xeon 4210, 96GB RAM
  - 25Gb ethernet + 100 Gb/s infiniband network to EPN

CR1, suspended in access shaft

Inside CR1, 180° panoramic view

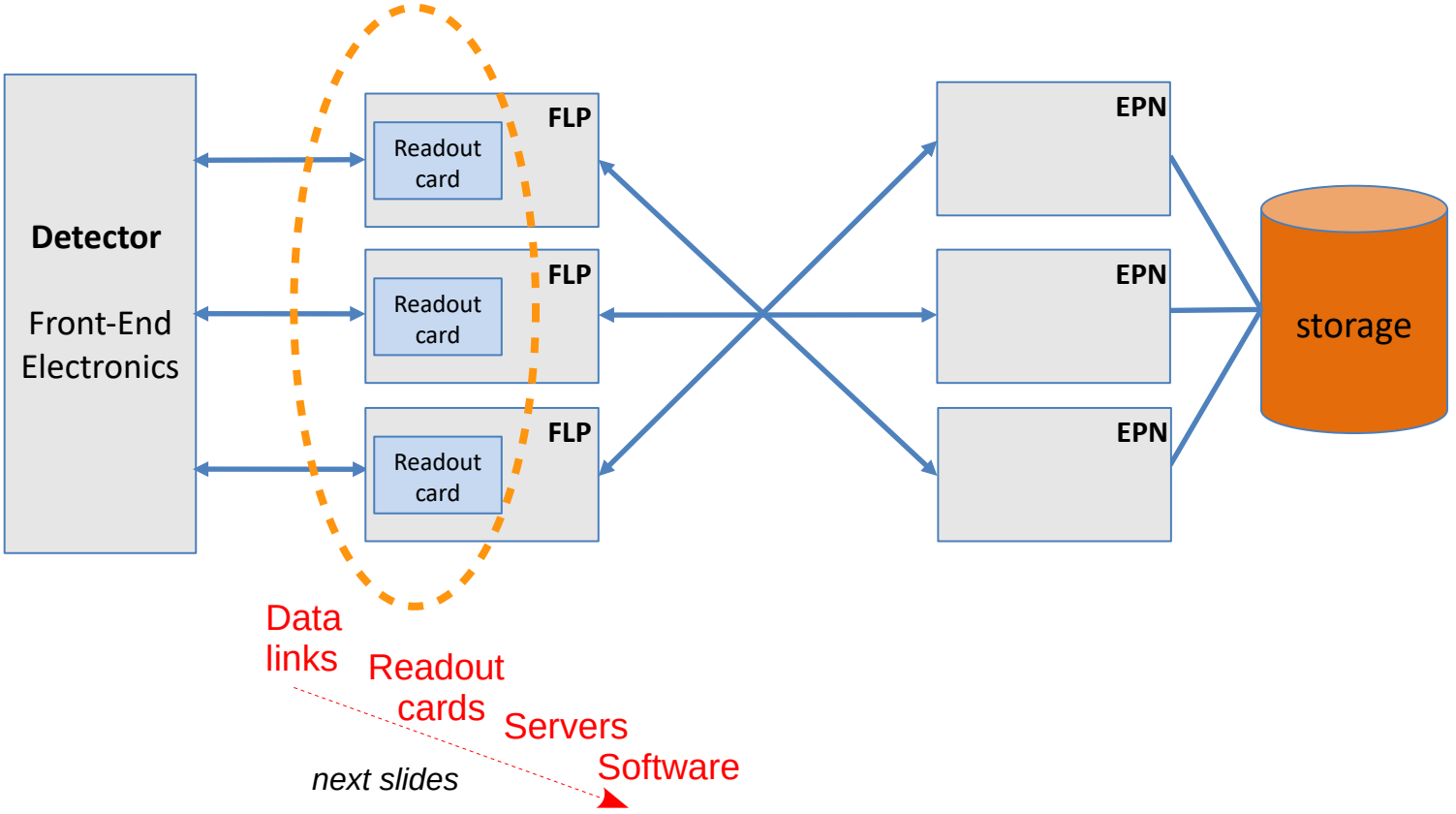
EPN farm

- Air cooled containers
- 250 servers
  - 64 cores, 8 GPUs
  - 100 Gb/s infiniband network
- 2x 1Tb ethernet to CC



CR0, inside & outside views

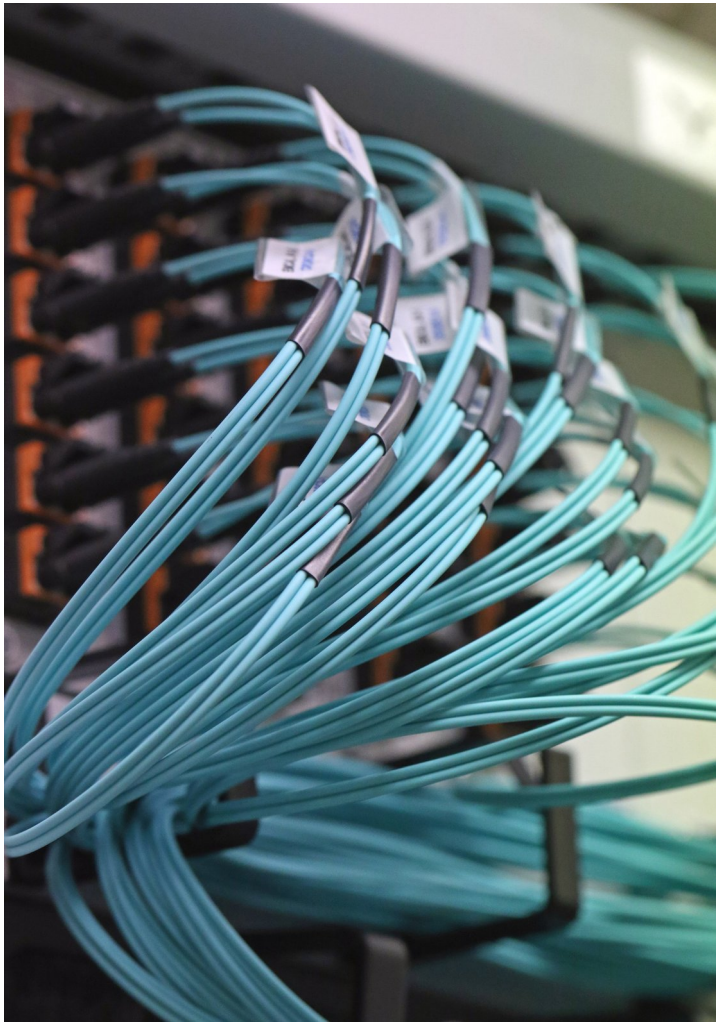
Detector readout



Detector Data Links

- Mainly GBT
  - radiation-hard bi-directional 4.8 Gb/s optical fiber link between counting room and experiment
  - It delivers/receives : DATA, TRIGGER and SLOW CONTROL.
- Support for ALICE custom link DDL1 and 2 (used during Run 1 and Run 2)
  - DDL1, 2.125 Gb/s
  - DDL2, max 5.3 Gb/s
- ~8’000 links in total

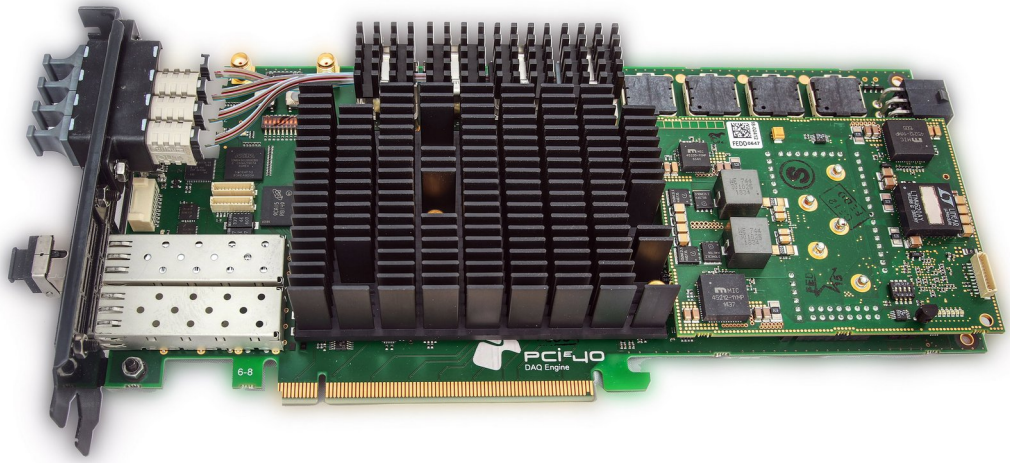
=> need for high-density readout system !



Detector links patch panel in CR1

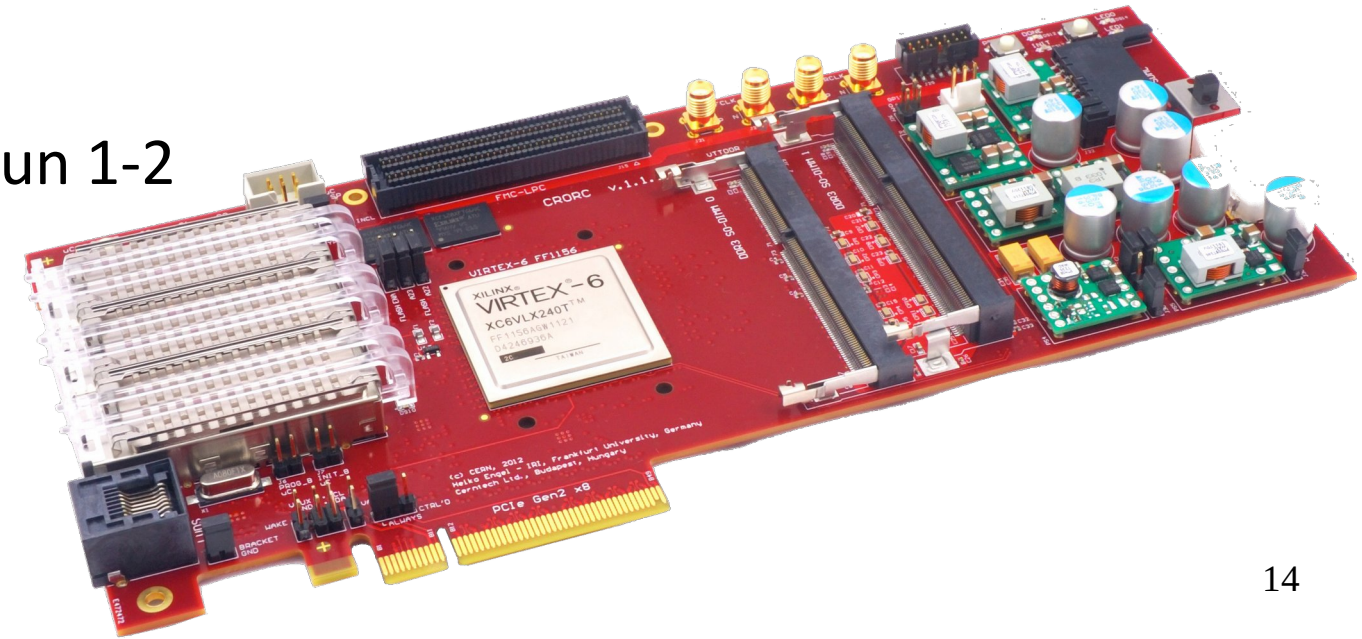
Readout hardware - GBT detector

- CRU (a.k.a. LHCb PCIE40)
  - FPGA: Intel Arria10
  - Maximum 48 GBT links bidirectional
- PCIe gen.3 x16
  - Dual DMA engine Gen3 x8
  - Typical throughput: 110 Gb/s
- Firmware allows dedicated user logic for on-board compression



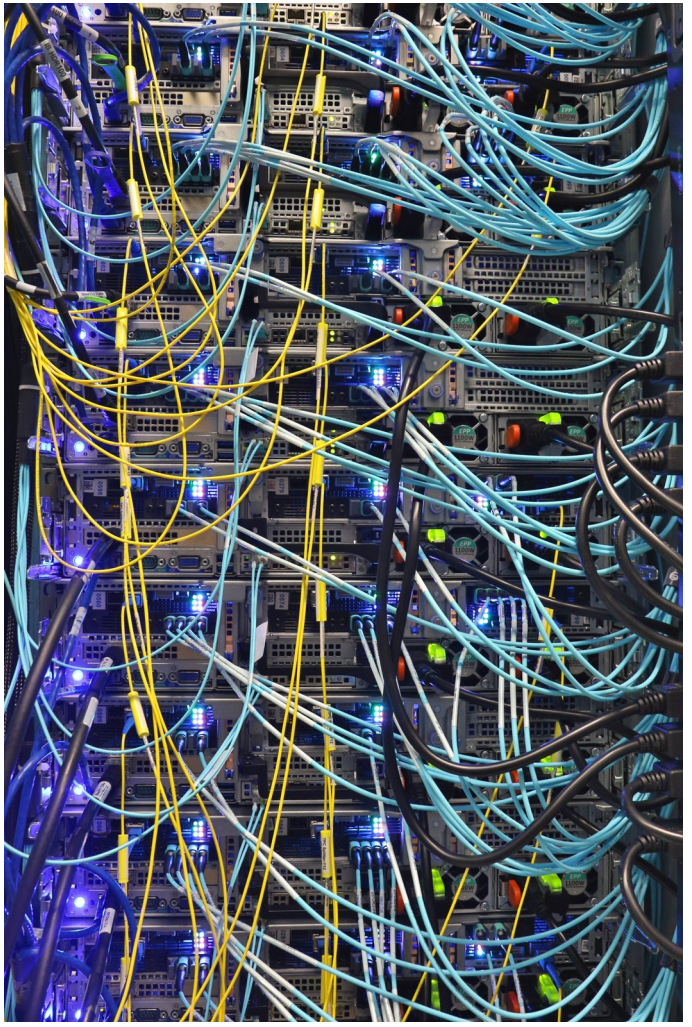
Readout hardware - DDL detector

- C-RORC
  - FPGA: Xilinx VIRTEX6
  - 6 DDL links
  - PCIe gen.2 x8
  - In use in ALICE Run 1-2



FLP Servers

- Packing I/O cards into small box (<3U)
  - Input: 3x CRU / C-RORC
  - Output: 1x infiniband HDR100
- Extensive testing and selection procedure
  - Throughput performance with realistic dataflow
  - Cards cooling
- DELL PowerEdge R740



Stack of FLP servers showing connectivity

Readout software

- Move the data from detector electronics into memory of PC
- Initialize hardware: CRU, C-RORC using common ROC (Read-Out Card) driver interface
- Allocate memory buffers for DMA
- Provide data pages to be filled by PCIe device
- Aggregate and slice data input
  - Trigger and continuous detectors, data grouped in chunks of 128 LHC orbits = 1 timeframe
- Check data consistency
  - Raw Data Headers from the incoming payload provide trigger counters, detector status bits, structure sizes, etc
- Distribute data to consumers
  - Formatting into O2 messages and adding toplevel headers
  - Forwarding to Data Distribution software in charge of pipeling local processing and sending to EPNs
- Report performance and errors
- Special features for commissioning / debugging: on-the-fly LZ4 compression, recording to disk, simulated data file player.

O2 software stack

- Provides framework for baseline features
  - Data processing: same code for synchronous / asynchronous processing
  - Data transport / inter-process messaging: provided by FairMQ library
- Other tools and services:
  - Control, data Quality Control, Monitoring, Logging, Configuration, Bookkeeping
  - Interface with Detector Control System and Central Trigger Processor

Commissioning

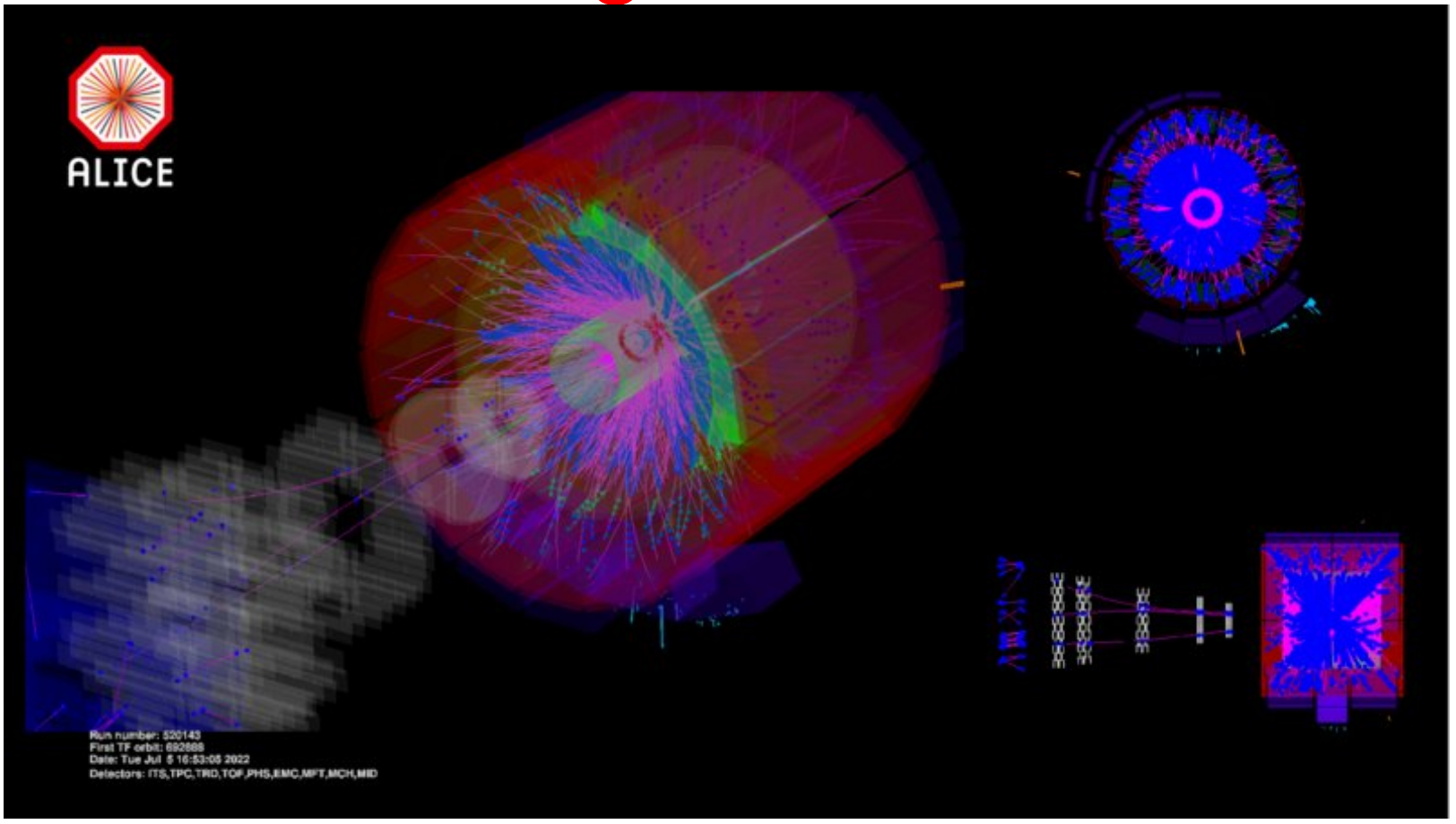
- Lots of work done in 2022 to get the system ready for LHC restart

  Over 2000 PB have been readout in the past 200 days (test + physics data)

  i.e. ~0.6 GB/s/FLP on a 6 months period

ALICE Control Room

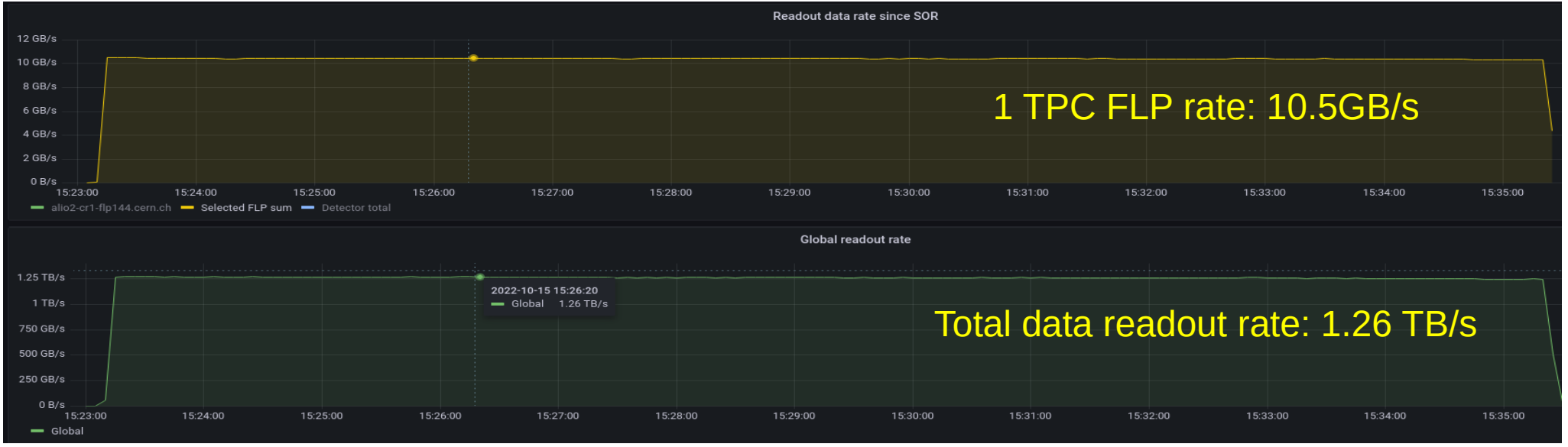
Commissioning



5 July 2022: ALICE first 13.6 TeV collisions of LHC Run 3.

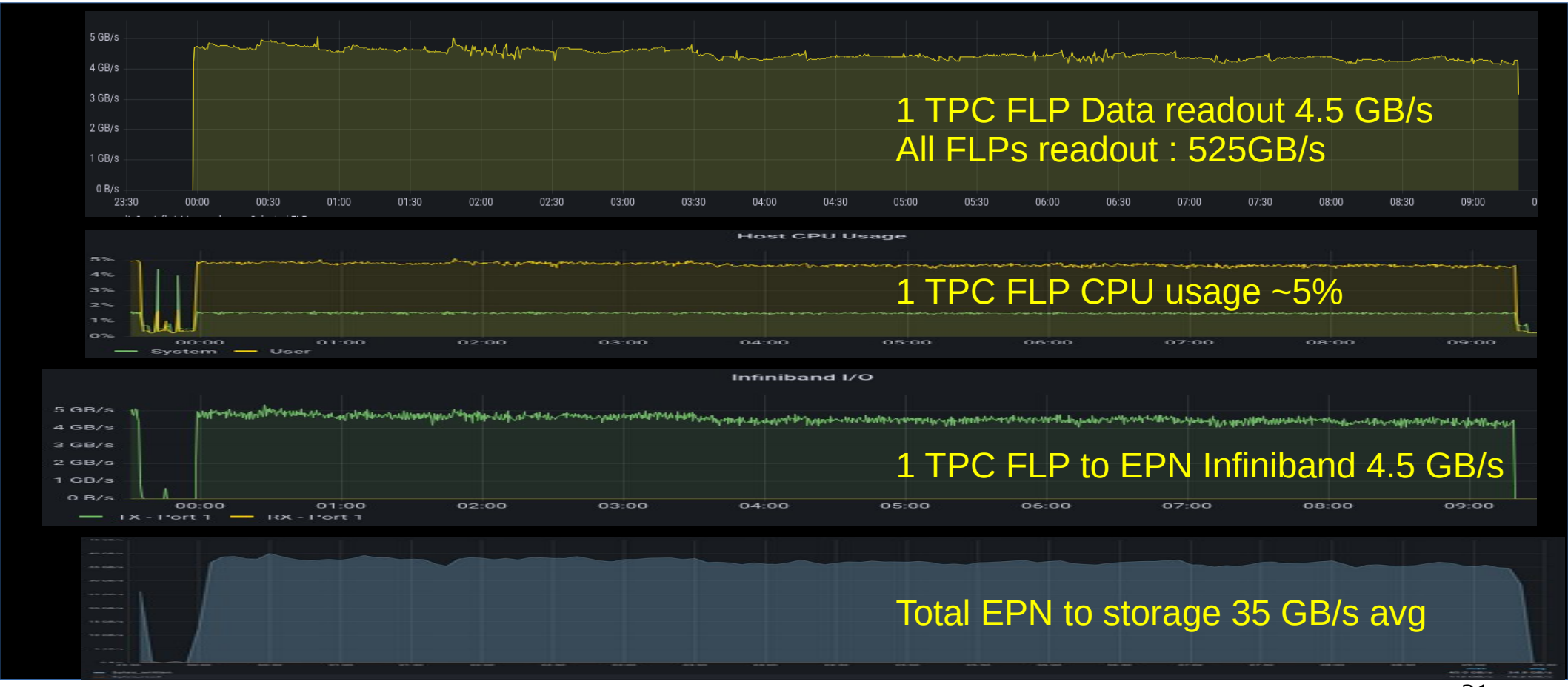
DAQ Performance

- Running routinely above the design validation criteria of 70Gb/s per FLP for testing
- Readout data flow exercised in demanding and changing conditions
  - Optimization of memory buffers (best page size depends on detector payload)
  - Little NUMA effects seen, plenty of headroom on QPI links



Run 527397 2MHz pp – no EPN – readout test at high rate

DAQ Performance



Physics run 527446 – duration 9h20m50s

Outlook

- DAQ system doing fine
  - Design margins find their use to accommodate detector evolving needs
  - Hardware does the job
  - Ongoing software developments: always more tools and features needed
    - Ease support and operations

- ALICE started Run 3 with success
  - Actively taking p-p physics data
  - Exercising data flow and processing with p-p at high rates
  - Waiting for HI collisions