ALICE Upgrade for Run3: Computing

HL-LHC Trigger, Online and Offline Computing Working Group Topical Workshop

Sep 5th 2014

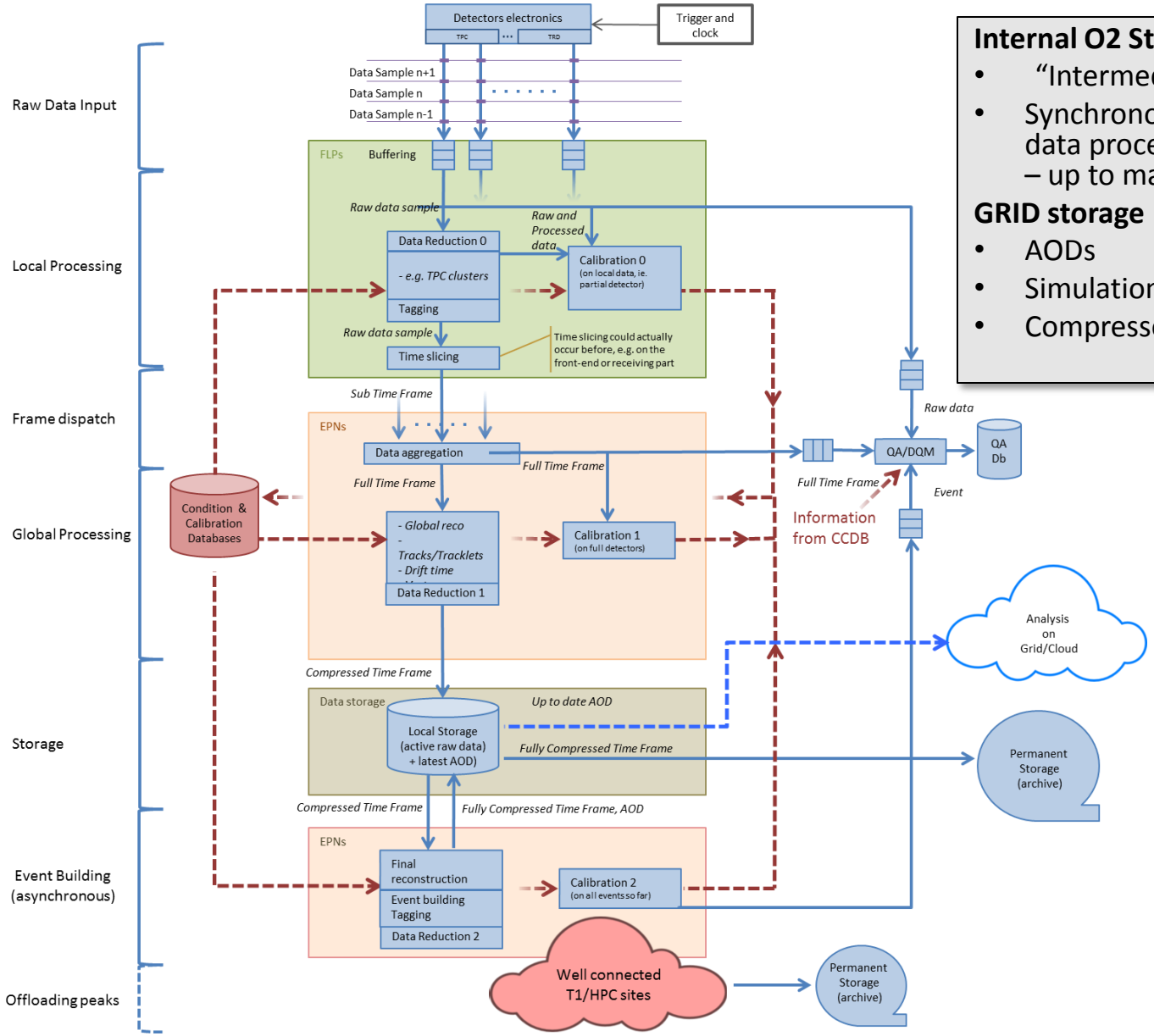
Timeline and resources consideration

- ALICE upgrade in Run3
  - CPU needs proportional to track multiplicity + pileup
  - Storage needs proportional to accumulated luminosity
- ALICE upgrade basic estimates
  - Event rate 50KHz (Pb-Pb), 200KHz (p-p, p-Pb)
  - Event size 1.1TB/sec from detector; 13GB/sec average processed and compressed to storage

Data processing and systems consolidation

- RAW data rates and volume necessitate the creation of an online-offline facility (02) for data compression, incorporating
  - DAQ functionality – detector readout, data transport and event building
  - HLT functionality – data compression, clustering algorithms, tracking algorithms
  - Offline functionality – calibration, full event processing and reconstruction, up to analysis objects data

Data flow and compression in O2



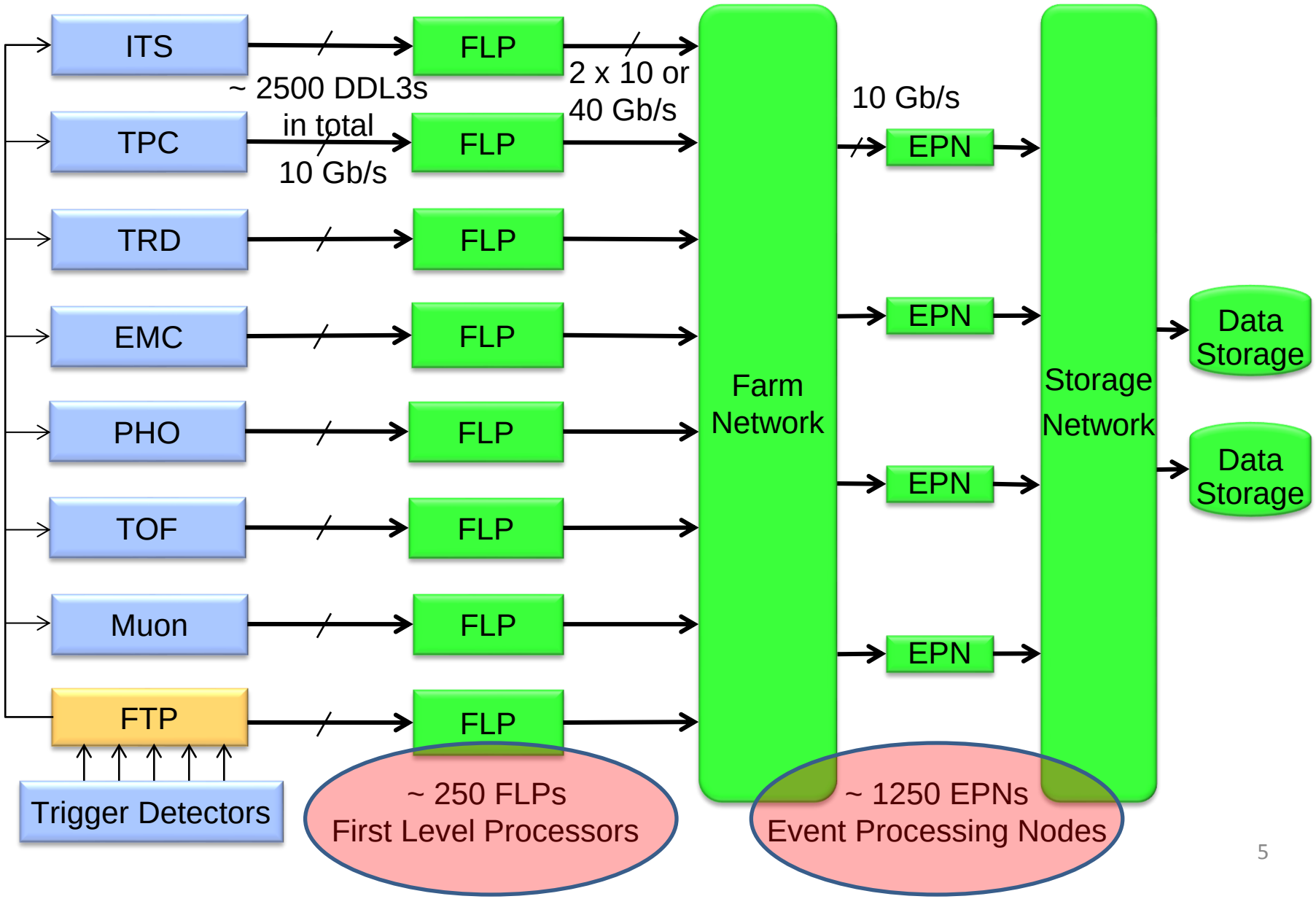
Internal O2 Storage

- "Intermediate" format:
- Synchronous and asynchronous RAW data processing with increased precision – up to max compression

GRID storage

- AODs
- Simulations
- Compressed RAW (custodial)

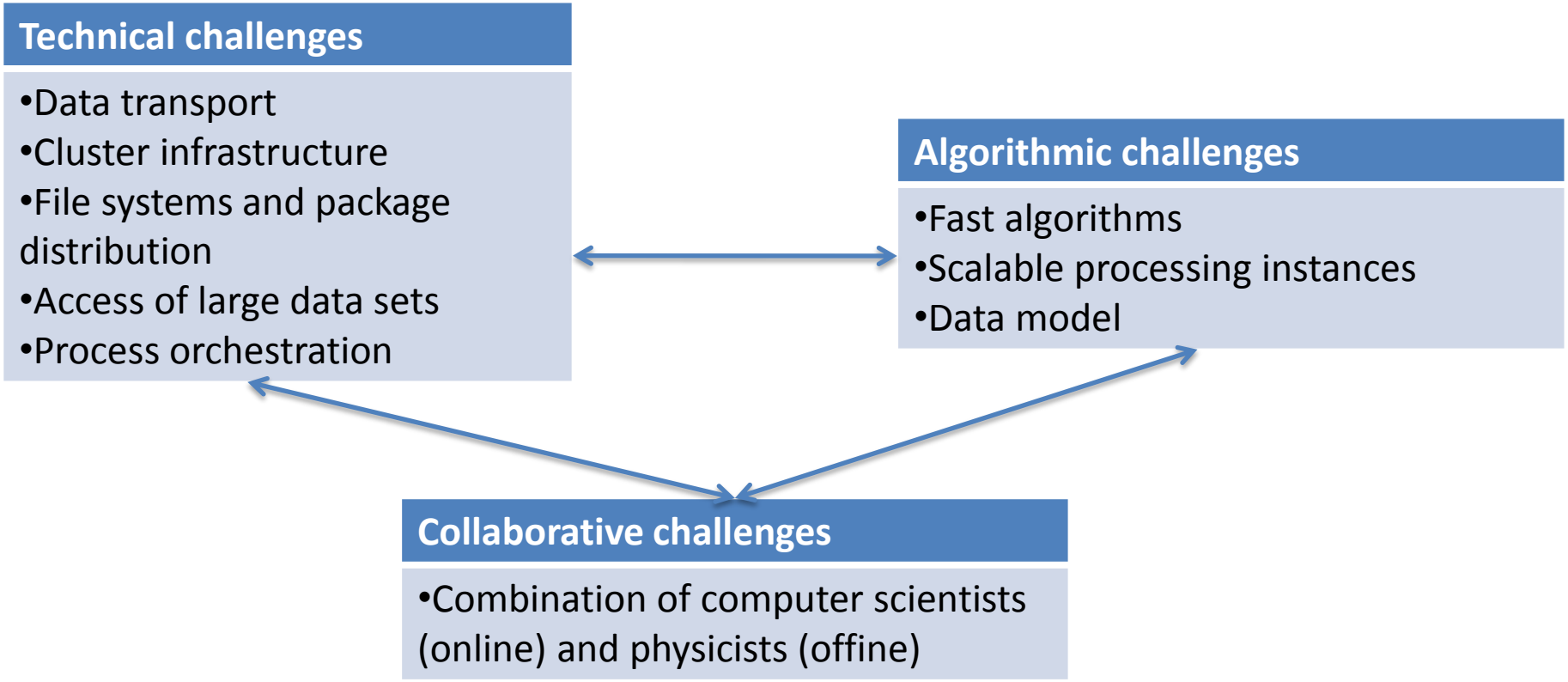
Hardware architecture of O2



Software and process improvements for O2

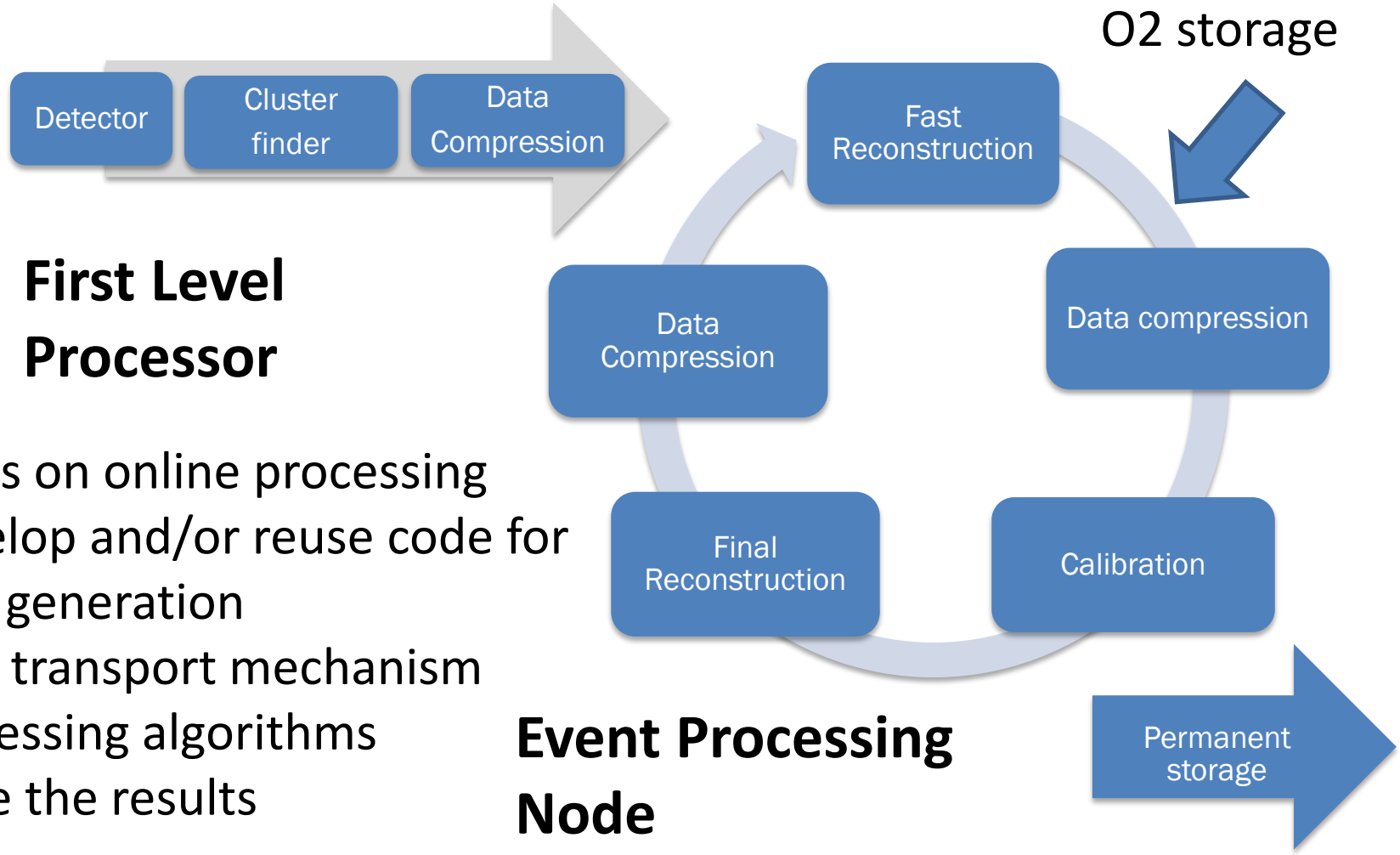
- ‘Offline quality’ calibration
  - Critical for the data compression
  - Compressed data in format allowing reprocessing, i.e. finer-grain calibration is still possible
- Use of FPGAs, GPUs and CPUs in combination
  - Software uses specific advantages of each
  - A well-tested approach in production (current HLT)
- New framework to incorporate all tasks
  - ALFA (ALICE-FAIR) developed in collaboration with the FAIR collaboration at GSI Darmstadt

ALFA – online system for data processing



- Technical/Algorithmic challenges are addressed in already formed Computing Working Groups

Simplified data processing scheme

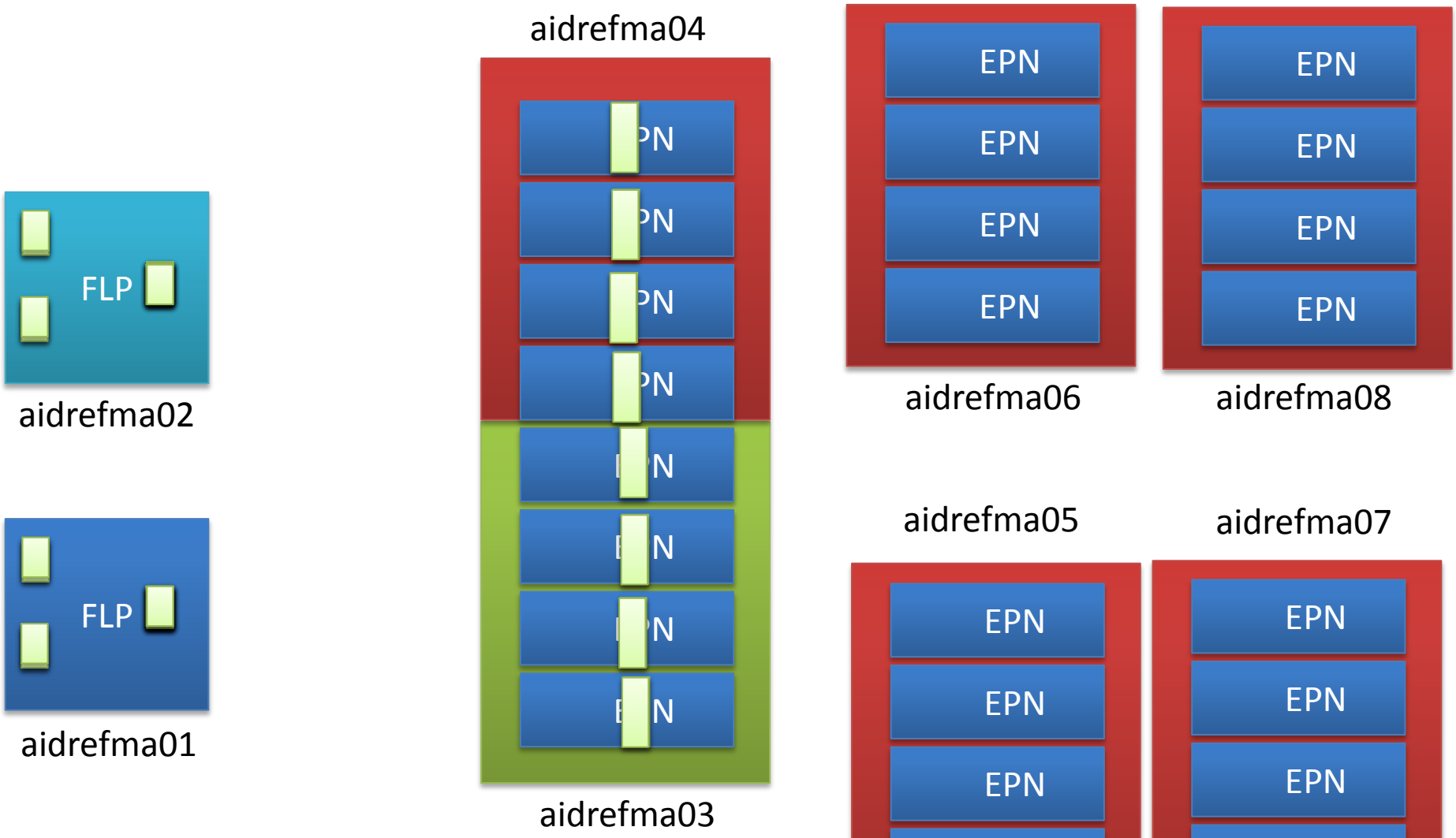


First Level Processor

- Focus on online processing
- Develop and/or reuse code for data generation
- Data transport mechanism
- Processing algorithms
- Store the results

Event Processing Node

Data transport prototype



Existing and tested to expected throughput

Preparations for Run3 – tests of processing chain

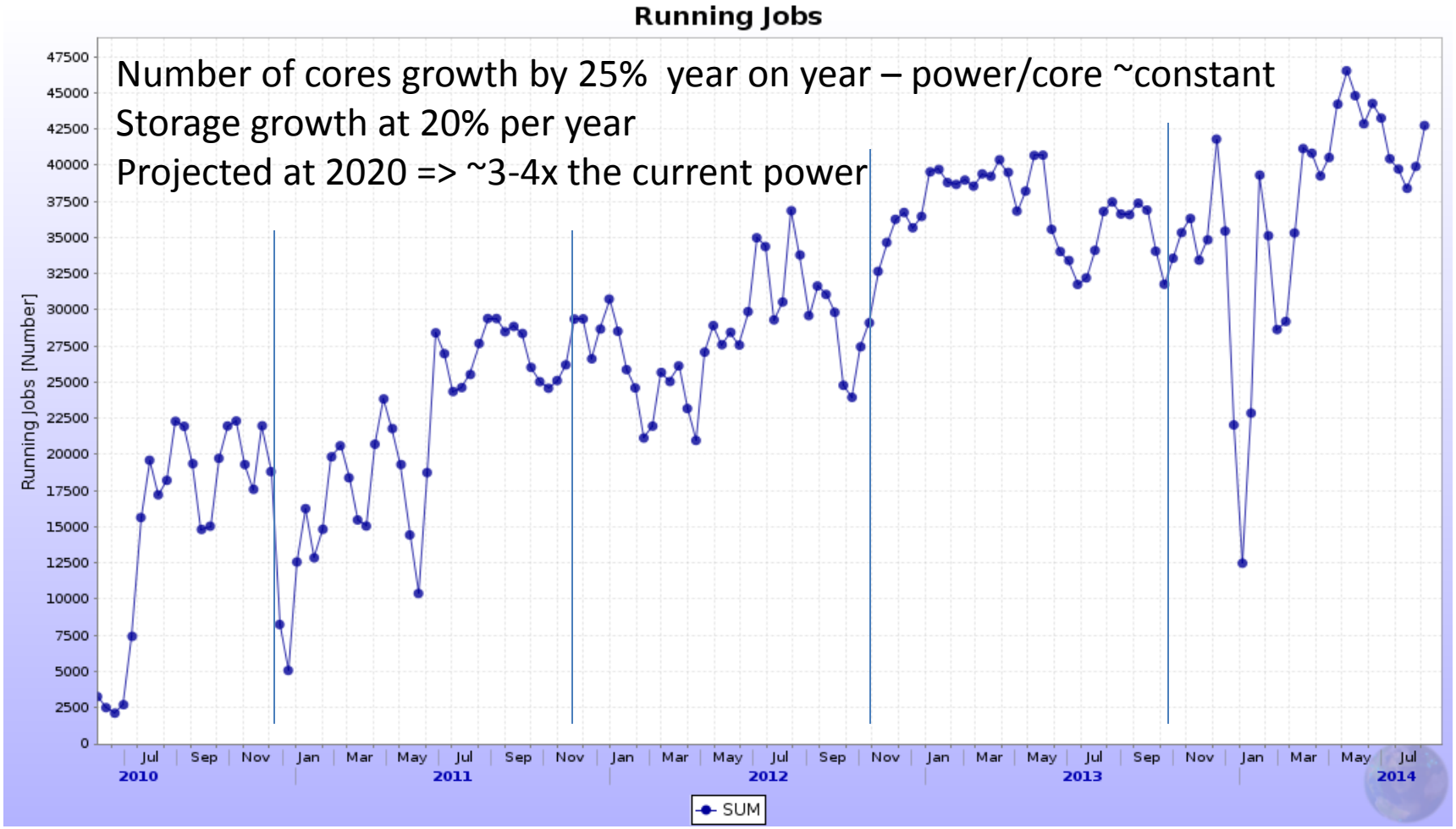
- Cluster finding in HLT for all detectors
  - New, more powerful, HLT cluster
  - System validated for TPC over 2 years of data taking in Run1
  - Largest data compression factor (x4)
- Detector calibration online
  - Many detectors (currently offline) calibration algorithms in HLT
  - Critical step for good quality tracking and data compression
- Preparation of O2 test cluster @10% of expected size
  - Testbed for ALFA development

The Grid

- General assumption and roles
  - Will continue to do what it does now
  - MC simulation, end-user and organized data analysis, raw data reconstruction (reduced)
  - Custodial storage for RAW data – output of O2 compressed data stream
- Expected that the Grid will grow with the same rate
- Search for additional resources
  - New sites
  - Contributed resources from existing sites

Growth expectations

- Based on ‘flat budget’ and modified Moore’s law



Grid upgrades – data processing

- Assume sites will continue offering standard CPUs
  - 'GPU ready' algorithms have to cope
- Growth beyond 'flat budget' scenario necessitates finding new resources
  - This is an ongoing process
- Contributed cycles from existing, non-standard sources
  - Supercomputer back-fill – this can potentially bring equivalent of tens of thousand CPU cores for MC
  - Ongoing collaborative work with ATLAS (PanDa)
- Support for various cluster configurations will continue and expand, as needed
  - Standard batch/whole node/AI-cloud-on demand
  - Increases flexibility, efficiency

Grid upgrades – data storage

- Grid storage is expected to hold the analysis object data and compressed RAW (custodial)
- The data volume to the Grid will be reduced to reasonable size by the O2 facility
- General expectation for cheaper storage driven by the needs of the ‘big players’ (insert Facebook here)
  - May modify upward the current 20% yearly growth
  - Perhaps slower media – need for more sophisticated data access models
  - Biggest impact is on the cost of storage for the O2 facility

Summary

- ALICE upgrade for LHC Run3 (2020 and beyond)
  - Event rate x 100, data volume x 10 (of current rates)
- Data reduction to manageable levels - dedicated O2 computing facility, incorporating DAQ/HLT and Offline functions
  - Heterogeneous hardware (CPU/GPU/FPGA)
  - Data volume reduction by x10 to 13GB/sec, up to analysis-ready data containers
  - Framework development, software demonstrators, test cluster in preparation
- Grid resources are expected to grow with the usual rate
  - Intensive programme to incorporate new sites and contributed computing power on existing facilities