DAQ<->Offline interface

R.Itoh, KEK

Discussion items:

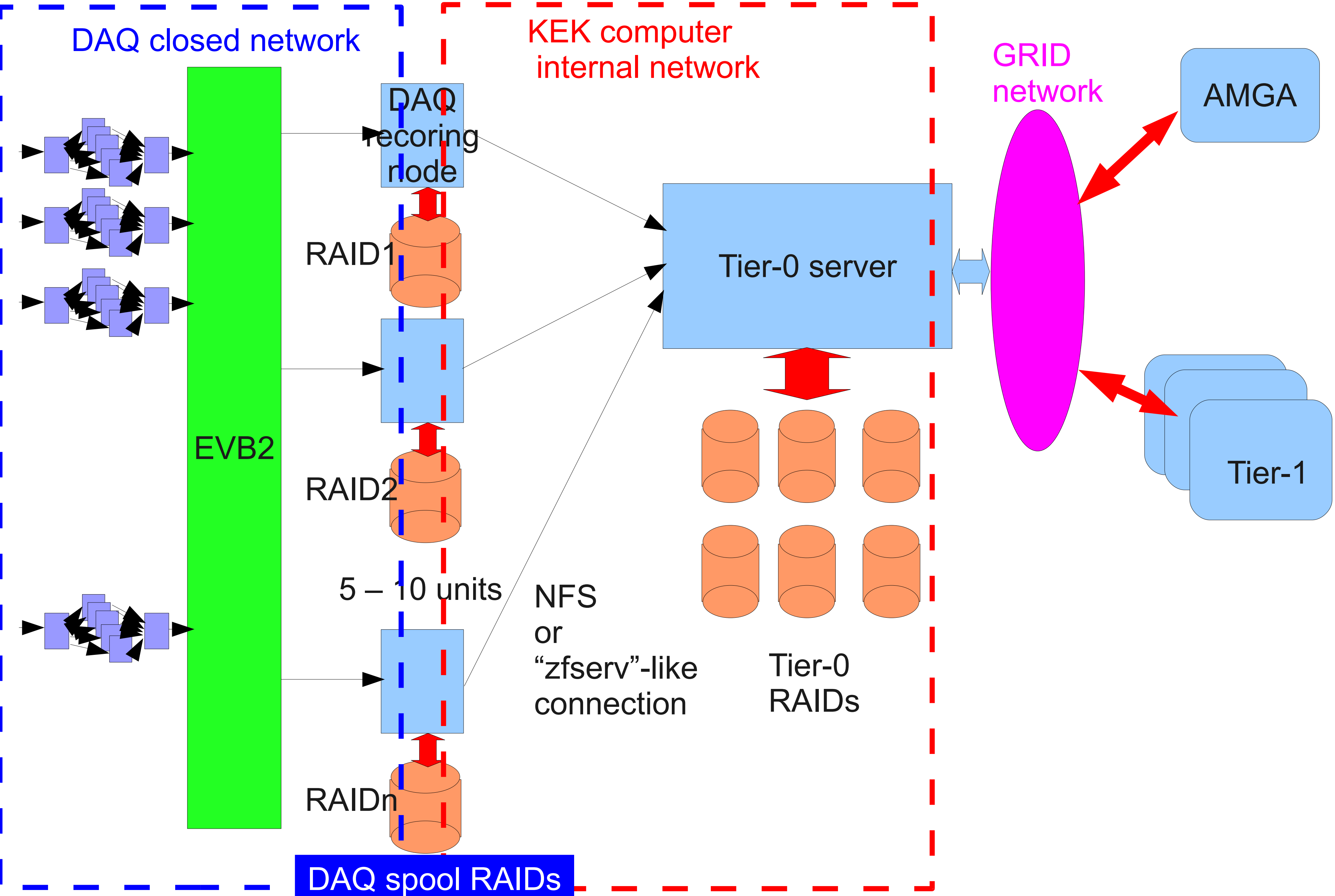
- Clarify the sequence to transfer raw data files from DAQ spool RAID to Tier-0.

- Method of data quality monitoring (DQM/QAM)
  => Proposal of “Prompt Reco”

- Method to access Tier-0 raw data from DAQ side

1. Transfering raw data from DAQ to Tier-0

My understanding on the network connection



Procedure of raw data copy

0. DAQ recording nodes write raw data in multiple consecutive files in DAQ spool disk.
   - One data file is blocked around 4 or 8 GB.
   - The file is written in "SeqRoot" format.
   - File name could be ennnnnnrmmmmmm-raidx.sroot-yyyy

     x: raid number, yyy: sequence number of blocked file

DAQ concern

1. At run end, DAQ recording node generates a "spool file" in which
   - data file names (+directory names) in the DAQ spool disk
   - run information (# of events, HLT process info, etc.)

   are written. It is placed in the "spool directory" of each recording node disk.

2. Tier-0 server periodically fetches the spool file directory (via NFS) and initiates the data file copy whenever a new spool file is detected. The files are transferred to Tier-0 RAID, together with the spool file.

3. Tier-0 server then generates "meta-data" for AMGA containing the place of raw data files in Tier-0 RAID and "run info" taken from the spool file. The meta-data is finally registered in AMGA through GRID network.

   * Note: step2 and step3 should be separated since the related network is different.

     -> Isolate AMGA from raw data copy sequence

Questions

- Layout of Tier-0 servers. How are they connected to DAQ recording nodes?
  - * We have 5~10 recording nodes, and Tier-0 servers are supposed to be connected to these servers (almost) directly.
  - * The recording bandwidth is supposed to be 100-200MB/sec/RAID server through a dedicated 10GbE link/server.
    - -> Similar (or faster) bandwidth is required for data copy to Tier-0. Need to implement another (or multiple) 10GbE NIC(s) in DAQ RAID server.
      - => How many NICs? Layout of RAIDs to achieve required performance?

- The size of DAQ spool RAID.
  - * We assume the size to keep raw data files for a week.
    - -> Assume total recording bandwidth of 1.2GB/sec,
  - * The required size for a day is 100TB. -> 700TB/week.
  - * If we record the stream in 10 RAID servers, one server need to have ~100TB/server. (Could be smaller at t=0. 1/5 or less).
    - => Is it enough?

2. DQM (Data Quality Monitor) / QAM (Quality Assurance Monitor) issue

- DQM / QAM is primarily provided by HLT. We will implement a real time monitoring of histograms accumulated in HLT. The design of DQM/QAM is in progress (by me).

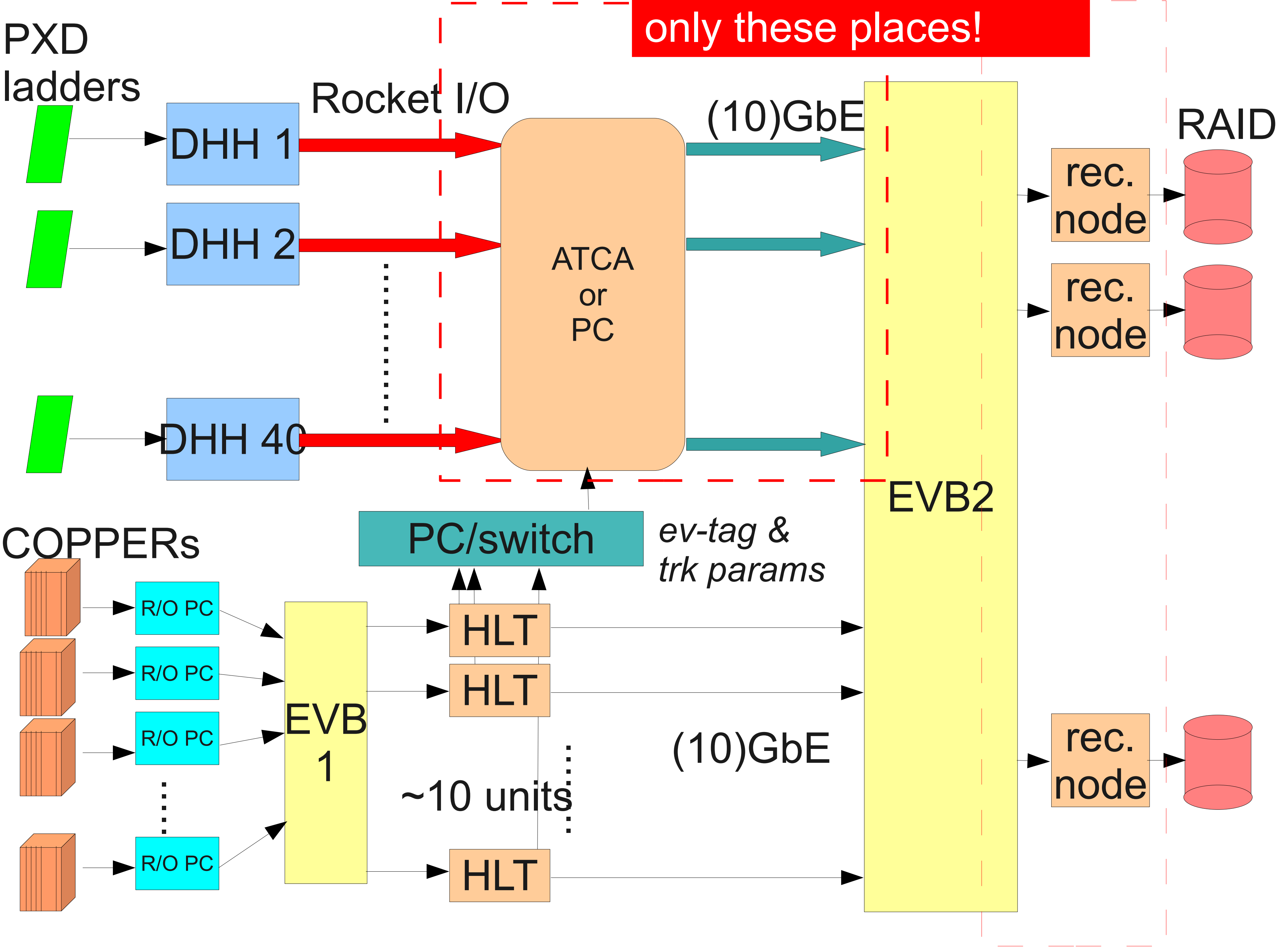
- However, PXD data are merged AFTER HLT processing. It means there is no particular platform to monitor PXD data in DAQ data flow for now.
  * Of course, there is a possibility to run DQM on PXD readout box (ATCA or PC), but it cannot use other detector information.

- Another idea is to run the monitor on recording node server, where full events are built. But the CPU power of the recording node may not be enough.

- Better method is to monitor the quality in offline full reconstruction. However, the offline reconstruction is supposed to be GRID based and a large time-lag is expected to monitor the quality.

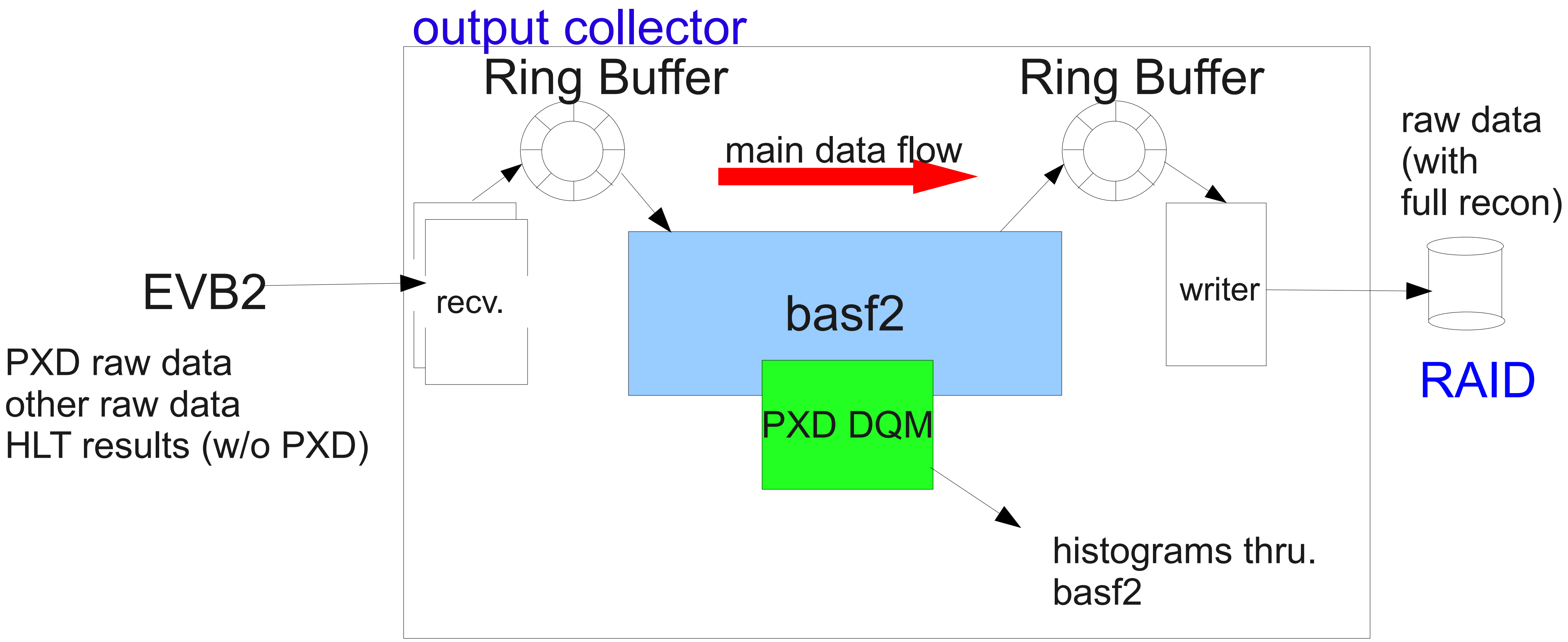
Proposal of “Prompt Reco” scheme

PXD integration



PXD data are available only these places!

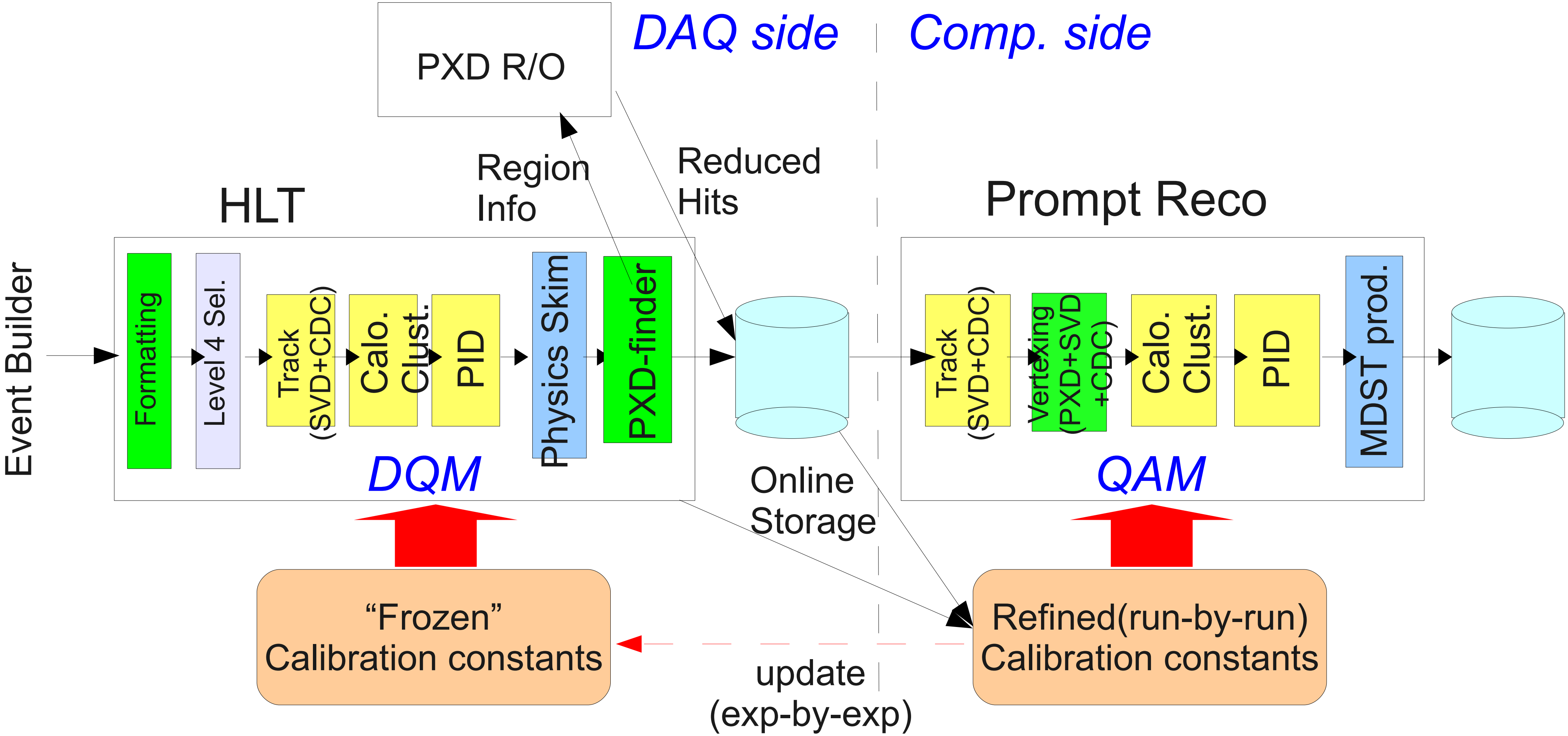
PXD-DQM on DAQ recording node



* Processing power of recording node is limited.
* May not be possible to monitor all events.
Only for sampled events.......

HLT Event Processing chain

🟨 The same software shared by HLT and offline



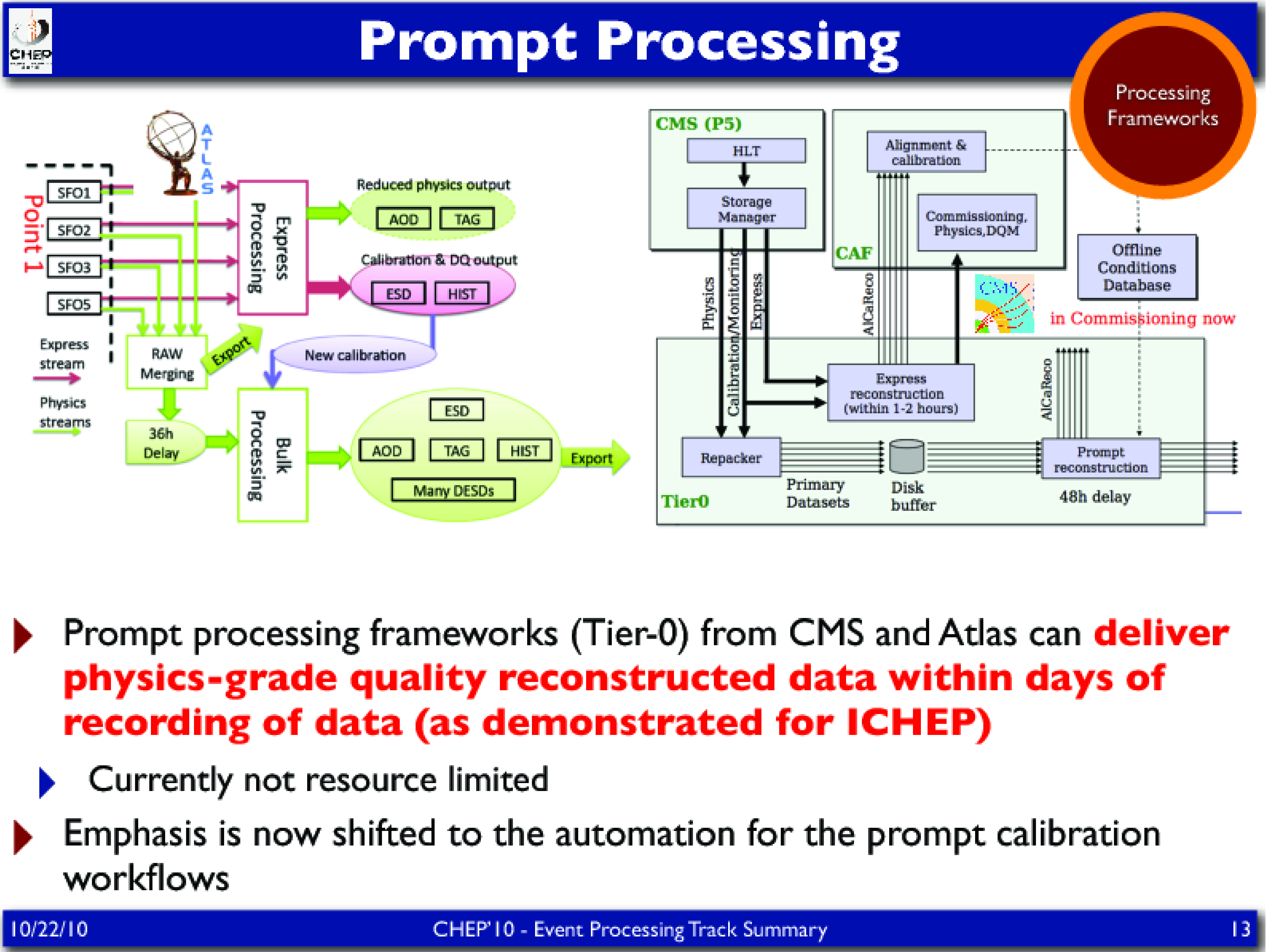
* Idea is, to run offline reconstruction instead of raw data copy right after a run is completed.

* The refined constants can be generated in HLT processing.

* The same HLT parallel processing scheme can be recycled for the "prompt reco" scheme.

* This scheme was actually adopted by Belle already for data skimming.

- Prompt Reco is already used in ATLAS and CMS.



Prompt Processing

- Prompt processing frameworks (Tier-0) from CMS and Atlas can **deliver physics-grade quality reconstructed data within days of recording of data (as demonstrated for ICHEP)**
  - Currently not resource limited
- Emphasis is now shifted to the automation for the prompt calibration workflows

- Prompt Reco can produce “physics level” processing results which can be directly placed on Tier-1 as DST files.

- The CPU power for DST production is supposed to be centralized in KEK even in Belle II and the implementation seems to be straight-forward using HLT technology.

- Prompt Reco can generates QAM histograms for all detectors and physics modes within a few hours delay.

- Drawbacks are
    * DST production (at least first step) cannot utilize GRID CPUs.
        => But the use of GRID for 1st step DST production is not so much desired, anyway.
    * A large fraction of KEK CPUs have to be configured outside GRID.
        => Maybe a good way to share with GRID.....
    * Man power problem
        => The work can be shared with HLT development.

3. Method to access raw data in Tier-0 from DAQ side

- Sometimes DAQ people need to access recorded raw data for the DAQ debugging.

- The raw data are supposed to be placed in Tier-0, where the access through GRID is assumed.

- But we need a quick and interactive access to the raw data. GRID based access does not fit for this purpose.

- The raw data are supposed to be placed in KEK computer facility, anyway.

- Any good way to access raw data inside KEK bypassing GRID?