



Contribution ID: 219

Type: oral presentation

THE DAQ NEEDLE IN THE BIG-DATA HAYSTACK

Tuesday 14 April 2015 14:00 (15 minutes)

Technology convergences in the post-LHC era

In the course of the last three decades HEP experiments have had to face the challenge of manipulating larger and larger masses of data from increasingly complex and heterogeneous detectors with hundreds of millions of electronic channels. The traditional approach of low-level data reduction using ad-hoc electronics working on fast analog signals, followed by global readout and digitisation, and a final stage of centralised processing in a more or less monolithic system has reached its limit before the LHC era. LHC experiments have been forced to turn to a distributed approach, leveraging the appearance of high speed switched networks developed for digital telecommunication and the internet. This has led to a generation of experiments where the use of custom electronics, analysing coarser-granularity analog or digital “fast” data, is limited to the first phase of triggering, where predictable latency and real time processing, as well as reliable, low-jitter clock and trigger distribution, are a necessity dictated by the limits of the front end readout buffers. Low speed monitoring (temperatures, pressures, etc.) and controls (thresholds, calibrations, voltages, etc.) have remained decoupled and considered an altogether separate realm in the detector design and operation.

We believe that it is now time for the HEP community to prepare for the next “revolution”. Already, the mass of persistent data produced by e.g. the LHC experiments means that multiple pass end-to-end offline processing is becoming increasingly burdensome. Some experiments (e.g. ALICE) are moving towards a single-pass system for data reduction, relying on fast calibration feedback loops for zero suppression and low-level pattern recognition into the online system. The pristine “raw” channel readouts become thus volatile and no longer permanently stored. Others (e.g. LHCb) read out every channel for each beam crossing and delegate the entirety of the data reduction, reconstruction and selection to a fully software system. The latter approach is particularly attractive if low power techniques can be developed to counter the negative effects of the consequent increase in material budget for services and cooling in the active areas of the detector. Further developments can be envisaged.

On the one hand very large scale integration paired with progress in radiation hard technologies, as well as the appearance of high-bandwidth bidirectional optical links, both on and off-silicon, could make intelligent very-front-end electronics and high-speed low-power readout a possibility already in the next decade, thus lifting strict latency limitations. At the same time, the traditional distinction between readout, trigger and control data channels will become increasingly artificial, paving the way to the possibility of running fully programmable algorithms at on- or near-detector electronics.

On the other hand, boosted by the “big data” industry, massively parallel and distributed analysis of unstructured data has become ubiquitous in commercial applications. Apart for their attractiveness for use in monitoring of both detector parameters and data flow, as well as data analysis, these technologies indicate in our opinion a possible evolutionary path for future DAQ and trigger architectures. In particular, a new trend is emerging from the data mining and analytics world which consists in “bringing the algorithm to the data”. For HEP experiments, this might mean to abandon the consolidated paradigm represented by the triad low-level trigger - event building - high level trigger.

How close can we bring our algorithms to the detector ? Can we take advantage of the ideas, software (and hardware) technologies developed for data mining and search engines ? Can we imagine a future detector with extremely deep multi-stage, asynchronous or even virtual pipelines, where data streams from the various detector channels are analysed and indexed in quasi-real-time, and the final selection is operated as a distributed “search for interesting event parts”?

Can we push these ideas even further, removing the inflexible notion of pre-processed datasets and paving

the way to completely new forms of selection and analysis, that can be developed, tested and implemented “online” as aggregation and reduction algorithms making use of the full, unstructured information from the experiment and directly returning the high-level physics quantities of interest ?
We investigate the potential impact of these different developments in the design of detector readout, trigger and data acquisition systems in the next decades.

Author: MESCHI, Emilio (CERN)

Presenter: MESCHI, Emilio (CERN)

Session Classification: Track 1 Session

Track Classification: Track1: Online computing