The DAQ needle in the big-data haystack

E Meschi
CERN, Geneva, Switzerland
E-mail: emilio.meschi@cern.ch

Abstract. In the last three decades, HEP experiments have faced the challenge of manipulating larger and larger masses of data from increasingly complex, heterogeneous detectors with millions and then tens of millions of electronic channels. LHC experiments abandoned the monolithic architectures of the nineties in favor of a distributed approach, leveraging the appearance of high speed switched networks developed for digital telecommunication and the internet, and the corresponding increase of memory bandwidth available in off-the-shelf consumer equipment. This led to a generation of experiments where custom electronics triggers, analysing coarser-granularity “fast” data, are confined to the first phase of selection, where predictable latency and real time processing for a modest initial rate reduction are “a necessary evil”. Ever more sophisticated algorithms are projected for use in HL-LHC upgrades, using tracker data in the low-level selection in high multiplicity environments, and requiring extremely complex data interconnects. These systems are quickly obsolete and inflexible but must nonetheless survive and be maintained across the extremely long life span of current detectors.

New high-bandwidth bidirectional links could make high-speed low-power full readout at the crossing rate a possibility already in the next decade. At the same time, massively parallel and distributed analysis of unstructured data produced by loosely connected, “intelligent” sources has become ubiquitous in commercial applications, while the mass of persistent data produced by e.g. the LHC experiments has made multiple pass, systematic, end-to-end offline processing increasingly burdensome.

A possible evolution of DAQ and trigger architectures could lead to detectors with extremely deep asynchronous or even virtual pipelines, where data streams from the various detector channels are analysed and indexed in situ quasi-real-time using intelligent, pattern-driven data organization, and the final selection is operated as a distributed “search for interesting event parts”. A holistic approach is required to study the potential impact of these different developments on the design of detector readout, trigger and data acquisition systems in the next decades.

1. Introduction: A bit of history
It is interesting, for the purpose of our discussion, to start with a quick look at the past history of data acquisition systems for particle and nuclear physics experiments. Figure 1 shows what we believe to be one of the first examples (if not the very first) of a data acquisition system.

The system features all the main components constituting an experiment and its acquisition. The detector (E) is connected to a programmable trigger (HR) and analog front-end electronics (G) is acquired (R) and recorded on digital media (P). At the time, using the words of the author, it was perceived that “...visual or audible methods of counting are quite trying on the nerves because close attention is necessary on account of the probability law of emission of the particles; and long periods of counting are essential because the method is a statistical one. A
self-recording device would therefore be an obvious improvement”. At a time when experiments were quite simple, the need for “automated recording” was mainly related to the “human factor”: how to best free the researcher brain from a menial task and put it to work on what the brain does best. It is fair to say that the logical structure of a HEP experiment did not vary much for many decades. While the scale of experiments increased, and the technologies improved, the guiding principle was always that of providing data reduction through fast “automated” signal processing (applying thresholds etc.) and storing the result in some digital or analog form for offline analysis. The appearance of digital signal processing and modular electronics allowed physicist to build more flexible and complex experiments with many acquisition channels, but it is only with the advent of colliders, with their inherently high interaction rates, that a new concept, that of “event building” was introduced.

Since the early days of high-energy physics, acquisition of subsequent events was disabled when the readout of an interesting event was initiated and until readout was complete. Other factors in an experiment contributed to this deadtime, but readout time was the dominating factor. The high interaction rates, and consequent high trigger rates, generated at colliders, meant that the readout time was an even larger fraction of the time between triggers. Buffering events at or near the detector during triggering, such that when an accept occurred, buffered data could be quickly read out without disabling the trigger, became a necessity. A common denominator of collider experiments of the pre-LHC era was the multi-stage pipeline/trigger structure outlined in figure 2. Usually, only a fraction of the total data from each event was made available for use in the trigger decision. The remaining data remained scattered over many front-end buffers. An accept at the lowest trigger level caused the corresponding data to be held in the pipeline. Subsequent stages of trigger/pipeline could apply increasingly sophisticated algorithms while the rate was reduced, and a final software algorithm of some kind selected events to be stored. The corresponding data had then to be collected in one place for this high level trigger to work on. The event builder function is to connect the individual data sources (detector front-end electronics) and the data destinations (high-level trigger and online mass...
Regardless of the implementation, the event builder function is to multiplex data. If rates are low, this multiplexing operation can take place over a single time-shared bus using a software-controlled selection of source and destination. If necessary, a bus can be extended indefinitely using repeaters; it is simple to control and relatively inexpensive. In this scheme, a single processor controls the bus at any given point in time and reads data from all the buffers. This was the technique used in the majority of data acquisition systems prior to LHC [2, 3], and at the dawn of LEP we read: “typical detectors are more than 100 000 electronics channels. Raw data per event can reach several megabytes due to more common use of time-sliced digitising [forcing] data compaction in situ, to keep the transported data to a level acceptable to storage devices... the purpose... becomes gathering all events of physics interest with minimum background [and] minimum deadtime. One should have in mind that the data acquisition system (not including front-end electronics) amounts to about 5% of the detector cost” [4]. Two new keywords appear at this time: deadtime (in the form of actual dead time of the DAQ system OR of unwanted background), and cost. As a matter of fact, the human factor persists too. The same author observes later, of the then-being-assembled LEP experiments, that concerning single-board computers to control the busses, “Surprisingly few experiments envisage the purchase of commercial boards but the situation is evolving fast”. This “sociological” resistance of the field to adopt standards is an important point to which we will return later.

Great effort was put into the development and use of data buses for high energy physics applications. CAMAC and FASTBUS, very popular in the seventies and eighties respectively, were later abandoned for the computer industry standard VMEbus. This was at a time when commercial multiprocessor systems used themselves a shared bus architecture. At this stage, limited local cache memory on a module meant no need for continuous bus activity, and the same held for data acquisition systems where the time to read out an event was usually much shorter than the processing time. The use of VLSI front-end circuitry changed this radically, increasing the acceptable trigger rates by a factor of at least 1000. Only five years later it was observed: “Similarly, the performance of high-level processors and the density of on-line data storage have both improved by a factor of almost 1000 over the last fifteen years. Unfortunately, the speeds of standard busses used for event building have improved by only a factor of ten in the same time period. The event builder has become the bottleneck” [5]. The situation can be visualized as in figure 3. The use of modular electronics standards in the data acquisition has declined over the history of the field to reach the current state of affairs at the LHC experiments. A parallel event builder was clearly necessary for the much higher throughput required by LHC experiments. With several orders of magnitude more electronic channels than the previous generation, and beam crossing times of few tens of nanoseconds with subsequent very high trigger rates, data acquisition systems required an interconnect with no inherent blocking, in addition to much larger amounts of pre-building buffering, in order to achieve minimal deadtime. The ever-increasing channel count could now be counted among the recurrent keywords that characterize the evolution of HEP DAQ systems. Altogether, this led to the adoption of the well-known common event builder architecture with a large switch fabric and distributed processing which is now mainstream across the field. It is fair to say that at the time of designing the experiments, the corresponding technology was far from given, both in terms of complexity and cost. Fortunately, another field, digital telecommunications and the internet, was driving the market and pushing these very same high-bandwidth switched networks forth. At the same time, with personal computers hitting the consumer market and the gaming community, high memory bandwidth, unthinkable only a decade earlier, became available off-the-shelf to profit of these new interconnects, and cheap computing farms to run large distributed tasks finally had reason of the mainframes and expensive proprietary workstation clusters.

The diversely bold approach with which the different LHC experiments embraced these new
possibilities is reflected today in the different architectures of their data acquisition systems, and also recalls that, in a relatively closed community of fairly specialized people, resistance to drop proprietary, home-grown solutions in favor of standards, as already mentioned above, is to be expected. Related to this, recent HEP experiments, from conception to completion of their physics program, now often last several decades. Fundamental design choices may reflect on the life of more than one generation of physicist both in requiring maintenance of legacy system and in influencing what may or may not be upgraded/modified at a subsequent stage. This last keyword - longevity - concludes our brief historical discussion of DAQ for HEP.

2. New challenges
One of the most important trends that emerged from the LHC designs, as already mentioned, was the widespread adoption of standard networking hardware derived mostly from telecommunications and cluster interconnect applications. Low-latency, high bandwidth cluster interconnects turned out to be a winning solutions in many cases with respect to expensive telecommunications-class hardware. Another important new aspect, which went mostly unnoticed, was the progressive move away from the complex multi-stage trigger systems of the previous decade. CMS, for example, adopted an architecture with a single-level hardware trigger and a large “filter farm” of linux computers to carry out the final online selection off the full Level-1 trigger rate. As noted above, this was a bet at the time since the necessary networking and computing hardware was not (yet) completely available [6].

2.1. The LHC upgrades
More than for the technical aspects, the CMS choice is important because it marks the beginning of a clear trend to progressively reduce the role of custom hardware triggers in favor of faster readout, and software triggers. The constraints of this approach are dictated by the limits of the front-end readout buffers, the complexity and power consumption of the optical links required to extract the enormous amount of data generated by tracking detectors, the consequent need for cooling which increase the dead material budget and may end up resulting in an unacceptable loss of resolution (figure 4). The brute-force readout of the upgraded LHC detectors at the full
machine clock speed, however, is still technically challenging and only LHCb, with its specific partially open structure, is currently planning to upgrade its DAQ system to read out every channel for each beam crossing, and delegate the entirety of the data reduction, reconstruction and selection to a fully software system [7]. To cope with the unprecedented data-rates and volumes generated by HL-LHC collisions in the upgraded detectors, while taking into account the previous considerations concerning readout links, other experiments are planning to include tracking data in their low-level triggers to help achieve the coverage and resolution necessary to keep the rate under control in the most difficult physics channels [8, 9]. The corresponding hardware, however, poses big challenges in terms of connectivity between the components, and the bandwidth required out of the detector to feed data into these systems is comparable or larger than the projected readout bandwidth. To mitigate these problems, intelligent data reduction at the front-end is being studied. Aside from considerations related to the long term maintenance of complex trigger hardware (see longevity above), these ideas are interesting in their own right as they indicate, in the author’s opinion, a potential evolution in detector design that might, in the long term, pave the way to full readout at the crossing rate or other applications. As an example, CMS is planning to introduce “trigger modules” in the tracker which autonomously identify high-$P_t$ track stubs by doing a coarse-grained pattern recognition (fig. 5) 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, taking advantage of special processor architectures (GPGPU) which lend themselves especially well to treat TPC data [11]. The pristine raw channel readouts become thus volatile.

2.2. High-speed links, silicon photonics, and all that
Brute-force full readout at the crossing rate 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. In this respect, silicon photonics promises unprecedented performance, if radiation-hard devices apt to work in the high-radiation environments of collider inner trackers can be devised. Both commercial and specific developments indicate that the field is highly active and could become relevant in the timescale of the LHC upgrades. Big industrial actors such as Intel or Molex are launching their silicon photonics products with speeds up to 1.6 Tbs (see for example [12]). At the same time, interest in the HEP community is growing for the potential of these technologies and

![Figure 5. 2S scheme for the CMS upgraded tracker: high $P_t$ tracks (stubs) can be identified if the cluster center in the top layer lies within a search window in $r - \phi$ (rows) [10].](image)
rad-hardness characterization studies begin to appear. These seem to indicate that commercial bundles using standard CMOS processes provide only moderate overall radiation tolerance, but transceivers themselves are likely to tolerate doses in excess of 1 Mrad [13].

These high-bandwidth links could make low-power readout at the crossing rate a possibility already in the next decade. The fact that they are also bidirectional would make it possible to bring significant bandwidth into the detector, like calibration constants and fast controls, thus enabling the operation of intelligent front-end electronics. The traditional distinction between readout, trigger and control data channels would then become increasingly artificial, paving the way to the possibility of running fully programmable algorithms at on- or near-detector electronics. As an example, the “retina” algorithm discussed in [14] lends itself particularly well to be implemented as an intelligent interconnect among tracker layers. On the other hand, only significant progress in radiation-hard very large scale integration will allow to fully exploit this opportunity.

Meanwhile, other technologies for the readout are being studied. Free-space optical links, for example, using on- or off-silicon photonics, offer some advantages for interlayer communication, by removing, among others, the problem of long term radiation damage of the optical fibers and have been proposed, together with opto-electromechanical devices, for an elegant implementation of intelligent front-end interconnects [15]. Use of 60 GHz wireless data transfer is also being considered. For example [16, 17] propose to use solid-state antennae connecting within the detector free space. One of the many technical challenges for these techniques is the minimization of cross-talk due to reflections. Even more exotic, an example of a prototype detector with wireless powering and wireless readout can be found in [18].

2.3. Are our data big data?

However much one wants to make of buzzwords, the “big data” industry seems to be here to stay. Massively parallel and distributed analysis of unstructured data has become ubiquitous in commercial applications. HEP is since a while no longer alone (or perhaps no longer altogether?) at the forefront of technology in manipulating large amounts of data, and perhaps not even any longer in analyzing them. Yet the future challenges faced by the field are formidable and must be confronted with limited resources. Apart for their attractiveness for use in monitoring of both detector parameters and data flow, as well as data analysis, big-data technologies indicate a possible evolutionary path for future DAQ and trigger architectures, and one that could enable the best use of available resources by focusing them on the aspects specific to our field.

As an example, one of the data mining and analytics mottos is “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 software (and hardware) technologies developed for data mining and search engines ? The principle could be applied to:

- build online “event directories” for quick scrutiny of data
- search the data we want in the detector (compare to “partial event building” of old)
- execute complex algorithms (e.g. pattern recognition) in parallel on a “distributed, searchable front-end”

Pushing these ideas even further, the inflexible notion of pre-processed datasets might give way to new forms of selection and analysis, 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.
3. Technology convergence in the post-LHC era

As discussed in the previous sections, major recent progress in data acquisition systems is characterized by the adoption of new trending technologies derived from other fields, and also by the adoption and use of conventional old and new technologies in slightly unconventional ways. In the past decade, HEP experiment have vastly profited from the development of large switch fabrics and interconnects for telecommunications. Currently, progress in areas such as high-bandwidth data transport and real-time analysis is driven by commercial applications in data mining/analytics and networks of machine learning. One of the most important applications which is surfacing in this area, albeit admittedly still in its infancy, carries the somewhat pompous monicker of “Internet of Things” (IoT). Quoting Wikipedia, this is “…the network of physical objects or ‘things’ embedded with electronics, software, sensors and connectivity to enable it to achieve greater value and service by exchanging data with the manufacturer, operator and/or other connected devices. Each thing is uniquely identifiable through its embedded computing system but is able to interoperate within the existing Internet infrastructure”. By and large, it consists of acquiring information from a large set of loosely interconnected, heterogeneous, intelligent sources, organising the said information in new ways, structured and unstructured, and analysing it to extract added value in the form of correlations, trends, patterns, which can be reused to improve the functioning of the intelligent objects themselves and to understand the system formed by the objects and their users in new ways. Spelled like this, the resemblance of IoT to our everyday job of acquiring information from many heterogeneous subdetectors, with their more or less intelligent front-end electronics, organising it in a coherent and intelligible fashion, using it to better calibrate the detectors, storing it and making it available for analysis, is striking. Figure 6 shows a fully data-driven triggerless DAQ architecture based on internet hardware and software services. Pattern-guided real time indexing takes place in the experiment “front-end”. The task of selecting useful data, calibrating the detector, and doing physics analysis are no longer distinct and are delocalised over large transnational networks of data centers. The HLT, for example, does not require a strictly controlled synchronous process and is limited only by the available bandwidth, not the local processing power. As such, researchers interested in a particular type of physics can submit their own particular selection as a “search” on the data acquisition system. The performances required for such a usage pattern would be anticipated by the data processing and data communication trends.

3.1. A holistic approach to detector and DAQ design

The ideas underpinning IoT are not all immediately applicable. It has been pointed out, among others [19], that embedded devices which can stand our radiation environments are not likely to hit the market, or have a market for that matter, any time soon. Which is why it is important to carefully consider the value of the approach and, if considered worthwhile, invest research and development efforts in areas, like this one, which are specific to the field. Clearly, in order to enable new approaches to all kind of problems, including the DAQ design, one cannot consider the problem in isolation. The design of a novel DAQ architecture for a future detector cannot be abstracted from the design of the detector readout, and ultimately from the design of the detector itself.

4. Conclusions

At the HEP energy frontier, the scale of new projects has never ceased to increase. At the same time, projects like the LHC span, nowadays, several generations of physicists. But the LHC would not exist if even earlier generations had not started thinking about it long before it became a realistic possibility. It would also not be a success if the people who imagined first and then realized the machine and the experiments had not been bold enough to risk novel solutions and unpaved paths. The next HEP project, if it is realized, will be even more challenging and
Figure 6. A cartoon of a triggerless data-driven DAQ architecture based on internet hardware and services.

will require even longer study, more preparation, and more faith. It is hard to predict if the technologies and ideas discussed here will have an impact on it, but exploring new ways without being certain of the success is, after all, what makes our job so much fun.

References

[1] Kovaric A 1919 *Physical Review* **13** 272
[2] Sinervo P K et al. 1989 *IEEE Trans. Nucl. Sci.* **36** 440–445
[3] Haynes W 1992 *DESY-92-129, C92-09-21*
[4] Gavillet P 1985 *Nucl. Instrum. Meth. A* **235** 363–379
[5] Barsotti E et al. 1990 *AIP Conf. Proc.* **209** 82
[6] Cittolin S et al. (CMS Collaboration) 2002 *CERN-LHCC-2002-026*
[7] Campana P et al. (LHCb Collaboration) 2014 *CERN-LHCC-2014-016*
[8] Shochet M et al. (ATLAS Collaboration) 2013 *CERN-LHCC-2013-007*
[9] Jeitler M et al. 2014 *JINST* **9** C08002
[10] Hall G et al. 2014 *Journal of Instrumentation* **9** C10034
[11] Buncie P et al. (ALICE Collaboration) 2015 *CERN-LHCC-2015-006*
[12] URL http://www.intel.com/content/www/us/en/research/intel-labs-silicon-photonics-mxc-connector.html
[13] Drake G et al. 2014 *Journal of Instrumentation* **9** C10037
[14] Abba A et al. 2015 *Journal of Instrumentation* **10** C03018
[15] Fernando W et al. 2012 *Phys. Procedia* **37** 1805–1812
[16] Dittmeier S et al. 2014 *JINST* **9**
[17] Pelikan D et al. 2014 *JINST* **9**
[18] De Lurgio P et al. arXiv:1310.1098v1
[19] Neufeld N *these proceedings*