The LHCb Online System

Niko Neufeld

CERN EP Division, Geneva Switzerland

For the LHCb Collaboration

Abstract

The online system of the LHCb experiment will be responsible for the acquistion of data, the distribution of trigger data and the control, configuration and monitoring of the whole experiment. The design of the system being completed, its main components and the technologies chosen for their implementation will be described.

The LHCb experiment at the LHC, CERN is a forward spectrometer dedicated to the study of the decay of B-mesons, which are produced mainly in the forward region in pp collisions at the LHC. Efficient triggering on B-mesons requires the reconstruction of detached decay vertices of these mesons. A four level trigger scheme has been devised therefore, which consists of a hardware trigger from calorimetry and high p_T muons (L0), a dedicated software vertex trigger (L1) and two higher level triggers, which are running on a large event filter CPU-farm. More on the trigger can be found in [1]. For optimal efficiency the average acceptance rate of the L1 trigger has been set to 40 kHz, which is the rate at which the whole detector has to be read out by the data acquisition system (DAQ).

The trigger decisions from the two first levels must be distributed within a fixed, short time to the front-end electronics buffers, where the event data are temporarily stored. This is the task of the timing and fast control system (TFC), which is the second main part of the LHCb online system.

Finally all the components of the detector, including DAQ and TFC systems must be configured, controlled and monitored. A single integrated system is responsble for this in LHCb, the Exeperiment Control System (ECS), which is the third main part of the online system.

The design of the system has been finished last year with the publication of the Technical Design Report (TDR) [2]. This paper reviews the main components, their functionalities and the technologies chosen for their implementation.

1 Data Acquistion System

The whole online system is depicted in Figure 1. The DAQ system occupies the central part of the figure. Upon a positive trigger decision all front-end electronics boards (top) send their data in the front-end multiplexer stage, where data are merged to optimise the load on the data-links. Data are pushed onwards to the read-out units (RU), which do further merging and send the fragments in a round-robin fashion to one of some 60 entry points in the event-filter CPU-farm. This is done by means of a standard Gigabit Ethernet switching network, the central element of which is a large Gigabit Ethernet switch, capable of handling over 120 Gigabits/s. The fragments belonging to a specific event are guaranteed to reach the same entry point into the event filter farm (EFF), the sub-farm controller (SFC). There they are assembled into one whole and sent off to one of the worker CPUs of the sub-farm. In the worker CPU the higher-level trigger algorithms operate now on the entire event and reduce the data to a finally accepted rate to storage of some 200 Hz. Conservatively assuming the raw data size as well as the size of the additional information from parial reconstruction to be 100 kB, we arrive at an average of 40 MByte/s, which permanently stored for offline analysis.

The protocol used for the data transport is a simple push-through protocol. There is no lateral synchronisation and no central event-mangager like entity. Every source sends data as soon as it is ready to do so, and it is assumed that the destination is always ready to receive. To avoid random loss of data due to buffer overflow, a throttling mechanism has been introduced. Components of the system can disable the trigger by sending a signal to the TFC system, which will stop further events from entering the system, until the condition clears. It is important to notice that this does not impose any bias, because the event, which caused the problem will still go through in any case - only later events will be prevented from being recorded.

The merging of data fragments, storing and forwarding them requires a technology capable of very efficiently manipulating Ethernet packets (frames). In LHCb network processors will be used for this purpose, which are very powerful new devices, developed for the networking industry to do packet processing at very high speeds in large backbone routers. LHCb has chosen a specific NP, the IBM NP4GS3 [3], which turns out to be excellently suited for the purposes of our system [4]:

- Merging up to seven incoming streams of fragments, each at 40 kHz, into one larger one, preserving the order, stripping of unecessary headers, and adding data transport history if necessary, while observing the constraints resulting from the usage of Ethernet
- Concatenating all ~ 60 fragments belonging to one event to one contiguous block of data

- Operate as a basic element of the central switching network, which can be built from network processors, should ecommical or technical reasons prevent us from using a commercial switch[1].

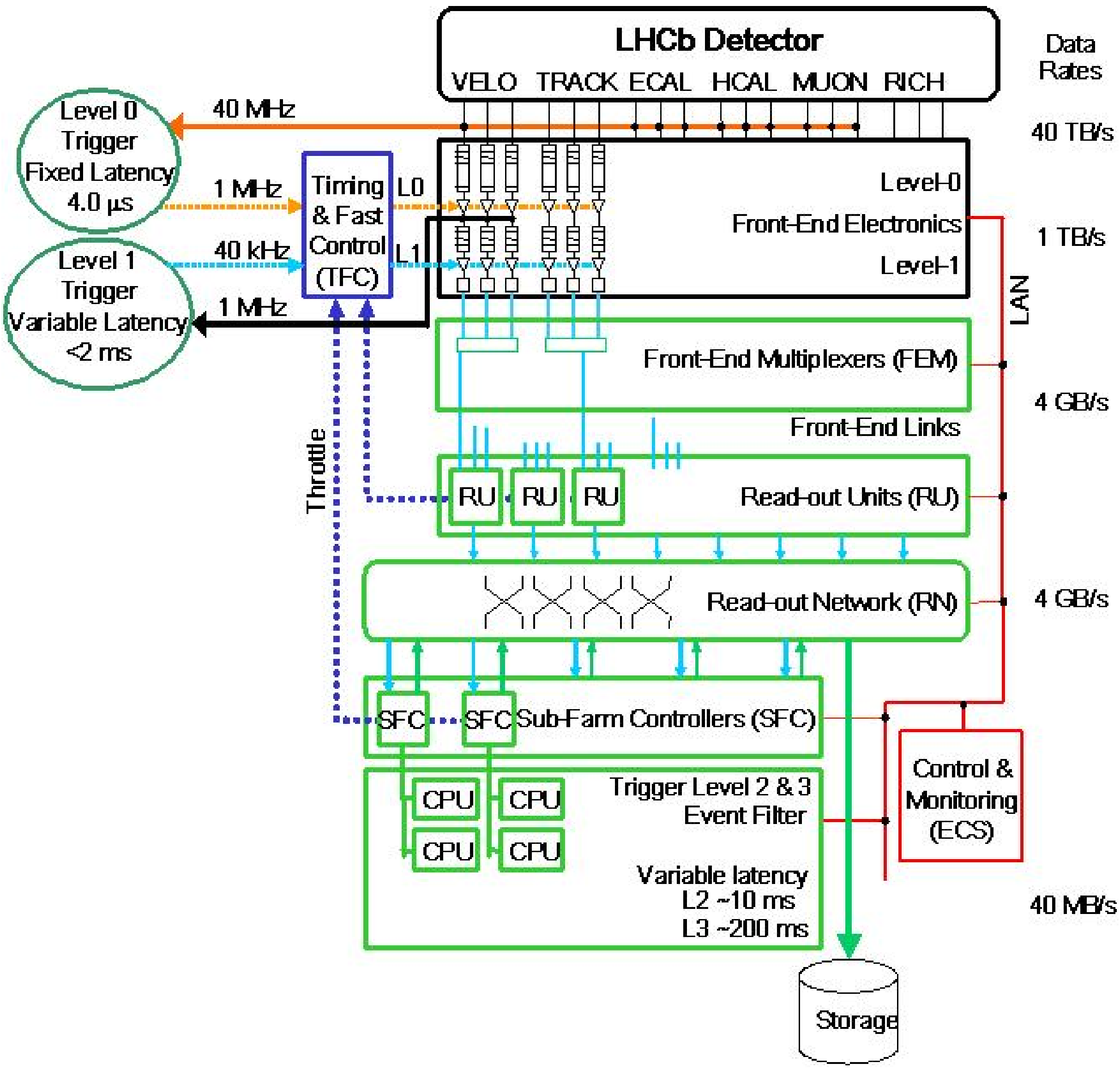

Figure 1: Schematic drawing of the LHCb online system

2 Timing and Fast Control System

The TFC system has two main tasks. First of all it has to distribute the trigger decisions to the front-end buffers at a rate of 1.1 MHz. To this end it makes use of the TTC components, which have been developed for timing and trigger distribution of the LHC experiments at CERN [5]. Secondly it protects

[1]The specific pattern of our network traffic, namely the arrival of many packets at almost the same moment at a single output port (because they all have the same destination), differs quite a bit from the standard "random" traffic, internet routers are geared at. Preliminary studies show that some commercial switches drop packets in such a situation

the system from buffer overflows. Its central module, the Readout Supervisor (RS), keeps track of the buffer occupancy of the front-end electronics and disables the trigger in time to avoid any overflow at this stage. Other parts of the system, which are less deterministic in their nature and can hence not be directly taken into account by the Readout Supervisor have access to a central throttle signal (by means of programmable ORs), which is then causing the RS to disable the trigger.

Last but not least, the TFC system also collects data about the triggering itself, which is added to the data-stream.

3 Experiment Control System

The Experiment Control System will be in charge of monitoring, configuring and controlling the entire experiment, which includes the control of the DAQ ("run-control"), the detector hardware ("slow-control") but also the event-filter farm, the environment at the experimental site (temparature, humidity, airflow...), the communication with outside systems like the accelerator, the safety system and so on. To have only one system for all these tasks obviously facilitates maintainance because common structures appear over and over again, usage by a usually non-expert shift-crew, because user interfaces and system response are coherent throughout and also construction because it maximises the possibility of re-use of software and hardware components.

LHCb participates in the common project for control systems for the LHC experiments at CERN, JCOP [6]. One of the main parts of JCOP is to provide a common software framework to build control systems. The backbone of this framework is a commercial SCADA (Supervisory Control and Data Acquisition) system, PVSS-II. This is a networked, multi-platform system, which allows the implementation of a large distributed control system with the necessary user interfaces, data bases and communications mechanisms.

The ECS is a hierarchical system which implements each controled entity as a potentially self-contained unit endowed with local intelligence and capable of independent control. The possibility to separate the control of certain parts of the detector from the rest of the hierarchy is essential for *partioning*. Partitioning is the indendent, concurrent operation of sub-systems of the experiment, usually (combinations of) sub-detectors. Partitioning is essential for quick problem isolation and tracking without compromising overall operational efficiency of the experiment.

The hierarchical structure of the ECS is depicted in Figure 2. Starting from the bottom one notices the leaf nodes, which are called *devices*. These devices communicate directly either with hardware (power supplies, valves, etc.) or to software processes (e.g. a trigger algorithm). They collect the status informantion of their attached system publish it to their parent nodes and transmit commands. Devices can be grouped for convenience or logically ("the gas-system of sub-detector X"). Each sub-system node is able to sequence and automate operations, it provides automatic error recovery, it filters and handles alarms, and it allows a user interface to be attached directly. Further up in the hierarchy we find the detector control and data acquistion of entire sub-detectors, which are again controlled by parents, which are the detector control and data acquistion of the entire experiment. As can be seen, these two together with a set of external systems interface ultimately to a unique controlling entity. Throughout the systems alarms and status information are transmitted from the bottom of the hierarchy upwards and commands are flowing downwards.

The front-end hardware of the ECS differs with the place in the experiment. Equipment in radiation sensitive areas is controlled by the radiation hard SPECS [7]. Equipment in areas, where radiation is not an issue, are controlled by tiny embedded PCs, which are interfaced to the ECS via a standard Ethernet network. Like in the data acquisition system there are no shared bus systems anywhere, which, as experience from previous experiments shows, greatly facilitates diagnostics and keeps problems with the controls hardware isolated.

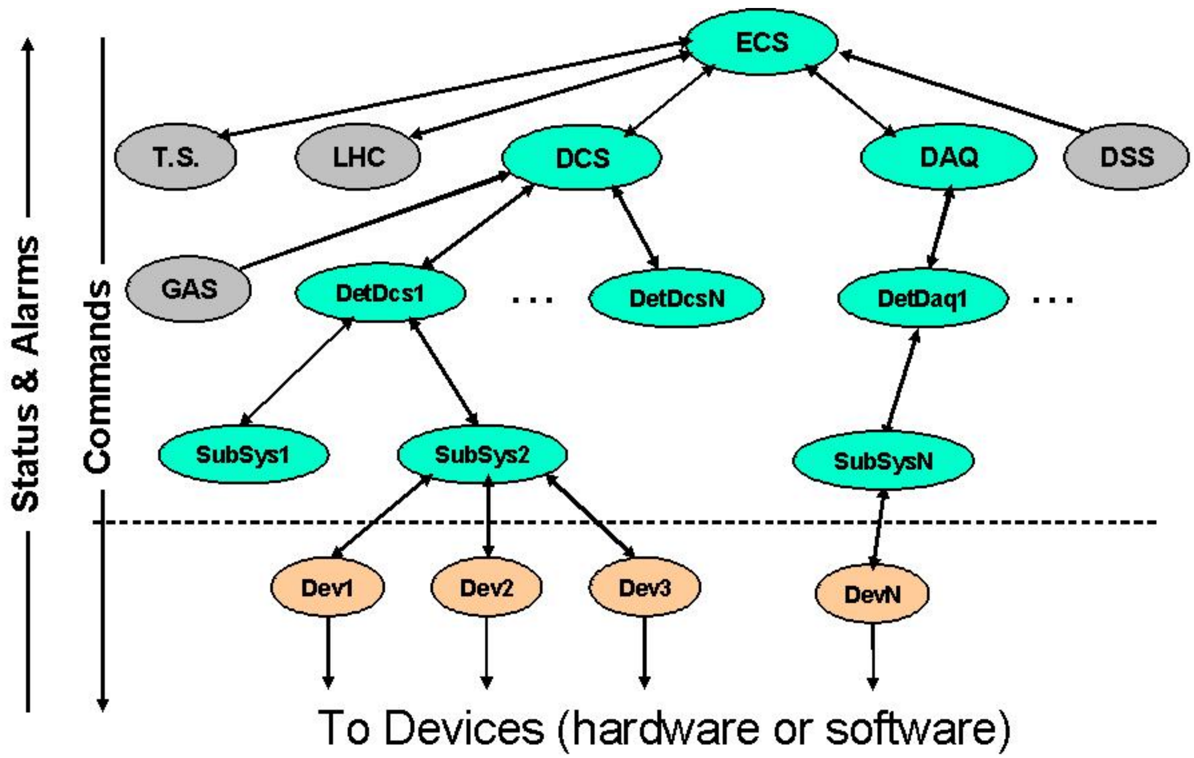

Figure 2: The hierarchical structure of the Experiment Control System.

4 Summary

The LHCb online system has three main components, the DAQ, TFC and ECS. In design great care has been taken to obey the following principles

- Uniformity and homogeneity
- Strict separation of data and control paths
- Usage of commercially available components whereever possible
- Usage of cheap, standard technologies like Ethernet and PCs
- Simple protocols over point-to-point links
- Scalability by avoiding a single orchestrating element

While all the key components exist or can be bought already today, the entire system will be fully operational for the commissioning of the detector before the startup of the LHC in 2007.

References

[1] H. Dijkstra, *The LHCb Trigger System*, these proceedings

[2] LHCb Coll. *LHCb Online System TDR.* CERN/LHCC 2001-040, Geneva

[3] IBM PowerNP NP4GS3 Datasheet, online

`http://www-3.ibm.com/chips/techlib/techlib.nsf/products/IBM_PowerNP_NP4GS3`

[4] B. Jost and N. Neufeld, *A versatile Network Processor based electronics module for the LHCb Data Acquisition system* LHCb internal note, LHCb 2001-132

[5] The Timing and Trigger Control System for the LHC Detectors, home-page http://ttc.web.cern.ch/TTC/intro.html

[6] LHC Experiments Joint COntrols Project home-page http://itcowww.cern.ch/jcop/

[7] SPECS home-page http://www.lal.in2p3.fr/technique/se/pub/lhcb/PCItoSPECS/index.html