Large-scale DAQ tests for the LHCb upgrade

Antonio Falabella, *INFN CNAF Bologna,* Matteo Manzali, *Università di Ferrara,* Francesco Giacomini, *INFN CNAF Bologna,* Umberto Marconi, *INFN Bologna,* Balasz Voneki, *Cern,* Niko Neufeld, *Cern,* Sebastian, Valat *Cern*

***Abstract*—The Data Acquisition (DAQ) of the LHCb experiment[1] will be upgraded in 2020 to a high-bandwidth trigger-less readout system. In the new DAQ event fragments will be forwarded to the to the Event Builder (EB) computing farm at 40 MHz. Therefore the front-end boards will be connected directly to the EB farm through optical links and PCI Express based interface cards. The EB is requested to provide a total network capacity of $32\mathrm{Tb/s}$, exploiting about 500 nodes. In order to get the required network capacity we are testing various technology and network protocols on large scale clusters. We developed on this purpose an Event Builder implementation designed for an InfiniBand interconnect infrastructure. We present the results of the measurements performed to evaluate throughput and scalability measurements on HPC scale facilities.**

I. Introduction

The LHCb experiment is one of the four main experiments performed with LHC at Cern. It aims at the study of CP violation and precision measurements of rare decays of b and c quark hadrons.
A major upgrade of the detector is foreseen during the second long shutdown of the LHC (2018-2019). A new trigger-less[2] readout system will operate at the bunch crossing frequency of $40\mathrm{MHz}$. The Event Builder (EB) is the part of the readout system responsible for the collection of the event fragments from the sub-detector readout boards. These fragments are then combined to form the full event. To operate at the aforementioned frequency the EB will be redesigned. The idea is to implement it with an high throughput LAN using off-the-shelf hardware. Possible technologies are InfiniBand, Ethernet or Omni-Path.

II. Readout system architecture

The schema of the upgraded readout system is shown in 1. The front-end boards are connected to the EB nodes through optical links. The EB LAN consists of ~ 500 nodes connected through an high-throughout fabric technology. Each node must be able to communicate at $100\mathrm{Gbit/s}$ full-duplex. The event fragments from the sub-detectors are collected to create the full event ready to be sent to storage for further processing.

A. Event builder design

The relevant numbers to design the event builder are summarized in I In order to test the possible implementation of the EB using the InfiniBand standard we developed prototype software evaluators in a collaboration between INFN and Cern. The idea is to compare the performance of the two different implementation and compare their scalability. In 2 the main blocks of the DAQPIPE implementation developed at Cern are shown. They consist of a Readout Unit(RU) which emulates the front-end boards and consumes the event fragments generated by the Generator, the Builder Unit(BU) is then selected by Event Manager(EM) to assemble the event fragments into the complete event. Except for the EM each

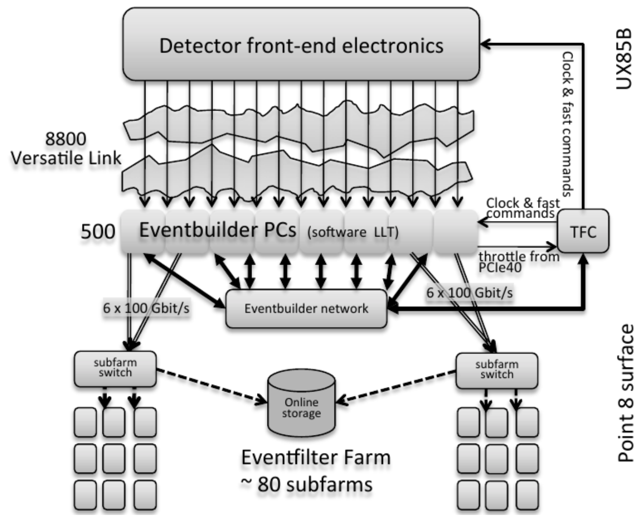

Fig. 1. The architecture of upgraded LHCb readout system

TABLE I
Relevant numbers to consider for the EB design

Event frequency	30 MHz
Average event size	100 KBytes
I/O of single node	100 Gbit/s (full-duplex)

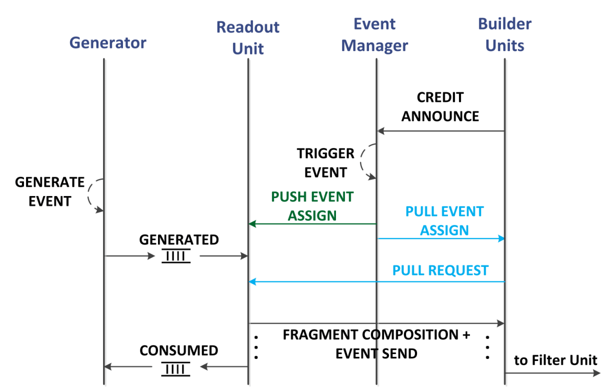

Fig. 2. LHCb DAQPIPE main blocks

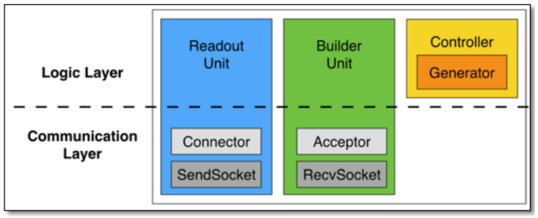

Fig. 3. LSEB main blocks

EB node runs any of the this elements. A different approach is used in the LSEB implementation which implement both the RU the BU but instead of choosing the BU with a EM a round-robin mechanism is chosen. The building blocks are shown in 3. Both implementations separate cleanly the EB logic from the transport in order to allow to test different technologies.

III. Scalability tests

As mentioned above the EB software must operate at the scale of ~ 500 nodes so, in order to exploit the performances of our implementation we tested them at the CINECA consortium Galileo cluster[5]. The main features of this cluster are summarized in II. The plot in 4 shows the average bandwidth measured by the BU for an increasing number of nodes. The test has been performed up to 128 nodes. As can be seen the performance are similar for the two implementations, the bandwidth being $> 50\%$ for every test. A slight drop of performances can be accounted to the concurrent usage of the network of the other processes running on the cluster.

TABLE II
Relevant numbers to consider for the EB design

Nodes	516
Processors	8-cores Intel Haswell 2.40 GHz (2 per node)
Network	InfiniBand QDR

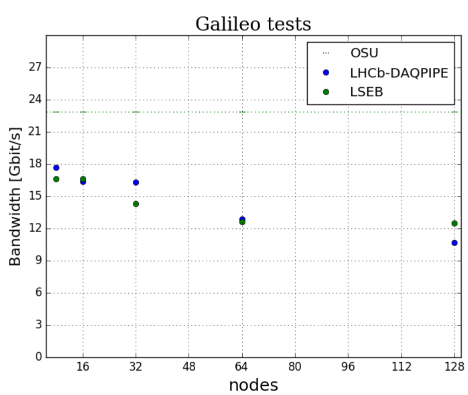

Fig. 4. Bandwidth performance comparison between the two implementations.

IV. Conclusion

The LHCb will under a major upgrade during the second long shutdown. Apart from detector upgrades also the readout system will redesigned in order to allow for a trigger-less data acquisition. The EB can be realized using commercial network technologies such as InfiniBand. We realized performances evaluators in order to exploit the design possibilities of the EB. The scalability tests performed with the Galileo cluster show that the implementations we realized are viable. The slight decrease of the performances can be accounted to the fact that the cluster is a production cluster, so the network usage is not exclusive.

References

[1] LHCb Collaboration, *The LHCb detector at the LHC* JINST,3,2008,S08005
[2] LHCb Collaboration, *LHCb Trigger and Online Upgrade Technical Design Report* CERN,LHCC,2014,C10026
[3] LHCb Collaboration, *The Upgrade of the LHCb Trigger System* JINST,9,2014,C10026,
[4] Antonio Falabella, *The 40 MHz trigger-less DAQ for the LHCb Upgrade* Nuclear Inst. and Methods in Physics Research,A,2016,pp. 280-283
[5] http://www.hpc.cineca.it/content/galileo