DAQ Architecture for the LHCb Upgrade

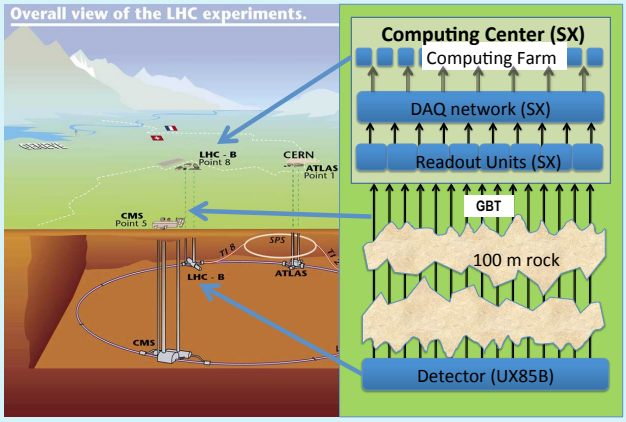
Guoming Liu, Niko Neufeld
CERN, Switzerland
niko.neufeld@cern.ch

Introduction to LHCb Upgrade

The LHCb collaboration has proposed a major upgrade to allow operation at the a luminosity of $2x10^{33}$ $cm^{-2}s^{-1}$ after 2018. One of the major upgrades is to allow readout of the entire detector at 40 MHz. In order to improve the trigger efficiencies, LHCb adopts a flexible, full software-based trigger solution. All data need to be sent to a large computing farm for event building and filtering. According to the simulation, the total data rate to the farm is about 38.4 Tb/s.

DAQ Upgrade

The DAQ system is based on a local area network. For each collision (we call each collision an event), each Readout Unit (RU) reads out data from the front-end electronics through direct GigaBit Transceiver (GBT) links. A RU is normally implemented in a custom electronics board based on an FPGA . The RU sends the data fragment to a Builder Unit (BU) through the DAQ network. The BU assembles all fragments belonging to this event and sends the complete event to a Filter Unit (FU) for event filtering. Finally the FU sends the selected events to the storage system. Normally, BU and FU are implemented as software processes in the computing farm.

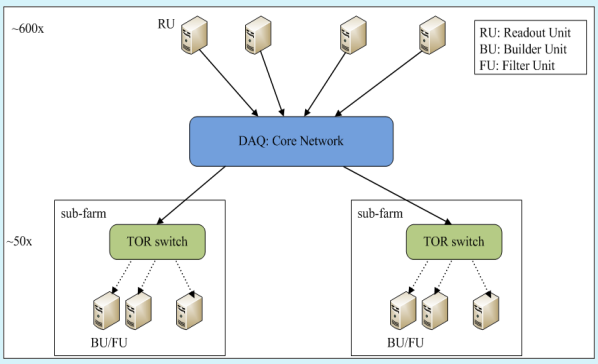


Several solutions are feasible to build such a large network. Studies are required to build a cost-effective and reliable DAQ system. A few principles shall be followed:

- Simple single-stage data-flow
- Minimize the number of expensive ports in the core network layer.
- Deploy all components (RU, BU, FU) close to each other in the computing center as much as possible to keep technological options open and minimize the number of expensive optical components.
- Use the most efficient technology for different connections.

Solution 1: unidirectional

- BU & FU are implemented in the same server
- RUs are connected to the core network
- Data flow in the core network is unidirectional

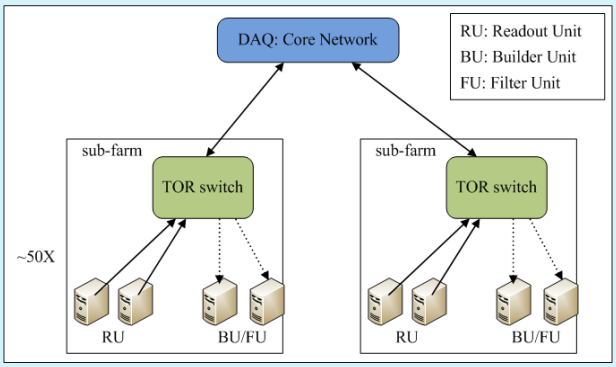


Advantages:
- Simple architecture: unidirectional data flow.
- High reliability: simple function in each components

Disadvantages:
- High cost: the core network devices are very expensive

Solution 2: bidirectional data flow

- BU & FU are implemented in the same server
- RUs and BU/FU servers are connected to the access layer i.e. TOR switch
- Data flow in the core network is bidirectional

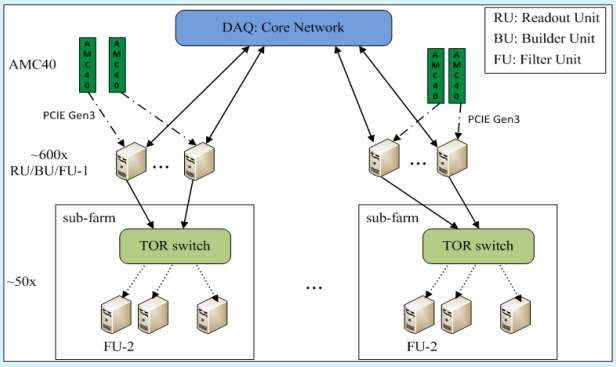


Advantages:
- makes use of bisectional bandwidth of the core network, can save up to 50% of bandwidth and ports in the core network. The cost per port in the core network is usually 3 ~ 4 times more expensive than in a TOR switch

Disadvantages:
- RUs and BU/FUs need to be close enough to connect the same TOR switch.

Solution 3: bidirectional with uniform RU/BU

- RU, BU and part of the FU (FU-1) are implemented in the same server
- The server receives data from the readout board AMC40 via PCIe bus.
- The server is equipped with 2 network interfaces: one is connected to the core network for event building, the other one is connected to the TOR switch for event filtering.
- Data flow in the core network is bidirectional



Advantages:
- makes use of bisectional bandwidth of the core network
- Network technologies for event building and event filtering can be different, which allows to choose the best solution and make the decision as late as possible. A combination of high-speed InfiniBand for the event building and lower speed 10Gbase-T Ethernet for event filtering is the most cost effective solution today.
- Allows simple architecture with minimum buffering in the readout board AMC40. Some complexities have been shifted to the event-building servers.

Disadvantages:
- Increases the complexity in the event-building servers
- Lots of I/O required in the event-building servers

Conclusion

Several network architectures have been identified. The bidirectional solution with uniform RU/BU currently looks like the most cost-effective architecture.