

# Technical workshop on InfiniBand for Trigger/DAQ

CERN Jan 14<sup>th</sup> to 15<sup>th</sup>, 2013

## ***Introduction & Welcome***

Lorne Levinson (Weizmann Institute of Science)

Niko Neufeld (CERN)

# Welcome

- About 50 registered participants
  - from all LHC experiments, the CERN IT department and from “outside” CERN
- Experts from Mellanox, the “parents” of InfiniBand – thanks for coming to Geneva for two days

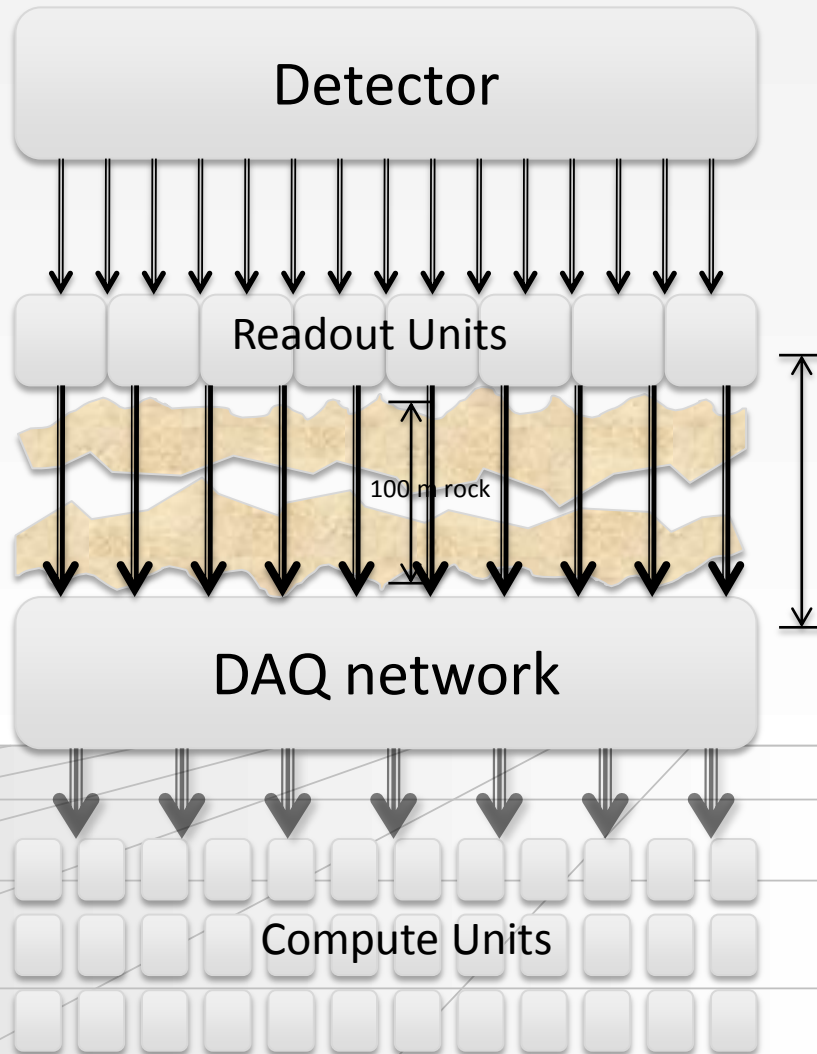
# Some goals of this work-shop

- We want to learn (more) about InfiniBand
- We want to understand if and how it can be used for building large network-based DAQ systems
- We want to exchange ideas about ways how we think IB could be used, how IB is already used and how IB will be used
- We want to explain our challenge to the makers of InfiniBand (Mellanox), in particular how we are different from traditional users of IB in the HPC community

# Format

- Two sessions, both with technical presentations by Mellanox and presentations about (present and future) applications in HEP
- The work-shop dinner
- Time for ad-hoc meetings with Mellanox experts on Tuesday

# The DAQ challenge



- Collecting data from the detector into multiple devices (RUs)
- Sending pieces from all RUs to the same unit in a server-farm
- While doing so cover a significant physical distance (more than in a typical data-center)
- And of course: as cheap and as reliable as possible
- Aggregated amount of data
  - present  $O(100)$  GB/s
  - future  $O(10)$  TB/s

# Why InfiniBand?

- It's... well, cheap – price for FDR switch-port ~ 300 USD (street-price) today
- It has high bandwidth (about 52 Gbit/s usable for FDR)
- It has a reliable, message-oriented remote DMA layer, which is very low-cost on the sending/receiving hosts, if used properly (low cost in terms of memory bandwidth (0-copy) and CPU load (offload by the IB-adapter))
- It's used widely in HPC, which is a community not entirely different from "us" (quite a few HEP labs / universities, not CERN, actually are involved in HPC as well)

# The devil's advocate: why **not** InfiniBand?



- InfiniBand is (almost) a single-vendor technology
- In a very bursty traffic typical for a DAQ system, can the InfiniBand flow-control cope with it while preserving high bandwidth utilization
- InfiniBand is mostly used in HPC, where hardware is regularly and radically refreshed (unlike IT systems like ours which are refreshed and grow over time in an evolutionary way). Is there not a big risk that the HPC guys jump onto a new technology and InfiniBand is gone?
- The InfiniBand software stack seems awfully complicated compared to the Berkeley socket API in TCP/IP. What about tools, documentation, tutorial material to train engineers, students, etc...?

# Practicalities

- Introductory material:
  - <https://indico.cern.ch/conferenceDisplay.py?confId=218156> (at the bottom of the page: “Material”)
  - Verbs programming:  
<http://thegeekinthecorner.wordpress.com/category/infiniband-verbs-rdma/>
- Access to LHCb IB test-cluster: use login and password sent to you by email. In case you need one and/or have problems please contact [niko.neufeld@cern.ch](mailto:niko.neufeld@cern.ch) or [enrico.bonaccorsi@cern.ch](mailto:enrico.bonaccorsi@cern.ch)
- Workshop dinner: Monday, Jan 14<sup>th</sup> 19:30
  - Restaurant – Pizzeria de la Place  
286, rte de Meyrin, Meyrin  
3<sup>rd</sup> tram stop after CERN, Tram 18, “Meyrin Village”