

LHCb Run3 Online Storage

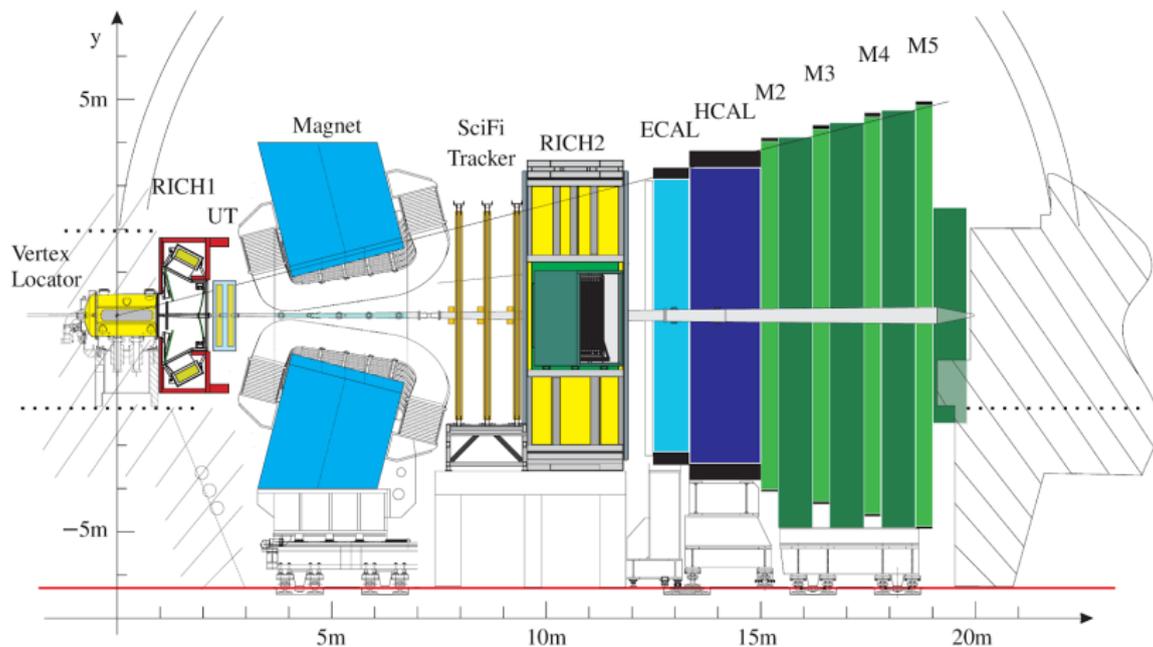


Beat Jost
Tommaso Colombo
Francesco Sborzacchi
Rainer Schwemmer
Niko Neufeld

CERN, EP

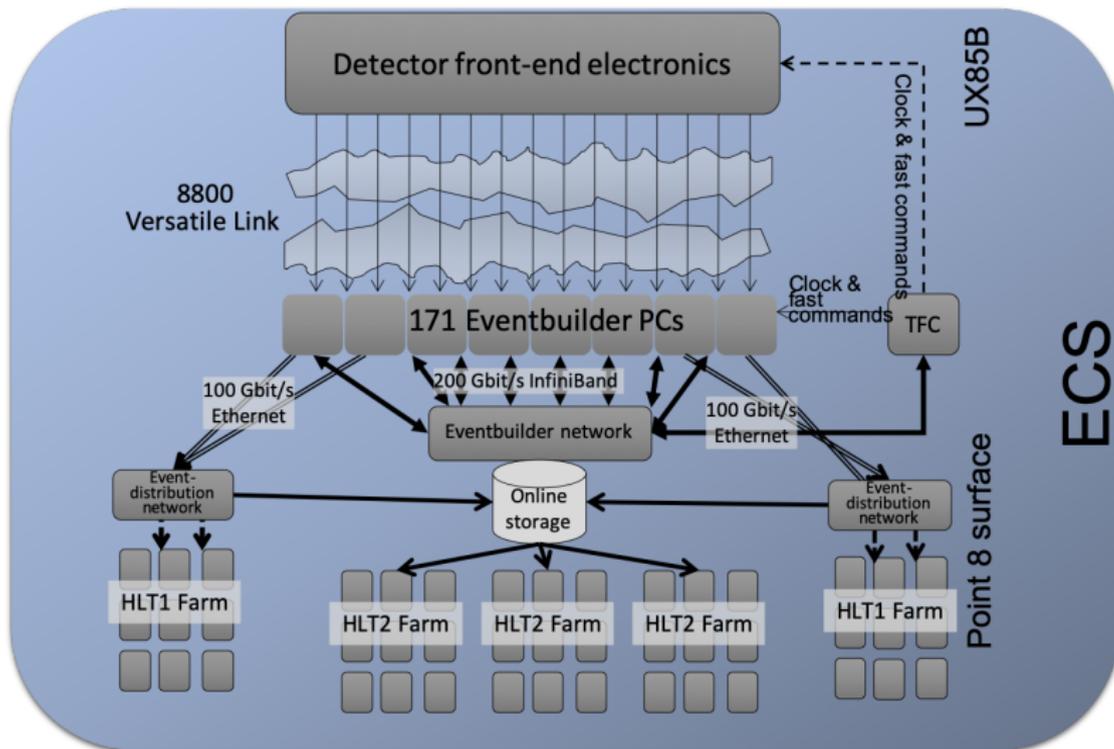
EOS Workshop
Feb. 3rd to 5th 2020
CERN, Geneva, Switzerland

The LHCb Experiment

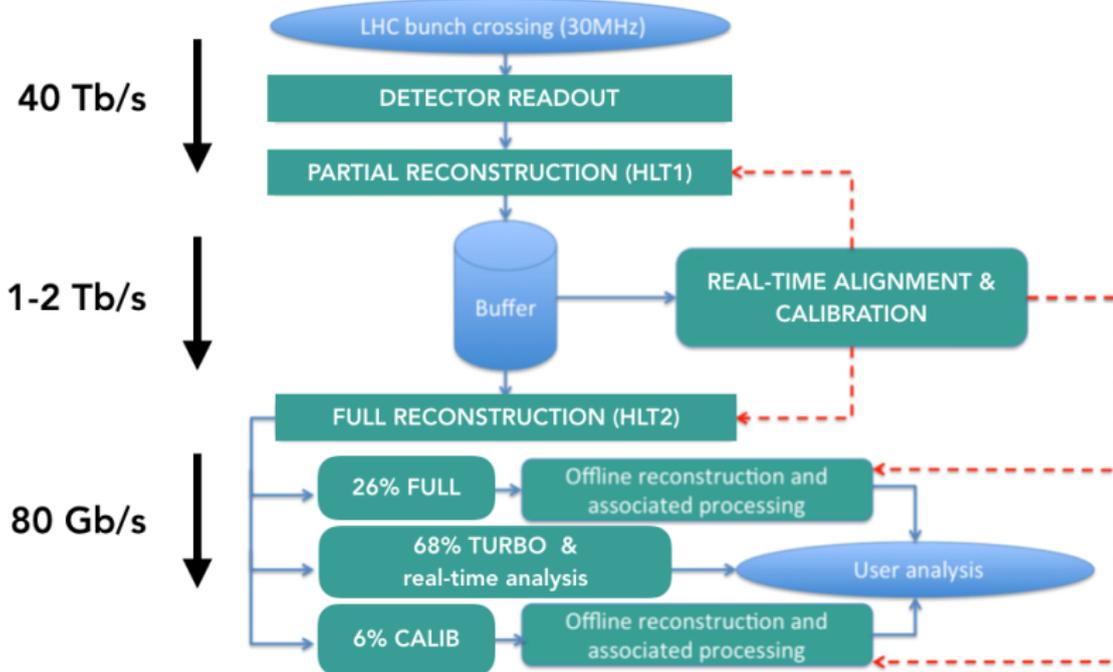


- ▶ In the LHC proton-bunches collide every 25 ns
- ▶ LHCb will read out the *entire* detector for every collision
- ▶ Aggregated data from one collision are approximately 100 kB in size
- ▶ The data arrive on ~ 10000 optical fibres
- ▶ Data from all fibres need to be collected into a single location ("event-building")
- ▶ There is a two-stage software filtering call High Level Trigger 1, HLT1 and HLT2
- ▶ HLT1 and HLT2 are coupled by a large disk-buffer (\rightarrow this talk)

Hardware-centric view of LHCb data-flow



Logical view of LHCb data-flow and -filtering



- ▶ Sequential read / write, no updates
- ▶ in principle: write once, read once, delete reliably
- ▶ Data reside on buffer for at least 2 hrs
- ▶ Minimum buffer depth is for 3 days
- ▶ Deeper buffer allows better use of CPU resources over the year
- ▶ No redundancy required
- ▶ No need for full POSIX semantics
- ▶ Initially 500 kHz HLT1 output rate corresponding to 50 GB/s (2021) written to storage
- ▶ Upgrade-able to 2 MHz corresponding to 200 GB/s written to storage
- ▶ Buffer must be available at all times to avoid back-pressure on DAQ (writing) and optimally use CPU resources (reading)
- ▶ Mechanisms should be in place to avoid reading data twice by mistake

- ▶ A priori a use-case for spinning drives (or even tapes - not economical), SMR is fine?
- ▶ Number of spindles drives system-size
- ▶ Hardware will be NAS boxes with JBODs
- ▶ Easiest would be a cluster-file system which "can just do it", but we would settle for a common name-space
- ▶ Is a flash-pool useful, and if useful is it possible?
- ▶ Storage doesn't need to be "on-site" at Point-8
- ▶ Can we do something together?