Event Building
Event Building: collection and formatting of all the data elements of an event into a single unit

- normally last step before high-level trigger or storage
- can be implemented on buses, can use custom interconnects, can be based on (Ethernet) network

Network-based EB is choice of all LHC experiments and a case study for networking in DAQ
Network switch: crossbar

➔ Each input port can potentially be connected to each output port

➔ At any given time, only one input port can be connected to a given output port

➔ Different output ports can be reached concurrently by different input ports
Network switch: crossbar

➔ Ideal situation → all inputs send data to different outputs

No interference (Congestion)

All input ports send data concurrently
Crossbar switch: event building

➔ EB workload implies converging data flow
  • all inputs want to send to same destination at the same time

➔ “Head of line blocking”
  • congestion
Congestion

- Well known phenomena..
  - in Geneva and other cities
- Differently from road traffic, Ethernet HW is allow to “drop” packets
  - Higher level protocols have to take care of re-sending
  - Possibly important performance impacts
Adding input and output FIFO dramatically improve the EB pattern handling

EB workload anyway problematic

- FIFO size is limited, variable data size
- Limited internal switching speed

Traffic shaping or Network over-sizing
LHC experiments
Sometime impossible to take a proper decision in a single place

- too long decision time
- too far
- too many inputs

Distribute the decision burden in a hierarchical structure

- usually $\tau_{N+1} \gg \tau_N$, $f_{N+1} \ll f_N$

At the DAQ level, proper buffering must be provided for every trigger level

- absorb latency
- de-randomize
LHC DAQ phase-space
LHC experiments have $O(10^7)$ channels operating at 40 MHz (25 ns) → 40 TB/s

In addition, interesting phenomena are extremely rare

$$\sigma_H / \sigma_{Tot} \sim O(10^{-13})$$

Events are complex

- significant number of overlapping collisions (pile-up $\mu$)

Experiments are large (O(10 m))

Multi-level trigger system and …
LHC L1 Trigger and FE electronics

- Particle time of flight $>> 25$ ns
- Cable delays $>> 25$ ns

Dedicated synchronization, timing and signal distribution facilities

- Typical L1 decision latency is $O(\mu s)$
  - dominated by signal propagation in cables

Digital/analog custom front-end pipelines store information during L1 trigger decision.
LHC: After L1?

Custom hardware L1 trigger and front-end electronics followed by network-based High-Level Trigger farm(s)

- commercially available HW organized in a farm
  - events are independent

Connection between custom section and the network-based one achieved via dedicated HW and point-to-point connectivity

- electrical or optical, standard or custom
### Read-out links at the LHC (in Run 1)

<table>
<thead>
<tr>
<th>Link Type</th>
<th>Bandwidth</th>
<th>Approx. Links</th>
<th>Flow Control</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SLINK</strong></td>
<td>Optical: 160 MB/s</td>
<td>≈ 1600 Links</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Receiver card interfaces to PC.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **SLINK 64**      | LVDS: 400 MB/s (max. 15m)   | ≈ 500 links   | yes          |
|                   | (FE on average: 200 MB/s to readout buffer) |               |              |
|                   | Receiver card interfaces to commercial NIC |               |              |
|                   | (Network Interface Card)     |               |              |

| **DDL**           | Optical 200 MB/s            | ≈ 500 links   | yes          |
|                   | Half duplex: Controls FE (commands, Pedestals, Calibration data) |               |              |
|                   | Receiver card interfaces to PC |               |              |

| **TELL-1 & GbE Link** | Copper quad GbE Link         | ≈ 400 links   | no           |
|                       | Protocol: IPv4 (direct connection to GbE switch) |               |              |
|                       | Forms “Multi Event Fragments” |               |              |
|                       | Implements readout buffer    |               |              |
ATLAS HLT Farm
CMS
Long Shutdown 1: TDAQ Perspective

<table>
<thead>
<tr>
<th>Phase 0</th>
<th>LS1</th>
<th>Run 2</th>
<th>LS2</th>
<th>Phase I</th>
<th>LS3</th>
<th>Phase II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run 1</td>
<td></td>
<td>(Prepare Run 2)</td>
<td></td>
<td>Run 3</td>
<td></td>
<td>Run 4</td>
</tr>
<tr>
<td>(Prepare Run 2)</td>
<td></td>
<td>Ultimate luminosity</td>
<td></td>
<td>(Prepare Phase I)</td>
<td></td>
<td>(Prepare Phase II)</td>
</tr>
<tr>
<td>Consolidation</td>
<td></td>
<td>$\sqrt{s} = 13\text{~to}14 \text{ TeV}$</td>
<td></td>
<td>Ultimate luminosity</td>
<td></td>
<td>HL-LHC</td>
</tr>
<tr>
<td>$L_{\text{inst}} \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$</td>
<td></td>
<td>25 ns bunch spacing</td>
<td></td>
<td>$L_{\text{inst}} \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu \sim 27$</td>
<td></td>
<td>$L_{\text{inst}} \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$</td>
<td></td>
<td>$\mu \sim 55\text{~to}81$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\int L_{\text{inst}} \sim 50 \text{ fb}^{-1}$</td>
<td></td>
<td>$\int L_{\text{inst}} &gt; 350 \text{ fb}^{-1}$</td>
<td></td>
<td>$\int L_{\text{inst}} \sim 3000 \text{ fb}^{-1}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

⇒ LHC data-acquisition system backbones installed ~2007

- during Run 1 → stability, efficiency, performance reach and optimization

⇒ LS1 was occasion to

- upgrade core systems and review architectures
- introduce new technologies, retire obsolete ones
- follow changes on the detector side
- prepare for challenges of Run2 (and Run3)
Long Shutdown 1: TDAQ Perspective

<table>
<thead>
<tr>
<th>Year</th>
<th>Phase 0 Run 1</th>
<th>LS1</th>
<th>Run 2</th>
<th>LS2</th>
<th>Phase I Run 3</th>
<th>LS3</th>
<th>Phase II Run 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>(Prepare Run 2)</td>
<td></td>
<td>(Prepare Phase I)</td>
<td></td>
<td>(Prepare Phase II)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>Consolidation</td>
<td>LS1</td>
<td></td>
<td>LS2</td>
<td></td>
<td>LS3</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td></td>
<td>2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td></td>
<td>2015</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td></td>
<td>2017</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td></td>
<td>2019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2021</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2022</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2023</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2024</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2025</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2026</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2027</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2028</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2029</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2030</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- LHC data-acquisition system backbones installed ~2007
  - during Run 1 → stability, efficiency, performance reach and optimization

- LS1 was occasion to
  - upgrade core systems and review architectures
  - introduce new technologies, retire obsolete ones
  - follow changes on the detector side
  - prepare for challenges of Run2 (and Run3)
Run 2: Challenges

➔ Increased pile-up

- larger data size → bandwidth and storage
- more complex events → increased computing needs, trigger efficiency and rejection power

<table>
<thead>
<tr>
<th># Vertexes</th>
<th>CMS</th>
<th>ATLAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(Some) Run 2 updates

➔ Merge L2 and L3 into a single HLT farm
  • preserve Region of Interest, but diluted the farm separation and fragmentation
  • increased flexibly, computing power efficiency

➔ No architectural changes, but
  • all network technologies replaced
    – Myrinet → Ethernet
    – Ethernet → Infiniband
  • file-based event distribution in the farm
    – achieve full decoupling between DAQ and HLT
Looking forward to LS2 and beyond

On the long term, all experiments looking forward to significant increase in L1 trigger rate and bandwidth. ALICE and LHCb will pioneer this path during LS2.

➔ First level trigger for Pb-Pb interactions $500 \text{ Hz} \rightarrow 50 \text{ kHz}$

➔ 22 MB/event
  - 1 TB/s readout $\rightarrow 500 \text{ PB/month}$

➔ Data volume reduction
  - on-line full reconstruction
  - discard raw-data

➔ Combined DAQ/HLT/offline farm
  - COTS, FPGA and GPGPU

➔ 1 MHz $\rightarrow 40 \text{ MHz}$ readout and event building $\rightarrow$ trigger-less
  - trigger support for staged computing power deployment

➔ 100 kB/event
  - on-detector zero suppression $\rightarrow$ rad-hard FPGA
  - 4 TB/s event-building
Almost The End
What I did not talk about …

➔ Many many topics

• Run Control → Steering the DAQ, Finite State Machine
• Configuration → Storing, distributing and archiving SW, HW and trigger configuration
• Monitoring → The quality of the data, the state of the detector, the functionality of the DAQ

➔ Your chance of hearing about these and much more and learn through practice …
Sixth edition of the International School of Trigger and Data Acquisition will be held in February 2016 and hosted by Weizmann Institute.

http://isotdaq.web.cern.ch/isotdaq/isotdaq/Home.html
The End
References

➔ Lectures and papers from H. Spieler
  • http://www-physics.lbl.gov/~spieler/

➔ Lecture at ISOTDAQ schools
  • http://isotdaq.web.cern.ch/isotdaq/isotdaq/Home.html

➔ Of course, previous Summer Student courses
  • http://indico.cern.ch/scripts/SSLPdisplay.py?stdate=2011-07-04&nbweeks=7

➔ DAQ@LHC Workshop
  • http://indico.cern.ch/scripts/SSLPdisplay.py?stdate=2011-07-04&nbweeks=7