Overview of the *Phase II* DAQ Architectures of ATLAS and CMS

G. Lehmann Miotto CERN/EP-DT
**DAQ System Overview**

- Provide an overall framework for configuration, control and monitoring of the experiment (online software)
- Distribute timing, trigger and fast commands to detector front-end
- Readout detector data after hardware trigger
- Build events from fragments based on unique event identifier
- Store data for further selection
- Serve data to the High Level Trigger (HLT)
- Forward accepted events to permanent storage
  - Build data files
ATLAS and CMS in Phase II

- High luminosity LHC increases the number of interactions per bunch crossing to ~200
- Hardware based, fast triggers loose part of their selectivity
  - Increased rate of accepted events
- New/upgraded detectors
  - More readout channels
  - Harsher radiation environment
  - Larger event size
The HL-LHC DAQ Challenge

- Scale up to control/monitoring infrastructure by ~1 order of magnitude
- Increase readout rate to 750-1000 kHz
  - Today 100 kHz
- Increase the overall throughput to ~50 Tb/s
  - Today ~2 Tb/s
- Increase rate of data to permanent storage to ~7.5-10 kHz
  - Today ~1 kHz
- ATLAS and CMS both need a substantial upgrade of their respective DAQ systems
  - Relying on experience gained during Run 1-3
High level view of CMS DAQ
High level view of ATLAS DAQ
ATLAS Detector Interface

- Support for different protocols
- Hooks for detector specific FW
- Hooks for detector specific SW
CMS Detector Interface

Common FW blocks in detector specific BE
## The readout

<table>
<thead>
<tr>
<th>CMS</th>
<th>ATLAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATCA based</td>
<td></td>
</tr>
<tr>
<td>BE boards + DAQ &amp; Timing hub</td>
<td></td>
</tr>
<tr>
<td>Data sent to surface over reliable protocol (e.g. TCPIP) to I/O processors</td>
<td></td>
</tr>
<tr>
<td>Detector specific BE boards embed DAQ firmware for data transmission via front panel</td>
<td></td>
</tr>
<tr>
<td>PCIe based</td>
<td></td>
</tr>
<tr>
<td>FELIX cards in servers + data handler servers</td>
<td></td>
</tr>
<tr>
<td>Integrated interface to event building</td>
<td></td>
</tr>
<tr>
<td>Limited hooks for detector specific firmware blocks in FELIX card; Data Handler software application embedding detector specific software</td>
<td></td>
</tr>
</tbody>
</table>
Timing, control and BUSY

<table>
<thead>
<tr>
<th>CMS</th>
<th>ATLAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Leaf of distribution over DTH</td>
<td>- Leaf of distribution over FELIX</td>
</tr>
<tr>
<td>- Timing &amp; control distributed over backplane through custom serial data stream</td>
<td>- Timing &amp; control distributed over point-to-point optical links (lpGBT) to detector electronics</td>
</tr>
<tr>
<td>- Status monitoring for BUSY logics collected over the backplane</td>
<td>- BUSY ON/OFF collected over point-to-point optical links</td>
</tr>
</tbody>
</table>
Layout of CMS DAQ

- Trigger Processors
  - Global Trigger
- TTC/TTS
  - TCDS/EVM
- 560x100 Gbs data links
- 40 32x100 Gbs switches
- 500 IO servers
- ~500x200 Gbs switch
- ~100 Tbs bisection bandwidth
  - 64 32x100 HLT switches
- ~5 PB local storage
  - 60 GBs access
- Storage
- Event networks
  - Data to Surface 200m fibers
  - Data to Surface routers
- DTH
- 120 ATCA crates
- Trigger and detector data: ~50,000 x 1-10 Gbps GBT links
- SMTS
- CDR/T0
- HLT PC farms/clouds
  - ~9.2 MHS06
Layout of ATLAS DAQ
Event building and temporary storage

**CMS**
- Baseline 200 Gb/s interconnects (e.g. Infiniband HDR)
- I/O servers receiving data from readout (TCP), acting both as event building sources and destinations
  - Folded event builder
- Complete events stored in high speed storage to serve HLT (~1 min)

**ATLAS**
- Baseline 100 Gb/s interconnects (e.g. Ethernet)
- Data injected directly from readout
- Introduction of logical event building, i.e. data fragments stored in large distributed high performance storage (~1h)
  - Event builder is the bookkeeper
High Level Trigger

- Using respective offline sw frameworks
  - Algorithms are outside DAQ scope

- Manage large, heterogeneous farm
  - In ATLAS additional hardware based tracking for trigger
  - Possible introduction of accelerators

- Estimate power, space & network needs
  - Based on estimates of required CPU HS06
    - ATLAS 4.5 MHS06 + HTT, CMS 9.2 MHS06
    - Assume different models of evolution of computing power

- Continuous re-assessment and simulations
  - Large impact on need for refurbishing data center infrastructure at experiments
Storage of accepted events

- Both ATLAS and CMS have modular, scalable systems in place to aggregate accepted events into files and transfer them to Tier0
  - Based on Hard Disk Drives

- ATLAS foresees to use the same distributed storage system used for temporary buffering before HLT for this task in Phase II

- CMS will decide on the implementation of this subsystem later
Online software

- Both ATLAS and CMS have mature online sw frameworks
  - General evolution of sw model, introduction of new technologies and third party tools

CMS
- Increase homogeneity across systems
- Creation a uniform stack providing DAQ and TCDS functionality for test-bench setups, standalone single-crate systems, small scale DAQ systems (aka miniDAQs), and global DAQ

ATLAS
- Introducing containers
- New cluster orchestrator for HLT farm
- Service oriented operational monitoring
- Upgrade of event sampling system
Summary

- The DAQ systems for Phase II rely on the experience accumulated over many years of data taking
  - Hardware technologies
  - Communication protocols
  - Software designs & tools

- The main difference in approach between the two experiments is at the level of the detector interfaces

- Most components need to be re-implemented to sustain the increased event size and HW trigger rate
  - Active R&D is ongoing in ATLAS and CMS, sometimes in common (e.g. large distributed key-value store as temporary storage before HLT filtering)

- Only the baseline ideas were presented for both ATLAS and CMS
  - Looking into open options there is even more commonality