Operational performance of the ATLAS trigger and data acquisition system and its possible evolution

Andrea Negri\textsuperscript{1}

on behalf of the ATLAS TDAQ collaboration\textsuperscript{2}

\textsuperscript{1}University and INFN Pavia

\textsuperscript{2}https://cdsweb.cern.ch/record/1386334

May 21, 2012
TDAQ Architecture

- Three selection levels
  - Level 1 on custom h/w
  - High Level Triggers (Level 2 & Event Filter) on computer farms
On Level 1 accept (latency 2.5 $\mu$s):
- Data pushed to buffers hosted on ReadOut System PCs (ROS)
- Region Of Interest sent to L2
- **Level 2 (latency \(\sim 40\) ms)
  - Selection based on Region of Interest concept
  - Only few % of event data pulled via Data Collection network

**Diagram:**
- **Event rates design (2011 peak):**
  - 40 MHz (20 MHz)
  - 75 kHz (~65 kHz)

**Trigger**
- **Level 1**
  - Custom Hardware
- **Regions of Interest**
  - ~5000 Processing Unit

**Level 1 Accept**
- **ROI data**
- **DAQ**
  - Calo/Muon
  - Other
  - Other

**Data rates design (2011 peak)**
- ATLAS Event 1.5MB/25 ns
  - (1.2 MB/50 ns)

**Data Flow**
- \(~110\) GB/s
  - (~ 85 GB/s)
Event Builder

- Pull data from Data Collection network
- Output full events to Back-End network

Event rates design (2011 peak)

- 40 MHz (20 MHz)
- 75 kHz (~65 kHz)
- 3 kHz (~5.5 kHz)

Region of Interest

Custom Hardware

Level 1

< 5000

Processing Unit

Level 2

ROI data

L2 Accept

DC

Level 1 Accept

OD

Event Builder

Data rates design (2011 peak)

- ATLAS Event 1.5MB/25 ns (1.2 MB/50 ns)
- ~ 110 GB/s (~ 85 GB/s)
- ~ 4.5 GB/s (~ 6.5 GB/s)
Event Filter (latency \( \sim 1 \) s)
- Full event reconstruction
- Accepted events sent to Data Logger farm
TDAQ Architecture

- **Data Logger**
  - Save events in streams (files)
  - Files asynchronously transferred to Tier 0
In 2011 some systems running beyond design specification
- Event Builder
- Data Logger

Event rates design (2011 peak)

40 MHz (20 MHz)
75 kHz (~65 kHz)
3 kHz (~5.5 kHz)
~ 200 Hz (~600 Hz)

Data rates design (2011 peak)

ATLAS Event 1.5MB/25 ns (1.2 MB/50 ns)
~ 110 GB/s (~85 GB/s)
~ 4.5 GB/s (~6.5 GB/s)
~ 300 MB/s (~700 MB/s)
Data taking 2011

- 31 weeks of p-p operations
- \( \sqrt{E} = 7 \) TeV
- Continuous luminosity increase
  - \( \mathcal{L}_{\text{peak}} = 3.42 \times 10^{33} \text{cm}^{-2} \text{s}^{-1} \)
  - \( \mathcal{L}_{\text{int}} = 4.9 \text{ fb}^{-1} \)
- Bunch cross every 50 ns instead of 25
  - Higher pile-up
- Overall TDAQ efficiency \( \sim \) 94%
  - 2.7 PB of data recorded (5.7M files)

- 4 weeks of Pb-Pb operation
  - \( \mathcal{L}_{\text{peak}} = 5.12 \times 10^{26} \text{cm}^{-2} \text{s}^{-1} \)
  - \( \mathcal{L}_{\text{int}} = 160 \mu \text{b}^{-1} \)
31 weeks of p-p operations

$\sqrt{E} = 7$ TeV

Continuous luminosity increase

- $\mathcal{L}_{\text{peak}} = 3.42 \times 10^{33} \text{cm}^{-2}\text{s}^{-1}$
- $\mathcal{L}_{\text{int}} = 4.9 \text{fb}^{-1}$

Bunch cross every 50 ns instead of 25

Higher pile-up

Overall TDAQ efficiency $\sim 94\%$

- 2.7 PB of data recorded (5.7M files)

4 weeks of Pb-Pb operation

- $\mathcal{L}_{\text{peak}} = 5.12 \times 10^{26} \text{cm}^{-2}\text{s}^{-1}$
- $\mathcal{L}_{\text{int}} = 160 \mu\text{b}^{-1}$
TDAQ operation 2011

- HLT farm increased with LHC performance
  - 16 new racks (+50%)
- Balance issues promptly addressed
  - To hide h/w heterogeneity, EB-EF system configuration moved from a sliced system to a flat (random) mapping of EF nodes to EB ones
- L2 vs EF rack sharing configurable run-by-run

Sharing tool developed

(Poster session, id 91)
Preventative maintenance
- Replaced all Event Builder nodes
- Rolling replacement of ROS MBs (75/153)

Major 2011 operational issue
- Network cards failures in replaced ROS nodes
- Workaround: installed different network cards

New functionalities
- E.g.: Missing $E_T$ at L2
  - Special request for calo ROSes
  - Extracting missing $E_T$ information directly from front end boards
  - Expensive requests for ROS

Run control (Details in previous talk)
- Improved automation of our DAQ monitoring and control system
- Improved stop-less and automatic recovery procedures
Preventative maintenance
- Replaced all Event Builder nodes
- Rolling replacement of ROS MBs (75/153)

Major 2011 operational issue
- Network cards failures in replaced ROS nodes
- Workaround: installed different network cards

New functionalities
- E.g.: Missing $E_T$ at L2
  - Special request for calo ROSes
  - Extracting missing $E_T$ information directly from front end boards
  - Expensive requests for ROS

Run control (Details in previous talk)
- Improved automation of our DAQ monitoring and control system
- Improved stop-less and automatic recovery procedures
TDAQ operation 2011

- Preventative maintenance
  - Replaced all Event Builder nodes
  - Rolling replacement of ROS MBs (75/153)

- Major 2011 operational issue
  - Network cards failures in replaced ROS nodes
  - Workaround: installed different network cards

- New functionalities
  - E.g.: Missing $E_T$ at L2
    - Special request for calo ROSes
    - Extracting missing $E_T$ information directly from front end boards
    - Expensive requests for ROS

- Run control (Details in previous talk)
  - Improved automation of our DAQ monitoring and control system
  - Improved stop-less and automatic recovery procedures
2011/2012 shutdown activities

- ROS rolling replacement continued
  - HLT farm
    - 12 new racks replaced 16 old ones
    - Now: ~ 1600 nodes (Mother-Boards)
    - Most racks (36) configurable as L2 or EF on run by run basis
  - Back-End network upgraded
    - Installed second core router for redundancy
    - As for DC network
  - Tests
    - At the peak operating conditions expected during 2012
    - Predict possible bottlenecks

<table>
<thead>
<tr>
<th>CPU model</th>
<th>Cores/node</th>
<th>Racks</th>
<th>Nodes</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>E5420</td>
<td>8</td>
<td>11</td>
<td>341</td>
<td>L2/EF</td>
</tr>
<tr>
<td>E5540</td>
<td>8</td>
<td>14</td>
<td>448</td>
<td>EF</td>
</tr>
<tr>
<td>X5650</td>
<td>12</td>
<td>25</td>
<td>904</td>
<td>L2/EF</td>
</tr>
</tbody>
</table>
2011/2012 shutdown activities

- ROS rolling replacement continued
- HLT farm
  - 12 new racks replaced 16 old ones
  - Now: ~1600 nodes (Mother-Boards)
  - Most racks (36) configurable as L2 or EF on run by run basis

- Back-End network upgraded
  - Installed second core router for redundancy
  - As for DC network

- Tests
  - At the peak operating conditions expected during 2012
  - Predict possible bottlenecks

<table>
<thead>
<tr>
<th>CPU model</th>
<th>Cores/node</th>
<th>Racks</th>
<th>Nodes</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>E5420</td>
<td>8</td>
<td>11</td>
<td>341</td>
<td>L2/EF</td>
</tr>
<tr>
<td>E5540</td>
<td>8</td>
<td>14</td>
<td>448</td>
<td>EF</td>
</tr>
<tr>
<td>X5650</td>
<td>12</td>
<td>25</td>
<td>904</td>
<td>L2/EF</td>
</tr>
</tbody>
</table>
2011/2012 shutdown activities

- ROS rolling replacement continued
- HLT farm
  - 12 new racks replaced 16 old ones
  - Now: \(\sim 1600\) nodes (Mother-Boards)
  - Most racks (36) configurable as L2 or EF on run by run basis

- Back-End network upgraded
  - Installed second core router for redundancy
  - As for DC network

- Tests
  - At the peak operating conditions expected during 2012
  - Predict possible bottlenecks

<table>
<thead>
<tr>
<th>CPU model</th>
<th>Cores/node</th>
<th>Racks</th>
<th>Nodes</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>E5420</td>
<td>8</td>
<td>11</td>
<td>341</td>
<td>L2/EF</td>
</tr>
<tr>
<td>E5540</td>
<td>8</td>
<td>14</td>
<td>448</td>
<td>EF</td>
</tr>
<tr>
<td>X5650</td>
<td>12</td>
<td>25</td>
<td>904</td>
<td>L2/EF</td>
</tr>
</tbody>
</table>
2011/2012 shutdown activities

- ROS rolling replacement continued
- HLT farm
  - 12 new racks replaced 16 old ones
  - Now: ~1600 nodes (Mother-Boards)
  - Most racks (36) configurable as L2 or EF on run by run basis

- Back-End network upgraded
  - Installed second core router for redundancy
  - As for DC network

- Tests
  - At the peak operating conditions expected during 2012
  - Predict possible bottlenecks

<table>
<thead>
<tr>
<th>CPU model</th>
<th>Cores/node</th>
<th>Racks</th>
<th>Nodes</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>E5420</td>
<td>8</td>
<td>11</td>
<td>341</td>
<td>L2/EF</td>
</tr>
<tr>
<td>E5540</td>
<td>8</td>
<td>14</td>
<td>448</td>
<td>EF</td>
</tr>
<tr>
<td>X5650</td>
<td>12</td>
<td>25</td>
<td>904</td>
<td>L2/EF</td>
</tr>
</tbody>
</table>
\[ \sqrt{E} = 8 \text{ TeV} \]

**Impressive LHC start-up**
- 80\% of the expected peak luminosity in few weeks

<table>
<thead>
<tr>
<th>( \mathcal{L} ) ( \times 10^{33} \text{ cm}^{-2} \text{s}^{-1} )</th>
<th>Done</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta^* [m] )</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>( \text{Bunches} )</td>
<td>1082</td>
<td>1331</td>
</tr>
<tr>
<td>( p/\text{bunch} \times 10^{11} )</td>
<td>1.2</td>
<td>1.65</td>
</tr>
<tr>
<td>(&lt; \mu &gt;)</td>
<td>29.8</td>
<td>35</td>
</tr>
</tbody>
</table>

- Overall TDAQ efficiency 93.6\%
  - Comparable to last year
- Bunch crossing still 50 ns
  - Pile-up a major concern
Pile-up 2012: CPU usage

- Processing time linear scaling verified up to $\langle \mu \rangle \sim 22$
- Extrapolating to $\langle \mu \rangle = 35$
  - 25% CPU margin shared across L2&EF
- Extrapolation uncertainties: trigger menu, ROS collection time
- CPU usage evolution is being surveyed
Pile-up 2012: CPU usage

- Processing time linear scaling verified up to $\langle \mu \rangle \sim 22$
- Extrapolating to $\langle \mu \rangle = 35$
  - 25% CPU margin shared across L2&EF
- Extrapolation uncertainties: trigger menu, ROS collection time
  - CPU usage evolution is being surveyed

![Average Pileup vs. Average L2 Processing Time (ms)](chart1)

![Average Pileup vs. Average EF Processing Time (ms)](chart2)
Pile-up 2012: event size

- Pile-up dependency for some detectors (E.g.: Inner)
  - Evolution largely linear
    - Future deviations cannot be excluded
  - Extrapolation for $\langle \mu \rangle$ up to 35
    - Event size up to $\sim 1.8$ MB
- We may face limited operational margins at peak luminosity
- Additional EB capacity to be deployed to meet peak demand
- Data Logger capacity to be increased
  - additional h/w or
  - increase b/w into existing h/w
Pile-up 2012: event size

- Pile-up dependency for some detectors (e.g.: Inner)
- Evolution largely linear
  - Future deviations cannot be excluded
- Extrapolation for $\langle \mu \rangle$ up to 35
  - Event size up to $\sim 1.8$ MB
- We may face limited operational margins at peak luminosity
- Additional EB capacity to be deployed to meet peak demand
- Data Logger capacity to be increased
  - additional h/w or
  - increase b/w into existing h/w
Pile-up 2012: event size

- Pile-up dependency for some detectors (E.g.: Inner)
- Evolution largely linear
  - Future deviations cannot be excluded
- Extrapolation for $\langle \mu \rangle$ up to 35
  - Event size up to $\sim 1.8$ MB
- We may face limited operational margins at peak luminosity
  - Additional EB capacity to be deployed to meet peak demand
  - Data Logger capacity to be increased
    - additional h/w or
    - increase b/w into existing h/w
Pile-up 2012: event size

- Pile-up dependency for some detectors (E.g.: Inner)
- Evolution largely linear
  - Future deviations cannot be excluded
- Extrapolation for $\langle \mu \rangle$ up to 35
  - Event size up to $\sim 1.8$ MB
- We may face limited operational margins at peak luminosity
- Additional EB capacity to be deployed to meet peak demand
- Data Logger capacity to be increased
  - additional h/w or
  - increase b/w into existing h/w

Andrea Negri (ATLAS TDAQ)
Pile-up 2012: ROSes

- ROS performance can be limited by:
  - Access rate
  - Bandwidth
  - Load (not a problem for new h/w)

- ROS parameters are being surveyed

- Motherboards of Transition Radiation Tracker ROSes recently replaced

![Diagram showing request rate, max rate, max load, and max bandwidth for different devices]
**Pile-up 2012: ROSes**

- ROS performance can be limited by:
  - Access rate
  - Bandwidth
  - Load (not a problem for new h/w)

- ROS parameters are being surveyed
  - Motherboards of Transition Radiation Tracker ROSes recently replaced

---

**Diagram:**
- **Max Rate [kHz]**
- **Max Load**
- **Max Bandwidth [MB/s on DC1/2]**

Andrea Negri (ATLAS TDAQ)
Long shutdown

- First occasion for major hardware and software upgrades
- Define a s/w scalable model to be used in 2014 and beyond
- Profit from experience from past and ongoing data-taking
  - Build-in further scalability and flexibility

Current assumptions for 2014
- 100 kHz L1 rate
- 1 kHz average physics output rate
  - Extension of the Data Logger capacity
  - Provide online data compression for a more efficient use of resources
- 25 ns bunch crossing
  - But be prepared for 50 ns:
    learn as much as possible this year on high pile up operation
Long shutdown

- First occasion for major hardware and software upgrades
- Define a s/w scalable model to be used in 2014 and beyond
- Profit from experience from past and ongoing data-taking
  - Build-in further scalability and flexibility

Current assumptions for 2014
- 100 kHz L1 rate
- 1 kHz average physics output rate
  - Extension of the Data Logger capacity
  - Provide online data compression for a more efficient use of resources
- 25 ns bunch crossing
  - But be prepared for 50 ns:
    learn as much as possible this year on high pile up operation
Data taking confirmed the success of the current design
... and stimulated interest to explore possible evolutions
- Simplify CPU and network resources balancing
- Reduce complexities
- Simplify HLT steering
## Data Flow Evolution

- Merge L2, EB, EF within a single homogeneous system
  - A single farm
  - In each node:
    - RoI based processing $\rightarrow$ event building $\rightarrow$ full event processing
  - Possibility to have a single network

---

**Event rates design (2011 peak)**

- 40 MHz (20 MHz)
- 75 kHz (~65 kHz)
- 3 kHz (~5.5 kHz)
- ~ 200 Hz (~600 Hz)

---

**Trigger**

- Level 1: Custom Hardware
  - ~5000 processing units
  - ROI data
  - L2 Accept

- Level 2: Event Filter
  - ~15000 processing units
  - Full Event
  - EF Accept

---

**DAQ**

- Calo/Muon
- Other
- ROD

---

**Data Flow**

- ~110 GB/s (~85 GB/s)
- ~4.5 GB/s (~6.5 GB/s)
- ~300 MB/s (~700 MB/s)
- Merge L2, EB, EF within a single homogeneous system
  - A single farm
  - In each node:
    - RoI based processing → event building → full event processing
  - Possibility to have a single network
Data Flow Evolution

- A single HLT homogeneous farm
- On each HLT node
  - One Data Collection Manager (DCM) in charge of data collection, caching and integrity
  - Multiple Processing Units (HLTPUs) in charge of event selection
  - Communication via shared memories
- A single SuperVisor (HLTSV) distributes L1 results to HLT nodes
  - Must sustain 100 kHz (otherwise multiple HLTSVs)
  - Possibility to merge HLTSV with a s/w based RoIB under evaluation
- Data Loggers receive events from DCMs and store them to disk
- ROS application unchanged
Data Flow Evolution

- A single HLT homogeneous farm
- On each HLT node
  - One Data Collection Manager (DCM) in charge of data collection, caching and integrity
  - Multiple Processing Units (HLTPUs) in charge of event selection
  - Communication via shared memories
- A single SuperVisor (HLTSV) distributes L1 results to HLT nodes
  - Must sustain 100 kHz (otherwise multiple HLTSVs)
  - Possibility to merge HLTSV with a s/w based RoIB under evaluation
- Data Loggers receive events from DCMs and store them to disk
- ROS application unchanged
Data Flow Evolution

- A single HLT homogeneous farm
- On each HLT node
  - One Data Collection Manager (DCM) in charge of data collection, caching and integrity
  - Multiple Processing Units (HLTPUs) in charge of event selection
  - Communication via shared memories
- A single SuperVisor (HLTSV) distributes L1 results to HLT nodes
  - Must sustain 100 kHz (otherwise multiple HLTSVs)
  - Possibility to merge HLTSV with a s/w based RoIB under evaluation
- Data Loggers receive events from DCMs and store them to disk
- ROS application unchanged
Data Flow Evolution: advantages

- **Simpler Data Flow configuration**
  - Only 5 application types (were 9)

- Automatic CPU balance on each HLT node

- Automatic HLT system balance
  - No need to pre-determine the L2/EF sharing

- No additional contributions to fragment lifetime inside Read Out Buffers
  - ROS cleared after RoI based processing or EB

- Reduced ROS load
  - All event fragments only requested once from a ROS
  - Less network connections (one per HLT node)

- HLT selection still based on RoI

- A single HLT steering instance
Data Flow Evolution: advantages

- Simpler Data Flow configuration
  - Only 5 application types (were 9)
- Automatic CPU balance on each HLT node
- Automatic HLT system balance
  - No need to pre-determine the L2/EF sharing
- No additional contributions to fragment lifetime inside Read Out Buffers
  - ROS cleared after RoI based processing or EB
- Reduced ROS load
  - All event fragments only requested once from a ROS
  - Less network connections (one per HLT node)
- HLT selection still based on RoI
- A single HLT steering instance
Data Flow Evolution: advantages

- Simpler Data Flow configuration
  - Only 5 application types (were 9)
- Automatic CPU balance on each HLT node
- Automatic HLT system balance
  - No need to pre-determine the L2/EF sharing
- No additional contributions to fragment lifetime inside Read Out Buffers
  - ROS cleared after RoI based processing or EB
- Reduced ROS load
  - All event fragments only requested once from a ROS
  - Less network connections (one per HLT node)
- HLT selection still based on RoI
- A single HLT steering instance
Data Flow Evolution: advantages

- Simpler Data Flow configuration
  - Only 5 application types (were 9)
- Automatic CPU balance on each HLT node
- Automatic HLT system balance
  - No need to pre-determine the L2/EF sharing
- No additional contributions to fragment lifetime inside Read Out Buffers
  - ROS cleared after RoI based processing or EB
- Reduced ROS load
  - All event fragments only requested once from a ROS
  - Less network connections (one per HLT node)
- HLT selection still based on RoI
- A single HLT steering instance
No need to create and transport L2 Result
ROS access and data unpacking done only once
Flexibility for HLT strategies and to exploit DF resources
Different strategies under evaluation (depending on the needs)
  - Minimize L2 latency (giving time to more complex algorithms)
    - Change the chains/steps execution model and re-order the chains
  - Minimize ROS access rate, by optimizing EB request
    - Choose the best time of EB moving algorithms between L2 & EF
Data Flow Evolution: single HLT steering

- No need to create and transport L2 Result
- ROS access and data unpacking done only once
- Flexibility for HLT strategies and to exploit DF resources
- Different strategies under evaluation (depending on the needs)
  - Minimize L2 latency (giving time to more complex algorithms)
  - Change the chains/steps execution model and re-order the chains
  - Minimize ROS access rate, by optimizing EB request
  - Choose the best time of EB moving algorithms between L2 & EF
Data Flow Evolution: single HLT steering

- No need to create and transport L2 Result
- ROS access and data unpacking done only once
- Flexibility for HLT strategies and to exploit DF resources
  - Different strategies under evaluation (depending on the needs)
    - Minimize L2 latency (giving time to more complex algorithms)
      - Change the chains/steps execution model and re-order the chains
    - Minimize ROS access rate, by optimizing EB request
    - Choose the best time of EB moving algorithms between L2 & EF
Data Flow Evolution: single HLT steering

- No need to create and transport L2 Result
- ROS access and data unpacking done only once
- Flexibility for HLT strategies and to exploit DF resources
- Different strategies under evaluation (depending on the needs)
  - Minimize L2 latency (giving time to more complex algorithms)
    - Change the chains/steps execution model and re-order the chains
  - Minimize ROS access rate, by optimizing EB request
    - Choose the best time of EB moving algorithms between L2 & EF
Data Flow Evolution: design

- Design phase ongoing
  - First implementation to be ready for the end of the run
- Looking for common solutions, minimizing code duplication
  - A common framework for all the applications
- Different s/w technologies under evaluation
  - Profit from experience
  - But with an open attitude toward new ideas and views
- Prototype available for testing design and spot problems
  - Current applications adapted to the proposed design
  - Developed 2 years ago and integrated in the current release
  - Tested on ATLAS TDAQ system
Design phase ongoing
- First implementation to be ready for the end of the run

Looking for common solutions, minimizing code duplication
- A common framework for all the applications

Different s/w technologies under evaluation
- Profit from experience
- But with an open attitude toward new ideas and views

Prototype available for testing design and spot problems
- Current applications adapted to the proposed design
- Developed 2 years ago and integrated in the current release
- Tested on ATLAS TDAQ system
Data Flow Evolution: design

- Design phase ongoing
  - First implementation to be ready for the end of the run
- Looking for common solutions, minimizing code duplication
  - A common framework for all the applications
- Different s/w technologies under evaluation
  - Profit from experience
  - But with an open attitude toward new ideas and views
- Prototype available for testing design and spot problems
  - Current applications adapted to the proposed design
  - Developed 2 years ago and integrated in the current release
  - Tested on ATLAS TDAQ system
Scalability validated up to \( \sim 1200 \) HLT nodes (\( \sim 13k \) HLTPUs)

- Traffic shaping strategy allows to prevent network congestions
  - In each DCM, limit the number of concurrent requests
  - A similar algorithm is being used in EB nodes of the current system
- A single HLTSV able to sustain more than 100 kHz
- Overhead of s/w RoIB to be evaluated
Scalability validated up to $\sim 1200$ HLT nodes ($\sim 13k$ HLTPUs)

Traffic shaping strategy allows to prevent network congestions
- In each DCM, limit the number of concurrent requests
- A similar algorithm is being used in EB nodes of the current system

A single HLTSV able to sustain more than 100 kHz

Overhead of s/w RoIB to be evaluated
Scalability validated up to $\sim 1200$ HLT nodes ($\sim 13k$ HLTPUs)

Traffic shaping strategy allows to prevent network congestions
- In each DCM, limit the number of concurrent requests
- A similar algorithm is being used in EB nodes of the current system

A single HLTSV able to sustain more than $100 \text{ kHz}$

Overhead of s/w RoIB to be evaluated
Conclusions

- Data taking 2011
  - Smooth TDAQ operation: \( \sim 94\% \) run efficiency
  - Extended HLT farm in course of operations
  - Stable and reliable data collection system
  - Excellent operational stability of control, configuration and monitoring
  - Improved automation: monitoring and recovery procedures

- Data taking 2012
  - Overall smooth and quick start up
  - High pileup effects under control

- Data Flow evolution
  - Merge L2, EB, EF within a single homogeneous system
  - Prototype studies did not spot problems
  - Design phase ongoing
  - To be ready at the beginning of 2013
Traffic shaping strategy allows to prevent network congestions

- In each DCM, limit the number of concurrent requests
- A similar algorithm is being used in EB nodes of the current system

**Graphs:**
- L2 acceptance = 5%
- Full event size = 1300 kB
- RoI DC size = 20 kB
- L2 Proc. time = 40 ms
- EF Proc. time = 2000 ms
- Automatic load balance inside each node
  - System promptly reacts to operation condition changes
  - System always capable of sharing CPU resources between the L2 and EF algorithms
Test in realistic operational conditions: fixed L1 rate

- As long the CPUs are not saturated the throughput rate is stable with increasing L2 processing time
- After saturation performance decreases as expected
Comparison between old and new architecture

- Same setup: 23 XPU racks to be shared between L2 and EF