A New Tool for Monitoring CMS Tier 3 LHC Data Analysis Centers

In Cooperation With:
The Texas A&M Tier 3 CMS Grid Site on the Brazos Cluster

Texas A&M University:
  David Toback
  Guy Almes
  Steve Johnson
  Vaikunth Thukral
  Daniel Cruz

Sam Houston State University:
  * Joel Walker
  Jacob Hill
  Michael Kowalczyk
Motivation

• LHC data analysis is distributed globally across a grid of computing centers
• Requires cooperation of complex machines under diverse management
• Cutting edge technology always has many possible points of failure
• Enormous data transfer and storage needs
• Vast quantities of jobs must be run
• Site availability and optimization depend on accurate and current monitoring feedback
Monitoring Goals

• Monitor data transfers, data holdings, job status, and site availability
• Optimize for a single CMS Tier 3 site
• Provide a convenient and broad view
• Unify grid and local cluster diagnostics
• Give current status and historical trends
• Realize near real-time reporting
• Email administrators about problems
• Improve the likelihood of rapid resolution
Implementation Goals

• Host monitor online with public accessibility
• Provide rich detail without clutter
• Favor graphic performance indicators
• Merge raw data into compact tables
• Avoid wait-time for content generation
• Avoid multiple clicks and form selections
• Harvest plots and data with scripts on timers
• Automate email and logging of errors
Email Alert System Goals

• Operate automatically in background
• Diagnose and assign a “threat level” to errors
• Recognize new problems and trends over time
• Alert administrators of threats above threshold
• Remember mailing history and avoid “spam”
• Log all system errors centrally
• Provide daily summary reports
Monitor Workflow Diagram

**CMS Dashboard**
- Job Submission & Runtime plots
- Job Outcome & Efficiency plots
- Failed job attribution plots
- Per user job status overview
- Hourly, Daily & Weekly views
- Service Availability (SAM) tests
- Best practice test job outcomes

**Local Cluster Environment**
- Disk usage distribution
- Node, CPU & RAM occupancy
- Queue usage & Job summary
- Detailed per user job status
- Load averages & Cluster stress
- Essential “heartbeat” functions
- Self locking stops redundancy

**Monitoring Software**
- Parallel data queries triggered by cron
- Recursively deep data harvesting
- Automated quality analysis of plots
- Archiving of plots with thumbnails
- Generation of deeply nested tables
- Generation of graphical data displays
- Preservation of system state data
- Email of relevant system alerts
- Consolidation of unified web display

**CMS PhEDEx**
- Transfer Rate & Quality plots
- Queued & Resident data plots
- Production data & Load tests
- Hourly, Daily & Weekly views
- Link Status & Transfers per site
- Subscribed & Resident data sums
View the working development version of the monitor online at:

collider.physics.tamu.edu/tier3/mon/

The next five slides provide a tour of the website with actual graph and table samples
Monitoring Category I:

**Data Transfers to the Local Cluster**

- Do we have solid links to other sites?
- Is requested data transferring successfully?
- Is it getting here fast?
- Are we passing load tests?

<table>
<thead>
<tr>
<th>Production</th>
<th>PMDGx Data Transfers</th>
<th>Load Test</th>
<th><strong>Select</strong></th>
<th>Hour</th>
<th>Day</th>
<th>Week</th>
<th>Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link Status</td>
<td>Limited Time</td>
<td>Link Status</td>
<td>Rate</td>
<td>Bytes</td>
<td>Files</td>
<td>Failed</td>
<td>Error</td>
</tr>
<tr>
<td>Valid</td>
<td>T1_CH_VEN_H Buffer</td>
<td>Valid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Valid</td>
<td>T1_DE_VTI Buffer</td>
<td>Valid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Valid</td>
<td>T1_E8_PIC Buffer</td>
<td>Valid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Valid</td>
<td>T1_FR_CONV32 Buffer</td>
<td>Valid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Valid</td>
<td>T1_FR_ICN Buffer</td>
<td>Valid</td>
<td>15.9 MB/s</td>
<td>13 TB</td>
<td>577</td>
<td>830</td>
<td>1034</td>
</tr>
<tr>
<td>Valid</td>
<td>T1_P1_1200G Buffer</td>
<td>Valid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Valid</td>
<td>T1_UK_PAL Buffer</td>
<td>Valid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Valid</td>
<td>T1_UK_FMN Buffer</td>
<td>Valid</td>
<td>8.2 MB/s</td>
<td>608.1 GB</td>
<td>450</td>
<td>0</td>
<td>311</td>
</tr>
<tr>
<td>Valid</td>
<td>T2 FR_Pisa</td>
<td>Valid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Valid</td>
<td>T2 US_Caltech</td>
<td>Valid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Valid</td>
<td>T2 US_Florida</td>
<td>Valid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Valid</td>
<td>T2 US_Nebraska</td>
<td>Valid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Valid</td>
<td>T2 US_SanJose</td>
<td>Valid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Valid</td>
<td>T2 US_UCED</td>
<td>Valid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Valid</td>
<td>T2 US_Vandenberg</td>
<td>Valid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Valid</td>
<td>T2 US_Wisconsin</td>
<td>Valid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Valid</td>
<td>T3 US_Colorado</td>
<td>Valid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Valid</td>
<td>T3 US_Rice</td>
<td>Valid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Valid</td>
<td>T3 US_TTU</td>
<td>Valid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Totals</td>
<td>26.0 MB/s</td>
<td>2.1 TB</td>
<td>1032</td>
<td>630</td>
<td>2320</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Monitoring Category II:
Data Holdings on the Local Cluster

• How much data have we asked for? Actually received?
• Are remote storage reports consistent with local reports?
• How much data have users written out?
• Are we approaching disk quota limits?

<table>
<thead>
<tr>
<th>Directory</th>
<th>Bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>PhEDEx Monte Carlo</td>
<td>19.1 TIB</td>
</tr>
<tr>
<td>PhEDEx CMS Data</td>
<td>530.0 GB</td>
</tr>
<tr>
<td>PhEDEx Load Tests</td>
<td>49.8 GB</td>
</tr>
<tr>
<td>User Output</td>
<td>3.8 TIB</td>
</tr>
<tr>
<td>Indara Suarez</td>
<td>2.5 TIB</td>
</tr>
<tr>
<td>Diamond</td>
<td>528.7 GB</td>
</tr>
<tr>
<td>Alonzo Furillo</td>
<td>475.1 GB</td>
</tr>
<tr>
<td>Roy J. Mendalvo</td>
<td>140.0 GB</td>
</tr>
<tr>
<td>Tai Sakuma</td>
<td>53.6 GB</td>
</tr>
<tr>
<td>Alfredo Gurrola</td>
<td>45.4 GB</td>
</tr>
<tr>
<td>Ji Eun Kim</td>
<td>19.2 GB</td>
</tr>
<tr>
<td>Valknuth Thukral</td>
<td>43.7 MB</td>
</tr>
<tr>
<td>Daniel J. Cruz</td>
<td>6.9 MB</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>556.0 MB</td>
</tr>
<tr>
<td>Total</td>
<td>23.5 TIB</td>
</tr>
</tbody>
</table>
Monitoring Category III:
Job Status of the Local Cluster

• How many jobs are running? Queued? Complete?
• What percentage of jobs are failing? For what reason?
• Are we making efficient use of available resources?
• Which users are consuming resources? Successfully?
• How long are users waiting to run?
Monitoring Category IV: Site Availability

- Are we passing tests for connectivity and functionality?
- What is the usage fraction of the cluster and job queues?
- What has our uptime been for the day? Week? Month?
- Are test jobs that follow “best practices” successful?
Monitoring Category V:
Alert Summary

- What is the individual status of each alert trigger?
- When was each alert trigger last tested?
- What are the detailed criteria used to trigger each alert?

<table>
<thead>
<tr>
<th>Alert System Test</th>
<th>Alert has NULL Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Transfer Quality</td>
<td>Test Status: UNKNOWN (No Result for 24 Hours)</td>
</tr>
<tr>
<td>Load Test Transfer Quality</td>
<td>Test Status: FAILED (2012-08-24 16:05 UTC)</td>
</tr>
<tr>
<td>Load Tests Missing</td>
<td>Test Status: UNKNOWN (No Result for 24 Hours)</td>
</tr>
<tr>
<td>PhEDEx Transfer</td>
<td>Test Status: PASSED (2012-08-24 16:07 UTC)</td>
</tr>
<tr>
<td>Disk Usage</td>
<td>Test Status: PASSED (2012-08-24 11:42 UTC)</td>
</tr>
<tr>
<td>Disk Quota</td>
<td>Test Status: PASSED (2012-08-24 16:07 UTC)</td>
</tr>
<tr>
<td>Disk Permissions</td>
<td>Test Status: PASSED (2012-08-24 11:42 UTC)</td>
</tr>
<tr>
<td>PhEDEx Mismatch</td>
<td>Test Status: PASSED (2012-08-24 16:07 UTC)</td>
</tr>
<tr>
<td>Torque Stalled</td>
<td>Test Status: PASSED (2012-08-24 16:05 UTC)</td>
</tr>
<tr>
<td>Condor Stalled</td>
<td>Test Status: PASSED (2012-08-24 16:05 UTC)</td>
</tr>
<tr>
<td>SAM Failed</td>
<td>Test Status: UNKNOWN (No Result for 24 Hours)</td>
</tr>
<tr>
<td>SAM Missing</td>
<td>Alert is Disabled</td>
</tr>
<tr>
<td>CATS Failed</td>
<td>Test Status: PASSED (2012-08-24 16:05 UTC)</td>
</tr>
<tr>
<td>CATS Missing</td>
<td>Test Status: PASSED (2012-08-24 16:05 UTC)</td>
</tr>
<tr>
<td>Cluster Heartbeat</td>
<td>Test Status: PASSED (2012-08-24 16:05 UTC)</td>
</tr>
<tr>
<td>Cluster Web Server</td>
<td>Test Status: PASSED (2012-08-24 16:05 UTC)</td>
</tr>
<tr>
<td>Cluster Server Error</td>
<td>Test Status: PASSED (2012-08-24 16:05 UTC)</td>
</tr>
<tr>
<td>Cluster Server Hibernating</td>
<td>Test Status: PASSED (2012-08-24 11:42 UTC)</td>
</tr>
<tr>
<td>Cluster Server Timeout</td>
<td>Alert is Disabled</td>
</tr>
</tbody>
</table>
Distribution Goals

• Make the monitor software freely available to all other interested CMS Tier 3 Sites
• Globally streamline away complexities related to organic software development
• Allow for flexible configuration of monitoring modules, update cycles, site details and alerts
• Package all non-minimal dependencies
• Single step “Makefile” initial installation
• Build locally without root permissions
Distribution Prerequisites

- A clean account on the host cluster
- Linux shell: /bin/sh & /bin/bash
- Apache web server with .ssi enabled
- Perl and cgi-bin web directory
- Standard build tools, e.g. make, cpan, gcc
- Access to web via lwp-download or wget, etc.
- Group access to common disk partition
- Job scheduling via crontab
Distribution Timetable

- The primary software authoring and beta testing phases at Texas A&M are now complete.
- The LHC Physics Center (LPC) at Fermilab is sponsoring a month-long residency for Joel Walker to work on distribution of the monitor.
- The transition from a single example installation to a generalized package is well underway.
- The goal is a 2-hour top-to-bottom install.
- The target date for a first release satisfying this criterion is Monday, September 3, 2012.
Distribution Details

• The software repository will be held at github: https://github.com/joelwwalker/cms-site-monitor (this first upload is not suitable for cross-site use)
• Volunteer sites are needed to begin test installs as the distribution version is released
• The developer is eager to consult closely during the early installation trials
• This will be a recursive procedure of debugging and improving the distribution package
• Feedback on the usefulness and portability of the new monitoring software is appreciated
Ongoing Work

• Enhancement of content and real-time usability
• Vetting for robust operation and completeness
• Expanding implementation of the alert layer
• Development of suitable documentation
• Distribution to other University Tier 3 sites
• Improvement of portability and configurability
• Seeking out a continuing funding source
Conclusions

- New monitoring tools are uniquely convenient and site specific, with automated email alerts
- Remote and Local site diagnostic metrics are seamlessly combined into a unified presentation
- Early deployment at Texas A&M has already improved rapid error diagnosis and resolution
- We are engaged in a new phase of work to bring the monitor to other University Tier 3 sites
We acknowledge the Norman Hackerman Advanced Research Program, The Department of Energy ARRA Program, and the LPC at Fermilab for prior support in funding.