# 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**

# First There Was the 30 Minute Meal

# After that … a bit of an Arms Race

# And Now, Presenting …

# Why Should You Care About this Project?

- It is (mostly) Ready
- It is (mostly) Working
- It is (completely) Free
- It is very Flexible
- It is very Easy
- It makes your job Easier
- ~~You can trust me~~
- You don't need to trust me

  (installs 100% locally as an unprivileged user)

# A Small Cheat: The “Mise En Place”

# In other Words, 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
- ~ 100K file inodes and ~ 2GB of disk

# Ok, Let's Start Cooking

- wget http://www.joelwalker.net/code/brazos/brazos.tgz
- tar –xzf brazos.tgz
- cd brazos
- ./configure.pl  (answer two questions)
- make (this takes a while)  … What is it doing?
    - setting up your environment ( .bashrc, etc. )
    - building local /bin, /lib, /include, perl5
    - compiling and linking libraries ( zlib, libpng, gd, etc. )
    - bootstrapping "cpanm" to load Perl modules & dependencies
    - creating the directory structure & moving files into place
- exec bash
- edit local.txt, modules.txt, alert.txt, users.txt in ~/mon/CONFIG
- Test modules and set crontab to run:

```
* * * * * . ${HOME}/.bashrc && ${BRAZOS_BASE_PATH}${BRAZOS_CGI_PATH}/_Perl/brazos.pl > /dev/null 2>&1
```

# While that Simmers … Monitoring Goals

- Monitor data transfers, data holdings, job status, and site availability
- Optimize for a single CMS Tier 3 (or 2?) 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

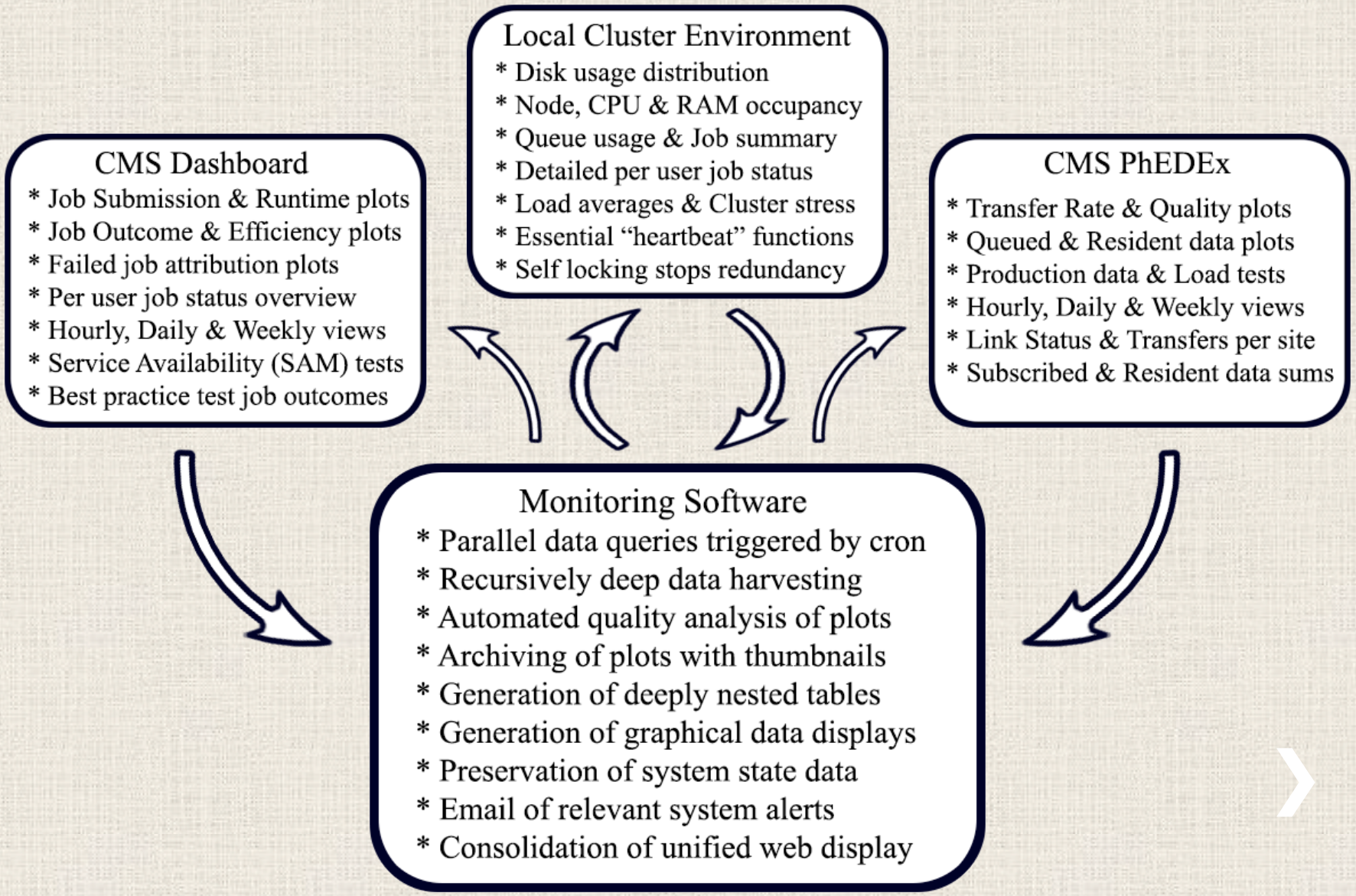
# Implementation Goals

- 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



**View the working development version of the monitor online at:**

**brazos.tamu.edu/~ext-jww004/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?**

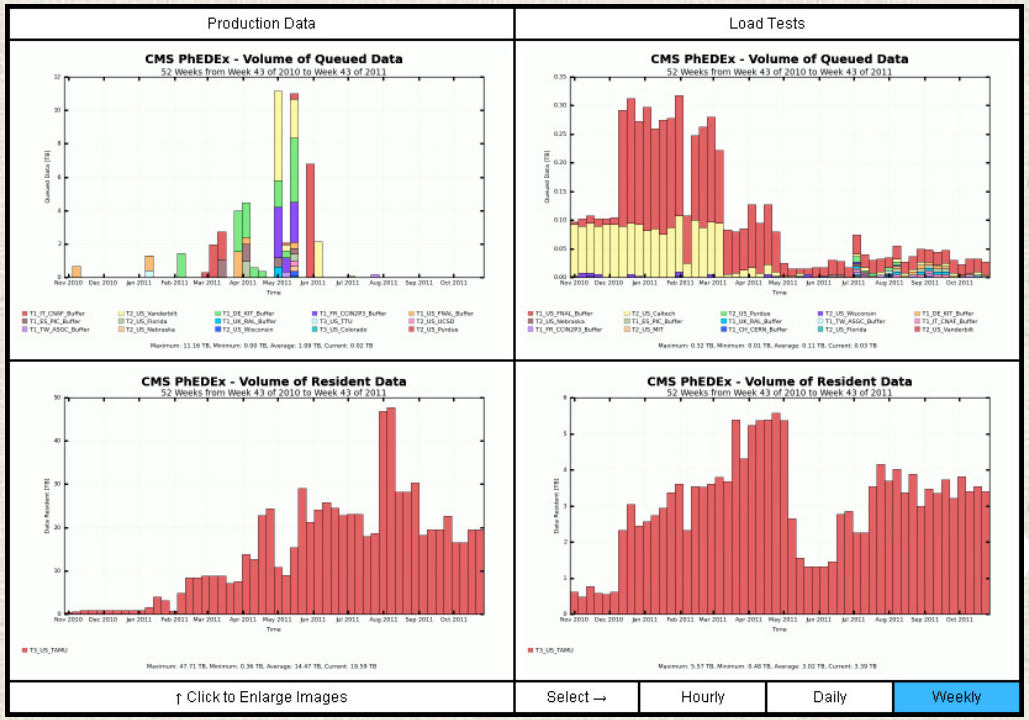
| ↓ Production | PhEDEx Data Transfers | Load Test ↓ | ← Select → | Hour | Day | Week | Month |
|---|---|---|---|---|---|---|---|
| Link Status | Linked Node | Link Status | Rate | Bytes | Files | Expired | Errors |
| Valid | T1_CH_CERN_Buffer | Valid | - | - | - | - | - |
| Valid | T1_DE_KIT_Buffer | Valid | - | - | - | - | - |
| Valid | T1_ES_PIC_Buffer | Valid | - | - | - | - | - |
| Valid | T1_FR_CCIN2P3_Buffer | Valid | - | - | - | - | - |
| Valid | T1_IT_CNAF_Buffer | Valid | 15.9 MiB/s | 1.3 TiB | 577 | 638 | 1834 |
| Valid | T1_TW_ASGC_Buffer | Valid | - | - | - | - | - |
| Valid | T1_UK_RAL_Buffer | Valid | - | - | - | - | - |
| Valid | T1_US_FNAL_Buffer | Valid | 8.2 MiB/s | 688.1 GiB | 459 | 0 | 311 |
| Valid | T2_IT_Pisa | Valid | - | - | - | - | - |
| Valid | T2_US_Caltech | Valid | - | - | - | - | - |
| Valid | T2_US_Florida | Valid | - | - | - | - | - |
| Valid | T2_US_MIT | Valid | - | - | - | - | - |
| Valid | T2_US_Nebraska | Valid | - | - | - | - | - |
| Valid | T2_US_Purdue | Valid | 1.9 MiB/s | 159.5 GiB | 16 | 0 | 175 |
| Valid | T2_US_UCSD | Valid | - | - | - | - | - |
| Valid | T2_US_Vanderbilt | Valid | - | - | - | - | - |
| Valid | T2_US_Wisconsin | Valid | - | - | - | - | - |
| Valid | T3_US_Colorado | Valid | - | - | - | - | - |
| Valid | T3_US_Rice | Valid | - | - | - | - | - |
| Valid | T3_US_TTU | Excluded | - | - | - | - | - |
| Totals | | | 26.0 MiB/s | 2.1 TiB | 1052 | 638 | 2320 |

## 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?

| HEPX Disk Store Usage | |
|---|---|
| Directory | Bytes |
| → PhEDEx Monte Carlo | 19.1 TiB |
| → PhEDEx CMS Data | 539.8 GiB |
| → PhEDEx Load Tests | 49.8 GiB |
| ↓ User Output | 3.9 TiB |
| Indara Suarez | 2.6 TiB |
| Diamond | 529.7 GiB |
| Alexx Perloff | 475.1 GiB |
| Roy J. Montalvo | 146.0 GiB |
| Tai Sakuma | 58.6 GiB |
| Alfredo Gurrola | 45.4 GiB |
| Ji Eun Kim | 10.2 GiB |
| Vaikunth Thukral | 43.7 MiB |
| Daniel J. Cruz | 6.9 MiB |
| → Miscellaneous | 556.0 KiB |
| Total | 23.5 TiB |
| ↑ Click to Expand or Collapse 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?

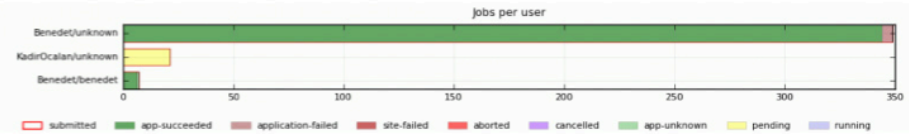
Current TORQUE Batch Queue Job Status ( 2011-10-20 13:20 UTC )

| User Name | Queue | Run Status | Processors | Memory | CPU Hours | Run Hours | Queue Hours |
|---|---|---|---|---|---|---|---|
| Totals | - | - / - | - | - | - | - | - |
| No Jobs Are Currently Queued | | | Select Sorting By → | | User | Queue | Status |

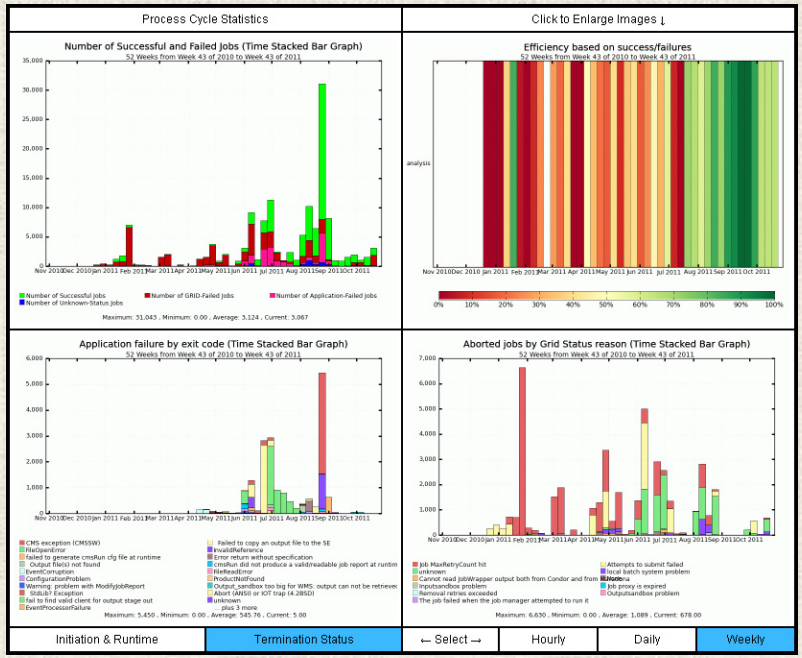
Current Condor Batch Queue Job Status ( 2011-10-20 13:20 UTC )

| User Name | Universe | Run Status | Processors | Memory | CPU Hours | Run Hours | Queue Hours |
|---|---|---|---|---|---|---|---|
| Totals | - | - / - | - | - | - | - | - |
| No Jobs Are Currently Queued | | | Select Sorting By → | | User | Universe | Status |

Recent Brazos Cluster Job Activity by User




| User Name | Pending | Running | Terminated | App Failed | Grid Failed | CPU Usage | Run Hours |
|---|---|---|---|---|---|---|---|
| Benedet | 0 | 0 | 356 | 1.4 % | 0.0 % | 94.0 % | 921.2 |
| Kadir Ocalan | 21 | 0 | 0 | 0.0 % | 0.0 % | 100.0 % | 0.0 |
| Totals | 21 | 0 | 356 | 1.4 % | 0.0 % | 94.0 % | 921.2 |
| ↑ Click User Row for Job Details | | | | Select → | Hour | Day | Week |



## 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?**

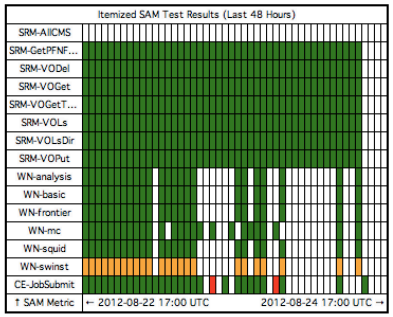
| Service Availability Percentage | | |
|---|---|---|
| Day | Week | Month |
| 94 % | 93 % | 98 % |

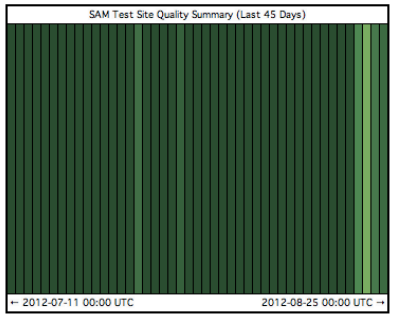
| Brazos Cluster Heartbeat Tests ( 2012-08-24 16:05 UTC ) | | | | |
|---|---|---|---|---|
| SSH Link to Hurr | FData Filesystem Mount | FData Partition Size | "DU" Query Status | "DU" Query Timer |
| Pass | Pass | 98.2 TiB | Pass | 0.1 Seconds |

| Brazos Cluster Load Statistics ( 2012-08-24 16:05 UTC ) | | | | |
|---|---|---|---|---|
| Brazos Head Load | Hurr Head Load | GridFTP Processes | Filesystem CPU Use | Filesystem Memory Use |
| 0.64 | 0.27 | 0 | 0.0 % | 0.0 % |
| Occupied Nodes | Occupied Processors | Load Average per CPU | Physical Memory Use | Virtual Memory Use |
| 100.0 % of 312 | 82.3 % of 2,840 | 72.2 % of 2,840 | 18.5 % of 8.91 TiB | 0.1 % of 20.2 TiB |
| Queue | Job Capacity | Jobs Running | Jobs Queued | Other Jobs |
| GRID | 256 | 346 | 0 | 0 |
| HEPX | 320 | 302 | 18 | 0 |
| HEPX ROUTING | 0 | 0 | 0 | 0 |
| BGSC | 320 | 320 | 0 | 0 |
| BGSC ROUTING | 0 | 0 | 25 | 0 |

| ↓ Grid Client | CRAB Analysis Test Suite (CATS) Completed Job Status | | | |
|---|---|---|---|---|
| Output Host → | Local: Brazos Cluster | | Remote: Fermi National Laboratory | |
| Output Size → | Small | Large | Small | Large |
| gLite | Pass:10 Fail:0 Other:0<br>2012-08-24 10:02 UTC | Pass:10 Fail:0 Other:0<br>2012-08-24 16:02 UTC | Pass:10 Fail:0 Other:0<br>2012-08-23 23:12 UTC | Pass:10 Fail:0 Other:0<br>2012-08-23 04:02 UTC |
| CondorG | Pass:10 Fail:0 Other:0<br>2012-08-24 09:03 UTC | Pass:10 Fail:0 Other:0<br>2012-08-24 13:02 UTC | Pass:10 Fail:0 Other:0<br>2012-08-24 16:59 UTC | Pass:10 Fail:0 Other:0<br>2012-08-24 01:02 UTC |
| PBS | Pass:10 Fail:0 Other:0<br>2012-08-24 09:07 UTC | | | |
| ↑ Test Results Link to Job Details ↑ | | | | |

Service Availability Monitoring (SAM) Tests



## 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?**

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

# 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

# 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**

**Special Thanks to: Dave Toback, Guy Almes,**
**Rob Snihur, Oli Gutsche, and David Sanders**