LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-407208

Completion Report for Multi-Site Incentive MRT 2779 Implement ASC Tripod Initiative by 30SEP08

D. East, J. Cerutti, J. Noe, K. Cupps, J. Loncaric, J. Sturtevant

September 24, 2008

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

TABLE OF CONTENTS

Executive Summary 1
Establishment of Executive and Steering Committees 2
Initial TOSS Delivery and Deployment 3
TOSS Deployment 4
Positioning of TOSS for Future TLCC Procurements 6
Appendix A — TripodOS Proposal 7
Appendix B — TripodOS Governance 12
Appendix C — Clarification of Intent and Completion Criteria .. 19
Appendix D — Support Websites Screenshots 21
Appendix E — MRT 2779 Criteria Sign-off Sheets 24
Appendix F — FY08 Final Report: Tripod Working Groups 30

Executive Summary

This report provides documentation and evidence for the completion of the deployment of the Tripod common operating system (TripodOS, also known as and generally referred to below as TOSS). Background documents for TOSS are provided in Appendices A and B, including the initial TOSS proposal accepted by ASC HQ and Executives in July 2007 and a Governance Model defined by a Tri-Lab working group in September 2007. Appendix C contains a document that clarifies the intent and requirements for the completion criteria associated with MRT 2779. The deployment of TOSS is a Multi-Site Incentive from the ASC FY08-09 Implementation Plan due at the end of Quarter 4 in FY08.

> **Milestone**: Multi-Site Incentive MRT 2779
> **Title**: Implement ASC Tripod Initiative by 30SEP08
> **Category**: Advanced Simulation and Computing
> **ASC Program Element**: CSSE, FOUS

The milestone definition is:

> *By 31DEC07, implement the Tripod configuration management structure, including establishment of Tripod common operating system (TripodOS) Executive and Steering Committees. By 31MAR08, demonstrate the TripodOS stack on the initial hardware of the ASC-funded Tri-lab Linux Capacity Clusters (TLCCs). By 30SEP08, deploy and operate all ASC TLCC platforms that will be delivered before mid-August, under the generally available (GA) release of TripodOS. Furthermore, position TripodOS for future deployment on all ASC Linux capacity platforms delivered in FY09 and beyond.*

This milestone consists of four distinct elements each of which will be discussed in the indicated sections of this completion report:

1. *Establishment of Executive and Steering Committees* (Release Planning and Support Committees) by 31 December 2007.
2. *Initial TOSS Delivery and Deployment* demonstrating a working TOSS deployment on initial delivery TLCC hardware by 31 March 2008.
3. *TOSS Deployment* by 30 September 2008 on all TLCC hardware delivered to the Laboratories before mid-August 2008, commonly understood to be 15 August 2008.
4. *Positioning of TOSS for Future TLCC Procurements* in FY09 and beyond.

The first three elements (those with due dates) were completed before the specific dates in the milestone definition. The fourth element, which speaks to a programmatic positioning, was also accomplished as documented below. Also provided in this report, since it is relevant to the TOSS deployment, is a summary of activities and deliveries associated with TLCC (see *TOSS/TLCC Build and Delivery Summary*). Lastly, the Tripod organization includes other smaller projects that are not directly TOSS-related. These projects are conceived and mentored by working groups established as part of Tripod. Since this MRT is tied organizationally to these other

Tripod projects, it was determined that a year-end summary of the other Tripod projects would be appropriate to include in this report. This additional information can be found in Appendix F.

Establishment of Executive and Steering Committees

Upon development and agreement of the Governance Model (Appendix A), the required three committees—Executive, Releasing Planning, and Support—were formed and the Release Planning Manager and Support Manager named as of 25 September 2007. Table 1 lists the committee memberships.

Committee	LANL	LLNL	SNL	HQ
Executive	John Cerutti	Doug East	John Noe	Thuc Hoang / Ted Blacker
Release Planning	Josip Loncaric Dave Montoya	Trent D'Hooge Pam Hamilton (1)	Josh England Curt Janssen	N/A
Support	Jim Lujan	Mark Grondona (2)	Steve Monk	N/A

Notes
(1) Release Planning Manager
(2) Support Manager

Table 1. TOSS committee membership

TOSS Executive Committee members participate in a monthly Tripod Steering Committee telecon that started on 11 September 2007 that has served in lieu of a quarterly TOSS Executive Committee meeting. In addition, ad hoc TOSS-specific telecons were held on 21 January 2008 and 6 February 2008, and a TOSS Executive session held at the ASC PI meeting in Monterey, CA on 25 February 2008.

Release Planning and Support Committee telecons have been combined to date. This is due to the exigent circumstances of the initiation of TOSS and the work towards the first release in March 2008 as well as an update in September 2008. The first of the combined Release Planning and Support telecons was on 20 September 2007.

Also of significance for support of TOSS is the establishment of the TOSS Software Repository, TOSS Bug Tracking System, and TOSS Documentation Wiki. Table 2 lists the respective hosting sites and availability dates of these components for Tri-Lab access. Appendix D contains screen shots of the top level web pages for these components.

TOSS Support Component	Hosting Site	Available for Tri-Lab Access
Software Repository	LLNL	17 October 2007
Bug Tracking System	LLNL	30 October 2007
Documentation Wiki	SNL	4 October 2007

Table 2. TOSS support components

The Executive, Release Planning, and Support Committees were fully formed and functioning well before 31 December 2007 as were the critical support components listed in Table 2. The completion criterion for this milestone element is therefore satisfied.

Initial TOSS Delivery and Deployment

TOSS 1.0[1] was made *generally available* on 3 March 2008 and a minor release (TOSS 1.0-1) issued on 25 March 2008 (see Figure 1).

```
Subject: [release] chaos-release-4.0-1
Date: Tuesday 25 March 2008 13:22
From: Trent D'Hooge <tdhooge@llnl.gov>
To: chaos-announce@lists.llnl.gov, tos-devel@lists.llnl.gov, LC System
Administration Group <lc-sag@lists.llnl.gov>

New chaos 4.0 minor release with security patches, bug fixes and a few
additional packages as requested by users.

2008-3-25 tdhooge
* tagging chaos-release-4.0-1 / TOSS-1.0-1

 .
 . [additional details of update removed for brevity]
 .
```

Figure 1. TOSS 1.0-1 release announcement

TOSS 1.0-1 was installed and in use on Hype at Synnex[2], the TLCC initial delivery test bed hardware, when it was announced (see Figure 3) on 26 March 2008 that the Synthetic Work Load (SWL) was completed and a decision to go forward with TLCC deliveries would be submitted to the ASC Executives on 27 March 2008 for their concurrence, which was ultimately given. Figure 2 shows the summary of the SWL run on Hype that completed on 26 March 2008. The failure rate is 1.29 percent, which less than the 2 percent failure rate considered acceptable.

Summary

Start Time: Thu Mar 20 08:55:05 PDT 2008
Job Totals

Passed:	1144
Failed:	7
Unknown:	8
Total:	1159
Failure Rate::	1.29%

Run Times

Wall Clock Run Time:	709.24 hrs.
Node Run Time:	5408.50 node-hrs.
Node Utilization:	29.71%.

Excessive Run-time Variation Job Count: 40
End Time: Wed Mar 26 07:09:26 PDT 2008

Figure 2. SWL summary on Hype at TOSS 1.0-1

[1] TOSS 1.0 is equivalent to CHAOS 4.0 as found in some correspondence and evidence.
[2] The system integration site in Fremont, CA used by Appro, the TLCC contract awardee.

Recall that the criterion for completion of this element of MRT 2779 is "*By 31MAR08, demonstrate the TripodOS stack on the initial hardware of the ASC-funded Tri-lab Linux Capacity Clusters (TLCCs).*" The completion of a full-scale SWL on Hype (the initial TLCC hardware), as announced on 26 March 2008, satisfies this criterion.

```
Date: Wed, 26 Mar 2008 08:35:45 -0700
To: Mike McCoy <mgmccoy@llnl.gov>, Doug East <dre@llnl.gov>,
        Terri Quinn <quinn1@llnl.gov>, kimcupps@llnl.gov
From: Mark Seager <seager@llnl.gov>
Subject: TLCC07 PLAN A GO DECISION
Cc: "Gary M. Ward" <ward31@llnl.gov>, "Trent D'Hooge" <tdhooge@llnl.gov>,
        "Sheila A. Faulkner" <saf@llnl.gov>,
        Matt Leininger <leininger4@llnl.gov>, Brent Gorda <gorda1@llnl.gov>,
        Pam Hamilton <hamilton5@llnl.gov>

All,
      As you are aware, the Hype cluster at Synnex was upgraded a couple of
weeks ago with B3 Barcelonas.  Sheila and Trent have been running a number of
tests including the infamous STRIDE benchmark that uncovered the Errata 298
problem with B2c Barcelonas.  As far as we can determine we have not had an
reoccurrence of the Errata 298 problem.  Also, the SWL testing went extremely
well.  Hype (after some initial fallout from the processor upgrade) appears
to be very stable.  So as of this morning, the Tri-Lab community will
recommend to the ASC Exec's tomorrow on ASC Exec call to GO WITH PLAN A!
This is great news and quite a relief.  AMD, Appro and Synnex are all geared
up to begin building 8 SU immediately and ship Hype to LLNL starting after
the ASC Exec call.  AMD has moved up the delivery of 6,000 Barcelona parts to
make this feasible.
      I want to personally thank Sheila and Trent for all the hard work
getting over the Hype rebuild hump and completing the testing early.
      Regards,
      ++Mark

----------
Mark K. Seager  SCCD ADH          |
V-925-423-3141  LLNL              | The dust of exploded beliefs may make a
P-800-265-8691  POBOX 808, L-554  | fine sunset. -Geoffrey Madan,
F-925-422-3887  7000 East Ave.    | writer  (1895-1947)
seager@llnl.gov Livermore, CA 94551|
```

Figure 3. E-mail announcing SWL completion on Hype

TOSS Deployment

Completion of the TOSS Deployment criteria for MRT 2779 is inextricably linked to the build, vendor integration, and delivery of the TLCC hardware (scalable units or SUs). The contract for the TLCC clusters was awarded to Appro on 17 September 2007. Two problems vexed the build and delivery of the TLCC hardware. The first problem is referred to as the *Errata 298 Problem*; the second is the *VDDIO Regulation Problem*. Both problems were due to different aspects of the transition from dual to quad core processors, summarized below.

Errata 298 Problem
Early AMD Barcelona quad-core processor testing demonstrated a known problem (Errata 298) to be a major issue with ASC codes. This problem involved a translation lookaside buffer (TLB) L3 cache coherency race condition that would cause frequent node crashes using select ASC code kernels. The severity of this problem was confirmed in late-November 2007 during testing of Hype, the one SU test bed cluster to be sited at LLNL. One proposed fix to this problem, a Linux kernel patch proposed by AMD, was rejected as being too complex and very difficult to support, given that AMD refused to provide ongoing support for the patch. As a result, the preferred plan of action (plan A) was simply to wait for the next (B3) revision of the AMD Barcelona processor that purportedly fixed this problem. This required AMD to produce the new processor revision and build up shippable quantities by the March 2008 time frame.

AMD was able to deliver the revised B3 processors "on time," which were retrofitted into Hype at Synnex and shown to fix this problem. The decision to proceed with B3 revision processors was made by the ASC Executives on 27 March 2008.

VDDIO Regulation Problem
As builds progressed on larger clusters in the May time frame, unexplained node hangs became more frequent. In early-June, a hang reproducer using the MATMULT code was found by Sandia and investigation into the root cause of the hangs ensued. These hangs were observed in approximately 40 percent of the compute nodes. This problem required almost two months of investigation and cooperation of the Tri-Lab community and all involved vendors and component suppliers: Appro, Supermicro, and AMD before a fix was finally found.

In late July, AMD convinced all involved that the issue was excessive voltage variation in the VDDIO circuitry under a quad core workload. These power variations affect motherboards depending on the margins of the electronic circuitry; i.e., a number of motherboards function without error because the operating margins of the components happen to fall in such a way that the variations do not cause error. Supermicro developed a solution for the problem, which consisted of soldering four additional capacitors to the motherboard to extend each voltage regulator bandwidth from 30 kHz to 55 kHz. The decision to retrofit existing systems (at all four Labs and those being built at Synnex) was made 4 August 2008 and a schedule for field repairs was developed and executed within two weeks with good results; (see Table 3).

TOSS/TLCC Build and Delivery Summary
As a result of the two problems discussed above, the hardware delivery schedules slipped significantly. In particular, the VDDIO regulation problem resulted in delivery delays for five clusters to be pushed out beyond the mid-August criterion of this MRT 2779 element; see Table 3. Based on this, the third MRT 2779 criterion "*By 30SEP08, deploy and operate all ASC TLCC platforms that will be delivered before mid-August, under the generally available (GA) release of TripodOS*" applies to four clusters: Hype and Juno (LLNL), Lobo (LANL), and Unity (SNL-NM).

Table 3 provides the date that the criteria (see Appendix C) were passed for MRT 2779 for each applicable cluster. Appendix E contains sign-off sheets as evidence of completion of the MRT 2779 criteria for each of the applicable clusters.

Cluster	Site	Scalable Units	Delivery Date	VDDIO Rework Date Complete	Passed MRT 2779 Criteria
Hype	LLNL	1	31 Mar 2008	9 Aug 2008	14 Apr 2008
Lobo	LANL	2	21 May 2008	13 Aug 2008	28 May 2008
Unity	SNL-NM	2	19 May 2008	15 Aug 2008	2 Jun 2008
Juno	LLNL	8	[3]5 May 2008	9 Aug 2008	29 Aug 2008
Whitney	SNL-CA	2	20 Aug 2008	6 Aug 2008	N/A
Eos	LLNL	2	20 Aug 2008	6 Aug 2008	N/A
Hurricane	LANL	2	25 Aug 2008	6 Aug 2008	29 Aug 2008
Glory	SNL-NM	2	mid-Sep 2008	Factory Parts	N/A
Hera	LLNL	6[4]	mid-Sep 2008	Factory Parts	N/A

Table 3. TLCC deliveries and MRT 2779 pass dates[5]

Positioning of TOSS for Future TLCC Procurements

TOSS is successfully operating on four clusters and 13 SUs to date, with an additional five clusters and 12 SUs coming on-line shortly. In addition, TOSS is installed on seven of LLNL's non-TLCC Linux clusters. The various TOSS committees are functioning and the software repository, bug tracking, and other support mechanisms are in place and undergoing iterative improvement. A new revision of TOSS (1.0.6) was released on 25 August 2008 that contained numerous RedHat security patches, SLURM fixes, and an out of memory work around. An in-depth TOSS technical discussion including both Release Planning and Support Committee members is scheduled for October 2008. All this is further evidence that the TOSS release process is fully functional and TOSS is well positioned to be the operating system software stack for future TLCC procurements. TOSS is testimony to what can be accomplished via Tri-Lab cooperation.

[3] First four SUs, four more SUs delivered 3 June 2008.
[4] Four SUs are ASC funded, two SUs are LLNL Institutional funded.
[5] Gray cells are systems not subject to MRT 2779 due to their delivery being after 15 Aug 2008.

Appendix A — TripodOS Proposal

Proposal for a
Tripod Operating System (TripodOS) Software Stack and
Configuration Management Process

Submitted by LLNL and SNL

Introduction

LLNL and SNL commit to deliver a fully functional cluster software stack and operational model that will support Tri-lab production workloads at scale on TLCC07 systems. To accomplish this while minimizing risk, LLNL's existing CHAOS[6] distribution will be used to provide a successful, known, and highly scalable solution. CHAOS will be evolved into TripodOS to meet LANL and SNL operational requirements. Also proposed is a model by which TripodOS will be further evolved to meet the long-term needs of the ASC Tri-lab community. This proposal provides the ASC Program with the best opportunity to achieve Tripod objectives with a solution that has a long-term record of success and that will be well supported on TLCC07 hardware at the earliest possible time.

Background

LLNL has deployed a common Linux cluster software stack, known as CHAOS, on nearly ten thousand Linux cluster nodes, spanning four generations of COTS hardware platforms over the past six years. CHAOS currently runs in production on 48 clusters ranging in size from 6 to 1,152 nodes. CHAOS is more than just a *product*; it is also a repeatable *process* by which the software is defined, assembled, tested, delivered, supported, and measured. CHAOS utilizes local expertise, a well-defined schedule, and long-term strategic planning.

CHAOS is developed and maintained at LLNL by a staff of system programmers that are focused specifically on production solutions and who embrace the open source model. They collaborate with vendor partners and the open source community, and have established reputations as respected contributors of HPC-focused enhancements. CHAOS starts with open source products and local modifications are kept to a minimum. Local enhancements or fixes are pushed upstream whenever possible. All locally developed packages are released as open source with the intent of encouraging external collaboration. This same development and support philosophy will be applied to TripodOS. Appendix A provides additional detail about CHAOS and its applicability as the foundation for TripodOS. Appendix B contains a list of documents that provide evidence of the guiding philosophies of CHAOS.

Stack Definition

TripodOS will be based on the Red Hat Enterprise Linux (RHEL) distribution integrated with other components needed for HPC clusters. By starting with RHEL, Red Hat's complete development, testing, and release cycle management processes are leveraged, as well as the full range of support services that are provided as part of their enterprise product offering. The existing DOE site-wide Red Hat contract allows RHEL to be purchased at a very low price compared to UNIX-based offerings from Sun, IBM, or HP. By providing TripodOS to the Tri-lab, the impact of LLNL's highly successful Linux development effort is extended as well as the expertise of LLNL's full-time on-site Red Hat analyst who will be a contributing member of the TripodOS development team and an advocate within Red Hat on the Tri-lab's behalf.

Fundamental components of the TripodOS stack are:

- A complete RHEL distro[7] augmented as required to support targeted HPC (e.g., TLCC) hardware and cluster computing in general.
- A RHEL kernel that is optimized and hardened to support large scale cluster computing, including EDAC[8] support for all TLCC platforms.

[6] Clustered High Availability Operating System
[7] A distro is a specific Linux distribution created by an organization or group of individuals.
[8] Error Detection and Correction

- The OFED[9] Infiniband software stack including MVAPICH and OpenMPI libraries, and Subnet Manager (SM) scalable to full size TLCC systems.
- The SLURM resource manager with support for both MVAPICH and OpenMPI over Infiniband, full NUMA awareness, scalability to 18,432 processors, and a compatibility library to support TORQUE job submission command syntax.
- Fully integrated Lustre and Panasas parallel file system clients and Lustre server software.
- Scalable cluster administration tools to facilitate installation, configuration (including BIOS setup/upgrade), and remote lights-out management.
- An extensible cluster monitoring solution with support for both in-band[10] and out-of-band (e.g., IPMI) methods.
- A PAM[11] authentication framework for OTP and Kerberos and an access control interface to SLURM.
- A test framework for hardware and operating system validation and regression testing, extensible to include Tri-lab tests.
- GNU C, C++ and Fortran90 compilers integrated with MVAPICH and OpenMPI.

This software stack will form the foundation upon which other collaborative solutions can be layered, including the Tripod Application Development Tool set and the Tri-lab Workload Manager (Moab).

Packaging, Delivery and Support Model

The TripodOS support model resembles that of many commercial software providers and also, where applicable, incorporates elements of the open source support model. Key elements of this model are described below.

Packaging. TripodOS will be delivered on DVD and will also be available on a Tri-lab YUM repository containing RPMs and buildable SRPMs for all components. All software releases, including minor bug fix releases, will be provided in RPM format. Baseline configuration files will be provided as a starting point for site-specific customization and to document what was tested at LLNL.

Release cycle. TripodOS releases correspond to a particular set of package versions that have been installed on reference TLCC hardware, frozen, and tested with the TripodOS test suite. TripodOS will track Red Hat's 18-month major release and quarterly minor release cycle. Bug fixes (including security fixes) will be made available as RPM package updates in the repository as they become available. These updated packages will be periodically incorporated into fully tested intermediate releases. TripodOS minor releases will be supported for six months; i.e., support will always be available for the current minor release and the previous one. This allows sites some flexibility in local scheduling of updates while minimizing the costs associated with supporting older releases.

QA validation. All TripodOS releases are subject to extensive validation, stress, and regression testing executed on dedicated QA clusters at LLNL and SNL (including SNL's 128-node Talon test bed cluster). As TripodOS is deployed across the Tri-lab, it will be necessary to perform local validation after the software is integrated into the local environments. Further, the QA test suite will grow over time as Tri-lab issues are identified and resolved. The complete TripodOS test environment (test framework and full Tri-lab test suite) will be included in the distribution to facilitate localized testing. We anticipate using the collaborative LANL/SNL Gazebo/Cbench test harness when it is available. Until then, the existing LLNL test framework will be used.

Support model. TripodOS will utilize a tiered support model as follows.

- Tier 1 support will be provided by each lab's local support staff. The availability of Tier 1 support (number of staff and hours of support) is determined by local site requirements. Tier 1 staff will respond to trouble reports from local users and attempt to resolve problems. If a problem cannot be resolved locally, the Tier 1 support representative will perform additional diagnostic and information gathering procedures then open a ticket in the TripodOS bug reporting system (see *Issue tracking* below), including the specification of a severity level using established Tri-lab severity definitions.

---

[9] OpenFabrics Enterprise Distribution

[10] In-band monitoring uses coordinated, parallel SNMP GETBULK from a lightweight SNMP agent to minimize the impact on MPI synchronization.

[11] Pluggable Authentication Module

- Tier 2 support will initially be provided by LLNL but with the expectation that over time the ownership and support responsibility for some TripodOS packages will transition to other labs. Tier 2 support personnel will monitor the Tri-lab bug tracking system M-F 0730 through 1700 Pacific Time and will provide an initial response to the customer in the timeframe defined for the specified severity level. Proposed severity levels and response times are provided in Appendix C of this document. Subsequent to providing an initial response to the customer, Tier 2 support staff will continue to pursue problem resolution in a manner consistent with the problem severity and as determined by priorities established in weekly TripodOS issue tracking telecons.

- Tier 3 support is provided by vendor partners (e.g. Red Hat, CFS, Mellanox) and other upstream code providers in the open source community. LLNL will engage Tier 3 support providers as required if problems cannot be resolved locally. (In the case of SLURM, LLNL is both the Tier 2 and Tier 3 support provider.) In some cases, LLNL may be able to issue a locally developed fix and then work with the Tier 3 provider to get the fix pushed upstream. In other cases, it may be desirable or necessary to get a fix directly from the Tier 3 support provider. In either case, LLNL will facilitate the process and to pursue getting fixes into the supported code base so as to minimize local patches in TripodOS.

Issue tracking. LLNL will provide a Tri-lab bug tracking system (Bugzilla) allowing each lab to open tickets and track problems on an ongoing basis. Tier 2 staff will be automatically notified when tickets are opened or modified by customers, and customers will be automatically notified as their tickets are updated. Over time, the bug tracking system will also become a knowledge base that can be used to research similar problems, and to provide data for error trending and related QA activities.

As mentioned previously, Tier 1 support personnel will define an initial problem severity when they submit a ticket. The severity will be used determine the initial response time by Tier 2 support and will generally drive the problem resolution process. The LLNL Support Manager will facilitate a weekly Tri-lab issue tracking telecon for the purposes of reviewing status of open tickets, refining priorities, and overall management of the problem resolution process. In the event that a Tri-lab customer is dissatisfied with the handling of a problem through the normal process, a formal escalation to the Tripod Steering Committee (TSC) can be invoked. Additionally, the LLNL Support Manager will provide the TSC with a quarterly out brief of all support issues that were addressed.

Integration assistance. A Tri-lab integration team will be established to provide on-site assistance with initial installation and configuration of TripodOS on TLCC07 systems, and to provide ongoing consultation and training as appropriate to help Tri-lab staff during the transition to this software environment.

Documentation. In addition to Release Notes provided with TripodOS software packages, LLNL will provide a wiki or equivalent to house documentation (e.g., HOWTOs and best practices) and to facilitate information sharing within the Tri-lab community.

TripodOS Evolution

TripodOS is well suited to meet the long-term needs of the ASC Tri-lab community. Process maturity is a key to success. Over the past six years, LLNL has developed and repeatedly evolved CHAOS to support new hardware, integrate new functionality, improve operational capabilities, and accommodate evolving user requirements. While these processes are currently optimized to meet specific LLNL requirements, these are fundamentally a collection of standardized models and best practices that will be extended to TripodOS and accommodate Tri-lab participation. The success of this model has also been demonstrated at SNL. Over the last two years, SNL has developed a common user environment across multiple generations of clusters. The proposed TripodOS model has been proven and is working at both LLNL and SNL.

LLNL, working with SNL, must initially undertake many of the development responsibilities for TripodOS in order to meet the near term requirement to provide a solution for TLCC07. Going forward, however, TripodOS will evolve to a model that establishes package owners at other laboratories. Package owners will have the ability to modify directly the source code for the packages each maintains and each will have the opportunity and indeed the obligation to influence the overall functionality and technical direction of their assigned packages. As protocols for interaction and evolution develop, developers at all three laboratories will assume direct write access to the main TripodOS source code repository.

The TripodOS release management strategy balances the need to provide bug fixes and new capabilities against the need to provide users with stability and long-term compatibility. Since TripodOS is derived from RHEL, a basic three-year TripodOS release roadmap is defined to coordinate with the RHEL product release cycle. Actual TripodOS releases will be determined by requirements to support new TLCC hardware, need to address security vulnerabilities, and new feature requests from the Tri-lab community.

LLNL will facilitate quarterly TripodOS release planning meetings, at which Tri-lab feature requests and other issues pertaining to release content will be discussed and prioritized. Release planning meetings will consider short-term needs for upcoming releases and will also refine the long-term TripodOS roadmap. Tri-lab release planning representatives should be prepared to discuss the needs of their development, support, and end-user communities in this forum. In the event that consensus on an issue cannot be achieved by release planning participants, formal escalation to the Tripod Configuration Management Steering Committee can be invoked. Figure 1 shows the TripodOS lifecycle with Tri-lab participation in all stages of the process.

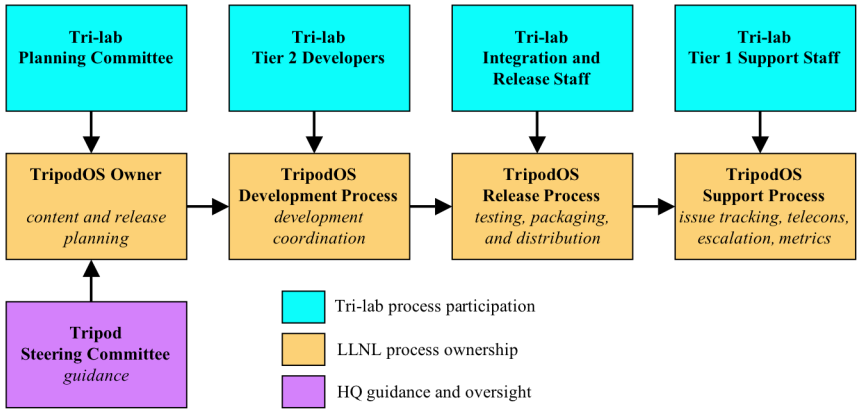


Figure 4. TripodOS lifecycle with Tri-lab participation

Backporting TripodOS to run on existing Tri-lab ASC Linux clusters is technically feasible though not always justifiable from a cost standpoint. As is the case at LLNL, SNL plans to support one software environment across all ASC and institutional clusters, SNL will upgrade the Thunderbird cluster to TripodOS after it is ready for TLCC07. However, old SNL clusters that are nearing end-of-life will maintain their current software environment due to the unjustified costs of upgrading on the systems with a limited lifetime. TripodOS currently supports all major COTS architectures (x86, x86_64, Itanium) and most cluster interconnects (Ethernet, Elan3, Elan4, Infiniband) but may lack support for specific hardware such as motherboard chipsets. Depending on the similarity between existing Tri-lab clusters and systems already deployed at LLNL, the effort may be very large and the impact on users may be significant. As requested, LLNL will work with other labs to evaluate the feasibility of and level of effort required to backport TripodOS to run on existing Linux clusters.

Metrics for Success

Metrics will be used to evaluate the quality of the TripodOS product and the quality of delivery and operation of this product. The following metrics will be used.

Product quality

Given the initial lack of a Tri-lab QA validation suite, the following metrics will be used during Year 1 to verify the compliance and stability of the TripodOS software stack for initial installations on TLCC07.

Metric	Description
Y1PQ1	Successful execution of the TripodOS configuration verification utility.
Y1PQ2	Successful execution of three mixed MPI/OpenMP jobs (sPPM, UMT2K, and LINPACK) sequentially or simultaneously across 90% of the SU compute nodes for at least four hours with correct results and without software failure.
Y1PQ3	Successful execution of the Presta MPI stress test sequentially or simultaneously across 90% of the SU compute nodes for four hours without software failure.

As TripodOS is deployed across Tri-lab TLCC clusters, the QA validation suite will evolve to where it can provide verification of TripodOS compliance and stability on major releases. Expected for year 2 and beyond, the following metrics will be used.

Metric	Description
PQ1	Successful execution of the TripodOS configuration verification utility.
PQ2	Successful execution of the TripodOS QA validation suite.

Operational excellence

The following operational excellence metrics will be tracked and provided to the Tri-lab by the TripodOS Support Manager on a quarterly basis and on demand.

Metric	Description
OE1	Tier 2 support will meet committed customer response times 90% of the time.
OE2	Bug tracking system and RPM repository will be 98% available.
OE3	Wiki and other documentation repositories will be 98% available.
OE4	Distribution of RHEL security patches for local or remote root exploits will occur within one business day 90% of the time.

Service Level Agreement

Upon acceptance of this proposal, LLNL will work with SNL, LANL, and HQ to put into place a Tri-lab service level agreement (SLA) consistent with the proposal.

Schedule

ASC benefits from the significant CHAOS development and integration work that has already been done for the Peloton procurement, which enables delivery of an accelerated solution for TLCC07. A release of TripodOS with support for TLCC07 hardware will be available within eight weeks of vendor selection. TripodOS will be available for use during build, pre-ship and post-ship testing, and acceptance of TLCC07 hardware at all Tri-lab sites.

Budget

The budget for the TripodOS development, support, and management effort is $4M per year, the detail of which is provided in Appendix D. Per year funding is split $2M for LLNL and $1M each for SNL and LANL. LLNL has an increased funding requirement due to its responsibilities for the distro release repository, testing, Tier 2 support, vendor Tier 3 support coordination, etc. SNL and LANL funding provides for Tier 1 and 2 support, RHEL licenses, site-specific integration, and other Tripod software development efforts. It is assumed that the budget corresponding to existing efforts at the labs that are supplanted by TripodOS will in turn be used to provide support for TripodOS efforts; i.e., the proposed budget does not cover all direct costs that are incurred by the labs using TripodOS but that any gaps are filled by budgets from defunct activities.

Appendix B — TripodOS Governance

TripodOS Software (TOSS) Governance

Lawrence Livermore National Laboratory
Los Alamos National Laboratory
Sandia National Laboratories
NNSA Advanced Simulation and Computing Program

13 September 2007

Introduction

The Tripod Operating System Software (TOSS) is the Tri-laboratory software stack born of an Advanced Simulation and Computing (ASC) Program Headquarters policy decision to utilize a common software stack across all newly procured Linux capacity clusters, initiating with Tri-laboratory Capacity Computing (TLCC) platforms in FY08. The TOSS governance model presented in this document outlines both the management of day-to-day support as well as an orderly process for the evolution of the stack to meet the growing needs of the NNSA and ASC Tri-lab community.

The goal of the TOSS effort is to increase efficiencies in the ASC Tri-lab community with respect to both the utility and the cost of the capacity computing environment. The process to secure these benefits on a continuing basis is described by this governance model. The objective of this governance model is the establishment and continuation of a fully collaborative and collegial relationship with full participation of all the partner laboratories, as required for achieving the ultimate goal of delivering and supporting a single software stack for ASC capacity systems now and into the future.

Described next are the three committees key to governance of TOSS. The Release Planning Committee is responsible for guiding the strategic evolution of the TOSS product. The Support Committee is responsible for prioritization of work directed at bugs and other performance issues and ensuring fixes are rolled into appropriately spaced software revisions. There is no intended hierarchy between the Release Planning and Support Committees. The TOSS Executive committee is primarily an arbitration and high-level management committee responsible for the resolution of issues not resolvable at lower levels. The Executive Committee also provides Program guidance as necessary. Figure 1 depicts the relationships between these committees.

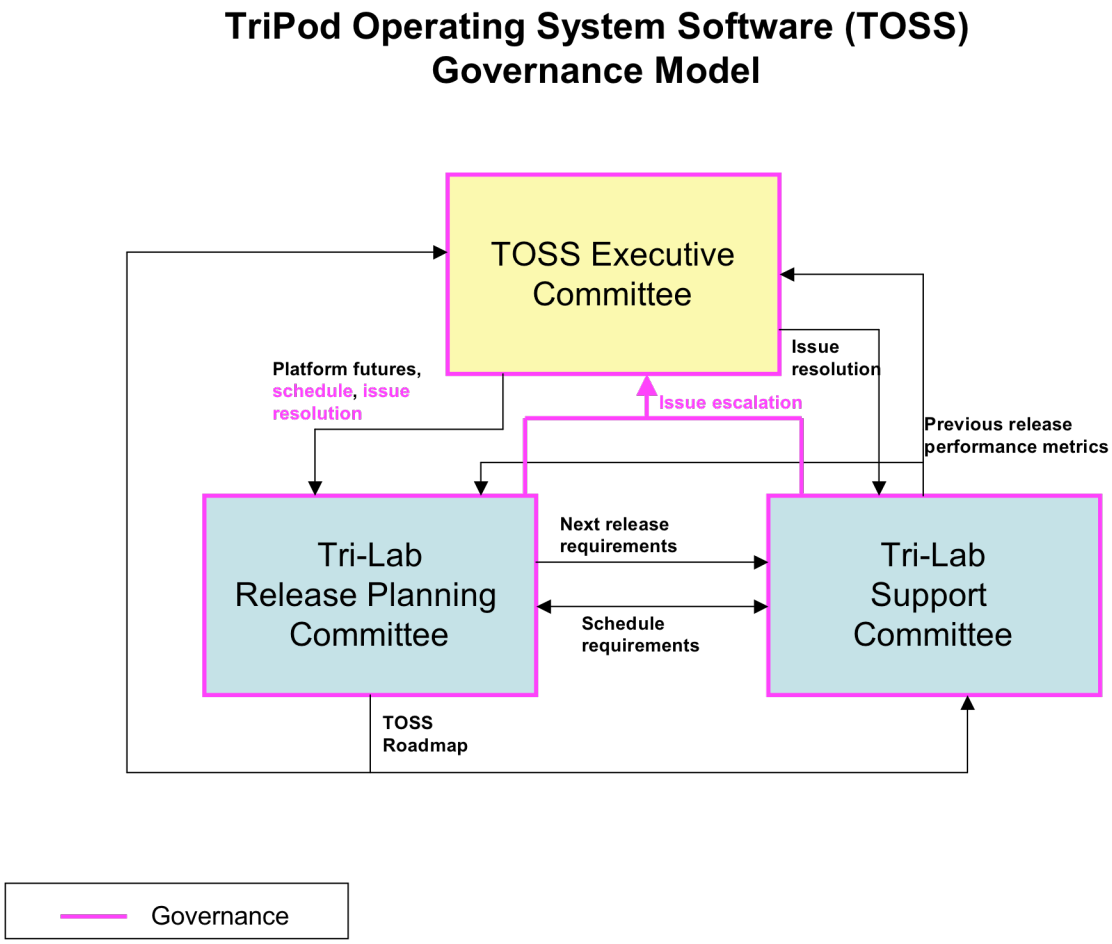


Figure 1: TOSS Governance Model

TOSS Release Planning Committee

Role

This group meets *at least* monthly for software stack content, release tracking and planning purposes. The more frequent meetings (frequency to be decided by the Release Manager with input from the Committee) will consider short-term needs and status of software development to be included in upcoming releases. Quarterly meetings (or as required) will refine the long-term TOSS roadmap, including ongoing alignment with the selected Linux distribution roadmap, currently RHEL (Red Hat Enterprise Linux), and any guidance from the TOSS Executive Committee regarding new platform procurements, budget restrictions, policy refinements, etc. Tri-lab feature requests and other issues pertaining to release content will be discussed and prioritized. It is the responsibility of members of this committee to bring forward requests coming from their laboratory, weigh objectively any global issues, and prioritize work effort for the common good. It is also the members' responsibility to understand the status of all products under development at their respective laboratory, and to accurately communicate issues (e.g., resource constraints, security significance, etc.) and progress in a timely manner in both directions. In the event that consensus or compromise cannot be achieved in the area of future

directions or release schedules, formal escalation to the TOSS Executive Committee can be invoked.

Membership
This is a six person committee consisting of two representatives from each laboratory, including a TOSS Release Planning Manager (see Appendix A) responsible for coordinating meetings, setting agendas, insuring product owners are on schedule, identifying as early as possible any issues affecting the release schedule, and leading interactions with vendors. The members are chosen by each laboratory's ASC Executive. The members of this group must have detailed technical experience in this area and, optimally, each should be involved in the development of TOSS. The members should be prepared to discuss the needs of their development, support, and end-user communities in this forum; in addition, members should be collaborative and communicate well.

TOSS Support Committee

Role
The TOSS Support Committee meets at least weekly and is responsible for reviewing status of open tickets, refining priorities, and overall management of the problem resolution process. The Support Committee is also responsible for providing the Release Planning Committee members with information they need, such as support performance metrics, in order to help plan the evolution of TOSS. In the presumably extreme case where a Tri-lab customer is dissatisfied with the handling of a problem through the normal process, formal escalation to the TOSS Executive Committee can be invoked.

Membership
This is a three person committee. There will be one member from each laboratory, which includes the TOSS Support Manager (see Appendix A) responsible for weekly meeting coordination and bug ticket routing and follow-up. The members are chosen by each laboratory's ASC Executive. The members of this group must be technically qualified and involved in the development, deployment, or support of TOSS software at their respective lab. The members should be prepared to discuss the details of reported bugs and must have a through understanding of platform specifics. In addition, members should be collaborative and communicate well.

TOSS Executive Committee

Role
The TOSS Executive Committee communicates quarterly with the Release Planning committee and meets a minimum of once per year to discuss new platform futures, budget projections, and other related high level information that may effect the evolution of TOSS. Further, the Executive Committee members are responsible for resolution of issues escalated to them by the Release Planning and Support Committees.

Membership
This four person committee consists of one mid- to high-level manager from each laboratory as well as HQ's TriPod executive. The members are appointed by their laboratory's ASC executive.

TOSS Ongoing Support Model—A Summary

Below is a summary of the TOSS Ongoing Support Model that is described in detail in the TOSS Transition Plan document. A summary is provided here for convenience of the reader.

On a day-to-day basis, Tier 1 support representatives at each laboratory provide operational support for local ASC capacity systems and users. In the event that Tier 1 support cannot resolve a user issue, or if some system-level or other operational deficiency in TOSS is encountered, Tier 1 will escalate the problem by opening a ticket in the TOSS bug tracking system. The ticket will include, among other information, the following information

- Specification of a severity level using established Tri-lab severity definitions.
- Categorization of the issue; e.g., bug report, feature request, informational, etc.
- Identification of affected component; e.g., kernel, IB, MPI, resource manager, etc.

Tier 2 support staff will monitor the TOSS bug tracking system during an established principal period of maintenance (PPM) and will provide an initial response to the customer in the timeframe defined for the specified severity level. Subsequent to providing an initial response to the customer, Tier 2 support staff will continue to pursue problem resolution in a manner consistent with the problem severity and as determined by priorities established in weekly TOSS Support Committee issue tracking telecons. At times, and in the interests of keeping production systems highly available, it may be necessary for a laboratory to apply an available fix or existing solution before an official patch is released. In these cases, it is expected that the laboratory will apply with the official patch once it becomes available.

The Support Manager will facilitate the weekly TOSS issue tracking telecon for the purposes of reviewing status of open tickets, refining priorities, and overall management of the problem resolution process. In the presumably rare case where a Tri-lab customer is dissatisfied with the handling of a problem through the normal process, formal escalation to the TOSS Executive Committee can be invoked. As depicted in Figure 1, the Support Manager will also provide the TOSS Release Planning Committee with metrics to measure both the quality of the TOSS product as well as the support.

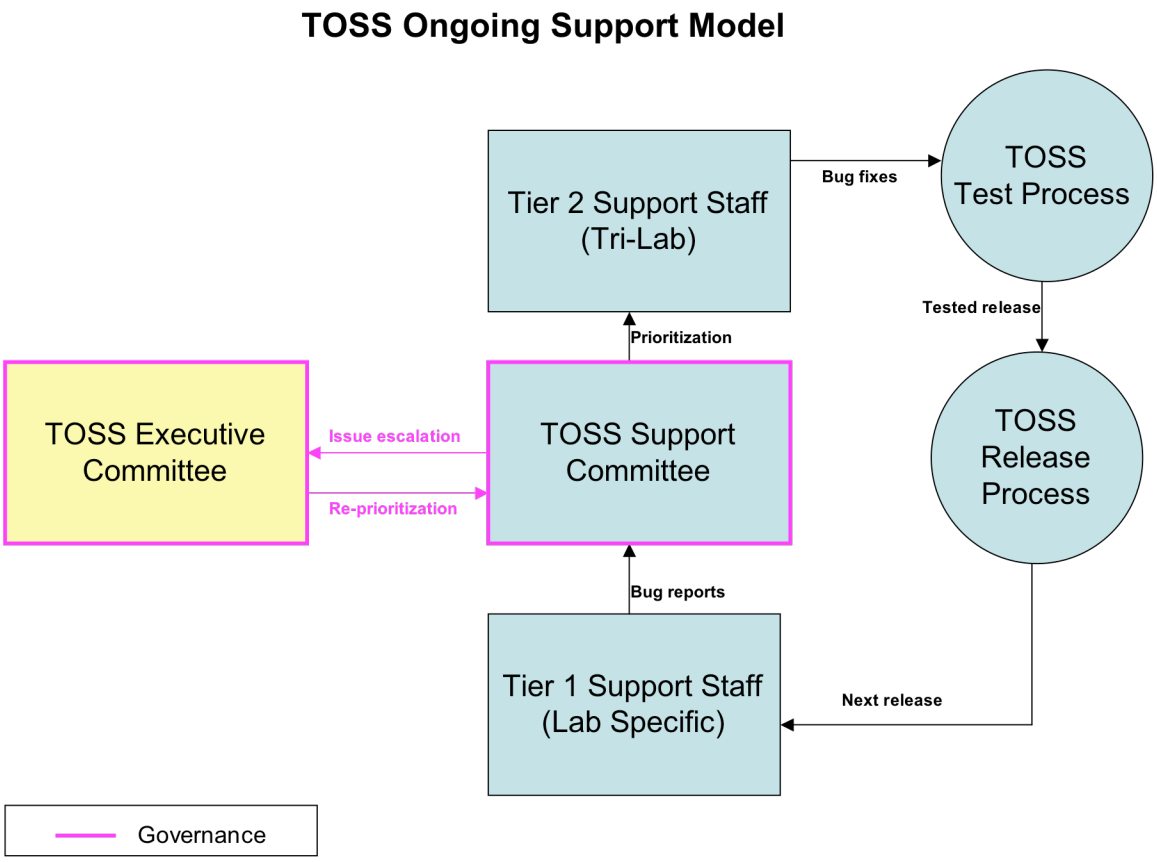


Figure 2: TOSS Ongoing Support Model

Ongoing Support Severity Levels and Tier 2 Response Times
The principal period of maintenance (PPM) is M-F 0730 through 1700 Pacific Time excluding LLNL holidays. Response times are in reference to the PPM.

Severity Level	Definition	Response Time
1	System is down or otherwise non-functional. Impact is global (affects all users).	1 business hour
2	System is severely degraded in terms of performance and/or functionality. Impact is global (affects all users).	2 business hours
3	System is impaired but usable. Impact is localized (only affects some users or may affect all users at a minimal level).	4 business hours
4	System is minimally impaired. This severity level includes informational reports, feature requests, etc.	1 business day

Tiger Team Support
There may be cases where a severity level 1 or 2 problem is encountered at a specific laboratory and the cause of the issue is difficult to diagnose. At the discretion of the Support Manager and at the request of the affected laboratory, a Tiger Team of appropriate personnel may be deployed to troubleshoot the issue locally rather than remotely.

Reporting Metrics
Metrics measuring *product quality* and *operational excellence* in supporting TOSS will be provided by the Support Manager to the TOSS Release Planning Committee.

Product quality

Given the initial lack of a Tri-lab QA validation suite, the following metrics will be used during Year 1 to verify the compliance and stability of initial installations of TOSS.

Metric	Description
Y1PQ1	Successful execution of the TOSS configuration verification utility.
Y1PQ2	Successful execution of mixed MPI/OpenMP jobs (sPPM, UMT2K, LINPACK, QCD, and LAMMPS) sequentially or simultaneously across 90% of the SU compute nodes for at least four hours with correct results and without software failure.
Y1PQ3	Successful execution of the Presta MPI stress test sequentially or simultaneously across 90% of the SU compute nodes for four hours without software failure.
Y1PQ4	Successful execution of Lustre and Panasas parallel file system tests.

As TOSS is deployed, the QA validation suite is expected to evolve to where it can provide verification of TOSS compliance and stability on major releases. Expected for year 2 and beyond, the following metrics will be used.

Metric	Description
PQ1	Successful execution of the TOSS configuration verification utility.
PQ2	Successful execution of the TOSS QA validation suite.

Operational excellence

The following operational excellence metrics will be tracked and provided to the TOSS Release Planning Committee by the Support Manager on a quarterly basis and on demand.

Metric	Description
OE1	Tier 2 support will meet committed customer response times 90% of the time.
OE2	Bug tracking system and RPM repository will be 98% available.
OE3	Wiki and other documentation repositories will be 98% available.
OE4	Distribution of RHEL security patches for local or remote root exploits will occur within one business day 90% of the time.
OE5	Number of software bugs reported by priority level.
OE6	Number of software bugs resolved by priority level.
OE7	Average time to resolve software bugs by priority level.
OE8	Maximum time to resolve software bugs by priority level.

Appendix A
Definitions, Roles, and Responsibilities

Tri-lab Tier 1 support staff: These individuals are the support representatives at each laboratory responsible for providing operational support for the local ASC capacity systems and customers by investigating trouble reports, and diagnosing and resolving problems.

Tri-lab Tier 2 staff: These individuals contribute to the development and ongoing evolution of TOSS and represent the next level of response, should Tier 1 fail to fully resolve an issue. They participate in monitoring the TOSS bug tracking system, debugging issues, and developing solutions or pursuing solutions from upstream providers as required.

Tri-lab integration and release staff: These individuals are responsible for coordination of local integration, testing, and deployment of new TOSS releases, bug fixes, etc. These individuals may or may not be the same people providing Tier 1 support but should have strong operational involvement in the production capacity environment. While integration staff normally works locally, it is presumed that members who are either highly skilled or who possess unique specialties will be willing to be dispatched to sister laboratories for short periods to help either with initial integration or to address subtle and insidious problems.

TOSS Support Manager: This individual has overall responsibility for ensuring that the operational support model for TOSS is working effectively. The Support Manager proactively monitors the trouble ticket system to maintain awareness of the Tier 2 work flow, including high priority issues being worked, tickets in need of assignment or escalation, resource conflicts or constraints, etc. The Support Manager also facilitates a weekly TOSS issue tracking telecon and provides quarterly statistics on support-related activities.

TOSS Release Planning Manager: This individual has overall accountability for ensuring that TOSS delivers the capabilities described in the proposal and that it meets the ongoing needs of the ASC Tri-lab community. The Release Planning Manager oversees the development and release planning processes, and coordinates with the Support Manager to ensure that operational issues and trends are appropriately considered in the release management strategy. The Release Planning Manager is also responsible for coordination with product owners across the laboratories, ensuring development efforts are on schedule, identifying and resolving release schedule issues as early as possible, and leading interactions with vendors. The Release Planning Manager is responsible for coordination of regular telecons and quarterly strategic planning meetings.

Appendix C — Clarification of Intent and Completion Criteria

Clarification of Intent and Completion Criteria for Multi-Site Incentive MRT #2779

NNSA Advanced Simulation and Computing Program
Lawrence Livermore National Laboratory
Los Alamos National Laboratory
Sandia National Laboratory

20 March 2008

This document clarifies the intent and associated completion criteria in reference to the "deploy and operate" clause within the current TriPoD common operating system (TripodOS or TOSS) Multi-Site Incentive (MRT #2779) and has been agreed to by the respective Laboratory representatives and the Advanced Simulation and Computing (ASC) Program Office. Multi-Site Incentive MRT #2779 states that:

By 31DEC07, implement the Tripod configuration management structure, including establishment of Tripod common operating system (TripodOS) Executive and Steering Committees. By 31MAR08, demonstrate the TripodOS stack on the initial hardware of the ASC-funded Tri-lab Linux Capacity Clusters (TLCCs). By 30SEP08, deploy and operate all ASC TLCC platforms, that will be delivered before mid August, under the generally available (GA) release of TripodOS. Furthermore, position TripodOS for future deployment on all ASC Linux capacity platforms delivered in FY09 and beyond.

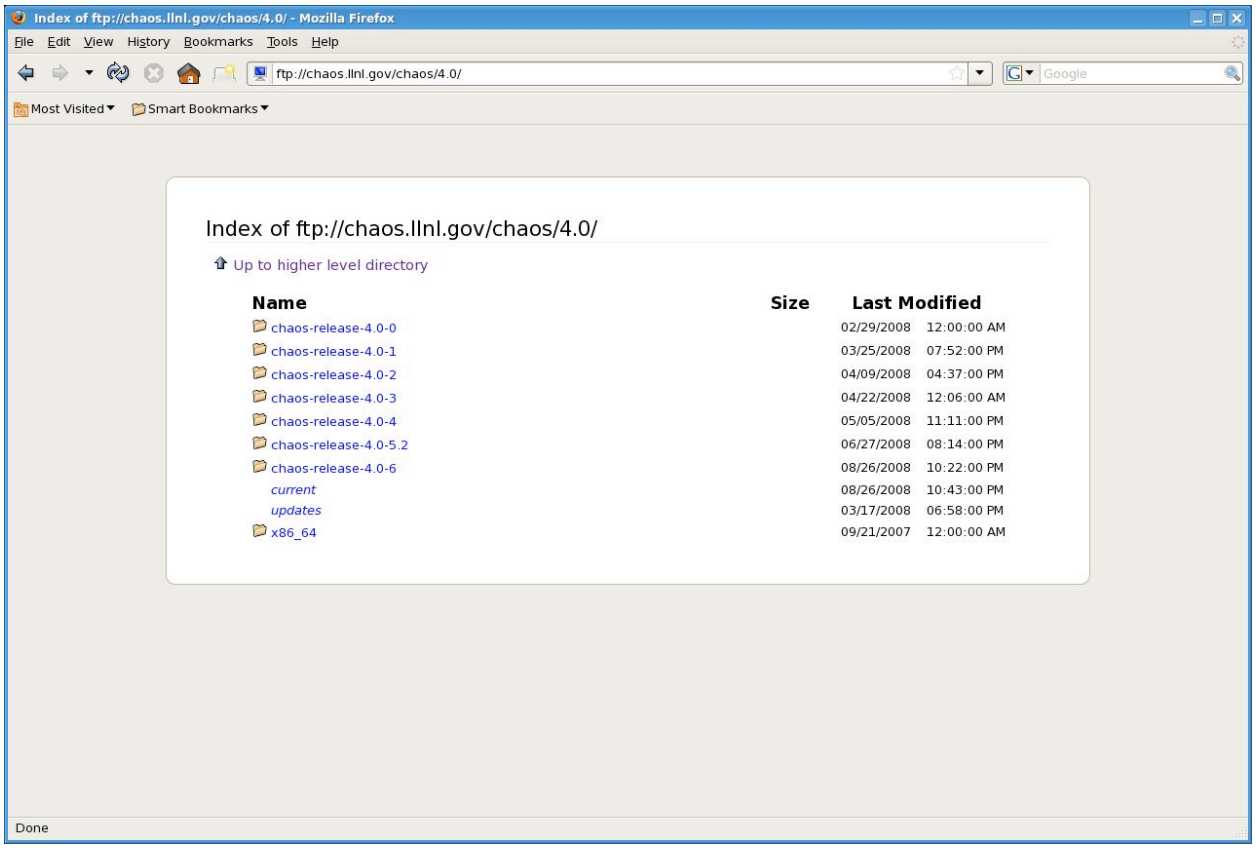
The intent and completion criteria associated with the "deploy and operate" clause are:

- Hardware (nodes, switches, etc.) delivered before mid-August has been installed.
- TripodOS GA (including subsequent associated patches or releases) software stack has been installed.

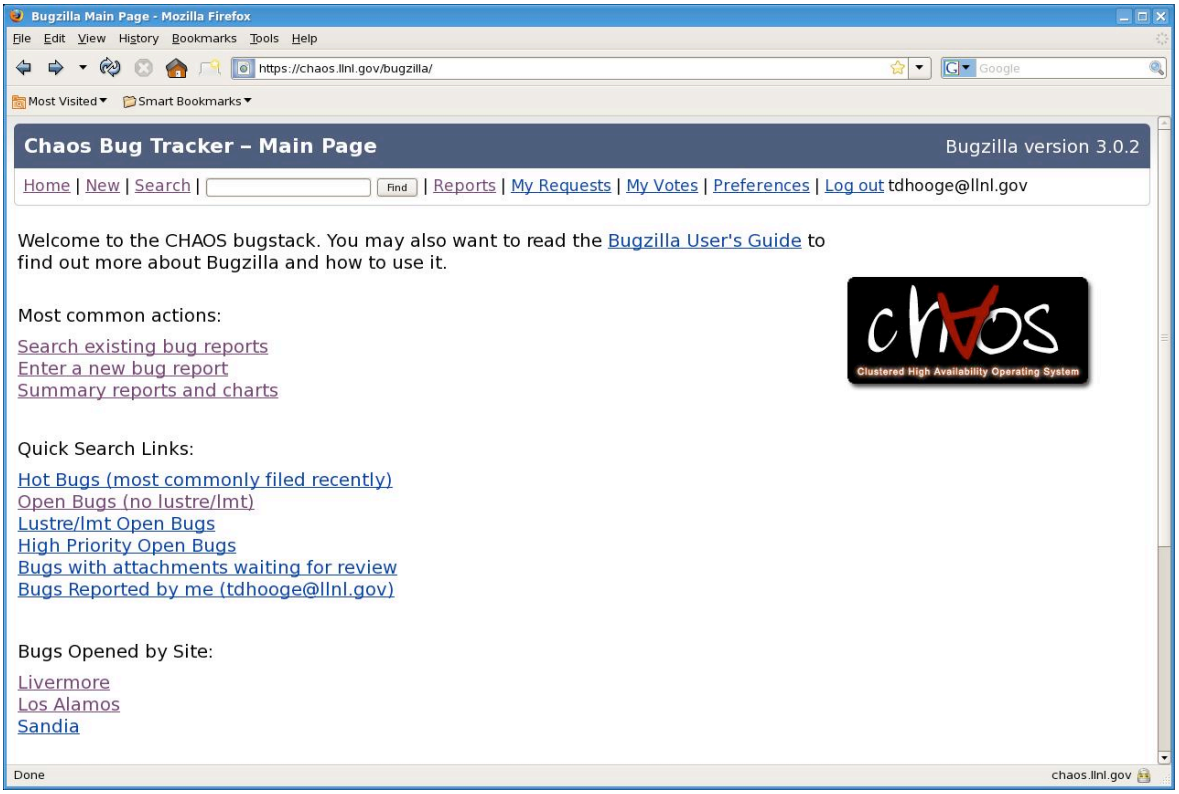
- Application development environment (compilers, libraries, debuggers, etc.) necessary for major targeted applications for the platform has been installed.
- Basic functionality tests can be run across the platform, though an entire Synthetic Work Load (SWL) need not be run. Basic functionality tests include the ability to launch parallel jobs through SLURM (not necessarily through MOAB). Functionality tests do not include performance or stability tests.
- Operation and deployment as described above does not have to be accomplished on the final intended network. That is, basic functionality testing of hardware, OS stack, and the application environment can be accomplished on an unclassified network even if the system's final residence will be on a classified network.

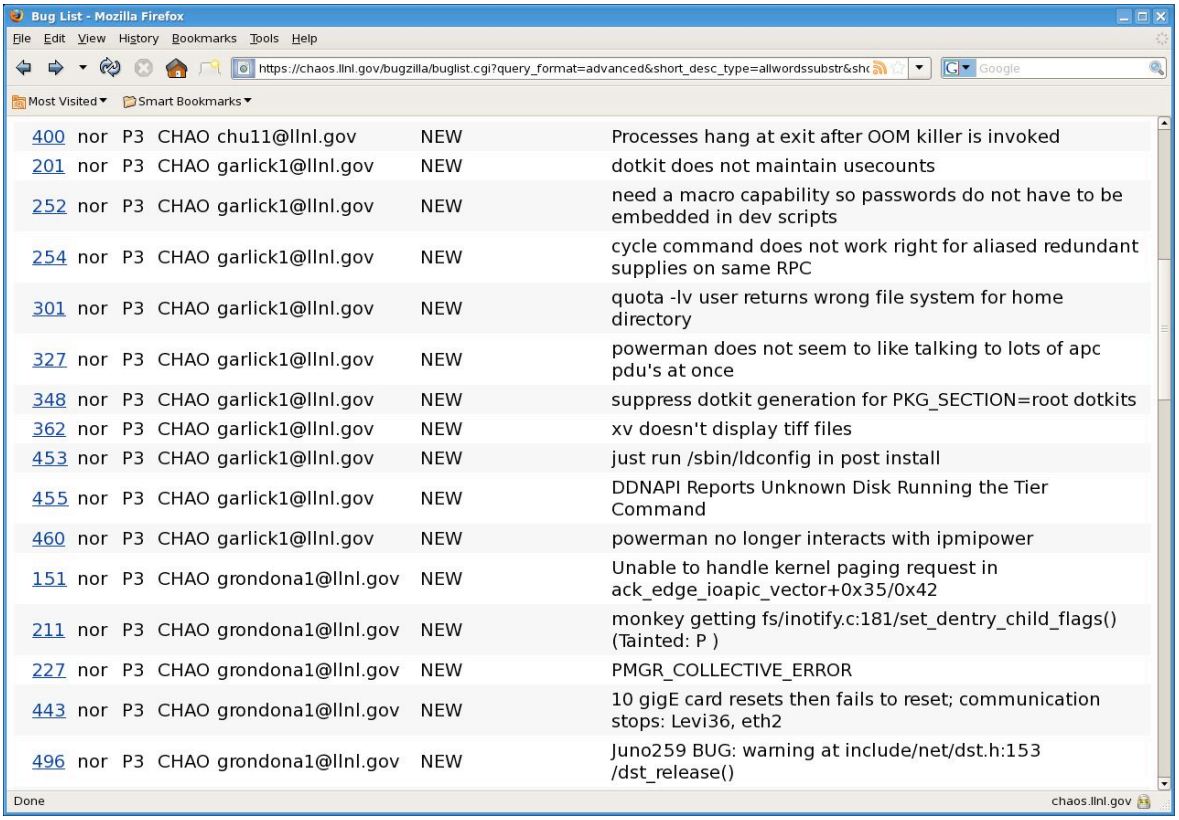
Appendix D — Support Websites Screenshots

Software Repository

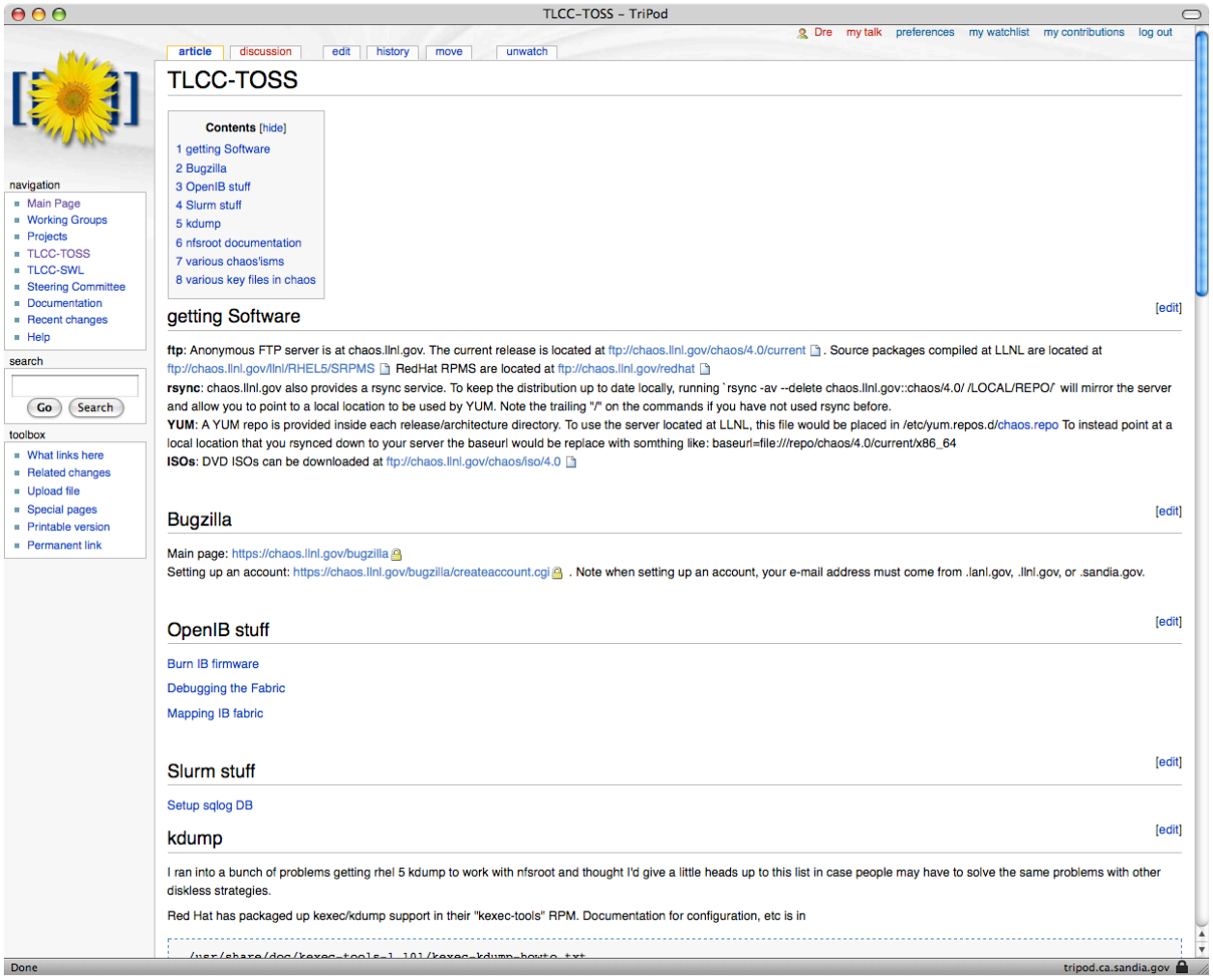


Bug Tracking System





Documentation Wiki



Appendix E — MRT 2779 Criteria Sign-off Sheets

LANL Clusters (Lobo, Hurricane [not needed for MRT 2779 completion])

Sign-off Sheet for Completion Criteria for Multi-Site Incentive MRT #2779

Cluster Name: Lobo

Completion Date: May 28, 2008

☒ Hardware (nodes, switches, etc.) delivered before mid-August has been installed.

☒ TripodOS GA (including subsequent associated patches or releases) software stack has been installed. [Sample: Run "uname –r".]

☒ Application development environment (compilers, libraries, debuggers, etc.) necessary for major targeted applications for the platform has been installed. [Sample: Run each tool and collect version information.]

☒ Basic functionality tests can be run across the platform, though an entire Synthetic Work Load (SWL) need not be run. Basic functionality tests include the ability to launch parallel jobs through SLURM (not necessarily through MOAB). Functionality tests do not include performance or stability tests. [Sample: Launch a parallel job through SLURM, verify that SWL Functional Tests pass.]

NOTE: Operation and deployment as described above does not have to be accomplished on the final intended network. That is, basic functionality testing of hardware, OS stack, and the application environment can be accomplished on an unclassified network even if the system's final residence will be on a classified network.

Tested by:

Timothy O. Harrington, LANL — Date: 9-8-08

Concurrence:

Josip Loncaric, LANL TLCC Technical POC — Date: 9/8/08

Sign-off Sheet for Completion Criteria for Multi-Site Incentive MRT #2779

Cluster Name: Hurricane

Completion Date: August 29, 2008

- [x] Hardware (nodes, switches, etc.) delivered before mid-August has been installed. NOTE: Hurricane was delivered 8-25-08.
- [x] TripodOS GA (including subsequent associated patches or releases) software stack has been installed. [Sample: Run "uname -r".]
- [x] Application development environment (compilers, libraries, debuggers, etc.) necessary for major targeted applications for the platform has been installed. [Sample: Run each tool and collect version information.]
- [x] Basic functionality tests can be run across the platform, though an entire Synthetic Work Load (SWL) need not be run. Basic functionality tests include the ability to launch parallel jobs through SLURM (not necessarily through MOAB). Functionality tests do not include performance or stability tests. [Sample: Launch a parallel job through SLURM, verify that SWL Functional Tests pass.]

NOTE: Operation and deployment as described above does not have to be accomplished on the final intended network. That is, basic functionality testing of hardware, OS stack, and the application environment can be accomplished on an unclassified network even if the system's final residence will be on a classified network.

Tested by:

Timothy O. Harrington, LANL — Date: 9-8-08

Concurrence:

Josip Loncaric, LANL TLCC Technical POC — Date: 9/8/08

LLNL (Hype, Juno)

Sign-off Sheet for Completion Criteria for
Multi-Site Incentive MRT #2779

Cluster Name: Hype

Completion Date: 14 April, 2008

- θ Hardware (nodes, switches, etc.) delivered before mid-August has been installed.
- θ TripodOS GA (including subsequent associated patches or releases) software stack has been installed. [Sample: Run "uname -r"]
- θ Application development environment (compilers, libraries, debuggers, etc.) necessary for major targeted applications for the platform has been installed. [Sample: Run each tool and collect version information, build a parallel job then launch it through SLURM.]
- θ Basic functionality tests can be run across the platform, though an entire Synthetic Work Load (SWL) need not be run. Basic functionality tests include the ability to launch parallel jobs through SLURM (not necessarily through MOAB). Functionality tests do not include performance or stability tests. [Sample: SWL Functional Tests pass].
- θ Operation and deployment as described above does not have to be accomplished on the final intended network. That is, basic functionality testing of hardware, OS stack, and the application environment can be accomplished on an unclassified network even if the system's final residence will be on a classified network.

Tested by:

Trent D'Hooge, LLNL — Date: 5-Sept-2008

Concurrence:

Mark Seager, LLNL TLCC Technical POC — Date: 5 SEPT 2008

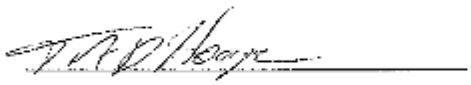
Sign-off Sheet for Completion Criteria for
Multi-Site Incentive MRT #2779

Cluster Name: Juno

Completion Date: 29 August, 2008

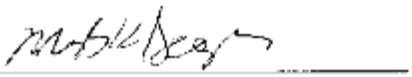
- θ Hardware (nodes, switches, etc.) delivered before mid-August has been installed.
- θ TripodOS GA (including subsequent associated patches or releases) software stack has been installed. [Sample: Run "uname –r"]
- θ Application development environment (compilers, libraries, debuggers, etc.) necessary for major targeted applications for the platform has been installed. [Sample: Run each tool and collect version information, build a parallel job then launch it through SLURM.]
- θ Basic functionality tests can be run across the platform, though an entire Synthetic Work Load (SWL) need not be run. Basic functionality tests include the ability to launch parallel jobs through SLURM (not necessarily through MOAB). Functionality tests do not include performance or stability tests. [Sample: SWL Functional Tests pass].
- θ Operation and deployment as described above does not have to be accomplished on the final intended network. That is, basic functionality testing of hardware, OS stack, and the application environment can be accomplished on an unclassified network even if the system's final residence will be on a classified network.

Tested by:



Trent D'Hooge, LLNL — 5-Sept-2008 — Date

Concurrence:



Mark Seager, LLNL TLCC Technical POC — 9 SEPT 2008 — Date

SNL-NM (Unity)

Sign-off Sheet for Completion Criteria for
Multi-Site Incentive MRT #2779

Cluster Name: Unity
Completion Date: June 2, 2008

- [x] Hardware (nodes, switches, etc.) delivered before mid-August has been installed.
- [x] TripodOS GA (including subsequent associated patches or releases) software stack has been installed. [Sample: Run "uname –r".]
- [x] Application development environment (compilers, libraries, debuggers, etc.) necessary for major targeted applications for the platform has been installed. [Sample: Run each tool and collect version information.]
- [x] Basic functionality tests can be run across the platform, though an entire Synthetic Work Load (SWL) need not be run. Basic functionality tests include the ability to launch parallel jobs through SLURM (not necessarily through MOAB). Functionality tests do not include performance or stability tests. [Sample: Launch a parallel job through SLURM, verify that SWL Functional Tests pass.]

NOTE: Operation and deployment as described above does not have to be accomplished on the final intended network. That is, basic functionality testing of hardware, OS stack, and the application environment can be accomplished on an unclassified network even if the system's final residence will be on a classified network.

Tested by:

Marcus Epperson, <SNL> — Date September 10, 2008

Concurrence:

Robert A. Ballance, Acting

John Noe, <SNL> TLCC Technical POC — Date September 10, 2008

Appendix F — FY08 Final Report: Tripod Working Groups

FY08 Final Report
Tripod Working Groups and Projects
September, 2008

Executive Summary

Our Tripod mission continues to be, to enable a common capacity computing environment in FY08 and an ASC common capability computing environment in 2013. To achieve these goals, in FY07, Tripod working groups identified and documented technology gaps that were used as the basis for Tripod "areas of concern" in a Tripod Call for FY08 Project Proposals.

Tripod FY08 Projects were selected that clearly addressed the Tripod goals and the following areas of concern:

- Application Performance
- Standardization
- Shared Environment
- Cost Minimization
- System Resilience
- User Support

In FY08, the Working Groups were assigned responsibility for guidance and oversight of the funded projects. The relationships between Working Groups, with their areas of interest, and Funded Projects can be summarized as the following:

- Compilers and Application Development Tools Working Group (includes Application Development Tools, interfaces to Visualization tools, HPSS, File Systems)
    - Open|SpeedShop Project
- Resource Management Working Group (includes MOAB; resource management tools; common scheduling software (SLURM); resource usage reporting)
    - Workload Characterization Project
- HPC Monitoring & Metrics Working Group (includes monitoring, resource usage statistics, RAS metrics)
    - Application Monitoring Project
- Tri-lab Environment Working Group (configuration management; license sharing & management; tri-lab WAN issues; multi-domain security issues)
    - Shared Work Space Project

There are inter-project relationships as well:

- TripodOS Software Stack (TOSS) Project
    - Gazebo Test & Analysis Suite Project

Tripod Projects for the next FY are proposed by the Tripod Working Groups to address new technology gaps identified and documented by the Working Groups.

The following sections of text discuss each Project's goals and FY08 accomplishments, as well as reports from each Working Group. FY09 Tripod Projects are documented in the FY09 ASC Implementation Plan as part of a new CSSE product, Common Computing Environment (CCE).

Common Computing Environment

The goal of the Common Computing Environment (CCE) product is to enable such an environment across the tri-labs which will initially be deployed on the TLCC systems. The scope of this product includes funded R&D projects to address gap areas identified by the tri-lab technical Working Groups.

The CCE Working Groups and Projects focus on a common software stack to include, but not be limited to, operating system software; application development tools; resource management; HPC monitoring and metrics; and common tri-lab environment issues of configuration management, licenses, WAN access, and multi-realm security, to name a few.

FY08 Key Accomplishments for Common Computing Environment (CCE)

- Released TOSS version 1 on schedule, March 3, 2008, for general availability (GA). All three labs passed the synthetic workload acceptance tests for their respective TLCC systems using the TOSS software stack.
- Developed requirements and prototypes for application monitoring tool.
- Defined tri-lab requirements for the shared workspace environment.
- Used the Gazebo acceptance test package as the underlying infrastructure for testing each of the new TLCC systems. During the TLCC acceptance phase, the test package was used to quantify system node utilization and test coverage, leading to improved system stabilization.

Common Computing Environment Critical Deliverables Planned for FY09

- Deliver tri-lab Level 2 milestone "Deploy Tripod capabilities for capacity computing environment".
- Deploy TOSS 1.2 (Based on RHEL 5.3).
- Implement consistent configuration management for TOSS, including evidence of a single code repository, use of this repository to manage releases of TOSS, and use of an appropriate release process including adequate testing.
- Deploy tri-lab productivity on demand (Tripod) capabilities, including O|SS, Shared Work Space, and Gazebo Test and Analysis Suite.
- Deploy tri-lab performance monitoring tools that provide consistent analysis and reporting of system usage and workload characterization across the tri-labs.
- Implement production baseline of application monitoring tool with prototype of advanced implementation.

Tripod Operating System Software (TOSS)

The TOSS is the tri-lab software stack to run across all newly procured Linux capacity clusters, initiating with TLCC platforms delivered in FY08. The goal of the TOSS project is to increase efficiencies in the ASC tri-lab community with respect to both the utility and the cost of the CCE. This project delivers a fully functional cluster operating system (kernel, Linux distribution, IB stack and related libraries, and resource manager) capable of running MPI jobs at scale on TLCC hardware. The system is to meet Tripod requirements for proving a common software environment on TLCC hardware across the tri-lab complex, now and into the future.

TOSS provides a complete product with full life-cycle support. Well-defined processes for release management, packaging, QA testing, configuration management and bug tracking are used to ensure that a production-quality software environment can be deployed across the tri-lab in a consistent and manageable fashion.

FY08 Accomplishments:

TOSS version 1 was released on schedule, March 3, 2008, for general availability. All three labs passed the synthetic workload acceptance tests for their respective TLCC systems using the TOSS software stack.

Open|SpeedShop *(O|SS)*

O|SS is targeted to be the main performance analysis tool set for the tri-lab ASC capacity machines. It is being developed jointly between the tri-lab partners and the Krell Institute. It provides many typical performance analysis steps in a single environment, including basic profiling in various forms, as well as MPI, I/O, and floating pointing exception tracing.

The O|SS efforts within Tripod include both maintenance and stability improvements to ensure the toolset's usability by our tri-lab user community as well as the design, implementation, and integration of new tools and scalability enhancements matching the demands of both tri-lab applications and TLCC hardware. The scalability enhancements will allow the project to investigate O|SS for ASC capability systems.

FY08 Accomplishments:

- Installation and Support
  - New, unified installation support for base components, libraries and the main tool developed and integrated into mainline code.
  - Successful port to Chaos 4 (offline and MRNet versions) and new TLCC clusters.
  - Extensive testing of newly developed offline and MRNet versions, which are crucial for planned scaling work.
- User Training and Education
  - Held several tutorials at several events.
  - Organized workshop with tool developer community to discuss future directions and leveraging of external tool sets.
  - Direct user support at all three laboratories.
- Plugin Development
  - Completed prototype port and integration of MPIP into O|SS offline infrastructure.
  - Redesign of Javelina to match internal structure of O|SS.

Workload Characterization

The Workload Characterization project will develop a tri-lab common reporting interface for compute resource requirements for current and future use (with programmatic characterization of the work), and for platform usage data, tied to the programmatic characterization of the work.

Development and integration of tri-lab performance monitoring tool(s) may include: developing new functionality in SNL's HPC Estimations and Requirements Tool (HERT); modifying existing local laboratory tools; and integrating HERT with the Moab resource manager and, possibly, local laboratory tools and databases.

SNL's HERT web-based tool is viewed as the prototype for collecting and reporting, current and future requirements for compute resources, with programmatic characterization of the work. Additional development is needed in the areas of validation, test suites, and common reporting capability; as well as a more general mechanism that can interface to multiple existing databases at each of the tri-labs.

Resource Management/Moab development will be used to tie HERT estimates, with their respective workload characterization, to job requests and resulting platform usage data.

Application Monitoring

The Application Monitoring project will develop tools that facilitate automated monitoring of production applications on ASC systems.

The tools should provide basic information about a user's job, to answer questions such as the following:

- Is the job making progress?
- How frequently is the job interrupted?
- What are causes and symptoms of interruptions to the job?
- Should the system intervene (for example, to kill or restart the job)?
- Should the system operators or the user be notified?
- How much time and storage is spent preparing for job restarts?

The project will develop a basic set of monitoring tools, along with their system and application interfaces. The result will be an extensible tool that can serve as a framework for future application monitoring functionality.

FY08 Accomplishments:

- Publish design document outlining baseline and advanced functionality along with man pages specifying functionality.
- Review and revise application monitoring database schema for tracking and updating application progress.
- Develop and demonstrate functional prototype for baseline application monitoring tool along with command line and web interfaces.
- Provide interfaces and programming notes to allow status checking and action responses through either built-in defaults or custom, user-supplied scripts.

Shared Work Space

This project will deploy a collaborative on-line environment in which team areas can be created, with tools for communication, document reference, and other project centric tools to support planning and implementation. Ease of access from all of the tri-lab member locations is required.

Capabilities include ability to create team areas; posting of documents; documentation, plan, and reference material; code source repository access; task manager; access via current lab crypto-type cards; and Wiki, forums, email tracking and other tools.

FY08 Accomplishments:

- Currently we plan to establish 2 levels with associated products for tri-pod collaboration:
  - Developer team collaboration: Gforge (100 user support ordered).
  - Work group collaboration: Plone (open-source) - need to still assess sharepoint viability.
- Access issues: We have been testing and discussing with network groups the issues regarding ability to access the tools with respective crypto cards from each lab. This capability is available. We have had successful tests with SNL and it can be done at LANL within the Turquoise network. There is still work needed to formalize and define the breadth that we can make it work.
- Server establishment: Still need to assess further. Recently discovered that SNL has a gforge service. We are working with SNL staff and plan to test accessibility. In addition also found out that SNL has sharepoint open to external users with no license cost - need to assess. Initial plans have us setting up a server at LANL on the Turquoise network unless the SNL connections work out.
- Intellectual Property: An aspect of this project was to predefine a process for how co-development would be impacted by IP issues. We have had initial discussion with tech transfer organizations at one of the labs and it all seems do-able and this effort is what they would like to see occur. We need to write up an initial process and work it through the other labs.

Gazebo Test and Analysis Suite

Gazebo is a collection of software components used to test, monitor, and analyze the health of a HPC system. With Gazebo, suites of system and application tests are run on an HPC system through either a web-based interface or from the system's master control node.

Test results are stored to a file system, and optionally to a database, and "normalized" so that a known set of timings and results establish a baseline for a healthy system. Through a set of tools, system analysts can monitor the health of the target running system and easily detect anomalous behavior.

Capabilities include results and coverage reporting tools (Command Line Interface only); database server and results schema; simple client-server communication protocol for network interaction; server daemon mythd (my test harness daemon); limited web client used for proof of concept; and acceptance test package (Command Line Interface).

Fully integrate CBENCH suite of tests into Gazebo. CBENCH is SNL's suite of test programs and scripts which interrogate and report the status of individual hardware components comprising the cluster.

FY08 Accomplishments

- Used the Gazebo acceptance test package as the underlying infrastructure for testing each of the new TLCC systems. During the TLCC acceptance phase, the test package was used to quantify system node utilization and test coverage, leading to improved system stabilization.

Gazebo scripts were modified to support not only MOAB/Torque based systems, but also to support MOAB/Slurm based systems.

Details of FY08 Accomplishments

LANL component:

- Coordinated LLNL and SNL interaction with Gazebo development.
- Incorporated Slurm support into Gazebo
- Added updates and made bug fixes as necessary to support TriPod usage
- Added node coverage and utilization test reporting to package which is relied upon heavily for the TLCC acceptance testing.
- Code base now under CVS control
- Investigating Ruby on Rails as web interface design mechanism for future requirements for LANL and LLNL.

LLNL component:

Developed interface between Synthetic Work Load (SWL) test suite and Gazebo for Moab job submission. Both direct job submission through Gazebo and indirect submission through Gazebo from the native SWL job management scripts were supported. Used the indirect job

submission strategy in all the LLNL Phase 1 TLCC Pre-Ship and Post-Ship SWL stability testing.

SNL component:

Developed a non-intrusive integration of the Cbench Cluster Testing and Benchmarking Toolkit into Gazebo. The full suite of Cbench testset capabilities (or any subset thereof) can be integrated into the Gazebo testing harness framework such that Gazebo can use the Cbench tests natively. The connection is highly dynamic and supports as many uniquely named Gazebo job submission configs as needed, as well as easy updates to existing submit configs. Cbench test results also are integrated with the Gazebo "trend data" capabilities. Cbench testing that is run via the Gazebo mechanisms also has the ability to be analyzed by native Cbench utilities. In addition, some tweaks to Gazebo were prototyped to make it run successfully across a larger footprint of Trilab cluster systems.

Compilers and Application Development Tools Working Group

Team Focus Areas:

Compilers

In early FY08 we had a cross lab review of compilers in use, licenses, support contracts and approaches to local support. We found that there are areas where we could benefit through added processes. These included:

- Access to each other's vendor bug reported issues and resolution
- Establishing a cross lab level of support with vendors applicable with our users needs. Currently varies by lab / vendor.
- Sharing test suites
- Assessing each others user support levels and capabilities.

Although there are opportunities in this area there are no over-riding concerns that would drive immediate action. We will do an update this year and assess if a project should be established to implement one of the opportunities.

Debuggers

In mid FY08 we had a cross lab review of debuggers in use, licenses, support contracts and approaches to local support. In addition we have discussed issues such as scalability and lightweight debugger support. Currently LLNL is driving a scalability project with TotalView and is prototyping a light weight debugger (STAT). In addition we are starting to discuss these issues with other labs such as ORNL. This area is more dynamic and we plan on a more detailed review session this year.

Message Passing Libraries

We have not had a planning session in this area. Current TLCC implementations and prior use at the labs is going to push the priority of a SLURM/Open MPI integration project. Currently LLNL has focused on MVAPICH and LANL/SNL on Open MPI as their general purpose MPI implementation. All agree however to use the other MPI as a secondary MPI for problem resolution issues. We expect to further detail plans in this area this year.

Performance Analysis Tools

We have not had a formal planning session in this area but have all been closely involved through the Open|SpeedShop projects and have collaborated on a publication, tutorials and various workshops.[12] Much effort has been focused on establishing the O|SS framework as the base infrastructure for performance analysis needs. This is close to being a production-established framework. Efforts are in place to add additional plug-ins; Javelina, mpiP (via Tripod) and develop proposals for Office of Science submittal for memory tool plug-ins. A concern that has been surfaced is the ensured sustainability of

[12] M. Schulz, J. Galarowicz, D. Maghrak, W. Hachfeld, D. Montoya, S. Cranford, "Open|SpeedShop: An Open Source Infrastructure for Parallel Performance Analysis", Special Issue of Scientific Programming on Large-Scale Programming Tools and Environments, 2008, to appear.

the tri-lab tool capability that we have established with Krell Institute through continued funding.

Interfaces to systems: Visualization, HPSS, filesystems

This is a new area of investigation for this working group. We have not had a chance to digest what this means or what can be done in these areas. More to come.

Working Group Management/Oversight for Open|SpeedShop

This is a shared project across the labs and Krell Institute. The focus is to further establish as a Tri-lab model for a development/implementation project with shared development resources and collaboration on extended tool additions (plug-ins). This is targeted to be a production-ready framework this year. Efforts (Tripod project) are in place to add additional plug-ins; Javelina, mpiP, OTF-based I/O tracing. We plan on coordinating other project additions to the framework for FY09. It has been instrumental in creating a shared tool development/maintenance capability through Krell Institute that is cost beneficial to the labs that we need to ensure is sustainable. Models for continuation are being discussed as part of the Sustainability project below and we are requesting Tripod support.

Open Source Tools Capability Sustainability

This is a key area for focus for FY09. Through the past year discussions in HPC Tools workshops have included consolidation of both infrastructure frameworks and models for continued sustainability for open source tool providers. Issues have ranged from how to place prototype tools such as STAT into more of a production maintenance organization and broader issues such as the maintenance model for tools such as Open|SpeedShop and Open MPI. This project expects to propose models for production maintenance that would include increased community support and an external responsible agent (such as Krell Institute) with the goal of developing an external partner that can provide a tool development and maintenance capability that can also grow with community funding. As part of our planning efforts, we are looking at the possibility of establishing a model that includes service maintenance contracts to cover specific levels of support and training. In the near term we are requesting Tripod funding support until the model is sustainable.

Resource Management Working Group

Progress towards our original goal of fostering tri-lab collaboration and uniformity in Moab / resource management has been strong. LLNL and SNL staff met at the Cluster Resources Inc. (CRI) Moab Conference at the end of May.

We have written tutorials on a number of Moab-related topics and have shared these with the tri-labs: https://computing.llnl.gov/jobs/moab

As each lab submits bug reports / questions to CRI, we continue to stay informed as to the progress and issues each lab is having. We also chime in frequently to help each other address the submitted issues.

We had another tri-lab meeting at SNL on August 20th to get an update on our collective progress towards deploying Moab.

HPC Monitoring & Metrics Working Group

The primary work of the Monitoring and Metrics Working group in FY08 was to oversee the work of the Application Monitoring project. Their work is documented above.

Work on system monitoring is expected to resume with the roll-out of TLCC.

Tri-lab Environment Working Group

The tri-lab environment working group was restructured in FY08 to focus on configuration management; license sharing & management; tri-lab WAN issues; and multi-domain security issues. The working group is also responsible for oversight and guidance of the Shared Work Space Project.