



**Completion Report for
MRT 3243**
TLCC Clusters in Production at LLNL

**Milestone Completion
Date of Report**

**15 October 2008
7 January 2009**

**LLNL-TR-409843
K. Cupps
Lawrence Livermore
National Laboratory**

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

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

Executive Summary

This report provides documentation and evidence for the completion of the deployment of LLNL's Tri-Lab Capacity Computing (TLCC) clusters *juno* and *eos* on the classified network at LLNL. The deployment of these clusters is an L2 Milestone in the ASC FY09-10 Implementation Plan due at the end of Quarter 1 in FY09.

Milestone: MRT 3243

Title: TLCC Clusters in Production at LLNL

Category: Advanced Simulation and Computing

ASC Program Element: CSSE, FOUS

The milestone definition is:

This milestone is the culmination of the TLCC platform procurement and will be satisfied when all classified scalable units are functioning properly on the classified network and science codes are running at scale on the classified side. This is the progression past the acceptance phase of the hardware and software stack on the unclassified side and encompasses running NIF and SSP science at scale on the classified network. Both the 8 Scalable Unit Juno system for weapons applications and the 2 Scalable Unit Eos system for NIF applications are part of this Milestone.

This goal of the TLCC procurement was to build a common capacity hardware environment across the three NNSA laboratories, Livermore, Los Alamos, and Sandia and achieve a reduced Total Cost of Ownership (TCO) using a common approach. The Tripod Operating System Software (TOSS) is the tri-lab software environment which runs across all newly procured TLCC machines. Using a common software stack on common hardware reduces system administration and debugging costs across the three Labs as well as increases Tri-Lab application portability. In fact, the TLCC/TOSS strategy has been leveraged by all the intended Scalable Unit (SU) recipients (LLNL, LANL, SNL-NM, SNL-CA) as well as by the Kansas City Plant and a SCIF at LLNL. What started out as a contract for 21 SUs at four sites became thirty-one scalable units for six separate customers. This Milestone was completed on schedule; however the original deployment schedule for the TLCC scalable units was delayed due to issues in vendor hardware. This report outlines the hardware issues that caused initial delays as well as the final deployment schedule of the hardware.

TLCC Hardware Issues and Deployment Schedule

The contract for the TLCC clusters was awarded to Appro on 17 September 2007. The TLCC SU deliveries to the three Labs were originally scheduled to occur between December 15th and early June of 2008. Two major hardware issues came up that significantly affected the original schedule. 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, and are 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 1).

TOSS/TLCC Build and Delivery Summary

As a result of the two problems discussed above, the hardware delivery schedules slipped significantly. Table 1 provides the date that the clusters were delivered to the respective Labs, the date the re-work was done and for LLNL clusters only, the date the cluster became “Generally Available” to users. The first four clusters in Table 1 were delivered before the VDDIO issue was found. The last five clusters (in gray in Table 1) were not delivered until the VDDIO work had been completed at the integrator site.

Cluster	Site	Scalable Units	Delivery Date	VDDIO Rework Date Complete	Generally Available
Hype	LLNL	1	31 Mar 2008	9 Aug 2008	14 August 2008
Lobo	LANL	2	21 May 2008	13 Aug 2008	
Unity	SNL-NM	2	19 May 2008	15 Aug 2008	
Juno	LLNL	8	¹ 5 May 2008	9 Aug 2008	15 October 2008
Whitney	SNL-CA	2	20 Aug 2008	6 Aug 2008	
Eos	LLNL	2	20 Aug 2008	6 Aug 2008	17 September 2008
Hurricane	LANL	2	25 Aug 2008	6 Aug 2008	
Glory	SNL-NM	2	mid-Sep 2008	Factory Parts	
Hera	LLNL	6	mid-Sep 2008	Factory Parts	26 November 2008

Table 1. TLCC Cluster Delivery, Re-work and GA Dates

The Eos cluster was delivered at LLNL on August 20th and was immediately integrated on the classified network. By September 17th, it was made available to all NIF users. Juno was delivered and integrated on the unclassified side and experienced a large lag from delivery to final classified integration due to the VDDIO troubleshooting. Once Juno had its VDDIO re-work completed the integration and testing was completed on the unclassified side in less than a month. Juno moved to the classified network on September 16th and a series of user runs began. The NIF team ran a very successful laser plasma simulation for four days continuously on 1024 nodes of Juno using 16 tasks per node. The email below is from Denise Hinkel, the NIF laser scientist who ran the simulation. Her job was killed at 8AM on October 15, 2008 so that Juno could be made generally available to all ASC users.

```
Return-Path: <hinkel1@llnl.gov>
Received: from mail-2.llnl.gov ([unix socket]) by mail-2.llnl.gov (Cyrus
v2.2.12) with LMTPA; Tue, 14 Oct 2008 17:06:00 -0700
Received: from nspiron-1.llnl.gov (nspiron-1.llnl.gov [128.115.41.81]) by
mail-2.llnl.gov (8.13.1/8.12.3/LLNL evision: 1.7 $) with ESMTP id
m9F05tep003244 for <futral2@mail.llnl.gov>; Tue, 14 Oct 2008 17:06:00 -0700
X-Attachments: None
X-IronPort-AV: E=McAfee;i="5300,2777,5405"; a="29101026"
X-IronPort-AV: E=Sophos;i="4.33,411,1220252400"; d="scan'208";a="29101026"
Received: from sandridge.llnl.gov (HELO [128.115.43.17]) ([128.115.43.17])
by nspiron-1.llnl.gov with ESMTP; 14 Oct 2008 17:05:57 -0700
Mime-Version: 1.0
Message-Id: <p06240805c51ae3881f92@[128.115.43.17]>
Date: Tue, 14 Oct 2008 17:07:10 -0700
To: "Winfield S. Futral" <futral2@llnl.gov>
From: Denise Hinkel <hinkel1@llnl.gov>
Subject: Completion of our Juno run
Cc: still1@llnl.gov
Content-Type: text/plain; charset="us-ascii" ; format="flowed"
Content-Transfer-Encoding: 7bit
```

Hi Scott,

¹ First four SUs, four more SUs delivered 3 June 2008.

We will shut the job down tomorrow morning by 8 a.m. Many thanks to you, Adam Moody, and the many others who have made this run so successful.

Ideally, we need one weekend to complete this simulation. (Can you believe that it has now been running for 81 hours?). Also, ideally, we would like that weekend to be as soon as possible, because (1) Juno is right now working beautifully for us ; and (2) we need to complete post-processing based on one more weekend of running.

So -- is there any hope of getting a DAT on Juno this weekend? If so, how do I request it? If not, when do you foresee our getting such a DAT?

Thank you for all of your help in making this simulation happen.

Regards,
Denise

Figure 1. E-mail of NIF simulation success as precursor to Juno GA

Message-ID: <4935BC18.2050703@llnl.gov>
Date: Tue, 02 Dec 2008 14:52:08 -0800
From: LC Hotline <lc-hotline@llnl.gov>
User-Agent: Thunderbird 2.0.0.18 (Macintosh/20081105)
MIME-Version: 1.0
To: ocf-status@llnl.gov
Subject: Hera system is now GA
Content-Type: text/plain; charset=ISO-8859-1; format=flowed
Content-Transfer-Encoding: 7bit

Attention OCF Users:

LC is happy to announce that on Wednesday, 11/26, the Hera machine officially became "Generally Available", or "GA". Hera is a large capacity OCF computing resource with 864 quad socket/quad core nodes shared by M&IC and ASC, and is targeted to run small to moderate parallel jobs.

Information about the Hera system can be found in the alphabetical list of OCF systems at:
https://computing.llnl.gov/?set=resources&page=OCF_resources

WCI users have been given accounts on the system, as have the Alliance users from ALC and the Institutional users (Grand Challenge and LDRD) from Thunder. As always, Foreign National users must have their Security Plan's updated to reflect Hera before access can be granted.

--

If you have any questions, please contact the LC Hotline.

Timothy Fahey
Livermore Computing Hotline Lawrence Livermore National Laboratory
Phone: 925-422-4531 Customer Services Group
FAX: 925-422-0592 P.O. Box 808 L-63
Email: lc-hotline@llnl.gov Livermore, CA 94551

Figure 2. Email announcing Hera General Availability