Cray XT3/XT4 Software:

Status and Plans

David Wallace
Cray Inc

ABSTRACT: : *This presentation will discuss the current status of software and development plans for the CRAY XT3 and XT4 systems. A review of major milestones and accomplishments over the past year will be presented.*

KEYWORDS: 'Cray XT3', 'Cray XT4', software, CNL, 'compute node OS', Catamount/Qk

1. Introduction

The theme for CUG 2007 is "*New Frontiers*," which reflects upon how the many improvements in High Performance Computing have significantly facilitated advances in Technology and Engineering. UNICOS/lc is moving to new frontiers as well. The first section of this paper will provide a perspective of the major software milestones and accomplishments over the past twelve months. The second section will discuss the current status of the software with respect to development, releases and support. The paper will conclude with a view of future software plans.

2. Major Accomplishments

UNICOS/lc software has made significant progress over the last year. And while it still have a ways to go, customers and users are getting excellent results from the Cray XT3/XT4 hardware and software. During the first half of 2006, the primary focus was completing the development and release for UNICOS/lc version 1.4. This release added support for dual core AMD Opteron

processors, support for larger system configurations (five rows and larger), a new version of SuSE Linux and CFS Lustre as well as a host of other new features and fixes to software deficiencies. Much has been done in terms of new software releases and adding support for new hardware.

2.1 UNICOS/lc Releases

The last twelve months have been very busy with respect to software releases. Two major releases (UNICOS/lc 1.4 and 1.5) and almost weekly updates have provided a steady stream of new features, enhancements and fixes to our customers. UNICOS/lc 1.4 was released for general availability in June 2006. A total of fourteen minor releases have been released since last June. The current version of 1.4 is 1.4.50. This is likely to be the last update to 1.4.

UNICOS/lc version 1.5 was released for limited availability in August 2006. A number of lessons were learned with version 1.4 and consequently, extensive testing on a large Cray XT3 system was commenced shortly after the 1.5 Limited Availability release. A new general availability release criteria was added to version 1.5 which required a customer to be running the 1.5

release in production on a large system. Oak Ridge National Laboratory (ORNL) graciously agreed to participate in the 1.5 Customer Test Program, offering multiple weekends of dedicated test time to test UNICOS/lc 1.5. The release criteria were completed and the ORNL system Jaguar began running 1.5 in production in November 2006. UNICOS/lc version 1.5 was released for general availability in December 2006. Since the release of 1.5, a total of seven minor updates have been released. The current 1.5 version is 1.5.45.

Figure 1 Oak Ridge Cray XT4 (courtesy of Oak Ridge National Laboratory)

2.2 New Product Support

Support for the Cray XT4 (see Figure 1) was the major theme for UNICOS/lc 1.5. Three Cray XT4 systems have been installed with a combined total of 181 cabinets. The 1.5 release also added support for dual core processors on SIO nodes. This 1.4 feature was delayed for a variety of reasons. Once the feature was completed, customers were required to upgrade to UNICOS/lc 1.5. An incompatibility between the Cray SeaStar and the AMD dual core Opteron on SIO nodes resulted in Cray implementing a replacement program for customers with dual core SIO nodes and Cray SeaStar 1.2s.

3. Current Development Activities

3.1 UNICOS/lc 2.0

Software development for the Cray XT3 and Cray XT4 is currently focused on the next release of UNICOS/lc. Version 2.0 will introduce a new compute node operating (Compute None Linux) in addition to supporting Catamount/Qk. Compute Node Linux (CNL) is the next generation of lightweight kernels for compute nodes on the CRAY XT3 and Cray XT4 computer systems. The CNL operating system provides a runtime environment based on the SUSE SLES distribution and the SuSE Linux kernel. Modifications are being made to the system software to match the legacy performance and scaling characteristics of Catamount/QK as well as address new marketing and customer requirements.

Major enhancements to UNICOS/lc version 2.0 will include new versions of PBS Pro (version 8.1) from Altair Computing, Lustre file system (version 1.4.9) from Cluster File Systems, and a Linux kernel (version 2.6.16) on the Service nodes. Linux 2.6.16 will also be used for CNL. The Linux kernel upgrade is from the SuSE SLES 10 software distribution. Only the kernel is being upgraded as part of UNICOS/lc 2.0. An upgrade to the full SuSE SLES 10 distribution is planned for a future release.
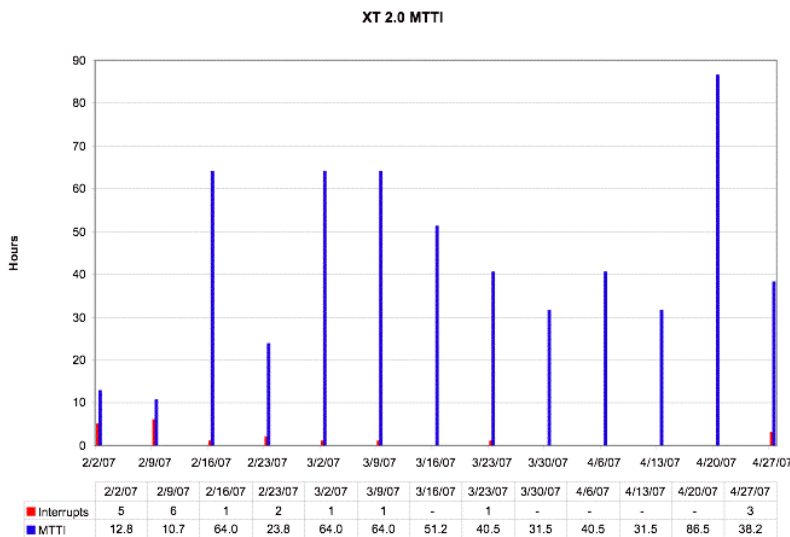
Software development continues to improve the software installation and upgrade tools. New features have been added that will greatly reduce the amount of dedicated time required to upgrade the system software. Beginning with version 2.0, system administrators may now install released software on an alternate boot root while the system is in operation. Support for maintaining multiple versions of the system software has also been added. A 'system set' is created by the system administrator which contains all of the files and boot images required to boot the system. Each system set can be labelled to aid in identification. Tools are provided that facilitate easy switching being different major and minor versions of UNICOS/lc installed in the system sets.

With each major release, Software Division is elevating the criteria for releasing the software. Below is a summary of the release criteria for the UNICOS/lc 2.0 release:

- Feature content is complete

- Reliability metrics have been met. The key metrics are shown below:

    • LA: Internal system SMTTI will be >40 hours, w/ no service node failures, passes 72 hour reliability run

    • GA: Passes 72 hour reliability run in production at a customer site with at least 42 cabinets

    • SPR

        ▪ LA: <10 Critical SPRs (relative to this release)

        ▪ GA: No Critical or Urgent (relative to this release)

- Test plan has been executed and received a passing score

- Performance metrics have been met. Performance targets have been established for Lustre file I/O, latency and bandwidth, memory used by the compute node OS, and other aspects of the system.

- Installability: this refers to a set of test installations that have been completed successfully using the documented procedures and tools

- Supportability: documentation and tools are available to Software Product Support.

Figure 2 shows the current system interrupt and MTTI metrics for the UNICOS/lc 2.0 system. This chart shows the aggregate values for the systems running UNICOS/lc 2.0 during shared batch hours.

XT 2.0 MTTI

	2/2/07	2/9/07	2/16/07	2/23/07	3/2/07	3/9/07	3/16/07	3/23/07	3/30/07	4/6/07	4/13/07	4/20/07	4/27/07
Interrupts	5	6	1	2	1	1	-	1	-	-	-	-	3
MTTI	12.8	10.7	64.0	23.8	64.0	64.0	51.2	40.5	31.5	40.5	31.5	86.5	38.2

Figure 2 UNICOS/lc 2.0 reliability metrics

3.2 UNICOS/lc 2.0 Release Schedule

UNCOS/lc 2.0 will be released in two phases. The first phase will be a limited availability (LA) release intended for customers with small systems or test systems. The second phase will be the general availability release aimed at supporting production workloads.

The two phase approach was adapted after the UNICOS/lc 1.4 release. The UNICOS/lc 1.4 release was thoroughly tested on the in-house systems and passed the release criteria. At the time, the largest in-house system was four cabinets. This version of UNICOS/lc was quite stable on small systems but serious problems related to scaling were encountered by customers with large systems. Steps were taken with the UNCOS/lc 1.5 release

to incorporate testing the LA release on a large system and addressing any problems found prior to releasing the software for general availability. The release criterion now stipulates that 1) the candidate release software must pass a 72-hour reliability run on the customer system, and 2) the candidate release software must be running in production on a customer system on at least forty-two cabinets.

The planned schedule for the release of UNICOS/lc 2.0 is shown below:

- UNICOS/lc 2.0 LA: 2Q07

- UNICOS/lc 2.0 GA: 4Q07

UNICOS/lc 2.0 with support for both Catamount/Qk and Compute Node Linux adds significantly to the logistics of supporting, testing and exposing multiple versions of the software on the limited number of internal platforms available to Software Development. Figure 3 shows matrix of the internal development, pre-production and production systems and the different versions of software that is installed on the systems.

	M	T	W	Th	F
Guppy	1.5	1.5	1.5	1.5	1.5
Perch	1.4	2.0Cat	2.0Cat	2.0	2.0
Seal	1.5	1.5	1.5	1.5	2.0Cat
Salmon	2.0	2.0	2.0	1.5	1.4
Bass (Dev)	Catamount	CNL	Catamount	CNL	Catamount
Shark			CNL		CNL

Figure 3 UNICOS/lc on internal Cray XT3/4 systems

Below is a color key for Figure 3:

- V1.4 (black)
- V1.5 (blue)
- V2.0 CNL (turquoise)
- V2.0 Catamount/Qk (orange)
- Catamount/Qk-development (purple)
- CNL-development (green)

As you can see from the figure, the available time is quite limited and in some cases, a specific version of the software may only run on a given system for just hours at a time as these systems are also used for dedicated usage (benchmarking, support and development).

4 New Frontiers

4.1 New Product Support

Beyond the UNICOS/lc 2.0 release, Software Development for the Cray XT3 and XT3 is pushing the frontiers of new capabilities, reliability, performance and scaling. On the horizon, support for the following hardware upgrades will be available in the near future:
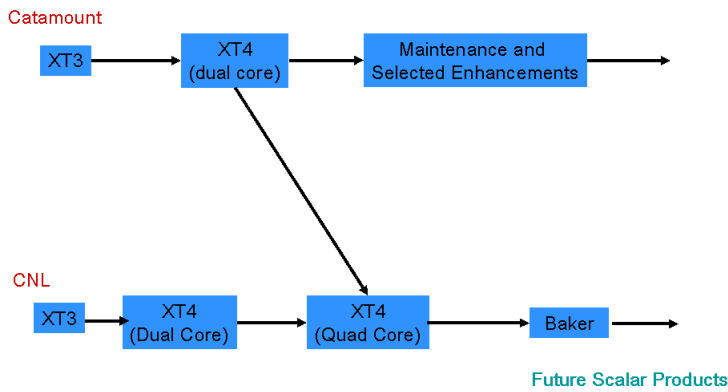
- Cray XT4 Quad Core
- XT4-FPGA
- PCI-Express

- Cray XT4 future upgrades

4.2 Cray XT4 Software Direction

The operating system for new Cray Scalar products will be based on a SuSe Linux kernel on both the Service and Compute nodes. With the introduction of the Cray XT4 with quad core processors, Compute Node Linux will be the only compute node OS supported. Figure 4 gives a view of the compute node OS support road map.

Catamount

| XT3 | → | XT4 (dual core) | → | Maintenance and Selected Enhancements | → |

CNL

| XT3 | → | XT4 (Dual Core) | → | XT4 (Quad Core) | → | Baker | → |

Future Scalar Products

Figure 4 Compute node OS support path

Customers planning to upgrade their systems to quad core processors will have to plan to convert from Catamount/Qk to CNL. Cray plans to provide aids to assist customers in migrating to CNL.