XT9? Integrating and Operating a Conjoined XT4+XT5 System

Don Maxwell, Josh Lothian, Richard Ray, Jason Hill,
and **David Dillow**, *ORNL*; and **Jeff Becklehimer** and
Cathy Willis, *Cray Inc*.

ABSTRACT: *The National Center for Computational Sciences at Oak Ridge National Laboratory recently acquired a Cray XT5 capable of more than a petaflop in sustained performance. The existing Cray XT4 has been connected to the XT5 to increase the system's computing power to a peak of 1.64 petaflops. The design and implementation of the conjoined system will be discussed. Topics will include networks, Lustre™, the Cray software stack, and scheduling with Moab™ and TORQUE.*

KEYWORDS: XT4, XT5, CNL, ALPS, Moab, TORQUE, Infiniband, External Login

1. Introduction

The Cray XT line of products has been the workhorse of the Leadership Computing Facility (LCF) operated by the National Center for Computational Sciences (NCCS) at the Oak Ridge National Laboratory (ORNL) since the formation of the facility in 2004. Upgrades to these machines have been implemented in steps bringing the current theoretical peak of the XTs operated by NCCS to 1.64 Petaflops. Currently, two large XTs provide compute cycles in support of the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program established and operated by the U.S. Department of Energy Office of Science. While each of these machines provides a very powerful resource to the INCITE program, the combination of the two provides a unique resource for capability codes within the Office of Science.

Combining two XT machines requires engineering on several different levels both in hardware and software. Not only do the two machines have to be physically connected via some network, other areas such as filesystems, scheduling, and external login services have to be considered. While many codes may never have the need or desire to span XTs due to communication issues inherent in connecting two large machines with commodity networking or simply due to architectural differences in processors and memory, a few codes are able to take advantage of this resource. The work necessary to combine the machines also offers users who do want to continue running on only one XT for individual jobs some advantages as well. External login nodes provide users with one-stop shopping for access to both machines' filesystems, batch interfaces and programming environments. In addition, an external login node can be configured with additional hardware to provide a much more capable platform for compiling, file transfers, and other activities needed to build and maintain production codes. Login nodes attached directly to the XT suffer from processor and memory limitations and also have the added disadvantage of being unavailable when the XT itself is unavailable.

2. Hardware

2.1 Cray XTs

The subject XTs being combined are a Cray XT4 and a Cray XT5 named Jaguar. Jaguar XT4 has been in production in its current form since May 2008 when it was upgraded to AMD Opteron 2.1 GHz quad-core processors. Jaguar XT5 was the first Cray XT to break the petaflop barrier in both sustained and theoretical peak performance. It is currently undergoing a transition to operations period that not only provides several projects selected by the Office of Science with the ability to scale their codes to the size of the machine but also allows the machine to stabilize as hardware and software issues

unique to a machine of this size are discovered and fixed. The table below provides a comparison of the two machines.

	Jaguar XT5	Jaguar XT4
Cabinets	200	84
Processors	AMD Opteron 2.3 GHz quad-core	AMD Opteron 2.1 GHz quad-core
Compute Cores	149,504	31,328
Memory (TB)	300	62
Links	115,200	48,384
Theoretical Peak Performance (TFLOPS/s)	1,375	263
I/O Capacity (TB)	4,100*	700
I/O Bandwidth (GB/s)	100*	40
Service Nodes	256	116

* The current filesystem on Jaguar XT5 is an Infiniband direct-attached configuration using roughly half of the available storage capacity available. The other half is being used for development of a Lustre routed filesystem called Spider [1]. The two halves will be merged into a Spider configuration which will be mounted center wide during the next few months.

2.2 External Login Nodes

While the specifications for the final production nodes have not been finalized at this time, the current prototype nodes being used for proof of concept have the following configuration:

Processors	Quad socket AMD Opteron 2.0 GHz quad-core
Memory (GB)	32

External connectivity is provided by a 1Gb Ethernet device and filesystem connectivity which will be discussed later is provided by Infiniband.

As mentioned previously, one of the goals of the external login nodes is to provide a much more capable platform for software development than the current service nodes inside the Cray XT can provide, so the final specification of these nodes will attempt to meet that goal.

2.3 Networks

Providing connectivity from the compute nodes on one XT to the compute nodes on another XT obviously requires an intermediate network. NCCS has chosen Infiniband as the network to provide access to the center wide Spider filesystem, so it was the obvious choice to provide connectivity between the XTs as well. The Infiniband network dubbed SION (Scalable I/O Network) is comprised of two core Cisco DDR switches connected by a Cisco DDR aggregation switch. Another Cisco DDR switch provides connectivity on a different floor for the external login nodes for a combined total of approximately 1,100 ports for all four switches.

Below is a diagram of the Infiniband network connectivity.

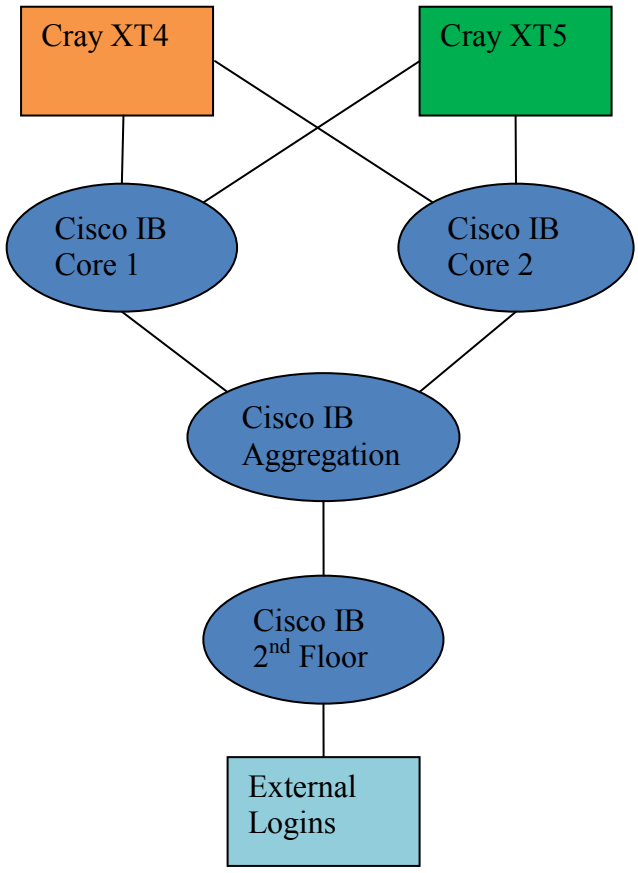

Figure 1. Infiniband Network

The XT compute nodes are connected to service nodes inside the machine via the Cray Seastar network. The service nodes provide the Infiniband connectivity with Jaguar XT4 containing 48 Infiniband cards and Jaguar XT5 containing 192 Infiniband cards. Routing between the machines requires some algorithm for determining which Infiniband nodes will route to each other as well as which Inifiniband router each compute node will use. With the disparity in the number of

Infiniband nodes on each machine, the simple choice to make is a one-to-one mapping of the 48 Infiniband nodes on Jaguar XT4 to 48 Infiniband nodes on Jaguar XT5. That is the current strategy being used to route between the machines. Based on the fact that there was a choice of 48 routers among 192 inside Jaguar XT5, routers were chosen based on torus location inside the Seastar network to attempt to spread the load as much as possible inside that machine. Furthermore, routers between machines were chosen based on location in the Infiniband network. Referring to Figure 1, a Jaguar XT4 router attached to Cisco IB Core 1 always routes to a Jaguar XT5 router on Cisco IB Core 1 and the same holds true for Cisco IB Core 2. Another interesting experiment would be a many-to-one mapping utilizing all of the Infiniband nodes on Jaguar XT5. The obvious question is where are the bottlenecks and does the oversubscription of the Infiniband nodes on Jaguar XT4 do more harm than good. Some work has already been done to characterize the bottlenecks in the network. Those results will be presented in another paper [1] at this conference.

Each compute node also has to determine which Infiniband router it will use to get to the other machine. A simple calculation is used to determine the distance inside the torus to each router to provide the optimal router for that node. If a predetermined maximum number of clients is already using that router, the router with the next shortest distance to the node will be used. This is a static calculation stored in a file that is then read at boot time by the compute nodes and used to setup the routes to the compute nodes on the other XT.

2. Software

2.1 Lustre Filesystems

Combining two machines presents several challenges particularly when each machine has local filesystems that are critical to job execution. Both Jaguar XT4 and Jaguar XT5 currently have local Lustre filesystems that must be presented to users on external login nodes. This presents problems for the external login nodes since the inherent instability of supercomputers means that filesystems are disappearing more often than desired. The center wide Spider filesystem will provide the external login nodes with a much more stable filesystem since it will be served outside of the XTs. Program development can continue on external login nodes with a center wide filesystem when XTs are down. However, the current environment dictates that the local Lustre filesystems must be accommodated, so using the Lustre LNET routing code, the filesystems are being presented on the external login nodes via SION.

Users are provided with a Lustre working area on each XT at /tmp/work/$USER. This is a link that points to an area in a variety of Lustre filesystems. The link provides a single point of reference for the user's working area regardless of how filesystems might change, users might get moved around, etc. No scripts should need to be changed if the link is used.

With multiple XT filesystems now available on the external login nodes, the user needs a mechanism to choose the filesystem pointed to by /tmp/work/$USER. The Modules package provides this functionality based on user selection of module xt4 or module xt5. There is further discussion of the xt modules in a later section. Creating the link, however, becomes an issue when the target filesystem is unavailable. Timeout logic was added to the modules to prevent user hangs and to provide a user-friendly error message.

Filesystem hangs based on XT availability have been a major issue during the development of the external login nodes. Due to the fact that hangs can be experienced by users when a XT disappears taking its filesystem down, a script was developed to run on each external login node to monitor the filesystems. The script attempts to ping each filesystem's metadata server, and if unsuccessful, attempts to umount the filesystem. Upon a successful ping, the filesystem is remounted if not already mounted. Several issues have arisen during the development of this script. While a forced unmount of the filesystems prevents hangs for the users, it causes issues with the mounted filesystems table (/etc/mtab). Filesystems remain in /etc/mtab so that a subsequent mount fails. Automounted NFS filesystems have experienced issues related to locking /etc/mtab when attempting to manipulate the file to fix the forced umount issue. While most of the XT reboots find that the monitoring script is working, edge cases mean that this development is ongoing. As stated earlier, a more stable filesystem that doesn't rely on XT servers will alleviate this issue.

2.2 Batch Scheduling

Scheduling jobs is one of the key factors in providing users with the capability of accessing resources both on independent XTs and potentially across XTs. NCCS developed a partnership with Cluster Resources Inc.™ (CRI) late in 2005 that led to the porting of the Moab Workload Manager® to the Cray XT platform. Many of the reasons for choosing Moab™ are discussed in an earlier CUG paper [2]. Suffice it to say that the rich features and flexibility of the scheduler have not only served existing workloads well but also given NCCS the flexibility to add new capabilities in a short amount of time.

The development of a unified scheduler with knowledge of two independent XTs required some development within both Moab and TORQUE which is an open source resource management package maintained in part by CRI.

Two paths were possible when considering the correct course to take in development of this new feature – modifying the native XT resource manager or using the Moab Workload Manager for Grids model. After a requirements discussion with CRI, the decision was made to proceed with modifications to the native XT resource manager.

This model requires two resource managers to be defined inside Moab pointing to each XT. Given the fact that the Cray Application Level Placement Scheduler (ALPS) is required to launch jobs on a XT and that is has no knowledge of multiple XTs, the model seems to fit the current architecture. TORQUE is an integral part of the interaction with ALPS and is therefore also unique to each XT. This means that the TORQUE suite of tools continue to see independent XTs. The particular TORQUE configuration seen by the end user from the external login nodes is controlled via the xt modules which will be discussed in more detail in the next section.

While both TORQUE and ALPS maintain their view of independent XTs, Moab sees both XTs as a single resource with the capability of controlling resources based on both a global view and an independent view. Many policies inside Moab can be controlled using partitions, and each XT is defined as a partition in this configuration. Partitions can support independent users, priorities, and job templates just to name a few of the parameters that have been explored at NCCS. Using these capabilities, jobs can be directed to different resources based on size, user, etc. Conversely, the entire resource consisting of multiple partitions can be treated as a single entity when configuring these same parameters. This provides the flexibility to pick and choose parameters based on policies important to a computing center. Since Moab is multi-partition aware, only one instance of the daemon needs to be running, and since it is scheduling both XTs, it would stand to reason that it should not run on either XT. Instead, an independent server should be selected to allow scheduling to continue independent of either XT being unavailable. This does create an issue since Moab, like TORQUE, needs to interact with ALPS. To solve this problem, an account was setup to provide passwordless ssh access to each XT, and with the help of sudo, given the permissions needed to interact with ALPS on the XTs.

While the traditional job launch mechanism on the XTs at NCCS has been provided with the qsub command provided by TORQUE, msub provided with Moab not only provides the same capability but in addition gives the user the flexibility of not specifying a particular resource as the target. With msub, users can specify a particular target partition but with no specification, a resource is chosen for the user based on a load balancing algorithm. Currently, a very simple algorithm is being used based purely on availability of resources at the time of job launch, but CRI is open to expanding that algorithm to include other factors. It might be desirable, for instance, to consider other policies such as queue depth, historical utilization, or other factors when selecting a resource. It could also be beneficial to delay resource selection until runtime providing the best choice for the current workload. These areas will be explored as the rollout of the external login nodes progresses.

Another area that may require changes in the future is priorities. With the native XT resource manager model, priorities for individual partitions are implemented using fairshare policies. Fairshare by definition has a dynamic factor associated with its calculation based on historical usage. Given that NCCS has traditionally setup priorities based on units of days, a dynamic aspect in the priority may lead to some confusion. If NCCS policy dictates that priorities are to be global in nature, the problem becomes moot since the fairshare implementation is only needed for different priorities in different partitions. However, if not, the Moab Workload Manager for Grids model may need to be revisited.

2.3 Cray XT Software

Providing the Cray XT software environment on the external login nodes is also critical for making the nodes a productive resource to users. Compilers, debuggers, libraries and all the tools necessary for building codes must be available on the external login nodes just as they are on the XT login nodes. While it would be possible to allow the versions of software on each XT to be independent, it seems much more manageable to keep them in step with each other. After all, a job that could potentially be submitted to either XT will need to have access to the same software on each machine to run correctly. With this in mind, a decision was made early in the design of the external login nodes to create a common default environment. Regardless of which machine is being targeted from the external login nodes, the default software stack is the same. This does not mean that the XTs themselves necessarily have to stay in sync with each other – they just have to be able to support the common environment defined on the external login nodes. In other words, new modules could be installed on either XT and the external login nodes and would therefore be available to users, but they would not be in the default environment until both XTs were capable of supporting the software.

The concept of a sharedroot much like the XT sharedroot supports the XT software stack. All external login nodes mount a common NFS area where the XT software stack is installed. A separate RPM database provides the capability of installing the XT rpm packages on only one machine. Through a series of links that point to the XT sharedroot, the software appears on each external login node as it does on the XTs themselves. This configuration eases the management burden of maintaining the XT software stack on multiple external login nodes.

As referenced above, establishing an environment for a particular XT on the external login nodes is necessary

for filesystems and software that are dependent on each individual XT. An xt module customized for each XT sets up the environment necessary to interact with Cray systems software such as apstat needed to query ALPS application status, xtnodestat to determine the layout of jobs based on cabinets, TORQUE tools, etc. These utilities have no knowledge of other XTs and therefore must be directed to an individual XT. Users are provided passwordless ssh access into the XTs once authenticated on the external login nodes, so executing these XT commands is simply a matter of establishing an ssh session to the correct XT and issuing the command remotely. This is all done via a wrapper for common XT commands used by end users.

2.4 Jobs Spanning XTs

As stated in the introduction, many codes will never have the need or desire to span XTs, but some codes with well-defined communication patterns can take advantage of combined XTs. While Moab does have the capability of spanning partitions, neither ALPS nor Cray's MPI currently supports access to multiple XTs. Therefore, a job that spans XTs today must establish jobs on each side and then use some other method outside the current Cray software stack to communicate among compute nodes. Both Cray and NCCS have been working on MPI implementations to support this capability. NCCS has been working with the Open MPI Project [3] to provide this functionality.

Conclusion

Real benefits can be provided to users when combining the resources of both the Cray XT4 and Cray XT5 into a common entity accessible from a single source. External login nodes provide a platform for interaction with the XT machines both individually and as one. They also provide a much more capable platform for development of codes and a more stable environment outside of the XTs. Users have experienced compilation bottlenecks on the existing XT login nodes. Taking advantage of these new resources should provide users with a more productive resource for accomplishing the mission of leadership computing within NCCS.

Acknowledgments

This work was sponsored by the U.S. Department of Energy's Office of Advanced Scientific Computing Research and performed at the Oak Ridge National Laboratory, managed by UT-Battelle, LLC under contract number DE-AC05-00OR22725.

About the Authors

Don Maxwell is a Senior System Administrator at Oak Ridge National Laboratory primarily focused on the Cray XT series. He has been a key member of past teams in bringing up new supercomputers for the NCCS. He can be reached at maxwellde@ornl.gov.

Josh Lothian is a Senior System Administrator at Oak Ridge National Laboratory working as a member of the team responsible for the NCCS infrastructure and networking. He can be reached at lothian@ornl.gov.

Richard Ray is a System Administrator at Oak Ridge National Laboratory working as a member of the team responsible for the NCCS infrastructure. He can be reached at raray@ornl.gov.

Jason Hill is a System Administrator at Oak Ridge National Laboratory working primarily with Lustre and HPSS™. He is a member of the team that is bringing up the ORNL center wide filesystem "Spider". He can be reached at hilljj@ornl.gov.

David Dillow is a Linux Developer at Oak Ridge National Laboratory. As a member of the Technology Integration Group, he evaluates and develops new technology to meet the needs of the NCCS. He is currently focused on I/O. He can be reached at dillowda@ornl.gov.

Jeff Becklehimer is a Principal Engineer with Cray Inc. He resides at Oak Ridge National Laboratory and is involved in all aspects of the Cray XT computers. He can be reached at jlbeck@cray.com.

Cathy Willis is a Systems Engineer IV with Cray Inc. She is assigned as a site analyst at Oak Ridge National Laboratory and is involved in all aspects of support of the Cray XT computers. She can be reached at willis@cray.com.

References

[1] Shipman, G. M., et al. "The Spider Center Wide Filesystem; From Concept to Reality", *Proceedings of CUG 2009*, Atlanta, May 2009.

[2] Maxwell, D., Jackson, M., et al. "Moab Workload Manager on Cray XT3", *Proceedings of CUG 2006*, Lugano, May 2006.

[3] “Open MPI: Open Source High Performance Computing", http://www.open-mpi.org/.