Cray Operating Systems Road Map

Charlie Carroll, *Cray Inc.*

ABSTRACT:. *This paper discusses the Cray Operating Systems road map. In addition, the rationale for coming changes is discussed.*

KEYWORDS: Operating systems, releases

1. Introduction

The Cray Software Operating Systems and I/O (OSIO) group provides key infrastructure and service components of the software stack.

These components include:

- Compute node kernels
  - XT CNL
  - XT Catamount
  - X2 CNl
  - XMT
- Service node kernel
  - Supports all compute node types
- File systems
  - Lustre
  - DVS (Data Virtualization Service)
- Networking
  - Portals
  - TCP/IP
  - uGNI and DMAPP
- Operating system services
  - Checkpoint / restart
  - Node health daemon
  - CSA (Comprehensive System Accounting)
- System management
  - Interface to system data
  - ALPS (Application Level Placement Scheduler)
    - Interfaces to PBS Pro, Moab/Torque and LSF
  - Command interface

This paper discusses the main themes to be emphasized in upcoming OSIO software releases, followed by specific features to be delivered in these releases.

2. Release Themes

Upcoming OSIO releases will emphasize certain broad themes. Before getting into specifics, we will take a look at the big picture.

2.1. System stability

Stability and robustness are important to any customer system. They are especially important in large supercomputers with millions of separate components.

Failover has been added to Lustre servers. This will allow a Cray system to ride through the failure of an individual Lustre OSS.

Portals and Lustre are two areas responsible for a sizable percentage of current system failures. To improve reliability in these areas, OSIO is: fixing bugs; increasing our test coverage; improving our ability to make a small test system look “big” through clever programming; reviewing codes and designs more thoroughly; and in some cases, rewriting sections of code.

A node health checker has been added. The fundamental idea is to verify that a node has the basic services needed to run jobs. The node health checker runs at various times—for example, after a job exits with a non-zero return code, or after all jobs, or after system reboot.

2.2. Performance

Cray's Compute Node Linux (CNL) implementation performs extremely well, largely because we have limited what services and features run on compute nodes. Over time, customers have asked for additional features to be added. We will do so only if we can see a clear benefit for the performance penalty incurred. If we're not vigilant, we will break the camel's back one straw at a time.

CLE 2.2 will contain a job-placement feature which will tend to place applications on nodes which are "close" together, which lowers the communication time. On dedicated systems, we've seen 10% or more improvement in application running time.

2.3. System Management

With CLE 2.2 and SMW 4.0, Cray is introducing unified interface for accessing system data. ALPS, the event subsystem, login nodes and system functions will send data to a centralized data store, which can be accessed by administrators with a variety of tools.

Scaling to large systems with hundreds of thousands of cores will stress existing technology in many ways, system management in particular. Cray will, as much as possible, hide scale issues behind the system management API through the clever use of caches and other data handling / reduction techniques.

2.4. Hardware Support

Part of OSIO's mission is to support new Cray hardware as it becomes available. Shanghai and Istanbul processors will be introduced in 2009. As much as possible, OSIO will introduce software changes in advance of hardware availability. Done this way, the software stabilizes early and the transition to new hardware is smoothed.

2.5. Lustre

The Lustre file system is a major component of Cray's software stack. We are working with Sun, the primary developers of Lustre, to improve both its functionality and robustness.

Testing a large, distributed piece of software, particularly one which can only be stressed at large scale, requires careful thought. Cray is working with Sun to design test plans and paradigms to ensure we deliver a quality product.

Cray is taking a more considered approach toward new Lustre features. Each one needs to be evaluated to compare the benefits against the costs and risks. We may decide not to release or support certain features deemed not robust enough to meet the scale of Cray's large systems.

2.6. Internal Infrastructure

OSIO is working to improve its internal infrastructure for building, testing and delivering software. While the direct effects will be mostly transparent to our customers, they will see improved robustness and reliability. In addition, the mechanics of certain operations such as delivering source code will become smoother.

3. Upcoming Releases

Cray released CLE 2.2 (code-named Congo) in 2009. Cray will release Danube in 2010. Danube will support only Gemini-based systems. (Gemini is Cray's new network card.)

3.1. Congo Features

The Node Health Checker (NHC) will increase the number of error conditions which can be detected. As discussed in section 2.1, the goal of node health is to ensure that all jobs start on healthy nodes, that is, nodes able to provide all the services expected of them. In addition, administrator configurability will allow the control of when the node health checker is run and how errors are handled.

Congo (with the corresponding SMW 4.0 release) will see the first phase of features supporting the single administrative API. This includes log consolidation—along with features to improve log performance—and a database with system information.

DVS (Data Virtualization Service) will be supported . DVS projects any VFS-based, POSIX-compliant file system to the compute nodes. New features in Congo are load balancing (useful for projecting a read-only file system from multiple servers) and cluster parallel mode.

Checkpoint / restart will be officially supported in Congo. This feature is one more tool to make Cray XT systems more robust.

Congo will contain a job-placement feature which will tend to place applications on nodes which are "close" together, which lowers the communication time. On

dedicated systems, we've seen 10% or more improvement in application running time.

External services are increasingly supported in Congo. External Lustre servers are the norm for most new systems. Some customers are working with Cray Custom Engineering to install external login nodes.

Lustre failover will be supported in CLE 2.2. See Nic Henke's paper at this conference for a more complete description of this feature.

3.2. Features Being Discussed

There are a number of features which will possibly / likely make it into future releases. However, the details are too murky at this point to say in which release they might appear or to give details about the implementation. Because of the significance of these features, we discuss them here to prepare customers for what might be coming.

Application resiliency will be increasingly important as systems get larger and larger. As the number of hardware components—CPUs, memory DIMMs, voltage regulators, et cetera—increases, the chance of a failure increases. Applications need to be able to survive hardware failures.

Achieving resilient applications will require changes at many levels of the systems. The MPI standard must be adapted to support resiliency. Cray's MPI implementation must be made robust in the face of network failures. (This will happen in Gemini-based systems.) ALPS must be able to heal its fan-out tree after a node dropout. Applications must be notified of failures, and then take corrective actions.

Cray is exploring ways to reduce the effect of compute node noise or jitter on applications. We are building a prototype to move overhead work (interrupts, daemon execution) to a single core. This frees the other cores on a node for application work. Noise-sensitive applications appear to benefit from this specialization of cores. POP on 8K cores showed a 23% improvement.

Not all applications benefit. Because of this, core specialization (if released in a future version of CLE) will be selectable on a job basis.

Dedicating one core to overhead is a special case of the general idea of specializing cores. Other instantiations are possible.

3.3. Baker-Gemini Features

Gemini will offer a new and improved software interface to the NIC. The User Gemini Network Interface (uGNI) will allow MPI applications access to Gemini. Support for PGAS languages will be through the Distributed Memory Applications.

Through a combination of hardware and software improvements, Baker-Gemini will be able to survive the loss of an interconnect link, that is, a single hardware link failure will not take the entire system, although some applications may be terminated.

Gemini will increase performance, notably in the injection count—millions of messages per second instead of a few hundred thousand—and ping-pong latency—about 1.7 microseconds.

Gemini will cause changes in many parts of the software stack. Gemini will be supported starting with the Danube (2010) release.

4. Conclusion

This paper has presented specific features which will be coming in 2009 and 2010 releases of Cray's operating system. In addition, we have discussed the themes and thought processes behind our plans.

Acknowledgments

The author would like to thank his colleagues and development team at Cray. Their commitment to producing the world's best supercomputers makes it a pleasure to come to work every day, as well as making this paper possible.

About the Author

Charlie Carroll is Director, OS and I/O with Cray, Inc. If you have comments on our road map, he would love to hear from you at charliec@cray.com.