IBM @server

Using IBM Cluster Systems Management (CSM) for Linux in xCAT Environments

White Paper

Janet Ellsworth

Vallard Benincosa

Bruce Potter

November, 2003

Introduction

IBM® Cluster Systems Management (CSM) can help reduce the total cost of ownership for cluster environments by streamlining and simplifying the management of large numbers of servers from a single point of control. This is of particular value in large data centers or for customers who are running complex high-performance workloads, where managers are constantly challenged to meet ever changing business demands and improve service levels, at the same time that IT budgets are shrinking. As customers evolve from using clusters for single-purpose workloads to employing cluster technology in grids or utility models, CSM can help provide a path for growth and enhanced value.

CSM for Linux has been under development since early 2000 and was first released as a product in June, 2001. The original requirement for this product was to provide common cluster management for Linux and AIX®, leveraging IBM's existing cluster software portfolio to work with Open Source packages and best practices from the Linux cluster market. While CSM provides basic cluster management functions like hardware control and operating system and software installation, it also provides significant value-added functions such as configuration file management, a robust monitoring infrastructure, automated event management, diagnostic probes, and common management of AIX and Linux clusters.

As CSM was being developed, the Extreme Cluster Administration Toolkit (xCAT), a script-based package, was developed by IBM's advanced technical sales support team and provided to customers purchasing Linux clusters based on IBM @server® xSeries®, IBM @server BladeCenter™ HS20 and IBM @server 325 (e325) servers and used by IBM Global Services, to address their need for tools to deploy and manage Linux clusters. In the interim, xCAT became a proving ground for new concepts and practices centered on the management of Linux clusters, and can be a model for collaborative development between IBM and its customers.

CSM is now available in its third release with expanded scalability and hardware support.[1] Many xCAT functions have been integrated into CSM, along with utilities to automate the process of migrating from xCAT to CSM. Through the use of CSM, customers who have implemented xCAT can now have many of these functions with full product support from IBM.

The key benefits of CSM can help to reduce costs, provide higher availability of the cluster for productive use, and improve system utilization. Customers who have existing AIX operating system-based cluster systems can leverage those skills to manage their Linux clusters. System administrators can automate repetitive installation and configuration tasks, automate problem determination and recovery, monitor and report health information and resource utilization, and automatically recover from node, storage or network failures. This can lead to overall simplification of cluster administration.

[1]This paper addresses **CSM v1.3.2 (made available on September 30th, 2003)**, unless otherwise noted.

What's New for CSM v1.3.2?

- CSM runs on the new AMD Opteron™ machine – IBM @server 325
  (as of CSM v1.3.2.1 service update 11/03)

- CSM has now been tested on Red Hat 9

- The install mechanism for SuSE and SuSE Linux Enterprise Server (SLES) distributions has been changed. Previously CSM made use of the Open Source package System Installation Suite (SIS). CSM will now use AutoYaST, the native SuSE / SLES installer. Feedback from customers indicated a preference for using the native install mechanism that came with the operating system rather than having to download a separate package. (Note that CSM already uses the native install mechanism on Red Hat - Kickstart).

- New diagnostic probes to help customers to determine that certain things are correct before installing the nodes in the cluster. These probes are designed to verify that the management server has been set up correctly, and that the hardware control points, console servers, and associated information are all configured correctly.

- CSM now provides backup and restore scripts designed to save and restore critical CSM management server data. These scripts can be used in a variety of ways such as setting up new clusters using configuration data from existing customers, moving a cluster's management server to a backup server, or saving critical data prior to a management server operating system upgrade and then restoring it.

- A new command called csmstat that will provide a snapshot of the cluster when run. It can be run across all the nodes in the cluster, or for a particular set of nodes or node groups. It will return for each node the network reachability of the node, power status, status of network interfaces on the nodes and LED/LCD display. For customers familiar with PSSP, this functionality is very similar to what the PSSP command "spmon -d" was often used for.

IBM Linux Cluster Management Strategy

Overview

Customers deploying Linux clusters require tools for managing the hardware, installing operating systems and other software, managing configurations, and automatically monitoring and processing events. In addition to these functional requirements, a systems management solution must be flexible and scalable. Use of and participation in Open Source projects is particularly important with many customers deploying Linux.

IBM offers a full range of systems management solutions for Linux clusters, including Cluster Systems Management for Linux, IBM Director, and General Parallel File System for Linux (GPFS). CSM for Linux v1.3.2, available today, supports up to 512 managed nodes (with larger systems available by special bid) and provides many functions equivalent to those in xCAT, plus additional features such as monitoring and event management that are not available in xCAT. In addition, CSM v1.3.2 provides management of both xSeries, BladeCenter HS20 and e325 servers running Linux and IBM @server pSeries® servers running AIX 5L™ and/or Linux from a single management server.

CSM can come with both the IBM @server Cluster 1350 (an integrated cluster offering based on xSeries, BladeCenter HS20 and e325 servers running Linux) and the IBM @server Cluster 1600 (clusters based on pSeries servers running AIX 5L and/or Linux), and is also available as a stand-alone software product. CSM contains many of the functions in xCAT, and more of these features are planned to be incorporated into CSM in upcoming releases; the detailed mapping between xCAT and CSM features is described later in this document.

IBM Director is designed to provide extensive hardware management for xSeries servers, a graphical user console, and upward integration modules for enterprise management solutions such as Tivoli®. IBM Director is designed to work more broadly in both cluster environments and loosely connected work groups and enterprises. CSM can also be used to extend IBM Director with functions specific to managing Linux clusters. The integration between CSM and IBM Director first made available in CSM v1.3.1 in May 2003 allows the CSM-based cluster nodes to be displayed in the IBM Director console, and ties the event management infrastructure of CSM and IBM Director together (i.e., events that are being monitored by CSM will appear in the IBM Director console).

IBM offers General Parallel File System for Linux (GPFS), a high-performance, shared-disk file system for Linux cluster environments. It is based on GPFS for AIX, which has been widely accepted in large scale clusters on IBM RS/6000® SP™ and pSeries systems. GPFS is designed to provide data access from all nodes in a Linux cluster. Parallel and serial applications can readily access shared files using standard UNIX® file system interfaces, and the same file can be accessed concurrently from multiple nodes. GPFS provides high availability through logging and replication, and can be configured for failover from both disk and server malfunctions. GPFS provides an enhanced layer of scalability and manageability for Linux clusters.

Development Model

The development model for IBM's cluster systems management products has been adapted for flexibility. Periodic service updates of CSM are planned to pick up new hardware, new Linux distributions and potentially new requirements in between releases and the development teams have planned to provide early access to releases in response to such requirements.

The Enhanced Cluster Tools (ECT) site on IBM alphaWorks®[2] provides tools that are available at no cost on an "as-is" basis. These tools add to existing CSM functionality, such as setting up HPC tools and Myrinet™ switches, configuring terminal servers and Remote Service Attachments (RSA's), consuming SNMP events and turning them into CSM events, and a Web browser interface for CSM. IBM also provides an xCAT-to-CSM migration tool on this site as well. The purpose of ECT is to rapidly release new CSM functions "as-is" in advance of these functions being available in the base product, and to provide additional complementary tools that can be frequently updated.

CSM also leverages certain Open Source tools such as rdist, AutoUpdate and Conserver, with the added robustness and stability of a fully supported product.

In summary, IBM has an extensive set of tools available for managing Linux clusters that leverages Open Source as well as IBM's significant investments in AIX 5L and Linux clustering, xSeries platform management and xCAT. Customers can begin to migrate from xCAT to CSM now, and can then be positioned to take advantage of a full range of enhancements and service offerings, including integration with IBM Director, common AIX 5L and Linux cluster management, advanced monitoring and automated operations, and continued support for new xSeries hardware and additional Linux distributions. Longer term, customers may be able to take advantage of planned investments in areas such as high-availability cluster solutions and automated provisioning using CSM's robust event monitoring capabilities.

CSM Overview

Customers face a variety of challenges with managing their cluster systems. These may include installing and updating software on all or selected nodes of the cluster, detecting and diagnosing errors before a major problem occurs, handling a problem that occurs after hours, and remotely controlling or rebooting servers without having to go out to each and every one.

CSM provides a robust, powerful and centralized way to manage large numbers of xSeries, e325, BladeCenter HS20 and pSeries servers from a single point-of-control. CSM may help lower the overall cost of IT ownership by simplifying the tasks of installing, operating and maintaining clusters. Clusters can be scaled up easily without necessarily increasing administrative support. CSM enables quick cluster setup and can help save time making configuration changes and upgrades. The highly reliable infrastructure and automated event monitoring capabilities can lead to problem avoidance, rapid resolution and recovery, thereby helping improving the time the cluster is available for productive use. The efficient monitoring and reduced network traffic can improve overall performance.

CSM has been designed to be flexible and modular, and to leverage leading systems management technologies available from IBM. Following are the key design points which drive the development of CSM, and the benefits that are derived.

---

[2] http://www.alphaworks.**ibm.com**/tech/ect4linux

- Make use of Open Source systems management tools where applicable.[this is problematic considering the issues we faced on the recent release. I think we should temper these statements]
- Provide a modular architecture, so that administrators can choose the parts of CSM they wish to use.
- Leverage proven cluster technology from the RS/6000 SP (Parallel Systems Support Programs (PSSP) for AIX) and xCAT, and provide a transition for both of these sets of customers.
- Integrate with IBM Director for a unified management solution on Linux, in which CSM and/or Director can be used independently, but when used together they are designed to provide the capabilities needed to manage the cluster hardware in a coordinated fashion.
- Provide very efficient monitoring with minimum use of cycles on the managed servers.
- Provide mechanisms intended to integrate with high availability clusters and autonomic computing.
- Provide command line access to all cluster operations and data, enabling administrators to utilize scripting for cluster administration.
- Respond rapidly to feedback and requirements from customers.

CSM Features

CSM provides a rich set of features designed to manage many aspects of the cluster. This includes installing and updating machines remotely, continuously monitoring all machines in the cluster, remotely powering on/off and rebooting nodes, managing configuration files for groups of nodes or for the entire cluster, establishing automated responses that can be run any time a problem occurs to provide notification or take corrective action, and running probes to diagnose problems. These features are described in more detail in the sections which follow.

Manage Node and Node Group Information

The administrator can add nodes to the cluster, remove nodes from the cluster, and change the information for each node in the cluster, all from the management server. Many run-time attributes of all the nodes can be queried in real time from the management server with a single command. Efficient heart-beating provides a status for each node.

Node groups can be defined to manage and monitor collections of cluster nodes as a single entity. Node groups can be static or dynamic; for example, there is a predefined dynamic node group for AIX 5L nodes and a predefined node group for Linux nodes. As nodes are added to the cluster, they will automatically become part of one of those node groups. The dynamic node groups are based on SQL select string queries.

Install, Set Up, and Maintain Software on the Nodes

The administrator can define the nodes in the cluster and then CSM can remotely perform network installation of the nodes, including installation of the full operating system, CSM, and other required software. CSM will update software on the nodes, such as new CSM versions, and new kernels (including custom kernels). Different node types can be maintained uniquely based on node groups defined by the administrator. CSM will automatically set up the security configuration for the underlying cluster infrastructure, called the Reliable Scalable Cluster Technology (RSCT).

Control Hardware

CSM provides hardware control capability to power on, off, reboot, and query nodes in the cluster. CSM uses the integrated service processor that comes with the xSeries, e315 and BladeCenter HS20 servers to communicate with each node in the cluster. CSM can bring up a remote hardware console for any node, and use this console to collect MAC addresses. CSM can manage a variety of hardware attributes, including configuring SNMP alerts and collecting error logs from the service processors, and collecting environmental information, vital product data, and hardware inventory information. The CSM architecture is structured so that additional power methods can be added.

Distributed Command Execution

An administrator can run commands in parallel across nodes or node groups in the cluster and gather the output using the dsh (distributed shell) command. A sophisticated fan-out algorithm allows the customer to control the level of parallelism. The dsh command can use rsh (UNIX basic remote shell) or ssh (secure shell) - the administrator can decide which one to use. The dshbak command can format the output returned from dsh if desired; for example, collapsing identical output from more than one node so that it is displayed only once. A graphical user interface for dsh, the distributed command execution manager, is also provided with CSM to separate output and make it easier to see the status and results of each node.

Configuration File Management

A configuration file manager (CFM) is provided to synchronize and maintain the consistency in files across nodes in the cluster. This saves the administrator from having to copy files manually across the nodes in the cluster. The particular files can be changed on the management server and then distributed to all the nodes or node groups in the cluster. Different versions of files can be distributed to groups of nodes to facilitate maintaining different node types. CFM can use make use of meta variables for IP address or hostname substitution in files being transferred. Scripts can also be run for processing before and after a file is copied (e.g., to stop/start daemons). CFM makes use of rdist for file transfer. Rdist can use rsh (UNIX basic remote shell) or ssh (secure shell) - the administrator can decide which one to use.

Event Monitoring and Automated Responses

An administrator can set up monitoring for various conditions across nodes or node groups in the cluster, and have actions run in response to events that occur in the cluster. Conditions that can be monitored include:

- network accessibility
- power status
- whether applications or daemons that are running on the node are up or down
- CPU, memory and file system utilization

Additional conditions can be customized. Actions can be programmed to occur in response to a condition, including commands to be run on the management server or on any node of the cluster, or notification actions such as logging, e-mail or paging. SNMP traps can also be generated in response to events in the cluster.

Predefined *conditions* are shipped with CSM for every type of event that can be monitored, so that monitoring can be started "out of the box." Predefined *responses* for e-mail notification, SNMP traps, logging and displaying a message to a console are provided as well. A user can take a condition, quickly associate it with a response, and start monitoring. The administrator can easily customize these conditions and responses, or tailor them to fit their own needs. Conditions can be created with various severity levels; for example, when an informational event is triggered, one response can be run (such as logging), when a warning event is triggered for that same condition, another response can be run (such as e-mail), and when a critical event is triggered for that same condition, yet another response can be run (such as paging the administrator). In addition to notification actions, administrator-defined recovery actions can also occur. These can include cleaning up file systems that are filling up, taking actions to help restart a critical application that went down, and so on. The types of resources that can be monitored or controlled can be extended by an administrator providing their own sensors.

The monitoring capability for CSM was designed to be extremely efficient. Monitoring is done only if requested by the administrator and if the resources are available; each resource manager determines the best way to monitor its set of resources - polling is avoided if at all possible.

Security

The cluster security infrastructure provides an "out-of-the-box" security authentication mechanism with host-base public key/private key setup. The security infrastructure allows for pluggable security mechanisms; for example, Kerberos V5 **can be** supported in the future. CSM is designed to automatically set up and exchange public keys between the management server and the managed nodes for both RMC and openssh.

IBM Director Integration with CSM

CSM provides integration with IBM Director so that both products can be used together to manage machines from a single graphical console, sharing information and events with each other. Features include recognition of CSM management servers and nodes during system discovery, representation of CSM-based cluster nodes in the IBM Director console as manageable entities, the ability to view CSM node attribute data using the console's inventory browser, and the ability to create dynamic groups of CSM nodes based on CSM node attribute data (as well as any other system inventory data).

The Director integration for CSM also allows the administrator to build event action plans based on CSM conditions (as well as other Director event types), and correspondingly the ability to trigger Director actions based on RMC events. CSM and other Linux commands can be run on the cluster nodes or management server through the Director console. Director provides additional hardware control features beyond those available in CSM, including RAID configuration and rack manager, and also runs in both Windows® and Linux environments.

Interoperability between CSM for Linux and AIX 5L

CSM is **also** available for both Linux **on pSeries** and AIX 5L. This enables clusters of both xSeries, e325 and BladeCenter HS20 servers running Linux, and pSeries servers running AIX 5L **or Linux**, to be managed from a single point-of-control, using an AIX 5L Version 5.2 operating system-based management server. This is of particular value to customers who have xSeries, e325, BladeCenter HS20 and pSeries servers in their infrastructure, and who wish to harness the power of both platforms while using common management software to control the systems as a single managed entity.

Additional Features

CSM includes diagnostic tools **called probes** to examine daemons and software components on the management server and cluster nodes to aid in customer problem determination – some of the types of things these probes check for are network health, NFS health, status of daemons that CSM provides, and determining whether everything is set up correctly to install the cluster nodes.

Additional Information

CSM supports various hardware (including several xSeries servers, BladeCenter HS20 and e325) as well as various Linux distributions which are documented in its online documentation at http://www.**ibm.com**/servers/eserver/clusters/library/linux.html.

Enhanced Cluster Tools (ECT) on IBM alphaWorks

The Enhanced Cluster Tools (ECT) package on IBM alphaWorks was created to provide additional functions and utilities to enhance CSM. Software is available on an "as-is" (unsupported) basis and may be downloaded free of charge. Some of the tools that are available on ECT are:

- Tools to set up the HPC stack (such as openpbs, maui, mpich, as well as setting up Myrinet switches)
- Scripts to configure terminal servers and RSA's,
- Tools for consuming SNMP events and turning them into CSM events
- A Web browser interface for CSM.
- An xCAT-to-CSM migration tool.

A Mapping of xCAT Function to CSM

CSM and xCAT can coexist on a cluster, and share many of the same features. The following table describes the mapping between xCAT and CSM commands.

Category	Function	xCAT Command	CSM Command
Hardware Control	Power control (both RSA and APC)	rpower, rreset	rpower
	Hardware inventory and VPD	rinv	lshwstat
	Hardware environmentals	renv	lshwstat
	RSA event log access	reventlog	reventlog
	List nodes attached to a service processor	mpascan	lshwinfo
	Reset the RSA	mpareset	rpower, resetsp_hcp, resetsp_host
Remote Consoles	ELS, ESP, CPS, Computone, iTouch	rcons, wcons	rconsole
	MAC address collection via console	getmacs	getmacs, getadapters
	Choosing which serial port to use	site.tab (serialmac), tty.tab	ConsoleSerialDevice node attribute
Node Installation	Network install nodes	mkks, nodeset, rinstall, winstall	csmsetupks, csmsetupsis, installnode
	Apply updated RPMs and kernels	post directory	updates directory
Administration	Set up ssh keys or rsh .rhosts	makesshgkh, genrhosts	updatenode
	Run command on nodes	psh	dsh
	Copy files to nodes	prsync, prcp	/cfmroot, updatenode
	Get status of nodes	pping, nodestat	lsnode -p
	Display information about a node	nodels	lsnode

xCAT Features Incorporated into CSM and ECT

The following table indicates when the functional gaps between xCAT and CSM are expected to be filled. Some of these features are available now or planned to be available in CSM, while others are planned to be made available on the ECT site.

xCAT Function	CSM Release Available In
Scaling to 512 nodes	CSM v1.3.0 in 12/02 (512+ by special bid)
Scalability to thousands of nodes via staged installation and multiple tftp, NFS servers	CSM v1.3 and future
Support multiple install servers	CSM v1.3
Automated parallel configuration of service processors and automated setup of management processor adapters	Configuring SNMP alerts in CSM v1.3; configuration of userid/pw/ip address is future
Support installation of nodes without putting CSM code on them	CSM v1.3.1 in 5/03
Get MAC addresses from the Ethernet switch (Cisco and Extreme) without terminal servers	ECT
Remote BIOS and SP firmware updates (rflash)	ECT
Set up HPC tools including automatic build, installation and configuration of PBS/Maui scheduler and integration with PBS commands, setup of MPICH, MPICH-GM and LAM-MPI	ECT
Support node range syntax	CSM v1.3
Support disk cloning install method (rclone, wclone, getclone, clonestat)	ECT
Non-node device support (power ctrl for terminal servers and serial reset support for RSAs)	ECT
Commands for automated setup of terminal servers	ECT
Set up DNS, NTP, and NIS	DNS & NTP set up is in ECT & CSM (respectively) today, NIS will be in ECT in the future
Automatic build of Myrinet driver and installation of Myrinet software, automatic generation of Myrinet routes, checking Myrinet route integrity (troubleshooting)	Future
Automatic dhcpd.conf generation	CSM v1.2
Effective ssh host key management	CSM v1.3
Install updated kernels	CSM 1.3.0.10
Parallel copy of files	ECT
Add cluster user	CSM 1.3.1

xCAT Function	CSM Release Available In
Tile multiple rconsole windows	CSM 1.3.2
Catching SNMP alerts	ECT
Cloning (via SystemImager)	ECT
rvid, wvid (console via service processor)	CSM v1.3 (BladeCenter management module only)

Running CSM and xCAT Together

CSM and xCAT can be run together on the same cluster. This is useful for customers who are making the transition from xCAT to CSM, or who may want to use features of xCAT for some functions that CSM does not yet provide. CSM has an additional RPM named "csm.ect" that may be downloaded off of the IBM alphaWorks site:
http://www.alphaworks.ibm.com/tech/ect4linux
There is also a document on that site named CSM ECT HOWTO.html that explains in detail how to use CSM ECT.

CSM ECT ships most of the base code from xCAT. It can be used to either migrate xCAT on an existing system to CSM, or to add xCAT functions to an existing CSM cluster.

To co-exist xCAT and CSM on a system with xCAT installed:

Install CSM on the head node of an existing xCAT cluster
Install CSM ECT rpm on the head node
Run /opt/csm/ect/bin/xcat2csm to copy node information from the xCAT tables into the CSM database
Run "updatenode -a" to install CSM on all nodes
xCAT and CSM can be run side by side
If nodes are added to the xCAT tables, xcat2csm can be run again to copy just the new nodes into the CSM database

To co-exist xCAT and CSM on a system with CSM only installed:

Install CSM ECT rpm on the head node
Most of the xCAT code will be installed with this rpm in the /opt/csm/ect/xcat subdirectories.
Run /opt/csm/ect/bin/csm2xcat to populate xCAT tables. Edit as appropriate.

Read the CSM ECT HOWTO.html file available from the alphaWorks Web page or inside the /opt/csm/ect/doc directory for more information.

Extending CSM

Interfaces exist that allow additional functions to be added to CSM, including function that is currently in xCAT. These interfaces are the CSM command line interface (documented in the

Administration Guide), and plug-in interfaces for console methods, power methods, and additional event sensors. CSM uses C/C++ for the event subsystem and the node information database. The remaining code is perl (scripts) that tie together those components with the Open Source tools. The value of CSM can be further enhanced by contributions from the user community; within IBM these may be submitted to bp@us.ibm.com; a process for contributions from outside IBM is planned in the future.

Conclusion

IBM has well over a decade of experience in delivering robust, scalable, and highly manageable clusters, starting with the RS/6000 SP and evolving to the wide range of cluster software now available for Linux. This heritage of production-ready technology, and IBM's demonstrated expertise in delivering large and powerful integrated cluster solutions**,** position IBM as a leader in Linux clusters.

CSM and xCAT represent two key contributions to IBM's cluster leadership. With many of the xCAT features now available or planned to be available in CSM, customers can leverage the combined strength and flexibility of these offerings in a fully supported product. CSM, along with IBM Director, the Enhanced Cluster Tools extensions on IBM alphaWorks, result in a comprehensive solution for cluster management.

Resources and Information

CSM external mailing list	http://www.**ibm.com**/developerworks/oss/mailman/listinfo/csm
IBM clusters information	http://www.**ibm.com**/servers/eserver/clusters/
Cluster software (including information about new releases of CSM and Try-and-Buy feature)	http://www.**ibm.com**/servers/eserver/clusters/software/
CSM FAQ	http://techsupport.services.ibm.com/server/cluster/tips/csm_faq.html
Service Web page (to obtain information about CSM updates)	http://techsupport.services.**ibm.com**/server/cluster
CSM documentation (Administration Guide, Planning and Installation Guide, Hardware Planning and Control Guide)	http://www.**ibm.com**/servers/eserver/clusters/library/
Enhanced Cluster Tools (ECT) on IBM alphaWorks	http://www.alphaworks.**ibm.com**/tech/ect4linux
xCAT	http://xcat.org
Man pages	Available for all CSM commands
Redbook on Linux Clustering with CSM and GPFS (SG24-6601)	http://www.redbooks.**ibm.com**/

IBM Corporation
Marketing Communications
Systems Group
Route 100
Somers, New York 10589

Produced in the United States of America
11-03