Platform HPC

Focus on science, not computer science
with the easiest and most complete HPC cluster solution

A Platform Computing Technology White Paper
November 2010

Platform Computing

Table of Contents

1. Introduction

As many IT managers know, cost and complexity often go hand in hand. There are few software environments more complex to manage and utilize than modern high performance computing (HPC) clusters. Therefore, addressing the problem of complexity in cluster management is a key aspect of containing cost.

Linux® operating system-based clusters have become increasingly prevalent at large supercomputing centers and continue to make significant in-roads in commercial and academic settings. This is primarily due to their superior price/performance and flexibility, as well as the availability of commercial applications that are based on the Linux OS.

Ironically, the same factors that make Linux a clear choice for high performance computing often make the operating system less accessible to smaller computing centers. These organizations may have Microsoft Windows® administrators on staff, but have little or no Linux or cluster management experience. The complexity and cost of cluster management often outweigh the benefits that make open, commodity clusters so compelling. Not only can HPC cluster deployments be difficult, but the ongoing need to deal with heterogeneous hardware and operating systems, mixed workloads, and rapidly evolving toolsets make deploying and managing an HPC cluster a daunting task.

These issues create a barrier to entry for scientists and researchers who require the performance of an HPC cluster, but are limited to the performance of a workstation or a single. This is why ease of use is now mandatory for HPC cluster management. This paper reviews such an easy to use cluster management solution, which is now commercially available from Platform Computing.

2. The cluster management challenge

To provide a proper HPC application environment, system administrators need to provide a full set of capabilities to their users, as shown in Figure 1. These capabilities include cluster provisioning and node management, application workload management, and an environment that makes it easy to develop, run and manage distributed parallel applications.

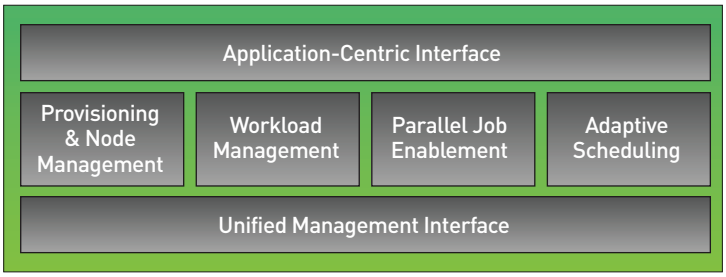

Figure 1. Essential components of an HPC cluster solution

Modern application environments tend to be heterogeneous; some workloads require Windows compute hosts while others require particular Linux operating systems or versions. The ability to change a node's operating system on-the-fly in response to changing application needs — referred to as adaptive scheduling — is also important since it allows system administrators to maximize resource use, and present what appears to be a larger resource pool to cluster users.

Learning how to use a command line interface to power-up, provision and manage a cluster is time consuming is extremely time-consuming. Administrators therefore need remote, web-based access to their HPC environment that makes it easier for them to install and manage an HPC cluster. An easy-to-use application-centric web interface can have tangible benefits including improved productivity, reduced training requirements, reduced errors rates, and secure remote access.

While there are several cluster management tools that address parts of these requirements, few address them fully, and some tools are little more than collections of discrete open-source software components.

Cluster toolkits such as ROCKS, xCAT and Scyld Clusterware™ focus largely on the problem of cluster provisioning and management. While they clearly simplify cluster deployment, administrators wanting to make changes to node configurations or customize their environment will quickly find themselves hand-editing XML configuration files or writing their own shell scripts. Third-party workload managers and various open-source MPI[1] libraries might be included as part of a distribution. However, these included components are loosely integrated and often need to be managed separately from the cluster manager. As a result the cluster administrator needs to learn how to utilize each additional piece of software in order to manage the cluster effectively.

Other HPC solutions are designed purely for application workload management. These include solutions such as Condor®, Oracle Grid Engine, MOAB Cluster Suite®, and PBS Professional™. While these are all capable workload managers, most do not address at all the issue of cluster management, application integration, or adaptive scheduling. If such capabilities exist they usually require the purchase of additional software products.

Parallel job management is also critical. One of the primary reasons that customers deploy HPC clusters is to maximize application performance. Processing problems in parallel is a common way to achieve performance gains. The choice of MPI, its scalability, and the degree to which it is integrated with various OFED[2] drivers and high performance interconnects has a direct impact on delivered application performance.

[1] MPI refers to the message-passing interface, a widely deployed solution in HPC environments that enables distributed parallel applications by facilitating communications between processes running on different machines.

[2] OFED refers to the OpenFabrics Enterprise Distribution

Furthermore, if the workload manager does not incorporate specific parallel job management features, busy cluster users and administrators can find themselves manually cleaning up after failed MPI jobs or writing their own shell scripts to do the same.

Complexity is a real problem. Many small organization or department grapple with a new vocabulary full of cryptic commands, configuring and troubleshooting Anaconda[3] kick start scripts, finding the correct OFED drivers for specialized hardware, and configuring open source monitoring systems like Ganglia or Nagios®. Without an integrated solution administrators may need to deal with dozens of distinct software components, making managing HPC cluster implementations extremely tedious and time consuming.

Cluster Management Challenges:

- Using a command-line interface for cluster administration
- Provisioning and managing clusters remotely
- Sourcing multiple tools for cluster management, workload management & application integration
- Integrating, installing & supporting multiple software components Application integration using complex scripts
- Parallel job management Quickly re-provisioning nodes based on workload

3. Re-thinking HPC clusters

Clearly these challenges demand a fresh approach to HPC cluster management. Platform HPC represents a "re-think" of how HPC clusters are deployed and managed. Rather than addressing only part of the HPC management puzzle, Platform HPC addresses all facets of cluster management. It provides:

- Complete, easy-to-use cluster management solution
- Integrated application support
- User-friendly, topology aware workload management
- Robust workload and system monitoring and reporting
- Dynamic operating system multi-boot (adaptive scheduling)
- GPU scheduling
- Robust commercial MPI library (Platform MPI)
- Web-based interface for access anywhere

3.1 Most complete HPC cluster management solution

Platform HPC makes it easy to deploy, run and manage HPC clusters while meeting the most demanding requirements for application performance and predictable workload management. It is a complete solution that provides a robust set of cluster management capabilities; from cluster provisioning and management to workload management and monitoring. The easy to use unified web portal provides a single point of access into the cluster, making it easy to manage your jobs and optimize application performance.

As the table in Appendix A illustrates, other HPC cluster software are incomplete combinations of multiple tools and interfaces, which are not integrated, certified, or tested together. Platform HPC is more than just a stack of software; it is a fully integrated and certified solution designed to ensure ease of use and simplified troubleshooting.

3.2 Integrated application support

High performing, HPC-optimized MPI libraries come integrated with Platform HPC, making it easy to get parallel applications up and running. Scripting guidelines and job submission templates for commonly used commercial applications simplify job submission, reduce setup time and minimize operation errors. Once the applications are up and running, Platform HPC improves application performance by intelligently scheduling resources based onworkload characteristics.

[3] Anaconda is the installation program used by Fedora, Red Hat Enterprise Linux and other Linux

3.3 Fully certified and supported

Platform HPC unlocks cluster management to provide the easiest and most complete HPC management capabilities while reducing overall cluster cost and improving administrator productivity. It is based on the industry's most mature and robust workload manager, Platform LSF, making it the most robust and reliable solution on the market.

Other solutions are typically a collection of open-source tools, which may also include pieces of commercially developed software. They lack key HPC functionality and vendor support, relying on the administrator's technical ability and time to implement. Platform HPC is a single product with a single installer and a unified web-based management interface. With the best support in the HPC industry, Platform HPC provides the most complete solution for HPC cluster management.

4. Complete solution

Platform HPC provides a complete set of HPC cluster management features. In this section we'll explore some of these unique capabilities in more detail.

4.1 Easy to use, cluster provisioning and management

With Platform HPC administrators can quickly provision and manage HPC clusters with unprecedented ease. It ensures maximum uptime and can transparently synchronize files to cluster nodes without any downtime or re-installation.

Fast and efficient software Installation – Platform HPC can be installed on the head node takes less than one hour using three different mechanisms:

- Onsite using third-party provisioning tools
- In the factory or onsite by the hardware vendor using the hardware vendor or a third party's provisioning tools
- Onsite by the customer using provisioning tools installed on existing hardware

Installing software on cluster nodes is simply a matter of associating cluster nodes with flexible node group definitions through the web-based interface.

Flexible provisioning – Platform HPC offers multiple options for provisioning Linux operating environments that include:

- Package-based provisioning
- Image based provisioning
- Diskless node provisioning

Mixed clusters comprised of Linux, Microsoft Windows and Windows HPC nodes can also be deployed and, using Platform HPC node personalities, can change dynamically based on application requirements.

Large collections of hosts can be added as a node group. Platform HPC automatically manages details such as IP address assignment and node naming conventions that can reflect the position of cluster nodes in data center racks.

Unlike competing solutions, Platform HPC can deploy multiple operating systems and OS versions to a cluster simultaneously. This includes Red Hat® Enterprise Linux, CentOS, Scientific Linux, and SUSE® Linux Enterprise Server. This provides administrators with greater flexibility in how they serve their user communities and means that HPC clusters can grow and evolve incrementally as requirements change.

Simplified configuration changes – Platform HPC simplifies administration and increases cluster availability by allowing changes such as new package installations, patch updates, and changes to configuration files to be propagated to cluster nodes automatically without the need to re-install cluster nodes. It also provides a mechanism whereby experienced administrators can quickly perform operations in parallel across multiple cluster nodes.

Repository snapshots / trial installations – Upgrading software can be risky, particularly in complex environments. If a new software upgrade introduces problems, administrators often need to rapidly "rollback" to a known good state. With other cluster managers this can mean having to re-install the entire cluster. Platform HPC incorporates repository snapshots, which are "restore points" for the entire cluster. Administrators can snapshot a known good repository, make changes to their environment, and easily revert to a previous "known good" repository in the event of an unforeseen problem. This powerful capability takes the risk out of cluster upgrades.

New hardware integration – When new hardware is added to a cluster it may require new or updated device drivers that are not supported by the OS environment on the installer node. This means that a newly updated node may not network boot and provision until the head node on the cluster is updated with a new operating system; a tedious and disruptive process. Platform HPC includes a driver patching utility that allows updated device drivers to be inserted into existing repositories, essentially future proofing the cluster and providing a simplified means of supporting new hardware without needing to re-install the environment from scratch.

Software updates with no re-boot – Some cluster managers always re-boot nodes when updating software, regardless of how minor the change. This is simple way to manage updates. However, scheduling downtime can be difficult and disruptive. Platform HPC performs updates intelligently and selectively so that compute nodes continue to run even as non-intrusive updates are applied. The repository is automatically updated so that future installations include the software update. For changes that require the re-installation of the node (e.g. upgrading an operating system) these changes can be made in a "pending" state until downtime can be scheduled.

Meta-installer with optional provisioning mechanisms – In order to ensure compatibility with diverse compute environments, Platform HPC cluster can self-provision or be installed using a third-party cluster manager. This makes it easy to install the software on clusters that already have another third-party cluster manager installed.

4.3 User-friendly, topology aware workload management

Platform HPC includes a robust workload scheduling capability, which is based on Platform LSF — the industry's most powerful, comprehensive, policy driven workload management solution for engineering and scientific distributed computing environments. By scheduling workloads intelligently according to policy, Platform HPC improves end user productivity with minimal system administrative effort. In addition, it allows HPC user teams to easily access and share all computing resources, while reducing time between simulation iterations.

GPU scheduling – Platform HPC provides the capability to schedule jobs to GPUs as well as CPUs. This is particularly advantageous in heterogeneous hardware environments as it means that administrators can configure Platform HPC so that only those jobs that can benefit from running on GPUs are allocated to those resources. This frees up CPU-based resources to run other jobs. Using the unified management interface, administrators can monitor the GPU performance as well as detect ECC errors.

Figure 2. Resource monitoring and reporting

4.2 Unified management interface

Competing cluster management tools either do not have a web-based interface or require multiple interfaces for managing different functional areas. In comparison, Platform HPC includes a single unified interface through which all administrative tasks can be performed including node-management, job-management, jobs and cluster monitoring and reporting. Using the unified management interface, even cluster administrators with very little Linux experience can competently manage a state of the art HPC cluster.

Job management – While command line savvy users can continue using the remote terminal capability, the unified web portal that make it easy to submit, monitor, and manage jobs. As changes are made to the cluster configuration, Platform HPC automatically re-configures key components, ensuring that jobs are allocated to the appropriate resources. The web portal is customizable and provides job data management, remote visualization and interactive job support.

Workload/system correlation – Administrators can also correlate workload information with system load, so that they can make timely decisions and proactively manage compute resources against business demand. When it's time for capacity planning, the management interface can be used to run detailed reports and analyses which quantify user needs and remove the guess work from capacity expansion.

Simplified cluster management – The unified management console is used to administer all aspects of the cluster environment. It enables administrators to easily install, manage and monitor their cluster. It also provides an interactive environment to easily package software as kits for application deployment as well as pre-integrated commercial application support.

One of the key features of the interface is an operational dashboard that provides comprehensive administrative reports. As Figure 2 illustrates, Platform HPC enables administrators to monitor and report on key performance metrics such as cluster capacity, available memory and CPU utilization. This enables administrators to easily identify and troubleshoot issues.

The easy to use interface saves the cluster administrator time, and means that they do not need to become expert in the administration of open-source software components. It also reduces the possibility of errors and time lost due to incorrect configuration. Cluster administrators enjoy the best of both worlds – easy access to a powerful, web-based cluster manager without the need to learn and separately administer all the tools that comprise the HPC cluster environment.

4.4 Robust Commercial MPI library

Platform MPI – In order to make it easier to get parallel applications up and running, Platform HPC includes the industry's most robust and highest performing MPI implementation, Platform MPI. Platform MPI provides consistent performance at application run-time and for application scaling, resulting in top performance results across a range of third-party benchmarks.

Open Source MPI – Platform HPC also includes various other industry standard MPI implementations. This includes MPICH1, MPICH2 and MVAPICH1, which are optimized for cluster hosts connected via InfiniBand, iWARP or other RDMA based interconnects.

4.1 Integrated application support

Job submission templates – Platform HPC comes complete with job submission templates for ANSYS Mechanical, ANSYS Fluent, ANSYS CFX, LS-DYNA, MSC Nastran, Schlumberger ECLIPSE, Simulia Abaqus, and Blast (Figure 3). By configuring these templates based on the application settings in your environment, users can start using the cluster without writing scripts.

Scripting Guidelines – Cluster users that utilize home-grown or open-sources applications, can utilize the Platform HPC scripting guidelines. These user friendly interfaces help minimize job submission errors. They are also self-documenting, enabling users to create their own job submission templates.

Benchmark tests – Platform HPC also includes standard benchmark tests to ensure that your cluster will deliver the best performance without manual tuning.

4.5 Flexible OS provisioning

Platform HPC can deploy multiple operating systems versions concurrently on the same cluster and, based on job resource requirements, dynamically boot the Linux or Windows operating system required to run the job. Administrators can also use a web interface to manually switch nodes to the required OS to meet application demands, providing them with the flexibility to support special requests and accommodate unanticipated changes. Rather than being an extra-cost item as it is with other HPC management suites, this capability is included as a core feature of Platform HPC.

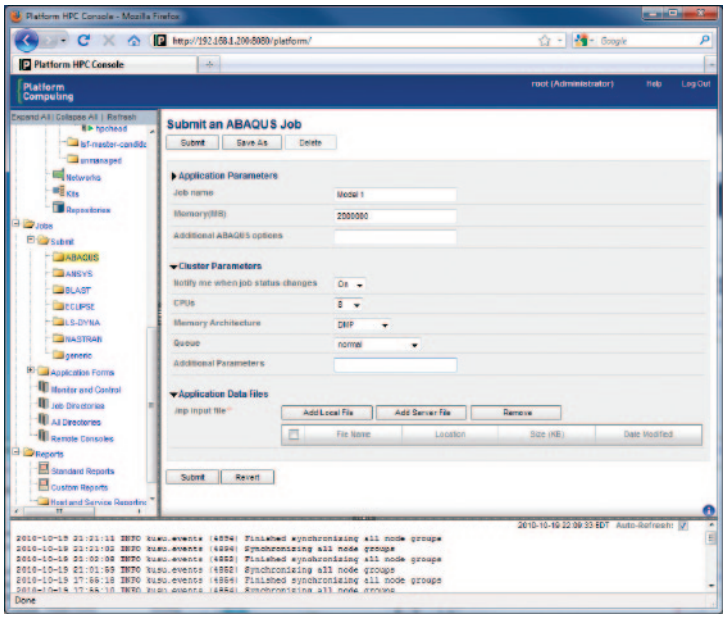

Figure 3. Job submission templates

4.6 Commercial Service and support

Certified cluster configurations – Platform HPC is tested and certified on all partner hardware platforms. By qualifying each platform individually and providing vendor-specific software with optimized libraries and drivers that take maximum advantage of unique hardware features, Platform Computing has essentially done the integration work in advance. As a result clusters can be deployed quickly and predictably with minimal effort. As a testament to this, Platform HPC is certified under the Intel Cluster Ready program.

Enterprise class service and support – Widely regarded as having the best HPC support organization in the business, Platform is uniquely able to support an integrated HPC platform. Because support personnel have direct access to the Platform HPC developers, Platform Computing is able to offer a higher level of support and ensure that any problems encountered are resolved quickly and efficiently.

5. Summary

Platform HPC is the ideal solution for deploying and managing state of the art HPC clusters. It makes cluster management simple, enabling analysts, engineers and scientists from organizations of any size to easily exploit the power of Linux clusters. Unlike other HPC solutions that address only parts of the HPC management challenge, Platform HPC uniquely addresses all aspects of cluster and management including:

- Easy to use, cluster provisioning and management
- User-friendly, topology aware workload management
- Unified management interface
- Robust commercial MPI library
- Integrated application support
- Flexible OS provisioning
- Commercial HPC service and support

By providing simplified management over the entire lifecycle of a cluster, Platform HPC has a direct and positive impact on productivity while helping to reduce complexity and cost. The comprehensive web-based management interface, and features like repository snapshots and the ability to update software packages on the fly means that state-of-the-art HPC clusters can be provisioned and managed even by administrators with little or no Linux administration experience.

Appendix A. Comparison of Platform HPC vs. Open Source Tools

Capability / Feature	Platform HPC	ROCKS*	xCAT*	MOAB Cluster Suite*	Grid Engine*	PBS Pro*	Bright Cluster Manager
Cluster Provisioning and Management							
Initial cluster provisioning	✓	✓	✓	*			✓
Multiple provisioning methods	✓		✓				
Web-based cluster management	✓			✓			
Node updates with no re-boot	✓						
Repository snapshots	✓						
Flexible node templates	✓	weak	✓	*			✓
Multiple OS and OS versions	✓	weak	✓	*			✓
Workload Management & Application Integration							
Integrated workload management	✓	✓		✓	✓	✓	✓
HPC libraries & toolsets	✓	✓					
NVIDIA CUDA SDK support	✓	✓	*	*	*	*	
Web-based job management	✓			✓	✓	*	
Web-based job data management	✓					*	
Multi-boot based on workload	✓			*			
Advanced parallel job management	✓			✓		✓	
Commercial application integrations	✓					*	
MPI Libraries							
Commercial grade MPI	✓						
Workload and system monitoring, reporting and correlation							
Workload monitoring	✓			✓	✓	✓	✓
Workload reporting	✓			*	✓	*	
System monitoring & reporting	✓	✓	✓	*			✓
Workload and system load correlation	✓					*	

* Item marked with an asterisk indicate that the feature is available but requires a separately installed add-on package

+ Comparisons based on web contents and public documents

Platform Computing

Platform Computing is the leader in cluster, grid and cloud management software - serving more than 2,000 of the world's most demanding organizations for over 18 years. Our workload and resource management solutions deliver IT responsiveness and lower costs for enterprise and HPC applications. Platform has strategic relationships with Cray, Dell™, HP, IBM®, Intel®, Microsoft®, Red Hat®, and SAS®. Visit www.platform.com.

World Headquarters
Platform Computing Corporation
3760 14th Avenue
Markham, Ontario
Canada L3R 3T7
Tel: +1 905 948 8448
Fax: +1 905 948 9975
Toll-free Tel: 1 877 528 3676
info@platform.com

Sales - Headquarters
Toll-free Tel: 1 877 710 4477
Tel: +1 905 948 8448

North America
New York: +1 212 888 6270
San Jose: +1 408 392 4900

Europe
Bramley: +44 (0) 1256 883756
London: +44 (0) 20 3206 1470
Paris: +33 (0) 1 41 10 09 20
Düsseldorf: +49 2102 61039 0
info-europe@platform.com

Asia-Pacific
Beijing: +86 10 82276000
Xi'an: +86 029 87607400
asia@platform.com
Tokyo: +81(0)3 6302 2901
info-japan@platform.com
Singapore: +65 6307 6590
wliaw@platform.com