HP XC Cluster Product Summary

Introduction

Once found primarily in university or scientific research settings, the Linux compute cluster has become a preferred production technology for the enterprise because of its openness, flexibility, performance, low cost, and reliability. In fact, compute clusters are fast replacing the traditional shared-memory parallel systems used in complex design, simulation, and modeling applications. Multiple servers in a cluster can run portions of the application in parallel, dramatically increasing performance as more compute nodes are added. Using clusters allows scientists and engineers to scale out, achieve higher performance, advance computing technology, and strive for bleeding-edge results—all with an affordable price tag.

However, there are numerous obstacles that hinder the broad deployment and use of Linux clusters. Integration and support can be complex. Applications may not be available. Access to data can be slowed due to interconnect bottlenecks. Typical user assembly of Linux cluster components from multiple sources requires time-consuming and costly integration, testing, and porting of target software stack, drivers, and libraries. Pre-integrated, yet incomplete cluster solutions require the user to independently add and validate key components. Limited choices of platform components often forces users to select solutions that do not fully meet their needs.

HP, the leader in HPC clustering technology, has solved these challenges and provides high performance computing (HPC) users with an easier way to increase performance and lower costs. The HP XC Cluster now offers even more power and choice for Linux clusters. The HP XC Cluster delivers the simplicity of single system administration combined with robust, industrial-grade Linux solutions. The availability of a supported, complete, and flexible solution removes the typical risks for organizations that require a solid, production ready Linux cluster resource.

This white paper describes the HP XC Cluster, a family of integrated, scalable Linux clusters that combine XC System Software V2.1 with defined implementations of HP Cluster Platforms. A general overview of the HP XC Cluster is first discussed, providing basic system descriptions, benefits, and components. More detailed product information is then explained platform components, software, and services. Concluding the paper, we present a comprehensive portrayal of HP's leadership in HPC and Linux technology.

System Overview

The HP XC Cluster is an industrial-grade, production-ready system that represents the next generation of cluster system architecture for high-performance technical computing. HP XC Clusters provide a comprehensive, supported solution for serial and parallel applications, complemented by a robust application development environment and extensive ISV application portfolio.

The design of the HP XC System architecture and the selection of key components support deployment of the HP XC Cluster as a shared HPC utility: a manageable and adaptive production resource capable of supporting a dynamic, workload environment with multiple users and multiple applications, across multiple departments. The HP XC Cluster environment provides high system utilization and a comprehensive, manageable central resource. The cluster environment features comprehensive support, ease of administration, flexible and adaptive system architecture, resource management tools, breadth of applications and workload type, and high-performance global I/O capability.

The HP XC Cluster can scale well beyond that of conventional symmetric multiprocessing (SMP) systems while retaining the manageability, usability, and reliability required by production sites. The HP XC Cluster is designed to deliver improved overall system utilization and lower cost of ownership when compared with alternative solutions.

While capable of supporting supercomputing-scale workloads, the XC Cluster provides a simple, easily managed and maintained environment that supports deployment in small departmental sites, reducing the administrative burden often associated with clusters.

The XC System Software is based on a standard Linux distribution combined with several open source packages. This open source base combined with technology from Hewlett-Packard and its partners achieves a powerful and complete solution for development, execution, and management of simultaneous parallel and serial applications.

The HP XC Cluster environment transforms a collection of servers into a single production system with unified scheduling and a single point of administration and control. The HP XC System provides users with enhanced ease-of-use with integration of leadership resource management, scheduling and job launch technology from Platform Computing and the open source-based Simple Linux Utility for Resource Management (SLURM). Access to a broad applications catalog is provided by preserving the underlying Linux API/ABI and integration of HP-MPI to provide an interconnect-neutral (but MPICH compatible) application MPI library.

With client support for the open source Lustre file system, HP XC Clusters provide high performance, highly available, global I/O capability. HP's leadership implementation of Lustre, HP StorageWorks Scalable File Share (SFS), fully supports HP XC Clusters. HP XC System Software also supports NFS, including both server and client functionality.

The XC System Software is available on the HP family of integrated clusters systems, the HP Cluster Platforms. These platforms are customer-ready solutions consisting of a broad choice of processors, interconnects, and middleware that enable rapid and confident deployment of HPC clusters. The clusters are built to uniform, worldwide specifications and are fully integrated with HP warranty and support. Available in configurations from 5 nodes to 512 nodes, the HP Cluster Platforms also includes packaging options for density or expandability. The HP Cluster Platforms offer multiple operating systems and middleware options. In addition to support for these standard HP cluster offerings, the HP XC System software can be delivered on customized hardware configurations assembled with similar base components.

The HP XC Cluster features HP Integrity and ProLiant servers and best-in-class, high-performance, industry-standard components. The following hardware platforms are cluster solutions supported by Version 2 of HP XC System Software:

Figure 1. Summary of platforms currently offered with the HP XC System Software

	Processor	Currently offered cluster nodes	Cluster interconnects
HP Cluster Platform 3000	64-bit Intel® Xeon™ processor	ProLiant DL140, ProLiant DL360, ProLiant DL380 (head node)	InfiniBand; Myrinet; Gigabit Ethernet
HP Cluster Platform 4000	AMD Opteron™	ProLiant DL145; ProLiant DL385 (head node); ProLiant DL585	Quadrics, InfiniBand, Myrinet, and Gigabit Ethernet
HP Cluster Platform 6000	Intel® Itanium® 2 processor	Integrity rx1620 and rx2620	Quadrics, InfiniBand, and Gigabit Ethernet

Key features

- Scalability and manageability, with support for industry-standard hardware platforms ranging in size from 5 to 512 nodes1
- Delivered as a comprehensive, integrated, supported offering
- Access to extensive catalog of community developed and commercial application software
- Price/performance that leverages latest industry-standard technologies and performance enhancements from HP and its partners
- Support for high performance, global I/O with HP StorageWorks SFS
- State-of-the-art workload management (parallel and serial jobs, multiple applications)
- Support for high-bandwidth, low-latency interconnects

HP expertise in offering total solution and system support on Linux clusters is leveraged throughout the delivery and deployment phases. The HP XC Clusters are staged and assembled in three steps. First, the system is integrated and tested on the factory floor. Then, the interconnect and cabinets are reconfigured for transport and shipped. When the system arrives onsite, HP field service engineers take the lead in finishing the hardware installation. The field engineers handle cabinet placement and perform the onsite cabling of the high-speed interconnect, an Ethernet-based administrative network, and power cables, when required. Third, HP's skilled onsite consultants are available to perform software startup and ongoing system administration activities. The tri-team work of our factory specialists, onsite field engineers, and onsite consultants facilitates a seamless implementation.

Simple, complete solution

Mirroring their Linux and standards-based technology foundations, Linux clusters typically are assembled as a collection of components from multiple developers, either by an integrator, OEM, or the user. When assembled by an integrator, the resulting deployment may lag the pace of open source and standards development, especially as proprietary technologies are integrated. Moreover, many cluster solutions are incomplete, requiring the user to obtain the operating system and to add functionality such as job scheduling. Additionally, support for layered applications and critical development tools are uncertain. When assembled by a user, the mix of

[1] Larger configurations are supported as custom configurations

components often requires integration and testing of the target software stack and porting of drivers and libraries (requiring initial investment and on-going maintenance). Also, a custom implementation can preclude the use of tools and applications that may require porting or other modification to support the custom design. The lack of a supported and structured cluster reference design presents risks for users who require a solid, production resource.

The HP XC Cluster addresses these shortfalls. The XC System Software includes, within one single environment, a full Linux operating system, cluster management capabilities for installation, on-going system administration, robust resource management and scheduling, and integrated HP-MPI and HP Math libraries. The software has been tuned to specifics of HP hardware platforms and leverages the console and system monitoring features of the HP ProLiant and Integrity servers. The software has been thoroughly tested on HP Cluster Platforms and validated with a broad range of cluster and HPTC development tools. Applications from the leading developers of software for CAE, EDA, Life and Material Sciences, and other disciplines are available on XC Clusters V2. HP supports the complete offering, providing a single vendor interface for all hardware and software within the XC Cluster.

Flexibility

To support the rapid pace of technology development, especially with regard to core processor and networking components, the HP XC offers structured flexibility in both hardware and software perspectives. Structured flexibility provides support for new technologies within a defined platform, as well as the capability to grow and/or modify components over time.

For software, a modular design enables fast transition to and validation of new packages and updates. Moreover, the XC System Software is an implementation of an open source reference architecture that is used by HP as a base for advanced technology development and custom deployments. These deployments provide a test bed for new technologies that become candidates for inclusion in future HP XC software releases.

For hardware, the HP XC Cluster is implemented on the HP Cluster Platform, which provides a choice of packaging styles (high density or expandable, modular design), and a node count range from 5 to 512 nodes. In a specified fabric of high-speed interconnects and an Ethernet-based administrative network developed for modular arrays of HP cabinetry, HP Cluster Platforms currently support a choice of Intel® Itanium®, AMD Opteron™ and 64-bit Intel® Xeon™ processors.

Areas of design

The design of the XC Cluster was developed to accommodate the more diverse system configurations and user workloads becoming common in HPTC production environments and to support the required system configuration, system administration, workload management, and system resiliency.

Specific features of this system design include:

- Optimized application performance
- Scalable cluster management
- Rapid and automated system initialization
- Robust, advanced resource management
- Utilization and co-development of open source cluster technologies
- Support for industry-standard hardware platforms and software interfaces
- Support for global file system with high performance I/O

These areas and their user benefits are discussed in the following sections.

Optimized application performance

The HP XC Cluster is designed to deliver leadership cluster performance and to support high user workload throughput. At the resource provisioning level, the SLURM offers highly reliable, efficient, and available cluster management and resource scheduling, at scale. The combination of LSF and SLURM supports scheduling of parallel and serial jobs and enhances overall job throughput by matching available resources to job requirements.

High performance is also enabled through integration of HP-MPI V2.1. HP-MPI is a high performance, robust, high-quality, native MPI implementation and provides optimized point-to-point and collective communication routines.

Local storage access is supported for application file I/O. In addition to the standard NFS support, the HP XC Cluster also supports Lustre and StorageWorks Scalable File Share (SFS), HP's high performance and high availability implementation of Lustre. With the use of a separate interconnect for administration and I/O, user application traffic can be isolated from administrative operations. With this separation, application I/O performance and process communication can be made more predictable while still enabling administrative operations to proceed. Alternatively, the XC System administrator may route I/O across the administrative network so that the I/O has less impact on the synchronicity of parallel applications.

Scalable cluster management

The HP XC System Software provides single-point of control for installation, node power management, monitoring, and general systems administration.

The HP XC Cluster includes the following scalable cluster management capabilities:

- *Console control*: remote power management and on-going health monitoring that leverages the embedded system management capabilities for the ProLiant and Integrity servers;
- *Configuration and Management Database*: Maintains network topology and node information;
- *Extensive monitoring suite*: Nagios, a user-friendly system and network monitoring application, watches hosts and services specified by the administrator and provides alerts and event handlers as specified. SuperMon, a highly scalable, high-speed cluster monitoring system, provides all required node statistics to the Nagios subsystem.
- *Pdsh,* a multithreaded remote secure shell client, is used to execute commands on multiple remote hosts in parallel. In addition to support for typical shell commands, pdsh can be used as a tool to repair and/or examine nodes outside of normal functions.

Rapid and automated system initialization

The HP XC System Software installation is designed to provide the maximum amount of automation possible while allowing for high levels of flexibility. Software installation is done first on the cluster's head node. Next, HP XC software is used to collect node information during a topology discovery stage. Using a simple user interface, the administrator assigns roles for various nodes (such as log-in or I/O) as required. Each node is booted over the network with an autoinstall kernel, and then a Base Image is distributed, along with node-specific configuration information. The popular and robust SystemImager open source software is employed for installation.

Robust and advanced resource management

LSF HPC from Platform Computing Corporation is the premier resource manager on the market today, and is deployed within the HP XC Cluster for managing batch and interactive jobs. In addition to launching jobs, LSF provides extensive job management and information capabilities. LSF offers a rich set of scheduling policies for scheduling and prioritizing jobs, based on combinations of static and dynamic system attributes. An extremely reliable and richly

configurable network-based queuing system provides maximum flexibility and central administration of all work in the cluster. As noted above, LSF is integrated with an underlying resource manager, SLURM.

Customers requiring alternative resource management (for instance, mandated corporate standards) can install and utilize other schedulers to submit jobs and manage queues, as the XC cluster supports Linux conventions.

Utilization and co-development of open source technologies

HP XC System Software is an HP-supported and defined implementation of an open source architecture for high performance clusters. By utilizing open source components, users are able to benefit from the experience of multiple sites, which have fostered the highly capable and functional technologies incorporated in the HP XC System Software. Open source technologies develop at a pace rarely matched by proprietary solutions. By choosing to build upon and integrate with, rather than compete with, the established cluster technologies that comprise the HP XC Cluster solution, HP and its collaborative partners extend the functionality of clusters, instead of merely replicating it.

Support for industry-standard hardware platforms and software interfaces

HP is the worldwide leader for industry-standard servers with ProLiant and Integrity systems and the number one supplier of Linux servers. The HP XC Cluster utilizes these servers as cluster nodes and benefits from HP's industry leadership expertise and from partnerships with major suppliers of networking technologies. HP systems support Linux as a tier one operating system, with full qualification and testing of Linux with our partners, and incorporating server manageability capabilities for Linux. The HP XC System Software continues that commitment to Linux and standards, with support for Linux application interfaces and standard protocols.

Support for global file system with high performance I/O

The HP XC System Software supports NFS 3 but is also enabled for Lustre client services for high-performance and high-availability file I/O. Lustre is an open, high-performance, highly scalable file system engineered to provide extremely fast, scalable, and reliable I/O for Linux clusters. HP augments the Lustre software with unique usability features on StorageWorks hardware. Using rigorous product development and testing methods, HP delivers a fully supported, integrated file system, HP SFS. The combination of HP XC Clusters and SFS provides a powerful match of computing capability and high performance I/O that overcomes the traditional I/O bottlenecks that can be encountered with Linux clusters.

System architecture

The HP XC Cluster architecture is designed to provide a flexible infrastructure to perform the administrative tasks of a production-grade cluster, while enabling optimal application performance.

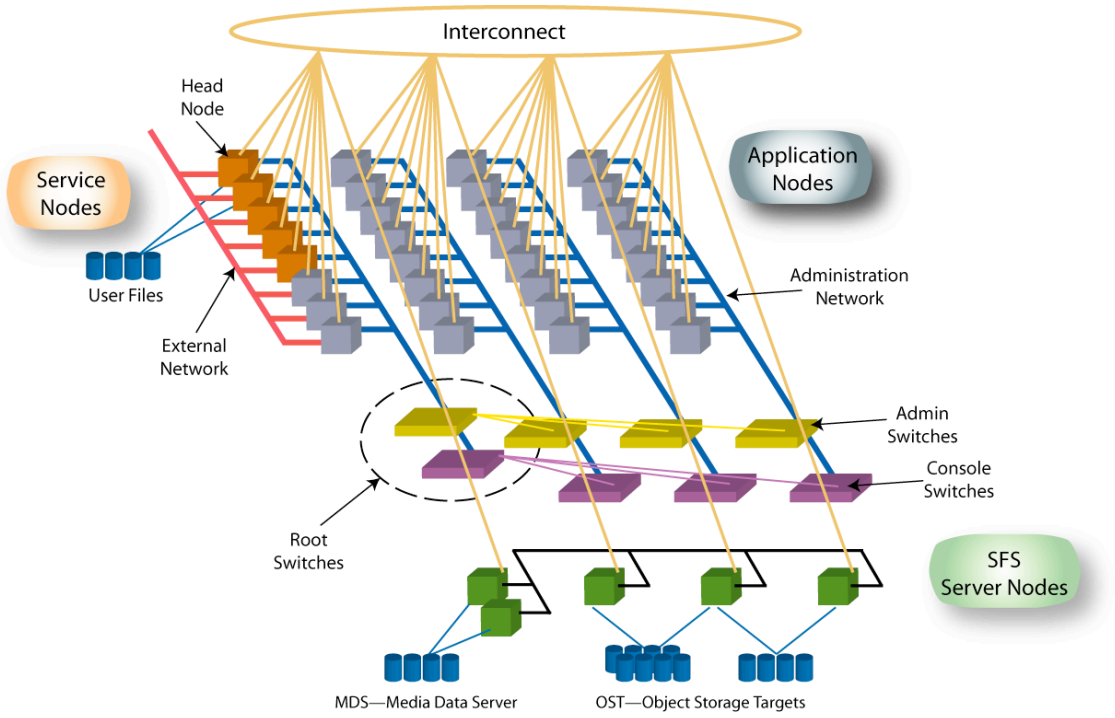
While all nodes can be deployed for computational jobs, specific nodes are enabled to perform non-computational cluster tasks (e.g., I/O). These enabled nodes, referred to as Service Nodes, may have external connections to enable I/O and login services. For large clusters, service nodes may be assigned tasks for managing groups of nodes and for filtering and aggregating system information. One service node serves as the cluster control (or head) node and is used for initial software installation. The head node is capable of providing system services through hosting daemons such as the SLURM and LSF masters. It is also possible to configure the HP XC Cluster so that system services are distributed across several nodes, rather than centralized in a single head node. This distribution of services may be selected for reasons of systems availability, scalability, and system load. However, all the non-computational functions can be delivered by a single head node.

To support cluster administration on the service nodes, as well as to facilitate the automated discovery and initialization process, the HP XC System Architecture has defined the network fabric (and provided specific manufacturing and integration specifications for cabling). The network fabric is comprised of an administrative network and an HP XC System Interconnect for computational and I/O traffic. The HP XC System Interconnect is a private, flat network with each cluster node connected to the interconnect switch.

The administration network has two branches: one, based on Gigabit Ethernet, handles the administration and operations traffic for the cluster and can be used as well to route I/O traffic. The other branch of the administrative network is based on 10/100 Ethernet and is the Console Branch, with direct connection to the embedded management processors within each node. This console branch is used for system power on and power off and for collecting and monitoring a server's health attributes.

The same Gigabit Ethernet network can be used for the administration network and the System Interconnect. This is recommended, however, only for smaller clusters (32 compute nodes or less).

Figure 1. High-level XC system architecture



Service nodes typically will be co-located in the same rack(s), have external connections, and can be directly connected into two root switches within the administration network. Non-service (application) nodes can be assembled into compute building blocks (if the HP Compute Platform Modular Packaging Design is selected), in which case these nodes will be connected to the administrative network through a topology of branch switches.

The architecture shares the HP XC System Interconnect with the HP StorageWorks SFS servers. Each HP XC Cluster node is enabled as an HP SFS client and can directly access the HP SFS servers.

It is important to note that all cluster nodes, both service nodes and application nodes, can be used for computational jobs to run applications. The specific terminology is used to help define which cluster nodes have been configured to support multiple administration tasks, in addition to computational tasks.

HP Cluster Platforms

The XC System Architecture is a supported implementation for HP's new family of packaged clusters, the HP Cluster Platform. The HP Cluster Platform provides the customer with range of options, such as the number, type, and system configuration of utility nodes and application nodes, a choice of HP XC System Interconnects, system storage, and external connections.

The **HP Cluster Platform 6000** features HP Integrity servers with Intel Itanium 2 processors and a choice of Quadrics, InfiniBand, or Gigabit Ethernet interconnects. Co-developed by HP and Intel, the revolutionary Intel Itanium processor architecture reduces platform costs and enables higher performance and scalability. The Itanium 2 processors include extensive instruction-level parallelism and chip sets that offer massive on-chip resources (large number of registers, and large L2 cache on die), thereby reducing memory operations for complex workloads. Integrity servers deliver superior 64-bit performance and offer optimal performance in complex technical applications, such as seismic analysis, molecular modeling, and structural mechanics. The HP XC Systems based on the HP Cluster Platform 6000 utilize the HP Integrity rx1620 and rx2620 servers as service and application nodes.

The **HP Cluster Platform 4000** features ProLiant servers with AMD Opteron™ processors, a choice of Myrinet or Ethernet interconnects. The HP Cluster Platform 4000 is an excellent performer in environments where the workload is a mix of 32-bit and 64-bit applications, such as electronic design automation, life sciences, and geosciences. Applications that are compute intensive and run well on clusters of IA-32-based systems, such as many computational fluid dynamics applications, will also see advantages with AMD Opteron™ processor-based clusters. The HP XC Systems deployed on the HP Cluster Platform 4000 currently are available with the ProLiant DL385 for the service node and the ProLiant DL145 and DL585 as the application nodes.

The **HP Cluster Platform 3000** features the world's best selling ProLiant DL140, DL360 and DL380 servers available with the latest 64-bit Intel Xeon processors for 64-bit applications. New Intel® EM64T technology allows users to run either 32-bit or 64-bit operating systems on the servers while improving the memory address space. This means businesses will be able to move memory-starved 32-bit applications and memory-intensive applications onto the servers. To support visualization capability, the Cluster Platform is also being extended to support HP workstations, beginning with the Intel Xeon processor-based 8200.

These systems are integrated and mounted in HP 10000 series 42U cabinets with power distribution units (PDUs) for power management and fans for adequate cooling in a typical computer environment.

HP Cluster Platforms are fully configured, integrated, and tested at the factory prior to shipment and are disassembled into individual cabinets with shock pallets for shipment. This process ensures that the system is complete and allows for quick, easy installation of the equipment at the customer's site.

The system can be expanded, upon request, to increase the number of compute nodes in the initial configuration or modify its memory, disk, and network configuration to meet the customer's application requirements. A storage subsystem design can also be provided upon request.

Packaging styles (dense or modular configurations)

The HP Cluster Platforms are offered in two packaging styles: bounded systems in dense configurations (smallest number of racks possible) or module solutions that can be expanded to 512 interconnected nodes.

The bounded configurations for the Cluster Platform are optimized around the available InfiniBand, Myrinet-2000 and Quadrics QsNet II interconnect sizes and range from bounded systems up to 32-nodes, 64-nodes, and 128-nodes. Smaller 8 port and 16 port configurations are available by request. For Cluster Platforms with Gigabit Ethernet as the system interconnect, the 24 port ProCurve 2824 and the 48 port ProCurve 2848 are utilized, along with Myrinet 64-port and 128-port switches. Configurations can be assembled around these interconnects, limited by the physical dimensions of the racks and the selected nodes.

The modular HP Cluster Platforms are built using a "building block" approach, separating the compute resources from the interconnect and console tasks and the storage. Each building block is an individual rack, integrated with the specified components. There are four types of racks that can be regarded as building blocks:

- Compute Building Blocks (CBB): a rack providing the majority of nodes in a system (application nodes)
- Interconnect Building Blocks (IBB): a rack providing the interconnect switches
- Utility Building Block (UBB): a rack primarily containing components to support the cluster infrastructure, including the control node and other service nodes. In addition, to fully utilize the rack infrastructure, application nodes can be integrated into the UBB. (Any node placed in the UBB or UBR is considered a 'utility node'.)
- Utility Expansion Rack (UXR): Supports requirements for additional service nodes beyond the capacity of the UBB (and as with the UBB, can also house application nodes to allow full utilization of the rack.)

Compute Building Blocks (CBB) are assembled with application nodes, network switches for connection of the nodes to the administration network, and PDUs. The number of nodes supported in the rack will depend upon the physical size of the node. The CBB can accommodate up to 38 1U compute nodes and up to 19 2U compute nodes.

The **Interconnect Building Block (IBB)** consists of the rack, necessary power distribution units, and slide kits with the high-speed interconnect switches (chassis and switch cards). The cluster nodes in the CBB and UBB racks are connected to the switch cards in the IBB rack. IBBs are not needed for configurations that require only one interconnect switch, as that can be integrated into the Utility Building Block. The IBB is typically positioned in the center of the cluster to facilitate cabling between cabinets.

The **Utility Building Block (UBB)** consists of at least one service node (serving as the cluster's head node), power distribution units, optional rackmount KVM (keyboard/video monitor), optional MSA1000 storage subsystem, administration network switches, and as noted above, the cluster interconnect for clusters that only require one system interconnect. The **Utility Expansion Rack** may be fully populated with service and application nodes and does not require administration switches (as it can utilize the switches in the UBB).

HP XC Cluster product information

As noted in the System Overview section, the HP XC Cluster is delivered as an integrated hardware and software solution supported by HP Services. The following sections provide additional information on the hardware components of the HP Cluster Platform and discuss key elements of the HP XC System Software.

HP Cluster Platform Components

HP Cluster Platforms are built using the best-in-class, high-performance, industry-standard components offered on the market for cluster solutions.

Cluster nodes

Each node is a HP rack-optimized server and can be tailored to specific customer requirements for memory and internal storage. Standard configurations for the HP Cluster Platforms can support from 5 to 512 two-processor application nodes. (Note that support for more than 512 nodes will be standard post version 2 and is available as a custom offering.)

Detailed product descriptions on systems are available at www.hp.com and in Reference Guides for the HP Cluster Platforms.

High-speed interconnect network

A high-speed interconnect supporting a private network within the HP XC Cluster is a key ingredient used for application traffic, and as required, used to support NFS or SFS traffic. Within the HP XC Cluster, all nodes have a direct connection to the high-speed switch. The Itanium 2–based HP Cluster 6000 clusters configurations utilize Quadrics QsNet II switches, with Quadrics QM500 PCI-X Network Adapter. The ProLiant-based HP Cluster 4000 clusters utilize Myricom Myrinet 2000 switches and the Myrinet-2000 PCI module Rev D or Rev E adapters. Both HP Cluster Platforms are also available with Voltaire InfiniBand or Gigabit Ethernet as the system interconnect.

Quadrics

Quadrics' scalable, high-performance elements combine the QsNet interconnect with standard microprocessor and operating system platforms. QsNet enables large clusters of these processing units to be connected together. With extremely high bandwidth and ultra-low latency, it is a crucial component for building scalable systems. The current generation, QsNet^{II}, is designed for high performance PCI-X interfaces, using parallel copper interconnects to deliver over 900Mbytes/s of user space to user space bandwidth. The network adapters are built using Quadric's Elan4 network chips and support some of the highest performance currently available in cluster networking systems.

The components of the QsNet^{II} switch network can be combined in a 'fat tree' network that scales, in powers of 4, up to many thousands of nodes. Fat tree networks have many properties that make them attractive for high performance switch fabric. Most importantly, the bisectional bandwidth of the network scales linearly with growth in network size. The topology is also inherently highly resilient with large amounts of redundancy in the higher levels of the switch.

As noted earlier, bounded HP Cluster Platforms are available with a range of Quadrics switches, including 32-port, 64-port, and 128-port switches. Smaller configurations are available by request using the smaller 8 and 16-port switches.

Myrinet

The Myricom Myrinet 2000 switch provides a high-speed, low-latency interconnect for the cluster nodes, and is available with three different types of chassis:

- 9U 128-port maximum switch (for all modular solutions and 3 to 4 cabinet dense solutions)
- 5U 64-port maximum switch (for two cabinet dense solutions only)
- 3U 32-port maximum switch (for one cabinet dense solutions only)

These chassis house and provide power for fiber-optic switch cards and a monitor card. Each switch port is connected to a PCI-X adapter card in each node via a fiber-optic cable. For modular configurations of up to128 server nodes, a Myrinet 2000 128-port switch may be selected to easily accommodate future expansion. The Myricom switch architecture can be expanded to scale and support cluster configurations with more than 128 nodes. A monitor card enables the system administrator to monitor the switch. HP Cluster Platforms support two types of PCI-X adapters, the single port RevD adapter, and the dual-port RevE adapter

InfiniBand

The InfiniBand interconnect expands the options available for HP Cluster Platforms. The HP Cluster Platforms feature InfiniBand switches from Voltaire. These interconnects combine open standards with high bandwidth, low latency performance. The product is available in two switch sizes and can be configured to support up to 512 nodes in a federated design:

- 288 10 Gbps port 14U enclosure provides full bisectional bandwidth for large clusters;
- 24 10 Gbps port 1U enclosure provides economical small clusters and configuration flexibility.

The switch enclosure can be connected to the clustered nodes using either PCI-X or PCI-E adapters.

Gigabit Ethernet Interconnect

For Cluster Platforms with Gigabit Ethernet as the system interconnect, HP's ProCurve family of switches are deployed either as a single cluster switch or in tiered hierarchy utilizing federations of larger ProCurve switches, configured with smaller switches directly connected to nodes.

For smaller configurations and as node-level switches, the 24 port HP ProCurve 2824 and the 48 port ProCurve 2848 are utilized.

The HP ProCurve 5308xl switch can be used as a standalone switch for configurations up to 32 nodes or in tree topologies with configurations up to 128 nodes

For larger configurations, higher bandwidth is available with topologies utilizing the ProCurve 930xM switch and multiple ProCurve 2848 node level switches. For full bisectional bandwidth, Myricom 128-port switches equipped with Gigabit Ethernet line cards are available.

Ethernet administration network

An administrative network handles management and installation tasks and isolates administration traffic from application activity. Application nodes are connected to the administration network via Ethernet switches local to the cabinet they reside in. This minimizes inter-cabinet cabling and forms a local admin subnet. The management branch of the administration network is based on HP ProCurve 2848 and 2824 Gigabit Ethernet switches. The console branch utilizes HP ProCurve 2650 Ethernet 10/100 network switches.

Storage

Each node must have an internal disk drive. Each node may optionally be attached to an HP Storage Works Modular Smart Array 1000 (MSA1000). In addition, the HP XC System supports

high performance file I/O with the HP SFS. The HP XC System can be connected to HP SFS over external Ethernet connections or over the XC System Interconnect. When configuring the HP XC System to connect to the HP SFS using the HP XC System Interconnect, there must be enough switch ports available to accommodate the number of nodes in the HP XC System, as well as the number of nodes required by the HP SFS (see the HP Storage Works SFS Reference Guide). The switches for the combination HP XC/HP SFS system will be included as part of the Cluster Platform 4000 or Cluster Platform 6000 configuration, rather than as part of the HP SFS.

Cabinet

The system components are integrated and mounted in the HP 10000 series 42U cabinets with power distribution units (PDUs) for power management and fans for adequate cooling in a typical computer environment.

A rack-mounted foldout flat-panel display with keyboard and mouse is provided for direct local system access. The display and keyboard/mouse are connected to the video and keyboard/mouse interface of the service administration node. The system can also be accessed with an external connection via the administration switch.

The cabinets can accommodate up to 42U of vertical mounting space and will provide mounting space, AC power distribution, and fan cooling for the application and service nodes, switches, and storage. The system nodes and Ethernet switches are configured to package the application and service nodes into compute cabinets. The high-speed interconnect switches are packaged into their own cabinet to facilitate installation and cabling.

User documentation

The HP Cluster Platform documentation set includes available user documentation of the system subassemblies (e.g., rx2620 servers, ProLiant servers, Myrinet switches, monitors, etc.) In addition, the following Cluster Platform-specific documentation is provided:

- HP Cluster Platform System Overview and Site Planning Guide
- Installation and Operations Guide

Packaging

Cabinets loaded with equipment will be shipped to customer sites using proper material (cushioned pallet, crated, etc.)

Software

The HP XC System Software includes HP-MPI and other HP proprietary and open source software that supplies the infrastructure for optimal interconnect performance. In addition, HP XC System Software includes a Linux distribution compatible with Red Hat Enterprise Linux 3.0 Update 4, and software that provides users and system administrators with single-system attributes for managing and using HP XC System resources. Also, HP XC System Software includes technology from Platform Computing Inc. for job management and optimized throughput of parallel applications. HP will provide patches and updates to the HP XC System Software as part of the standard HP support process.

Version 2 of HP XC System Software was released in November 2004. The next release, which is the primary focus of this whitepaper, is Version 2.1 available June 2005. HP will provide patches and updates to HP XC System Software as part of the standard HP support process.

The System Overview outlined specific features and functionality provided by the HP XC System Software. Additional information on the software components used within the HP XC System, as well as additional capabilities, is contained in the following sections.

XC Reference Architecture for High Performance Clusters

The HP XC System Software is a supported and integrated implementation of a reference architecture for HPTC, defined by HP and based on its experience in delivering and developing cluster solutions. The HP XC Cluster architecture utilizes premier components from leading suppliers of cluster technologies. In selecting these components, several key criteria have been employed:

- Capability—the components are capable of providing, in total, the features required for a complete high performance cluster solution
- Provability—the components have been widely used and proven in production facilities
- Modularity—the components are capable of being integrated into a single management and deployment environment with uniform semantics, and without multiple conflicting methods for job management or execution
- Portability—the components can be combined to produce a whole system that can assure application portability across the platform family, and can adopt to future technologies

Table 1. Definition of XC Reference Architecture for Clusters

Function	Implementation
Distribution / Kernel	Red Hat 3.0 update 4/2.4.21 kernel compatible
Cluster Alias	LVS—Linux Virtual Server
Outbound Net	NAT—Linux Network Address Translation
Batch	Platform LSF V6.0
Launch	SLURM - Simple Linux Utility for Resource Management
System Files Management	SystemImager +limited config tools +database
Console	Console Management Facility (CMF)
Operations	Pdsh, syslog-ng
Monitoring	Nagios [with plug ins]—Open source service and networking monitoring program SuperMon - High performance cluster monitoring
MPI	HP-MPI v2.1

User view

HP XC System Software provides end users with a single system view of the underlying hardware:

- Single login
- Integrated program development environment
- Integrated job submission system
- Single file system namespace providing coherent access to user data

The Linux Virtual Server (LVS) is used to present a single host name to users so that they see the HP XC System as a single system for login access to prepare and launch high performance applications. LVS directs each user login session to one of several potential login nodes. The user is presented with a Linux system command interface. A user submits an application to the queuing system and receives the output when the job completes. MPI is the primary programming model used for parallel applications.

Workload and resource management

Effective resource management utilizes an efficient queue management system that enables jobs, both batch and interactive, to be presented to the system based on site-defined policies and priorities. It also utilizes a scheduling system that can make the most efficient use of the hardware resources available.

From the view of users, the primary interface for job submittal is Platform Computing's LSF (Version 6). Platform LSF is used to schedule parallel workloads, such as those used in solving grand challenge problems, while simultaneously scheduling serial workloads to utilize remaining compute resources. Its policy driven scheduler enables full utilization of high performance network interconnects available on clustered systems and supercomputers. With more than 1,600 customers and an extensive library of third-party application integrations, Platform LSF is the leading commercial solution for production-quality workload management and is deployed on several of the top 10 supercomputers.

LSF enables scheduling and prioritizing jobs based on combinations of static and dynamic system attributes, such as the number of processors or job attributes, or time limits and user attributes including uid, gid, account id, project id, and priority-based quotas for size and time allocation. LSF policies include first-come first-served, fairshare, hierarchical fairshare, deadline-constrained, exclusive, adaptive dispatch, backfill, preemptive, and job slot reservation. The extremely reliable network-based queuing mechanism provides maximum flexibility and central administration of all work in the cluster while ensuring that all work submitted to the HP XC System runs to completion. An extensible LSF checkpoint interface supports application-level check-pointing on the HP XC System, when supported by the application. In addition, LSF tracks all usage on the HP XC System and provides an accounting log for system utilization and charge back analysis.

To provide efficient resource management and utilization, XC System Software has integrated SLURM. SLURM is an open source, fault-tolerant, highly scalable cluster management and job scheduling system for Linux clusters. Components include machine status, partition management, job management, scheduling, and stream copy modules. SLURM allows the administrator to treat the total set of CPUS, providing by all nodes, as a single pool for computational jobs. Alternatively, this set can be partitioned for use by LSF. SLURM provides common underpinning for LSF and other high level schedulers and has plug-in support for the Maui scheduler.

SLURM can run reliably and efficiently on Linux clusters as large as several thousand nodes. It is well suited for large-scale, high-performance computing environments, and its design avoids known weaknesses (such as inflexibility or fault intolerance) in available commercial resource management products. The integration of LSF with SLURM creates a high-quality, comprehensive workload and resource management system on the HP XC Cluster.

When required, users can install other resource managers (and opt to turn off SLURM), as required for support of application-specific implementations or to comply with mandated policies.

Configuration and administration

The configuration tools support all of the tasks needed for system initialization, routine (re)configuration, and boot-up. Administrative tools are available for control of the system as it operates in production. The system can detect a node failure. Detected node failures include a node not responding, fan failure, and temperature out of range.

The HP XC Cluster software exploits the integrated remote console management capability of both the HP ProLiant and HP Integrity servers. The remote management ports support power on/off and system boot via the console branch of the administration network.

As noted earlier, the HP XC System supports rapid and automated system installation, using a simple user interface. SystemImager is utilized to distribute the HP XC Base Image, and node-specific configuration information is maintained in a configuration database.

System monitoring

The HP XC System monitoring is built upon the Nagios system, a system and network monitoring application. It uses a tiered setup to allow for varying levels of granularity in logging and monitoring information. Nodes are monitored using Nagios. Nagios watches hosts and services specified by the administrator and provides alerts when problems emerge or when problems subside. Some key features of Nagios include:

- Monitoring of network services such as SMTP
- Monitoring of host resources such as processor load
- Simple plug-in design that allows administrators to easily develop their own service checks
- Parallelized service checks
- Contact notifications when service or host problems occur and get resolved
- Ability to define event handlers to be run during service or host events for proactive problem resolution
- Support for implementing redundant monitoring hosts
- Optional web interface for viewing current network status, notification and problem history, and log file

Nagios is used to report information collected via the infrastructure of SuperMon. The data collected by SuperMon includes both system performance metrics, such as CPU and memory utilization, as well as environmental data, such as fan, temperature, and power-supply status. This data is collected on a regular basis via a call to the SuperMon daemon on the head node, which communicates with daemons on the other nodes.

Modules

The HP XC System Software includes the HP XC modules package (not to be confused with kernel modules), which provides for the dynamic modification of a user's environment via module files. The module command processes the information contained in the module file to alter or set the shell environment variables, such as PATH and MANPATH. Users can add their own module files to further modify their environment. The module files that are automatically installed and are specific to HP XC System Software include Intel-compilers, mkl, and mpi.

MPI

A key obstacle for widespread adoption of Linux clusters has been the difficulty for commercial ISVs to release MPI versions of their applications. The flexibility and breadth of options for clusters has created an application Tower of Babel, with developers selecting different versions of Linux, different implementations of MPICH, different compilers, etc. A key feature of the HP XC System Software stack is the incorporation of the HP-MPI libraries and drivers, which enable software developers, including major ISVs, to develop a single set of executables across interconnects. By providing a common reference design (HP-MPI with Red Hat 3.0 based) and defining a full environment, a single cluster is able to support a mix of ISV applications. This approach has been enthusiastically endorsed by the ISV community, resulting in rapid availability of an extensive portfolio of commercial applications.

With HP XC System Software Version 2, HP-MPI V2.1 is fully integrated with SLURM resource management. HP-MPI complies with the MPI-1.2 and MPI-2 standards and is a high performance, robust, high quality, native implementation. HP-MPI enhancements provide optimized point-to-point and collective communication routines. HP-MPI takes advantage of shared memory for intra-node communication and the RDMA technology for inter-node communication whenever possible. HP-MPI supports 32- and 64-bit applications, single- and multi-threaded, and provides tools to debug and instrument MPI execution. Also offered is MPICH 1.2.5 compatibility when an application is built shared with HP-MPI. Support for Quadrics Ltd. QsNet® and GigaBit Ethernet

is included in HP-MPI on Integrity systems. Myrinet GM2.1 and GigaBit Ethernet support is included on HP ProLiant, including standard TCP/IP support. Applications built with HP-MPI are interconnect-independent; it is not necessary to recompile or relink an application in order for it to operate on a different HP XC System Interconnect.

For applications not available with HP-MPI support, alternative MPIs (such as MPICH) can be installed and run with the XC System Software. The MPI installed should support Red Hat EL 3.0, and the System Interconnect on the target cluster.

Grid Enablement

XC clusters are grid-enabled. HP tests the Globus Toolkit, the industry standard middleware for grids, on XC clusters to ensure that Globus runs well on these systems. Versions of Globus that have been tested include both version 2.4, which is currently deployed in many production grids, and version 4.0, which provides support for the new web-services standards for grids.

Serviceability

HP XC System Software includes some core components that facilitate the serviceability of the HP XC System and the HP XC System Software. For detection of potential problems relating to the HP XC System Interconnect, there are interconnect-specific diagnostic tools as well as generic tests that stress the network. An HP XC Installation and Operational Verification procedure can be used to check whether system installation has been done correctly and whether the system behaves correctly. During normal operation of the HP XC System, various facilities are provided to help detect and diagnose problems. Crash dump support is provided as a way to trace the origins of software errors (which, in some cases, could be triggered by hardware failures). The sys_check utility is able to collect data and log files to further facilitate problem analysis. Hardware failures can be anticipated or detected by environmental monitoring using Nagios and Supermon.

Availability

For increased redundancy and availability, the HP XC System may be configured with the administration and login services available on two different nodes (administration nodes). The HP XC System may also include separate I/O nodes with shared access to SAN storage. Should a node fail running one of these services (administration, login, or I/O), the service can be manually resumed on its alternate administration node. In addition, there are some key redundancy features at the application level that help to limit single points of failure. The SLURM and LSF job launch mechanisms, as well as scheduling and resource management, provide application level failover support between the two administration nodes. For example, the SLURM control daemon can be optionally configured with a backup to allow failover in the event of a node failure. If the administration node running the SLURM and LSF services fails, this failure will be detected. The services will then resume on the alternate administration node (e.g. an LSF daemon will be started on that node) with no disruption to users, or to pending and running jobs.

Software documentation

The HP XC software documentation set includes the following:

- HP XC System Software Release Notes
- HP XC System Software Hardware Preparation Guide
- HP XC System Software Installation Guide
- HP XC System Software Administration Guide
- HP XC System Software User's Guide

In addition, Platform's LSF documentation set will be included on the HP XC Documentation CD-ROM along with Reference Manuals from the open source documentation for SLURM and Nagios and a Linux Administration Handbook.

Compilers and tools

HP XC System Software can support most development and performance tools available today for Linux clusters, subject to the limitations of the tool with regard to processor types and interconnects.

HP works with the software development community to ensure that a full complement of tools are available for HP XC Cluster application development and deployment. For example, HP is working closely with Intel on development tools and compilers in support of Intel processor-based Linux clusters. Several open source and commercially available software packages have been tested and/or are supported with the HP XC System Software, and are listed in Table 2.

Table 2. Application Development Tools Supported on XC System Software V2.

Function	XC Cluster Supported development software
Debugging	Allinea Software DDT Etnus TotalView Gdb PathScale Pathdb
High Perf I/O	HP Scalable File Share Client (Lustre 1.2.x client)
Profiler and performance tools	Intel® Trace Analyzer and Collector (Intel Xeon and Itanium processors) Intel® VTune™ Performance Analyzer Oprofile Profiling tool for Linux (open source) PAPI (open source) Paratools TAU PathScale OptiPath MPI Acceleration Tools
Compilers	Intel® C++, and Fortran compilers (Intel Xeon and Itanium processors, AMD Opteron™ processors) GNU Compiler Collection—GNU C++, GCC, GNU Fortran (g77) – open source Pathscale EKOPath Compiler Suite (Intel Xeon and AMD Opteron processors) The Portland Group C, C++, Fortran compilers/Debugger/Profiler (Intel Xeon & AMD Opteron Processors) Scientific Computing Associates 's TCP Linda (Pre-processor)
Libraries and math environments	ACML—AMD Core Math Library (Intel Xeon & AMD Opteron Processors) Insightful S-Plus (Intel Xeon & AMD Opteron Processors) Intel Math Kernel Library (MKL) (Intel and AMD processors) Intel® Cluster Math Kernel Library MATLAB and Distributed Computing Toolbox and Engine from The Mathworks (Intel Xeon and AMD Opteron processors) NAG Fortran and C Libraries Quadrics SHMEM, as part of QsNet II user libraries, on Itanium processor-based systems connected with the Quadrics QsNet II switch Visual Numerics IMSL Fortran and C Libraries Wolfram Research Mathematica and GridMathematica
Support for 3rd party solutions not integrated within XC software provided by ISV or open source project	

Through support for standard interfaces and cluster practices, it is expected that many other development tools should operate in the HP XC System Software Environment.

Services

This section provides the service descriptions for the level of software maintenance and onsite hardware support.

XC warranty description

Hardware Warranty: HP Cluster Platforms carries the underlying warranty and enhanced service options for the underlying components. For instance, the HW Warranty service for the DL585 nodes features a 3 year HW support onsite with next business day onsite response. It includes parts, remote diagnosis and support, and onsite labor. For the DL145, the standard warranty is 1-year parts only. The basic warranty for major interconnects includes a 1 year HW onsite support service with next business day onsite response.

Software Warranty: The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. For example, HP warrants only that the Software media will be free of physical defects for a period of ninety (90) days from delivery. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

HP Cluster Platforms includes staging, integration, and installation. Consultant-level outside support is recommended to help address the planning and introduction of the solution into the user environment (a full range of training courses is available for an additional charge.)

Optional integration and installation is available for quotation on options not included in the HP Cluster Platform. Additionally, optional custom site planning, custom freight arrangements, VIS engineering team onsite installation of the HP XC System/HP Cluster Platform, and tailored customer configuration documentation are also available for quotation.

Optional customer services

HP has experience servicing systems, networks, storage devices, and software for HPC deployments and environments of all sizes. A wide range of purchasing options and solutions built from standards-based service modules ensure the service coverage and the level of expert assistance desired.

Optional services may be purchased at the time of the product purchase as HP Care Pack services, easy to order and easy to use service packages. Optional services may also be purchased at a later point of time as HP Contractual or Per-Event Services to extend and/or expand and/or complement the standard product warranty.

Hardware Support: Service capabilities include total lifecycle support, preventive and diagnostic services, access to a global service network, and support for an array of third-party peripherals.

Hardware Support Services offer a full range of high-quality remote and onsite hardware support options to meet specific response or repair-time objectives. These support options range from next-day or 4-hour onsite response to 6-hour call-to-repair time commitments. To accommodate specific hours of operations, coverage window options range from standard business hours (9x5), Monday—Friday 8am—5pm local time up to 24x7.

Software Support: Standard software support includes 9x5 telephone support and rights to new versions for the specified period of time (one year or three years). 24x7 software support extends the telephone support window to round-the-clock access. Standard software support includes customer access to technical resources during standard hours (see later), problem analysis, escalation management, and resolution. HP also provides unlimited access to an electronic facility that includes a knowledge database with known symptoms and solutions, software product descriptions, specifications, and technical literature. In addition, HP will also make available certain software patches, including security patches, to the HP XC System Software. With

standard software support, customers can access technical resources via telephone, electronic communications, or FAX (where available), during standard business hours on standard business days, including the hours of 8:00 am and 5:00 pm, Monday through Friday excluding HP holidays. 24x7 software support extends the access-window to 24 hours a day, from Monday through Sunday, including holidays. Business terms and conditions governing software services can be found at the HP website http://legal.hp.com/legal/files/Standard_Agreements.asp.

Software support contracts can be obtained to meet the needs of customers, including remedial technical remote support along with migration and upgrade planning and a full suite of proactive deliverables.

Mission-critical services help keep systems up and running. For high availability, performance, and security, mission-critical services include Mission-Critical Partnerships, Critical Services, and Proactive 24 Service.

Recover-All service provides priority repair or replacement of HP equipment damaged in a disaster to maintain business continuity and protect corporate assets.

Consulting and integration

Consulting results in lower total cost of ownership due to the following: quicker integration of technology into a customer's computing environment, greater reliability and utilization of the system, efficient application migration, optimal application performance, and knowledge transfer to the customer.

Each HP XC Cluster includes a Quickstart service from HP Consulting and Integration (C&I) Services, purchased as a separate but required option. Quickstart includes onsite training and consulting to enable speedy deployment and facilitate knowledge transfer.

C&I offers the following optional services for HP XC Cluster deployments:

- Onsite system software support
- Onsite application software support
- Implementation management
- Customer education

In addition, C&I Services bundles Quickstart with every system.

Why HP?

Combining the innovations and talent of four historical HPC powerhouses—Hewlett Packard, Compaq Computer Corporation, Digital Equipment Corporation and Convex Computer—into one powerful market force, HP's HPC business brings decades of expertise and a broad range of choices to customers across the globe. From blades and departmental servers, to high-end systems and clusters designed for the engineering enterprise and supercomputer centers, the HP portfolio delivers more power and choice, more scalability and manageability, more expertise and innovation, and more investment protection.

According to IDC, HP leads the industry in HPC revenue worldwide and is the market share leader in the enterprise, divisional, and departmental segments—the largest and fastest growing segments in the overall HPC market.

Committed to HPC

HP's mission is to exercise our technology leadership and market strength to satisfy our customers' most demanding compute and data-intensive requirements. We are committed to HPC as a growth business and recognize that innovations developed for HPC customers today will be used by mainstream customers tomorrow.

Our strategy focuses on achieving excellence in four areas:

- leadership performance and scalability with systems that feature the latest microprocessor technologies in scalable configurations
- partner and customer leadership (with HP Labs and collaborations with partners and customers) in key HPC technologies, including cluster interconnects, software, high performance I/O, and computational grids
- complete solutions across a choice of operating systems through an extensive portfolio of market-leading applications software from ISVs
- lasting value and investment protection through leveraging the benefits of commercial, industry standard technologies; HP's certified configurations; and worldwide services and support capabilities

HP provides solutions, not just systems. HP's expert end-to-end HPC services provide a single point of accountability ensuring successful implementation—starting with planning, continuing through migration and transition, and extending to ongoing maintenance and optimization. HP provides senior-level expertise in system's integration and administration, application's migration and optimization, system's architecture, and customer education, as well as implementation program management.

HP offers a comprehensive choice of cluster solutions for high performance computing, featuring UNIX®, Linux, and Microsoft Windows operating systems. As noted earlier, HP Cluster Platforms provide a cluster-ready platform for popular third-party and open source cluster software as well as HP XC System Software. For customers who require custom Linux solutions, HP combines its supercomputing expertise and Linux cluster technology expertise with the open source flexibility of Linux to deliver specifically tailored cluster solutions.

HP delivers on its HPC commitment with a worldwide organization including the HPC Division of the Enterprise Servers and Storage Group, regional HPC business groups in the Americas, Europe, Japan and Asia-Pacific, and the worldwide HPC practice within the Consulting and Integration Services organization. In addition, HP Labs provides focus on generating new

technologies for technical computing that will help maintain HP's leadership in HPC for many years to come.

Linux @HP

HP is strongly established as the leading Linux vendor. According to first quarter 2005 figures released recently by IDC, HP again leads the fast-growing worldwide Linux server market in factory revenue and units, HP maintained its number-one spot in the Linux server market with 27.7% market share in terms of revenue.

Linux qualification and certification on industry-standards-based ProLiant technology is complemented by HP's strong support for the Linux kernel development on Itanium processors and support for open source development in general (see hp.opensource.com).

HP's Linux operating system support spans its entire product line of servers based on Intel architecture (IA-32, Itanium processors, and 64-bit Intel Xeon processors), and AMD Opteron processors.

HP's software commitment is equally strong. HP has ported its manageability, high-availability, quality of service, Telco, and security software to Linux. In addition, HP made its Linux printer drivers available to the open source community and continues to evaluate its software portfolio, looking for open source candidates where it makes business sense.

The HP HPC Linux portfolio

HP has extensive expertise and experience in delivering high performance compute clustering solutions based on both Linux and UNIX. The HPC portfolio is built upon a scalable framework with the common base of HP Cluster Platforms supporting a choice of operating systems, as well as layered middleware and applications. In addition to support for the HP XC System Software, these platforms include solutions developed in collaboration with partners such as Scali and based on open source collaborations such as OSCAR and ROCKS.

For customers who require custom Linux solutions, HP combines its supercomputing expertise and Linux cluster technology expertise with the open source flexibility of Linux to deliver specifically tailored cluster solutions. In such instances, HP can rapidly deploy solutions based on major deployments. HP supported Sandia in the deployment of the Cplant installation, one of the first major Linux cluster deployments in production and recently worked with that site on deployment of its 512-node and 128-node Institutional Linux Clusters, utilizing HP ProLiant servers and Myricom Myrinet high-speed interconnect. HP is also assisting Pacific Northwest National Laboratories in its deployment of the top-performing Linux-based cluster, with more than 1900 Itanium 2 processor–based systems and a Quadrics-based interconnect fabric.

Working with the open source community

HP has extensive experience working in the open source community. The following are joint projects between HP and the Linux community:

- HP is a corporate member of the Open Source Development Lab. Martin Fink of HP is the VP for OSDL and has been since its creation. HP leads the Datacenter Working group and is an active participant in the Carrier grade working group within OSDL.
- For the last decade, HP has invested in 64-bit computing and co-developed the Itanium architecture with Intel. HP Labs scientists maintain the Itanium processor Linux kernel working with the open source community.
- HP is a charter member of the Free Standards Group and an active participant in the LSB (Linux Standard Base) specification. HP chairs the Futures group for LSB.

- Lustre file system project—HP is working closely with CFS and the DOE Trilabs on development of the high-performance cluster file system.
- Open Printing Workgroup within FSG (formerly started as HP Open Source Printing Initiative)—HP is working with the open source community to advance the state of Linux printing. HP is implementing broad support for HP hardcopy devices employing open source technologies, which includes the LSB printing initiative and hosting the OSDN printer summit.
- Several core Debian team members are on the HP staff. Debian is a non-commercial Linux distribution preferred by some customers and developers. With HP's efforts, Debian was one of the first distributions available for Itanium 2 processor–based systems from HP.
- Samba is the leading UNIX-based file/print software that interoperates with Microsoft® Windows® desktops and servers. Several principal Samba developers are on the HP staff to further optimize Samba for HP customers.
- HP is a charter member of the Open Software Developers Network (OSDN), an industry-wide organization. HP also has a relationship with a large-scale open-source community service provider, SourceForge. HP sponsors the open source clustering foundry and hosts the group's internally developed open source projects at SourceForge.
- HP ported Linux to the iPAQ Pocket PC and sponsors the HandHeld.org open source project to foster innovation on handheld devices.
- HP released its single-system image clustering technology to open source, as well as many other projects, including the Solaris to Linux Threads porting libraries.
- HP is a major contributor to Apache 2.0, with more than 15 people on its programming staff.
- The Gelato Federation, co-founded by HP and seven leading research institutions, develops software to enable researchers to advance their studies using Linux and Itanium-based systems.
- HP supports the Open Source Software Institute, which focuses on research and education to help accelerate the use of open source software in government information technology.
- HP is a founding member of the GNOME foundation and has committed to making GNOME the default desktop for its UNIX workstations.
- HP is also a founding member of the KDE to ensure its customers have a choice of desktops.

More description of HP's contributions to the open source community can be found at: http://opensource.hp.com.

For more information

www.hp.com/go/hptc
HP's High Performance Computing home site.

www.hp.com/go/clusters
HP's Clustering home site.

4AA0-0815ENW, 06/2005

invent