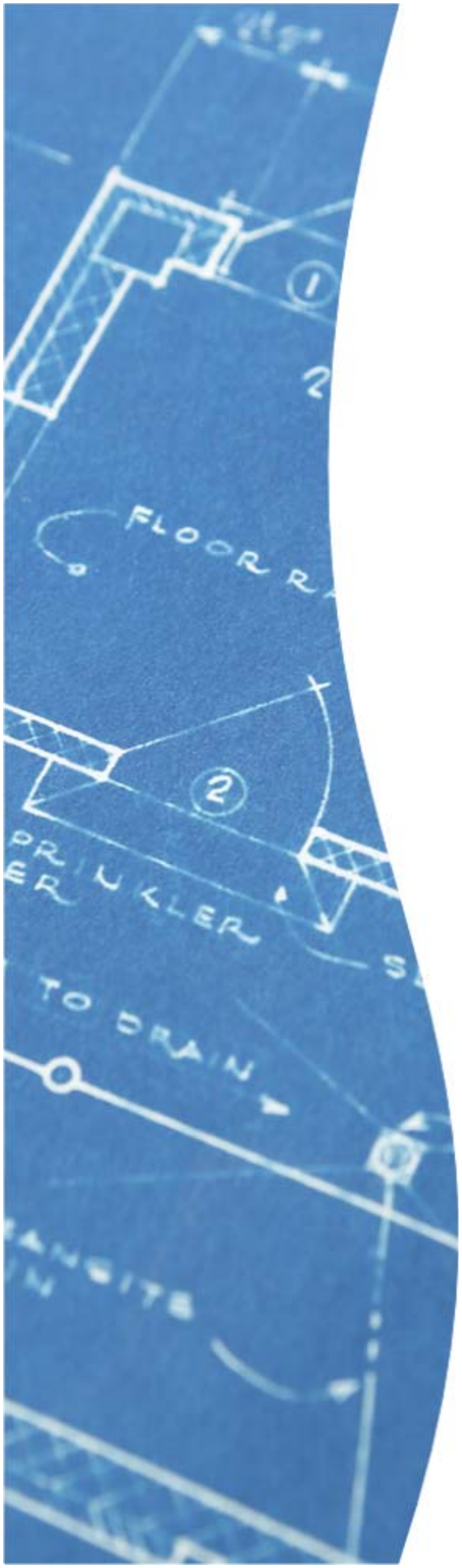


SUN™ CUSTOMER READY HPC CLUSTER:

REFERENCE CONFIGURATIONS WITH SUN FIRE™ X4100, X4200, AND X4600 SERVERS

Jeff Lu, Systems Group

Sun BluePrints™ OnLine — April 2007

Part No 820-1270-11
Revision 1.1, 4/18/07
Edition: April 2007

Table of Contents

Sun Customer Ready HPC Cluster: Reference Configurations with Sun Fire X4100, X4200, and X4600 Servers

The Sun™ Customer Ready HPC Cluster includes servers, networking, interconnects, and software installed in a Sun Rack cabinet. The Sun Customer Ready HPC Cluster with Sun Fire™ X4100/X4100 M2, X4200/X4200 M2, and X4600/X4600 M2 servers is optimized for High-Performance Computing (HPC) applications such as scientific research, mechanical computer-aided engineering (MCAE), electronic design automation (EDA), financial analysis, and any other compute-intensive application. These configurations focus on delivering computing power with high space density, in a cluster or grid architecture.

All Sun Customer Ready HPC Clusters are configured and tested by the Sun℠ Customer Ready program, which offers factory-integrated solutions using Sun and third-party products. Such factory integrated systems help enable businesses to deploy IT solutions in a simpler, safer, and swifter manner. As a result, customers gain on productivity and save on deployment costs and risk. (For more information, consult `http://www.sun.com/servers/cr`.)

Warranty and post-sales support for Sun Customer Ready HPC Cluster systems using select, Sun price-listed components is provided through Sun's Service Plans for Systems. For Sun Customer Ready HPC Cluster components that are not on Sun's price list, post-sales support may be contracted through Sun's Managed Services.

Reference Configurations

The reference configurations described in this paper are starting points for building Sun Customer Ready HPC Clusters configured with the Sun Fire X4100, X4200, and X4600 families of servers. The configurations define how Sun Systems Group products can be configured in a typical grid rack deployment. This document describes configurations using Sun Fire X4100 and X4100 M2 servers with a Gigabit Ethernet data fabric and with a high-speed InfiniBand fabric. In addition, this document describes configurations using Sun Fire X4200, X4200 M2, X4600, and X4600 M2 servers with an InfiniBand data fabric. These configurations focus on single rack solutions, with external connections through uplink ports of the switches.

These reference configurations have been architected using Sun's expertise gained in actual, real-world installations. Within certain constraints, as described in the later sections, the system can be tailored to the customer needs. Certain system components described in this document are only available through Sun's factory integration. Although the information contained here could be used during an integration on-site, the optimal benefit is achieved when the Sun Customer Ready program integrates the system.

The remaining sections of this document provide the following information:

- High level system architecture
- Interconnect diagrams
- Management node strategy and recommended management software
- Rack limits, size, airflow, weight, and power specifications
- Configuration rules
- List of components
- Default network and disk partition information

Sun Customer Ready HPC Cluster Architecture

The Sun Customer Ready HPC Cluster consists of a set of components integrated in a Sun Rack 1000-42/38 rack, with interconnections and management utility to deliver compute and system management operations. The Sun Customer Ready HPC Cluster provides an Ethernet management fabric for the management of components, connected to the rack management node. The applications are distributed to run on the Sun Fire servers, delivering results through the data fabric, which consists of either Gigabit Ethernet switches or higher performance InfiniBand switches. An optional provisioning fabric for operating systems and applications can also be configured through Gigabit Ethernet switches in the rack.

Figure 1 illustrates the overall layout of the Sun Customer Ready HPC Cluster components, including the optional Keyboard Monitor and Mouse (KMM) for local interface and the grid master node for scheduling grid computing resources.

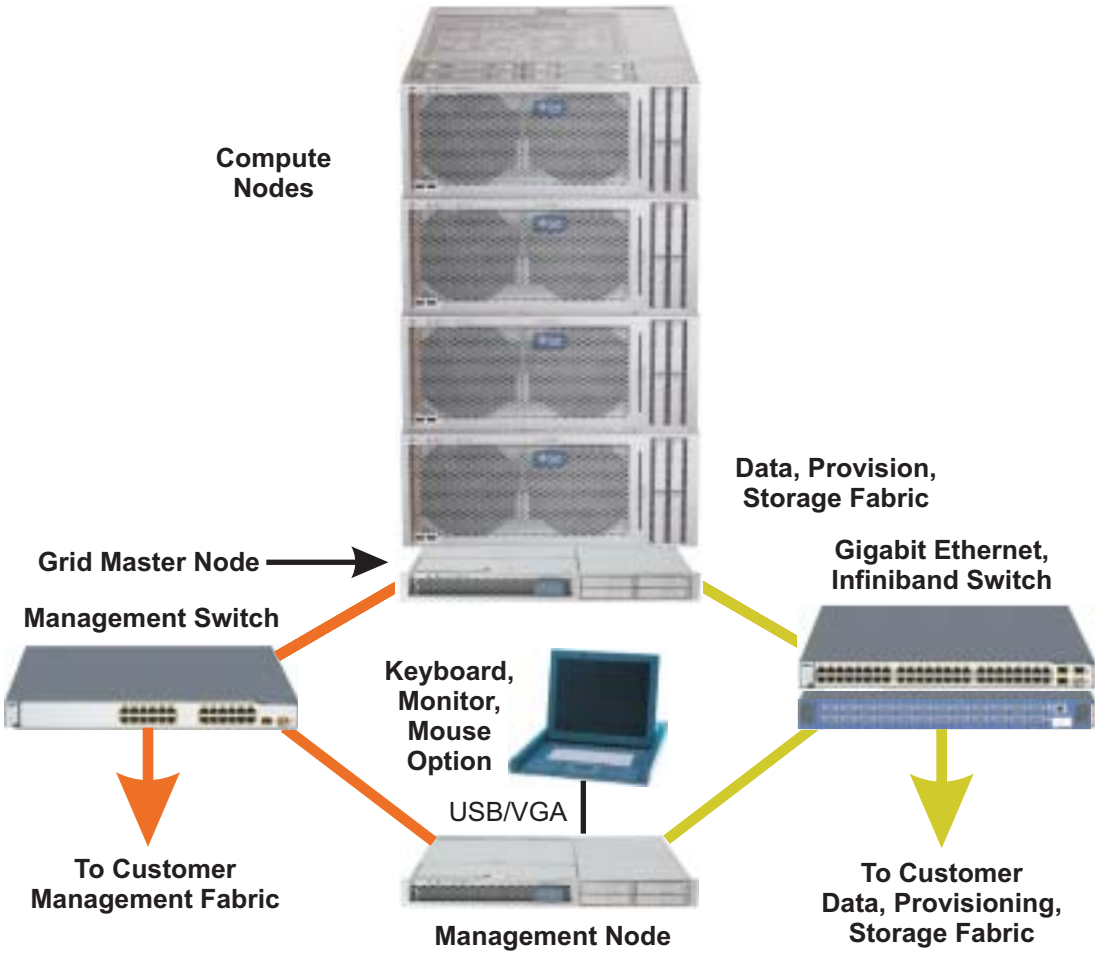


Figure 1. Sun Customer Ready HPC Cluster Architecture.

Sun Customer Ready HPC Cluster Hardware Components

Each Sun Customer Ready HPC Cluster contains Sun Fire servers configured as compute nodes, a management node, and a grid master node.

- *Compute Nodes* — The compute nodes in the system are the primary engines performing the computation in the rack. Typical usage in HPC involves provisioning of the OS and tasks on to each compute node. Smaller workloads can be loaded on each node's local disk. Larger data sets are delivered through external storage connections.
- *Management Node* — The management node in the Sun Customer Ready HPC Cluster runs the Sun N1™ System Manager software. Connecting through the management port, the management node communicates to each of the server nodes and switches in the rack for setup and monitoring. The management node handles DHCP, PXE booting, OS provisioning, and application provisioning tasks.

- *Grid Master Node* — The grid master node handles task scheduling and communication between compute nodes for a grid environment. The grid master node can also communicate with external compute nodes through uplinks in the switches. Typically, a separate server is configured as the grid master node using the Sun N1™ Grid Engine software; this node communicates with the computing nodes through the data fabric and the provisioning fabric.

In addition, each Sun Customer Ready HPC Cluster contains the following hardware components that provide system interconnects; rack, power, and cabling; and optional KMM and terminal server functionality:

- *Gigabit Ethernet Data Switch* — The rack system uses either a 24-port or 48-port Gigabit Ethernet switch as the data switch. The data switch serves as the primary interconnection for the compute nodes and the external data interface.
- *Management Switch* — The management switch is connected to the management node, the Service Processor port of each of the compute nodes, and the management port of the components in the Sun Customer Ready HPC Cluster. Uplink ports are used to connect to an external management network.
- *InfiniBand Switch* — An InfiniBand switch serves as the main connection between the compute nodes and external InfiniBand fabric. The switch supports 4X InfiniBand, with non-blocking features between ports. A 24-port 1U switch as well as a 96-port 6U switch offer different InfiniBand topologies for the rack, depending on the site requirements.
- *Storage Data Switch* — A separate Gigabit Ethernet storage data path is recommended for servers connecting to Network Attached Storage (NAS) devices. Additional high performance storage architecture options, such as Fibre Channel or InfiniBand are possible for custom configuration for each server through additional option cards and switching components.
- *Rack and Power* — All components are pre-installed in either a Sun Rack 1038 or 1042. The rack is equipped with front and rear doors, and side covers. The Sun Customer Ready HPC Cluster is equipped with a full configuration of power distribution units, which provide sufficient power outlets to cable all installed nodes. Each component has a separate AC power cord. The power cords are routed from the rear, with side-mounted outlet strips to minimize the impact on rear access to the components in the rack.
- *Cabling and Interconnect* — All cable harnesses are precut, routed, and terminated in locations that provide minimal excess cable without impeding component access or replacement. For a Sun Customer Ready HPC Cluster configuration with all component types, the rack can be configured with 3 separate types of harnesses. (Data fabric: yellow; Management fabric: orange: Serial: white.) InfiniBand cables are handled with cable management attachment to ensure proper installation and operation.

- *Keyboard, Monitor, Mouse* — A keyboard, video monitor, and mouse (KMM) unit in the rack system is cabled to the management node by default, and can be used to connect to each of the compute nodes for local interfacing. if required.
- *Terminal Server* — All nodes in the rack and the network switches have the option of connecting via their serial console ports to a 48-port terminal server. The terminal server ports use consistent RJ-45 serial port wiring standards. The terminal server can then be used to perform individual configuration functions remotely, if necessary. The terminal server provides normal telnet and secure shell (SSH) access to a serial port using the telnet IP-address port-number protocol.

Component Configuration Rules

Table 1 lists the minimum and maximum number of components in a single Sun Customer Ready HPC Cluster.

Table 1. Component Configuration Rules for Sun Customer Ready HPC Clusters.

Component	Minimum	Maximum
1U compute nodes (Sun Fire X4100/X4100 M2 servers)	2	40[a]
2U compute nodes (Sun Fire X4200/X4200 M2 servers)	2	20
4U compute nodes (Sun Fire X4600/X4600 M2 servers)	2	10
Management node	0	1
Grid master node	0	1
Gigabit Ethernet switch	0	4
InfiniBand 24-port switch	0	2
InfiniBand 96-port switch chassis	0	1
Keyboard, Monitor, Mouse (KMM) unit	0	1
Terminal server	0	1

a. Maximum depends on server configuration.

Reference Architecture Components

Each Sun Customer Ready HPC Cluster contains integrated Sun Fire splerver compute nodes, rack system management infrastructure, a Gigabit Ethernet or InfiniBand data fabric, and options for the storage interface. Multiple operating systems, including the Solaris™ Operating System (Solaris OS) and Linux, are supported. In addition, the Sun Customer Ready HPC Cluster is configured with the Sun N1 System Manager and the Sun N1 Grid Engine software.

Sun Fire Server Compute Nodes

The Sun Fire X4100/X4100 M2, X4200/X4200 M2, and X4600/X4600 M2 families of servers are the primary compute nodes in the system configurations described in this

article. The key difference between the M2 and the non-M2 offerings is the Rev-F Opteron™ CPU implementation in the M2 systems. The M2 systems also use DDR2 memory in the server, which provides memory performance improvement in each node. In addition, the Sun Fire X4100 M2 and Sun Fire X4200 M2 systems also include PCI-Express expansion slots for improved I/O throughput. The M2 systems have no changes to the disk and chassis features, compared to the non-M2 systems.

- *Sun Fire X4100 and X4100 M2 Servers*
  The Sun Fire X4100 and X4100 M2 servers offer two Opteron sockets in a 1U form factor. With a maximum of 64 GB of memory, each server with two dual-core CPUs running at 2.6 GHz can deliver up to 20.8 GFLOPS of theoretical peak computing. Each server has four Gigabit Ethernet connections, and can have up to two SAS disks for local storage. There are two PCI-X expansion slots in the Sun Fire X4100 server, and two PCI Express slots in the Sun Fire X4100 M2 server.
- *Sun Fire X4200 and X4200 M2 Servers*
  The Sun Fire X4200 and X4200 M2 servers offer two Opteron sockets in a 2U form factor. With a maximum of 64 GB of memory, each server with two dual-core CPUs running at 2.6 GHz can deliver up to 20.8 GFLOPS of theoretical peak computing. Each server has four Gigabit Ethernet connections, and can have up to four SAS disks for local storage. The Sun Fire X4200 server has five PCI-X expansion slots, while the Sun Fire X4200 M2 server has four PCI Express slots and one PCI-X slot for I/O expansion.
- *Sun Fire X4600 and X4600 M2 Servers*
  The Sun Fire X4600 and X4600 M2 server are eight-socket Opteron platforms, offering up to 16 computing cores in a 4U density. With a maximum of 128 GB of memory, each Sun Fire X4600 server with 8 CPUs (16 cores) running at 2.6 GHz can deliver up to 83 GFLOPS of theoretical peak computing. Each server has four Gigabit Ethernet connections, and can have up to four SAS disks for local storage. In addition, there are six PCI Express and two PCI-X expansion slots for I/O expansion, allowing a wide array of clustering and storage interface configurations to meet the deployment requirements.

The server management connection to the Integrated Lights Out Management (ILOM) is through a dedicated Ethernet port. The ILOM of each server provides control of remote power on/off and status, and offers browser and command line interfaces through standard industry protocols, including HTTPS, IPMI 2.0, SSH, and SNMP. The ILOM also provides full remote keyboard, mouse, video, and storage emulation as an USB device for full feature control from a separate management node.

Rack System Management Infrastructure

A robust management architecture has been defined to manage all of the components within the rack. All compute and switch components in the rack are connected to the

management Ethernet switch through the management port. The management switch also provides uplink ports to the management network outside of the single rack.

The recommended management node within the rack is a Sun Fire X4100/X4100 M2 server, running the Sun N1 System Manager software on the Solaris or Red Hat Linux operating system, which provides monitoring and provisioning functions for the rack. The management node is connected to the management switch for full access to the management ports of all components. This management node can also manage additional rack resources, with connections through the management switch. The management fabric can be implemented on the same physical switch unit as the data/provisioning switch, using Virtual Local Area Network (VLAN) functionality to establish different network segments.

A Keyboard, Monitor, and Mouse (KMM) option is available to provide a rack user interface access to all nodes. The KMM is connected to the management node by default.

A terminal server in the rack is not needed, given the KVM (keyboard, video, and mouse) over IP management feature in the Sun Fire server. However, if a customer requires serial connections to all nodes as a standard deployment policy, a terminal server is available.

Data Fabric

The standard data fabric in Sun Customer Ready HPC Clusters is a Gigabit Ethernet connection to each compute node, with options of multiple Gigabit Ethernet connections for higher bandwidth or redundancy. A separate provisioning data fabric can also be configured using the additional Gigabit Ethernet port of the compute node and a data switch in the rack.

InfiniBand (IB) is a switched data fabric that provides high performance and low latency. InfiniBand 4X Single Data Rate (SDR) connections can achieve a theoretical maximum transmission of 10 Gigabits per second. Each compute node is configured with an InfiniBand host channel adapter (HCA) connected to an InfiniBand switch in the rack. Multiple InfiniBand configurations are possible to achieve a wide bandwidth data connection, depending on the deployment requirements.

Storage Interface

Various data storage architectures can be implemented with a Sun Customer Ready HPC Cluster. The standard storage interface is through a Gigabit Ethernet storage data fabric to each of the servers. A Network Attached Storage (NAS) device can be configured to this storage data fabric as required. Higher performance storage connections through Fibre Channel or InfiniBand to external storage area network

(SAN) devices are accomplished through option cards in the server with additional switches.

This document does not address the storage network topology or storage component options, as storage components are not part of the initial Sun Customer Ready HPC Cluster offering.

Operating System

The computing nodes in a Sun Customer Ready HPC Cluster support the following operating system (or later) revisions:

- Solaris 10 OS Update 2, 64-bit
- Red Hat Enterprise Linux (RHEL) 4.0 U4 64-bit
- SUSE Linux Enterprise Server 9 SP3 64-bit

The recommended operating system for the management node and the grid master node is the Solaris 10 OS.

System Management Software

The recommended management software for Sun Customer Ready HPC Clusters is the Sun N1 System Manager software, v1.3.2, installed on the management node with either the Solaris or Red Hat Linux OS. The Sun N1 System Manager software performs bare metal hardware discovery for new nodes, provisions the OS and firmware, monitors hardware and software, and integrates with lights-out management features in Sun x64 servers.

Grid Computing Management

The Sun N1 Grid Engine software running on the Solaris, Red Hat, or SUSE Linux operating system is the recommended grid computing software in a Sun Customer Ready HPC Cluster. The Sun N1 Grid Engine software manages and dispatches jobs onto the compute grid, with policy-based workload management and dynamic provisioning of application workloads.

Typically, a separate Sun Fire server is configured as the master node for the N1 Grid system; this server communicates with the compute nodes through the data fabric and the provisioning fabric. The master node handles task scheduling and communication between compute nodes for a grid environment. The master node can also communicate with external compute nodes through uplink ports in the switches.

Reference Configurations

The following sections provide details on five distinct reference configurations for Sun Customer Ready HPC Clusters using the Sun Fire X4100, X4200, and X4600 families of servers:

- Sun Fire X4100/X4100 M2 Servers with Gigabit Ethernet (page 10)
- Sun Fire X4100/X4100 M2 Servers with InfiniBand (page 14)
- Sun Fire X4100/X4100 M2 Servers with Non-Blocking InfiniBand (page 18)
- Sun Fire X4200/X4200 M2 Servers with InfiniBand (page 22)
- Sun Fire X4600/X4600 M2 Servers with InfiniBand (page 26)

The first three reference configurations illustrate various ways of configuring a Sun Customer Ready HPC Cluster with Sun Fire X4100/X4100 M2 servers. Given the lower latency and wider, faster data path of the InfiniBand data fabric, the InfiniBand configurations are expected to deliver a higher system efficiency as compared to a configuration using a Gigabit Ethernet fabric. The two InfiniBand configurations illustrate examples of Sun's blocking and non-blocking offerings. The first InfiniBand configuration is a blocking configuration, using two 24-port switches and delivering a 2:1 ratio for bisectional bandwidth between nodes.[1] The non-blocking InfiniBand configuration uses a larger, more expensive switch to achieve a fully non-blocking 1:1 ratio for bandwidth between nodes.

The reference configurations for the Sun Fire X4200 and Sun Fire X4600 families of servers illustrate the use of 2U and 4U systems, respectively. Both of these example configurations feature InfiniBand connections, with fully non-blocking connections within the rack. The Sun Fire X4600 servers are connected with two host channel adapter (HCA) cards per server node, illustrating an alternate InfiniBand installation that can be used to achieve balanced connections between nodes.

1. A 2:1 ratio indicates that between certain nodes there is a 50% chance that the connection is available, while a 3:1 ratio indicates that 33% of the time the connection is available. The higher the blocking ratio, the lower the performance of the connection fabric. A 1:1 ratio indicates a fully non-blocking topology.

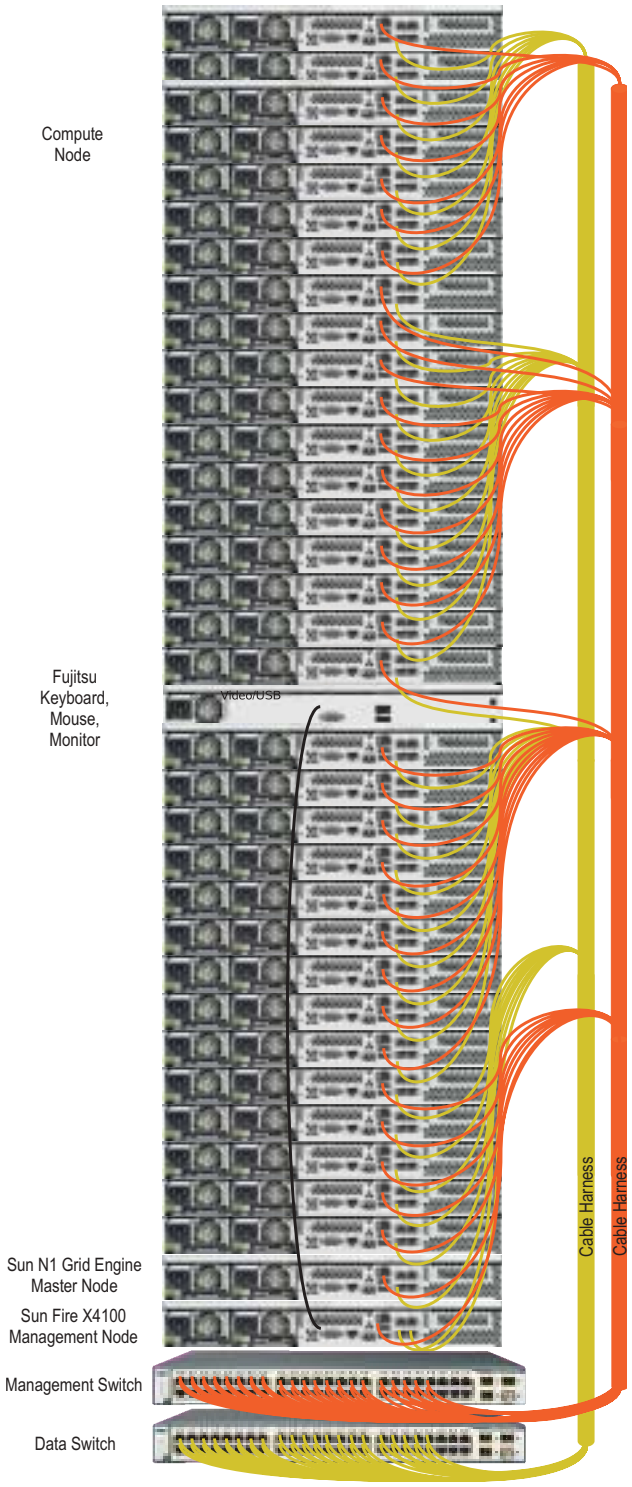
Sun Fire X4100/X4100 M2 Servers with Gigabit Ethernet

This reference configuration consists of a single Sun Rack 1000-42 rack containing 32 dual-socket Sun Fire X4100/X4100 M2 servers used as compute nodes, one 48-port management fabric switch, and one Gigabit Ethernet data fabric switch. This configuration also includes a Sun Fire X4100/X4100 M2 server with a KMM for use as the management node, and a second Sun Fire X4100/X4100 M2 server as the master node for the Sun N1 Grid Engine software. With a configuration of 2.6 GHz dual-core Opteron processors, this rack can deliver an estimated 0.67 TFLOPS of theoretical peak computation.

Configuration and Cabling Diagram

Figure 2 shows a high level wiring diagram of the rack. The included table details how each component is arranged in the rack, and lists the cable harness mapping to each port of the component. A Gigabit Ethernet port from the compute nodes is connected to the 48-port Gigabit Ethernet data switch, with uplinks for external connections. The same Gigabit Ethernet fabric is used for the provisioning network and storage network in this configuration.

A Sun Fire X4100/X4100 M2 server, running the Solaris OS and the Sun N1 System Manager software, is configured as the management node. This management node is not required if an external system management node already exists. A second Sun Fire X4100/X4100 M2 server is setup as the master node for the Sun N1 Grid Engine software, providing task scheduling for HPC grid computing.



Loc	Device	F	Dev Port	Data Harness Label	SW	Dev Port	Mgmt Harness Label	SW
U42				U-42,Rack<->SW-1,Prt-42	42		U-42,Rack<->SW-2,Prt-42	42
U41				U-41,Rack<->SW-1,Prt-41	41		U-41,Rack<->SW-2,Prt-41	41
U40				U-40,Rack<->SW-1,Prt-40	40		U-40,Rack<->SW-2,Prt-40	40
U39	X4100	C	0	U-39,Rack<->SW-1,Prt-39	39	SP	U-39,Rack<->SW-2,Prt-39	39
U38	X4100	C	0	U-38,Rack<->SW-1,Prt-38	38	SP	U-38,Rack<->SW-2,Prt-38	38
U37	X4100	C	0	U-37,Rack<->SW-1,Prt-37	37	SP	U-37,Rack<->SW-2,Prt-37	37
U36	X4100	C	0	U-36,Rack<->SW-1,Prt-36	36	SP	U-36,Rack<->SW-2,Prt-36	36
U35	X4100	C	0	U-35,Rack<->SW-1,Prt-35	35	SP	U-35,Rack<->SW-2,Prt-35	35
U34	X4100	C	0	U-34,Rack<->SW-1,Prt-34	34	SP	U-34,Rack<->SW-2,Prt-34	34
U33	X4100	C	0	U-33,Rack<->SW-1,Prt-33	33	SP	U-33,Rack<->SW-2,Prt-33	33
U32	X4100	C	0	U-32,Rack<->SW-1,Prt-32	32	SP	U-32,Rack<->SW-2,Prt-32	32
U31	X4100	C	0	U-31,Rack<->SW-1,Prt-31	31	SP	U-31,Rack<->SW-2,Prt-31	31
U30	X4100	C	0	U-30,Rack<->SW-1,Prt-30	30	SP	U-30,Rack<->SW-2,Prt-30	30
U29	X4100	C	0	U-29,Rack<->SW-1,Prt-29	29	SP	U-29,Rack<->SW-2,Prt-29	29
U28	X4100	C	0	U-28,Rack<->SW-1,Prt-28	28	SP	U-28,Rack<->SW-2,Prt-28	28
U27	X4100	C	0	U-27,Rack<->SW-1,Prt-27	27	SP	U-27,Rack<->SW-2,Prt-27	27
U26	X4100	C	0	U-26,Rack<->SW-1,Prt-26	26	SP	U-26,Rack<->SW-2,Prt-26	26
U25	X4100	C	0	U-25,Rack<->SW-1,Prt-25	25	SP	U-25,Rack<->SW-2,Prt-25	25
U24	X4100	C	0	U-24,Rack<->SW-1,Prt-24	24	SP	U-24,Rack<->SW-2,Prt-24	24
U23	X4100	C	0	U-23,Rack<->SW-1,Prt-23	23	SP	U-23,Rack<->SW-2,Prt-23	23
U22	X4100	C	0	U-22,Rack<->SW-1,Prt-22	22	SP	U-22,Rack<->SW-2,Prt-22	22
U21	KMM		0	U-21,Rack<->SW-1,Prt-21	21		U-21,Rack<->SW-2,Prt-21	21
U20	X4100	C	0	U-20,Rack<->SW-1,Prt-20	20	SP	U-20,Rack<->SW-2,Prt-20	20
U19	X4100	C	0	U-19,Rack<->SW-1,Prt-19	19	SP	U-19,Rack<->SW-2,Prt-19	19
U18	X4100	C	0	U-18,Rack<->SW-1,Prt-18	18	SP	U-18,Rack<->SW-2,Prt-18	18
U17	X4100	C	0	U-17,Rack<->SW-1,Prt-17	17	SP	U-17,Rack<->SW-2,Prt-17	17
U16	X4100	C	0	U-16,Rack<->SW-1,Prt-16	16	SP	U-16,Rack<->SW-2,Prt-16	16
U15	X4100	C	0	U-15,Rack<->SW-1,Prt-15	15	SP	U-15,Rack<->SW-2,Prt-15	15
U14	X4100	C	0	U-14,Rack<->SW-1,Prt-14	14	SP	U-14,Rack<->SW-2,Prt-14	14
U13	X4100	C	0	U-13,Rack<->SW-1,Prt-13	13	SP	U-13,Rack<->SW-2,Prt-13	13
U12	X4100	C	0	U-12,Rack<->SW-1,Prt-12	12	SP	U-12,Rack<->SW-2,Prt-12	12
U11	X4100	C	0	U-11,Rack<->SW-1,Prt-11	11	SP	U-11,Rack<->SW-2,Prt-11	11
U10	X4100	C	0	U-10,Rack<->SW-1,Prt-10	10	SP	U-10,Rack<->SW-2,Prt-10	10
U09	X4100	C	0	U-09,Rack<->SW-1,Prt-09	9	SP	U-09,Rack<->SW-2,Prt-09	9
U08	X4100	C	0	U-08,Rack<->SW-1,Prt-08	8	SP	U-08,Rack<->SW-2,Prt-08	8
U07	X4100	C	0	U-07,Rack<->SW-1,Prt-07	7	SP	U-07,Rack<->SW-2,Prt-07	7
U06	X4100	M	0	U-06,Rack<->SW-1,Prt-06	6	SP	U-06,Rack<->SW-2,Prt-06	6
U05	X4100	M	0	U-05,Rack<->SW-1,Prt-05	5	1	U-05,Rack<->SW-2,Prt-05	5
U04	Mgmt 48P switch	X		U-04,Rack<->SW-1,Prt-04	4		U-04,Rack<->SW-2,Prt-04	4
U03	GigE 48P switch	X		U-03,Rack<->SW-1,Prt-03	3		U-03,Rack<->SW-2,Prt-03	3
U02	MPS 60A	P		U-02,Rack<->SW-1,Prt-02			U-02,Rack<->SW-2,Prt-02	
U01	MPS 60A	P						

KEY: P MPS T Terminal Server
X Data Switch M Management Node
X Mgmt Switch C Compute Node

Figure 2. Rack and Cabling Mapping: Sun Fire X4100 Servers with Gigabit Ethernet.

Power Cord Connections

The Sun Fire X4100/X4100 M2 server provides power redundancy with an optional second power supply unit (PSU). For HPC, only one power supply unit is configured. Servers are connected in an interleaved arrangement between two circuits in the rack, ensuring half of the servers can continue operation in the case of one circuit failure. Table 2 shows AC power cord connections for the reference configuration. The current draws are distributed to ensure they do no exceed the limits of each plug, strip pair, and circuit maximum specifications. The notation is in the format of Uxx.y, where Uxx describes the component mounted in the location starting at rack unit number xx, and y is the AC power cord number of the unit.

Table 2. AC Power Cord Connections for Sun Fire X4100 Servers and Gigabit Ethernet Configurations.

Circuit A	Strip						
	0	3	1	4	2	5	Plug
	U03.1	U07.1	U15.1	U23.1	U31.1	U39.1	0
	U05.1	U08.1	U16.1	U24.1	U32.1		1
	U06.1		U17.1	U25.1	U33.1		2
			U18.1	U26.1	U34.1		3
							4
							5
Circuit B	Strip						
	0	3	1	4	2	5	Plug
	U04.1	U09.1	U11.1	U19.1	U27.1	U35.1	0
	U05.2	U10.1	U12.1	U20.1	U28.1	U36.1	1
	U06.2		U13.1	U21.1	U29.1	U37.1	2
			U14.1	U22.1	U30.1	U38.1	3
							4
							5

Site Planning Considerations for Sun Fire X4100 Servers with Gigabit Ethernet

Table 3 summarizes the size, airflow, power, and weight of the entry level reference configuration, based on preliminary component specification, and compares them against the rack limits.

Table 3. Site Planning Considerations: Sun Fire X4100 Servers with Gigabit Ethernet Configurations.

		Size (RU)		Air Flow (CFM) @35C		Power (W)		Weight (kg, lb)			
Description	**Total Units**	**Unit RU**	**Total RU**	**Unit CFM**	**Total CFM**	**Unit W**	**Total W**	**Unit kg**	**Total kg**	**Unit lb**	**Total lb**
Sun Fire X4100 compute node	32	1	32	96	3072	550	17600	19.6	627.2	43.1	1379.8
Sun Fire X4100 management node, grid master node	2	1	2	96	192	550	1100	19.6	39.2	43.1	86.2
1U 24-port 4X IB switch	0	1	0	15	0	96	0	7.7	0.0	16.9	0.0
1U Ethernet switch	2	1	2	15	30	160	320	6.4	12.8	14.1	28.2
KMM (keyboard, monitor, mouse)	1	1	1	0	0	22	22	12.4	12.4	27.3	27.3
Cable harness (GigE), IB cables	2	0	0	0	0	0	0	7.5	15.0	16.5	33.0
Rack cabinet with MPS	1	2	2	0	0	25	25	195.0	195.0	429.0	429.0
Total Reference Config System			39		3294		19067		901.6		1983.5
Max rack limits: 1042MPS60A			42				32448		920.8		2025.7
Max rack limits: 1042MPS32A			42				39936		920.8		2025.7

Sun Fire X4100/X4100 M2 Servers with InfiniBand

This reference configuration illustrates how a full rack of Sun Fire X4100/X4100 M2 servers can be configured with a high performance data fabric using InfiniBand technology. Given the lower latency and a wider, faster data path, the InfiniBand data fabric is expected to deliver a much higher system efficiency over a Gigabit Ethernet fabric configuration.

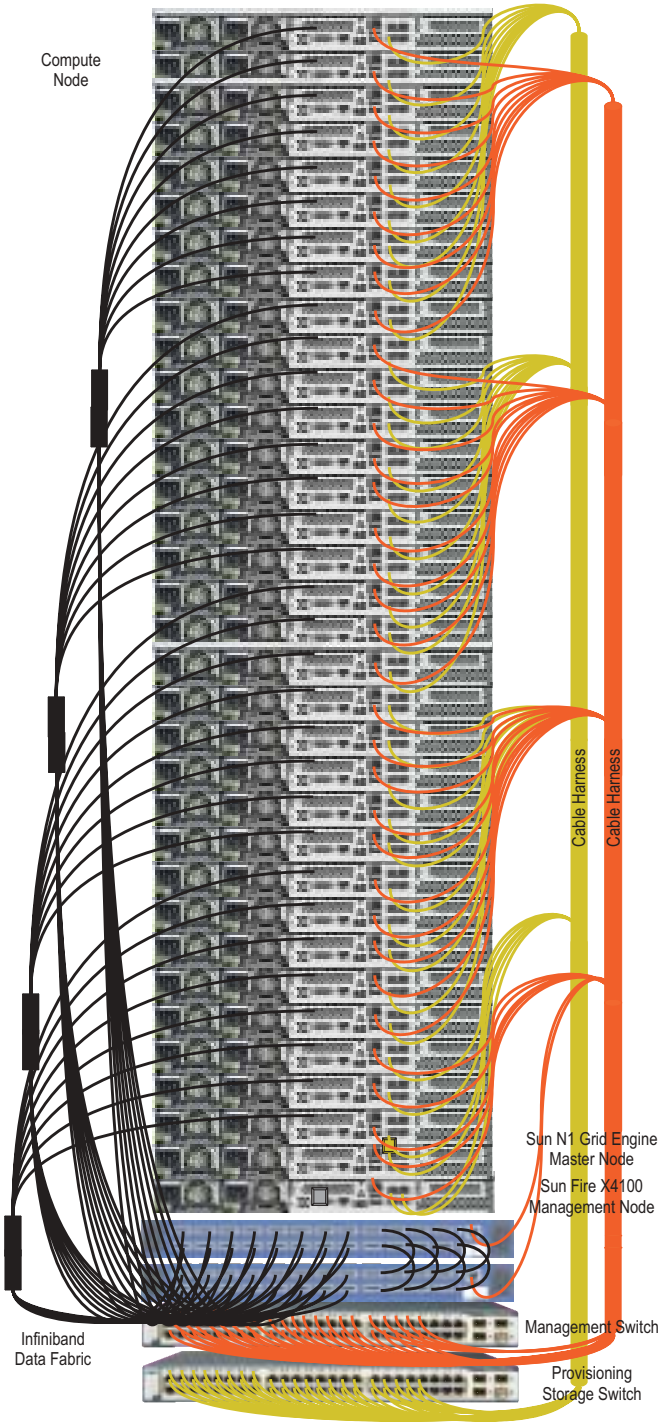
Thirty-two dual-socket Sun Fire X4100/X4100 M2 servers, each with 2.6 GHz dual-core CPUs, are configured in a Sun Rack 1000-1042 rack as compute nodes, with an estimated 0.67 TFLOPS of theoretical peak computational performance. A Sun Fire X4100/X4100 M2 server is connected to the management fabric and serves as the management node, and a second Sun Fire X4100/X4100 M2 server is configured as the grid master node.

Configuration and Cabling Diagram

Figure 3 shows a high level wiring diagram of the rack. The included table details how each component is arranged in the rack, and lists the cable harness mapping to each port of the component. Each of the 32 compute nodes is configured with a single port Voltaire 4X InfiniBand HCA, wired to a Voltaire 24-port InfiniBand switch in the rack. The 4X HCA card is installed in the expansion slot in the Sun Fire X4100/X4100 M2 server, supporting InfiniBand Single Data Rate (SDR) mode. With SDR, the 4X HCA offers 10 Gigabits per second of connection bandwidth between nodes. Each of the two 24-port Voltaire InfiniBand switches provides non-blocking connections between 16 compute nodes in the rack. Two InfiniBand switches are connected together with eight InfiniBand ports, giving a 2:1 bisectional bandwidth ratio within the rack between all 32 nodes.

A Gigabit Ethernet port of the compute nodes is connected to a 48-port Gigabit Ethernet switch, configured as the provisioning and the storage data fabric for the rack. The management network is on a separate 48-port switch, with dedicated uplink ports to external management connections.

A Sun Fire X4100/X4100 M2 server is setup as the management node, running the Solaris OS and the Sun N1 System Manager software. A second Sun Fire X4100/X4100 M2 server is setup as the master node for the Sun N1 Grid Engine software, providing task scheduling for HPC grid computing.



Loc	Device	F	Dev Port	Data Harness Label	SW	Dev Port	Mgmt Harness Label	SW
U42				U-42,Rack<->SW-1,Prt-42	42		U-42,Rack<->SW-2,Prt-42	42
U41				U-41,Rack<->SW-1,Prt-41	41		U-41,Rack<->SW-2,Prt-41	41
U40	X4100	C	0	U-40,Rack<->SW-1,Prt-40	40	SP	U-40,Rack<->SW-2,Prt-40	40
U39	X4100	C	0	U-39,Rack<->SW-1,Prt-39	39	SP	U-39,Rack<->SW-2,Prt-39	39
U38	X4100	C	0	U-38,Rack<->SW-1,Prt-38	38	SP	U-38,Rack<->SW-2,Prt-38	38
U37	X4100	C	0	U-37,Rack<->SW-1,Prt-37	37	SP	U-37,Rack<->SW-2,Prt-37	37
U36	X4100	C	0	U-36,Rack<->SW-1,Prt-36	36	SP	U-36,Rack<->SW-2,Prt-36	36
U35	X4100	C	0	U-35,Rack<->SW-1,Prt-35	35	SP	U-35,Rack<->SW-2,Prt-35	35
U34	X4100	C	0	U-34,Rack<->SW-1,Prt-34	34	SP	U-34,Rack<->SW-2,Prt-34	34
U33	X4100	C	0	U-33,Rack<->SW-1,Prt-33	33	SP	U-33,Rack<->SW-2,Prt-33	33
U32	X4100	C	0	U-32,Rack<->SW-1,Prt-32	32	SP	U-32,Rack<->SW-2,Prt-32	32
U31	X4100	C	0	U-31,Rack<->SW-1,Prt-31	31	SP	U-31,Rack<->SW-2,Prt-31	31
U30	X4100	C	0	U-30,Rack<->SW-1,Prt-30	30	SP	U-30,Rack<->SW-2,Prt-30	30
U29	X4100	C	0	U-29,Rack<->SW-1,Prt-29	29	SP	U-29,Rack<->SW-2,Prt-29	29
U28	X4100	C	0	U-28,Rack<->SW-1,Prt-28	28	SP	U-28,Rack<->SW-2,Prt-28	28
U27	X4100	C	0	U-27,Rack<->SW-1,Prt-27	27	SP	U-27,Rack<->SW-2,Prt-27	27
U26	X4100	C	0	U-26,Rack<->SW-1,Prt-26	26	SP	U-26,Rack<->SW-2,Prt-26	26
U25	X4100	C	0	U-25,Rack<->SW-1,Prt-25	25	SP	U-25,Rack<->SW-2,Prt-25	25
U24	X4100	C	0	U-24,Rack<->SW-1,Prt-24	24	SP	U-24,Rack<->SW-2,Prt-24	24
U23	X4100	C	0	U-23,Rack<->SW-1,Prt-23	23	SP	U-23,Rack<->SW-2,Prt-23	23
U22	X4100	C	0	U-22,Rack<->SW-1,Prt-22	22	SP	U-22,Rack<->SW-2,Prt-22	22
U21	X4100	C	0	U-21,Rack<->SW-1,Prt-21	21	SP	U-21,Rack<->SW-2,Prt-21	21
U20	X4100	C	0	U-20,Rack<->SW-1,Prt-20	20	SP	U-20,Rack<->SW-2,Prt-20	20
U19	X4100	C	0	U-19,Rack<->SW-1,Prt-19	19	SP	U-19,Rack<->SW-2,Prt-19	19
U18	X4100	C	0	U-18,Rack<->SW-1,Prt-18	18	SP	U-18,Rack<->SW-2,Prt-18	18
U17	X4100	C	0	U-17,Rack<->SW-1,Prt-17	17	SP	U-17,Rack<->SW-2,Prt-17	17
U16	X4100	C	0	U-16,Rack<->SW-1,Prt-16	16	SP	U-16,Rack<->SW-2,Prt-16	16
U15	X4100	C	0	U-15,Rack<->SW-1,Prt-15	15	SP	U-15,Rack<->SW-2,Prt-15	15
U14	X4100	C	0	U-14,Rack<->SW-1,Prt-14	14	SP	U-14,Rack<->SW-2,Prt-14	14
U13	X4100	C	0	U-13,Rack<->SW-1,Prt-13	13	SP	U-13,Rack<->SW-2,Prt-13	13
U12	X4100	C	0	U-12,Rack<->SW-1,Prt-12	12	SP	U-12,Rack<->SW-2,Prt-12	12
U11	X4100	C	0	U-11,Rack<->SW-1,Prt-11	11	SP	U-11,Rack<->SW-2,Prt-11	11
U10	X4100	C	0	U-10,Rack<->SW-1,Prt-10	10	SP	U-10,Rack<->SW-2,Prt-10	10
U09	X4100	C	0	U-09,Rack<->SW-1,Prt-09	9	SP	U-09,Rack<->SW-2,Prt-09	9
U08	X4100	M	0	U-08,Rack<->SW-1,Prt-08	8	SP	U-08,Rack<->SW-2,Prt-08	8
U07	X4100	M	0	U-07,Rack<->SW-1,Prt-07	7	1	U-07,Rack<->SW-2,Prt-07	7
U06	IB 24P switch	X		U-06,Rack<->SW-1,Prt-06	6		U-06,Rack<->SW-2,Prt-06	6
U05	IB 24P switch	X		U-05,Rack<->SW-1,Prt-05	5		U-05,Rack<->SW-2,Prt-05	5
U04	Mgmt 48P switch	X		U-04,Rack<->SW-1,Prt-04	4		U-04,Rack<->SW-2,Prt-04	4
U03	GigE 48P switch	X		U-03,Rack<->SW-1,Prt-03	3		U-03,Rack<->SW-2,Prt-03	3
U02	MPS 60A	P		U-02,Rack<->SW-1,Prt-02			U-02,Rack<->SW-2,Prt-02	
U01	MPS 60A	P						

KEY: P MPS; X Data Switch; X Mgmt Switch; T Terminal Server; M Management Node; C Compute Node

Figure 3. Rack and Cable Mapping: Sun Fire X4100 Servers with InfiniBand.

Power Cord Connections

The Sun Fire X4100/X4100 M2 server provides power redundancy with an optional second power supply unit (PSU). For HPC, only one power supply unit is configured. Servers are connected in an interleaved arrangement between two circuits in the rack, ensuring half of the servers can continue operation in the case of one circuit failure. Table 4 shows AC power cord connections for the reference configuration. The current draws are distributed to ensure they do no exceed the limits of each plug, strip pair, and circuit maximum specifications. The notation is in the format of Uxx.y, where Uxx describes the component mounted in the location starting at rack unit number xx, and y is the AC power cord number of the unit.

Table 4. AC Power Cord Connections for Configurations with Sun Fire X4100 Servers and InfiniBand.

Circuit A	Strip						
	0	3	1	4	2	5	Plug
	U03.1	U08.1	U15.1	U23.1	U31.1	U39.1	0
	U05.1	U09.1	U16.1	U24.1	U32.1	U40.1	1
	U06.1		U17.1	U25.1	U33.1		2
	U07.1		U18.1	U26.1	U34.1		3
							4
							5
Circuit B	Strip						
	0	3	1	4	2	5	Plug
	U04.1	U10.1	U11.1	U19.1	U27.1	U35.1	0
	U05.2	U8.2	U12.1	U20.1	U28.1	U36.1	1
	U06.2		U13.1	U21.1	U29.1	U37.1	2
	U07.2		U14.1	U22.1	U30.1	U38.1	3
							4
							5

Site Planning Considerations for Sun Fire X4100 Servers with InfiniBand

Table 5 summarizes the size, airflow, power, and weight of the entry level reference configuration, based on preliminary component specification, and compares them against the rack limits.

Table 5. Site Planning Considerations: Sun Fire X4100 Servers with InfiniBand.

		Size (RU)		Air Flow (CFM) @35C		Power (W)		Weight (kg, lb)			
Description	**Total Units**	**Unit RU**	**Total RU**	**Unit CFM**	**Total CFM**	**Unit W**	**Total W**	**Unit kg**	**Total kg**	**Unit lb**	**Total lb**
Sun Fire X4100 compute node	32	1	32	96	3072	550	17600	19.6	627.2	43.1	1379.8
Sun Fire X4100 management node, grid master node	2	1	2	96	192	550	1100	19.6	39.2	43.1	86.2
1U 24-port 4X IB switch	2	1	2	15	30	96	192	7.7	15.4	16.9	33.9
1U Ethernet switch	2	1	2	15	30	160	320	6.4	12.8	14.1	28.2
KMM (keyboard, monitor, mouse)	0	1	0	0	0	22	0	12.4	0.0	27.3	0.0
Cable harness (GigE), IB cables	4	0	0	0	0	0	0	7.5	30.0	16.5	66.0
Rack cabinet with MPS	1	2	2	0	0	25	25	195.0	195.0	429.0	429.0
Total Reference Config System			40		3324		19237		919.6		2023.1
Max rack limits: 1042MPS60A			42				32448		920.8		2025.7
Max rack limits: 1042MPS32A			42				39936		920.8		2025.7

Sun Fire X4100/X4100 M2 Servers with Non-Blocking InfiniBand

This reference configuration illustrates how a full rack of Sun Fire X4100/X4100 M2 servers can be configured with a high performance InfiniBand data fabric using a single non-blocking InfiniBand switch. Given the lower latency and a wider, faster data path, the InfiniBand data fabric is expected to deliver a much higher system efficiency over a Gigabit Ethernet fabric configuration.

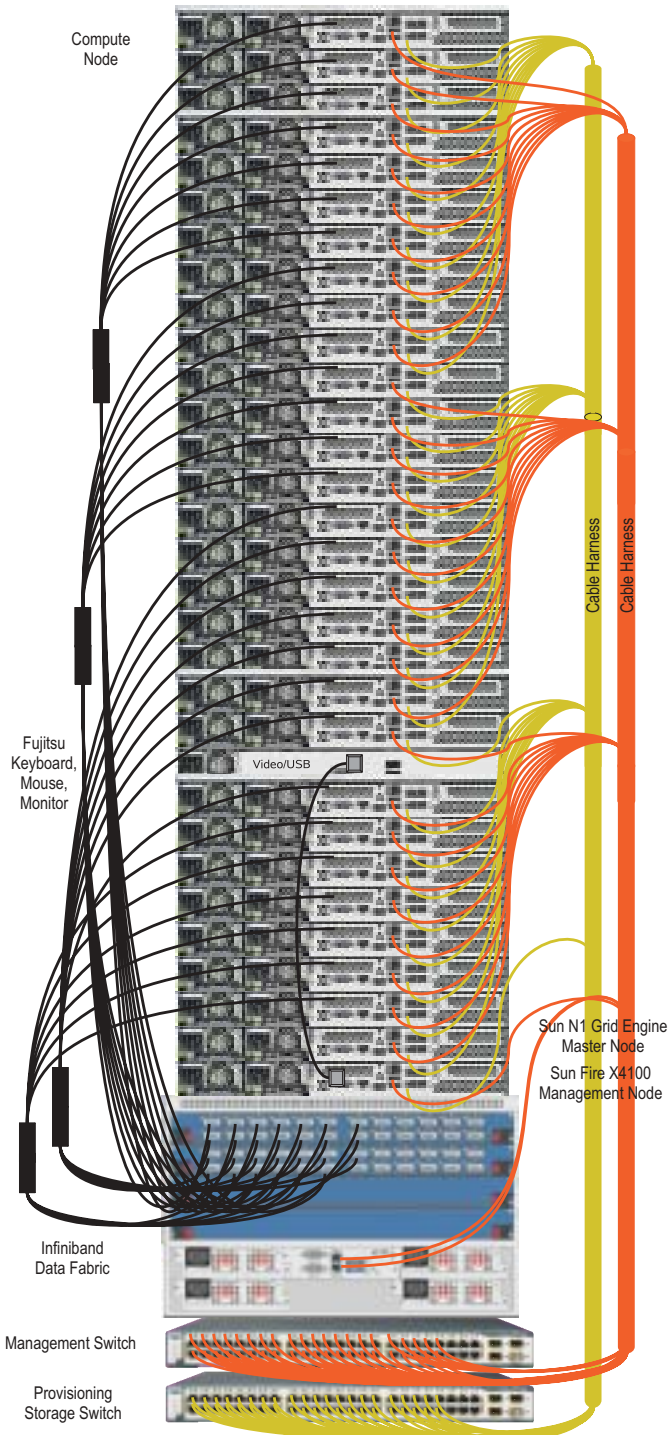
Twenty-eight dual-socket Sun Fire X4100/X4100 M2 servers, each with 2.6 GHz dual-core CPUs, are configured in a Sun Rack 1000-1042 rack as compute servers, with an estimated 0.58 TFLOPS of theoretical peak computational performance. A Sun Fire X4100/X4100 M2 server is connected to the management fabric and serves as the management node, and a second Sun Fire X4100/X4100 M2 server is configured as the grid master node.

Configuration and Cabling Diagram

Figure 4 shows a high level wiring diagram of the rack. The included table details how each component is arranged in the rack, and lists the cable harness mapping to each port of the component. Each of the 28 compute nodes is configured with a single-port Voltaire 4X InfiniBand HCA, wired to a Voltaire 96-port InfiniBand switch in the rack. The 4X HCA card is installed in the expansion slot in the Sun Fire X4100/X4100 M2 server, supporting InfiniBand Single Data Rate (SDR) mode. With SDR, the 4X HCA offers 10 Gigabits per second of connection bandwidth between nodes. The Voltaire 9096 chassis is populated with two 24-port Voltaire InfiniBand switch modules, providing non-blocking InfiniBand connections between all InfiniBand ports. The InfiniBand switch modules are connected together with a non-blocking backplane within the rack between all 28 nodes.

A Gigabit Ethernet port of the compute nodes is connected to a 48-port Gigabit Ethernet switch, configured as the provisioning and storage data fabric for the rack. The management network is on a separate 48-port switch, with dedicated uplink ports to external management connections.

A Sun Fire X4100/X4100 M2 server is setup at the management node, running the Solaris OS and the Sun N1 System Manager software. A second Sun Fire X4100/X4100 M2 server is setup as the master node for the Sun N1 Grid Engine software, providing task scheduling for HPC grid computing.



Loc	Device	F	Dev Port	Data Harness Label	SW	Dev Port	Mgmt Harness Label	SW
U42	X4100	C	0	U-42,Rack<->SW-1,Prt-42	42	SP	U-42,Rack<->SW-2,Prt-42	42
U41	X4100	C	0	U-41,Rack<->SW-1,Prt-41	41	SP	U-41,Rack<->SW-2,Prt-41	41
U40	X4100	C	0	U-40,Rack<->SW-1,Prt-40	40	SP	U-40,Rack<->SW-2,Prt-40	40
U39	X4100	C	0	U-39,Rack<->SW-1,Prt-39	39	SP	U-39,Rack<->SW-2,Prt-39	39
U38	X4100	C	0	U-38,Rack<->SW-1,Prt-38	38	SP	U-38,Rack<->SW-2,Prt-38	38
U37	X4100	C	0	U-37,Rack<->SW-1,Prt-37	37	SP	U-37,Rack<->SW-2,Prt-37	37
U36	X4100	C	0	U-36,Rack<->SW-1,Prt-36	36	SP	U-36,Rack<->SW-2,Prt-36	36
U35	X4100	C	0	U-35,Rack<->SW-1,Prt-35	35	SP	U-35,Rack<->SW-2,Prt-35	35
U34	X4100	C	0	U-34,Rack<->SW-1,Prt-34	34	SP	U-34,Rack<->SW-2,Prt-34	34
U33	X4100	C	0	U-33,Rack<->SW-1,Prt-33	33	SP	U-33,Rack<->SW-2,Prt-33	33
U32	X4100	C	0	U-32,Rack<->SW-1,Prt-32	32	SP	U-32,Rack<->SW-2,Prt-32	32
U31	X4100	C	0	U-31,Rack<->SW-1,Prt-31	31	SP	U-31,Rack<->SW-2,Prt-31	31
U30	X4100	C	0	U-30,Rack<->SW-1,Prt-30	30	SP	U-30,Rack<->SW-2,Prt-30	30
U29	X4100	C	0	U-29,Rack<->SW-1,Prt-29	29	SP	U-29,Rack<->SW-2,Prt-29	29
U28	X4100	C	0	U-28,Rack<->SW-1,Prt-28	28	SP	U-28,Rack<->SW-2,Prt-28	28
U27	X4100	C	0	U-27,Rack<->SW-1,Prt-27	27	SP	U-27,Rack<->SW-2,Prt-27	27
U26	X4100	C	0	U-26,Rack<->SW-1,Prt-26	26	SP	U-26,Rack<->SW-2,Prt-26	26
U25	X4100	C	0	U-25,Rack<->SW-1,Prt-25	25	SP	U-25,Rack<->SW-2,Prt-25	25
U24	X4100	C	0	U-24,Rack<->SW-1,Prt-24	24	SP	U-24,Rack<->SW-2,Prt-24	24
U23	X4100	C	0	U-23,Rack<->SW-1,Prt-23	23	SP	U-23,Rack<->SW-2,Prt-23	23
U22	X4100	C	0	U-22,Rack<->SW-1,Prt-22	22	SP	U-22,Rack<->SW-2,Prt-22	22
U21	KMM		0	U-21,Rack<->SW-1,Prt-21	21		U-21,Rack<->SW-2,Prt-21	21
U20	X4100	C	0	U-20,Rack<->SW-1,Prt-20	20	SP	U-20,Rack<->SW-2,Prt-20	20
U19	X4100	C	0	U-19,Rack<->SW-1,Prt-19	19	SP	U-19,Rack<->SW-2,Prt-19	19
U18	X4100	C	0	U-18,Rack<->SW-1,Prt-18	18	SP	U-18,Rack<->SW-2,Prt-18	18
U17	X4100	C	0	U-17,Rack<->SW-1,Prt-17	17	SP	U-17,Rack<->SW-2,Prt-17	17
U16	X4100	C	0	U-16,Rack<->SW-1,Prt-16	16	SP	U-16,Rack<->SW-2,Prt-16	16
U15	X4100	C	0	U-15,Rack<->SW-1,Prt-15	15	SP	U-15,Rack<->SW-2,Prt-15	15
U14	X4100	C	0	U-14,Rack<->SW-1,Prt-14	14	SP	U-14,Rack<->SW-2,Prt-14	14
U13	X4100	M	0	U-13,Rack<->SW-1,Prt-13	13	SP	U-13,Rack<->SW-2,Prt-13	13
U12	X4100	M	0	U-12,Rack<->SW-1,Prt-12	12	1	U-12,Rack<->SW-2,Prt-12	12
U11	Voltaire 96P IB	X		U-11,Rack<->SW-1,Prt-11	11		U-11,Rack<->SW-2,Prt-11	11
U10				U-10,Rack<->SW-1,Prt-10	10		U-10,Rack<->SW-2,Prt-10	10
U09				U-09,Rack<->SW-1,Prt-09	9		U-09,Rack<->SW-2,Prt-09	9
U08				U-08,Rack<->SW-1,Prt-08	8		U-08,Rack<->SW-2,Prt-08	8
U07				U-07,Rack<->SW-1,Prt-07	7		U-07,Rack<->SW-2,Prt-07	7
U06				U-06,Rack<->SW-1,Prt-06	6	SP	U-06,Rack<->SW-2,Prt-06	6
U05				U-05,Rack<->SW-1,Prt-05	5		U-05,Rack<->SW-2,Prt-05	5
U04	Mgmt 48P switch	X		U-04,Rack<->SW-1,Prt-04	4		U-04,Rack<->SW-2,Prt-04	4
U03	GigE 48P switch	X		U-03,Rack<->SW-1,Prt-03	3		U-03,Rack<->SW-2,Prt-03	3
U02	MPS 60A	P		U-02,Rack<->SW-1,Prt-02			U-02,Rack<->SW-2,Prt-02	
U01	MPS 60A	P						

KEY: P MPS T Terminal Server
X Data Switch M Management Node
X Mgmt Switch C Compute Node

Figure 4. Rack and Cable Mapping: Sun Fire X4100 Servers with Non-Blocking InfiniBand.

Power Cord Connections

The Sun Fire X4100/X4100 M2 server provides power redundancy with an optional second power supply unit (PSU). For HPC, only one power supply unit is configured. Servers are connected in an interleaved arrangement between two circuits in the rack, ensuring half of the servers can continue operation in the case of one circuit failure. Table 6 shows AC power cord connections for the reference configuration. The current draws are distributed to ensure they do no exceed the limits of each plug, strip pair, and circuit maximum specifications. The notation is in the format of Uxx.y, where Uxx describes the component mounted in the location starting at rack unit number xx, and y is the AC power cord number of the unit.

Table 6. AC Power Cord Connections for Configurations with Sun Fire X4100 Servers and Non-Blocking InfiniBand.

Circuit A	Strip						
	0	3	1	4	2	5	Plug
	U03.1	U12.1	U13.1	U23.1	U31.1	U39.1	0
	U06.1		U16.1	U24.1	U32.1	U40.1	1
			U17.1	U25.1	U33.1	U41.1	2
			U18.1	U26.1	U34.1	U42.1	3
							4
							5
Circuit B	Strip						
	0	3	1	4	2	5	Plug
	U04.1	U12.2	U13.2	U19.1	U27.1	U35.1	0
	U06.2		U14.1	U20.1	U28.1	U36.1	1
			U15.1	U21.1	U29.1	U37.1	2
				U22.1	U30.1	U38.1	3
							4
							5

Site Planning Considerations for Sun Fire X4100 Servers with Non-Blocking InfiniBand

Table 7 summarizes the size, airflow, power, and weight of the entry level reference configuration, based on preliminary component specification, and compares them against the rack limits.

Table 7. Site Planning Considerations: Sun Fire X4100 Servers with Non-Blocking InfiniBand.

		Size (RU)		Air Flow (CFM) @35C		Power (W)		Weight (kg, lb)			
Description	**Total Units**	**Unit RU**	**Total RU**	**Unit CFM**	**Total CFM**	**Unit W**	**Total W**	**Unit kg**	**Total kg**	**Unit lb**	**Total lb**
Sun Fire X4100 compute node	28	1	28	96	2688	550	15400	19.6	548.8	43.1	1207.4
Sun Fire X4100 management node, grid master node	2	1	2	96	192	550	1100	19.6	39.2	43.1	86.2
96-port IB switch, 6U + 1U space	1	7	7	190	190	950	950	38.0	38.0	83.6	83.6
1U Ethernet switch	2	1	2	15	30	160	320	6.4	12.8	14.1	28.2
KMM (keyboard, monitor, mouse)	1	1	1	0	0	22	22	12.4	12.4	27.3	27.3
Cable harness (GigE), IB cables	4	0	0	0	0	0	0	7.5	30.0	16.5	66.0
Rack cabinet with MPS	1	2	2	0	0	25	25	195.0	195.0	429.0	429.0
Total Reference Config System			42		3100		17817		876.2		1927.6
Max rack limits: 1042MPS60A			42				32448		920.8		2025.7
Max rack limits: 1042MPS32A			42				39936		920.8		2025.7

Sun Fire X4200/X4200 M2 Servers with InfiniBand

This reference configuration illustrates how a full rack of Sun Fire X4200/X4200 M2 servers can be configured with a high performance data fabric using InfiniBand technology. Given the lower latency and a wider, faster data path, the InfiniBand data fabric is expected to deliver a much higher system efficiency over a Gigabit Ethernet fabric configuration.

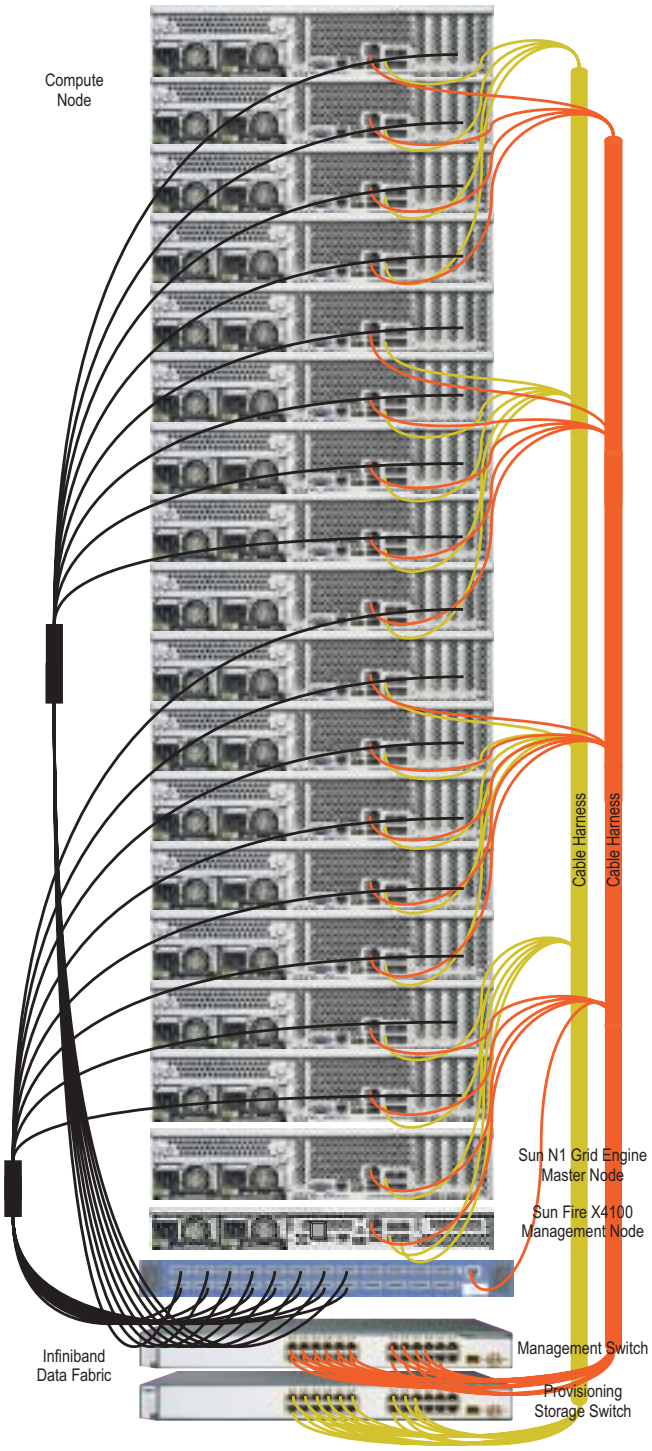
Sixteen dual-socket Sun Fire X4200/X4200 M2 servers, each with 2.6 GHz dual-core CPUs, are configured in a Sun Rack 1000-1042 rack as compute nodes, with an estimated 0.33 TFLOPS of theoretical peak computational performance. A Sun Fire X4100/X4100 M2 server is connected to the management fabric and serves as the management node. A Sun Fire X4200/X42100 M2 server is setup as the master node for the Sun N1 Grid Engine software.

Configuration and Cabling Diagram

Figure 5 shows a high level wiring diagram of the rack. The included table details how each component is arranged in the rack, and lists the cable harness mapping to each port of the component. Each of the 16 compute nodes is configured with a single-port Voltaire 4X InfiniBand HCA, wired to a Voltaire 24-port InfiniBand switch in the rack. The 4X HCA card is installed in the expansion slot in the Sun Fire X4200/X4200 M2 server, supporting InfiniBand Single Data Rate (SDR) mode. With SDR, the 4X HCA offers 10 Gigabits per second of connection bandwidth between nodes. Each of the two 24-port Voltaire InfiniBand switches provides non-blocking connections between 16 compute nodes in the rack. Eight InfiniBand ports are available for external connections, giving an external bisectional bandwidth ratio of 2:1.

A Gigabit Ethernet port of the compute nodes is connected to a 24-port Gigabit Ethernet switch, configured as the provisioning and the storage data fabric for the rack. The management network is on a separate 24-port switch, with dedicated uplink ports to external management connections.

A Sun Fire X4100/X4100 M2 server is setup as the management node, running the Solaris OS and the Sun N1 System Manager software. A Sun Fire X4200/X4200 M2 server is setup as the master node for the Sun N1 Grid Engine software, providing task scheduling for HPC grid computing.



Loc	Device	F	Dev Port	Data Harness Label	SW	Dev Port	Mgmt Harness Label	SW
U42				U-42,Rack<->SW-1,Prt-42	42		U-42,Rack<->SW-2,Prt-42	42
U41				U-41,Rack<->SW-1,Prt-41	41		U-41,Rack<->SW-2,Prt-41	41
U40	X4200			U-40,Rack<->SW-1,Prt-40	40		U-40,Rack<->SW-2,Prt-40	40
U39		C	0	U-39,Rack<->SW-1,Prt-39	39	SP	U-39,Rack<->SW-2,Prt-39	39
U38	X4200			U-38,Rack<->SW-1,Prt-38	38		U-38,Rack<->SW-2,Prt-38	38
U37		C	0	U-37,Rack<->SW-1,Prt-37	37	SP	U-37,Rack<->SW-2,Prt-37	37
U36	X4200			U-36,Rack<->SW-1,Prt-36	36		U-36,Rack<->SW-2,Prt-36	36
U35		C	0	U-35,Rack<->SW-1,Prt-35	35	SP	U-35,Rack<->SW-2,Prt-35	35
U34	X4200			U-34,Rack<->SW-1,Prt-34	34		U-34,Rack<->SW-2,Prt-34	34
U33		C	0	U-33,Rack<->SW-1,Prt-33	33	SP	U-33,Rack<->SW-2,Prt-33	33
U32	X4200			U-32,Rack<->SW-1,Prt-32	32		U-32,Rack<->SW-2,Prt-32	32
U31		C	0	U-31,Rack<->SW-1,Prt-31	31	SP	U-31,Rack<->SW-2,Prt-31	31
U30	X4200			U-30,Rack<->SW-1,Prt-30	30		U-30,Rack<->SW-2,Prt-30	30
U29		C	0	U-29,Rack<->SW-1,Prt-29	29	SP	U-29,Rack<->SW-2,Prt-29	29
U28	X4200			U-28,Rack<->SW-1,Prt-28	28		U-28,Rack<->SW-2,Prt-28	28
U27		C	0	U-27,Rack<->SW-1,Prt-27	27	SP	U-27,Rack<->SW-2,Prt-27	27
U26	X4200			U-26,Rack<->SW-1,Prt-26	26		U-26,Rack<->SW-2,Prt-26	26
U25		C	0	U-25,Rack<->SW-1,Prt-25	25	SP	U-25,Rack<->SW-2,Prt-25	25
U24	X4200			U-24,Rack<->SW-1,Prt-24	24		U-24,Rack<->SW-2,Prt-24	24
U23		C	0	U-23,Rack<->SW-1,Prt-23	23	SP	U-23,Rack<->SW-2,Prt-23	23
U22	X4200			U-22,Rack<->SW-1,Prt-22	22		U-22,Rack<->SW-2,Prt-22	22
U21		C	0	U-21,Rack<->SW-1,Prt-21	21	SP	U-21,Rack<->SW-2,Prt-21	21
U20	X4200			U-20,Rack<->SW-1,Prt-20	20		U-20,Rack<->SW-2,Prt-20	20
U19		C	0	U-19,Rack<->SW-1,Prt-19	19	SP	U-19,Rack<->SW-2,Prt-19	19
U18	X4200			U-18,Rack<->SW-1,Prt-18	18		U-18,Rack<->SW-2,Prt-18	18
U17		C	0	U-17,Rack<->SW-1,Prt-17	17	SP	U-17,Rack<->SW-2,Prt-17	17
U16	X4200			U-16,Rack<->SW-1,Prt-16	16		U-16,Rack<->SW-2,Prt-16	16
U15		C	0	U-15,Rack<->SW-1,Prt-15	15	SP	U-15,Rack<->SW-2,Prt-15	15
U14	X4200			U-14,Rack<->SW-1,Prt-14	14		U-14,Rack<->SW-2,Prt-14	14
U13		C	0	U-13,Rack<->SW-1,Prt-13	13	SP	U-13,Rack<->SW-2,Prt-13	13
U12	X4200			U-12,Rack<->SW-1,Prt-12	12		U-12,Rack<->SW-2,Prt-12	12
U11		C	0	U-11,Rack<->SW-1,Prt-11	11	SP	U-11,Rack<->SW-2,Prt-11	11
U10	X4200			U-10,Rack<->SW-1,Prt-10	10		U-10,Rack<->SW-2,Prt-10	10
U09		C	0	U-09,Rack<->SW-1,Prt-09	9	SP	U-09,Rack<->SW-2,Prt-09	9
U08	X4200			U-08,Rack<->SW-1,Prt-08	8		U-08,Rack<->SW-2,Prt-08	8
U07		M	0	U-07,Rack<->SW-1,Prt-07	7	SP	U-07,Rack<->SW-2,Prt-07	7
U06	X4100	M	0	U-06,Rack<->SW-1,Prt-06	6	1	U-06,Rack<->SW-2,Prt-06	6
U05	IB 24P switch	X		U-05,Rack<->SW-1,Prt-05	5	SP	U-05,Rack<->SW-2,Prt-05	5
U04	Mgmt 24P switch	X		U-04,Rack<->SW-1,Prt-04	4		U-04,Rack<->SW-2,Prt-04	4
U03	GigE 24P switch	X		U-03,Rack<->SW-1,Prt-03	3		U-03,Rack<->SW-2,Prt-03	3
U02	MPS 60A	P		U-02,Rack<->SW-1,Prt-02			U-02,Rack<->SW-2,Prt-02	
U01	MPS 60A	P						

KEY:
P MPS
X Data Switch
X Mgmt Switch
T Terminal Server
M Management Node
C Compute Node

Figure 5. Rack and Cable Mapping: Sun Fire X4200 Servers with InfiniBand.

Power Cord Connections

The Sun Fire X4200/X4200 M2 server provides power redundancy with an optional second power supply unit (PSU). For HPC, only one PSU is configured. Servers are connected in an interleaved arrangement between two circuits in the rack, ensuring half of the servers can continue operation in the case of one circuit failure. Table 8 shows AC power cord connections for the reference configuration. The current draws are distributed to ensure they do no exceed the limits of each plug, strip pair, and circuit maximum specifications. The notation is in the format of Uxx . y, where Uxx describes the component mounted in the location starting at rack unit number xx, and y is the AC power cord number of the unit.

Table 8. AC Power Cord Connections for Configurations with Sun Fire X4200 Servers and InfiniBand.

Circuit A	Strip						
	0	3	1	4	2	5	Plug
	U03.1		U15.1	U23.1	U31.1	U39.1	0
	U05.1	U09.1					1
	U06.1		U17.1	U25.1	U33.1	U41.1	2
	U07.1						3
							4
							5
Circuit B	Strip						
	0	3	1	4	2	5	Plug
	U04.1		U11.1	U19.1	U27.1	U35.1	0
	U05.2						1
	U06.2		U13.1	U21.1	U29.1	U37.1	2
	U07.2						3
							4
							5

Site Planning Considerations for Sun Fire X4200 Servers with InfiniBand

Table 9 summarizes the size, airflow, power, and weight of the entry level reference configuration, based on preliminary component specification, and compares them against the rack limits.

Table 9. Site Planning Considerations: Sun Fire X4200 Servers with InfiniBand.

		Size (RU)		Air Flow (CFM) @35C		Power (W)		Weight (kg, lb)			
Description	**Total Units**	**Unit RU**	**Total RU**	**Unit CFM**	**Total CFM**	**Unit W**	**Total W**	**Unit kg**	**Total kg**	**Unit lb**	**Total lb**
Sun Fire X4200 compute node, grid master	17	2	34	151	2567	550	9350	23.6	401.2	51.9	882.6
Sun Fire X4100 management node	1	1	1	96	96	550	550	19.6	19.6	43.1	43.1
1U 24-port 4X IB switch	1	1	1	15	15	96	96	7.7	7.7	16.9	16.9
1U 24-port Ethernet switch	2	1	2	15	30	100	200	5.5	11.0	12.1	24.2
KMM (keyboard, monitor, mouse)	0	1	0	0	0	22	0	12.4	0.0	27.3	0.0
Cable harness (GigE), IB cables	4	0	0	0	0	0	0	7.5	30.0	16.5	66.0
Rack cabinet with MPS	1	2	2	0	0	25	25	195.0	195.0	429.0	429.0
Total Reference Config System			40		2708		10221		664.5		1461.9
Max rack limits: 1042MPS60A			42				32448		920.8		2025.7
Max rack limits: 1042MPS32A			42				39936		920.8		2025.7

Sun Fire X4600/X4600 M2 Servers with InfiniBand

This reference configuration illustrates how a full rack of Sun Fire X4600/X4600 M2 servers can be configured with a high performance InfiniBand data fabric and InfiniBand switch. Given the lower latency and a wider, faster data path, the InfiniBand data fabric is expected to deliver a much higher system efficiency over a Gigabit Ethernet fabric configuration.

With the InfiniBand switch and management components, nine Sun Fire X4600/X4600 M2 servers can fit in a Sun Rack 1000-1042 cabinet. These servers, each with eight CPUs and dual-core 2.6 GHz processors, are configured as compute nodes and provide an estimated 0.75 TFLOPS of theoretical peak computational performance for the rack.

A Sun Fire X4100/X4100 M2 server is connected to the management fabric and serves as the management node. A second Sun Fire X4100/X41100 M2 server is setup as the master node for the Sun N1 Grid Engine software.

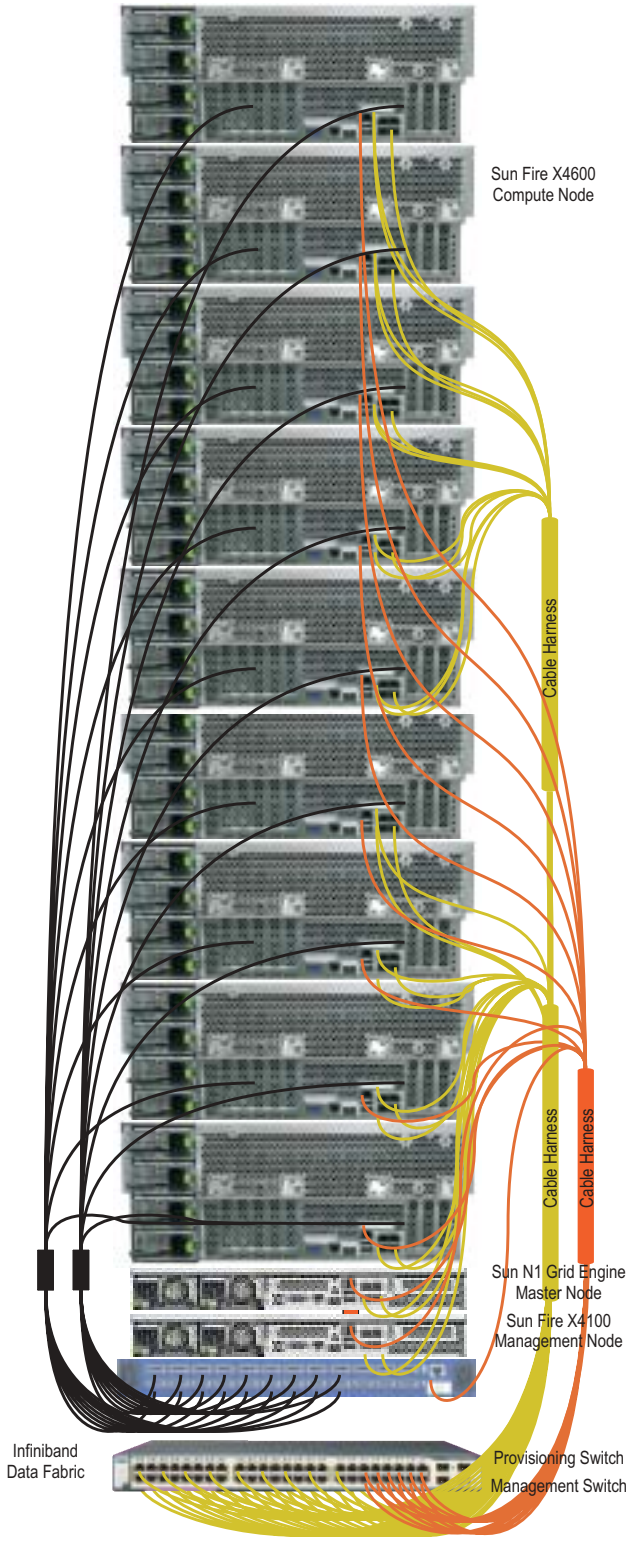
Configuration and Cabling Diagram

Figure 10 shows a high level wiring diagram of the rack. The included table details how each component is arranged in the rack, and lists the cable harness mapping to each port of the component.

Each compute node is configured with two single-port Voltaire 4X InfiniBand HCA cards, wired to a Voltaire 24-port InfiniBand switch in the rack. The 4X HCA cards are installed in slot-2 and slot-5 of the PCI-E slots in the Sun Fire X4600/X4600 M2 server to ensure independent I/O connections, supporting full bandwidth for InfiniBand SDR mode. With SDR, the dual 4X HCA offers 10 Gigabits per second of connection bandwidth between ports. A 24-Port Voltaire InfiniBand switch provides non-blocking connections between each InfiniBand port and each compute node within the rack. With six additional ports for external InfiniBand interface, a 3:1 bisectional bandwidth ratio for each port can be achieved with this InfiniBand configuration, connecting to an external InfiniBand fabric.

Three of the Gigabit Ethernet ports of Sun Fire X4600/X4600 M2 servers are connected to a 48-port switch, configured as the provisioning and the storage data fabric for the rack. Uplink ports provide external connections to the rack. The management network is on a separate VLAN on the same 48-port Gigabit Ethernet switch, with one dedicated up link port to external management connection.

A Sun Fire X4100/X4100 M2 server is setup as the management node, running the Solaris OS and the Sun N1 System Manager software. A second Sun Fire X4100/X4100 M2 server is setup as the master node for the Sun N1 Grid Engine software, providing task scheduling for HPC grid computing.



Loc	Device	F	Dev Port	Data Harness Label	SW	Dev Port	Mgmt Harness Label	SW
U42				U-42,Rack<->SW-1,Prt-42				
U41	X4600		2	U-41,Rack<->SW-1,Prt-41	34			
U40			1	U-40,Rack<->SW-1,Prt-40	33			
U39		C	0	U-39,Rack<->SW-1,Prt-39	32	SP	U-39,Rack<->SW-2,Prt-39	48
U38				U-38,Rack<->SW-1,Prt-38				
U37	X4600		2	U-37,Rack<->SW-1,Prt-37	31			
U36			1	U-36,Rack<->SW-1,Prt-36	30			
U35		C	0	U-35,Rack<->SW-1,Prt-35	29	SP	U-35,Rack<->SW-2,Prt-35	47
U34				U-34,Rack<->SW-1,Prt-34				
U33	X4600		2	U-33,Rack<->SW-1,Prt-33	28			
U32			1	U-32,Rack<->SW-1,Prt-32	27			
U31		C	0	U-31,Rack<->SW-1,Prt-31	26	SP	U-31,Rack<->SW-2,Prt-31	46
U30				U-30,Rack<->SW-1,Prt-30				
U29	X4600		2	U-29,Rack<->SW-1,Prt-29	25			
U28			1	U-28,Rack<->SW-1,Prt-28	24			
U27		C	0	U-27,Rack<->SW-1,Prt-27	23	SP	U-27,Rack<->SW-2,Prt-27	45
U26				U-26,Rack<->SW-1,Prt-26				
U25	X4600		2	U-25,Rack<->SW-1,Prt-25	22			
U24			1	U-24,Rack<->SW-1,Prt-24	21			
U23		C	0	U-23,Rack<->SW-1,Prt-23	20	SP	U-23,Rack<->SW-2,Prt-23	44
U22				U-22,Rack<->SW-1,Prt-22				
U21	X4600		2	U-21,Rack<->SW-1,Prt-21	19			
U20			1	U-20,Rack<->SW-1,Prt-20	18			
U19		C	0	U-19,Rack<->SW-1,Prt-19	17	SP	U-19,Rack<->SW-2,Prt-19	43
U18				U-18,Rack<->SW-1,Prt-18				
U17	X4600		2	U-17,Rack<->SW-1,Prt-17	16			
U16			1	U-16,Rack<->SW-1,Prt-16	15			
U15		C	0	U-15,Rack<->SW-1,Prt-15	14	SP	U-15,Rack<->SW-2,Prt-15	42
U14				U-14,Rack<->SW-1,Prt-14				
U13	X4600		2	U-13,Rack<->SW-1,Prt-13	13			
U12			1	U-12,Rack<->SW-1,Prt-12	12			
U11		C	0	U-11,Rack<->SW-1,Prt-11	11	SP	U-11,Rack<->SW-2,Prt-11	41
U10				U-10,Rack<->SW-1,Prt-10				
U09	X4600		2	U-09,Rack<->SW-1,Prt-09	10			
U08			1	U-08,Rack<->SW-1,Prt-08	9			
U07		C	0	U-07,Rack<->SW-1,Prt-07	8	SP	U-07,Rack<->SW-2,Prt-07	40
U06	X4100	M	0	U-06,Rack<->SW-1,Prt-06	7	SP	U-06,Rack<->SW-2,Prt-06	39
U05	X4100	M	0	U-05,Rack<->SW-1,Prt-05	6	1	U-05,Rack<->SW-2,Prt-05	38
U04	IB 24P switch	X		U-04,Rack<->SW-1,Prt-04	5	SP	U-04,Rack<->SW-2,Prt-04	37
U03	GigE 48P switch	X		U-03,Rack<->SW-1,Prt-03	4			
U02	MPS 60A	P		U-02,Rack<->SW-1,Prt-02	3			
U01	MPS 60A	P						

KEY: P MPS T Terminal Server
X Data Switch M Mgmt Node
X Mgmt Switch C Compute Node

Table 10. Rack and Cable Mapping: Sun Fire X4600 Servers with InfiniBand.

Power Cord Connections

The Sun Fire X4600/X4600 M2 server provides 3+1 power redundancy within the server. Servers are connected in an interleaved arrangement between two circuits in the rack, ensuring half of the servers can continue operation in the case of one circuit failure. Table 11 shows AC power cord connections for the reference configuration. The current draws are distributed to ensure they do no exceed the limits of each plug, strip pair, and circuit maximum specifications. The notation is in the format of Uxx . y, where Uxx describes the component mounted in the location starting at rack unit number xx, and y is the AC power cord number of the unit.

Table 11. AC Power Cord Connections for Configurations with Sun Fire X4600 Servers and InfiniBand.

Circuit A	Strip						
	0	3	1	4	2	5	Plug
	U03.1	U07.1	U15.1	U23.1	U31.1	U39.1	0
	U04.1	U07.2	U15.2	U23.2	U31.2	U39.2	1
	U05.1	U07.3	U15.3	U23.3	U31.3	U39.3	2
	U06.1	U07.4	U15.4	U23.4	U31.4	U39.4	3
							4
							5
Circuit B	Strip						
	0	3	1	4	2	5	Plug
	U04.2	U11.1	U19.1		U27.1	U35.1	0
	U05.2	U11.2	U19.2		U27.2	U35.2	1
	U06.2	U11.3	U19.3		U27.3	U35.3	2
		U11.4	U19.4		U27.4	U35.4	3
							4
							5

Site Planning Considerations for Sun Fire X4600 Servers with InfiniBand

Table 12 summarizes the size, airflow, power, and weight of the entry level reference configuration, based on preliminary component specification, and compares them against the rack limits.

Table 12. Site Planning Considerations: Sun Fire X4600 Servers with InfiniBand.

		Size (RU)		Air Flow (CFM) @35C		Power (W)		Weight (kg, lb)			
Description	**Total Units**	**Unit RU**	**Total RU**	**Unit CFM**	**Total CFM**	**Unit W**	**Total W**	**Unit kg**	**Total kg**	**Unit lb**	**Total lb**
Sun Fire X4600 compute node	9	4	36	474	4266	1715	15435	48.1	432.9	105.8	952.4
Sun Fire X4100 management node, grid master	2	1	2	96	192	550	1100	19.6	39.2	43.1	86.2
1U 24-port 4X IB switch	1	1	1	15	15	96	96	7.7	7.7	16.9	16.9
1U Ethernet switch	1	1	1	15	15	160	160	6.4	6.4	14.1	14.1
KMM (keyboard, monitor, mouse)	0	1	0	0	0	22	0	12.4	0.0	27.3	0.0
Cable harness (GigE), IB cables	3	0	0	0	0	0	0	7.5	22.5	16.5	49.5
Rack cabinet with MPS	1	2	2	0	0	25	25	195.0	195.0	429.0	429.0
Total Reference Config System			42		4488		16816		703.7		1548.1
Max rack limits: 1042MPS60A			42				32448		920.8		2025.7
Max rack limits: 1042MPS32A			42				39936		920.8		2025.7

Default Operating System Configuration

The following section includes information on the default configuration when using the Solaris OS. Similarly, “Red Hat Linux Operating System” on page 31 and “SUSE Linux Operating System” on page 32 contain information on the default configurations using Red Hat and SUSE Linux Operating Systems.

Solaris Operating System

Table 13 lists the Solaris OS image settings used for the Sun Customer Ready HPC Cluster reference configurations.

Table 13. Solaris OS Image Settings.

Version	Distribution	Locale	Timezone
Solaris 10 U2 64b	SUNWXcall	en_US	EDT
Solaris 10 U3 64b	SUNWXcall	en_US	EDT

File System Layout for Solaris OS

The default file system layout for a compute node and the Sun N1 Grid Engine master node is shown in Table 14. Free disk space can be mapped to unassigned slices by the user depending on the usage.

Table 14. File System Layout for Compute and Grid Master Nodes.

Slice	Tag	Mount	Size	Logging	Purpose
0	root	/	10 GB	Yes	/ UFS
1	swap		8 GB	NA	SWAP space
2	backup		Entire disk	NA	Entire disk
3	unassigned		0	NA	
4	unassigned		0	NA	
5	var		6 GB	Yes	Variable data
6	unassigned		0	NA	
7	unassigned		0	NA	
8	boot		7.84 MB	NA	Boot kernels
9	alternates		15.69 MB	NA	Use default

Table 15 shows the default file system layout for the management node running the Sun N1 System Management software. Because the Sun N1 System Management software loads the operating system and application images in the `/var` partition, this partition is configured as the remaining disk space.

Table 15. File System Layout for the Management Node.

Slice	Tag	Mount	Size	Logging	Purpose
0	root	/	10 GB	Yes	/ UFS
1	swap		8 GB	NA	SWAP space
2	backup		Entire disk	NA	Entire disk
3	unassigned		0	NA	
4	unassigned		0	NA	
5	var		Free	Yes	OS and application images for provisioning
6	unassigned		0	NA	Boot kernels
7	unassigned		0	NA	
8	boot		7.84 MB	NA	Boot kernels
9	alternates		15.69 MB	NA	Use default

Red Hat Linux Operating System

Table 16 lists the Red Hat Enterprise Linux OS image settings used for the Sun Customer Ready HPC Cluster reference configurations.

Table 16. Red Hat Enterprise Linux OS Image Settings.

Version	Distribution	Locale	Timezone
RHEL AS 4 U4 64b	Red Hat	en_US	EDT

File System Layout for Red Hat Linux OS

The default file system layout for a compute node and the Sun N1 Grid Engine master node is shown in Table 17.

Table 17. File System Layout for Compute and Grid Master Nodes.

Tag	Mount Point	Size	File System Type	Purpose
root	/	Free	Ext3	Main space and application space
swap		8 GB	swap	SWAP space
var		6 BG	Ext3	Variable data
boot	/boot	100 MB	Ext3	Boot kernels

Table 18 shows the default file system layout for the management node running the Sun N1 System Management software. Because the Sun N1 System Management

software loads the operating system and application images in the /var partition, this partition is configured as the remaining disk space.

Table 18. File System Layout for Management Node.

Tag	Mount Point	Size	File System Type	Purpose
root	/	10 GB	Ext3	Root and applications
swap		8 GB	swap	SWAP space
var		Free	Ext3	OS and application images for provisioning
boot	/boot	100 MB	Ext3	Boot kernels

SUSE Linux Operating System

Table 19 lists the SUSE Linux OS image settings used for the Sun Customer Ready HPC Cluster reference configurations.

Table 19. SUSE Linux OS Image Settings.

Version	Distribution	Locale	Timezone
SUSE 9 SP3 64b	SUSE	en_US	EDT

File System Layout for the SUSE Linux OS

The default file system layout for a compute node and the Sun N1 Grid Engine master node is shown in Table 20.

Table 20. File System Layout for Compute and Grid Master Nodes.

Tag	Mount Point	Size	File System Type	Purpose
root	/	Free	Reiserfs3	Main space and root
swap		8 GB	swap	SWAP space
var		6 GB	Reiserfs3	Variable data

Default Networking and Access Configuration

The following sections include information on the default networking and access configuration, including hostname and IP address assignments.

Default Network Assignment

The default data network in a Sun Customer Ready HPC Cluster is assigned to the 10.10.0.0/16 segment, while the management network is assigned to the 10.11.0.0/16 segment. The management node and the Sun N1 Grid Engine external connection are left unassigned.

Table 21. Default Network Assignment.

Network	Address	Netmask	Comment
External	<IP Address>	<Netmask>	For connecting to external management network
Data	10.10.0.0	255.255.0.0	Data network
Management	10.11.0.0	255.255.0.0	Management network

Default Hostname and IP Address Assignments

While there are numerous hostname and IP address naming conventions for data center equipment, the default scheme for the Sun Customer Ready HPC Cluster is defined for easy tracking of components. For hostnames, the components use a naming convention associated with the function, rack number, the rack unit (RU) location, and the interface:

`<function>-<rack number>-<RU location>-<Interface>`

The `<function>` describes the component's role and the `<rack number>` reflects the rack number in the data center configuration. Starting at the bottom of the rack, the `<RU location>` is the lowest rack unit location the component occupies in the rack. For example, a 4-RU component mounted in the location 15 to 18 RU will have a 15 in its assignment. The `<interface>` maps to a port of the component.

The IP Address uses the `<rack number>` and `<RU location>` as part of the assignment.

Management Node Assignment

Table 22 lists the hostname and IP address information for the management node.

Table 22. Management Node Assignment.

Hostname	Address	Interface	Netmask	Comment
manage-1-07-d2	<IP Address>	<int>2	<Netmask>	For connecting to external management network
manage-1-07-d0	10.10.1.7	<int>0	255.255.0.0	Data network
manage-1-07-d1	10.11.1.7	<int>1	255.255.0.0	Management network

Sun N1 Grid Engine Master Node Assignment

Table 23 lists the hostname and IP address information for the Sun N1 Grid Engine master node.

Table 23. Sun N1 Grid Engine Master Node Assignment.

Hostname	Address	Interface	Netmask	Comment
n1ge-1-08-d2	<IP Address>	<int>2	<Netmask>	For connecting to external user network, submitting jobs
n1ge-1-08-d0	10.10.1.8	<int>0	255.255.0.0	Data network
n1ge-1-08-sp	10.11.1.8	sp	255.255.0.0	Management network

Compute Node Hostnames and IP Addresses

Table 24 lists the hostname and IP address information for the compute nodes. The data port on each compute node has this naming convention, starting at the bottom of the first rack.

Table 24. Compute Node Assignment.

Hostname	Address	Interface	Netmask	Comment
comp-1-09-d0	10.10.1.9	<int>0	255.255.0.0	Data network default
comp-1-10-d0	10.10.1.10	<int>0	255.255.0.0	Data network default
...	...			
comp-1-42-d0	10.10.1.42	<int>0	255.255.0.0	Data network default

Service Processor Hostnames and IP Addresses

Table 25 lists the service processor hostname and IP address information. The service processor in each server uses this convention, starting at the bottom of the first rack.

Table 25. Service Processor Assignment.

Hostname	Address	Interface	Netmask	Comment
comp-1-09-sp	10.11.1.9	sp	255.255.0.0	Management network default
comp-1-10-sp	10.11.1.10	sp	255.255.0.0	Management network default
...	...			
comp-1-42-sp	10.11.1.42	sp	255.255.0.0	Management network default

Switches and Terminal Server Hostnames and IP Addresses

Table 26 lists the hostname and IP address information for the switches and terminal server. The terminal server, network switches, and InfiniBand switches use this naming convention starting at the bottom of the first rack.

Table 26. Switches and Terminal Server Assignment.

Hostname	Address	Interface	Netmask	Comment
datasw-1-03-d	10.11.1.3	Enet port	255.255.0.0	Data switch default
managesw-1-04-m	10.11.1.4	Enet port	255.255.0.0	Management switch default
termserver-1-05-sp	10.11.1.5	sp	255.255.0.0	Terminal server default
ibsw1-1-06-sp	10.11.1.6	sp	255.255.0.0	InfiniBand switch default

Service Processor Access

Table 27 lists the default root password for the Service Processor.

Table 27. Service Processor Access.

User Name	Password
root	changeme

Sun N1 Grid Engine Software Defaults

Table 28 lists the default password information for the Sun N1 Grid Engine software.

Table 28. Root Password for the Sun N1 Grid Engine Software.

User	User Name	Password
root	root	admin
grid engine admin user	sgeadmin	sgeadmin

Default Values

The Sun N1 Grid Engine software is installed on the head node in the `/opt/N1GE` directory. This directory is shared out to the compute nodes using NFS.

The following default values are used to configure the Sun N1 Grid Engine software on the master and compute nodes:

```
SGE_ROOT="/opt/N1GE"
SGE_QMASTER_PORT="536"
SGE_EXECD_PORT="537"
CELL_NAME="default"
ADMIN_USER=""
QMASTER_SPOOL_DIR="/opt/N1GE/default/spool/qmaster"
EXECD_SPOOL_DIR="/opt/N1GE/default/spool"
GID_RANGE="20000-21000"
SPOOLING_METHOD="berkeleydb"
DB_SPOOLING_SERVER="none"
DB_SPOOLING_DIR="/opt/N1GE/default/spooldb"
ADMIN_HOST_LIST=""
SUBMIT_HOST_LIST=""
EXEC_HOST_LIST=""
EXECD_SPOOL_DIR_LOCAL=
HOSTNAME_RESOLVING="true"
SHELL_NAME="ssh"
DEFAULT_DOMAIN="none"
ADMIN_MAIL="none"
ADD_TO_RC="true"
SET_FILE_PERMS="true"
RESCHEDULE_JOBS="wait"
SCHEDD_CONF="1"
SHADOW_HOST=
EXEC_HOST_LIST_RM=""
REMOVE_RC="true"
WINDOWS_SUPPORT="false"
WIN_ADMIN_NAME="Administrator"
```

Parts List

Table 29 lists the major components and parts used in the reference configurations. For an updated and complete list of Sun Customer Ready HPC Cluster supported parts, please refer to the Sun Web site.

Table 29. Parts List.

Description	Configuration Rules	Quantity per HPC Reference Configuration				
		X4100 GigE	X4100 IB	X4100 IB NB	X4200 IB	X4600 IB
Sun Rack 1000-42, MPS 32 Amps (EMEA, APAC - except Japan and Taiwan)	One rack with MPS per config	1	1	1	1	1
Sun Rack 1000-42, MPS 60 Amps (Americas, Japan, Taiwan)	One rack with MPS per config	1	1	1	1	1
Sun Fire X4600: 8x885 (2.6 GHz Dual-Core),16x2GB, 2x73GB SAS, DVD, 4xPSU	9 max SR1042, 8 max SR1038	0	0	0	0	9
Sun Fire X4200: 2x285 (2.6 Ghz), 4x2GB, 2x73GB SAS, DVD, dual PSU	18 max SR1042 9 max SR1038	0	0	0	17	0
Sun Fire X4100: 2x285 (2.6 Ghz), 4x2GB, 2x73GB SAS, DVD, dual PSU	compute node	32	32	28	0	0
Sun Fire X4100 server, 2 Opteron 252 CPUs, 4 GB, 73 GB SAS, DVD, 2 PSU	1 per mgmt, 1 per Sun N1 Grid Engine	2	2	2	1	2
Slide rail kit for X4600	1 per X4600	0	0	0	0	9
Slide rail kit for X2100 and X4x00	1 per server	34	34	30	18	2
Cisco Catalyst 3750 48-port Gigabit Ethernet switch	1 for mgmt, 1-2 for data ports	2	2	2	0	1
Cisco Catalyst 3750 24-port Gigabit Ethernet switch	1 for mgmt, 1-2 for data ports	0	0	0	2	0
Keyboard/Monitor/Mouse shelf unit (1RU). Fujitsu model NC14010-B452/US.	1 per rack, optional	1	1	1	0	0
Management Cable Harness, includes 12 orange Ethernet CAT5e cables	1 per rack	0	0	0	0	1
Management Cable Harness, includes 40 orange Ethernet CAT5e cables	1 per rack	1	1	1	1	0
Data Cable Harness, includes 43 yellow Ethernet CAT5e cables	Up to 2 per rack	1	1	1	1	1
Voltaire HCA 410 PCIX (for X4100/X4200 PCIX slot)	1 per IB port	0	32	28	16	0
Voltaire HCA 410ex PCIe (for X4100 M2, X4200 M2, X4600/X4600 M2)	1 per IB port	0	32	28	16	18
Voltaire 24-port IB Switch module	4 module max	0	0	2	0	0
Voltaire sFB-4 Fabric module	4 modules max	0	0	4	0	0
Voltaire ISR power supply unit	2 module max	0	0	2	0	0
Voltaire ISR9096S-M 96-port IB Switch Chassis	1 per rack	0	0	1	0	0
Voltaire ISR9024S-M 24-port IB Switch	2 max per rack	0	2	0	1	1

Description	Configuration Rules	Quantity per HPC Reference Configuration				
		X4100 GigE	X4100 IB	X4100 IB NB	X4200 IB	X4600 IB
IB cable, 4X	1 per IB port	0	39	28	16	18
Solaris 10 OS DVD Multilingual Media Kit, no hard copy doc or license included	1 per server, OS option	1	1	1	1	1
Sun N1 System Manager license for server with up to 4 sockets	1 per server, management option	34	34	30	18	11
Sun N1 Grid Engine 6 License for Maximum of 51-250 CPUs	1 per management cluster, HPC grid option	1	1	1	1	1
Sun Customer Ready HPC Cluster Ship Kit including documentation for any grid rack	1 per rack	1	1	1	1	1

Ordering Information

As a starting point to configure and get budgetary references, account teams, customers, and partners can use the external online Sun Customer Ready HPC Cluster configuration tool at http://www.sun.com/servers/sungridracksystem/configtool/

All prices and discounts are based on US List Prices and subject to change at any time. They are provided for budgetary reference only, not as a quotation.

To get an official Customer Ready program quotation using only Sun's price-listed products, contact a Sun Center representative who will use the standard WebDesk or PartnerWebDesk process.

About the Author

Jeff Lu is a Senior Staff Engineer in the Systems Group and a technical lead for the Sun Customer Ready HPC Cluster program. Jeff has over twenty years of experience in computing systems product design and development with Digital, Dell, and IBM prior to joining Sun in 2005. Jeff holds degrees in MSEE and MBA.

Acknowledgements

The author would like to thank the Sun Customer Ready HPC Cluster team for reviewing and validating many details in the Reference Configurations. A special thanks to Mark Lamourine for laying out the default network naming convention, and Craig Snow for providing the default OS partitioning template.

References

1. Sun Customer Ready HPC Cluster.
   `http://www.sun.com/servers/cr/hpccluster`
2. Sun Customer Ready HPC Cluster documentation.
   `http://www.sun.com/products-n-solutions/hardware/docs/Servers/grid_rack_system/grid_rack_system/index.html`
3. Documentation for Sun server products, including the Sun Fire X4100, X4200, and X4600 families of servers.
   `http://www.sun.com/products-n-solutions/hardware/docs/Servers/`
4. Sun Customer Ready Program.
   `http://www.sun.com/servers/cr`
5. Online Sun Customer Ready HPC Cluster Configuration Tool.
   `http://www.sun.com/servers/sungridracksystem/config-tool/`

Ordering Sun Documents

The SunDocs℠ program provides more than 250 manuals from Sun Microsystems, Inc. If you live in the United States, Canada, Europe, or Japan, you can purchase documentation sets or individual manuals through this program.

Accessing Sun Documentation Online

The `docs.sun.com` web site enables you to access Sun technical documentation online. You can browse the `docs.sun.com` archive or search for a specific book title or subject. The URL is

`http://docs.sun.com/`

Sun Microsystems, Inc. 4150 Network Circle, Santa Clara, CA 95054 USA **Phone** 1-650-960-1300 or 1-800-555-9SUN (9786) **Web** sun.com

Printed in USA 04/07