



Configuration Guide

Aspen Systems Inc
© Aspen Systems 2009

Table of Contents

<u>Configuration Guide</u>	1
<u>Introduction</u>	1
<u>Configuration Type and Distribution Selection</u>	2
<u>Operating System/Distribution Selection</u>	2
<u>Codes and Testing</u>	6
<u>Remote System Testing</u>	6
<u>Standard Build Selection Option</u>	7
<u>Naming and IP Addressing</u>	8
<u>"aspensys" Account</u>	8
<u>Compiler Installs</u>	8
<u>Cluster/System Naming</u>	9
<u>Secondary Interfaces on Computers</u>	10
<u>IP Addressing</u>	10
<u>File Systems and File Sharing</u>	12
<u>File Sharing</u>	14
<u>Network Configuration</u>	16
<u>E-Mail Configuration</u>	16
<u>Network Time Protocol</u>	16
<u>Ganglia / Web Portal</u>	17
<u>Aspen Utilities Package</u>	17
<u>Master Firewall/NAT Server</u>	18
<u>Run Levels</u>	18
<u>Services and Additional Configuration Options</u>	20
<u>Scheduler</u>	21
<u>Environment Modules</u>	21
<u>ABC and IPMI Packages</u>	23
<u>Aspen Beowulf Cluster Management System</u>	24
<u>Interconnects and MPI Implementations</u>	25
<u>Facilities and Contact Information</u>	29
<u>Aspen Contact & Customer Experience Level</u>	30
<u>Warranty and Additional Support Options</u>	32
<u>Additional Support Options</u>	32
<u>Comments</u>	33

Configuration Guide

Introduction

Specifying and purchasing a High Performance Computing (HPC) cluster that meets your needs can be time consuming and labor intensive. To help with this process, Aspen provides you with four tools that can help you at different periods in your procurement. This is the Configuration Guide, which you can use to learn about the choices that you will make when filling out your Statement of Work.

- **Buyers Guide (PDF) (One Online Page)** – Your sales engineer may have you look at the Buyers Guide on-line while he or she discusses your options and requirements.
- **Detailed Buyers Reference (PDF) (One Online Page)** – Useful information that will help you better understand your needs and how Aspen HPC solutions can be used to meet them. This is the detailed explanation of choices outlined in this Buyers Guide and is available via our web site or as a downloaded document.
- **Configuration Guide (PDF) (One Online Page) (this document)** – The Configuration Guide is a detailed explanation of specific information we will need in your Statement of Work. Most sections of the Configuration Guide correspond to a section of the Statement of Work. Aspen will use your completed Statement of Work and your quotes to perform a final engineering review of your system before it is built. We ask that you provide us a completed Statement of Work before or with your Purchase Order so that we may complete your final engineering review as expeditiously as possible. If you have any issues filling out your Statement of Work, your sales engineer or Aspen production engineers will help you.

Use the Configuration Guide, this document, to understand your choices. Then, when you decide to purchase, use the SOW to select your system configurations and record your special configuration requirements or desires. We want to make sure your delivered system meets your specific needs, so please take some time to read this document, and contact your sales engineer to discuss any questions you may have. We will be happy to arrange a conference call or begin an email dialog to address your questions.

Configuration Type and Distribution Selection

There are several types of configurations you may purchase from us. Each of the configuration types requires a different level of customization and technical interaction to complete. You will be asked to choose which type of configuration you require when you fill out your SOW.

Bare Metal configurations have no Operating System (O.S.) installed or configured, and require you to install all software after delivery. We install an operating system for testing and qualification assurance purposes, then delete the O.S. before we ship the system(s) to you.

Base O.S. configurations have our test O.S. installed when we ship the system(s) to you, but we don't configure any additional software. You may also select the specific O.S. you wish to have installed.

A **Cluster Expansion** configuration consists of nodes you are adding to an existing cluster. We will need to contact you to verify our original disaster recovery image validity or retrieve a current image from your cluster. A great deal of additional configuration may be required if the your new nodes differ in major respects; for instance, mixing 32 and 64 bit architectures. If your current cluster O.S. is too old, it may not include drivers for newer hardware or contain current security or performance enhancements. While it will incur additional cost, an entire cluster O.S. refresh might be beneficial to you in some cases, and we can work with you to determine if that is applicable to your situation, and if so, how best to implement that upgrade.

A **Full Cluster** configuration is a fully functional cluster. We perform all unit tests as well as configure and test the entire system as a cluster. The remainder of this guide deals with the “Full Cluster” configuration.

Operating System/Distribution Selection

Many cluster vendors offer only one operating system choice to their customers. Aspen offers several choices. We understand that you may have specific requirements for a particular distribution. Perhaps you have a commercial application that is only certified to function correctly on a particular O.S. or kernel, or you have a code that has only been ported to one particular distribution. Perhaps you simply prefer one O.S. over another because you are familiar with it, like its features, or prefer to administer that distribution. We offer you a choice of several different default distributions, and other operating systems can be loaded at your request.

You may choose from our default list when you fill out your SOW, entering a specific version of your selected O.S if you have one. If you require an O.S. that is not on our default list, we will need to discuss your specific requirements and the trade-offs associated with your choice before your systems are built.

If you choose a commercial distribution, you must either possess or acquire the appropriate quantity and type

of licenses necessary to install your cluster, and have them available for our use during the install of your system(s), or you must have those licenses included in your quote so that we may purchase and install them for you.

Aspen offers WareWulf on CentOS for diskless and single image clusters, [a management system used to provision diskless compute nodes]. We can also configure WareWulf with individual node disks for swap and scratch space as well. Some training is necessary to utilize a single image system, and, while scalable, this configuration is not right for everyone. For optimum performance in fully diskless node operation, you should ensure that your code(s) will execute within physical memory without the use of swap space. In WareWulf, the O.S. resides in physical memory as well. In order to minimize memory footprint, few development or user tools are installed on your node image(s).

If you are considering a single image system, please be aware that many of the shortcomings of a more traditional disked cluster, such as failed node re-imaging, node upgrades, and configuration consistency, are eliminated by the “aspenutil” automated imaging and restore tools we provide with our default O.S. offerings. These tools are installed on all Full Cluster configurations and are free of charge to you as an Aspen Systems customer. Our default O.S. offerings are listed below.

Aspen Default Operating System / Distribution Offerings

<u>O.S.</u>	<u>Remarks</u>
RedHat Enterprise Server (version 4 or newer)	License(s) required for this selection.
SUSE Linux Enterprise Server (version 10 or newer)	License(s) required for this selection.
CentOS (version 4.4 or newer)	No license(s) required.
Fedora Core (Core 6 or newer)	No license(s) required.
OpenSUSE (version 10 or newer)	No license(s) required.
Warewulf (diskless, based on CentOS)	No license(s) required, but an engineer will contact you to discuss your specific requirements.

You are never without support, Aspen supports the O.S. or Distribution you choose for the life of your support contract with us. However, the normal life cycle of patches for the O.S. you choose may limit your future hardware compatibility or performance. The product patch life cycle of each distribution is shown below.

Generalized Product Support Life Cycle

<u>O.S.</u>	<u>Life Cycle</u>
RedHat Enterprise Server (version 4 or newer)	7 years
SUSE Linux Enterprise Server (version 10 or newer)	various
CentOS (version 4.4 or newer)	Follows RHEL matched version
Fedora Core (Core 6 or newer)	1 year
OpenSUSE (version 10 or newer)	2 years
Warewulf (diskless, based on CentOS)	Follows CentOS

A cluster does not normally need to be managed as you would the same number of individual enterprise nodes. The master is treated as a single node for update purposes, and any specialty nodes such as io or admin are also treated as individual nodes. The compute nodes are, however, generally treated as a single entity for upgrade purposes, and Aspen provides tools on our default O.S. choices to facilitate that approach. Compute nodes in a typical HPC cluster are not exposed directly to outside networks. Access is usually given only through a master or gateway node, which is hardened against outside attacks by by a firewall and a restricted set of externally accessible services.

Patch life cycle can also be much less of an issue on a Full Cluster configuration as it might be on, say, an enterprise web server. Many of our more experienced HPC customers stabilize their system(s) on an O.S. and optimized code-base which changes very little throughout the life cycle of the system(s). When a customer does budget an O.S. upgrade from Aspen, it is typically because they have expanded the cluster ("*Cluster Upgrade" configuration type*), or re-purposed an older system after purchasing its replacement.

Sometimes your O.S. or distribution choice might be driven by the hardware you and your sales engineer have selected, or by the capabilities you require. Fedora and OpenSUSE are used to deploy the latest open source utilities and applications for their respective sponsors, and are used to drive upgrades to their enterprise counterparts, so these distributions will often support the latest hardware and performance enhancements.

Fedora has a wide following in the Beowulf community, incorporating many HPC utilities and supporting many others easily. OpenSUSE is targeted more toward the beginning Linux desktop user. Some general recommendations are:

- If you require commercial patch support, select RedHat Enterprise or Novell SUSE Enterprise.
- If you want the latest and greatest hardware support and performance enhancements, select Fedora or, perhaps, OpenSUSE.
- If you have no specific preference, we recommend CentOS. It has a wide user base with a plethora of ported applications and utilities, is extremely stable in our experience, and has an extended patch support life cycle. CentOS is a free clone of RedHat Enterprise Linux.

- If you require diskless or single image capabilities, select our WareWulf offering.

If you chose an O.S. or distribution that is not on the default list, many of the utilities we describe in later sections of the Configuration Guide will not be installed on your delivered cluster. Additionally, the O.S. or distributions may impose specific configuration limitations or guidelines. We will discuss those configurations with you, but your selection of that O.S. or distribution will indicate your consent to those limitations and to our delivery of a system that has less capabilities than the standard Aspen system(s) offerings described in this guide.

[<< Previous](#) | [Next >>](#)

Codes and Testing

Aspen will install up to two (2) additional commercial or open source applications or codes on your cluster free of charge. If the application is commercially licensed, we will require you to provide the licenses and installation packages, or have the application included in your quote so that we may purchase and install it for you. The applications must be supported on your selected distribution or O.S..

We will install additional applications for you, subject to additional charges. Contact your sales engineer to discuss pricing and options. If you choose to have us install any additional applications, we will require you to perform Remote System Testing to ensure that the application(s) is/are installed and functioning correctly prior to shipping.

Remote System Testing

Aspen offers Remote System Testing for your convenience. We highly recommend that you take advantage of this capability.

Aspens experienced engineers at work

After our build, unit, quality assurance, and system(s) testing phases are complete, your system(s) can be connected to a special network that will allow SSH, HTTP, and HTTPS access from the Internet. We will provide a host name for you to use to connect to your system(s), and you can use your Aspen Support Account (ASA) login (or the root password we will provide you via phone) to ssh to your system(s) on our production floor.

This allows you to build and run your code on your cluster before it ever ships. If you have problems, appropriate Aspen engineering resources can then be used to help configure, build, test, tune, or run your code as necessary. This way, your cluster is often ready to go when it is delivered. All you have to do is unpack it, connect power and network, login, and begin computing!

Another major advantage to our remote testing approach is the capability to change software and hardware configurations before you receive the final product. Customers often don't know all their configuration parameters or requirements in advance, and we understand that. Running the codes at our facility prior to shipping will allow us to adjust your cluster environment as necessary to get the best performance, and also familiarize you with your new purchase and how best to operate it.

Standard Build Selection Option

Like most HPC vendors, Aspen offers a standardized build and package selection that follows HPC best practices. However, unlike some other HPC vendors, we also offer you the opportunity to configure your cluster hardware and software the way you want it, with options and capabilities tuned to your needs and your environment.

If you wish a standardized software image that is configured with all recommended options, Aspen can configure your system using best practices outlined in the Detailed Buyers Reference and this Configuration Guide. To select an Aspen standard build for your system(s), determine your hardware configuration with your sales engineer, then fill out your Statement of Work with

- your distribution selection
- any additional applications you require,
- your remote systems testing choice,
- check the “standard build” selection and submit the SOW.

Aspen will configure your system with a software environment that supports your application(s) well and provide long and trouble free service. Many customers value our standard cluster configuration to serve their HPC computing needs, and purchase that option from us again and again.

Naming and IP Addressing

Aspen will configure your system(s) root password(s) and all support equipment logins with the password "Password" if your system(s) will not undergo Remote System Testing before being shipped to you. **You should change this password before you connect your new system(s) to the network at your location.**

If your system(s) will undergo Remote System Testing, a more secure password is configured. We will communicate that password to you on the phone or via other secure means. E-Mail will not be used to give you this root and support equipment password.

"aspensys" Account

Aspen will configure your system(s) with an "aspensys" account that we may use to remotely log in to your system after installation for support, if allowed. Aspen utilizes this account for remote assistance, and it is configured with "sudo" access by default. If remote access is not allowed, you may remove this account after the system(s) arrive(s) at your location. If your cluster has the Aspen Beowulf Cluster Management System installed, the "aspensys" account is required for it to properly operate.

Compiler Installs

Aspen will install any compilers included in your quote for you. We will always install GCC and the standard GNU Fortran compiler(s) included with your distribution. All commercial compilers are installed in "/usr/local/compiler-name" by default.

If you have a license for a compiler that is not included on your quote and wish us to install that compiler for you, we will require that your license be available for installation on your system(s) when we build, and we may require a copy of your compiler installation package if the version is too old or out of production.

Cluster/System Naming

Host names are configured in many different services and applications throughout a cluster, as well as being used to acquire commercial licenses for compilers, schedulers, and other utilities. We configure your system(s) with your selected host name(s) prior to customer test and delivery.

Correct naming is critical in another way as well. Aspen tracks everything you purchase by Job Number and host name, and that data is entered into our production testing suite during unit testing. We also label each system with its name for you. After delivery the Job Number and system name is used to uniquely identify that unit for support should you require it.

Aspen standard naming is "master" for the master node, "node1" for the first node, "node2" for the second node, and so on. The next table outlines our standard name for each type of unit you purchase. You may choose to use our standard naming convention, or enter your own convention for each type of system you are purchasing when you fill out the SOW.

Support equipment in your cluster is also named. That includes all peripheral equipment that can be network connected. Our standard for support equipment is to include the rack number and unit number of the system within the cluster. "r3raid4" means that this unit is your 4th external Raid chassis system in the cluster, and it is located in Rack #3.

Aspen Standard Names

Computers

- **Master**
- **io** (*dedicated NFS / file system server*)
- **admin** (*dedicated administration node*)
- **login** (*dedicated login / compile node*)
- **nodeX** (*X is node number*)

Support Equipment

- **rXupsY** (*UPS #Y, located in Rack #X*)
- **rXraidY** (*Raid #Y, located in Rack #X*)
- **rXeswitchY** (*Ethernet Switch #Y, located in Rack #X*)
- **rXiswitchY** (*Infiniband Switch #Y, located in Rack #X*)
- **rXmswitchY** (*Myrinet Switch #Y, located in Rack #X*)
- **rXkswitchY** (*Remote IP KVM Switch #Y, located in Rack #X*)

Secondary Interfaces on Computers

- **mmaster** (*Myrinet interface on master*)
- **emaster** (*2nd Ethernet interface on master*)
- **imaster** (*InfiniBand interface on master*)
- **master-ipmi** (*IPMI interface on master*)
- **mio** (*Myrinet interface on io*)
- **iio** (*InfiniBand interface on io*)
- **eio** (*2nd Ethernet Interface on io*)
- **io-ipmi** (*IPMI interface on io*)
- **madmin** (*Myrinet interface on admin*)
- **iadmin** (*InfiniBand interface on admin*)
- **eadmin** (*2nd Ethernet Interface on admin*)
- **admin-ipmi** (*IPMI interface on admin*)
- **mlogin** (*Myrinet interface on login*)
- **ilogin** (*InfiniBand interface on login*)
- **elogin** (*2nd Ethernet Interface on login*)
- **login-ipmi** (*IPMI interface on login*)
- **mnodeX** (*Myrinet interface on nodeX*)
- **inodeX** (*InfiniBand interface on nodeX*)
- **enodeX** (*2nd Ethernet interface on nodeX*)
- **nodeX-ipmi** (*IPMI interface on nodeX*)

IP Addressing

Normally, Full Cluster configurations are configured with at least one internal network that connects the nodes and the master and is used for cluster only traffic. In many cases, clusters have more than one internal network dedicated to different purposes.

The master node then functions as a Network Address Translation gateway for the other nodes in the cluster, with no direct access from your organizational network to your compute nodes.

The default IP spaces Aspen utilizes for each particular type of cluster network is shown below. Aspen may configure each network on your system(s) with one of the two address ranges shown below for that particular network.

You may also choose your own IP space if direct routing is needed or required by your organization or application requirements.

Aspen Standard Private Internal IP Address Spaces

- 10.10.0.0/16 (255.255.0.0), or 10.20.0.0/16 (255.255.0.0)
(*first IP Subnet – Administrative and/or NFS Ethernet*)
- 10.11.0.0/16 (255.255.0.0), or 10.21.0.0/16 (255.255.0.0)
(*2nd IP Subnet – Ethernet, Myrinet, InfiniBand*)
- 10.12.0.0/16 (255.255.0.0), or 10.22.0.0/16 (255.255.0.0)
(*3rd IP Subnet – Ethernet, Myrinet, InfiniBand*)
- 10.13.1.0/16 or 10.13.1.0/24, or 10.23.1.0/16 or 10.23.1.0/24
(*IPMI / Management subnet*)

By default, we configure your master nodes external interface to request DHCP information to configure its external IP address. You may enter specific IP information for that interface in the SOW, and we will configure your IP addresses for you prior to shipping.

If you wish us to configure your system(s) external IP address for you, we will require the IP address, its subnet mask, and the gateway address you wish the system(s) to be configured with. If you have internal DNS server(s), we recommend you enter those in the SOW at that time as well.

[<< Previous](#) | [Next >>](#)

File Systems and File Sharing

Aspen Systems provides a default RAID and file system layout, and you may choose to accept that configuration, or provide your own in the SOW.

Aspen Systems highly recommends that for Full Cluster configurations your master node O.S./boot disk be a hardware or software RAID1 (mirrored) configuration. The master node contains all the utilities necessary to restore any other node in the cluster¹⁹ in our default configuration, but manual intervention and quite a bit of labor is necessary to configure another node to restore a failed master.

Not using a mirrored configuration for your cluster master disk can result in downtime for your cluster and a great deal of effort on your part to restore that master should the single boot disk fail. All disks are mechanical, and all mechanical devices will eventually fail.

Aspen images your system(s) prior to delivery, and we store those images to allow for disaster recovery purposes in case something should happen to your cluster²⁰. Our disaster recovery image may be old, so newer applications, configuration changes, or users you have added since the disaster recovery image was last taken will not be restored. Your sales engineer can discuss the disadvantages of not having a mirrored master boot/O.S. disk with you in detail, and your quote will probably have a mirrored master O.S./boot configuration.

You may also have either internal RAID configuration(s) or external RAID chassis(s) that you wish to use for application and user data. Aspen recommends that you utilize a least one hot spare disk for all RAID levels other than RAID1. Disks normally account for approximately 50% of the failures in a cluster, and disks from the same lot or of the same type often have failures at nearly the same time.

Standard RAID5, for instance, protects you from a single disk failure, and your RAID system will be configured to notify you and Aspen in case of a disk failure²¹. Given the tendency of disks to fail within a short time of each other, use of one, or better still, multiple RAID hot spares protects you in these situations.

In the past, smaller disks and easily corruptible file systems drove many original Unix administrators to implement multiple file system mounts to minimize collateral damage in the event of a run-away process that filled a file system, or to decrease file system check time in the event of a corrupted file system during the boot process.

Aspen provides a default file system layout that takes advantage of modern file system reliability and the larger disk sizes available today. Our default configuration is shown next.

Default Configuration

<u>Device</u>	<u>FS</u>	<u>Mount</u>	<u>Size</u>	<u>Remarks</u>
/dev/md0 (software RAID)				<i>If your system is configured with software RAID, /dev/md0 will consist of the 1st partitions on the first two disks. If your system has hardware RAID, a single device will be presented to the O.S., and the first partition on that device will be partitioned as /boot.</i>
/dev/sda1 (SATA,SAS,SCSI)	ext3	/boot	128MB	
/dev/hda1 (PATA)				
/dev/md1 (software RAID)				<i>If your system is configured with software RAID, /dev/md1 will consist of the 2nd partitions on the first two disks. If your system has hardware RAID, a single device will be presented to the O.S., and the 2nd partition on that device will be partitioned as swap.</i>
/dev/sda2 (SATA,SAS,SCSI)	swap	N/A	4 GB or 1.5x Mem	
/dev/hda2 (PATA)				
/dev/md2 (software RAID)				<i>If your system is configured with software RAID, /dev/md2 will consist of the 3rd partitions on the first two disks. If your system has hardware RAID, a single device will be presented to the O.S., and the 3rd partition on that device will be partitioned as /.</i>
/dev/sda3 (SATA,SAS,SCSI)	ext3	/	remainder of device	
/dev/hda3 (PATA)				

RAID device on master (if so equipped)

<u>Device</u>	<u>FS</u>	<u>Mount</u>	<u>Size</u>	<u>Remarks</u>
/dev/sdb1 (SATA,SAS,SCSI)	ext3	/home	entire device	<i>You might also choose to use a single RAID1 software configuration (2 disks) for your data directory. In that case, /home would be /dev/md3.</i>
/dev/md3 (software RAID)				

Node disk(s) (if so equipped)

<u>Device</u>	<u>FS</u>	<u>Mount</u>	<u>Size</u>	<u>Remarks</u>
/dev/md0 (software RAID)				<i>If your system is configured with software RAID, /dev/md0 will consist of the 1st partitions on the first two disks. If your system has hardware RAID, a single device will be presented to the O.S., and the first partition on that device will be partitioned as /boot.</i>
/dev/sda1 (SATA,SAS,SCSI)	ext3	/boot	128MB	
/dev/hda1 (PATA)				

/dev/md1 (software RAID)					<i>If your system is configured with software RAID, /dev/md1 will consist of the 2nd partitions on the first two disks. If your system has hardware RAID, a single device will be presented to the O.S., and the 2nd partition on that device will be partitioned as swap.</i>
/dev/sda2 (SATA,SAS,SCSI)	swap	N/A	4 GB or 2x Mem		
/dev/hda2 (PATA)					
/dev/md2 (software RAID)					<i>If your system is configured with software RAID, /dev/md2 will consist of the 3rd partitions on the first two disks. If your system has hardware RAID, a single device will be presented to the O.S., and the 3rd partition on that device will be partitioned as /.</i>
/dev/sda3 (SATA,SAS,SCSI)	ext3	/	remainder of device		
/dev/hda3 (PATA)					

File Sharing

In Full Cluster configurations, Aspen configures your master node to NFS export the following directories to the nodes. If you have an io node, it will be configured to export one or more of these file systems. If you have an admin node, that node will export /aspendata to the nodes.

<u>Export</u>	<u>Mounted on nodes</u> <u>as</u>	<u>Remarks</u>
/usr/local	/usr/local	<i>Compilers, libraries, and many utilities are installed in the master /usr/local directory. /usr/local is mounted on the nodes so that they have access to these libraries and utilities.</i>
/home	/home	<i>/home is mounted across all nodes so that user data is available on every node.</i>
/aspendata	/aspendata	<i>/aspendata contains source code for installed applications, node images, and aspen utilities for node imaging..</i>

If you are going to use a parallel file system in your system(s), many other considerations will need to be taken into account. Depending on the parallel file system chosen and your requirements, adjustments to your quote and your configuration may be necessary. Some parallel file systems are commercial products, and may require licensing as well. If you are thinking of utilizing a parallel system in your purchase, discuss it with your sales engineer.

Network Configuration

Aspen configures all Full Cluster configurations to allow host based secure shell (SSH) authentication between all nodes for all users. All Aspen installed tools utilize secure shell to perform their functions. The "r" commands are used on some clusters by users or applications, but we do not enable them by default due to possible security risks.

If you require host based remote shell (RSH,REXEC,RLOGIN) between nodes due to your application, users, or administration preferences, you may select that option in your SOW.

E-Mail Configuration

Aspen recommends that you allow the master node in a Full Cluster configuration to e-mail both you and Aspen Systems so that monitoring applications can notify you, and us, of events such as disk failures. Depending on your internal network, firewall, and e-mail server configurations, this may require an internal mail relay host on your network. You may choose the e-mail configuration for your system and enter specific E-Mail settings in your SOW.

Network Time Protocol

Aspen configures your system(s) with the network time protocol (NTP) daemon included with your distribution, and maintains ntp1.aspsys.com and ntp2.aspsys.com for your system(s) to use as time synchronization sources.

In a Full Cluster configuration, all nodes are configured to synchronize from the master node, which in turn receives its time information from Aspen and two other public NTP servers. You may also choose time servers inside your network or publicly available NTP servers, or any combination of Aspen, public, and internal servers as your time source in your SOW.

Ganglia / Web Portal

Aspen installs Ganglia on Full Cluster configurations by default. The Ganglia install can be reached by "http://your-cluster-external-ip-address/ganglia" after the system is delivered, or via "http://cXX.aspsys.com/ganglia" while in Remote System Testing.

Aspen also installs the Apache Web server on your master with a simple web portal that consolidates all the various web applications on your master into a single page with links for your convenience. The web portal is at the document root directory of the server, and can be reached at "http://your-cluster-external-ip-address/" or via "http://cXX.aspsys.com/" while in Remote System Testing.

Aspen Utilities Package

Aspen installs the "aspenutil" package on all Full Cluster configurations by default. We highly recommend that you select this package in your SOW. This package will allow you to copy any node to a re-usable "image" and then restore that image onto the same or a different node. Aspenutil can also perform "bare metal restores", where the failed or new node has no preexisting configuration at all. This capability is used to restore an existing failed node disk, for instance, or to add additional nodes to your cluster.

It can also be used to upgrade your entire cluster by updating a single node, copying it, then restoring every other node of that type with the newer image.

The images contain meta-data for file system re-creation, and our utility changes IP addresses and host names, so no additional configuration is necessary after the restore is complete. It is common to entirely re-image a 256 node cluster in less than an hour using this technology. Installation and operation of "aspenutil" also requires that your cluster be configured with an internal cluster-only network. Use of "aspenutil" does require certain services to be enabled on the master node, listed below.

- DHCP Server – the master node is configured with a DHCP server that serves the configured IP address(s) and host names for all nodes. A standard network layout that utilizes an internal cluster network is required to prevent the master node from serving DHCP requests on your external network. The DHCP server is enabled only on your master internal interface(s), not on its external network interface.
- TFTP Server – the master node is configured with a tftp server that is configured to service tftp requests on the internal cluster network.

- NFS Server – the master, or admin, node is configured to serve the /aspendata/ directory, via nfs, to the internal cluster network.

Master Firewall/NAT Server

Aspen configures all Full Cluster configuration master nodes with a basic "iptables" packet firewall package. If your O.S. or distribution includes packages to manage this firewall via a GUI, that package will be used to configure the firewall. The firewall allows basic services such as SSH (port 22), HTTP (port 80), and HTTPS (port 443), and may allow other port access as required by your configuration.

The master node is also normally configured as a NAT server that masquerades all internal node IP addresses on the internal network so that they may communicate with the outside world. The default configuration does not allow an internal host to act as a server to external nodes.

Run Levels

Most Linux systems utilize run levels to control what services are started after a node is booted. Generally, run level 0 is halted, 1 is single user mode, 2 is single user with networking, 3 is multiuser mode without any X Windows Systems UI started, 4 is unused, 5 is multi-user with X Windows started, and run level 6 is used to reboot the system.

Aspen configures your master node to boot into run level 5 (X windows GUI started) and your nodes to boot into run level 3 (multi-user, no X Windows GUI) unless you select differently in your SOW. The following table shows the default Aspen run-levels by node type.

Node Type	Default Run Level
master	5
io	3
admin	3
node	3

[<< Previous](#) | [Next >>](#)

Services and Additional Configuration Options

Aspen installs and enables the appropriate monitoring on Full Cluster system(s) if that particular unit type is included in your quote. We recommend that you learn and utilize these monitoring tools, but you may choose not to have them installed. All support equipment that is network connected will be configured with internal cluster network addresses by default. Many units have their own HTTP, SSH, and telnet servers that can be used for management and troubleshooting.

- **3Ware 3DM** – If your quote contains a 3Ware RAID controller card, we install and configure the 3DM utility. It is a daemon that runs on the node containing the 3Ware card, and provides both command line (`tw_cli`) and HTTP (`http://your-external-master-name:888/`) connectivity. You may use either utility to check and configure your RAID subsystem, and the daemon also provides email notification to configured addresses on RAID events if e-mail is properly configured and enabled on your master node. RAID disks don't fail that often, but when they do you may not notice the first failure unless you have monitoring enabled. The 3DM web utility will be configured with the system(s) root password for configuration access.
- **Infortrend RaidWatch** – If your quote contains an external Infortrend RAID subsystem, we install and configure the RAIDWatch utility. It is a daemon that runs on the master or admin node, and provides java interfaces to configure RaidWatch (`/usr/local/infortrend/configuration.sh`) or configure your RAID subsystem itself (`/usr/local/infortrend/raidwatch.sh`). After installation, the RaidWatch daemons monitor the Infortrend RAID subsystem for RAID events, and e-mails them to you and to us if the system has e-mail configured and enabled. Newer versions of RaidWatch run directly on the RAID subsystem itself, and can be configured to e-mail alerts if the cluster has e-mail configured and enabled.
- **APCUPSD** – If your quote contains an APC UPS system, we install and configure the "APCUPSD" utility. It is a daemon that runs on the master, and provides both command line (`apcaccess`) and HTTP (`http://your-external-ip-address/apcupsd/`) connectivity. You may use either utility to check and configure your UPS, and the daemon provides power event notification and automatic shutdown of your cluster. APCUPSD is normally used to monitor only one UPS. If you have multiple UPS units to protect your entire Full Cluster system(s), you may choose to monitor only the UPS connected to the master, and configure APCUPSD to shutdown the entire cluster based on master power events. Monitoring multiple UPS systems may require that the software be installed on multiple hosts, and the configuration for automated shutdown with multiple UPS units can become quite complex.
- **PowerAlert** – If your quote contains a Triplite UPS system, we install and configure the "PowerAlert" daemon. PowerAlert is not accessible as a web page, It is a java application that is installed on the master node, and monitors and e-mails power alerts to you.
- **Managed Switch(es)** – If your quote contains any managed switch such as InfiniBand, Myrinet, or Ethernet, the switch(es) will be configured with an internal cluster network IP to be used for network access. These units can be accessed by utilizing a VNC session to the master to display a web browser that can access the internal IP space, or, if ABC is installed, through the "tools" menu on the ABC web page.

Scheduler

Aspen installs and enables the Torque Resource Manager and Maui Cluster scheduler by default on all Full Cluster configurations. For more complex scheduling requirements, Aspen recommends Moab, a commercial offering from the company that supports Torque and Maui, and we can install and configure PBS Pro, SLURM, Sun Grid Engine, LoadLeveler, N1GE, or Platform LSF as well.

Moab, PBS Pro, LoadLeveler, and Platform LSF are commercial products and you must supply licenses for the installation system(s), or your quote must include the software so that we may purchase and install them for you.

Environment Modules

Aspen installs and configures environment modules for your use in your Full Cluster system(s) by default. Environment modules are used to change any particular cluster users search paths and environment variables to support their specific models or applications.

A users path depends on what modules are loaded in that users shell configuration files. This provides a standard way to handle multiple compilers and MPI implementations needed for different codes by different users.

For instance, you might require a particular MPI, let's say Intel optimized 64 bit InfiniBand Open MPI, to run a particular code. The module name for this MPI would be called "openmpi-ib-intel", and you could;

- **module load openmpi-ib-intel** *#places the environment module in your current shell*
- **module initadd openmpi-ib-intel** *#places the environment module in your startup for all shells*

That command would modify your "PATH" and "LD_LIBRARY_PATH" variables, modify your man path so that "man mpiexec" would show the proper manual page, and set a \$MPIHOME variable so that your make files could include "-I\$MPIHOME/include -L\$MPIHOME/lib -lmpi" to allow proper compilation. "module unload" and "module initrm" would be used to unload this module from the current and all succeeding shells.

ABC and IPMI Packages

Intelligent Platform Management Interface, or IPMI, is a standard that allows for full remote node control, configuration, and monitoring. We highly recommend this capability for system(s) installed in remote locations, lights out facilities, or any location where you will have difficulty physically accessing your system(s) after installation. IPMI capabilities allow you to much more easily manage your cluster in almost all situations, so Aspen recommends that IPMI be installed on your cluster if your budget allows. IPMI is required for ABC installations.

When IPMI is installed, each system has a Baseboard Management Controller (BMC) that utilizes the node power supply, connects to the nodes motherboard for power control, monitoring, and console access, and has its own dedicated Ethernet connection to the internal cluster network or vamps from the nodes first Ethernet interface. Each BMC has a discrete IP address assigned, and can be thought of as a small host computer in and of itself, whose sole job is to manage the node it is installed in.

If IPMI is included in your quote, your system will be configured with management tools to flash the BMCs, set their IP address while the system is booted into its normal O.S., and command line and GUI tools that can be used from the master to remotely control power on the node(s), connect to the SOL (serial over LAN) console if that has been enabled on the host, or pull remote sensor data for node troubleshooting.

Some of our IPMI BMCs provide remote KVM over LAN capabilities, serving a web page that allows full remote video, keyboard, and mouse access to the system(s).

IPMI capabilities may be especially useful to you if,

- you are a less experienced HPC administrator, and
- will allow Aspen Systems Technical Support to have remote access to your cluster so that we may remotely assist you, and
- believe that you will utilize our support services often, perhaps due to lack of familiarity with the system(s) or cluster administration in general.

Given remote access and full IPMI KVM over LAN capabilities on your system(s), Aspen can help with virtually any troubleshooting and fault correction that may occur. If your system(s) do(es) not have IPMI, we may require you to assist us in actions we cannot perform remotely. If we are not allowed remote access to your cluster, you will be forced to perform all software and hardware troubleshooting with us on the phone. On average, case resolution time is significantly quicker if our Technical Support staff has remote access to your IPMI enabled system(s).

IPMI BMCs can be connected to the cluster internal network and IP addressed appropriately, or they can be connected to your network, and assigned an IP address from your organizational space.

When configured with internal cluster only IP addresses and connected to a dedicated management network or the cluster internal network, system IPMI interfaces will not be reachable from the external network, and require that you first connect to the master node via its external interface, then utilize tools on the master node to connect to all other nodes.

Some customers connect only their master node IPMI BMC to the external network so that they may remotely connect to it in event of its failure, and connect all other nodes interfaces to the cluster only internal network. The master node is necessary for proper cluster functioning, so if your master node is down, the cluster is non-functional in most cases.

If you wish your IPMI interfaces on all nodes to be connected to the organizational network, you must provide IP addresses, subnet mask(s), and gateway information to us, as well as coordinate for at least one external network connection to be connected to your IPMI Ethernet switch. In that case, please contact your Sales Engineer to discuss your options and configuration in more detail.

Aspen Beowulf Cluster Management System

The Aspen Beowulf Cluster (ABC) Management System⁴¹ is our commercial cluster management tool. It is designed to allow simplified cluster administration via a secure web server from anywhere in the world. ABC requires that your cluster be configured with IPMI.

ABC has a myriad of features, some of which are remote console via web browser, automated monitoring and fault notification, GUI based node re-imaging⁴², cluster package synchronization, a PAM based web scheduler interface⁴³, full node remote control for remote or lights out operation⁴⁴, resource graphing, SNMP trap and event management, and the ability to interface with your UPS and PDU systems. Contact your sales engineer for a real time demo of ABC on one of our test clusters.

If ABC is included in your quote, it will be configured and installed on your Full Cluster system(s).

Interconnects and MPI Implementations

Aspen supports Ethernet, InfiniBand, Myrinet 2G, and 10G interconnects, and installs the appropriate low level drivers for the interconnects included in your quote.

On InfiniBand, Aspen recommends SDR and DDR Qlogic and Infinipath PCI-E and HyperTransport adapters teamed with Qlogic InfiniBand switches. The appropriate stack will be installed to support your chosen MPI(s), either OFED, Quicksilver, or InfiniPath. On Myrinet we install the Myrinet Express (MX) driver software. Most MPI implementations include drivers for Gigabit Ethernet by default.

Your choice of MPI implementation(s) can greatly affect your application efficiency. Aspen Systems will install, configure, test, and support up to three (3) MPI implementations on your Full Cluster system(s) for free, with additional MPI implementations available at additional cost⁴⁸.

For each commercial compiler quoted on your system, as well as GNU GCC, g++, or gfortran, Aspen will install and configure, if applicable, the MPI implementation(s) you select for every interconnect on the system.

So, for instance, if you have an Intel compiler and an InfiniBand interconnect and selected MPICH and Open MPI as your MPI implementations, you would have;

- MPICH (*built with GNU for Ethernet*)
- MPICH (*built with Intel for Ethernet*)
-
- **....something missing??** (*see below*)
- Open MPI (*built with GNU for Ethernet*)
- Open MPI (*built with Intel for Ethernet*)
- Open MPI (*built with GNU for InfiniBand*)
- Open MPI (*built with Intel for InfiniBand*)

Note that MPICH isn't being built for InfiniBand (*..something missing??*) in the example above, because MPICH does not support the InfiniBand interconnect. MVAPICH is the port of MPICH to InfiniBand, and is a totally separate MPI implementation.

Both 32 and 64 bit builds are possible, although perhaps not recommended, for each of these options when your quoted system(s) are X86_64 architecture.

As you can see, a simple set of 3 MPI implementations using 2 different compiler suites (GNU and Intel) on an x86_64 architecture Full Cluster configuration with 2 interconnects (Ethernet and a low latency interconnect such as InfiniBand or Myrinet), could result in more discrete MPI installations than one might immediately expect. The formula looks like this;

$$\begin{aligned} & \mathbf{3} \text{ MPI implementations} \times \mathbf{2} \text{ Compilers (GCC, Intel)} \times \mathbf{2} \text{ Interconnects (Ethernet, InfiniBand)} \times \mathbf{2} \\ & \text{architectures (i386, x86_64)} = \\ & \mathbf{24} \text{ separate MPI installations (for you to choose from on your system(s))} \end{aligned}$$

It is important to understand that *MPI installations are compiler, bit architecture, and Interconnect specific in most cases*. Also, specific MPI Implementations are either MPI 1, or MPI 2, standards compliant. This also helps explain why we recommend environment modules, and believe that they should become an integral part of your user environment management strategy. Modules are needed in some cases to manage the large number of MPI implementations, compilers, and tools that may be installed on your Full Cluster configuration.

Your choice of MPI implementation is driven by your Interconnect, your applications porting status (*what MPI implementation your code is known to work with*), and your compiler choice. Please research to understand what MPI implementation you require to successfully compile and run your code(s). If you have any questions at all, please contact your Aspen Sales Engineer to help you with your MPI Implementation choice.

If the MPI implementation you've chosen is commercial, you must provide the license(s) to us before we install your system(s), or have them included in your quote so that we may install them for you.

The table below lists most common MPI implementations that you may choose from for your system(s).

Common MPI implementations

<u>MPI</u>	<u>Commercial</u>	<u>MPI Standard</u>	<u>Supported Interconnects</u>
MPICH	no	1	Ethernet, shared memory
MPICH2	no	2	Ethernet, shared memory
LAM/MPI	no	1	TCP/IP, Ethernet, InfiniBand (<i>mVAPI</i>), shared memory
LA-MPI	no	1.2	TCP, UDP, Ethernet, Myrinet (<i>GM only</i>), InfiniBand, Quadrics, HiPPI
MPICH-MX	yes, but free	1	Myrinet (<i>MX only</i>)
MVAPICH	no	1	InfiniBand, iWARP, RDMA
MVAPICH2	no	2	InfiniBand, iWARP, RDMA
HP-MPI	yes	2	TCP/IP
Scali MPI	yes	2	Ethernet, InfiniBand, InfiniPath, Myrinet
OpenMPI	no	2	Ethernet, TCP/IP, InfiniBand, InfiniPath, Myrinet

[<< Previous](#) | [Next >>](#)



Facilities and Contact Information

Aspen provides racked system(s) in 42U or 25U racks by default, although we can provide other rack solutions on request. Our 42U rack is 6' 7" tall and has rollers to allow it to be moved from your unloading area to its final location. Your facility must have adequate clearance through all doorways and hallways between the unloading area and final destination. Please check for obstacles such as door jambs and automatic door mechanisms which may cause clearance problems. A minimum of two (2) personnel are normally required to move a fully loaded 42U rack.

Your delivery and unloading area may require special consideration as well. If your system(s) are configured in 42U rack(s) and your facility has a loading dock, your rack(s) will require a 8' door opening to be safely unloaded from the delivery vehicle. We can coordinate a lift-gate equipped delivery vehicle to ameliorate this issue.

A fully loaded 42U rack may weigh as much as 2600 pounds, and normally requires a freight elevator to move between floors. Please check with your facility administrator or research to find out what weights can be accommodated for all areas between your unloading area and your final location. Your final installation location must be capable of sustaining the static load of the entire installed system as well.

Aspen ships any racked system(s) with all systems(s) installed and internally wired. If your Full Cluster configuration requires more than one rack, each rack will be fully self contained, requiring only power connections to be connected, and the specially wired cable bundles to be ran between racks and connected for system(s) power on and operation.

Aspen ships all racked system(s) on pallets which require pallet jacks to move. Each rack is wrapped in a shipping box and plastic, banded, and bolted to the pallet. Additional unit or node shipping cartons are wrapped with your rack(s), and may contain accessories and other miscellaneous items. Retain these shipping cartons in case you should need to return a unit to us for repair in future.

If you are receiving system(s) without racks, each node is normally individually packaged in a unit shipping carton. If justified by quantity of individually shipped systems or total weight, your individual system(s) will be palletized and banded at our manufacturing facility, and require a pallet jack to move as well.

Cooling is very important for high performance computing systems. Aspen recommends that the ambient temperature of your installed location, after installation and while the system(s) is/are operating under load, never exceed 90° Fahrenheit or 32° Celsius. Lower temperatures are highly desirable, and will decrease unplanned failures and increase your system(s) longevity and reliability.

Your sales engineer can prepare a heat load calculation for your delivered system(s) that you can use to determine any additional cooling that might be necessary. Your sales engineer can also work with you to design special cooling solutions, such as totally enclosed rack cooling systems, cold aisle / hot aisle environments, and supplemental cooling to meet your needs. Speak to your sales engineer about what options would be best for your installation.

Most system(s) offered by Aspen Systems are server class, and as such may contain additional cooling fans operating at high RPM and generating a high ambient noise level while they are operating. If noise abatement is a requirement or concern in your situation or installed location, work with your sales engineer to design a system that will meet your specific noise requirements.

Your sales engineer will work with you to determine your system(s) power requirements, and the most effective way to power your new system(s). Installation location modifications may be necessary to accommodate the number and type of electrical outlets necessary to power your new system(s). Your quote should reflect the configuration you and your sales engineer have agreed upon after these discussions.

Your sales engineer can discuss Un-Interruptible Power Supply (UPS) options and implementation as well if applicable to your situation. Some facilities already have UPS protection, while other customers may require only the master and/or data servers to be UPS equipped. Still other customers require all system(s) to be UPS protected.

We encourage you to discuss any and all facility or delivery questions with your sales engineer prior to submitting a P.O.. Many of these modifications can result in quote modifications to accommodate your specific requirements. We will ask you to outline your facility and delivery requirements in your SOW so that we may better serve you.

Aspen Contact & Customer Experience Level

Aspen will attempt to assign a specific Production Engineer as well as a primary Support Engineer to you. These engineers, just as your Sales Engineer does for your account and sales needs, will function as primary Points of Contact for your system(s) production, installation, and support phases.

In the course of the production and support of your system(s), you will almost invariably get the personal e-mail address of your Production and Support Engineer(s). In order to help us service your needs better,

please

- utilize the *personal* e-mail address of your ***Sales Engineer*** to contact him or her, and their *personal* cell phone or extension.
- Utilize the support@aspsys.com e-mail address to contact your ***Production Engineer*** or ***Support Engineer***, and call them on the *main Aspen toll free number*.

Your Production or Support Engineer may be on a site install or upgrade visit, or otherwise unavailable. All Production and Support Engineering personnel receive, and monitor, e-mail sent to support@aspsys.com.

A trouble ticket number e-mail with an appropriately modified subject line is automatically emailed back to you when you send an initial e-mail to support@aspsys.com. Simply reply to that e-mail with the subject line unchanged to ensure that all correspondence on that issue remains on that ticket. All e-mail to support that has that ticket number in the subject line is kept in that ticket for reference and historical purposes. That ticket number can be used by any available engineer as a reference about your particular issue. We want to make sure that we take care of you promptly.

We will assign your Support engineer(s) based on our understanding of your HPC or clustering experience and requirements. For example, if you are an extremely experienced HPC administrator, we will attempt to pair you with engineers that know best how to work with your experience level. If you are a less experienced administrator or HPC user, we will pair you with engineers best fitted for your needs at that experience level.

We will ask you to tell us your experience level in your SOW so that we can match you with engineers who will best fit your particular needs.

[<< Previous](#) | [Next >>](#)

Warranty and Additional Support Options

Aspen provides Bronze (1 year), Silver (2 years), and Gold (3 year) hardware warranties. Fourth and fifth year warranties can be purchased as well.

Aspen offers advance replacement parts, but you must obtain a Return Material Authorization (RMA) number from Aspen to get an advance replacement. Aspen pays for shipping both ways when an RMA is processed, shipping the new part to your organization with an RMA and return shipping label already included.

Prior to your first advance replacement part delivery, we will ask you to submit our Aspen Support Agreement, which outlines our parts advance policy.

Aspen can also process an RMA for you to return an entire node or peripheral to Aspen for repair or replacement should that be needed. Aspen pays for standard ground shipping both ways in this case as well, and this service is available for the term of your hardware warranty.

Aspen also offers on-site hardware support through select hardware maintenance partners. We can tailor your on-site hardware maintenance coverage days, hours, and response times to fit your budget and needs. Aspen recommends the use of on-site spares kits for all organizations who opt for on-site support. Talk to your sales engineer for more information.

Additional Support Options

Aspen can work with you to tailor a support contract that best fits your needs. Some of the additional support options Aspen can provide you are;

- **Blocks of support time**

These are 1, 5, 10, 15, and 25 hour blocks of time that you may pre-purchase, then use only as you need them, to have an Aspen engineer complete any cluster administration or upgrade tasks you wish

- **On-site visit**

On-site visits are normally used to accomplish a specific task, or perhaps to perform additional informal on-site training for users or administrators after or at the same time your cluster is delivered.

- **Full support contract**

Aspen can support your entire cluster administration needs as well. If you do not have an on-site

administrator, and do not wish to administer your cluster yourself, Aspen offers full cluster administration and support contracts. These will normally include a on-site hardware maintenance contract as well, and require Aspen remote access to your cluster.

In all cases, Aspen recommends that you allow remote access to your cluster from Aspen support engineers if possible or allowed. While this is not possible in all situations due to security or organizational requirements, remote access will make your life easier in the event of a problem occurring, and allow your Aspen support engineers to work directly with you on problems to quickly solve them.

Aspen can provide a VPN (Virtual Private Network) client on your cluster that connects back to dedicated Aspen support systems after the cluster is installed and operational. This VPN client can be turned on or off at will by you to permit Aspen access only when you wish, or can be left on and connected to Aspen at all times. Many customers have found this “call home” solution more flexible and easier to implement than making the local security configuration changes necessary to allow incoming connections.

Comments

We have attempted to cover most of the salient configuration issues for your new system(s), but we know that we may not have covered everything you need. Aspen is committed to building your system(s) in a way that satisfies you, and needs your input to do so.

Please be sure to note any special requirements or questions that you have in your SOW, or better yet, discuss them now with your sales engineer. Some configuration changes may drive quote changes to accommodate your specific needs.

The design, purchase, and deployment of High Performance Computing System(s) can be time consuming. Your contacts at Aspen are there to help you through this process. Please contact your sales engineer, he or she will be happy to work with you to design the perfect solution for your needs.

[<< Previous](#) | [Back to Configuration Guide Start](#)
