Table of Contents

Preface .................................................................................................................................i

Feedback ...............................................................................................................................i

1. Scyld Beowulf System Overview .................................................................................. 1
   System Components and Layout ...................................................................................... 1
   Assembling the Cluster .................................................................................................... 2
   Software Components .................................................................................................... 3

2. Quick Start Installation ................................................................................................. 5
   Install onto the Head Node ............................................................................................ 5
   Boot and Configure the Compute Nodes .......................................................................... 8

3. Graphical Installation on the Head Node ....................................................................... 11
   Starting the Graphical Installation .................................................................................. 11
   Booting the Head Node .................................................................................................. 11
   Welcome to Scyld Beowulf ............................................................................................ 13
   Language Selection ....................................................................................................... 15
   Keyboard Configuration ................................................................................................. 15
   Choosing Mouse Type .................................................................................................... 16
   Disk Partitioning ............................................................................................................ 17
      Automatic Partitioning ............................................................................................... 19
      Manual Partitioning with Disk Druid ......................................................................... 20
         Disk Druid’s Action Buttons ..................................................................................... 21
      Minimal Partitioning ................................................................................................... 22
   Partitioning Problems ..................................................................................................... 22
   Bootloader Configuration ............................................................................................... 23
   Network Configuration ................................................................................................... 26
   Network Security Configuration ..................................................................................... 27
   Additional Language Support ....................................................................................... 29
   Setting the Time Zone .................................................................................................... 30
   root Password Selection ................................................................................................. 31
   Selecting Package Groups .............................................................................................. 32
      Unresolved Dependencies ........................................................................................... 34
   About to Install ............................................................................................................... 34
   Graphical Interface (X) Configuration ............................................................................ 34
   Monitor Configuration .................................................................................................... 35
   Customize Graphics Configuration .................................................................................. 36
   Reboot the System .......................................................................................................... 37
   Welcome ......................................................................................................................... 37
   License Agreement .......................................................................................................... 37
   Date and Time .................................................................................................................. 38
   System User ..................................................................................................................... 38
   Beowulf Cluster ............................................................................................................... 38
   Sound Card ...................................................................................................................... 38
   Additional CDs ............................................................................................................... 38
   Finish Setup ..................................................................................................................... 39
4. Installation of the Compute Nodes

Compute Node Boot Media ................................................................................. 41
- PXE Network Boot ....................................................................................... 41
- PXE Media Boot ........................................................................................... 41
Beosetup ............................................................................................................. 41
- Starting the BeoSetup Tool ........................................................................... 42
The Main Window .............................................................................................. 44
- Apply and Revert buttons ............................................................................. 44
- Short Cuts ....................................................................................................... 44
- Pop-up Menus ................................................................................................ 44
Node Floppy button ............................................................................................ 44
Node CD button .................................................................................................. 45
Booting the Compute Nodes ............................................................................. 45
Incorporating the Compute Nodes ................................................................. 46
Optional Compute Node Disk Partitioning .................................................. 46
Reboot the Compute Nodes ............................................................................ 47
BeoBoot Floppy or CD-ROM ........................................................................... 47
Congratulations! .................................................................................................. 47

5. Cluster Verification Procedure ........................................................................ 49
bpstat .................................................................................................................. 49
beostatus ............................................................................................................ 49
bpsh ..................................................................................................................... 51
linpack ................................................................................................................ 51
mpi-mandel ........................................................................................................ 51

6. Troubleshooting a Scyld Beowulf Installation ................................................. 53
Failing PXE Network Boot .............................................................................. 53
Mixed Uniprocessor and SMP Cluster Nodes ................................................. 55
Mixed 32 and 64-bit cluster nodes ................................................................... 56
Device Driver Updates ...................................................................................... 56
Finding Further Information ............................................................................ 56

A. Compute Node Disk Partitioning ................................................................. 59
Architectural Overview ...................................................................................... 59
Operational Overview ...................................................................................... 59
Partitioning Disks ............................................................................................ 59
Default Partitioning ......................................................................................... 60
Mapping Compute Node Partitions .................................................................. 60
Generalized, User-Specified Partitions ............................................................ 61
Unique Partitions .............................................................................................. 61
Preface

Congratulations on purchasing Scyld Beowulf™, the most scalable and configurable Linux clustering software on the market. This guide describes how to install the Scyld Beowulf software and have a scalable cluster running in just a few minutes.

NOTE: Your reseller may have preinstalled the software on the cluster. If so, you may skip the installation instructions.

While this document contains all of the information needed to get your system up and running, additional guides and reference manuals are available. The Administrator's Guide and User's Guide describe how to configure, use, maintain and update the cluster. The Programmer's Guide and Reference Guide describe the commands, architecture, and programming interface for the system. All of the documentation may be viewed using a web browser from the last CD in the set. (This is an autorun CD; it should bring up a browser on either Windows or Linux).

Feedback

We welcome any reports on errors or difficulties that you may find. We also would like your suggestions on improving this document. Please direct all comments and problems to: support@scyld.com.

When writing your e-mail please be as specific as possible, especially with errors in the text. Please include the chapter and section information. Also please mention in which version of the manual you found the error. This version is Series 30cz-1, published Feb 2006.
Chapter 1. Scyld Beowulf System Overview

Scyld Beowulf Series 30cz-1 streamlines the processes of configuring, running, and maintaining a Linux cluster using a group of commodity off-the-shelf (COTS) computers connected through a private network. The front-end "head node" computer in the cluster distributes computing tasks to the other machines (known as the "compute nodes") in a parallel architecture.

System Components and Layout

The head node is configured with a full Linux installation. Each machine in the cluster is installed with a Network Interface Controller (NIC) connected to a private network. In order for the head node to communicate with an outside network, it needs two NICs: one for the private internal cluster network, and the other for the outside network. It is suggested that the head node be connected to an outside network so multiple users can access the cluster from remote locations, as shown in Figure 1-1.

![Figure 1-1. Cluster Configuration](image)

In addition to the NIC(s), a booting mechanism is needed. The compute nodes do not boot a full distribution themselves, rather they boot across the network from a computer which hosts boot images (PXE booting), or, optionally, from a CD or floppy disk that contains a small PXE boot image (see the Administrator’s Guide for a list of compute node boot options).

Hardware selection for a Beowulf system is based upon the price/performance ratio. Scyld recommends the following components for use with this release of the Scyld Beowulf distribution:

**Recommended Components**

**Processors**
Intel® Pentium® IV, Intel® Xeon®, AMD Opteron™, single- or multi-core

**Architecture**
one, two, or four sockets per motherboard

**Physical Memory**
2,048 MBytes (2 GByte) or more preferred, minimum 512 MBytes

**Operating System**
Scyld Beowulf (this release)
Chapter 1. Scyld Beowulf System Overview

Network Interface Controllers (NIC)

Gigabit Ethernet (Fast Ethernet, at minimum) PCI-X or PCI-Express adapters (with existing Linux driver support) in each node for the internal private IP network. The head node requires an additional NIC for connecting the cluster to the external network. This NIC should be selected based on the network infrastructure (e.g., Fast Ethernet if the external network you are connecting the cluster to is Fast Ethernet). For a list of the latest supported NICs contact Scyld.

Network Switch

All compute nodes, and the head node private network NIC, must be connected to a non-blocking Gigabit Ethernet switch for the internal private network. At a minimum, the network switch must match the speed of the network cards. Note that the switch is a critical component for the correct operation and performance of the cluster. In particular, the switch must be able to handle all of the network traffic over the private interconnect, including cluster management traffic, migrating processes, transferring libraries, and storage traffic. It must also properly handle DHCP and PXE.

Note: it is sometimes confusing to identify which NIC is connected to the private network. Take care to connect the head node to the private switch through the NIC with the same or higher speed than the NICs in the compute nodes.

Drives

For the head node we recommend using either Serial ATA (SATA) or SCSI disks in a RAID 1 (mirrored) configuration. The operating system on the head node requires approximately 3 GB of disk space. We recommend configuring the compute nodes without local disks (diskless).

If the compute nodes do not support PXE boot, a floppy (32-bit architectures) or bootable CD-ROM drive (32- and 64-bit architectures) is required. If local disks are required on the compute nodes, we recommend using them for storing data that can easily be re-created, such as scratch storage or local copies of globally-available data.

If you plan to create boot CDs for your compute nodes, your head node requires a CD-RW or writable DVD drive.

In the default configuration, /home on the master node is exported to the compute nodes (other file systems may be exported as well). If you expect heavy file system traffic, we recommend you provide a second pair of disks in a RAID 1 (mirrored) configuration for these exported file systems. Otherwise, it is possible for access to the exported file systems to interfere with the head node accessing its system files, thus affecting the head node’s ability to launch new process and manage the cluster.

Optional Hardware Components

Gigabit Ethernet with a non-blocking switch serves most users. However, some applications benefit from a lower-latency interconnect. Infiniband is an industry standard interconnect providing low-latency messaging, IP, and storage support. Infiniband can be configured as a single universal fabric serving all of the cluster’s interconnect needs. More information about Infiniband may be found at the Infiniband Trade Association web site (http://www.infinibandta.org). Scyld supports Infiniband as a supplemental messaging interconnect in addition to Ethernet for cluster control communications.

Assembling the Cluster

The full Scyld Beowulf Scalable Computing Distribution is installed only on the head node. A graphic utility (BeoSetup) is included on the head node to aid in the cluster-configuration process.
Chapter 1. Scyld Beowulf System Overview

Most recent hardware supports network boot (PXE boot) which Scyld recommends for booting compute nodes. Compute nodes that do not support network boot require a floppy disk or CD-ROM drive, with suitable boot media inserted before being powered up.

Software Components

A brief description of the major portions of the Scyld Beowulf distribution is given below. For more information, see the Administrator’s Guide and the User’s Guide.

- **BeoSetup**: A graphic utility for configuring and administering the Scyld Beowulf cluster.
- **BeoStatus**: A graphic utility for monitoring the status of the Scyld Beowulf cluster.
- **Scyld BeoMaster** The Scyld BeoMaster software is an integral part of the Scyld Beowulf distribution. It allows processes to be started on compute nodes in the cluster and tracked in the process table on the head node. BeoMaster also provides process migration mechanisms to help in the creation of remote processes and removes the need for most binaries on the remote nodes.
- **MPICH/LAM**: Message Passing Interface, modified to work with Scyld Beowulf cluster software.
Chapter 2. Quick Start Installation

It is assumed that appropriate staff have installed the cluster hardware, all are connected to electrical power and the proper public and cluster private network, and that the head node and cluster nodes are clearly marked.

The Scyld Beowulf distribution is provided as a set of three CD-ROM discs which include the basic Linux operating system distribution as well as Scyld Beowulf cluster software. The first CD in the series (disc 1) is bootable, and is used to initiate the installation on the head node. The last disc in the series contains product documentation (which may be read directly from the disk on any running PC or workstation).

This chapter describes installing Scyld Beowulf on a head node where the compute nodes fall into one of these categories:

1. compute nodes are network-bootable (via PXE)
2. compute nodes booted from local media (which may be created using the BeoSetup utility on the head node)

Refer to Chapter 3, Graphical Installation for other scenarios.

The compute nodes join the cluster without any explicit installation. Having obtained a boot image, via either PXE or local media, they're converted to a Scyld-developed network boot system (and seamlessly appear as part of a virtual parallel computer).

Install onto the Head Node

If you need more information on any of the following steps, the installation procedure is fully documented in Chapter 3, Graphical Installation.

1. Boot the front-end (head node) machine from the Scyld Beowulf distribution CD-ROM labeled "Disc 1." The graphical installation process starts after 20 seconds. Hit Enter to start the process immediately.
2. Follow the on-screen instructions. The first few screens set basic elements, including the default language, keyboard, and mouse. For most screens you may accept the default values offered.
3. If you need to set up hard disk partitions or bootloader values other than the defaults, please see the detailed instructions in Chapter 3, Graphical Installation.
4. You must configure the networks as described below (and also in Chapter 3, Graphical Installation).

Tip: To proceed with configuring the network, you must know which interface is connected to the public network and which is connected to the private network.
Chapter 2. Quick Start Installation

For eth0 (or the interface connected to the public network) DHCP is selected by default. If your external network is set up to use static IP addresses, select this interface and click Edit; your network administrator should provide you with the IP address. Set the IP Address and Netmask, then click OK. If you set a static IP address for the public interface you must also click the radio button labelled manually for Set the hostname and provide a hostname, gateway, and primary DNS IP addresses.

For eth1 (or the interface connected to the internal private cluster network), you must choose a static IP address. We recommend choosing a non-routable address (such as 192.168.x.x or 10.x.x.x).

Select eth1 (or the interface connected to the internal private cluster network) and click Edit. In the Edit interface dialog box, check Activate on Boot and un-check Configure using DHCP. Specify the IP address in the four IP Address boxes. Once you have specified the IP Address, you must also set your Netmask based on the address.
Click **OK** to return to the screen described in Figure 2-1.

---

**Caution**

*Note: For eth1 (or the device connected to the private cluster network), *do not* select DHCP. You must select and edit this interface and manually set the IP address (see Figure 2-2). Also check the *Activate on Boot* checkbox to make the specific network device initialized at boot-time.*

Note that the head node also functions as a PXE and DHCP server for the cluster. On the Firewall page, following, ensure that the interface is set as a trusted interface (If you would like to allow all traffic from a device, otherwise the cluster will not boot.

Configure the network settings for all of the devices listed. Click **Next** to continue.

---

5. Configure your firewall settings. If you choose to install a firewall, you must enable all traffic on eth1 (or the interface connected to the internal private cluster network) to pass through the firewall. Check the box next to the private network interface under *Trusted Devices.*

Click **Next** to continue.

---

6. Proceed through the dialog boxes to configure additional language support and your timezone.

7. You must supply a root password. Refer to Chapter 3, Graphical Installation. An alphanumeric password of at least 8 characters, with special characters to thwart attacks, is recommended.

8. Review the packages to be installed. We recommend accepting the defaults unless you are an experienced Linux system administrator and have specific requirements.

9. The installation begins. You are prompted to insert additional disks, as necessary.

10. When the installation is finished you are prompted to verify the identification of the video hardware (located by the installer when your system was probed). If the default identification is incorrect, see Chapter 3, Graphical Installation.

11. Upon reboot, you are greeted with a *Welcome* screen emblazoned with the Scyld Software logo. The subsequent screen presents the Scyld End User License Agreement (EULA); you must accept the EULA to proceed with the installation.

12. The next screen enables you to to set up your Scyld Beowulf cluster by choosing the private cluster interface, establishing the initial IP address for compute nodes. Also specify the number of nodes in your system. Note that these parameters can be changed later with *BeoSetup.*

---

*Figure 2-3. Set Up Private Cluster Interface*
Chapter 2. Quick Start Installation

13. Continue through the installation screens to set the date and time, a system-level user account, and to verify the sound card (if one is installed). Scyld Beowulf is now installed and set up on your system. Next, boot and configure the compute nodes to get your cluster up and running.

Note: If prompted to restart Beowulf services, click Yes. Any warning about rebooting the cluster nodes can safely be ignored at this point.

Boot and Configure the Compute Nodes

1. If you are not logged in as root already, log into the head node using the root username and password set up earlier. Start the cluster configuration tool, beosetup, by clicking on the Beosetup "Building Blocks" icon at the bottom of the screen (hover the cursor over the tray icon that looks like triangle of yellow blocks, then click). If beosetup fails to start check the syslog for possible errors. You can manually start it by typing /usr/bin/beosetup in a terminal window.

2. If your compute nodes can’t boot across the network via PXE, or if for some reason you don’t want them to use PXE, you may create compute node disks. You only need to perform this step if you want to boot compute nodes from boot media; BeoSetup does not have to be run unless you want to create boot media, or you want to monitor the booting process.

   You may create either boot floppy disks or boot CD-ROM disks. You need a CD-RW drive installed on the head node to create boot CDs and CD-ROM drives on each compute node in order to boot from the boot CDs.

   a. Click Node Floppy or Node CD in the BeoSetup window.
   b. Click OK in the Create BeoBoot Floppy or Create BeoBoot CD Image window to create the node boot image. This image consists of a basic boot image which first boots the node and then downloads the full compute node boot image.

   Note that Create BeoBoot Floppy writes the image directly to the disk in the floppy drive. Create BeoBoot CD Image creates an ISO file on your system without writing anything to the CD drive; you must then write the ISO file to a blank CD-R using cdrecord (see the man page for cdrecord) or a similar utility. This ISO image can be copied to as many CDs as necessary for your cluster.
   c. Click Close to close the window.
   d. Write the boot image to blank media and move it to a compute node. Repeat for each compute node.

3. First, be sure to set the BIOS on each compute node to boot either via PXE or via the physical medium you have chosen.

   Boot the compute nodes by powering them up in the order you want them to be numbered, typically one-by-one from the top of a rack, downwards (or from the bottom, up). Node numbers are assigned in order of connection with the head node. You can reorder nodes later as desired; see the Administrator’s Guide).

   Compute nodes are listed by beosetup by node number and Ethernet Station (MAC) Addresses.

4. The nodes transition through the boot phases and, after a few seconds, are shown by beosetup to be in the up state. The cluster is now fully operational with diskless compute nodes.
beosetup shows the compute nodes’ status. All nodes show up when ready for use. A state of error may result from a partition table lock.
Chapter 3. Graphical Installation on the Head Node

This chapter guides you through the graphical installation of the Scyld Beowulf software. This software installation is intended for the first computer ("node") of the Scyld Beowulf cluster, which functions as the "head node" in the cluster, controlling and monitoring other nodes as well as distributing jobs. It is assumed that you are familiar with the concepts outlined in the previous chapters and that you have correctly assembled the hardware for your Scyld Beowulf cluster. If this is not the case, please refer to the previous chapters to acquaint yourself with the Scyld Beowulf software and then verify that your hardware configuration is set-up properly.

Each screen displayed within the graphical installation sequence has an online help panel on the left and an installation dialog panel on the right. You can scroll through the online help before answering the questions.

Starting the Graphical Installation

This section helps you get your Scyld Beowulf graphical installation running. If your machine does not boot from the CD-ROM, you must change the settings in the machine’s BIOS. Refer to your computer’s reference manual if you need instruction regarding how to change settings in the BIOS.

Booting the Head Node

Caution

Installing Scyld Beowulf over another Linux or Scyld Beowulf installation does not preserve any existing system files or data. You must exit the Scyld Beowulf installer, and backup any important data which you wish to save by a suitable means. Or, if you are experienced, when asked in the the Section called Disk Partitioning, select “Manually partition” as your disk partitioning strategy.

Running the graphical installation on the head node is just one of the uses for the Scyld Beowulf CD-ROM installation set. It may also be used to perform a text-mode installation. When you boot the head node with the Scyld Beowulf CD-ROM (Disc 1), you have 20 seconds to select the preferred installation mode, graphical or text. The CD-ROM defaults to installing in graphical mode. The selection screen is shown in Figure 3-1.
Chapter 3. Graphical Installation on the Head Node

Selecting the Graphical Installation

1. Insert the Scyld Beowulf CD-ROM into the CD-ROM drive on the head node. Restart the system by powering it on or resetting it.

2. When the installation selection screen appears (see Figure 3-1) and prompts you with boot:, type Enter to start the graphical installation. Linux boots from the CD-ROM. The installer may prompt you to test the installation media, or to skip the test. You can make either choice; if you choose Ok, the test takes a few minutes to complete.

   The Anaconda-based installer runs. This may take a few minutes. Text messages scroll on the screen, then the video system is probed, which may cause the screen to flash.

   Finally, the Welcome screen appears.

There are several options that may be selected from this screen other than the graphical install documented in this chapter. It is normally not required that these options be used. They are provided for experienced Linux administrators, or for use at the direction of your support representative.

- text mode: to select text mode install, type linux text.
- F1–Main: returns you to the installation selection screen.
- F2–Options: describes some Installer Boot Options. Some of the options available are:
  - To disable hardware probing, type linux noprobe at the boot: prompt. This is useful if installation fails because the installer can’t probe the hardware correctly.
  - To test your installation media, type linux mediacheck at the boot: prompt. This is useful if installation fails because the installer can’t read the installation CD.
Chapter 3. Graphical Installation on the Head Node

- To enter rescue mode, type `linux rescue` at the `boot:` prompt. This is more fully documented under the `F5–Rescue` item below.

- If you have a driver disk, type `linux dd` at the `boot:` prompt. This is useful if you must provide a driver that is not included on the standard Scyld installation media, for example for a hard disk or RAID controller on which you want to install the operating system.

- To prompt for the install method being used, type `linux askmethod` at the `boot:` prompt.

- To use an installer update disk, type `linux updates` at the `boot:` prompt.

- To force the use of the lowest graphical screen resolution, type `linux lowres` at the `boot:` prompt. This is useful if your monitor, video cable, or a monitor connection involving a KVM can’t handle a higher screen resolution even though the video card and monitor report they can handle a higher resolution.

- `F3–General` provides some additional advice for working around installation failures.

- `F4–Kernel` provides some help with parameters that may be passed to the kernel as it boots. The options described under `F2` above are specific examples of the general method described here. Take care using this option unless you are an experienced Linux administrator.

- `F5–Rescue` invokes the Scyld Beowulf rescue mode. Rescue mode includes some utilities that are useful if the head node does not boot after installation, such as an editor you can use in examining and editing configuration files, and tools to work with hard drives. To invoke this option, type `linux rescue` at the `boot:` prompt.
Welcome to Scyld Beowulf

The Scyld Beowulf cluster software provides its functionality through extensions to the Linux kernel and additional system libraries. The Scyld Beowulf advancements are seamlessly incorporated so that there is no loss of previous Linux capabilities. However, a full installation of the Scyld Beowulf overwrites and replace any existing Linux installation, meaning that any previous settings are lost.

The Welcome screen is shown in Figure 3-2. You do not need to supply any information on this screen.

This installation program is based on the "Anaconda" installer, which is used by many Linux distributions. The Scyld Beowulf head node interface is designed to appear as a standard Linux installation, thus most of the questions are the same as a workstation Linux installation. Most cluster-specific questions are at the end of this process, but there are a few that must be handled during the regular installation process, networking in particular. All Scyld Beowulf specific items are fully detailed in this installation guide.

You can look at the release notes by clicking on the Release Notes button on the left panel of the Welcome screen. You may also browse the documentation directly from the last of the installation CDs on any Linux or Windows workstation or PC. This Autorun CD launches a browser from which the entire documentation set is available.
Chapter 3. Graphical Installation on the Head Node

The left-hand side of each screen presented contains Online Help for that specific screen. It is recommended that you read the information presented, especially if you are unsure of your selections. If you do not wish to see the Online Help, you may click the Hide Help button, which is located under the Online Help frame. To bring the Online Help back, click the Show Help button.

Click on Next when you are ready to proceed to the next installation screen.

Language Selection

![Language Selection Screen]

Figure 3-3. Choosing Language

Choose the language for this installation. The default is English. Click Next when you are ready to proceed to the next installation screen.
Keyboard Configuration

Choose the layout type for the keyboard (for example, U.S. English) that you would like to use for the system.

Figure 3-4. Choosing Keyboard Configuration

The installer probes your hardware and selects the keyboard language configuration best suited to what it finds. You can change its choice by scrolling through the choices in the windows and clicking a different language. Generally, the keyboard configuration selected by the installer is the correct one.

Click Next when you are ready to proceed to the next installation screen.
Chapter 3. Graphical Installation on the Head Node

Choosing Mouse Type

The installer probes your hardware and selects the mouse type best suited to what it finds. You can change its choice by scrolling through the choices in the windows and clicking a different manufacturer and/or mouse type. Note also that the connection method (PS/2, serial, or USB) is also among the choices. Generally, the mouse selected by the installer is the correct one.

Click Next when you are ready to proceed to the next installation screen.
Chapter 3. Graphical Installation on the Head Node

Disk Partitioning

The purpose of this step is to establish where the information you selected in the previous step is installed on your hard drive. This involves partitioning the data into several sections, and formatting your drive accordingly using a disk partition program called Disk Druid. You have the following options:

Automatic Partition

With this option, the installer uses a pre-configured format to determine how best to utilize the machine’s drive capacity, and presents the result with Disk Druid. Scyld recommends allowing the installer to automatically partition your hard drive.

Note that if you wish to preserve existing data, this option is usually unacceptable, and you should choose a manual installation instead. If you choose this method, proceed with the Section called Automatic Partitioning.

Manual with Disk Druid

Disk Druid provides a graphical interface with which you can choose your own partition scheme. This is recommended for advanced users only. If you choose this method, proceed with the Section called Manual Partitioning with Disk
Chapter 3. Graphical Installation on the Head Node

Druid.

Automatic Partitioning

The automatic partition option lets you decide what device(s) are used for the installation, and what part of each device is be used. The default option is to erase any existing partition on the selected device(s) and perform an installation using the entire drive. Alternatively, you may choose to perform an installation on the currently existing available space, or on the space that would be made available by removing existing Linux partitions (but leaving non-Linux partitions, such as Windows).

The "Review" option is also selected by default, so that you may accept or refuse the configuration made by the installer prior to committing the changes to the device. If you select this option, the next screen is that given by Disk Druid, Figure 3-9, showing the suggested partitioning scheme. You can change the automatic partitioning results if desired.

You are prompted to select the hard drive to use for the installation, and to direct Disk Druid’s behavior with respect to existing partitions. Check the Review box to review all partitions before they are created. Click Next when you are ready to proceed to the next installation screen.

Clicking Next may display a Warning dialog box like Figure 3-8 asking you to confirm that you want to remove disk partitions. If you are sure you can safely delete the disk partitions, click Yes to proceed. Otherwise click No to return to
Chapter 3. Graphical Installation on the Head Node

Figure 3-7. You then see a Disk Druid screen showing the default partitioning.

![Disk Druid Screen](image)

Figure 3-8. Warning — deleting all data from disk

If you have any doubts about whether the contents of the disk can be destroyed, click No to return to the screen in Figure 3-6. Otherwise, click Yes to erase any old data, and create a new partition table.

**Manual Partitioning with Disk Druid**

Manual partitioning gives you the greatest flexibility, but requires some knowledge on the part of the user. Partitioning the device amounts to creating various partitions, keeping in mind two issues:

**Partition type and name**

You must define some core partitions with specific names and type, such as /boot, /home, / (root partition) and a swap partition.

**Partition size**

Your partitions must respect some minimal size constraints so that they can hold the data, and have room for potential growth.
Chapter 3. Graphical Installation on the Head Node

Figure 3-9. Partitioning with Disk Druid

Disk Druid’s Action Buttons

Displayed between the "Partitions" section and the "Drive Summaries" are the Disk Druid action buttons. They are used to add and delete partitions, change partition attributes, accept changes and exit the program. Each button is listed below with its functionality.

New

Used to request a new partition. When selected, a dialog box appears requesting mount point and size for the partition. See Figure 3-10.

Edit

Used to modify attributes of the partition currently highlighted in the "Partitions" section. When selected, a dialog box appears with changeable fields. Which fields are changeable depends on whether the partition information has already been written to disk.

Delete

Used to delete the partition currently highlighted in the "Current Disk Partitions" section. You are prompted to confirm your intention to delete the partition.
Chapter 3. Graphical Installation on the Head Node

Reset

Used to restore the partition table settings to its original state. Any changes that you made are lost.

RAID

Used to provide redundancy to any or all disk partitions. This should only be used if you have experience using RAID.

LVM

Used to configure the logical volume manager (LVM). This should only be used if you have experience with LVM setup.

Minimal Partitioning

Disk Druid displays the hard disks found on the system, along with the existing partitions on those disks. Select the first hard disk, usually /dev/hda or /dev/sda, and click New. A screen appears that is similar to Figure 3-10. Type /boot in the Mount Point box. Provide a Size of at least 100 MB. Check the box Force to be a primary partition. It's a good idea to also check Check for bad blocks. Click OK to create the partition.

Repeat this process to create a swap partition (select swap as the File System Type and choose a size at least two times your computer's RAM. Do not check Force to be a primary partition. Click OK to create the swap partition.

Repeat the above for Mount Points / and /home, choosing ext3 or another file system other than swap. You may also create mount points for /var and /usr, or let Linux create these within the / (root) partition. Generally, you need at least 4GB of space in /, of which at least 3GB is needed for /usr.

To configure hardware RAID, refer to your RAID vendor's documentation. To configure software RAID or LVM, refer to Administrator's Guide or a good Linux reference. The rest of the space may be allocated to /home.

Figure 3-10. Add Partition in Disk Druid
Partitioning Problems

If Disk Druid can not allocate a partition, a dialog box appears which lists the unallocated partition and the reason for the failure. Lack of disk space is the most common reason. To remedy the situation, you have a few choices. You may move the partition to another drive that has sufficient space available, resize the partition (by deleting it and then re-adding a smaller one), or just delete the partition entirely. To modify the partition, double-click on it or use the Edit button.

Bootloader Configuration

![Bootloader Configuration](image)

Figure 3-11. Bootloader Configuration

This section assists you in configuring the method used to boot your system. The bootloader is a piece of software that starts the Linux kernel or other operating systems.

The first decision is whether or not to install a boot loader. One may already exist on your harddrive, or you may want to use another device (such as a floppy) to boot the Scyld Beowulf OS. In these cases you do not want to replace the existing bootloader.

By default, GRUB is chosen as the boot loader (see Figure 3-11). LILO is the legacy bootloader, while GRUB is a newer bootloader which may be easier for new users. This may be changed by clicking the Change boot loader button (see Figure
3-12). You may select *LILO* or *GRUB*, or elect to not install a boot loader. (Most users should install a bootloader.)

![Change Boot Loader](image1)

**Figure 3-12. Change Boot Loader**

Note that, if you select *Do not install a boot loader* and *OK* you see a confirmation dialog box (Figure 3-13) reminding you that you need to use removable boot media (floppy or CD) to boot your master node if you proceed. You may cancel this dialog box and return to Figure 3-12.

![No boot loader warning](image2)

**Figure 3-13. No boot loader warning**

Requiring a password for the bootloader provides a higher level of system security. To set a password, select the **Use a boot loader password** checkbox. Enter a password and confirm it (see Figure 3-14).
You can gain more control over the boot process by checking the *Configure advanced boot loader options* box before clicking **Next** (see Figure 3-15).

The bootloader can be installed either in the master boot record (MBR) of the device you are booting from, or in the first sector of the boot partition. For help on deciding which is right for your system, refer to the online help during installation. Generally, installing on the MBR is recommended.

The **Force LBA32** option should only be used if you experience problems with large drives (see the online help).
This screen also provides a place to enter kernel parameters, which the boot loader passes to the kernel upon boot. These parameters depend on the specific kernel you are booting, and should be changed only by experienced users.

Click **Next** when you are ready to proceed to the next installation screen.

**Network Configuration**

A typical Scyld Beowulf cluster has one interface dedicated to the private cluster network, and one to the external network. The following setup is assuming that eth0 is the interface connected to the external network and eth1 is the interface connected to the private network. You can configure your network settings for each network device on your system.

**Tip:** To proceed with configuring the network, you must know which interface is connected to the public network and which is connected to the private network.

For eth0 (or the interface connected to the public network), **DHCP** is selected by default. If your external network is set up to use static IP addresses, select this interface and click **Edit**—your network administrator should provide you with the IP
address and netmask. Set the IP Address and Netmask, then click OK (see Figure 3-17). If you set a static IP address for the public interface, you also must click Edit for Set the hostname and provide a hostname, gateway, and primary DNS IP addresses.

**Caution**
For eth1 (or the device connected to the private cluster network), you must configure the network interface manually by clicking the Edit radio button and assigning a static IP address and netmask to the private network interface.

For eth1, check the Activate on Boot box to make the specific network device initialized at boot-time.

For eth1 (or interface connected to the internal private cluster network), you must un-check configure using DHCP and manually set up a static IP address. We recommend choosing a non-routable address (such as 192.168.x.x or 10.x.x.x). Once you specify the IP Address, you must also set your Netmask based on the address. Click OK to return to the screen in Figure 3-16.

Configure the network settings for all of the devices listed. Click Next when you are ready to proceed to the next installation screen.
Network Security Configuration

This step of the installation process allows you to customize several aspects of the firewall that protects your cluster from possible network security violations. Please note that the security features provided with this system don’t guarantee a completely secure system.

The first item to consider is the Security Level. You may choose:

**High**

All incoming requests are blocked, isolating the cluster from the rest of the network.

**Medium**

Blocks any incoming requests from ports below 1023 as well as the NFS server port and ports used for remote X clients.

**No Firewall**

All connections are allowed. This option is not recommended unless you plan to configure your firewall after the installation.
The rest of the configuration is only available if you click the Customize radio button. Be aware that, from this screen, you can not specify different rules for different interfaces, other than trusted or not trusted. All untrusted interfaces allow the same incoming traffic.

You can select which network devices are trusted. For these devices, all traffic is accepted without regard to the content. It is not recommended to use a trusted device for an interface with a public network.

**Tip:** You must select your private network interface as a Trusted Device in order for the cluster to operate.

---

**Caution**

The *private* interface used for your cluster, usually eth1, must be set up as a Trusted Device in order for the cluster software to work properly. It must be checked in the list titled *If you would like to allow all traffic from a device.*

Selecting a *public* network interface, usually eth0, as a Trusted Device may compromise the security of your cluster. In addition to the security considerations, selecting to allow DHCP on any interface other than the private interface can lead to conflicts with DHCP servers setup for those other networks.

---

Select services for which you want to allow possible connections.

Select additional ports. Empty default is fine for Scyld Beowulf to run properly. However, if you are planing to run services such as SSH or FTP between the public network and the master node, these services must be explicitly allowed.

Click **Next** to continue.
Chapter 3. Graphical Installation on the Head Node

Additional Language Support

Some applications can display messages in a variety of languages. In this section, you need to select the default language used by your system, and the additional languages it can support. Language support affects not only the content, but also the format of messages, including date, monetary values, etc.

Click Next to continue.
Setting the Time Zone

Select the appropriate time zone for the location of your cluster. You have the option of setting this according to location or your time zone’s offset from Universal Coordinated Time (UTC).

First, click the **Location** or the **UTC** tab.

- For the **Location** tab, select a city on the map or in the text listing below the map. You can change the map that appears by changing the geographical area listed in the **View** menu.
- For the **UTC** option, select the appropriate offset from those listed.

In either case, highlight the box labeled **System clock uses UTC** if this is true of your system clock. The time zone selection you make here should match your system hardware clock. Click **Next** to continue.
root Password Selection

You must choose a password for root the user. An alphanumeric password of at least 8 characters, with special characters is recommended.

Click Next to continue.

Selecting Package Groups

This section enables you to select the particular software packages that you wish to install. The default package selections should be good for most users. To examine or change package details, check Customize the set of packages to be installed before clicking Next, otherwise leave the default Accept the current package list.

Tip: Ensure that X Software Development and GNOME Software Development are checked. These packages are required in order to run the cluster management tools.
If you chose to customize (or examine) packages, the next screen shows the details of all packages to be installed (see Figure 3-23).

If you want to know more about a particular package, select it and click on the *Details* link (see Figure 3-23).
Chapter 3. Graphical Installation on the Head Node

Figure 3-23. Package Details

Click Next to continue.

Unresolved Dependencies

You may encounter unresolved dependencies when installing the packages for the Scyld Beowulf. That is, some software packages depend upon others for the system to function properly. If any required packages are missing, you have the opportunity to rectify the situation. Simply select the Install packages to satisfy dependencies button in the ensuing dialog box.

About to Install

All of the information required for installation of Scyld Beowulf has been collected. To start the process of formatting your disks and installing the Scyld Beowulf software, click on the Next button.

Now sit back and relax while the installer does the work. This may be a lengthy process depending on your computer and the number of packages you chose to install. A progress indicator is displayed so that you may monitor the time remaining. The installation may stop at some point requiring another disk to proceed. When this happens, insert the next disc into the CD-ROM drive, and click OK.
Graphical Interface (X) Configuration

Figure 3-24. Video Configuration

The installer automatically probes your system to find the best match for your video card and memory. If it fails to detect them automatically, you must choose from the list of video cards.

If you do not see your card listed, it may be because XFree86 does not support it. However, if you are technically knowledgeable, you may choose Unlisted Card and attempt to configure it by matching the card’s video chipset with one of the available X servers. You are also prompted for the amount of video memory installed on your video card—consult the video card documentation and enter the accurate amount of memory for XFree86 to work properly. If your video card has a video clock chip, choose No Clockchip Setting to let XFree86 automatically detect the proper clockchip, which works in most cases.

You can also choose to skip this step by checking the box labelled Skip X configuration. Click Next to proceed.
Monitor Configuration

The installer probes your monitor and normally identifies it correctly. If probing fails to correctly identify your monitor, or if you wish to change the settings, select the appropriate monitor type. Do not select a monitor that is merely "similar" to yours, unless you are positive that the selected monitor does not exceed the capabilities of your monitor. With older monitors there is a possibility of physical damage if you choose a more capable monitor type.

Clicking the **Restore original values** button reverts to the values discovered by probing the monitor. Click **Next** to proceed.
Customize Graphics Configuration

On this screen, select what color depth and resolution you would like for running X-Windows. Reasonable defaults are chosen by the installer based on your video card and monitor. From this screen you may change the Color Depth and Screen Resolution. Click Next to continue.

Reboot the System

Congratulations, you have successfully installed Scyld Beowulf on the head node of your Scyld Beowulf cluster. Remove the Scyld Beowulf CD-ROM from your drive and click Next to reboot your machine in order to finalize the installation and set up the cluster software.

Welcome

After the system reboots from its own hard disk, you are presented with a Welcome screen. The following steps enable you to finalize your system configuration and to install Scyld Beowulf software.
Chapter 3. Graphical Installation on the Head Node

License Agreement

Verify the standard license agreement by checking Yes, I agree to the License Agreement. Click Next to continue.

Date and Time

Set the date and time for your system. If you are connected to the Internet and want your computer to synchronize its clock with a remote time server using the Network Time Protocol (NTP), check the box labelled Enable Network Time Protocol and choose or provide an NTP server. Click Next to continue.

System User

Although most cluster administration activities require root access, you may wish to set up at least one non-administrative system user. On this page you can provide a username, full name, and password. To enable network login facilities such as Kerberos or NIS, click the Use Network Login button and configure remote authentication. Click Next to continue.

Beowulf Cluster

Figure 3-27. Beowulf Cluster

This page is very important. Under Cluster Interface choose the private ethernet interface which you set up in the Section called Network Configuration. Under Number of nodes, use the number of licensed compute nodes you have in your cluster. For the beginning IP address, choose an address that corresponds to the manual IP configuration you chose for the private network (see Figure 3-17). For example, if you chose 192.168.104.1 as the address for eth1, you could use 192.168.104.10 as the beginning IP address for your cluster.

Sound Card

If the system detects a sound card, you are given an opportunity to play a test sound to ensure that the card is configured correctly. Click Play test sound to play a test sound through all available channels. Click Next to continue.
Additional CDs

Nothing needs to be done for this page. Click Next to continue.

Finish Setup

Congratulations! The head node installation is complete, and the system has been configured. When you click Next, the Welcome series ends and you are presented with a standard Linux login screen.

Now that you have a functioning head node, it is time to boot and configure the compute nodes. Please go on to the next chapter.
Chapter 4. Installation of the Compute Nodes

In Scyld Beowulf clusters, no explicit installation is required on the compute nodes. The head node controls booting, provisioning, and operation of the compute nodes using the configuration solely from the head node.

Compute Node Boot Media

One of the innovations of Scyld Beowulf is the ability to boot compute nodes using a variety of boot mechanisms, yet always use a consistent run-time environment for applications, provisioned dynamically from the head node. This is accomplished without changing the administrative procedures or end-user interface. A second innovation is an architecture that provisions machines as operational compute nodes in as little as one second, even when they have not been previously configured.

PXE Network Boot

The easiest and recommended boot mechanism is PXE, the Preboot eXecution Environment. PXE is a network boot protocol that is nearly ubiquitous on current machines. Older machines may be inexpensively retrofitted by replacing the NIC or adding a boot ROM.

Using direct PXE boot has several advantages over using other boot media. The most significant is that the driver needed to support the specific NIC is included with the hardware. While this driver is not suitable for its run-time use, it eliminates the need to install and update network drivers in two places. A second advantage is speed: PXE boot is faster than using spinning disks, especially floppy disks. For these reasons we recommend using PXE boot whenever it is available.

PXE Media Boot

For network hardware that does not support PXE booting, Scyld supports booting from removable media by installing a special PXE environment to the media.

Beosetup

Scyld Beowulf includes the BeoSetup cluster configuration tool for simplifying the installation of the compute nodes. To begin configuring your compute nodes, you must log in as root user using the password you set in the Section called root Password Selection in Chapter 3. If you chose GNOME for your desktop (the default), you will see a graphical desktop
Figure 4-1. Gnome Desktop

Starting the BeoSetup Tool

BeoSetup is used to execute the installation of the compute nodes in the cluster. These sections provide descriptions of some of the basic functionality. For a detailed description of BeoSetup, see the Administrator’s Guide.

To start BeoSetup, click the link to this tool on the tray in the GNOME desktop (see Figure 4-1). If it is not, you may start the BeoSetup GUI from a terminal window with the command,

`bash$ beosetup`
The BeoSetup program (BeoSetup) is a graphical front-end for configuring and controlling a Scyld Beowulf cluster. It may be run by any user to monitor cluster node state, run commands, and read node logs, but the full functionality is only available to the ‘root’ user. When you start this tool as a user other than root, you are asked for the root password. If you don’t supply it, functionality is limited. For this reason, Scyld recommends running Beosetup as root.

The BeoSetup program is a thin layer over the underlying cluster functionality, not the cluster interface itself. Every operation that it performs and every status that it reports is available from the command line, using scripts and with a library interface. Most of the configuration settings are written to the configuration file /etc/beowulf/config. Many of the actions, such as generating a boot floppy, report the command and options used to accomplish the task.

The first time you run BeoSetup, you see a dialog box asking if you want to Auto Activate nodes as they appear to the head node. Normally you answer Yes, and not have to take any action to add nodes other than powering them on. Answering No requires you to manually activate nodes as described below.
Chapter 4. Installation of the Compute Nodes

The Main Window

The main window contains three panes with Ethernet hardware (MAC) addresses, uniquely identifying machines. The Configured Nodes pane contains machines assigned a node number, along with the relevant state. The other two panes contain a list of MAC addresses. The Ignored pane lists machines that should never be added to this cluster, even though they have requested an IP address from the DHCP service, or a PXE image. The Unknown pane lists machines that have requested an IP address or PXE image, but have not yet been assigned to either of the other two lists.

Addresses may be moved between lists by dragging an address with the left (first) mouse button or by right (third button) clicking on the address with the mouse and choosing the appropriate pop-up menu item. Configured nodes may only be moved if they are in the Down state.

Note that, if you elect to automatically add new nodes to the cluster, or manually configure nodes to be added as described below, nodes do not appear in the Unknown or Ignored lists unless the maximum number of nodes is already connected to the master.

Apply and Revert buttons

After moving addresses between lists, the Apply button must be clicked for changes to take effect. Clicking on the Apply button saves the changes to the configuration file and signals the Beowulf daemons to re-read the configuration file.

Revert re-reads the existing Beowulf configuration file. This has the effect of undoing any undesired changes that have not yet been applied or synchronizing beosetup with any changes that have been made to the configuration file by an external editor.

Short Cuts

Next to the Apply and Revert buttons are short-cut buttons for generating node boot images, Node Floppy and Node CD, generating a new Config Boot, and changing Configuration settings or Preferences. These items are also accessible through the File Menu and Settings Menus.

Pop-up Menus

Each list item has a pop-up menu associated with it that may be accessed by right-clicking while pointing the cursor to the list item. Only those functions in the pop-up menu which may be applied to the highlighted line are clearly visible. Some operations are invalid at certain times, and are "grayed out" (not selectable).

Node Floppy button

If you plan to boot the compute nodes from floppy disk, you may use BeoSetup to create node floppy disks (recommended 1 disk per node) using the following procedure:

1. Click on the Node Floppy button in the main window.
2. Insert a floppy disk into the floppy drive of the head node.
3. Type in any Kernel boot flags required for your nodes. Normally, you need not change these.
4. Click OK to write the boot image to the floppy disk (/dev/fd0).
Chapter 4. Installation of the Compute Nodes

Figure 4-3. Creating BeoBoot Disks

**Node CD button**

If you plan to boot the compute nodes from CD-ROM, you may use BeoSetup to create node CDs (recommended 1 disc per node) using the following procedure:

1. Click on the **Node CD** button in the main window.
2. Type in any **Kernel boot flags** required for your nodes. Normally, you need not change these.
3. Click OK to write the boot image to an ISO image. Writing the image to a CD-R would normally be performed by `cdrecord`.

Figure 4-4. Creating BeoBoot CDs

Booting the Compute Nodes

Boot the compute nodes by powering them on, using the method selected at the beginning of this section (PXE or boot media). As the compute nodes boot, they are listed in BeoSetup by Ethernet Station (MAC) Addresses in the order they connect to the cluster. You may change this order, but it is easiest to power them up in order. The nodes appear in the **Configured Nodes** by default.
Chapter 4. Installation of the Compute Nodes

Incorporating the Compute Nodes

Drag compute node MAC Addresses to the Configured Nodes pane; click Apply. This assigns the nodes to the cluster, using numbers 0 through \( N-1 \), where \( N \) is the maximum number of compute nodes configured.

Optional Compute Node Disk Partitioning

If your compute nodes are diskless, you may skip this section.

Compute node hard disks may be remotely partitioned from the master machine. If the compute node hard disks have not been previously partitioned, you may use `beofdisk` to generate default partition tables for the compute node hard disks. For more details and options regarding the following steps, or to create custom partitioning, see Appendix A.

1. On the master machine, in the directory, capture partition tables for the nodes.
Chapter 4. Installation of the Compute Nodes

bash$ beofdisk -q

With the --q parameter, beofdisk queries all compute nodes. For the first drive found with a specific geometry (cylinders, heads, sectors), it reads the partition table and records it in a file (see Appendix A for more details). If the compute node disk has no partition table, a default table is created. Note: if the partition table on the disk is empty or invalid, a default partition table is not created. The command indicates whether it is creating a default partition table. If the partition table is invalid or empty, you must create a default partition using the --d parameter:

bash$beofdisk -d

2. For each drive of a specific geometry found on any compute node, write the appropriate partition table.

bash$ beofdisk -w

This technique is useful, for example, when you boot a single compute node with a local hard disk that is already partitioned, and you want the same partitioning applied to all compute nodes. You would boot the prototypical compute node, capture its partition table, boot the remaining compute nodes and write that prototypical partition table to all nodes.

3. If needed, update the file /etc/beowulf/fstab on the head node to record the mapping of the partitions on the compute node disks to the filesystems.

Reboot the Compute Nodes

As the root user, reboot all of the compute nodes using these steps:

1. Select the node in the Configured Nodes pane of BeoSetup.
2. Right-click the mouse.
3. Select Change Node State, and select Reboot.

If you use bpctl to reboot the compute nodes, mounted partitions on the compute nodes may not be dismounted properly.

BeoBoot Floppy or CD-ROM

Compute nodes that do not implement PXE require a BeoBoot initial image to boot and operate as a member of the cluster. This BeoBoot image may be created using the BeoSetup tool. You may copy this image onto floppy disk or CDROM, one for each compute node.

• For the floppy or CD boot, see the Section called Node Floppy button or the Section called Node CD button.

Congratulations!

This completes the installation of the compute nodes and your entire Scyld Beowulf cluster.
Chapter 4. Installation of the Compute Nodes
Chapter 5. Cluster Verification Procedure

Once the head and compute nodes have been configured and rebooted we run through the cluster verification to identify common software and hardware configuration problems. (Cluster verification is generally required by reseller technical support when starting on a new issue.) When you call your reseller for support, they will require that you have completed the cluster verification procedures outlined in this chapter and that you capture information using the **beosi** script.

**bpstat**

**bpstat**, run at a shell prompt on the head node, shows a table of status information for each node in the cluster. (You do not need to be a privileged user to use this command.) An example of using this command is shown below.

```
[root@cluster root]# bpstat
Node(s) Status Mode User Group
5-9 down ---------- root root
4    up    ---x--x--x any  any
0-3 up    ---x--x--x root root
```

Ensure that each node is listed as **up**. The node count is based on the number of nodes specified by the `iprange` (see the Preference Settings in **beosetup**). Nodes that have not yet been configured are marked as **down**. Nodes currently booting are temporarily shown with a status of **boot**. Wait 10-15 seconds and try again. A node shows status **error** when it has experienced an initialization problem. As a first step to remedy the problem, right-click on the node entry in the **BeoSetup** display and select **View Syslog**; check for error messages. Typical problems are failing network connections, unpartitioned hard disks, or unavailable network file systems.

**beostatus**

**Beostatus** is a graphical user interface (GUI) program. Start it by either clicking on the icon on the desktop, Figure 5-1, or by typing the command **beostatus** at a terminal windows on the head node.

![Beostatus Icon](image)

**Figure 5-1. Beostatus Icon**

(You do not need to be a privileged user to use this command.) The **beostatus** window is shown in Figure 5-2.
Chapter 5. Cluster Verification Procedure

Figure 5-2. BeoStatus

The default Beostatus GUI mode is known as the "Classic" display. This displays specific state and resource usage information on a per-node table format.

Each row in the beostatus window corresponds to a different node in the cluster. The following are the columns shown:

Node
The node’s assigned node number, starting at zero. Node -1, if shown, is the head node. The total number of node entries shown is set by the iprange or nodes keywords in the file /etc/beowulf/config, rather than the number of detected nodes. Inactive node entries display the last reported data in a faded, grayed-out, row.

Up
A graphical representation of the node’s status. A green checkmark is shown if the node is up and available. Otherwise, a red ‘X’ is shown.

State
The node’s last known state. This should agree with that reported by both bpstat and BeoSetup.

CPU ‘x’
The CPU loads for the node’s processors; at minimum the CPU load for the first processor in each node. Since it is possible to mix uni- and multi-processor machines in a Scyld Beowulf, the number of CPU load columns is equal to the maximum number of processors for any node in your cluster; the label N/A will be shown for nodes with less than the maximum processors.

Memory
The node’s current memory usage.

Swap
The node’s current swap space (virtual memory) usage.
Chapter 5. Cluster Verification Procedure

Disk

The node’s hard disk usage. If a RAMdisk is used, the maximum value shown is one-half the amount of physical memory. As the RAMdisk competes with the kernel and application processes for memory not all the RAM may be available.

Network

The node’s network bandwidth usage. The total amount of bandwidth available is the sum of all network interfaces for that node.

Sanity-check the information shown by beostatus. The configured nodes that are powered up (those with a green checkmark in the Up column) should show expected values in the subsequent usage columns. Assuming no active jobs on your cluster, the CPU and Network usage columns should be fairly close to zero. The memory usage columns (Memory, Swap, and Disk) should be showing reasonable values.

bpsh

The bpsh command is the Beowulf shell command. Analogous in functionality to both the rsh (remote shell) and ssh (secure shell) commands, it’s used to execute commands on the cluster’s compute nodes from the master. For example, this command will execute on node number 3:

```
[root@cluster root]$ bpsh 3 ls -al /tmp
```

linpack

HPL is a portable version of the High Performance Linpack benchmark. Run it against all available nodes using the following shell script; wait up to a minute to see its complete output.

```
[root@cluster root]$ linpack
```

<table>
<thead>
<tr>
<th>Caution</th>
</tr>
</thead>
<tbody>
<tr>
<td>The linpack script runs a non-optimized version of the HPL benchmark, and is intended for verification purposes only. Do not use the results for performance characterization. It may take several minutes on a large cluster.</td>
</tr>
</tbody>
</table>

mpi-mandel

The mpi-mandel program is a visualizer for the Mandelbrot set. The following command is an example of how to run this program using 4 processors:

```
[root@cluster root]$ NP=4 mpi-mandel --demo \ 
/usr/share/doc/mpi-mandel-1.0.20a/mandel.fav
```
Chapter 6. Troubleshooting a Scyld Beowulf Installation

Failing PXE Network Boot

If a compute node fails to join the cluster, there are several places to look:

- First rule out physical problems, like disconnected ethernet cables, malfunctioning network equipment, etc.
- Next, check the log for the compute node. In BeoSetup select the node in question and right-click. Select View Syslog to see the node’s syslog messages. (Select View BeoBoot Log to view the node’s node_up initialization log.) An alternative is to run the System Logs tool from the standard Linux System Tools menu. Select the system log from the list of logs in the left panel and scroll near the end to see errors that may have been reported while the node was booting. The advantage of using BeoSetup’s View Syslog is that it extracts all the node specific information from the syslog into a single view.
- If a node fails to appear either initially (upon power-up) or subsequently disappears, or fails to appear in either the main or Unknown panels of BeoSetup, the node may unable to find the head node’s DHCP server. Another DHCP server may be answering and supplying IP addresses.
- Verify that the Beowulf services are started on the head node by opening the Services applet (see Figure 6-1) and observing whether the Beowulf service is started. Services can be restarted by clicking the restart button of the services applet or from BeoSetup, by choosing Start Cluster from the File menu.

![Figure 6-1. Starting Services Applet](image.png)

If not, click StartCluster item under the File menu (see Figure 6-2). Then power-cycle a compute node to see if it now joins the cluster.
• If you are unable to start the cluster services (or Service Reconfigure from the File in BeoSetup), verify that the head node’s network interface is properly set using the Configuration button on the Network Properties tab (see Figure 6-3). Then start or reconfigure cluster services again.

• Try booting your compute nodes again.

Figure 6-3. Checking Master network

• Verify that the Beowulf services have started (see Figure 6-4). Checking the boxes next to the beostat and beowulf services (in the Service Configuration applet) ensures that these services start at boot time. Be sure to click Save before exiting the applet.
• If the compute nodes fail to boot immediately after power-up but successfully boot later, the problem may lie with the configuration of a managed switch.

Some Ethernet switches delay forwarding packets for approximately one minute after link is established, attempting to verify that no network loop has been created (“spanning tree”). This delay is longer than the PXE boot timeout on some servers.

Disable the spanning tree check on the switch; the parameter is typically named "fast link enable".

Mixed Uniprocessor and SMP Cluster Nodes

The Scyld Beowulf system architecture eliminates the problem of unintentionally running different versions of a program over the cluster’s compute nodes.

The cluster nodes are required to run the same kernel version, typically with the same features and optimization enabled. Uniprocessor machines can run the SMP kernel and SMP machines can run the uniprocessor kernel (although only one processor will be used). The best choice for a mixed cluster is to run the SMP kernel.

The kernel selection is handled at master installation time; based upon the hardware detected. An SMP kernel is installed if the master is detected as an SMP. A similar decision is made based on the processor generation; for instance a kernel compiled to use Intel® Pentium® IV features do not work on a Transmeta(tm) Crusoe(tm) processor.

Note that Intel’s EM64T architecture only supports a single kernel, since even a single CPU utilizes Hyper-Threading. While the name of the kernel may indicate that it is infact a uniprocessor kernel, it is SMP aware.

If you installed a specialized kernel on a master that now needs to support compute nodes with a different set of features (e.g. uniprocessor master with SMP compute nodes), insert the first CD in the Scyld distribution. Linux should automatically launch a package update program. If the CD fails to autorun, follow these steps to install specialized kernels:

1. Mount the Scyld Beowulf CD-ROM on the head node.
2. Change to the directory mount-point/Scyld/RPMS, where mount-point is typically /mnt/cdrom.
3. Copy the kernel RPM, kernel-smp-kernel-version to the head node.
Chapter 6. Troubleshooting a Scyld Beowulf Installation

4. Install this kernel on the head node, by executing:

   bash$ rpm -i kernel-smp-version

5. Reboot the head node and select the name of the SMP kernel from the boot loader prompt or GUI.

6. Make a new Phase 2 image by executing:

   bash$ /usr/bin/beoboot -2 -n

Note that rebooting the head node automatically reboots the compute nodes, causing them to automatically use the updated kernel.

Mixed 32 and 64-bit cluster nodes

Mixing 32 and 64-bit nodes is not possible. The head node is migrating processes to the compute nodes. All nodes in the cluster must have the same CPU architecture. If you want to mix Opteron nodes and IA32 (Pentium or Xeon), you must boot the Opteron in 32-bit mode.

<table>
<thead>
<tr>
<th>Caution</th>
</tr>
</thead>
<tbody>
<tr>
<td>While both Intel EM64T and AMD Operton technologies are x86_64 architectures, combining both in a single cluster is not supported.</td>
</tr>
</tbody>
</table>

Device Driver Updates

Scyld Beowulf releases are tested on many different machine configurations, but it’s impossible to provide device drivers for hardware unknown at release time.

Most unsupported hardware, or device-specific problems, are resolved by updating to a newer device driver. Some devices may not yet be supported under Linux. Check with your hardware vendor.

The Scyld Beowulf architecture makes most driver updates simple. Drivers are installed and updated on the head node exactly as with a single machine installation. The new drivers are immediately available to compute nodes, although already-loaded drivers are not replaced.

There are two irregular device driver types that require special actions: disk drivers and network drivers, both of which apply to the compute nodes. In both cases the drivers must be available to load additional drivers and programs, and are thus packaged in initial ramdisk images.

Another irregular instance is where drivers must execute scripts when they load. (One example is Infiniband.) Contact the hardware vendor or Scyld support if you have difficulty with the script that loads the driver.

Finding Further Information

If you encounter a problem installing your Scyld Beowulf cluster, and you find this guide cannot help you, check the following sources for pertinent information:

- See Installation Guide, Graphical Install of Front-End Node on the head node or on the Scyld disc in the installation kit for detailed installation instructions.
• The *Administrator's Guide* - a description of more advanced administration and setup options - is available on the head node or on the last disc of the Scyld CD set.

• The *Reference Guide* - a complete technical reference to the Scyld Beowulf software - is available on the head node or on the last disc of the Scyld CD set.

• Run the *BeoSetup* application for access to detailed error info regarding the status of booting the compute nodes.

Please visit the Scyld MasterLink™ website at http://www.scyld.com/support.html for the most up-to-date product documentation and other helpful information about your Scyld Beowulf software.
Chapter 6. Troubleshooting a Scyld Beowulf Installation
Appendix A. Compute Node Disk Partitioning

Architectural Overview

The Scyld Beowulf system uses a "diskless administration" model for compute nodes. This means that the compute nodes boot and operate without the need for mounting any file system, either on a local disk or a network file system. By using this approach, the cluster system does not depend on the storage details or potential misconfiguration of the compute nodes, instead putting all configuration information and initialization control on the master.

This does not mean that the cluster cannot or does not use local disk storage or network file systems. Instead it allows the storage to be tailored to the needs of the application rather than the underlying cluster system.

The first operational issue after installing a cluster is initializing and using compute node storage. While the concept and process is similar to configuring the master machine, the "diskless administration" model makes it much easier to change the storage layout on the compute nodes.

Operational Overview

Compute node hard disks are used for three primary purposes:

Swap Space:
expanding the Virtual Memory of the local machine

Application file storage:
providing scratch space and persistent storage for application output

System caching:
increasing the size and count of executables and libraries cached by the local node

In addition, local disk may be used to hold a cluster file system (for use when the node acts as a file server to other nodes).
To make this possible, Scyld provides programs to create disk partitions, a system to automatically create and check file systems on those partitions, and a mechanism to mount file systems.

Partitioning Disks

Deciding on a partitioning schema for compute node disks is no easier than with the head node, but at least it may be more easily changed.

Compute node hard disks may be remotely partitioned using the beofdisk command. The beofdisk command automates the partitioning process, allowing all compute node disks with a matching hard drive geometry (cylinders, heads, sectors) to be partitioned simultaneously.

The beofdisk command may also be used to read an existing partition table on a compute node hard disk, as long as that disk is properly positioned in the cluster. The command captures the partition table of the first hard disk of its type and geometry (cylinder, heads, sectors) in each position on a compute node’s controller (e.g., sda or hdb). The script sequentially queries the compute nodes numbered, 0 through N-1, where N is the number of nodes currently in the cluster.

The default partition table allocates three partitions: a BeoBoot partition equal to 2 MB (currently unused), a swap partition equal to two times the node’s physical memory, and a single root partition equal to the remainder of the disk. The partition table for each disk geometry is stored in the directory /etc/beowulf/fdisk on the master machine with the filename

59
Appendix A. Compute Node Disk Partitioning


While it is not possible to predict every configuration that might be desired, the typical procedure to partition node disks is as follows:

1. Capture partition tables for the nodes. Note, if the nodes’ drives have no partition tables, this command creates a default partition set (and reports this activity to the console). If there is an empty partition table, or an invalid partition table, it is captured and recorded as described, but no default partition set is created. See the Section called Default Partitioning to set up default partitions.
   
   bash$ beofdisk -q

2. Write the appropriate partition table to every drive on every node.
   
   bash$ beofdisk -w

3. Reboot all compute nodes using BetSetup to make the partitioning effective.

Default Partitioning

To apply the recommended default partitioning to all of your disks follow all of these steps:

1. Generate default partition maps to /etc/beowulf/fdisk:
   
   bash$ beofdisk -d

2. Write these out to the nodes:
   
   bash$ beofdisk -w

You must reboot the compute nodes before the new partitions are usable. Rebooting should be done using BeoSetup.

Mapping Compute Node Partitions

If your compute node hard disks are already partitioned, edit the file /etc/beowulf/fstab on the head node to record the mapping of the partitions on your compute node disks to your filesystems. This file contains example lines (commented out) showing mapping of file systems to drives (read the comments in the fstab file for guidance. First query the disks on the compute nodes to determine how that are partitioned.

bash$ beofdisk --q

This creates a partition file in /etc/beowulf/fdisk with a name similar to sda:512:128:32, containing lines similar to:

[root@cluster root]# cat sda:512:128:32
/dev/sda1 : start= 32, size= 8160, id=89, bootable
/dev/sda2 : start= 8192, size= 1048576, Id=82
/dev/sda3 : start= 1056768, size= 1040384, Id=83
/dev/sda4 : start= 0, size= 0, Id=0
Appendix A. Compute Node Disk Partitioning

Read the comments in /etc/beowulf/fstab. Add the lines to the file to use the devices named in the sda file:

```
# This is the default setup from beofdisk
#/dev/hda2  swap swap defaults 0 0
#/dev/hda3  /  ext2  defaults 0 0
/dev/sda1   /boot  ext3  defaults 0 0
/dev/sda2   swap swap defaults 0 0
/dev/sda3   /scratch ext3 defaults 0 0
```

After saving fstab, you must reboot the compute nodes for the changes to take affect.

Generalized, User-Specified Partitions

To create a unique partition table for each disk type/position/geometry triplet, remotely run the `fdisk` command on each compute node where the disk resides:

```
bash$  bpsh n fdisk device
```

where \( n \) is the node number or the first compute node with the drive geometry you want to partition, and `device` is the device you wish to partition (e.g., `/dev/sda`, `/dev/hdb`). One you have created the partition table and written it to the disk using `fdisk`, capture it and write it to all disks with the same geometry using `beofdisk -q`.

```
bash$  beofdisk -w
```

Reboot the compute nodes using BeoSetup before the partitioning is effective.

You must then map filesystems to partitions as described in the Section called Mapping Compute Node Partitions.

Unique Partitions

To generate a unique partition for a particular disk, first partition your disks using one of the above scenarios. Then, from the head node, remotely run `fdisk` on the appropriate compute node to re-create a unique partition table using:

```
bash$  bpsh n fdisk device
```

where \( n \) is the compute node number for which you wish to create a unique partition table and `device` is the device you wish to partition (e.g., `/dev/sda`). You will then need to map file systems to partitions as described in the Section called Mapping Compute Node Partitions.