# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td></td>
</tr>
<tr>
<td>1. Scyld ClusterWare System Overview</td>
<td>1</td>
</tr>
<tr>
<td>System Components and Layout</td>
<td>1</td>
</tr>
<tr>
<td>Recommended Components</td>
<td>1</td>
</tr>
<tr>
<td>Assembling the Cluster</td>
<td>3</td>
</tr>
<tr>
<td>Software Components</td>
<td>3</td>
</tr>
<tr>
<td>2. Quick Start Installation</td>
<td>5</td>
</tr>
<tr>
<td>Introduction</td>
<td>5</td>
</tr>
<tr>
<td>Network Interface Configuration</td>
<td>5</td>
</tr>
<tr>
<td>Cluster Public Network Interface</td>
<td>5</td>
</tr>
<tr>
<td>Cluster Private Network Interface</td>
<td>6</td>
</tr>
<tr>
<td>Network Security Settings</td>
<td>7</td>
</tr>
<tr>
<td>Red Hat Installation</td>
<td>8</td>
</tr>
<tr>
<td>Scyld ClusterWare Installation</td>
<td>8</td>
</tr>
<tr>
<td>Install Yum Command</td>
<td>9</td>
</tr>
<tr>
<td>Configure Yum To Support ClusterWare</td>
<td>9</td>
</tr>
<tr>
<td>Install ClusterWare</td>
<td></td>
</tr>
<tr>
<td>Trusted Devices</td>
<td>10</td>
</tr>
<tr>
<td>Compute Nodes</td>
<td>11</td>
</tr>
<tr>
<td>3. Detailed Installation Instructions</td>
<td>13</td>
</tr>
<tr>
<td>Red Hat Installation Specifics</td>
<td>13</td>
</tr>
<tr>
<td>Network Interface Configuration</td>
<td>13</td>
</tr>
<tr>
<td>Cluster Public Network Interface</td>
<td>13</td>
</tr>
<tr>
<td>Cluster Private Network Interface</td>
<td>16</td>
</tr>
<tr>
<td>Network Security Settings</td>
<td>18</td>
</tr>
<tr>
<td>Package Group Selection</td>
<td>19</td>
</tr>
<tr>
<td>Updating Red Hat or CentOs Installation</td>
<td>21</td>
</tr>
<tr>
<td>Updating RHEL4 using up2date</td>
<td>21</td>
</tr>
<tr>
<td>Updating RHEL4 using media</td>
<td>22</td>
</tr>
<tr>
<td>Updating CentOS 4 Using Yum Update</td>
<td>22</td>
</tr>
<tr>
<td>Updating Using Media</td>
<td>22</td>
</tr>
<tr>
<td>Scyld ClusterWare Installation</td>
<td>22</td>
</tr>
<tr>
<td>Install Yum Command</td>
<td>22</td>
</tr>
<tr>
<td>Configure Yum To Support ClusterWare</td>
<td>23</td>
</tr>
<tr>
<td>Install ClusterWare</td>
<td>23</td>
</tr>
<tr>
<td>Trusted Devices</td>
<td>25</td>
</tr>
<tr>
<td>Enabling Access to External License Servers</td>
<td>26</td>
</tr>
<tr>
<td>Setting the Library Path for LAM</td>
<td>26</td>
</tr>
<tr>
<td>Scyld ClusterWare Updates</td>
<td>26</td>
</tr>
<tr>
<td>Updating ClusterWare</td>
<td>27</td>
</tr>
<tr>
<td>4. Installation of the Compute Nodes</td>
<td>29</td>
</tr>
<tr>
<td>Compute Node Boot Media</td>
<td>29</td>
</tr>
<tr>
<td>PXE Network Boot</td>
<td>29</td>
</tr>
<tr>
<td>PXE Media Boot</td>
<td>29</td>
</tr>
<tr>
<td>The BeoSetup Tool</td>
<td>29</td>
</tr>
</tbody>
</table>
Starting BeoSetup ........................................................................................................... 30
BeoSetup Node Panels ................................................................................................... 31
BeoSetup Action Buttons ............................................................................................... 31
BeoSetup Short-Cuts ....................................................................................................... 32
BeoSetup Pop-Up Menus ................................................................................................. 32
Creating a Node Floppy ................................................................................................. 32
Creating a Node CD ....................................................................................................... 33
Booting the Compute Nodes ......................................................................................... 33
Incorporating the Compute Nodes ................................................................................. 34
Optional Compute Node Disk Partitioning ................................................................. 34
Rebooting the Compute Nodes ..................................................................................... 34
Congratulations! ............................................................................................................. 35

5. Cluster Verification Procedures .................................................................................. 37
   Monitoring Cluster Status .......................................................................................... 37
      bpstat ...................................................................................................................... 37
      BeoStatus ............................................................................................................. 37
   Running Jobs Across the Cluster ............................................................................. 38
      Directed Execution with bpsh .............................................................................. 39
      Dynamic Execution with beorun and mpirun ....................................................... 39
      linpack .................................................................................................................. 39
      mpi- mandel ......................................................................................................... 40

6. Troubleshooting ClusterWare .................................................................................. 43
   Failing PXE Network Boot ...................................................................................... 43
   Mixed Uni-Processor and SMP Cluster Nodes ......................................................... 45
   IP Forwarding ........................................................................................................... 45
   SSH Traffic .............................................................................................................. 45
   Device Driver Updates ............................................................................................. 45
   Finding Further Information .................................................................................... 46

A. Compute Node Disk Partitioning ............................................................................. 47
   Architectural Overview .............................................................................................. 47
   Operational Overview ............................................................................................... 47
   Disk Partitioning Procedures .................................................................................. 47
      Typical Partitioning ............................................................................................. 48
      Default Partitioning ............................................................................................ 48
      Generalized, User-Specified Partitions ................................................................. 48
      Unique Partitions .................................................................................................. 49
   Mapping Compute Node Partitions ......................................................................... 49

B. Changes to Configuration Files .............................................................................. 51
   Changes to Red Hat Configuration Files .................................................................. 51
   Possible Changes to ClusterWare Configuration Files ........................................... 51
Preface

Congratulations on purchasing Scyld ClusterWare, the most scalable and configurable Linux Cluster Software on the market. This guide describes how to install Scyld ClusterWare using Penguin’s installation repository. You should read this document in its entirety, and should perform any necessary backups of the system before installing this software. You should pay particular attention to keeping a copy of any local configuration files.

The Scyld ClusterWare documentation set consists of:

- The Installation Guide containing detailed information for installing and configuring your cluster.
- The Release Notes containing release-specific details, potentially including information about installing or updating the latest version of Scyld ClusterWare.

These product guides are available in two formats, HTML and PDF. You can browse the documentation on the Penguin Computing Support Portal at http://www.penguincomputing.com/support using the username and password you received when you purchased Scyld ClusterWare.

Once you have completed the Scyld ClusterWare installation, you can view the PDF documentation in /usr/share/doc/PDF/scyld-doc/, or launch your Mozilla browser and go to the default page, file:///usr/share/doc/HTML/index.html. In the Featured Links section, click on the ClusterWare Documentation link.

Note: If your reseller pre-installed Scyld ClusterWare on your cluster, you may skip these installation instructions and visit the User’s Guide and Reference Guide for helpful insights about how to use Scyld ClusterWare.

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@penguincomputing.com.

When writing your email, 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 Scyld ClusterWare HPC, Revised Edition, published April 22, 2011.

Notes

1. http://www.penguincomputing.com/support
2. file:///usr/share/doc/HTML/index.html
Chapter 1. Scyld ClusterWare System Overview

System Components and Layout

Scyld ClusterWare 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 "master node" in the cluster is configured with a full Linux installation, distributing computing tasks to the other "compute nodes" in the cluster. Nodes communicate across a private network and share a common process execution space with common, cluster-wide process ID values.

A compute node is commonly diskless, as its kernel image is downloaded from the master node at node startup time using the Preboot eXecution Environment (PXE), and libraries and executable binaries are transparently transferred from the master node as needed. A compute node may access data files on locally attached storage or across NFS from an NFS server managed by the master node or some other accessible server.

In order for the master node to communicate with an outside network, it needs two network interface controllers (NICs): one for the private internal cluster network, and the other for the outside network. It is suggested that the master node be connected to an outside network so multiple users can access the cluster from remote locations.

Figure 1-1. Cluster Configuration

Recommended Components

Hardware selection for a ClusterWare system is based on the price/performance ratio. Scyld recommends the components listed below:

Processors

64-bit Intel® or AMD™ x86_64 architecture required, single-core or multi-core
Chapter 1. Scyld ClusterWare System Overview

**Architecture**
1, 2, or 4 sockets per motherboard

**Physical Memory**
4096 MBytes (4 GBytes) or more preferred, minimum 2048 MBytes (2GBytes)

**Operating System**
Red Hat Enterprise Linux 4 (RHEL4) or CentOS 4 **required**
The *Release Notes* state the specific version and update of Red Hat or CentOS required to support the ClusterWare release you are installing.

**Network Interface Controllers (NIC)**
Gigabit Ethernet (Fast Ethernet at a minimum) PCI-X or PCI-Express adapters (with existing Linux driver support) in each node for the internal private IP network.
The master node typically employs 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).

**Network Switch**
The master node private network NIC and all compute nodes should be connected to a non-blocking Gigabit Ethernet switch for the internal private network. At a minimum, the network switch should match the speed of the network cards.
The switch is a critical component for correct operation and performance of the cluster. In particular, the switch must be able to handle all network traffic over the private interconnect, including cluster management traffic, process migration, library transfer, and storage traffic. It must also properly handle DHCP and PXE.

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

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

If the compute nodes do not support PXE boot, a bootable CD-ROM drive is required. If local disks are required on the compute nodes, we recommend using them for storing data that can be easily 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 master 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 that you provide a second pair of disks in a RAID 1 (mirrored) configuration for these exported file systems. Otherwise, it is possible for accesses to the exported file systems to interfere with the master node accessing its system files, thus affecting the master node’s ability to launch new processes 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 at 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 ClusterWare Cluster Virtualization Software and the underlying Linux operating system are installed only on the master node. A graphic utility, BeoSetup, is included on the master node to aid in the cluster-configuration process.

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

Software Components

The following are integral components of Scyld ClusterWare:

- **BeoSetup**: A graphic utility for configuring and administering a Scyld cluster.
- **BeoStatus**: A graphic utility for monitoring the status of a Scyld cluster.
- **Scyld BeoMaster**: Allows processes to be started on compute nodes in the cluster and tracked in the process table on the master node. Scyld BeoMaster also provides process migration mechanisms to help in creating remote processes, and removes the need for most binaries on the remote nodes.
- **MPICH, MPICH2, MVAPICH, MVAPICH2, LAM, and OpenMPI**: Message Passing Interfaces, customized to work with Scyld ClusterWare.

For more detailed information on these software components, see the Administrator’s Guide and the User’s Guide.

Notes

Chapter 2. Quick Start Installation

Introduction

Scyld ClusterWare is supported on Red Hat Enterprise Linux 4 (RHEL4) and CentOS 4. This document describes installing on Red Hat, though installing on CentOS will be identical, except where explicitly noted. Scyld ClusterWare is installed on the master node after installing a RHEL4 or CentOS 4 base distribution. You must configure your network interface and network security settings to support Scyld ClusterWare.

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

This chapter introduces you to the Scyld ClusterWare installation procedures, highlights the important steps in the Red Hat installation that require special attention, and then steps you through the installation process. Installation is done using the /usr/bin/yum command, installing from a repository of rpms, typically across a network connection. See Chapter 3 for more detailed instructions. Refer to the Red Hat documentation for information on installing RHEL4.

Network Interface Configuration

Tip: To begin, you must know which interface is connected to the public network and which is connected to the private network. Typically, the public interface is eth0 and the private interface is eth1.

It is important to properly configure the network interfaces to support Scyld ClusterWare. The Network Configuration screen is presented during the RHEL4 installation; it can be accessed post-installation via the Applications -> System Settings -> Network menu options.

Cluster Public Network Interface

For the public network interface (typically eth0), the following settings are typical, but can vary depending on your local needs:

• DHCP is the default, and is recommended for the public interface.
• If your external network is set up to use static IP addresses, then you must configure the public network interface manually. Select and edit this interface, setting the IP address and netmask as provided by your Network Administrator.
• If you use a static IP address, the subnet must be different from that chosen for the private interface. You must set the hostname manually and also provide gateway and primary DNS IP addresses.

Tip: When configuring the network security settings (see the Section called Network Security Settings), Scyld recommends setting a firewall for the public interface.
Chapter 2. Quick Start Installation

Figure 2-1. Public Network Interface Configuration

Cluster Private Network Interface

Caution
For the private network interface (typically eth1), DHCP is shown as default, but this option cannot be used. The configuration tool beonetconf requires a static IP address for the private interface. Therefore, you must configure the network interface manually and assign a static IP address and netmask.

Caution
The cluster will not run correctly unless the private network interface is trusted. You can set this interface as a "trusted device" when configuring the network security settings post-installation; see the Section called Trusted Devices.

For the cluster private interface (typically eth1), the following settings are required for correct operation of Scyld ClusterWare:

- Do not configure this interface using DHCP. You must select this interface in the Network Configuration screen and edit it manually in the Edit Interface dialog (see Figure 2-2).
- Set this interface to "activate on boot" to initialize the specific network device at boot-time.
- Specify a static IP address. We recommend using a non-routable address (such as 192.168.x.x, 172.16.x.x to 172.30.x.x, or 10.x.x.x).
- If the public subnet is non-routable, then use a different non-routable range for the private subnet (e.g., if the public subnet is 192.168.x.x, then use 172.16.x.x to 172.30.x.x or 10.x.x.x for the private subnet).
- Once you have specified the IP address, set the subnet mask based on this address. The subnet mask must accommodate a range large enough to contain all of your compute nodes.
Chapter 2. Quick Start Installation

Tip: You must first select the private interface in the Network Configuration screen, then click Edit to open the Edit Interface dialog box.

Tip: Although you can edit the private interface manually during the Red Hat installation, making this interface a "trusted device" must be done post-installation.

Network Security Settings

Caution
The security features provided with this system do not guarantee a completely secure system.

The Firewall Configuration screen presented during the RHEL4 installation applies to the public network interface and should be set according to your local standards.

The RHEL4 installer allows you to select some, but not all, of the security settings needed to support Scyld ClusterWare. The remaining security settings must be made post-installation; see the Section called Trusted Devices.

Scyld has the following recommendations for configuring the firewall:

- Set a firewall for the public network interface (typically eth0).
- If you chose to install a firewall, you must make the private network interface (typically eth1) a "trusted device" to allow all traffic to pass to the internal private cluster network; otherwise, the cluster will not run correctly. This setting must be made post-installation.
- The Red Hat installer configures the firewall with most services disabled. If you plan to use SSH to connect to the master node, be sure to select SSH from the list of services in the Firewall Configuration screen to allow SSH traffic to pass through the firewall.
Chapter 2. Quick Start Installation

Red Hat Installation

Caution
Scyld ClusterWare depends on certain Red Hat packages, and the Scyld installation may fail if the necessary Red Hat packages are not installed. Therefore, Scyld recommends that you install all Red Hat packages.

To install all packages during a Red Hat installation, check the Everything box in the Package Group Selection screen.

To update an existing Red Hat installation to include all packages, insert the first Red Hat CD and invoke the Red Hat update program. Check the Everything box in the Package Group Selection screen, then continue with the update process.

Caution
Check the Release Notes for your release and see whether you must update your Red Hat or CentOS base installation. If you are not familiar with the up2date and yum commands, see the Section called Updating Red Hat or CentOS Installation in Chapter 3 for details on the update procedures. Before running up2date, make sure you have at least version up2date-4.4.67–4 installed on your Red Hat master node.

Scyld ClusterWare Installation

Scyld ClusterWare is installed using the Penguin Yum repository on the Penguin Computing Support Portal. Each Scyld ClusterWare release is continuously tested with the latest patches from Red Hat and CentOS. Before installing or updating your master node, be sure to visit the Support Portal to determine if any patches should be excluded due to incompatibility with ClusterWare. Such incompatibilities should be rare. Then, update RHEL4 or CentOS 4 on your master node before proceeding (excluding incompatible packages if necessary) with installing or updating your Scyld ClusterWare

Install Yum Command

Scyld ClusterWare uses the yum command to install and update software. CentOS 4 includes yum, but RHEL4 does not. If yum is not installed on your master node, then log into MasterLink at http://www.penguincomputing.com/support using the login name and password that was previously communicated to you from Penguin Computing. Click on Yum For RedHat, and download the following RPMs to your system:

- yum-2.4.3-1.noarch.rpm
- python-elementtree-1.2.6-4.2.1.x86_64.rpm
- python-sqlite-1.1.7-1.2.x86_64.rpm
- python-urlgrabber-2.9.8-2.noarch.rpm
- sqlite-3.3.3-1.2.x86_64.rpm

Install the rpms with the following command:

[root@scyld ~]# rpm -ivh yum-2.4.3-1.noarch.rpm \  > python-elementtree-1.2.6-4.2.1.x86_64.rpm \

Chapter 2. Quick Start Installation

> python-sqlite-1.1.7-1.2.x86_64.rpm \\
> python-urlgrabber-2.9.8-2.x86_64.rpm \\
> sqlite-3.3.3-1.2.x86_64.rpm

Configure Yum To Support ClusterWare

The Yum repo configuration file for Scyld ClusterWare must be downloaded from the Penguin Computing Support Portal and properly configured:

1. Login to the Support Portal at http://www.penguincomputing.com/support
2. Click on Download your Yum repo file to download this clusterware.repo file and place the it in the /etc/yum.repos.d/ directory.
3. Set the permissions:
   ```
   [root@scyld ~]# chmod 644 /etc/yum.repos.d/clusterware.repo
   ```

With this setup complete, your master node is ready to retrieve Scyld ClusterWare installations and updates.

Install ClusterWare

You can use Yum to install ClusterWare and all updates up to and including the latest ClusterWare release, assuming you have updated your RHEL4 or CentOS 4 base distribution as prescribed in the ClusterWare Release Notes.

1. Verify the version you are running with the following:
   ```
   [root@scyld ~]# cat /etc/redhat-release
   ```
   This should return a string similar to “Red Hat Enterprise Linux AS release 4 (Nahant Update n)” or “Red Hat Enterprise Linux ES release 4 (Nahant Update n)”, where “n” is the update specified in the Release Notes.
   
   Then reboot your system.
2. Clear the Yum update list and cached packages with the command:
   ```
   [root@scyld ~]# yum clean all
   ```
3. Remove base distribution lam and openMPI files that will cause yum groupinstall to fail, using the command:
   ```
   [root@scyld ~]# yum remove {openmpi,lam}{,devel,-libs}{.i386,.x86_64}
   ```
4. Perform the Yum install:
   ```
   [root@scyld ~]# yum groupinstall Scyld-ClusterWare
   ```
   which prompts you to accept the list of rpms before actually updating the system.
   
   If the Yum install completes successfully, a Complete! message is displayed to the terminal.

   If Yum fails with a Transaction Check Error that complains that a base distribution rpm is newer than the Scyld ClusterWare rpm that is attempting to replace it, then you must manually install the downlevel Scyld ClusterWare rpm(s). For example, if the complaint is about the kernel rpm, then do:
   ```
   [root@scyld ~]# cd /var/cache/yum
   [root@scyld ~]# ls cw-*/packages/kernel-*
   ```
   and locate the newest ClusterWare kernel, which should reside in either cw-core/packages or cw-updates/packages. Then install that newest kernel, e.g.
Chapter 2. Quick Start Installation

[root@scyld ~]# rpm -i --oldpackage cw-updates/packages/kernel-*

and repeat the groupinstall:

[root@scyld ~]# yum groupinstall Scyld-ClusterWare

which should now Complete! successfully.

The latest ClusterWare kernel should now be installed on your system. Edit /etc/grub.conf, if necessary, to make this newest ClusterWare kernel the default for subsequent reboots.

5. Run the beonetconf graphical utility to configure the private network for the Scyld ClusterWare cluster.

In the Cluster Configuration screen, choose your cluster private interface (typically eth1) from the dropdown, then enter the number of licensed compute nodes you have in your cluster and the beginning IP address for the compute nodes. Then click OK.

![Cluster Configuration Screen](image)

**Figure 2-3. Cluster Configuration Screen**

**Tip:** The beginning IP address is chosen by the system as station 100, based on the IP address you chose when you manually configured the private interface. For example, if you chose 172.20.20.1 as the address for eth1, the system would use 172.20.20.100 as the beginning IP address for your compute nodes.

**Tip:** For a Class C subnet, which allows only 253 station addresses, the default starting address of 100 will result in a maximum of 153 nodes in your cluster. To configure more than 153 nodes for a Class C subnet, choose a different starting address and/or a subnet mask that allows more station addresses.

This utility modifies the /etc/beowulf/config file. You can also modify /etc/beowulf/config using the beosetup utility or by directly editing the file. The changes will take effect after the system reboots.

6. Reboot your system.

7. To verify that ClusterWare was installed successfully, do the following:

   [root@scyld ~]# uname -r

   The result should match the specific ClusterWare kernel version noted in the Release Notes.
Chapter 2. Quick Start Installation

Trusted Devices

If you chose to install a firewall, you must make the private network interface (typically eth1) a "trusted device" to enable all traffic on this interface to pass through the firewall; otherwise, the cluster will not run properly. This must be done post-installation.

1. After you have installed Red Hat and Scyld ClusterWare, reboot the system and log in as "root".
3. In the Security Level Configuration dialog box, make sure the private interface is checked in the "trusted devices" list, then click OK.

Tip: If you plan to use SSH to connect to the master node, be sure that SSH is checked in the "trusted services" list.

You are now ready to boot and configure the compute nodes, as described in the next section.

Compute Nodes

In a Scyld cluster, the master node controls booting, provisioning, and operation of the compute nodes. You do not need to explicitly install Scyld ClusterWare on the compute nodes.

Scyld recommends configuring your compute nodes to boot via PXE and using the auto-activate node options, so that each node can automatically join the cluster as it powers on. Nodes do not need to be added manually, unless your compute nodes
Chapter 2. Quick Start Installation

are incapable of PXE booting. If your private network interface doesn’t support PXE booting, or if for some reason you
don’t want to use PXE, you may create compute node disks.

You can boot and configure your compute nodes using the procedure outlined here. See Chapter 4 for more detailed instruc-
tions and additional information for users who want more control over how compute nodes are configured and added to the
cluster.

1. If you are not already logged in as root, log into the master node using the root username and password.
2. Optional step: To get a graphical view of the nodes as they join the cluster, start the BeoSetup tool. Click the BeoSetup
   icon on your desktop or enter beosetup in a terminal window.

   An alternative to using BeoSetup is to enter the command bpstat -U in a terminal window on the master node to view a
   continuously updated table of node status information.
3. Set the BIOS on each compute node to boot either via PXE or via the physical medium you have chosen. Using the
   auto-activate option with PXE booting allows each node to automatically boot and join the cluster as it powers on.
4. Node numbers are initially assigned in order of connection with the master node. 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). You can reorder nodes later as desired; see the Administrator’s Guide.
   If you have BeoSetup open, it will list the compute nodes by node number and Ethernet station or Media Access Control
   (MAC) addresses.
5. The nodes transition through the boot phases. As the nodes join the cluster and are ready for use, they will be shown as
   "Up" by BeoSetup (if open) or in your terminal window if you used bpstat -U.

The cluster is now fully operational with disk-less compute nodes. See Chapter 5 for more about bpstat and node states.

Notes

1. http://www.penguincomputing.com/support
Chapter 3. Detailed Installation Instructions

This chapter provides detailed instructions for installing Scyld ClusterWare. This software installation is intended for the first computer ("node") of the cluster, which functions as the "master node" to control and monitor other nodes and distribute jobs.

Scyld ClusterWare is installed on the master node that is running with a base distribution of RHEL4 or CentOS 4.

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 cluster. If this is not the case, please refer to the previous chapters to acquaint yourself with Scyld ClusterWare, and then verify that your hardware configuration is set up properly.

Red Hat Installation Specifics

During a RHEL4 installation, you have the option to configure various aspects of the installation to support Scyld ClusterWare. Important points include the following:

- **Disk partitioning** — Scyld recommends letting the installer automatically partition the disk; refer to the Red Hat documentation if you plan to manually partition instead.

- **Network interface configuration** — To support your Scyld cluster, you need to configure one interface dedicated to the external public network (typically eth0) and one to your internal private cluster network (typically eth1). Detailed instructions are provided in the section on Network Interface Configuration later in this chapter.

- **Network security settings** — You can configure some of your firewall settings during a RHEL4 installation. Other settings needed to support a Scyld cluster must be made post-installation. Detailed instructions are provided in the sections on Network Security Settings and Trusted Devices later in this chapter.

- **Package group selection** — Scyld recommends installing all Red Hat packages. See the Section called Package Group Selection later in this chapter.

The following sections provide instructions and/or recommendations for specific portions of the RHEL4 installation that are relevant to an optimal Scyld ClusterWare installation. This guide does not cover all steps in the RHEL4 installation; you should refer to the Red Hat documentation for more complete information.

Network Interface Configuration

**Tip:** To begin, you must know which interface is connected to the public network and which is connected to the private network. Typically, the public interface is eth0 and the private interface is eth1.

A typical Scyld cluster has one interface dedicated to the external public network (typically eth0) and one dedicated to your internal private cluster network (typically eth1). It is important to properly to configure both of these interfaces to support your Scyld ClusterWare installation.

The network interface configuration screen will be presented to you during a RHEL4 installation. For an existing Red Hat installation, you can access the network configuration screens through the Red Hat Applications -> System Settings -> Network menu options.
Cluster Public Network Interface

DHCP is selected by default for all network devices, as shown below in the Red Hat Network Configuration Screen. For the public network interface (typically eth0), this option is recommended.

![Network Configuration Screen](image)

Figure 3-1. Public Network Interface (DHCP Default is Recommended)

However, if your external network is set up to use static IP addresses, then follow these steps to manually configure the interface:

1. In the Network Configuration screen, select the public network interface (typically eth0) in the Network Devices list, then click **Edit** to open the Edit Interface dialog box.
2. In the Edit Interface dialog box:
   a. Select the Activate on boot checkbox to initialize the specific network device at boot-time.
   b. Specify the IP address and netmask provided by your network administrator.

When you have completed these settings, click OK to return to the Network Configuration screen.

3. In the Set the hostname area of the Network Configuration screen, select the manually radio button and provide a host name.

4. In the Miscellaneous Settings area of the screen, enter the gateway and primary DNS IP addresses provided by your Network Administrator.
Cluster Private Network Interface

**Caution**

For the private network interface (typically eth1), DHCP is shown as default, but this option cannot be used. The configuration tool `Beonetconf` requires a static IP address for the private interface. Therefore, you must configure the network interface manually and assign a static IP address and netmask.

The cluster will not run correctly unless the private network interface is trusted. You can set this interface as a "trusted device" when configuring the network security settings post-installation; see the Section called `Trusted Devices`.

1. In the Network Configuration screen, select the private network interface (typically eth1) in the Network Devices list, then click **Edit** to open the Edit Interface dialog box.
2. In the Edit Interface dialog box:
   a. Select the *Activate on boot* checkbox to initialize the specific network device at boot-time.
   b. Specify a static IP address. We recommend using a non-routable address (such as 192.168.x.x, 172.16.x.x to 172.30.x.x, or 10.x.x.x).
   c. If the public subnet is non-routable, then use a different non-routable range for the private subnet (e.g., if the public subnet is 192.168.x.x, then use 172.16.x.x to 172.30.x.x or 10.x.x.x for the private subnet).
   d. Once you have specified the IP address, set the subnet mask based on this address. The subnet mask must accommodate a range large enough to contain all of your compute nodes.

   When you have completed these settings, click **OK** to return to the Network Configuration screen.

3. In the *Set the hostname* area of the Network Configuration screen, you have the option to set the hostname automatically via the DHCP server or to provide one manually; this can be done according to your local standards.

The following figure illustrates a completed typical configuration for both the public and private network interfaces.
Chapter 3. Detailed Installation Instructions

Network Security Settings

Caution
The security features provided with this system do not guarantee a completely secure system.

The Firewall Configuration screen presented during the RHEL4 installation applies to the public network interface and should be set according to your local standards. This screen allows you to customize several aspects of the firewall that protects your cluster from possible network security violations.

The RHEL4 installer allows you to select some, but not all, of the security settings needed to support Scyld ClusterWare. The remaining security settings must be made post-installation; see the Section called Trusted Devices.
Scyld recommends setting a firewall for the public network interface (typically eth0). You can configure the following security settings during the Red Hat install:

1. Select from the following firewall options:
   a. *No Firewall* — Allows all connections to your system and does no security checking. This option is not recommended unless you plan to configure your firewall after the installation.
   b. *Enable Firewall* — Blocks any connections to your system that are not defaults or explicitly defined by you. By default, connections are allowed in response to outbound requests, such as DNS replies or DHCP requests.

2. Select services for which you want to allow possible connections. You can select any combination of the services listed.
   
   **Tip:** If you plan to use SSH to connect to the master node, be sure that SSH is checked in the *Trusted Services* list.

3. Set the *Enable SELinux?* dropdown to "Disabled".

If you chose to install a firewall, you must make the private network interface (typically eth1) a "trusted device" to enable all traffic on this interface to pass through the firewall. See the Section called *Trusted Devices*. 

---

*Figure 3-6. Security Settings During Installation*
Chapter 3. Detailed Installation Instructions

Package Group Selection

Caution
Scyld ClusterWare depends on certain Red Hat packages, and the Scyld installation may fail if the necessary Red Hat packages are not installed. Therefore, Scyld recommends that you install all Red Hat packages.

The Red Hat package selection screens enable you to select the particular software packages that you wish to install.

1. In the Package Installation Defaults screen, select the Customize... option.

   Figure 3-7. Customize Package Installation

2. In the Package Group Selection screen, scroll down to the Miscellaneous section. Select the Everything checkbox, then continue the installation process.
Tip: To update an existing Red Hat installation to include all packages, insert the first Red Hat CD and invoke the Red Hat update program. Check the Everything box in the Package Group Selection screen, then continue with the update process.

Updating Red Hat or CentOS Installation

Updating RHEL4 using up2date

Customers who are registered with the Red Hat Network (RHN) and who have up2date installed on their system can use this feature to update their system to the latest RHEL4. Alternatively, you can use Red Hat media. To use up2date, first check your system for the up2date RPM by doing:

[root@scyld ~]# rpm -q up2date

This should return the version of up2date currently installed on the system, such as up2date-4.4.67-4 (or a newer version). Update your system to the latest RHEL4 packages using the command:

[root@scyld ~]# up2date
Chapter 3. Detailed Installation Instructions

See the up2date man page for instructions on installing and using up2date. The Red Hat web site also provides useful information at http://kbase.redhat.com/faq/FAQ_58_2599.shtm. If the update completes successfully, a complete! message is displayed as the last output to your terminal.

**Updating RHEL4 using media**

If you update your system via the Red Hat distribution media, be sure to select an "upgrade install" rather than a "full install", then follow the instructions provided with the Red Hat media.

Now you need to install and configure the Yum repo file for ClusterWare. Go to the section called "Installing and Configuring Yum to Support ClusterWare Updates".

**Updating CentOS 4 Using Yum Update**

Update CentOS 4 either using yum, or using CentOS distribution media. Note that Penguin continually tests ClusterWare with new patches from Red Hat and CentOS. Visit the Penguin Computing Support Portal at http://www.penguincomputing.com/support to see the most recent errata fix tested with ClusterWare, and see any cautions about updated packages which may cause problems with ClusterWare.

You can use any CentOS 4 mirror site. Use the following command:

```
[root@scyld ~]# yum update --disablerepo=cw*
```

(---disablerepo-cw* is used above in case the ClusterWare repo is already installed in /etc/yum.repos.d, you must exclude it during the yum update). You can also exclude other packages using the --exclude=$package parameter. See the yum man page for instructions on using yum. The CentOS web site also provides an online manual for yum at http://www.centos.org/docs/4/html/yum/.

**Updating Using Media**

If you update your system via distribution media, be sure to select an "upgrade install" rather than a "full install", then follow the instructions provided with the media.

*Tip:* The just-installed newest base distribution kernel becomes the default in /etc/grub.conf. However, the Scyld ClusterWare distribution includes a customized kernel that must be the kernel that is booted when running Scyld ClusterWare HPC.

**Scyld ClusterWare Installation**

Scyld ClusterWare is installed using the Penguin Yum repository on the Penguin Computing Support Portal. Each Scyld ClusterWare release is continuously tested with the latest patches from Red Hat and CentOS. Before installing or updating your master node, be sure to visit the Support Portal to determine if any patches should be excluded due to incompatibility with ClusterWare. Such incompatibilities should be rare. Then, update RHEL4 or CentOS 4 on your master node before proceeding (excluding incompatible packages if necessary) with installing or updating your Scyld ClusterWare.
**Install Yum Command**

Scyld ClusterWare uses the `yum` command to install and update software. CentOS 4 includes `yum`, but RHEL4 does not. If `yum` is not installed on your master node, then log into MasterLink at http://www.penguincomputing.com/support using the login name and password that was previously communicated to you from Penguin Computing. Click on *Yum For RedHat*, and download the following RPMs to your system:

- `yum-2.4.3-1.noarch.rpm`
- `python-elementtree-1.2.6-4.2.1.x86_64.rpm`
- `python-sqlite-1.1.7-1.2.x86_64.rpm`
- `python-urlgrabber-2.9.8-2.noarch.rpm`
- `sqlite-3.3.3-1.2.x86_64.rpm`

Install the rpms with the following command:

```bash
[root@scyld ~]# rpm -ivh yum-2.4.3-1.noarch.rpm \
> python-elementtree-1.2.6-4.2.1.x86_64.rpm \
> python-sqlite-1.1.7-1.2.x86_64.rpm \
> python-urlgrabber-2.9.8-2.x86_64.rpm \
> sqlite-3.3.3-1.2.x86_64.rpm
```

**Configure Yum To Support ClusterWare**

The Yum repo configuration file for Scyld ClusterWare must be downloaded from the Penguin Computing Support Portal and properly configured:

1. Login to the Support Portal at http://www.penguincomputing.com/support
2. Click on *Download your Yum repo file* to download this `clusterware.repo` file and place the it in the `/etc/yum.repos.d/` directory.
3. Set the permissions:
   ```bash
   [root@scyld ~]# chmod 644 /etc/yum.repos.d/clusterware.repo
   ```

With this setup complete, your master node is ready to retrieve Scyld ClusterWare installations and updates.

**Install ClusterWare**

You can use Yum to install ClusterWare and all updates up to and including the latest ClusterWare release, assuming you have updated your RHEL4 or CentOS 4 base distribution as prescribed in the ClusterWare *Release Notes*.

1. Verify the version you are running with the following:
   ```bash
   [root@scyld ~]# cat /etc/redhat-release
   ```
   This should return a string similar to “Red Hat Enterprise Linux AS release 4 (Nahant Update n)” or “Red Hat Enterprise Linux ES release 4 (Nahant Update n)”, where “n” is the update specified in the *Release Notes*.
   Then reboot your system.
Chapter 3. Detailed Installation Instructions

2. Clear the Yum update list and cached packages with the command:
   
   ```
   [root@scyld ~]# yum clean all
   ```

3. Remove base distribution lam and openMPI files that will cause `yum groupinstall` to fail, using the command:
   
   ```
   [root@scyld ~]# yum remove {openmpi, lam}{,devel,-libs}{.i386,.x86_64}
   ```

4. Perform the Yum install:
   
   ```
   [root@scyld ~]# yum groupinstall Scyld-ClusterWare
   ```
   
   which prompts you to accept the list of rpms before actually updating the system.

   If the Yum install completes successfully, a `Complete!` message is displayed to the terminal.

   If Yum fails with a `Transaction Check Error` that complains that a base distribution rpm is newer than the Scyld ClusterWare rpm that is attempting to replace it, then you must manually install the downlevel Scyld ClusterWare rpm(s). For example, if the complaint is about the `kernel` rpm, then do:
   
   ```
   [root@scyld ~]# cd /var/cache/yum
   [root@scyld ~]# ls cw-*/*packages/kernel-*
   ```
   
   and locate the newest ClusterWare kernel, which should reside in either `cw-core/packages` or `cw-updates/packages`. Then install that newest kernel, e.g.:
   
   ```
   [root@scyld ~]# rpm -i --oldpackage cw-updates/packages/kernel-*
   ```
   
   and repeat the groupinstall:
   
   ```
   [root@scyld ~]# yum groupinstall Scyld-ClusterWare
   ```
   
   which should now `Complete!` successfully.

   The latest ClusterWare kernel should now be installed on your system. Edit `/etc/grub.conf`, if necessary, to make this newest ClusterWare kernel the default for subsequent reboots.

5. Run the `beonetconf` graphical utility to configure the private network for the Scyld ClusterWare cluster.

   In the Cluster Configuration screen, choose your cluster private interface (typically eth1) from the dropdown, then enter the number of licensed compute nodes you have in your cluster and the beginning IP address for the compute nodes. Then click **OK**.

   ![Figure 3-9. Cluster Configuration Screen](image-url)
**Tip:** The beginning IP address is chosen by the system as station 100, based on the IP address you chose when you manually configured the private interface. For example, if you chose 172.20.20.1 as the address for eth1, the system would use 172.20.20.100 as the beginning IP address for your compute nodes.

**Tip:** For a Class C subnet, which allows only 253 station addresses, the default starting address of 100 will result in a maximum of 153 nodes in your cluster. To configure more than 153 nodes for a Class C subnet, choose a different starting address and/or a subnet mask that allows more station addresses.

This utility modifies the /etc/beowulf/config file. You can also modify /etc/beowulf/config using the beosetup utility or by directly editing the file. The changes will take effect after the system reboots.

6. Reboot your system.

7. To verify that ClusterWare was installed successfully, do the following:

   ```bash
   [root@scyld ~]# uname -r
   ``

   The result should match the specific ClusterWare kernel version noted in the Release Notes.

**Trusted Devices**

If you chose to install a firewall, you must make the private network interface (typically eth1) a "trusted device" to enable all traffic on this interface to pass through the firewall; otherwise, the cluster will not run properly. This must be done post-installation.

1. After you have installed Red Hat and Scyld ClusterWare, reboot the system and log in as "root".

2. Access the security settings through the Red Hat **Applications -> System Settings -> Security Level** menu options.

3. In the Security Level Configuration dialog box, make sure the private interface is checked in the "trusted devices" list, then click **OK**.

   **Tip:** If you plan to use SSH to connect to the master node, be sure that SSH is checked in the "trusted services" list.
Chapter 3. Detailed Installation Instructions

You are now ready to boot and configure your compute nodes; see Chapter 4.

Enabling Access to External License Servers

1. Enable ipforward in the /etc/beowulf/config file. The line should read as follows:
   ```
   ipforward yes
   ```
2. Reboot the compute nodes and restart the cluster services. To do so, run the following two commands as "root" in quick succession:
   ```
   [root@scyld ~]# bpctl -S all -R
   [root@scyld ~]# /etc/init.d/beowulf restart
   ```

Setting the Library Path for LAM

To run LAM, the user must set LD_LIBRARY_PATH to include the compiler specific libraries; otherwise, LAM executables will fail when run. For PGI compilers, set the path as follows:

```
[root@scyld ~]# export LD_LIBRARY_PATH=/usr/lib64/LAM/pgi:$LD_LIBRARY_PATH
```

For Intel compilers, repeat the command for setting the path, and substitute "intel" instead of "pgi" in the directory path.
Scyld ClusterWare Updates

You can use Yum update to update ClusterWare once you have upgraded your RHEL4 or CentOS 4 base distribution. See the Section called Updating Red Hat or CentOs Installation for details on updating your base distribution, and the Section called Scyld ClusterWare Installation for how to install the yum, if necessary, and how to set up the Yum repo configuration files.

To verify which distribution you are currently running, do the following:

[root@scyld ~]# cat /etc/redhat-release

Updating ClusterWare

1. Clear the Yum update list with the command:
   [root@scyld ~]# yum clean all

2. Update ClusterWare as follows:
   [root@scyld ~]# yum groupupdate Scyld-ClusterWare

   This will show you the list of rpms to be updated on your system, and prompt you to accept the list before actually updating the system. If there are any packages you do not want to install, you can add the --exclude=$package parameter to the yum groupupdate command. We do not recommend excluding any ClusterWare packages.

   If the Yum update completes successfully, a Complete! message is displayed as the last output to the terminal.

   If Yum fails with a Transaction Check Error that complains that a base distribution rpm is newer than the Scyld ClusterWare rpm that is attempting to replace it, then you must manually install the downlevel Scyld ClusterWare rpm(s). For example, if the complaint is about the kernel rpm, then do:

   [root@scyld ~]# cd /var/cache/yum
   [root@scyld ~]# ls cw-*/packages/kernel-*

   and locate the newest ClusterWare kernel, which should reside in either cw-core/packages or cw-updates/packages. Then install that newest kernel, e.g.:

   [root@scyld ~]# rpm -i --oldpackage cw-updates/packages/kernel-*

   and repeat the groupupdate:

   [root@scyld ~]# yum groupinstall Scyld-ClusterWare

   which should now Complete! successfully.

3. Reboot your system.

4. To verify that ClusterWare was installed successfully, do the following:
   [root@scyld ~]# uname -r

   The result should match the ClusterWare kernel version noted in the Release Notes.

5. Restart the compute nodes.

Notes

Chapter 3. Detailed Installation Instructions

Chapter 4. Installation of the Compute Nodes

In a Scyld cluster, the master node controls booting, provisioning, and operation of the compute nodes. You do not need to explicitly install Scyld ClusterWare on the compute nodes.

Scyld recommends configuring your compute nodes to boot via PXE and using the auto-activate node options, so that each node can automatically join the cluster as it powers on. Nodes do not need to be added manually, unless your compute nodes are incapable of PXE booting. If your private network interface doesn’t support PXE booting, or if for some reason you don’t want to use PXE, you may create compute node disks.

This chapter provides detailed instructions and additional information for users who want more control over how compute nodes are configured and added to the cluster.

Compute Node Boot Media

One of the innovations of Scyld ClusterWare 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 master 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. See the sections on creating node floppies and creating node CDs later in this chapter.

The BeoSetup Tool

Scyld ClusterWare includes the BeoSetup cluster configuration tool, a graphical front-end for simplifying installation and control of the compute nodes in the cluster. Running BeoSetup isn’t required for adding compute nodes; they will join the cluster by default. However, running BeoSetup as you initially power on the compute nodes gives you a graphical view of the nodes as they join the cluster.

BeoSetup is a thin layer over the underlying cluster functionality, not the cluster interface itself. Every operation that BeoSetup performs and every status that it reports is available from the command line, using scripts and with a library
Chapter 4. Installation of the Compute Nodes

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.

BeoSetup may be run by any user to monitor cluster node state, run commands, and read node logs; however, the full functionality is available only 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".

This section provides descriptions of some of the basic BeoSetup functionality. For more detailed information on BeoSetup, see the *Administrator’s Guide*.

**Starting BeoSetup**

If you chose GNOME (the default) for your desktop, you will see a graphical desktop when you log into Scyld.

![Figure 4-1. Scyld Graphical Desktop](image)

To start the BeoSetup cluster configuration tool, click the BeoSetup icon on the Scyld graphical desktop or enter `beosetup` in a terminal window.

The BeoSetup main window is shown below and described in detail in the following sections. When this window first appears, it will contain no compute nodes.
Chapter 4. Installation of the Compute Nodes

An alternative to using BeoSetup is to enter the command `bpstat -U` in a terminal window on the master node to view a continuously updated table of node status information.

![BeoSetup Window (Initial Instance)](image)

**Figure 4-2. BeoSetup Window (Initial Instance)**

**BeoSetup Node Panels**

The BeoSetup main window contains 3 panes with Ethernet station or Media Access Control (MAC) addresses, uniquely identifying machines as follows:

- **Configured Nodes** — This pane lists machines that have been assigned a node number, along with the relevant state.
- **Ignored Nodes** — This 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.
- **Unknown** — This 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.

You can move nodes between lists by dragging an address with the left mouse button. Alternatively, right-click the address and choose the appropriate pop-up menu item. Configured nodes may be moved only when they are in the "Down" state.

**Tip**: Whether you are automatically adding new nodes to the cluster or manually configuring nodes to be added as described below, nodes do not appear in the "Unknown" or "Ignored" lists unless the maximum number of nodes are already connected to the master.

**BeoSetup Action Buttons**

Above the **Configured Nodes** panel are two action buttons, specifically:

- **Apply** — If you move nodes between lists, you must click the **Apply** button for the changes to take effect. Clicking this button saves the changes to the configuration file and signals the ClusterWare daemons to re-read the configuration file.
Chapter 4. Installation of the Compute Nodes

- **Revert** — Clicking the **Revert** button causes Scyld to re-read the existing configuration file. This will undo any undesired changes that have not yet been applied. This can also be used to synchronize BeoSetup with any changes you have made to the configuration file with an external editor.

**BeoSetup Short-Cuts**

To the right of the **Apply** and **Revert** action buttons is a row of short-cut buttons, specifically:

- **Node Floppy** — The **Node Floppy** short-cut allows you to write the boot image directly to a floppy disk for booting your compute nodes. Details are provided below.
- **Node CD** — The **Node CD** short-cut allows you to write the boot image to your hard drive so that it can be burned to a CD-ROM. Details are provided below.
- **Config Boot** — The **Config Boot** short-cut allows you to access the cluster boot options.
- **Configuration** — The **Configuration** short-cut allows you to access the cluster configuration settings, including network properties, node boot options, node file system options, and PXE network boot settings.
- **Preferences** — The **Preferences** short-cut allows you to choose which columns to display in the **Configured Nodes** panel and to change the node state colors.

All of the short-cut items are also accessible through the **File** and **Settings** menus.

**BeoSetup Pop-Up Menus**

Each listed node has an associated pop-up menu. Click the node to highlight it, then right-click to open the pop-up menu. Only those functions in the pop-up menu that may be applied to the highlighted node are clearly visible. Some operations are invalid at certain times, and are "grayed out" (not selectable).

**Creating a Node Floppy**

> **Caution**
> Boot floppies are supported only for i386 (32-bit) clusters, and cannot be used for x86_64 clusters.

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 this procedure:

1. Click the **Node Floppy** button in the BeoSetup main window to open the Create PXE Boot Floppy dialog box.
2. Insert a floppy disk into the floppy drive of the master node. If the correct drive name does not appear in the dialog box, then use the **Browse** option to locate it. The default is `/dev/fd0`, which specifies the first floppy drive in a standard desktop computer.
3. Click **OK** to write the boot image to the floppy disk.
Chapter 4. Installation of the Compute Nodes

Creating a Node CD

If you plan to boot the compute nodes from CD-ROM, you may use BeoSetup to create node CDs (recommended 1 disc per node). The following procedure creates an ISO image on the hard drive; you can then burn this image to a CD.

1. Click the **Node CD** button in the BeoSetup main window to open the Create PXE Boot CD Image dialog box.
2. If the correct ISO does not appear in the dialog box, then use the **Browse** option to locate it.
3. Click **OK** to write the ISO image to your hard drive.
4. Writing the ISO image from your hard drive to a CD-ROM would normally be performed by the **cdrecord** command.

Booting the Compute Nodes

Boot the compute nodes by powering them on, using the method you previously selected (PXE boot or boot media). As the compute nodes boot, they are activated automatically and added to the cluster in the order they are powered on.

The BeoSetup main window will list the nodes in the **Configured Nodes** panel. By default, this list is ordered by the node number in the ID column. You can reorder this list by Media Access Control (MAC) address or IP address by clicking the appropriate column header.

If you want to add nodes manually, set the appropriate preference before powering on the nodes. In this case, the node MAC addresses will be listed in the **Unknown** panel. See the **Administrator’s Guide** for details on manually adding nodes to the cluster.
Chapter 4. Installation of the Compute Nodes

Incorporating the Compute Nodes

Nodes are assigned to the cluster using the numbers 0 through $N - 1$, where $N$ is the maximum number of compute nodes configured. Scyld recommends that you power up the nodes in order (for example, from the top of the rack to the bottom), so that the node numbers bear some correlation to the physical location of the node.

Optional Compute Node Disk Partitioning

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 the `beofdisk` command to generate default partition tables for the compute node hard disks. For more details, see Appendix A.

Rebooting the Compute Nodes

As the "root" user, reboot all of the compute nodes using these steps:
1. Select the node in the *Configured Nodes* panel of the BeoSetup window, then right-click with the mouse to access the pop-up menu.

2. In the pop-up menu, hover over *Change Node State*, and then select *Reboot* from the submenu.

If you use the *bpctl* command to reboot the compute nodes, mounted partitions on the compute nodes may not be dismounted properly. Use the *node_down* script instead. See the *Reference Guide* for more information on the Scyld commands.

**Congratulations!**

This completes the installation of the compute nodes and your entire Scyld cluster.
Chapter 5. Cluster Verification Procedures

Once the master node and compute nodes have been configured and rebooted, you should run through the cluster verification to identify common software and hardware configuration problems. This chapter describes the Scyld ClusterWare tools for monitoring cluster status and running jobs across the cluster.

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.

Also see the Administrator’s Guide and the User’s Guide for more detailed information.

Monitoring Cluster Status

You can monitor the status of the nodes in your cluster using the bpstat command or the BeoSetup tool, as described in this section.

bpstat

The bpstat command, run at a shell prompt on the master 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.

Following is an example of the outputs from bpstat for a cluster with 10 compute nodes.

```
[root@cluster ~]# 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
```

Some things to keep in mind for bpstat:

- Ensure that each node is listed as "up". The node count is based on the number of nodes specified by the "iprange" parameter; see the network properties in the BeoSetup Cluster Configuration screen.
- 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.
- The "error" status indicates a node initialization problem. Check for error messages by viewing /var/log/beowulf/node.N (where N is the node number). Typical problems are failing network connections, unpartitioned hard disks, or unavailable network file systems.

BeoStatus

The BeoStatus tool is a graphical user interface (GUI) program. You can start it by clicking the BeoStatus icon on the desktop.
Chapter 5. Cluster Verification Procedures

Alternatively, type the command `beostatus` in a terminal window on the master node; you do not need to be a privileged user to use this command.

You can also view the status of the cluster in text mode by typing the command `beostatus -c` at a terminal window on the master node.

The default BeoStatus GUI mode (shown below) is a tabular format known as the "Classic" display. Each row corresponds to a different node, with specific state and resource usage information displayed per node.

![Figure 5-1. BeoStatus in the "Classic" Display Mode](image)

You should sanity-check the information shown in the BeoStatus window. The configured nodes that are powered up (those with a green checkmark in the "Up" column) should show expected values in the associated usage columns. When there are no active jobs on your cluster, the CPU and Network columns should be fairly close to zero. The memory usage columns (Memory, Swap, and Disk) should be showing reasonable values.

- **Node** — The node’s assigned node number, starting at zero. Node -1, if shown, is the master 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. The entry for an inactive node displays the last reported data in a 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 the state reported by both the `bpstat` command and in the BeoSetup window.
- **CPU "X"** — The CPU loads for the node’s processors; at minimum, this indicates the CPU load for the first processor in each node. Since it is possible to mix uni-processor and multi-processor machines in a Scyld cluster, 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 number of processors.
- **Memory** — The node’s current memory usage.
- **Swap** — The node’s current swap space (virtual memory) usage.
- **Disk** — The node’s hard disk usage. If a RAM disk is used, the maximum value shown is one-half the amount of physical memory. As the RAM disk 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.
Running Jobs Across the Cluster

Jobs can be executed on a Scyld cluster using either "directed execution" with the **bpsh** command or "dynamic execution" with the **beorun** or **mpprun** commands.

**Directed Execution with bpsh**

In the directed execution mode, the user explicitly defines which node (or nodes) will run a particular job. This mode is invoked using the **bpsh** command, the ClusterWare shell command analogous in functionality to both the **rsh** (remote shell) and **ssh** (secure shell) commands. Following are some examples of using **bpsh**:

- This example runs **hostname** on the compute node and writes the output back to the user’s screen from compute node 0:

  ```bash
  [user@cluster user]$ bpsh 0 /bin/hostname
  .0
  ```

- The following example runs the **uptime** utility on node 0, assuming it is installed in /usr/bin:

  ```bash
  [user@cluster user]$ bpsh 0 /usr/bin/uptime
  12:56:44 up 4:57, 5 users, load average: 0.06, 0.09, 0.03
  ```

**Dynamic Execution with beorun and mpprun**

In the dynamic execution mode, Scyld decides which node is the most capable of executing the job at that moment in time. Scyld includes two parallel execution tools that dynamically select nodes, **beorun** and **mpprun**. They differ only in that **beorun** runs the job on the selected nodes concurrently, while **mpprun** runs the job sequentially on one node at a time.

The following example shows the difference in the amount of time the system uses to run a command with **beorun** vs. **mpprun**:

```bash
[user@cluster user]$ date;beorun -np 8 sleep 1;date
Fri Aug 18 11:48:30 PDT 2006
Fri Aug 18 11:48:31 PDT 2006

[user@cluster user]$ date;mpprun -np 8 sleep 1;date
Fri Aug 18 11:48:46 PDT 2006
Fri Aug 18 11:48:54 PDT 2006
```

**linpack**

**Caution**

The **linpack** script is a portable, non-optimized version of the High Performance Linpack (HPL) benchmark. It is intended for verification purposes only, and the results should not be used for performance characterization.
Chapter 5. Cluster Verification Procedures

The **linpack** script can be run against all available nodes using the following shell script:

```
[root@cluster ~]# linpack
```

You may need to wait for up to a minute to see its complete output; it may take several minutes on a large cluster. While running **linpack**, you can view BeoStatus to verify that all nodes are taking part in the application. The following figure shows **linpack** running with BeoStatus open in both the GUI mode and the text mode.

![Figure 5-2. linpack with BeoStatus in GUI and Text mode](image)

**mpi-mandel**

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

```
[root@cluster ~]# NP=4 mpi-mandel --demo /usr/share/doc/mpi-mandel-1.0.20a/mandel.fav
```

While running **mpi-mandel**, you can view the BeoStatus display to verify that all nodes are taking part in the application. The following figure shows **mpi-mandel** running with BeoStatus open in the GUI mode.

Note that **mpi-mandel** does not automatically terminate. You can terminate the program manually as follows:

1. Click the **Stop** button in the display window button bar.
2. Select **File -> Exit** from the display window menu to exit the program and close the display window.
Figure 5-3. mpi-mandel with BeoStatus in GUI mode
Chapter 6. Troubleshooting ClusterWare

Failing PXE Network Boot

If a compute node fails to join the cluster when booted via PXE network boot, there are several places to look, as discussed below.

Rule out physical problems.
Check for disconnected Ethernet cables, malfunctioning network equipment, etc.

Check the compute node’s system log.
There are several ways to do this:

- Open the BeoSetup window, select the node in question, and right-click to access the pop-up menu. Select View Syslog to see the master node’s /var/log/messages file, filtered for messages about the selected node. Alternatively, select View BeoBoot Log from the pop-up menu to view the selected node N’s boot/status log, /var/log/beowulf/node.N.
- Run the standard Linux System Logs tool by selecting the System Tools -> System Logs from the desktop menu to open the System Logs window. Select the System Log from the list of logs in the left panel, then scroll near the end to see errors that may have been reported while the node was booting.
- View /var/logs/messages. Viewing with an editor provides a static display of the error messages. Other methods for accessing the system logs may update the display in real time as messages are written to the log.

The advantage of using BeoSetup’s View Syslog option is that it extracts all the node-specific information from the syslog into a single view.

Check for the correct DHCP server.
If a node fails to appear initially (on power-up), subsequently disappears, or fails to appear in either the Configured Nodes or Unknown panels of the BeoSetup window, then the node may be unable to find the master node’s DHCP server. Another DHCP server may be answering and supplying IP addresses.

To check whether the master is seeing the compute node’s DHCP requests, or whether another server is answering, use the Linux tcpdump utility. The following example shows a correct dialog between compute node 0 (10.10.100.100) and the master node.

```
[root@cluster -]# tcpdump -i eth1 -c 10
Listening on eth1, link-type EN10MB (Ethernet),
capture size 96 bytes
18:22:07.901571 IP master.bootpc > 255.255.255.255.bootps:
  BOOTP/DHCP, Request from .0, length: 548
18:22:07.902579 IP .-1.bootps > 255.255.255.255.bootpc:
  BOOTP/DHCP, Reply, length: 430
18:22:09.974536 IP master.bootpc > 255.255.255.255.bootps:
  BOOTP/DHCP, Request from .0, length: 548
18:22:09.974882 IP .-1.bootps > 255.255.255.255.bootpc:
  BOOTP/DHCP, Reply, length: 430
18:22:09.977268 arp who-has .-1 tell 10.10.100.100
18:22:09.977285 arp reply .-1 is-at 00:0c:29:3b:4e:50
18:22:09.977565 IP 10.10.100.100.2070 > .-1.tftp: 32 RRQ
  "bootimg::loader" octet tsize 0
18:22:09.978299 IP .-1.32772 > 10.10.100.100.2070:
```
Chapter 6. Troubleshooting ClusterWare

UDP, length 14
10 packets captured
32 packets received by filter
0 packets dropped by kernel

Check the network interface.
Verify that the master node’s network interface is properly set up. From the BeoSetup window, click the Configuration button to open the Cluster Configuration window. Then check the network interface settings in the Network Properties tab of this window. Then start or reconfigure cluster services again.

![Figure 6-1. BeoSetup Network Properties](image)

Verify that ClusterWare services are running.
Choose System Settings -> Server Settings -> Services from the desktop menu to open the Service Configuration applet. Make sure that the beowulf checkbox is checked in the left panel of this window.
Alternatively, you can check the status of ClusterWare services by entering the following command in a terminal window:

```
[root@cluster ~]# service beowulf status
```

To restart ClusterWare services, click the beowulf entry to highlight it, and then click the Restart button in the icon bar. Be sure to click the Save button before exiting the applet.

You can also restart ClusterWare services from the BeoSetup window by choosing File -> Start Cluster or File -> Service Reconfigure from the menu options at the top of the window. Then power-cycle a compute node to see if it now joins the cluster.

Alternatively, restart ClusterWare services from the command line using either this command:

```
[root@cluster ~]# /etc/init.d/beowulf restart
```

or this command:

```
[root@cluster ~]# service beowulf restart
```
Chapter 6. Troubleshooting ClusterWare

Check the switch configuration.

If the compute nodes fail to boot immediately on 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 Uni-Processor and SMP Cluster Nodes

The Scyld ClusterWare 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. Uni-processor machines can run the SMP kernel. The best choice for a mixed cluster is to run the SMP kernel. Beginning with CW4.1.1, support for uniprocessor kernels was dropped.

IP Forwarding

If IP forwarding is enabled in /etc/beowulf/config but is still not working, then check /etc/sysctl.conf to see if it is disabled.

Check for the line "net.ipv4.ip_forward = 1". If the value is set to 0 (zero) instead of 1, then IP forwarding will be disabled, even if it is enabled in /etc/beowulf/config.

SSH Traffic

The Red Hat installer configures the firewall with most services disabled. If SSH traffic isn’t passing through the firewall, then check your firewall settings to make sure SSH is selected as a trusted service.

To do this, log in as a root user and choose the Red Hat Applications -> System Settings -> Security Level menu option to open the Security Level Configuration window. Then make sure that SSH is selected in the list of trusted services.
Chapter 6. Troubleshooting ClusterWare

Device Driver Updates

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

Most problems with 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 ClusterWare architecture makes most driver updates simple. Drivers are installed and updated on the master 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 RAM disk 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 cluster and find that this Installation Guide cannot help you, the following are sources for more information:

- See the Release Notes for special installation or upgrade procedures that must be taken for your particular version of ClusterWare. It is available on the master node or on the documentation CD included in the Scyld installation kit.
- See the Administrator's Guide, which includes descriptions of more advanced administration and setup options. It is available on the master node or on the documentation CD included in the Scyld installation kit.
- See the Reference Guide, a complete technical reference to Scyld ClusterWare. It is available on the master node or on the documentation CD included in the Scyld installation kit.
- Run the beosetup application for access to detailed error information for the compute nodes, including syslog messages and the node_up initialization log.

For the most up-to-date product documentation and other helpful information about Scyld ClusterWare, visit the Scyld Customer Support website at http://www.penguincomputing.com/support.

Notes

1. http://www.penguincomputing.com/support
Appendix A. Compute Node Disk Partitioning

Architectural Overview

The Scyld ClusterWare system uses a "disk-less 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 "disk-less 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** — Expands the Virtual Memory of the local machine.
- **Application File Storage** — Provides scratch space and persistent storage for application output.
- **System Caching** — Increases the size and count of executables and libraries cached by the local node.

In addition, a local disk may be used to hold a cluster file system (used 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.

Disk Partitioning Procedures

Deciding on a partitioning schema for the compute node disks is no easier than with the master node, but it can be changed more easily.

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

If the compute node hard disks have not been previously partitioned, you can use **beofdisk** to generate default partition tables for the compute node hard disks. The default partition table allocates three partitions, as follows:

- A BeoBoot partition equal to 2 MB (currently unused)
- A swap partition equal to 2 times the node’s physical memory
- A single root partition equal to the remainder of the disk
Appendix A. Compute Node Disk Partitioning

The partition table for each disk geometry is stored in the directory /etc/beowulf/fdisk on the master node, with the filename specified in nomenclature that reflects the disk type, position, and geometry. Example filenames are hda:2495:255:63, hdb:3322:255:63, and sda:2495:255:63.

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.

Typical 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. From the master node, capture partition tables for the compute nodes:

   [root@cluster ~]# 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. If the compute node disk has no partition table, this command creates a default partition set and reports the activity to the console.

   If the partition table on the disk is empty or invalid, it is captured and recorded as described, but no default partition set is created. You must create a default partition using the beofdisk $-d$ command; see the Section called Default Partitioning.

2. Based on the specific geometry of each drive, write the appropriate partition table to each drive of each compute node:

   [root@cluster ~]# 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. Reboot all compute nodes using the beosetup command to make the partitioning effective.

4. If needed, update the file /etc/beowulf/fstab on the master node to record the mapping of the partitions on the compute node disks to the file systems.

Default Partitioning

To apply the recommended default partitioning to each disk of each compute node, follow these steps:

1. Generate default partition maps to /etc/beowulf/fdisk:

   [root@cluster ~]# beofdisk -d

2. Write the partition maps out to the nodes:

   [root@cluster ~]# beofdisk -w

3. You must reboot the compute nodes before the new partitions are usable. Rebooting should be done using beosetup.
Appendix A. Compute Node Disk Partitioning

Generalized, User-Specified Partitions

To create a unique partition table for each disk type/position/geometry triplet, follow these steps:

1. Remotely run the `fdisk` command on each compute node where the disk resides:
   
   ```bash
   [root@cluster ~]# 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`).

2. Once 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:

   ```bash
   [root@cluster ~]# beofdisk -w
   ```

3. You must reboot the compute nodes before the new partitioning will be effective. Rebooting should be done using `beosetup`.

4. You must then map file systems to partitions as described later in this chapter.

Unique Partitions

To generate a unique partition for a particular disk, follow these steps:

1. Partition your disks using either default partitioning or generalized partitions as described above.

2. From the master node, remotely run the `fdisk` command on the appropriate compute node to re-create a unique partition table using:

   ```bash
   [root@cluster ~]# 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`).

3. You must then map file systems to partitions as described below.

Mapping Compute Node Partitions

If your compute node hard disks are already partitioned, edit the file `/etc/beowulf/fstab` on the master node to record the mapping of the partitions on your compute node disks to your file systems. This file contains example lines (commented out) showing the mapping of file systems to drives; read the comments in the `fstab` file for guidance.

1. Query the disks on the compute nodes to determine how they are partitioned:

   ```bash
   [root@cluster ~]# beofdisk -q
   ```

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

   ```bash
   [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

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

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

3. After saving `fstab`, you must reboot the compute nodes for the changes to take affect.
Appendix B. Changes to Configuration Files

Changes to Red Hat Configuration Files

An installation of Red Hat sets a default configuration optimized for a stand-alone server. Installing ClusterWare on a Red Hat installation changes some of these default configuration parameters to better support a cluster. The following sections describe the changes the ClusterWare installation automatically makes to the Red Hat configuration. Any of these may be reversed; however, reversing them may adversely affect the operation of the ClusterWare cluster.

1. /etc/grub.conf has been modified.

   After ClusterWare has been installed, the default boot becomes the newest ClusterWare smp kernel.

2. NFS Services default configuration has been modified.

   By default, Red Hat configures NFS to "off" for security reasons. However, most cluster applications require that at least the home directory of the master node be accessible to the compute nodes. NFS services on the master are set with the default to "on" for run levels 3, 4, and 5.

   The default out-of-box chkconfig for NFS on RHEL4 is as follows:

   ```bash
   [root@scyld ~]# chkconfig --list nfs
   nfs 0:off 1:off 2:off 3:on 4:on 5:on 6:off
   ```

   ClusterWare has changed the default to the following:

   ```bash
   [root@scyld ~]# chkconfig --list nfs
   nfs 0:off 1:off 2:off 3:on 4:on 5:on 6:off
   ```

   To get NFS to mount directories from the master to the compute nodes, the file /etc/exports needs one entry per line for each file system to export from the master to the compute nodes (the RHEL4 default is a blank/non-existent file). ClusterWare creates this file if it didn't already exist, and adds several new entries of the form:

   ```bash
   ExportedDirectoryPathname @cluster (accessMode, syncMode, no_root_squash)
   ```

   The export for /home from the master is configured with an accessMode of rw (read-write) and a syncMode of sync by default for data reliability reasons, and the non-/home directories are exported ro (read-only) for security reasons and async for performance reasons.

   See the ClusterWare Release Notes for details about which directories are added by Scyld.

3. /etc/sysconfig/syslog has been modified.

   Compute nodes will forward messages to the master node’s syslogd daemon, which places them in /var/log/messages. In order for this to function correctly, ClusterWare modifies the /etc/sysconfig/syslog file by adding the -r option to the SYSLOGD_OPTIONS line:

   ```bash
   SYSLOGD_OPTIONS="-m 0 -r"
   ```

Possible Changes to ClusterWare Configuration Files

A clean install of ClusterWare introduces various ClusterWare configuration files that include default settings that a local sysadmin may choose to modify. A subsequent upgrade from one ClusterWare release to a newer release will avoid replacing these potentially modified files. Instead, an update installs a new version of the default file as a file of the form CWconfigFile.rpmnew. Therefore, after a ClusterWare upgrade, the sysadmin is encouraged to compare each such exist-
Appendix B. Changes to Configuration Files

ing CWconfigFile with the new default version to ascertain which of the new default entries are appropriate to manually merge into the preexisting CWconfigFile file.

1. /etc/beowulf/config and config.rpmnew

   ClusterWare specifies additional libraries for compute nodes that may help various applications and scripts execute out-of-the-box.

2. /etc/beowulf/fstab and fstab.rpmnew

   ClusterWare specifies additional /dev devices and NFS-mounted directories for compute nodes that may help various applications and scripts execute out-of-the-box.