SGI® InfiniteStorage Appliance Manager User's Guide

007–4699–005

New Features in This Guide

This revision contains the following:

- Support for NFS using Remote Direct Memory Access (NFS-RDMA) over InfiniBand. See:
  - "NFS-RDMA Over InfiniBand" on page 6
  - "InfiniBand Network Interfaces" on page 23
  - "NFS-RDMA Client Packages" on page 51
  - Chapter 5, "Troubleshooting NFS-RDMA Problems" on page 93
- Support for NFS version 4 (NFSv4). See "Global Options" on page 47.
- Support for backups using Network Data Management Protocol (NDMP). See:
  - "NDMP Configuration" on page 55
  - "NDMP" on page 90
- Support for high-availability (HA) clusters using Linux-HA (Heartbeat) from the High Availability Linux project (http://linux-ha.org/). See:
  - "System Management and Monitoring with Appliance Manager" on page 1
  - "Network Interface" on page 12
  - "Time and Date" on page 15
  - "System Name" on page 56
  - "HA" on page 82
  - "Setting Up an HA Cluster after Reinstalling" on page 103
- The new CIFS **Latencies** graph is the equivalent of the NFS **Service Times** graph. It displays the average amount of time each SMB request took to service. The **Operations by Type** graph has also been made more accurate, and the IOPS graph has also been updated as a result. See "CIFS" on page 84.

- The ability to control whether or not symbolic links made by NFS users that point outside of the Samba share will be followed. See "CIFS Configuration" on page 51.

Caution: This feature is a performance/security tradeoff that is only interesting for sites running both CIFS and NFS from the same filesystem. Allowing linking could be a security risk if, for example, an NFS user created a symbolic link to `/etc/passwd`. However, unchecking the box will cause a decrease in performance.

- The ability to specify the NetBIOS workgroup to which the machine should belong (for sites using CIFS). See and "System Name" on page 56.
- Removal of support for the remote replication of filesystems through the use of NetVault:Replicator and dedicated interfaces.
- Removal of the screen that reports HTTP metrics for default installations of the Apache server.

Record of Revision

Version	Description
001	September 2004 Original publication
002	December 2004 Documents SGI InfiniteStorage NAS Manager version 2
003	October 2006 Documents SGI InfiniteStorage NAS Manager version 3.2
004	January 2007 Documents SGI InfiniteStorage Appliance Manager version 4.0
005	September 2007 Documents SGI InfiniteStorage Appliance Manager version 4.1

Contents

Figures

Tables

About This Guide

This manual describes the operation of SGI InfiniteStorage Appliance Manager. It discusses the following:

- Chapter 1, "Overview" on page 1, describes the tasks you can accomplish with Appliance Manager and introduces the interface
- Chapter 2, "Initial System Setup" on page 9, describes how to use the Setup Wizard to perform your initial system configuration.
- Chapter 3, "Server Configuration and Management" on page 19, describes how to use Appliance Manager to configure the various components of your system and perform general system administration and configuration.
- Chapter 4, "Performance Monitoring" on page 67, describes the current and historical views of the state and the performance of a storage server.
- Chapter 5, "Troubleshooting NFS-RDMA Problems" on page 93, discusses the problems that might be seen with network filesystem remote direct memory access (NFS-RDMA) protocol over InfiniBand
- Appendix A, "How SGI InfiniteStorage Appliance Manager Configures Filesystems" on page 95, describes how Appliance Manager constructs a filesystem and provides an overview of the underlying volume and RAID device configuration that the system uses to lay out the filesystem.
- Appendix B, "How SGI InfiniteStorage Appliance Manager Configures the CXFS Cluster" on page 101, describes how Appliance Manager constructs a CXFS cluster.
- Appendix C, "Reinstalling Appliance Manager" on page 103, describes the procedure to reinstall Appliance Manager after the network is configured and how to set up an HA cluster after reinstalling.

In addition, this document includes a glossary of terms.

Related Publications

For more information, see the following SGI publications:

- *SGI InfiniteStorage Software Platform Release Notes*
- *DMF Administrator's Guide for SGI InfiniteStorage*
- *DMF Filesystem Audit Guide for SGI InfiniteStorage*
- *CXFS Administration Guide for SGI InfiniteStorage*
- *CXFS MultiOS Client-Only Guide for SGI InfiniteStorage*
- *OpenVault Operator's and Administrator's Guide*
- *TMF Release and Installation Guide*
- *TMF User's Guide*
- *TMF Administrator's Guide*
- *XVM Volume Manager Administrator's Guide*

For information about the Linux-HA (Heartbeat) from the High Availability Linux project, see the following website:

http://linux-ha.org/

Obtaining Publications

You can obtain SGI documentation in the following ways:

- See the SGI Technical Publications Library at http://docs.sgi.com. Various formats are available. This library contains the most recent and most comprehensive set of online books, release notes, man pages, and other information.
- You can also view man pages by typing `man` <*title*> on a command line.

Conventions

The following conventions are used throughout this publication:

Convention	Meaning
`command`	This fixed-space font denotes literal items such as commands, files, routines, path names, signals, messages, and programming language structures.
variable	Italic typeface denotes variable entries and words or concepts being defined.
`user input`	Bold, fixed-space font denotes literal items that the user enters in interactive sessions. (Output is shown in nonbold, fixed-space font.)
Menu item	Bold font indicates a menu item or button in the graphical user interface (GUI).
...	Ellipses indicate that a preceding element can be repeated.
`manpage`(*x*)	Man page section identifiers appear in parentheses after man page names.

Reader Comments

If you have comments about the technical accuracy, content, or organization of this publication, contact SGI. Be sure to include the title and document number of the publication with your comments. (Online, the document number is located in the front matter of the publication. In printed publications, the document number is located at the bottom of each page.)

You can contact SGI in any of the following ways:

- Send e-mail to the following address:

  techpubs@sgi.com

- Contact your customer service representative and ask that an incident be filed in the SGI incident tracking system.

- Send mail to the following address:

  SGI
  Technical Publications
  1140 East Arques Avenue
  Sunnyvale, CA 94085–4602

SGI values your comments and will respond to them promptly.

Appliance Manager Comments

If you have comments about using the Appliance Manager software, please send email to SGI engineering at `appman-feedback@sgi.com`. You can also access this from the following menu selection:

Help
> About

Overview

This chapter discusses the following:

System Management and Monitoring with Appliance Manager

Appliance Manager is a web-based interface that lets you configure, manage, and monitor a storage server.

You can use Appliance Manager to do the following:

- Perform initial system configuration using the Setup Wizard.
- Configure the system components.
- Perform general system administration tasks.
- Monitor the state and performance of the storage server, including the following: disk utilization, CPU utilization, network throughput, and services.
- Review historical data describing the state and performance of the storage server.
- View connected clients and determine how each of these contribute to the current workload.
- Detect and investigate problems.
- Set up a two–node cluster to provide high-available (HA) resources that survive a single point of failure. A *resource* is a service associated to an IP address and managed by Linux-HA (Heartbeat). Heartbeat starts, monitors and stops resources. Heartbeat uses the IP address of a resource to redirect clients to the node currently running the resource. Each HA resource is actively owned by one node. If that node fails, another node restarts the HA applications of the failed node. To application clients, the services on the backup node are indistinguishable

from the original services before failure occurred. It appears as if the original member has crashed and rebooted quickly.

- Create CXFS filesystems, add and delete CXFS client-only nodes, monitor CXFS filesystems and nodes, and download CXFS client software to client-only nodes. CXFS 4.1 and later is supported by Appliance Manager. CXFS requires a license.
- Monitor the Data Migration Facility (DMF) and perform certain DMF configuration tasks.

Note: DMF is a hierarchical storage management system for SGI environments. DMF version 3.4 and later is supported by Appliance Manager. Consult the release notes for the procedure to activate the DMF monitoring screens. DMF support requires an additional license, in addition to those for DMF and the base Appliance Manager products themselves.

Appliance Manager Interface

To access the Appliance Manager features, click one of the menu options displayed across the top of the Appliance Manager screen. As you page through Appliance Manager, your location is shown below the menu options. You can also click an item in this path to directly access that location. For example, Figure 1-1 shows the screen you would see if you selected the **CPU Utilization** item from the **Resources** category on the **Monitoring** menu page. This is menu path is shown in this guide as the following:

Monitoring
> Resources
> CPU Utilization

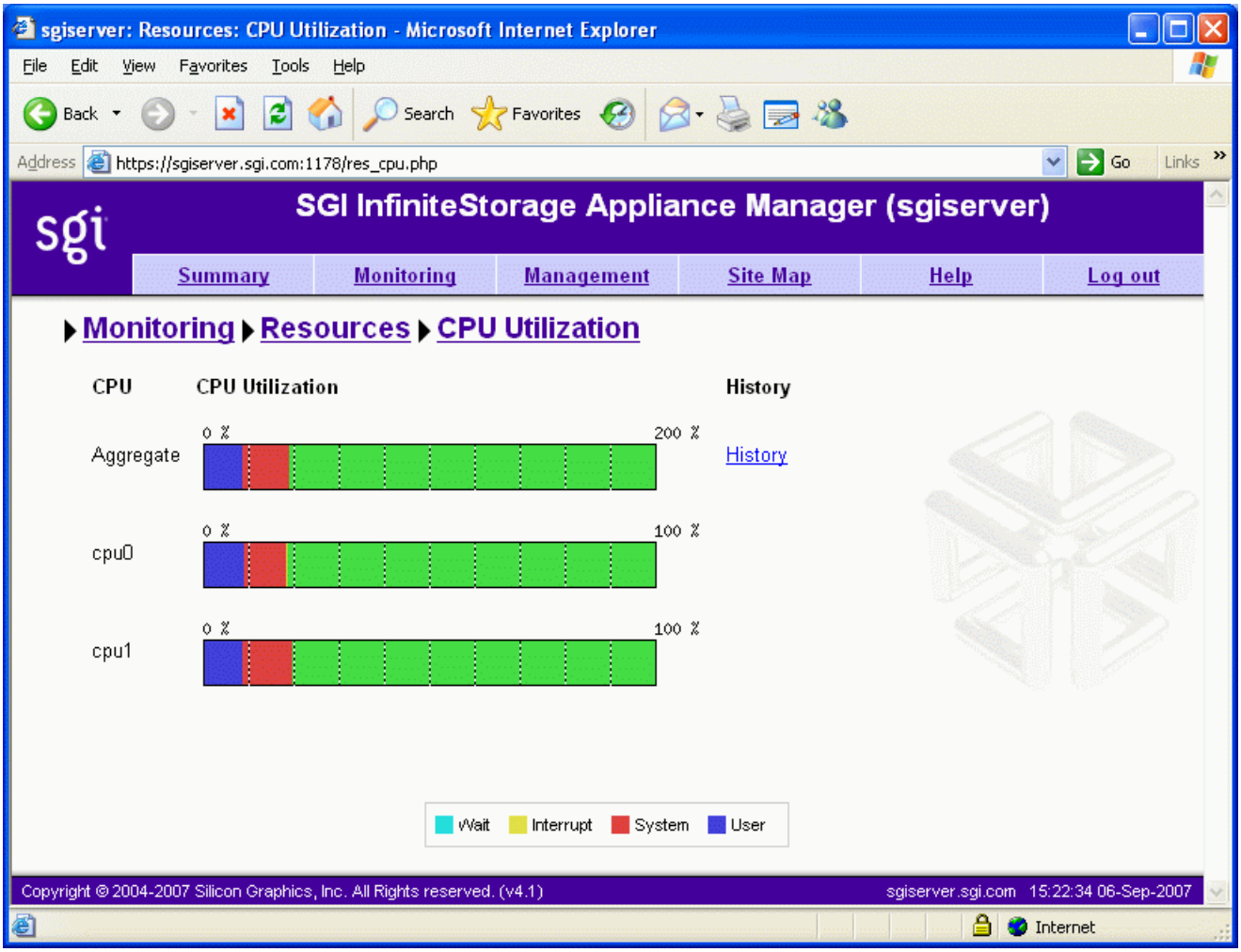

Figure 1-1 Appliance Manager Interface

The menu options are as follows:

Summary Displays a graphic summary of system utilization, including CXFS filesystem and node status (if CXFS is licensed and installed), number of alerts, CPU usage, disk space, disk throughput, network throughput,

current clients, and uptime. See "System Summary" on page 70.

Monitoring

Lets you monitor features in the following categories:

- **Alerts** displays messages from the system logs. See "System Alerts" on page 73.
- **Resources** groups a list of system resources that Appliance Manager monitors. Select a resource (such as **Disk Space**) to display its status. See "Resources" on page 73.
- **Services** groups a list of services provided by the storage server. Select a service (such as **HA**) to display its status. You can also display the current versions of installed software. See "Services" on page 81.
- **Clients** displays various I/O criteria by which to display information about the storage server's clients. See "Clients" on page 91.

Management

Lets you perform tasks in the following categories:

- **Resources** groups a list of system resources that you can configure using Appliance Manager. Select a resource (such as **Network Interfaces**). See:
  - "Network Interface Configuration" on page 21
  - "Storage Configuration" on page 27
  - "DMF Configuration" on page 42
  - "User and Group Configuration" on page 45
- **Services** groups a list of services that you can configure using Appliance Manager. See:
  - "Creating Filesystems" on page 29
  - "CIFS Configuration" on page 51
  - "CXFS Configuration" on page 54

	• **Global Configuration** groups a list of options for various general system administration tasks. See "Global Configuration" on page 56. • **Operations** lets you save and restore the system configuration files that you create with Appliance Manager, gather support and performance data, and shut down or reboot the system. See "Operations" on page 64.
Site Map	Displays an index of direct links to each screen that Appliance Manager displays.
Help	Displays this guide, the release notes, and legal information about Appliance Manager.
Log In	Displays the management log-in screen, in which you enter the administration password that enables you to make changes with Appliance Manager and use the **Management** screens. (No password is required to use the **Monitoring** screens.) You must also enable cookies.
Log Out	Allows you to exit from the management function but still access monitoring functions. (After you have logged in, the menu selection changes to **Log Out**.)

XVM Snapshots

The XVM snapshot feature provides the ability to create virtual point-in-time images of an XFS filesystem without causing a service interruption.

Note: The snapshot feature does not apply to CXFS or DMF filesystems.

The snapshot feature requires a minimal amount of storage because it uses a copy-on-write mechanism that copies only the data areas that change after the snapshot is created.

Snapshot copies of a filesystem are virtual copies, not actual media backup for a filesystem. You can, however, use a snapshot copy of a filesystem to create a backup dump of a filesystem, allowing you to continue to use and modify the filesystem while the backup runs.

You can also use a snapshot copy of a filesystem to provide a recovery mechanism in the event of data loss due to user errors such as accidental deletion. A full filesystem backup, however, is necessary in order to protect against data loss due to media failure.

Note: Use of the snapshot feature requires a license.

Creating filesystem snapshots requires that you first create a snapshot repository in which original copies of regions of data that have changed on the filesystem are stored. If you plan to use the snapshot feature, you must allow room for the snapshot repository on the RAID when you create the filesystems.

NFS-RDMA Over InfiniBand

SGI storage servers support the NFS remote direct memory access (NFS-RDMA) protocol over InfiniBand.

NFS-RDMA requires the following:

- The IP over InfiniBand (IPoIB) server interface must be configured to the same subnet as the client.
- The InfiniBand interfaces on the server must be on different subnets in order to achieve correct routing of packets and good network performance. Do the following

  1. Verify that the IP addresses are different. For example, the following are acceptable (the third number in the set is different, as highlighted):

     ```
     192.168.0.22
     192.168.1.22
     ```

  2. Verify that the IP addresses are still different when the netmasks for their interfaces are applied. In practice, this means that the position where the IP addresses differ must use a value of `255` for the corresponding netmask. For example, given the above IP addresses that differ in the third position, the netmask must be as follows (using `255` in the third position, as highlighted):

     ```
     255.255.255.0
     ```

- For an NFS-RDMA client to mount the NFS-RDMA server, do the following:
  1. Install the NFS-RDMA client packages as described in"NFS-RDMA Client Packages" on page 51.
  2. Use the `rdma` mount option. For example:
     - If the mount server mount address is the Internet Protocol over InfiniBand (IPoIB) address:

       `mount -t nfs -o rdma` *`servername`*`:/mnt/`*`directory`*

       For example:

       ```
       mount -t nfs -o rdma server-ib0:/mnt/data
       ```

     - If the mount server address is not the IPoIB address:

       `mount -t nfs -o rdma=`*`server_IPoIB_address servername`*`:/mnt/`*`directory`*

       For example:

       ```
       mount -t nfs -o rdma=192.168.0.22 server-eth:/mnt/data
       ```

       For more information, see the `mount.nfs` man page.

Initial System Setup

This chapter describes how to use the Setup Wizard to perform the initial system configuration:

- "Accessing the Setup Wizard" on page 9 discusses the Ethernet connections that must be in place in order to run the Setup Wizard.

  Note: Before running the Setup Wizard, ensure that the hardware setup instructions have been completed and verified and that the machine has been powered up. For information on system hardware setup, see your system's *Quick Start* guide.

- "Using the Setup Wizard to Configure the System" on page 10 steps you through the screens of the Setup Wizard.
- "Customizing Your Installation" on page 17 provides an overview of the configuration tasks you may need to perform in order to customize the system for your specific needs after you have finished using the Setup Wizard for initial configuration.

If you must reinstall SGI InfiniteStorage Appliance Manger, see Appendix C, "Reinstalling Appliance Manager" on page 103.

Accessing the Setup Wizard

To access the wizard, do the following:

1. Connect a cross-over Ethernet cable from a laptop or PC to the primary Ethernet port on the storage server as described in the *Quick Start Guide*.
2. Launch a web browser to the following URL:

   https://192.168.9.9:1178

   Note: You may need to temporarily reset the IP address of the laptop or PC to 192.168.9.1 for this to work correctly.

Appliance Manager supports the following browsers:

- Internet Explorer 6.0 or later
- Mozilla Firefox 1.0 or later

The first time you boot the system, the web browser presents the Setup Wizard. After you have completed initial system configuration with the Setup Wizard and restarted the system, the web browser presents the Appliance Manager summary screen, from which you can access all of the product features.

Note: Until you have run through the Setup Wizard, you will not be able to access the rest of Appliance Manager.

When using the Setup Wizard, you may see warning or error messages when you click **Next** after filling in the fields for a particular page. This happens when the system detects a problem in what you have configured. When a warning message appears, the system will still proceed to the next screen. When an error message appears, the system remains on the current screen.

All of the initial system configuration you perform through the Setup Wizard can be later modified using Appliance Manager, as described in Chapter 3, "Server Configuration and Management", in the section "Global Configuration" on page 56 in particular.

Using the Setup Wizard to Configure the System

The initial Setup Wizard screen is the **Introduction** screen, as shown in Figure 2-1. The box at the left of the screen shows the steps that will be covered in order by the Setup Wizard and your location within the steps.

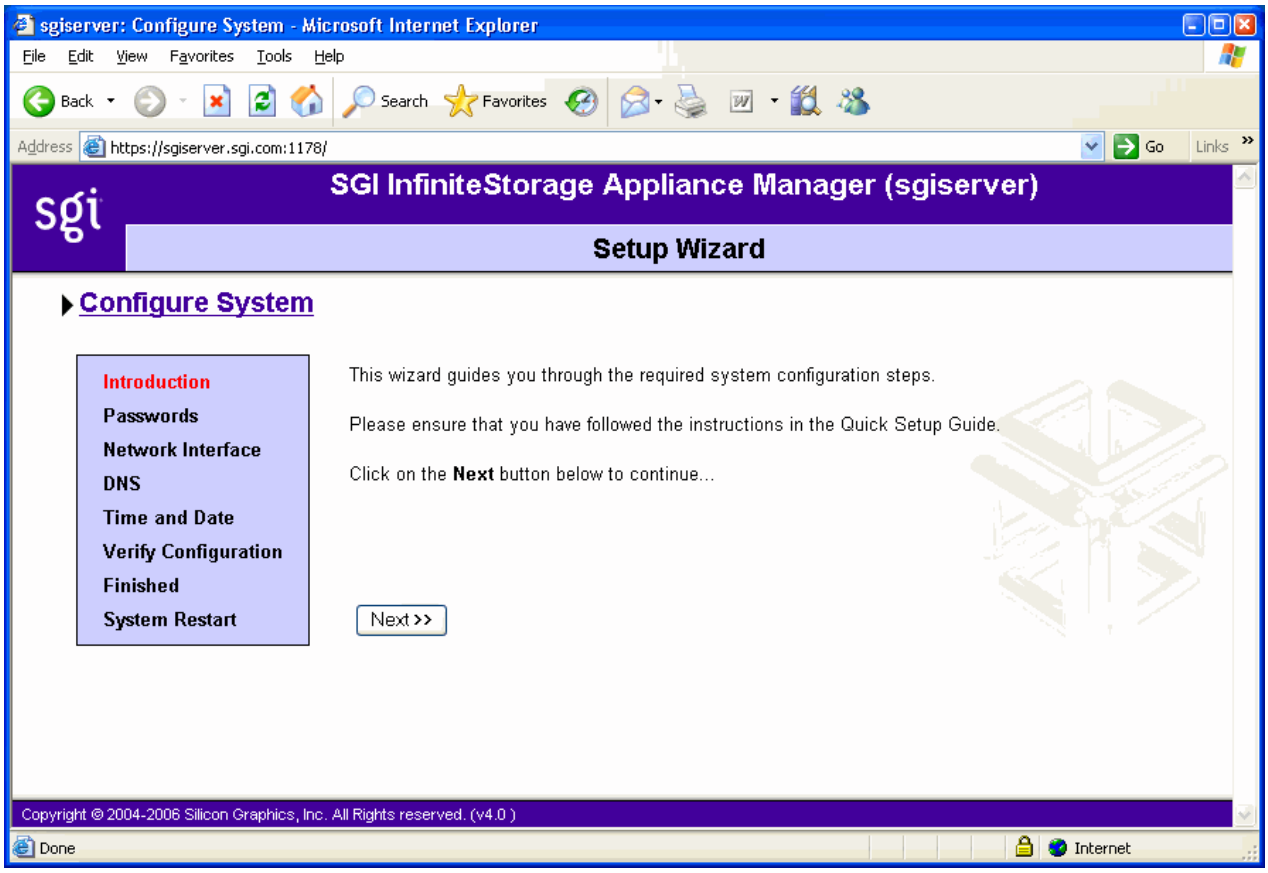

Figure 2-1 Setup Wizard

The system comes preconfigured with a single filesystem that takes up the entire space of the RAID device. On completion of the Setup Wizard, the factory preconfigured filesystem will automatically be configured to be exported and shared via NFS and CIFS. Additionally, on SAN systems, the filesystem will be made available via CXFS. The CXFS cluster is automatically created for you during the setup process. To configure the client-only nodes, see "CXFS Configuration" on page 54.

Note: There are situations that will require a different filesystem configuration than the one that is preinstalled, which takes up the entire space of the RAID device. If you plan to use the XVM snapshot feature or iSCSI targets, for example, you must reserve space on the RAID device. For further information on creating filesystems, see "Filesystems" on page 28.

Click **Next** to display the **Passwords** screen.

Passwords

On initial setup, the **Passwords** screen displays two sets of text boxes, allowing you to enter and confirm the following:

- Appliance Manager administration password. This is the password you must enter in order to perform web-based server configuration and management. The password is not required to view the system monitoring screens. The default administration password is `INSECURE`.
- Command-line configuration password. This is the `root` password for the system

  Note: You must enter the indicated passwords in order to continue with the setup process.

Click **Next** to display the **Network Interface** screen.

Network Interface

The **Network Interface** screen lets you configure the network management interface (`eth0`) for the system and, for HA clusters, the *management IP address*, which is the interface via which Appliance Manager will be accessible. For information, see "Network Interface Configuration" on page 21.

Caution: If you configure an incorrect IP address for the management interface, you can render the system inaccessible from the network.

The system is shipped with `eth0` preconfigured as the management interface and a static IP address of `192.168.9.9`. This lets you plug a laptop into the storage server. For information on other system settings (such as the default gateway), see "Global Configuration" on page 56.

To configure network interfaces in addition to the management interface and to configure aggregated (bonded) interfaces, you must complete the initial system setup and customize your installation, as described in "Customizing Your Installation" on page 17.

Configure the following fields:

System name

Specifies the fully qualified domain name (FQDN) for this storage server. The default hostname is `sgiserver`.

Note: On SAN systems, you cannot change the hostname via Appliance Manager after the **Setup Wizard** completes because changing the hostname in a CXFS cluster is disruptive. To change the hostname later, you must use various CXFS and Linux command-line tools.

Management IP address

(HA clusters only) Specifies the IP address via which Appliance Manager will be accessible (that is, the IP address of the UI resource, which is Appliance Manager)

Workgroup

Specifies the NetBIOS workgroup to which the machine should belong. The default is `WORKGROUP`. If you are not using CIFS, you can ignore this setting.

Default gateway

Specifies the *default network gateway*, which is the IP address of the router that this system should use to communicate with machines that are outside of its subnet. See "System Name" on page 56.

This field can be left blank if either of the following is true:

- The default gateway is supplied by a DHCP server
- All the machines that need to access the system are in the same subnet

IP address	Specifies the IP address for the system if you are not using DHCP. If you have an HA cluster, you must provide the IP address for each node in the cluster.
Subnet mask	Specifies the subnet mask to use for the system if you are not using DHCP. If you have an HA cluster, you must provide the IP address for each node in the cluster.
Use DHCP	Specifies when checked that dynamic host configuration protocol (DHCP) will be used to configure the Ethernet interface. (Another system must be the DHCP server.) For information on DHCP, see "Ethernet Network Interfaces" on page 22. If you require a particular IP address for the system, leave this box unchecked to use static IP addressing.

Click **Next** to display the **DNS** screen.

DNS

If you do not have a DNS server and use only an `/etc/hosts` file, you can leave the fields on this screen blank and use Appliance Manager to modify or import a host file. You can do this after you have completed the initial system setup and restarted the system, as described in "Customizing Your Installation" on page 17. For information on `/etc/hosts` files, see "DNS and Hostnames" on page 62.

Configure the following fields:

Domain search	Specifies the domain name or names that the DNS servers will use to resolve partial name queries. If you have multiple domains, list them in the order you want to use for lookup. This is important in cases where there are two machines with the same name, each on a different domain.
Nameserver #	Specifies up to three IP addresses for the DNS name servers to use. If an address you specify is down, the system will use the next one.

Note: If you specify one or more servers for DNS, all name resolution will be provided by a the specified DNS servers (plus the contents of `/etc/hosts`). If you do not specify a server, only `.local` names will be resolvable via multicast DNS (plus the contents of `/etc/hosts`). You cannot use both DNS to resolve names and multicast DNS to resolve `.local` domain names.

Click **Next** to display the **Time and Date** screen.

Time and Date

Use the **Time and Date** screen to set the following:

Timezone Specifies the local time zone for Appliance Manager.

NTP enable Enables automatic time synchronization with Network Time Protocol (NTP). If the server has Internet access, see the following website for information about using the public NTP timeserver:

http://www.pool.ntp.org/

Note: In order to set the time, you must complete the setup and reboot the system, as described in "Customizing Your Installation" on page 17.

NTP servers Specifies the NTP servers. In an HA cluster, NTP is required and you must specify at least one NTP server.

Click **Next** to display the **Verify Configuration** screen.

Verify Configuration

The **Verify Configuration** screen provides a summary of the configuration information you have entered on the previous screens. For the SAN server, it also shows the CXFS private network that is configured by default.

Note: At this point in the process, the passwords you provided on the **Passwords** screen have been implemented. If you click **Previous** to page back through the screens in order to edit any of the information, the text boxes on the **Passwords** screen will no longer be visible.

Click **Next** to apply this configuration. It takes several seconds for the configuration to be applied; after the configuration changes are is complete, the **Finished** screen will appear.

Finished

The **Finished** screen indicates that the configuration information you entered in the Startup Wizard has been applied. After the software setup phase has completed, Appliance Manager will require a restart.

If you need to modify the custom installation (for example, to add name services or reconfigure the preinstalled filesystem), you will be able to do so after restarting the system.

Click **Next** to restart the system and display the **System Restart** screen.

System Restart

The **System Restart** screen displays as the system is restarting and indicates the Appliance Manager license entitlements and the browser address from which to access Appliance Manager. Point your browser to the following address:

https://*YOUR_SERVER*:1178/

As the system is restarting, you should remove the cross-over cable and connect the management interface into the local network.

Note: After you complete the initial hardware setup and reboot the system, you can customize the installation as described in "Customizing Your Installation" on page 17.

Customizing Your Installation

After completing your system setup and restarting your system, you may need to modify or complete your system installation through configuration procedures that you perform directly with Appliance Manager.

The following aspects of system configuration require that you use Appliance Manager to customize your system:

- Creating a different filesystem configuration than the one that is preinstalled. This will be necessary if you plan to use the following features:
  - XVM snapshots
  - iSCSI targets
  - Any filesytem configuration that requires more than a single filesystem that fills the RAID device

  To configure the system to use these filesystems and files, you must destroy the preconfigured filesystem and create new filesystems. For information on destroying and creating filesystems, see "Filesystems" on page 28.
- Configuring network interfaces in addition to the management interface. For information on configuring network interfaces, see "Network Interface Configuration" on page 21.
- Configuring aggregate interfaces. An *aggregate (bonded) interface* is a virtual network interface that consists of real interfaces working in tandem. A virtual interface can provide the aggregated bandwidth of all of the interfaces that you used to create it. For information, see "Aggregated Network Interfaces" on page 24.
- Modifying the `/etc/hosts` file. For information on `/etc/hosts` files, see "DNS and Hostnames" on page 62.
- Configuring authentication services. For information on configuring Active Directory, LDAP, or NIS for the system, see "Name Service Client" on page 57.
- Configuring local users and groups, as described in "User and Group Configuration" on page 45.
- Setting the time directly, as described in "Time and Date" on page 63.
- Setting the email gateway and the administrator email address to which system messages should be forwarded, as described in "System Name" on page 56.

Server Configuration and Management

This chapter describes how to use SGI InfiniteStorage Appliance Manager to configure the various components of your system and perform general system administration:

- "Network Interface Configuration" on page 21 describes how to configure and modify the network interfaces for the system
- "Storage Configuration" on page 27 describes how to configure filesystems, filesystem snapshots, remote replication, and iSCSI targets
- "DMF Configuration" on page 42 describes the Data Migration Facility (DMF) tasks that you can perform
- "User and Group Configuration" on page 45 describes how to configure a name service client, local users, local groups, and user and group quotas
- "NFS Configuration" on page 47 describes how to configure NFS to share filesystems
- "CIFS Configuration" on page 51 describes how to configure CIFS to share filesystems
- "CXFS Configuration" on page 54 describes how to configure CXFS client-only nodes and manage the CXFS cluster
- "NDMP Configuration" on page 55 describes how to configure Network Data Management Protocol (NDMP) for backups
- "Global Configuration" on page 56 describes how to perform various general administration functions
- "Operations" on page 64 describes how to save changes to the configuration files and restore them, how to gather support and performance data, and shut down the system

Figure 3-1 shows the top level **Management** screen.

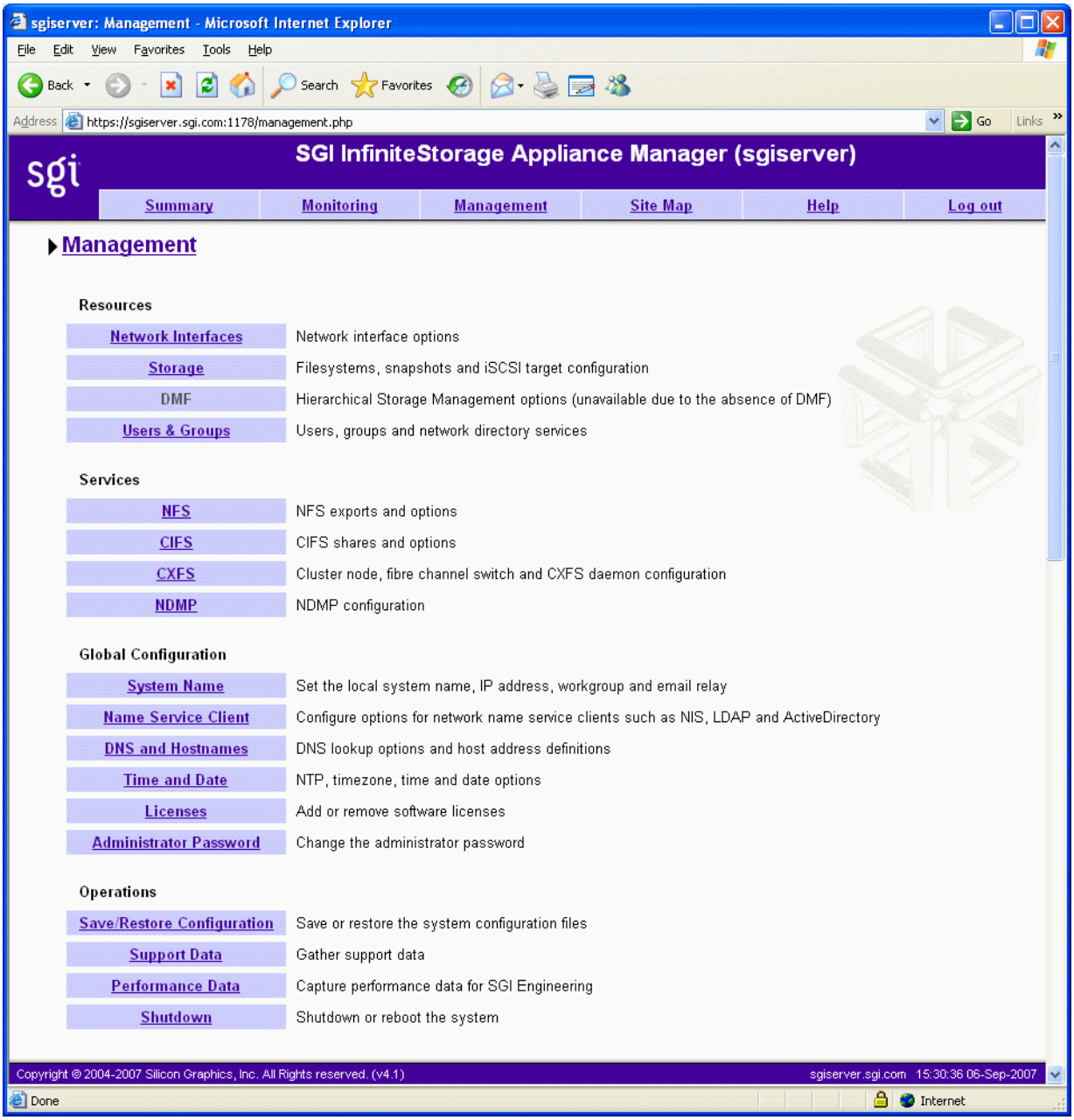

Figure 3-1 Management Screen

Network Interface Configuration

You can use Appliance Manager to configure and modify the network interfaces for the system. When configuring the system, you must consider the difference between the management interface and the remainder of the interfaces in the system.

The management interface is the first interface in the machine, `eth0`, which is dedicated for use by Appliance Manager. On a NAS system, the remainder of the interfaces in the system are used for fileserving. On a SAN system, the remainder of the interfaces are preconfigured for the CXFS private network and connection to the Fibre Channel switch.

Caution: Changing the network interface configuration for a SAN system can leave the CXFS cluster inoperative. If you are required to change the configuration, you must do so carefully by using the `cxfs_admin` command, or the CXFS GUI. For more information, see Appendix B, "How SGI InfiniteStorage Appliance Manager Configures the CXFS Cluster" on page 101.

You can configure these ports as individual, standalone ports or you can group these ports together into an *aggregated (bonded) network interface*.

Bonding interfaces together gives you the aggregated bandwidth for multiple clients of all of the interfaces that constitute the aggregated interface. For most systems, this can significantly increase performance over a system in which all of the interfaces are configured as individual network ports.

For further information, see:

- "Management Interface" on page 22
- "Ethernet Network Interfaces" on page 22
- "InfiniBand Network Interfaces" on page 23
- "Aggregated Network Interfaces" on page 24

Caution: Ensure that the hardware settings are correct before you configure the network interfaces. For information on hardware setting, see the *Quick Start Guide* for your system.

Management Interface

When the system is shipped from the factory, the management interface has a preconfigured IP address. When using the Setup Wizard, you connect a laptop to that interface in order to perform the initial setup tasks. For information on the Setup Wizard, see Chapter 2, "Initial System Setup" on page 9.

The management interface is always configured as an individual network interface and cannot be part of an aggregated interface.

You can modify the management interface by selecting `eth0` from the following screen:

Management
> Resources
> Network Interfaces
> Modify

For information on the network configuration parameters you can modify, see "Ethernet Network Interfaces" on page 22.

Caution: If you configure an incorrect IP address for the management interface, you can make Appliance Manager inaccessible.

Ethernet Network Interfaces

To see the available Ethernet network interfaces and change their parameters, select the following:

Management
> Resources
> Network Interfaces
> Modify

To modify an interface, select it. You can change the following fields:

Enabled	Enables the interface. You cannot disable the management interface.
Speed	Displays the port speed of the Ethernet card, which is usually **Autonegotiate**.

Duplex	Displays the duplex of the Ethernet connection, which is usually **Autonegotiate**.
Automatic discovery by DHCP	Specifies that dynamic host configuration protocol (DHCP) will be used to configure the Ethernet interface. (Another system must be the DHCP server.)
Static	Specifies that a particular IP address is required for the network interface. If you select this, you must provide the IP address and subnet mask.

InfiniBand Network Interfaces

To see the available InfiniBand network interfaces and change their parameters, select the following:

Management
> Resources
> Network Interfaces
> Modify

Note: For NFS-RDMA to function correctly, the IP over InfiniBand (IPoIB) interface must be configured to the same subnet as the clients. See "NFS-RDMA Over InfiniBand" on page 6.

To modify an interface, select it. You can change the following fields:

Enabled	Enables the interface.
Speed	(Does not apply to InfiniBand.)
Duplex	(Does not apply to InfiniBand.)
Automatic discovery by DHCP	(Not supported in this release.)

Static	Specifies that a particular IP address is required for the network interface. If you select this, you must provide the IP address and subnet mask.

Aggregated Network Interfaces

An aggregated (bonded) interface is a virtual network interface that consists of real interfaces working in tandem. You use bonded interfaces on NAS systems to increase bandwidth to NFS and CIFS clients. (It does not apply to CXFS clients because they are connected via Fibre Channel.)

A virtual interface can provide the aggregated bandwidth of all of the interfaces that you used to create it.

Note: Any single client can achieve the bandwidth of only a single interface at a time. An aggregated interface increases the aggregate bandwidth for multiple clients.

For example, if you have three interfaces each with a bandwidth of 10, the aggregate bandwidth is 30. For an individual client, however, the maximum bandwidth remains 10. When additional clients access the aggregated interface, the clients are assigned to the subinterfaces, and up to three clients can use a bandwidth of 10 at the same time. Thus multiple clients accessing the system increase the aggregate bandwidth, improving the performance to a maximum bandwidth of 30.

For example, Figure 3-2 shows a configuration in which all clients connect to a single IP address (`192.168.0.3`). The switch is responsible for sharing the load across 4 aggregated interfaces (`eth1--eth4`). Therefore, 4 times as many clients can communicate with the same server without a loss in overall performance.

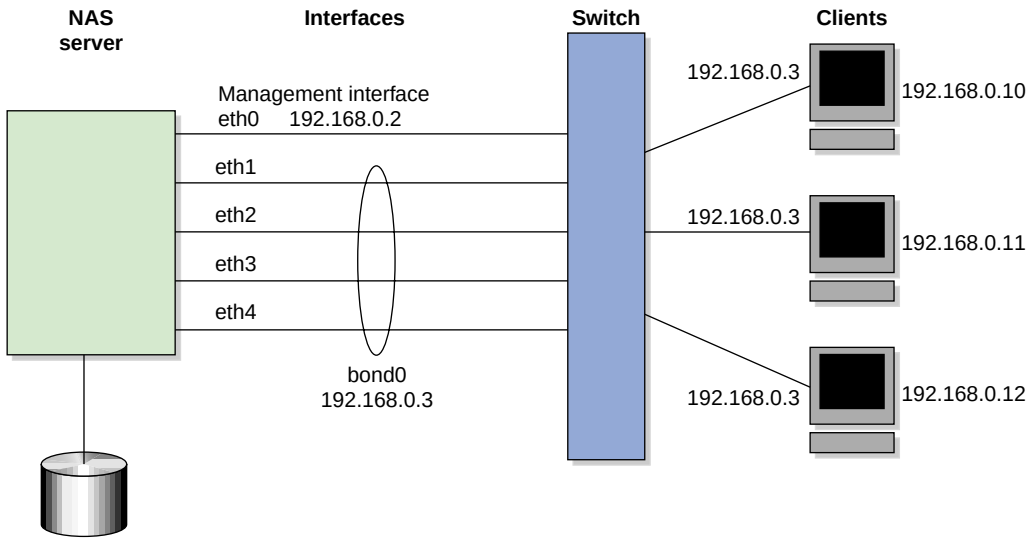

Figure 3-2 Aggregated Network Interfaces

Output load balancing controls how the server chooses which subinterface to send replies. *Input load balancing* controls how clients are assigned to subinterfaces, and how and when clients are moved from one subinterface to another. Load balancing happens on a per-packet basis. When a client sends a packet, it traverses a switch, which determines at which subinterface the packet arrives. Input load balancing ensures that each client arrives at a different subinterface. The clients see only one interface because the balancing is done by the system.

In addition to configuring an aggregated interface in Appliance Manager, you must configure the ports on the switch so that they use either static trunking or 802.3ad dynamic trunking. For more information, refer to the user manual for your switch.

To create an aggregated interface, select the following:

Management
> Resources
> Network Interfaces
> Create an aggregated interface

The available interfaces are displayed for selection.

When you configure an aggregated interface, you specify the following:

Bonding mode

Selects a bonding mode that governs the relation of the subinterfaces to a switch and defines the protocol that is used for assigning network switch ports to an aggregated interface:

- *Dynamic bonding* uses the 802.3ad protocol to communicate with the switch and automatically bond the appropriate switch ports together. You may need to configure your switch to enable the 802.3ad protocol on a range of switch ports or for the switch as a whole.
- *Static bonding* requires that the switch be manually configured to bond specific switch ports together.

Your choice depends upon what your switch supports:

- If your switch supports the 802.3ad protocol, choose dynamic bonding.
- If your switch only supports manually grouping ports together in a bond, choose static bonding.
- If your switch does not support any bonding, you must configure all your network interfaces as separate individual interfaces.

Output Load Balancing

Specifies how the server chooses which subinterface to send replies:

- **Layer 3 (IP header)** specifies that the server and client are on different subnets.
- **Layer 2 (MAC address)** specifies that all packets sent to the clients use separate MAC addresses. This option is more efficient than **Layer 3 (IP header)**. Use this option only if the clients are in the same broadcast domain as the server.

Note: Do not select this option if the switch immediately upstream of the server is acting as a router rather than a switch (that is, making packet routing decisions at Layer 3 rather than Layer 2) or if the clients are in a different subnet and you have another router between the server and clients.

IP address Specifies the IP address of the new aggregated interface.

Note: The IP address for an aggregated interface must be configured statically. Appliance Manager does not support DHCP and dedicated IP addresses for aggregated interfaces.

Subnet mask Specifies the subnet mask of the new aggregated interface.

Click **Create** to create the bond.

Storage Configuration

You can use Appliance Manager to configure the following:

- XFS or clustered CXFS filesystems
- iSCSI targets
- XVM filesystem snapshots

These features are available under the following menu selection:

Management
> Resources
Storage

The following sections describe these features:

- "Filesystems" on page 28
- "iSCSI Targets" on page 34

Filesystems

This section describes how to list, create, grow, and destroy filesystems on an SGI RAID device:

For background information about how Appliance Manager works, see Appendix A, "How SGI InfiniteStorage Appliance Manager Configures Filesystems" on page 95.

Listing Filesystems

To display a brief description of the RAID to which Appliance Manager is connected, use the **List** option:

Management
> Resources
> Storage
> Filesystems
> List

This includes the worldwide name (WWN) of the RAID device and an indication of the RAID status, which will be **ONLINE** unless a hardware or software failure mode has prevented communication between Appliance Manager and the array firmware (such as if the array is powered down or a cable has been pulled out).

Note: Appliance Manager displays the mounted filesystems that you have created through Appliance Manager.

The **Type** field on this screen indicates whether the listing is a filesystem, a snapshot repository, or iSCSI storage.

If you have created a snapshot repository but have not scheduled any snapshots to be taken and stored on that repository, its size will appear as 0 on this display.

Creating Filesystems

The **Create** option steps you through a filesystem creation wizard. The steps that the wizard will take are listed in a box to the left of the screen, with the current step highlighted.

The filesystem creation procedure is mostly automatic. You provide the name, size, and general characteristics of the filesystem to create and Appliance Manager determines the underlying layout of the filesystem on the disk devices. For information on how Appliance Manager calculates how to allocate the disk resources, see Appendix A, "How SGI InfiniteStorage Appliance Manager Configures Filesystems" on page 95.

There is a limit to the number of filesystems on a particular array. This limit is less than 30 filesystems for a 4-tray array, but it can be smaller on large arrays (because each filesystem will use 2 or 3 of the total 254 LUN numbers per tray of disks in the array). For this reason, SGI recommends that you use as few filesystems as possible on a disk array. The number of filesystems and repositories that you can create depends on the make and model of the storage arrays that are connected. Some arrays are capable of supporting up to 254 LUN numbers, but others support only 31 or fewer. The number of LUN numbers consumed by a filesystem/repository depends upon the number of disks and the size of the disks and trays that are connected to the storage array. SGI recommends that you create as few filesystems as possible in order to save LUN numbers (which can later be utilized to grow the filesystem) and because the storage subsystem performs better with fewer filesystems configured.

Note: When you create the filesystem, the system detects whether the disk configuration is supported and issues a warning if it is not. You can continue to create the filesystem under these circumstances, but the filesystem will not be an efficient one.

You can grow an XFS filesystem after you have created it, by whatever size you choose. It is most efficient, however, if you create a filesystem that fills the disk array and add additional disks if you need to grow the filesystem, filling those disks when you do.

Perform the following steps to create a filesystem:

1. Select the **Create** option:

   Management
   > Resources
   > Storage
   > Filesystems
   > Create

2. Appliance Manager searches for the RAID arrays on the system and displays them on the **Arrays** screen. If you have more than one storage array, a list of arrays will be presented and you can chose on which arrays the filesystem should be created. The resulting filesystem will be spanned across multiple storage arrays. Spanning filesystems across multiple arrays is possible only for external storage arrays (the SGI InfiniteStorage series). Click **Next**.

3. The **Options** screen displays the filesystem configuration options. These are based on the devices that are available to the system and include the following categories:

   Drive Type — The drive type may be Serial-attached SCSI (SAS), Serial ATA (SATA) or Fibre Channel (FC). You cannot create a filesystem that spans multiple types of disks.

   Goal — You can select a filesystem optimized for performance or capacity (if appropriate for your system). If you select for capacity, Appliance Manager will use all the available disk space to create the filesystem, although this will usually come at the cost of slower performance.

   Workload — You can select a filesystem optimized for bandwidth or for I/O per second (IOPS). Select **Bandwidth** when you will have a small set of files and you must perform streaming reads and streaming writes as fast as possible. Select **IOPS** when you will be performing random reads and writes to different sets of files. Normally, **IOPS** will be the better choice.

If you are optimizing for IOPS, it is best to build one large filesystem. In general, there is a cost to having multiple filesystems.

Available Space This displays the available space in gigabytes.[1]

Click **Next**.

4. On the **Purpose** screen, select whether the filesystem will be a clustered CXFS filesystem or an XFS filesystem. The **Purpose** screen will appear if Appliance Manager is managing a SAN (CXFS) system or if DMF is installed. Depending on the existence of CXFS and DMF, you will be asked if you want to create a clustered CXFS filesystem or a local XFS filesystem, and with or without DMF support. DMF filesystem option will create the filesystem with 512–byte inodes, and the `dmapi` and `mtpt` mount options as required for DMF support. (It will not add the filesystem to the DMF configuration file; you must do this later manually.) For more information about DMF , see "DMF Configuration" on page 42.

5. On the **Name & Size** screen, enter the following:

   - Filesystem mount point (must be begin with `/mnt/` as shown)
   - Filesystem size in gigabytes [2] The default filesystem size is the size of a filesystem that will completely fill the disk devices. If you choose less than this maximum size, the filesystem will be divided up among the disks. For example, if you create a filesystem that is 20% of the maximum size, it will be spread out among the first 20% of each disk. If you create a second filesystem that is also 20% of that maximum size, it will be spread out among the second 20% of each disk.

   Note: If you plan to use the XVM snapshot feature, you must leave enough room on the RAID disks for the snapshot repository. For example, if you have 3000 GiB of space on the RAID, you should not create a filesystem of more than 2400 GiB so that you can later create a repository of 600 GiB (20% of the size of the base filesystem). For information, see "XVM Snapshots" on page 5.

   You cannot use snapshot on CXFS or DMF filesystems or on high-availability (HA) clusters.

   ---

1 GiB, 1024 megabytes
2 GiB, 1024 megabytes

- Optional snapshot repository size. (This does not apply to HA clusters.) The size of the repository that you will need depends on several factors:
  - The size of the filesystem for which you are creating a snapshot. A repository that is approximately 10% of this size is a reasonable starting estimate.
  - The volatility of the data in the volume. The more of the data that changes, the more room you will need in the repository volume.
  - The snapshot frequency. (More frequent snapshots results in smaller individual snapshots.)

  Click **Next**.

6. The **Confirmation** screen summarizes the filesystem options you have selected. Click **Next** to confirm you choices and create the filesystem.
7. The **Create filesystem** screen displays a "please wait" message and transitional status during the filesystem creation process. Click **Next** after the operation is finished and the completion message displays.
8. The **Create repository** screen (if you have chosen to create a snapshot repository) displays a "please wait" message and transitional status during the filesystem creation process. Click **Next** after the operation is finished and the completion message displays.
9. The **Resource Group** screen (for HA clusters) lets you create a new HA resource group or select an existing resource group for the filesystem. To create a new resource group, specify the following:

IP address	Specifies the IP address that clients will connect to in order to access the filesystem.
Name	Specifies the hostname (do not use the fully qualified domain name) that is visible to CIFS clients for this resource group.
Priority	Specifies the priority with which the resource group will run on each node, in the range 0 through 10. The priority determines the processing order of dependencies and more importantly which resources will be left out if all dependencies cannot be satisfied. Use 10 for the node that should run the resource most often (the preferred node), and

lower numbers for other nodes. There must be at least two nodes running each resource.

10. The **NFS and CIFS** screen lets you configure the filesystem so that it can be exported with NFS or CIFS network protocols. (If you NFS export and/or CIFS share a CXFS filesystem, it will only be exported/shared from the CXFS metadata server, not from CXFS clients.) [3] For information, see "NFS Configuration" on page 47 and "CIFS Configuration" on page 51. Click **Next**.
11. The **Finished** screen indicates that the filesystem has been created. Click **Done**.

Growing Filesystems

Note: You cannot use Appliance Manager to grow a CXFS filesystem.

You can use a filesystem normally as you grow it. (You do not need to disable access or unmount it, or take any other special actions before growing the filesystem.)

To increase the size of an existing XFS filesystem, do the following:

1. Select the **Grow** option:

   Management
   > Resources
   > Storage
   > Filesystems
   > Grow

2. The **Filesystem** screen in the wizard lists the current filesystems along with their usage and size. Select the filesystem you want to grow and click **Next**.
3. The **Size** screen lets you enter the size in gigabytes[4] by which the filesystem should be grown and click **Next**.
4. The **Confirmation** screen displays the current size of the filesystem and the amount to grow the filesystem. Click **Next**.
5. The **Growing** screen displays a "please wait" message during the growing process. Click **Next** after the operation is finished and the completion message displays.

---

3 *Metadata* is information that describes a file, such as the file's name, size, location, and permissions.The *metadata server* is the node that coordinates updating of metadata on behalf of all nodes in a cluster.

4 GiB, 1024 megabytes

6. The **Finished** screen indicates that the larger filesystem is available. Select **Done**.

Destroying Filesystems

To delete a filesystem, do the following:

1. Select **Destroy**:

   Management
   > Resources
   > Storage
   > Filesystems
   > Destroy

   This screen displays a list of the existing filesystems.

2. Select a filesystem from the list. A message indicates that all data on the specified filesystem will be destroyed.

3. Confirm that you want to destroy the filesystem and select **Yes, destroy the filesystem**.

On completion, a `SUCCEEDED` message appears.

iSCSI Targets

Internet Small Computer Systems Interface (iSCSI) is a protocol that is used to transport SCSI commands across a TCP/IP network. This allows a system to access storage across a network just as if the system were accessing a local physical disk. To a client accessing the iSCSI storage, the storage appears as a disk drive would appear if the storage were local.

In an iSCSI network, the client accessing the storage is called the *initiator* and runs iSCSI Initiator software. The remote storage that the client accesses is called the *target*, which is what appears to the initiator as a disk drive.

A common application of an iSCSI network is to configure an Exchange Server as an iSCSI initiator that uses an iSCSI target as its mail store.

Figure 3-3 illustrates iSCSI storage. Each client (initiator) is configured to connect to a specific iSCSI target (an area allocated in the RAID iSCSI storage pool), and views this target as if it were a local disk. The lines in Figure 3-3 indicate data flow.

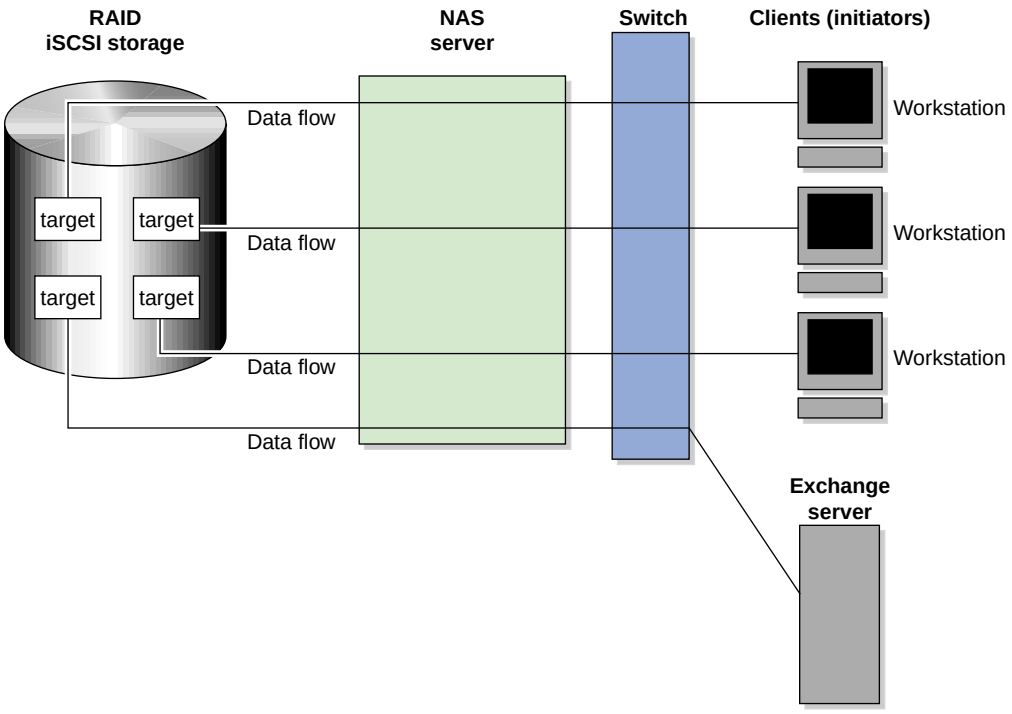

Figure 3-3 iSCSI Storage

You can use Appliance Manager to create iSCSI targets on the RAID storage. An iSCSI initiator will be able to connect to the system and access those targets, format them as NTFS, and use the targets as it would use a disk drive. After the target is formatted by the initiator, it will appear by default as a drive letter.

You cannot configure Appliance Manager itself as an initiator, and you cannot re-export iSCSI targets with NFS, CIFS, or CXFS. In addition, you cannot export existing filesystems that you have created with Appliance Manager as iSCSI targets; you can create filesystems and configure them to be exported by NFS , CIFS, or CXFS, but you must configure iSCSI targets separately on the RAID device.

Creating iSCSI Pool and Targets

You create iSCSI targets with a creation wizard, just as you create filesystems.

Perform the following steps to create an iSCSI target:

1. Select the **Create Target** option:

   Management
   > Resources
   > Storage
   > iSCSI
   > Create Target

2. If this is the first target, the system will display a message indicating that you must create the iSCSI storage pool before you can create a target.

   Note: Although you can grow this storage pool at a later time when you create additional targets, SGI recommends that you create a storage pool that is large enough to contain all of the targets that you will need. Creating the iSCSI storage pool can be a slow process, but once you have created the pool, creating the targets themselves is a fast process.

   If you have previously created iSCSI storage, you can grow the storage at this time; in this case, the screen displays how much storage you have available.

   To create or grow iSCSI storage, click **Next** and proceed to step 3 below. If you do not need to create or grow iSCSI storage, select **Skip this step** and proceed to step 8 below.

3. Appliance Manager searches for the RAID arrays on the system and displays them on the **Arrays** screen. Click **Next.**

4. The **Options** screen displays the iSCSI storage configuration options. For information, see "Creating Filesystems" on page 29.

5. In the **Size** screen, enter the size in gigabytes[5] for the iSCSI storage pool. Click **Next**.

6. The **Confirmation** screen summarizes the options you have selected. Click **Next** to confirm your choices and create the pool.

[5] GiB, 1024 megabytes

7. The **Creating** screen displays a "please wait" message during the target creation process. Click **Next** after the operation is finished and the completion message displays.

8. The **Target Name** screen lets you specify the target information. Enter the domain and optional identifier for the iSCSI name and the size of the target in the following fields:

Domain	Specifies an iSCSI qualified name (which is a unique name that starts with `iqn`), then a year and month, then an internet domain name in reverse order. A default name appears based on the current system configuration. If in doubt, leave this field as is.
Identifier	Specifies a string that will be used to uniquely identify the target. If you create only one target, this is optional. If you create more than one target, each must have a unique identifier. By default, a unique target identifier is provided for you.
Target Size (GiB)	Specifies the size, in gigabytes, of the target.

   Click **Next**.

9. The **Target Options** screen defines access to the target. You must specify at least one authentication option:

   Note: If more than one initiator were to write to the same target at the same time, there is a high risk of data loss. By using one or more authentication options, you ensure that only one client (initiator) can access an individual target at a time.

- Authentication:

  Initiator IP Address — Specifies the IP addresses of the initiators that will be allowed access to this target

- Challenge Handshake Authentication Protocol (CHAP) authentication, in which the initiator will supply the following information to the target:

  Target Username — Specifies the username that the initiator must supply to connect to the target using CHAP authentication. (This is not the username with which you logged in to Appliance Manager; it is specific to the iSCSI target you are defining.)

  Target CHAP Secret — Specifies the password that the initiator must supply to connect to the target using CHAP authentication. It must be in the range from 12 through 16 characters. (This is not the password with which you logged in to Appliance Manager; it is specific to the iSCSI target you are defining.)

  Re-enter Target CHAP Secret — Verifies the CHAP secret.

- Mutual CHAP authentication, in which the target will supply the following information to the initiator:

  Mutual Username — Specifies the target username for mutual CHAP authentication. With mutual CHAP authentication, after the initiator supplies a username, the target must supply a username and password back to the initiator. If you leave the **Mutual Username** field blank, it defaults to the target username.

  The mutual name is usually ignored by initiators, which only care about the mutual secret. When the client connects to a target, the iSCSI initiator software verifies that the mutual secret specified in Appliance Manager matches the secret specified in the initiator.

  Mutual CHAP Secret — Specifies the mutual CHAP secret.

Re-enter Mutual CHAP Secret	Verifies the mutual CHAP secret.

You must enter the CHAP username and secret specified on this screen in the iSCSI initiator software on the client in order for the initiator to be able to authenticate with and connect to the target. For a Windows client, this is the username and secret you enter in Microsoft's iSCSI Initiator program.

10. The **Confirmation** screen summarizes the target options you have selected. Click **Next** to confirm your choices and create the iSCSI target.
11. The **Finished** screen indicates that the iSCSI target has been created. Select **Done**.

After you have created iSCSI targets, select the following to see what initiators are connected to what targets:

Monitoring
> Clients
> iSCSI

The iSCSI Initiator

Appliance Manager lets you configure iSCSI targets for use by an iSCSI initiator, such as Microsoft's free iSCSI initiator or the iSCSI initiator included with various Linux and UNIX distributions.

After you have created an iSCSI target, you must configure the initiator on the client system that will connect to the target. You must specify the following:

- Hostname of the storage server
- Target identifier
- Any CHAP authentication details you configured when creating the target (for specific instructions, see the documentation supplied with your iSCSI initiator)

After the iSCSI initiator has connected to the target, the target will appear as a disk drive on the client system and can then be formatted using the tools supplied with the client operating system.

The following is an example of configuring a Windows client (it assumes that you have already created a target or targets):

1. Download the iSCSI Initiator from Microsoft's web site (http://www.microsoft.com/) and install it on the Windows client.

2. Open the **iSCSI Initiator Control Panel** applet.
3. Add the storage server to the list of **Target Portals**.
4. Select the iSCSI target to connect to from the **Targets** list and click **Log On**.
5. Specify CHAP authentication details in the **Advanced** settings.
6. Use the following tool to partition and format the target and assign a drive letter:

   Start Menu
   > Administrative Tools
   > Computer Management
   > Disk Management

Note: If more than one initiator were to write to the same target at the same time, there is a high risk of data loss. By using one or more authentication options, you ensure that only one client (initiator) can access an individual target at a time.

Miscellaneous iSCSI Management

The **iSCSI** menu also provides the following management options:

List Targets	Lists the existing iSCSI targets.
Modify Target	Modifies the authentication settings you defined on the **Target Options** screen when you created an iSCSI target.
Destroy Target	Destroys an existing iSCSI target.
Destroy Storage Pool	Destroys the iSCSI storage pool on the RAID device and all existing targets.
Stop/Start	Stops or starts the iSCSI service. If you are backing up the system, taking iSCSI services offline ensures that the data is in a consistent state.

Scheduling Snapshots

To schedule how often the system will create a snapshot of a filesystem, do the following:

1. Select the **Schedule Snapshots** menu:

   Management
   > Resources
   > Storage
   > Snapshots
   > Schedule Snapshots

2. Select the filesystem for which you want to schedule snapshots.

3. Specify the following options:

   Scheduled?

   Specifies that a snapshot will take place for the filesystem.

   Scheduled Snapshot Times

   Specifies the hours at which a snapshot should take place. You can select multiple boxes.

   Custom Time Specification

   Specifies the times and frequency that a snapshot should take place. You can specify this value using one of the following forms:

   - Spelled out using the following syntax:

     `every` *XX* `minute/s|hour/s|day/s [from` *XX*`:`*XX* `to` *XX*`:`*XX*`]`

     The specific times listed within brackets are optional. For example:

     ```
     every 1 hour
     every 20 minutes from 8:00 to 22:00
     every 4 days
     ```

   - Exact times. For example:

     ```
     12:45 23:00 9:30
     ```

   - The word `never`, which disables snapshots.

   Maximum number of snapshots

   Specifies the maximum number of snapshots that will be retained in the repository before the oldest snapshot is deleted when a

new snapshot is taken. By default, the system will retain 32 snapshots. The maximum number is 256. SGI recommends that you use the default.

Repository

Specifies the repository name.

Click **Schedule snapshots** to apply your settings.

4. Verify that you want to update the snapshot schedule by clicking **Yes**. (To return to the previous screen, click **No**.)

Note: The system will delete the oldest snapshot if it determines that repository space is running low.

Snapshots are made available in the `/SNAPSHOTS` directory of the base filesystem. They are named according to the date and time at which they are taken. For example, a snapshot might be named as follows:

```
/mnt/data/SNAPSHOTS/2006_07_30_113557_Sun
```

DMF Configuration

The **DMF Resources** screens let you do the following:

- Stop and start DMF and tape daemons
- Enable/disable tape drives
- Import/export volumes from an OpenVault library (but not the Tape Migration Facility, TMF)
- Empty a lost or damaged DMF tape
- Alter DMF configuration parameters
- Audit the databases

Tape Volume and Drive Screens

Appliance Manager supports most common DMF configurations. There are some limitations to this support, however. Specifically, the following are assumed to be true:

- The OpenVault mounting service is preferred. Ejection and injection of tape volumes from and into a tape library is disabled if TMF is in use, but the other functions are supported for both OpenVault and TMF.
- All tapes that are ejected and injected using Appliance Manager are for use by a DMF volume group or allocation group. Other tapes may reside in the library, but they cannot be managed by Appliance Manager.
- Each DMF library server manages only a single tape library. Appliance Manager refers to the library by using the name of the library server. Use of more than one tape library is not supported.
- Each DMF drive group is associated with an OpenVault drive group or a TMF device group of the same name.

Emptying a Lost or Damaged Tape Volume

The **Empty Tape Volume** screen uses the `herr`, `hvfy`, and `hlock` DMF database flags to record the progress of the emptying procedure. If you use the `dmvoladm`(8) command to inspect the database entry for a tape while it is being emptied, you may see unexpected settings of these flags. Appliance Manager's use of these flags does not interfere with DMF's.

Appliance Manager does not make any use of the VOL database flags reserved for site use, although the **Import** and **Export** screens do allow you to manipulate them.

The **Empty Tape Volume** screen's **Empty Volume**, **Remove Volume** and **Reuse Volume** options cannot remove soft-deleted files from a tape volume, unlike the **Merge Volume** button. You must wait until they have been hard-deleted by the scheduled `run_hard_deletes.sh` task or by the `dmhdelete`(8) command.

Also, these three buttons may need access to the output file from the previous run of the scheduled `run_filesystem_scan.sh` task or the `dmscanfs`(8) command. If it cannot be found or is older than the files remaining on the tape, some files may be misreported in the **Alerts** screen as soft-deleted and remain on the tape as described above. Trying again after the next run of `run_filesystem_scan.sh` is likely to succeed in this case.

For more information, see the `dmemptytape`(8) man page for more information.[6]

DMF Configuration Screens

You can use the **DMF Configuration** screens to inspect and modify various DMF parameters:

Management
> Resources
> DMF
> Configuration

For initial configuration of DMF, use the **Edit** link:

Management
> Resources
> DMF
> Configuration
> Edit

This link allows you to directly modify the configuration file or import another configuration file.

Caution: You must ensure that the changes you make are safe. For more information, see the `dmf.conf`(5) man page and the *DMF Administrator's Guide for SGI InfiniteStorage*.)[7]

The **Check** link allows you to perform syntax and sanity checks on the current configuration of DMF:

Management
> Resources
> DMF
> Configuration
> Check

6 You can access man pages from the SGI Technical Publications Library at http://docs.sgi.com.
7 You can access man pages and books from the SGI Technical Publications Library at http://docs.sgi.com.

SGI recommends that you use the **Check** link after making any modification to ensure that the changes are safe.

The **Global** link displays parameters for all of DMF:

Management
> Resources
> DMF
> Configuration
> Global

If you click **Switch to Expert Mode** on the **Global** page, Appliance Manager presents more parameters. You should use expert mode with care. To return to normal mode, click **Switch to Normal Mode**. Excluded from both modes are parameters that are:

- Deprecated
- Specific to the Resource Scheduler or Resource Watcher stanzas

To work around these restrictions, the **Edit** link allows you to edit the DMF configuration file directly.

The other links provide quick access to commonly altered parameters of already-configured features. You should make changes with care. Parameters that can be dangerous to change are displayed but may not be altered; tis includes those parameters that control the search order of volume groups and media-specific processes (MSPs) when recalling files.

Note: On the **DMF Configuration** screens, disk sizes use multipliers that are powers of 1000, such as kB, MB and GB. This is for consistency with the DMF documentation and log files. However, the rest of Appliance Manager, including the DMF **Monitoring** screens, use multipliers that are powers of 1024, such as kiB, MiB and GiB.

User and Group Configuration

Appliance Manager lets you configure local users, local groups, and user and group quotas.

Local Users and Groups

Appliance Manager can create and add local user and group accounts to access the storage server locally. This is a local database only; these users and groups do not interact with the users and groups provided by the name server. If you search the site directory and do not find the user or group data you are looking for, the system searches this local database. The local user accounts will be used for authentication for CIFS shares if you are not using LDAP or Active Directory authentication.

Caution: If you create a local user and subsequently add that user in the sitewide directory, access problems may result. For example, if you create local user `Fred` with a UID of `26`, `Fred` will be able to create local files. But if you subsequently add a user `Fred` on a sitewide name services directory with a different UID, user `Fred` will be unable to access those local files because the system will use the sitewide name and UID first.

If you are using LDAP or Active Directory as a name service client, a user must be present in LDAP or Active Directory and you will not be able to authenticate local users and groups. In this case, adding local users and groups may be useful for ID mapping, but authentication does not use the local password files.

When you select the **Import** option for either **Local Users** or **Local Groups**, you can choose among the following actions:

- Add the new users and groups. If there is an existing user or group with one of the names you are adding, keep the existing user or group.
- Add the new users. If there is an existing user or group with one of the names you are adding, replace the existing user or group with the new user or group.
- Replace all current unrestricted users or groups with the new users or groups.

Accounts with a UID or GID of less than `1000` are considered restricted and are not imported or replaced.

If you use a *shadow file*, which is a file that is protected from all access by non-`root` users and stores the encrypted passwords, then you can use the **Import Users** screen to import this file as well as the password file itself.

Quotas

You can use Appliance Manager to set and modify user and group quotas for filesystems if you have mounted the filesystem with quotas enabled.

You cannot enable quotas on a filesystem directory through Appliance Manager. You must use the `ssh` command to log in to the system, edit the `fstab` file, and remount the filesystem.

NFS Configuration

To configure filesystems so that they are available for network clients by means of the NFS network protocol, select the following:

Management
> Services
> NFS

This screen displays a link for **Global Options** and all of the filesystems that have been created with Appliance Manager, whether or not they have been enabled for export. To specify NFSv4 options, select **Global Options**. To change the export options, select an individual filesystem name or **All Filesystems**. See:

- "Global Options" on page 47
- "Export Options" on page 49
- "NFS-RDMA Client Packages" on page 51

Global Options

The **Global Options** screen lets you specify the following:

Enable NFSv4	Specifies whether NFSv4 is enabled (checked) or not. If enabled, an NFS exported filesystem will be accessible via both NFSv3 and NFSv4. The following fields are only relevant if you have enabled NFSv4.
NFS serving domain	Specifies the serving domain. If NFSv4 is enabled, the mapping of user/group IDs between the client and server requires both to belong to the same NFS serving domain.

Enable Kerberos	Specifies whether Kerberos is enabled (checked) or not. Enabling Kerberos forces encrypted authentication between the NFS client and server. Furthermore, the NFS exported filesystems will only be accessible to a Kerberos enabled client via NFSv4. The following fields are only relevant if you have enabled Kerberos.
	Note: Appliance Manager supports Kerberos 5. You must use a mechanism to synchronize the time between all systems.
Realm	Specifies the Kerberos realm in which the NFSv4 server operates.
Domain	Specifies the DNS domain name that corresponds to the realm.
KDC	Specifies the key distribution center (KDC). In most cases, the KDC will be the same system as the Kerberos admin server. However, if the admin server in your Kerberos environment is not used for granting tickets, then set the KDC to the system that grants tickets.
Admin Server	Specifies the server containing the master copy of the realm database.
Keep Existing Keytab	Select this radio button to keep the existing keytab without changes
Update Keytab	Select this radio button to change the principal user and password for the existing keytab
Principal	Specifies a user that belongs to the Kerberos server with sufficient privileges to generate a keytab for the NFS server.
Password	Specifies the principal's password.
Upload Keytab	Copies the selected file to `/etc/krb5.keytab` on the NFS server. Click **Browse** to see a list of available files.

Verify Keytab	Specifies that the keytab should be verified. This is not supported by Active Directory.

Export Options

You can choose to export or not export a filesystem by clicking the check box. When you enable a filesystem for export, you can do one of the following:

- "Use Export Options" on page 49
- "Use a Custom Definition" on page 50

After specifying the configuration parameters, click **Apply changes** at the bottom of the screen.

Use Export Options

If you select **Use export options**, you must specify the following:

Read-only	Specifies that the client has access to the filesystem but cannot modify files or create new files.
Asynchronous writes	Specifies whether or not to use asynchronous writes. Data that is written by the client can be buffered on the server before it is written to disk. This allows the client to continue to do other work as the server continues to write the data to the disk. By default, writes are performed synchronously, which ensures that activity on the client is suspended when a write occurs until all outstanding data has been safely stored onto stable storage.
Allow access from unprivileged ports	Allows access for Mac OS X clients or other NFS clients that initiate mounts from port numbers greater than 1024. If there are no such clients on your network, leave this option unchecked.
All hosts	Allows connections from anywhere on a network.
Restrict to Kerberos aware clients	Allows connections only to those systems that are Kerberos aware (if Kerberos is enabled in "Global Options" on page 47)

Local subnet

Allows connections from the indicated subnet. You can select any subnet from those that have been defined for the network interfaces.

Restrict to hosts

Specifies the set of hosts that are permitted to access the NFS filesystem. You can specify the hosts by hostname or IP address; separate values with a space or tab. For example, you could restrict access to only the hosts on a Class C subnet by specifying something like the following:

```
150.203.5
```

To allow hosts of IP address `150.203.5.*` and `myhost.mynet.edu.au`, specify the following:

```
150.203.5. myhost.mynet.edu.au
```

You can also specify hosts by network/subnet mask pairs and by netgroup names if the system supports netgroups.

To allow hosts that match the network/subnet mask of `150.203.15.0/255.255.255.0`, you would specify the following:

```
50.203.15.0/255.255.255.0
```

To allow two hosts, `hostA` and `hostB`, specify the following:

```
hostA hostB
```

Note: Access still requires suitable user-level passwords. The localhost address `127.0.0.1` will always be allowed.

Use a Custom Definition

If you select **Use custom definition**, you can enter any NFS export options that are supported in the Linux `/etc/exports` file.

For example, the following entry gives `192.168.10.1` read-write access, but read-only access to all other IP addresses:

```
192.168.10.1(rw) *(ro)
```

Note: There cannot be a space between the IP address and the export option.

For information on the `/etc/exports` file, see the `exports`(5) man page. [8]

NFS-RDMA Client Packages

You must download and install all of the NFS-RDMA packages for the architecture of your client machine. Select:

Management
> Services
> NFS
> NFS-RDMA Client Packages

The resulting screen shows the packages available for download. Select the appropriate OS version for your site.

CIFS Configuration

To configure filesystems so that they are available for network clients by means of the CIFS network protocol, select the following:

Management
> Services
> CIFS

All of the filesystems created with Appliance Manager are displayed on this screen, whether or not they have been enabled for sharing. To share a file, select it and click the **Shared?** box.

Specify the following **Share Options**:

Share name — Specifies the name under which the filesystem will appear to a Windows client, as displayed in its Network Neighborhood.

[8] You can access man pages from the SGI Technical Publications Library at http://docs.sgi.com.

Comment — Specifies an arbitrary string to describe the file.

Read-only — Specifies that the client has access to the filesystem but cannot modify files or create new files.

Allow guest users — Specifies that users can gain access to the CIFS filesystem without authenticating. Uncheck this option to allow connections only to valid users.

The CIFS protocol requires a password for authentication. If you are configured as an Active Directory client, then the authentication is distributed. See "Active Directory" on page 58.

Always synchronize writes — Ensures that activity on the client is suspended when a write occurs until all outstanding data has been safely stored onto stable storage. If you do not check this box, data that is written by the client can be buffered on the server before it is written to disk. This allows the client to continue to do other work as the server continues to write the data to the disk. This is the faster write option and is recommended.

Allow symbolic linking outside of the share — Specifies that symbolic links made by NFS users that point outside of the Samba share will be followed.

Caution: This feature is a performance/security tradeoff that is only interesting for sites running both CIFS and NFS from the same filesystem. Allowing linking could be a security risk if, for example, an NFS user created a symbolic link to `/etc/passwd`. However, unchecking the box will cause a decrease in performance.

All hosts — Allows connections from anywhere on a network.

Local subnets — Allows connections from the indicated subnet. You can select one subnet in this field and you must choose it from the available interfaces as set in the **Network Interfaces** screen.

Restrict to hosts

Specifies the set of hosts that are permitted to access the CIFS filesystem. You can specify the hosts by name or IP number; separate values by a space or tab. For example, you could restrict access to only the hosts on a Class C subnet by specifying something like the following:

```
150.203.5
```

To allow hosts of IP address `150.203.5.*` and `myhost.mynet.edu.au`, specify the following:

```
150.203.5. myhost.mynet.edu.au
```

You can also specify hosts by network/subnet mask pairs and by netgroup names if the system supports netgroups. You can use the `EXCEPT` keyword to limit a wildcard list.

For example, to allow all IP address in `150.203.*.*` except one address (`150.203.6.66`), you would specify the following:

```
150.203. EXCEPT 150.203.6.66
```

To allow hosts that match the network/subnet mask of `150.203.15.0/255.255.255.0`, you would specify the following:

```
50.203.15.0/255.255.255.0
```

To allow two hosts, `hostA` and `hostB`, specify the following:

```
hostA, hostB
```

Note: Access still requires suitable user-level passwords. The localhost address `127.0.0.1` will always be allowed.

After specifying the configuration parameters, select **Apply changes** at the bottom of the screen.

CXFS Configuration

To manage a CXFS cluster, select the following:

Management
> Services
> CXFS

This lets you choose the following options:

Cluster Nodes	Adds, enables, disables, and deletes client-only nodes and displays node status. To add a client-only node, you must specify the node's hostname, CXFS private network IP address, and operating system: AIX IRIX Linux[9] Mac OS X Solaris Windows For the specific operating system release levels supported, see the CXFS release notes. When you add a new node, it is automatically enabled and able to mount all CXFS filesystems.
Switches	Displays Fibre Channel switches. To fence/unfence ports on a switch, select the switch's IP address then select the ports to fence/unfence.
Stop/Start	Displays the status of CXFS cluster daemon and lets you start, restart, or stop all of the CXFS daemons.
Client Packages	Provides access to CXFS client packages for each client platform, which may be downloaded to the clients via Appliance Manager.

To create a CXFS filesystem, see "Creating Filesystems" on page 29.

9 Red Hat Enterprise Linux, SUSE Linux Enterprise Server 9 or 10, SGI ProPack running SLES 10

NDMP Configuration

The storage server administered by Appliance Manager acts as a network data management protocol (NDMP) server; that is, it processes requests sent to it from a data migration application (DMA) in order to transfer data to/from a remote NDMP tape/data server.

In order to perform backups of the storage server using NDMP, you will need a DMA (such as Legato Networker) and a separate NDMP tape server.

The **NDMP** configuration screen in Appliance Manager allows you to configure your system such that it will communicate with your DMA and your NDMP tape server. For information on initiating backup/restore operations, refer to the documentation that came with your DMA software.

To administer NDMP for backups, select the following

Management
> Services
> NDMP

The **NDMP** screen lets you configure the following parameters:

Protocol	Select the version of NDMP you want to use. (Protocol version 4 is the default. Protocol version 3 is provided for backward compatibility. If in doubt, use version 4.)
New Sessions	Select whether new NDMP sessions are allowed or disallowed lets you stop backup clients from connecting to the NDMP server or allow the connection. With **Allowed**, authorized backup clients may connect and initiate backup sessions. With **Disallowed**, no new client sessions may be established (existing sessions will not be affected).
Interfaces	Select the individual interfaces where the `ndmp` server will listen for connections. To use all interfaces, leave all interfaces unselected.
Authorized Clients	Specify the IP address of those clients that are authorized to access NDMP. If you want all clients to have access, leave this field blank.
Username	Enter the username that NDMP clients will use to establish sessions with the NDMP server.

New Password	Set the password for the username
Confirm New Password	Reenter the password for the username

Note: When backing up a full filesystem, the quota and `mkfs` information will be backed up the following `tar` file in the `root` directory of the filesystem:

`.volume_info_`*date*

For example, the following file was backed up on August 6th 2007 at 2:45 PM:

`.volume_info_200708061445`

The information will be placed in the `root` directory of the filesystem if it is restored. However, the quotas and `mkfs` options will not be applied on restoration; the administrator may choose to apply them if desired.

Global Configuration

The following sections describe the following aspects of system administration that you can perform with Appliance Manager:

System Name

Use the **System Name** screen to set the following system components:

System name	Specifies the fully qualified domain name (FQDN) for this storage server. The default hostname is `sgiserver`. You cannot change the default hostname for a SAN server.

Management IP address	Specifies the IP address of the management interface
Workgroup	Specifies the NetBIOS workgroup to which the machine should belong. The default is `WORKGROUP`. If you are not using CIFS, you can ignore this setting.
Email gateway	Specifies the gateway server where email should be directed. The email gateway is also known as the *smart host*. Email sent from this system should not be sent to the storage server except for system warnings. If you do not specify an email gateway in this field, email sent on a system will collect locally in `/var/spool/mail`.
Administrator email address	Specifies the address where system email, such as system alerts, will be sent. If you do not specify an address, the messages go to `root`. This field can be any valid email address.
	Note: The administrator email address is used only for forwarding system notices. It is not sent to SGI or used for any other purposes.
Default network gateway	Specifies the IP address of the router that this system should use to communicate with machines that are outside of its subnet.
***HA_Node* IP address**	Specifies the IP address of an HA node
***HA _Node* Subnet mask**	Specifies the subnet mask of an HA node
Use DHCP	Specifies whether or not to use dynamic host configuration protocol (DHCP) .

You can also use the **Network Interfaces** screen for `eth0` to configure or modify the management interface. For information on these options, see "Ethernet Network Interfaces" on page 22.

Name Service Client

The **Name Service Client** screen lets you specify a *name service* (or *directory service*) for the system. A name service is the application that manages the information

associated with the network users. For example, it maps user names with user IDs and group names with group IDs. It allows for centralized administration of these management tasks.

You can specify whether you are using local files (if you have no sitewide protocol and names and IDs are kept locally on server), Active Directory services, lightweight directory access protocol (LDAP), the sitewide network information service (NIS).

Note: When specifying servers on the **Name Service Client** screen, you must use IP addresses rather than hostnames, because the system may require a name service client to determine the IP address from the hostname.

Local Files Only

The **Local Files Only** selection specifies that an external name server will not be used. All user and group name to ID mapping will be done using local users and groups. See "Local Users and Groups" on page 46.

Active Directory

Active Directory is a directory service that implements LDAP in a Windows environment. It provides a hierarchical structure for organizing access to data. CIFS authentication will automatically use the Active Directory service.

Note: The **Active Directory** section is disabled if there are no Active Directory DNS servers specified. See "DNS and Hostnames" on page 62.

The following Active Directory components appear on the **Name Service Client** screen:

Active Directory domain
: Specifies the full domain name of the Active Directory.

  Note: If you later change the server hostname on which Appliance Manager runs, you must rejoin the Active Directory domain because the Active Directory Security Identifier (SID) will be changed.

Domain Controller
: Specifies a domain controller.

Administrative user
: Specifies the user with administrator privileges.

Password

Specifies the password for the administrator user. For security reasons, the Active Directory password cannot contain the following characters:

```
;*&|'<>?[]
```

Re-enter password

Verifies the password for the administrator user.

UID/GID Mapping

Lets you manage UNIX user ID (UID) and group ID (GID) mapping on the Active Directory server, using one of the following:

- **RFC 2307 (Microsoft Windows Server 2003 R2)**. In order for this to function correctly:
  - The Active Directory domain controller must be running Microsoft Windows Server 2003 R2.
  - The Identity Management for UNIX service must be installed on the domain controller.
  - You must use the **UNIX Attributes** tab in Active Directory user management to set up UIDs and GIDs for all users requiring access to this system.
- **Microsoft Windows Services For UNIX**. In order for this to function correctly:
  - Microsoft Windows Services for UNIX must be installed on the Active Directory domain controller.
  - You must use the **UNIX Attributes** tab in Active Directory user management to set up UIDs and GIDs for all users requiring access to this system.
- **Automatic assignment in range 10000-20000**. In this mode, UIDs and GIDs in the range 10000 through 20000 will be automatically assigned to Active Directory users on a first-come, first-served basis. (This choice is not recommended because it may affect interoperability.)

Note: For best interoperability, SGI strongly recommends that you choose either **RFC 2307 (Microsoft Windows Server 2003 R2)** or **Microsoft Windows Services For UNIX**, as appropriate for your environment.

Click **Apply changes**.You will then be presented with a confirmation screen that allows you to verify whether or not you want to commit the changes.

Caution: Depending on your environment, making changes to the UID/GID mapping may result in ownership changes of user files.

LDAP

Lightweight directory access protocol (LDAP) is a networking protocol that organizes access to data in a directory tree structure. Each entry in the tree has a unique identifier called the *distinguished name.*

The default LDAP server IP address is the local host. You will probably need to specify a different IP address.

Fields:

LDAP server	Specifies the IP address of the LDAP server.
Base	Specifies the distinguished name of the base of the subtree you will be searching.
Root binddn	Specifies the distinguished name of the user to whom you are assigning `root` privileges for administration. This is expressed as a node in the directory tree that refers to a user account.
Password	Specifies the password that will be required to authenticate against the LDAP server. For security reasons, the LDAP password cannot contain the following characters: `;*&|'<>?[]`
Re-enter password	Verifies the password that will be required to authenticate against the LDAP server.

To use LDAP for CIFS authentication, you must configure the LDAP server to use the RFC2307/bis or NIS schema to supply POSIX account information. In addition, you must add a Samba schema to the LDAP database. These schemas specify how the user and group data is organized in the database. The database must be organized using these particular schemas so that the CIFS authentication mechanism is able to extract the data it needs.

For a description of how to add the Samba schema to a Fedora Directory Server, see:

http://directory.fedora.redhat.com/wiki/Howto:Samba

For a description of how to add the samba schema to an OpenLDAP Server, see:

http://www.samba.org/samba/docs/man/Samba-HOWTO-Collection/passdb.html#id327194

The following website provides another description of an OpenLDAP configuration:

http://www.unav.es/cti/ldap-smb/ldap-smb-3-howto.html

For other LDAP servers (such as the Sun Directory Server, Novell's eDirectory, and IBM's Tivoli Directory Server) the above information may be useful; however, please refer to the relevant documentation for your server product more information.

NIS

Network information service (NIS) is a network lookup service that provides a centralized database of information about the network to systems participating in the service. The NIS database is fully replicated on selected systems and can be queried by participating systems on an as-needed basis. Maintenance of the database is performed on a central system.

Note: NIS cannot be used for CIFS authentication.

Specify the following:

Domain name — Specifies the NIS domain name for this system.

NIS server — Specifies the IP address of the NIS server. If the NIS server is on the same subnet as Appliance Manager, Appliance Manager finds the NIS server IP address and provides it as a default. If you are not on the same subnet, you must enter the address in this field.

DNS and Hostnames

You can use the **DNS and Hostnames** screen to specify how to map hostnames to IP addresses for the system. Click **Edit local hosts table** to access the **Hosts** screen, where you can edit the `/etc/hosts` file that contains local mappings or import the contents of a file you specify. For information on the `/etc/hosts` file, see the `hosts`(5) man page. [10]

You can also specify the DNS servers to map hostnames to IP addresses and to resolve hostnames that are incomplete.

Domain Search — Specifies the domain name or names of the DNS servers that the system uses to provide hostname-to-IP-address translation.

If you have multiple domains, list them in the order you want to use for lookup. This is important in cases where have two machines with the same name, each on a different domain, to establish the lookup priority.

10 You can access man pages from the SGI Technical Publications Library at http://docs.sgi.com.

Nameserver *N*	You can specify up to three IP addresses for the DNS name servers to use. If an address you specify is down, the system will use the next one

Note: If you specify one or more DNS servers, SGI InfiniteStorage Appliance Manager adds `mdns off` to the `/etc/host.conf` file in order to force resolution of `.local` names to go to the DNS server rather than using multicast DNS.

If you later remove the DNS servers, the value of `mdns off` in `/etc/host.conf` remains the same.

If you manually edit `/etc/host.conf` to force `mdns on`, Appliance Manager will not change this setting provided that you do not specify DNS servers via "DNS and Hostnames" on page 62.

Time and Date

Use the **Time and Date** screen to set the following.

Timezone	Sets the local time zone for Appliance Manager. You can choose a time zone from a drop-down list of options, or you can set a custom time zone. For example, the following specifies what that name of the time zone is for both standard and daylight savings periods, and when the change-over is from daylight to standard and back again (going from standard to daylight on the 10th month and the 5th Sunday, and back again on the 4th month and the first Sunday): `AEST-10AEDT,M10.5.0,M4.1.0` For more information about custom timezone format, see the `tzfile` man page.[11]
NTP Time Synchronisation	Enables automatic time synchronization with Network Time Protocol (NTP). The NTP protocol is used to synchronize clocks on computer systems over a network. Select **Apply NTP changes** keep the system's time in synchronize with an NTP server or **Set time**

11 You can access man pages from the SGI Technical Publications Library at http://docs.sgi.com.

	from NTP server to go off and synchronize it now once only. If the server has Internet access, see the following website for information about using the public NTP timeserver: http://www.pool.ntp.org/
Set Time from NTP Server	Sets the system date and time directly instead of using NTP time synchronization.

Licenses

The **Licenses** screen provides a text box in which you can type in or paste licenses obtained from SGI. Some licenses, such as the license for XVM snapshot, will not take affect until you reboot the system. For an HA cluster, you can add or delete licenses for one node at a time.

Administrator Password

The **Administrator Password** screen changes the Appliance Manager *administration password*, which is the password required to perform server configuration and management. This password is not required to view the Appliance Manager monitoring screens.

Operations

The following sections describe other operations you can perform with Appliance Manager:

- "Save/Restore Configuration" on page 65
- "Support Data" on page 65
- "Shutdown" on page 65

Save/Restore Configuration

The **Save/Restore Configuration** screen screen lets you save the current Appliance Manager configuration or restore a previously saved version. The configuration information saved includes how the interfaces are configured and what files should be mounted. You may find this useful if you have made an error in the present configuration and you wish to return to a previously configured state.

Caution: This procedure does not provide a system backup and specifically does not save or restore user data; it provides a snapshot record of the configuration.

This screen lists previously saved configurations, labeled by date. After restoring a configuration, you should restart the system.

Support Data

If there is a problem with the system, SGI Call Center Support may request support data in order to find and resolve the problem. The **Gather Support Data** screen lets you generate an archive containing copies of the storage server's software and hardware configuration and log files.

To collect the data, select **Yes, gather information**. This process can take more than 30 seconds on large RAID configurations and requires at least 200 MB of free space in `/tmp`.

Shutdown

From the **Shutdown** screen, you can specify to reboot or shut down the system in a specified number of seconds.

Performance Monitoring

SGI InfiniteStorage Appliance Manager provides current and historical views of the state and the performance of a storage server. This includes CPU usage, disk and network throughput, and many other metrics. It also allows you to view connected clients and determine how each of these contribute to the current workload.

This chapter does not describe all of the details of each Appliance Manager monitoring screen, because most screens are quite straightforward. Instead, it attempts to explain why the displayed information matters and how it can be sensibly interpreted.

This chapter discusses the following:

Figure 4-1 shows the top-level **Monitoring** screen.

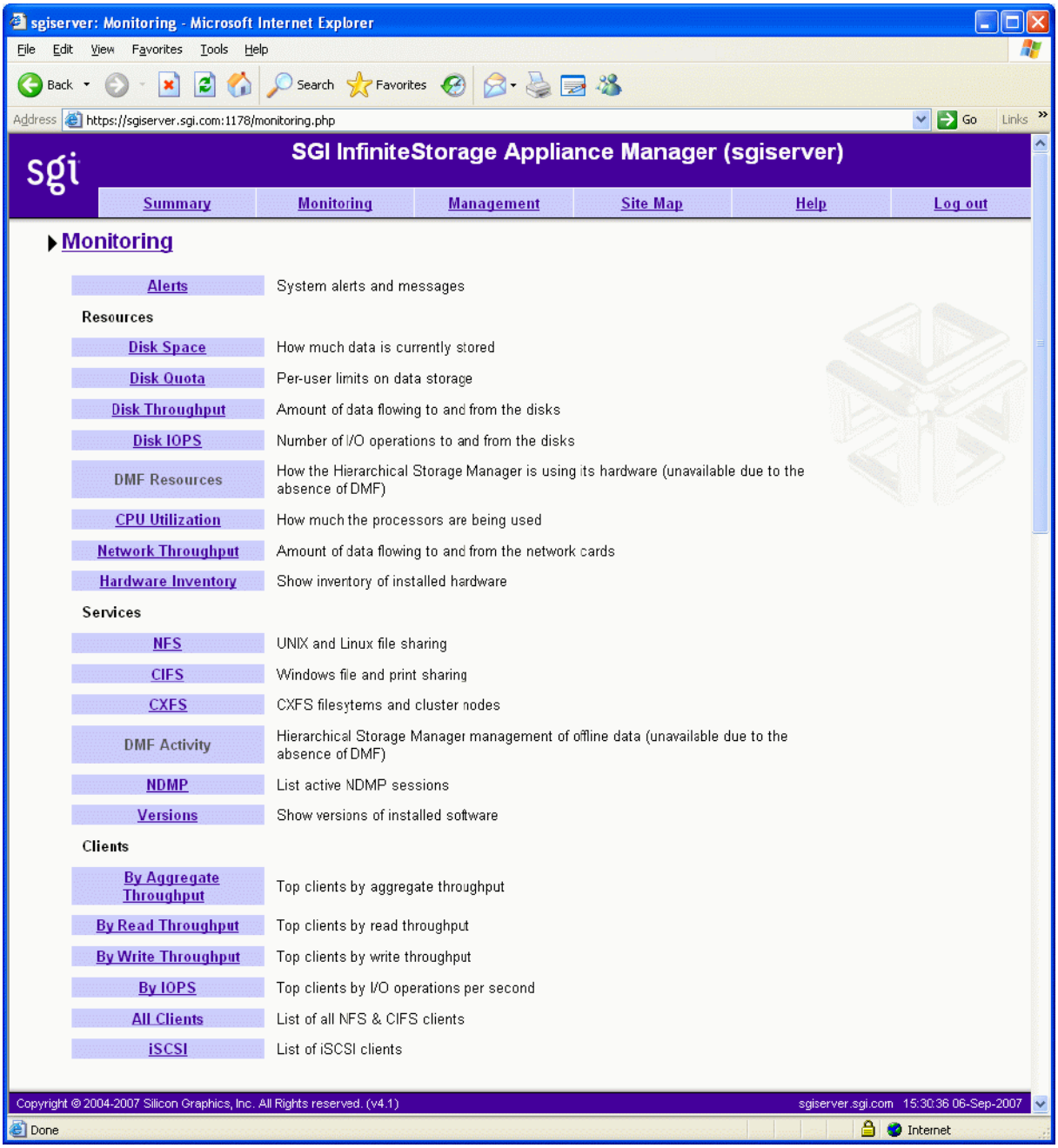

Figure 4-1 Monitoring Screen

Metrics Collected

The information provided by Appliance Manager can be roughly broken down into "who" and "how much". Appliance Manager continuously gathers performance metrics and stores them in archives in `/var/lib/appman/archives`. Each month, a data reduction process is performed on the metric gathered for the month. This reduces the size of the archives while retaining a consistent amount of information.

Although the size of metric archives has a bounded maximum, this can still be quite large depending on the configuration of the server and how many clients access it. For example, a server with a large number of filesystems could generate up to 100 Mbytes of archives per day. You should initially allow around 2 Gbytes of space for archive storage and monitor the actual usage for the first few weeks of operation.

Note: Appliance Manager uses the International Electrotechnical Commission's International Standard names and symbols for binary multiples of units. In particular, this means that 1 MiB/s is 2^{20} = 1048576 Bytes per second. For more information on this standard, see the National Institute of Standards & Technology information about prefixes for binary multiples at:

http://physics.nist.gov/cuu/Units/binary.html

Appliance Manager distinguishes between *current* and *historic* time. Current metrics are either drawn live from the server or are taken from the last few minutes of the metric archives. Historic metrics are taken exclusively from the metric archives. Appliance Manager is able to display this historical information for three time periods:

- Last hour
- Last day (the previous 24 hours)
- Last month (the previous 30 days)

Within bar graphs, Appliance Manager uses color-coding to display the direction of data flow:

- Red represents write and receive data flow
- Blue represents read and send data flow

Figure 4-2 describes how Appliance Manager color-codes the direction of data flow graphs. For an example of the result in a graph, see Figure 4-3 on page 72.

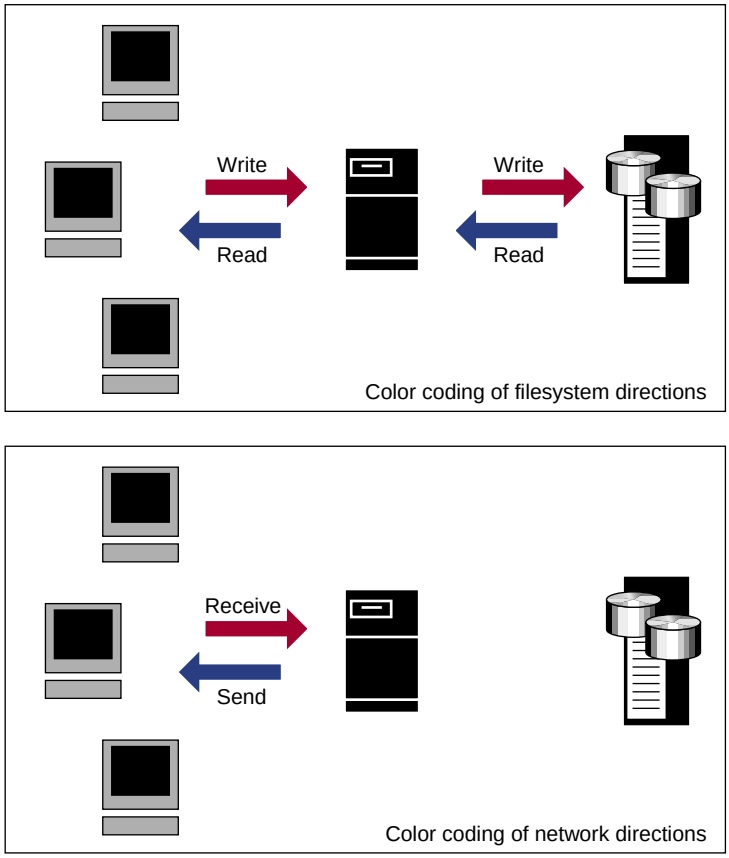

Figure 4-2 Color-Coding the Direction of Data Flow

System Summary

Appliance Manager provides a **Summary** menu selection at the top of the screen. This screen displays the following on a per-node basis:

- CXFS filesystem and node status (For details, see "CXFS" on page 85):
  - If all CXFS filesystems are stable (or if no filesystems exist), the **Filesystems** indicator will be green, and it will say **Stable**

  - If all cluster nodes are stable, the **Nodes** indicator will be green it will say **Stable**
  - If any of the filesystems or nodes are inactive or in a transient state (such as mounting filesystems), the indicators will be red and appropriate status text will be displayed

- CPU utilization
- Disk space
- Disk throughput
- Network throughput
- InfiniBand throughput (if installed)
- The number of NFS, CIFS, and iSCSI clients (if iSCSI targets have been created)
- System uptime
- Number of users
- Load average

You can drill down to more detailed status by clicking the headings to the left of the graphs.

Click **History** to view the historical status of a parameter.

The screen displays ticks along the status bars labeled **d** (day) and **h** (hour). These represent the average value over the past day or hour, rather than the immediate value that is shown by the graph.

Figure 4-3 shows an example **Summary** screen.

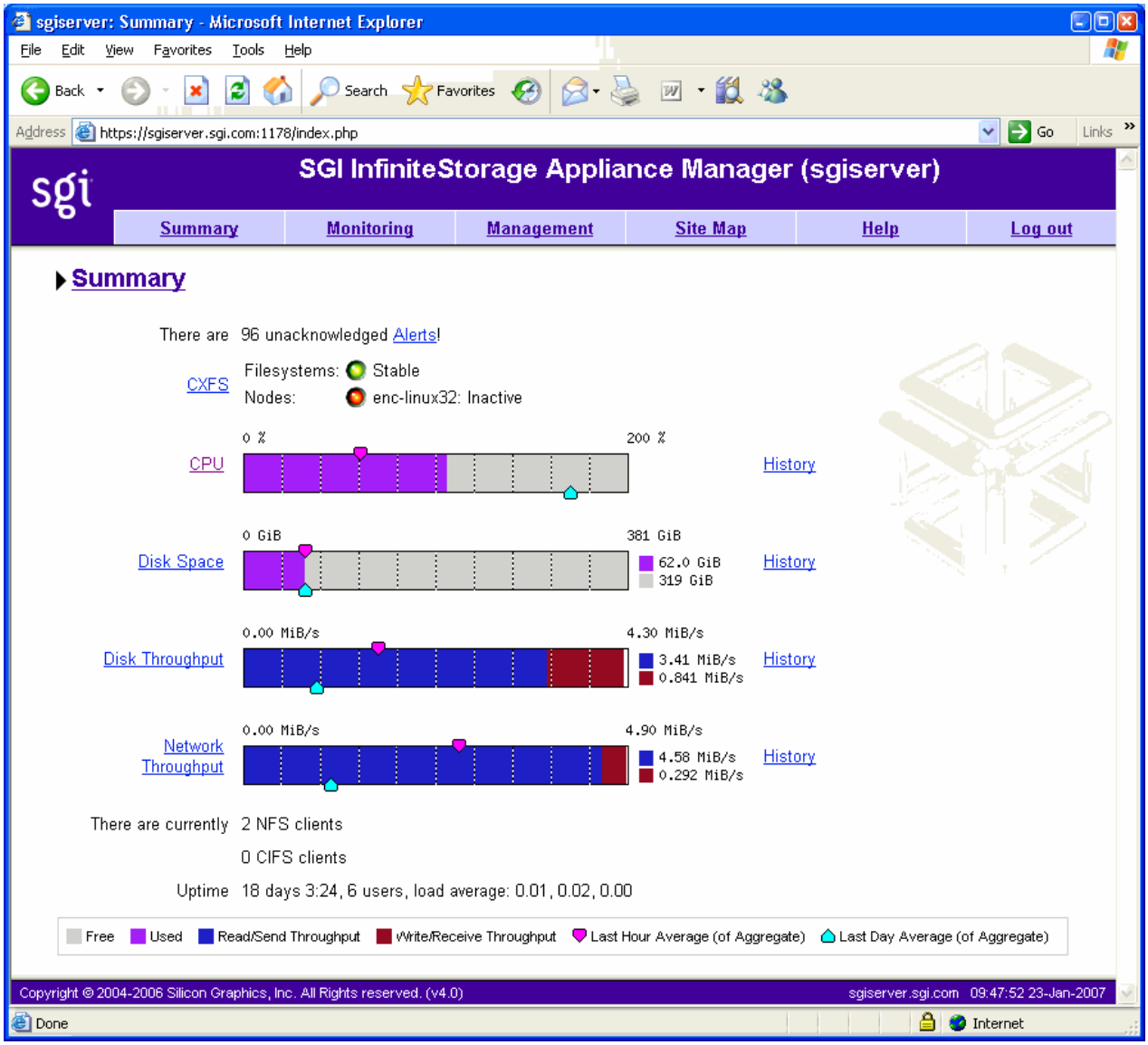

Figure 4-3 Summary Screen

In Figure 4-3, the bar graph for **Disk Throughput** shows 3.41 MiB/s of data read/sent (the blue part of the graph) and 0.841 MiB/s of data written/received (the red part of the graph). If you were sending and receiving data at the same rate, there would be equal amounts of red and blue in the graph. For more information, see Figure 4-2 on page 70.

For an HA cluster, the **Summary** screen will let you select the node for which you want to display information.

System Alerts

The **Alerts** screen displays messages from the system logs. These provide informative messages, notifications of unusual events, and error conditions.

Only unacknowledged alerts are displayed unless you click **Show Acknowledged**. You must log in in order to acknowledge alerts.

After a period of time, alerts are archived and will not be redisplayed. Acknowledged alerts are archived after 2 days and unacknowledged alerts are archived after 7 days. The `/var/lib/appman/alerts/archive` file contains all the archived alert messages.

Resources

Appliance Manager contains a separate screen to display the utilization of each resource.

The following sections provide details about the resources:

Where multiple physical resources are bonded into a single logical resource (for example, load-balanced NICs and RAID volumes in a filesystem), Appliance Manager shows the structure of the aggregated resource, and (where possible) shows metrics for both the aggregate and the component resources.

Disk Space

The **Disk Space** screen shows the number of bytes available on each filesystem. If the amount of disk space appears low on a filesystem on which disk quotas are enabled, you can use the **Disk Quota** screen to find out who is using the most disk space.

Disk Quota

Disk quotas provide limits on the number of files and the amount of disk space a user is allowed to consume on each filesystem. A side effect of this is that they make it possible to see how much each user is currently consuming.

Because quotas are applied on a per-filesystem basis, the limits reported in the **All Filesystems** screen are not additive. This means that if a user has a 500-MiB disk space limit on filesystem A and a 500-MiB limit on filesystem B, the user cannot store a 1-GiB file because there is no single filesystem with a large-enough space allowance.

However the current usage shown in the **used** column on the **All Filesystems** screen is additive, so you can use this screen to determine the users who are currently consuming the most disk space. The **All Filesystems** screen highlights users who have exceeded the quota on any filesystem on which they have been allocated a quota.

Disk Throughput and Disk IOPS

Disk operations occur when the result of a file operation is committed to disk. The most common types of disk operation are data reads and writes, but in some types of workload, metadata operations can be significant. *Metadata operations* include the following:

- Truncating and removing files
- Looking up filenames
- Determining the size and types of files

Disk operations are measured in I/O per second (IOPS).

Disk throughput is the amount of data that is transferred to and from the disks. This is predominantly the result of reading and writing data.

The **Disk Throughput** and **Disk IOPS** screens display a bar graph for each active filesystem. For RAID filesystems, a separate graph is displayed for each volume element.

If the cache hit rate is low and the network throughput is high, the disk throughput should be high. Usually, the disk throughput would be steady somewhere a little under the maximum bandwidth of the disk subsystem. If the disk throughput is consistently too high relative to the network throughput, this might indicate that the server has too little memory for the workload.

Under heavy loads, a storage server must be able to sustain a high rate of disk operations. You can use the disk operations metrics in conjunction with other metrics to determine the characteristics of a workload so that you can tune the server can be tuned. For example, a high utilization of NICs but few IOPS could indicate that a workload is coming straight from the cache. A large number of IOPS but low throughput (either disk or network) indicates a metadata-dominated load. You can determine the contributing operations or clients from the **NFS** screen, **CIFS** screen, and the various screens under the **Clients** category.

DMF Resources

The **DMF Resources** screens show how DMF is using its hardware, as described in the following sections:

- "OpenVault Tape Libraries" on page 76
- "Tape Drives" on page 76
- "Tape Volumes" on page 77
- "DMF-Managed Filesystems" on page 77
- "Disk Caches" on page 78
- "DMF Error Messages" on page 78

For information about solving problems, see "DMF Error Messages" on page 78. For information on how Appliance Manager displays user-generated DMF activity, see "DMF Activity" on page 89.

OpenVault Tape Libraries

The following displays the *tape library slot usage*, which is the number of slots used by DMF, other applications, or vacant):

Monitoring
> Resources
> DMF
> Tape Libraries

The **Tape Libraries** screen is available only if the OpenVault tape subsystem is in use. This screen is unavailable if you are using Tape Management Facility (TMF). (You must choose a single method for handling tapes, either OpenVault or TMF.)

Tape Drives

The following shows information about tape drives:

Monitoring
> Resources
> DMF
> Tape Drives

The **Tape Drives** screen provides information for each tape drive concerning its current state:

- Idle
- Busy
- Unavailable

When the drive is in use, it also shows the following:

- Activity (such as waiting)
- Purpose (such as recall)
- Details of the tape volume (such as volume name)

Note: This information is available only for DMF's tapes. Any other use, such as filesystem backups or direct tape use by users, is not shown; any such drives appear to be idle on this screen.

This screen also includes a link to the **Reservation Delay History** screen, which indicates when demand for tape drives exceeds the number available. This is purely a relative indication, to be compared visually with the equivalent indicator at other times; it has no useful numerical value.

Tape Volumes

The following shows the number of tape volumes in various states according to volume group (VG):

Monitoring
> Resources
> DMF
> Tape Volumes

Those volume groups that share an allocation group are shown together inside a box that indicates the grouping.

Because of their normally large number, full volumes are only shown numerically. Those in other states (such as empty) are shown graphically. History links show trends over time.

DMF-Managed Filesystems

The following shows the proportions of files on DMF-managed filesystems that are migrated and not migrated:

Monitoring
> Resources
> DMF
> Filesystems

Gathering this information is expensive, so the screen is only updated when DMF needs this information for other purposes, and the time at which it was accurate is displayed. Updates are caused by the execution of the following DMF programs: `dmfsfree`, `dmdaux` (run at DMF startup), `dmaudit`, `dmscanfs`, `dmhdelete`, and `dmselect`.

The screen also displays the amount of offline data related to the filesystems and the over-subscription ratios (which are typically in the range of 10:1 to 1000:1, although they vary considerable from site to site). As this is being viewed from the filesystem perspective, the fact that migrated files may have more than one copy on the back-end

media is not considered. That is, this is a measure of data that could be on disk but is not at the time, rather than a measure of the amount of back-end media being used.

Disk Caches

The following shows Disk Cache Manager (DCM) disk caches:

Monitoring
> Resources
> DMF
> Caches

DCM disk caches have similar issues to filesystems with regard to the frequency of updates as described in "DMF-Managed Filesystems" on page 77.

Dual-resident refers to cache files that have already been copied to back-end tape and can therefore be quickly removed from the cache if it starts filling. *Non-dual-resident* files would have tape copies made before they could be removed, which is much slower.

DMF Error Messages

This section describes problems you may encounter when monitoring DMF with Appliance Manager.

DMF Statistics are Unavailable or DMF is Idle

```
This screen requires statistics from DMF that are unavailable;
check that DMF is running, including the "pmdadmf2"process.
Make sure the DMF "EXPORT_METRICS" configuration parameter is enabled.
```

This message indicates that DMF is idle. When this occurs, perform the following procedure:

1. Check the version of DMF by running the `dmversion` command. It should report version 3.4.0.0 or later.
2. Check that the `EXPORT_METRICS on` line has been added to `/etc/dmf/dmf.conf` after the `TYPE base` line.

   Run `dmcheck` to search the DMF configuration file for syntactic errors.

3. Check that DMF has been restarted after the change to `/etc/dmf/dmf.conf` was made in step 2.
4. Check that the data is being exported by DMF by running the following command:

   ```
   # dmarenadump -v
   ```

   If it is not, run the following commands as `root`:

   ```
   # cd /dmf/spool  # or equivalent at your site
   # rm base/arena
   # /etc/init.d/dmf restart
   # /etc/init.d/pcp stop
   # /etc/init.d/pcp start
   # /etc/init.d/appman restart        # if necessary
   ```

5. Check that the data is passing through PCP by running the following command:

   ```
   # pminfo -f dmf2
   ```

   If it is not, run the following commands as `root`:

   ```
   # cd /var/lib/pcp/pmdas/dmf2
   # ./Remove
   # ./Install
   # /etc/init.d/appman restart
   ```

OpenVault Library Is Missing

```
No OpenVault-controlled library found.
```

This indicates that OpenVault is not running. Run the following command to verify that the `ov_stat` command is available:

```
# ls -lL /usr/bin/ov_stat
-rws--x--x 1 root sys 322304 Jul 22 2005 /usr/bin/ov_stat
```

If the file permissions are not `-rws--x--x` as shown above, run the following command to change the permissions:

```
# chmod 4711 /usr/bin/ov_stat
```

CPU Utilization

Serving files places demands on the storage server CPU as well as the I/O subsystem. The CPU helps with copying data to and from disks, calculating checksums, and other tasks. Table 4-1 shows the CPU metrics that Appliance Manager reports.

Table 4-1 CPU Metrics Reported by Appliance Manager

CPU Metric	Description
Wait time	Time when a CPU was forced to do nothing while waiting for an event to occur. Typical causes of wait time are filesystem I/O and memory swapping.
Interrupt time	Time the CPU spent processing requests from I/O devices. In a storage server context, these are almost exclusively generated by disk operations or network packets and by switching between processes.
System time	Time the CPU spent executing kernel code. This is usually dominated by NFS file serving and accessing data from disks.
User time	Time when the CPU is devoted to running ordinary programs. The biggest consumers of user time in a storage server would usually be the CIFS server, HTTP server, or FTP server.

CPU time is displayed as a percentage, where 100% is the total time available from a single CPU. This means that for an 8–CPU server, the total available CPU time is 800%.

In general, NFS workloads consume more system time, whereas CIFS, HTTP, and FTP workloads consume more user time. The Appliance Manager performance monitoring infrastructure consumes only a small amount of user time.

The most useful problem indicator is consistently having little or no idle time. This can mean that the server is underpowered compared to the workload that is expected of it.

Network Throughput

The **Network Throughput** screen displays the amount of data transferred through each network interface card (NIC).

If an interface is load-balanced, Appliance Manager displays throughput for both the aggregated (bonded) interface and its constituent interfaces.

Note: The throughput displayed is total network throughput (which includes protocol headers), so real data transfer will be somewhat lower than this value. The **Services** category screens show the amount of real data transferred from a variety of perspectives.

Hardware Inventory

The hardware inventory is a summary of the hardware configuration, including the CPUs, I/O controllers, memory, network controllers, and SCSI disks. The list of SCSI disks includes both the system `root` disk and the configured RAID logical units (LUNs).

Services

A *service* is a task that is performed by the storage server. While the primary service is fileserving, Appliance Manager breaks this down by the different methods of accessing the server. The services known to Appliance Manager are high availability (HA), NFS, CIFS, CXFS, DMF, and NDMP.

This section discusses the following screens available under the **Services** category:

HA

The **HA** monitoring screen displays an overview of the number of nodes and resources configured for high availability (HA). For each resource, you can see the resource's IP address, the node that currently owns the resource, the filesystems attached to the resource, and whether the resource is started or stopped.

NFS

Note: The **NFS** screen is available only if SGI Enhanced NFS is installed.

NFS traffic is a major contributor to storage server utilization. NFS services report statistics aggregated across all exports/shares as well as statistics for each export/share.

Table 4-2 describes the statistics reported by both the **NFS** and **CIFS** screens. Table 4-3 and Table 4-4 describe additional information that is reported.

Note: There is not a one-to-one correspondence between CIFS and NFS IOPS. The former measures operations that are received from a network client, the latter measures operations that are sent to a local filesystem.

Table 4-2 Statistics Reported by **NFS** and **CIFS** Screens

Graph	Description
Throughput	Current incoming and outgoing traffic for the export/share (the NFS service **Throughput** graph includes all types of operations, whereas the CIFS graph only shows actual data transfer)
Operations by Type	Export/share operations by class
Read Block Sizes	Reads by size
Write Block Sizes	Writes by size

Table 4-3 Additional Information Reported by the **NFS** Screen

Category	Description
IOPS	I/O per second for TCP and for UDP
Service Times	Number of operations falling into each service time band as tracked by the NFS server for each operation

Table 4-4 Additional Information Reported by the **CIFS** Screen

Category	Description
IOPS	Number of SMB operations per second
Latencies	Number of SMB operations falling into each service time band

NFS services gather like operations into a smaller number of operation classes. Table 4-5 summarizes these classes. (The NFS operation statistics measure classes of NFS protocol operations sent by clients.)

Table 4-5 NFS Operation Classes

Class	Description
access	File accessibility tests; checks whether a client can open a particular file
commit	Commit request; requests that the server flush asynchronously written data to stable storage
fsinfo	Filesystem statistics and information requests, `pathconf` calls, and service availability tests
getattr	File attribute retrieval operations
inode_mods	New file or directory creation, hard and symbolic link creation, file renaming, and device file creation operations
lockd	General lock operations not covered by other classes
lockd_granted	Number of lock granting operations
lockd_share	Number of export/share reservation operations

Class	Description
lookup	Operations that result in filename translations; that is, operations that are applied to a filename rather than to a file handle, such as `open`
read	File read operations and symbolic link resolution operations
readdir	Directory entry listing operations
readdirplus	Extended directory entry listing operations; returns the attributes of the directory entries as well as their names
remove	File deletion operations
setattr	File attribute setting operations, which include file truncations and changing permissions and ownership
write_async	Asynchronous writes; the written data may be cached and scheduled for writing at a later time
write_sync	Synchronous write; these do not complete until the data is written to stable storage
xattr	Operations that manipulate XFS extended attributes

CIFS

Note: The **CIFS** screen is available only if the SGI Samba packages are installed.

CIFS traffic is a major contributor to storage server utilization. CIFS services report statistics aggregated across all exports/shares as well as statistics for each export/share.

Table 4-2 describes the statistics reported by both the **NFS** and **CIFS** screens.

CIFS services gather like operations into a smaller number of operation classes. While these classes are largely similar, there are some differences. Table 4-6 summarizes these classes.

Note: Clients can perform file operations in a variety of different ways, which can result in similar logical operations being recorded as differing sets of CIFS operations depending on the application.

Table 4-6 CIFS Operation Classes

Class	Description
cancel	Cancel current activity operations
change/notify	Operations requesting notification of changes to a file or in a directory
close	File close operations
create/open	File and directory create and open operations
delete/remove	File deletion and directory removal operations
findfirst/next	Operations searching for files or scanning a directory
flush	Operations requesting a flush to disk of buffered data
getattr	Operations requesting file and directory attributes, such as access times
getsecurity	Operations requesting file access permissions
ioctl	Operations performing special filesystem features, such as sparse file handling
lock/unlock	File locking and unlocking operations
misc	All other operations, including infrequent filesystem features
move	File and directory move and rename operations
read	File read operations
setattr	Operations setting file and directory attributes, such as hidden file status
setsecurity	Operations setting file access permissions
write	File write operations

CXFS

The **CXFS** screen reports the status of CXFS filesystems and cluster nodes.

Filesystem information:

- Filesystem name.

- A **Usage** bar that shows the amount of disk space used on the filesystem. The numbers to the right of the bar show used space and filesystem size in gigabytes.
- **Stable** indicator, which is either green if the current state of the system **matches the expected state** of the system or red if it does not. For example, a filesystem is considered stable if it has been successfully mounted by **all nodes that are capable of mounting it**. If one or more nodes are currently trying to mount a filesystem, its stable indicator will be red and the **Status** text will be similar to **hostname: trying to mount**. After all nodes have mounted the filesystem, the indicator will be green.
- The most common **Status** states for filesystems include:
  - **Mounted**: All enabled nodes have mounted the filesystem
  - **Unmounted**: All nodes have unmounted the filesystem

Node information:

- Hostname.
- Node type, which is either **server** for the metadata server or **client** for a client-only node.[1]
- Cell ID, which is a number associated with a node that is allocated when a node is added into the cluster definition. The first node in the cluster has cell ID of 0, and each subsequent node added gets the next available (incremental) cell ID. If a node is removed from the cluster definition, its former cell ID becomes available.
- **Connected** indicator, which is one of the following colors:
  - Green if the node is physically plugged in, turned on, and accessible via the private network and Fibre Channel switch
  - Red if the node is not accessible
  - Gray if the node has been intentionally disabled by the administrator
- **Stable** indicator, which is one of the following colors:
  - Green if the node has joined the cluster and mounted the clustered filesystems
  - Red if the node has not joined the cluster and mounted the filesystems

1 *Metadata* is information that describes a file, such as the file's name, size, location, and permissions.The *metadata server* is the node that coordinates updating of metadata on behalf of all nodes in a cluster.

  - Gray if the node has been intentionally disabled by the administrator

  When a node comes online, the **Connected** indicator should always be green, with the **Stable** indicator red while the node is establishing membership, probing XVM volumes, and mounting filesystems. After these processes complete, both indicators will be green.

- The most common **Status** states for nodes include:
  - **Disabled**: The node is intentionally not allowed to join the cluster
  - **Inactive**: The node is not in cluster membership
  - **Stable**: The node is in membership and has mounted all of its filesystems

Any other filesystem or node status (not mentioned above) requires attention by the administrator.

Figure 4-4 the following shows the following:

- `sgiserver` has **Connected**=green, **Stable**=green, and **Status**=**Stable**, indicating everything is fine.
- `enc-linux64` and `enc-linux32` both have **Connected**=red, **Stable**=green, and **Status**=**Disabled**. This means that both systems are either powered down or not plugged in (**Connected**=red), but are considered stable (**Stable**=green) because the administrator disabled them via the CXFS management pages.
- `enc-mac` is powered down or not plugged in (**Connected**=red), but is enabled; it is therefore expect it to be up, hence the **Status**=**Inactive** state and **Stable**=red indicator.
- Because `sgiserver` and `enc-win` are the only nodes in the cluster that are actually operating correctly, they are the only nodes that have mounted the filesystem `/mnt/clufs`. All the other nodes are inactive or disabled, so they cannot mount that filesystem. However the filesystem itself is stable, and its status is therefore **Mounted**.

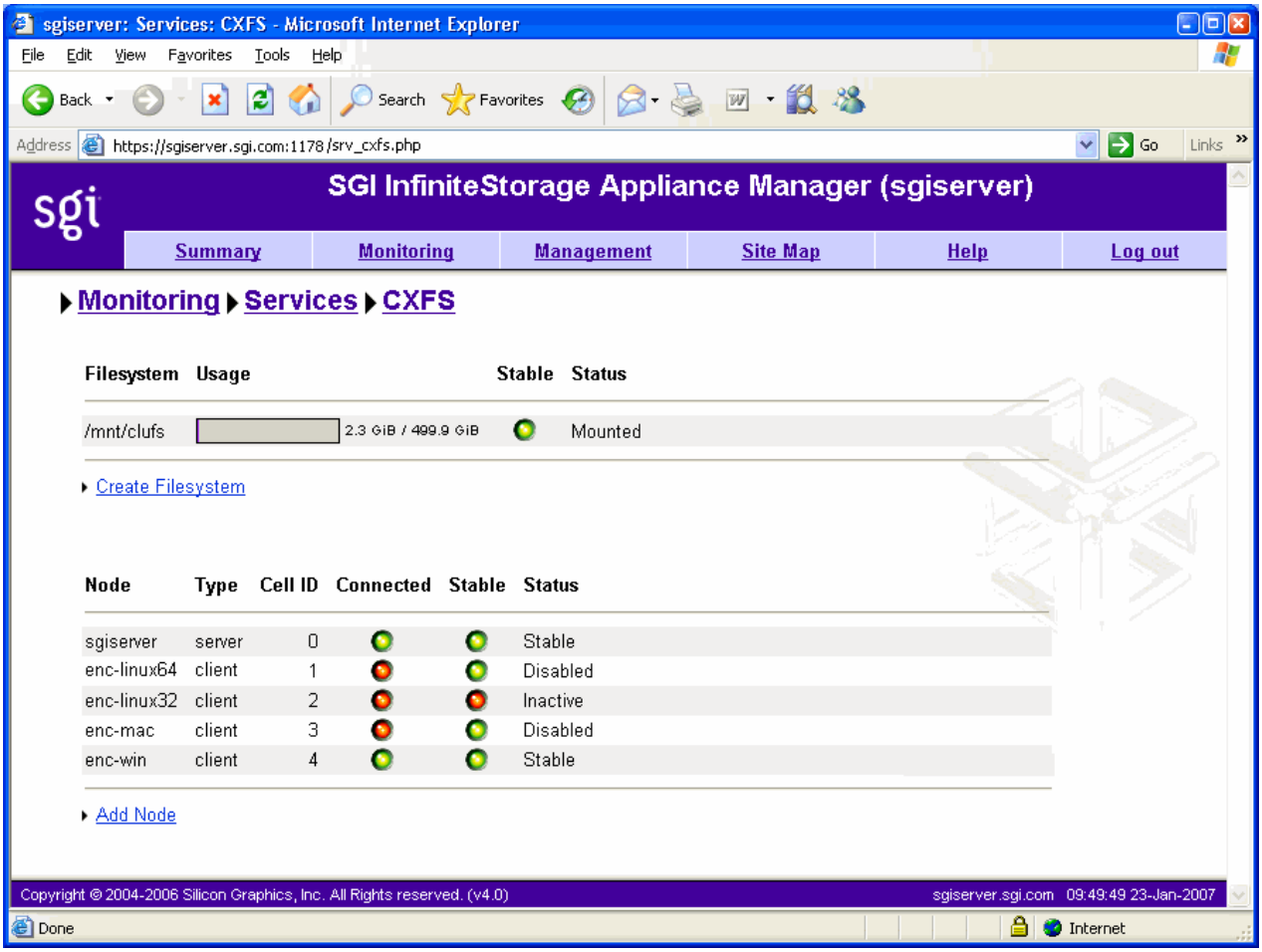

Figure 4-4 CXFS Monitoring Example

You can use the status of the nodes and filesystems as they appear on the **CXFS** screen to help diagnose issues. For example, when a client node is trying to mount a clustered filesystem, that client's status will be **Mounted 0 of 1 filesystems**. The filesystem's status will be ***client* trying to mount**. After a few seconds, the client should mount the filesystem and then both client and filesystem will be shown as **Stable** again.

However, if the client appears stuck on **Mounted 0 of 1 filesystems** for an extended period of time, this indicates there is a problem. In this case, do the following:

1. Check the status of the metadata server and the other clients. If other nodes are stable, it indicates that the filesystem and RAID are operating correctly and have been mounted by those other nodes.
2. Check the CXFS log file on the client for mounting-related errors. For example:

   ```
   cis_fs_mount ERROR: Illegal logbsize: 64 (x == 16k or 32k)
   cis_fs_mount ERROR: logbsize must be multiple of BBSIZE: 64
   op_failed ERROR: Mount failed for data3 (data3 on /mnt/data3)
   ```

   In this example, the client is unable to mount the filesystem due to one of the filesystem's mount options. In this case, you must use `cxfs_admin` to adjust the filesystem's mount options appropriately.
3. If no other nodes are stable (that is, all are trying to mount the filesystem and have been stuck in that state for an extended period), check the Appliance Manager **Alerts** page and the CXFS log files on the metadata server.

   See the following for more information about CXFS log files and tools:

   - *CXFS Administration Guide for SGI InfiniteStorage*
   - *CXFS MultiOS Client-Only Guide for SGI InfiniteStorage*

DMF Activity

The **DMF Activity** screen shows user-generated DMF activity from two points of view:

- Number of requests being worked on (the **Requests** screen)
- Rate of data throughput resulting from those requests (the **Throughput** screen)

Note: Values shown on the **Requests** and **Throughput** screens are averaged over the previous few minutes, so they are not necessarily integers as would be expected. This process also causes a slight delay in the display, which means that the values of **DMF Activity** screens do not necessarily match the current activity on the system, as seen in the DMF log files.

There are two distinct types of requests that are reflected in these screens:

- Requests from the user to the DMF daemon. These are presented as an aggregate across the DMF server, and on a per-filesystem basis, using the label of **Filesystems**.
- Requests from the DMF daemon to the subordinate daemons managing the back-end storage, the caches, the volume groups (VGs), and the media-specific processes (MSPs). Technically, caches are a variant of MSP despite their different purpose, hence the description **Non-Cache MSP** in the Appliance Manager screens.

Sometimes, there is a 1:1 correspondence between a daemon request and a back-end request by cache, volume group, or MSP (such as when a file is being recalled from back-end media back to the primary DMF-managed filesystem), but this is frequently not the case. For example, migrating a newly created file to back-end media will result in one back-end request per copy, but deleting a migrated file results in a single daemon request but no back-end request at that time. Tape merges may cause a lot of activity within a volume group but none at the daemon level.

On the top-level requests and throughput screens, and their associated **History** screens, for the sake of clarity the different types of requests are not distinguished from each other. However, if you zoom in (via one of the **Filesystems**, **Caches**, **Volume Groups**, or **MSPs** links on the left-hand side), the resulting screen shows the broad categories as well as by filesystem or by back-end storage group, as appropriate. This also applies to the related **History** screens.

NDMP

The **NDMP** screen shows the following information about the NDMP backup operations that are currently running:

Session ID — Displays the process ID of the NDMP session

Type — Displays the type of NDMP session. There are three major types of possible session:

- A `DATA` session writes/reads data to/from a network mover
- A `MOVER` session receives/sends data from/to a network NDMP data server
- A `LOCAL` session writes data to a locally attached backup device

Start Time	Displays the time that the backup began in seconds since 00:00:00 UTC, January 1, 1970
DMA Host	Displays the IP address of the data mover agent (DMA) host
DATA Host	Displays the IP address of the data host
GiB	Displays the number of gigabytes[2] that have been transferred
Throughput MiB/s	Displays the speed of throughput for the backup in megabytes[3] per second

To stop a backup, select it and click **Terminate Selected**. To select all backups, click the box in the table header.

To reset the page, select **Clear Selection.**

Versions

The **Versions** screen displays the version numbers of key software packages that have been installed.

Clients

A *NAS client* is a computer running a program that accesses the storage server. NAS clients are known to Appliance Manager by their IP address; if multiple accessing programs are running on the same computer, they are all counted as a single client.

Note: Client information is gathered only for CIFS and NFS protocols.

The **All Clients** screen will not be available if neither SGI Samba nor SGI Enhanced NFS are installed.

The **All Clients** screen displays the NAS clients sorted according to hostname. The other selections sort according to the chosen selection (such as by aggregate throughput).

2 GiB, 1024 megabytes
3 MiB, 1024

From each of these screens, you can change the sorted display of the data without returning to the **Monitoring** screen.

Displaying the NAS clients in this fashion is useful for pinpointing how the current set of clients are contributing the workload profile. For example, upon noticing an unusually large amount of network traffic on the **Network Throughput** screen, changing to display the clients in order of aggregate throughput will quickly identify the contributing clients.

From the list of clients, you can display a detailed view of the NFS and CIFS traffic generated by a particular client. This is useful when trying to diagnose problems that affect only a single client or type of client. For example, by viewing the client detail, it may be obvious that throughput is limited by the client using very small read and write sizes. Continuing from the client details to the client history screen can help diagnose problems, such as hung NFS mounts.

The **iSCSI** screen displays a list of the connected iSCSI initiators are connected and their targets.

Troubleshooting NFS-RDMA Problems

This chapter discusses the following problems that might be seen with network filesystem remote direct memory access (NFS-RDMA) protocol over InfiniBand:

Incorrect Routing of Packets or Poor Network Performance

If NFS-RDMA over InfiniBand results in incorrect routing of packets or poor network performance, verify that the IP addresses are different and that they remain different when the netmasks for their interfaces are applied. See "NFS-RDMA Over InfiniBand" on page 6.

NFS-RDMA Server Will Not Start

If the NFS-RDMA server will not start, verify that the `svcxprt_rdma` kernel module has been loaded and that the `/etc/sysconfig/nfs` file contains the following line:

```
SGI_NFS_RDMA_SERVER="yes"
```

NFS-RDMA Error Messages

```
Permission denied
```

Make sure that the server has the NFS-RDMA license installed.

How SGI InfiniteStorage Appliance Manager Configures Filesystems

This appendix describes how SGI InfiniteStorage Appliance Manager constructs a filesystem and provides an overview of the underlying volume and RAID device configuration that the system uses to lay out the filesystem:

For information about creating filesystems via Appliance Manager, see "Creating Filesystems" on page 29. The system uses the options you provide to create the underlying filesystems automatically.

Filesystem Creation Goals

Appliance Manager creates a filesystem with the goal of generalizing optimization for a variety of fileserver workloads.

When you create a filesystem, you choose whether to optimize for performance or capacity. If you select for capacity, Appliance Manager will use all the available disk space to create the filesystem, although this may come at the cost of slower performance. You also select a filesystem optimized for bandwidth or for I/O per second (IOPS). Select for bandwidth when you will have a small set of files and you must perform streaming reads and streaming writes as fast as possible. Select for IOPS when you will be performing random reads and writes to different sets of files. In general, selecting for IOPS will be the better choice.

In conjunction with these options, Appliance Manager attempts to provide a balance among these factors:

- Performance
- Manageability

- Reliability

When a filesystem is configured efficiently on a NAS system, you can support a great deal of data traffic at full disk-performance capacity.

Disk Striping

To optimize performance, Appliance Manager configures the filesystem so that the data is striped across multiple disk drives. The data is written in units to each drive in turn, in a round-robin fashion. This allows multiple reads and writes to take place in parallel, increasing IOPS performance.

To achieve maximum striping, the underlying RAID disk devices in a NAS system are grouped together into *physical volume elements* that combine multiple drives into a single logical unit. On top of that, the software groups the physical volume elements together into stripes, which together form a single concatenated volume element per filesystem. Figure A-1 describes this.

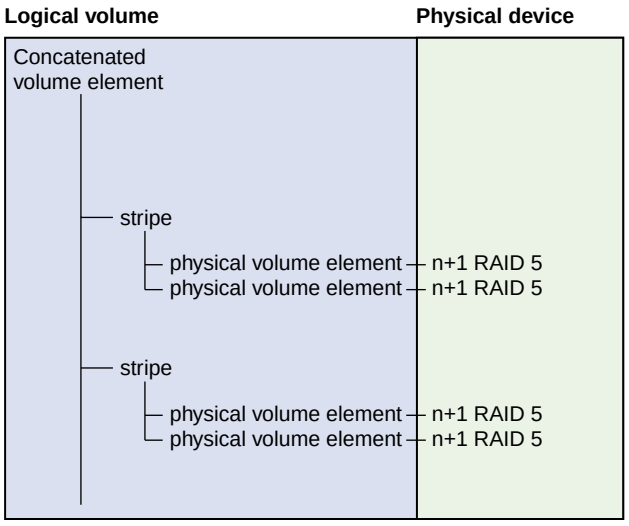

Figure A-1 Filesystem Structure

Appliance Manager uses RAID 5 devices. With RAID 5, parity is spread across the disks in an array. Because of this, you can lose one of the disks in the array without losing your data; the RAID device can still reconstruct the data. Where the disks in the RAID array are all the same size, the usable RAID capacity is the total number of disks in the array minus one.

When you create a filesystem, the system determines how much capacity the RAID devices provide and how the RAID devices can be arranged into stripes. From this, the system determines how many stripes the software will use to create the filesystem of the size you defined. If the number of RAID units in the system allows it, the system builds stripes that are two RAID units wide. If possible, the system builds RAID stripes that are four units wide.

The longer the stripe, the better the performance. After you have created the filesystem, however, you can add new disks to the system only in numbers that correspond to the stripe unit. For example, if the system's stripe unit is a four-way stripe of 4+1 RAID devices, then you must add 20 disks at a time if you need to grow the filesystem, as illustrated by Figure A-2, in which a stripe consists of 4 physical volume elements, each of which requires 5 disks (4 disks plus 1 parity disk). In this case, optimizing for performance entirely would cause you to lose manageability in terms of growing the filesystem at a later time.

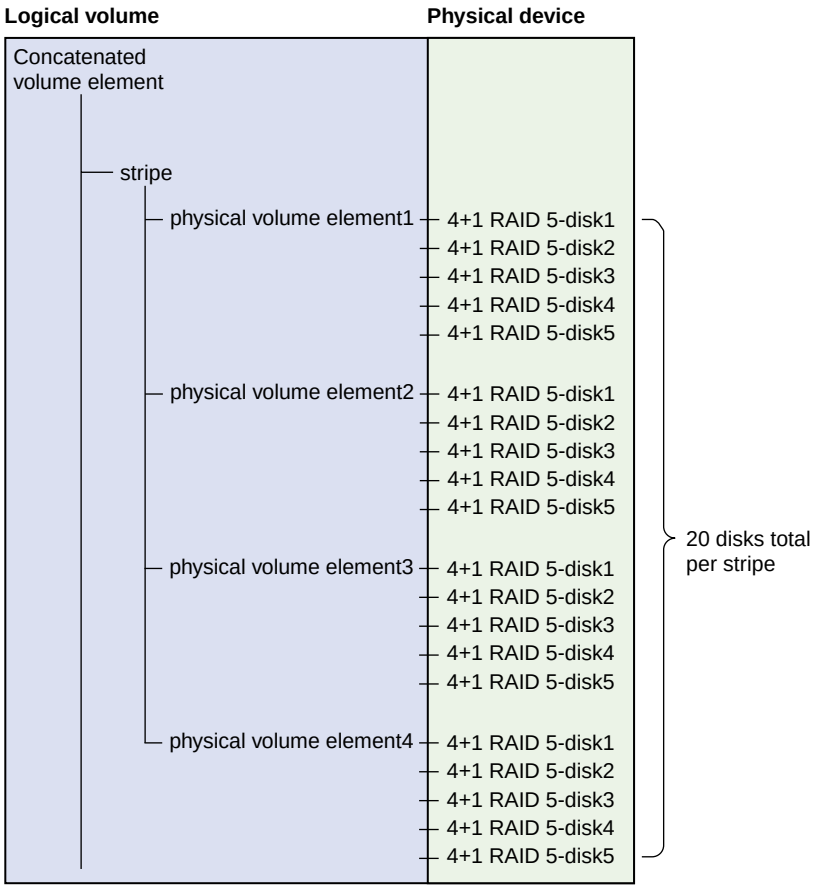

Figure A-2 Four-Way Stripe

Filesystem Configuration Factors

In determining the underlying filesystem configuration, Appliance Manager uses internal performance models that evaluate in a numerical fashion, according to multidimensional criteria, how suitable a RAID configuration will be.

When determining the filesystem configuration, Appliance Manager considers the following inputs:

- Whether you selected for capacity or performance
- Whether you selected for bandwidth or IOPS optimization
- How many disks the system has in its array

The underlying configuration will be different depending on the number of disks and whether that number divides evenly. In general, the system tries to use the most number of disks possible. When you create the filesystem, the system calculates the following:

- Stripe units
- Stripe width alignment down to the hardware level
- Header alignment on stripe boundaries

Disk Allocation

There is a fixed number of I/O per second (IOPS) that can be performed at the same time for each disk; if more than one filesystem shares the same disk, they share the IOPS for that disk. If there is only one filesystem on the disk, you get the performance for the entire array. If there are two filesystems on the disk, that performance is divided in two (and not always in a predictable way).

If the workload is more than 3 streaming reads or writes (for example, for media or satellite ingest), you should select for IOPS. For increased performance, contact SGI Professional Services.

Hot Spare Assignment

To increase reliability of a RAID system, a RAID array is often configured with a certain number of disks assigned as hot spares. A hot spare is a drive within a system that is not used unless a drive within that system fails, at which point it is added to the system to replace the failed drive without shutting the system down or interrupting service.

When creating filesystems with Appliance Manager, the assignment of hot spares is automatic. By default, the storage server has a single filesystem and hot spares

assigned. If you destroy that filesystem, the system will create the hot spares it determines are necessary for system reliability when you create your first new filesystem on the empty array. If you manually assign hot spares, the system will leave those disks as hot spares and create more hot spares if it determines that you need them.

How SGI InfiniteStorage Appliance Manager Configures the CXFS Cluster

SGI InfiniteStorage Appliance Manager **Setup Wizard** automatically creates the CXFS cluster if CXFS is licensed and installed. After the **Setup Wizard** is complete, the cluster will contain the following:

- One CXFS metadata server node, on which Appliance Manager runs [1]
- One Fibre Channel switch
- One clustered filesystem

You will add client nodes later using the **CXFS Cluster Nodes** management pages.

The cluster name, private network IP address, and Fibre Channel switch IP address are all pre-set. By default, the cluster name is `sgisan`. The CXFS private network is in the `10.`*x.x.x* range; one of the ethernet ports on the metadata server is assigned an address within this range, typically `eth2`. Another ethernet port (typically `eth3`) is configured for direct connect to the Fibre Channel switch.

Note: Ethernet port assignments may vary, depending on your system hardware.

Changing the Network Configuration

If you have a site-specific reason that requires you to change the cluster name or private network IP address, you can do so using the CXFS tools described in *CXFS Administration Guide for SGI InfiniteStorage*. However, both these operations are disruptive to the cluster and must be done with care.

For example, to change the private network IP address by using the `cxfs_admin` tool:

1. Use `cxfs_admin` to disable the metadata server and change the `private_net` IP address.

[1] *Metadata* is information that describes a file, such as the file's name, size, location, and permissions.The *metadata server* is the node that coordinates updating of metadata on behalf of all nodes in a cluster.

2. Use Appliance Manager to reconfigure the appropriate network interface with the new private network IP address.
3. Use `cxfs_admin` to reenable the metadata server.

Note: While the metadata server is disabled, the CXFS management and monitoring pages in Appliance Manager will display the error message `Unable to connect to cluster`. These pages will return to normal after the metadata server is enabled and has reestablished membership, which can take several seconds. For more information, see "Cluster Connection Issues" on page 102.

To change the cluster name, you must completely destroy and re-create the cluster using the CXFS tools.

Cluster Connection Issues

The message `Unable to connect to cluster` may appear on the **Summary** page or on the CXFS management or monitoring pages for the following possible reasons:

- The metadata server is currently establishing membership in the CXFS cluster. It can take several seconds for the metadata server to establish membership. Wait a few seconds and reload the page.
- The CXFS cluster daemons are not running. Check the daemon status on the following page:

  Management
  > Services
  > CXFS
  > Start/Stop

  Start the cluster daemons if necessary.
- The CXFS cluster is misconfigured. You can use the `cxfs_admin` and `cxfs-config` tools to further diagnose cluster configuration issues. For more information, see *CXFS Administration Guide for SGI InfiniteStorage*.

Reinstalling Appliance Manager

For information about installing Appliance Manager onto a clean machine, see the *SGI InfiniteStorage Software Platform Release Notes*. This appendix discusses the following:

Reinstalling After the Network is Configured

If you are reinstalling from CD after your network has already been configured, you must still run through the Setup Wizard in order for the system to operate correctly. If networking has already been configured, replace `https://192.168.9.9:1178/` in the instructions above with `https://`*YOUR_SERVER*`:1178/` in order to access the Setup Wizard (where *YOUR_SERVER* is the hostname or IP address of your system).

Setting Up an HA Cluster after Reinstalling

If you use a high-availability (HA) cluster, you must do the following:

1. Do the following on the first node:
   a. Run the following script:

      ```
      /usr/lib/appman/ha/appman-ha-setup stage1 0 2
      ```

   b. Press **Enter** to acknowledge that this will destroy data on your server.
   c. Provide the hostname or IP address of an NTP server (this will be used to synchronize the time, it will not be written to the NTP configuration file.
   d. Press **Enter** to reboot.
2. Do the following on the second node:
   a. Run the following script:

      ```
      /usr/lib/appman/ha/appman-ha-setup stage1 1 2
      ```

b. When prompted, press **Enter** to acknowledge that this will destroy data on your server.

c. Provide the hostname or IP address of an NTP server (this will be used to synchronize the time, it will not be written to the NTP configuration file.

d. Press `Enter` to reboot.

Note: You must wait until both nodes have rebooted and are sitting at the login prompt before continuing to step 3.

---

3. On the first node, do the following:

   a. Login via the L2, because IP addresses have been set back to defaults.

   b. Run the following script:

      ```
      /usr/lib/appman/ha/appman-ha-setup stage2
      ```

   c. When prompted, press **Enter** to acknowledge that this will destroy data on your server.

   d. Supply the `root` password for each system.

   e. Specify a partition for DRBD, for example `/dev/sda5`. The same partition must exist on both nodes. All data on this partition on both nodes will be destroyed.

   f. Specify the size for the DRBD partition or press `Enter` to use the default of 20GB. The size specified must be less than or equal to the physical size of the block device you specified.

   g. Wait while DRBD synchronizes. If you like, you can log into the first node and check the synchronization progress:

      ```
      cat /proc/drbd
      ```

   h. Wait for a few minutes in order to get a cluster connection.

   i. When prompted, specify the L2 password. This can be empty (just press `Enter`). All L2s in the cluster must have the same password.

   j. Specify the IP address for each node's L2 (depending on your configuration, all systems might be on the same L2).

k. Press `Enter` to reboot.

Note: After the `stage2` setup is complete, you will see `network failed` messages during system boot. You can ignore these messages.

4. Plug a laptop into eth0 on the first node and run the Setup Wizard:

   `https://192.168.9.9:1178`

   See Chapter 2, "Initial System Setup" on page 9.

5. Plug the first node's eth0 into the network and wait for it to reboot.

Glossary

Active Directory

A directory service that implements *LDAP* in a Windows environment. It provides a hierarchical structure for organizing access to data.

administration password

The password required to log into the **Management** screens of SGI InfiniteStorage Appliance Manager

aggregate (bonded) network interface

Virtual network interface that consists of real interfaces working in tandem. A virtual interface can provide the aggregated bandwidth of all of the interfaces that you used to create it.

cell ID

A number associated with a node that is used by the CXFS software and appears in messages.

CHAP

Challenge Handshake Authentication Protocol is a means of authentication used between a client and server where the password is sent over the wire in a form that is impossible to discover and impossible to replay. Both client and server must know what the original password is, but someone snooping on wire traffic cannot recover the password and cannot later send the original (snooped upon) authentication packet to the server in an attempt to try to trick it into letting them authenticate as a valid client.

CIFS

Common internet filesystem. This protocol is usually used by Microsoft Windows clients.

client-only node

A node in a CXFS cluster that does not run cluster administration daemons and is not capable of coordinating CXFS metadata.

cluster

A *cluster* is the set of systems (nodes) configured to work together as a single computing resource. A cluster is identified by a simple name and a cluster ID. In CXFS, a cluster running multiple operating systems is known as a *multiOS cluster*.

csync2

A tool that is used to synchronize configuration files across the HA cluster. See http://oss.linbit.com/csync2/.

current metric

Metric drawn live from the server or taken from the last few minutes of the metric archives.

CXFS

Clustered XFS filesystem.

DCM

Disk cache manager, which lets you configure the DMF disk MSP to manage data on secondary storage, allowing you to further migrate the data to tape as needed.

default network gateway

The IP address of the router that this system should use to communicate with machines that are outside of its subnet.

DHCP

Dynamic host configuration protocol (DHCP) allows one or more server systems to dynamically distribute network IP addresses and site configuration parameters to new or requesting client systems. By using DHCP, a site with only a few available addresses can serve a large number of hosts that connect to the network only occasionally, or a large site can manage the permanent assignment of addresses with a minimum of administrative attention. The NAS server can be configured as a DHCP client.

directory service
See *name service*.

disk IOPS
Disk I/O per second.

disk striping
Writing data in units to multiple disks in a round-robin fashion, increasing IOPS performance.

disk throughput
The amount of data that is transferred to and from disks.

distinguished name
A unique identifier for an entry in an LDAP directory tree structure.

DMF
Data Migration Facility, a hierarchical storage management system for SGI environments.

DRDB
Distributed Replicated Block Device. See http://www.drbd.org/.

dual-resident file
In DCM, a cache-resident copy of a migrated file that has already been copied to tape, and can therefore be released quickly in order to prevent the cache filling, without any need to first copy it to tape

FC
Fibre Channel storage interface connection.

fence
The isolation of a problem node so that it cannot access I/O devices, and therefore cannot corrupt data in the shared CXFS filesystem.

FQDN

Fully qualified domain name.

filesystem resource

An HA resource providing one or more filesystems grouped together and shared by NFS and CIFS.

gigabyte

1024 megabytes (also known as *gibibyte*). On the **DMF Configuration** screens, disk sizes use multipliers that are powers of 1000, such as kB, MB and GB. This is for consistency with the DMF documentation and log files. However, the rest of Appliance Manager, including the DMF **Monitoring** screens, use multipliers that are powers of 1024, such as kiB, MiB and GiB.

HA

High availability

HA cluster

High-availability cluster. HA resources survive a single point of failure. In an HA cluster, each HA resource is actively owned by one node. If that node fails, another node restarts the HA applications of the failed node. To application clients, the services on the backup node are indistinguishable from the original services before failure occurred. It appears as if the original member has crashed and rebooted quickly.

Heartbeat

Infrastructure from the High Availability Linux Project (http://linux-ha.org)

historic metric

Metric taken exclusively from the metric archives.

hot spare

Disk drive within a RAID array that is not used unless another drive within the RAID array fails, at which point it is added to the filesystem to replace the failed drive without shutting the filesystem down or interrupting service.

idle time

Time that remained when the CPU could not find any tasks to run

initiator

The client accessing the storage in an iSCSI network.

interrupt time

Time the CPU spent processing requests from I/O devices. In a storage server context, these are almost exclusively generated by disk operations or network packets and by switching between processes.

I/O fencing

See *fence*.

IOPS

I/O per second.

IPoIB

IP over InfiniBand

iSCSI

Internet Small Computers Systems Interface is a protocol that is used to transport SCSI commands across a TCP/IP network. This allows a system to access storage across a network just as if the system were accessing a local physical disk. In an iSCSI network, the client access the storage is called the *initiator*. The remote storage that the client accesses is called the *target*.

LDAP

Lightweight directory access protocol (LDAP) is a networking protocol that organizes access to data in a directory tree structure.

metadata

Information that describes a file, such as the file's name, size, location, and permissions.

metadata server

The node that coordinates updating of metadata on behalf of all nodes in a CXFS cluster.

MSP

Media-specific process, the daemon-like process in DMF by which data blocks are copied onto alternate media, and which assigns keys to identify the location of the migrated data.

name service

Application that manages the information associated with network users.

NAS client

Computer running a program that accesses the storage server.

NFS

Network file system.

NIC

Network interface card.

NIS

Network information service (NIS) is a network lookup service that provides a centralized database of information about the network to systems participating in the service.

node

A *node* is an operating system (OS) image, usually an individual computer. (This use of the term *node* is different from the NUMA definition for a brick/blade on the end of a NUMAlink cable.)

See also *client-only node*.

non-dual-resident file

A file in DCM that is not a cache-resident copy of a migrated file. It must be migrated to tape before it can be removed.

NTP

Network Time Protocol.

physical volume element

The combination of multiple RAID disk drives into a single logical unit.

RAID

Redundant array of independent disks.

RAID 5

A level of RAID that uses block-level striping and distributed parity

remote replication

Duplication of local filesystem writes on a separate machine. The local system is the replication *source*, the system on which the writes are duplicated is the replication *destination*.

resource

In the context of high availability, a service associated to an IP address and managed by Linux-HA (Heartbeat). Heartbeat starts, monitors and stops resources. Heartbeat uses the IP address of a resource to redirect clients to the node currently running the resource. Each HA resource is actively owned by one node. If that node fails, another node restarts the HA applications of the failed node. To application clients, the services on the backup node are indistinguishable from the original services before failure occurred. It appears as if the original member has crashed and rebooted quickly. Resources can have a *resource IP address* associated with them that is used to redirect clients to the node that is currently running the resource.

In the context of the Appliance Manager interface, a resource is something that is monitored and managed by Appliance Manger (such as network interfaces or DMF).

resource group

A set of related HA resources that are started together on the same node. See http://www.linux-ha.org/ResourceGroup

resource IP address

An IP address that represents an HA resource and moves between nodes in the cluster as required.

serial ATA (SATA)

Serial advanced technology attachement storage interface connection.

service

Task performed by the storage server.

shadow file

A file that is protected from all access by non-`root` users and stores the encrypted passwords

`ssh`

A tool that is used to communicate between nodes in the cluster. See http://www.openssh.com/

smart host

The gateway server where email should be delivered.

snapshot

See *XVM snapshot*.

system time

Time the CPU spent executing kernel code. This is usually dominated by NFS file serving and accessing data from disks.

tape library slot usage

The number of slots used by DMF, other applications, or vacant.

target
The storage that appears to the initiator as a disk drive in an iSCSI network.

TMF
Tape Management Facility.

VG
Volume group, one of the components of a DMF library server. A volume group is responsible for copying data blocks onto alternate media.

UI resource
The Heartbeat resource providing the Appliance Manager user interface (UI) service

wait time
Time when a CPU was forced to do nothing while waiting for an event to occur. Typical causes of wait time are filesystem I/O and memory swapping.

XFS
The SGI filesystem.

XVM snapshot
Virtual point-in-time image of a filesystem. Snapshot copies are not actual media backup for a filesystem.

YaST
An operating system setup and configuration tool.

Index

G

H

I

L

M

N

O

P

Q

R

S

T