



---

# SGI NAS

## User Guide

---

Copyright © 2012 SGI. All rights reserved; provided portions may be copyright in third parties, as indicated elsewhere herein. No permission is granted to copy, distribute, or create derivative works from the contents of this electronic documentation in any manner, in whole or in part, without the prior written permission of SGI.

SGI reserves the right to make changes to this document at any time without notice and assumes no responsibility for its use.

Refer to the latest product announcement or contact your local SGI representative for information on feature and product availability.

This document includes the latest information available at the time of publication.

#### TRADEMARKS AND ATTRIBUTIONS

SGI, Silicon Graphics, and the SGI logo are trademarks or registered trademarks of Silicon Graphics International Corp. or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in several countries. Novell and SUSE are registered trademarks of Novell, Inc., in the United States and other countries. Red Hat and all Red Hat-based trademarks are trademarks or registered trademarks of Red Hat, Inc. in the United States and other countries. Microsoft and Windows are registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries. All other trademarks mentioned herein are the property of their respective owners.

Document Number: 007-5860-001

## Table of Contents

1 Overview .....	9
1.1 Purpose .....	9
1.2 Audience .....	9
1.3 Document conventions.....	9
2 Introduction.....	11
2.1 Terminology.....	11
2.2 Functional Block Diagram.....	15
2.3 Storage Limits.....	16
3 NMC Overview.....	17
3.1 Accounts.....	17
3.2 Command Completion.....	17
3.3 Command Summary.....	18
3.4 Scripting.....	19
4 NMV Overview.....	21
4.1 Accounts.....	21
4.2 Login.....	21
4.3 Navigation.....	21
4.4 Terminal Access.....	22
4.5 View Log.....	22
5 Initial Setup.....	23
5.1 Installation wizard I.....	23
5.2 Installation wizard II.....	26
6 Managing Data Volumes.....	33
6.1 Data Redundancy.....	33
6.2 Create Data Volume.....	34
6.2.1 De-duplication.....	36
6.2.2 Auto-expand.....	37
6.3 Creating various RAID configurations.....	38

6.4 View Status.....	38
6.5 Edit Properties.....	39
6.6 Expand Data Volume.....	41
6.7 Destroy Data Volume.....	43
6.8 Export/Import Data Volumes.....	43
6.8.1 Export.....	44
6.8.2 Import.....	44
6.9 Scrub.....	47
6.10 Emergency reserve of volume space.....	48
7 Disk Management.....	49
7.1 Locating Disks.....	49
7.2 JBOD Management.....	50
7.2.1 About JBOD.....	50
7.2.2 NMV JBOD Management.....	50
7.2.3 NMC JBOD Management.....	54
7.3 Viewing Disk Status .....	56
7.4 Adding Spares to a Data Volume.....	58
7.5 Adding Global Spares.....	57
7.6 Adding Cache Devices.....	61
7.7 Adding Log Devices.....	62
7.8 Removing a Device.....	63
7.9 Replacing a Disk.....	63
7.10 Taking a disk offline.....	64
7.11 Recovering a previously disconnected disk.....	64
7.12 Replacing a Redundancy Group.....	64
7.13 Creating a Mirror.....	65
7.14 Detaching a Mirror.....	65
7.15 Re-attaching a Mirror.....	65
8 Managing Folders.....	67
8.1 Create Folders.....	67



8.2 View Status.....	68
8.3 Edit Properties.....	68
8.3.1 Logbias property.....	70
8.4 Destroy Folder.....	70
8.5 Search & Indexing.....	71
8.6 Sharing Folders.....	71
8.6.1 Sharing Folders with NFS and CIFS.....	72
9 NFS File Sharing.....	73
9.1 Create NFS Share.....	73
9.2 Edit NFS Folder Properties.....	73
9.3 Mounting on Linux.....	73
10 CIFS File Sharing.....	75
10.1 Introduction.....	75
10.2 Configuring CIFS server.....	75
10.2.1 Create CIFS share.....	78
10.3 Other sharing options.....	81
10.3.1 WebDAV.....	81
10.3.2 RSYNC.....	83
11 Managing Snapshots.....	85
11.1 Create Snapshot.....	85
11.2 Creating Checkpoints.....	86
11.3 Setup Periodic Snapshots.....	87
11.4 View Snapshots.....	88
11.5 View Scheduled Snapshots.....	89
11.6 Recover Snapshot.....	89
11.7 Delete Snapshot.....	94
11.8 Other operation with snapshots.....	95
12 SCSI Target (Managing Blocks).....	97
12.1 Create zvol.....	97
12.2 View Zvol Properties.....	99

12.3 Destroy a Zvol.....	99
12.4 Add initiators and targets.....	99
12.5 Create initiator group.....	99
12.6 Create target group.....	100
12.7 Create LUN mappings.....	101
13 Managing iSCSI.....	103
13.1 Add remote initiator.....	103
13.2 Create iSCSI target.....	103
13.3 Create iSCSI target portal group.....	104
13.4 Setting up CHAP Authentication.....	105
14 Asynchronous Replication.....	109
14.1 Auto-Sync.....	112
14.2 Auto-Tier.....	112
14.2.1 The table of auto-tier properties.....	113
14.2.2 Auto-tier protocols.....	115
14.2.3 Setting up auto-tier with 'rsync+nfs' transport protocol.....	115
15 Synchronous Replication (Auto-CDP).....	117
15.1 Installation.....	118
15.2 Getting Started.....	118
15.3 The alternative hostname.....	120
15.4 Enabling Auto-CDP service instance.....	120
15.5 Volume operations and Auto-CDP.....	122
15.6 Service monitoring.....	123
15.7 Auto-CDP configuration properties.....	124
15.8 Service States.....	124
15.9 Troubleshooting.....	125
15.10 Creating Auto-CDP – example.....	126
15.11 Reverse mirroring – example.....	128
15.12 How to establish SSH-binding.....	129
15.13 SSH-bind troubleshooting .....	132

16 Using NDMP with SGI NAS.....	135
16.1 Setting up NDMP server.....	136
16.2 Three-way copying with ndmpcopy.....	137
16.3 NDMP log – NDMP device – NDPM session.....	138
16.4 Performance and considerations.....	139
16.5 Known issues.....	140
17 Operations and Fault Management.....	141
17.1 Runners.....	141
17.2 Triggers.....	143
17.3 Handling an Unrecoverable I/O Error.....	145
17.4 Handling a System Failure.....	146
18 Analytics.....	147
18.1 DTrace.....	147
18.1.1 DTrace command line.....	147
18.2 NMV Analytics.....	148
18.3 I/O Performance.....	149
18.4 Performance Benchmarks.....	150
18.4.1 I/O performance benchmark.....	150
18.4.2 Network performance benchmark.....	151
19 Managing the Users.....	155
19.1 Adding Local Appliance Users.....	155
19.2 Local Appliance Groups.....	156
19.3 LDAP.....	157
19.4 ACLs.....	159
19.5 User Quotas.....	161
19.6 Group Quotas.....	161
20 Managing NMS Properties.....	163
20.1 Multi-nms.....	164
20.2 Concept Diagram.....	166
20.3 Known issues.....	166

21 Managing the Network.....	169
21.1 Changing Network Interface Settings.....	169
21.2 Link Aggregation.....	169
21.3 VLAN.....	170
21.4 IP Aliasing.....	173
21.5 TCP Ports used by SGI NAS.....	173
22 Managing the Appliance.....	177
22.1 Secure Access.....	177
22.2 Registering the Commercial Version.....	179
22.3 Installing/ Removing plug-ins.....	179
22.4 Saving and Restoring Configurations.....	181
22.5 Upgrades.....	183
23 Contact information.....	185
23.1 Support request .....	185
23.2 Other resources.....	186

# 1 Overview

## 1.1 Purpose

This guide is intended to describe all the aspects of work with SGI NAS. For the step-by-step installation instructions read the SGI NAS Quick Start Guide. The SGI NAS User Guide contains basic SGI NAS terminology, NMV and NMC managing instructions, and provides information about basic SGI NAS functionality.

## 1.2 Audience

The guide's audience is the SGI NAS administrators, system administrators, users or any other involved parties.

## 1.3 Document conventions

- SGI NAS Management Console (NMC) commands:

```
nmc:/$
```

- UNIX shell commands:

```
#
```

- A note or another piece of important information:





## 2 Introduction

SGI NAS is a software-based storage appliance based on the Zetta File System (ZFS) from OpenSolaris. SGI NAS supports file and block storage and a variety of advanced storage features such as replication between various storage systems and virtually unlimited snapshots and file sizes.

The product supports direct-attached SCSI, SAS, and SATA disks, and disks remotely connected via iSCSI, FibreChannel, or AoE protocols. Networking support includes 10/100/1G BaseT and many 10G Ethernet solutions, as well as aggregation (802.3ad) and multi-path I/O. For most installations, we recommend 100Mbps Ethernet at a minimum.

An in-kernel CIFS stack is provided and NFS v3 and v4 are supported. For easy access from Windows, WebDAV offers another file sharing option. The product also makes use of rsync, ssh, and zfs send/receive, CIFS and NFS transports for tiering and replication. Block level replication (remote mirroring) is provided as an optional module.

Directory services such as Active Directory and LDAP are supported, including UID mapping, netgroups, and X.509 certificate based client authentication.

### 2.1 Terminology

Term	Comment
<b>G;</b> = <b>B5G</b>	SGI NAS System.
<b>SA-API</b>	Storage Appliance API. NMS (see next) is a sole provider of SA-API. The API provides access to the appliance's management objects and services. All client management applications use the same API (namely, SA-API) to monitor and administer the appliance. This ensures consistent view of the appliance from all clients, transactional behavior of all management administrative and monitoring operations, and easy third-party integrations. Read full description of SA-API in <a href="#">SA-API</a> documentation
<b>NMS</b>	SGI NAS Management Server. There is only one server instance per appliance. The server provides public and documented Storage Appliance API (SA-API) available to all

Term	Comment
	appliance management and monitoring clients, remote and local, including (but not limited to) NMC.
<b>NMC</b>	SGI NAS Management Console. NMC can be used universally to view and configure every single aspect of the appliance: volumes and folders, storage and network services, fault triggers and statistic collectors. NMC communicates with the local NMS (see previous) and remote management consoles and management servers to execute user requests. Multiple NMC instances can be running on a given appliance. NMC is a single-login management client with a capability to manage multiple appliances and groups of appliances.
<b>NMV</b>	SGI NAS Management View. Web client uses the same SA-API (above) to communicate with the NMS. NMV shows status of all appliances on the network, displays graphical statistics collected by 'statistic collectors' (see below), and more.  SGI NAS management software is further illustrated in Section "Functional Block Diagram" below.
<b>Volume</b>	SGI NAS volume is a ZFS pool (a. k. a. zpool), with certain additional attributes. There is a one-to-one relationship between a volume and the underlying ZFS pool.
<b>Folder</b>	SGI NAS folder is a ZFS filesystem.
<b>Auto-Snap</b>	A type of appliance's storage service. The <b>auto-snap</b> service enables easy management of snapshots, providing regular multiple period scheduling on a per-folder or per-volume basis (with or without recursion into nested folders/filesystems). In addition, auto-snap allows to define a certain snapshot-retention policy. Snapshots can be kept for years, and/or generated frequently throughout the day.
<b>Auto-Tier</b>	A type of the appliance's storage services. The <b>auto-tier</b> (or simply, "tiering") service can regularly and incrementally copy data from one host (local or remote, appliance or non-appliance) to a destination, local or remote, again of any type. SGI NAS <b>auto-tier</b> service runs on a variety of transports, and can use snapshots as its replication sources. This solution fits the more common backup scenarios found in disk-to-disk backup solutions. However, unlike regular backup solutions with only the latest copy available on the backup destination, this solution provides the advantage of both 'the latest copy' and a configurable number of previous copies.



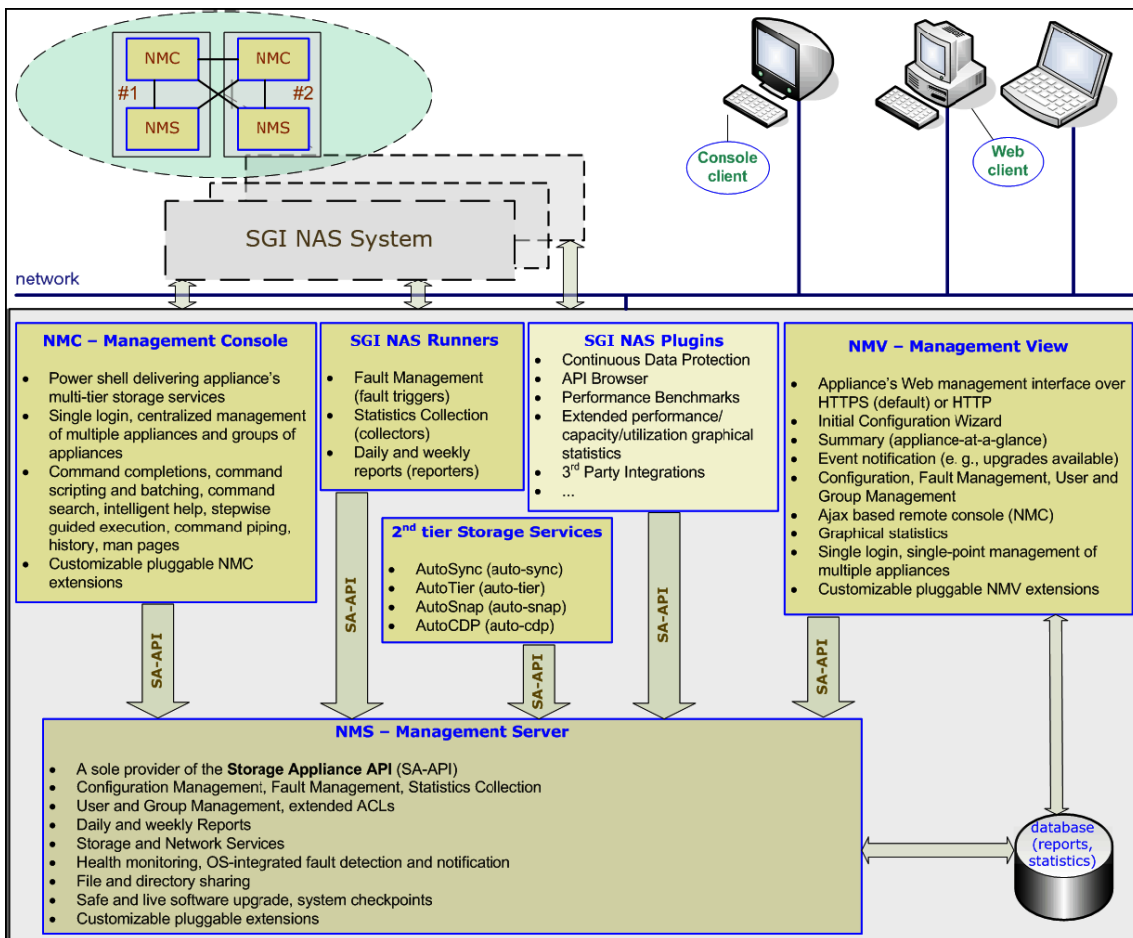
Term	Comment
<b>Auto-Sync</b>	A type of the appliance's storage services. The <b>auto-sync</b> (or simply, "syncing") service will maintain a fully synchronized copy of a given volume or folder on another SGI NAS System. Where tiering provides a copy, SGI NAS <b>auto-sync</b> service provides a true mirror, inclusive of all snapshots. The major difference between <b>auto-tier</b> (see previous) and <b>auto-sync</b> services that the latter transfers both data <u>and</u> filesystem metadata from its source to its (syncing) destination. This allows for standby hosts, as well as image-perfect recovery sources for reverse mirroring in case of a failure in the primary storage.
<b>Auto-CDP</b>	Automatic Continuous Data Protection (CDP) service. SGI NAS <b>auto-cdp</b> service provides remote mirroring capability. The service allows to replicate disks between two different appliances in real time, at a block level. Conceptually, the service performs a function similar to local disk mirroring scheme of RAID 1 except that in the case of <b>auto-cdp</b> this is done over IP network.  Auto-CDP is distributed as plug-in (see below).
<b>Trigger</b>	Fault Triggers, or simply 'triggers', are the appliance's primary means of fault management and reporting. Each fault trigger is a separate (runtime-pluggable) module that typically runs periodically at a scheduled interval and performs a single function, or a few related functions. Triggers actively monitor appliance's health, state of all its services and facilities, including hardware. See also 'SGI NAS Runners' below.
<b>Collector</b>	Statistic Collectors, or simply 'collectors' are, as the name implies, the appliance's means to collect network and storage statistics. A large number of network and storage IO counters is collected on a regular basis and recorded into SQL database. The data is then used to generate daily and weekly reports, and (via NMV - see above) various performance/utilization graphs and charts. The available collectors include ' <b>volume-collector</b> ', ' <b>nfs-collector</b> ', ' <b>network-collector</b> '. See also 'SGI NAS Runners' below.
<b>Reporter</b>	Yet another type of pluggable module tasked to generate periodic reports. The available reporters include ' <b>network-reporter</b> ', ' <b>nfs-reporter</b> ', ' <b>volume-reporter</b> ', ' <b>services-reporter</b> '. See also 'Runners' below.
<b>Indexer</b>	Indexer is a special runner that exists for a single purpose: to index a specified folder, or folders. Once a folder is indexed, it can be searched for keywords, and the search itself takes almost no time. In a way, Indexers provide functionality similar to Internet search engines (think 'Google'). However, in addition to searching the most recent raw and structured data, Indexer will allow you to search back in history - as long as there are snapshots available (that is, retained according to the <b>auto-sync/tier/snap</b> policies) to keep this history.

Term	Comment
<b>Runner</b>	Triggers, Collectors, Reporters, and Indexers - also commonly called 'Runners' - are pluggable modules that perform specific Fault Management, Performance Monitoring, Reporting, and archive Indexing tasks. All appliance's runners use the same SA-API (see above) provided by NMS (see above). The runners can be easily added – they are the source of future customizations in the product.
<b>COMSTAR</b>	Common Multiprotocol SCSI Target. In addition to providing support for the iSCSI and Fibre Channel protocols, COMSTAR addresses an overall design goal of making it possible to build a fully compliant (in the strict T10 standards sense) block level storage target.  SGI NAS can export ZFS storage as fully virtualized thin provisioned FC or iSCSI LUNs. For more information, please refer to the Section 'SCSI Target'. Support for Fibre Channel as a target is available from the optional Target FC plug-in.
<b>LUN</b>	Physical and logical drives, attached to the appliance directly or via iSCSI or FC SAN, are commonly called LUNs. The terms "LUN", "hard drive" and "disk" are used interchangeably.  See also <a href="http://en.wikipedia.org/wiki/Logical_Unit_Number">http://en.wikipedia.org/wiki/Logical_Unit_Number</a>
<b>Zvol</b>	Emulated (virtual) block device based on a given appliance's volume. Can be used as additional swap partition but the primary usage: easy iSCSI integration. Zvol is a powerful and flexible tool also because of its tight integration with the appliance's storage services. Zvol can be thin provisioned, and can be grown over time, both in terms of its effective and maximum size. Thin provisioned (also called 'sparse') zvol does not allocate its specified maximum size. At creation time thin provisioned zvol actually allocates only a minimum required to store its own metadata. You can grow both the effective (actually used) size of the zvol by storing more data on it, and the maximum size of the zvol, by incrementing its property called 'volsize'.
<b>Plug-in</b>	SGI NAS extension module that can be easily added (installed) and removed. plug-in uses the same SA-API (see above) as all the rest software components, and implements a certain well-defined (extended) functionality. At installation time, plug-in integrates itself with the appliance's core software. Many plug-ins are integrated with NMC and NMV and add new menus and commands.
<b>System checkpoint</b>	System checkpoint (or simply 'checkpoint') is a bootable snapshot of the appliance's operating system. SGI NAS provides a reliable and secure software upgrade mechanism that relies on system checkpoints.  Prior to any software upgrade, the current working root filesystem is snapshot-ed and the resulting snapshot is then converted into a bootable system checkpoint, visible via GRUB boot menu.  System checkpoint is automatically created when you upgrade the base appliance

Term	Comment
	software and/or install additional (pluggable) modules. For details on the appliance's safe and live upgrade mechanisms, please see Section ' <a href="#">Upgrades</a> '.

## 2.2 Functional Block Diagram

The following block diagram illustrates the main components of the SGI NAS management architecture. The management software includes SGI NAS Management Server and its clients: NMC, NMV, SGI NAS runners (see Section 2.1, Terminology, above), SGI NAS plug-ins, 2<sup>nd</sup> tier storage services.



## 2.3 Storage Limits

With ZFS many of the limits inherent in the design of other storage systems go away. Here is a summary of some of the key threshold limits in ZFS.

Description	Limit
Number of files in a directory	$2^{48}$
Maximum size of a file system	$2^{64}$ bytes
Maximum size of a single file	$2^{64}$ bytes
Number of snapshots of any file system	$2^{64}$
Number of file systems in a pool	$2^{64}$
Maximum pool size	$2^{78}$ bytes

## 3 NMC Overview

The SGI NAS Management Console (NMC) provides a complete set of operations for managing the storage appliance. NMC also includes wizards and the ability to record and replay commands across all deployed SGI NAS instances. Command completion is provided to guide you through the interface. Using `'help keyword'` is another way to learn the available commands.

### 3.1 Accounts

SGI NAS provides a “root” and “admin” user account. In NMC, the “root” user account has rights to perform all actions. The default passwords are “nasnas” and should be changed immediately after system installation. The passwords can be changed using the NMC command:

```
nmc:/$ setup appliance password
```

### 3.2 Command Completion

You can interchangeably use the “TAB-TAB” approach for command completion, type command names or partial command actions to enter a menu driven mode, or add “-h” as necessary to most secondary commands for full usage statements and examples.

Whichever way you use to enter commands, NMC will present a number of (completion) choices. To quickly find out the meaning of all those multiple options, type '?' and press Enter. For instance, type `show appliance`, and press TAB-TAB or Enter:

```
nmc:/$ show appliance
```

In response NMC will show a number of options - in this particular case appliance's services and facilities that can be 'shown'. Note that `<?>` is part of the `show appliance` completion set - its presence indicates availability of brief per-option summary descriptions. Next:

- type '?'

- observe brief descriptions
- decide which (completion) option to use
- repeat the sequence, if needed

### 3.3 Command Summary

The commands available in NMC are shown in the following table.

show	display any given object, setting or status
setup	create or destroy any given object; modify any given setting
query	advanced query and selection
switch	manage another SGI NAS System or a group of systems
destroy	destroy any given object: volume, folder, snapshot, storage service, etc.
create	create any given object: volume, folder, snapshot, storage service, etc.
run	execute any given runnable object, including storage services: auto-snap auto-scrub auto-sync auto-tier
share	share (via NFS, CIFS, RSYNC, FTP and WebDAV [31]) a volume or a folder.  Share zvol (Section "Terminology") via iSCSI.
unshare	unshare a volume or a folder
record	start and stop NMC recording sessions
help	SGI NAS manual pages

Of these, the primary commands are `setup` and `show`. You can run `setup usage` or `show usage` to get a comprehensive usage guide for these commands. Search the result using `'/` (forward search) and `'?` (backward search).

By running `setup`, you can see the available options.

```
nmc:/$ setup
Option ?
<?> appliance auto-scrub auto-snap auto-sync auto-tier collector
delorean diagnostics folder group inbox indexer iscsi lun mypool/
network plug-in recording reporter script-runner snapshot storagelink
trigger usage volume zvol
-----
Navigate with arrow keys (or hjkl), 'q' or Ctrl-C to quit
Summary information: short descriptions and tips
```

By running `show`, you can see the available options.

```
nmc:/$ show
Option ?
all appliance auto-scrub auto-snap auto-sync auto-tier collector
faults folder group inbox indexer iscsi lun mypool/ network
performance plug-in recording reporter script-runner scsi-target share
snapshot trigger usage version volume zvol
-----
Navigate with arrow keys (or hjkl), 'q' or Ctrl-C to quit
Appliance at a glance: show appliance's network and
```

## 3.4 Scripting

NMC is easily scriptable, and can be used to quickly create custom scripts that run periodically, on event, or on-demand.

For more information, run in the NMC:

```
nmc:/$ create script-runner -h
```

```
nmc:/$ record -h
```





## 4 NMV Overview

SGI NAS Management View (NMV) is SGI's Web-based GUI. Nearly all administrative functions can be performed using this GUI.

### 4.1 Accounts

SGI NAS provides a “root” and “admin” user account. In NMV, the “admin” user account has rights to perform all actions. The default passwords are “nasnas” and should be changed immediately after system installation. The passwords can be changed on the '**Settings**' tab under the '**Appliance**' heading.

### 4.2 Login

The default management port is 2000. Both HTTP and HTTP/s access are supported.

### 4.3 Navigation

The primary tabs in NMV are:

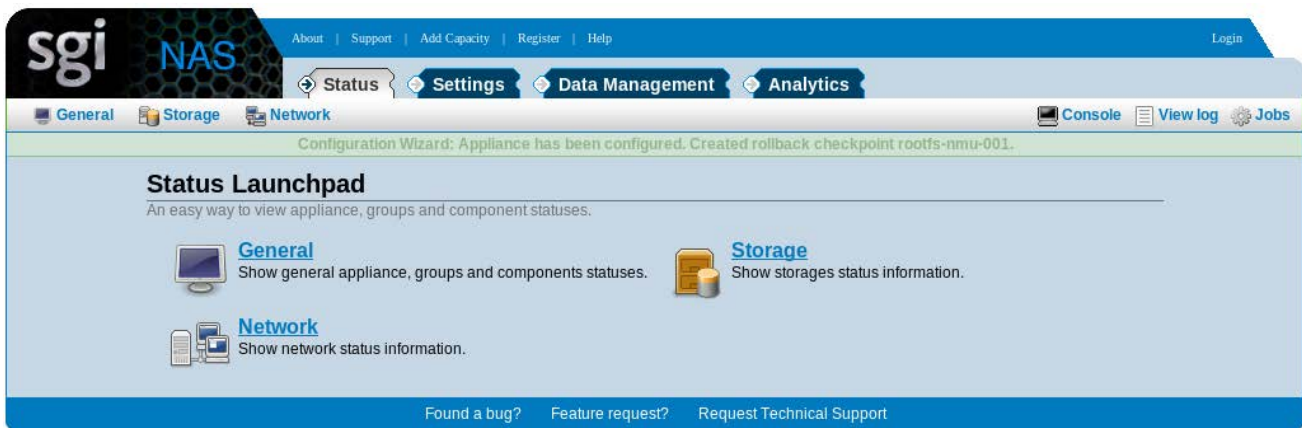
- Status
- Settings
- Data Management
- Analytics

The **Status** pages give you status on the appliance, network, and storage.

The **Settings** pages allow you to make configuration changes to the appliance.

The **Data Management** pages allow you to administer data volumes and folders.

The **Analytics** pages allow you to see storage and network performance trends over time.



### 4.4 Terminal Access

Note that you can access NMC from within the Web browser by clicking the Console icon.

### 4.5 View Log

The results of administrative actions are shown on a status bar at the top of the page. However it quickly disappears. To see the results of actions that have taken place in this Web session you can click on 'View log' in the upper right corner of the GUI.

## 5 Initial Setup

During the installation process you register your software on the Web and receive a license in email. Next you enter some basic network configuration information such as the default gateway to be used.

After the network configuration is setup the Web server can start. To connect to NMV from a Web browser enter the configured network address in the browser and use port 2000.

Once connected to the Web server for the first time, the installation wizard will lead you through a few basic installation steps.

### 5.1 Installation wizard I

Step 1. Provide some basic information about the appliance, such as host and domain name, timezone, NTP Server, keyboard layout and language.

The screenshot displays the SGI NAS installation wizard interface. The top left corner features the 'sgi NAS' logo. A 'View log' link is located in the top right corner. The main content area is titled 'STEP #1: BASIC CONFIGURATION' and contains several configuration fields:

- Host Name:** A text input field containing 'myhost'. Below it, a note states: 'Default appliance's host name. It must begin and end with ASCII letter or digit, and can only contain ASCII letters, digits, and hyphens (-)'.
- Domain Name:** A text input field containing 'mydomain.com'. Below it, a note states: 'Default appliance's domain name. Domain name is a group of labels delimited by dots (.). Labels can only contain ASCII letters, digits, underscores (\_), and hyphens (-). Each label must not have more than 63 characters and must not begin or end with a hyphen (-). It is desirable for the domain name to contain a Top Level Domain (.com, .net, etc.), but you can use any domain name.'
- Time Zone:** A dropdown menu with 'America' selected, and a secondary dropdown menu with 'United States - US/Pacific' selected. Below it, a note states: 'Geographical time zone for selected continent and country.'
- NTP Server:** A text input field containing 'pool.ntp.org'. Below it, a note states: 'Default NTP server address.'
- Keyboard Layout:** A dropdown menu with 'US-English' selected. Below it, a note states: 'Default keyboard layout.'
- Language:** A dropdown menu with 'English' selected. Below it, a note states: 'User Interface Language'

On the left side of the interface, there is a sidebar with the following menu items: 'Basic Configuration', 'Admin Passwords', 'Notification System', and 'Save Configuration?'. The 'Basic Configuration' item is currently selected. At the bottom of the main configuration area, there is a 'Next' button.



GMT, UTC, UCT, Zulu, Universal timezones are specified by selecting 'Etc' in left drop-down 'Time Zone' list and choosing required in the right box.

Step 2. Specify root and admin user passwords.

The screenshot shows the SGI NAS configuration interface for Step 2: ADMIN PASSWORDS. The left sidebar contains a navigation menu with 'Admin Passwords' selected. The main content area has four password fields: 'Password for root', 'Repeat root password', 'Password for admin', and 'Repeat admin password'. Each field is followed by a small instruction: 'Change root password.', 'Re-enter root password.', 'Change admin password.', and 'Re-enter admin password.' respectively. At the bottom, there are 'Back' and 'Next' buttons.

Step 3. On the next screen provide notification information. Specify SMTP server information to enable automatic issue reporting to SGI Support, requesting additional capacity, etc.

The screenshot shows the SGI NAS configuration interface for Step 3: NOTIFICATION SYSTEM. The left sidebar has 'Notification System' selected. The main content area contains several fields for SMTP configuration: 'SMTP Server' (localhost), 'SMTP User', 'SMTP Password', 'SMTP Send Timeout' (30), 'SMTP Authentication' (Plain), 'From E-Mail Address' (myhost-noreply@mydomain.com), 'E-Mail Addresses' (root@localhost), 'E-Mail Addresses for faults', and 'E-Mail Addresses for statistics'. Each field has a small explanatory text below it. At the bottom, there are 'Back', 'Test', and 'Next' buttons.

You can specify separate e-mail addresses for statistics (Volumes reports, Network Statistics, NFS Statistics), notifications (system reports with status 'NOTICE' and 'INFO') and faults (system reports with status 'WARNING' and 'CRITICAL').

After completing the installation steps, you'll be asked to confirm the settings and save the configuration.

After saving the configuration you are taken to a second installation wizard.

Setting	Value
<b>Basic Configuration (6 items)</b>	
Host Name	myhost
Domain Name	mydomain.com
Time Zone	US/Pacific
NTP Server	pool.ntp.org
Keyboard Layout	US-English
Language	English
<b>Admin Passwords (2 items)</b>	
Password for root	*****
Password for admin	*****
<b>Notification System (9 items)</b>	
SMTP Server	localhost
SMTP User	
SMTP Password	*****
SMTP Send Timeout	30
SMTP Authentication	Plain
From E-Mail Address	myhost-noreply@mydomain.com
E-Mail Addresses	root@localhost
E-Mail Addresses for faults	
E-Mail Addresses for statistics	

\* Values marked in bold have been changed during Wizard 1

Back Save

## 5.2 Installation wizard II

Step 1. Here you can add or delete network interfaces, setup primary interface, aggregate interfaces, configure, setup default gateway and name servers.

**Network**

Please verify networking configuration. Primary Interface must be configured; it will be used for management access to this storage node. You may also want to create separate aggregated interface(s) for large data transfers.

ISCSI Initiator

Disks

Volumes

Folders and Shares

Review Changes and Exit

**STEP #1: NETWORK**

**Network Interfaces:**

Interface	Type	Configuration	Primary	Actions
e1000g0	physical	Using DHCP as 172.16.157.150/255.255.255.0	<input checked="" type="radio"/>	

**Operations:**

**Add Interface**

Interface Type:  (Type of new interface: single or aggregated.)

All Available Devices:

Available capable interfaces. Select just one for single type or two or more for aggregation. To select/deselect items use CTRL key.

**Edit Interface**

Configured Interfaces:  (All available configured interfaces. For aggregation select two or more interfaces.)

Configuration Method:  (Network Interface configuration method; static or dynamic (via DHCP).)

**Change Default Gateway**

Default Gateway:  (Default networking gateway IPv4 address in a dot-decimal notation (#.#.#.#).)

**Change Name Servers**

Name Server 1:  (Primary naming server IPv4 address in a dot-decimal notation (#.#.#.#).)

Name Server 2:  (Secondary naming server IPv4 address in a dot-decimal notation (#.#.#.#).)

Name Server 3:  (Additional naming server IPv4 address in a dot-decimal notation (#.#.#.#).)

Step 2. Configure iSCSI Initiator to use virtual disks exported via VMware or another iSCSI target. The appliance supports all 3 types of iSCSI discovery.

**STEP #2: ISCSI INITIATOR**

**ISCSI Initiator parameters:**

**Initiator Name**   
iSCSI initiator node name. Maximum of 223 characters.

**Initiator Alias**   
iSCSI initiator node alias. Maximum length of 223 characters.

**Authentication Method**   
Authentication mode: none or CHAP.

**Number of Sessions**   
The number of configured iSCSI sessions that will be created for each iSCSI target to utilize I/O multipathing feature.

**Header Digest Method**   
Enable or disable CRC32 check for SCSI packet headers (may affect performance).

**Data Digest Method**   
Enable or disable CRC32 check for SCSI data transfers (may affect performance).

**RADIUS Server Access**   
Enable or disable RADIUS server to access and verify authorization.

**Configured ISCSI discovery methods:**

Parameter	Type	Enabled	Delete

**Additional ISCSI discovery method:**

**ISCSI Discovery Method**   
Method to discover new iSCSI targets: SendTargets, static address or iSNS server.

**ISCSI Target IP Address**   
iSCSI Target IPv4 address and optional port number in form #.#.#:[port].

Step 3. Allows you to review the available disks. If any new disks were added, click on refresh and they will appear in the list:

The screenshot shows the SGI NAS web interface at the 'STEP #3: DISKS' configuration stage. The interface has a blue header with the 'sgi NAS' logo and a 'View log' link. A left sidebar contains navigation tabs: Network, iSCSI Initiator, Disks, Volumes, Folders and Shares, and Review Changes and Exit. The 'Disks' section is active, showing a table of available disks. Below the table is a 'Refresh...' button and navigation buttons for '<< Previous Step' and 'Next Step >>'. The table lists five disks with their respective device names, types, sizes, and models.

Disk	Device	Type	Size	Volume	Attach	Model
c0t1d0	sd2	disk	68.49 GB		mpt	FUJITSU, Rev. 0104
c0t2d0	sd1	disk	232.89 GB	syspool	mpt	Hitachi, Rev. V5DOA73A
c0t3d0	sd3	disk	68.49 GB		mpt	FUJITSU, Rev. 0104
c0t4d0	sd4	disk	68.49 GB		mpt	FUJITSU, Rev. 0104
c0t5d0	sd5	disk	68.49 GB		mpt	FUJITSU, Rev. 0104



Step 4. Next, the Wizard will help you to create or import data volumes:

**sgl NAS**

View log

**STEP #4: VOLUMES**

**Volumes:**

No available volumes.

**Operations:**

**Add New Volume**

**Performance considerations**

For mirrored configurations:  
 Random read performance scales linearly with the number of disks; write performance scales linearly with the number of mirror sets.  
 Sequential read throughput scales linearly with the number of disks; write throughput scales linearly with the number of mirror sets.

For parity (RAID-Z, RAID-Z2) configurations:  
 Random read and write performance scales linearly with the number of RAID sets.  
 Sequential read and write throughput scales linearly with the number of data (non-parity) disks.  
**Caution!** It is NOT recommended to use non-redundant device configuration within a ZFS volume

**Volume Configuration:**

Available Disks	Redundancy Type	Final Volume Configuration
<ul style="list-style-type: none"> <li>c0t1d0 : mpt(disk) : 68.49 GB</li> <li>c0t3d0 : mpt(disk) : 68.49 GB</li> <li>c0t4d0 : mpt(disk) : 68.49 GB</li> <li>c0t5d0 : mpt(disk) : 68.49 GB</li> </ul>	<p>▼</p> <p>Add to pool &gt;&gt;</p> <p>Add to spare &gt;&gt;</p> <p>Add to log &gt;&gt;</p> <p>Add to cache &gt;&gt;</p> <p>Add to selected &gt;&gt;</p> <p>&lt;&lt; Remove selected</p> <p>&lt;&lt; Remove all</p>	<div style="background-color: #eee; height: 100px;"></div>
Available physical and logical (virtual) disks. Mounted/Slices/Partitions are not allowed.	Add or Remove disks and groups	This configuration will be used to create volume.

**Volume Properties:**

**Name**

Volume name must begin with a letter and can only contain alphanumeric characters (a-z, A-Z, 0-9) in addition to the following three special characters: underscore (\_), hyphen (-) and period (.). Volume name has the following restrictions: the beginning sequence c[0-9] is not allowed; the name log is reserved; a name that begins with mirror, raidz, or spare is not allowed because these name are reserved. In addition, volume name must not contain a percent sign (%).

**Description**

Optional volume description. Maximum length is 255 characters.

**Deduplication**

Controls the deduplication option for the volume. If enabled, it will optimize use of duplicate copies of data. Default is off.

**Compression**

Controls the compression algorithm used for this dataset. Default is "on". Setting compression to "on" uses the lzjb compression algorithm. The lzjb compression algorithm is optimized for performance while providing decent data compression. Currently, "gzip" is equivalent to "gzip-6".

**Autoexpand**

Controls automatic pool expansion when the underlying LUN is grown.

**Sync**

Controls synchronous requests (standard - ensure all synchronous requests are written to stable storage; always - every file system transaction will be written and flushed to stable storage by system call return; disabled - synchronous requests are disabled). Default is standard.

**Force creation**

Forces use of (virtual or physical) disks (LUNs) even if they appear to be in use.

**Import Volume**

No exported or destroyed volumes found.

## Step 5. Create folders.

**sg NAS**

View log

**STEP #5: FOLDERS AND SHARES**

**Folders:**

No available folders.

**Operations:**

**Create New Folder**

**Volume:** vol1  
Folder's volume.

**Folder Name:** fol1  
Each folder pathname's component delimited by backslash (?) can only contain alphanumeric characters (a-z, A-Z, 0-9) in addition to the following three special characters: underscore (\_), hyphen (-) and period (.). Folder pathname must begin with an alphanumeric character and not contain a percent sign (%).

**Description:**  
Human-readable description for this folder.

**Record Size:** 128K  
Specifies a suggested block size for files in the folder. Default is 128K.

**Log Bias:** latency  
Provide a hint to ZFS about handling of synchronous requests in this dataset. If logbias is set to latency (the default), ZFS will use pool log devices (if configured) to handle the requests at low latency. If logbias is set to throughput, ZFS will not use configured pool log devices. ZFS will instead optimize synchronous operations for global pool throughput and efficient use of resources.

**Deduplication:** off  
Controls the deduplication option for this dataset. If enabled, it will optimize use of duplicate copies of data. Default is "off".

**Compression:** on  
Controls the compression algorithm used for this dataset. Default is "on". Setting compression to "on" uses the lzjb compression algorithm. The lzjb compression algorithm is optimized for performance while providing decent data compression. Currently, "gzip" is equivalent to "gzip-6".

**Number of Copies:** 1  
Controls the number of copies of data stored for this dataset. Default is "1".

**Case Sensitivity:** mixed  
Indicates whether the file name matching algorithm used by the file system should be case-sensitive, case-insensitive, or allow a combination of both styles of matching. Use "mixed" if the folder is planned to be shared via CIFS and NFS at the same time. Default is "mixed".

**Unicode Only:** off  
Enable it if you want to exclude non-Unicode file names creation for this folder. If set, this option will ensure better inter-client operability. Make sure this option is enabled if you are planning to use this folder as a share for MacOS X, Windows and Linux clients. Default is "off".

**Sync:** standard  
Controls synchronous requests (standard - ensure all synchronous requests are written to stable storage; always - every file system transaction will be written and flushed to stable storage by system call return; disabled - synchronous requests are disabled). Default is standard.

Create

<< Previous Step    Next Step >>

While walking Wizard-guided steps, pay attention to system notices. When creating a volume, please keep in mind that redundant configurations typically improve performance and reliability but reduce effective storage capacity.

Step 6. Finally, with its last screen the Wizard invites to review and save all changes:

**STEP #6: REVIEW CHANGES AND EXIT**

**Network**

Interface	Type	Configuration
e1000g0	physical	Configured as 150.166.43.141/255.255.255.0 with mtu 1500
e1000g1	physical	Unconfigured

**Disks**

Disk	Device	Type	Size	Volume	Attach	Model
c0t1d0	sd2	disk	68.49 GB	vol1	mpt	FUJITSU, Rev. 0104
c0t2d0	sd1	disk	232.89 GB	syspool	mpt	Hitachi, Rev. V5DOA73A
c0t3d0	sd3	disk	68.49 GB	vol1	mpt	FUJITSU, Rev. 0104
c0t4d0	sd4	disk	68.49 GB		mpt	FUJITSU, Rev. 0104
c0t5d0	sd5	disk	68.49 GB		mpt	FUJITSU, Rev. 0104

**Volumes**

Volume	Configuration	Size	Allocated	Free	Capacity	Dedup Ratio	State
vol1	raidz1 group: 1, devices: 2	136.00 GB	356.00 KB	136.00 GB	0%	1.00x	ONLINE

**Folders and Shares**

Folder	Refer	Used	Avail	CIFS	NFS	FTP	RSYNC	WebDAV	Index
vol1/fo11	31.00 KB	31.00 KB	66.90 GB	-	-	-	-	-	-

Results 1 - 1 (all)

**Warning:** Optimize appliance's I/O performance by disabling ZFS cache flushing. While providing a considerable performance improvement in certain scenarios (in particular those involving CIFS, NFS or iSCSI) - this settings may be unsafe, in terms of application-level data integrity. It is strongly recommended to use this feature if and only if your storage is NVRAM protected, and the hardware platform is connected to Uninterrupted Power Supply (UPS). Default setting: unchecked (disabled).

Optimize I/O performance for CIFS/NFS/iSCSI UPS-backed deployments?

Create periodic scrubbing service (auto-scrub) for system volume weekly, every Sunday at 3am ?

<< Previous Step    Start NMV

The Wizard will also recommend to create system checkpoint (see above) – “a snapshot of the freshly installed and initially configured appliance”.



## 6 Managing Data Volumes

This section describes the administration of data volumes. SGI NAS allows you to first aggregate your available disks into data volumes, and then to allocate file or block-based storage from the data volume. The data volume provides a storage pooling capability, so that file systems or blocks can have room to expand without being over-provisioned.

### 6.1 Data Redundancy

The data volume provides redundancy capabilities similar in concept to the RAID features of other storage systems. Redundancy options are: none, mirrored, RAID-Z1 (single parity), RAID-Z2 (double parity), and RAID-Z3 (triple parity). It is recommended that you always choose some form of redundancy for your pool.

The redundancy options in SGI NAS may sound familiar to other standard RAID options, but there are some important differences. For example, SGI NAS always relies on checksums to determine, if data is valid instead of assuming that devices will report an error on the read quest.

For RAID-1 the assumption is that either side of the mirror is equally current and correct. With mirroring in SGI NAS, checksums always validate the data and in the event of conflicts the most recent data with a valid checksum is used.

With RAID-5, if the data being written is smaller than the stripe width then multiple I/O operations are needed (read the data, modify it, write it). With SGI NAS RAID-Z1, all writes are full stripe writes. This helps to ensure, that data is always consistent on disk (even in the event of power failures, etc.).

When multiple redundancy groups are in the data volume, SGI NAS will dynamically stripe writes across them. However unlike RAID-0 stripes, the disks participating in the write are dynamically determined and there are no fixed length sequences.

### **A Note On Redundant Configurations**

A mirrored volume pool consists of matched drives or drive groups, where by data always has a redundant copy on the mirrored set of disks. Mirroring can make use of other pooled technologies such as parity, allowing multiple groups of disks to be setup each with one primary array and one secondary, mirrored array. In most cases, for best reliability and performances, administrators would setup a combined or striped set of mirrored devices (sometimes referred to as **RAID 10**). In the case of two-way mirrors RAID 10 will halve your overall storage capacity, but will provide the best read/write performance, as reads are striped across all of the primary disks, and writes only require a single duplication of each write to a secondary drive. At any time, any number of failed drives are permitted, as long as no two drives in a paired set fail at the same time.

Parity based RAID volumes make use of one or two dedicated drives to maximize capacity without reducing redundancy of stored data. Each write is committed across all drives in a group, including the parity devices, and they further take some penalty in calculating the parity. The reverse is equally true, as reads must combine data and parity across all devices in a group. To improve performance, it is generally recommended to also stripe multiple parity based RAID groups together to allow parallel reads/writes to the disk. This is commonly referred to as RAID 50. Up to one drive in a RAIDZ1 group, or two drives in a RAIDZ2 group can fail at a time without losing data. In the RAID50 setup, you both allow for future expansion with new parity groups, as well as allow for more drive failures, limited still to at most two per group.

In both mirrored and parity based RAID volumes, you should establish multiple spare devices equal to the size of each member drive. Redundant, striped arrays of either variety, with sufficient spare disks, allow one to achieve the greatest level of reliability on commodity disks. As disk capacity grows and gets ever cheaper, you can expand on these striped volumes. The ZFS based filesystem allows for continuous volume growth, but consistent disk group sizing across a striped array. is recommended. Therefore, as disk sizes increase, it is considered a good practice to create disk sub-groups of as close to an equal size as possible.

Redundant configurations improve not only reliability of your SGI NAS system but performance as well. For mirrored configurations:

- Random reads scale linearly with the number of disks; writes scale linearly with the number of mirror sets.
- Read throughput scales linearly with the number of disks; write throughput scales linearly with the number of mirror sets.

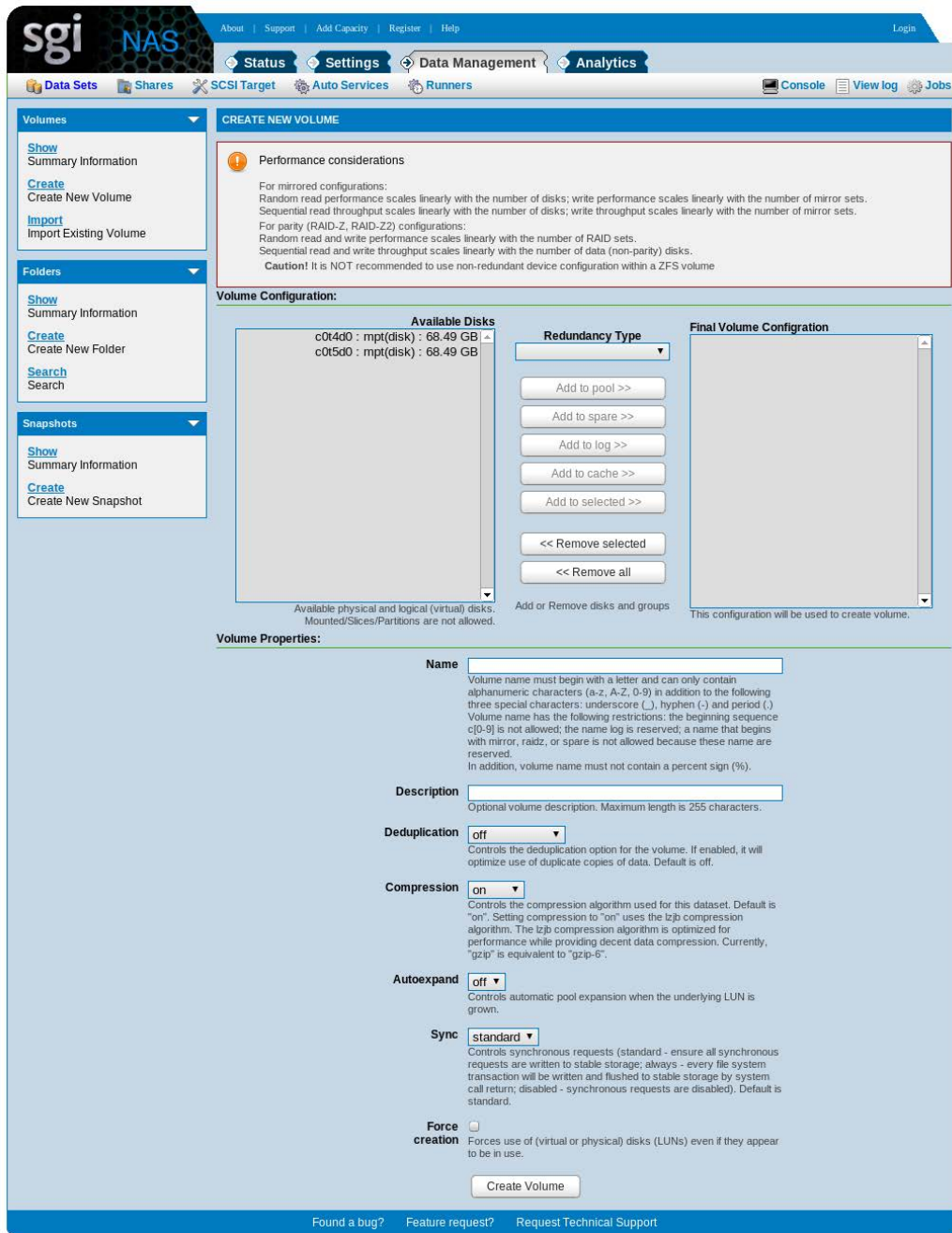
For parity (RAID-Z, RAID-Z2) configurations:

- Random I/O reads and writes scale linearly with the number of RAID sets.
- Sequential I/O throughput scales linearly with the number of data (non-parity) disk

## **6.2 Create Data Volume**

Data volumes are managed from the **Data Management** tab in NMV. When the **Data Management** tab is selected there is a link to **Data Sets**. From this page you can create a volume. Here is an example of the screen where a volume can be created.

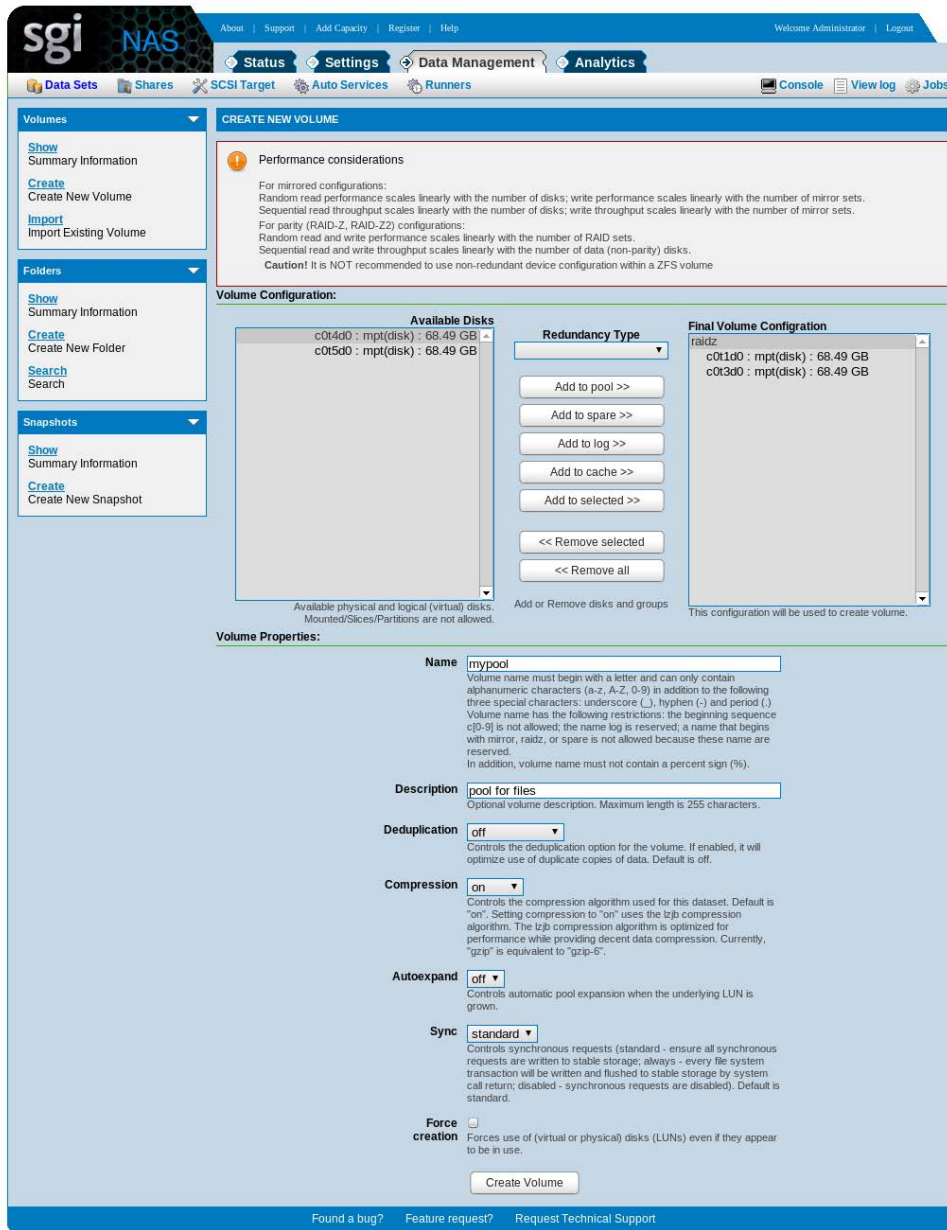




The redundancy options include mirroring, RAID-Z1 (single parity), RAID-Z2 (double parity), and RAID-Z3 (triple parity).

All disks that are not already contained in data volumes are shown. Note that this may include disks mapped to the SGI NAS system from other storage systems.

Here is an example of selecting two available disks for a RAID-Z1 configuration.



In addition to choosing the disks, you can also specify the volume name and various properties such as a description, de-duplication, and auto-expand.

### 6.2.1 De-duplication

De-duplication is a technique for increasing the effective storage capacity within a data volume. Data is examined, when it is being written to non-volatile storage. Hashes of the data blocks are compared to entries in the de-duplication table and if there are matches then the



existing data block's reference count is incremented instead of creating a new data block.

The following de-duplication options are available:

- SHA-256
- SHA-256 verify
- Off

De-duplication can save storage capacity and I/O bandwidth, but it will also increase latency. To minimize the performance impact, make sure that the de-duplication table fits in RAM. To estimate the size you can use this formula:

$$( (\text{Size of pool} / \text{average block size}) * (270 \text{ bytes}) ) / \text{estimated dedup ratio}$$

Since there is a performance penalty with de-duplication it is off by default. Turn this option on, if you think you will have a lot of redundant data blocks in the pool. This can be true in virtualized environments or backened storage for email systems.

When de-duplication is turned on, the SHA-256 algorithm (a cryptographic hash algorithm from NIST) is used.

There is a one in  $2^{256}$  chance that SHA-256 will report a hash match even though the two blocks being compared are not the same. To ensure this is not an issue, you can include verify as an option which will read the data blocks after a hash match to ensure the blocks are the same.

The de-duplication is also available when using the auto-synch service. See the section on "Asynchronous Replication" for details.

### 6.2.2 Auto-expand

Auto-expand will automatically try to expand the size of the data volume when a new disk is added. This is another option that is off by default. One reason for this is to ensure that spare devices don't unexpectedly increase the size of your data volume when they are temporarily

activated in response to a disk failure. Once the volume size is expanded it can't be shrunk.

### 6.3 Creating various RAID configurations

When you create a data volume you choose a redundancy type such as RAID-Z1 for each group of disks in the volume. There is a penalty for putting too many disks in a RAID-Z1 (or – Z2 or –Z3) group such as slow re-silvering times. For larger volume sizes you would instead split the disks between multiple redundancy groups. SGI NAS will then essentially stripe writes across the redundancy groups.

Here is an example using NMC to create something similar to a RAID50 configuration. Make sure you have at least six available disks. Assume their names are: c1t1d0 c1t2d0 c1t3d0 c1t4d0 c1t5d0 c1t6d0. In NMC:

```
nmc:/$ setup volume create my-notexactly-raid50
Group of devices: c1t1d0, c1t2d0, c1t3d0
Group redundancy type: raidz1
```

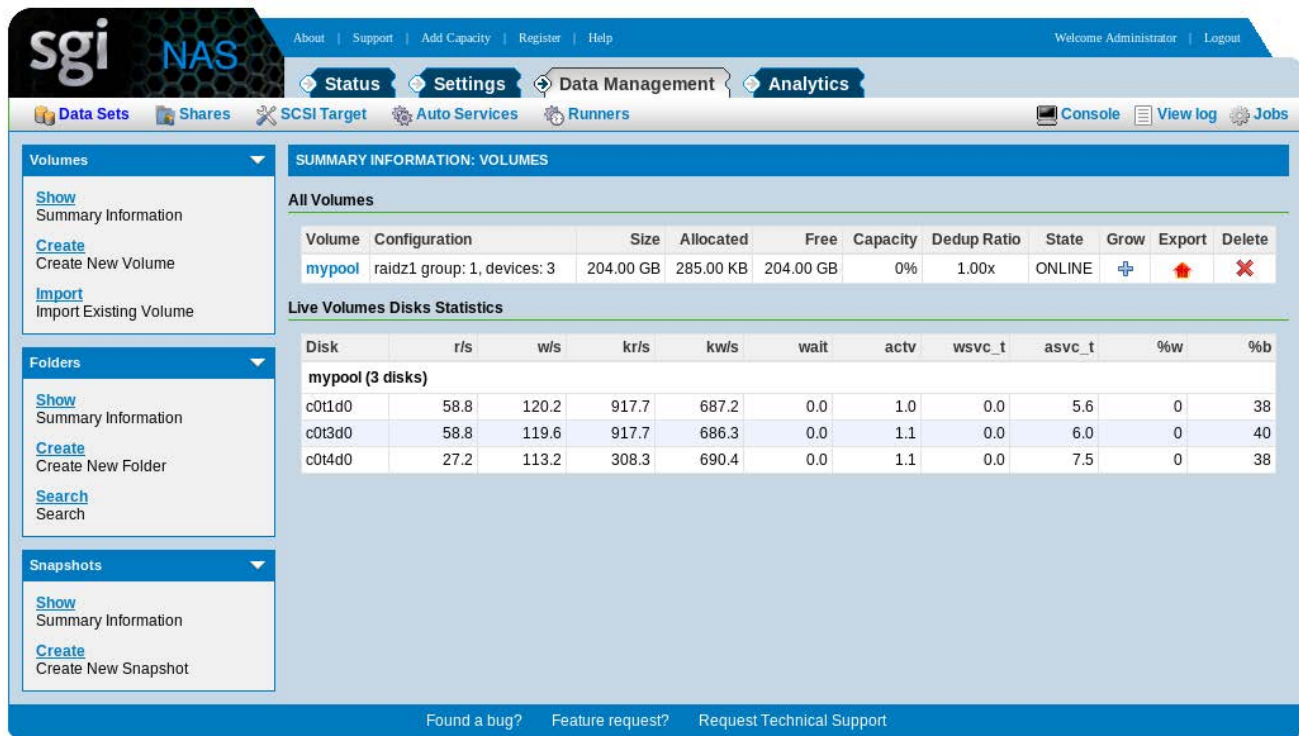
You are then asked 'Continue adding devices to the volume 'my-notexactly-raid50'?

Type 'y'.

```
Group of devices: c1t4d0, c1t5d0, c1t6d0
Group redundancy type: raidz1
Create volume 'my-notexactly-raid50'? y
```

### 6.4 View Status

You can view the status of all of your data volumes by selecting the Show link under the Volumes heading. Here is a sample screen:



From this view you can take actions on the volume such as expanding it, exporting it, deleting it, or editing its properties.

In NMC you can get the status of the data volume using `show volume volumename status`. This will list the status of each device in the pool, and any I/O or checksum errors that have occurred.

## 6.5 Edit Properties

There are a few data volume properties that can be edited after the volume is created. To edit the data volume properties, click on the name of the data volume in the summary view. In particular you can change the de-duplication, sync and auto-expand properties. Two of them are already described earlier in this section. Sync option is described below.

[About](#) | [Support](#) | [Add Capacity](#) | [Register](#) | [Help](#)

[Welcome Administrator](#) | [Logout](#)

Status
Settings
Data Management
Analytics

Data Sets
 Shares
 SCSI Target
 Auto Services
 Runners

Console
 View log
 Jobs

Volumes

[Show](#)  
Summary Information  
[Create](#)  
Create New Volume  
[Import](#)  
Import Existing Volume

VOLUME: MYPOOL

**Disks:**

Disk	Size	Status	Errors (R/W/C)
<b>raidz1-0 (3 devices)</b>			
c0t1d0	68.49 GB	ONLINE	0 / 0 / 0
c0t3d0	68.49 GB	ONLINE	0 / 0 / 0
c0t4d0	68.49 GB	ONLINE	0 / 0 / 0

**Hot Spares (only in available state):**

Disk	Size	Volume
c0t5d0	68.49 GB	

**Configurable Properties:**

**Description**   
Optional volume description. Maximum length is 255 characters.

**Deduplication**  Controls the deduplication option for the volume. If enabled, it will optimize use of duplicate copies of data. Default is off.

**Compression**  Controls the compression algorithm used for this dataset. Default is "on". Setting compression to "on" uses the lzjb compression algorithm. The lzjb compression algorithm is optimized for performance while providing decent data compression. Currently, "gzip" is equivalent to "gzip-6".

**Autoexpand**  Controls automatic pool expansion when the underlying LUN is grown.

**Sync**  Controls synchronous requests (standard - ensure all synchronous requests are written to stable storage; always - every file system transaction will be written and flushed to stable storage by system call return; disabled - synchronous requests are disabled). Default is standard.

**Static Properties:**

Property	Value
allocated	192K
altroot	-
autoexpand	off
autoreplace	on
available	134G
bootfs	-
cachefile	-
capacity	0%
dedupditto	0
dedupratio	1.00x
delegation	on
failmode	continue
free	204G
guid	14461877147504694059
health	ONLINE
info	raidz1 group: 1, devices: 3
name	mypool
size	204G
sync	standard
used	128K
version	28

[Found a bug?](#)   [Feature request?](#)   [Request Technical Support](#)

In NMC you can see all volume properties using the command:

```
nmc:/$ show volume <volumename> property
```

To change a property use:

```
nmc:/$ setup volume <volumename> property
show          devices          quota          sync
aclinherit    exec            readonly       userquota
atime         groupquota     recordsize    version
autoreplace   logbias        refquota      vscan
canmount      mslslabel      reservation    xattr
checksum      nbmand         reservation    <Enter>
compression   nms:description secondarycache <help|-h>
copies        nms:ustatus    setuid
dedup         primarycache   snapdir
```



Starting from v3.0.4 a new property which is worth to be specifically mentioned is added. '**Sync**' property provides the possibility to manage synchronous requests per filesystem. It is specially intended to be used, when synchronous requests are need to be disabled in order to increase the performance. However, it should be used very carefully. There are three possible statuses of the property:

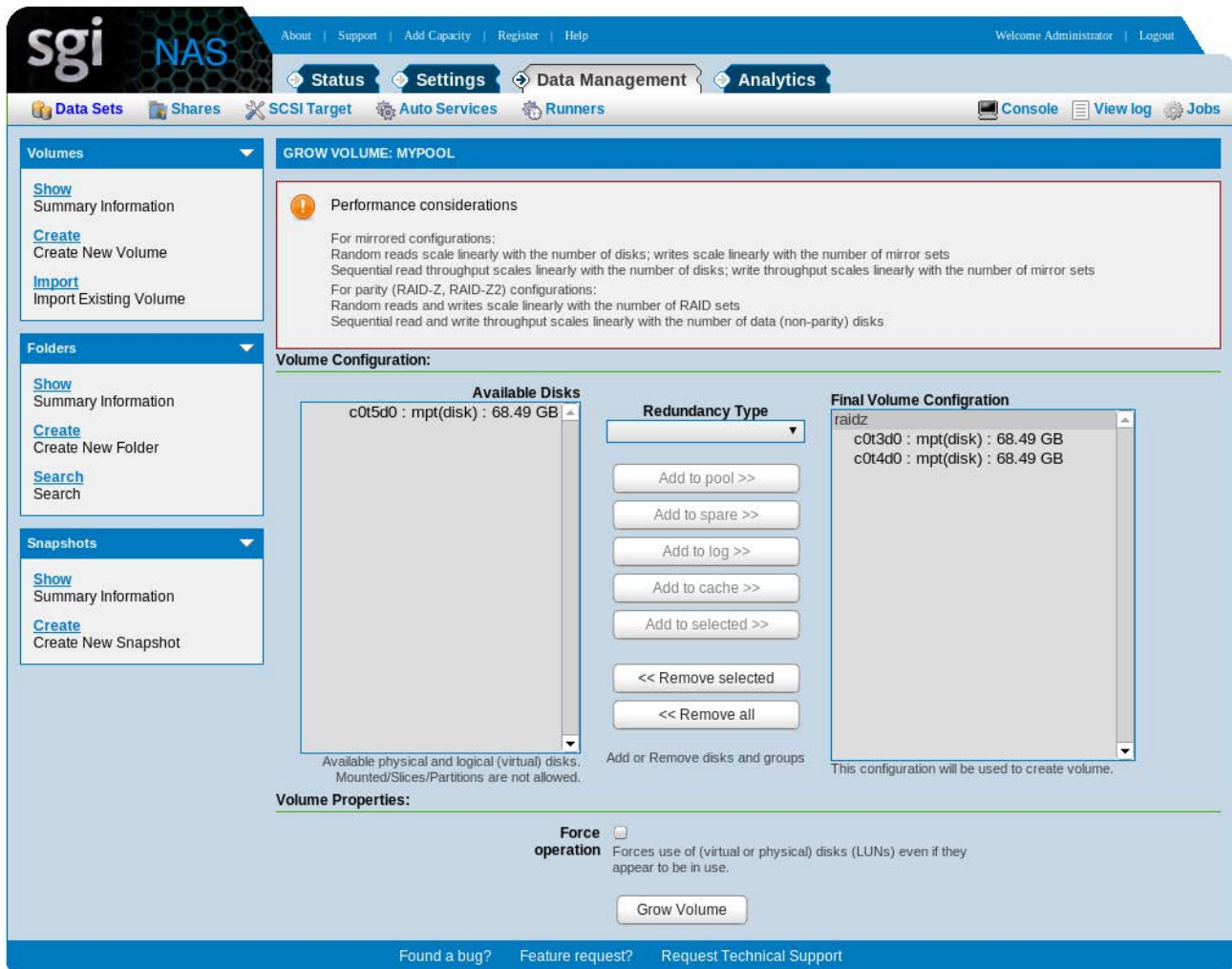
**sync=standard** Synchronous file system transactions are written out to ZIL and then all devices written are flushed to ensure the data is stable. This status is set by default.

**sync=always** Every file system transaction would be written and flushed to stable storage by system call return. Setting up this option decrease performance dramatically and is recommended to be used only in case additional precautions are required.

**sync=disabled** Disable all synchronous requests. It means, that filesystem transactions are committed to the stable storage only on the next data manipulation unit transaction, which may take long period of time. It results in a great performance improvement, but it's strictly not recommended in terms of data safety.

## 6.6 Expand Data Volume

You can expand a data volume by selecting “Grow” on the summary page. You will see a page similar to the view when creating a volume. You are adding a new disk group at this point. Select the redundancy level, select the available disks to use, and add them to the pool. Here is an example of adding a second RAID-Z1 group to a pool to create a configuration roughly similar to RAID 5:



After selecting “Grow Volume” the disk group is added to the data volume.

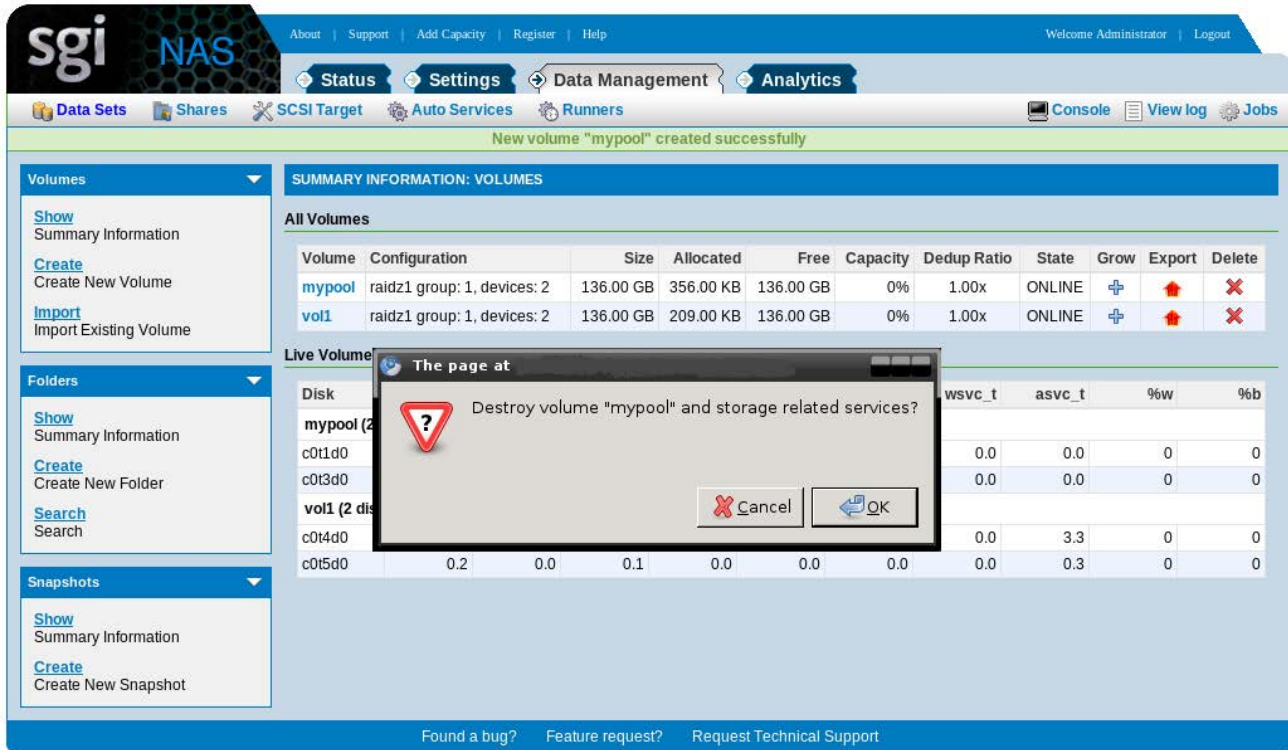
After adding this new redundancy group to the data volume, SGI NAS will favor writing to this newer group. The goal is to balance the writes across all the redundancy groups over time.

To expand a volume in NMC, use the command :

```
nmc:/$ setup volume <volumename> grow
```

## 6.7 Destroy Data Volume

A volume can be deleted by selecting the delete icon on the summary page. A dialog box will appear to confirm the request before the volume is actually destroyed.



Note that you will lose all the data in the volume at this point so make sure it is what you want to do!

Here is an example destroying a volume named "myvolume" using NMC:

```
nmc:/$ destroy volume <myvolume>
```

## 6.8 Export/Import Data Volumes

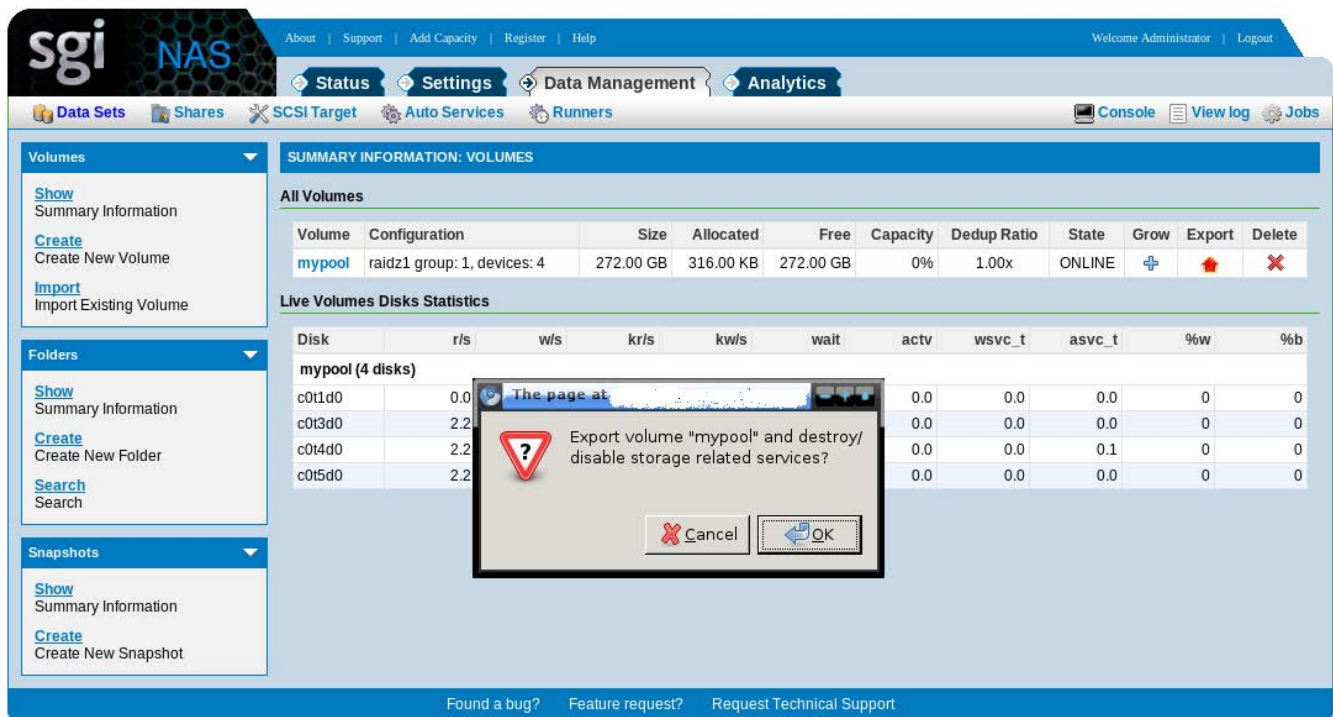
If you are going to perform a system or software upgrade, you should consider exporting the volumes first. Exporting the volume protects the underlying physical drives from I/O activity.



### 6.8.1 Export

The export will unmount any datasets in the volume. The volume meta-data is persistent. After the export, you can then import the volume into a new system and any datasets and ZFS configuration will be restored.

In NMV, the export option can be found on the summary page. A dialog box will appear to confirm your decision to export the volume, as shown below.



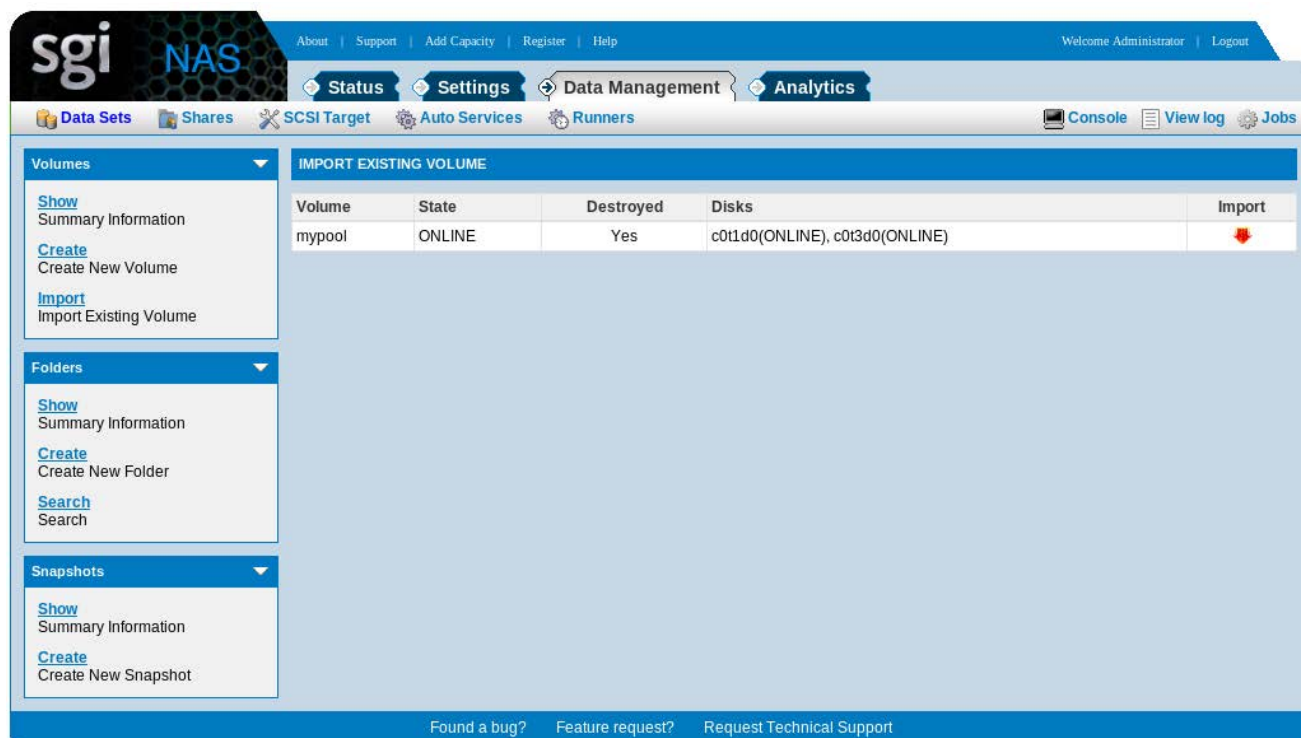
You can perform an export in NMC with the command:

```
nmc:/$ setup volume <volumename> export
```

### 6.8.2 Import

By default, SGI NAS will import existing accessible data volumes when a system starts. You can also import data volumes manually.





The import option is available under the Volumes heading. Selecting “Import” will show the volumes that can be imported.

The syntax for setup volume import in NMC is:

```
nmc:/$ setup volume import [-D] [-f] [-s] [vol-name] [new-name]
```

'**vol-name**' is the name of the exported or destroyed volume.

You can use new-name to provide a new name for the imported volume so that it won't conflict with any existing volumes.

The '**-D**' option is needed to import a destroyed volume.

The '**-f**' option forces the import, even if the system thinks the volume is already active.

The '**-s**' option applies the default auto-snap snapshot policy to the imported volume.

The NMC command `show volume import` will show data volumes that can be imported.

Volume names are shown along with their GUID. The volume's globally unique identifier may need to be used, for example, if two volumes have been exported or destroyed that used the same name. In this case the syntax would be:

```
nmc:/$ setup volume import
```

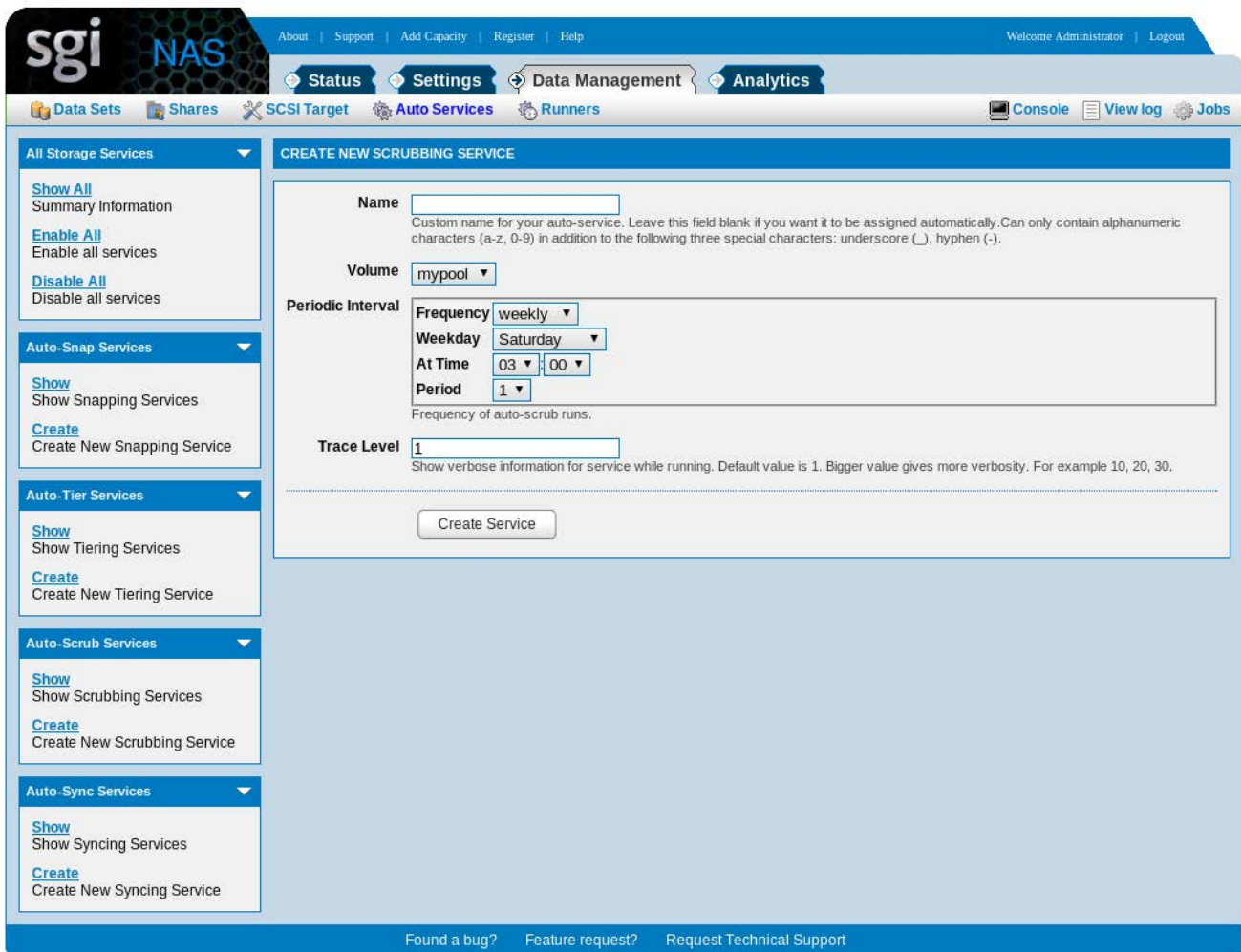
```
myvol:380744323214575787.
```

To recover properly from failures or unclean shutdowns, import will replay any transactions in the ZFS Intent Log (ZIL). This occurs for regular or forced imports. If a separate ZIL is being used and it is unavailable, then the import will fail. Be sure to use a mirrored ZIL if using a separate log device to protect against this scenario.

## 6.9 Scrub

SGI NAS can periodically check the contents of the data volume. Scrubbing the data volume will read the data blocks checking for errors. If there is redundancy in the pool configuration then SGI NAS can correct any errors it finds.

To enable periodic scrubs for a data volume, go to **Data Management** → **Auto Services**. Choose [Create](#) under the '**Auto –Scrub Services**' heading and you will see the following screen:



Choose an existing data volume from the pull-down list and define scrub schedule. Note that scrubbing is resource-intensive, so it is preferable to perform it during the maintenance time

window, if it's possible.

Corresponding NMC command:

```
nmc:/$ setup auto-scrub create
```

This command will create a scheduled auto-scrub for all appliance's volumes.

If you need to set up scrubbing service for a specific volume, run:

```
nmc:/$ create auto-scrub volume
```

and choose a corresponding volume.

## 6.10 Emergency reserve of volume space

SGI NAS allows to make a reserve volume space in order to solve out-of space issues.

It is done by NMS property 'autoreserve\_space' which is enabled by default and volume-check runner. 'autoreserve\_space' property is managed by the following NMC command:

```
nmc:/$ setup appliance nms property autoreserve_space
```

**Caution:** It is strongly recommended to keep `autoreserve_space` option enabled.

To free reserved space, run:

```
nmc:/$ setup volume <volname> free-reserve
```

This functionality will help to avoid out-of-space issues, when volume is stuck and no operation (e.g. export, destroy, etc) can be performed.



Note, that NMS creates free-space reserve after volume is 50% full. In order to alarm NMS about the occupied space 'volume-check' runner must be enabled. Read more about runners in section [Runners](#).

## 7 Disk Management

### 7.1 Locating Disks

SGI NAS supports physical slots to LUNs mapping based on on-disk GUID. SGI NAS slot mapping utility produces a map { disk GUID  $\Leftrightarrow$  slot number }.

Existing drive  $\Leftrightarrow$  slot mapping can be modified and additional mappings can be added. You can also make a given drive's LED blink, to identify its exact location in the appliance.

Please use the following NMC commands to view and administer slot mapping:

```
nmc:/$ show lun slotmap
```

```
nmc:/$ setup lun slotmap
```



NMC also provides implementation of 'lspci' command which lists all pci buses and devices in the system:

```
nmc:/$ lspci
```

## 7.2 JBOD Management

### 7.2.1 About JBOD.

JBOD, also referred to as "just a bunch of disks or drives," is also known as *spanning*, which refers to a computer's hard disks that have not yet been configured according to the RAID (redundant array of independent disks) system to increase fault tolerance and improve data access performance.

The RAID system stores the same data redundantly on multiple disks that appear to the operating system as a single disk. Although JBOD can also make disks appear to be a single one, it accomplishes this task by combining the drives into one large logical one. JBOD does not deliver any advantages over using separate disks independently and does not provide any of the fault tolerance or performance benefits of RAID.

However, the SGI NAS JBOD Managing interface adds the following features:

- Slot mapping
- Disk blinking
- Devices' state monitoring and warning notification

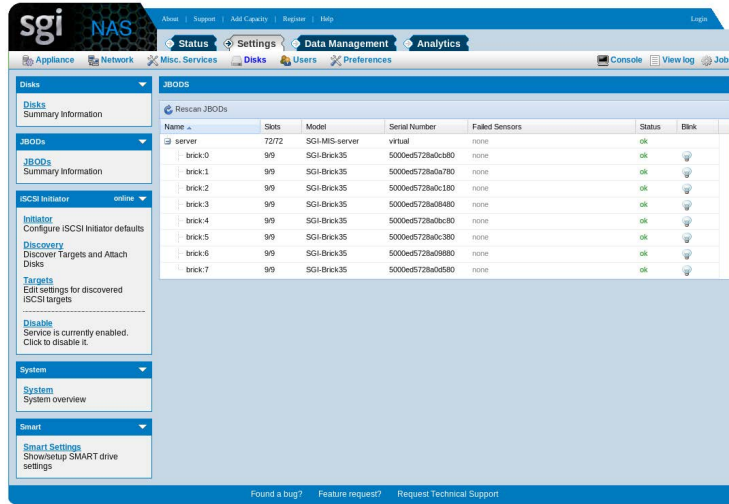
### 7.2.2 NMV JBOD Management

SGI NAS does not enable you to manage disks using JBOD unless a disk drive is already connected.

To view JBODs' states open **Settings** → **Disks** → **JBODs**

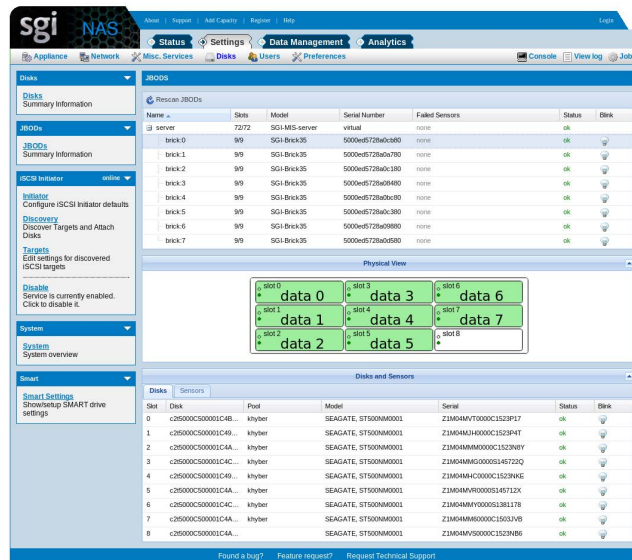
JBODs web page lists all the attached JBODs:

# SGI NAS User Guide



Note that full functionality is only available with the supported JBODs. The list of compatible JBODs is available in [Hardware Compatibility List](#).  
Not supported JBODs will appear in the list of JBODs and statistic information will be shown, but there will be no image of JBOD.

Click on JBOD's name under the list of JBODs to see the list of disks and other devices:



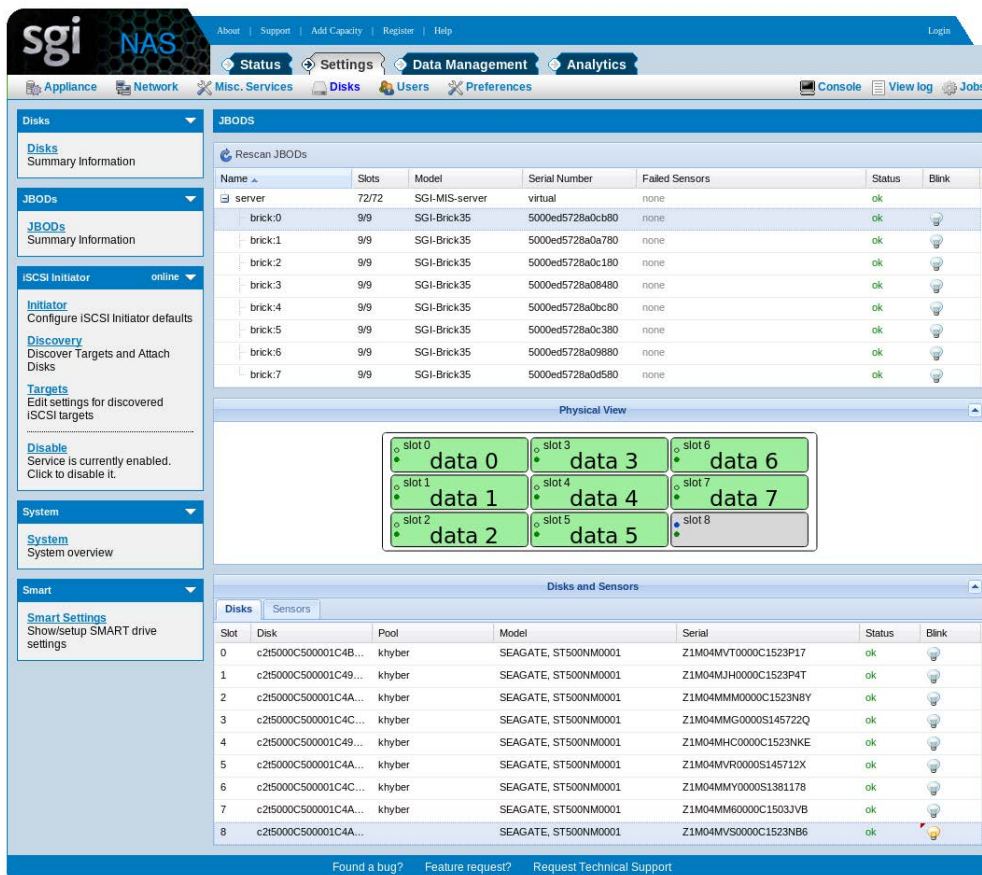
A logical view of the corresponding JBOD shows the disk-to-slot mapping. The list of disks includes the following:

- Disk number
- Pool to which the disk belongs (optional)

- Disk model
- Serial number
- Status.

The green dot indicates that the disks have an 'OK' status.

The red dot indicates you that the disk has failed or is corrupted. To locate and replace a failed disk, click on the light bulb to make a LED on a physical disk blink. In turn, the light bulb turns yellow and the unfilled dot on the JBOD screen also blinks blue:



After blinking is turned on you need to check the blinking device (usually blue or orange bulb).

To see the list of devices, sensors, their readings and states, click on '**Sensors**' tab:



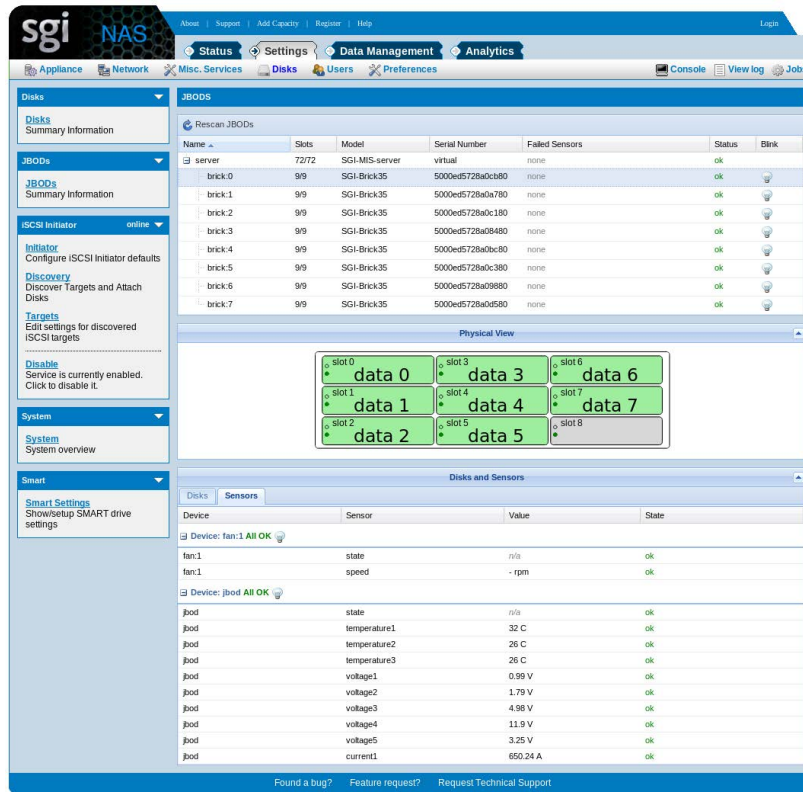
Device	Sensor	Value	State
⊕ Device: fan:0 All OK			
⊕ Device: fan:1 All OK			
⊕ Device: fan:2 All OK			
⊕ Device: fan:3 All OK			
⊕ Device: jbod All OK			
jbod	state	n/a	ok
jbod	temperature0	36 C	ok
jbod	temperature1	32 C	ok
jbod	temperature3	27 C	ok
jbod	voltage0	12 V	ok
jbod	voltage1	5.04 V	ok
jbod	voltage3	3.3 V	ok
jbod	voltage5	1.19 V	ok
jbod	voltage7	1 V	ok
⊕ Device: psu:0 All OK			
⊕ Device: psu:1 All OK			

Note that the list of devices contains the statistics not only for JBODs, but also for fans and power supply units.

If some parameter is in wrong state, an e-mail will be sent to specified address\* and failed sensors will appear in NMV\*:

\* Mailer must be configured. Read more about mailer settings in [Step 3 of Installation wizard 1](#)

\* Note that for every vendor each sensor (i.e. jbod\_temperature1 or jbod\_voltage3, etc.) may refer to a different parameter. For details, please refer to vendor's JBOD technical documentation or product support.



## 7.2.3 NMC JBOD Management

NMC provides the functionality as NMV.

1. To see the list of JBODs in a system, run:

```
nmc:/$ show jbod
```

NAME	MODEL	SLOTS	SERIAL	FAILED SENSORS
jbod:1	LSI-DE1600-SAS	5/12	50080e521e49b000	-
jbod:2	LSI-DE1600-SAS	2/12	50080e521e497000	-

2. To see specified JBOD slot mapping, run:

```
nmc:/$ show jbod jbod:1 slotmap
```

```

+-----+
/                                             /|
/                                             /|
+-----+ |
| [ slot 01 ] [ slot 02 ] [ slot 03 ] [ slot 04 ] | +
| [ slot 05 ] [ slot 06 ] [ slot 07 ] [ slot 08 ] | /
| [ slot 09 ] [ slot 10 ] [ slot 11 ] [ slot 12 ] | /

```

```
+-----+
```

SLOT	LUN	Pool	Model
1	-	-	-
2	-	-	-
3	c4t5000C50025A1050Bd0	-	SEAGATE, ST31000424SS
4	c4t5000C5002109AD93d0	-	SEAGATE, ST3500414SS
5	-	-	-
6	-	-	-
7	c4t5000C500212ABCFBd0	-	SEAGATE, ST31000424SS
8	c4t5000C5002109B99Bd0	-	SEAGATE, ST3500414SS
9	-	-	-
10	-	-	-
11	-	-	-
12	c4t5000C5002109B08Bd0	-	SEAGATE, ST3500414SS

3. To make JBOD blink:

```
nmc:/$ show jbod <jbod:id> blink
```

4. To make psu or fan blink:

```
show jbod <jbod:id> <psu|fan> <id> blink
```

5. To make slot blink:

```
show jbod <jbod:id> lun <disk> blink
```

6. To list all the sensors readings:

```
nmc:/$ show jbod jbod:1 sensors
```

7. To see the failed sensors in NMC, run:

- for all JBODs:

```
nmc:/$ show jbod alerts
```

- for specific JBOD:

```
nmc:/$ show jbod <jbod:id> alerts
```

8. To see all the current readings, slot mapping table, sensors and alerts for all JBODs in the system, run:

```
nmc:/$ show jbod all
```

### 7.3 Viewing Disk Status

By selecting the **Settings** tab in NMV, and the **Disks** sub-tab, you can see a list of disks, various disk properties, and whether they are already associated with data volumes. An example screen is shown below:

The screenshot shows the SGI NAS web interface. The top navigation bar includes links for About, Support, Add Capacity, Register, and Help. The user is logged in as Administrator. The main menu has tabs for Status, Settings, Data Management, and Analytics. The left sidebar shows the 'Disks' section with options for Summary Information, iSCSI Initiator (online), Discover, Targets, Disable, and Smart Settings. The main content area is titled 'SUMMARY INFORMATION: DISKS AND VIRTUAL LUNS' and contains two tables.

**All Disks**

Disk	Device	Type	Size	Volume	Attach	Model	Blink
<a href="#">c0t1d0</a>	sd2	disk	68.49 GB	mypool	mpt	FUJITSU, Rev. 0104	
<a href="#">c0t2d0</a>	sd1	disk	232.89 GB	syspool	mpt	Hitachi, Rev. V5DOA73A	
<a href="#">c0t3d0</a>	sd3	disk	68.49 GB	mypool	mpt	FUJITSU, Rev. 0104	
<a href="#">c0t4d0</a>	sd4	disk	68.49 GB		mpt	FUJITSU, Rev. 0104	
<a href="#">c0t5d0</a>	sd5	disk	68.49 GB		mpt	FUJITSU, Rev. 0104	

**Live Disks Statistics**

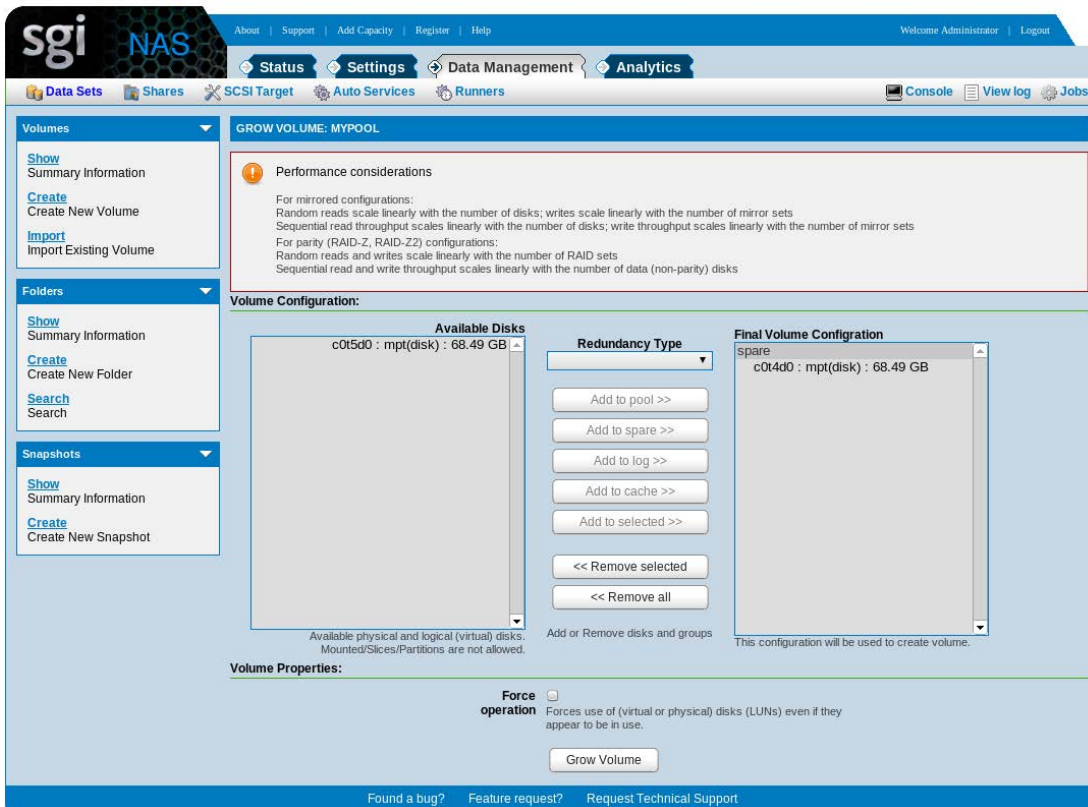
Disk	r/s	w/s	kr/s	kw/s	wait	actv	wsvc_t	asvc_t	%w	%b
c0t1d0	2.0	0.8	6.8	0.0	0.0	0.0	0.0	0.0	0	0
c0t2d0	1.2	12.4	0.6	30.1	0.0	0.5	0.0	36.7	0	11
c0t3d0	2.0	0.8	6.8	0.0	0.0	0.0	0.0	0.0	0	0
c0t4d0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	0
c0t5d0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	0

If the disk already belongs to a data volume, the volume name will show in the Volume column. If you click on the disk name, you can see additional properties about the disk.

If disks have recently been mapped to this host, or if you suspect the configuration information is out-of-date, you can update the information using the Refresh button. Note that re-synchronizing the system with the disk configuration can take some time.

## 7.4 Adding Spares to a Data Volume

In NMC you can add one or more devices to a data volume by clicking 'Grow' in the data volume summary view. The Grow Volume page will appear showing you the available disks. Select one or more disks and then click the button “Add to spare”. You will see a view similar to the following screenshot:



Select “Grow Volume” to add the spare devices to the volume.

## 7.5 Adding Global Spares

SGI NAS allows you to have hot spares for your volumes. If a device in the pool fails, the system will detect the failure and activate the spare device automatically. However, if you have multiple volumes you may not want to dedicate a spare device to each one. With global hot spares, one device can be a spare for multiple volumes. If there is a failure in any of the volumes, the spare can then be activated.

To set up a device that serves as a spare for multiple volumes, first create a volume and add devices to it. Note that in this example we are setting up a mirror but other redundancy options such as RAID-Z1 would also work.

```
nmc:/$ create volume my-mirror
Group of devices : c1t1d0, c1t2d0
Group redundancy type : mirror
Continue adding devices to the volume 'my-mirror'? (y/n) y

After setting up the mirror, add the spare device.

Group of devices : c1t3d0
Group redundancy type : spare
Continue adding devices to the volume 'my-mirror'? (y/n) n

Create volume 'my-mirror'? (y/n) y
```

Now create a second volume.

```
nmc:/$
    create volume my-mirror2
Group of devices : c1t4d0, c1t5d0
Group redundancy type : mirror
Continue adding devices to the volume 'my-mirror2'? (y/n) n

Create volume 'my-mirror2'? (y/n) y

Now you have two volumes. To allow for c1t3d0 to be a spare for the second
volume, in addition to the first volume, do the following:

nmc@myhost:/$ setup volume my-mirror2 grow spare c1t3d0
```

At this point each volume is using the device 'c1t3d0' as a spare. If a device fails in either pool, the spare will be activated.

NMV provides the same functionality. Note the following examples.

In NMV, go to **Data Management** → **Data Sets** and click on the volume to which you want add a hot spare. Check the square under **Hot spares** title and click '**Allow**'. In the example below we added a hot spare to '**pool1**':

# SGI NAS User Guide

The screenshot shows the SGI NAS web interface for configuring 'VOLUME: POOL1'. The interface includes a navigation menu on the left with options like 'Volumes', 'Folders', and 'Snapshots'. The main content area is divided into several sections:

- Disks:** A table showing disk details for 'pool1 (device)'.

Disk	Size	Status	Errors (RW/C)
pool1 (device)			
c01d0	68.49 GB	ONLINE	0 / 0 / 0
- Hot Spares (only in available state):** A table showing hot spare disks.

Disk	Size	Volume
c04d0	68.49 GB	pool1
c05d0	68.49 GB	
- Configurable Properties:** A form with various settings:
  - Description: Optional volume description. Maximum length is 255 characters.
  - Deduplication:  off
  - Compression:  on
  - Autosexpand:  off
  - Sync:  standard
- Static Properties:** A table of system properties.

Property	Value
allocated	151K
altroot	-
autosexpand	off
autoreplace	on
available	66.9G
bootfs	-
cache	-
capacity	0%
dedupratio	0
deduplicate	1.00x
delegation	on
failmode	continue
free	66.9G
guid	2377614082784282037
health	ONLINE
info	pool: 1, spare: 1, devices: 2
name	pool1
size	68G
sync	standard
used	110K
version	28

Next, do the same to other volume to which you want to add the same hot spare. In the example below it is 'pool2':

The screenshot shows the SGI NAS web interface for configuring 'VOLUME: POOL2'. The interface is similar to the previous one, but with the following differences:

- Disks:** A table showing disk details for 'pool2 (device)'.

Disk	Size	Status	Errors (RW/C)
pool2 (device)			
c0c00	68.49 GB	ONLINE	0 / 0 / 0
- Hot Spares (only in available state):** A table showing hot spare disks.

Disk	Size	Volume
c04d0	68.49 GB	pool1
c05d0	68.49 GB	pool2
- Configurable Properties:** A form with various settings:
  - Description: Optional volume description. Maximum length is 255 characters.
  - Deduplication:  off
  - Compression:  on
  - Autosexpand:  off
  - Sync:  standard
- Static Properties:** A table of system properties.

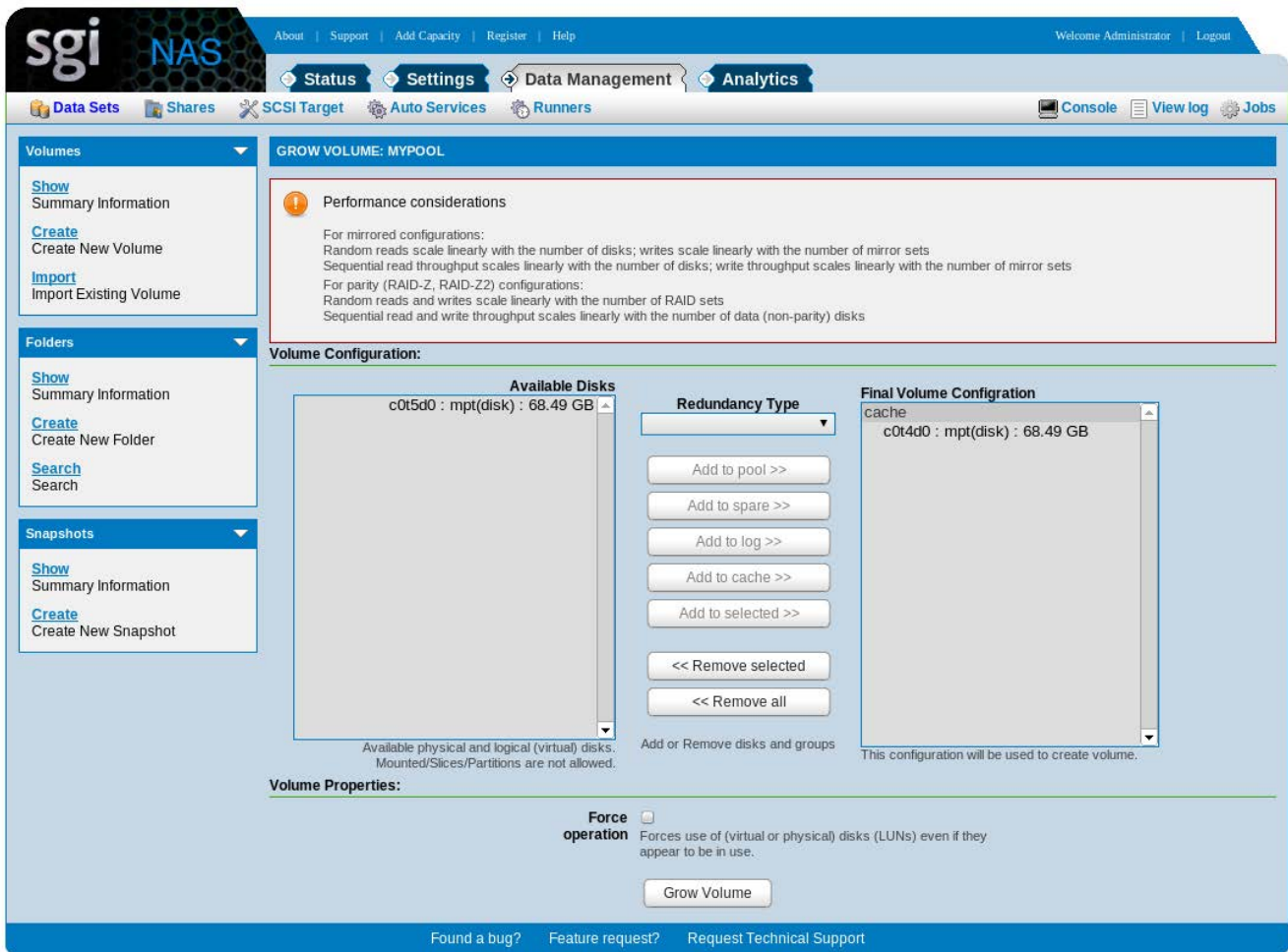
Property	Value
allocated	148K
altroot	-
autosexpand	off
autoreplace	on
available	66.9G
bootfs	-
cache	-
capacity	0%
dedupratio	0
deduplicate	1.00x
delegation	on
failmode	continue
free	66.9G
guid	15412524897888266270
health	ONLINE
info	pool: 1, spare: 1, devices: 2
name	pool2
size	68G
sync	standard
used	108K
version	28



## **7.6 Adding Cache Devices**

ZFS provides the Adaptive Replacement Cache (ARC) from main memory. The ARC is shared across all data volumes. Additional cache devices, also referred to as L2ARC, are assigned to a specific data volume.

In NMC you can add one or more cache devices to a data volume by clicking Grow in the data volume summary view. The Grow Volume page will appear showing you the available disks. Select one or more disks and then click the button “Add to cache”. You will see a view similar to the following screenshot:

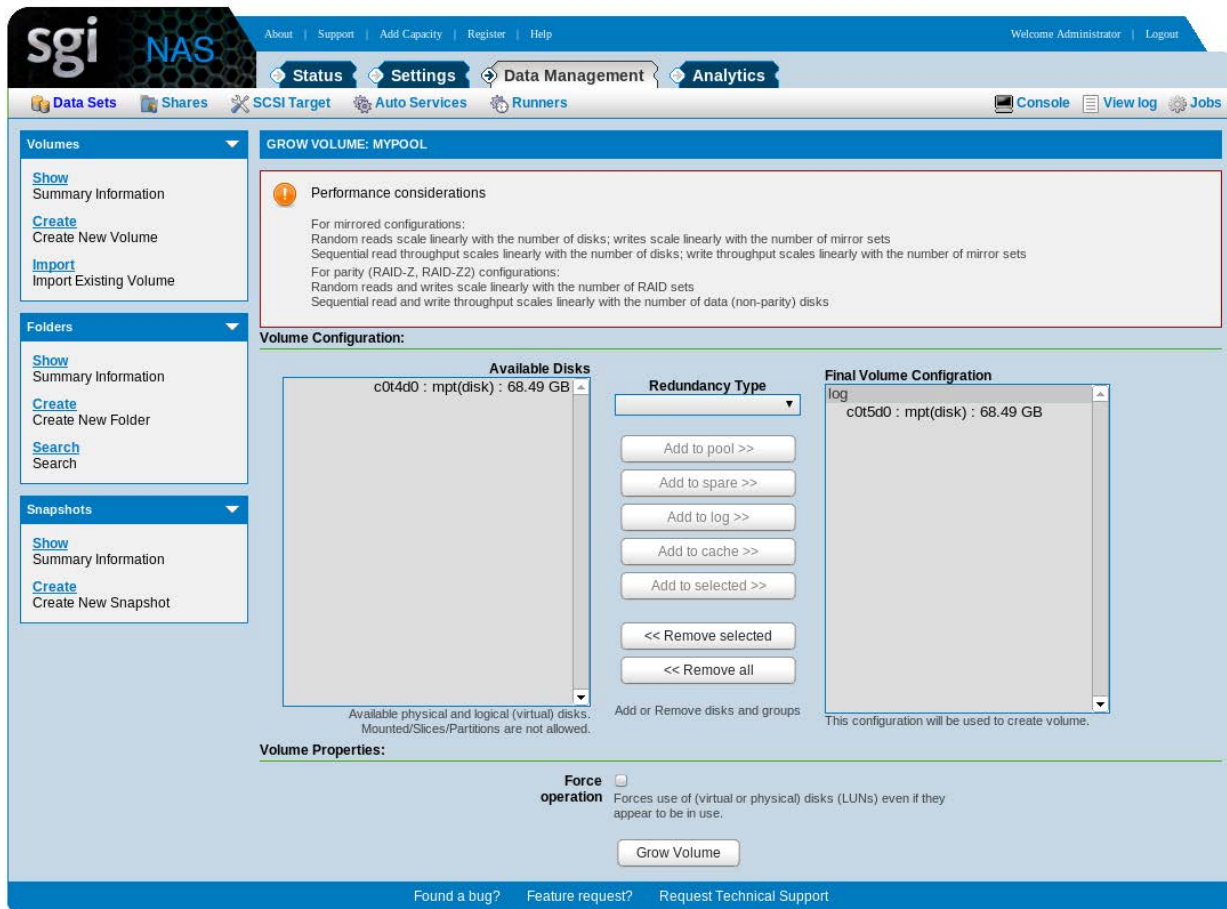


You can then click “Grow Volume” to add the cache device to the volume. Cache devices can improve your read performance for random I/O workloads.

## 7.7 Adding Log Devices

SGI NAS uses an intent log to meet POSIX requirements for handling synchronous writes. By default, the intent log is a part of the main data volume, but you may be able to improve performance by moving it to a separate device such as a solid-state disk (SSD).

In NMC you can add one or more log devices to a data volume by clicking 'Grow' in the data volume summary view. The Grow Volume page will appear, showing you the available disks. Select one or more disks and then click the button “Add to log”. You will see a view similar to the following screenshot:



You can then click “Grow Volume” to add the log device to the data volume.

Note, that it is recommended that you mirror your log device. Other types of redundant configurations are not supported.

## 7.8 Removing a Device

Cache, spare, and separate log devices can be removed from a data volume. To do this in NMC, use the command:

```
nmc:/$ setup volume <volumename> remove-lun
```

## 7.9 Replacing a Disk

If a disk in a data volume fails, you can replace it with another available disk. The following example shows how to do this using NMC:

```
nmc:/$ setup volume
Option ? mypool
Option ? replace-lun
LUN to replace      : c1t1d0
LUN to use as a replacement : c1t4d0
Replace 'c1t1d0' with 'c1t4d0' in the volume 'mypool' ? Yes
```

This involves re-silvering the disk and can take some time. Using the NMC command:

```
nmc:/$ show volume <volumename> status
```

you will be able to tell, if the resilver is done or in-progress.

## 7.10 Taking a disk offline

If a disk in the data volume is having problems, it can be taken offline with the NMC command:

```
nmc:/$ setup volume <volumename> offline-lun
```

## 7.11 Recovering a previously disconnected disk

If a disk in the data volume was taken offline, but is now ready to be returned to the volume, then you can add it back with the NMC command :

```
nmc:/$ setup volume <volumename> online-lun.
```

Note that resilvering has to complete before the disk is fully online.

## 7.12 Replacing a Redundancy Group

You can replace a redundancy group with another group that is the same size or larger. If the data volume's autoexpand property is on, then replacing all the groups with larger groups will expand the volume size.

To replace a redundancy group in NMC, use the command:

```
nmc:/$ setup volume <volumename> replace-lun.
```

## 7.13 Creating a Mirror

If you created a data volume without redundancy, you can later create a mirrored configuration. This is also true for a non-redundant separate log device. You can also add a mirror to a mirrored data volume.

The command to do this is:

```
nmc:/$ setup volume <volumename> attach-lun.
```

Note that the device being added must be at least as large as the existing device or redundancy group.

## 7.14 Detaching a Mirror

You can remove a device from a mirror by detaching it. To do this in NMC, use the command

```
nmc:/$ setup volume <volumename> detach-lun.
```

## 7.15 Re-attaching a Mirror

If you detached a disk temporarily from a mirror you can add it back in NMC using the command:

```
nmc:/$ setup volume <volumename> attach-lun.
```



## 8 Managing Folders

### 8.1 Create Folders

To create a folder in NMC use the command:

```
nmc:/$ create folder
```

To create a folder in NMV you can select the [Create](#) link under the **'Folders'** heading. You will see the screen below where you enter information such as the folder name, description, record size, and other properties.

The screenshot shows the SGI NAS web interface. The top navigation bar includes links for About, Support, Add Capacity, Register, and Help. The main navigation bar has tabs for Status, Settings, Data Management, and Analytics. The left sidebar shows a tree view with Volumes, Folders, and Snapshots. The main content area is titled 'CREATE NEW FOLDER' and contains a form with the following fields:

- Volume:** A dropdown menu set to 'mypool'. Below it, the text reads 'Folder's volume.'
- Folder Name:** A text input field containing 'folder1'. Below it, a note states: 'Each folder pathname's component delimited by backslash (?) can only contain alphanumeric characters (a-z, A-Z, 0-9) in addition to the following three special characters: underscore (\_), hyphen (-) and period (.). Folder pathname must begins with an alphanumeric character and not contains a percent sign (%).'
- Description:** A text input field with the placeholder text 'Human-readable description for this folder.'
- Record Size:** A dropdown menu set to '128K'. Below it, the text reads 'Specifies a suggested block size for files in the folder. Default is 128K.'
- Log Bias:** A dropdown menu set to 'latency'. Below it, a note states: 'Provide a hint to ZFS about handling of synchronous requests in this dataset. If logbias is set to latency (the default), ZFS will use pool log devices (if configured) to handle the requests at low latency. If logbias is set to throughput, ZFS will not use configured pool log devices. ZFS will instead optimize synchronous operations for global pool throughput and efficient use of resources.'
- Deduplication:** A dropdown menu set to 'off'. Below it, the text reads 'Controls the deduplication option for this dataset. If enabled, it will optimize use of duplicate copies of data. Default is "off".'
- Compression:** A dropdown menu set to 'on'. Below it, a note states: 'Controls the compression algorithm used for this dataset. Default is "on". Setting compression to "on" uses the lzjb compression algorithm. The lzjb compression algorithm is optimized for performance while providing decent data compression. Currently, "gzip" is equivalent to "gzip-6".'
- Number of Copies:** A dropdown menu set to '1'. Below it, the text reads 'Controls the number of copies of data stored for this dataset. Default is "1".'
- Case Sensitivity:** A dropdown menu set to 'mixed'. Below it, a note states: 'Indicates whether the file name matching algorithm used by the file system should be case-sensitive, case-insensitive, or allow a combination of both styles of matching. Use "mixed" if the folder is planned to be shared via CIFS and NFS at the same time. Default is "mixed".'
- Unicode Only:** A dropdown menu set to 'off'. Below it, a note states: 'Enable it if you want to exclude non-Unicode file names creation for this folder. If set, this option will ensure better inter-client operability. Make sure this option is enabled if you are planning to use this folder as a share for MacOS X, Windows and Linux clients. Default is "off".'
- Sync:** A dropdown menu set to 'standard'. Below it, a note states: 'Controls synchronous requests (standard - ensure all synchronous requests are written to stable storage; always - every file system transaction will be written and flushed to stable storage by system call return; disabled - synchronous requests are disabled). Default is standard.'

At the bottom of the form is a 'Create' button. Below the form, there are links for 'Found a bug?', 'Feature request?', and 'Request Technical Support'.

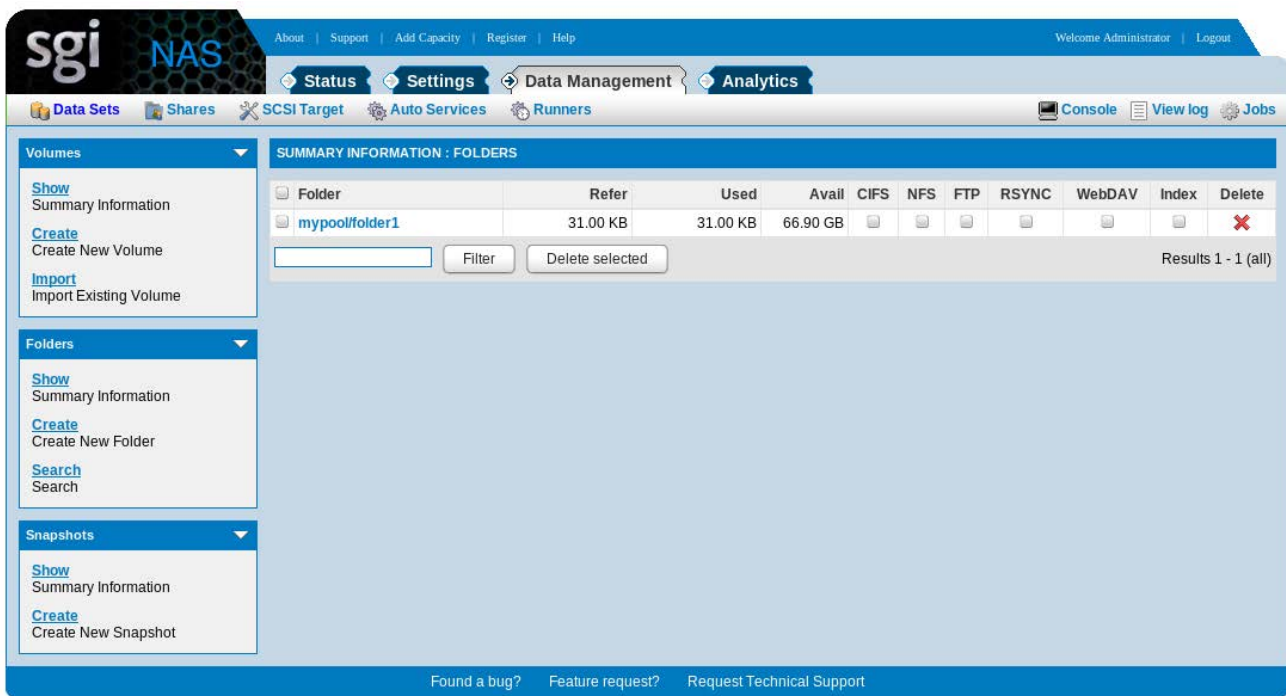
De-duplication is off for the folder by default because it has a performance impact. Turn it on

only if you expect to have duplicate blocks in this file system.

For case-sensitivity, the default choice is “mixed” which is optimal if the folder is going to be used in mixed CIFS and NFS environments.

## 8.2 View Status

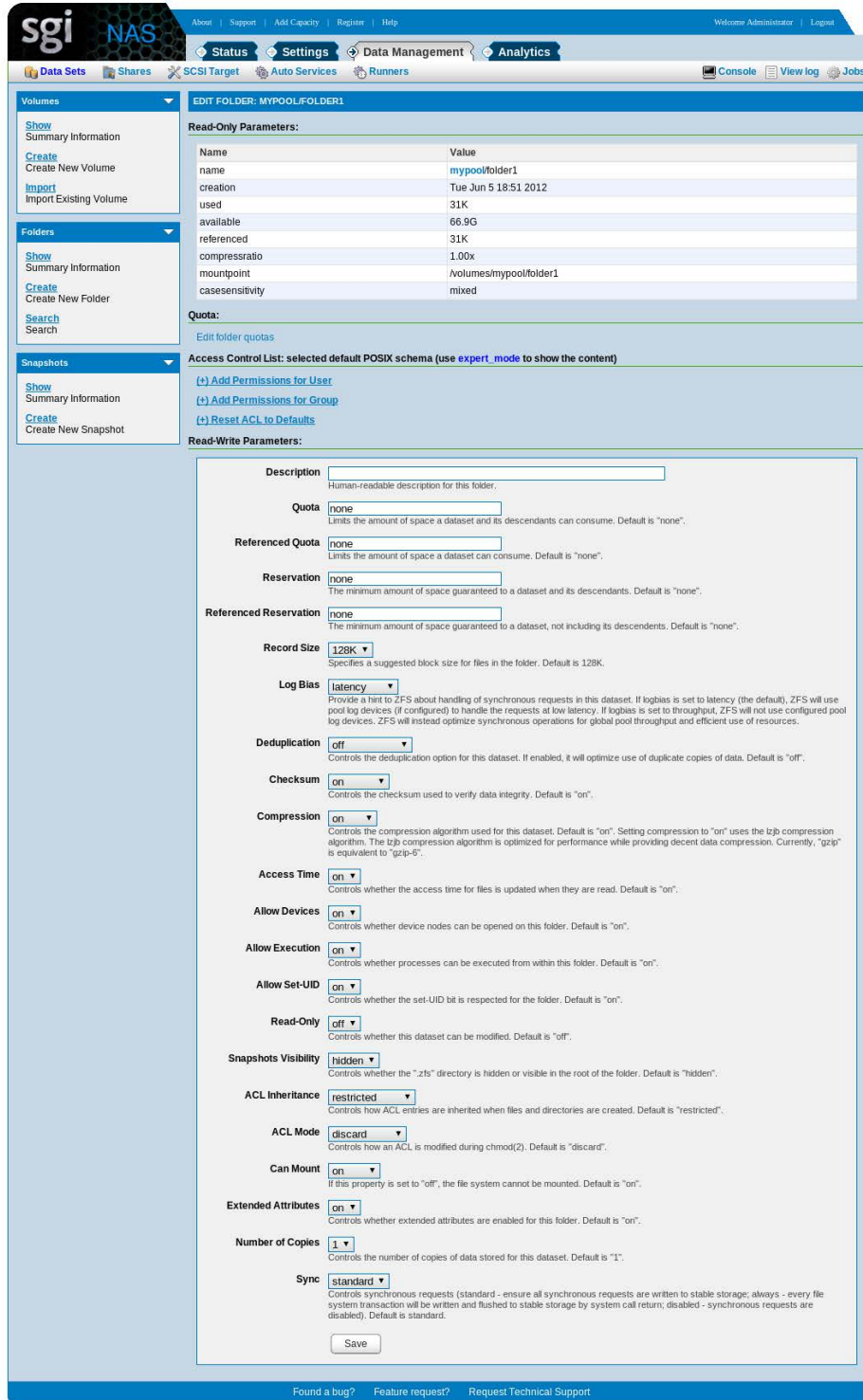
After creating a folder you can monitor its status on the folder summary view. This view will show how much space is used and available, and also indicate if the folder is being shared via any sharing protocol.



## 8.3 Edit Properties

After a folder is created you can edit the properties by clicking on the folder’s name in the summary view.





In NMC you can see all folder properties using the command:

```
nmc:/$ show folder <foldername> property
```

To change a property use:

```
nmc:/$ setup folder <foldername> property <propertyname>
```

### 8.3.1 Logbias property

The 'logbias' property is specifically intended to improve database performance. The logbias property provides a hint to ZFS on how to handle synchronous requests. Note that database engines typically employ synchronous I/O when writing transaction logs. If logbias is set to 'latency' (the default), ZFS will use the volume's log devices (if available in the volume) to handle the requests at the lowest possible latency.

Typically, database transaction logs need the shortest latency. Therefore, use 'logbias=latency' on the corresponding SGI NAS folder or zvol that holds the database data. On the other hand, database data files need to be optimized for throughput. So, the appropriate setting is: 'logbias=throughput'.

If 'logbias' is set to 'throughput', ZFS will not use configured log devices.

This property can be set in NMC, as shown below:

```
nmc:/$ setup folder <foldername> property logbias  
logbias : throughput
```

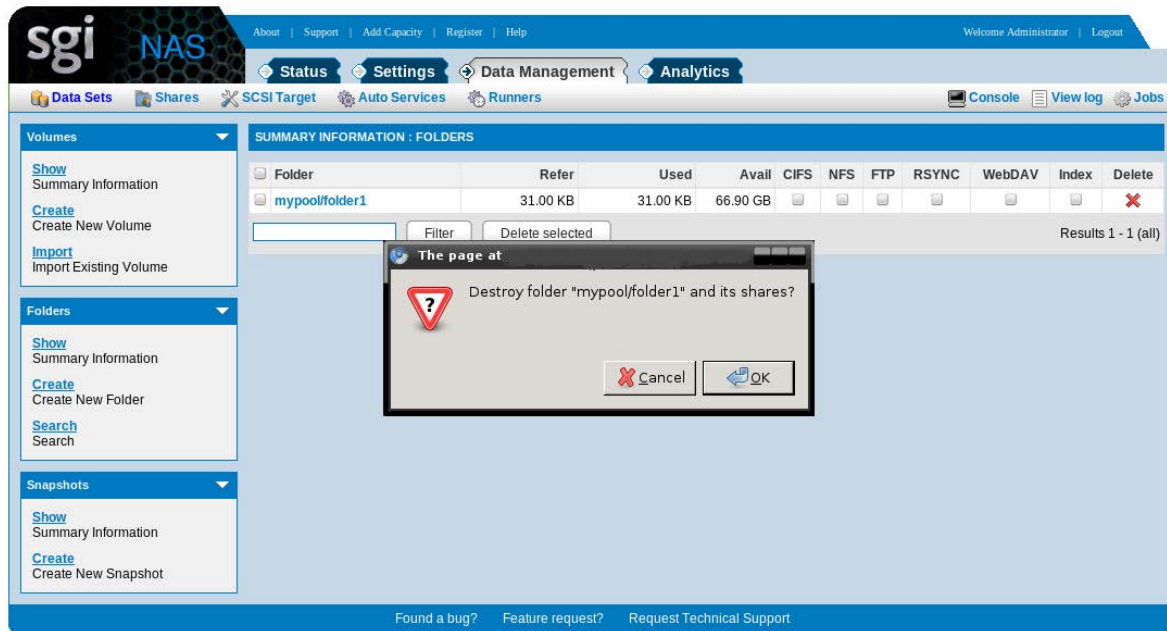
## 8.4 Destroy Folder

To delete a folder in NMC, use the command:

```
nmc:/$ destroy folder
```

A folder can be deleted in NMV from the folder summary view. Click on the delete icon to remove the folder.

This operation can be done in NMV as well. Go to **Data Management** → **Data sets** → **Folders** → [Show](#).



If a few folders need to be deleted, you can either select, or filter them by special parameter.

- Select the folders and click on 'Delete selected' to delete selected folders.
- Filter the folders and click on 'Delete filtered'. Note, that this operation deletes all the filtered folders, even if they are not shown on the current page.

## 8.5 Search & Indexing

SGI NAS supports the ability to index and later search folders and their snapshots. To start indexing a folder, click the Index checkbox on the folder summary view.

After selecting indexer, a dialog box will appear confirming that you want to create an indexer. Note that the indexer runs at a scheduled time, so searching immediately may not work.

## 8.6 Sharing Folders

SGI NAS can share folders using a variety of protocols including CIFS, NFS, WebDAV, RSYNC, and FTP. CIFS and NFS sharing are describing in the following sections.

### **8.6.1 Sharing Folders with NFS and CIFS**

Sharing a folder with NFS or CIFS are each described in separate sections. If you plan to share a folder using both protocols, there are a couple property settings to be aware of.

When you create a new folder the default setting for casesensitivity is “mixed”. This will ensure the proper behavior if the folder will be shared via CIFS and NFS. You can change this property only at folder creation time.

Another important property is “nbmand”. SGI NAS will check this property when you share a folder via NFS that has already been shared with CIFS, or vice versa. If the property is off, you will be asked to change it to on.

CIFS protocol assumes mandatory locking and UNIX traditionally uses advisory locking so it is recommended to set the property 'nbmand' to 'on' in order to enforce mandatory cross-protocol share reservations and byte-range locking in a mixed NFS/CIFS environment.

For 'nbmand' property changes to take effect, the folder need to be remounted. Unmounting and mounting the folder again may cause a temporary loss of client connections. Note that you can remount manually any time later.

## 9 NFS File Sharing

### 9.1 Create NFS Share

You can use the NMC `share` command to share a folder via NFS.

```
nmc:/$ share mypool/folder1 nfs
rw      : group-engineering:10.16.16.92
ro      : group-marketing
root    : admin
extra-options:
```

### 9.2 Edit NFS Folder Properties

Click on [Edit](#) link to manage NFS share properties.

Folder	Refer	Used	Avail	CIFS	NFS	FTP	RSYNC	WebDAV	Index	Delete
mypool/folder1	31.00 KB	31.00 KB	66.90 GB		<input checked="" type="checkbox"/> Edit					<input type="checkbox"/>

### 9.3 Mounting on Linux

Note, that child file systems do not get mounted automatically. You need to mount each ZFS file system you're exporting via NFS separately.



## 10 CIFS File Sharing

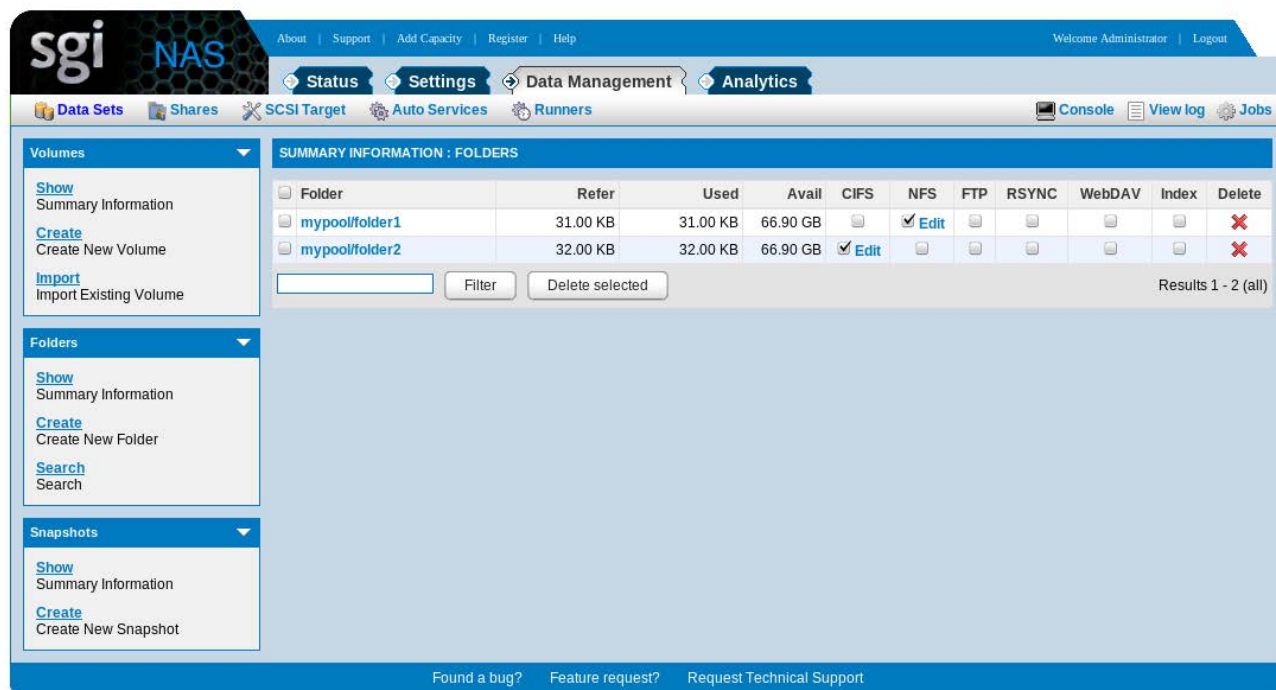
### 10.1 Introduction

SGI NAS provides one of the best existing kernel and ZFS-integrated CIFS stacks, with native support for Windows Access Control Lists (ACL).

See the CIFS User Guide to study all the aspects of using the CIFS protocol for sharing folders in SGI NAS for domain or workgroup, anonymously and as authorized user, including Active Directory integration, identity mapping instructions and troubleshooting tips. This section explains how to use CIFS capabilities to share SGI NAS folders for:

### 10.2 Configuring CIFS server

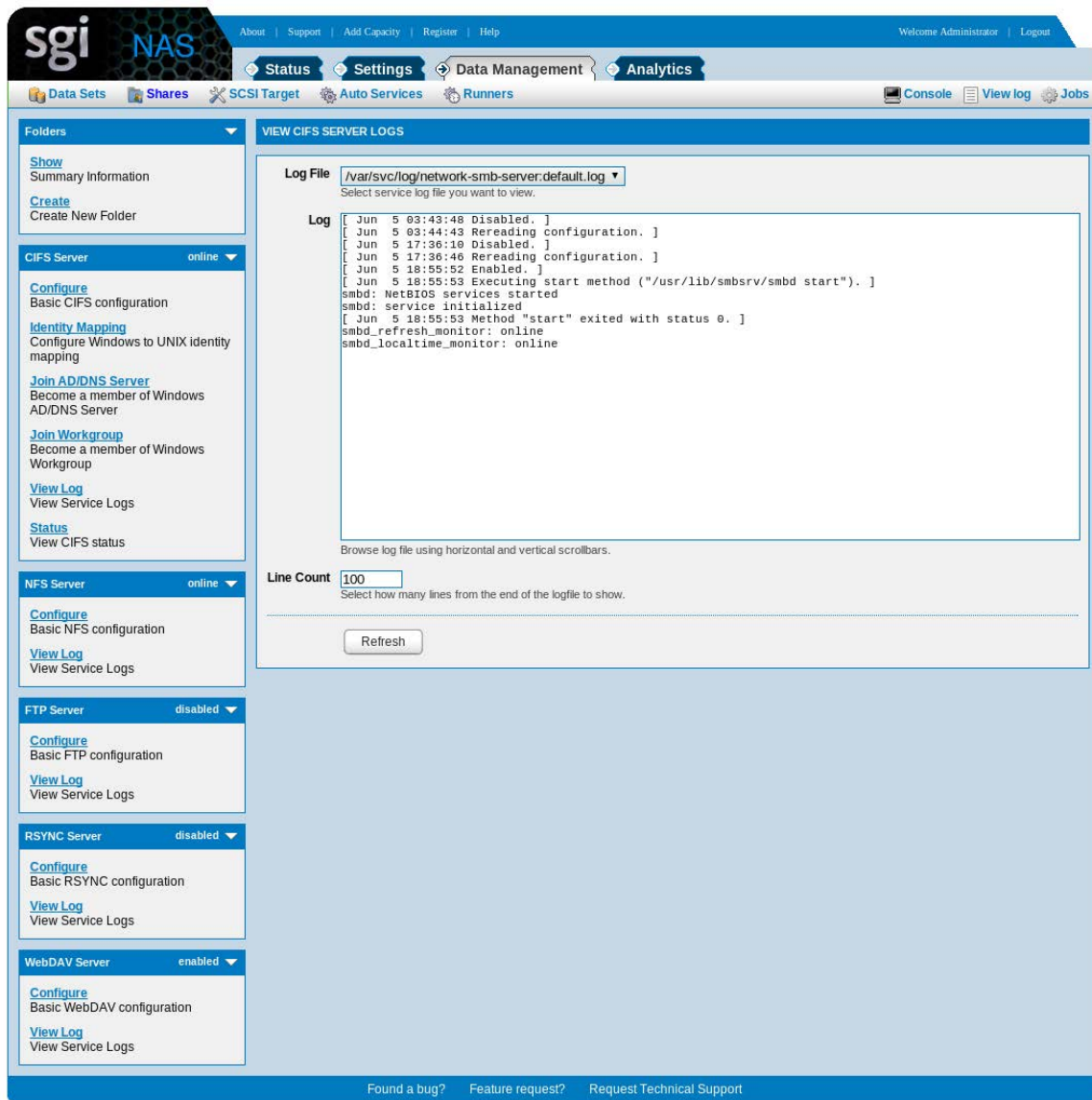
NMV provides a page to configure all network services, including CIFS (**Data Management** → **Shares**):



In NMV, you will find on this page a number of related links to configure, join workgroup, join active directory, unconfigure, and view the log file (see above).

The following screenshot illustrates viewing CIFS logfile:





In NMC, network services are configured via '**setup network service**'.

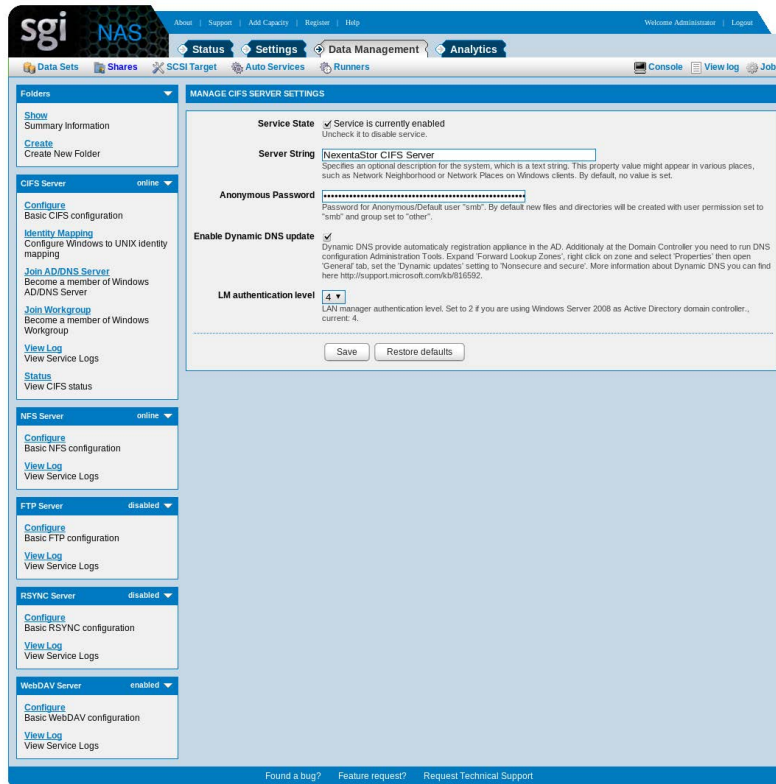
The corresponding NMC command to view the log would be, respectively:

```
nmc:/$ show network service cifs-server log
```

The important screen, however, is '**CIFS Server Settings**', which you get by clicking on the link denoted as **Configure**. In NMC, the corresponding command would be (see footnote<sup>A</sup>):

```
nmc:/$ setup network service cifs-server configure
```

A As always, there is no need to remember this command. Simply enter **setup**, and then keep pressing TAB-TAB and making selection.



Here, make sure that the server is enabled, and specify a password for anonymous access.



The default password is "nasnas".

It is important to change the default pre-configured password for anonymous access.

## 10.2.1 Create CIFS share

Use regular 'share/unshare' functionality to share (and unshare) appliance's folders for Windows users.

In NMC you would use the same generic 'show share', 'share' and 'unshare' commands. For example:

```
nmc@testbox1:/$ show share

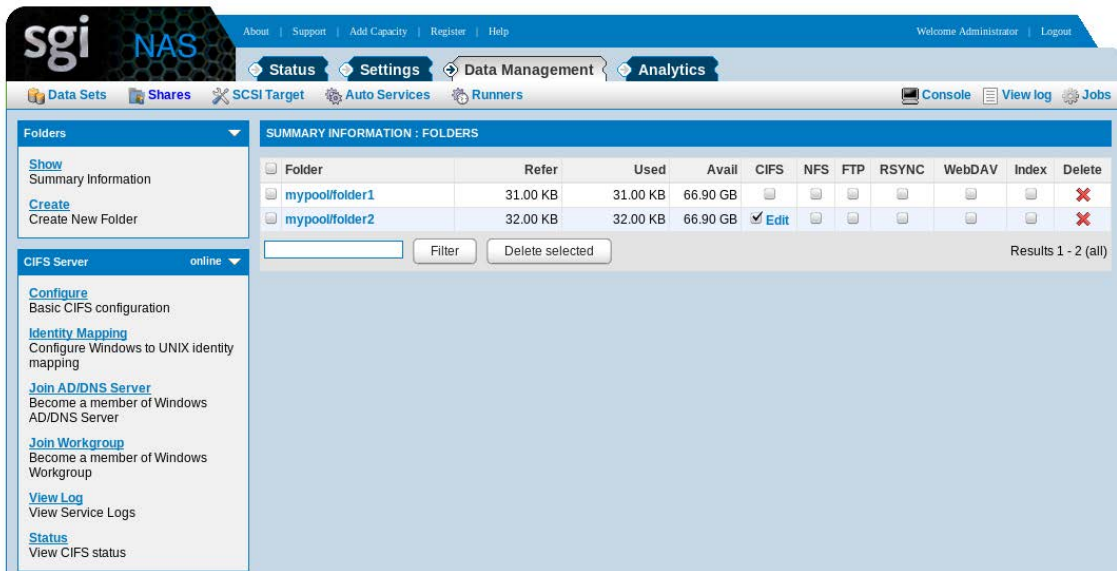
FOLDER                CIFS  NFS   RSYNC  FTP   WEBDAV  SAMBA
vol1/a                 Yes   Yes   -       -     -       -
vol1/a/b               Yes   Yes   -       -     -       -

nmc@testbox1:/$ show folder mypool/a share cifs -v

PROPERTY              VALUE
```

```
folder          mypool/a
share_name      mypool_a
comment         '
anonymous_rw    true
PROPERTY       VALUE
folder          mypool/a/b
share_name      mypool_a_b
comment         '
anonymous_rw    true
```

In NMV, to display or change existing CIFS shares, or add new ones, go to **Data Management**  
→ **Shares**:



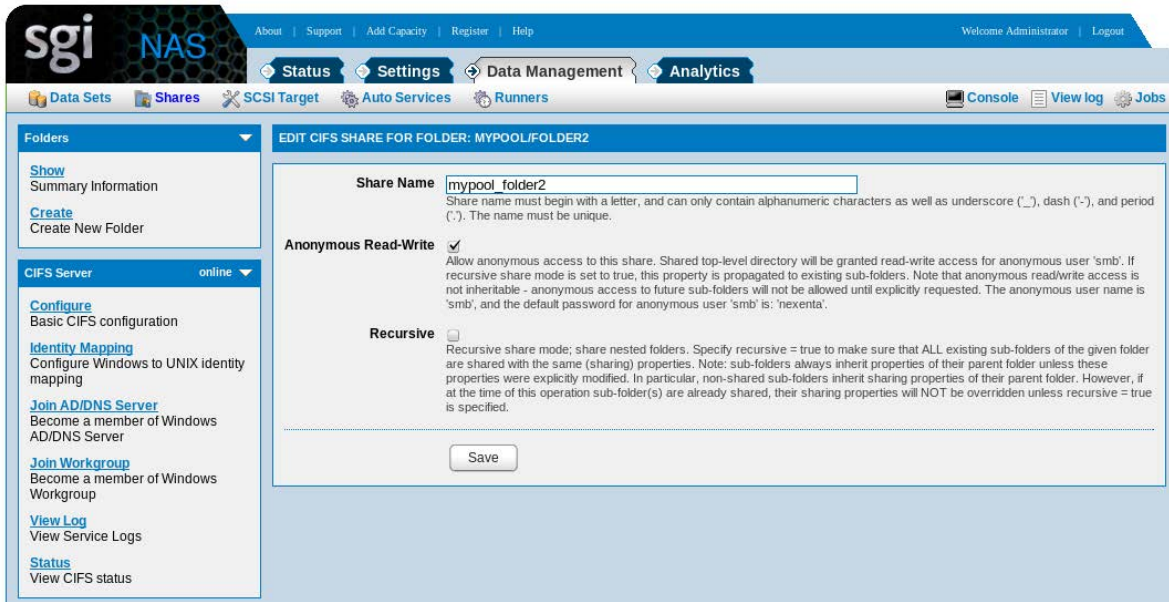
NMV page **Data Management** → **Shares** is the single point of control that allows to create folders with a given set of properties, and destroy existing folders.

Folders can be filtered by name – see the [Filter](#) button at the bottom of the screenshot (below). Folders can be shared via CIFS (as well as NFS, FTP, RSYNC, and WebDAV).

In addition, the same page is used to view and configure CIFS server settings (see left panel below). All this power and flexibility is available via NMV **Data Management** → **Shares** page:



To configure any given share, first enable it (checkbox in the CIFS column above). This will automatically share the folder using default system settings. You can view and modify those settings by clicking on the [Edit](#) link to the right of the checkbox, as shown below:



This is just a basic overview for CIFS sharing. To get more information on accessing CIFS shares with anonymous or authorized user in domain or workgroup mode, see the CIFS User Guide.

## 10.3 Other sharing options

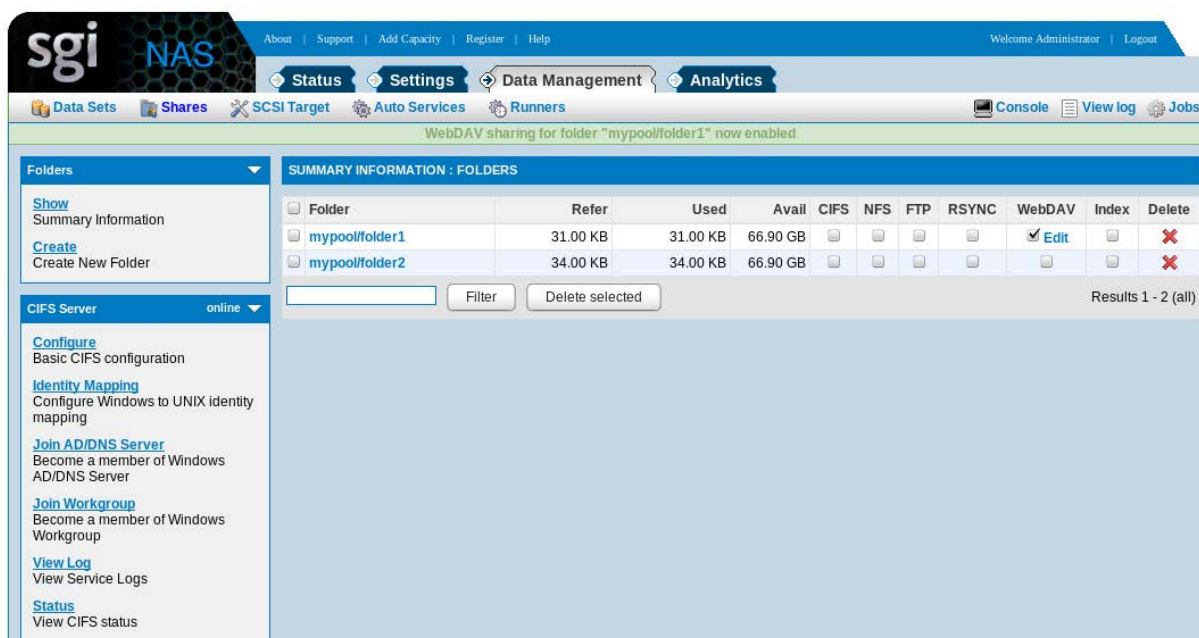
### 10.3.1 WebDAV

WebDAV is a set of methods based on the Hypertext Transfer Protocol (HTTP) that facilitates collaboration between users in editing and managing documents and files stored on World Wide Web servers<sup>A</sup>.

Follow the steps bellow to create and access WebDAV share:

1. In NMV, go to **Data Management** → **Shares** and check WebDAV checkbox to share a folder. On the following screen 'vol1/folder1' was shared:

<sup>A</sup> <http://en.wikipedia.org/wiki/WebDAV>



2. Use web-browser to access the share at: `https://<appliance_address>/mypool/folder1`

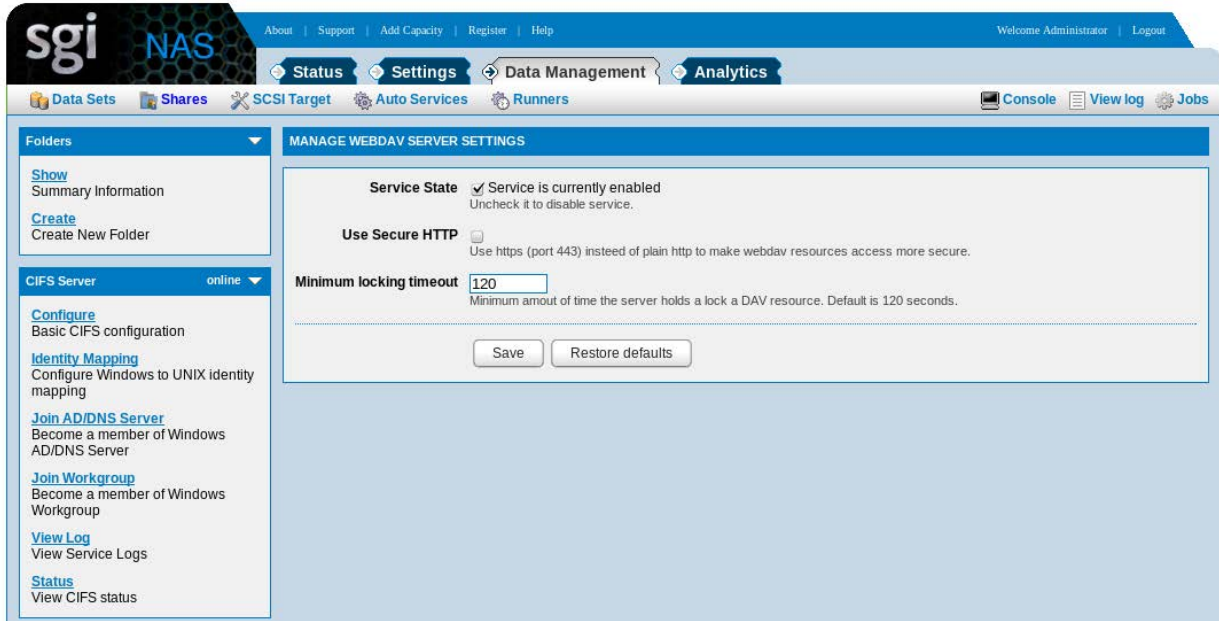


To access WebDAV share use the following credentials:

```

user: webdav
password: nasnas
    
```

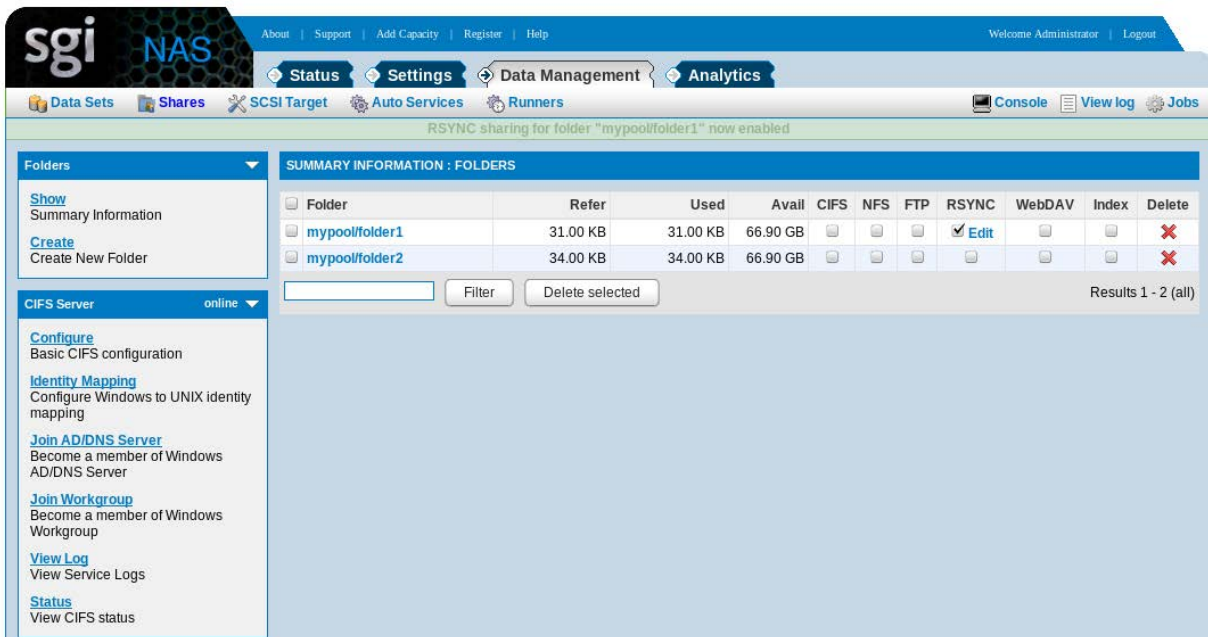
By default, WebDAV uses port 443 to access the shares. It is https-based connection. To enable http-based connection via port 80, uncheck the '**Use secure HTTP**' option in WebDAV server settings. In NMV go to: **Data Management** → **Shares** → **WebDAV Server** → [Configure](#)



### 10.3.2 RSYNC

SGI NAS provides the possibility to share a filesystem via RSYNC:

1. In NMV, go to **Data Management** → **Shares** check RSYNC to share the chosen folder. In the example above folder 'vol1/folder1' is shared via RSYNC.



2. Depending on your OS use the corresponding RSYNC client to access the share (i.e. [DeltaCopy](#), [Syncrify](#) etc.).



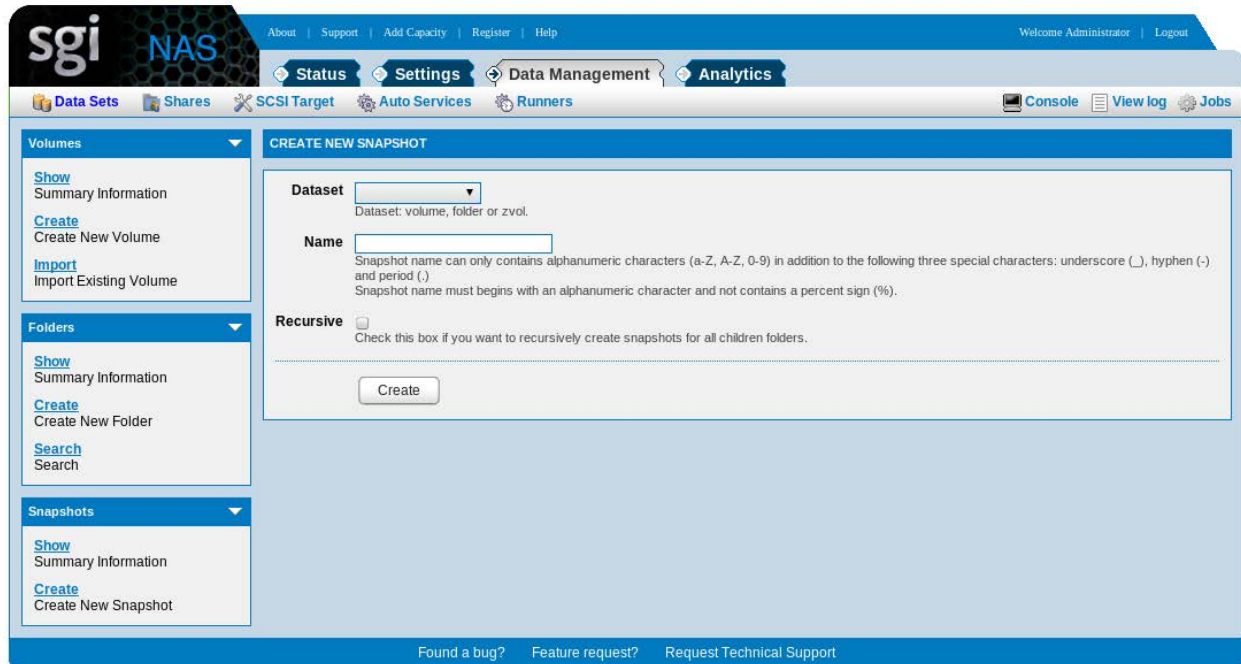


## 11 Managing Snapshots

Snapshots are read-only, point-in-time representations of a file system or zvol. Because of the copy-on-write nature of ZFS, snapshots can be created instantaneously.

### 11.1 Create Snapshot

Before making significant changes to a folder, you may want to create a snapshot so that you can recover to this point in time if anything goes wrong. You can create a new snapshot in NMV from the **Data Sets** subtab under the **Data Management** tab. Under the **'Snapshots'** heading choose [Create](#) and you will see the following screen:



Enter the name of the snapshot (e.g. mypool/myfolder@now) and indicate whether child folders should be included, and then click [Create](#). The snapshot is taken immediately, and will appear if you then click [Show](#) under the Snapshots heading on this page.

## 11.2 Creating Checkpoints

Checkpoint is a system snapshot to which system can rollback. Read the full definition [here](#). Checkpoint can be created from any syspool's snapshot by running the following NMC command:

```
nmc:/$ setup appliance checkpoint create-from <snapshot_name>
```

SGI NAS keeps a certain number of checkpoints. Default value is 8. To set up the number of checkpoints to be kept on the appliance, run:

```
nmc:/$ setup appliance nms property upgrade_rootfs_keep
```

If any checkpoint needs to be stored on appliance permanently, run:

```
nmc:/$ setup appliance <checkpoint_name> hold
```

This will mark checkpoint as to be kept permanently. To remove this mark, run:

```
nmc:/$ setup appliance <checkpoint_name> release
```

### 11.3 Setup Periodic Snapshots

SGI NAS can automatically take snapshots of a folder or zvol on a periodic basis. To set up a snapshot schedule in NMV, use the **Auto Services** subtab under the **Data Management** tab. Then choose [Create](#) under the 'Auto-Snap Services' heading and you will be presented with a page similar to the following:

The screenshot displays the SGI NAS web interface. The top navigation bar includes links for 'About', 'Support', 'Add Capacity', 'Register', and 'Help'. The user is logged in as 'Administrator'. The main navigation tabs are 'Status', 'Settings', 'Data Management', and 'Analytics'. Under 'Data Management', there are sub-tabs for 'Data Sets', 'Shares', 'SCSI Target', 'Auto Services', and 'Runners'. The 'Auto Services' sub-tab is active, showing a sidebar with categories: 'All Storage Services', 'Auto-Snap Services', 'Auto-Tier Services', 'Auto-Scrub Services', and 'Auto-Sync Services'. The 'Auto-Snap Services' section is expanded, showing 'Show Snapping Services' and a 'Create' link. The main content area is titled 'CREATE NEW SNAPPING SERVICE' and contains the following form fields:

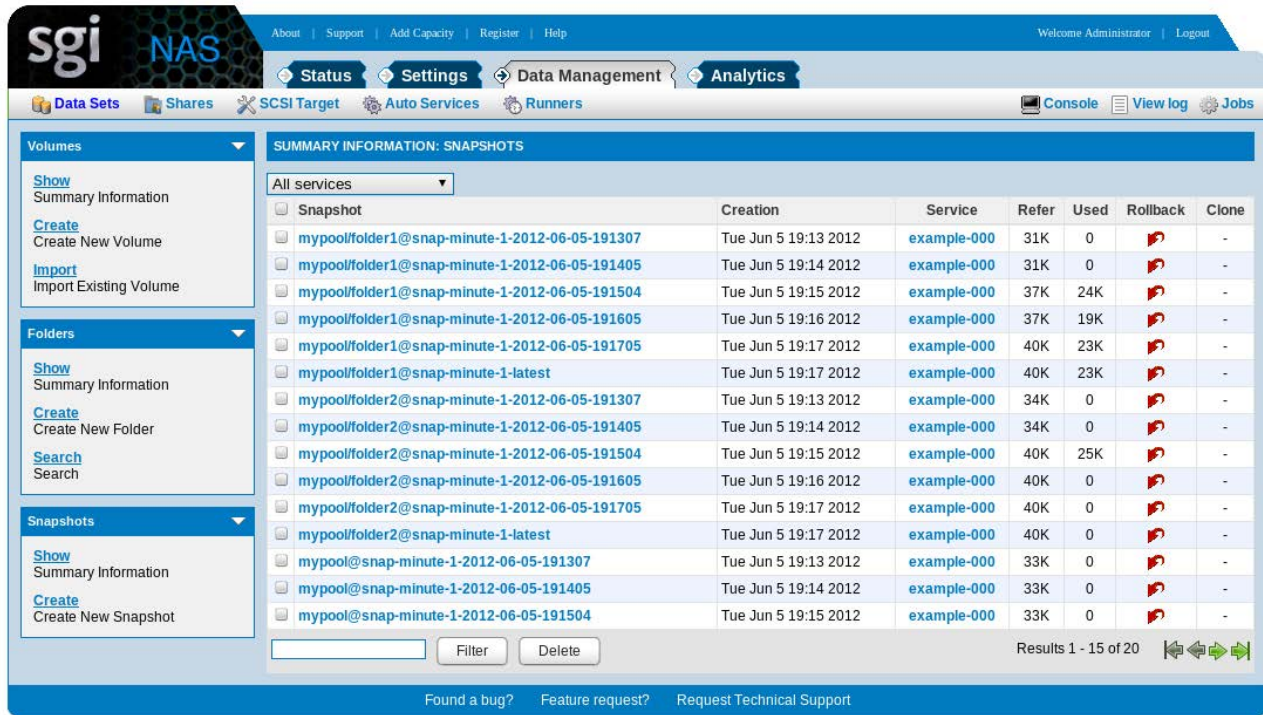
- Name:** A text input field with a placeholder: 'Custom name for your auto-service. Leave this field blank if you want it to be assigned automatically. Can only contain alphanumeric characters (a-z, 0-9) in addition to the following three special characters: underscore (\_), hyphen (-).'.
- Dataset:** A dropdown menu with 'mypool' selected.
- Periodic Interval:** A form with 'Frequency' set to 'hourly' and 'Period' set to '1'.
- Keep Days:** A text input field with '7' entered. Below it: 'Snapshot retention policy. Parameter defines number of days to keep snapshots on the appliance.'
- Mark latest as:** A text input field with 'latest' entered. Below it: 'Custom suffix to mark latest snapshots.'
- Recursive:** A dropdown menu with 'Yes' selected. Below it: 'Snapshot nested folders.'
- Exclude folders:** A text input field. Below it: 'Optional: exclude folders matching the specified UNIX pattern. Only ASCII characters are allowed.'
- Trace level:** A text input field with '1' entered. Below it: 'Show verbose information for service while running. Default value is 1. Bigger value gives more verbosity. For example 10, 20, 30.'

A 'Create Service' button is located at the bottom of the form.

Here you can chose the name of an existing dataset, and specify the schedule: periodic interval, number of days to keep the snapshots, exceptions, trace level. If you need to include subfolders to the snapshots, check recursive. After selecting the desired options, click [Create](#) Service.

## **11.4 View Snapshots**

To view all existing snapshots in NMV, go to Data Sets page and click on [Show](#) under the '**Snapshots**' heading. You will see a page similar to the following:



From this view you can clone a snapshot or rollback a folder or zvol to this point in time.

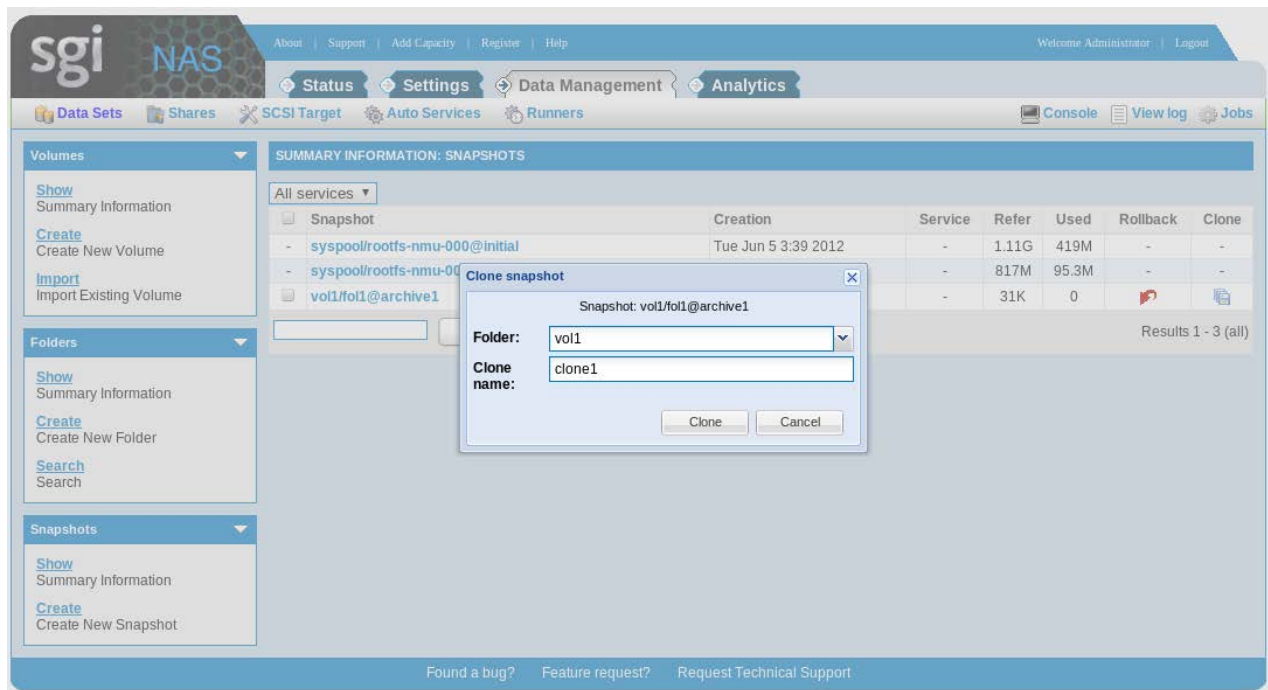
## 11.5 View Scheduled Snapshots

To ensure that you scheduled your snapshots correctly in NMV, you can view all the scheduled snapshots and verify that your new entry is present. To see the summary list, select [Show](#) under 'Auto-Snap Services'.

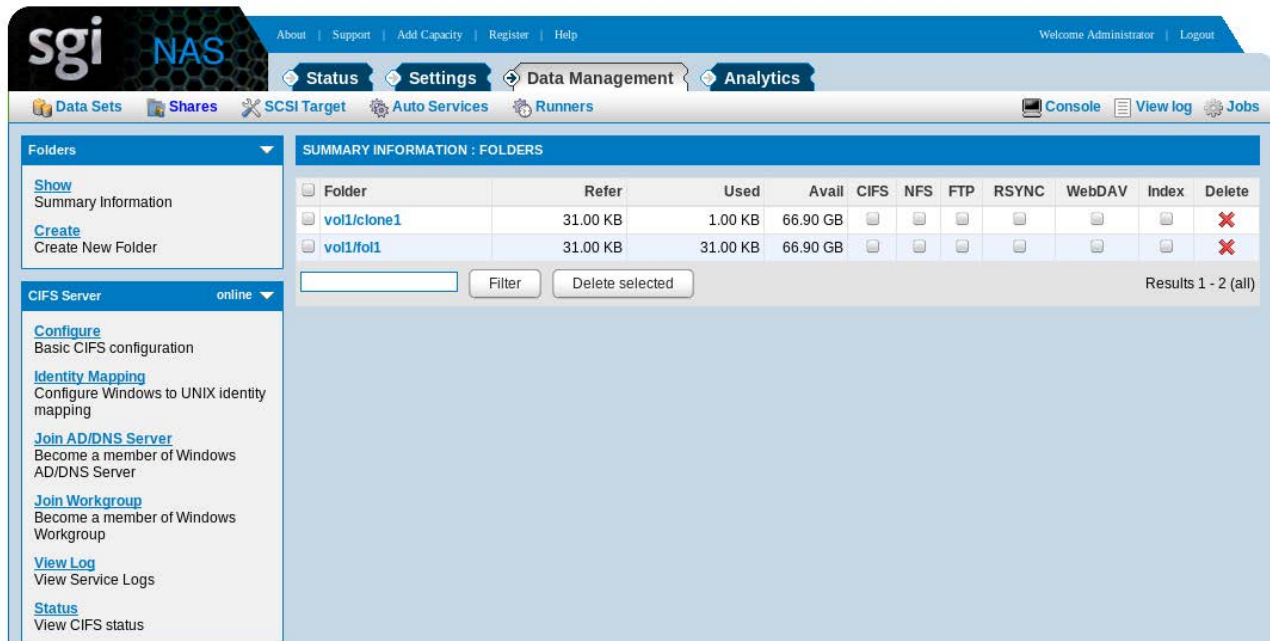
From this view you can delete the periodic snapshot or edit the snapshot service. One useful option available from the edit screen is to take a snapshot immediately.

## 11.6 Recover Snapshot

To recover a file system from a snapshot you first create a clone. Clones are read-write copies of a file system and are based on a snapshot. The origin snapshot cannot be destroyed, if a clone exists. In the example above snapshot is cloned to 'vol1/fol1' as 'clone1'.



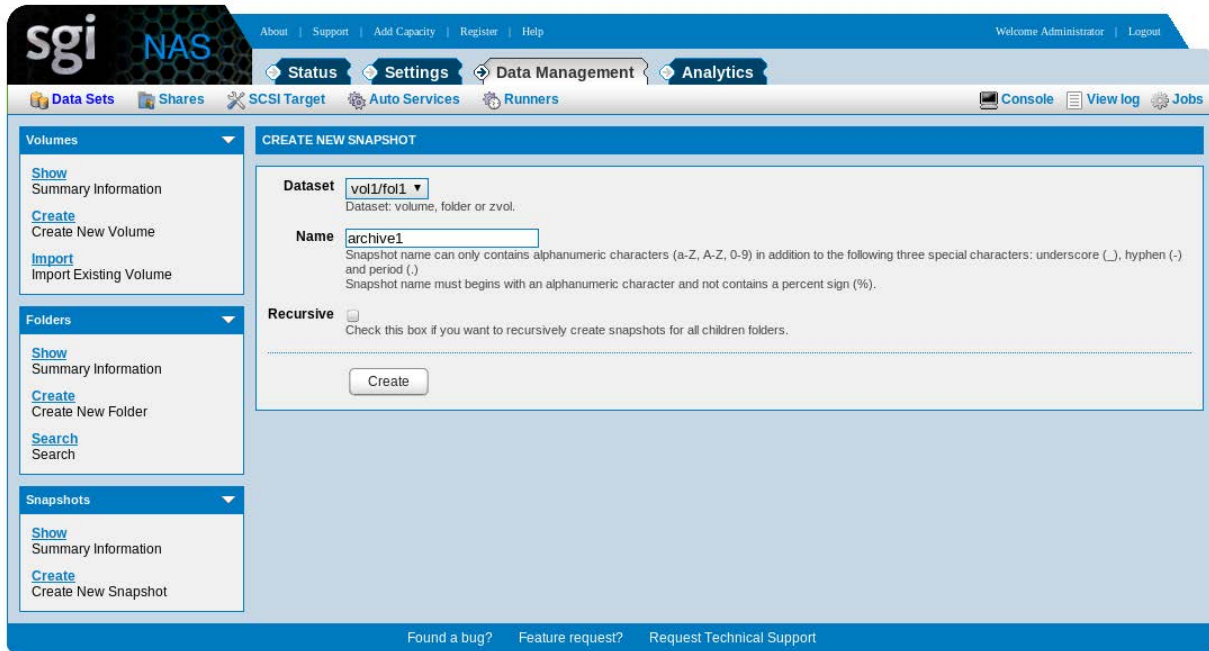
It can be found on 'Folder's summary' page.



There is another simple way to recover data from the snapshot.

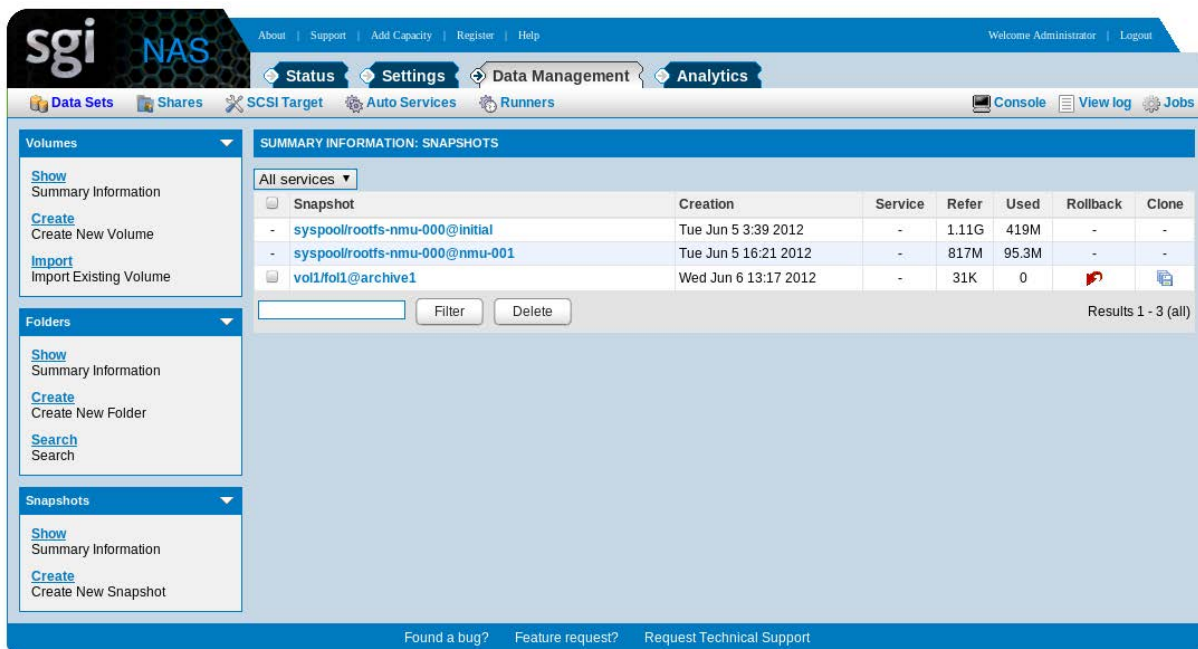
1. For example, a snapshot of 'vol1/fo11' named 'archive1' was created:





It appears on the 'Snapshot summary information' page.

2. Share 'vol1/fo1' via CIFS, if it wasn't previously shared.
3. Click on the folder to edit the properties. Change parameter 'Snapshot visibility' from 'hidden' to 'visible'.



# SGI NAS User Guide

The screenshot displays the SGI NAS web interface for editing folder parameters. The interface is organized into several sections:

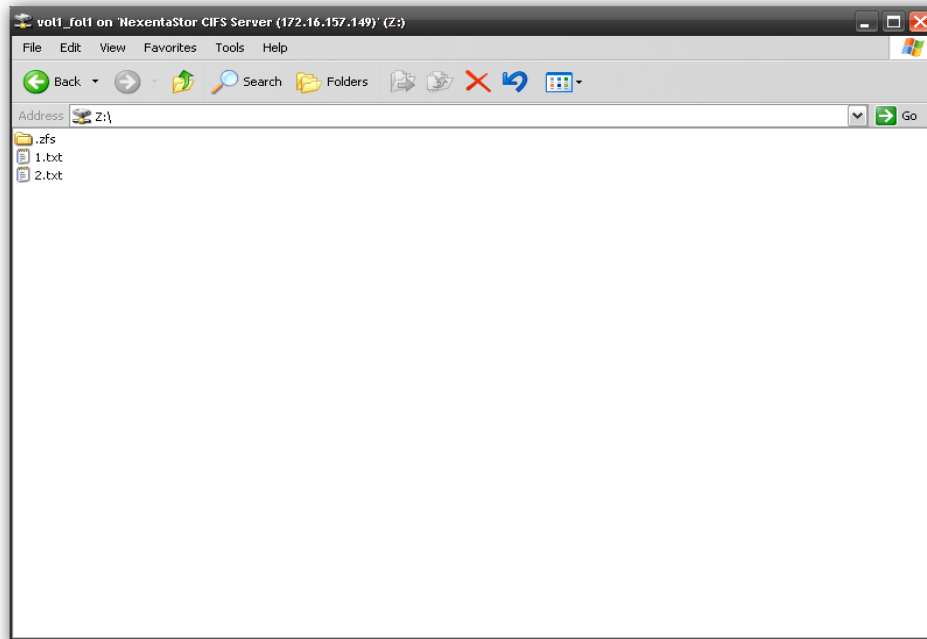
- Navigation Menu (Left):** Contains links for Folders, CIFS Server (online), NFS Server (online), FTP Server (disabled), RSYNC Server (online), and WebDAV Server (enabled). Each server section includes options for configuration, identity mapping, joining servers/workgroups, and viewing logs.
- Top Bar:** Shows navigation tabs for Status, Settings, Data Management, and Analytics, along with utility icons for Data Sets, Shares, SCSTarget, Auto Services, and Runners. A user profile for 'Welcome Administrator' is visible.
- Folder Information (Top Right):** 'EDIT FOLDER: VOL1/FOL1'. A table lists Read-Only Parameters:
 

Name	Value
name	vol1.fol1
creation	Wed Jun 6 13:16 2012
used	60K
available	66.9G
referenced	32K
compressratio	1.00x
mountpoint	/volumes/vol1/fol1
casesensitivity	mixed
- Access Control (Middle):** Shows an Access Control List with one entry:
 

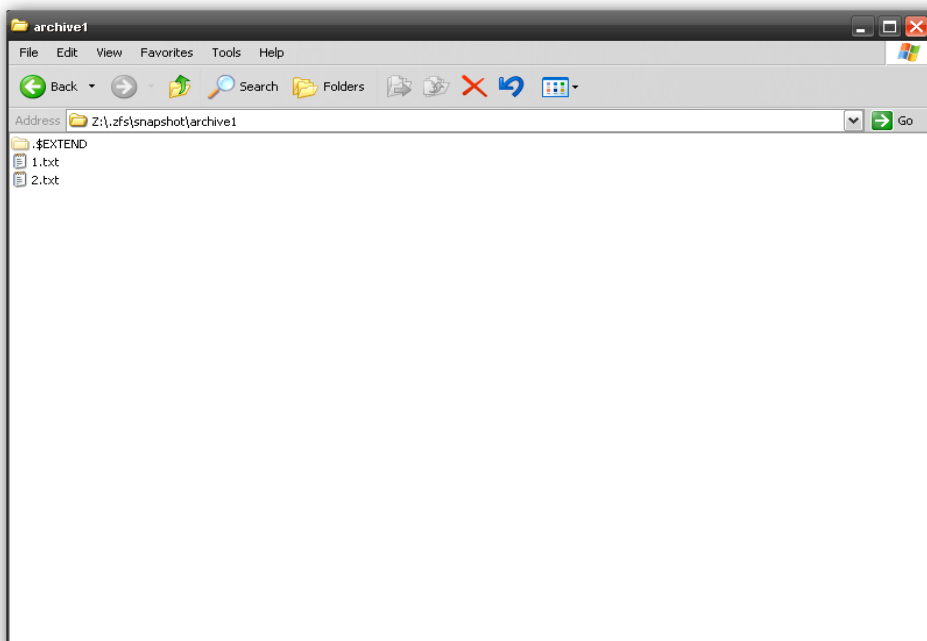
Entity	Allow	Deny	Delete
user: smb	list_directory, read_data, add_file, write_data, add_subdirectory, append_data, read_xattr, write_xattr, execute, delete_child, write_attributes, write_acl, write_owner, synchronize		X
- Read-Write Parameters (Bottom):** A large form with various settings:
  - Description:** Text input field.
  - Quota:** none
  - Referenced Quota:** none
  - Reservation:** none
  - Referenced Reservation:** none
  - Record Size:** 128K
  - Log Bias:** latency
  - Deduplication:** off
  - Checksum:** on
  - Compression:** on
  - Access Time:** on
  - Allow Devices:** on
  - Allow Execution:** on
  - Allow Set-UID:** on
  - Read-Only:** off
  - Snapshots Visibility:** visible
  - ACL Inheritance:** restricted
  - ACL Mode:** discard
  - Can Mount:** on
  - Extended Attributes:** on
  - Number of Copies:** 1
  - Sync:** standard



4. Map Network drive in Windows (read more about CIFS sharing in [9.3 Anonymous access](#)).
5. Access the share and see '.zfs' directory, which was previously hidden.



6. Open \.zfs\snapshot\archive1 and find the read-only copies of the files.



From here the files may be copied and the copies can be later modified.

This method has a limitation and it doesn't work for zvols.

## 11.7 Delete Snapshot

When viewing existing snapshots you can select a snapshot and then delete it. Note that because of the copy-on-write nature of SGI NAS and ZFS, deleting a snapshot will not necessarily free additional space in the folder.

To delete a snapshot go to **Data Management** → **Data Sets** → **'Snapshots'** → [Show](#).

Snapshot	Creation	Service	Refer	Used	Rollback	Clone
mysnapshot						
myspool@archive1	Tue Jun 5 19:25 2012	-	50K	0		
myspool@clone1	Tue Jun 5 19:19 2012	-	33K	0		
myspool@snap-minute-1-latest	Tue Jun 5 19:29 2012	example-000	34K	0		
- syspool/rootfs-nmu-000@initial	Tue Jun 5 3:39 2012	-	1.11G	419M	-	-
- syspool/rootfs-nmu-000@nmu-001	Tue Jun 5 16:21 2012	-	817M	95.3M	-	-

If a few snapshots need to be deleted, you can either select, or filter them by special parameter. The same way as folders.

- Select the snapshots and click on 'Delete selected' to delete selected snapshots.
- Filter the snapshots and click on 'Delete filtered'. Note, that this operation deletes all the filtered snapshots, even if they are not shown on the current page.

Corresponding NMC comand:

```
nmc:/$ setup snapshot <snapshotname> destroy
```

## 11.8 Other operation with snapshots

To rename snapshot, run:

```
nmc:/$ setup snapshot <snapshot_name> rename
```

Example:

```
nmc@myhost:/$ setup snapshot syspool/rootfs-nmu-000@nmu-003 rename
Input new name for snapshot 'syspool/rootfs-nmu-000@nmu-003' : upgradel
Snapshot 'syspool/rootfs-nmu-000@nmu-003' was successfully renamed to 'syspool/r
ootfs-nmu-000@upgradel'
```



## 12 SCSI Target (Managing Blocks)

A SCSI Target is a generic term used to represent different types of targets such as iSCSI or Fibre Channel. SCSI Target accesses all of the different types of targets in the same way and hence allows the same zvol to be exported to any type of target (or to multiple targets at once). Configuring a target means making it available to the system. This process is specific to the type of target being configured.

### 12.1 Create zvol

A zvol is an emulated block device contained within a data volume. Zvols provide an easy way to expose SCSI Targets to hosts. For example, a zvol can serve as the backing store for an iSCSI target. A zvol can also be used as a swap partition.

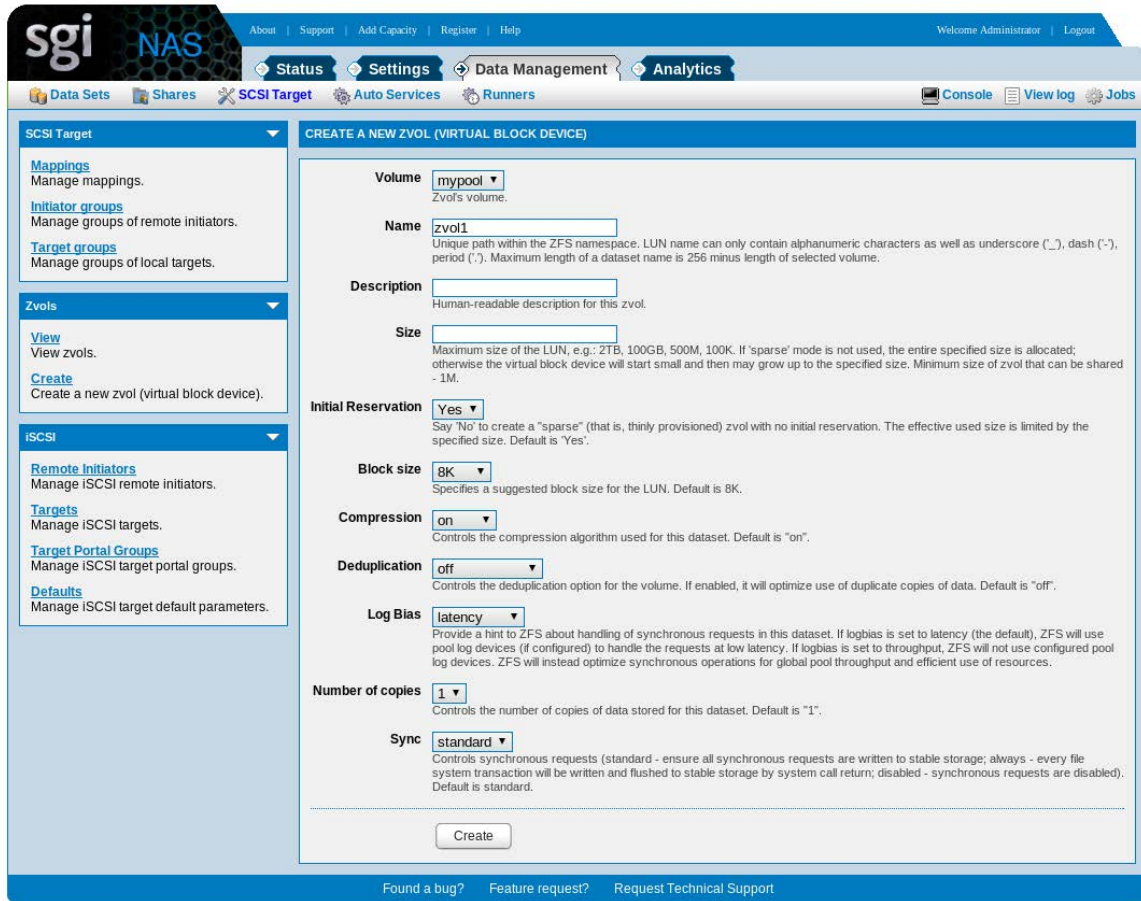
Storage services such as snapshotting and replication can be used with zvols.

Thin provisioning is supported for zvols, meaning that storage space is allocated on-demand. Here is an example using NMC to create a 5TB zvol named zvol1 within the data volume vol1:

```
nmc:/$ create zvol vol1/zvol1 -S -s 5TB
```

Alternatively, you can type 'create zvol' and follow the prompts to complete the request.

In NMV you can create a zvol on the “SCSI Target” page. You will be prompted for the data volume that will contain the new zvol, the zvol name, an optional description, and whether the zvol will have space initially reserved. The block size and maximum size is also specified. You can indicate whether the zvol data should be compressed on the backend storage and how many redundant copies should be stored.



Here is an example in NMC of setting up periodic snapshots for the zvol:

```
nmc:/$ create auto-snap zvol vol1/zvol1
```

You will then be asked to provide the snapshot frequency, retention policy, etc.

Zvol can be thin provisioned, and can be grown over time, both in terms of its effective and maximum size. A thin provisioned (also called 'sparse') zvol does not allocate its specified maximum size. At creation time a thin provisioned zvol actually allocates only a minimum required to store its own metadata.

You can grow both the effective (actually used) size of the zvol by storing more data on it, and the maximum size of the zvol, by incrementing its property called '**volsize**'. In NMC, the latter is done via:

```
nmc:/$ setup zvol <zvol-name> property volsize
```

You can change 'volsize' property anytime.

Be careful shrinking a zvol. If you indicate a zvol size less than it is currently using, it may cause a data loss.

A similar function is available via the NMV web GUI.

## 12.2 View Zvol Properties

If a zvol is being shared over iSCSI and/or FC as a SCSI disk, the writeback caching for that disk can be Enabled or Disabled. When writeback caching is enabled, the disk performs better on writes but the data is not flushed to the backing store of the zpool before a write I/O is completed to the initiator. Disabling writeback caching will always ensure that data is flushed to stable storage before a write is completed. But doing so will reduce the disk write performance.

To control writeback caching select 'SCSI Target → 'View (Zvols)'. Click on the zvol name and its properties will show up. Select the desired writeback caching mode from the drop down list.

## 12.3 Destroy a Zvol

To destroy a zvol in NMC, use the command:

```
nmc:/$ destroy zvol
```

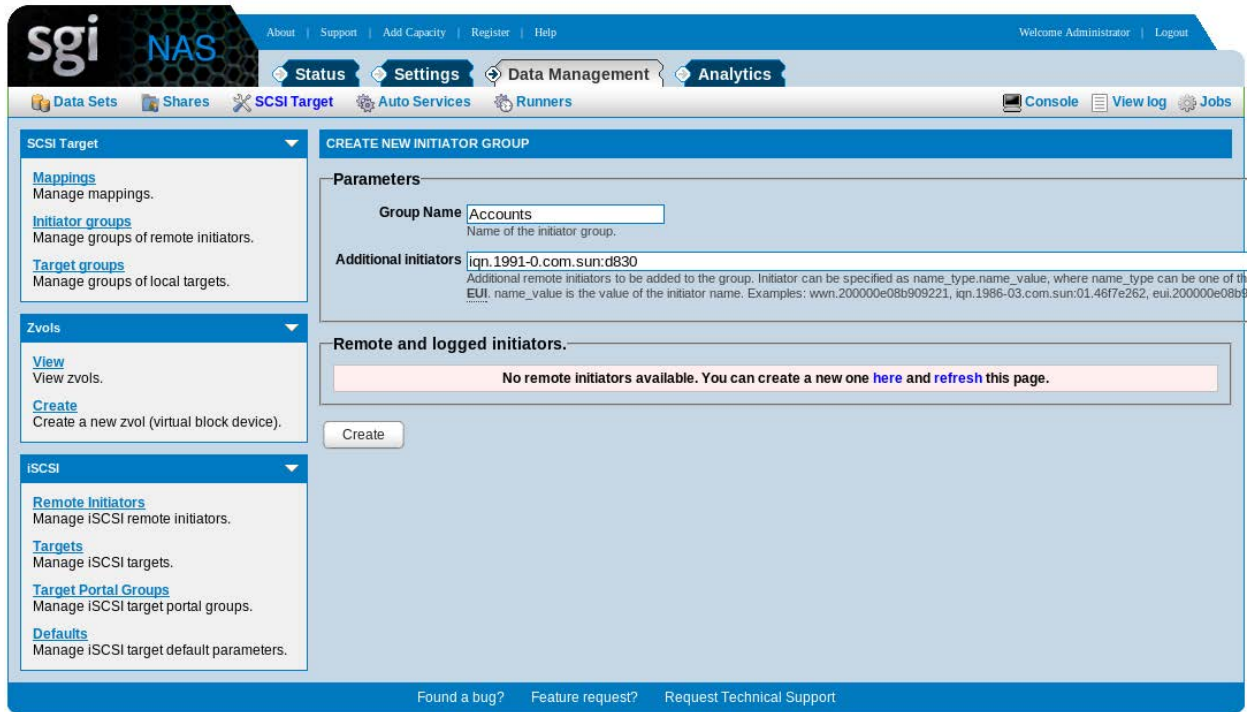
## 12.4 Add initiators and targets

The configuration of initiators and targets is protocol-specific. iSCSI is shipped by default with SGI NAS and is described in the section “Managing iSCSI”. Other protocols are beyond the scope of this document.

## 12.5 Create initiator group

You can share a zvol with all remote initiators. In this case you do not need to create any initiator groups. If you want to control which initiators can see a zvol, then you need to create

one or more initiator groups. Even if you intend to associate only a single initiator with a zvol, the initiator needs to be in an initiator group.



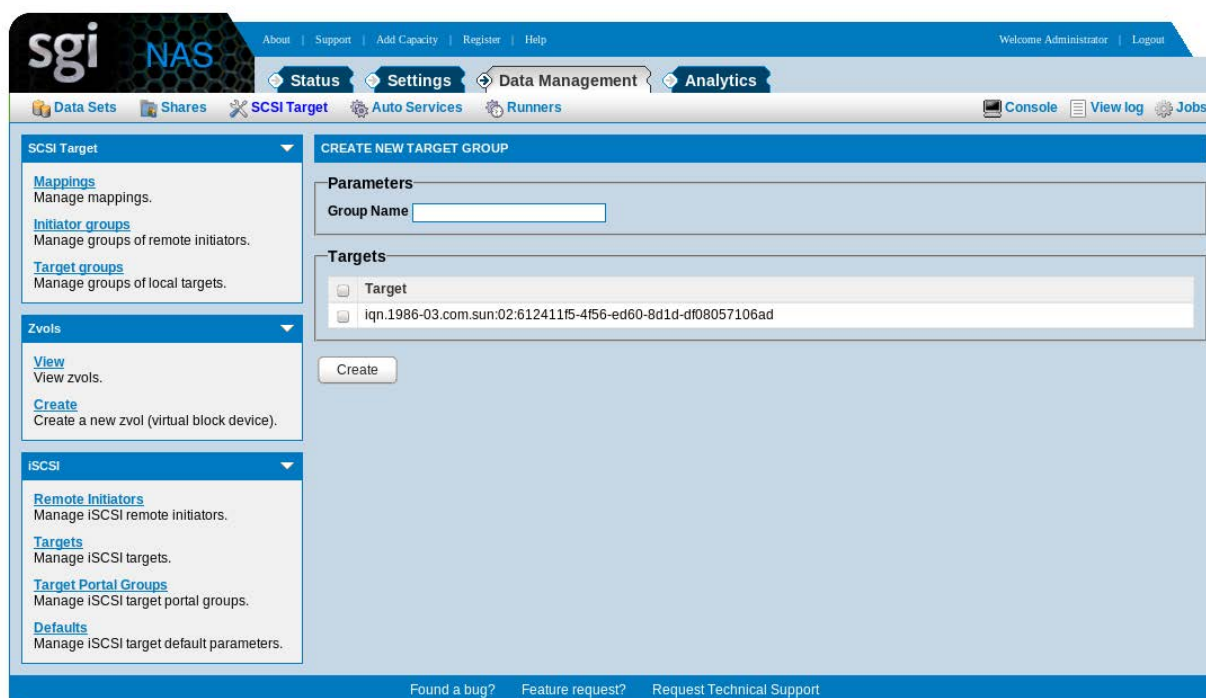
To create an initiator group in NMV, click the link [Initiator Groups](#).

Provide a group name and a list of remote initiators for this group, and then click [Create](#).

## 12.6 Create target group

You can associate a zvol with a set of targets by putting the targets in a target group. Target groups are not required. The following screen in NMV shows how to create a target group. You simply choose a group name and select the targets to be in the group.





## 12.7 Create LUN mappings

LUN mappings allow you to control which remote initiators can see a zvol. A zvol is not accessible over the SAN until it has been mapped.

Here is an example in NMV of creating a LUN mapping for a zvol:

Instead of defining and choosing initiator and target groups, you can simply select “All”. However, remote iSCSI initiators will not find this target if you haven’t defined at least one iSCSI target.

When creating a LUN mapping you can choose a specific LUN id or let SGI NAS assign one automatically.

The screenshot displays the SGI NAS web interface. At the top, there is a navigation bar with the SGI NAS logo and links for 'About', 'Support', 'Add Capacity', 'Register', and 'Help'. On the right side of the navigation bar, it says 'Welcome Administrator' and 'Logout'. Below the navigation bar, there are tabs for 'Status', 'Settings', 'Data Management', and 'Analytics'. A secondary navigation bar contains icons for 'Data Sets', 'Shares', 'SCSI Target', 'Auto Services', and 'Runners'. On the far right of this bar are 'Console', 'View log', and 'Jobs'.

The main content area is divided into two sections. On the left is a sidebar with a 'SCSI Target' dropdown menu. Under this menu, there are three sub-sections: 'Mappings' (Manage mappings), 'Initiator groups' (Manage groups of remote initiators), and 'Target groups' (Manage groups of local targets). Below these are 'Zvols' (with 'View' and 'Create' links) and 'iSCSI' (with 'Remote Initiators', 'Targets', 'Target Portal Groups', and 'Defaults' links).

The right section is titled 'MANAGE MAPPINGS' and contains a 'Mappings' sub-section. A pink message box states: 'No LUN mappings available. You can create a new one here.' Below this message is a 'Create new mapping' dialog box. The dialog box has the following fields: 'Zvol:' (dropdown menu with 'mypool/zvol1' selected), 'Initiator group:' (dropdown menu with 'All' selected), 'Target group:' (dropdown menu with 'All' selected), and 'LUN #:' (text input field with '<AUTO>' entered). At the bottom of the dialog box are 'Create' and 'Cancel' buttons.

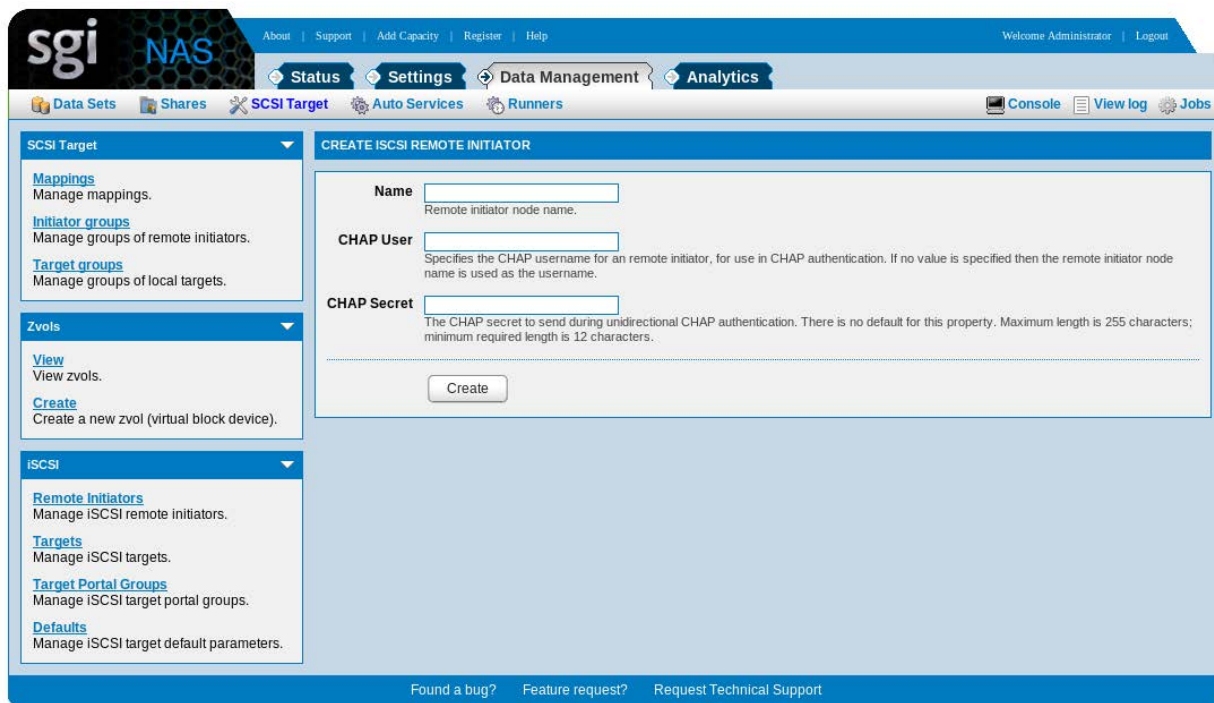
At the bottom of the interface, there are three links: 'Found a bug?', 'Feature request?', and 'Request Technical Support'.

## 13 Managing iSCSI

You can set default authentication and registration options by clicking the “Defaults” under the iSCSI heading. For authentication, you can choose CHAP, RADIUS, or none. You can also indicate whether configured targets should be registered with iSNS.

### 13.1 Add remote initiator

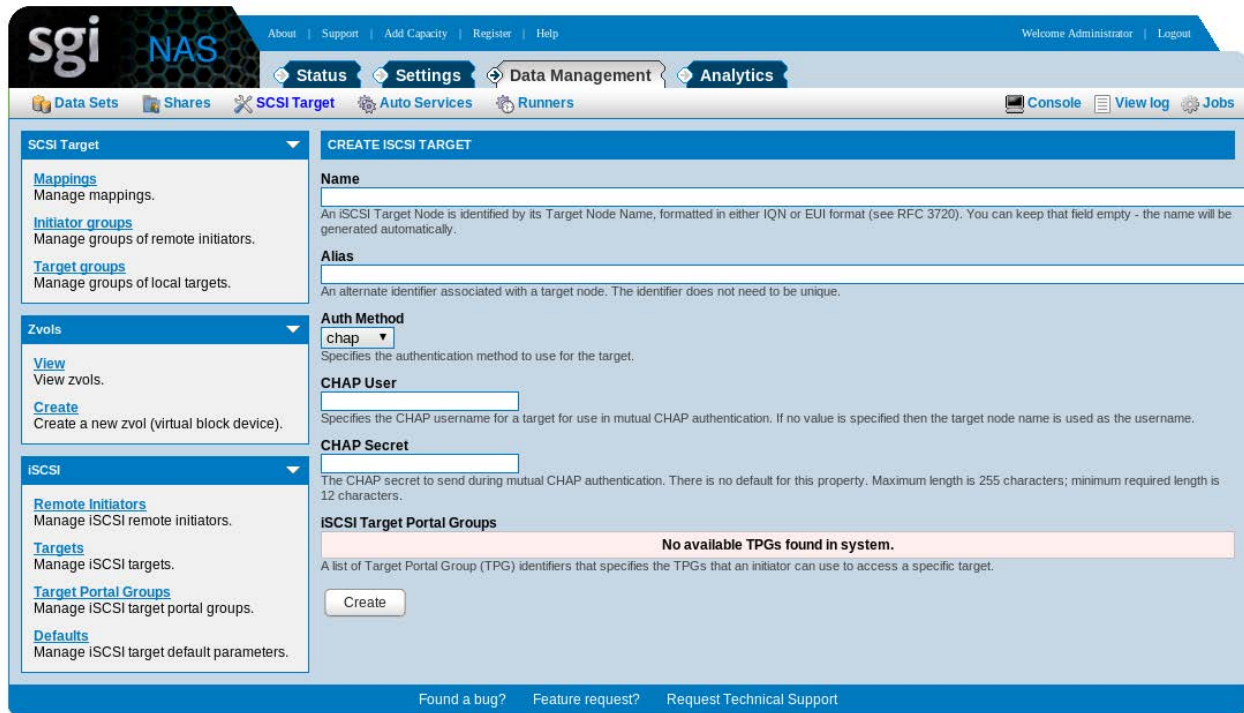
Add remote initiators if you want to control which initiators can see individual zvols. You can specify CHAP authentication information for each remote initiator.



### 13.2 Create iSCSI target

If you map a zvol for iSCSI, the remote initiator won't see it unless you have defined at least one iSCSI target.

In NMV, you can create an iSCSI target by selecting [Targets](#) under the 'iSCSI' heading on the 'SCSI Target page', and then clicking Create. You will see this page:



If the name is left blank then an IQN will be automatically assigned by SGI NAS.

Other fields such as Alias and a CHAP user/secret can also be entered. The CHAP user/secret specified here is used for bidirectional CHAP only. For the non-bidirectional CHAP authentication (The usual case), the CHAP parameters are specified for the initiator on the Initiators page and not here. Finally, clicking on '**Create**' button will create the target.

Unless Multiple Targets are needed for a more advanced configuration, this step is only needed once. All the SCSI LUNs created and exported afterwards, will be exposed on the iSCSI SAN via this target.

### 13.3 Create iSCSI target portal group

A target portal group can be associated with a target to further specify the IP address and port that should be used for a target.

## 13.4 Setting up CHAP Authentication

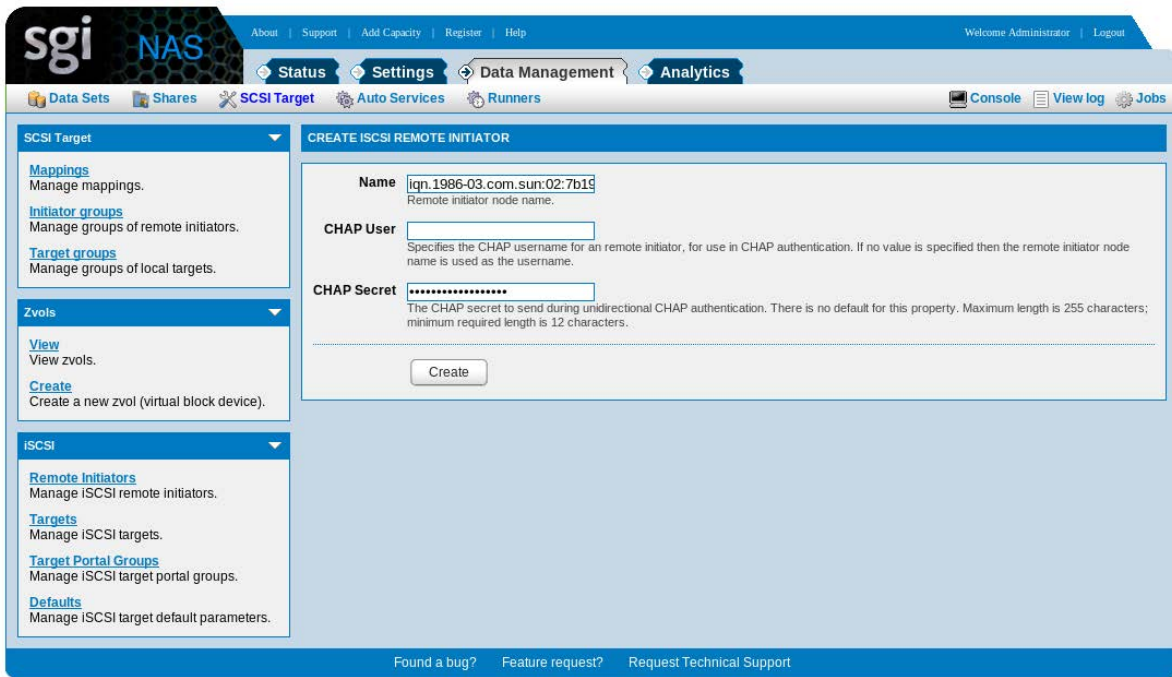
The Challenge-Handshake Authentication Protocol (CHAP) is used for authentication of iSCSI initiators and targets. In standard CHAP, the target authenticates the initiator. With bi-directional CHAP (aka mutual CHAP), the initiator also authenticates the target.

For SGI NAS iSCSI target, when using standard CHAP, the CHAP secret is set on a per initiator basis. That is, for every initiator logging into a SGI NAS iSCSI target (CHAP enabled), the user needs to create an initiator and set its CHAP secret. The CHAP secret for target is only set when using bidirectional CHAP authentication.

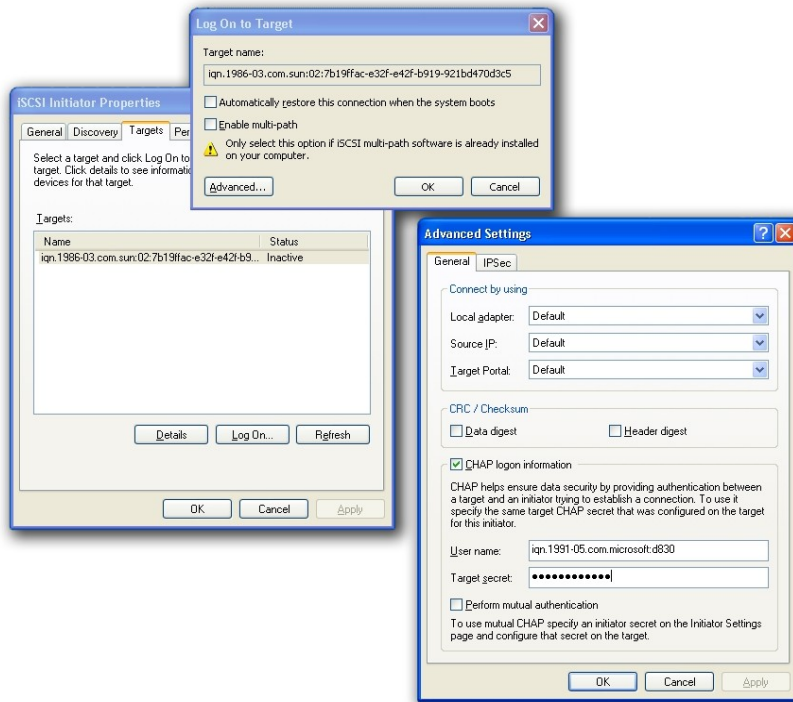
The use case shown below shows how to set the CHAP secret for a Microsoft iSCSI initiator. Here are the steps:

1. Create a target with auth method set to chap. Or if the target is already created, update it and set auth method to chap.
2. Create a remote initiator and provide its CHAP secret (you can also specify its CHAP user name. However for most cases that is not required).
3. Set the same CHAP secret (as set in step #2 above) on the initiator side.
4. To create a target with CHAP on, Create the target as described before and select 'Auth Method' chap from the drop down list. Do not enter CHAP secret here unless you want to use bidirectional CHAP authentication.

Next create the initiator by selecting '**SCSI Target**' → [Remote Initiators](#). Enter the iSCSI name of the initiator. For Microsoft iSCSI initiator, the initiator name is available under 'General' tab of the Microsoft initiator UI. Now enter a CHAP secret for this initiator. The CHAP secret has to be a minimum of 12 characters. Click '**Create**' to create the initiator with CHAP secret.



Now go to the Microsoft Initiator UI and discover this target (This is mostly done by entering the IP address of the SGI NAS appliance under the 'discovery → 'target portal' and then going under the 'Target' tab and clicking 'Refresh').



Select the newly discovered target and click 'LogOn'. In the 'Log On to Target' screen, click 'Advanced...'. Now select Chap Log on information and enter the target secret (same as what was set for the initiator created above). Now, click 'OK' and again 'OK' to logon to the target. At this point the Initiator should be able to logon to the target using CHAP Authentication.





## 14 Asynchronous Replication

SGI NAS provides various types of replication and backup solutions. In the following two section we will review:

- Asynchronous replication (Auto-Tier, Auto-Sync services)
- Synchronous replication (Auto-CDP)

This section describes asynchronous replication.

The continuing growth of disk based storage has had two primary affects. The amount of data to backup is increasingly difficult to fit onto tape or within a backup window, and the costs of capacity of storage makes it feasible to build online backups out of disk subsystems themselves. One of SGI NAS's primary uses is in this new digital archiving role. Whereas tapes will always find use, the development of disk based backup systems regulates tape to the final tier of archiving, where offline preservation is the requirement. You will find this product fits many roles, including primary storage, secondary storage to any primary storage array, and even remote site replication and archival.

What makes multi-tier storage possible in SGI's solution is the '**auto-tier**' service, which can regularly copy data from one source, local or remote of any nature, to a destination target again of any type. The only limitation is that at least one of either the source or destination must be local. In large arrays where the appliance encompasses both first tier and second tier storage, you'll even see local-to-local tiering. Tiering is accomplished by taking a given filesystem or share, breaking into smaller manageable chunks, and replicating that data at that point in time to another volume. Using snapshots at the target end, you can maintain a full efficient backup of the primary storage at unique intervals typical of backups. Whereas you may have hourly and daily snapshots on your primary NAS, auto-tiering with snapshots will generally have daily, monthly, and even yearly snapshot points, with the same policies for retention of any given periodicity.

Commonly tiering service is set up locally, over NFS or CIFS, or using rsync directly with or

without SSH (See '[14.12. How to establish SSH-binding?](#)'). A simple example of tiering data from an NFS file server to our example volume is to create a filesystem to tier to and then to setup an auto-tier from our source NFS server.



Both **auto-sync** and **auto-tier** are schedulable, fault-managed, tunable SGI NAS Data Replication services that can be used in a variety of backup, archiving, and DR scenarios.

Auto-tier and auto-sync are not limited just to the first two tiers, as tertiary tiering for more critical data is also common. As legal and business drivers dictate, tiering will also include access policy enforcement, limiting data access to restricted personnel over longer periods of time.

As either a primary or secondary storage server, SGI NAS can pull or push data regularly at arbitrary intervals, transferring only the periodic changes seen. This can be done as frequently or as sparingly as required, thus being ideal for both large tiering as well as replication needs or for providing WAN-base off site mirroring.

SGI NAS provides the complete range of data replication services:

**Auto-Tier** - In the case of 'auto-tier' (or simply, tiering) service, SGI NAS makes use of snapshots and user definable source and destination points to regularly replicate a single copy of a file system to another storage pool, whether local or remote. Using snapshots on the target end, the tiered copy may have arbitrarily different retention and expiration policies and can be administered separately. SGI NAS tiering service runs on a variety of transports, and can use snapshots as its replication sources. This solution fits the more common backup scenarios found in disk-to-disk backup solutions. The auto-tier service is not limited just to the first two tiers, as tertiary tiering for more critical data is also common. As legal and business drivers dictate, tiering will also include access policy enforcement, limiting data access to restricted personnel to over longer periods of time.

**Auto-Sync** - Another option provided is the '**auto-sync**' (or simply, **syncing**) service, which will maintain a fully synchronized copy of a given volume, file system, or emulated block device (a. k. a. **zvol**, Section "Using ZVOLs") on another NAS. Where tiering provides a copy, auto-sync provides a true mirror, an exact replica of data, inclusive of all snapshots and file system properties. Auto-sync uses the built-in ZFS snapshot capability to efficiently identify and replicate only changed blocks. This allows central mapping of multiple

snapshots of a file system onto remote storage, all the while maintaining control of the retention and expiration of that data at the replication source. This facility is the most ideal for full disaster recovery.



Both auto-sync and auto-tier are designed from ground up to use a **variety of transports** (a.k.a. protocols), which provides required flexibility to execute over Internet and Intranet, from behind a firewall and in the environment that requires extra security.

- Both auto-sync and auto-tier support **any schedule**. You can schedule the services to run every minute, every hour at a given minute of the hour, every few hours, every day at a certain time, etc. You can schedule services to run once a year, or at certain day of every second month, and so on.
- Both auto-sync and auto-tier support all **3 possible directions** of the replication: local to local (**L2L**), local to remote (**L2R**), and remote to local (**R2L**). When replicating to or from remote host, the latter does not necessarily need to be an SGI NAS system, although in the auto-sync case it must be another ZFS based system.



- Both auto-sync and auto-tier provide a combined **replication + snapshots** capability. You can tier from a given source (for instance, from a given snapshot or a directory), and generate snapshot at the remote or local destination every time the replication has run.

As of the version 1.1.6 of the appliance software:

- the services can be set up to **run only once** - at a given scheduled time.
- auto-sync can execute in a **daemon mode** and run incremental replications every second or every few seconds.
- auto-sync can be used to replicate locally or remotely the appliance's system folder (a.k.a. **root filesystem**) that contains appliance's Operating System and configuration. The replication destination may or may be another SGI NAS system, and - in the case when it is an appliance - may or may not reside on appliance's system volume. The equivalent tiering capability is not being planned.

**Hybrid-tier/sync** - SGI NAS provides a hybrid tiering-syncing service which enables a history of changes on the tiering destination. Unlike regular backup solutions with only the latest copy available on the backup target, this solution would have the advantage of both 'the latest copy' and a configurable number of previous copies - the latter in accordance with the retention policy.

In addition, you could tier from a snapshot, which provides the best combination of transactional snapshot at the source combined with potentially faster transport to copy the data without risking that it is being modified concurrently.



The primary difference between auto-sync and auto-tier is threefold:

1. **Data and meta-data.** Auto-sync transfers not only data (files, directories) but filesystem **meta-data** as well, including snapshots.
2. **Folder and Directory.** Auto-tier can have a **directory** within a filesystem as its top level source, while auto-sync cannot. To be able to transfer meta-data, auto-sync must have a **folder** (filesystem) as its top level source.
3. **Copying over.** When executing the very first time, auto-tier can write over the existing files and directories at the destination. When executing the very first time, auto-sync cannot copy over an **existing** destination - it will create new folder(s) at the destination, and keep those folders fully in-sync with the source folders after each subsequent scheduled run of the service. Those new folders will be complete clones of the folders at the source.

Independently of its transport, auto-sync always re-creates source snapshots at the destination.

## 14.1 Auto-Sync

The SGI NAS auto-sync service transfers snapshots between storage systems. The service is built on the ZFS send/receive capability. It assumes that the source and target systems are using ZFS.

Auto-sync replicates dataset snapshots and can be configured to send only the incremental changes. A key advantage is that it also replicates the dataset properties (metadata).

To setup an auto-sync service in NMV, select 'Auto Services' under the **Data Management** tab and then select [Create](#) under the '**Auto-Sync Services**' heading.

The auto-sync service includes a de-duplication capability for the data transfer. Both the sending and receiving systems need to support de-duplication. By using de-duplication you may be able to reduce the amount of data sent across the network. As blocks are sent, if they are a duplicate then only a reference is sent instead of the full data block. This can be especially beneficial over slow or expensive network links. De-duplication is managed separately for each auto-sync replication stream.

## 14.2 Auto-Tier

SGI NAS auto-tier (tiering) service regularly and incrementally replicates a given source directory (and all nested sub-directories except those that are excluded) to another storage pool, whether local or remote. By using snapshots on the destination, this tiered copy may

have arbitrarily different retention and expiration policies and can be administered separately from the source file system. To setup an auto-tier service in NMV, select **Auto Services** under **Data Management** tab and then select [Create](#) under the '**Auto-Tier Services**' heading.

The same operation is available in NMC:

```
nmc:/$ setup auto-tier create
```

### 14.2.1 The table of auto-tier properties

Properties	Description
<b>Name</b>	Optional parameter. User-friendly service name.
<b>Direction</b>	There are three possible directions of replication flow: Local-to-Local (L2L), Local-to-Remote (L2R), Remote-to-Local(R2L).  Note, that in order to create L2R or R2L auto-tier service two hosts need to be SSH-bound to each other beforehand. Read about establishing ssh-bind in SGI NAS User Guide
<b>Destination Folder</b>	Depending on replication direction local or remote folder.
<b>Replicate content*</b>	Replicate content of the folder, but not the folder itself.  This parameter means replicate only source folder's subfolders.
<b>Periodic interval</b>	Consists of two parameters: Period and Frequency.  Frequency values depend on period. For example: Period: hourly, frequency: 4 means run service every 4 hours. Period: daily, Frequency: 05, means every 5 <sup>th</sup> day, etc.
<b>Transport protocol</b>	There four possible options: RSYNC, RCYNC+ssh, RSYNC+tcp, RSYNC+nfs. Read more in Section <a href="#">14.2.1. Auto-tier transport protocols</a>

\* 'Replicate content' property of auto-tier is the same as of auto-sync.

<b>Remote source host</b>	IP address or hostname of ssh-bound destination appliance.
<b>Remote source folder</b>	Destination folder, e.g.: vol1/fo11
<b>Keep days</b>	Number of days to keep snapshots on local appliance.
<b>Trace level</b>	The level of debug information to be shown in auto-tier log file.
<b>Rate limit</b>	Network traffic limitation. Kb/s
<b>RSYNC FanOut/depth</b>	This parameter determines the depth of the auto-sync recursion.
<b>Tiering Snapshot</b>	This parameter lets to start replication from any specified snapshot.
<b>Exclude folders</b>	<p>This option let you expel folders from replication stream. For example, we have vol1/fo11 which contains subfolders 'a', 'b', and 'c'. To exclude 'a' and 'c' from the replication stream, type in specified field:</p> <div style="border: 1px solid gray; padding: 5px; width: fit-content; margin: 10px auto;"> <p>Exclude folders <input type="text" value="a c"/></p> </div> <p>(with quotes!), which will exclude 'a' and 'c' folders from the replication stream.</p>
<b>RSYNC Options</b>	Extra RSYNC options. List of RSYNC arguments which are used to start RSYNC process. Type ' <b>man rsync</b> ' in NMC to get more information (for advanced usage only).
<b>Service retry</b>	This option enables retry the attempt to connect to remote host in case of lost connectivity prior to expiration of the next scheduled interval
<b>Copy ACLs</b>	In some cases auto-tier can copy folder's ACL. This option works only when RSYNC or RSYNC+ssh protocol is used. In case of L2R or R2L replication both hosts must be SGI NAS systems.
<b>Symlink as destination</b>	Enable if you really want to use symlink as destination

	folder. In this case all files in the directory the symlink points to will be overwritten!
--	--



If replicated folder has ACL settings for a custom user (not '**root**' or '**admin**'), this user must exist on both appliances, otherwise the ACLs won't be copied.

## 14.2.2 Auto-tier protocols

Auto-tier uses the following protocols for replication.

- RSYNC

RSYNC is used for local-to-local replication. In that case RSYNC simply synchronizes the files from local directory to another.

- RSYNC+ssh

Used for local-to-remote and remote-to-local replication. During the synchronization local host connects to remote via ssh and starts RSYNC on the remote host. Then local and remote RSYNCs synchronize the directories.

- RSYNC+tcp

Used for local-to-remote and remote-to-local replication. Connection is established with remote RSYNC server. Replication performs via RSYNC-shared folders.

- RSYNC+nfs

Used for local-to-remote and remote-to-local replication. During the synchronization remote nfs-share is mounted on the local appliance. Read more about replication via 'rsync+nfs' in next section.

## 14.2.3 Setting up auto-tier with 'rsync+nfs' transport protocol.

Follow the steps below to set up auto-tier using 'rsync+nfs':

1. Create NFS share on SGI NAS system, where NFS server is set up.

Root

Only root users from the hosts specified in this list have root access. For this option to work, appliance domain name MUST match client's domain name. By default, no host has root access, so root users are mapped to an anonymous user 'nfs:nobody'. The list contains zero or more colon-delimited netgroups, hosts and/or networks/subnets, for instance: netgroup-engineering:10.16.16.92,@nas1.local.net,@192.168.0.1,@192.168.1.0/24. Multiple, individual IP addresses would be specified, for example, as @172.16.132.20:@172.16.134.20.

2. Go to **Data Management** → **Shares** → **'NFS'** → **Edit**. In 'Root' field write the IP address of SGI NAS system which will be used as NFS client to connect to current appliance.
3. Create auto-tier service on SGI NAS NFS client. Indicate the share created in entry 2 as destination folder.
4. Click on created auto-tier service and push **'Run now'** button.

There is also an alternative way to set auto-tier service with 'rsync+nfs' transport protocol:

1. Go to **Data Management** → **Auto-Services** → **'Auto-Tier Services'** → **Edit**, click on **'More Options'**
2. Remove last 'o' in **'RSYNC Options'** field (marked with red square on the picture below):

RSYNC Options

Extra RSYNC options for rsync\*:// protocol.



## 15 Synchronous Replication (Auto-CDP)

SGI NAS provides synchronous replication through an option plug-in module called Auto-CDP (Automatic Continuous Data Protection). Remote mirroring is provided at the block level between two SGI NAS systems.

The following table provides a feature summary:

Feature	Comment
Link neutral	Can use any network adapter that supports the TCP/IP protocol
Reverse replication	The direction of replication can be reversed at any time. The operation is also known as reverse update. The typical scenario includes: 1) failure of a primary volume 2) importing and continued usage of the secondary volume 3) reverse synchronization secondary => primary
Active logging	Continue logging operations whenever the Remote Mirror software is disabled or interrupted.
Multihop sets	Replicate data from one primary volume to a secondary volume; the secondary volume then replicates the data again to another secondary volume, and so on, in daisy-chain fashion.
Mutual backup	Concurrently transmit data copies to and receive data copies from a remote volume. Also known as a <i>bilateral relationship</i> .
Optimized resynchronization	Resynchronize volumes following disk, link, system, and storage platform outages; you only replicate those blocks that were modified since the last synchronization
RAID support	Use RAID volumes as part of your Remote Mirror software strategy. Volumes can be any RAID level.
Well known port and firewall	Port 121 is the default TCP/IP port used by the Remote Mirror software. The firewall must be opened to allow RPC traffic to/from this well-known port address.

## 15.1 Installation

Before installing the Auto-CDP plug-in, please make sure that the data volume that you intend to replicate exists. Also make sure that both appliances are SSH-bound (See '[15.12.How to establish SSH-binding?](#)') and networking connectivity is properly setup.



Even if your data volume is unused, the initial syncing will take a significant amount of time because of the block-level sector-by-sector transfer of all of its (unused) blocks over the IP network.

To verify that Auto-CDP plug-in is available for installation, run the following command:

```
nmc:/$ show plug-in remotely-available
```

To install Auto-CDP plug-in, please run the following command:

```
nmc:/$ setup plug-in install autocdp
```

Note that plug-ins can also be installed using the SGI NAS Web GUI.

The installation will require NMS restart and NMC re-login. After installation, use NMV or NMC to verify that the plug-in is successfully installed. In NMC, that corresponding operation would be:

```
nmc:/$ show plug-in autocdp
```

The command will display the plug-in version, as well as other useful information.

## 15.2 Getting Started

Creating Auto-CDP service instances involves the following steps:

1. Select local (**primary**) data volume to replicate. The name of the service instance is in form ":volname".



Assuming there is a volume named '**vol1**', the corresponding Auto-CDP service will be named **:vol1**.

Here, and throughout the rest of this document, **:volname** indicates the name of the volume to block-mirror using Auto-CDP, and simultaneously, the name of the corresponding Auto-CDP service.

Note, Auto-CDP service cannot be created for **syspool** – the appliance's system volume.

1. Select remote appliance. Specifying existing SSH-bound appliance registered on local appliance-creator;
2. Select disks on the remote appliance to serve as block-level replicas of the disks of the local (primary) volume.

As always, to carry out the 1-2-3 steps, NMC provides a guided multiple-choice interactive environment. As always, the same steps can be executed via command line, using the options specified above.

Once initiated, Auto-CDP will transfer the local (primary) volume's metadata, which will effectively create a secondary (remote) volume out of the corresponding remote disks.

The appliance's Auto-CDP will keep both data and ZFS metadata on the replicated disks in-sync, at all times.

Note: Auto-CDP requires using either DNS hostname for the local and remote appliances, or their 'replacement' via local host tables. See 'setup appliance hosts -h' for more information.

The following NMC wizard command can be used for service instance creation:

```
nmc:/$ create auto-cdp
```

For details on service creation and all supported command line options, please see the corresponding man page:

```
nmc:/$ create auto-cdp -h
```

To modify/show parameters of newly created Auto-CDP service use the following commands:

```
nmc:/$ setup auto-cdp <volname>
```

where “:**volname**” is a service instance, and simultaneously, the name of the volume that is being block-mirrored via (or by) Auto-CDP.

### 15.3 The alternative hostname

The alternative local hostname can be specified instead of the one automatically selected by the Auto-CDP service.

#### **Background:**

Auto-CDP identifies the primary <=> secondary nexus by the specified IP addresses and their corresponding fully qualified hostnames. In a simple case there is a single networking interface and appliance's fully qualified domain name (FQDN). If the appliance has multiple IP interfaces, an attempt is made to find the matching primary IP to be used for the specified secondary IP address. The logic to find the best matching pair of interfaces for the CDP data transfers is built-in.

To override the default logic and specify an alternative hostname, use the **-H** option as shown below during service creation command:

```
nmc:/$ create auto-cdp -H althost
```

where “althost” is the alternative hostname to be used for the Auto-CDP service instance created.

### 15.4 Enabling Auto-CDP service instance

After Auto-CDP service instance was successfully created the wizard will ask permission on service enable. Service can be enabled/disabled at any given time using the following command:

```
nmc:/$ setup auto-cdp :volname enable
```

This enables the remote mirror replication for the primary volume and also uses the remote mirror scoreboard logs to start the resynchronization process so that the corresponding secondary data volume becomes a full replica of the primary volume.

Sizes of the remote (**secondary**) disks (a. k. a. LUNs) must be equal or greater than the corresponding primary disks that are being replicated.

Once enabled, SGI NAS Auto-CDP service instance will update a remotely mirrored data volume. Only the blocks logged as changed in the remote mirror scoreboards are updated.

Use '-f' (force) option when the primary and the secondary volumes/luns might be different and no logging information exists to incrementally resynchronize the volumes/luns.

Reverse synchronization and DR (disaster recovery)

At some point in time the secondary setup will be used as a disaster recovery (DR) site. There are two scenarios which needs to be considered while failing over to secondary setup:

1. Primary site is still active and you just want to manually switch to the secondary for primary maintenance operations. The assumption is that Auto-CDP service instance was in replication mode. In this case, the following command needs to be executed on primary:

```
nmc:/$ setup volume volname export
```

The command above will gracefully disable instance “:volname” after export completes. After command is complete, the following command needs to be executed on secondary:

```
nmc:/$ setup volume import volname
```

The “:volname” instance will stay in logging mode and once primary site is up the export/import operations can be repeated. After import command is complete on primary, execute the following command to enable reverse synchronization back from secondary to primary:

```
nmc:/$ setup auto-cdp :volname enable -r
```

The '-r' option (reverse) used to reverse the direction of synchronization: that is, synchronize

from the secondary volume to the primary. With the '-r' option the primary volume becomes a passive destination while the secondary volume is considered active source (of the changes).

2. Primary site is down and you need to forcefully switch to secondary. The graceful Auto-CDP service instance disable is not possible in this case and loss of data may occur. However, the filesystem on a secondary setup is going to be always in consistent state due to transactional nature of ZFS and synchronous mode of SNDR operation. On the secondary setup, execute the following command to forcefully import data volume:

```
nmc:/$ setup volume import volname -f
```

The '-f' option will disregard host checking and forcefully import data volume on a secondary setup. The rest of operations could be done similar to (1).

The reverse operation '-r' then resumes Remote Mirror replication of new updates from the primary volume to the secondary volume automatically so that the volume sets remain synchronized. We recommend to quiesce the workload to the volume sets during the restore/refresh operation. This action ensures that the primary and secondary volumes match before replication of new updates resumes.

## 15.5 Volume operations and Auto-CDP

Auto-CDP service tightly integrated with main NMC/NMS management commands which simplifies service maintenance. The following data volume operations supported:

1. Volume grow. Add new LUN or group of LUNs to an existing data volume will interactively invoke Auto-CDP NMC wizard for disk pair setup;
2. LUN removal or replacement will ensure that appropriate Auto-CDP NMC wizard is invoked
3. Volume export will ensure that corresponding Auto-CDP instance is disabled if desired;
4. Volume import will ensure that corresponding Auto-CDP instance is enabled if desired;
5. Volume destroy will ensure that corresponding Auto-CDP instance is removed;
6. Simple-Failover service ensures that Auto-CDP configuration is securely transferred to

all machines with-in the simple-failover group and automatically activates Auto-CDP service on failover machine.

## 15.6 Service monitoring

There are number of ways to monitor Auto-CDP service health and statuses:

1. Automatic SGI NAS trigger 'autocdp-check'. As any other SGI NAS trigger it will send notification events (e-mails) if some of services instances suddenly enters logging mode or when initial syncing enters into normal replication mode;
2. AVS level monitoring of service status and I/O progress. The following command can be used:

```
nmc:/$ show auto-cdp :volname iostat
name                t s   pct   role rkps rtps wkps wtps
v/rdsk/c2t0d0s0    P L  0.05 net    8   3    0    0
v/rdsk/c2t1d0s0    P L  0.40 net    0   0   345    0
```

The output includes the following information:

```
t          Volume type
s          Volume status
pct        Percentage of volume requiring sync
role       Role of the item being reported
rtps       Number of reads
wtps       Number of writes
rkps       Kilobytes read
wkps       Kilobytes written
```

3. Service-generated socket-level network traffic. The following command can be used:

```
nmc:/$ show auto-cdp :volname stats
```

4. Service logs can be viewed by executing the following command:

```
nmc:/$ show auto-cdp :volname log
```

5. To monitor logging activity interactively, execute the following command:

```
nmc:/$ show auto-cdp :volname logtail -f
```

## 15.7 Auto-CDP configuration properties

Auto-CDP service provides the following configurable properties:

1. **cdproot** - AutoCDP internal database root folder. Location where NMS and Auto-CDP service will store its internal data (AVS bitmaps devices). Default is "syspool/.cdp". This option allows user to create dedicated volume to store zvol-emulated bitmap devices and isolate system volume;
2. **cdp\_scoreboard\_log\_size** - specifies Auto-CDP scoreboard log size. Default is '1G' (one gigabyte). The AVS bitmap devices. Zvol-emulated and thin-provisioned.

## 15.8 Service States

The following are the states that can be set to the operational service instance:

- **volume failed (VF)** - an I/O operation to the local data volume has failed ;
- **bitmap failed (BF)** - an I/O operation to the local bitmap volume has failed ;
- **need sync (SN)** - a sync to this volume has been interrupted. It needs to be completed. The direction of the data flow must not be changed until one or the other is done;
- **need reverse sync (RN)** - a reverse sync to this volume has been interrupted. It needs to be completed (or restored via Point-in-Time Copy). The direction of the data flow must not be changed until one or the other is done;
- **logging (L)** - incoming writes are logged in the bitmap devices only. Data is not replicated to the remote site. **need sync** and **need reverse sync** are all sub-states of **logging** such that writes are logged in the bitmap, but not replicated;
- **reverse syncing (RS)** - a secondary to primary copy is in progress;
- **syncing** – a primary to secondary copy is in progress.



## 15.9 Troubleshooting

To troubleshoot, execute either one of the following commands:

1. To re-enable the entire service corresponding to instance “:volname”:

```
nmc:/$ setup auto-cdp :volname enable
```

2. To re-enable the entire service and fully resynchronize the associated primary volume 'volname' to secondary volume. Beware, this operation may take a long time to complete:

```
nmc:/$ setup auto-cdp :volname enable -f
```

3. To re-enable specific <LUN> pair for service instance “:volname”:

```
nmc:/$ setup auto-cdp :volname lun <LUN> enable
```

4. To re-enable and fully resynchronize the specific <LUN> pair for service instance “:volname”:

```
nmc$ setup auto-cdp :volname lun <LUN> enable -f
```

5. To reset all Auto-CDP services and re-initialize AVS databases on both sides active and passive:

```
nmc$ setup auto-cdp reset
```



WARNING! This operation will reset Auto-CDP service to its initial (post-creation) state.

6. As a troubleshooting example, to replace failed drive just run standard volume command:

```
nmc:/$ setup volume volname replace-lun
```

The major difference between all of the commands listed above is: **granularity**.

The first two commands (1,2) execute on a level of the entire service instance, with the corresponding action applied to all associated disk pairs. Use command (1) if you want to move service from “logging” mode back to “replication”.

The second pair (3,4) of troubleshooting actions is LUN specific. These two commands (3,4) are especially useful when a single or a few specific pairs of syncing LUNs appear to be stuck in a so called 'logging' mode and will not change states. Another relevant scenario is related to importing of the secondary volume. If the newly imported mirrored volume shows faulted drive(s), use the LUN specific re-synchronization to troubleshoot.

The reset command (5) is plug-in/service wide and will affect all instance on both active and passive sides.

The disk replacement, as well as all other disk management operations, tightly integrated with the service and the right action will be taken if corresponding Auto-CDP instance is present. Simply execute normal volume management operation and do not worry about the complexity associated with AVS disk set management.

## 15.10 Creating Auto-CDP – example

In the following example appliance 'testbox1' is a primary, 'testbox2' is a secondary. This example includes all NMC prompts – it is a complete demonstration of auto-cdp creation:

```
nmc@testbox1:/$ create auto-cdp
Remote appliance : 192.168.37.128
Remote for c2t1d0 : c2t1d0
Remote for c2t0d0 : c2t0d0
Creating new Auto CDP service 'auto-cdp:vol1', please wait...
Successfully created service 'auto-cdp:vol1'
Enable it now? Yes
Enabling service, please wait...
PROPERTY          VALUE
name              :vol1
max_q_fbas        16384
autosync          off
max_q_writes      4096
async_threads     2
state             syncing
to_host           testbox2
from_host         testbox1
```

```

type                active
TESTBOX1            TESTBOX2
c2t1d0              =>    c2t1d0
c2t0d0              =>    c2t0d0
The local host is 'active' auto-cdp node.

```

Once the initial synchronization between a pair of active (primary) and passive (secondary) volumes commences, you can monitor it by the following NMC commands:

```
nmc:/$ show auto-cdp <name> stats
```

or

```
nmc:/$show auto-cdp <name> iostat
```

In fact, these two commands are always useful, in terms of monitoring the data replication traffic, whether it is auto-cdp, auto-sync or auto-tier service. However, auto-cdp traffic monitoring is particularly useful at the time of the initial block-level [syncing](#).



It is recommended not to use the primary (active) volume during the very first (the initial) CDP synchronization. Any updates on the primary during this period of time may considerably delay the initial synchronization. Note also that during this phase a major part of the available I/O bandwidth is used by the auto-cdp service, which is yet another reason to let it run through as soon as possible.

See '`show auto-cdp <name> stats`' for more information.

Following is an example of the output:

```

nmc@testbox1:/$ show auto-cdp :vol1 stats -i 1

TCP CONNECTIONS                SNEXT      RNEXT      TRANSFER
192.168.37.128.1022-192.168.37.134.121  1313611534 3140553278 1.60 MB
192.168.37.128.1022-192.168.37.134.121  1314180374 3140554114 569.68 KB
192.168.37.128.1022-192.168.37.134.121  1314838374 3140554994 658.88 KB
192.168.37.128.1022-192.168.37.134.121  1316976874 3140557854 2.14 MB
192.168.37.128.1022-192.168.37.134.121  1321352574 3140563706 1.25 MB
192.168.37.128.1022-192.168.37.134.121  1327471974 3140571890 955.38 KB
192.168.37.128.1022-192.168.37.134.121  1328722174 3140573562 1.25 MB...

```

Once the traffic stops, you'll be able to see the block-level replicated volume on the remote side:

```
nmc@testbox2:/$ show auto-cdp :vol1 -v
PROPERTY          VALUE
name              :vol1
max_q_fbas        16384
autosync          off
max_q_writes      4096
async_threads     2
state             logging
to_host           testbox2
from_host         testbox1
type              passive
TESTBOX1          TESTBOX2
c2t1d0            =>    c2t1d0
c2t0d0            =>    c2t0d0
The local host is 'passive' auto-cdp node.
```

## 15.11 Reverse mirroring – example

In the following 6-steps example appliance 'testbox1' is again a primary, and 'testbox2' is a secondary. The reverse mirroring starts from exporting a volume from the primary appliance (Step #1)...



One critically important guideline in re CDP:

It is recommended NOT to have the primary and secondary volume imported simultaneously. In fact, SGI NAS software will prevent this from happening.

Also note: the remotely mirrored volume may be imported only at one side, primary or secondary, at any given moment.

In short, several preparation steps need to be performed before actually enabling reverse mirroring from 'testbox2' to 'testbox1' (Step #5 below):

```
Step #1. testbox1 (primary): first, export vol1
nmc@testbox1:/$ setup volume vol1 export
Export volume 'vol1' and destroy all associated shares ? Yes

Step #2. testbox2 (secondary): import vol1
```

```
nmc@testbox2:/$ setup volume import voll
volume: voll
state: ONLINE
scrub: none requested
config:
  NAME          STATE      READ WRITE CKSUM
  voll          ONLINE     0    0    0
  mirror        ONLINE     0    0    0
    c2t0d0      ONLINE     0    0    0
    c2t1d0      ONLINE     0    0    0

Step #3. ...using secondary volume until (and if) the problem with primary is
resolved...

Step #4. testbox2 (secondary): export voll
nmc@testbox2:/$ setup volume voll export
Export volume 'voll' and destroy all associated shares ? Yes

Step #5. testbox1 (primary): reverse syncing
nmc@testbox1:/$ setup auto-cdp :voll enable -r
Enable reverse synchronization for auto CDP service 'voll'? Yes
Enabling service, please wait...
Auto CDP service ':voll' enabled.

Step #6. testbox1 (primary): import voll
nmc@testbox1:/$ setup volume import voll
volume: voll
state: ONLINE
```

## 15.12 How to establish SSH-binding

Tiering over SSH, syncing over SSH, and managing remote appliance over SSH – these are the features in use by the appliance that rely on pre-existing trusts being established between multiple devices, whether source or destination data volumes, or other appliances. SSH binding allows for remote management as well as strong cryptographic verification of multi-host trusts, forming a network of ssh-bound hosts.

The following commands manage hosts binding and unbinding:

```
nmc:/$ setup network ssh-bind
```

and

```
nmc:/$ setup network ssh-unbind
```

To view currently bound hosts, run:

```
nmc:/$ show network ssh-bindings
```

To see more detailed description including all ssh-bound appliances hostnames, run this command in verbose mode:

```
nmc:/$ show network ssh-bindings -v
```

A simple example to bind to **server2** is:

```
nmc:/$ setup network ssh-bind server2
```

resulting in a prompt for server's root password. You'll be prompted in a similar way to unbind a host. If the host is no longer up, unbind can be forced with '-f' flag:

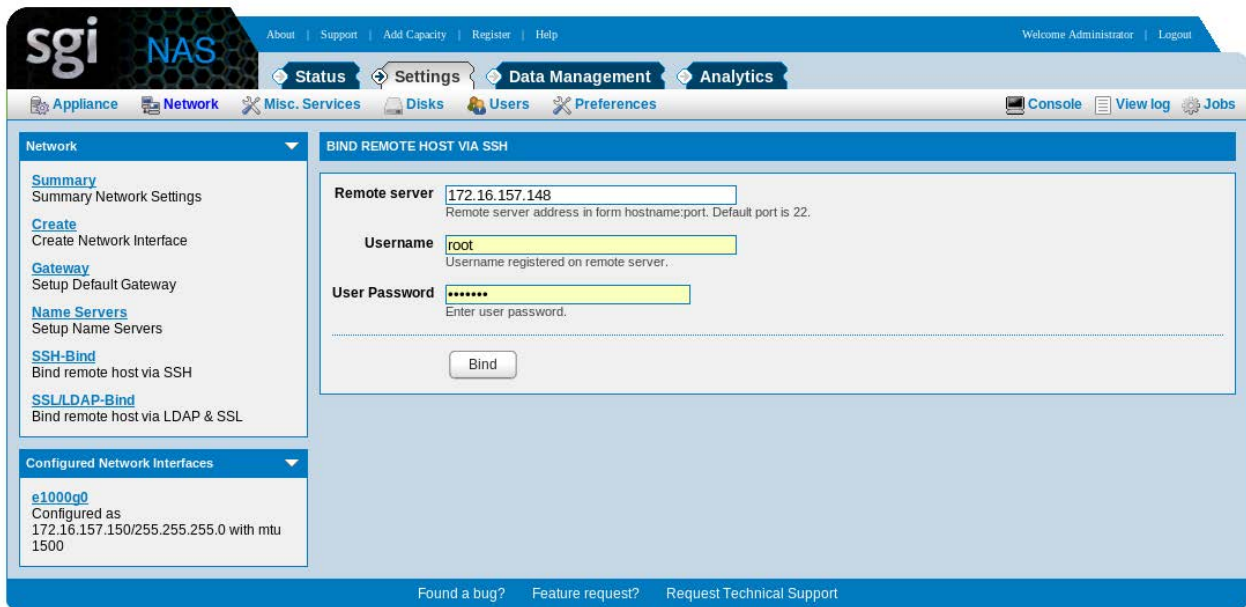
```
nmc:/$ setup network ssh-unbind server2 -f
```

For more information, run

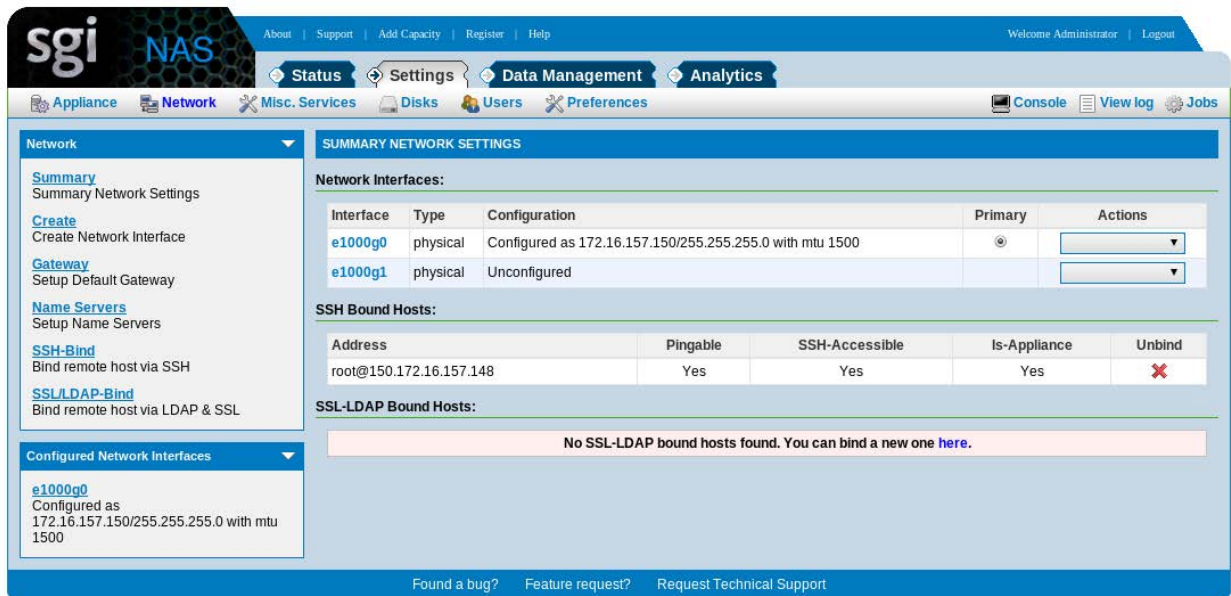
```
nmc:/$ setup network ssh-bind -h
```

NMV provides the same functionality. Go to **Settings** → **Network** and click on [SSH-bind](#) in the left '**Network**' pane:

# SGI NAS User Guide



Specify remote server IP address, username and password and click 'Bind'. On the next screen remote host 172.16.157.148 appears in the list of SSH bound hosts:



## 15.13 SSH-bind troubleshooting

If binding fails, you may receive one of the following error messages:

```
com.sginas.nms.ConnectFailure:unknown host 'host.example.com'
```

Possible problem:

- Name of the remote server is invalid.
- DNS server is down. To change a DNS server, run the following command:

```
nmc:/$ setup network nameservers
```

To find an equivalence between host name and IP address, run:

```
nmc:/$ setup appliance hosts
```

```
com.sginas.nms.ConnectFailure: host 'host.example.com' is unreachable
```

Possible problem:

- Remote host is down.
- Firewall on the remote host is blocking all the incoming connections.

```
com.sginas.nms.ConnectFailure: host 'host.example.com' is alive but drops incoming TCP connections on port 22
```

Possible problem:

- Firewall on the remote host or on the gateway is blocking the requests to the specified port.
- SSH-daemon accepts connection from the other port. In that case, run the following NMC command to set up specified port:

```
nmc:/$ setup network ssh-bind host.example.com:portnumber
```

```
com.sginas.nms.ConnectFailure: unable to connect to host 'host.example.com' port '22': Connection refused
```

Possible problem:

- Ssh daemon is not running on the remote server.



- Different port is used for incoming connections. In that case run the following NMC command to set up needed port:

```
nmc:/$ setup network ssh-bind host.example.com:portnumber
```

```
com.sginas.nms.OperationFailed:timed out waiting for ssh server on host  
'$host' to respond ssh connection
```

- To troubleshoot this error, add the following line to `/etc/ssh/sshd_config` (Warning! Only if GSSAPI is not used):

```
GSSAPIAuthentication no
```

```
com.sginas.nms.AuthenticationFailed: permission denied while accessing  
root@host.example.com:~/ssh
```

Possible problem:

- Wrong password was entered.
- Remote host is wrongly configured. Check if root login is permitted in `/etc/ssh/sshd_config`: 'PermitRootLogin yes'.
- No write permission to `.ssh` on the remote host or to file `.ssh/authorized_keys`. To check it, use the following NMC command:

```
nmc:/$ option expert_mode = 1  
nmc:/$ !ssh [-p 2222] jack@host.example.com 'ls -la ~/.ssh'  
total 36  
drwx----- 2 root root 4096 Jan 7 01:13 .  
drwx----- 37 root root 4096 Mar 31 19:05 ..  
-rw----- 1 root root 1769 Mar 17 01:04 authorized_keys  
-rw----- 1 root root 1675 Nov 5 10:50 id_rsa  
-rw-r--r-- 1 root root 393 Nov 5 10:50 id_rsa.pub  
-rw-r--r-- 1 root root 13428 Mar 19 20:29 known_hosts
```

Note, that all the information mentioned above is relevant for SGI NAS version 3.0.4 and higher.



## 16 Using NDMP with SGI NAS

NDMP, or **Network Data Management Protocol**, is a networking protocol (and simultaneously, an open standard – see NDMP the Open Standard web site) for backing up data in a heterogeneous environment.

The task of NDMP is to back up networks of heterogeneous file servers, including dedicated NAS systems, under control of any NDMP compliant backup applications. NDMP supports a three-way copy: that is, copying data between two remote servers directly, without having the data 'traveling' through controlling backup management system.

SGI NAS provides a fully NDMP-compliant implementation. The supported NDMP versions include: NDMP v2, v3, and v4.

You can use any NDMP management client software\* to make backups from SGI NAS systems to tape.

The list of such backup management software includes\*:

- EMC/Legato Networker
- Symantec Veritas NetBackup
- Arkeia Network Backup Suite
- IBM Tivoli Storage Manager
- Bakbone NetVault

NDMP transfers all filesystem level information and meta-information, including attributes, owner, group owner Ids, and ZFS ACLs. NDMP correctly recognizes soft and hard links, which is why no duplicate information is copied. When initiating backup, NDMP server makes a temporary snapshot of a given ZFS volume.

---

\* Also sometimes referred to as Data Management Application (DMA)

\* Interoperability with 3rd party backup management software is not fully tested. The work is ongoing to test and certify SGI NAS with all major NDMP based backup solutions.

## 16.1 Setting up NDMP server

The only necessary thing to setup NDMP server is to define user and password. It is recommended that user name is at least 4 characters long. Password must contain at least 8 characters, including uppercase symbols and numbers. Beware: giving user/password information to third parties provides unrestricted access to all files on all appliance's volumes.



SGI NAS doesn't provide NDMP management client software. SGI NAS only provides the support of the NDMP protocol and enables NDMP compliant software to copy SGI NAS datasets to tape drives or virtual tape devices.

In order to use NDMP protocol and backup your SGI NAS datasets, you need to establish a third-party NDMP server.

The list of available NDMP compliant applications is available at:

<http://www.ndmp.org/products/#backup>

If you have enough RAM (3GB or more), it would be practical to move folder for temporary files to RAM disk, which mounted as /tmp in SGI NAS. Usually NDMP server uses no more than 130 MB for temporary files.

If you have several network cards and want NDMP server to listen just one of them, you should enter IP address for 'mover-nic' option. Otherwise leave it empty.

To start NDMP server configuration, run:

```
nmc:/$ setup network service ndmp-server configure
```

For example:

```
nmc:/$ setup network service ndmp-server configure
Enter username           : tmpuser
Enter password           : tMppassword777
Enter path for temporary files and logs : /var/ndmp
IP address for connections : 192.168.1.177
```

To return default settings and prohibit access to ndmp-server, run:

```
nmc:/$ setup network service ndmp-server unconfigure
```

## 16.2 Three-way copying with ndmcopy

SGI NAS includes an NDMP management utility that can be used to perform NDMP transfers between two NDMP capable (and NDMP compliant) servers. The utility is called 'ndmcopy'.

At runtime 'ndmcopy' establishes connections to NDMP server processes on the source and destination machines, performs authentication to the servers, and initiates NDMP backup on the source machine and NDMP restore on the destination machine.

Prior to starting NDMP transfer with 'ndmcopy', prepare the following:

- Source and destination hosts. This can be appliance's IP addresses or hostnames. For example: 'hostone', 'hosttwo'.
- Source and destination folder names. In the example below it is '/opt/a' and '/opt/b', respectively.
- Username and password information for both servers. In our example (below): 'tmpuser', and 'tMppassword777' are the username/password pair defined on both source and destination appliances.

To start copying, run:

```
nmc:/$ ndmcopy hostone:/opt/a hosttwo:/opt/b
      -sa tmpuser:tMppassword777
      -da tmpuser:tMppassword777
```

The following shows an actual ndmcopy execution:

```
nmc:/$ ndmcopy -v -v hostone:/opt/a hosttwo:/opt/b
      -sa tmpuser:tMppassword777
      -da tmpuser:tMppassword777
```

Creating NDMP connection.

```
Authorizing connection with TEXT authorization.
      user = tmpuser, password = tMppassword777.
Authenticated root with NDMP_AUTH TEXT.
```

```
Creating NDMP connection.
Authorizing connection with TEXT authorization.
    user = tmpuser, password = tMppassword777.
Authenticated root with NDMP_AUTH_TEXT.

RESTORE: Sending recover request.
RESTORE: Done! returning.
NDMP_NOTIFY_DATA_READ
RESTORE: after, check_notification
DUMP: Done! returning.
NDMP_NOTIFY_DATA_HALTED
```

## 16.3 NDMP log – NDMP device – NDPM session

NDMP server uses two log files:

- 'system-ndmpd:default.log' – NDMP service lifecycle and service level logging
- 'ndmplog.0' – NDMP traces and details

To check the log files, use:

```
nmc:/$ show network service ndmp-server log
```

To list active NDMP sessions:

```
nmc:/$ show network service ndmp-server sessions
```

Following is a sample output of this command:

```
nmc:/$ show network service ndmp-server sessions
Session Id:          18
Protocol version:    4
Authenticated:       Yes
EOF:                 Yes
Client address:      192.168.1.177:34307
scsi.open:           No
scsi.adapter:        ""
scsi.valid target:   No
```

```
tape.device:          Not open
mover.state:         Idle
mover.mode:          Read
mover.pause reason:  N/A
mover.halt reason:   N/A
mover.record size:   64512
mover.record number: 0
mover.pos:           0
mover.win len:       -1
mover.win off:       0
data.operation:      Backup
data.state:          Active
data.halt reason:    N/A
data.data type:      Local
data.aborted:        No
data.read offset:    0
data.read length:    0
data.total size:     79673770
data.bytes processed: 81027072
data.env[0]:         UPDATE: "Y"
data.env[1]:         HIST: "Y"
data.env[2]:         FILESYSTEM: "/opt/a/"
```

To list local NDMP devices:

```
nmc:/$ show network service ndmp-server devices
```

Use this command (above) for tape libraries or any other NDMP-compliant external devices attached directly to SGI NAS systems.

## 16.4 Performance and considerations

Use the following command to monitor NDMP 'live' performance:

```
nmc:/$ show network service ndmp-server performance
```

For instance:

```
nmc:/$ show network service ndmp-server performance
[4:35:48, Sat May 23, 2009]
Session: 17 (Idle) Done: 66.7% of 79673770 bytes (53121024) ,
Speed: 16.9 MBytes/s
```

NDMP performance depends heavily on the average file size. Table below summarizes NDMP throughputs for different file sizes (left column) measured on a Sun Storage 7110 platform:

8KB	4.93 MB/sec
64KB	22.3 MB/sec
128KB	33.8 MB/sec
1MB	55.7 MB/sec
1GB	58.4 MB/sec

NDMP server loads directory structure and file information into appliance's operational memory during the backup. Typically, approximately 100 MB of memory is needed for every million of files to backup. If you have a huge amount of files to backup, try to subdivide backup into smaller portions, for example 10 million files per NDMP session.

NDMP server creates temporary files on disk during backup. It is possible to get better performance, if you move the folder for temporary files to RAM disk, for more information see [16.1 Setting up NDMP server](#).

### 16.5 Known issues

NDMP server requires at least 0,3-1 GB of free RAM for service to operate. Otherwise, it halts the session and does not write any information to log file.



## 17 Operations and Fault Management

### 17.1 Runners

To see existing runners in NMC use the command:

```
nmc:/$ show appliance runners
```

RUNNER	STATUS	STATE	SCHEDULED
memory-check	enabled	ready	every 12 minutes
runners-check	enabled	ready	every 12 hours
cpu-utilization-check	enabled	ready	every 15 minutes
nms-check	enabled	ready	not schedulable
services-check	enabled	ready	every 15 minutes
volume-check	enabled	ready	every 10 minutes
hosts-check	enabled	running	hourly
nms-fmcheck	enabled	ready	not schedulable
network-collector	enabled	ready	hourly
nfs-collector	enabled	ready	hourly
volume-collector	enabled	ready	hourly
volume-reporter	enabled	ready	weekly on Sat 04:00am
services-reporter	enabled	ready	daily at 05:00am
nfs-reporter	enabled	ready	weekly on Sat 03:00am
network-reporter	enabled	ready	weekly on Sat 02:00am
indexer:vtest/a	enabled	ready	daily at 01:00am
indexer:vtest/b	enabled	ready	daily at 01:00am

This shows:

several fault triggers (all with extension “check”), followed by statistic collectors, followed by storage and network service reporters, followed by two specific indexers with their associated folders.

In NMV, you can view runners by selecting **Runners** under the **Data Management** tab:

Runner	Type	Status	State	Scheduled
autosync-check	fault trigger	enabled	ready	every 10 minutes
comstar-trigger	fault trigger	enabled	ready	not schedulable
config-sd-reporter	reporter	enabled	ready	weekly on Sun 03:01am
cpu-utilization-check	fault trigger	enabled	ready	every 15 minutes
disk-check	fault trigger	disabled	ready	hourly
hosts-check	fault trigger	enabled	running	hourly
memory-check	fault trigger	enabled	ready	every 12 minutes
network-collector	statistics collector	enabled	ready	hourly
network-reporter	reporter	enabled	ready	weekly on Sat 02:02am
nfs-collector	statistics collector	enabled	ready	hourly
nfs-reporter	reporter	enabled	ready	weekly on Sat 03:00am
nms-autosmartcheck	fault trigger	enabled	ready	every minute
nms-check	fault trigger	enabled	ready	not schedulable
nms-fmcheck	fault trigger	enabled	ready	not schedulable
nms-rebootcheck	fault trigger	enabled	ready	not schedulable
nms-zfscheck	fault trigger	enabled	ready	every 12 minutes
runners-check	fault trigger	enabled	ready	every 12 hours
services-check	fault trigger	enabled	ready	every 15 minutes
services-reporter	reporter	enabled	ready	daily at 05:01am
ses-check	fault trigger	enabled	running	hourly
smart-collector	statistics collector	enabled	ready	hourly
volume-check	fault trigger	enabled	running	hourly
volume-collector	statistics collector	enabled	ready	hourly
volume-reporter	reporter	enabled	ready	weekly on Sat 04:03am

The appliance's framework allows you to add runners. SGI NAS runners have the advantage of exercising the entire NMS-provided SA-API (Section [2.1. Terminology](#); see also [21. References](#)), execute periodically, and/or on event, and/or run constantly in the background.

SGI NAS runners rely on the mailing facility which can be configured in NMC using the command:

```
nmc:/$ setup appliance mailer
```

All appliance's runners are runtime-configurable. Runners' times-to-run and other properties can be changed via:

```
nmc:/$ setup trigger
nmc:/$ setup collector
nmc:/$ setup reporter
```

```
nmc:/$ setup indexer
```

Each of the **setup** commands listed above has its **show** 'counterpart', to show the existing configuration and runtime status:

```
nmc:/$ show trigger
nmc:/$ show collector
nmc:/$ show reporter
nmc:/$ show indexer
```

For instance:

```
nmc:/$ setup trigger cpu-utilization
```

This can be used to disable, enable, run, and configure standard fault trigger that monitors CPU utilization. For instance, press TAB-TAB or Enter, type or select 'property', and view all 'cpu-utilization' properties available for tuning. You could change the alarm-generating thresholds (in this case - low and critically low idle CPU), make it run more or less frequent, etc.

```
nmc:/$ show trigger cpu-utilization -v
```

This will show the trigger's current runtime state, status and existing configuration in detail (notice the verbose -v option).

## 17.2 Triggers

Part of the SGI NAS Fault Management facility is realized through Fault Triggers. A fault trigger, or simply, a trigger, is a special kind of a pluggable runner module ('help runners') that performs a certain fault management and monitoring operation(s). Each trigger monitors one, or a few related conditions.

If any of the monitored conditions are violated, a fault trigger raises an alarm, which manifests itself in several ways:

- email notification to the administrator, with detailed description of the fault, including: severity, time, scope, suggested troubleshooting action, and often an excerpt of a related log with details.

- red color showing up via one of the NMC '**show**' operations detailed below.

show trigger all-faults

show trigger <name>

show appliance runners

show faults all-appliances

- and message posted to appliance's Inbox (see Section "Inbox").



Notifications of hardware faults are *immediate*. Unlike many other potentially faulty conditions that are getting periodically 'polled', any hardware fault itself triggers the appliance's fault management logic, that in turn includes email notification.

To see all available fault triggers in NMC, use the command `show trigger all`.

In all cases a trigger that 'carries' the alarm will be shown in **red**, assuming NMC colors are enabled. In addition, the faulted trigger will try to notify system administrator via appliance's mailing facility. Therefore, as already noted elsewhere, it is important to setup the appliance's mailer.

Trigger counts the fault conditions every time it runs. Typically, the fault trigger will send email once the faulty condition is observed a certain configurable number of times. Typically, after that the trigger itself goes into 'maintenance' state - it will still run and count the faulty conditions but it will not send email notification anymore - that is, until system administrator clears it from its maintenance state:

```
nmc:/$ setup trigger <name> clear-faults
```

Similar to the rest appliance's runners, triggers are flexible, in terms of their runtime behavior and trigger-specific conditions they monitor. For details on any specific fault trigger, run:

```
nmc:/$ show trigger <name> -v
```

where <name> stands for the trigger's name, and -v (verbose) is used to display details



The appliance includes one special fault trigger – 'nms-check'. This trigger performs fault management/monitoring function for the Fault Management facility itself. Nms-check tracks NMS connectivity failures and internal errors.

```
nmc:/$ show trigger nms-check -v
```

In presence of network failures, this will show all alarms (in detail) that the appliance failed to report.



```
nmc:/$ show faults all-appliances
```

This generates Fault Management summary report that includes all known (explicitly ssh-bound and dynamically discovered) SGI NAS systems.

Upon generating the summary, use a combination of NMC '**switch**' operation ()(Section "Command Reference - switch") and '**show faults**' - to 'zoom-in' into a particular ('faulted') appliance for details.

## 17.3 Handling an Unrecoverable I/O Error

Here is an example of an unrecoverable I/O error and possible resolutions:

```
FMA EVENT: ===== START =====
FMA EVENT: SOURCE: zfs-diagnosis
FMA EVENT: PROBLEM-IN: zfs://pool=iscsidisk/vdev=56f1d7932d4c039d FMA EVENT:
AFFECTS: zfs://pool=iscsidisk/vdev=56f1d7932d4c039d
FMA EVENT: ===== END =====
fault trigger 'nms-fmacheck (E1)' reached the configured maximum of 1 failure

FAULT: *****
FAULT: Appliance      : ups-sginas1 (OS v0.99.5b82, NMS v0.99.5)
FAULT: Machine SIG   : 1CG5KI
FAULT: Primary MAC    : 0:15:17:a:d1:fc
FAULT: Time           : Tue Mar 11 14:10:47 2008
FAULT: Trigger        : nms-fmacheck
FAULT: Fault Type     : ALARM
FAULT: Fault ID       : 1
FAULT: Fault Count    : 1
FAULT: Severity       : CRITICAL
FAULT: Description    : FMA Module: zfs-diagnosis, UUID:
FAULT:                 : 5bbb38fb-f518-4aa2-9018-8c2fe7e70360
FAULT: *****

Fault class : fault.fs.zfs.vdev.io
Description : The number of I/O errors associated with a ZFS device exceeded
acceptable levels. Refer to http://sun.com/msg/ZFS-8000-FD for more information.
```

The following is the type of email you might receive if you have FMA checks enabled and your appliance mailier is properly configured.

An unrecoverable I/O error scenario presents only two options:

- (1) Manually recover the faulted device. As specified in the fault report, it makes sense to review the posted URL (<http://illumos.org/msg/ZFS-8000-FD>) for the latest tips and guidelines.

In the case of FC/iSCSI/USB attached drives, please verify connectivity to the corresponding target(s).

Next, ssh into the appliance as root. At this point NMC will automatically determine the presence of a faulted condition and will prompt you to execute corrective action (you will simply need to press Enter).

- (2) The second option is simple: power cycle the appliance. This may cause an unrecoverable loss of data: the in-flight data that was not committed to stable storage at the time of the hardware failure will be lost. However, the existing data on the affected volume will **not** be corrupted. After power cycling, the entire faulted volume (that is, the volume that contains the faulted drive) will be marked 'offline' and inaccessible.

## **17.4 Handling a System Failure**

If the SGI NAS system fails and is restarted, then you will get an email after the restart , which includes information about date and time of the reboot. By default, last 20 lines from the system log file are included in the e-mail. To change this properties, run the following NMC command:

```
nmc:/$ setup trigger nms-rebootcheck
```

## 18 Analytics

### 18.1 DTrace

DTrace is a comprehensive dynamic tracing framework created by Sun Microsystems to analyze performance and troubleshoot problems on production systems in real time. For in-depth guide on DTrace language and details, please visit DTrace at [OpenSolaris.org](http://OpenSolaris.org).

DTrace can be used to generate performance profiles and analyze bottlenecks. DTrace can help to troubleshoot problems by providing detailed views of the system internals.

#### 18.1.1 DTrace command line

As with many other functional components, SGI NAS provides an **easy** DTrace integration. DTrace is integrated into the SGI NAS Management Console as one of its top level commands (Section 'Top-Level Commands').

To start using DTrace, type 'dtrace' at NMC prompt and use TAB-TAB to navigate, or simply press [Enter](#) and make a selection. DTrace is functionally sub-divided into sections, as follows:

```
nmc:/$ dtrace
Option ?
<?> IO  cpu  locks  memory  misc  network  report-all
-----
Choose one of the options above, 'q' or Ctrl-C to quit
Memory utilization, physical and virtual memory statistics
```

In most cases examples are provided; to see an example, select 'example' option. For instance:

```
nmc:/$ dtrace cpu cpuwalk example
```

To override the default behavior of any given dtrace utility, specify extra options in the command line, for instance:

```
nmc:/$ dtrace cpu cpwalk 5
```

This will run for 5 seconds (as opposed to running until Ctrl-C is pressed by default). Use TAB-TAB to navigate and make a selection. For details on particular command line options use help (-h), for instance:

```
nmc:/$ dtrace cpu cpwalk -h
```

## 18.2 NMV Analytics

To setup and view a performance chart in NMV, select the **Analytics** tab and the **Profiles** subtab. After selecting some metrics and creating the chart, the chart will appear at the bottom of the page. You can choose any number of metrics, but make sure that the scale for each metric is compatible or some series may be difficult to see. If you choose metrics from multiple statistics (top level tree nodes) you will get multiple charts.





You can close a chart by clicking the “x” icon in the upper right, or by clicking the “remove” button for the appropriate entry in the chart list at the top of the screen.

The header panel as well as any chart can be “shuttered” closed by using the triangle-shaped toggle button next to the x button.

To reorder charts, click in the blue heading (title) area and drag the chart where you wish to display it. The chart list will update to show the new order.

If the series lines of a chart are difficult to see in the default line chart presentation, you can click the bar chart icon in the upper left of the toolbar to change the view.

### 18.3 I/O Performance

You can check the real-time I/O performance of your data volume using NMC or NMV. In NMC,

the command:

```
nmc:/$ show volume <volumename> iostat
```

will show capacity, number of reads and writes, and read and write bandwidth.

## 18.4 Performance Benchmarks

The appliance includes a number of facilities to monitor and test its performance. This includes DTrace (Section [17.1 DTrace](#)). This includes NMV performance charts (Section [17.2.NMV Analytics](#)). This includes performance benchmarks intended to run stressful I/O and networking operations and show the corresponding statistics.

Performance benchmark functionality is included in a form of extensions – pluggable modules (plug-ins). These particular plug-ins are available to all users and can be installed into both Developer Edition and Commercial Editions.

SGI NAS includes currently two pluggable (micro-) benchmarks described in the subsequent sections: I/O benchmark and network performance benchmark.

To show all currently installed benchmarks:

```
nmc:/$ run benchmark
(and press TAB-TAB or Enter)
```

To list benchmarks (and other plug-ins) available in the remote central software repository:

```
nmc:/$ show plug-in remotely-available
```

### 18.4.1 I/O performance benchmark

```
Usage: [-p numprocs] [-b blocksize] [-q] [-s]
  -p <numprocs>    Number of process to run. Default is 2.
  -b <blocksize>   Block size to use. Default is 32k
  -s                No write buffering. fsync() after every write

This benchmark is using well known Bonnie++ tool, it is based on the Bonnie
benchmark written originally by Tim Bray.

Sequential Write (SEQ-WRITE):
  1. Block. The file is created using write(2). The CPU overhead
```

should be just the OS file space allocation.

2. Rewrite. Each <blocksize> of the file is read with read(2), dirtied, and rewritten with write(2), requiring an lseek(2).

**Sequential Read (SEQ-READ):**

Block. The file is read using read(2). This should be a very pure test of sequential input performance.

**Random Seeks (RND-SEEKS):**

This test runs SeekProcCount processes (default 3) in parallel, doing a total of 8000 lseek()s to locations in the file specified by random(). In each case, the block is read with read(2).

In 10% of cases, it is dirtied and written back with write(2).

**Example:**

```
nmc@thost:/$ run volume voll benchmark -p 2 -b 8192
Testing 'vol2'. Optimal mode. Using 1022MB files and 8192 blocks.
```

SEQ-WRITE	CPU	S-REWRITE	CPU	SEQ-READ	CPU	RND-SEEKS
162MB/s	8%	150MB/s	6%	188MB/s	9%	430/sec
158MB/s	7%	148MB/s	8%	184MB/s	7%	440/sec
-----	-----	-----	-----	-----	-----	-----
160MB/s	8%	149MB/s	7%	186MB/s	8%	435/sec

### 18.4.2 Network performance benchmark

Quoting Wikipedia lperf article (<http://en.wikipedia.org/wiki/lperf>):

'lperf' is a commonly used network testing tool that can create TCP and UDP data streams and measure the throughput of a network that is carrying them. lperf is a modern tool for network performance measurement written in C++.

lperf allows user to set various parameters that can be used for testing a network, or alternately for optimizing or tuning a network. lperf has a client and server functionality, and can measure the throughput between the two ends, either unidirectionally or bi-directionally. It is open source software and runs on various platforms including linux, unix and windows. It is supported by the National Laboratory for Applied Network Research.

Usage: [-s]

```
[-P numthreads] [-i interval] [-l length] [-w window] [-t time][hostname]
-s                run in server mode
-P numthreads    number of parallel client threads to run    (default = 3)
-i interval      seconds between periodic bandwidth reports (default = 3)
-l length        length of buffer to read or write           (default = 128KB)
-w window        TCP window size (socket buffer size)       (default = 256KB)
-t time          total time in seconds to run the benchmark (default = 30)
hostname         for the iperf client, you can optionally specify
                  hostname or IP address of the iperf server

Usage: [-s] [server-options]
       [-c] [client-options]

server-options   any number of valid iperf server command line option,
                  as per iperf documentation
client-options   any number of valid iperf client command line option,
                  as per iperf documentation
```

This plug-in is based on a popular Iperf tool used to measure network performance. The benchmark is easy to set up. It requires two hosts, one - to run iperf in server mode, another - to connect to the iperf server and run as a client. Use -s option to specify server mode.

The easiest way to run this benchmark is to select a host for the server and type 'run benchmark iperf-benchmark -s'. Next, go to the host that will run iperf client and type 'run benchmark iperf-benchmark'. You will be prompted to specify the server's hostname or IP address. See more examples below.

To run this benchmark, you can either:

- a) use built-in defaults for the most basic parameters, or
- b) specify the most basic benchmark parameters, or
- c) specify any/all iperf command line option, as per iperf manual page.

To display iperf manual page, run '!iperf -h'

### Examples:

#### Example 1.

Let's say, there are two appliances: hostA and hostB. On appliance hostA run:

```
nmc@hostA:/$ run benchmark iperf-benchmark -s
```

This will execute iperf in a server mode. On appliance hostB the iperf client connects to hostA and drives the traffic using default parameter settings:

```
nmc@hostB:/$ run benchmark iperf-benchmark hostA
```

#### Example 2.

Same as above, except that now we assign parameters such as:

```
* number of parallel client threads = 5
```

```
* seconds between periodic bandwidth reports = 10
* length of buffer to read or write = 8KB
* TCP window size = 64KB
```

```
nmc@hostA:/$ run benchmark iperf-benchmark -s
```

```
nmc@hostB:/$ run benchmark iperf-benchmark hostA -P 5 -i 10 -l 8k -w 64k
```

Notice that all these parameters are specified on the client side only. There is no need to restart iperf server in order to change window size, interval between bandwidth reports, etc.

**Example 3.**

Same as Example #1, except that iperf server is not specified in the command line. Instead, NMC will prompt you to select the server interactively from a list of all ssh-bound appliances:

```
nmc@hostA:/$ run benchmark iperf-benchmark -s
```

```
nmc@hostB:/$ run benchmark iperf-benchmark
```

**Example 4.**

\*\*\*\*\* Note: advanced usage only \*\*\*\*\*

You can specify any number of valid iperf server and/or client command line option, as per iperf documentation. Unlike the most basic command line options listed above, the rest command line options are not validated and do not have NMC provided defaults. Unlike the most basic command line options listed above, the rest command line options are passed to iperf AS IS.



## 19 Managing the Users

SGI NAS automatically syncs user ids in the /etc/passwd and /var/smb/smbpasswd files. Keep this in mind if you change a user ID on the appliance. You won't be able to access existing files owned by this user until you also change the file ownership to the new user id.

To change a user's user ID:

```
nmc:/$ setup appliance user jack property uidNumber
User id (uid) : 1001
```

To change the owner of the folder use the NMC command:

```
nmc:/$ setup folder <foldername> ownership
```

### 19.1 Adding Local Appliance Users

In NMV, you can add local appliance users by selecting **Users** under the **Settings** tab.

The screenshot shows the SGI NAS web interface. The top navigation bar includes 'About', 'Support', 'Add Capacity', 'Register', and 'Help'. The main navigation tabs are 'Status', 'Settings', 'Data Management', and 'Analytics'. The 'Settings' tab is active, and the 'Users' sub-tab is selected. The left sidebar shows a tree view with 'Users' and 'Groups' expanded. The main content area is titled 'CREATE NEW APPLIANCE USER' and contains the following form fields and sections:

- User Name:** Input field with value 'you'. Note: User name. Maximum length is 16 characters.
- Password:** Input field with masked characters '\*\*\*\*\*'. Note: User password. Maximum length is 40 characters.
- Password Confirm:** Input field with masked characters '\*\*\*\*\*'. Note: Password confirmation box.
- Display Name:** Input field. Note: The name which will be displayed on the top of the application. Maximum length is 255 characters.
- E-Mail Address:** Input field. Note: Optional e-mail address for this user.
- About:** Text area. Note: Optional additional information for this user.
- UNIX User:** Check box checked. Note: If unchecked, local UNIX user will not be created. Note: UNIX and/or LDAP users are required for integrated NFS/CIFS ACLs to work properly.
- UNIX Group:** Input field with value 'staff'. Note: The default group that this username belongs to. Groups provide a mechanism to assign permissions to a list of users all at once. Default is "staff".
- Home Folder:** Input field. Note: Optional. User's home folder pathname in the form of volume/folder. Example: home/je. If specified, new ZFS folder will be created and assigned as a home folder for new UNIX user with Read/Write permissions.
- NMV GUI PERMISSIONS:** A section with various checkboxes:
  - Can view:**  Can view confidential data
  - Can\_admin\_groups:**  Can add/delete/edit appliances groups
  - Can\_admin\_services:**  Can add/delete/edit periodic data services
  - Can\_admin\_runners:**  Can add/delete/edit runners services
  - Can\_admin\_datasets:**  Can add/delete/edit/export/import datasets
  - Can\_admin\_shares:**  Can share/unshare/edit folders share resources
  - Can\_admin\_users:**  Can add/delete/edit users
  - Can\_admin\_disks:**  Can add/delete/edit iSCSI/iFC resources
  - Can\_admin\_network:**  Can add/delete/edit network interfaces and services
  - Can\_admin\_appliance:**  Can upgrade/reboot/edit appliance
  - Can\_admin\_prefs:**  Can edit management server and Web GUI preferences
  - Can\_search\_folders:**  Can search in folders
  - Can\_use\_restapi:**  Can execute commands through rest api

A 'Create New' button is located at the bottom of the form.

Corresponding NMC command:

```
nmc:/$ setup appliance user create
New User      : tom
Home folder   : voll/a/b
Description   :
Default group : other
Password      : xxxxxxxx
Confirm password : xxxxxxxx
```

When you create a user, you need to specify user name, group and password. Optionally, you can indicate a home folder for this user and add a description.

## 19.2 Local Appliance Groups

You can also create user groups within SGI NAS. This is done in NMV on the Users page under the **Settings** tab. You can create a new group by selecting [New Group](#) under the 'Groups' heading.

The screenshot displays the SGI NAS web interface. At the top, there is a navigation bar with 'About', 'Support', 'Add Capacity', 'Register', and 'Help'. Below this is a secondary navigation bar with 'Status', 'Settings', 'Data Management', and 'Analytics'. The 'Settings' tab is active. The main content area is titled 'CREATE NEW APPLIANCE USER GROUP'. On the left, there is a sidebar with 'Users' and 'Groups' sections. The 'Groups' section is expanded, showing 'Summary' and 'New Group' options. The main content area has three columns: 'Available Users' with a list of 'admin', 'alice', and 'root'; 'Group Members' with a list of 'jon'; and a 'Group Name' field containing 'usergroup1'. There are 'Add >>' and '<< Remove' buttons between the columns, and a 'Create New' button at the bottom. A footer at the bottom of the page contains links for 'Found a bug?', 'Feature request?', and 'Request Technical Support'.

Corresponding NMC command:



```
nmc:/$ setup appliance usergroup create
```

## 19.3 LDAP

The Lightweight Directory Access Protocol (LDAP) is a common protocol interface to Network Directory Services. Widely deployed directory services are Domain Name Service (DNS), NIS (Network Information Service), etc. They provide the clients with information such as host IP addresses, usernames, passwords and home directories. LDAP is a widely-deployed, simple and efficient network protocol for accessing information directories. LDAP typically runs over TCP; it has the potential to consolidate existing Network Directory Services into a single global directory.

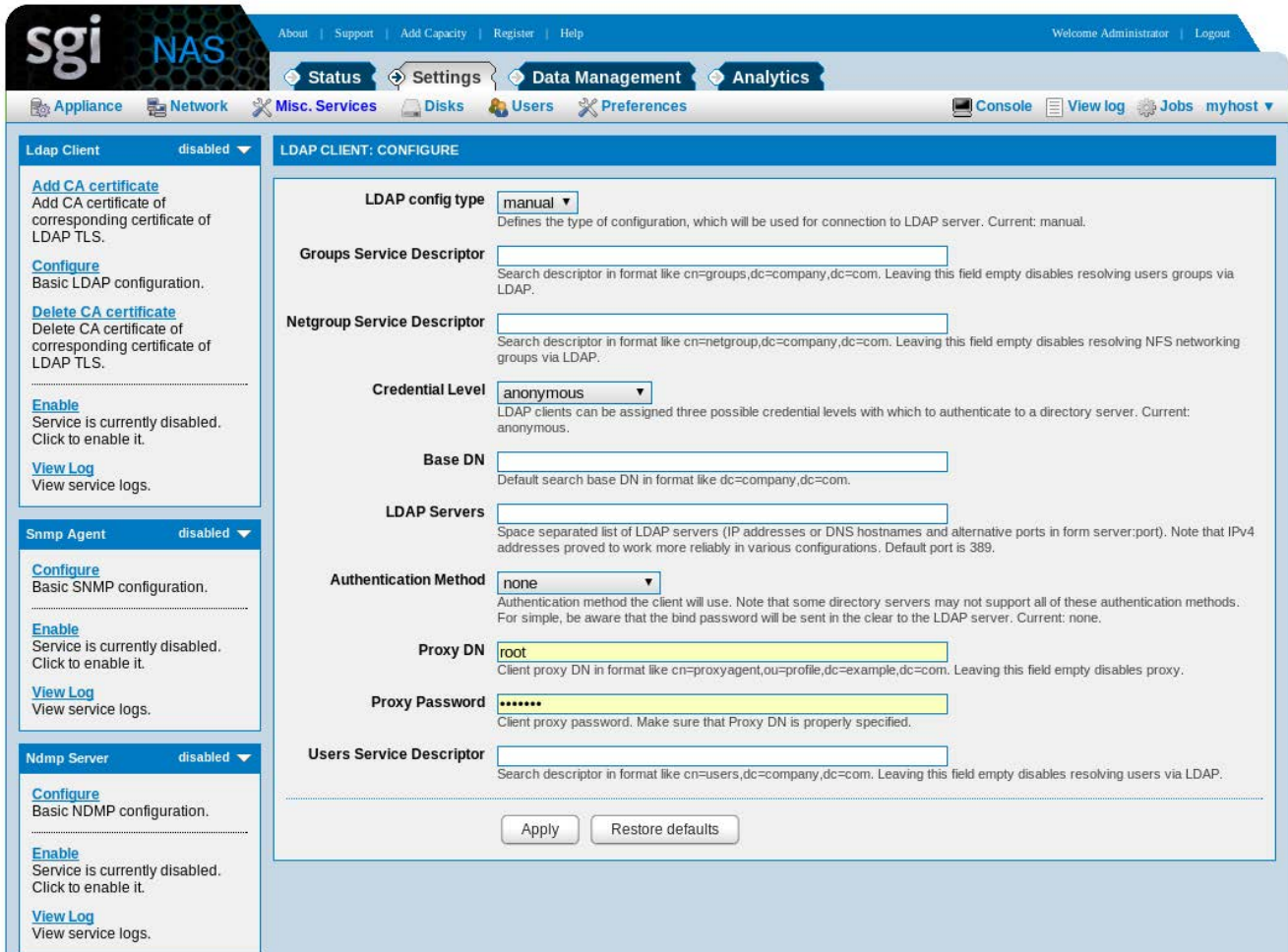
SGI NAS provides easy to use LDAP integration, specifically for usage in the NFS environments<sup>A</sup>. In addition, LDAP user and group management can be deployed with NFSv4 – the default NFS version provided by the appliance. In general, LDAP based user and group management is required to consistently utilize ZFS extended Access Control List (ACLs) across heterogeneous file services instead of POSIX permissions and attributes.

It is recommended that you use LDAP for centralized user management. SGI NAS is an LDAP client in this case. To use SGI NAS with LDAP server, make sure the server is available. You will need your base DN, with either anonymous or authenticated SASL bindings (the latter requiring account DN and password), and netgroup, user, and group subtree DNs if known. Netgroup (a group of hosts) is only necessary if currently supported by the LDAP server and is of interest.

---

A The work is underway to support CIFS workgroup mode with LDAP. As of the time of this writing, CIFS workgroup mode works with local Unix users and groups.

You define authentication information for communicating with an LDAP server within NMV on the **Settings** tab in **Misc. Services**.



Note that in addition to Unix based LDAP, SGI NAS provides Active Directory integration - an implementation of LDAP directory services by Microsoft for use primarily in Windows environments.

Finally, SGI NAS LDAP client provides integrated ability to authenticate itself using X.509 certificates. Management console and management UI both provide the corresponding interfaces.

## **19.4 ACLs**

SGI NAS provides native extended Access Control Lists (ACLs), capable of handling CIFS ACLs, as well as NFSv4 ACLs, as well as POSIX permissions natively in the same filesystem. The appliance supports full management of per-user, per-group, per-folder ACLs in its user interface, while also populating the system with accounts and groups that you may have already defined in Active Directory or other LDAP-based directory service.

SGI NAS User and Access Control management has the following characteristics:

- Support both local and LDAP (or AD) managed users and groups. In LDAP or Active Directory configurations, the local users and groups can be used to override centralized settings. After configuring LDAP client LDAP users automatically appear in **Settings** → **Users**.
- Native extended Access Control Lists (ACLs), that are both CIFS and NFSv4 compliant.

Following are two screenshots that show, first, appliance users (most of which are retrieved from LDAP in this case), and the management GUI capability to administer access control to a given folder (and its subfolders – all operations on ACLs are recursive, to reduce the amount of administration).

## SGI NAS User Guide

The screenshot shows the SGI NAS Users management interface. The main content area displays a table of users with the following data:

User	Group	Display Name	E-Mail Address	Created	Entities	Delete
root	root	Super-User			UNIX	-
admin	staff	Administrator	me@me.com	2012-06-05 03:45:10	NMV, UNIX	-
guest		Guest		2012-06-05 03:45:11	NMV	✗
user1		User1			LDAP	-
user2		User2			LDAP	-

At the bottom right of the table, it says "Results 1 - 5 (all)".

Notice, that in the case below a local 'test-user' and LDAP-defined 'rfgroup' are granted with special set of permissions:

The screenshot shows the SGI NAS folder management interface for 'mypool/folder1'. The 'Access Control List' section is expanded, showing the following permissions:

Entity	Allow	Deny	Delete
user:test-user	list_directory, read_data, execute		✗
group:rfgroup	list_directory, read_data, execute		✗
user:smb	list_directory, read_data, add_file, write_data, add_subdirectory, append_data, read_xattr, write_xattr, execute, delete_child, write_attributes, write_acl, write_owner, synchronize		✗

SGI NAS CLI management client provides the same capabilities command line. The users and groups can be retrieved (that is, 'shown'), created and deleted, extended permissions can be modified and all the rest related management operations can be executed using either NMV or/and NMC.

SGI NAS ACLs are native across ZFS, CIFS, and NFSv4, and as such have no conflict in

how they are operated on. Generally, one accomplishes ACL management via the following tasks:

- local user or LDAP configuration
- definition of per-user or per-group capabilities per volume or folder
- overall management of ACLs and ACEs system wide, allowing overriding of end user activity via CIFS/NFS



**A note on NFSv3 vs. ACL**

NFSv3 relies on POSIX permissions, which are a subset of ZFS extended ACLs. Thus, NFSv3 clients will only check with the POSIX level permissions.

However, even though POSIX permissions may otherwise grant a permission to a user, that will be nullified if the extended ACL on the server is defined and otherwise denies that access.

## 19.5 User Quotas

User quotas are defined in SGI NAS on a per-folder basis by setting the userquota folder property.

Here is an example in NMC of setting a quota for a user named “fred”:

```
nmc:/$ setup folder mypool/home property
Option ? userquota
User      : fred
userquota@fred : 2m
FOLD      PROPERTY      VALUE      SOURCE
mypool/home userquota@fred 2M        local
```

To view the current user quota for “fred” you can do this in NMC:

```
nmc:/$ show folder mypool/home property userquota@fred
FOLD      PROPERTY      VALUE      SOURCE
mypool/home userquota@fred 2M        local
```

## 19.6 Group Quotas

Group quotas are administered similarly to user quotas. The group should exist prior to administration.

```
nmc:/$ setup group mypool/home property
Option ? groupquota
Group          : staff
groupquota@staff : 100m
FOLD           PROPERTY      VALUE      SOURCE
mypool/home   groupquota@staff  100M      local
```

To view the current group quota for group 'staff' in NMC, run:

```
nmc:/$ show folder mypool/home property groupquota@staff
FOLD           PROPERTY      VALUE      SOURCE
mypool/home   groupquota@staff  100M      local
```

## 20 Managing NMS Properties

SGI NAS allows you to set up various NMS properties. To see the list of tunable NMS properties in NMV, go to **Setting** → **Preferences** → **Server**:

The screenshot displays the 'CUSTOMIZE NMS PROPERTIES' configuration page in the SGI NAS web interface. The page is organized into sections with various settings:

- Auto\_scrub\_takes\_lock:** No (dropdown). auto\_scrub service volume-locking policy: 0 (default) - do not lock the volume when scrubbing, 1 - do lock to prevent other auto-services from running when scrub is in progress.
- Auto\_sync\_takes\_lock:** Yes (dropdown). auto-sync service folder-locking policy: 0 (default) - do not lock the folder when syncing, 1 - do lock to prevent other auto-services from running when syncing is in progress.
- Auto\_tier\_takes\_lock:** No (dropdown). auto-tier service folder-locking policy: 0 (default) - do not lock the folder on local appliance when replicating, 1 - do lock to prevent other auto-services from running on the folder when tiering is in progress.
- Autoflush:** No (dropdown). Flush management server log upon every write. May slow down management server when trace\_level > 10.
- Automount\_timeout:** 600 (text input). Duration that a file system remains mounted when not in use (seconds). This applies to NFS and CIFS auto-mounts only.
- Autoreserve\_space:** Yes (dropdown). Automatically reserve space on every volume. NMS will create hidden folder '/nza-reserve' to reserve some space (from 10 to 100 MB) after volume is filled on 50%. Note, that 'volume-check' runner should be enabled. The reserve can be deleted if you get in troubles with volume without any space: 1 - enable (default), 0 - disable.
- Client\_trace\_exec\_output:** 0 (text input). Enable tracing of every system exec call on NMS client side. This option could produce very noisy output.
- Client\_trace\_level:** 1 (text input). Tracing levels: 0 - not tracing, 1 - enable tracing on NMS client side.
- Default\_ntp\_server:** pool.ntp.org (text input). Synchronize time on start using this NTP server.
- Disable\_plugins\_install:** No (dropdown). Disable plugins install.
- Disk\_write\_cache\_mode:** 0 (text input). Enable or disable disk write caches. Only certain SCSI devices supports cache settings. Modes: 0 - use current settings (default), 1 - try to globally enable disk write caches, 2 - try to globally disable disk write caches.
- Gid\_range\_max:** 60000 (text input). Maximum GID which can be used by the appliance.
- Gid\_range\_min:** 1000 (text input). Minimum GID which can be used by the appliance.
- Import\_caches\_timeout:** 3600 (text input). Volume import caches timeout. Default is 3600 seconds.
- Index\_lang:** english (text input). Set the stemming language for the snapshot indexer, the default is 'english'. Possible values: danish dutch english french german german2 hungarian italian korean german2 norwegian portuguese romanian russian spanish swedish turkish (pass 'none' to disable stemming).
- Indexroot:** /syspool/index (text input). Indexer's database root folder. Location where NMS and indexer will store its content metadata.
- Internal\_broadcast\_discovery:** No (dropdown). Enables internal broadcast discovery of appliances. Usually the work is done by HA agent; hosts-check. Recommended to keep this option disabled.
- Lunsync\_on\_start:** Yes (dropdown). Synchronizes LUNs at NMS startup. If specified, NMS executes lunsync command at startup, prior to synchronizing volumes. The option comes handy with certain types of SCSI and SATA devices that are not discovered properly by the kernel at boot time. Note also that lunsync can be executed from NMS client at any time.
- Object\_cache\_timeout:** 1200 (text input). NMS internal object cache timeout. The server will invalidate its internal per-object caches when the specified timeout (in seconds) expires (default is 30 secs).
- Rep\_rsync\_options:** -delete --exclude-from=/var/lib/nza/rep\_exc.txt (text input). Default rsync options used for mirroring. The system default is: -delete --exclude-from=/var/lib/nza/rep\_exc.txt --inplace --ignore-errors -HtpatpO.
- Rep\_rsync\_version:** 3 (text input). Default rsync version to be used for mirroring (2 or 3). The system default is 3. Note that some versions of rsync clients will not work with version 3, and it is recommended to upgrade client instead if possible.
- Rep\_ssh\_options:** (text input). Default SSH extra options used for mirroring/syncing. The system default is: -c aes128-cbc.
- Sampledoc\_maxsize:** 2097152 (text input). Maximum indexed document size used to generate samples while searching.
- Saved\_configroot:** /config (text input). Location to save to, and restore from, appliance configuration.
- Service\_log\_keep\_days:** 60 (text input). For how long to keep entries in the logdata per logging object.
- Srvpool\_affinity\_timeout:** 60 (text input). Time interval to retain client → NMS-from-the-pool relationship (connection affinity).
- Srvpool\_cnt\_initial:** 2 (text input). Initial size of the pool of Management Servers.
- Srvpool\_cnt\_max:** 10 (text input). Maximum size of the pool of Management Servers.
- Srvpool\_interface\_serialization\_timeout:** 60 (text input). Time interval to serialize API calls through a given interface object (provided by a given management server) after any write/delete/create operation on that interface (sticky write timeout).
- Srvpool\_port\_range\_min:** 2011 (text input). TCP port for the first Management Server in the pool to listen on.
- Timesync\_on\_start:** Yes (dropdown). Synchronize time using command 'ntpdate -bu pool.ntp.org' at NMS startup.
- Trace\_exec\_output:** 0 (text input). Enable tracing of every system exec call. This option could produce very noisy output.
- Trace\_level:** 1 (text input). Tracing levels: 0 - no tracing, 1 - default, 10 - corresponds to '-v' cmdline option (default for beta releases), 11 - TRACE entry level, 20 - corresponds to '-vv' cmdline option, 21 - backtrace when Exception is produced, 30 - corresponds to '-vvv' cmdline option.
- Uid\_admin:** 1000 (text input). UNIX default UID for admin user.
- Uid\_nfs:** 60001 (text input). UNIX default UID for nfs anonymous user.
- Uid\_range\_max:** 60000 (text input). Maximum UID which can be used by the appliance.
- Uid\_range\_min:** 1000 (text input). Minimum UID which can be used by the appliance.
- Uid\_smb:** 63001 (text input). UNIX default UID for smb anonymous user.
- Upgrade\_menu\_keep:** 8 (text input). How many system checkpoints to keep in the boot menu (default is 8).
- Upgrade\_proxy:** (text input). Proxy server through which upgrades will be downloaded in form of URL similar to: http://user:password@host:port.
- Upgrade\_roots\_keep:** 8 (text input). How many system checkpoints to keep on the system volume (default is 8).

At the bottom of the page, there is a 'Save' button and a footer with links: 'Found a bug?', 'Feature request?', and 'Request Technical Support'.

Corresponding NMC command:

```
nmc:/$ setup appliance nms property
Option ?
show <?> auto_scrub_takes_lock auto_sync_takes_lock auto_tier_takes_lock
autoflush automount_timeout autoreserve_space client_trace_exec_output
client_trace_level default_ntp_server delorean_event_level
delorean_fault_level disable_plugins_install disk_write_cache_mode
gid_range_max gid_range_min import_caches_timeout index_lang indexroot
internal_broadcast_discovery lunsync_on_start net_tcp_naglim_def
net_tcp_rcv_hiwat net_tcp_xmit_hiwat object_cache_timeout
rep_rsync_options rep_rsync_version rep_ssh_options sampledoc_maxsize
saved_configroot service_log_keep_days srvpool_affinity_timeout
srvpool_cnt_initial srvpool_cnt_max
srvpool_interface_serialization_timeout srvpool_port_range_min
sys_log_debug sys_snoop_interval sys_snooping sys_zfs_arc_max
sys_zfs_nocacheflush sys_zfs_resilver_delay sys_zfs_resilver_min_time
sys_zfs_vdev_max_pending timesync_on_start trace_exec_output trace_level
uid_admin uid_nfs uid_range_max uid_range_min uid_smb upgrade_menu_keep
upgrade_proxy upgrade_rootfs_keep volroot
```

## 20.1 Multi-nms

Multi-nms was introduced with SGI NAS. It lets you create few NMS instances which facilitates work for multiple NMS clients. Depending on how much processing is involved with a single request, how many requests a server will have to handle, single or multi NMS mode can be used.

Multi-NMS is enabled by default and starts two additional NMS instances. To see the current state of NMS, run:

```
nmc:/$ show appliance nms pool
```

ID	STATE	PORT	PID	cnt	min (ms)	max (ms)	%idle
0	active	2001	27691	84186	0.08	160587079.63	79
1	running	2011	27821	2583	0.56	31491.56	99
2	idle	2012	28002	2671	0.87	43409.44	99



Term	Definition
ID	<p>Identification number of NMS instance. NMS 0 – main SGI NAS Management Server which is always enabled. ID 1, 2 (and all the following) are additional management servers or pooled NMS.</p> <p>The number of additional NMS instances is defined by '<b>srvpool_cnt_initial</b>' property, that can be setup by running:</p> <pre>nmc:/S setup appliance nms property srvpool_cnt_initial</pre>
State	State of the NMS instance.
Port	Port on which DBUS of the corresponding NMS works.
PID	Process identification number
cnt	Number of the API requests
min, max	Minimum/maximum time of the operation execution (ms)
% idle	percent of time, when NMS instance is in 'idle' state.

One of the important property of multi-nms is '**srvpool\_cnt\_max**'. It determines the maximum number of additional management servers. To specify this property, run:

```
nmc:/S setup appliance nms property srvpool_cnt_max
```

To enable multi-nms, run:

```
nmc:/$ show appliance nms pool enable
```

To disable multi-nms, run:

```
nmc:/$ show appliance nms pool disable
```

To disable pooled NMS and then start them again, run:

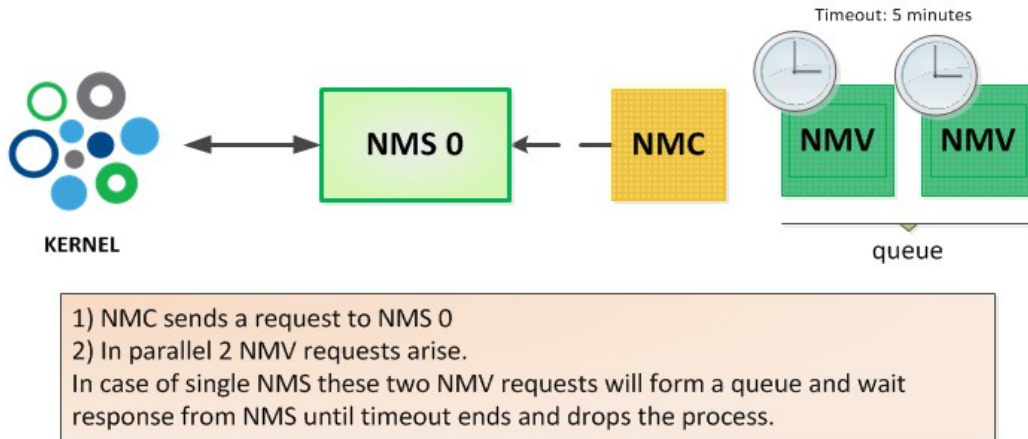
```
nmc:/$ setup appliance nms pool restart
```

Restart synchronizes all NMS instances.

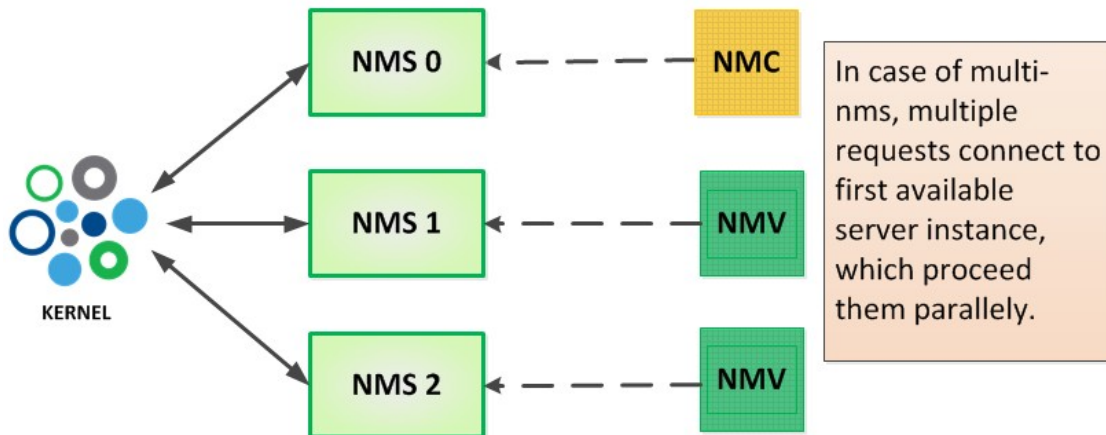
## 20.2 Concept Diagram

Observe the following diagrams to get the idea of single and multi-nms concepts:

### Example 1. Single NMS



### Example 2. Multi-NMS



## 20.3 Known issues

**Q.** After creating a new volume, I tried to view it with 'show volume' command, but it doesn't appear in the list of existing volume.

**A.** If multi-NMS is enabled, try to restart NMS to synchronize all NMS instances, by running the following NMC command:

```
nmc:/$ setup appliance nms restart
```

Same solution apply to the same error happen to any ZFS dataset.

**Q.** 1) After adding a new disk to the system, I tried to view it with 'show lun' command, but it doesn't appear in the list of existing LUNs.

**A.** If multi-NMS is enabled, try to restart NMS to synchronize all NMS instances, by running the following NMC command:

```
nmc:/$ setup appliance nms restart
```

In all cases, submit a support request, by running:

```
nmc:/$ support
```



Currently, multi-NMS mode doesn't work with HA Cluster and NS Cluster plugins. Multi-NMS mode automatically switches after cluster initialization.

This limitation is planned to be removed in future releases.



## 21 Managing the Network

### 21.1 Changing Network Interface Settings

For example to use jumbo frames you can adjust the Maximum Transmission Unit.

Settings can also be changed in NMC using the command:

```
nmc:/$ setup network interface
```

### 21.2 Link Aggregation

The appliance fully supports 802.3ad link aggregation, often referred to by its IEEE working group name of 'IEEE 802.3ad'. Other terms for link aggregation include 'Ethernet trunk', 'NIC teaming', 'port channel', 'port teaming', 'port trunking', 'link bundling', 'EtherChannel', 'Multi-Link Trunking (MLT)', 'NIC bonding', 'Network Fault Tolerance (NFT)'.

Link aggregation is used to combine multiple physical Ethernet links into one logical link to increase bandwidth and to protect against failures.

To create Link Aggregation in NMV go to **Settings** → **Network** → **Create**:

The screenshot displays the SGI NAS web interface. At the top, there's a navigation bar with 'About', 'Support', 'Add Capacity', 'Register', and 'Help'. Below that, a secondary navigation bar includes 'Status', 'Settings', 'Data Management', and 'Analytics'. The main interface is divided into a left sidebar and a main content area. The sidebar has a 'Network' section with a dropdown menu, and a 'Configured Network Interfaces' section showing 'e1000g0' configured with IP 172.16.3.100/255.255.0.0 and mtu 1500. The main content area is titled 'CREATE NETWORK INTERFACE' and contains the following configuration options:

- Interface Type:** Aggregated (dropdown menu). Description: Type of new interface: single or aggregated.
- All Available Devices:** A list box containing 'e1000g1' and 'e1000g2'. Description: Available capable interfaces. Select just one for single type or two or more for aggregation. To select/deselect multiple items use CTRL key.
- Configuration Method:** Static (dropdown menu). Description: Network interface configuration method: static or dynamic (via DHCP).
- IP Address:** 17.16.3.101 (text input). Description: Statically assigned IPv4 address in a dot-decimal notation (.#.#.#).
- Subnet Mask:** 255.255.0.0 (text input). Description: Statically assigned subnet mask in a dot-decimal notation (.#.#.#).

At the bottom of the configuration area is an 'Add Interface' button. The footer of the page contains links for 'Found a bug?', 'Feature request?', and 'Request Technical Support'.

Corresponding NMC command:

```
nmc:/$ create network aggregation
```

You can show existing link aggregates by running:

```
nmc:/$ show network aggregation
LINK          PORT          SPEED DUPLEX  STATE  ADDRESS          PORTSTATE
net1          --            1000Mb full  up     0:c:29:12:d2:93  --
              e1000g1       1000Mb full  up     0:c:29:12:d2:93  attached
              e1000g2       1000Mb full  up     0:c:29:12:d2:9d  attached
```

In this example, interface net1 is the aggregation of two physical network interfaces e1000g1 and e1000g2. The physical interfaces are then no longer visible for network administration and monitoring, unless you first destroy the aggregation using the NMC command:

```
nmc:/$ destroy network aggregation
```

Aggregation requires the switch to support Link Aggregation Control Protocol (LACP), which is a method to control the bundling of several physical ports together to form a single logical channel. LACP allows a network device to negotiate an automatic bundling of links by sending LACP packets to the peer (a directly connected device that also implements LACP).

Supported LACP modes:

Off mode	The default mode for SGI NAS aggregations. LACP packets are not generated
Active mode	The system generates LACP packets at regular intervals
Passive mode	The system generates LACP packets only when it receives an LACP packet from the switch. When both the aggregation and the switch are configured in passive mode, they cannot exchange LACP packets

## 21.3 VLAN

A virtual LAN, commonly known as a **VLAN**, is a group of hosts with a common set of requirements that communicate as if they were attached to the broadcast domain, regardless of their physical location. VLANs are created to provide the segmentation services traditionally provided by routers in LAN configurations. VLANs address issues such as scalability, security, and network management. The standard protocol used to configure virtual LANs is IEEE

## 802.1Q.

SGI NAS provides a fully compliant IEEE 802.1Q VLAN implementation.

To configure a virtual LAN from NMC, use the command `setup network interface`, as shown in this example:

```
nmc:/$ setup network interface e1000g1 vlan
VLAN Id :
-----
The ID associated with the VLAN. Press Ctrl-C to exit.
```

In this example a Virtual LAN is created with Ethernet frames carrying extra 4 bytes of VLAN header, as per 802.1Q specification. The VLAN header in turn will have the (12 bit) VLAN Id that was provided in the NMC dialog (above).



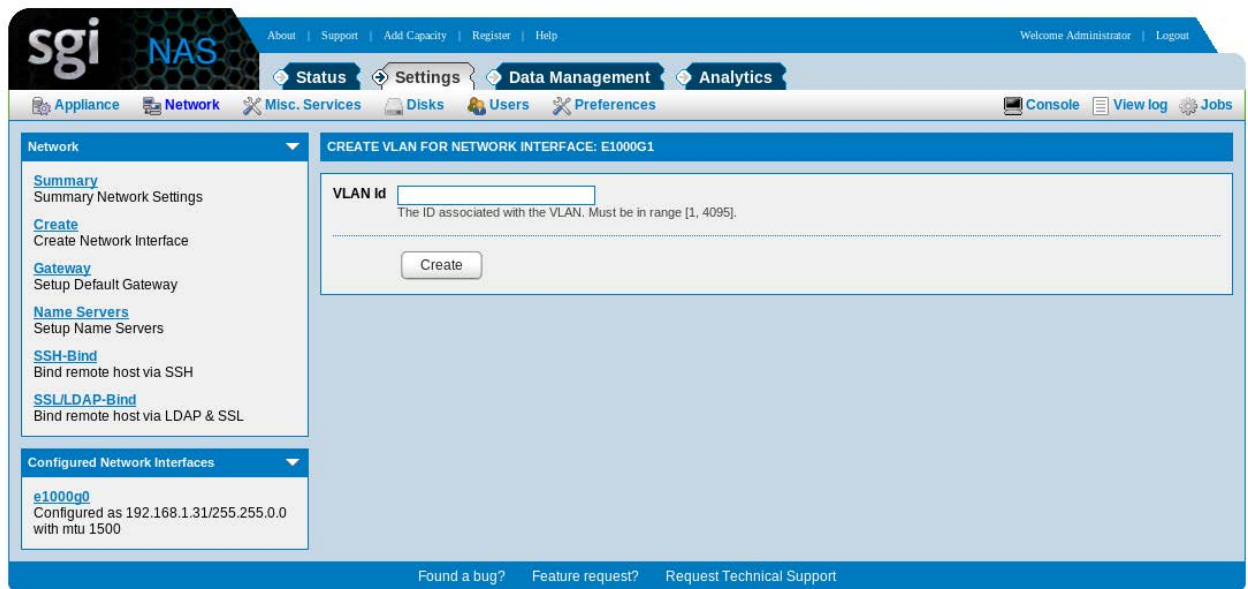
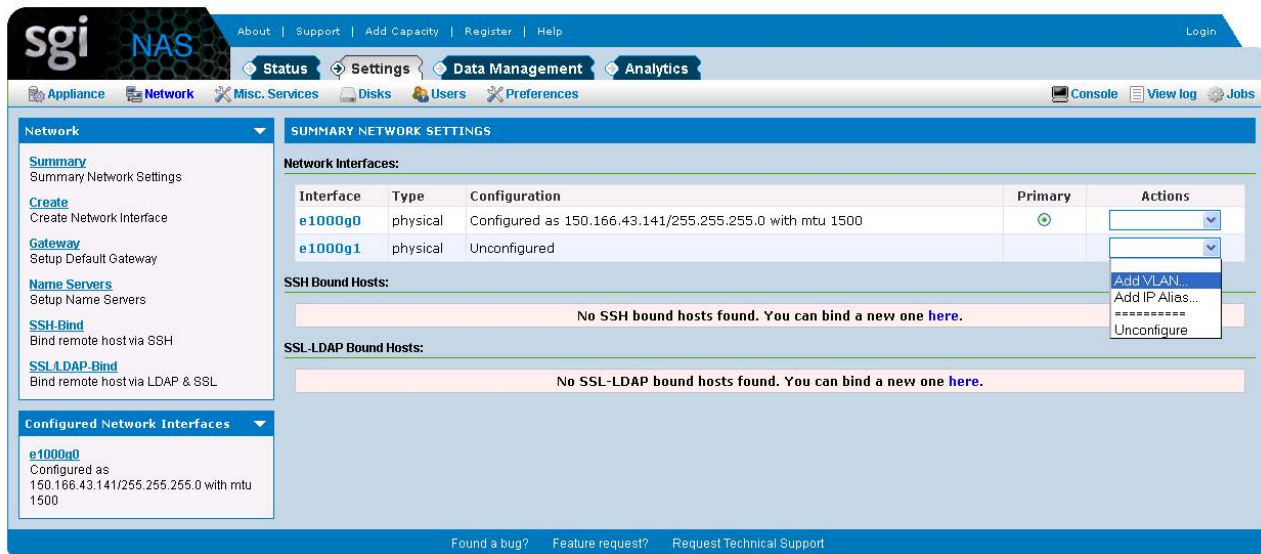
General information on VLAN and the 802.1Q standard is available on the web, for instance:

IEEE's 802.1Q standard 1998 version (2003 version)(2005 version)

Once created, a VLAN can be modified via DHCP or statically, exactly in the same way you would configure an existing physical networking interface or aggregated link. For instance:

```
nmc:/$ setup network interface vlan e1000g3001 static
```

The same ability to show and administer VLANs is available in the Web GUI.







VLANs can be provisioned over physical interfaces and aggregated links. Both options are supported.

## 21.4 IP Aliasing

IP aliasing associates more than one IP address with a given networking interface. Physical networking interfaces, VLANs, and aggregated links can be aliased. Use the NMC command

```
nmc:/$ setup network interface
```

to configure an IP alias.

For instance, the following configures an IP alias over the existing (physical) interface e1000g1:

```
nmc:/$ setup network interface e1000g1 ipalias
IP alias Id :
-----
The ID associated with the IP alias link. Press Ctrl-C to exit.
```

Once created, an IP-aliased interface can be configured via DHCP or statically, as you would configure an existing physical networking interface or aggregated link. Here is an example:

```
nmc:/$ setup network interface ipalias e1000g1:2
Option ?
destroy show dhcp static unconfigure
-----
Navigate with arrow keys (or hjkl), 'q' or Ctrl-C to quit
```

NMV can also be used to set up IP aliases.

## 21.5 TCP Ports used by SGI NAS

SGI NAS by default listens on the following management ports:

- 2000 – Web GUI (NMV)
- 2001 – SGI NAS Management Server (NMS)

- 2002 – SGI NAS Management Console daemon (NMCd)
- 2003 – SGI NAS Management DTrace daemon (NMDTrace)

• <b>21/tcp</b>	FTP
• <b>22/tcp</b>	SSH
• <b>443/tcp</b>	WebDAV (https)
• <b>111/tcp</b>	Sun RPC
• <b>139/tcp</b>	CIFS (netbios)
• <b>445/tcp</b>	CIFS
• <b>873/tcp</b>	RSYNC
• <b>2000/tcp</b>	Appliance's Web GUI (NMV)
• <b>2001/tcp</b>	NMS
• <b>2002/tcp</b>	NMC
• <b>2003/tcp</b>	NMDTRACE
• <b>2049/tcp</b>	NFS
• <b>4045/tcp</b>	NFS
• <b>10000/tcp</b>	NDMP server
• <b>9999/tcp</b>	Remote Replication

Disabling a network service closes the corresponding listening port. To disable a given service, please use the NMC command:

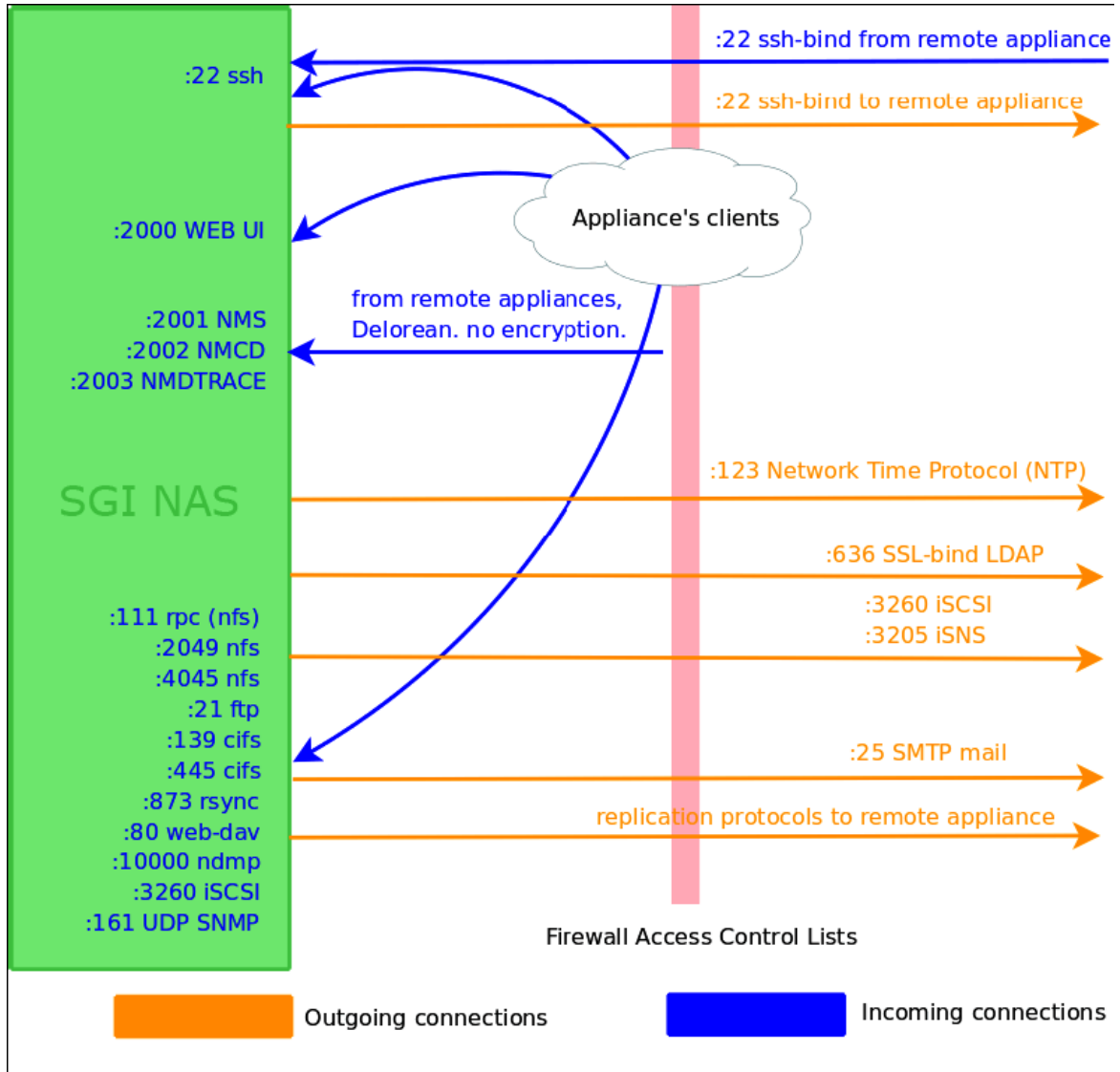
```
nmc:/$ setup network service
```

or the corresponding NMV page.

In addition to the ports open on the appliance itself, SGI NAS communicates to an outside TCP and UDP servers on the following IANA documented ports:

- 22/tcp – SSH (ssh-bind to remote appliances)
- 123/udp – NTP
- 636/tcp – LDAP
- 3260/tcp – iSCSI initiator
- 3205/tcp – iSNS
- 25/tcp – SMTP (fault reporting, tech support requests)

The following diagram provides a basic port coverage and commentary:





## 22 Managing the Appliance

You can adjust a variety of SGI NAS system settings using the NMC command:

```
nmc:/$ setup appliance
```

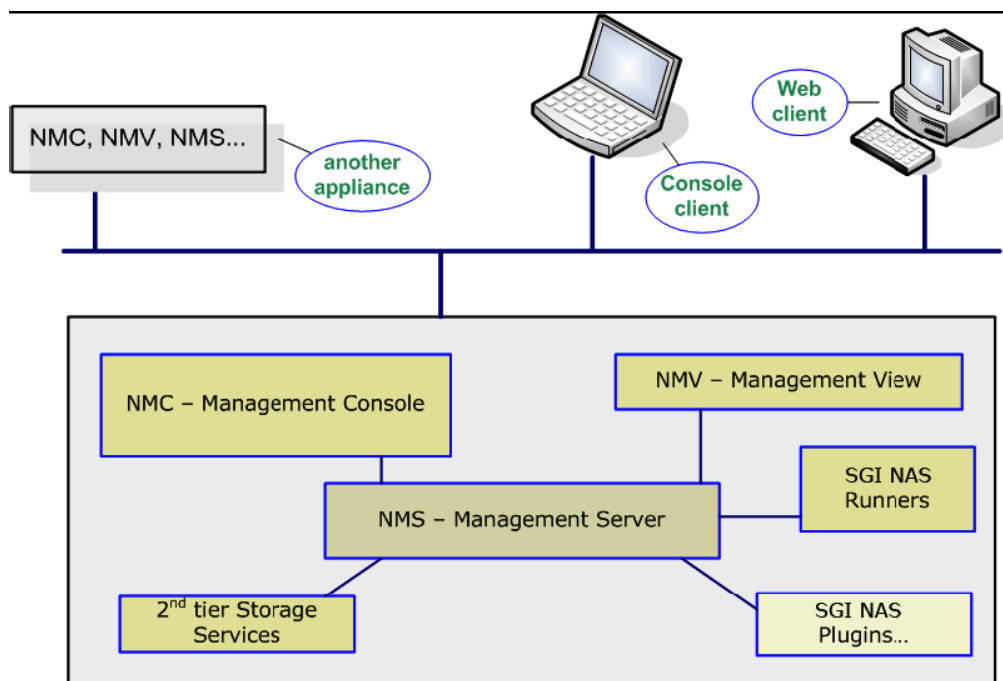
Here are some of the options:

- poweroff – power off the appliance
- checkpoint – take a system checkpoint
- domainname – set the domain name
- hostname – set the host name
- mailer – change settings for email notifications
- netmasks – set subnetwork masks
- reboot – restart the appliance
- timezone – change the timezone for the appliance
- user – edit appliance user information
- usergroup – edit appliance user group information

You can see the full list of available options when you type 'setup appliance' .

### 22.1 Secure Access

SGI NAS systems provides secure access to other SGI NAS systems as well as administrative management client applications on the network. The picture below illustrates an appliance (with its main functional blocks) being accessed from/by another appliance and two management clients. The inter-appliance access is executed either via SSH, or via SA-API (The SGI NAS communication API), or both. All management client applications, whether developed internally by Silicon Graphics International or by 3<sup>rd</sup> parties, access appliance via SA-API.



In all cases, access to appliance requires client authentication. SGI NAS supports two authentication mechanisms:

- via IP address of the client machine
- via ssh-keygen generated authentication keys.

The 2<sup>nd</sup>, ssh-keygen based, mechanism is the preferred one. This is the mechanism used by SGI NAS systems to communicate between themselves. The latter is required to run storage replication services, to execute in a group mode, to `switch` between appliances for the purposes of centralized management. To enable inter-appliance communication, simply use NMC 'ssh-bind' command (see '[14.12. How to establish SSH-binding](#)'). Once the appliances are ssh-bound, all the capabilities mentioned above are enabled automatically and executed in a secure way.

To use IPv4 address based authentication, simply make sure that IP address of your management client machine is recorded on the appliance, via NMC (for v3.0.3 and lower):

```
nmc:/$ setup appliance authentication iptable
```

For SGI NAS v3.0.4 or higher use the following NMC command:

```
nmc:/$ setup appliance authentication dbus-iptables
```

Administrative access to the appliance is required to perform this command. Alternatively, to use ssh-keygen generated authentication keys with your management application running on Windows, Linux or any other platform, use the same NMC command:

```
nmc:/$ setup appliance authentication keys
```

## 22.2 Registering the Commercial Version

You can display licensing information in NMV by selecting the '**About**' link or use the following NMC command:

```
nmc:/$ show appliance license
```

This will indicate whether you are using the trial or commercial edition, and how many days are left in a trial.

After obtaining the commercial license, you can register in NMC using the command

```
nmc:/$ setup appliance register
```

or click on the '**Register**' link at the top of the page in NMV. In NMV a form similar to the following will appear, where you can enter the new license key:

You can request additional capacity using the '**Add Capacity**' link in NMV. This will also require you to update the license key. Capacity is based on raw disk drive capacity, and log, cache, and spare devices are excluded from the calculation.

## 22.3 Installing/ Removing plug-ins

SGI NAS extension, a pluggable module or plug-in, can be easily added and removed. SGI NAS plug-in implements a certain well-defined extended functionality and uses the same Storage Appliance API (SA-API) as all the rest software components including 2<sup>nd</sup> tier storage services, fault management and statistic collection *runners*, and the management console and management web GUI. At installation time, plug-in integrates itself with the appliance's core software.

The currently available plug-ins include:

- Auto-CDP (Continuous Data Protection). Must be installed on a pair of (replicating) appliances.
- NMV based API browser
- I/O and network performance benchmarks
- network traffic probe
- HA plug-in called simple-failover. Must be installed on each appliance - member of a simple-failover group.
- virtualization management plug-in VM DataCenter
- Target FC
- WORM (Write Once, Read Many)

A complete list of SGI NAS plug-ins is available from your SGI sales representative.

Note that plug-ins are **not** downloadable from the website. Pluggable modules are distributed exactly in the same way as SGI NAS software upgrades and updates: via built-in reliable transactional upgrade mechanism (see SGI NAS [overview](#), Section [20.5. Upgrades](#)). To list already installed plug-ins, as well as plug-ins available for installation, run:

```
nmc:/$ show plug-in
nmc:/$ show plug-in remotely-available
```

To administer an existing plug-in or install a new one in NMC, type:

```
nmc:/$ setup plug-in
```

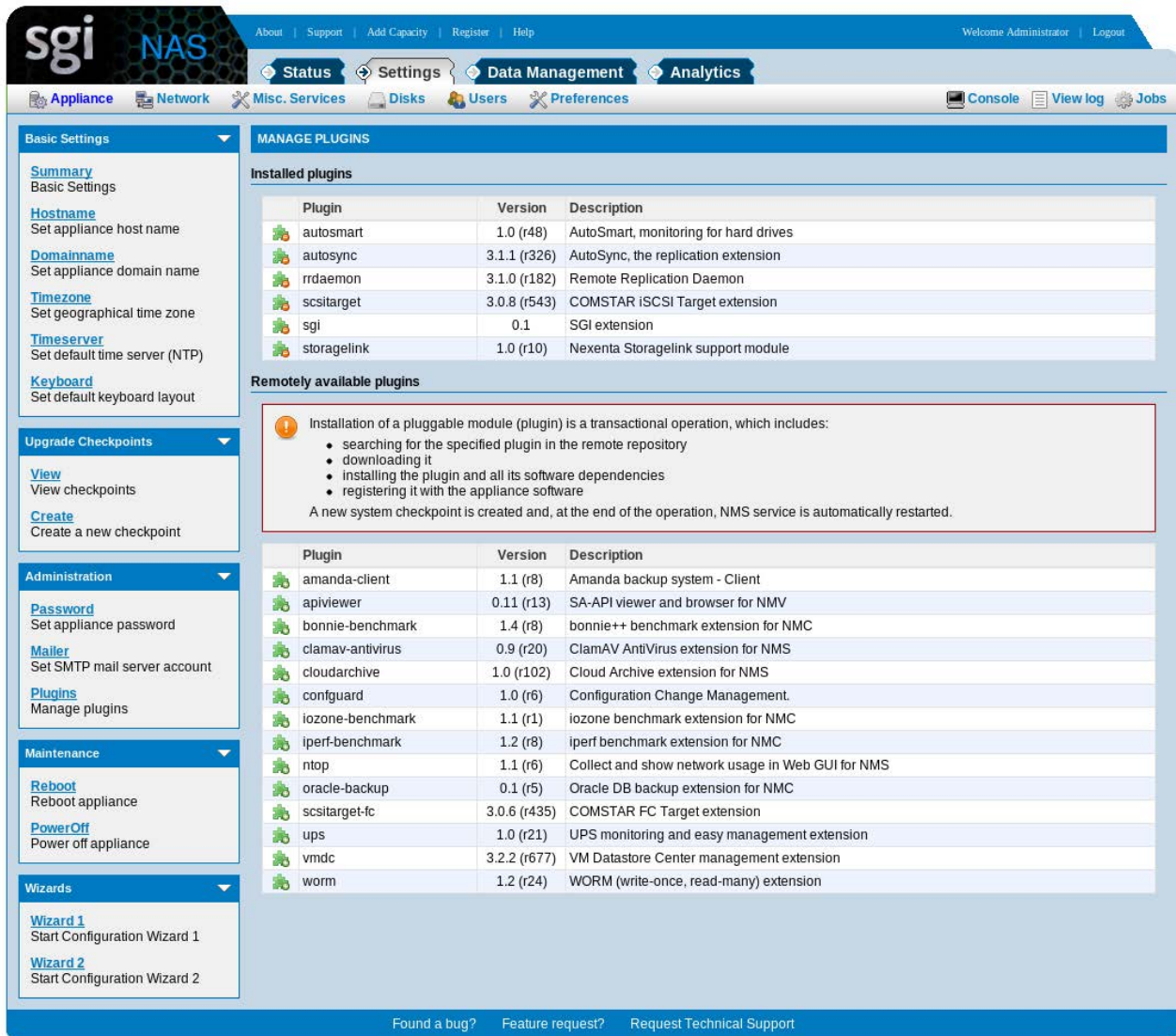
Alternatively, you can view, install and uninstall the SGI NAS extension modules using the system web GUI.

Free Trial users - please note that commercial plug-ins are available **upon request**. When requesting, please specify:

- Your SGI NAS license key
- Pluggable module name

SGI NAS pluggable extensions can be viewed and inspected:





## 22.4 Saving and Restoring Configurations

An operational appliance may have multiple running auto-services: auto-tiers, auto-scrubs, auto-snaps and auto-syncs. The appliance saves auto-services configuration along with some appliance's settings periodically and keeps up to three saved configurations. Sometimes, it is required to execute the save action manually or to restore all or a part of configuration. These actions can be performed with the NMC command:

```
nmc:/$ setup appliance configuration
```

## Common tasks

- Manually save the configuration:

```
nmc:/$ setup appliance configuration save
```

- Manually restore all or part of the configuration:

```
nmc:/$ setup appliance configuration restore
```

- Restore all auto-service configuration of a volume (the appliance and other volumes configurations are not changed):

```
nmc:/$ setup appliance configuration restore -V <volume-name>
```

- Save appliance configuration to the specified destination folder:

```
nmc:/$ setup appliance configuration save -F <directory-name>
```

- Restore configuration from specified destination folder:

```
nmc:/$ setup appliance configuration restore -F <directory-name>
```

The directory name can be relative or absolute. Using either choice has some implications for SGI NAS configurations:

- Relative: Then an appliance-specific configuration (mailer, plug-ins, hostname settings, etc.) is saved on the syspool and auto-services configurations are saved on volumes (each volume contains configuration of it's own auto-services only).
- Absolute: All configuration is saved to the given directory. Inside it sub-directories for each volume are created. Use this option to make a backup of the configuration.

Running the save or restore command without using the '-F' parameter makes appliance use a default directory name. To display the current value, run:

```
nmc:/$ setup appliance configuration location
```

After executing 'configuration restore' command you may choose from the list of available options:

```
nmc:/$ setup appliance configuration restore
all          auto-tier      nmc
auto-scrub   basic           nms
auto-snap    mailer          shared-services
```

Choosing 'all' restores all the mentioned option, including: Active Directory info, ID mapping info, network IP address, iSCSI targets, NFS and CIFS shares, share permission settings, etc.

## 22.5 Upgrades

You can upgrade the SGI NAS system using the NMC command:

```
nmc:/$ setup appliance upgrade
```

Sometimes upgrade cache files need to be cleaned up before the upgrade. Especially, if you get the following message: WARNING: The following packages cannot be authenticated!

You can run:

```
nmc:/$ setup appliance upgrade -C
```

or answer 'Yes', when you are asked: "Cleanup upgrade caches?".

To upgrade the data volume to the latest version, use the NMC command:

```
nmc:/$ setup volume <volumename> version-upgrade
```

To update all folders to the latest version, use the NMC command:

```
nmc:/$ setup folder version-upgrade.
```

To upgrade the version of the only one specified folder, run

```
nmc:/$ setup folder <foldername> version-upgrade
```

To upgrade specified folder and its subfolders, run:

```
nmc:/$ setup folder <folder name> version-upgrade -r
```

SGI NAS provides the possibility to upgrade locally, without an Internet connection. In order to upgrade locally, send a request to SGI Support for upgrade ISO with your license keys. After receiving upgrade ISO, burn a CD, insert it to your appliance CD-ROM drive and

run the following command:

```
nmc:/$ setup appliance upgrade -c
```

This command will automatically issue the local upgrade procedure.

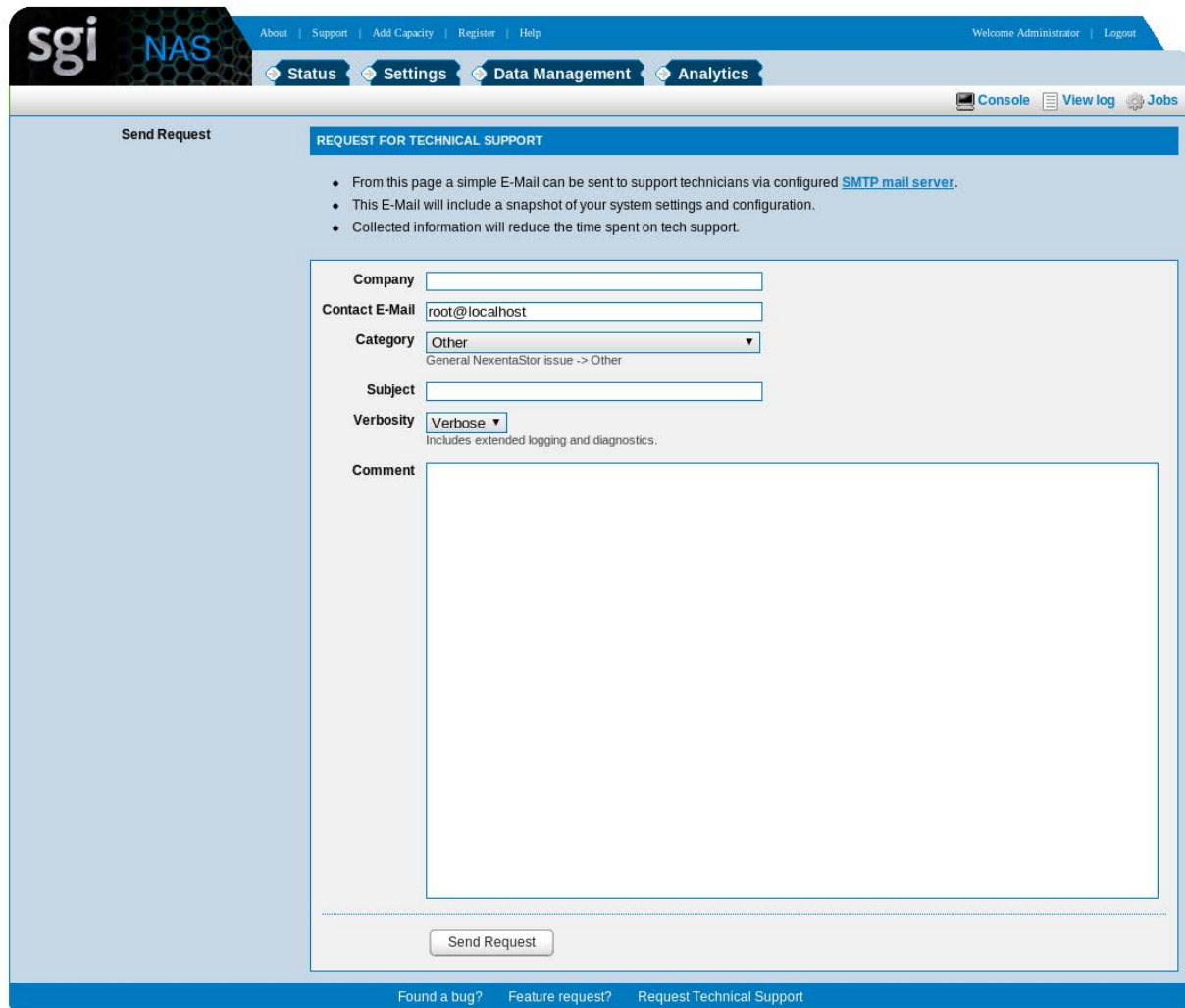
You can also save Upgrade ISO to USB flash drive and follow the Local upgrade instruction in NMC help text. To see these instructions, run:

```
nmc:/$ setup appliance upgrade -h
```

## 23 Contact information

### 23.1 Support request

To contact support at SGI, click on 'Support' :



The screenshot shows the SGI NAS web interface. The top navigation bar includes links for 'About', 'Support', 'Add Capacity', 'Register', and 'Help'. The user is logged in as 'Administrator'. The main menu has 'Status', 'Settings', 'Data Management', and 'Analytics'. The 'Send Request' page is titled 'REQUEST FOR TECHNICAL SUPPORT' and contains the following information:

- From this page a simple E-Mail can be sent to support technicians via configured [SMTP mail server](#).
- This E-Mail will include a snapshot of your system settings and configuration.
- Collected information will reduce the time spent on tech support.

The form fields are:

- Company:**
- Contact E-Mail:**
- Category:**  (Dropdown menu with options: General NexentaStor issue -> Other)
- Subject:**
- Verbosity:**  (Dropdown menu with options: Includes extended logging and diagnostics.)
- Comment:**

A 'Send Request' button is located at the bottom of the form. At the bottom of the page, there are links for 'Found a bug?', 'Feature request?', and 'Request Technical Support'.

or type the following NMC command:

```
nmc:/$ support
```

which will then prompt for a subject and message.

## **23.2 Other resources**

For licensing questions, please contact your SGI sales or support representative.