SGI NAS

Namespace Cluster User Guide

Release 3.1.x

007-5903-001

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, Supportfolio 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. 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. VMware is a registered trademark of VMware, Inc.
All other trademarks mentioned herein are the property of their respective owners.

Document Number: 007-5903-001

Table of Contents

1 Overview....5
1.1 Purpose....5
1.2 Audience....5
1.3 Document conventions....5
2 Introduction....7
2.1 Features....8
2.1.1 Ease of use....9
2.1.2 Quick start-up and namespace server autonomy....9
2.1.3 Single point of control....9
2.1.4 NFS v4.1 support....9
2.1.5 Live migration....10
2.1.6 Scalability and load balancing....10
2.1.7 Multi namespace links....10
2.1.8 Just-in-time graphics....10
2.1.9 Statistics: CPU, IOPS and bandwidth allocation, total and per folder....10
2.2 Terminology....11
2.3 How does Namespace link work?....11
3 Installation....15
3.1 Requirements....15
3.2 Installation....15
3.3 Icons legend....16
4 NMV management....17
4.1 Creating new Namespace Cluster....17
4.2 Adding a namespace server....18
4.3 Adding a logical group....20
4.4 Adding namespace link....20
4.5 Graphics....23
5 Migration....25

5.1 Overview....25
5.2 Live migration....26
5.3 Off-line migration....26
5.4 Migration management in NMV....27
5.5 Replication....30
6 Mounting the share via NFSv3....31
7 Fault management....33
7.1 Troubleshooting tips:....34
7.2 Known limitations....34
8 Contact information....35
9.1 Support request....35
9.2 Other resources....36

1 Overview

1.1 Purpose

This guide presents SGI NAS Namespace Cluster plug-in, which main goal is to provide management of multiple NFS servers, as well as advanced graphics and statistics. Namespace Cluster allows you to extend existing infrastructure with new servers and spread the workload between few servers instead of centralizing it on a single server. On the other hand, Namespace Cluster lets you manage all the NFS servers from every node.

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

One of the main problems of modern network infrastructures is the limitations created by the initial architecture design. As far as the number of users and applications simultaneously accessing the data constantly increase, any given storage server potentially becomes a bottleneck, in terms of available I/O bandwidth, CPU, memory, network and disk I/O resources. The solution for all mentioned issues involve various techniques of spreading the I/O workload over multiple storage servers, which can be integrated to the existing network infrastructure.

Namespace Cluster (SGI NAS) plug-in is designed to solve two major problems:

- Simplify management of multiple NFS servers, by isolating clients from actual physical locations of the shared server-based filesystems
- Remove the 'single server' bottleneck resulting from any/all of the following:
  - filesystem growing over time to the size that is difficult to manage via a single server 'resource'
  - increasing number of NFS clients simultaneously accessing shared file data
  - increasingly powerful and demanding applications on the client side that access shared file data

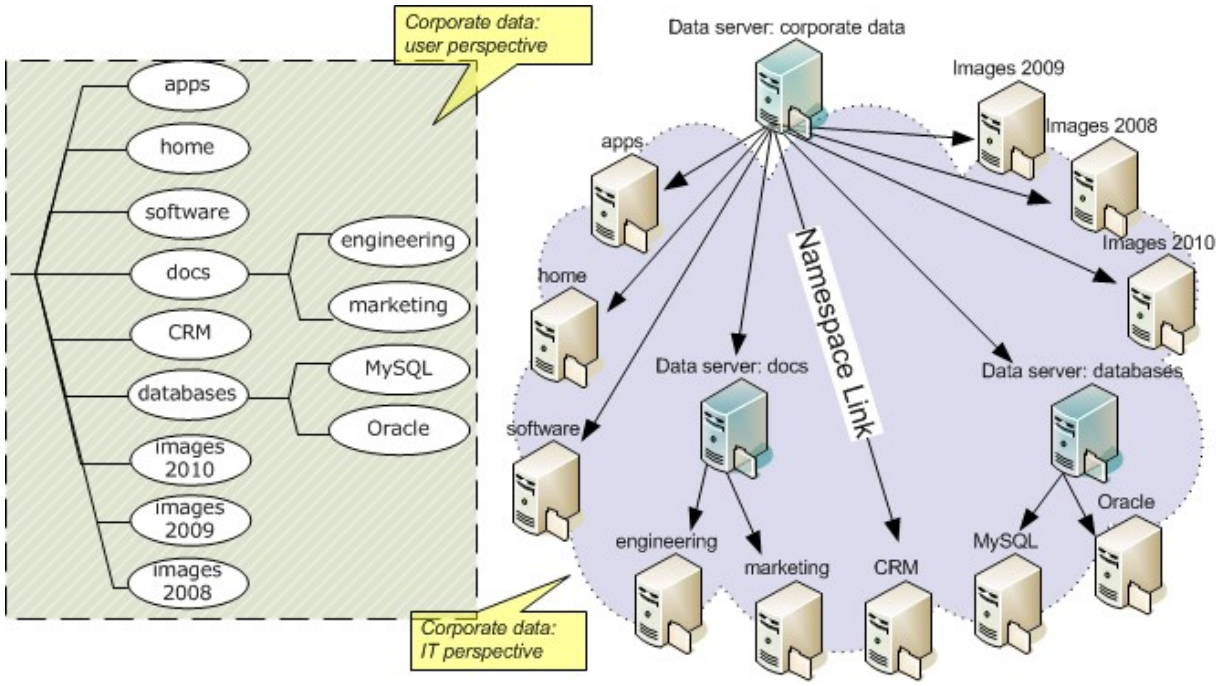
NAS clients have always been able to mount folders exported by multiple NAS servers wherever they desired in their local file system. However, maintaining a uniform naming scheme across multiple clients requires extra work from system administrators. The tools available to perform that work do not scale well.

Namespace cluster provides an easy and intuitive way to build a global namespace of any size, where directories and files are transparently distributed over any number of servers. The primary mechanism used to build a global namespace is called 'namespace link'. Namespace link has a name and two components: namespace server and a folder on this namespace server. For instance, if a link named 'corp/docs' has a value 'other-server:real-docs', clients accessing \corp/docs' would be automatically and transparently redirected to the corporate

documents residing on the 'other-server' in its 'real-docs' folder.

Storage administrator achieves various capabilities with Namespace Cluster. For end-users and client side applications, the names of the files and directories (that is, the 'name space') remains the same, independently of where data is physically stored. This allows administrators to manipulate the data according to system needs, without requiring users or applications to know the physical location of the different file systems and without incurring the overhead of updating user's machines. To users, all tiers of the namespace appear as a single large (possibly, extremely large) file system.

The following diagram represents the principal of work of Namespace Cluster from two different points of view: user and IT infrastructure:



Namespace cluster is a 'forest' of servers containing multiple namespace servers.
SGI NAS must be used as a namespace server. More than a single location of any given user folder WILL be supported as well.

2.1 Features

- Ease of use

- Quick start-up and namespace server autonomy
- Single point of control
- NFS v4/NFS v4.1 support
- Live migration
- Scalability and load balancing
- Real-time performance statistics (graphics)
- Statistics: CPU, IOPS and bandwith allocation, total and per shared folder

2.1.1 Ease of use

Namespace Cluster provides user-friendly interface that makes it easy and intuitive to configure (or reconfigure) the cluster of appliances of any size, and monitor its health and performance.

2.1.2 Quick start-up and namespace server autonomy

Each node in a cluster is equal to others and **autonomous** . You can access and perform operations from any cluster node.

2.1.3 Single point of control

All the namespace servers can be managed from one point. All the changes performed on one namespace server will be automatically implemented on the others.

2.1.4 NFS v4.1 support

Namespace Cluster supports both a uniform name space implemented by multiple file servers (also called in this document 'namespace servers'), and enables migration of files and directories from one server to another. The current implementation relies on NFSv4.1; future releases will include support for Microsoft’s Distributed File System (DFS).

2.1.5 Live migration

Possibility to transfer the data to new location incrementally, without unmounting shared folder.

2.1.6 Scalability and load balancing

The plugin allows (and makes it easy) to distribute NFS workload between namespace servers. After redirecting the client to the referred data folder, namespace server “removes” itself from the data path, and all the subsequent communications are happening directly between the client and the corresponding namespace server.

2.1.7 Multi namespace links

Namespace Cluster provides possibility to indicate more than one location of the required file system. It can be used in emergency scenario, when one of the locations is unreachable. This 'multi-location' functionality is not yet supported, and is currently being considered for future releases.

2.1.8 Just-in-time graphics

Namespace Cluster provides advanced monitoring system for NFS folders and namespace servers activity. See section 3.4 Graphics for more information.

2.1.9 Statistics: CPU, IOPS and bandwidth allocation, total and per folder

Namespace Cluster collects statistic information about folders and namespace servers activity which is shown on the graphics.

Generally, Namespace Cluster dashboard provides all tools to build and monitor a unified multi-server namespace:

- multi-view (User, Admin) control panel at the left;
- migration/replication monitor;

- list, status, and capability to add/delete namespace servers;
- NFS (un)sharing;
- basic capacity management;
- integrated real-time graphics.

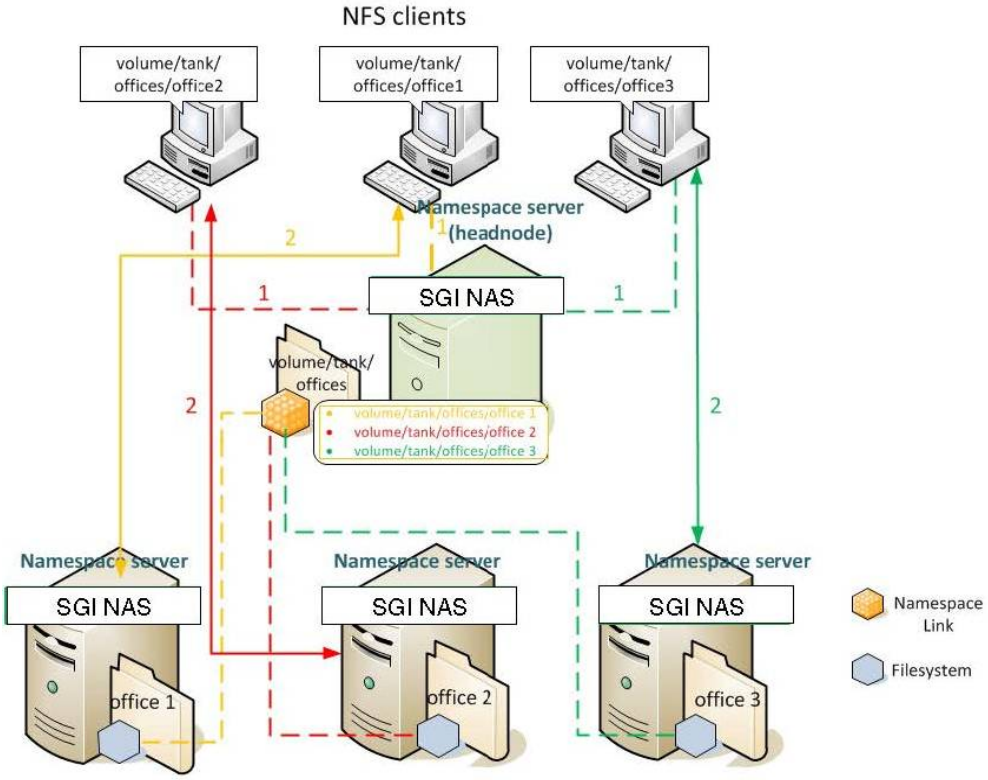
2.2 Terminology

Term	Definition
Namespace server	Namespace server is any appliance that is installed with the namespace cluster plugin and that is intended to be used as a node in the namespace cluster. Namespace server may provide NFS shares containing user data and referenced (via Namespace links) by other namespace servers.
Namespace link	Namespace linking allows to build a global namespace of any size, where directories and files are transparently distributed over any number of servers. Namespace link has a name and two components: namespace server and a folder on this namespace server. For instance, if a link named 'corp/docs' has a value 'other-server:real-docs', clients accessing 'corp/docs' would be automatically and transparently redirected to the corporate documents residing on the 'other-server' in its 'real-docs' folder.
Logical group	Logical group - provides for logical grouping of servers based on any criteria, e.g. same office, same subnet, same type of user or application data.

2.3 How does Namespace link work?

Following diagram illustrates the operational principle of Namespace link:

1. NFSv4 client traverses NFS share on a namespace server
2. NFSv4 client encounters namespace link
3. NFS client parses namespace link for NFS-accessible locations and mounts the first accessible.



For example, there is an NFS share on a Namespace server /volume/tank/offices, where namespace links to other physical servers are located. The servers represent the following directories: office1, office2 and office 3. When a client need to access /volume/tank/offices/office1, it addresses to namespace link in the share, which redirects it to the real dataset destination. This operation is transparent to users and appears as if all the required datasets are located on the same node.

Note that a namespace server takes part only in redirection procedure. After the redirection is done, the redirecting namespace server (sometimes called 'headnode') is not getting accessed anymore. Instead, the client accesses the referenced namespace server directly. It means that load is balanced between client and corresponding referenced namespace server and therefore namespace server is not overloaded.

Namespace server also takes part in relocation scenario. If primary location is down and there is more than one location defined for one share, namespace server redirects the client to another location[A].

A This functionality is not yet supported, and is currently being considered for future releases.

3 Installation

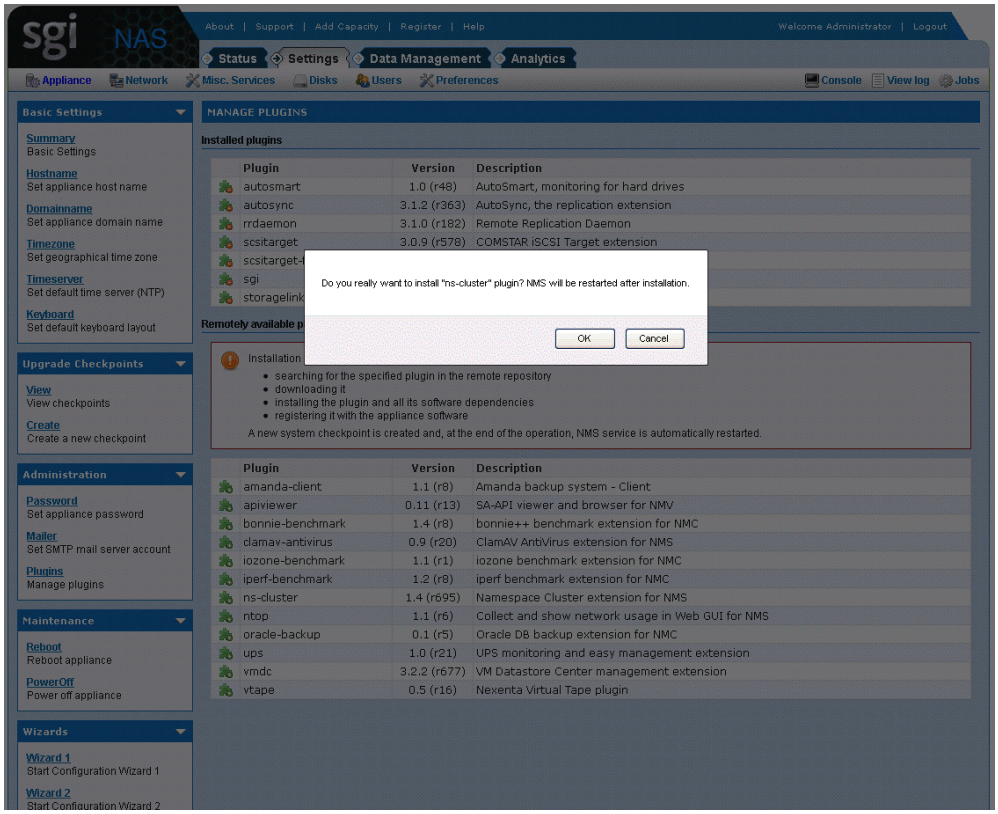
3.1 Requirements

- Namespace server: SGI NAS appliances with pre-installed '**ns-cluster**' plugin extension.
- Supported NFSv4.1 compliant client

3.2 Installation

'ns-cluster' plugin extension must be installed on all namespace servers in a cluster.

In NMV go to: **Settings** → **Appliance** and choose '**ns-cluster**' from the list of remotely-available plugins:

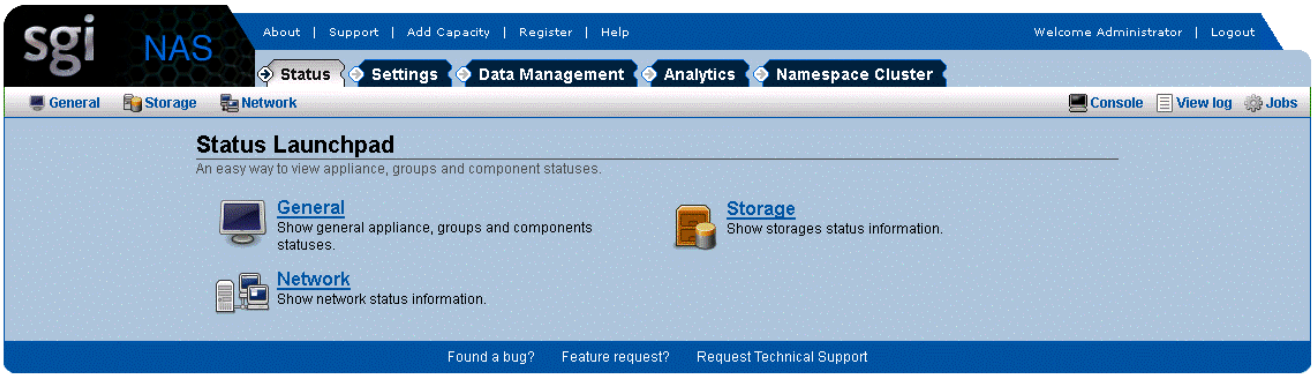


Installation requires NMS reboot.

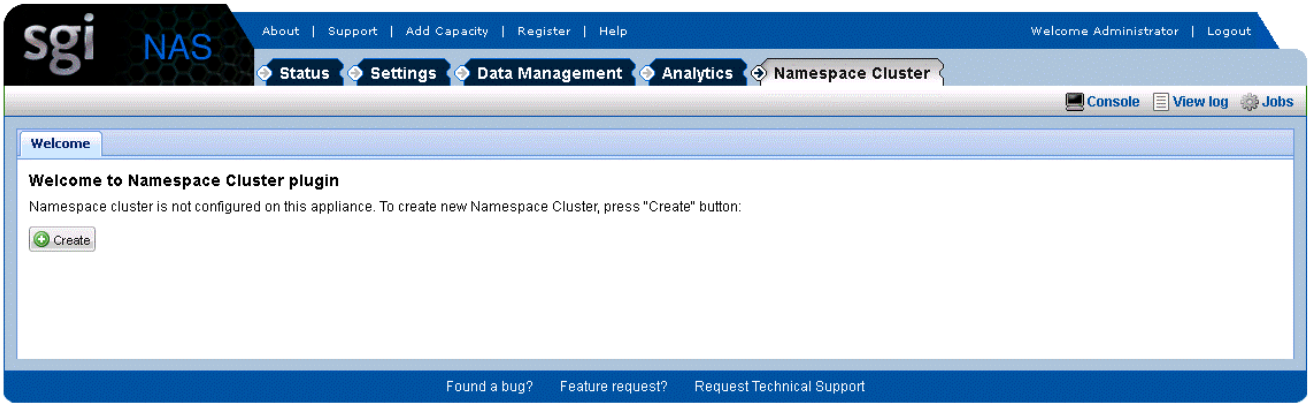
Corresponding NMC command:

```
nmc:/$ setup plugin install ns-cluster
```

After successful installation new 'Namespace Cluster' tab appears in NMV:



Click on NameSpace Cluster tab to enter Cluster menu:



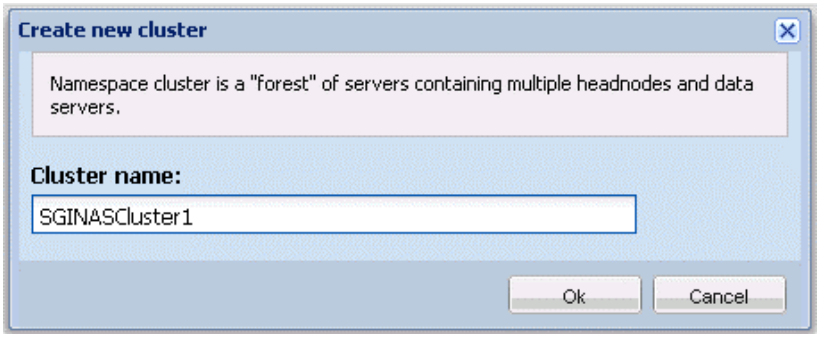
3.3 Icons legend

- - Namespace server
- - NFS shared folder
- - folder
- - Namespace link
- - Referenced data folder
- - Referenced data folder resides on a shared volume on HA-clustered appliances.

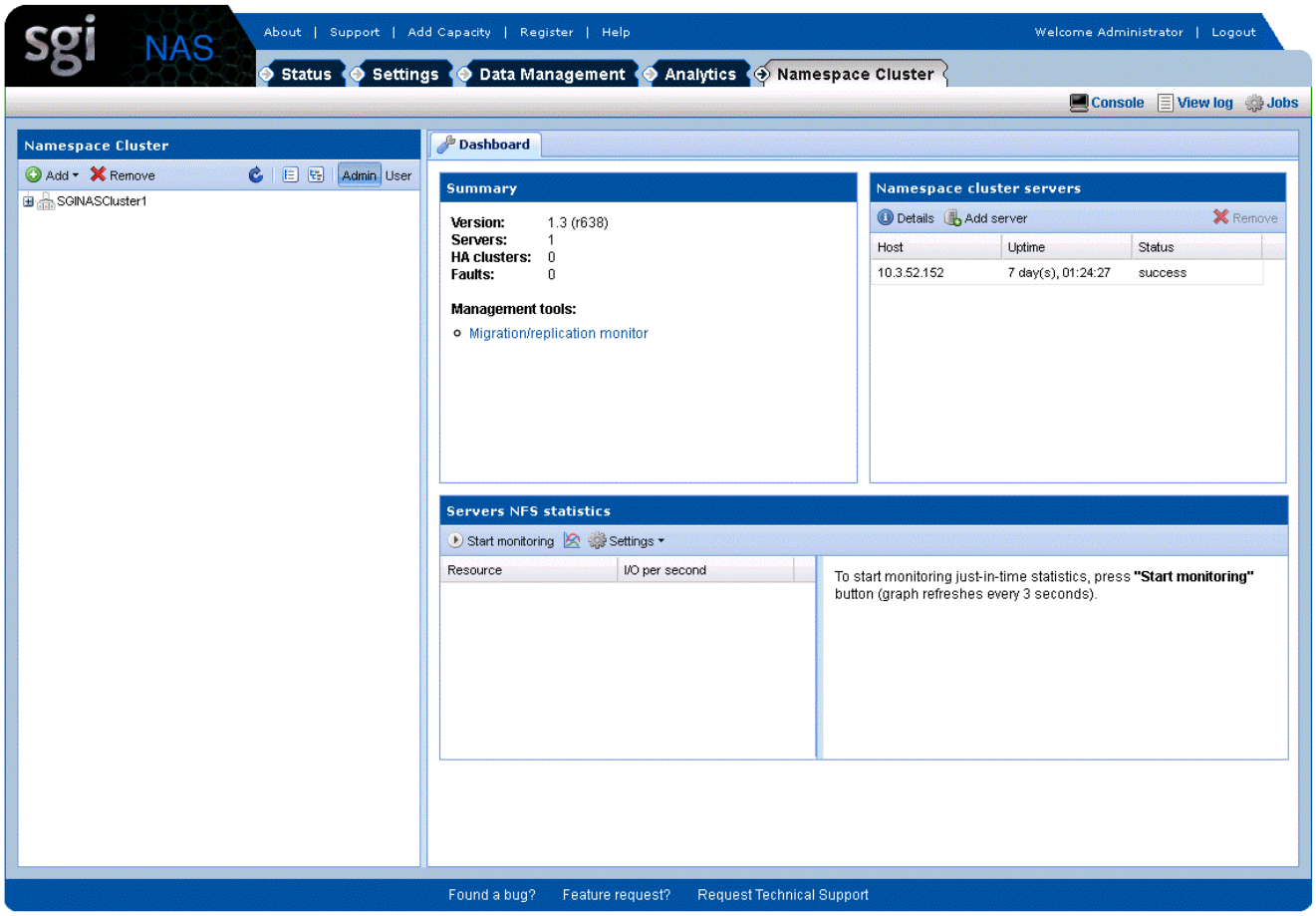
4 NMV management

4.1 Creating new Namespace Cluster.

To create new Namesapce Cluster click on '**Create**' button as it is shown on a picture below:

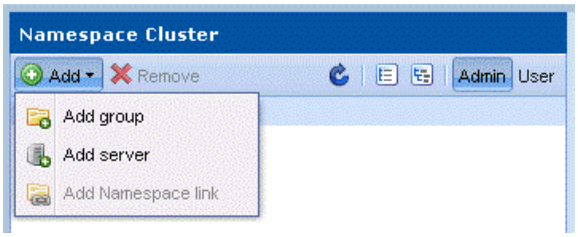


After successful initialization a screen with cluster status appears:

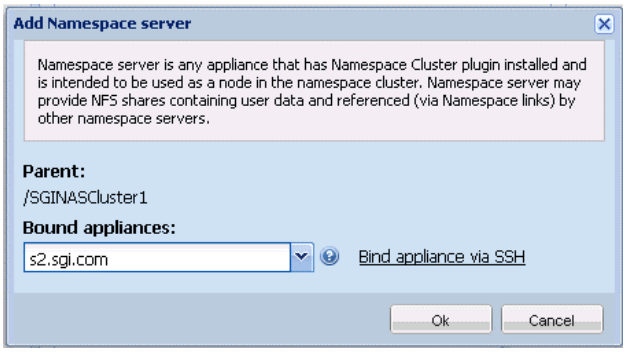


4.2 Adding a namespace server

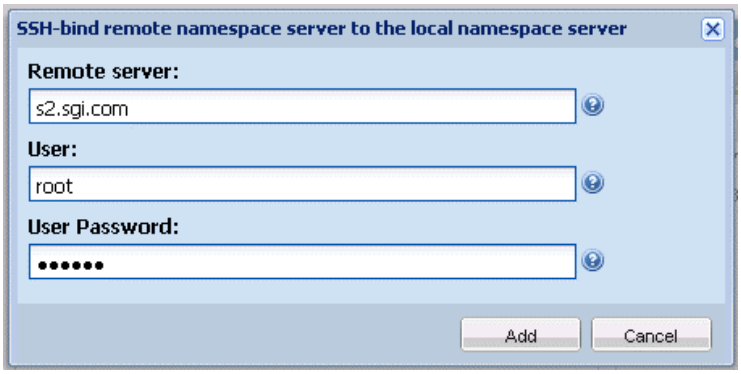
To add a new namespace server click on '**+Add**' and choose '**Add server**' from the drop-down list.



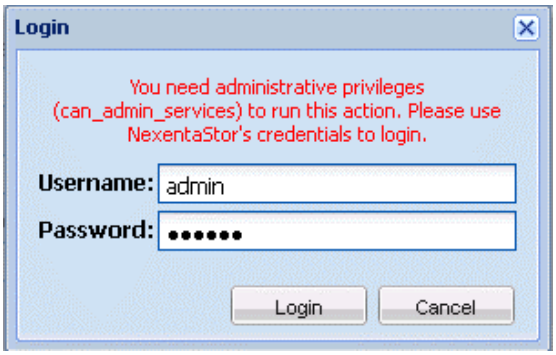
In a pop-up window choose a namespace server from the drop down list of bound appliances and Click '**Ok**'.



To add a new namespace server you need a remote appliance ssh-bound to current appliance. If it wasn't bind previously, it can be bound by clicking on '**Bind new appliance**'.



To complete the operation, you need to be logged as a user with '**can_admin_services**' permission. Otherwise, you'll be asked to provide the required credentials:



See section '**How to establish SSH-binding**' in the SGI NAS User Guide, publication number 007-5860-00x**.**

Users permissions can be specified in **Settings→Users.** Read more in the SGI NAS User Guide.

If adding procedure was performed successfully, new namespace server appears in '**Namespace tree**' with all NFS shared folders**:**

The namespace server that we just added contains one shared folder – 'data/corp'

Note:

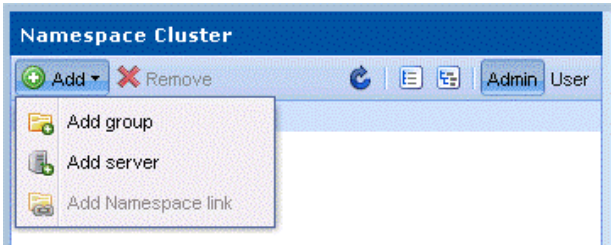
When a namespace server or a group, or namespace link is created on one namespace server in a cluster it is automatically synchronized with others (see section Single point of control)

Click on a shared folder to see the statistics.

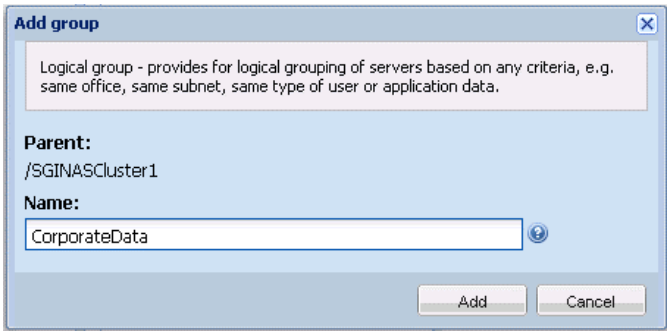
4.3 Adding a logical group

For ease of management namespace servers in a namespace tree can be unified in logical groups. You can specify any name for a logical group according to its assignment.

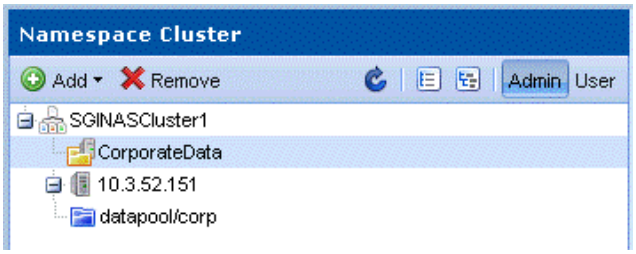
To create a new logical group, click on '+**Add**' and choose '**Add group**' from the the drop-down list:



In the pop-up window, specify '**Parent**', if there is any, or leave the line blank to place the group in the main tree. Click '**Add**':

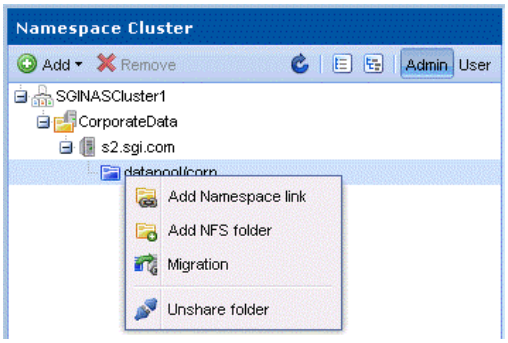


Newly created logical group should now appear in the '**Namespace tree**':



4.4 Adding namespace link

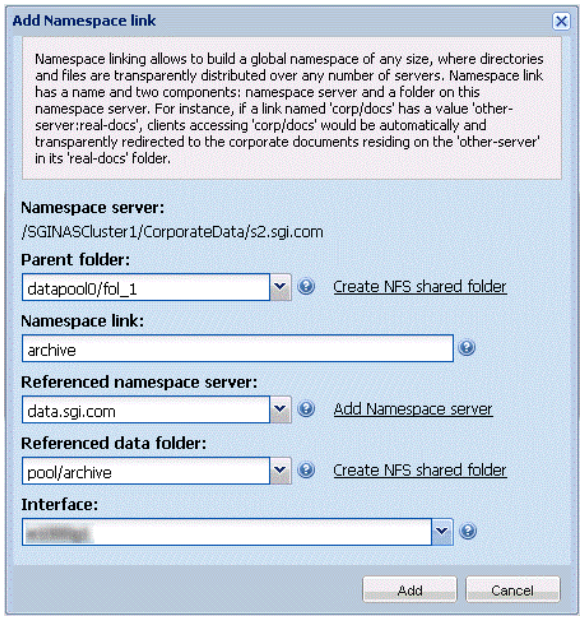
As soon as namespace server is created, namespace link can be added on this server. To add a new namespace link lick on '**+Add**' and choose '**Add namespace link**' from the drop-down list.



In the pop-up window, specify:

- A '**Namespace server**' where namespace link will be located
- In '**Parent folder**' choose a folder shared by NFS from the drop-down list where namespace link will be located (if it wasn't previously shared, it can be shared by clicking '**Create an NFS shared folder**').
- In '**Namespace link name**' field type the name of the namespace link.
- In '**Referenced namespace server**' choose a namespace server to which namespace link will be referenced
- In '**Referenced data folder**' choose a folder to which namespace link will be referenced

After filling all the required fields, click '**Add**'.



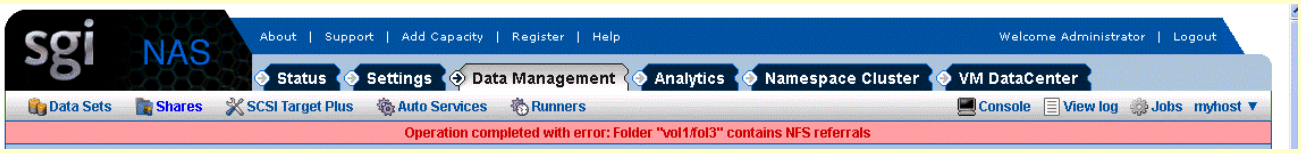
Several locations for one folder can be specified. In that case, if one of the locations is unreachable, it can be accessed via another. This functionality is not fully implemented yet. It will be scheduled for one of the future updating releases.

You can add new folder and share it via NFS automatically by clicking on '**Create an NFS shared folder**'.

If all the data is specified correctly, namespace link appears in the '**Namespace tree**'.

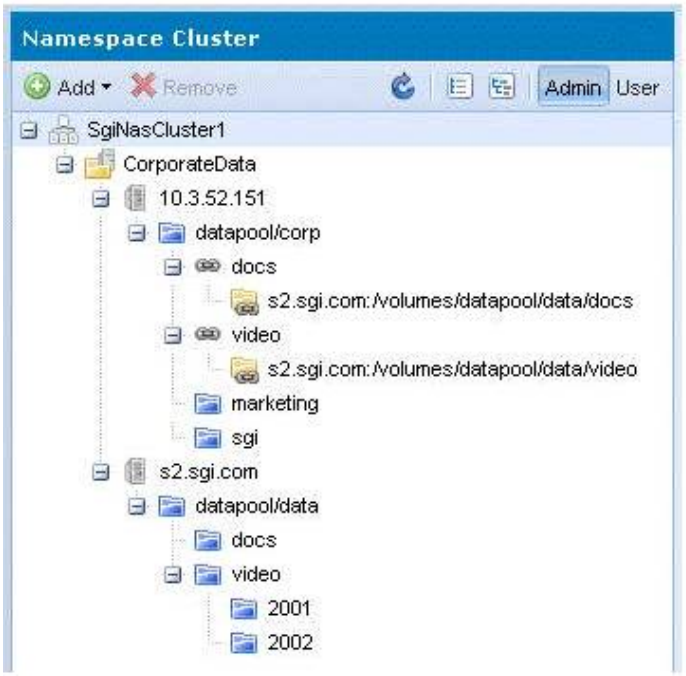
Note:

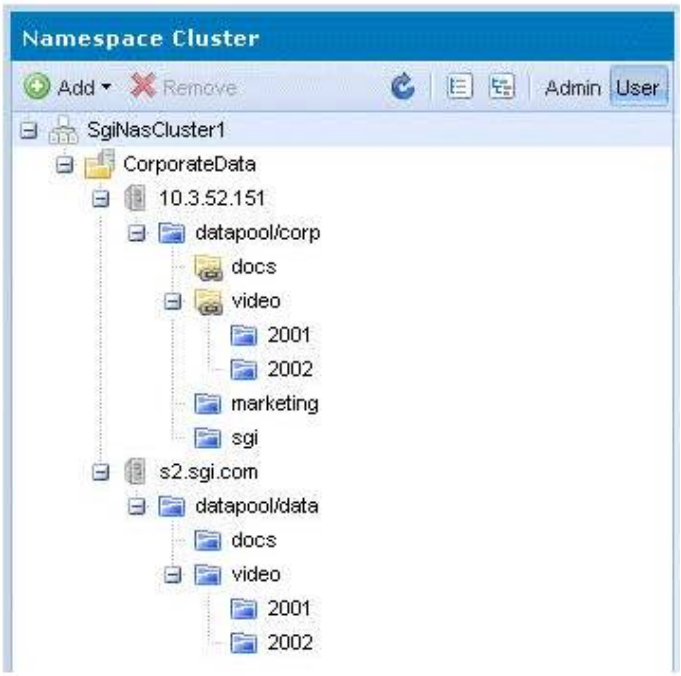
NFS share specified as location can't be unshared or removed while it contains namespace link:



To unshare or delete the folder, first remove all the namespace links it contains and then remove/unshare.

Keep adding required namespace links, groups and namespace servers to Namespace tree according to your configuration requires. On the following screen is the example of the tree structure. Note the difference in User and Admin view:

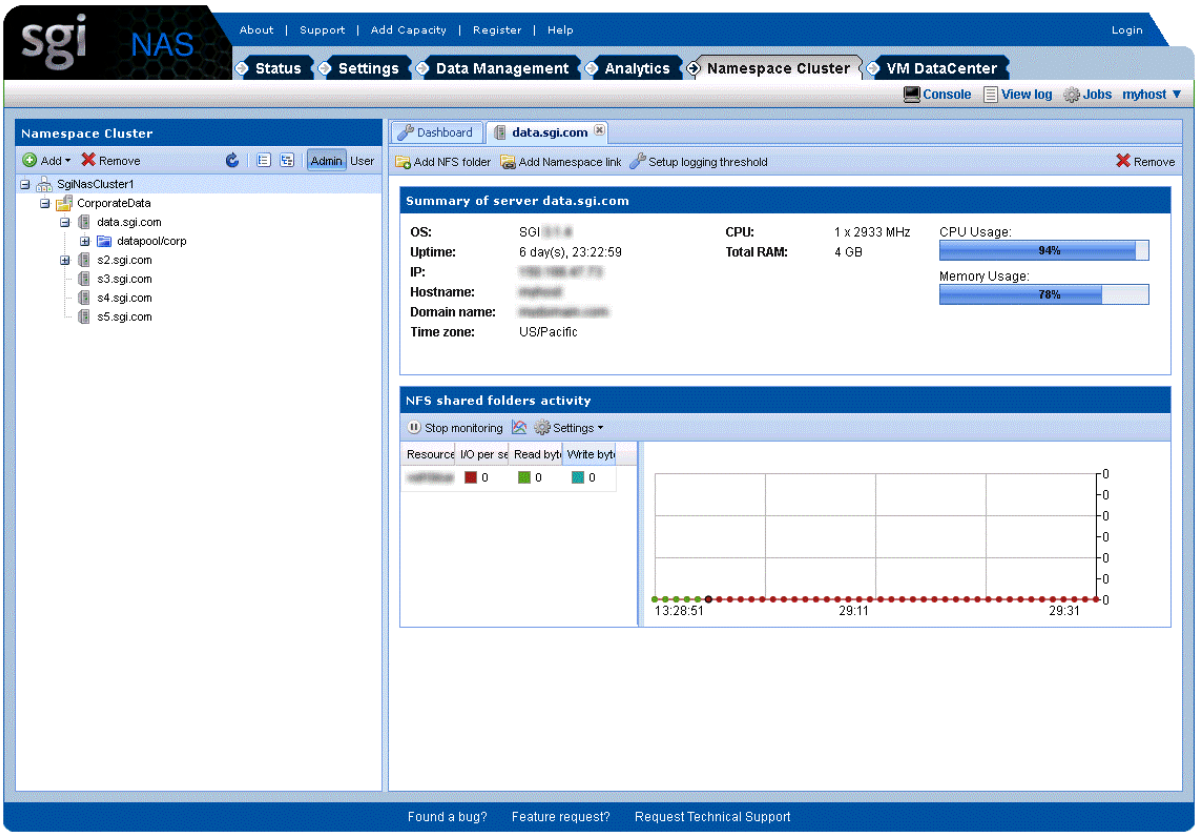




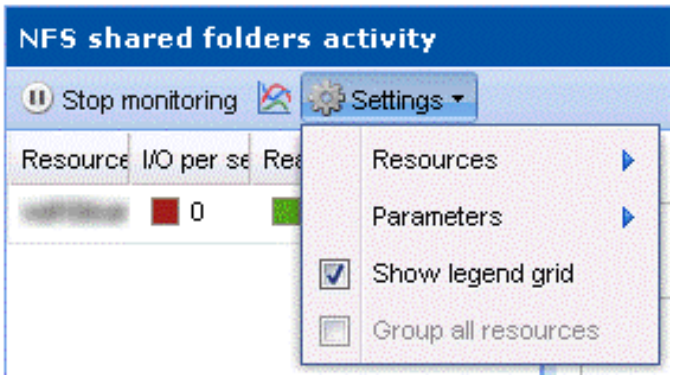
4.5 Graphics

Namespace cluster provides two types of graphics to monitor:

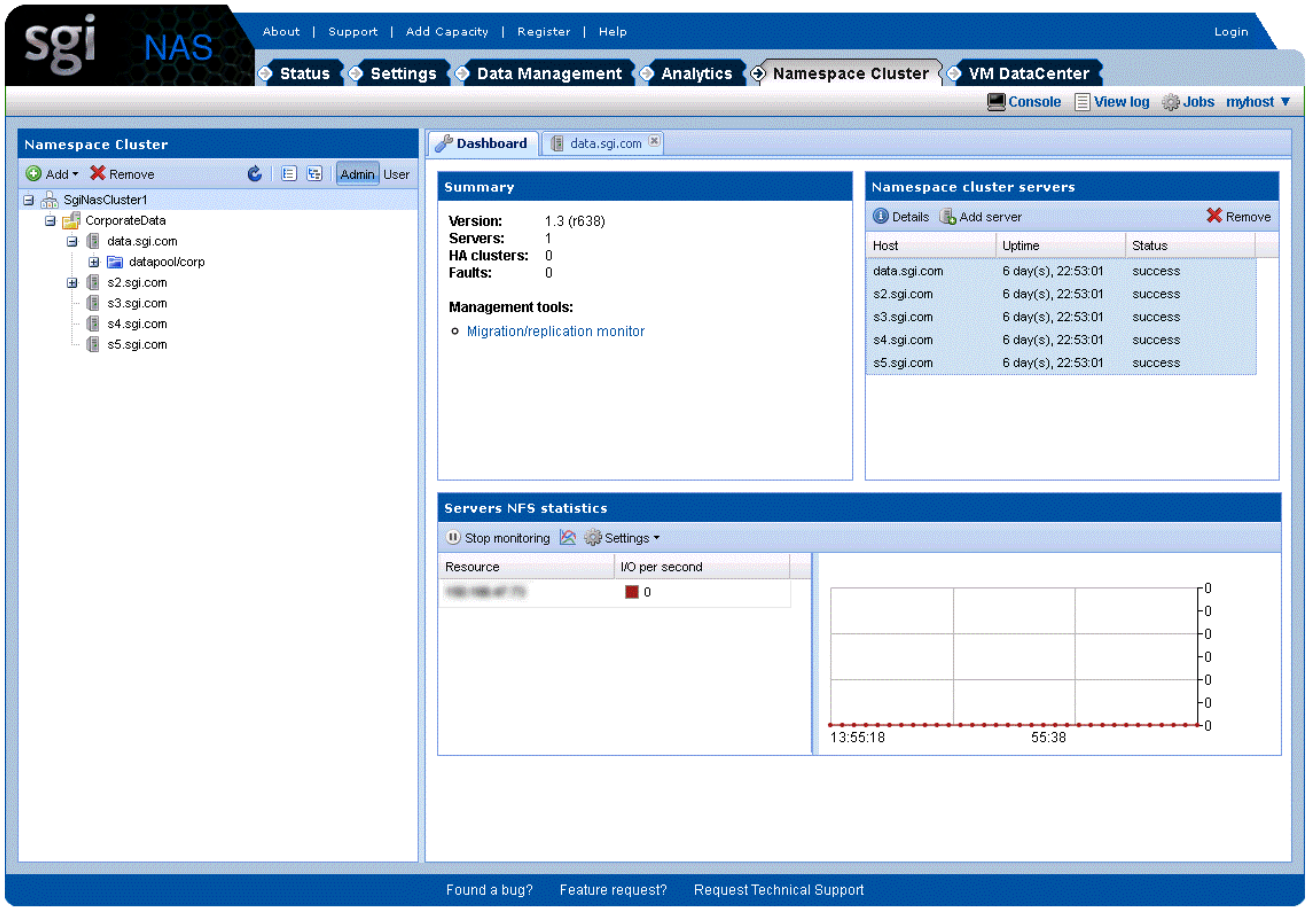
- NFS shares activity:



You can specify what parameters or resources to include to graph by clicking on 'Settings':



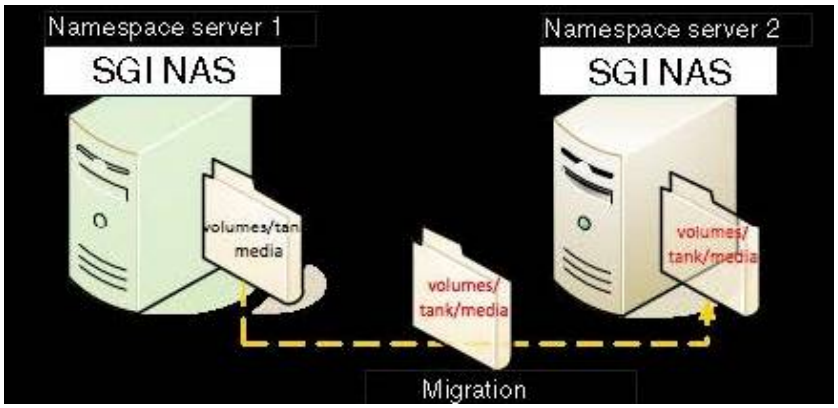
- Namespace servers activity can be seen on a dashboard:



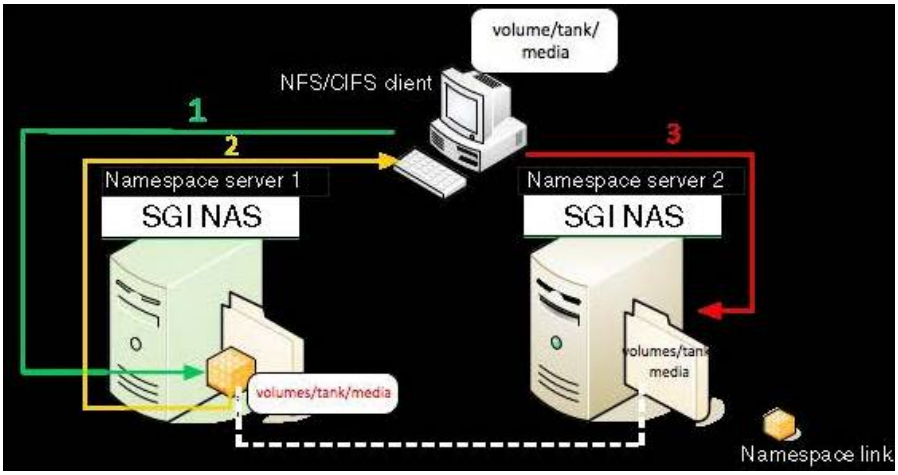
5 Migration

5.1 Overview

Namespace Cluster provides the possibility to perform data migration between namespace servers transparently to user. The operation performs the following: transfers the data from a namespace server to corresponding namespace server and then replaces the actual filesystem on the namespace server with namespace link. See the pictures below:



After creation of the namespace link the NFS clients is forwarded by the namespace link to the actual filesystem location. During migration between two namespace servers namespace link is created on source namespace server and NFS clients are redirected to the actual location. See the diagram below:



The whole procedure is absolutely transparent to the user, i.e. user obtains the access to **volume/tank/media** as if it was on the first Namespace Server.

Two types of migration can be associated with Namespace Cluster:

- Live migration
- Off-line migration

Note that migration is performed recursively: all the nested folders are moved to the destination.

5.2 Live migration

Live migration is a process of data relocation without switching off NFS share. It means that the data stays continuously accessible to users during migration. The mechanism of the operation is:

1. A snapshot of the folder is created and sent to corresponding namespace server.
2. After that the process of migration is started and continued by iterations. Every iteration the data is sent via Remote Replication transport protocol. All this time the share stays accessible to user and modifiable, so only the delta (the difference between the source and destination) is sent every iteration.
3. When delta is very small that it can be transferred for very short period of time (10 secs) Namespace Cluster unmounts the NFS share on the namespace server, substitutes it to namespace link and mounts the namespace link. Then the data becomes available again.

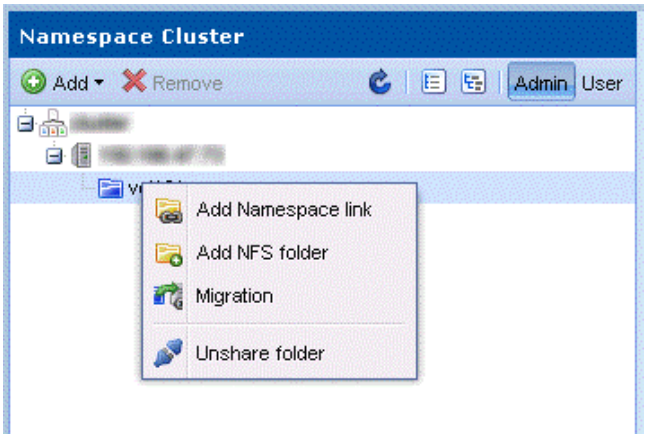
Live or incremental migration is used for big amount of data (for instance 10 TB), which migration process may take up to several hours or days.

5.3 Off-line migration

During off-line (non-incremental) migration, Namespace Cluster unmounts a share, transfers data to its new destination, substitutes the share with a namespace link to actual filesystem location and mounts a namespace link. This type of migration is useful in case of slow network connections between a source and a destination or to shares that don't require permanent availability.

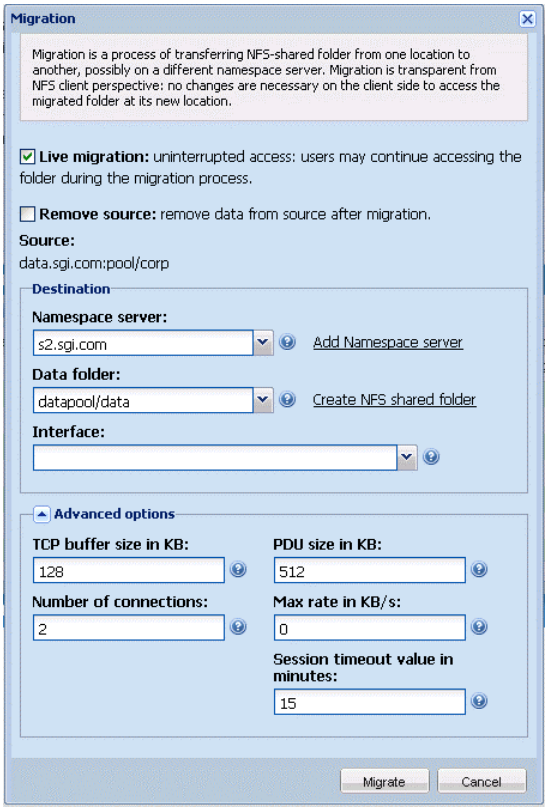
5.4 Migration management in NMV

To migrate a share from one namespace server to another and substitute it with Namespace Link, right-click on the share and choose 'Migration' as it's shown in the image below:



or click on share to open it in the right pane and click button.

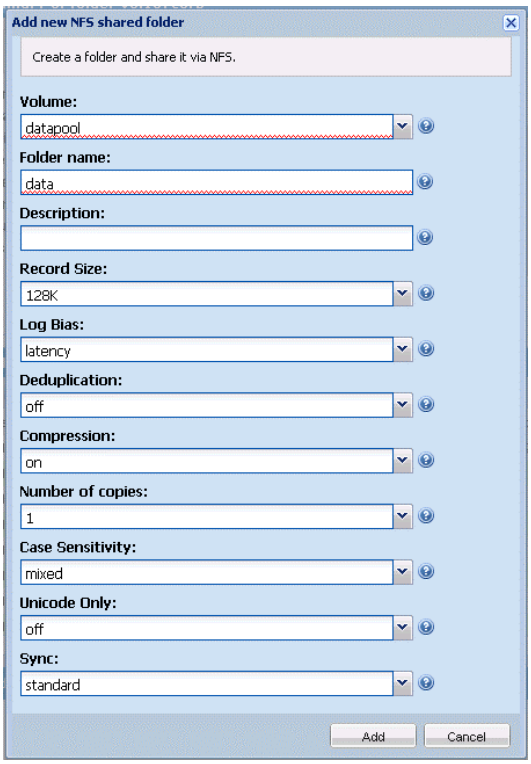
The following screen appears:



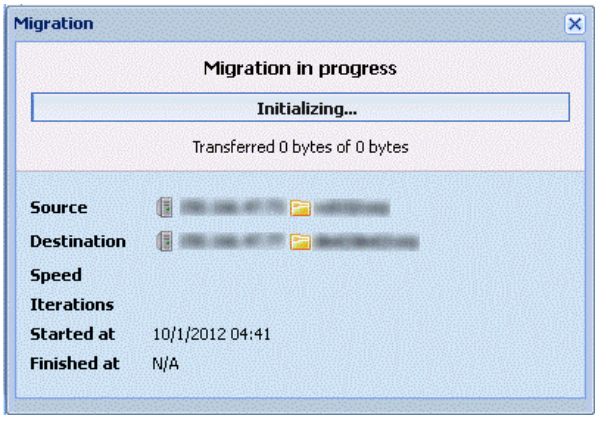
Note '**Remove Source**' checkbox. If you leave this checkbox unchecked, then a copy of original

dataset will remain on source. In our example it will be stored on source as 'pool/.corp'

If you don't have any NFS shares, click '**Create an NFS shared folder**' button to create new NFS share. The following screen will appear:



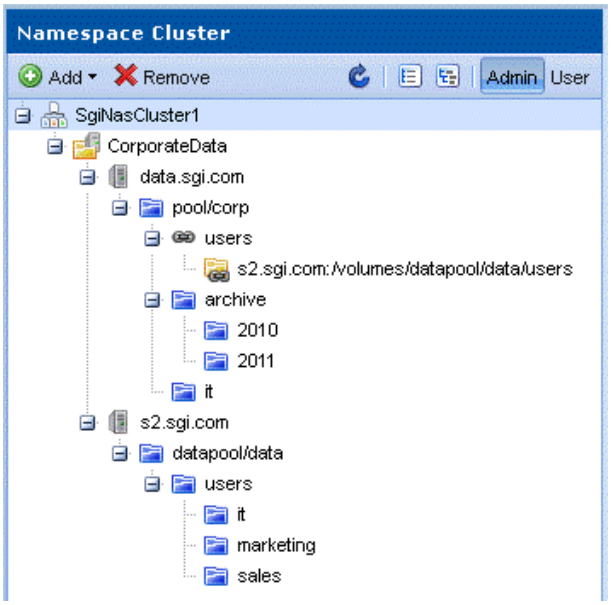
Fill the corresponding fields and click '**Add**':



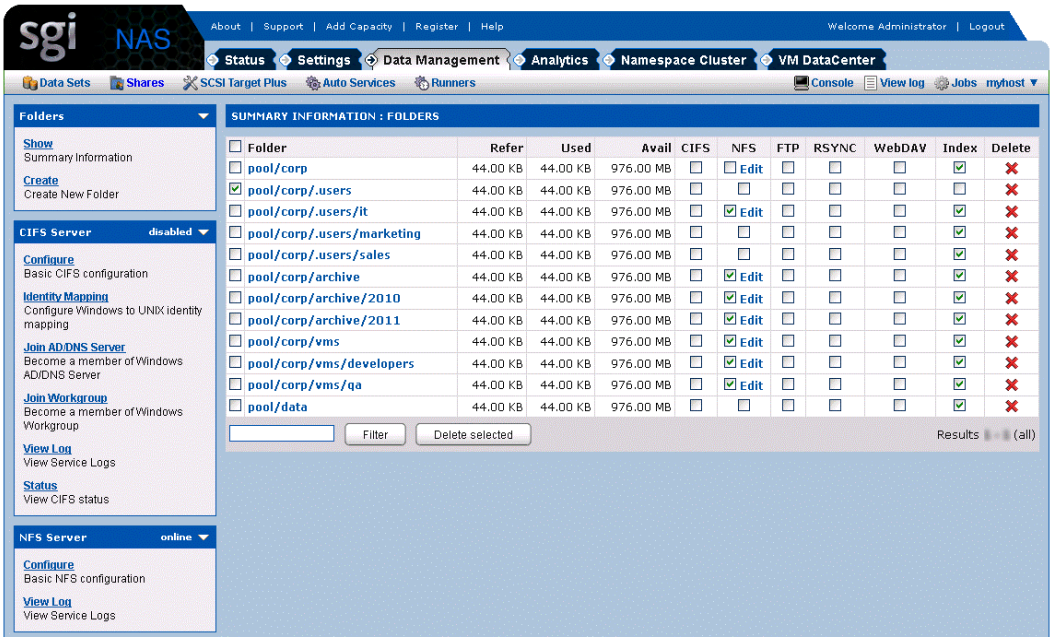
During the process of migration the share remains accessible for all the users.

After migration process is finished migrated share is located on the destination node. On the

place of share's original location appears Namespace Link. In our example it's NFS shared folder pool/corp/users:

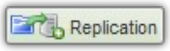


After migration is finished, you can go to **Data Management** → **Data Sets** → **Folder** → **Show** and delete folder pool/data/.users and all its subfolders to free space. This copy of shared folder is kept for safety reasons. In case migration procedure is not successful the data is safe. After successful migration procedure the copy of data can be left or removed according to administration needs. On the picture below the deletion of folder /pool/corp/.users is shown:

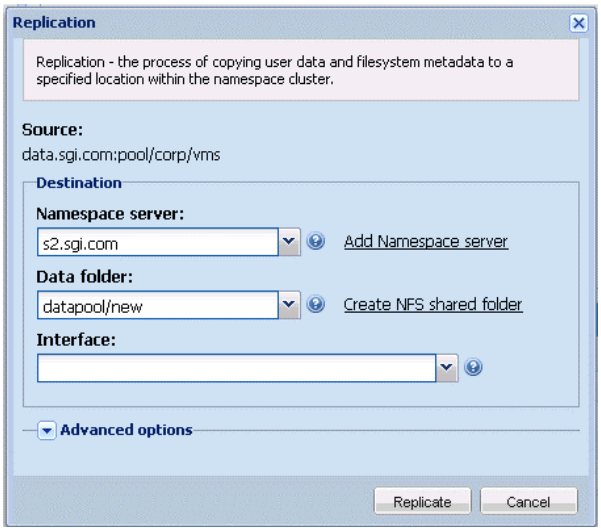


5.5 Replication

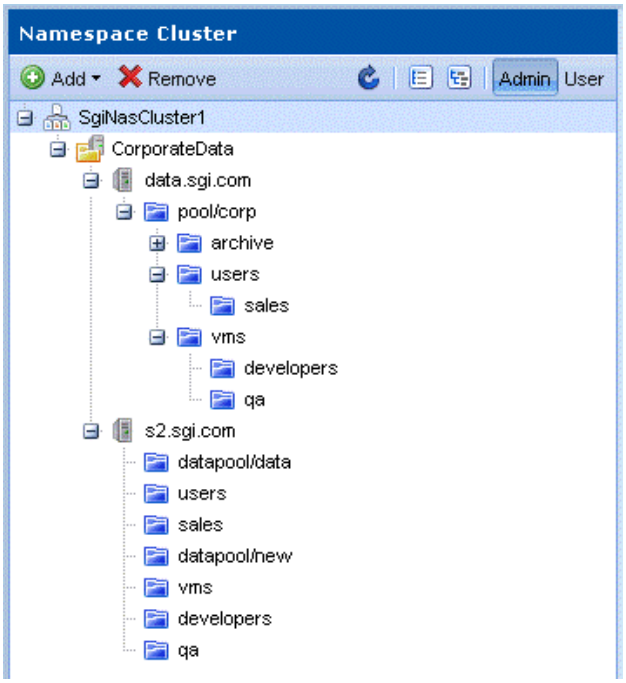
Replication procedure is a simple copying of data from one namespace server to another. No Namespace Link is created during this procedure.

Choose NFS shared folder and click 

button. Fill all the required fields on the replication screen and click 'Replicate'.



After successful replication the data appears on source and on destination:



6 Mounting the share via NFSv3.

The NFSv4.1 protocol provides a NFS referral mechanism that allows server to redirect the client to another server.

Client mounts a filesystem on server via NFS version 4.1. The server administrator can decide to move this exported filesystem to another server. The server must inform each client mounting filesystem on it that the data has moved.
The method used to communicate the migration event between client and server is the following: once the servers participating in the migration have completed the move of the filesystem, the error will be returned for subsequent requests received by the original server. Upon receiving the error, the client obtains the value of the specific attribute. The client then uses the contents of the attribute to redirect its requests to the specified server. This specific attribute is called NFS referral[A].

NFSv3 doesn't provide referral mechanism, but you can still mount the share and use the workaround below.

If your operating system doesn't support NFSv4.1 referrals, but has NFSv3 support and allows usage of 'amd' (automounter daemon), you can mount the share using NFSv3 protocol.

In this case SGI NAS will automatically transform referrals to symlinks to **/net/** directory, suitable for use by '**amd**'.

Centos 5.x:

Install *am-utils* using:

```
# yum install am-utils
```

No further configuration needed.

Debian 5.x/Ubuntu:

install *am-utils* using the following command

A http://nfsv4.bullopensource.org/doc/migration-and-replication-0.2.pdf

```
# apt-get install am-utils
```

No further configuration needed.

FreeBSD 8.x:

amd is shipped with the OS, to enable it add ***amd_enable='YES'*** to **/etc/rc.conf** and run

```
# /etc/rc.d/amd start
```

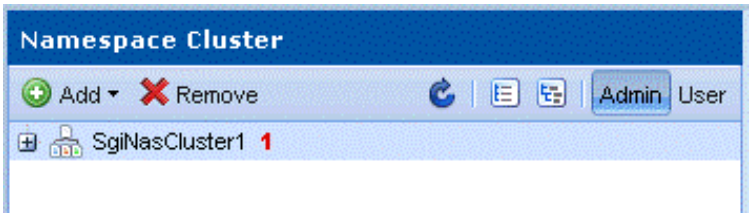
No further configuration needed.

Note, that all NFSv4.1 based clients are supported by Namespace Cluster out of the box (Linux kernel 2.6, VMware vSphere 5, etc.).

7 Fault management

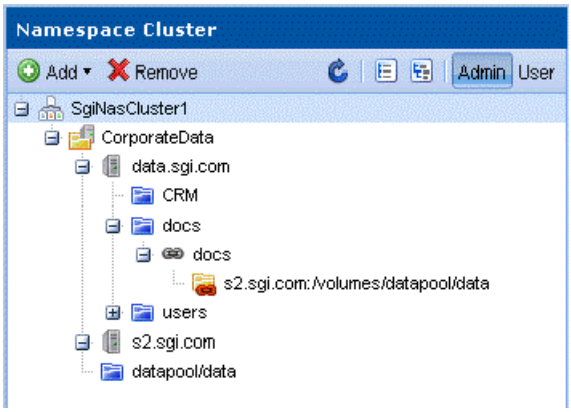
Namespace Cluster provides functionality to monitor namespace servers, and state of the shared filesystems (referenced data folder). The corresponding fault management is executed in a way consistent with the general appliance's fault management capability.

If a namespace server, referenced data folder or data server is faulted or unavailable, it appears in a namespace tree with a red alarm number, which specifies how many objects can be reached:

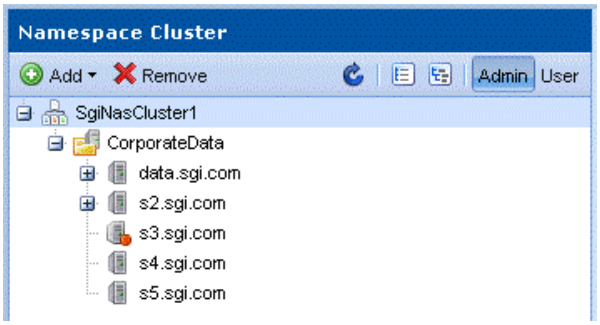


Expand the tree to see the details:

Observe how unavailable destination folder looks like :



Next screen shows unavailable server:



There are several node statuses:

success – a normal healthy state of a node.

warning – draws your attention to an event that may cause a failure or requires administration actions.

failure – namespace server is unreachable.

These statuses are the same for namespace links and folders.

7.1 Troubleshooting tips:

1. **Namespace server is unreachable.**

Ping the namespace server.

If it's not available:

- Check all the physical connections.
- Check if SGI NAS is up and running.

2. **Referenced data folder or namespace link is unreachable.**

- Check that corresponding disk is available.
- Check that corresponding NFS share wasn't forcefully unshared.

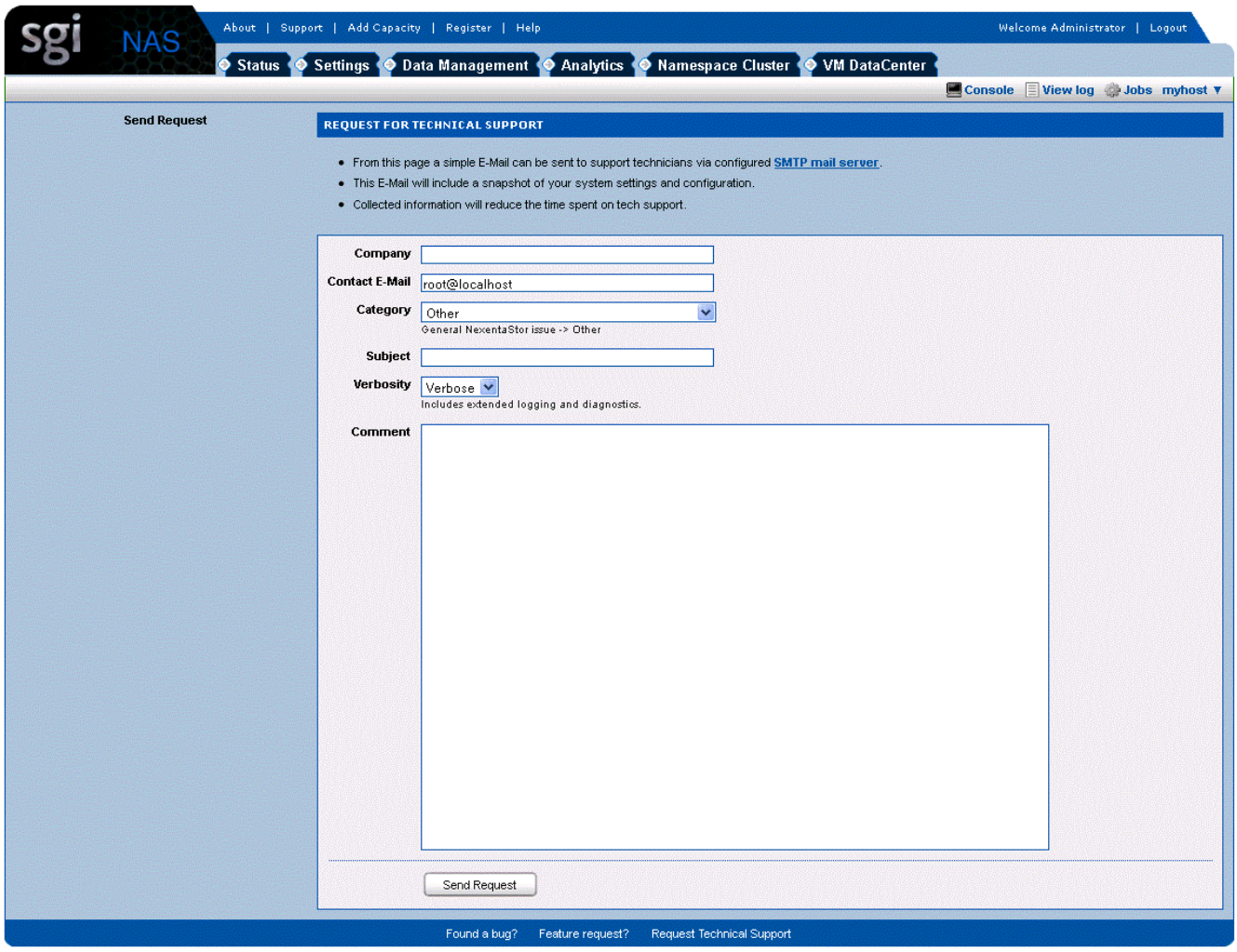
7.2 Known limitations

1. Namespace Cluster can't use CIFS protocol yet.

8 Contact information

8.1 Support request

To contact support at SGI, click on 'Support' in NMV screen shown below:



or type the following NMC command:

```
nmc:/$ support
```

which will then prompt for a subject and message.

8.2 Other resources

For licensing questions, please contact your SGI sales or support representative.