SUN BLUEPRINTS
APPROVED

Implementation Guide:
Sun™ Cluster 3.0 Series: Guide to Installation—Part 2

Chris Dotson, Enterprise Engineering

Sun BluePrints™ OnLine—May 2003

`http://www.sun.com/blueprints`

Sun Microsystems, Inc.
4150 Network Circle
Santa Clara, CA 95054 USA
650 960-1300
Part No.: 817-2016-10
Revision 1.0 3/25/03
Edition: May 2003

DOCUMENTATION IS PROVIDED "AS IS" AND ALL EXPRESS OR IMPLIED CONDITIONS, REPRESENTATIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE HELD TO BE LEGALLY INVALID.

Please
Recycle

Adobe PostScript

Module 2 Sun Cluster Intallation

Sun Cluster 3.0 Series: Guide to Installation—Part 2

This module reviews the Sun Cluster™ 3.0 concepts and components important to these specific procedures, as it describes the method of constructing a cluster. This is accomplished by installing the cluster software onto each node, then configuring the disks.

Note – Allow enough time to complete the cluster installation and configuration procedures. To build this Sun Cluster 3.0 two-node cluster, software can be installed using a variety of methods; however, the configuration procedures must be performed in the exact sequence shown. For this reason, the procedures included in later modules are intended to be performed consecutively, and following the prescribed sequence of operations.

Objectives

In this module, for each cluster node we will install and configure the Sun Cluster (SC) 3.0 software and apply the appropriate patches. As instructed during these exercises, you will confirm the correct configuration is achieved by performing the associated verification steps.

In this module, you will perform the following procedures in the prescribed sequence:

- Install the SC 3.0 software and patches on each cluster node in the specified sequence.
- Install and configure Solstice DiskSuite™ (SDS) 4.2.1 software for HA-NFS data service.
- Verify the auto-cluster formation and basic cluster operations
- Configure the SC 3.0 two-node quorum device

Prerequisites

This lab manual assumes you are a qualified Solaris™ network administrator. Additional pre-requisites include: ES-333, or equivalent. The intended audience for this module is sales engineers, system support engineers, professional service consultants, and system administrators. For installation queries, shell usage questions, patches and packages refer to the Sun Educational Services manuals for the “Solaris System Administration 1” and “Solaris System Administration 2” courses.

Major Resources

Hardware:

- 2x Sun Enterprise™ 220R servers (cluster node)
- 2x Sun StorEdge™ D1000 arrays [6 drives each] (shared storage)
- Netra™ T1 (admin workstation)
- Monitor (admin access)
- Annex Terminal Concentrator (console access)
- Ethernet hub (public network)
- 2x QFE I/O card (private interconnect)

Network: SWAN access

Software: SC 3.0 U3

Services: Sun Cluster

Introduction

This module describes the Sun Cluster software installation, and verifies the framework is configured correctly. This module also demonstrates key SC 3.0 features using the specific implementation to highlight Sun Cluster advantages. Throughout the procedure, key practices will be identified, and standard installation methods will be followed, ensuring each installation implements consistent practices, procedures, and documentation.

This BPLAB module requires a specific implementation of hardware and software. It is expected that the key practices and installation methods documented, herein, will be used for the majority of all cluster installations. For example, our configuration follows standard Sun Cluster installation practices, and implements best practices for administering a two-node cluster. SC 3.0 U3 will be installed, and a cluster will be formed, prior to configuring shared disks or data services.

Sun Cluster 3.0 Software

SC 3.0 is integrated with the Solaris operating environment (OE), and includes specialized software that implements highly available and scalable data services, and manages the Sun Cluster. SC 3.0 supports:

- Volume management software for administering shared data storage
- Software to enable all nodes to access all storage devices (even SC 3.0 nodes that are not directly connected to disks)
- Software to enable remote files to appear on every node as though they were locally attached to that node
- Software monitoring of user applications
- Software monitoring of cluster connectivity between nodes
- Software for configuring and creating highly available (HA) and scalable data services, including configuration files and management methods for starting, stopping, and monitoring both off-the-shelf, and custom applications using the application programmer interfaces (APIs)

Highly Available Data and Applications

The general design goal for the SunPlex™ platform is to reduce or eliminate system downtime due to a software or hardware failure, and to ensure data and applications are available even if an entire node (server) fails. SC 3.0 data services can be implemented to increase application performance (for example, scalable data services), and to provide enhanced availability of the system (for example, enabling maintenance of a node to occur without shutting down the entire cluster).

Failover and Scalable Services

SC 3.0 software provides high availability through application failover, which is the process by which a cluster automatically relocates a service from a failed primary node to a designated (that is, preconfigured) secondary node.

Scalable data services are designed to provide constant response times or throughputs, scaling to meet an increased workload. Each cluster node can process client requests and access shared data.

Topologies

SC 3.0 supports cluster-pairs and N+1 star topologies (or cluster interconnect schemes). Each supported topology requires careful consideration and planning to ensure failover (or scalability) can be achieved for each application hosted, and that systems, subsystems, and networks are configured to handle any additional workloads that can result. That is, alternate nodes have the performance and capacity to handle the increased workload.

Cluster Interconnect

The cluster interconnect is crucial to all cluster operations and should not be used to route any other traffic or data. This private network establishes exclusive use of preassigned (hard-coded) IP addresses.

The cluster interconnect supports cluster application data (shared nothing databases) and locking semantics (shared disk databases) between cluster nodes.

Redundant, fault-tolerant network links are implemented for SC 3.0, and all links are active at any given time. Adding links increases performance over the interconnect. Upon failure of a single link, failover is transparent and immediate. The application shall be the primary factor when determining actual performance (scaling) and which type of interconnects are supported.

Key Practice: Refer to Figure 1, and Table 1 through Table 5 of Module 1, (*Sun™ Cluster 3.0 Series: Guide to Installation—Part I*, Sun BluePrints OnLine, April, 2003), to configure additional private interconnects to increase interconnect performance and availability. For this configuration, `qfe2` and `qfe3` are shown as being 'unused.' Either of these connections (or both) can be configured, using Ethernet crossover cables between each cluster node. Sun Cluster software is then used to configure the additional private interconnects, before making the appropriate `/etc/inet/hosts` entries.

Cluster Membership Monitor

The Cluster Membership Monitor (CMM) is a distributed set of agents, one per cluster member. Agents exchange messages over the cluster interconnect to ensure valid cluster membership is preserved and maintained. The CMM drives the synchronized cluster reconfiguration process in response to changes in cluster membership, and handles cluster partitioning (for example, split-brain or amnesia). The CMM monitors all cluster members to ensure full connectivity.

Ultimately, the CMM ensures valid cluster membership and data integrity by maintaining a valid cluster quorum, and protects the cluster from partitioning itself into multiple, separate clusters in the event of cluster interconnect failure.

Failfast Mechanism

If the CMM detects a critical failure with a node, it calls upon the cluster framework to forcibly shut down (panic) the failing node and to remove it from the cluster membership.

Failfast will cause a node to shut down in two ways:

- If the node leaves the cluster and then attempts to start a new cluster without having a quorum, it will be "fenced" off from accessing shared data.
- If one or more cluster-specific daemons die (for example, `clexecd,` `rpc.pmfd`, `rgmd,` or `rpc.fed`) on a given node, that node will be forced to leave the cluster membership when CMM forces a panic.

Global Devices

SC 3.0 uses global devices to provide cluster-wide, highly available access to any device in the cluster.

- The cluster automatically assigns “globally” unique IDs to each disk, CD-ROM, and tape device within the cluster.
- This enables consistent access to each device from any node in the cluster.
- The global device name space is held in the `/dev/global` directory.

Disk Device Groups

For SC 3.0, all multihost disks are under the control of Sun Cluster software, for which you must perform the following:

- Create volume manager disk groups using the multihost disks.
- Register the volume manager disk groups as disk device groups (a type of global device).

SC 3.0 then registers every individual disk as a disk device group. After registration, the volume manager disk groups become accessible within the cluster. If more than one cluster node can read and write to a master disk device group, the data stored in that disk device group becomes highly available.

Note – Refer to the *Sun Cluster 3.0 Data Services Installation and Configuration Guide* for additional information about volume manager disk groups, and the association between disk device groups and resource groups.

Global Name Space

The SC 3.0 software mechanism that enables global devices is the global name space, which includes the `/dev/global` hierarchy and the volume manager namespaces. Normally, for SDS 4.2.1, the volume manager name spaces reside in the `/dev/md/diskset/[r]dsk` directories, and for Veritas Volume Manager (VxVM) 3.2, the name spaces normally reside in the `/dev/vx/[r]dsk` directories.

Within the Sun Cluster, each of the device nodes in the local volume manager name space are replaced by a symbolic link to a device node in the /global/.devices/node@nodeID file system, where nodeID is an integer value (for example, node@1 for clustnode1, and node@2 for clustnode2).

The global name space is automatically generated during the SC 3.0 installation procedure (scgdevs (1M)) and is updated during every reconfiguration reboot.

Cluster File Systems

A cluster file system is a proxy between the kernel and underlying file system on one cluster node, and the volume manager running on another cluster node that has been configured with a physical connection to the disks.

Programs can access a file in a cluster file system from any node in the cluster through the same file name (for example, /global/datafile1). Nodes do not require physical connection to the disks where the file is stored. A cluster file system is mounted on all cluster members, and can only see the underlying UFS file system.

Quorum and Quorum Devices

To ensure cluster and data integrity, it is important that a cluster never be allowed to split itself into separate, active partitions. The CMM guarantees that only one cluster is operational at a time and that cluster is able to access shared data. The majority number of votes (or “quorum”) is used to determine if an active partition will be allowed to form a cluster.

A quorum device is used to maintain a valid quorum vote count. A quorum device contributes to the vote count only if at least one of the cluster nodes to which it is currently attached is a valid cluster member. During cluster boot, a quorum device contributes to the vote count only if at least one of the nodes to which it is currently attached is booting and was a member of the most recently booted cluster at the time of shutdown.

A quorum device (for example, a dual-ported, shared disk device) is required for all two-node clusters since two or more quorum votes are required in order for a cluster to form. The two votes can come from the cluster nodes, or from just one node and a quorum device.

Key Practice: Protect against individual quorum device failures by configuring more than one quorum device between sets of cluster nodes. Use disks from different enclosures, and always configure an odd number of quorum devices between each set of nodes.

Cluster Reconfiguration

Cluster reconfiguration is performed to ensure a reliable and consistent cluster state. Each time a cluster is started (or when a node joins or leaves the cluster), the cluster framework performs a reconfiguration, which can be observed on the console.

Public Network Management (PNM)

The PNM monitors network interfaces and subnets of the system. Public network interfaces within an SC 3.0 configuration are assigned to a network adaptor failover (NAFO) group. Both primary and redundant interfaces are defined within the NAFO group.

Section 2.1: Install SC 3.0 On First Cluster Node—Without Reboot

Upon completion of this section, the first cluster node will be installed, but will NOT be rebooted. Choose “no” when asked about the automatic reboot option. After the SC 3.0 software is installed, you should observe numerous Sun Cluster alerts and messages.

For example, the console window displays numerous “NOTICE: CMM . .” messages, describing specific cluster events as they occur. Observe similar messages indicating the status of the cluster as it configures, such as the cluster “reconfiguration” number.

Cluster console events and messages, such as an “ERROR” or “FAIL” often require administrator intervention to verify the configuration is valid and operational.

Step 2.1.1

For local (manual) installations, insert the Sun Cluster 3.0 U3 CD into the `clustadm` workstation. Note that the `vold` (1M) daemon will automatically mount it using the `/cdrom` directory.

In order to make the contents of the CD-ROM available to the cluster nodes through the network, enter the following command into the `clustadm` workstation:

```
root@clustadm# share -F nfs -o ro,anon=0 /cdrom/cdrom0
```

Step 2.1.2

For local (manual) installations only, verify `/cdrom` has been shared correctly by entering the following command, on `clustadm`:

```
root@clustadm# share
-               /cdrom/ ro,anon=0   ""
root@clustadm#
```

Step 2.1.3

For local (manual) installations only, enter the following command on each cluster node:

```
# mkdir /cdrom
# mount –F nfs -o ro clustadm:/cdrom/suncluster_3_0 /cdrom
```

This example assumes the SC 3.0 admin workstation is `clustadm`, which can successfully share the contents of the CD-ROM drive, and is accessible from each cluster node as indicated.

Step 2.1.4

On the first cluster node only, execute `scinstall`, as indicated:

```
root@clustnode1#cd /cdrom/suncluster_3_0/SunCluster_3.0/Tools

root@clustnode1# ./scinstall
```

Caution – The next few steps install the first cluster node (`clustnode1`). When prompted at the last `scinstall` screen, enter **no** to `Automatic Re-Boot`. It is necessary to install the SC 3.0 patches first, prior to rebooting `clustnode1.`

- On the first cluster node only, begin the installation by choosing option 1 from the main menu to establish a new cluster.
- On the first cluster node only, continue to install the software packages. When prompted, continue to install software packages, including: `SUNWscr`, `SUNWscu`, `SUNWscdev`, `SUNWscgds`, `SUNWscman`, `SUNWscsal`, `SUNWscsam`, `SUNWscvm`, `SUNWscdm`, `SUNWscva`, `SUNWscvr`, and `SUNWscvw.`
- When prompted, continue the installation. Proceed to establish a new cluster named **`nhl`**, and continue the installation.
- When prompted by `Check`, run **`sccheck`** and verify it completes successfully.
- When prompted, enter **`clustnode2`** as the name of the other node planned for this cluster. Enter [**`Ctrl-D`**], to end the list. Then, proceed with the installation, as prompted.
- Do not enable DES Authentication.
- When prompted, enter **`y`** to accept the default network address, and enter **`y`** to accept the default netmask.
- Carefully read the next few prompts, observing each message. Enter **no** when asked if you are using transport junctions, and continue. When prompted from the list of available adapters, select `qfe0` as the first transport adapter. Next, select `qfe4` as the second transport adapter.
- When asked to create the directory name for the (local) global devices file system, enter **yes**, to use the default directory, `/globaldevices`.
- For the `Automatic Re-Boot` screen, enter **`no`** when asked, "`Do you want scinstall to reboot for you?`"

When prompted, press **Enter** to display the confirmation screen.

Step 2.1.5

Confirm the correct cluster information and installation packages are configured, before answering **yes** to each prompt, as indicated below:

```
 >>> Confirmation <<<

    Your responses indicate the following options to
scinstall:

scinstall -ik \

-C nhl \

-F \

-T node=clustnode1,node=clustnode2,authtype=sys \

-A trtype=dlpi,name=qfe0 -A trtype=dlpi,name=qfe4 \

-B type=direct



Are these the options you want to use [yes]? yes

Do you want to continue the install (yes/no) [yes]? yes
```

Verify the information is correct. The next few steps will complete this phase of the installation, where you will return to the `Main Menu`.

Step 2.1.6

Upon completion of the SC 3.0 software installation, if errors are encountered, note the pathname of the installation log file and determine the cause of any failures. For example: `/var/cluster/logs/install/scinstall.log.xxx`

From the `Main Menu`, enter **q** to quit `scinstall`.

Return to the `root@clustnode1#` shell prompt, where you will continue installing this first cluster node only.

Section 2.2: Identify the Device ID (DID) on the First Cluster Node

On the first cluster node, examine the `/etc/name_to_major` file, and record the DID number located at the end of the file. This value must be unique for the SunPlex platform. Furthermore, upon completion of the Sun Cluster software installation, the DID number configured in the `/etc/name_to_major` file should be identical on each cluster node.

Step 2.2.1

On the first cluster node, enter the following command to examine the `/etc/name_to_major` file. For later use, record the major device number entry during this installation for global device IDs (that is, DID `xxx`).

```
root@clustnode1# tail /etc/name_to_major

. . .

. . .

did 300
```

Note – Displaying the last few lines of this file should indicate the value for a global `did` of `xxx`. After the Sun Cluster software has been installed, each cluster node must have the identical value (for example, `did 300`). This file will be automatically modified, later, during VxVM installation.

Section 2.3: Verify DID is Available on Each Additional Cluster Node)

On each additional cluster node, examine the `/etc/name_to_major` file and verify that the DID number assigned by the first cluster node (`did 300`) is not already in use by another driver. For example, on `clustnode2` verify that `did 300` is not already in use.

Note – If these procedures have been followed, no conflict will occur. If the DID number is already in use on an additional cluster node, do not attempt to reconfigure the first cluster node at this time. Instead, you must reconfigure the conflicting driver on the additional cluster node to use a different major device number.

Step 2.3.1

On each additional cluster node, enter the following command to examine the `/etc/name_to_major` file. As indicated in the codebox below, verify there is no existing entry for `did 300`, on each additional cluster node:

```
root@clustnode2# grep 300 /etc/name_to_major

root@clustnode2#

root@clustnode2# grep did /etc/name_to_major

root@clustnode2#
```

Note – This file will be modified later during the VxVM installation.

Step 2.3.2

In preparation for upcoming patch installations, ensure the Sun Cluster 3.0 patches are accessible in single user mode, on the first cluster node. For local (manual) installations, patches can be made available by using the EIS CD-ROM.

For remote learning environments, make the SC 3.0 U3 patches available in single-user mode by copying all required patches to the local disk (c0t0), as in the example below. These will be removed later, after they have been applied.

On the first cluster node only, verify all required patches are successfully copied to the local disk. For example:

```
root@clustnode1# cd /cdrom/PATCHES/CLUSTER3.0U3

root@clustnode1# ls -lR

. . . {{output omitted}}. . .

root@clustnode1# mkdir -p /opt/PATCHES/CLUSTER3.0U3

root@clustnode1# cp -rp ./* /opt/PATCHES/CLUSTER3.0U3

root@clustnode1# ls -lR /opt/PATCHES

. . . {{output omitted}}. . .
```

Step 2.3.3

After veriyfing patches have been copied to local disk, reboot clustnode1 outside of the cluster before entering single-user mode. To do this, enter the following commands on the **first cluster node only**:

```
root@clustnode1# init 0

{{.... output omitted.....}}


ok boot -xs

{{.... output omitted.....}}
```

When prompted during the reboot, enter the root password to access the shell and perform system maintenance (single user mode).

The root password is: **abc**

Section 2.4: Install SC 3.0 Patches on the First Cluster Node

At this time, continue to install all required Sun Cluster 3.0 U3 patches on the first cluster node. This node should already be in single-user mode. In this section, after verifying patches have been applied, you will reboot the first cluster node, placing it back into multi-user mode.

Step 2.4.1

For local (manual) installations, obtain all required SC 3.0 U3 patches (Solaris 8 OE). We recommend accessing SunSolve^{SM} Online. To do this, go to `http://sunsolve.Sun.com` and click the Patches option on the left side column. Next, select `PatchPro`, followed by `SunCluster`, and specify all installed cluster components to generate a list of Sun Cluster Patches.

Note – SunSolve is a contract service from Sun Enterprise Services. It is a good idea to subscribe to this service, especially if you are running a production server.

Key Practice: Create a `/PATCHES` directory on a dedicated Management Server to store all patches. This enables centralized patch management. For example, the Sun BluePrints™ BPLAB hardware has been configured with a 'master' JumpStart server, which will serve all software binaries and patches, and act as the repository.

Key Practice: Refer to the individual patch `README` files to review any installation prerequisites before installing patches.

Step 2.4.2

At this time, recall that the SC 3.0 U3 patches were copied to the local `/opt` directory, and are accessible in single-user mode. Complete the SC 3.0 U3 patch installation on the first cluster node only. Enter the following, on `clustnode1`:

```
root@clustnode1# cd /opt/PATCHES/CLUSTER3.0U3

root@clustnode1# patchadd 110648-22

Checking installed patches......
```

Step 2.4.3

Verify the first patch installs, successfully. Then on the first cluster node, add the next patch:

```
root@clustnode1# patchadd 111554-09

Checking installed patches......
```

Step 2.4.4

Verify that all required patches install successfully. Some patches might already be installed, or might fail to install at this time.

Verify that any errors reported did not result in corrupt or missing packages required by the SC 3.0 software.

Review the list of installed patches by entering the following command into the first cluster node:

```
root@clustnode1# /usr/sbin/patchadd -p
```

Key Practice: Verify that all patches have been installed correctly by reviewing the patch installation logs, and resolve any installation errors or failures.

Step 2.4.5

After all required patches have been installed successfully, reboot the first cluster node:

```
root@clustnode1# init 6
```

The console window displays numerous messages indicating the status of the cluster.

During the reboot process, observe the console window for cluster messages appearing during the kernel and cluster initialization. Verify messages are displayed that indicate the first cluster node is “Booting as part of a cluster,” along with additional “NOTICE: CMM: ..” messages. At this time, you may ignore WARNING messages that instruct you to load the SunPlex Manager, or Apache software. These packages will be added in the future.

Section 2.5: Verify the Install Mode is Enabled

Step 2.5.1

The first cluster node is rebooted, after successful installation of Sun Cluster software and patches have been applied. In the example below, we verify that “Cluster install mode” is enabled on the first cluster node, prior to installing each additinal cluster node and establishing a cluster quorum. The first cluster node will be used to “sponsor” the additional nodes, as they are configured for operation within the cluster.

Note – For this two-node cluster, the “Cluster install mode” will be disabled and reset after a quorum device is configured and a valid cluster quorum is established.

On the first cluster node, enter the following:

```
root@clustnode1# scconf -p


Cluster name:                                      nhl
Cluster ID:                                        0x3D90B985
Cluster install mode:                              enabled
Cluster private net:                               172.16.0.0
Cluster private netmask:                           255.255.0.0
Cluster new node authentication:                   unix
Cluster new node list:                             clustnode1
clustnode2
Cluster nodes:                                     clustnode1

Cluster node name:                                 clustnode1
  Node ID:                                         1
  Node enabled:                                    yes
  Node private hostname:                           clusternode1-
priv
  Node quorum vote count:                          1
  Node reservation key:
0x3D90xxxxxxxxxxxx
  Node transport adapters:                         qfe0 qfe4

  Node transport adapter:                          qfe0
    Adapter enabled:                               no
    Adapter transport type:                        dlpi
    Adapter property:
device_name=qfe
    Adapter property:
device_instance=0
    Adapter property:
dlpi_heartbeat_timeout=10000

    Adapter property:
dlpi_heartbeat_quantum=1000
    Adapter property:
nw_bandwidth=80
    Adapter property:                              bandwidth=10
    Adapter port names:                            <NULL>

  Node transport adapter:                          qfe4
    Adapter enabled:                               no


{{..........output omitted.........}}
```

Section 2.6: Install SC 3.0 on Additional Cluster Nodes— Without Reboot

After the first cluster node has been installed successfully (and SC 3.0 patches have been applied), you may proceed to install the remaining cluster nodes.

Upon completion of this section, the additional cluster node(s) will be installed but will NOT be rebooted. Choose **no** when prompted to automatically reboot. For each additional cluster node, you will verify that SC 3.0 software has been installed successfully, before installing patches. Once this has been completed, each additional cluster node will be rebooted and will attempt to join the cluster.

Step 2.6.1

For local (manual) installations, insert the Sun Cluster 3.0 U3 CD into the `clustadm` workstation. Note that the `vold` (1M) daemon will automatically mount it in the `/cdrom` directory.

To make the contents of the CD-ROM available to the cluster nodes across the network, enter the following command into the `clustadm` workstation:

```
root@clustadm# share -F nfs -o ro,anon=0 /cdrom/cdrom0
```

Step 2.6.2

For local (manual) installations only, verify that `/cdrom` has been shared correctly by entering the following command on `clustadm`:

```
root@clustadm# share
-              /cdrom/ ro,anon=0   ""
root@clustadm#
```

Step 2.6.3

For local (manual) installations only, enter the following command into each cluster node:

```
# mount –F nfs -o ro clustadm:/cdrom/suncluster_3_0 /cdrom
```

This example assumes that the SC 3.0 administrative workstation is clustadm, which can successfully share the contents of the CD-ROM drive, and is accessible from each cluster node.

Step 2.6.4

At this time, continue installing the Sun Cluster software on each additional cluster node. Enter the following on clustnode2:

```
root@clustnode2# cd /cdrom/suncluster_3_0/SunCluster_3.0/Tools

root@clustnode2# ./scinstall
```

Step 2.6.5

Proceed with the SC 3.0 software installation on each additional cluster node.

Caution – On the last screen prompt, enter **no** to the Automatic Re-Boot option. It is necessary to install the SC 3.0 patches first, prior to rebooting clustnode2.

- On each additional cluster node, begin the installation by choosing option **2** from the Main Menu, and continue to add this machine to an established cluster. When prompted, continue the installation. Verify the following software packages are installed: SUNWscr, SUNWscu, SUNWscdev, SUNWscgds, SUNWscman, SUNWscsal, SUNWscsam, SUNWscvm, SUNWscdm, SUNWscva, SUNWscvr, and SUNWscvw.
- Continue the installation, and enter the name of the sponsoring node as **clustnode1**.
- Join the established cluster, named **nhl**.
- Continue the installation. When prompted, run **sccheck** and verify that it completes successfully.
- Enter **yes** to use autodiscovery, which will identify the cluster transport. Enter **yes** when each cluster interconnect, both qfe0 and qfe4, are “discovered” adding these connections to the configuration.
- When asked to create the name for the local global devices file system, enter **yes** to use the default /globaldevices.

- Enter **no** to the `Automatic Re-Boot` option.

When prompted, press **Enter** to continue.

Step 2.6.6

Confirm that the correct cluster information and installation packages are configured, before answering **yes** to each prompt, as indicated below:

```
 >>> Confirmation <<<

    Your responses indicate the following options to
scinstall:

scinstall -ik \

-C nhl \

-N clustnode1 \

-A trtype=dlpi,name=qfe0 -A trtype=dlpi,name=qfe4 \

-B type=direct

-m endpoint=:qfe0,endpoint=clustnode1:qfe0 \

-m endpoint=:qfe4,endpoint=clustnode1:qfe4



Are these the options you want to use [yes]? yes

Do you want to continue the install (yes/no) [yes]? yes
```

Note – After final confirmation of the installation parameters, `scinstall` proceeds to install the appropriate packages and prepare the node to become a member of the cluster. On the `clustnode2` console, verify the message `did driver major number set to xxx`, and verify the major number is correct, as indicated in previous steps (`clustnode1`, `did 300`).

On the `clustnode1` console, you will begin to observe numerous cluster messages, such as: “NOTICE: CMM: ..” and “WARNING:.. Path error ..” indicating `clustnode2` is attempting to reconfigure itself to become a cluster member.

When prompted, press Enter to continue, and return to the `Main Menu`.

Step 2.6.7

Upon completion of the SC 3.0 software installation, note the name of the installation log file, and identify any errors reported. For example: `/var/cluster/logs/install/scinstall.log.xxx`

Next, from the `Main Menu`, select **`q`** to quit `scinstall`.

This will return you to the `root@clustnode2#` shell prompt, where you will proceed to install the SC 3.0 patches.

Section 2.7: Install SC 3.0 Patches on Additional Cluster Nodes

In this section, you will install all the required Sun Cluster 3.0 U3 patches on each additional cluster node. Next, you will reboot each node, after all patches have been verified.

Step 2.7.1

For local (manual) installations, obtain all required SC 3.0 U3 patches (Solaris 8). We recommended accessing SunSolve Online to identify and obtain all required Sun Cluster patches.

Note – SunSolve is a contract service from Sun Enterprise Services. It is a good idea to subscribe to this service, especially if you are running a production server.

Key Practice: Create a /PATCHES directory on a dedicated Management Server to store all patches. This enables centralized patch management. For example, the Sun BluePrints™ BPLAB hardware has been configured with a ‘master’ JumpStart server, which will serve all software binaries and patches, and act as the repository.

Key Practice: Refer to the individual patch README files to review any installation prerequisites before installing patches.

Step 2.7.2

At this time, complete the patch installation by entering the following commands on `clustnode2`:

```
root@clustnode2# cd /cdrom/PATCHES/CLUSTER3.0U3

root@clustnode2# patchadd 110648-22

Checking installed patches......
```

Step 2.7.3

Verify the first patch installs successfully. Then add the next patch, on each additional cluster node:

```
root@clustnode2# patchadd 111554-09

Checking installed patches......
```

Step 2.7.4

Verify that all required patches are installed successfully. Several patches might already be installed, or might fail to install at this time.

Verify that any errors reported did not result in corrupt or missing packages required by SC 3.0 software. The following patches might not install: `111488-xx`, `111555-xx`, `112108-xx`, and `112866-xx`.

Review the list of installed patches by entering the following command into each additional cluster node:

```
root@clustnode2# /usr/sbin/patchadd -p
```

Key Practice: Verify that all patches have been installed correctly by reviewing the patch installation logs, and resolve any installation errors or failures.

Step 2.7.5

After all required patches have been installed successfully, reboot each additional cluster node:

```
root@clustnode2# init 6
```

Step 2.7.6

Wait for the reboot to finish, and the cluster to finish auto-formation. During the reboot process, observe the cluster console windows on each cluster node. Verify that cluster events are occurring on each cluster node, and that messages indicating auto-cluster formation are commencing.

For example, verify that `clustnode1` console "NOTICE:" messages appear, indicating state changes during cluster reconfiguration—for example, as each cluster interconnect path is verified and brought online.

Similarly, `clustnode2` console messages appear, indicating it is "Booting as part of a cluster." Additionally, note kernel and cluster initialization messages indicating "Configuring DID Devices" as all instances are being created, and the node obtains access to all attached disks.

Step 2.7.7

Wait for reboot to finish. Ensure that the cluster is stable, and all cluster interconnects and interfaces are "online" as per cluster console "NOTICE:" messages should indicate).

As `root`, execute the `/usr/cluster/bin/scstat` command and verify that each cluster node, and each interconnect path is `Online`.

```
# scstat
```

No errors, warnings, or degraded indicators should be indicated.

Prior to configuring the cluster quorum device, note the following example:

```
-- Cluster Nodes --

                    Node name           Status
                    ---------           ------
  Cluster node:     clustnode1 Online
  Cluster node:     clustnode2 Online

------------------------------------------------------------------

-- Cluster Transport Paths --

                    Endpoint            Endpoint            Status
                    --------            --------            ------
  Transport path:   clustnode1:qfe4 clustnode2:qfe4 Path online
  Transport path:   clustnode1:qfe0 clustnode2:qfe0 Path online
------------------------------------------------------------------
-- Quorum Summary --
  Quorum votes possible:      1
  Quorum votes needed:        1
  Quorum votes present:       1
-- Quorum Votes by Node --

                    Node Name           Present Possible Status
                    ---------           ------- -------- ------
  Node votes:       clustnode1          1        1       Online
  Node votes:       clustnode2          0        0       Online

-- Quorum Votes by Device --

                    Device Name         Present Possible Status
                    -----------         ------- -------- ------
```

Section 2.8: Establish the SC 3.0 Quorum Device - First Cluster Node Only

Step 2.8.1

On each cluster node, use the **`scdidadm`** command to verify the correct subsystem configuration has been established. Identify the DID number for the first shared disk which will be assigned as the quorum disk. The quorum devie will be physically located within shared storage. In our example, this should be **`d4`** and should correspond to **`/dev/rdsk/c1t0d0`** on each cluster node. To verify this, as `root`, enter:

```
# scdidadm -L
```

Note – Starting at Line 4 of the `scdidadm -L` output, notice the DIDs are displayed twice, once for each cluster node connection. Both nodes must share a connection to a quorum device, and is required for this two-node cluster. Verify that both nodes connect to the same physical device (for example, **`/dev/rdsk/c1t0d0)`** and have the same DID assignment (for example **`/dev/did/rdsk/d4`**). The DID assignments for all global devices should follow this example, and be identical on both cluster nodes.

Verify the following output, noting that lines 4 and 5 refer to the same physical disk or spindle:

```
1 clustnode1:/dev/rdsk/c0t0d0 /dev/did/rdsk/d1
2 clustnode1:/dev/rdsk/c0t1d0 /dev/did/rdsk/d2
3 clustnode1:/dev/rdsk/c0t6d0 /dev/did/rdsk/d3
4 clustnode1:/dev/rdsk/c1t0d0 /dev/did/rdsk/d4
4 clustnode2:/dev/rdsk/c1t0d0 /dev/did/rdsk/d4
5 clustnode1:/dev/rdsk/c1t1d0 /dev/did/rdsk/d5
5 clustnode2:/dev/rdsk/c1t1d0 /dev/did/rdsk/d5
6 clustnode1:/dev/rdsk/c1t2d0 /dev/did/rdsk/d6
6 clustnode2:/dev/rdsk/c1t2d0 /dev/did/rdsk/d6
7 clustnode1:/dev/rdsk/c1t8d0 /dev/did/rdsk/d7
7 clustnode2:/dev/rdsk/c1t8d0 /dev/did/rdsk/d7
8 clustnode1:/dev/rdsk/c1t9d0 /dev/did/rdsk/d8
8 clustnode2:/dev/rdsk/c1t9d0 /dev/did/rdsk/d8
9 clustnode1:/dev/rdsk/c1t10d0 /dev/did/rdsk/d9
9 clustnode2:/dev/rdsk/c1t10d0 /dev/did/rdsk/d9
10 clustnode1:/dev/rdsk/c2t0d0 /dev/did/rdsk/d10
10 clustnode2:/dev/rdsk/c2t0d0 /dev/did/rdsk/d10
```

```
11 clustnode1:/dev/rdsk/c2t1d0 /dev/did/rdsk/d11

11 clustnode2:/dev/rdsk/c2t1d0 /dev/did/rdsk/d11

12 clustnode1:/dev/rdsk/c2t2d0 /dev/did/rdsk/d12

12 clustnode2:/dev/rdsk/c2t2d0 /dev/did/rdsk/d12

13 clustnode1:/dev/rdsk/c2t8d0 /dev/did/rdsk/d13

13 clustnode2:/dev/rdsk/c2t8d0 /dev/did/rdsk/d13

14 clustnode1:/dev/rdsk/c2t9d0 /dev/did/rdsk/d14

14 clustnode2:/dev/rdsk/c2t9d0 /dev/did/rdsk/d14

15 clustnode1:/dev/rdsk/c2t10d0 /dev/did/rdsk/d15

15 clustnode2:/dev/rdsk/c2t10d0 /dev/did/rdsk/d15

16 clustnode2:/dev/rdsk/c0t0d0 /dev/did/rdsk/d16

17 clustnode2:/dev/rdsk/c0t1d0 /dev/did/rdsk/d17

18 clustnode2:/dev/rdsk/c0t6d0 /dev/did/rdsk/d18
```

Note – Verify local disk and CD-ROM devices are reported correctly in the output (for example, **d1**, **d2** and **d3** represent internal `clustnode1` devices, while **d16**, **d17**, and **d18** represent internal `clustnode2` devices).

Key Practice: Label devices and cables. Devices and cables that are easily identified make troubleshooting and maintenance easier. This is helpful even for systems that do not require high availability. Where high availability is a prerequisite, cable and device identification and labelling should be high priority. Verify disks (boot devices, mirrors, disk quorums, volume manager conifiguration databases, clones, hot-spares, and so on), CD-ROM drives, tape drives, and cable labels are correct. This enables service operations to easily correlate error messages, which may specify global DID numbers, metadevice names, or controller number and `sd/ssd` instances. Proper labelling can help when interpreting errors and to determine specific failed devices such as specific disk spindles and related components. Tape drives should be labelled with their `rmt` instance numbers.

Step 2.8.2

Prepare to establish a quorum device using **d4** as the quorum disk. Enter the following command and answer each prompt as indicated, on the first cluster node only:

```
root@clustnode1# scsetup

>>>> Initial Cluster Setup <<<<

{. . . . . output omitted. . . . . }

Is it okay to continue? yes

Do you want to add any quorum disks? yes

. . . {{output omitted}} . . .

Which global device do you want to use (d<N>)? d4

Is it okay to proceed with the update? yes
```

Observe cluster node console messages indicating cluster reconfiguration (reconfiguration #4) has completed.

Step 2.8.3

On the first cluster node press Enter when prompted.

Then, `scsetup` will prompt to add additional quorum disks. Enter **no**, as indicated:

```
{. . . . output omitted . . . }

Do you want to add another quorum disk (yes/no)? no
```

Step 2.8.4

Finally, reset the cluster “install mode,” as indicated on the first cluster node:

```
{. . . . output omitted . . . }

Is it okay to reset install mode? yes
```

Step 2.8.5

Verify the cluster initialization completes and that each cluster node reports console messages and cluster events, as they occur.

On the first cluster node, press Enter when prompted, to proceed to the main menu. Next, enter **`q`** to quit `scsetup`.

Note – Verify the cluster advances from the “install mode” state to operational status. Observe the `scconf -p` output, and verify that the **`install mode`** is ‘**`disabled`**’.

Step 2.8.6

Verify that the quorum device has been configured correctly as indicated by the command, **`scstat`**:

```
# scstat

-- Cluster Nodes --

                    Node name           Status
                    ---------           ------
  Cluster node:     clustnode1          Online
  Cluster node:     clustnode2          Online


-- Cluster Transport Paths --

                  Endpoint          Endpoint          Status
                  --------          --------          ------
  Transport path:  clustnode1:qfe4    clustnode2:qfe4    Path
online
  Transport path:  clustnode1:qfe0    clustnode2:qfe0    Path
online


-- Quorum Summary --

  Quorum votes possible:      3
  Quorum votes needed:        2
  Quorum votes present:       3


-- Quorum Votes by Node --

                   Node Name           Present Possible Status
                   ---------           ------- -------- ------
  Node votes:       clustnode1          1        1      Online
  Node votes:       clustnode2          1        1      Online


-- Quorum Votes by Device --

                   Device Name         Present Possible Status
                   -----------         ------- -------- ------
  Device votes:     /dev/did/rdsk/d4s2  1        1      Online


{ . . . output omitted . . . }
```

Section 2.9: Configure Additional Public Network Adapters - NAFO

Each cluster node should be connected to more than one subnet by configuring additional public network adapters for the secondary subnets.

In this section, configure NAFO software for each public network adapter, on each cluster node. This is done in preparation for creating the Resource Group.

Note – Refer to Figure 1, and Table 1 through Table 5 of Module 1, (*Sun™ Cluster 3.0 Series: Guide to Installation—Part I*, Sun BluePrints OnLine, April, 2003) for networking and cabling connections implemented in this configuration exercise.

NAFO configuration guidelines:

- Always maintain at least one public network adapter connection for each node in the cluster (SunPlex platform). A cluster node is inaccessible without this connection.
- Configure all public network adapters to belong to a NAFO group.
- Avoid unconfiguring (unplumbing) or bringing down the active adapter of a NAFO group without first switching over the active adapter to a backup adapter in the group.
- For any given node, there can be *at most* one NAFO group on a given subnet.
- All adapters in a given NAFO group must be connected to the same subnet.
- Only one adapter in a given NAFO group can have a host name associate, that is, an `/etc/hostname.`*`adapter`* file.
- A public network adapter can belong to only one NAFO group.

Note – Configure only public network adapters. Do not configure private network adapters.

Step 2.9.1

On each cluster node, use `vi` to create the **`/etc/hostname.hme0`** file, as indicated in the following codebox. After creating the file entry, verify `clustnode1` is configured correctly:

```
root@clustnode1# more /etc/hostname.hme0

clustnode1
```

Step 2.9.2

After creating the file entry, verify `clustnode2` is configured correctly by entering the following:

```
root@clustnode2# more /etc/hostname.hme0

clustnode2
```

Note – If a naming service is configured, this information should be in the naming service database.

Step 2.9.3

On each cluster node, configure a NAFO group using the available adapters, by entering the following on `clustnode1`:

```
# pnmstat -l

{{verify no existing nafo0 group exists, before proceeding}}

# pnmset -c nafo0 -o create hme0 qfe1

# pnmstat -l

group adapters status fo_time act_adp
nafo0 hme0:qfe1 OK NEVER hme0
```

Note – For local (manual) installations, verify additional failover capabilities on each cluster node. For example, verify that a public network failover occurs successfully between `hme0` and `qfe1`. That is, physically disconnect the public network hme0 cable connection, and manually verify each node is still "accessible" over the same public subnet, qfe1.

Section 2.10: Configure ntp.conf on Each Cluster Node

On each node of the cluster, the /etc/inet/ntp.conf file must be updated to reflect the actual cluster configuration. This helps ensure cluster nodes recover (that is, reboot and reconfigure) more rapidly, and does not attempt to establish a running cluster using nodes that do not exist (because cluster nodes 3 through 8 are NOT configured).

- Remove all entries for private host names that are not actively used in the cluster.
- Private host names are normally modified and configured using scsetup.
- An nsswitch.conf facility performs all lookups for private hostnames.

Note – Cluster nodes are not to be configured as an NTP server. Enter the designated NTP server in the /etc/inet/ntp.conf file. Also, if you have changed the private host names of the cluster nodes, update this file accordingly to reflect the customer-specific private hostnames. For additional information, refer to the *Network Time Protocol User's Guide*, and the ntpdate(1M) man pages, for invoking the NTP from within cron scripts.

Step 2.10.1

On each cluster node, create and configure the `/etc/inet/ntp.conf` file. First, copy `/etc/inet/ntp.conf.cluster` to `/etc/inet/ntp.conf.` Next, edit `/etc/inet/ntp.conf` to comment out the host name entries for any cluster nodes that are not actually configured in the cluster. Verify the entries, as indicated:

```
# more /etc/inet/ntp.conf

...{{output omitted}}
...
peer clusternode1-priv prefer
peer clusternode2-priv
#peer clusternode3-priv
#peer clusternode4-priv
#peer clusternode5-priv
#peer clusternode6-priv
#peer clusternode7-priv
#peer clusternode8-priv


. . . {{output omitted}}
```

Note – For this two-node configuration, only the first two cluster node entries are required.

Section 2.11: Verify `/etc/nsswitch` Entries

Step 2.11.1

Reverify the `/etc/nsswitch.conf` file, at this time. Ensure all site-specific settings are correct. Verify the file includes the following modifications as indicated below, on each cluster node:

```
# more /etc/nsswitch.conf
. . . {{output omitted}}. . .

group: files
hosts: cluster files [SUCCESS=return]
services: files
netmasks: cluster files
```

Key Practice: The cluster environment requires that local (`/etc`) files supporting network services are searched ahead of any naming services. This increases availability by not having to rely on an outside agent. For `hosts`, put `cluster files [SUCCESS=return]`, ahead of `dns` ,`nis`, etc. For `netmasks,` place `cluster files`, ahead of `dns`, `nis`, and so on.

Section 2.12: Update Private Interconnect Addresses on All Cluster Nodes

On each cluster node, confirm the IP Addresses assigned to the private intereconnects, and update the hosts file to reflect the actual configuration.

Step 2.12.1

On each cluster node, identify and record the IP address assignments for each private interface, as indicated:

```
# grep ip_address /etc/cluster/ccr/infrastructure

cluster.nodes.1.adapters.1.properties.ip_address xxx.xx.x.xxx
cluster.nodes.1.adapters.2.properties.ip_address xxx.xx.x.x
cluster.nodes.2.adapters.1.properties.ip_address xxx.xx.x.xxx
cluster.nodes.2.adapters.2.properties.ip_address xxx.xx.x.x
```

Step 2.12.2

Update `/etc/inet/hosts` to reflect the IP addresses assigned for each of the private interconnects, on each cluster node, as indicated in the codebox, below.

Next, verify that each cluster node is configured, as indicated in the following example (clustnode1):

```
root@clustnode1# more /etc/inet/hosts

. . . {{output omitted}}. . .

xxx.xxx.xx.xxx  clustnode1 loghost clustnode1.some.com
xxx.xxx.xx.xxx  clustadm
xxx.xxx.xx.xxx  clustnode2
xxx.xxx.xx.xxx  tc nhl-tc
xxx.xxx.xx.xxx  lh-hanfs

{{verify the private interconnect addresses are correct}}

xxx.xx.x.xxx clustnode1-priv-a
xxx.xx.x.x clustnode1-priv-b
xxx.xx.x.xxx clustnode2-priv-a
xxx.xx.x.x clustnode2-priv-b
```

Note – The entries must be unique, and must not conflict with any other host names in the domain.

Step 2.12.3

For local (manual) procedures, delete the patches that were temporarily copied to the first cluster node (clustnode1), under /opt/PATCHES/CLUSTER3.0U3. These patches are no longer required.

Section 2.13: Add Diagnostic Toolkit

Step 2.13.1

For local (manual) procedures, install the Diagnostic ToolKit (SUNWscdtk) package.

Caution – At this point in the installation, be careful when running diagnostics. DO NOT choose any tests that can destroy data already stored on disk, unless you intend to reinstall the systems.

End of Module 2

Module 2 is now complete. You have successfully performed the following procedures:

- Installed the cluster software
- Installed the patches for the cluster
- Configured the quorum device(s)
- Reset the cluster installation mode to an operational status
- Configured NAFO software for public network interfaces
- Updated the `/etc/nsswitch.conf` file to reflect the actual cluster configuration
- Updated the `/etc/hosts` file to reflect the IP addresses chosen for the private interconnects
- Installed the Diagnostic Toolkit package