sgi

SGI® Management Center Quick Start
Guide

Record of Revision

Version	Description
001	July 2010 Original publication

Contents

About This Guide

SGI Management Center software running on the master host provides a robust graphical interface for system configuration, operation, and monitoring.

This guide tells you how to get your SGI Management Center up and running:

- Chapter 1, "Networking Setup" on page 1

- Chapter 2, "Component Media Kit - Manufacturing Image" on page 5

- Chapter 3, "Rebuilding the Master Host After Failure" on page 19

Your SGI Management Center cluster may run one of the following operating systems:

- Red Hat Enterprise Linux (RHEL)

- SUSE Linux Enterprise Server (SLES)

Professional assistance is available through SGI Managed Services. If you ordered such service, please contact the SGI Customer Support Center to schedule the on-site visit.

Related Publications

For more information, see:

- *SGI Management Center Installation and Configuration*

- *SGI Management Center System Administrator's Guide*

- *SGI Management Center - Application Portal Install Guide*

Obtaining Publications

See the SGI Technical Publications Library at http://docs.sgi.com. Various formats are available. This library contains the most recent and most comprehensive set of online books, man pages, and other information.

Networking Setup

To use SGI Management Center, you must modify the network settings with site-specific information, while taking care not to disturb the required master host interfaces that were installed at the factory.

Required Interfaces for the Master Host

Note: Do not make any modifications to the interfaces described in this section except as directed in "Networking Changes You Must Make" on page 2.

The master host for your SGI Management Center has the following entries in `/etc/hosts` that were configured at the factory as part of the cluster integration process:

- An interface with the name `hostname.customer.domain`:

  ```
  0.0.0.0                 hostname.customer.domain
  ```

  You will change this to use your site-specific the IP address and hostname information as directed below in "Networking Changes You Must Make" on page 2.

- Default administration domain (required to match `/opt/sgi/sgimc/@genesis.profile` entries for *host*:*host* and *system.rna.host*:*host*). You must not change the host value:

  ```
  10.0.10.1       admin.default.domain     admin host loghost
  ```

- *(Optional)* InfiniBand:

  ```
  192.168.20.1    admin-ib.default.domain  admin-ib host-ib loghost-ib
  ```

- *(Optional)* A second Ethernet interface:

  ```
  172.16.10.1 admin-ge.default.domain admin-ge host-ge loghost-ge
  ```

Networking Changes You Must Make

⚠️ **Caution:** For SUSE Linux Enterprise Server (SLES) systems, you must first back up the /etc/hosts file if you want to use YaST to manipulate networking (because YaST will change all of the interfaces to match the external host name).

Using a text editor, make the following changes:

- Modify /etc/hosts to change the 0.0.0.0 entry to use your site-specific IP address and hostname information for the master host.

  For example, if the hostname.customer.domain value for your master host should be acmesgimc.acme.com and the IP address you want to use is 128.162.232.79, you would change from this:

  ```
  0.0.0.0              hostname.customer.domain
  ```

  to this:

  ```
  128.162.232.79     acmesgimc.acme.com
  ```

  You may add any other site-specific entries.

  You should make the same changes to /compute/etc/hosts and then check in the payload.

- Modify /etc/resolv.conf for site-specific search, domain, and nameserver entries, as applicable. If you use DNS, you will also later enter one of the nameserver values into SGI Management Center using the following menu:

  Edit
  > **Preferences**
  > **General Network**
  > **Domain Name Server**

  If DNS is not to be used, you must remove the existing nameservervalue.

- Modify the hostname value according to your operating system:

  - RHEL: set the HOSTNAME value in /etc/sysconfig/network

  - SLES: set the hostname in /etc/HOSTNAME

- Modify the required files to establish the gateway:
  - RHEL: set the GATEWAY field in
    /etc/sysconfig/network-scripts/ifcfg-eth0
  - SLES : set the default gateway in /etc/sysconfig/network/routes to the
    default gateway. For example, if the gateway is 137.38.141.254 then
    /etc/sysconfig/network/routes would contain the following:

    ```
    224.0.0.0   0.0.0.0   240.0.0.0   eth1   multicast
    default 137.38.141.254 - -
    ```

    Note: Preserve the multicast route.

Component Media Kit - Manufacturing Image

This chapter discusses the following:

- "Introduction to CMK-MI" on page 5

- "Application and Service Details" on page 8

- "Configuration Tools" on page 15

Introduction to CMK-MI

The *Component Media Kit - Manufacturing Image* (CMK-MI) is a preconfigured and validated high-performance computing (HPC) archive of a fully functional cluster. The archive contains all of the necessary packages, configuration management, drivers, applications, and functionality to enable quick integration and an overall reduced total cost of ownership. SGI Management Center (SMC) is the key software technology for easily installing, managing, and maintaining this HPC environment.

This section discusses the following:

- "Supported Hardware and Software" on page 6

- "Prerequisites" on page 6

- "Integrated Software" on page 7

- "ISO Media Kits" on page 7

- "User Account Information" on page 7

Supported Hardware and Software

CMK-MI supports the following:

- Hardware platform:

  - SGI Altix XE:

    - e1000 NIC: XE210, XE240, XE310, XE320

    - igb NIC: XE250, XE270, XE340, XE500

  - SGI Octane III

- SUSE Linux Enterprise Server 11 Service Pack 1 (SLES 11 SP1) operating system

- SGI Management Center

- SGI Management Center Application Portal

Prerequisites

CMK-MI has the following prerequisites:

- SGI Management Center requires an SGI LK license key

- The HPC Linux cluster must be configured properly per SGI Manufacturing's process documentation

- Master host cabling configuration:

  - `eth0` = External LAN

  - `eth1` = Management network

  - `ib0` = InfiniBand network (optional)

- Compute node cabling configuration:

  - `eth0` = Management network

  - `ib0` = InfiniBand network (optional)

- The Ethernet switch must support multicast Internet Group Management Protocol (IGMP) packets

Integrated Software

CMK-MI contains the following integrated and functional software:

- PBS Pro

- SLES 11 SP1 Native OFED

- Intel Cluster Checker

- Intel Cluster Runtime

- Java

- Java Runtime Environment (JRE)

- SGI LK

A large number of applications and services are preconfigured and validated within the CMK-MI archive. For more information, see "Application and Service Details" on page 8.

ISO Media Kits

Via YUM package management, CMK-MI supports ISO media kits for the following:

- SLES 11 SP1

- SGI Foundation Software

- SGI ProPack

- SGI Management Center

- SGI Management Center Application Portal

For details about the product offering, see the `/etc/sgi` file.

User Account Information

Table 2-1 shows the user logins provided in CMK-MI for SLES and SGI Management Center (SMC) users.

Table 2-1 CMK-MI User Accounts

Type	User	Password
SLES	root	sgisgi
SLES	sgi	sgisgi
SMC	root	root

Application and Service Details

CMK-MI has the following applications and services:

- "NAT" on page 9

- "NTP" on page 9

- "IGMP Routing" on page 9

- "Syslog Messages" on page 9

- "Modules" on page 10

- "Intel MPI Runtime Libraries" on page 10

- "GCC Compiler Suite" on page 10

- "SSH" on page 10

- "RSH" on page 10

- "PDSH" on page 10

- "PBS Pro" on page 11

- "LDAP" on page 11

- "NFS" on page 11

- "NIS" on page 12

- "AutoFS" on page 12

- "ConMan" on page 12

- "PowerMan" on page 13

- "Intel Cluster Checker" on page 13

- "YUM" on page 14

NAT

Network Address Translation (NAT) is implemented on the master host in `/etc/init.d/boot.local` with the following `iptables` rule:

```
/usr/sbin/iptables -t nat -A POSTROUTING -o eth0 -j MASQUERADE
```

NTP

Network Time Protocol (NTP) is implemented with compute nodes synchronized to the master host. The compute nodes use `broadcastclient` to find the server. The master host restricts which clients it will serve to `10.0.0.0/16` addresses.

IGMP Routing

An Internet Group Management Protocol (IGMP) route is established on the master host on the cluster private network (`eth1`). The `/etc/sysconfig/network/routes` file contains the following:

```
239.192.0.0    0.0.0.0    255.255.0.0    eth1    multicast
```

Syslog Messages

Syslog messages for compute nodes are forwarded to the master host. The `/etc/syslog-ng/syslog-ng.conf` file in the payload contains the following:

```
destination loghost { tcp("10.0.10.1" port(5140)); };
log { source(src); destination(loghost); };
```

The master host configuration allows logs in `/etc/syslog-ng/syslog-ng.conf`. The source `src {}` section contains the following:

```
udp(port(514));
tcp(ip("10.0.10.1") port(5140) keep-alive(yes));
```

Modules

The following modules are added to `/usr/share/modules/modulefiles` for the master host and the compute nodes:

```
compiler/gcc-4.3.3
mpi/intel-4.0
```

Intel MPI Runtime Libraries

Intel Message Passing Interface (MPI) and compiler tools libraries are installed in `/opt/intel` in the payload.

GCC Compiler Suite

GCC Compiler Suite is installed in the payload and on the master host.

SSH

Passwordless secure shell (`ssh`) is configured using the `sgi_node-config.py` tool. See "`sgi_node-config.py`" on page 15.

RSH

The remote shell (`rsh`) is configured on and compute nodes are added to `/etc/hosts.equiv` in the payload.

PDSH

Parallel distributed shell (`pdsh`) is configured with all compute nodes in `/etc/machines`.

PBS Pro

PBS Pro is configured with a short-term demonstration license. All the compute nodes run `pbs-mom` while the master host is both server and scheduler. A single queue is created. The `sgi` user account (`/home/sgi`) has a sample PBS script (`SGI_MPITEST.pbs`)

LDAP

A basic Lightweight Directory Access Protocol (LDAP) database is provided for the Application Portal. Due to issues with YaST, transport layer security (TLS) is disabled in the SLES 11 host clone image. An issue can occur if TLS is enabled when the host clone is created. To enable TLS again for LDAP, do the following:

1. Open `yast2` as `root`.

2. Open **LDAP Server** in the **Network Services** group.

3. Expand **Global Settings**.

4. Select **TLS Settings** (on the left) and put a checkmark in **Enable TLS** (on the right).

5. Click **OK**.

NFS

The master host exports `/home` to the world with the following options:

`rw,sync,no_root_squash,no_subtree_check`

The master host exports `/opt/intel` to the world with the following options:

`ro,sync,no_root_squash,subtree_check`

The compute nodes mount `/opt/intel` from the master host using `host:/opt/intel`.

NIS

Network Information Service (NIS) is configured with the master host as the NIS server restricting its service to `10.0.0.0/16`. The domain is `default.lan`. The maps are as follows:

```
auto.master
group.bygid
group.byname
passwd.byname
passwd.byuid
rpc.byname
rpc.bynumber
ypservers
```

The payload has `ypbind` on with `ypserver` set to the master host.

AutoFS

AutoFS is configured on in the payload including NIS-served `auto.master` (`+auto.master`) and `auto.home`, which mounts `/home` from `host:/home`.

ConMan

SGI Management Center (SMC) is configured to use the ConMan console manager (`conman`). All compute node consoles log to the following location:

`/var/log/conman/console.`*nodename*

The cluster is configured to match the node type.

Note: If new nodes are added, some IPMI configuration may be needed.

Depending on your node type, different IPMI parameters must be passed to the BMC. For example, an Altix XE 210 requires `-o intelplus`.

To accommodate this functionality within SMC, do the following:

1. Modify the `ipmitool_options.profile` file:

   # **vi /opt/sgi/sgimc/etc/ipmitool_options.profile**

2. Modify the following lines if required:

```
ipmitool.power._default_=-I lanplus
ipmitool.status._default_=-I lanplus
ipmitool.sol._default_=-I lanplus
```

For example, for an Altix XE 210, you would change the lines as follows:

```
ipmitool.power._default_=-I lanplus -o intelplus
ipmitool.status._default_=-I lanplus -o intelplus
ipmitool.sol._default_=-I lanplus -o intelplus
```

3. Restart the SMC daemon if you made a modification in step 2:

/etc/init.d/mgr restart

4. Modify the conman expect script as well, if it requires additional parameters:

vi /usr/lib/conman/exec/ipmitool.exp

5. Modify the following line if required:

```
set cmd "ipmitool -e \& -I lanplus -H $host -U $user -a"
```

For example, for an Altix XE 210, you would change the line to:

```
set cmd "ipmitool -e \& -I lanplus -o intelplus -H $host -U $user -a"
```

6. Modify the file /etc/freeipmi if required.

Some hardware may require workarounds such as the following to be present:

```
workaround-flags intel20
ipmiconsole-workaround-flags solpayloadsize
```

PowerMan

PowerMan is configured for all compute nodes in the cluster.

Intel Cluster Checker

Intel Cluster Checker is run across the cluster. The input XML file is
/home/sgi/cluster-check/clck-config-user.xml. The
/opt/intel/clck/1.5/nodelist file lists the compute nodes.

YUM

There is a local YUM repository configured for the installation of packages from the following sources:

- SLES 11 SP1

- SLES 11 SP1 SDK

- SGI Foundation Software

- SGI ProPack

The SLES 11 DVD ISO and the SDK are mounted automatically in `/etc/init.d/boot.local`. If you would like to install packages from the other locations, do the following:

1. Copy the ISO files to `/data1`.

2. Mount the ISO files. For example:

```
# mount -o loop /data1/foundation-2SP1-cd1-media-sles11-x86_64.iso /mnt/iso3
# mount -o loop /data1/propack-7SP1-cd1-media-sles11-x86_64.iso /mnt/iso4
```

3. List the repository availability using one of the following:

```
# yum repolist
# yum list   (lists all packages present in the repository)
```

Following are additional `yum` commands:

- Install packages onto the master host:

```
# yum install package1 package2 ...
```

- Remove packages from the master host:

```
# yum remove package1 package2 ...
```

- Install packages into the compute node payload:

```
# yum install --installroot=/compute package1 package2 ...
```

- Remove packages from the compute node payload:

```
# yum remove --installroot=/compute package1 package2 ...
```

You can also use the `rpm` command to perform the same process, but without automatic dependency checking:

- Install onto the master host:

    # **rpm -Uvh** *package*

- Install into the compute node payload:

    # **rpm --root /compute -Uvh** *package*

- Query software that has been installed on the master host:

    # **rpm -qa | grep -i** *package*

- Query software that has been installed in the compute node payload:

    # **rpm --root /compute -qa | grep -i** *package*

Note: After the software in the compute node payload has been installed or removed, you must check in the payload.

Configuration Tools

CMK-MI contains the following configuration tools:

- "`sgi_node-config.py`" on page 15
- "`sgi_cluster_import.pl`" on page 17

sgi_node-config.py

The `sgi_node-config.py` Python script configures the following files:

```
/etc/hosts
/etc/hosts.equiv
/root/.rhosts
/etc/skel/mpd.hosts
/etc/pdsh/machines
/etc/conman.conf
/etc/powerman/powerman.conf
/compute/etc/hosts
```

```
/compute/etc/hosts.equiv
/compute/root/.rhosts
/etc/skel/.rhosts
/home/sgi/.rhosts
/root/mpd.hosts
/home/sgi/mpd.hosts
/compute/root/mpd.hosts
/var/spool/PBS/server_priv/nodes
/opt/intel/clck/1.5/nodelist-eth
/opt/intel/clck/1.5/nodelist-ib
/opt/intel/clck/1.5/nodelist-ge
```

The script is located in:

```
/opt/sgi/Factory-Install/Integration/sgi_node-config.py
```

The script uses ssh keys and contains information for root and normal users.

The script arguments are as follows:

Option	Description
-h	Displays option help information
	Note: Alternate arguments, such as --help, are available. Use the -h option for more information.
-T *TEMP_DIR*	Runs in test mode and puts output files in a temporary directory (*TEMP_DIR*)
-v *level*	Outputs debug messages of the specified *level* or lower, from 1 (error messages) through 3 (error, warning, and informational messages)
-n *COUNT*	Specifies the number of compute nodes
-g *COUNT*	Specifies the number of graphics nodes
-S *COUNT*	Specifies the number of service nodes
--nco *OFFSET*	Specifies the number from which to start compute node numbering
--gnco *OFFSET*	Specifies the number from which to start graphics node numbering

`--snco` *OFFSET*	Specifies the number from which to start service node numbering
`-e`	Enables creation of dedicated Gigabit Ethernet entries
`-i`	Disables creation of InfiniBand node entries
`-b`	Disables creation of BMC entries
`-p`	Disables the fixing of user permissions
`--gs` *COUNT*	Specifies the number of Gigabit Ethernet switches in the cluster
`--gs_co` *OFFSET*	Specifies the number from which to start Gigabit Ethernet switch numbering
`--is` *COUNT*	Specifies the number of InfiniBand switches in the cluster
`--is_co` *OFFSET*	Specifies the number from which to start InfiniBand switch numbering
`-F`	Creates/updates core system and payload files
`-K`	Creates/updates system `ssh` keys
`-s` *STAGE_DIR*	Specifies the temporary staging directory (*STAGE_DIR*)
`--no_cleanup`	Leaves temporary files and directories in place

For example, the following creates a cluster configuration with 32 compute nodes, 2 Gigabit Ethernet switches, and 1 InfiniBand switch, and with SSH keys generated:

```
# ./sgi_node-config.py -n 32 --gs 2 --is 2 -F -K
```

`sgi_cluster_import.pl`

The `sgi_cluster_import.pl` Python script reads a file that contains MAC addresses and node names and writes out the following files that can be imported into SGI Management Center:

- `nodes.conf`
- `nodes.cwhost.sh`
- `nodes.dbix`

The script is located in:

`/opt/sgi/Factory-Install/Integration/CDT/sgi_cluster_import.pl`

The file is in the following format:

```
00:00:00:00:00:00 n001
00:00:00:00:00:00 n002
...
```

Note: The ： delimiters are optional.

The current nodes have already been imported into the cluster. The tool is left for future expansion.

Rebuilding the Master Host After Failure

If there has been severe failure, such as if your hard drive has become damaged or a filesystem has been corrupted, you can use the following procedure to return the master host to its original state as shipped from the factory. This procedure assumes that you have a working disk and a means of booting Linux based rescue media.

1. Back up the following files to another system:

   ```
   /data1/sgi_host_clone.log
   /data1/sgi_host_clone.md5sum
   /data1/sgi_host_clone.tar.gz
   ```

2. Boot the master host from the rescue media.

3. Choose the **Rescue System** mode from the boot list.

4. After the system boots, log in as root.

5. Verify that you can see the disks:

   # **cat /proc/partitions**

6. Use the parted(8) command to create the partitions for the filesystems:

   a. Enter the disk utility:

      # **parted**

   b. Create the following partition layout, changing the partition system IDs for swap (82) and linux (83):

      ```
      /boot     /dev/sda1    100MB
      /         /dev/sda2    100GB
      swap      /dev/sda3      8GB     (Or 1.5x the physical memory)
      /data1    /dev/sda4        *     (Remainder of the disk)
      ```

7. Make the following filesystems:

```
# mkfs.ext3 -L sgiboot /dev/sda1
# mkfs.xfs -f -L sgiroot /dev/sda2
# mkswap -L sgiswap /dev/sda3
# mkfs.xfs -f -L sgidata1 /dev/sda4
```

8. Make a temporary `rescue` directory:

```
# mkdir /tmp/rescue
```

9. Create and mount the filesystems onto their corresponding mount points:

```
# mount /dev/sda2 /tmp/rescue
# mkdir /tmp/rescue/boot
# mkdir /tmp/rescue/data1
# mount /dev/sda1 /tmp/rescue/boot
# mount /dev/sda4 /tmp/rescue/data1
```

10. Verify that the partitions are correctly mounted:

```
# df -h
```

11. Configure the Ethernet interface for local network connectivity:

```
# ifconfig interface external_IPaddress
```

12. Verify that the interface is configured properly by testing connectivity:

```
# ping  second_external_IPaddress
```

13. Copy the /data1/sgi_host_clone.tar.gz tarball (which contains the files required to rebuild the master node) to the `rescue` directory:

```
# cp /data1/sgi_host_clone.tar.gz /tmp/rescue
```

If the tarball is on a remote host, do the following instead:

```
# scp user@remote_host_IPaddress:/location_of_tarball /tmp/rescue/data1
```

If you do not have access to sgi_host_clone.tar.gz, contact SGI Support.

14. Extract the contents of the tarball to the parent directory of the newly created filesystem:

```
# tar xfz /tmp/rescue/data1/sgi_host_clone.tar.gz  -C /tmp/rescue
```

15. Change to a new root (the *"chrooted" environment*):

    # **chroot /tmp/rescue**

16. Install the GRUB boot loader:

    # **mount -t /proc**
    # **rm /boot/grub/device.map**
    # **grub-install /dev/sda**

    If the GRUB installation is successful, it will state that it installed the new boot
    loader.

17. Exit out of the "chrooted" environment:

    # **exit**

18. Reboot the master host:

    # **init 6**

19. Log in as `root`

20. Start SGI Management Center:

    # **mgrclient**

21. Select the compute nodes and select `provision` to recreate the distribution
    channels.

22. See Chapter 1, "Networking Setup" on page 1.