SGI® Management Center Quick Start Guide

007–5672–002

Record of Revision

Version	Description
001	July 2010 Original publication
002	August 2011 Revision

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, "SGI Management Center Setup" on page 1
- Chapter 2, "Applications and Services" on page 5
- Chapter 3, "Configuration Tools" on page 17
- Chapter 4, "Rebuilding the Master Host After Failure" on page 21

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 Installation 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 (SMC) Installation and Configuration*
- *SGI Management Center (SMC) System Administrator's 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.

SGI Management Center 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. This chapter discusses the following:

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*`:`*admin* and *system*`.`*rna*`.`*host*`:`*admin*). You must not change the `admin` value:

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

- *(Optional)* InfiniBand:

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

- *(Optional)* A second Ethernet interface:

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

Networking Changes You Must Make

This section discusses the networking changes you must make. Some instructions vary by operating system, either Red Hat Enterprise Linux (RHEL) or SUSE Linux Enterprise Server (SLES).

Caution: For 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).

Do the following:

- Modify the `eth0` interface definition file to use your site-specific IP address:
  - **RHEL:** `/etc/sysconfig/network-scripts/ifcfg-eth0`
  - **SLES:** `/etc/sysconfig/network/ifcfg-eth0`
- 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

  By default, SGI Management Center does not use a `nameserver` value.

- 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`

- Ensure that the multicast route is set. Do the following, assuming that `eth1` is the management interface:

  – **RHEL:** add the following to the `/etc/sysconfig/network-scripts/route-eth1` file:

  ```
  239.192.0.0/16 dev eth1
  ```

  – **SLES:** add the following to the `/etc/sysconfig/network/routes` file:

  ```
  224.0.0.0    0.0.0.0    240.0.0.0   eth1    multicast
  ```

- 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.

- Ensure that the Dynamic Host Configuration Protocol (DHCP) interface (for example, `eth1`) is present in the `/etc/sysconfig/dhcpd` file. For more information, see the `dhcpd`(5) man page.
- Ensure that Trivial File Transfer Protocol (`tftp`) is enabled to start on boot and is currently running:

```
# chkconfig tftp on
# /etc/init.d/xinetd restart
```

Security Changes You Must Make

You must also do the following:

- Ensure that there is no firewall on the management interface (such as 10.0.10.1 on the master host and 10.0.1.1 on the a compute node).
- Ensure that Security-Enhanced Linux (SELinux) in not enabled in `enforcing` mode on the master host

---

Note: SELinux `enforcing` mode may cause payload creation to fail. If you attempt to create a payload while `enforcing` mode is enabled, you will see a warning message.

Applications and Services

This chapter provides a summary of the applications and services provided with your SGI Management Center cluster:

NAT

Network Address Translation (NAT) is implemented on the master host as follows:

- **RHEL:** in `/etc/rc.d/rc.local` with the following `iptables` rule:

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

- **SLES:** 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.

Use the following to synchronize the master node with your time servers:

- **RHEL:** `/usr/bin/system-config-date`
- **SLES:** `/sbin/yast timezone`

IGMP Routing

An Internet Group Management Protocol (IGMP) route is established on the master host on the cluster private network (`eth1`):

- **RHEL:** The `/etc/sysconfig/network-scripts/route-eth1` file contains the following:

  ```
  239.192.0.0/16 dev eth1
  ```

- **SLES:** The `/etc/sysconfig/network/routes` file contains the following:

  ```
  239.192.0.0    0.0.0.0    255.255.0.0   eth1    multicast
  ```

System Log Messages

System log messages for compute nodes are forwarded to the master host using port 514 and the following files:

- **RHEL 5:** `/etc/syslog.conf`
- **RHEL 6:** `/etc/rsyslog.conf`
- **SLES:** `/etc/syslog-ng/syslog-ng.conf`

Modules

The following modules are added to the specified files for the master host and the compute nodes:

`compiler/gcc-4.`*`X.X`*
`mpi/intel-4.0`

- **RHEL:** `/usr/share/Modules/modulefiles`
- **SLES:** `/usr/share/modules/modulefiles`

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.

Shells

SMC has the following shells:

- Passwordless secure shell (`ssh`) is configured using the `sgi_node-config.py` tool. See "`sgi_node-config.py`" on page 17.
- The remote shell (`rsh`) is configured on and compute nodes are added to `/etc/hosts.equiv` in the payload.
- 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 `admin:/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 payload has `ypbind` on with `ypserver` set to the master host. The `home` filesystem is automounted based on the settings in `/etc/auto.master` and `/etc/auto.home` in the payload.

AutoFS

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

ConMan is configured to use the default BMC user name (`admin`) and password (`admin`). Following is an example of the `/etc/conman.conf` file using the default username:

```
CONSOLE name="gn001" dev="/usr/lib/conman/exec/ipmitool.exp gn001-bmc admin"
CONSOLE name="n001" dev="/usr/lib/conman/exec/ipmitool.exp n001-bmc admin"
CONSOLE name="n002" dev="/usr/lib/conman/exec/ipmitool.exp n002-bmc admin"
```

```
CONSOLE name="n003" dev="/usr/lib/conman/exec/ipmitool.exp n003-bmc admin"
CONSOLE name="n004" dev="/usr/lib/conman/exec/ipmitool.exp n004-bmc admin"
...
```

If you want to set a different user name and password, you must edit the `/etc/conman.conf` and `/etc/conman.pswd` files. For example, to use a username of `myuser` and a password of `mypass`:

- In `/etc/conman.conf`:

  ```
  CONSOLE name="gn001" dev="/usr/lib/conman/exec/ipmitool.exp gn001-bmc myuser"
  CONSOLE name="n001" dev="/usr/lib/conman/exec/ipmitool.exp n001-bmc myuser"
  CONSOLE name="n002" dev="/usr/lib/conman/exec/ipmitool.exp n002-bmc myuser"
  CONSOLE name="n003" dev="/usr/lib/conman/exec/ipmitool.exp n003-bmc myuser"
  CONSOLE name="n004" dev="/usr/lib/conman/exec/ipmitool.exp n004-bmc myuser"
  ...
  ```

- In `/etc/conman.pswd`:

  ```
  n[0-9]*-bmc : myuser : mypass
  gn[0-9]*-bmc : myuser : mypass
  ```

The cluster is configured to match the node type.

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

Special BMC Requirements

Depending on your node type, different IPMI parameters must be passed to the BMC. For example, an Altix XE 210 system 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 system, 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 system, you would change the line to:

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

6. Modify the /etc/freeipmi/freeipmi.conf file 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 to use freeipmi by default for all compute nodes in the cluster, using the default BMC user name (admin) and password (admin). If you want to set a different user name and password, you must edit the /etc/freeipmi/freeipmi.conf file. For example, to use a username of myuser and a password of mypass:

```
username myuser
password mypass
```

Following is an example of a `/etc/powerman.conf` file for 32 nodes:

```
include "/etc/powerman/ipmipower.dev"
device "ipmi1" "ipmipower" "/usr/sbin/ipmipower -h n[001-032]-bmc |&"
device "ipmi2" "ipmipower" "/usr/sbin/ipmipower -h gn[001-002]-bmc |&"
node "n[001-032]" "ipmi1" "n[001-032]-bmc"
node "gn[001-002]" "ipmi2" "gn[001-002]-bmc"
```

Intel Cluster Checker

Note: This applies only to Intel clusters.

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/1.6/nodelist` file lists the compute nodes.

Package Management

This section discusses the following:

Note: When installing a large number of packages at one time, `zypper` is more reliable on SLES than `yum`.

zypper (SLES Only)

On SLES systems, there is a local `zypper` repository configured for the installation of packages from the following sources:

- SLES
- SLES SDK
- SGI Foundation Software

The SLES 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:

```
sles# mount -o loop /data1/foundation-2.3-cd1-media-sles11-x86_64.iso /mnt/iso3
sles# mount -o loop /data1/accelerate-1.1-cd1-media-sles11-x86_64.iso /mnt/iso4
```

3. List the repository availability using the following:

   - SLES 11:

     ```
     sles11# zypper repos
     sles11# zypper packages (lists all packages present in the repository)
     ```

   - SLES 10:

     ```
     sles10# zypper service-list
     sles10# zypper search (lists all packages present in the repository)
     ```

See also "rpm" on page 15.

yum (Preferred for RHEL)

On RHEL systems, there is a local yum repository configured for the installation of packages from SGI Foundation Software.

The ISO is mounted automatically in /etc/rc.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:

   ```
   # . /etc/rc.local
   ```

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

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

Following are additional yum commands:

- Install packages onto the master host:

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

- Remove packages from the master host:

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

- Install packages into the compute node payload:

  rhel# **yum install --installroot=/compute *package1 package2* ...**

- Remove packages from the compute node payload:

  rhel# **yum remove --installroot=/compute *package1 package2* ...**

See also "rpm" on page 15.

rpm

You can also use the rpm command to perform the same process as listed above for zypper and yum, 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

This chapter discusses the following configuration tools:

`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
/opt/intel/clck/1.6/nodelist-eth
/opt/intel/clck/1.6/nodelist-ib
/opt/intel/clck/1.6/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
`-P`	Create/update nodes in PBS Pro
`-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:

   - SLES 10 or SLES 11:

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

   - RHEL 5.*x*:

     ```
     rhel5# mkfs.ext3 -L sgiboot /dev/sda1
     rhel5# mkfs.ext3 -L sgiroot /dev/sda2
     rhel5# mkswap -L sgiswap /dev/sda3
     rhel5# mkfs.ext3 -L sgidata1 /dev/sda4
     ```

   - RHEL 6.*x*:

     ```
     rhel6# mkfs.ext4 -L sgiboot /dev/sda1
     rhel6# mkfs.ext4 -L sgiroot /dev/sda2
     rhel6# mkswap -L sgiswap /dev/sda3
     rhel6# mkfs.ext4 -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 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 changed root 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, "SGI Management Center Setup" on page 1.