Platform Open Cluster Stack (OCS) Installation and Troubleshooting Notes

Version 4.1.1-2.0

October 26 2006

Platform Computing

Contents

- General Installation Notes
- Special Installation Notes
- Precautions
- Troubleshooting
- Learn About Platform Products
- Get Technical Support
- Copyright and Trademarks

[Top]

General Installation Notes

To install Platform OCS (Platform Open Cluster Stack) and the associated rolls, follow the installation instructions in the User Guide. The User Guide is available at the Platform Web site.

After installing Platform OCS, you should register your cluster on the Platform Web site: http://www.platform.com/Products/

The following topics cover likely tasks that you will routinely perform when installing or upgrading Platform OCS. We highly recommend that you familiarize yourself with these topics before you install or upgrade your version of Platform OCS.

- Minimum disk space requirements
- Reinstalling compute nodes that run older versions of Platform OCS
- Preserving existing partitions in a frontend installation
- Booting other operating systems in a new frontend installation
- Getting Enterprise Edition updates to the Red Hat Enterprise RPMs
- Changing the default partition size
- Installing an rsh server

Minimum disk space requirements

Platform OCS requires a minimum of 17 GB of disk space. Each specific partition has the following minimum requirements, with the default installation values in parentheses:

- Root partition: 6 GB (10 GB by default)
- Swap partition: 1 GB (4 GB by default)
- Export partition: 10 GB (Remainder of the first disk by default)

Reinstalling compute nodes that run older versions of Platform OCS

To reinstall compute nodes you must remove the `/.rocks-release` file from each node. This will ensure that a fresh installation instead of an upgrade is done on the nodes.

Before starting the frontend installation, log in as `root` and run the following command to reinstall all compute nodes:

```
# cluster-fork rm -fr /.rocks-release
```

Preserving existing partitions in a frontend installation

In this current version of Platform OCS, you can preserve the existing partitions on the frontend using the **Autopartition** option in Anaconda when performing a frontend installation. This is useful if you have another operating system installed on your frontend node and want to keep it.

In the **Disk Partitioning Setup** dialog, select **Autopartition**. You can choose to preserve all existing partitions, remove all Linux partitions, or remove all partitions. Regardless of the option you choose, Anaconda will preserve the Dell utility partition if it is installed on the first disk of your frontend node.

After setting up your partitions, the **Boot Loader Configuration** dialog displays your various boot partitions. Non-Linux partitions such as Windows use the `default label of "Other"`, which you should change to a more descriptive name. By default, the partition on which you install Platform OCS is selected as the default partition to boot.

Limitations

- You can only preserve partitions on frontend nodes. This feature is not supported on other appliance types, such as compute nodes.
- The Platform OCS installer is based on the Anaconda installer that is used by RedHat® Enterprise Linux (RHEL). The Anaconda installer does not detect Linux-based installations, such as other RHEL installations. You will need to manually add the appropriate entry into the Grub menu, as described in Booting other operating systems in a new frontend installation.
- Partitions using LVM and software RAID are not supported.

Dell utility partition

The **Autopartition** option automatically preserves the Dell utility partition (UP) even if you choose to remove all partitions. In addition,you cannot remove the Dell UP using the **Disk Druid** option.

If you install multiple versions of Platform OCS on the frontend node, the Dell UP boot loader will always boot the Platform OCS version installed in the partition located immediately after the UP on the disk. To boot other Platform OCS installations located in other partitions, you need to manually append the GRUB entries from those other partitions into the GRUB menu for the version of Platform OCS installed in the partition immediately after the UP.

For example, in SCSI disks, Dell UP is installed on `/dev/sda1`. This means the Dell UP bootloader will boot whatever operating system is installed in `/dev/sda2`.

Booting other operating systems in a new frontend installation

The Platform OCS installer does not automatically add entries for existing Linux operating systems to the GRUB configuration file of the new installation. Only non-Linux installations, such as Windows XP are added automatically. The following lists the steps for booting other Linux installations from the GRUB menu of a newly-installed Platform OCS cluster:

1. Boot into your newly installed Platform OCS frontend and log in as `root`.
2. Locate the boot partition containing the boot files for the selected Linux installation that you want to add:
   - For Platform OCS installations, the boot partition is usually the root partition (`/`).

- For other Linux installations such as RedHat Enterprise Linux, the boot partition is usually the boot partition (`/boot`), or the root partition (`/`).

3. Mount the device representing the boot partition for the chosen Linux installation.

   For example,

   ```
   # mkdir /tmp/mount
   # mount /dev/sda1 /tmp/mount
   ```

4. Locate the GRUB configuration file for the selected Linux installation:

   For example, if you mounted the boot partition on `/tmp/mount`,

   - For Platform OCS installations, this GRUB configuration file is `/tmp/mount/boot/grub/grub-orig.conf`
   - For other Linux installations such as RedHat Enterprise Linux, the configuration file is `/tmp/mount/boot/grub/menu.lst`

5. Locate the GRUB boot entry for booting the selected Linux installation:

   The sample entry below is for Platform OCS:

   ```
   title Platform OCS (2.6.9-34.EL)
           root (hd0,0)
           kernel /boot/vmlinuz-2.6.9-34.EL ro root=LABEL=/1 rhgb quiet
           initrd /boot/initrd-2.6.9-34.EL.img
   ```

6. Add your selected GRUB boot entry to the `/boot/grub/grub-orig.conf` file for the new Platform OCS installation. There are two ways to do this, depending on your desired boot order:
   - If you don't want your selected GRUB boot entry to be the default boot entry, just append it to the end of the `/boot/grub/grub-orig.conf` file.
   - If you want your selected GRUB boot entry to be the default boot entry, add the new entry to the beginning of the `/boot/grub/grub-orig.conf` file, before the first entry currently in this file.

7. Refresh the GRUB device map file for the new Platform OCS installation

   ```
   # rm -f /boot/grub/device.map
   # echo 'quit' | grub --batch --device-map=/boot/grub/device.map
   ```

8. Reboot the machine.
9. When the machine boots into GRUB, you will see your new GRUB boot entry added to the end of the list. To boot into it, select the boot entry and press **Enter**.

Getting Enterprise Edition updates to the Red Hat Enterprise RPMs

For Enterprise Edition, you can use the `rocks-dist` and `rocks-update` commands to synchronize the RPM repository with the Red Hat mirror. The RPMs on the SDSC cluster toolkit mirror are not officially supported Red Hat Enterprise 4 U3 RPMs.

Use the patch management tool, `rocks-update`, located in your Platform roll to install updates for your frontend and compute node operating systems.

More information is available in *Readme for Platform Roll* available once you have installed the Platform roll on the Platform OCS Base CD.

Changing the default partition size

The default root partition size for both frontend and compute nodes is currently 10 GB for root, and 4 GB for swap, with the remainder of the disk mounted as `/state/partition1` (that is, the export partition).

If you wish to increase the size of the root or swap space partition on the frontend, set the size during installation by selecting "manual partitioning" and creating the following partitions:

- `/`

  of type `ext3`, with a minimum size of 6 GB

- `swap`

  of type `linux swap`, with a size greater than the size of the RAM and a minimum of 1 GB.

- `/state/partition1`

  of type `ext3,` with a minimum size of 10 GB

If you want to change the size of the root or swap space for compute nodes, use the `custom-partition` tool as follows:

- Change the root partition size by running the following command:

  # **custom-partition -r** *<root_partition_size>*

  where *<root_partition_size>* is specified in MB and must be at least 6 000 MB.

- Change the swap partition size by running the following command:

  ```
  # custom-partition -s <swap_partition_size>
  ```

  where *<swap_partition_size>* is specified in MB and must be at least 1 000 MB.

- You can also change both the root and swap partition sizes simultaneously:

  ```
  # custom-partition -r <root_partition_size> -s <swap_partition_size>
  ```

Installing an rsh server

If you want to install remote shell services (for example, `rsh`, `rlogin`, `rexec`) on your compute nodes, the rsh server rpm must be merged with the Platform OCS distribution software on the frontend node before the compute nodes are installed. To set up `rsh` on your compute nodes, you must install an `rsh` server on the compute nodes. This can be done as a post-installation procedure.

We do not recommend rsh as it has known security flaws. We suggest you use ssh.

To install an rsh server after installing Platform OCS

In the following steps, replace *<arch>* with the architecture of your distribution: `i386` or `x86_64`.

---

1. Log in to the frontend node as `root`.
2. Copy the `rsh` server file to your frontend node.

   The rsh-server RPM can be found in your frontend node at `/home/install/rocks-dist/lan/`*<arch>*`/RedHat/`*RPMS/*

3. Install the rsh server by running the following command:

   ```
   # rpm -ivh rsh-server-0.17-25.3.<arch>.rpm
   ```

4. Add the package for `rsh` with `xinetd` to the compute nodes:
   a. Copy the `rsh-server-0.17-25.3.`*<arch>*`.rpm` file to the following directory:

```
/home/install/contrib/4.1.1/<arch>/RPMS/
```

b. Navigate and copy to the appropriate `xml` file for the type of compute nodes (**Appliance Type**) you are installing:

```
# cd /home/install/site-profiles/4.1.1/nodes/
```

For **compute**, run the following command:

```
# cp skeleton.xml extend-compute.xml
```

c. Add the package by adding the following lines to the `extend-compute.xml` file:

```
<package>rsh-server</package>
<package>xinetd</package>
```

d. Build a new Platform OCS distribution to distribute the `rsh` server to the compute nodes:

```
# cd /home/install
# rocks-dist dist
```

5. Reinstall the compute nodes (**Appliance Type**).
6. Enable the `rsh` server on the frontend node and compute nodes:
   a. Run the following commands:

```
# /sbin/chkconfig rsh on
# /sbin/chkconfig rlogin on
# /sbin/chkconfig rexec on
```

   b. Add the following lines to the `/etc/securetty` file:

```
rsh
rlogin
```

   c. Edit `/etc/pam.d/rsh`:

Set **pam_rhosts_auth.so** as follows:

```
auth required pam_rhosts_auth.so promiscuous
```

d. Edit `/etc/pam.d/rlogin`:

Set **pam_rhosts_auth.so** as follows:

```
auth sufficient pam_rhosts_auth.so promiscuous
```

e. To enable logins from other remote hosts, add the trusted host names to the `$HOME/.rhosts` file. For example:

```
hostname_1
hostname_2
...
hostname_n
```

f. Check that the permissions of the `$HOME/.rhosts` file are set to `600`.
g. Restart `xinetd` as follows:

```
service xinetd restart
```

Use `cluster-fork` to run commands across the nodes in your cluster.

[Top]

Special Installation Notes

The following topics are provided to assist you with special situations you will occasionally encounter that require manual steps before or during installation.

Platform Lava and Platform LSF HPC cannot run on the same node. The Platform LSF HPC roll will disable Platform Lava

---

- USB storage devices
- Secondary Fiber Channel/SCSI/SAS storage

USB storage devices

Platform OCS cannot be installed if certain USB storage devices are connected to the installation host (for example, a USB floppy or USB zip drive). You must disconnect these devices before starting the installation. This affects both frontend and compute node installations.

Installing Platform OCS via a USB drive will work if no other USB storage devices are connected.

Secondary Fiber Channel/SCSI/SAS storage

IMPORTANT

Failure to remove secondary SCSI/SAS storage attached to a server during OS installation may result in loss of data on the storage due to reformatting and repartitioning. To avoid any loss of data, detach the secondary storage before installation.

The Platform OCS 4.1.1-2.0 install kernel will not load modules to detect a Qlogic card. Thus secondary fiber storage attached via a QLogic card will not be detected or partitioned during installation.

Installation of any appliance (frontend, compute, or IBRIX node) that has a PCI-E-based Qlogic card results in the OS being installed on the fiber storage that is attached via the Qlogic card. This is because the qlogic modules load before the SCSI modules.

To avoid this problem, the Platform OCS 4.1.1-2.0 installation kernel does not contain qlogic modules and will therefore not detect any fiber storage connected to Qlogic cards during the installation of any appliance (frontend, compute, or IBRIX node). After installation, the server will boot up with the correct qlogic driver, and any connected fiber storage will be accessible by the server.

Disk Druid does not create a default partition layout for the user

If you select Disk Druid to manually partition your disk, there will be no default partition layout. You must create all the

partitions yourself, and supply Disk Druid with the correct partition layout. You cannot proceed until you create a root partition.

As an alternative to Disk Druid, Platform recommends auto-partitioning your disk instead. However, if you do choose to use Disk Druid, you must (at the very minimum) create the following partitions:

Partition	Mount point	File system type	Minimum partition size
Root	/	Ext3	6 GB
Swap	None	Swap	1 GB
Export	/state/partition1	Ext3	10 GB

See Changing the default partition size for an example on how to change default partition sizes.

Enabling BitTorrent file transfers

In Platform OCS, BitTorrent is not used to install nodes. SDSC cluster toolkit 4.1 uses `bittorrent` to transfer files to installing nodes. In tests it was found that installation was actually slower for reinstalling individual nodes, and for installing less than 32 hosts at a time. By default, Platform OCS uses `http`, not `bittorrent`, to transfer files to installing nodes. A node will require approximately three to five seconds of network bandwidth on Gigabit Ethernet.

You can enable BitTorrent file transfers by running the following command on the frontend:

```
# /opt/rocks/sbin/rocks-bittorrent on
```

You can restore the default setting (disable BitTorrent file transfers) by running the following command on the frontend:

```
# /opt/rocks/sbin/rocks-bittorrent off
```

Disabling kickstart caching

During installation of compute nodes, the frontend node must generate an Anaconda kickstart file. This file controls the partitioning, packages, and post-installation scripts that are applied to the booting node. Generating the kickstart file is very CPU-intensive for the frontend, and is a contributing factor to compute node installation failure. The differences between a kickstart file from one compute node to another is minimal.

Platform OCS uses a caching mechanism to reduce the load on the frontend node during kickstarting. This allows the frontend

node to install more compute nodes simultaneously. Caching is only used for compute nodes, and other appliances will not use caching.

Caching may not be compatible with all rolls, and for small clusters, it may not be necessary. In these cases, you can disable kickstart caching by running the following command on the frontend node:

```
# touch /home/install/sbin/cache/disable-cache
```

If the OCS database is changed manually, it is necessary to remove the kickstart cache file(s). Run:

```
# rm -rf /home/install/sbin/cache/ks.cache*
```

Precautions

The suggestions below provide pre- or post-installation precautions when installing or running Platform OCS. Be sure you are familiar with these suggestions before you install or run Platform OCS.

1. **Secondary fiber/SCSI/SAS storage**

   You must remove any secondary SCSI/SAS/FC storage attached to a compute node (or to any other appliance) before installing the OS. Failure to do so can result in loss of data on the storage due to reformatting and repartitioning.

   To avoid any loss of data, detach the secondary storage before installation.

   The Platform OCS 4.1.1-2.0 install kernel will not load modules to detect a Qlogic card. Thus secondary fiber storage will not be detected or partitioned during installation.

2. **PCI-E-based Qlogic cards**

   Installation of any appliance (frontend, compute, or IBRIX node) that has a PCI-E-based Qlogic card results in the OS being installed on the fiber storage that is attached via the Qlogic card. This is because the qlogic modules load before the SCSI modules.

   To avoid this problem, the Platform OCS 4.1.1-2.0 installation kernel does not contain qlogic modules and will therefore not detect any fiber storage connected to Qlogic cards during the installation. After installation, the server will boot up

with the correct qlogic driver, and any connected fiber storage will be accessible by the server.

3. **Eth1**: Not configuring `eth1` hinders deployment.

    Explanation: Platform OCS 4.1.1 requires eth1 on the frontend to be configured during installation. If `eth1` is not configured, the `httpd` service fails on the frontend and compute node deployment fails.

    Not configuring `eth1` during installation, or re-configuring `eth1` after installation also causes Myrinet jobs submitted from the frontend node to fail.

4. **Partitioning**: Changes to partitioning with Platform OCS 4.1.1-2.0 and later.

    Explanation: Some changes have been made to preserve for the Dell utility partition that affect partitioning. In the original Platform OCS 4.0, the `.rocks-release` file is created on the first partition of the first disk. The `.rocks-release` file is now in the root partition, `/.rocks-release`. During an upgrade of frontend and compute nodes, this `.rocks-release` file determines whether Platform OCS has previously been installed on a host.

    The `rocks-partition --list` option now displays the file system type, as well as the partition id/type, and the original device name and mount point information.

    The root partition size is 10 GB by default with a minimum size of 6 GB. The swap partition size is 4 GB by default with a minimum size of 1 GB.

    You should be aware of the following issues when you partition. They are not currently documented in the SDSC cluster toolkit 4.1 Users Guide:

    http://www.rocksclusters.org/rocks-documentation/4.1/

    Frontend upgrade: When upgrading a frontend, the installer requires you to select Disk Druid to set up the partitioning. In Disk Druid, you need to do the following in order to make the upgrade work:

    a. Specify the mount point for the root partition as "/", and reformat it.
    b. Specify the mount point for the /export partition as "/state/partition1", and do not reformat it.

    DB partitioning tables: You should never modify the partitions table in the Platform OCS mySQL database on the frontend node. Use `rocks-partition` to either list or delete the partitions for a specified node (using `--list` and `--delete`, respectively).

5. **Re-compiling applications for optimal performance**

Applications should be compiled using optimizations for their specific architecture. For instance: To take full advantage of the new Intel® Core Microarchitecture it is best to recompile your applications with compilers supporting the new architecture.

Troubleshooting

The suggestions below provide troubleshooting tips when installing or running Platform OCS.

1. **Symptom**: Using `add-extra-nic` to add a virtual NIC fails.

   Explanation: Platform OCS does not support virtual NICs.

   Solution: You need to manually add virtual NICs to each host they are needed on. You also need to recreate the virtual NIC every time you reinstall the node.

2. **Symptom**: You cannot use `rocks-update` to download and install a package that is not installed on the frontend node.

   Explanation: `rocks-update` can not install new packages. To install the new packages, you need to use the `yum` tool.

   Solution: Note that these solutions will not increment the repository install version on a compute or frontend node.

   ❍ To download and install a new package that is not installed on your frontend, use `rocks-update` to download the package and `yum` to install it:
      i. # **rocks-update -d** *<package_name>*
      ii. # **yum -y install** *<package_name>*
   ❍ To install a package in a new compute node, you need to configure Platform OCS to put the RPMs into an `extend-compute` node using the `rocks-compute` tool before running `rocks-update`:

   # **rocks-compute -a -p /var/cache/yum/***<package_name>***.rpm**

   ❍ To patch your currently-installed compute nodes, you can use the `yum` tool:

   # **cluster-fork 'yum -y install** *<package_name>***'**

This will install the package and any required dependencies.

3. **Symptom**: Customizations made to the Platform OCS distribution do not take effect after you reinstall a compute node.

   Explanation: Customizations made by adding XML files to or removing them from the `/export/home/install/site-profiles/4.1.1/nodes` directory require rebuilding the Platform OCS distribution. In Platform OCS 4.1.1, a new feature was introduced to cache kickstart files to improve the installation performance of compute nodes. If you do not rebuild the Platform OCS distribution, and you reinstall a compute node, a cached version of the kickstart file is used. This old cached version will not contain your new XML file changes.

   Solution: Always rebuild the Platform OCS distribution when you are customizing the distribution by adding or removing your own XML files. This will guarantee that the kickstart file used by the next compute node installation is based on the latest XML files in your Platform OCS distribution.

   ```
   # cd /export/home/install; rocks-dist dist
   ```

4. **Symptom:** When you download one or more RPMs and run **rocks-update -c** or **rocks-update -f** to patch, `rocks-update` reports that there are no updates to install.

   Explanation: The RPMs that `rocks-update` did not detect have been corrupted.

   Solution: Download the corrupted RPMs again.

5. **Symptom:** If the DNS has conflicting short names, you may experience the following symptoms:
   - When connecting from any compute node to the frontend via ssh, the ssh fails. You will get a "temporary failure in name resolution" error. `/var/log/message` will short an error message such as the one below:

     ```
     Jul 29 00:26:03 compute-850-0 sshd[22561]: reverse mapping checking getaddrinfo
     for cfg.local failed - POSSIBLE BREAKIN ATTEMPT!
     ```

   - Pinging the frontend from any compute node will also fail.

   Solution:

   - Look at the `/var/named/rocks.domain` file and check if the short name of the problematic compute node conflicts with any of the other short names.

     If it does, use **insert-ethers --remove** to remove the compute node and reinstall it with a unique rack and rank combination.

6. **Symptom:** When removing or upgrading a roll, packages and node/graph files in `site-profiles` may be executed on the frontend and compute nodes.

   Solution: If you made any customizations, you need to move RPM and XML files from `site-profiles` before removing or upgrading a roll.

   Check for any RPM and XML files inside the following directories and temporarily move these files to another location:

   - `/export/home/install/site-profiles/4.1.1`
   - `/export/home/install/contrib/4.1.1/x86_64/RPMS`

   You can now remove or upgrade a roll using `rollops`. When you are finished, you can move the files back to their original locations.

   Link the custom graph and node files into the distribution as follows:

   1. `# cd /export/home/install`
   2. `# rocks-dist dist`

7. **Symptom:** Configuration files on compute nodes for non-Platform rolls are correct for the first node installed, but wrong on the others.

   Explanation: Caching is used to speed the generation of the kickstart files. The kickstart files control the packages to install and the configuration files needed by the rolls. If a roll generates different configuration files depending on specific compute node configurations, the cached kickstart file will not reflect those differences. Rolls from Platform, IBRIX, Absoft, and NICE, are tested to work with kickstart caching. Other rolls have not been tested with kickstart caching.

   Solution: Turn off kickstart caching using the following command:

   `# touch /home/install/sbin/cache/disable-cache`

8. **Symptom:** When using `rocks-update`, you may encounter `up2date` errors.

   Explanation: These errors are generated by Red Hat's `up2date` utility.

   Solution: Contact Red Hat for support or go to http://www.redhat.com/ and search for your specific `up2date` error message.

9. **Symptom:** Frontend installation fails when installing from an external USB drive. After typing `boot: frontend` to perform a frontend install, the installation fails with the following message:

```
Cannot find file http://127.0.0.1/RedHat/base/updates.img
```

Explanation: If there is media inside the internal drive, the installer attempts to look for installation files from the internal drive first, even if you booted from the USB drive.

Solution: Eject the media from your internal drive before typing `boot: frontend`.

10. **Symptom:** If you select autopartitioning during a frontend install, the following error dialog appears:

```
-------------------
Error Partitioning
-------------------
Could not allocate requested partitions:

Partitioning failed: Could not allocate partitions as
primary partitions.
Press 'OK' to reboot your system.
```

If you press **OK**, you will reboot your system. After the reboot, the machine will start the installer again.

Explanation: This error indicates that you don't have enough disk space to create the partitions required by Platform OCS (root, swap, and export). The minimum amount of disk space required for a Platform OCS frontend installation is 17 GB.

To determine how much space you have left on your disk, follow these steps:

1. Select **Disk Druid** from the **Disk Partitioning Setup** screen
2. On the **Disk Druid** screen, look for the disk on which you want to install Platform OCS. To determine how much free disk space is available, look for **Free Space** under the **Device** column, and look across the row to find the size in MB.

Solution: If you don't have enough disk space, try any of the following suggestions to free up additional disk space:

- Select **Autopartition** from the **Disk Partitioning Setup** screen, and select either one of the following options from the **Prepare Disk** screen:
  - **Remove all Linux partitions on this system**
  - **Remove all partitions on this system**

○ Enter **Disk Druid** and manually clean up your partitions.
○ If your host has multiple disks, select a different disk.

11. **Symptom:** The RCS log files only contain one entry after installation.

Explanation: The kickstarting process has been optimized in Platform OCS 4.1.1-2.0. Systems install almost 1 minute faster with the new version. To do this, the files in RCS are only checked in once, during the first boot after installation.

Solution: You may restore the old behaviour.

For PXE booting nodes, edit the `/tftpboot/pxelinux/pxelinux.cfg/default` file and add "`logrcs`" to the end.

of the line beginning with "`append`".

12. **Symptom**: When installing a compute node with Platform OCS the installation may fail with the following message:

```
Error setting up Client Security Credentials
```

Explanation: Before a node can be reinstalled, its membership in the cluster must first be verified. To accomplish this, a cluster certificate is created on each node when it is first installed in the cluster. This certificate is presented to the frontend during the verification process.

When this verification process fails, you will get the above error. This error can occur due to the following:

○ Platform OCS has been installed on the compute node before, and the installer is using certificates from the older installation
○ The compute node has corrupt or mssing certificate files. This may occur when the compute node was rebooted during post-installation.

Solution:

Preferred solution: Boot into the compute node and delete the `/.rocks-releases` file, then reinstall your node. If this is not possible, boot into rescue mode by booting the node using a CD and using the rescue mode to get a shell. To delete the partition table, run the following command:

```
dd if=/dev/zero of=/dev/<your_primary_device> bs=1024
count=1024
```

Note that this will delete all partitions, including the Dell Utility partition if present.

Alternative solution (use the `dropcert` boot option): Open `/tftpboot/pxelinux/pxelinux.cfg/default` and add `dropcert` to the end of the "`append`" line, then reinstall your node.

13. **Symptom**: If you run "`411put *`" in any directory other than `/etc/411.d`, and then try to install a compute node, the node might fail to install.

    Explanation: The `kpp` script executed by the `kickstart.cgi` script might hang during the parsing of the `411-client.xml` file because of extraneous files in `/etc/411.d` that were added you ran "`411put *`".

    Solution: Remove all of the extraneous files from `/etc/411.d` and leave only the ones starting with "`etc.`". Refrain from running "`411put *`" in any directory, including `/root`.

14. **Symptom**: Nodes use only up to 496MB of memory during installation.

    Explanation: Segmentation faults can occur while installing nodes with large hard drives, such as a compute node installation on a server with two 250GB SATA drives and 1GB or more of RAM. To prevent these errors, the `mem` kernel parameter is set to a low value. Platform OCS sets this kernel parameter to 496MB (`mem=496M`).

    Solution: To use more memory during frontend installation, set the following parameter at the CD prompt during installation:

    frontend mem=<*preferred_memory_value*>

    To use more memory during compute node installation, edit the `/tftpboot/pxelinux/pxelinux.cfg/default` file on the frontened and add the following as a kernel parameter:

    mem=<*preferred_memory_value*>

    The nodes will now install using the amount of memory as specified by `<preferred_memory_value>`.

15. **Symptom**: When downloading RPMs from the Red Hat Network using `rocks-update`, you may encounter SSL errors, even when your host has the correct system time. This can sometimes result in the RPM not being downloaded, as shown in the following example:

    ```
    rocks-update: Downloading ===> gtk2-devel-2.4.13-18.x86_64.rpm
    There was an SSL error: (104, 'Connection reset by peer')
    A common cause of this error is the system time being incorrect. Verify that
    the time on this system is correct.
    ```

Solution: When running `rocks-update`, if you see an SSL error when downloading an RPM, check to see if that RPM was correctly downloaded. If not, run `rocks-update` to download the RPM again.

16. **Symptom**: After you build a frontend node from a central server, the frontend `rocks-update` relies on the central server to download updates.

    Solution: You need to convert your frontend node into a standalone frontend node so it can run `rocks-update` independently, as follows:

    1. Go to the frontend installation directory.

       ```
       # cd /export/home/install
       ```

    2. Create the installation directory structure.

       ```
       # mkdir -p
       /export/home/install/ftp.rocksclusters.org/pub/rocks/rocks-4.1.1/rocks-dist/rol
       ls/
       ```

    3. Move the central server installation directory to the new installaion directory.

       ```
       # mv
       /export/home/install/centralserver_hostname/install/rocks-dist/wan/IP_ADDRESS/r
       olls/*
       /export/home/install/ftp.rocksclusters.org/pub/rocks/rocks-4.1.1/rocks-dist/rol
       ls/
       ```

    4. Open the `/opt/rocks/etc/rocks-distrc` file and make the following changes in the `mirror` section:

In the `host` section, add the following line:

```
ftp.rocksclusters.org
```

In the `path` section, add the following line:

```
pub/rocks/rocks-4.1.1/rocks-dist
```

    5. Log in to the MySQL database and change the PublicKickstartHost value:

```
# mysql -u root -p cluster
Mysql> UPDATE app_globals SET Value="ftp.rocksclusters.org" WHERE
Component="PublicKickstartHost";
Mysql> quit
```

6. Delete the `/export/home/install/rocks-dist` file

```
# cd /export/home/install
# rm -rf /export/home/intsall/rocks-dist
```

7. Run `rocks-dist dist`

```
# ./rocks-dist dist
```

Note: Running `rocks-update -d` **packagename** prompts you to register with the Red Hat Network.

17. **Symptom**: Compute nodes crash after installing (or reinstalling) downloaded RPMs. This can also occur after installing (or reinstalling) a frontend node from a central server.

    Solution: Instead of reinstalling either the frontend or compute nodes, disable the updates roll on the frontend node before patching the frontend or compute nodes, as follows:

    1. In the frontend node, edit the `/opt/rocks/etc/rollopsrc` file in the and comment out the `updates roll` line.
    2. Turn off the updates roll using the `rollops` command.

    ```
    # rollops -p no -r updates
    ```

18. **Symptom**: The IBRIX node (or any other node) is patched after downloading updates to the frontend node.

    Solution: To prevent all nodes from patching the updates from the frontend node, disable the updates roll on the frontend node before patching the frontend or compute nodes, as follows:

    1. In the frontend node, edit the `/opt/rocks/etc/rollopsrc` file and comment out the `updates roll` line.
    2. In the frontend node, turn off the updates roll using the `rollops` command.

    ```
    # rollops -p no -r updates
    ```

19. **Symptom**: If you install Platform OCS to a frontend node and provide the PVFS2 roll during the initial installation, the

sample PVFS2 file system may not be available. To confim that it is missing, look for the `/mnt/pvfs2` sample directory.

Explanation: The `/etc/auto.master` file does not have an entry for the sample PVFS2 file system because it is missing a configuration line for PVFS2.

Solution: Log in to the frontend node as root and execute the following commands:

```
# echo "/mnt/pvfs2      /etc/auto.pvfs2     --timeout 60" >> /etc/auto.master
# cd /var/411
# make
# service autofs restart
```

20. **Symptom**: Some applications built with the Intel® compilers using -xP or /QxP optimization level may produce runtime errors.

    Explanation: The following message may be displayed when a program built with the switches "-xP" (on Linux*) or "/QxP" (on Windows*) is run on a system with an Intel® Core™ 2 Duo processor or an Intel® Xeon® processor 5100 series, previously codenamed "Woodcrest."

    "Fatal Error: This program was not built to run on the processor in your system."

    In addition, the program may not take the optimal execution path when it is built with the switches "-axP" (on Linux) or "/QaxP" (on Windows).

    Solution: Recompile/rebuild applications optimized for previous generation CPU architecture using latest Intel® Compilers that supports the new Intel® Core micro-architecture:

    ❍ Intel® C++ Compiler, for Linux* with Package ID l_cc_p_9.1.038 or higher
    ❍ Intel® C++ Compiler, for Linux* with Package ID l_cc_c_9.0.030 or higher
    ❍ Intel® C++ Compiler 8.1, for Linux* with Package ID l_cce_pc_8.1.036 or higher
    ❍ Intel® Fortran Compiler 9.1, for Linux* with Package ID l_fc_p_9.1.032 or higher
    ❍ Intel® Fortran Compiler 9.0, for Linux* with Package ID l_fc_c_9.0.031 or higher
    ❍ Intel® Fortran Compiler 8.1, for Linux* with Package ID l_fce_pc_8.1.036 or higher

21. **Symptom**: When booting multiple compute nodes, IP addresses are assigned randomly.

    Solution: Consider using Platform's Add-Hosts tool, which populates the Platform OCS database from an XML file, assigning the IP addresses you specify. For more information, see the Add-Hosts section of the *Readme for Platform Roll*.

Alternately, do not start multiple compute nodes simultaneously. To provide enough time for the IP addresses to be assigned in the correct order, start the nodes as follows:

1. Start the first node.
2. When the first node reaches the partitioning stage, start the second node.
3. Continue with the rest of the nodes--booting each node when the previous node reaches the partitioning stage.

22. **Symptom**: Cisco Topspin Infiniband drivers on all RHEL4 EM64T versions give the following warning message in lspci and when the self diagnostic script `hca_self_test` is run:

```
pcilib: Resource 5 in
/sys/bus/pci/devices/0000:00:1f.1/resource has a 64-bit
address, ignoring
```

Explanation: This is merely a warning message that will be addressed in future Red Hat updates and does not affect functionality or performance.

23. **Symptom**: If you change the `root` password for the operating system superuser after installing Platform OCS, you cannot log into the MySQL database with the new root password.

Explanation: When you install Platform OCS, the MySQL database installation is initially owned by the root user account. After installation, this becomes the MySQL `root` account. If you change the operating system superuser password, this has no impact on the login password of the MySQL `root` account. The original password you used when you installed Platform OCS is the valid MySQL `root` password.

24. **Symptom**: When running Linpack, a NULL output is returned with an mpirun message suggesting that the memory allocation should be increased. For example:

```
...
You can increase the amount of memory by setting the environment variable
P4_GLOBMEMSIZE (in bytes); the current size is 4194304
...
p4_error: alloc_p4_msg failed:0
...
```

Solution: Increase the value of the Linpack environment variable, `P4_GLOBMEMSIZE`. Choose an appropriate value (in bytes) based on your problem size. This value should be larger than the amount of memory requested by the problem.

To set the value of `P4_GLOBMEMSIZE`:

Add the following line to the `.bashrc` file in the home directory of the user running the parallel jobs:

`export P4_GLOBMEMSIZE=``<memsize>`

After increasing the value of the variable, make sure your free memory is as large as the value to which you set `P4_GLOBMEMSIZE`.

You can run `cleanipcs` to clean up any leftover semaphore arrays and shared memory segments from previously-running MPI jobs, as follows:

`% /opt/mpich/gnu/sbin/cleanipcs`

Note that you should run `cleanipcs` as a user rather than as root.
Also, `cleanipcs` by default is not set to execute. You will need to change the file permissions to make it executable.

---

25. **Symptom**: `mpi` application still launches in `ssh` after enabling `rsh`

    You enabled `rsh` as follows:

    1. Enabled `rsh` on the frontend node and the compute nodes.
    2. Set `RSHCOMMAND="rsh"` in `/opt/mpich/gnu/bin/mpirun`.

    However, the `mpi` application is still launched in `ssh`.

    Explanation: An additional step is required to enable `rsh`.

    Solution: In `/opt/mpich/gnu/bin/mpirun.ch_p4.args`, set `setrshcmd="yes"`.

    We do not recommend rsh as it has known security flaws. We suggest you use ssh.

---

26. **Symptom**: If a new user is added to a cluster, and additional hosts are added to the cluster afterwards, the OCS 411 feature might not propagate the new user's information to the newly added hosts.

    Solution: To synchronize the user information on all compute nodes, run:

```
# cluster-fork /opt/rocks/bin/411get --all
```

27. **Symptom**: When restarting daemons, the connection is refused. This occurs after running
    - ○ `lsadmin limstartup all`
    - ○ `lsadmin resstartup all`
    - ○ `badmin hstartup all`

    Solution: In `lsf.conf`, set `LSF_RSH=ssh`.

28. **Symptom**: When the SGE6 roll is installed and you install compute nodes, the compute nodes continuously restart.

    Solution: Stop `insert-ethers` on the frontend (F9 or F11 to exit). If `insert-ethers` is not running, try running **`insert-ethers --update`**.

29. **Symptom**: The `mpd` daemons are not started.

    Explanation: The hpc-ganglia RPM package is not installed by default on both frontend and compute nodes. This package contains `mpd`-related scripts that start up the `mpd` daemons. This means that the `mpd` daemons are not started by default.

30. **Symptom**: If a user installs a frontend from a central server and installs rolls from a CD, the Platform OCS installer (Anaconda) will fail with the following message:

    ```
    Unable to read header list. This may be due to a missing
    file or bad media. Press <return> to try again
    ```

    In the Alt-F3 virtual console, the following message is repeated 10 times:

    ```
    HTTPError: http://127.0.0.1/mnt/cdrom/RedHat/base/hdlist
    occurred getting HTTP
    Error 404: Not Found
    ```

    Explanation: The Platform OCS installer does not handle this case properly. The rolls downloaded from the central server are deleted when the rolls on the CD are copied over.

    Solution: As a workaround, copy the rolls you want to install via CD onto the central server. After you install your frontend, install those rolls from the central server.

    1. Insert the roll CD and mount it as /mnt/cdrom

2. Copy the desired roll from the CD

```
# rocks-dist --with-roll=<roll name> copyroll
```

3. Repeat with each roll that you want to copy from the CD
4. Unmount the CD.

```
# umount /mnt/cdrom
```

5. Repeat with any other roll CDs
6. Rebuild the Platform OCS distribution:

```
# cd /home/install
# rocks-dist dist
```

31. **Symptom**: mpirun on a single physical node with number of processes (np) greater than 5 results in "p4_error: semget failed for setnum" or "Killed by signal 2" error.

    Explanation: Mpirun allocates a semaphore array to each process for inter process communication. When a job is launched on a single node with number of processes greater than 5, the system limit for number of semaphore arrays is exceeded and the job exits with an error. The number of available arrays on a system can be obtained using the 'ipcs -l' command.

    Solution: Specify the number of processes per node (or number of cores per node) in the machine file that is passed to mpirun in the <node>:<number of processes> format. Example: compute-0-0:8 With this mpirun will only allocate one shared memory segment on compute-0-0 that is shared by all 8 MPI processes created as opposed to having each MPI process own a separate shared memory segment.

[Top]

Learn About Platform Products

World Wide Web

The latest information about all supported releases of any Platform product is available on the Platform Web site at http://www.platform.com. Forums, e-support, documentation and release information is available at http://my.platform.com/.

If you have problems accessing the Platform Web site or the Platform FTP site, contact support@platform.com.

Platform training

Platform's Professional Services training courses can help you gain the skills necessary to effectively install, configure and manage your Platform products. Courses are available for both new and experienced users and administrators at our corporate headquarters and Platform locations worldwide. Customized on-site course delivery is also available.

Find out more about Platform Training at http://www.platform.com/Services/Training, or contact trng@platform.com for details.

Technical support

Contact Platform Computing at support@platform.com for technical support. When contacting Platform, please include the full name of your company.

See the Platform Web site at http://www.platform.com/Company/Contact.Us.htm for other contact information.

Platform documentation

Roll readmes are packaged with the rolls, and are accessible from the Roll Call page in your frontend node: http://localhost/homepage/list-rolls.cgi.

[Top]

Get Technical Support

Contact Platform

Contact Platform Computing or your Platform OCS vendor for technical support. Use one of the following to contact Platform technical support:

Email

support@platform.com

World Wide Web

http://www.platform.com

Mail

Platform Support
Platform Computing Corporation
3760 14th Avenue
Markham, Ontario
Canada L3R 3T7

When contacting Platform, please include the full name of your company.

See the Platform Web site at http://www.platform.com/Company/Contact.Us.htm for other contact information.

Get patch updates and hotfixes

Obtain the latest patches and hotfixes for Platform OCS from the following page:

http://my.platform.com/products/platform-ocs

To obtain a user name and password, contact Platform Computing technical support at support@platform.com

[Top]

Copyright and Trademarks

IN NO EVENT WILL PLATFORM COMPUTING BE LIABLE TO ANYONE FOR SPECIAL, COLLATERAL, INCIDENTAL, OR CONSEQUENTIAL DAMAGES, INCLUDING WITHOUT LIMITATION ANY LOST PROFITS, DATA, OR SAVINGS, ARISING OUT OF THE USE OF OR INABILITY TO USE THIS PROGRAM.

Trademarks

This product includes software developed by the Rocks Cluster Group at the San Diego Supercomputer Center at the University of California, San Diego and its contributors.

® LSF and LSF HPC are trademarks or registered trademark of Platform Computing Corporation in the United States and in other jurisdictions.

TM ACCELERATING INTELLIGENCE, PLATFORM COMPUTING, and the PLATFORM and Platform OCS logos are trademarks of Platform Computing Corporation in the United States and in other jurisdictions.

® Linux is the registered trademark of Linus Torvalds in the U.S. and other countries.

Sun® Grid Engine (SGE) is an open source tool developed and maintained by Sun Microsystems, Inc.

® Sun is a registered trademark of Sun Microsystems, Inc. in the United States and other countries.

® Intel and Itanium are registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

® SGI and Altix are registered trademarks of SGI in the United States and other countries.

® Cisco Systems and Topspin are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the U.S. and in other countries

® Myrinet and Myricom are registered trademarks of Myricom, Inc. in the United States and other countries.

®TM Dell and Dell OpenManage are service marks or registered trademarks of Dell Inc. in the United States and in other countries.

Other products or services mentioned in this document are identified by the trademarks or service marks of their respective owners.

[Top]