Release Notes

Scyld ClusterWare Release 6.9.1-691g0000

About This Release

Scyld ClusterWare Release 6.9.1-691g0000 (released May 3, 2017) is the latest update to Scyld ClusterWare 6.

Scyld ClusterWare 6.9.1 expects to execute in a Red Hat RHEL6 Update 9 or CentOS 6.9 base distribution environment, each having been updated to the latest RHEL6/CentOS6 errata (https://rhn.redhat.com/errata/rhel-server-6-errata.html) as of the Scyld ClusterWare 6.9.1 release date. Any compatibility issues between Scyld ClusterWare 6.9.1 and RHEL6 are documented on the Penguin Computing Support Portal at http://www.penguincomputing.com/support.

Visit https://docs.redhat.com/docs/en-US/Red_Hat_Enterprise_Linux to view the Red Hat Enterprise Linux 6 *6.9 Release Notes* and other useful documents.

For the most up-to-date product documentation and other helpful information, visit the Penguin Computing Support Portal.

Important

Before continuing, make sure you are reading the most recent Scyld ClusterWare *Release Notes*, which can be found on the Penguin Computing Support Portal at http://www.penguincomputing.com/support/documentation. The most recent version will accurately reflect the current state of the Scyld ClusterWare yum repository of rpms that you are about to install. You may consult the *Installation Guide* for its more generic and expansive details about the installation process. The *Release Notes* document more specifically describes how to upgrade an earlier version of Scyld ClusterWare to Scyld ClusterWare 6.9.1 (see the Section called *Upgrading An Earlier Release of Scyld ClusterWare 6 to 6.9*), or how to install Scyld ClusterWare 6.9.1 as a fresh install (see the Section called *First Installation of Scyld ClusterWare 6 On A Server*).

Important for clusters using 3rd-party drivers or applications

Before installing or updating Scyld ClusterWare, if your cluster uses any 3rd-party drivers (e.g., Ethernet, InfiniBand, GPU, parallel storage) and if an install or update includes a new kernel, then verify that those 3rd-party drivers can be rebuilt or relinked to the new kernel. If an install or update involves upgrading to a new RHEL6 or CentOS6 base distribution, then verify that your cluster's 3rd-party applications are all supported by that new base distribution.

Important for clusters using Panasas storage

If the cluster uses Panasas storage, then you must ensure that a Panasas kernel module is available that matches the Scyld ClusterWare kernel you are about to install: 2.6.32-696.1.1.el6.691g0000. Login to your Panasas account at http://www.my.panasas.com, click on the *Downloads* tab, then click on *DirectFLOW Client for Linux* and then on *Search DirectFLOW Release*, and do a *Keyword* search for `691g0000`. If you find a Panasas rpm matching the to-be-installed 2.6.32-696.1.1.el6.691g0000 kernel, then download that rpm and continue with the Scyld ClusterWare update or install. Install the Panasas rpm after you finish installing the associated 2.6.32-696.1.1.el6.691g0000 kernel. If you do not find an appropriate Panasas rpm, then do not install this latest Scyld ClusterWare 6.9.1. The Panasas storage will not work with the 2.6.32-696.1.1.el6.691g0000 kernel without a matching Panasas kernel module.

First Installation of Scyld ClusterWare 6 On A Server

When installing Scyld ClusterWare 6 on a system that does not yet contain Scyld ClusterWare, you should perform the following steps:

1. The directory `/etc/yum.repos.d/` must contain active repo config files bearing a suffix of `.repo`. If there is no ClusterWare repo file, then you should download `clusterware.repo` that gives your cluster access to the customer-facing Scyld ClusterWare yum repos.

   To download a yum repo file that is customized to your cluster:

   a. Login to the Penguin Computing Support Portal at http://www.penguincomputing.com/support.

   b. Click on the tab labeled *Assets*, and then select a specific *Asset Name* in the list.

   c. In the *Asset Detail* section, click on *YUM Repo File*, which downloads an asset-specific `clusterware.repo` file, and move that file to the `/etc/yum.repos.d/` directory.

   d. Set the permissions: `chmod 644 /etc/yum.repos.d/clusterware.repo`

   e. The new `clusterware.repo` contains a *baseurl* entry that uses `https` by default. If your local site is configured to not support such encrypted accesses, then you must edit the repo file to instead use `http`.

   The file contains three sections, labeled *cw-core*, *cw-updates*, and *cw-next*. Generally, the *cw-next* repo should not be enabled unless so directed by Penguin Computing Support.

2. Examine `/etc/yum.repos.d/clusterware.repo` to ensure that it specifies the desired yum repository release version. Employ *$releasever* or *6* to use rpms from the latest Scyld ClusterWare release, which currently is 6.9. Alternatively, a more specific major-minor pair, e.g., *6.2*, limits the rpms to just that version, even as ClusterWare releases march forward to newer versions.

3. If updating using a Red Hat yum repo, then your Red Hat yum configuration file should also look in the Red Hat Server Optional repo to find rpms such as `compat-dapl-devel` and `sharutils`. The regular CentOS6 yum repo contains these rpms.

4. Install a useful Scyld ClusterWare script that simplifies installing (and later updating) software, then execute that script:

   ```
   yum install install-scyld
   install-scyld
   ```

5. **If the cluster uses Panasas storage**, then you should have already downloaded the Panasas rpm that matches the Scyld ClusterWare 6 kernel you have just installed. Now install the Panasas rpm using **rpm -i**.

6. Configure the network for Scyld ClusterWare: run **beonetconf**, or edit `/etc/beowulf/config`, to specify the cluster interface, the maximum number of compute nodes, and the beginning IP address of the first compute node. See the *Installation Guide* for more details.

7. If the private cluster network switch uses Spanning Tree Protocol (STP), then either reconfigure the switch to disable STP, or if that is not feasible because of network topology, then enable *Rapid STP* or *portfast* on the compute node and edge ports. See the Section called *Issues with Spanning Tree Protocol and portfast* for details.

8. Reboot the master node.

9. After rebooting the new kernel, and after installing any new kernel modules, you should rebuild the master node's list of modules and dependencies using **depmod**. See the Section called *Issues with kernel modules* for details.

Upgrading An Earlier Release of Scyld ClusterWare 6 to 6.9

If you wish to upgrade a RHEL5 (or CentOS5) or earlier base distribution to RHEL6/CentOS6, then we recommend you accomplish this with a full install of Release 6, rather than attempt to *update* from an earlier major release to Release 6. Visit https://docs.redhat.com/docs/en-US/Red_Hat_Enterprise_Linux for the Red Hat Enterprise Linux 6 *Installation Guide* for details. If you already have installed Scyld ClusterWare 5 (or earlier) on the physical hardware that you intend to convert to RHEL6/CentOS6, then we recommend that you backup your master node prior to the new installation of

RHEL6/CentOS6, as some of the Scyld ClusterWare configuration files may be a useful reference for Release 6, especially files in `/etc/beowulf/`.

When upgrading from an earlier Scyld ClusterWare 6 version to a newer Scyld ClusterWare 6, you should perform the following steps:

1. Examine `/etc/yum.repos.d/clusterware.repo` to ensure that it specifies the desired yum repository release version. Employ *$releasever* or *6* to use rpms from the latest Scyld ClusterWare release, which currently is 6.9. Alternatively, a more specific major-minor pair, e.g., *6.2*, limits the rpms to just that version, even as ClusterWare releases march forward to newer versions.
2. Consider whether or not to stop the cluster prior to updating software. Most updates can be made to a running cluster, although some updates (e.g., those affecting daemons that execute on the master node) require a subsequent restart of the ClusterWare service. Other updates require rebooting the master node, in particular when updating to a new kernel, and this obviously restarts the cluster nodes, too. The safest approach is to stop the cluster before updating the master node, and restart the cluster after the update completes.

   ```
   service beowulf stop
   ```
3. Update the software on the master node using the `install-scyld` script that guides you through the process, step by step. If this script doesn't exist on your system, then install it.

   ```
   yum install install-scyld    # if not already installed
   install-scyld -u
   ```
   The script first determines if it needs to update itself. If that self-update occurs, then the script exits and you should re-execute it.
4. **If the cluster uses Panasas storage**, then you should have already downloaded the Panasas rpm that matches the Scyld ClusterWare 6.9.1 kernel you have just installed. Now install the Panasas rpm using **rpm -i**.
5. Compare `/etc/beowulf/config`, which remains untouched by the Scyld ClusterWare update, with the new `config.rpmnew` (if that file exists), examine the differences:

   ```
   cd /etc/beowulf
   diff config config.rpmnew
   ```

   and carefully merge the `config.rpmnew` differences into `/etc/beowulf/config`. See the Section called *Resolve *.rpmnew and *.rpmsave configuration file differences* for details.

   Similarly, the preexisting `/etc/beowulf/fstab` may have been saved as `fstab.rpmsave` if it was locally modified. If so, merge those local changes back into `/etc/beowulf/fstab`.
6. If a new kernel has been installed, then reboot the master node. Otherwise, simply reboot the ClusterWare service:

   ```
   service beowulf restart
   ```
7. After rebooting a new kernel, and after installing any new kernel modules, you should rebuild the master node's list of modules and dependencies using **depmod**. See the Section called *Issues with kernel modules* for details.

Post-Installation Configuration Issues

Following a successful update or install of Scyld ClusterWare, you may need to make one or more configuration changes, depending upon the local requirements of your cluster. Larger cluster configurations have additional issues to consider; see the Section called *Post-Installation Configuration Issues For Large Clusters*.

Resolve **.rpmnew* and **.rpmsave* configuration file differences

As with every Scyld ClusterWare upgrade, after the upgrade you should locate any Scyld ClusterWare `*.rpmsave` and `*.rpmnew` files and perform merges, as appropriate, to carry forward the local changes. Sometimes an upgrade will save the locally modified version as `*.rpmsave` and overwrite the basic file with a new version. Other times the upgrade will keep the locally modified version untouched, installing the new version as `*.rpmnew`.

For example,

```
cd /etc/beowulf
find . -name \*rpmnew
find . -name \*rpmsave
```

and examine each such file to understand how it differs from the configuration file that existed prior to the update. You may need to merge new lines from the newer `*.rpmnew` file into the existing file, or perhaps replace existing lines with new modifications. For instance, this is commonly done with `/etc/beowulf/config` and `config.rpmnew`. Or you may need to merge older local modifications in `*.rpmsave` into the newly installed pristine version of the file. For instance, this is occasionally done with `/etc/beowulf/fstab.rpmsave`.

Generally speaking, be careful when making changes to `/etc/beowulf/config`, as mistakes may leave your cluster in a non-working state. In particular, take care when modifying the keyword entries for *interface*, *nodes*, *iprange*, and *nodeassign*. Those may be more safely manipulated by the **beonetconf** command. The *kernelimage* and *node* entries are automatically managed by ClusterWare services and should not be merged.

The remaining differences are candidates for careful merging. Pay special attention to merge additions to the *bootmodule*, *modarg*, *server*, *libraries*, and *prestage* keyword entries. New *nodename* entries for *infiniband* or *ipmi* are offsets to each node's IP address on the private cluster network, and these offsets may need to be altered to be compatible with your local network subnet. Also, be sure to merge differences in `config.rpmnew` comments, as those are important documentation information for future reference.

Contact Scyld Customer Support if you are unsure about how to resolve particular differences, especially with `/etc/beowulf/config`.

Disable SELinux and NetworkManager

Scyld ClusterWare execution currently requires that SELinux and NetworkManager services be disabled. The **install-scyld** script performs this disabling.

Disable library prelinking

Scyld ClusterWare migration between cluster nodes requires stable dynamic libraries. Edit `/etc/sysconfig/prelink` and ensure that *PRELINKING=no* is set. This will permanently block subsequent (usually daily) **prelink** operations. In addition, to immediately undo prelinking:

```
prelink --undo -all
```

See the *Administrator's Guide* for more details.

Optionally reduce size of /usr/lib/locale/locale-archive

Glibc applications silently open the file `/usr/lib/locale/locale-archive`, which means it gets downloaded by each compute node early in a node's startup sequence. The default RHEL6 `locale-archive` is about 100 MBytes in size,

thus consuming significant network bandwidth and potentially causing serialization delays if numerous compute nodes attempt to concurrently boot, and consuming significant RAM filesystem space on each node. It is likely that a cluster's users and applications do not require all the international locale data that is present in the default file. With care, the cluster administrator may choose to rebuild `locale-archive` with a greatly reduced set of locales and thus create a significantly smaller file. See the *Administrator's Guide* for details.

Optionally configure and enable compute node CPU speed/power management

Modern motherboards and processors support a degree of administrator management of CPU frequency within a range defined by the motherboard's BIOS. Scyld ClusterWare provides the `/etc/beowulf/init.d/30cpuspeed` script and its associated `/etc/beowulf/conf.d/cpuspeed.conf` configuration file to implement this management for compute nodes. The local cluster administrator is encouraged to review the *Administrator's Guide*'s *Configuring CPU speed/power for Compute Nodes* for details.

Optionally install a different TORQUE package

TORQUE is available in several versions: `torque-4-scyld` (which is the current default) and `torque-4-nocpuset-scyld` provide version 4, `torque-5-scyld` and `torque-5-nocpuset-scyld` provide version 5, and `torque-6-scyld` and `torque-6-nocpuset-scyld` provide version 6.

The `nocpuset` packages specifically disable the default *cpuset* functionality that optionally allows an application to constrain the movement of software threads between CPUs within a node in order to achieve optimal performance. See http://docs.adaptivecomputing.com/torque/4-1-4/help.htm#topics/3-nodes/linuxCpusetSupport.htm for details.

One, and only one, TORQUE must be installed at any one time. Since each TORQUE package specifies a list of package dependencies that should not be removed when uninstalling the existing TORQUE package, care must be taken to retain those dependencies when switching from one version of TORQUE to another. For example, to switch from `torque-4-scyld` to `torque-4-nocpuset-scyld`:

```
rpm -e --nodeps torque-4-scyld
yum install torque-4-nocpuset-scyld
```

Optionally enable job manager

The default Scyld ClusterWare installation includes two job managers: TORQUE and Slurm. TORQUE is available in several versions. See the Section called *Optionally install a different TORQUE package* for important details. Both Slurm and one, and only one, of these TORQUE versions must be installed on the master node, although only Slurm *or* one of the TORQUE versions may be enabled and executing at any one time.

To enable TORQUE, then after all compute nodes are up and running, you must first disable SLURM, then enable and configure TORQUE, then reboot all the compute nodes:

```
service slurm-scyld cluster-stop
chkconfig slurm-scyld off
beochkconfig 98slurm off
chkconfig torque on
beochkconfig 98torque on
service torque reconfigure
service torque start
bpctl -S all -R
```

and then after the compute nodes have rebooted, restart TORQUE cluster-wide:

```
service torque cluster-restart
```

To enable Slurm, then after all compute nodes are up and running, you must first disable TORQUE, then enable and configure Slurm, then reboot all the compute nodes:

```
service torque cluster-stop
chkconfig torque off
beochkconfig 98torque off
chkconfig slurm-scyld on
beochkconfig 98slurm on
```

Next, configure Slurm by generating `/etc/slurm/slurm.conf` and `/etc/slurm/slurmdbd.conf` from Scyld-provided templates:

```
service slurm-scyld reconfigure
```

Finally, start Slurm on the master node and reboot all compute nodes:

```
service slurm-scyld start
bpctl -S all -R
```

and then after the compute nodes have rebooted, restart Slurm cluster-wide:

```
service slurm-scyld cluster-restart
```

See the *Administrator's Guide* for more details about TORQUE configuration, and the *User's Guide* for details about how to use TORQUE.

Each Slurm user must setup the PATH and LD_LIBRARY_PATH environment variables to properly access the Slurm commands. This is done automatically for users who login when the *slurm* service is running and the *pbs_server* is not running, via the `/etc/profile.d/scyld.slurm.sh` script. Alternatively, each Slurm user can manually execute **module load slurm** or can add that command line to (for example) the user's `.bash_profile`.

See the *Administrator's Guide* for more details about TORQUE and Slurm configuration.

Optionally enable TORQUE scheduler

Scyld ClusterWare installs by default both the TORQUE resource manager and the associated Maui job scheduler. The Maui installation can coexist with an optionally licensed Moab job scheduler installation, although after the initial installation of either of these job schedulers, the cluster administrator needs to make a one-time choice of which job scheduler to employ.

If Moab is not installed, and if TORQUE is enabled as the operative job manager (see the Section called *Optionally enable job manager*), then simply activate Maui by moving into place two global profile files that execute **module load maui** and then start the *maui* service:

```
cp /opt/scyld/maui/scyld.maui.{csh,sh} /etc/profile.d
chkconfig maui on
service maui start
```

If Moab was previously installed, is currently active, and is the preferred job scheduler, then the cluster administrator can ignore the Maui installation (and any subsequent Maui updates) because Maui installs in a deactivated state and will not affect Moab.

If Maui is active and the cluster administrator subsequently installs Moab, or chooses to use an already installed Moab as the default scheduler, then deactivate Maui so as to not affect Moab:

```
rm /etc/profile.d/scyld.maui.*
chkconfig maui off
service maui stop
```

and then activate Moab as appropriate for the cluster.

Optionally enable Ganglia monitoring tool

To enable the Ganglia cluster monitoring tool,

```
chkconfig beostat on
chkconfig xinetd on
chkconfig httpd on
chkconfig gmetad on
```

then either reboot the master node, which automatically restarts these system services; or without rebooting, manually restart *xinetd* then start the remaining services that are not already running:

```
service xinetd restart
service httpd start
service gmetad start
```

See the *Administrator's Guide* for more details.

Optionally enable beoweb service

The beoweb service facilitates remote job submission and cluster monitoring (e.g., used by POD Tools). Beoweb version 2.0+ requires that the scyld-lmx license manager service be executing and able to access a valid license file at `/opt/scyld/scyld-lmx/scyld.lic`. If this file does not exist, then send your master node's MAC address to Penguin Computing Support to obtain a free license file.

When the license file is in place, start the scyld-lmx license manager, and enable and start beoweb:

```
/etc/init.d/scyld-lmx start
chkconfig beoweb on
service beoweb start
```

See the *Administrator's Guide* for more details.

Optionally enable NFS locking

If you wish to use cluster-wide NFS locking, then you must enable locking on the master node and on the compute nodes. First ensure that NFS locking is enabled and running on the master:

```
chkconfig nfslock on
service nfslock start
```

Then for each NFS mount point for which you need the locking functionality, you must edit `/etc/beowulf/fstab` (or the appropriate node-specific `/etc/beowulf/fstab.`*N* file(s)) to remove the default option *nolock* for that mountpoint. See the *Administrator's Guide* for more details.

Optionally adjust the size limit for locked memory

OpenIB, MVAPICH, and MVAPICH2 require an override to the limit of how much memory can be locked.

Scyld ClusterWare adds a *memlock* override entry to `/etc/security/limits.conf` during a Scyld ClusterWare upgrade (if the override entry does not already exist in that file), regardless of whether or not Infiniband is present in the cluster. The new override line,

```
*    -    memlock    unlimited
```

raises the limit to *unlimited*. If Infiniband is not present, then this new override line is unnecessary and may be deleted. If Infiniband is present, we recommend leaving the new *unlimited* line in place. If you choose to experiment with a smaller discrete value, then understand that Scyld ClusterWare MVAPICH requires a minimum of 16,384 KBytes, which means changing *unlimited* to *16384*. If your new discrete value is too small, then MVAPICH reports a "CQ Creation" or "QP Creation" error.

Optionally increase the max number of processes per user

RHEL6 defaults to a maximum of 1024 processes per user, as specified in `/etc/security/limits.d/90-nproc.conf`, which contrasts with the RHEL5 default of 16,384. If this RHEL6 value is too low, then override the *nproc* entry in that file, as appropriate for your cluster workload needs. Use a discrete value, not *unlimited*.

Optionally enable SSHD on compute nodes

If you wish to allow users to execute MVAPICH2 applications, or to use **/usr/bin/ssh** or **/usr/bin/scp** from the master to a compute node, or from one compute node to another compute node, then you must enable **sshd** on compute nodes by enabling the script:

```
beochkconfig 81sshd on
```

The cluster is preconfigured to allow user *root* ssh access to compute nodes. The cluster administrator may wish to configure the cluster to allow ssh access for non-root users. See the *Administrator's Guide* for details.

Optionally allow IP Forwarding

By default, the master node does not allow IP Forwarding from compute nodes on the private cluster network to external IP addresses on the public network. If IP Forwarding is desired, then edit `/etc/beowulf/config` to enable the directive *ipforward yes*, and ensure that the file `/etc/sysconfig/iptables` eliminates or comments-out the default entry:

```
-A FORWARD -j REJECT --reject-with icmp-host-prohibited
```

Optionally increase the nf_conntrack table size

Certain workloads may trigger a syslog message *nf_conntrack: table full, dropping packet*. At cluster startup, Scyld ClusterWare insures a NAT table max size of at least 524,288. However, this max value may still be inadequate for local workloads, and the *table full, dropping packet* syslog messages may still occur. Use:

```
sysctl -n net.nf_conntrack_max
```

to view the current max size, then keep manually increasing the max until the syslog messages stop occurring, e.g., use:

```
sysctl -w net.nf_conntrack_max=Nmax
```

to try new *Nmax* values. Make this value persist across master node reboots by adding:

```
net.nf_conntrack_max=Nmax
```

to `/etc/sysctl.conf`.

Optionally configure vm.zone_reclaim_mode on compute nodes

Because Scyld ClusterWare compute nodes are predominantly used for High Performance Computing, versus (for example) used as file servers, we suggest that the `/etc/beowulf/conf.d/sysctl.conf` file contain the line:

```
vm.zone_reclaim_mode=1
```

for optimal NUMA performance. Scyld ClusterWare's `node_up` script adds this line if it doesn't already exist, but will not alter an existing *vm.zone_reclaim_mode* declaration in that file. If the file `/etc/beowulf/conf.d/sysctl.conf` does not exist, then `node_up` creates it by replicating the master node's `/etc/sysctl.conf`, which may contain a *vm.zone_reclaim_mode=N* declaration that is perhaps not *=1* and thus not optimal for compute nodes, even if the value is optimal for the master node. In this case, the cluster administrator should consider manually editing `/etc/beowulf/conf.d/sysctl.conf` to change the line to *vm.zone_reclaim_mode=1*.

Optionally configure automount on compute nodes

If you wish to run automount from compute nodes, you must first set up all the necessary configuration files in `/etc/beowulf/conf.d/autofs/` before enabling the `/etc/beowulf/init.d/50autofs` script. These config files are similar to those normally found on a server in `/etc/`, such as `/etc/auto.master`, as the `50autofs` script copies the files in `/etc/beowulf/conf.d/autofs/` to each compute node's `/etc/`.

A default `/etc/beowulf/conf.d/autofs/auto.master` must exist. All automount config files that are listed in that `master.conf`, such as `/etc/auto.misc`, `/etc/auto.net`, etc., should also reside in `/etc/beowulf/conf.d/autofs/`.

Node-specific config files (`auto.master` and related `auto.*`) may reside in `/etc/beowulf/conf.d/autofs/$NODE/`. Those files override the default top level `/etc/beowulf/conf.d/auto.master`, etc., for the specific $NODE.

The `50autofs` script parses the config files as mentioned above. It creates mount point directories, installs the autofs4 kernel module, and starts **automount** on each booting compute node. The script exits with a warning if there are missing config files.

NOTE: This script does *not* validate the correctness of potential future automount mount requests (i.e., those described in the various `auto.*` config files). The cluster administrator should set up the config files, then enable `50autofs` and reboot one or a limited number of nodes and ensure that each potential automount will function properly prior to rebooting all

compute nodes. Common failures include naming an unknown server or attempting to mount a directory that has not been properly exported by the server. Mount failures will be syslogged in `/var/log/messages`.

Optionally reconfigure node names

You may declare site-specific alternative node names for cluster nodes by adding entries to `/etc/beowulf/config`. The syntax for a node name entry is:

```
nodename format-string [IPv4offset] [netgroup]
```

For example,

```
nodename node%N
```

allows the user to refer to node 4 using the traditional *.4* name, or alternatively using names like *node4* or *node004*. See **man beowulf-config** and the *Administrator's Guide* for details.

Post-Installation Configuration Issues For Large Clusters

Larger clusters have additional issues that may require post-installation adjustments.

Optionally increase the number of nfsd threads

The default count of 8 **nfsd** NFS daemons may be insufficient for large clusters. One symptom of an insufficiency is a syslog message, most commonly seen when you currently boot all the cluster nodes:

```
nfsd: too many open TCP sockets, consider increasing the number of nfsd threads
```

Scyld ClusterWare automatically increases the nfsd thread count to at least one thread per compute node, with a lowerbound of eight (for <=8 nodes) and an upperbound of 64 (for >=64 nodes). If this increase is insufficient, then increase the thread count (e.g., to 16) by executing:

```
echo 16 > /proc/fs/nfsd/threads
```

Ideally, the chosen thread count should be sufficient to eliminate the syslog complaints, but not significantly higher, as that would unnecessarily consume system resources. One approach is to repeatedly double the thread count until the syslog error messages stop occurring, then make the satisfactory value *N* persistent across master node reboots by creating the file `/etc/sysconfig/nfs`, if it does not already exist, and adding to it an entry of the form:

```
RPCNFSDCOUNT=N
```

A value *N* of 1.5x to 2x the number of nodes is probably adequate, although perhaps excessive. See the *Administrator's Guide* for a more detailed discussion of NFS configuration.

Optionally increase the max number of processID values

The kernel defaults to using a maximum of 32,768 processID values. Scyld ClusterWare automatically increases this default to 98,304 [= 3*32768], which likely is adequate for small- to medium-size clusters and which keeps pid values at a familiar 5-column width maximum. Because BProc manages a common process space across the cluster, even the increase to 98,304

may be insufficient for very large clusters and/or workloads that create large numbers of concurrent processes. The cluster administrator can increase the value further by using the **sysctl** command, e.g.,

```
sysctl -w kernel.pid_max=N
```

directs the kernel to use pid values up to *N*. The kernel (and BProc) supports an upperbound of 4,194,304 [= (4*1024*1024)]. To set a value *N* that persists across master node reboots, add an entry

```
kernel.pid_max=N
```

to `/etc/sysctl.conf`. NOTE: Even though `/etc/beowulf/conf.d/sysctl.conf` is referenced by the **sysctl** command that executes at boot time on each node, any `kernel.pid_max` entry in that file is ignored. The master node's `kernel.pid_max` value prevails cluster-wide for Scyld nodes.

Optionally increase the max number of open files

RHEL6 defaults to a maximum of 1024 concurrently open files. This value may be too low for large clusters. The cluster administrator can add a *nofile* override entry to `/etc/security/limits.conf` to specify a larger value. Caution: for *nofile*, use only a numeric upperbound value, never *unlimited*, as that will result in being unable to login.

Issues with Ganglia

The Ganglia cluster monitoring tool may fail for large clusters. If the `/var/log/httpd/error_log` shows a fatal error of the form *PHP Fatal error: Allowed memory size of 8388608 bytes exhausted*, then edit the file `/etc/php.ini` to increase the *memory_limit* parameter. The default is *memory_limit = 8M* can be safely doubled and re-doubled until the error goes away.

Post-Installation Release of Updated Packages

From time to time, Penguin Computing releases updated Scyld ClusterWare 6 rpms to track Red Hat kernel security or bug fix errata, or to fix Scyld ClusterWare problems or to introduce enhancements. Download the latest version of the Scyld ClusterWare 6 *Release Notes* from the Penguin Computing Support Portal (http://www.penguincomputing.com/services-support/documentation/) to ensure you have the latest guidance before updating your cluster.

First check for the availability of updated rpms:

```
yum check-update
```

and ascertain if the base distribution and/or Scyld ClusterWare would update to a newer kernel, or even more significantly to a new major-minor release. Upgrading the kernel will require updating, perhaps even rebuilding, any 3rd-party drivers that are installed and linked against the current kernel, and you should be prepared to do that if you proceed with the updates. Updating to a newer major-minor release may also affect 3rd-party applications that are validated only for the current base distribution release.

In general, if you choose to update software, then you should use:

```
install-scyld -u
```

and update all available packages.

If your cluster uses Panasas storage, then before updating Scyld ClusterWare you must ensure that a Panasas kernel module is available that matches the Scyld ClusterWare kernel that will be installed. See the section called *Important for clusters using Panasas storage* in the *About This Release* introduction for more information.

Notable Feature Enhancements And Bug Fixes

New in Scyld ClusterWare 6.9.1 - Scyld Release 691g0000 - May 3, 2017

1. The base kernel updates to 2.6.32-696.1.1.el6.691g0000. See https://access.redhat.com/errata/RHSA-2017:0892 for details.
2. The bproc *filecache* functionality now properly downloads files from the master node that were previously rejected because the files have restricted read access permissions. Now all files are downloaded to compute nodes - and, as always, downloaded files are given access permissions that are replicated from the master node.

New in Scyld ClusterWare 6.9.0 - Scyld Release 690g0000 - April 14, 2017

1. The base kernel updates to 2.6.32-696.el6.690g0000. See https://rhn.redhat.com/errata/RHSA-2017-0817.html for details. Scyld ClusterWare 6.9 expects to execute in a Red Hat RHEL6 Update 9 or CentOS6.9 environment.
2. TORQUE 6 updates to version 6.1.1, from www.adaptivecomputing.com/products/open-source/torque/. See www.adaptivecomputing.com/support/documentation-index/torque-resource-manager-documentation/ for details.
3. Scyld ClusterWare now distributes openmpi-2.1-scyld packages, which are initially version 2.1.0. Installation of openmpi-2.1 does not affect any earlier OpenMPI version. The libraries were built with Gnu version 4.4.7-18, Intel version 2013_sp1.3.174, and PGI version 14.6 compiler families. Scyld releases of OpenMPI derive from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.
4. Scyld ClusterWare now distributes Singularity, which is initially version 2.2.1. See the *User's Guide* for details.

New in Scyld ClusterWare 6.8.8 - Scyld Release 688g0000 - March 2, 2017

1. The base kernel updates to 2.6.32-642.15.1.el6.688g0000. See https://rhn.redhat.com/errata/RHSA-2017-0307.html for details.
2. The Slurm job manager updates to version 17.02.0, derived from http://slurm.schedmd.com. See the *User's Guide* Appendix G, *SLURM Release Information* for details.
3. The openmpi-1.10-scyld packages update to version 1.10.6, which by default update and replace only earlier 1.10.z packages and do not affect any other installed openmpi-x.y-scyld packages other than 1.10. See the Section called *Installing and managing concurrent versions of packages* for general issues about supporting multiple concurrent versions. The libraries were built with Gnu version 4.4.7-17, Intel version 2013_sp1.3.174, and PGI version 14.6 compiler families. Scyld releases of OpenMPI derive from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.

New in Scyld ClusterWare 6.8.7 - Scyld Release 687g0000 - February 10, 2017

1. The base kernel updates to 2.6.32-642.13.1.el6.687g0000. See https://rhn.redhat.com/errata/RHSA-2017-0036.html for details.
2. A new *install-scyld* package contains a script that greatly simplifies installing and updating software on the master:

   ```
   yum install install-scyld
   ```

   We strongly encourage using this script. See the Section called *First Installation of Scyld ClusterWare 6 On A Server* and the Section called *Upgrading An Earlier Release of Scyld ClusterWare 6 to 6.9* for details.
3. The igb Ethernet driver updates to version 5.3.5.4, derived from http://sourceforge.net/projects/e1000/.
4. The e1000e Ethernet driver updates to version 3.3.5.3, derived from http://sourceforge.net/projects/e1000/.
5. The Slurm job manager updates to version 16.05.8, derived from http://slurm.schedmd.com. See the *User's Guide* Appendix G, *SLURM Release Information* for details.
6. The openmpi-2.0-scyld packages update to version 2.0.2, which by default update and replace only earlier version 2.0 packages and do not affect any installed OpenMPI version 1.10 and earlier packages. See the Section called *Installing and managing concurrent versions of packages* for general issues about supporting multiple concurrent versions. The libraries were built with Gnu version 4.4.7-17,, Intel version 2013_sp1.3.174, and PGI version 14.6 compiler families. Scyld releases of OpenMPI derive from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.
7. Various scripts in `/etc/beowulf/init.d/` have been renamed with different numeric prefixes in order to adjust the execution ordering: `95sudo`, `98slurm`, and `98torque`. If any of these scripts has been copied and modified locally (see the Section called *Caution when modifying Scyld ClusterWare scripts* for details), then you should rename the local copy to match the new numeric prefix.

New in Scyld ClusterWare 6.8.6 - Scyld Release 686g0000 - December 9, 2016

1. The base kernel updates to 2.6.32-642.11.1.el6.686g0000. See https://rhn.redhat.com/errata/RHSA-2016-2766.html for details.
2. The Slurm job manager updates to version 16.05.6, derived from http://slurm.schedmd.com. See the *User's Guide* Appendix G, *SLURM Release Information* for details.
3. TORQUE version 6 updates to version 6.1.0, from www.adaptivecomputing.com/products/open-source/torque/. See www.adaptivecomputing.com/support/documentation-index/torque-resource-manager-documentation/ for details.
4. The script `/etc/beowulf/init.d/85run2complete` now supports optional `/etc/beowulf/config` overriding of the *idle_threshold* and *idle_time* values that were previously hardcoded in `85run2complete`. See the *r2c* comments in the `config` file.

New in Scyld ClusterWare 6.8.5 - Scyld Release 685g0000 - November 7, 2016

1. The base kernel updates to 2.6.32-642.6.2.el6.685g0000. See https://rhn.redhat.com/errata/RHSA-2016-2105.html for details.

   This kernel differs from the previous 2.6.32-642.6.1.el6.684g0000 only by the inclusion of a fix for the Red Hat CVE-2016-5195 ("kernel: mm: privilege escalation via MAP_PRIVATE COW breakage", aka "dirty COW") security exploit described by Red Hat Bugzilla #1384344.

New in Scyld ClusterWare 6.8.4 - Scyld Release 684g0000 - October 13, 2016

1. The base kernel updates to 2.6.32-642.6.1.el6.684g0000. See https://rhn.redhat.com/errata/RHSA-2016-2006.html for details.
2. The default `/etc/beowulf/fstab` no longer suggests mounting `/dev/cpuset` for TORQUE.
3. The *torque-scyld* and *torque-nocpuset-scyld* packages are replaced by *torque-4-scyld* and *torque-4-nocpuset-scyld* (still version 4.2.10). Also added to the Scyld ClusterWare distribution are *torque-5-scyld* and *torque-5-nocpuset-scyld* (version 5.1.3), and *torque-6-scyld* and *torque-6-nocpuset-scyld* (version 6.0.2), all from www.adaptivecomputing.com/support/download-center/torque-download/. Only one *torque* can be installed at any point in time. See the *User's Guide* Appendix B, *TORQUE and Maui Release Information* for details.

   NOTE: The first time updating from *torque-scyld* to the new packaging scheme, the cluster administrator must explicitly install one (and only one) of the *N* packages, e.g., **yum install torque-4-scyld**. That will both install the new package and remove the obsolete *torque-scyld* package. See the Section called *Issues with TORQUE* for details.
4. The Slurm job manager updates to version 16.05.5, derived from http://slurm.schedmd.com. See the *User's Guide* Appendix G, *SLURM Release Information* for details.
5. Scyld ClusterWare now distributes openmpi-2.0-scyld packages, which are initially version 2.0.1. Installation of openmpi-2.0 does not affect any earlier OpenMPI version.

   Additionally, the openmpi-1.10-scyld packages update to version 1.10.4, which by default update and replace only earlier version 1.10 packages and do not affect any installed OpenMPI version 1.8, 1.7, 1.6, or 1.5 packages. See the Section called *Installing and managing concurrent versions of packages* for general issues about supporting multiple concurrent versions. The libraries were built with Gnu version 4.4.7-17, Intel version 2013_sp1.3.174, and PGI version 14.6 compiler families. Scyld releases of OpenMPI derive from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.
6. MVAPICH2 updates to version 2.2 for the `mvapich2-psm-scyld` and `mvapich2-scyld` packages. See the Section called *Installing and managing concurrent versions of packages* for general issues about supporting multiple concurrent versions. This software suite derives from http://mvapich.cse.ohio-state.edu/. *NOTE:* MVAPICH2-2.1 introduced an algorithm to determine CPU topology on the node, and this new algorithm does not work properly for older Mellanox controllers and firmware, resulting in software threads not spreading out across a node's cores by default. See the Section called *Issues with MVAPICH2 and CPU Sets* or the *User's Guide* Appendix E, *MVAPICH2 Release Information* for details.

New in Scyld ClusterWare 6.8.3 - Scyld Release 683g0000 - September 8, 2016

1. The base kernel updates to 2.6.32-642.4.2.el6.683g0000. See https://rhn.redhat.com/errata/RHSA-2016-1664.html for details.
2. Fix another rare bproc bug that panics compute nodes with "soft lockup" or "hard lockup" messages.
3. Make additional bproc enhancements that improve the performance of multithreaded applications that concurrently execute multiple dozens of threads across multiple dozens of cores.
4. Introduce various "helper" routines in libbeoconfig.so that assist in parsing the `/etc/beowulf/config` *iprange* directive. The several consumers of that directive (*beonss*, *beoserv*, *bpmaster*) now use these helper routines to provide a consistent implementation. These changes should be transparent to users, although they serve as part of the foundation for upcoming enhancements to the *iprange* functionality and to Scyld ClusterWare's handling of very large clusters.

New in Scyld ClusterWare 6.8.2 - Scyld Release 682g0000 - July 26, 2016

1. The base kernel updates to 2.6.32-642.3.1.el6.682g0000. See https://rhn.redhat.com/errata/RHSA-2016-1406.html for details.
2. The openmpi-1.10-scyld packages update to version 1.10.3, which by default update and replace only earlier version 1.10 packages and do not affect any installed OpenMPI version 1.8, 1.7, 1.6, or 1.5 packages. See the Section called *Installing and managing concurrent versions of packages* for general issues about supporting multiple concurrent versions. The libraries were built with Gnu version 4.4.7-17, Intel version 2013_sp1.3.174, and PGI version 14.6 compiler families. Scyld releases of OpenMPI derive from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.
3. The Slurm job manager updates to version 16.05.0, derived from http://slurm.schedmd.com. See the *User's Guide* Appendix G, *SLURM Release Information* for details.
4. Eliminate a harmless error message that may be generated by the `/etc/beowulf/init.d/30cpuspeed` script.

New in Scyld ClusterWare 6.8.1 - Scyld Release 681g0000 - July 26, 2016

1. The base kernel updates to 2.6.32-642.1.1.el6.681g0000. See https://rhn.redhat.com/errata/RHBA-2016-1185.html for details.
2. Fix rare bproc bugs that panic compute nodes with "soft lockup" or "hard lockup" messages.
3. Improve bproc performance on compute nodes when handling processes with multi-gigabytes of allocated memory.

New in Scyld ClusterWare 6.8.0 - Scyld Release 680g0000 - June 3, 2016

1. The base kernel updates to 2.6.32-642.el6.680g0000. See https://rhn.redhat.com/errata/RHSA-2016-0855.html for details.
2. Supports the Intel Xeon E5-2600 "Broadwell" microarchitecture family.
3. The igb Ethernet driver updates to version 5.3.4.4, derived from http://sourceforge.net/projects/e1000/.

New in Scyld ClusterWare 6.7.7 - addendum to Release 677g0000 - November 7, 2016

1. The base kernel updates to 2.6.32-573.26.1.el6.677g0001. This kernel differs from the previous 2.6.32-573.26.1.el6.677g0000 only by the inclusion of a fix for the Red Hat Bugzilla #1384344 security exploit ("kernel: mm: privilege escalation via MAP_PRIVATE COW breakage"). This kernel, together with the matching kmod-* rpms which were built from the same source code files as they were in 677g0000, is currently only available in the Scyld ClusterWare 6.7 `updates.next` yum repo. NOTE: the matching Panasas kernel module is *not* yet available for this kernel, so do not install and use this kernel if your cluster employs Panasas storage.

   To install, ensure that the `/etc/yum.repos.d/clusterware.repo` file (or whatever the name of the ClusterWare repo file is being used) has URLs that refer to the `6.7` repo, then:

   ```
   yum --disablerepo=* --enablerepo=cw-next update
   ```

New in Scyld ClusterWare 6.7.7 - Scyld Release 677g0000 - May 31, 2016

1. The base kernel updates to 2.6.32-573.26.1.el6.677g0000. See https://rhn.redhat.com/errata/RHSA-2016-0715.html for details.
2. Introduce a new `/etc/beowulf/init.d/98entropy` script to optionally enable the **entropyd** daemon on compute nodes that adds entropy to /dev/random.
3. The **beoserv** daemon increases the max number of master nodes supported by the config file's `masterorder` directive from four to eight.
4. The **beosi** script is now more tolerant about network controller names. Previously, the script recognized only names beginning with *eth*, *lo*, and *virbr*.
5. The *scyld-release* rpm now installs the base distribution's *yum-plugin-priorities* rpm as a dependency. This supports adding the line `priority=3` to a Scyld ClusterWare yum repo config file, which assigns a higher priority to ClusterWare package names that are the same as base distribution package names, assuming that the base distribution yum repo config files use the default priority=99. (Lower priority values are higher priorities.) For example, this means that a base distribution's newer kernel-* rpms will not update an existing and older ClusterWare's kernel-* rpms, without needing to execute **yum** with a combination of **--disablerepo=* --enablerepo=cw-*** or **--disablerepo=cw* --exclude=kernel-*** arguments. See https://wiki.centos.org/PackageManagement/Yum/Priorities for details.

New in Scyld ClusterWare 6.7.6 - Scyld Release 676g0001 - May 9, 2016

1. Scyld ClusterWare now redistributes a non-default TORQUE package that does not employ the base distribution's cpuset functionality. See the Section called *Optionally install a different TORQUE package* for details.

New in Scyld ClusterWare 6.7.6 - Scyld Release 676g0000 - April 7, 2016

1. The base kernel updates to 2.6.32-573.22.1.el6.676g0000. See https://rhn.redhat.com/errata/RHSA-2016-0494.html for details.
2. The e1000e Ethernet driver updates to version 3.3.3, derived from http://sourceforge.net/projects/e1000/.
3. The `node_up` script now adds the line *vm.zone_reclaim_mode=1* to the file `/etc/beowulf/conf.d/sysctl.conf`, which gets populated to `/etc/sysctl.conf` for each booting compute node. See the Section called *Optionally configure vm.zone_reclaim_mode on compute nodes* for details.
4. Scyld ClusterWare now redistributes the Slurm job manager with the package name *slurm-scyld*, together with the Munge authentication plugin (package *munge-scyld*). This initial Slurm version is 15.08.6-1, derived from http://slurm.schedmd.com. The *slurm* service is initially chkconfig'ed *off*. The *torque-scyld* package continues to be distributed, and the *torque* service is also initially chkconfig'ed *off*. The two job management packages coexist on the master node, although only one of them should be enabled at any point in time. See the Section called *Optionally enable job manager* for details.

New in Scyld ClusterWare 6.7.5 - Scyld Release 675g0000 - March 8, 2016

1. The base kernel updates to 2.6.32-573.18.1.el6.675g0000. See https://rhn.redhat.com/errata/RHBA-2016-0150.html for details.
2. The igb Ethernet driver updates to version 5.3.3.5, derived from http://sourceforge.net/projects/e1000/.
3. Introduce a new `/etc/beowulf/init.d/97phi` script to optionally enable Intel Xeon Phi cards on compute nodes.
4. The openmpi-1.10-scyld packages update to version 1.10.2, which by default update and replace only earlier version 1.10 packages and do not affect any installed OpenMPI version 1.8, 1.7, 1.6, or 1.5 packages. See the Section called *Installing and managing concurrent versions of packages* for general issues about supporting multiple concurrent versions. The libraries were built with Gnu version 4.4.7-16, Intel version 2013_sp1.3.174, and PGI version 14.6 compiler families. Scyld releases of OpenMPI derive from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.

New in Scyld ClusterWare 6.7.4 - Scyld Release 674g0000 - December 29, 2015

1. The base kernel updates to 2.6.32-573.12.1.el6.674g0000. See https://rhn.redhat.com/errata/RHSA-2015-2636.html for details.
2. Introduce a new `/etc/beowulf/init.d/50autofs` script to optionally enable automount on compute nodes. See the Section called *Optionally configure automount on compute nodes* for details.
3. Beginning with *nodescripts-1.4.3-674g0001.x86_64.rpm*, Scyld ClusterWare now forceably enables the `/etc/beowulf/init.d/30cpuspeed` script. Penguin Computing has determined that optimal CPU performance requires that this script (or something like it) should be enabled. See the Section called *Optionally configure and enable compute node CPU speed/power management* and the comments inside the `30cpuspeed` script and inside the associated configuration file `/etc/beowulf/conf.d/cpuspeed.conf` for details.

New in Scyld ClusterWare 6.7.3 - Scyld Release 673g0000 - November 30, 2015

1. The base kernel updates to 2.6.32-573.8.1.el6.673g0000. See https://rhn.redhat.com/errata/RHBA-2015-1992.html for details.
2. Fix a rare bproc race condition that typically exhibits itself as a NULL pointer dereference doing a process exit on a compute node, which results in a kernel panic.
3. The openmpi-1.10-scyld packages update to version 1.10.1, which by default update and replace only earlier version 1.10 packages and do not affect any installed OpenMPI version 1.8, 1.7, 1.6, or 1.5 packages. See the Section called *Installing and managing concurrent versions of packages* for general issues about supporting multiple concurrent versions. The libraries were built with Gnu version 4.4.7-16, Intel version 2013_sp1.3.174, and PGI version 14.6 compiler families. Scyld releases of OpenMPI derive from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.
4. The MPICH3 mpich-scyld release updates to version 3.2, derived from http://www.mpich.org/. The libraries were built with Gnu version 4.4.7-16, Intel version 2013_sp1.3.174, and PGI version 14.6 compiler families. See the *User's Guide* Appendix F, *MPICH-3 Release Information* for details.
5. Populate compute nodes with standard /dev/std* devices.

New in Scyld ClusterWare 6.7.2 - Scyld Release 672g0000 - October 20, 2015

1. The base kernel updates to 2.6.32-573.7.1.el6.672g0000. See https://rhn.redhat.com/errata/RHBA-2015-1827.html for details.
2. Fix a compute node "soft lockup" that syslogs the offender as *filecache_sys_open*. This fix has also been applied to newer versions of the `kmod-filecache` rpm for Scyld ClusterWare releases 6.6.1 onward.
3. Fix a master node and compute node Out-Of-Memory (OOM) failure that is due to a bproc memory leak of the kernel's *size-512* dynamic memory slab. This leak occurs when a user program on the node executes the *execve()* or *execl()* intrinsic. Use **slabtop** to view the current usage of *size-512* dynamic memory, and compare that *SIZE* value to the total amount of physical memory on that node in order to gauge the current vulnerability to an OOM failure. Without the fix, the *size-512* size continually increases. This fix has also been applied to newer versions of the `kmod-bproc` rpm for Scyld ClusterWare releases 6.5.8 onward.
4. Introduce a new `/etc/beowulf/init.d/14rpc` script to manage startup of the *rpc.statd* daemon on a compute node, vs. the previous (and sometimes flawed) startup done by the `/usr/lib/beoboot/bin/node_up` script. Use **beochkconfig** to disable `14rpc` if *rpc.statd* is not needed.
5. Introduce a warning during **service beowulf start** and for each booting node (logged in `/var/log/beowulf/node.`*N*) to remind the cluster administrator that a *kernel.pid_max* entry in `/etc/beowulf/conf.d/sysctl.conf` is ignored, and that the master node's *kernel.pid_max* prevails cluster-wide.

New in Scyld ClusterWare 6.7.1 - Scyld Release 671g0000 - September 3, 2015

1. The base kernel updates to 2.6.32-573.3.1.el6.671g0000. See https://rhn.redhat.com/errata/RHSA-2015-1623.html for details.
2. Scyld ClusterWare now distributes openmpi-1.10-scyld packages, which are a redistribution of OpenMPI version 1.10 and derived from http://www.open-mpi.org/. These openmpi-1.10-scyld packages do not affect any installed OpenMPI version 1.8, 1.7, 1.6, or 1.5 packages. See the Section called *Installing and managing concurrent versions of packages* for general issues about supporting multiple concurrent versions. The libraries were built with Gnu version 4.4.7-16, Intel version 2013_sp1.3.174, and PGI version 14.6 compiler families. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.
3. The `/etc/beowulf/init.d/15openib` (sets up the Infiniband devices) node startup script updates to support additional QLogic devices and to use the generic udev functionality for a cleaner implementation.
4. The *beoserv* daemon's dhcp server functionality now accepts client packets that contain a `vendor_info` field as small as 8 bytes, vs. the previous minimum of 60 bytes, and thereby accepts client requests from some models of "smart" switches that were previously rejected.

New in Scyld ClusterWare 6.7.0 - Scyld Release 670g0000 - August 13, 2015

1. The base kernel updates to 2.6.32-573.1.1.el6.670g0000. See https://rhn.redhat.com/errata/RHSA-2015-1272.html and https://rhn.redhat.com/errata/RHBA-2015-1466.html for details.
2. The Scyld ClusterWare distribution of ganglia has been repackaged down from four rpms to two: `ganglia-scyld`, now updated to version 3.7.1-1, and `ganglia-web-scyld`, now updated to version 3.7.0-1.

New in Scyld ClusterWare 6.6.7 - addendum to Release 667g0000 - November 7, 2016

1. The base kernel updates to 2.6.32-504.30.3.el6.667g0001. This kernel differs from the previous 2.6.32-504.30.3.el6.667g0000 only by the inclusion of a fix for the Red Hat Bugzilla #1384344 security exploit ("kernel: mm: privilege escalation via MAP_PRIVATE COW breakage"). This kernel, together with the matching kmod-* rpms which were built from the same source code files as they were in 677g0000, is currently only available in the Scyld ClusterWare 6.6 `updates.next` yum repo. NOTE: the matching Panasas kernel module is *not* yet available for this kernel, so do not install and use this kernel if your cluster employs Panasas storage.

   To install, ensure that the `/etc/yum.repos.d/clusterware.repo` file (or whatever the name of the ClusterWare repo file is being used) has URLs that refer to the `6.6` repo, then:

   ```
   yum --disablerepo=* --enablerepo=cw-next update
   ```

New in Scyld ClusterWare 6.6.7 - Scyld Release 667g0000 - August 12, 2015

1. The base kernel updates to 2.6.32-504.30.3.el6.667g0000. See https://rhn.redhat.com/errata/RHSA-2015-1221.html for details.
2. The openmpi-1.8-scyld packages update to version 1.8.8, which by default update and replace only earlier version 1.8 packages and do not affect any installed OpenMPI version 1.7, 1.6, or 1.5 packages. See the Section called *Installing and managing concurrent versions of packages* for general issues about supporting multiple concurrent versions. The libraries were built with Gnu version 4.4.7, Intel version 2013_sp1.3.174, and PGI version 14.6 compiler families. Scyld releases of OpenMPI derive from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.
3. Relax a constraint in the part of the beoserv daemon that functions as the private cluster network's DHCP server. Previously, it silently rejected client requests that contain a *vendor_info* field shorter than 60 bytes. It now accepts a client packet with a *vendor_info* field as short as eight bytes, and it will issue an informative syslog warning about why a DHCP client request is being rejected, versus silently rejecting the packet for one of several possible reasons.

New in Scyld ClusterWare 6.6.6 - Scyld Release 666g0000 - June 29, 2015

1. The base kernel updates to 2.6.32-504.23.4.el6.666g0000. See https://rhn.redhat.com/errata/RHSA-2015-1081.html for details.
2. Fix the `/etc/beowulf/init.d/25cuda` script to correctly support MPI+CUDA functionality.

New in Scyld ClusterWare 6.6.5 - Scyld Release 665g0000 - June 4, 2015

1. The base kernel updates to 2.6.32-504.16.2.el6.665g0000. See https://rhn.redhat.com/errata/RHSA-2015-0864.html for details.
2. MVAPICH2 updates to version 2.1 for the `mvapich2-psm-scyld` and `mvapich2-scyld` packages. See the Section called *Installing and managing concurrent versions of packages* for general issues about supporting multiple concurrent versions. This software suite derives from http://mvapich.cse.ohio-state.edu/. *NOTE:* MVAPICH2-2.1 introduces an algorithm to determine CPU topology on the node, and this new algorithm does not work properly for older Mellanox

controllers and firmware, resulting in software threads not spreading out across a node's cores by default. See the Section called *Issues with MVAPICH2 and CPU Sets* or the *User's Guide* Appendix E, *MVAPICH2 Release Information* for details.

3. TORQUE updates to version 4.2.10, from www.adaptivecomputing.com/support/download-center/torque-download/. This release enables support of CPU Sets; see http://docs.adaptivecomputing.com/torque/4-1-4/help.htm#topics/3-nodes/linuxCpusetSupport.htm for details. Also, the Scyld ClusterWare `torque` rpm renames to `torque-scyld` and disallows the concurrent installation of the base distribution's `torque` packages.
4. The openmpi-1.8-scyld packages update to version 1.8.5, which by default update and replace only earlier version 1.8 packages and do not affect any installed OpenMPI version 1.7, 1.6, or 1.5 packages. See the Section called *Installing and managing concurrent versions of packages* for general issues about supporting multiple concurrent versions. The libraries were built with Gnu version 4.4.7, Intel version 2013_sp1.3.174, and PGI version 14.6 compiler families. Scyld releases of OpenMPI derive from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.

New in Scyld ClusterWare 6.6.4 - Scyld Release 664g0000 - April 1, 2015

1. The base kernel updates to 2.6.32-504.12.2.el6.664g0000. See https://rhn.redhat.com/errata/RHSA-2015-0674.html for details.
2. The `/etc/beowulf/init.d/15openib` script updates to support both QLogic and Mellanox Infiniband controllers. Previously, clusters with QLogic controllers have employed a custom init script that should now be explicitly disabled by the cluster administrator.
3. Introduce a new MVAPICH2 version 2.0.0 package `mvapich2-psm-scyld`, which employs the Performance Scaled Messaging (PSM) interface to provide superior performance for QLogic Infiniband controllers. We continue to distribute the `mvapich2-scyld` package (currently also version 2.0.0) that employs the traditional Verbs interface, which supports both Mellanox and QLogic controllers. This software suite derives from http://mvapich.cse.ohio-state.edu/. See the *User's Guide* Appendix E, *MVAPICH2 Release Information* for details.
4. The MPICH3 mpich-scyld release updates to version 3.1.4, derived from http://www.mpich.org/. The libraries were built with Gnu version 4.4.7, Intel version 2013_sp1.3.174, and PGI version 14.6 compiler families. See the *User's Guide* Appendix F, *MPICH-3 Release Information* for details.
5. `beoconfig-libs` updates to version 2.0.18 to fix an infrequent *glibc detected *** /usr/bin/python: double free or corruption* error, most often seen (if at all) when executing **yum**.

New in Scyld ClusterWare 6.6.3 - Scyld Release 663g0000 - February 11, 2015

1. The base kernel updates to 2.6.32-504.8.1.el6.663g0000. See https://rhn.redhat.com/errata/RHSA-2015-0087.html for details.
2. The igb Ethernet driver updates to version 5.2.15, derived from http://sourceforge.net/projects/e1000/.
3. The e1000e Ethernet driver updates to version 3.1.0.2, derived from http://sourceforge.net/projects/e1000/.
4. Updates the `/etc/beowulf/init.d/30cpuspeed` script that manages CPU speed/power on compute nodes. See the Administrator's Guide *Configuring CPU speed/power for Compute Nodes* for details.

New in Scyld ClusterWare 6.6.2 - Scyld Release 662g0000 - December 31, 2014

1. The base kernel updates to 2.6.32-504.3.3.el6.662g0000. See https://rhn.redhat.com/errata/RHSA-2014-1997.html for details.
2. The openmpi-1.8-scyld packages update to version 1.8.4, which by default update and replace only earlier version 1.8 packages and do not affect OpenMPI version 1.7, 1.6, or 1.5 packages. See the Section called *Installing and managing concurrent versions of packages* for general issues about supporting multiple concurrent versions. The libraries were built with Gnu version 4.4.7, Intel version 2013_sp1.3.174, and PGI version 14.6 compiler families. Scyld releases of OpenMPI derive from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.

New in Scyld ClusterWare 6.6.1 - Scyld Release 661g0000 - December 6, 2014

1. The base kernel updates to 2.6.32-504.1.3.el6.661g0000. See https://rhn.redhat.com/errata/RHSA-2014-1843.html for details.
2. Fix a compute node **bpslave** "soft lockup" that would occasionally occur during node boot.
3. The bproc *filecache* functionality now downloads to a compute node a mirror image of the master node's symlinks that follow a path to the final leaf file. For example, opening `/lib64/libcrypt.so.1` creates the symlink `/lib64/libcrypt.so.1` and downloads the leaf file `/lib64/libcrypt-2.12.so`. Previously, bproc *filecache* downloaded only the final leaf file and named it `/lib64/libcrypt.so.1`. This requires a coordinated update to the *beoserv*, *beoclient3*, and *bproc* packages.
4. The beonss **kickbackproxy** daemon that executes on each compute node now throttles its attempts to reconnect to the master node **kickbackdaemon** server if that connection has been lost. Previously, the **kickbackproxy** would rapidly attempt to reconnect, thereby keeping an otherwise idle orphaned compute node busy and thus constraining a run-to-completion reboot.

New in Scyld ClusterWare 6.6.0 - Scyld Release 660g0000 - November 17, 2014

1. The base kernel updates to 2.6.32-504.el6.660g0000. See https://rhn.redhat.com/errata/RHSA-2014-1392.html for details.
2. The Scyld ClusterWare igb Ethernet driver that we typically derive from http://sourceforge.net/projects/e1000/ has been temporarily removed from the Penguin Computing distribution until we can locate or craft a version that is compatible with RHEL6 Update 6 and CentOS 6.6. Meanwhile, the 2.6.32-504.el6.660g0000 kernel will use the native igb driver provided by Red Hat.
3. *IMPORTANT:* The Red Hat RHEL6 Update 6 and CentOS 6.6 base distributions now include an *mpich* version 3 package that conflicts with the Scyld ClusterWare *mpich* version 1.2.7p1 packages. See the Section called *Issues with Mpich* for details.

New in Scyld ClusterWare 6.5.8 - addendum to Release 658g0000 - November 7, 2016

1. The base kernel updates to 2.6.32-431.29.2.el6.658g0001. This kernel differs from the previous 2.6.32-431.29.2.el6.658g0000 only by the inclusion of a fix for the Red Hat Bugzilla #1384344 security exploit

("kernel: mm: privilege escalation via MAP_PRIVATE COW breakage"). This kernel, together with the matching kmod-* rpms which were built from the same source code files as they were in 658g0000, is currently only available in the Scyld ClusterWare 6.5 `updates.next` yum repo. NOTE: the matching Panasas kernel module is *not* yet available for this kernel, so do not install and use this kernel if your cluster employs Panasas storage.

To install, ensure that the `/etc/yum.repos.d/clusterware.repo` file (or whatever the name of the ClusterWare repo file is being used) has URLs that refer to the `6.5` repo, then:

```
yum --disablerepo=* --enablerepo=cw-next update
```

New in Scyld ClusterWare 6.5.8 - Scyld Release 658g0000 - October 17, 2014

1. The base kernel updates to 2.6.32-431.29.2.el6.658g0000. See https://rhn.redhat.com/errata/RHSA-2014-1167.html for details.
2. Populate each compute node at boot time by pushing the master node's file `/etc/beowulf/conf.d/limits.conf` to the the node as `/etc/security/limits.conf`. This master node's file is initially a concatenation of the master node's `/etc/security/limits.conf` and the files in the directory `/etc/security/limits.d/`. The cluster administrator may edit `/etc/beowulf/conf.d/limits.conf` as desired.
3. Fix a compute node hang that can occur when attempting to link an application that references a nonexistent library file.
4. Support bproc *filecache* pathnames that include embedded `/../` strings. Previously, these were rejected without resolving the true pathname.
5. Fix a rare bug that exhibits itself as a compute node that continually retries an unsuccessful boot, complaining that the communication **bpslave**-**bpmaster** communication (which defaults to port 932) cannot be established.
6. TORQUE updates to version 4.2.9, from www.adaptivecomputing.com/support/download-center/torque-download/.
7. The openmpi-1.8-scyld packages update to version 1.8.3, which by default update and replace only earlier version 1.8 packages and do not affect OpenMPI version 1.7, 1.6, or 1.5 packages. See the Section called *Installing and managing concurrent versions of packages* for general issues about supporting multiple concurrent versions. The libraries were built with Gnu version 4.4.7, Intel version 2013_sp1.3.174, and PGI version 14.6 compiler families. Scyld releases of OpenMPI derive from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.
8. The MPICH3 mpich-scyld release updates to version 3.1.3, derived from http://www.mpich.org/. The libraries were built with Gnu version 4.4.7, Intel version 2013_sp1.3.174, and PGI version 14.6 compiler families. See the *User's Guide* Appendix F, *MPICH-3 Release Information* for details.
9. Fix a problem in PVM that results in a hung application with unkillable threads.
10. NVIDIA K40 GPU now executes in *persistance* mode for quicker startup of GPU operations.

New in Scyld ClusterWare 6.5.7 - Scyld Release 657g0000 - August 18, 2014

1. The base kernel updates to 2.6.32-431.23.3.el6.657g0001. See https://rhn.redhat.com/errata/RHSA-2014-0924.html and https://rhn.redhat.com/errata/RHSA-2014-0981.html for details.
2. Fix a timing problem in **bproc** that can put a compute node's **bpslave** into a "soft lockup" state, with a stack traceback that identifies the culprit as *_spin_lock* called from *get_task_mm*.

3. Introduces a fix/workaround in bproc to avoid a timing problem that exhibits itself most frequently when doing a *bproc_move* from a compute node to another node, followed immediately by another *bproc_move* back to the same compute node. The workaround is to add a small delay prior to the second *bproc_move*. A complete fix will follow in a subsequent release.
4. The MPICH3 mpich-scyld release updates to version 3.1.2, derived from http://www.mpich.org/. The libraries were built with Gnu version 4.4.7, Intel version 2013_sp1.3.174, and PGI version 14.6 compiler families. See the *User's Guide* Appendix F, *MPICH-3 Release Information* for details.

New in Scyld ClusterWare 6.5.6 - Scyld Release 656g0000 - July 21, 2014

1. The base kernel updates to 2.6.32-431.20.3.el6.656g0000. See https://rhn.redhat.com/errata/RHSA-2014-0771.html for details.
2. **service beowulf reload** now re-reads the `/etc/beowulf/config` *libraries* entries and rebuilds the list of libraries managed by the bproc *filecache* functionality for the master node and all the *up* compute nodes.
3. TORQUE updates to version 4.2.8, from www.adaptivecomputing.com/support/download-center/torque-download/.
4. MVAPICH2 updates to version 2.0, derived from http://mvapich.cse.ohio-state.edu/. See the *User's Guide* Appendix E, *MVAPICH2 Release Information* for details.
5. The MPICH3 mpich-scyld release updates to version 3.1.1, derived from http://www.mpich.org/. See the *User's Guide* Appendix F, *MPICH-3 Release Information* for details.
6. The various MPI library suites (OpenMPI, MPICH, MPICH2, MVAPICH2, MPICH3) have been rebuilt with newer versions of the Gnu version 4.4.7, Intel version 2013_sp1.3.174, and PGI version 14.6 compiler families.

New in Scyld ClusterWare 6.5.5 - Scyld Release 655g0000 - June 10, 2014

1. The base kernel updates to 2.6.32-431.17.1.el6.655g0000. See https://rhn.redhat.com/errata/RHSA-2014-0475.html for details.
2. The Scyld ClusterWare igb Ethernet driver updates to version 5.2.5, derived from http://sourceforge.net/projects/e1000/.

New in Scyld ClusterWare 6.5.4 - Scyld Release 654g0000 - April 14, 2014

1. The base kernel updates to 2.6.32-431.11.2.el6.654g0000. See https://rhn.redhat.com/errata/RHSA-2014-0328.html for details.
2. Previously, Scyld ClusterWare distributed the **env-modules** package, which was a semi-customized redistribution of the RHEL6 **environment-modules** package. These two Scyld ClusterWare and RHEL6 packages could not co-exist, which meant that the various Scyld ClusterWare packages that employed environment modules (e.g., mpich2-scyld, mvapich2-scyld, mpich-scyld, openmpi-1.*-scyld) could not co-exist with RHEL6 packages that need the RHEL6 **environment-modules** (e.g., mpich2, mvapich2, openmpi, mpich). Beginning with Scyld ClusterWare 6.5.4, Scyld ClusterWare no longer distributes **env-modules**, and Scyld packages instead use the RHEL6 **environment-modules**, which now allows those Scyld ClusterWare and RHEL6 packages to co-exist.
3. TORQUE updates to version 4.2.7, from www.adaptivecomputing.com/support/download-center/torque-download/.

4. Scyld ClusterWare now distributes openmpi-1.8-scyld packages, which are a redistribution of OpenMPI version 1.8 and derived from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.
5. The openmpi-1.7-scyld packages updates to version 1.7.5, which by default update and replace only earlier version 1.7 packages and do not affect OpenMPI version 1.6 or 1.5 packages. See the Section called *Installing and managing concurrent versions of packages* for general issues about supporting multiple concurrent versions. Scyld releases of OpenMPI derived from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.
6. **service beowulf start** and **restart** now check the size of `/usr/lib/locale/locale-archive` and issue a warning if the file is huge and thus would impact cluster performance. See the Administrator's Guide for details.

New in Scyld ClusterWare 6.5.3 - Scyld Release 653g0000 - March 4, 2014

1. The base kernel updates to 2.6.32-431.5.1.el6.653g0000. See https://rhn.redhat.com/errata/RHSA-2014-0159.html for details.
2. The mpich-scyld release updates to version 3.1, derived from http://www.mpich.org/. See the *User's Guide* Appendix F, *MPICH-3 Release Information* for details.
3. The Scyld ClusterWare igb Ethernet driver updates to version 5.1.2, derived from http://sourceforge.net/projects/e1000/.
4. Scyld ClusterWare now redistributes an optional e1000e Ethernet driver version 3.0.4.1, derived from http://sourceforge.net/projects/e1000/. If a local cluster administrator wishes to update the default RHEL6.9/CentOS6.9 e1000e-2.3.2-k to the latest e1000e from SourceForge, then **yum install kmod-e1000e** and **depmod**.
5. Eliminate the unnecessary requirement that TORQUE Python libraries be installed in order for **beostatus** filtering to work.

New in Scyld ClusterWare 6.5.2 - Scyld Release 652g0000 - January 15, 2014

1. The base kernel updates to 2.6.32-431.3.1.el6.652g0000. See https://rhn.redhat.com/errata/RHBA-2014-0004.html for details.

New in Scyld ClusterWare 6.5.1 - Scyld Release 651g0000 - January 3, 2014

1. The base kernel updates to 2.6.32-431.1.2.el6.651g0000. See https://rhn.redhat.com/errata/RHSA-2013-1801.html for details. In addition to containing the usual Scyld "hooks", this Scyld ClusterWare kernel decreases the kernel's compiled-in TCP_TIMEWAIT_LEN timeout from its original 60 seconds down to 30 seconds. This reduces the potential for a node to be unable to allocate a socket due to an application voraciously creating and closing sockets so rapidly that all available sockets are either open or have been closed and are sitting in TIME_WAIT limbo state.
2. Fix a bug in **beosi** that causes the script to hang doing a **find /sys/class/infiniband** under some circumstances.
3. Fix various obscure timing problems in **bproc** that can panic the master node or a compute node, or hang a compute node, or lead to a kernel *soft lockup* state. These infrequent events would generally only occur following a network disconnect between a compute node and the master, or while shutting down the *beowulf* service, while a compute node's **bpslave** is mid-transaction with the master's **bpmaster**.

New in Scyld ClusterWare 6.5.0 - Scyld Release 650g0000 - December 9, 2013

1. The base kernel updates to 2.6.32-431.el6.650g0000. See https://rhn.redhat.com/errata/RHSA-2013-1645.html for details.
2. The Scyld ClusterWare igb Ethernet driver updates to version 5.0.6, derived from http://sourceforge.net/projects/e1000/.
3. TORQUE updates to version 4.2.6.1, from www.adaptivecomputing.com/support/download-center/torque-download/.
4. MVAPICH2 updates to version 2.0b, derived from http://mvapich.cse.ohio-state.edu/. See the *User's Guide* Appendix E, *MVAPICH2 Release Information* for details.
5. Bash *process substitution* now works, e.g., **diff <(sort file1) <(sort file2)**.
6. Improve the synchronization between the **beoserv** and **bpmaster** daemons with respect to what port number the latter wishes to use to communicate with the compute node **bpslave** daemons.

New in Scyld ClusterWare 6.4.7 - Scyld Release 647g0000 - October 28, 2013

1. The base kernel is updated to 2.6.32-358.23.2.el6.647g0000. See https://rhn.redhat.com/errata/RHSA-2013-1436.html for details.
2. Fix a bug in run-to-completion that was mistakenly introduced in Scyld ClusterWare 6.4.6 that caused *orphaned* compute nodes to always reboot after one hour.
3. The openmpi-1.7-scyld packages are updated to version 1.7.3, which by default update and replace only earlier version 1.7 packages and do not affect OpenMPI version 1.6 or 1.5 packages. See the Section called *Installing and managing concurrent versions of packages* for general issues about supporting multiple concurrent versions. Scyld releases of OpenMPI derived from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.
4. The **sendstats** daemon more reliably avoids being started more than once per compute node.

New in Scyld ClusterWare 6.4.6 - Scyld Release 646g0001 - September 27, 2013

1. TORQUE updates to version 4.2.5, from www.adaptivecomputing.com/support/download-center/torque-download/. This distribution also fixes problems that were introduced by the recent Scyld ClusterWare inclusion of the *Maui* scheduler, where the installation of *Maui* perturbed an optionally preexisting installation of the *Moab* scheduler. Both *Maui* and *Moab* can now coexist as installed packages, although the local cluster administrator must perform a one-time selection of which scheduler to use, if both are installed. See the Section called *Optionally enable TORQUE scheduler* for details.

New in Scyld ClusterWare 6.4.6 - Scyld Release 646g0000 - September 11, 2013

1. The base kernel is updated to 2.6.32-358.18.1.el6.646g0000. See https://rhn.redhat.com/errata/RHSA-2013-1173.html for details.
2. The MVAPICH2 release is updated to version 2.0a, derived from http://mvapich.cse.ohio-state.edu/. See the *User's Guide* Appendix E, *MVAPICH2 Release Information* for details.

New in Scyld ClusterWare 6.4.5 - Scyld Release 645g0001 - September 6, 2013

1. TORQUE updates to version 4.2.4.1, from www.adaptivecomputing.com/support/download-center/torque-download/. This TORQUE also fixes a *pbs_mom* security vulnerability that was announced by Adaptive Computing on Sept. 6, 2013, that afflicts all TORQUE releases to date.

New in Scyld ClusterWare 6.4.5 - Scyld Release 645g0000 - August 26, 2013

1. The base kernel is updated to 2.6.32-358.14.1.el6.645g0000. See https://rhn.redhat.com/errata/RHSA-2013-1051.html for details.
2. TORQUE updates to version 4.2.4, from www.adaptivecomputing.com/support/download-center/torque-download/. This Scyld ClusterWare distribution changes the default job scheduler from the problematic built-in *pbs_sched* to Adaptive Computing's *Maui*, currently version 3.3.1. Maui distributes as a separate rpm and is required by TORQUE 4.2.4. See the *User's Guide* Appendix B, *TORQUE and Maui Release Information* for details.
3. The mpich2, mvapich2, mpich-scyld, and openmpi-1.5, -1.6, and -1.7 packages have been rebuilt using newer Intel and PGI compiler suites: Intel composer_xe 2013.5.192 vs. composerxe-2011.4.191, and PGI 13.6 vs. 11.9.

New in Scyld ClusterWare 6.4.4 - Scyld Release 644g0000 - July 8, 2013

1. The base kernel is updated to 2.6.32-358.11.1.el6.644g0000. See https://rhn.redhat.com/errata/RHSA-2013-0911.html for details.
2. The Scyld ClusterWare packaging for OpenMPI has changed in order to more easily install and retain multiple co-existing versions on the master node. See the Section called *Installing and managing concurrent versions of packages* for general issues about supporting multiple concurrent versions. This Scyld release updates openmpi-1.7-scyld with OpenMPI version 1.7.2, which by default replaces the earlier version 1.7.1 rpms. This Scyld release also includes the first release of openmpi-1.6-scyld for OpenMPI version 1.6.5. These openmpi-1.6-scyld rpms will only update (and replace) earlier openmpi-1.6-scyld rpms and will not update any existing openmpi-scyld rpms, which will likely be version 1.6.4. Scyld releases of OpenMPI are derived from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.
3. Yet again improve the run-to-completion algorithm for determining if an orphaned compute node is effective idle (and thus can reboot). See the Section called *When Master Nodes Fail - With Run-to-Completion* in the *Administrator's Guide* for details, and **man bpctl** for a summary.

New in Scyld ClusterWare 6.4.3 - Scyld Release 643g0000 - June 13, 2013

1. The base kernel is updated to 2.6.32-358.6.2.el6.643g0000. See https://rhn.redhat.com/errata/RHSA-2013-0830.html for details.
2. The Scyld ClusterWare packaging for OpenMPI has changed in order to more easily install and retain multiple co-existing versions on the master node. The openmpi-scyld packaging, which last distributed version 1.6.4, is deprecated. It has been replaced by new packaging which incorporates the OpenMPI *x.y* family name into the package name, e.g., openmpi-1.7-scyld, openmpi-1.6-scyld, and openmpi-1.5-scyld. This Scyld release installs by default the openmpi-1.7-scyld version 1.7.1 rpms. The yum repo also contains (but does not install by default) openmpi-1.6-scyld version 1.6.4

and openmpi-1.5-scyld version 1.5.5 rpms, which may be manually installed if desired. The OpenMPI releases are derived from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for OpenMPI changelog details, and the Section called *Installing and managing concurrent versions of packages* for general issues about supporting multiple concurrent versions of OpenMPI.

3. The env-modules package now properly recognizes **module load** defaults that are declared in `.version` files found in the `/opt/scyld/modulefiles/` subdirectories.
4. Fix a rare deadlock of a master or compute node BProc I/O Daemon that can occur under very high workloads.
5. Suppress various redundant BProc syslog messages, e.g., a flurry of redundant ECONNREFUSED warnings.

New in Scyld ClusterWare 6.4.2 - Scyld Release 642g0000 - May 15, 2013

1. The base kernel is updated to 2.6.32-358.6.1.el6.642g0000. See https://rhn.redhat.com/errata/RHSA-2013-0744.html for details about 1.7.1, and .
2. The MVAPICH2 release is updated to version 1.9.0, derived from http://mvapich.cse.ohio-state.edu/. See the *User's Guide* Appendix E, *MVAPICH2 Release Information* for details.
3. The mpich-scyld release is updated to version 3.0.4, derived from http://www.mpich.org/. See the *User's Guide* Appendix F, *MPICH-3 Release Information* for details.
4. Supports forwarding compute node log messages to an alternative **syslogd** server other than to the default master node server. See the *Administrator's Guide* for details.

New in Scyld ClusterWare 6.4.1 - Scyld Release 641g0000 - April 10, 2013

1. The base kernel is updated to 2.6.32-358.2.1.el6.641g0000. See https://rhn.redhat.com/errata/RHSA-2013-0630.html for details.
2. TORQUE updates to version 4.2.2, from www.adaptivecomputing.com/support/download-center/torque-download/. See the *User's Guide* Appendix B, *TORQUE Release Information* for details.
3. The MVAPICH2 release is updated to version 1.9b, derived from http://mvapich.cse.ohio-state.edu/. See the *User's Guide* Appendix E, *MVAPICH2 Release Information* for details.
4. The mpich-scyld release is updated to version 3.0.3, derived from http://www.mpich.org/. See the *User's Guide* Appendix F, *MPICH-3 Release Information* for details.
5. A **service beowulf start** (or **restart**) and **reload** now saves timestamped backups of various `/etc/beowulf/` configuration files, e.g., `config` and `fstab`, to assist a cluster administrator to recover a working configuration after an invalid edit.

New in Scyld ClusterWare 6.4.0 - Scyld Release 640g0000 - March 13, 2013

1. The base kernel is updated to 2.6.32-358.0.1.el6.640g0001. See https://rhn.redhat.com/errata/RHSA-2013-0496.html and https://rhn.redhat.com/errata/RHSA-2013-0567.html for details.
2. The TORQUE release is updated to version 4.2.1, derived from http://www.adaptivecomputing.com/resources/downloads/torque/. See the *User's Guide* Appendix B, *TORQUE Release Information* for details.

3. The OpenMPI release is updated to version 1.6.4, derived from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.
4. Includes the first release of mpich-scyld, which is the Scyld ClusterWare distribution of mpich-3, version 3.0.2, derived from http://www.mpich.org/. See the *User's Guide* Appendix F, *MPICH-3 Release Information* for details.

New in Scyld ClusterWare 6.3.7 - Scyld Release 637g0000 - March 4, 2013

1. The base kernel is updated to 2.6.32-279.22.1.el6.637g0002. See https://rhn.redhat.com/errata/RHSA-2013-0223.html for details. This kernel includes built-in firmware to properly boot some compute node server models that employ a bnx2 Ethernet controller, in addition to the bnx2 server models previously supported by the 2.6.32-279.19.1.el6.636g0001 kernel, as well as some models that employ a cxgb3 Ethernet controller.
2. The TORQUE release is a "refresh" version 4.2.0, derived from http://www.adaptivecomputing.com/resources/downloads/torque/. Adaptive Computing refreshed their "Limited GA" 4.2.0 on February 14, 2013, and Scyld ClusterWare subsequently distributed it as torque-4.2.0-636g0001. See the *User's Guide* Appendix B, *TORQUE Release Information* for details.
3. `/etc/beowulf/config` supports a new directive, *firmware*, to assist in loading firmware for *bootmodule* drivers on compute nodes. See the Section called *Issues with bootmodule firmware* and the *Administrator's Guide* for details.

New in Scyld ClusterWare 6.3.6 - Scyld Release 636g0001 - February 8, 2013

1. The base kernel is updated to 2.6.32-279.19.1.el6.636g0001. See https://rhn.redhat.com/errata/RHSA-2012-1580.html for details. This kernel differs from the earlier 2.6.32-279.19.1.el6.636g0000 in that it includes built-in firmware to properly boot some compute node server models that employ a bnx2 Ethernet controller.
2. The Scyld ClusterWare igb Ethernet driver is version 4.1.2, derived from http://sourceforge.net/projects/e1000/.
3. Include a working **beonetconf** command.
4. Improve the run-to-completion algorithm for determining if an orphaned compute node is effective idle (and thus can reboot).

New in Scyld ClusterWare 6.3.6 - Scyld Release 636g0000 - January 2, 2013

1. The base kernel is updated to 2.6.32-279.19.1.el6.636g0000. See https://rhn.redhat.com/errata/RHSA-2012-1580.html for details.
2. The TORQUE release is updated to version 4.2.0, derived from http://www.adaptivecomputing.com/resources/downloads/torque/. See the *User's Guide* Appendix B, *TORQUE Release Information* for details.
3. Fix a problem with compute node server models that employ a *radeon* controller which failed to load firmware during node bootup.
4. Fix a BProc problem wherein the **/bin/ps** and **/bin/top** commands were not correctly reporting the CPU usage of processes executing on compute nodes.

New in Scyld ClusterWare 6.3.5 - Scyld Release 635g0000 - November 30, 2012

1. The base kernel is updated to 2.6.32-279.14.1.el6.635g0000. See https://rhn.redhat.com/errata/RHSA-2012-1426.html for details.
2. The TORQUE release is updated to version 4.1.3, derived from http://www.adaptivecomputing.com/resources/downloads/torque/. See the *User's Guide* Appendix B, *TORQUE Release Information* for details.

New in Scyld ClusterWare 6.3.4 - Scyld Release 634g0000 - November 5, 2012

1. The base kernel is updated to 2.6.32-279.11.1.el6.634g0000. See https://rhn.redhat.com/errata/RHSA-2012-1366.html for details.
2. Fix a BProc problem that panics the master node when bpmaster terminates (e.g., doing **service beowulf restart** or **stop**).
3. The OpenMPI release is updated to version 1.6.3, derived from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.

New in Scyld ClusterWare 6.3.3 - Scyld Release 633g0000 - October 24, 2012

1. The base kernel is updated to 2.6.32-279.9.1.el6.633g0001. See https://rhn.redhat.com/errata/RHSA-2012-1304.html for details.
2. The OpenMPI release is updated to version 1.6.2, derived from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.
3. The MPICH2 release is version 1.5, derived from http://www.mcs.anl.gov/research/projects/mpich2/. See the *User's Guide* Appendix D, *MPICH2 Release Information* for details.
4. The MVAPICH2 release is updated to version 1.8.1, derived from http://mvapich.cse.ohio-state.edu/. See the *User's Guide* Appendix E, *MVAPICH2 Release Information* for details. Since the Scyld ClusterWare 6 MVAPICH2 transport mechanism is *ssh*, the cluster administrator will likely need to configure the cluster to allow *ssh* access for non-root users. See the Section called *Optionally enable SSHD on compute nodes* and the *Administrator's Guide* for details.

New in Scyld ClusterWare 6.3.2 - Scyld Release 632g0000 - September 19, 2012

1. The base kernel is updated to 2.6.32-279.5.2.el6.632g0001. See https://rhn.redhat.com/errata/RHSA-2012-1156.html and https://rhn.redhat.com/errata/RHBA-2012-1199.html for details.
2. Fix a BProc problem that left a "lingering ghost" process on the master node that was not associated with any process on a compute node.
3. Add two *bootmodule* entries to `/etc/beowulf/config` to support the latest Penguin servers: `ahci` and `isci`.
4. Support the *nonfatal* mount option for harddrive entries specified in `/etc/beowulf/fstab` to more gracefully handle clusters that have some nodes with harddrives and some nodes without, thus perhaps avoiding needing node-specific `/etc/beowulf/fstab.`*N* file(s).

5. The OpenMPI release is updated to version 1.6.1, derived from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.

New in Scyld ClusterWare 6.3.1 - Scyld Release 631g0000 - August 13, 2012

1. The base kernel is updated to 2.6.32-279.2.1.el6.631g0000. See https://rhn.redhat.com/errata/RHBA-2012-1104.html for details.

New in Scyld ClusterWare 6.3.0 - Scyld Release 630g0000 - August 3, 2012

1. The base kernel is updated to 2.6.32-279.1.1.el6.630g0001. See https://rhn.redhat.com/errata/RHSA-2012-0862.html and https://rhn.redhat.com/errata/RHSA-2012-1064.html for details.
2. The Scyld ClusterWare igb Ethernet driver is version 3.4.8, derived from http://sourceforge.net/projects/e1000/. Most noticeably, this newer driver eliminates the "HBO bit set" syslogged messages that were introduced by version 3.4.7.
3. The OpenMPI environment modules now define *MPI_HOME*, *MPI_LIB*, *MPI_INCLUDE*, and *MPI_SYSCONFIG*.
4. The file `/etc/beowulf/conf.d/sysctl.conf` now gets copied at boot time to every compute node as `/etc/sysctl.conf` to establish basic **sysctl** values. The master node's `/etc/sysctl.conf` serves as the initial contents of `/etc/beowulf/conf.d/sysctl.conf`.
5. Scyld ClusterWare 6 compute nodes can now function properly as NFS clients of a Red Hat RHEL5 NFS server.

New in Scyld ClusterWare 6.2.1 - Scyld Release 621g0000 - June 18, 2012

1. The base kernel is updated to 2.6.32-220.17.1.el6.621g0000. See https://rhn.redhat.com/errata/RHSA-2012-0571.html for details.
2. The Scyld ClusterWare Adaptec aacraid driver is version 1.1.7-29100, useable for the 6805 controller, and is derived from http://www.adaptec.com/en-us/support/raid/sas_raid/sas-6805/.

New in Scyld ClusterWare 6.2.0 - Scyld Release 620g0000 - May 17, 2012

1. The base kernel is 2.6.32-220.13.1.el6.620g0001. See https://rhn.redhat.com/errata/RHSA-2012-0481.html for details.
2. The Scyld ClusterWare igb Ethernet driver is version 3.4.7, derived from http://sourceforge.net/projects/e1000/.
3. The Scyld ClusterWare Adaptec aacraid driver is version 1.1.7-28801, useable for the 6805 controller, and is derived from http://www.adaptec.com/en-us/support/raid/sas_raid/sas-6805/.
4. The TORQUE release is version 2.5.10, derived from http://www.adaptivecomputing.com/resources/downloads/torque/. See the *User's Guide* Appendix B, *TORQUE Release Information* for details.
5. The OpenMPI release is version 1.6, derived from http://www.open-mpi.org/. See the *User's Guide* Appendix C, *OpenMPI Release Information* for details.
6. The MPICH2 release is version 1.4.1p1, derived from http://www.mcs.anl.gov/research/projects/mpich2/. See the *User's Guide* Appendix D, *MPICH2 Release Information* for details.

7. The MVAPICH2 release is version 1.8, derived from http://mvapich.cse.ohio-state.edu/. See the *User's Guide* Appendix E, *MVAPICH2 Release Information* for details. Since the Scyld ClusterWare 6 MVAPICH2 transport mechanism is *ssh*, the cluster administrator will likely need to configure the cluster to allow *ssh* access for non-root users. See the Section called *Optionally enable SSHD on compute nodes* and the *Administrator's Guide* for details.

Known Issues And Workarounds

The following are known issues of significance with the latest version of Scyld ClusterWare 6.9.1 and suggested workarounds.

Issues with bootmodule firmware

RHEL6 introduced externally visible discrete firmware files that are associated with specific kernel software drivers. When **modprobe** attempts to load a kernel module that contains such a software driver, and that driver determines that the controller hardware needs one or more specific firmware images (which are commonly found in `/lib/firmware`), then the kernel first looks at its list of built-in firmware files. If the desired file is not found in that list, then the kernel sends a request to the **udevd** daemon to locate the file and to pass its contents back to the driver, which then downloads the contents to the controller. This functionality is problematic if the kernel module is an `/etc/beowulf/config` *bootmodule* and is an Ethernet driver that is necessary to boot a particular compute node in the cluster. The number of `/lib/firmware/` files associated with every possible *bootmodule* module is too large to embed into the `initrd` image common to all compute nodes, as that burdens every node with a likely unnecessarily oversized `initrd` to download. Accordingly, the cluster administrator must determine which specific firmware file(s) are actually required for a particular cluster and are not yet built-in to the kernel, then add *firmware* directive(s) for those files. See the *Administrator's Guide* for details.

Managing environment modules .version files

Several Scyld ClusterWare packages involve the use of environment modules. This functionality allows for users to dynamically set up a shell's user environment for subsequent compilations and executions of applications, and for viewing the manpages for commands that are associated with those compilations and executions.

The Scyld packages are found in the various `/opt/scyld/`*package/* subdirectories, and for each package there are subdirectories organized by package version number, compiler suite type, and per-version per-compiler subdirectories containing the associated scripts, libraries, executable binaries, and manpages for building and executing applications for that package. The `/opt/scyld/modulefiles/`*package/* subdirectories contain per-package per-version per-compiler files that contain various pathname strings that are prepended to the shell's $PATH, $LD_LIBRARY_PATH, and $MANPATH variables that properly find those `/opt/scyld/`*package/* scripts, libraries, executable files, and manpages.

For example, **module load mpich2/intel/1.5** sets up the environment so that the **mpicc** and **mpirun** commands build and execute MPI applications using using the Intel compiler suite and the `mpich2` libraries specifically crafted for mpich2 version 1.5. The **module load** command also understands defaults. For example, **module load mpich2/gnu** defaults to use the *gnu* compiler and the mpich2 version specified by the contents of the file `/opt/scyld/modulefiles/mpich2/gnu/.version` (if that file exists). Similarly, **module load mpich2** first looks at the contents of `/opt/scyld/modulefiles/mpich2/.version` to determine the default compiler suite, then (supposing *gnu* is that default) looks at the contents of `/opt/scyld/modulefiles/mpich2/gnu/.version` to determine which mpich2 software version to use.

As a general rule, after updating one of these Scyld packages that employs environment modules, the associated `/opt/scyld/modulefiles/`*package*'s subdirectories' `.version` files remain untouched. The responsibility for updating any `.version` file remains with the cluster administrator, presumably after consulting with users. If the contents

of a `.version` points to a compiler suite or to a package version number that no longer exists, then a subsequent **module load** for that package which expects to use a default selection will fail with a message of the form:

```
ERROR:105: Unable to locate a modulefile
```

The user must then perform **module load** commands that avoid any reference to the offending `.version`, e.g., use the explicit **module load mpich2/intel/1.5**, until the cluster administrator resets the `.version` contents to the desired default. Each module-employing Scyld package installs sample files with the name `.version.`*versionNumber*.

The openmpi packages manage defaults differently. Suppose `openmpi-2.0-scyld` is currently version 2.0.1 and is updating to 2.0.2. Just as the default update behavior is to replace all 2.0.1 packages with the newer 2.0.2 packages, this openmpi-2.0 update also silently changes the `gnu`, `intel`, and `pgi .version` files which happen to specify the same major-minor version, e.g., those that specify version 2.0.1 are silently updated to the newer 2.0.2. If, however, the current `.version` files specify an older major-minor release, e.g., 1.10.4, then updating `openmpi-2.0-scyld` does not change any of these older major-minor `.version` specifiers.

Additionally, each set of openmpi-x.y-scyld packages maintain a major-minor symlink that points to the newest major-minor-release module file. For example, when `openmpi-2.0-scyld` version 2.0.1 is currently installed, then the `/opt/scyld/modulefiles/openmpi/gnu/2.0` symlink changes to the `2.0.1` module file. When `openmpi-2.0-scyld` updates to 2.0.2, then `/opt/scyld/modulefiles/openmpi/gnu/2.0` changes that symlink to point to the `2.0.2` module file. This convenient symlink allows for users to maintain job manager scripts that simply specify a major-minor number, e.g., **module load openmpi/intel/2.0**, that survives updates from openmpi-2.0-scyld 2.0.1 to 2.0.2 to 2.0.3, et al, versus using scripts that contain the more specific **module load openmpi/intel/2.0.1** that break when 2.0.1 packages update to 2.0.2.

Note that each compiler suite can declare a different default package version, although most commonly the cluster administrator edits the `/opt/scyld/modulefiles/`*package/compiler/* `.version` files so that for a given *package*, all compiler suites reference the same default version number.

One method to check the current package defaults is to execute:

```
cd /opt/scyld/modulefiles
module purge
module avail
for m in $(ls); do module load $m; done
module list
module purge
```

and then verify each loaded default against the **module avail** available alternatives.

Installing and managing concurrent versions of packages

Scyld ClusterWare distributes various repackaged Open Source software suites, including several variations of "MPI", e.g., `openmpi`, `mpich-scyld`, `mpich2-scyld`, `mvapich2-scyld`. Users manage the selection of which software stack to use via the **module load** command. See the Section called *Managing environment modules .version files* for details.

By default, **install-scyld -u** updates each existing package with the newest version of that package by installing the newest version and removing all earlier (i.e., lower-numbered) versions, thereby retaining only a single version of each software suite. For example, the `openmpi-2.0-scyld` packages update to the latest 2.0.x version (major 2, minor 0, version x), and the `openmpi-1.10-scyld` packages update to the latest latest 1.10.y (major 1, minor 10, version y). Thus, a default update of package `openmpi-2.0` installs the newest version 2.0.x and removes earlier versions of 2.0, leaving versions 1.10.x, 1.8.x, 1.7.x, etc. untouched.

Because Scyld ClusterWare installs a package's files into unique `/opt/scyld/`*package/version* version-specific directories, this permits multiple versions of each major-minor package to potentially co-exist on the master node, e.g., openmpi versions 2.0.2 and 2.0.1. Each such *package/version* subdirectory contains one or more *compiler* suite subdirectories, e.g., `gnu`, `intel`, and `pgi`, and each of those contain scripts, libraries, executable binaries, and manpages associated with that particular package, version, and compiler suite.

Some customers (albeit rarely) may wish to install multiple concurrent x.y.z versions for a given x.y major-minor because specific applications might only work properly when linked to a specific version, or applications might perform differently for different versions. For example, to retain openmpi version 2.0.1 prior to using **install-scyld -u** or **yum update**, which might replace those 2.0.1 packages with a newer 2.0.z version, first edit `/etc/yum.conf` to add the line:

```
exclude=openmpi-2.0-scyld*
```

which blocks **yum** from updating any and all currently installed `openmpi-2.0-scyld` packages. If the cluster administrator wishes to install (for example) the 2.0.2 packages and not disturb the 2.0.1 installation, then temporarily comment-out that `exclude=openmpi-2.0-scyld*` line and execute:

```
yumdownloader openmpi-2.0-scyld-*2.0.2*
```

and then re-enable the `exclude=` line to again protect against any inadvertant `openmpi-2.0-scyld` updates. Manually install these additional downloaded rpms using **rpm -iv** -- and *not* use **rpm -Uv** or even **yum install**, as both of those commands will remove older `openmpi-2.0-scyld` packages.

Issues with OpenMPI

Scyld ClusterWare distributes repackaged releases of the Open Source OpenMPI, derived from http://www.open-mpi.org/. The Scyld ClusterWare distributions consist of a `openmpi-`*x.y*`-scyld` base package for the latest OpenMPI version *x.y.z*, plus several compiler-environment-specific packages for `gnu`, `intel`, and `pgi`. For example, the distribution of OpenMPI non-psm2 version 2.0.1 consists of the base rpm `openmpi-2.0-scyld-2.0.1` and the various compiler-specific rpms: `openmpi-2.0-scyld-gnu-2.0.1`, `openmpi-2.0-scyld-intel-2.0.1`, and `openmpi-2.0-scyld-pgi-2.0.1`.

Scyld ClusterWare distributes versions `openmpi-2.0-scyld`, `openmpi-1.10-scyld`, and `openmpi-1.8-scyld`, as well as `openmpi-psm2-2.0-scyld` and `openmpi-psm2-1.10-scyld` for clusters using the Intel Omni-Path Architecture (OPA) networking (which also requires `hfi1-psm` rpms from the Intel OPA software bundle).

A set of `openmpi-`*x.y*`-scyld` packages installs *x.y.z* version-specific libraries, executable binaries, and manpages for each particular compiler into `/opt/scyld/openmpi/`*version/compiler* subdirectories, and installs modulefiles into `/opt/scyld/modulefiles/openmpi/`*compiler/version* files. The directory `/opt/scyld/openmpi/`*version*`/examples/` contains source code examples. The `openmpi-psm2` packages similarly install into `/opt/scyld/openmpi-psm2/` and `/opt/scyld/modulefiles/openmpi-psm2/`.

The modulefiles appends the current shell's $PATH, $LD_LIBRARY_PATH, and $MANPATH with pathnames that point to the associated compiler-specific version-specific `/opt/scyld/openmpi/`*version/compiler/* (or `/opt/scyld/openmpi-psm2/`*version/compiler/*) subdirectories. This permits multiple versions to co-exist on the master node, with each variation being user-selectable at runtime using the **module load** command.

Many customers support multiple OpenMPI versions because some applications might only work properly when linked to specific OpenMPI versions. Sometimes an application needs only to be recompiled and relinked against a newer version of the libraries. Other applications may have a dependency upon a particular OpenMPI version that a simple recompilation won't fix. The cluster administrator can specify which compiler and version is the default by manipulating the contents of the various `.version` files in the `/opt/scyld/modulefiles/openmpi/` (or `openmpi-psm2`) subdirectories. For example, a **module load openmpi** might default to specify version 1.10.4 of the gnu libraries, while **module load openmpi-psm2** might default to specify version 2.0.1 of the intel libraries, while at the same time a version-specific **mod-**

ule load openmpi-psm2/gnu/1.10.4 or **module load openmpi/pgi/1.8.8** allows the use of different compilers and libraries for different OpenMPI versions.

The latest Open Source release of `openmpi-2.0-scyld` is a "mandatory" install, and `openmpi-1.10-scyld`, `openmpi-1.8-scyld`, `openmpi-psm2-2.0-scyld`, and `openmpi-psm2-1.10-scyld` are "optional" and can be manually installed by the cluster administrator using (for example) **yum install openmpi-psm2-1.10-scyld-***. A subsequent **yum update** will update each and every installed `openmpi-`*x.y*`-scyld` and installed `openmpi-psm2-`*x.y*`-scyld` to the latest available version *x.y.z*. If the cluster administrator wishes to retain additional *x.y.z* releases within an *x.y* family, then instead of doing **yum update**, the administrator should **yum update --exclude=openmpi*scyld-***, then download specific rpms from the yum repo as desired using **yumdownloader**, and then manually install (not update) the rpms using **rpm -i**. Note that the use of **yumdownloader** and **rpm -i** is necessary because doing a simple (for example) **yum install openmpi-1.10-scyld-1.10.4** will not, in fact, execute a simple *install* and retain older 1.10.z packages. Rather, it actually executes an *update* and removes any and all older installed versions of `openmpi-1.10-scyld-1.10.`*z* rpms.

Issues with Mpich

Beginning with Red Hat RHEL6 Update 6 and CentOS 6.6, the base distribution includes an *mpich* package, currently version 3. Scyld ClusterWare distributes three versions of mpich: *mpich* version 1.2.7p1, *mpich2-scyld* with some version 2 enhancements, and *mpich-scyld* version 3, which typically is a newer version than the RHEL6 or CentOS6 *mpich* version 3.

The RHEL6/CentOS6 *mpich* does not conflict with either Scyld ClusterWare *mpich2-scyld* or *mpich-scyld* because of their different names, but it does conflict with the Scyld ClusterWare *mpich*. Any attempt to install the new RHEL6 or CentOS6 *mpich* (version 3) triggers an update (and removal) of the older Scyld ClusterWare *mpich* (version 1.2.7p1), and that update fails because other Scyld ClusterWare packages have a dependency on Scyld ClusterWare mpich-1.2.7p1.

The cluster administrator can either manually remove all Scyld ClusterWare mpich-1.2.7p1 packages and proceed with a normal install or update of the RHEL6 or CentOS6 base distribution *mpich*, or the administrator can install or update RHEL6 or CentOS6 and explicitly exclude its *mpich* package. We recommend this latter approach for its simplicity, and because it retains the Scyld ClusterWare mpich-1.2.7p1 for users, and because there is no obvious need for the base distribution's *mpich* when the Scyld ClusterWare *mpich-scyld* version 3 will be the same or newer than the RHEL6/CentOS6 *mpich*. The straightforward way to exclude this is to edit the file `/etc/yum.conf` and add the line:

```
exclude=" mpich-3* "
```

NOTE: The Scyld ClusterWare mpich-1.2.7p1 packages are deprecated and will eventually be unavailable in future Scyld ClusterWare releases. We encourage user to migrate applications to *mpich-scyld* or to *openmpi*, both of which which support inter-thread communication using either Ethernet or Infiniband.

Issues with Scyld process migration in heterogeneous clusters

In a homogeneous cluster, all nodes (master and compute) are identical server models, including having identical amounts of RAM. In a heterogeneous cluster, the nodes are not all identical. The advantage of a homogeneous cluster is simplicity in scheduling work on the nodes, since every node is identical and interchangeable. However, in the real world, many if not most clusters are heterogeneous. Some nodes may have an attached GPU or different amounts of available RAM, or may even be different server models with different x86_64 processor technologies.

Scyld ClusterWare users have always needed to be aware of potential problems running applications on heterogeneous clusters. For example, applications expecting to employ a GPU have needed to take care to execute only on nodes with an attached GPU, and an application that is specifically compiled or linked to libraries that employ newer x86_64 instructions that are not universally understood by every x86_64 processor must ensure that the application only execute on the nodes with processors that understand those newer instructions.

However, RHEL6 heterogeneous clusters present a new set of challenges to users. The essence of the issue is this: when a software thread begins execution, some libraries (e.g., libc) make a one-time determination of which processor model is being used, and the library self-configures certain routines (e.g., strcmp) to use implementations that exploit processor model-specific instructions for optimal performance. However, if the software thread subsequently migrates to a different node in the cluster, then the thread's one-time determination state migrates to the destination node. If the destination node does not support the same x86_64 instructions that are supported by the original node, then the software thread will likely suffer a fatal "invalid opcode" trap if it attempts to execute one of these optimized library routines. Scyld ClusterWare performs such a thread migration through the use of the bproc_move() or bproc_rfork() library routines found in `libbproc`. These bproc routines are employed by the MPICH and MVAPICH libraries, and by the **bpcp** command.

One workaround to the problem is simple: use MPICH2 instead of MPICH, and use MVAPICH2 instead of MVAPICH, or use OpenMPI instead of either. None of those alternative MPI execution environments employ bproc_move() or bproc_rfork(). Another workaround is to execute all threads of a multithreaded application on identical nodes. More subtly, another workaround is to start executing the application on a node that employs the oldest processor technology; thus, any subsequent thread migration is guaranteed to find a node with a processor that supports a superset of the instructions supported by that initial node.

As for the **bpcp** command, in Scyld ClusterWare 6.3.0 and beyond, **bpcp** links with a special Scyld libc that uses only generic, universally acceptable x86_64 instructions. Users may similarly link applications to this special library by adding:

```
Xlinker -rpath=/lib64/scyld
```

as a linker option.

Issues with MVAPICH2 and mpirun_rsh or mpispawn

Scyld ClusterWare has applied a workaround to **mpiexec** to fix a problem with MPICH2 and MVAPICH2 exec'ing the application executable binary across NFS. The problem is *not* fixed for launching the application using **mpirun_rsh** or **mpispawn**, which likely will result in the application hanging as it attempts to execve() the application. We strongly encourage using only **mpiexec** to launch MPICH2 and MVAPICH2 applications.

Issues with MVAPICH2 and CPU Sets

MVAPICH2-2.1 introduces an algorithm to determine CPU topology on the node, and this new algorithm does not work properly for older Mellanox controllers and firmware, resulting in software threads not spreading out across a node's cores by default.

Prior to updating to MVAPICH2-2.1 or newer, the cluster administrator should determine the potential vulnerability to this problem. For each node that contains an Infiniband controller, execute **ibstat**, and if the first output line is:

```
CA 'mthca0'
```

then that node *may* exhibit the problem. The cluster administrator has two choices: either avoid updating the `mvapich2-scyld` packages (keeping in mind that the `mvapich2-psm-scyld` packages can be updated, as those packages are only used by QLogic Infiniband controllers, which don't have the problem); or update `mvapich2-scyld`, execute tests to determine if the problem exists for those Mellanox *mthca* nodes, and if the problem does exist, then instruct users to employ explicit CPU Mapping. See http://mvapich.cse.ohio-state.edu/static/media/mvapich/mvapich2-2.1-userguide.html#x1-540006.5 fo details.

Issues with beosetup

The **beosetup** tool is deprecated in Scyld ClusterWare 5 and is eliminated from Scyld ClusterWare 6.

Issues with xpvm

xpvm is not currently supported in ClusterWare 6.

Issues with ptrace

Cluster-wide **ptrace** functionality is not yet supported in Scyld ClusterWare 6. For example, you cannot use a debugger running on the master node to observe or manipulate a process that is executing on a compute node, e.g., using **gdb -p procID**, where *procID* is a processID of a compute node process. **strace** does function in its basic form, although you cannot use the **-f** or **-F** options to trace forked children if those children move away from the parent's node.

Issues with rsh

Currently, **rsh** is unavailable as a communication method between nodes. Consider **ssh** as an alternative.

Issues with IP Forwarding

If the *beowulf* service has started, then a subsequent **service iptables stop** (or **restart**) will hang because it attempts to unload the ipt_MASQUERADE kernel module while the *beowulf* service is using (and not releasing) that module. For a workaround, edit `/etc/sysconfig/iptables-config` to change:

```
    IPTABLES_MODULES_UNLOAD="yes"
to: IPTABLES_MODULES_UNLOAD="no"
```

Issues with kernel modules

The **modprobe** command uses `/usr/lib/`uname -r`/modules.dep.bin` to determine the pathnames of the specified kernel module and that module's dependencies. The **depmod** command builds the human-readable `modules.dep` and the binary `module.dep.bin` files, and it should be executed *on the master node* after installing any new kernel module.

Executing **modprobe** on a compute node requires additional caution. The first use of **modprobe** retrieves the current `modules.dep.bin` from the master node using bproc's *filecache* functionality. Since any subsequent **depmod** on the master node rebuilds `modules.dep.bin`, then a subsequent **modprobe** on a compute node will only see the new `modules.dep.bin` if that file is copied to the node using **bpcp**, or if the node is rebooted and thereby silently retrieves the new file.

In general, you should not execute **depmod** on a compute node, since that command will only see those few kernel modules that have previously been retrieved from the master node, which means the node's newly built `modules.dep.bin` will only be a sparse subset of the master node's full `module.dep.bin`. Bproc's *filecache* functionality will always properly retrieve a kernel module from the master node, as long as the node's `module.dep.bin` properly specifies the pathname of that module, so the key is to have the node's `module.dep.bin` be a current copy of the master's file.

Issues with port numbers

Scyld ClusterWare employs several daemons that execute in cooperating pairs: a server daemon that executes on the master node, and a client daemon that executes on compute nodes. Each daemon pair communicates using tcp or udp through a presumably unique port number. By default, Scyld ClusterWare uses ports 932 (*beofs2*), 933 (*bproc*), 3045 (*beonss*), and 5545 (*beostats*). In the event that one or more of these port numbers collides with a non-Scyld ClusterWare daemon using the same port number, the cluster administrator can override Scyld ClusterWare default port numbers to use different, non-colliding unused ports using the `/etc/beowulf/config` file's *server* directive. See **man beowulf-config** and `/etc/beowulf/config` for a discussion of the *server* directive.

The official list of assigned ports and their associated services is http://www.iana.org/assignments/port-numbers, and `/etc/services` is a list shipped with your base distribution. However, the absence in either list of a specific port number is no guarantee that the port will not be used by some software on your cluster. Use **lsof -i :***portNumber* to determine if a particular port number is in active use.

A common collision is with *beofs2* port 932 or *bproc* port 933, since the **rpc.statd** or **rpc.mountd** daemons may randomly grab either of those ports before ClusterWare can grab them. However, ClusterWare automatically recognizes the conflict and tries alternative ports until it finds an unused port. If this flexible search causes problems with other daemons, you can edit `/etc/beowulf/config` to specify a tentative override value using the *server beofs2* or *server bproc* directive, as appropriate.

Less common are collisions with *beonss* port 3045 or *beostats* port 5545. The *server beonss* and *server beostats* override values are used as-specified and not adjusted by ClusterWare at runtime.

Issues with TORQUE

Scyld ClusterWare repackages the TORQUE resource manager available from Adaptive Computing, http://www.adaptivecomputing.com/support/download-center/torque-download/. In every new TORQUE release, the Adaptive Computing developers fix bugs, add new features, and on occasion change configuration and scripting options. View the Adaptive Computing's TORQUE Release Notes and Changelog in the Scyld ClusterWare *User's Guide Appendix B. TORQUE Release Information*.

TORQUE version 4.2.0 exhibits a problem with **mpiexec** and MPICH. Currently, the only known workaround is to alternatively use MPICH2 or OpenMPI.

Beginning with version 684g0000 (ClusterWare 6.8.4), Scyld ClusterWare has changed its naming of TORQUE packages. The older *torque-scyld* and *torque-nocpuset-scyld* became version-specific *torque-4-scyld* and *torque-4-nocpuset-scyld*, *torque-5-scyld* and *torque-5-nocpuset-scyld*, and *torque-6-scyld* and *torque-6-nocpuset-scyld*. One and *only* one of these TORQUE packages *must* be installed on the master node at any point in time. If the older *torque-scyld* or *torque-nocpuset-scyld* is currently installed, then you must do an explicit one-time install of one of the newer package names, e.g., **yum install torque-4-scyld**, which will install the new package and remove the older *torque-scyld* package.

Issues with Spanning Tree Protocol and portfast

Network switches with Spanning Tree Protocol (STP) enabled will block packets received on a port for the first 30 seconds after the port comes online, giving the switch and the Spanning Tree algorithm time to determine if the device on the new link is a switch, and to determine if Spanning Tree will block or forward packets from this port. This is done to prevent "loops" which can cause packets to be endlessly repeated at a high rate and consume all network bandwidth. Each time the link goes down and comes back up, another 30-second blocking delay occurs. This delay can prevent PXE/DHCP from obtaining an IP address, or can prevent the node's initial kernel from downloading its initial root filesystem, which results in the node endlessly iterating in the early boot sequence, or can delay the node's ongoing *filecache* provisioning of libraries to the node.

We recommend disabling STP if feasible. If not feasible, then we recommend reconfiguring the switch to use *Rapid STP* or *portfast*, which avoids the 30-second delay, or employing some other port mode that will forward packets as a port comes up. There is no generic procedure for enabling these options. For Cisco switches, see http://www.cisco.com/en/US/products/hw/switches/ps700/products_tech_note09186a00800b1500.shtml. For other switch models, see the model-specific documentation.

If that reconfiguration is also not possible, you may need to increase the default Scyld ClusterWare timeout used by the node to a value safely greater than the STP delay: e.g., add *rootfs_timeout=120 getfile_timeout=120* to the `/etc/beowulf/config` *kernelcommandline* entry to increase the timeouts to 120 seconds.

Issues with Gdk

If you access a cluster master node using **ssh -X** from a workstation, some graphical commands or program may fail with:

```
Gdk-ERROR **: BadMatch (invalid parameter attributes)
  serial 798 error_code 8 request_code 72 minor_code 0
Gdk-ERROR **: BadMatch (invalid parameter attributes)
  serial 802 error_code 8 request_code 72 minor_code 0
```

Remedy this by doing:

```
export XLIB_SKIP_ARGB_VISUALS=1
```

prior to running the failing program. If this workaround is successful, then consider adding this line to `/etc/bashrc` or to `~/.bashrc`. See https://bugs.launchpad.net/ubuntu/+source/xmms/+bug/58192 for details.

Caution when modifying Scyld ClusterWare scripts

Scyld ClusterWare installs various scripts in `/etc/beowulf/init.d/` that **node_up** executes when booting each node in the cluster. Any site-local modification to one of these scripts will be lost when a subsequent Scyld ClusterWare update overwrites the file with a newer version. If a cluster administrator believes a local modification is necessary, we suggest:

1. Copy the to-be-edited original script to a file with a unique name, e.g.:

   ```
   cd /etc/beowulf/init.d
   cp 20ipmi 20ipmi_local
   ```

2. Remove the executable state of the original:

   ```
   beochkconfig 20ipmi off
   ```

3. Edit `20ipmi_local` as desired.

4. Thereafter, subsequent Scyld ClusterWare updates may install a new `20ipmi`, but that update will not re-enable the non-executable state of that script. The locally modified `20ipmi_local` remains untouched. However, keep in mind that the newer Scyld ClusterWare version of `20ipmi` may contain fixes or other changes that need to be reflected in `20ipmi_local` because that edited file was based upon an older Scyld ClusterWare version.

Caution using tools that modify config files touched by Scyld ClusterWare

Software tools exist that might make modifications to various system configuration files that Scyld ClusterWare also modifies. These tools do not have knowledge of the Scyld ClusterWare specific changes and therefore may undo or cause damage

to the changes or configuration. Care must be taken when using such tools. One such example is **/usr/sbin/authconfig**, which manipulates `/etc/nsswitch.conf`.

Scyld ClusterWare modifies these system configuration files at install time:

```
/etc/exports
/etc/nsswitch.conf
/etc/security/limits.conf
/etc/sysconfig/syslog
```

Additionally, Scyld ClusterWare uses **chkconfig** to enable *nfs*.

Running nscd service on master node may cause kickbackdaemon to misbehave

The **nscd** (Name Service Cache Daemon) service executes by default on the master node, and **/usr/sbin/nscd** executes by default on each compute node via `/etc/beowulf/init.d/09nscd`. However, if this service is also enabled and executes on the master node, then it may cause the Scyld ClusterWare name service **kickbackdaemon** to misbehave.

Accordingly, when the ClusterWare starts, if it detects that the **nscd** service is running on the master node, then ClusterWare automatically stops that service. ClusterWare does not permanently disable that service on the master node. To do that:

```
chkconfig nscd off
```

Note: even after stopping **nscd** on the master node,

```
service nscd status
```

will report that **nscd** is running because the daemon continues to execute on each compute node, as controlled by `/etc/beowulf/init.d/09nscd`.

Scyld ClusterWare MVAPICH CPU affinity management

The default MVAPICH behavior is to assign threads of each multithreaded job to specific CPUs in each node, starting with cpu0 and incrementing upward. Keeping threads pinned to a specific CPU may be an optimal NUMA and CPU cache strategy for nodes that are dedicated solely to a single job, it is usually suboptimal if multiple multithreaded jobs share a node, as each job's threads get permanently assigned to the same low-numbered CPUs. The default Scyld ClusterWare MVAPICH behavior is to not impose strict CPU affinity assignments, but rather to allow the kernel CPU scheduler to migrate threads as it sees fit to load-balance the node's CPUs as workloads change over time.

However, the user may override this default using:

```
export VIADEV_ENABLE_AFFINITY=1
```

Conflicts with base distribution of OpenMPI

Scyld ClusterWare 6.9.1 includes MPI-related packages that conflict with certain packages in the RHEL6 or CentOS6 base distribution.

If **yum** informs you that it cannot install or update Scyld ClusterWare because various **mpich** and **mpiexec** packages conflict with various **openmpi** packages from the base distribution, then run the command:

```
yum remove openmpi mvapich
```

to remove the conflicting base distribution packages, then retry the *groupupdate* of Scyld-ClusterWare.

Beofdisk does not support local disks without partition tables

Currently, **beofdisk** only supports disks that already have partition tables, even if those tables are empty. Compute nodes with preconfigured hardware RAID, where partition tables have been created on the LUNs, should be configurable. Contact Customer Service for assistance with a disk without partition tables.

Issues with bproc and the getpid() syscall

BProc interaction with *getpid()* may return incorrect processID values.

Details: The Red Hat's glibc implements the *getpid()* syscall by asking the kernel once for the current processID value, then caching that value for subsequent calls to *getpid()*. If a program calls *getpid()* before calling *bproc_rfork()* or *bproc_vrfork()*, then bproc silently changes the child's processID, but a subsequent *getpid()* continues to return the former cached processID value.

Workaround: do not call *getpid()* prior to calling *bproc_[v]rfork*.