Elastic Storage Server
Version 4.5

Quick Deployment Guide

SC27-8580-03

Elastic Storage Server
Version 4.5

Quick Deployment Guide

SC27-8580-03

Note
Before using this information and the product it supports, read the information in "Notices" on page 41.

This edition applies to version 4.5.x, 3.5, and 3.0 of the Elastic Storage Server (ESS) for Power®, and to all subsequent releases and modifications until otherwise indicated in new editions.

IBM welcomes your comments; see the topic "How to submit your comments" on page viii. When you send information to IBM, you grant IBM a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Contents

Tables

About this information

This information guides you in quickly installing, or upgrading to, version 4.5.x of the Elastic Storage Server (ESS).

For detailed ESS installation and upgrade information, see *Deploying the Elastic Storage Server*.

Who should read this information

This information is intended for experienced system installers and upgraders who are familiar with ESS systems.

Prerequisite and related information

ESS information

The ESS 4.5.x library consists of these information units:

- *Deploying the Elastic Storage Server*, SC27-6659
- *Elastic Storage Server: Quick Deployment Guide*, SC27-8580
- *IBM Spectrum Scale RAID: Administration*, SC27-6658

For more information, see IBM® Knowledge Center:

http://www-01.ibm.com/support/knowledgecenter/SSYSP8_4.5.0/sts45_welcome.html

For the latest support information about IBM Spectrum Scale™ RAID, see the IBM Spectrum Scale RAID FAQ in IBM Knowledge Center:

http://www.ibm.com/support/knowledgecenter/SSYSP8/sts_welcome.html

Related information

For information about:

- IBM Spectrum Scale, see IBM Knowledge Center:
  http://www.ibm.com/support/knowledgecenter/STXKQY/ibmspectrumscale_welcome.html
- IBM POWER8® servers, see IBM Knowledge Center:
  http://www.ibm.com/support/knowledgecenter/POWER8/p8hdx/POWER8welcome.htm
- The DCS3700 storage enclosure, see:
  - *System Storage® DCS3700 Quick Start Guide*, GA32-0960-03:
    http://www.ibm.com/support/docview.wss?uid=ssg1S7004915
  - *IBM System Storage DCS3700 Storage Subsystem and DCS3700 Storage Subsystem with Performance Module Controllers: Installation, User's, and Maintenance Guide* , GA32-0959-07:
    http://www.ibm.com/support/docview.wss?uid=ssg1S7004920
- The IBM Power Systems™ EXP24S I/O Drawer (FC 5887), see IBM Knowledge Center :
  http://www.ibm.com/support/knowledgecenter/8247-22L/p8ham/p8ham_5887_kickoff.htm
- Extreme Cluster/Cloud Administration Toolkit (xCAT), go to the xCAT website :
  http://sourceforge.net/p/xcat/wiki/Main_Page/

Conventions used in this information

Table 1 describes the typographic conventions used in this information. UNIX file name conventions are used throughout this information.

Table 1. Conventions

Convention	Usage
bold	**Bold** words or characters represent system elements that you must use literally, such as commands, flags, values, and selected menu options. Depending on the context, **bold** typeface sometimes represents path names, directories, or file names.
bold underlined	**bold underlined** keywords are defaults. These take effect if you do not specify a different keyword.
`constant width`	Examples and information that the system displays appear in `constant-width` typeface. Depending on the context, `constant-width` typeface sometimes represents path names, directories, or file names.
italic	*Italic* words or characters represent variable values that you must supply. *Italics* are also used for information unit titles, for the first use of a glossary term, and for general emphasis in text.
<*key*>	Angle brackets (less-than and greater-than) enclose the name of a key on the keyboard. For example, **<Enter>** refers to the key on your terminal or workstation that is labeled with the word *Enter*.
\	In command examples, a backslash indicates that the command or coding example continues on the next line. For example: `mkcondition -r IBM.FileSystem -e "PercentTotUsed > 90" \` `-E "PercentTotUsed < 85" -m p "FileSystem space used"`
{*item*}	Braces enclose a list from which you must choose an item in format and syntax descriptions.
[*item*]	Brackets enclose optional items in format and syntax descriptions.
<**Ctrl-***x*>	The notation <**Ctrl-***x*> indicates a control character sequence. For example, <**Ctrl-c**> means that you hold down the control key while pressing <**c**>.
item...	Ellipses indicate that you can repeat the preceding item one or more times.
|	In *synopsis* statements, vertical lines separate a list of choices. In other words, a vertical line means *Or*. In the left margin of the document, vertical lines indicate technical changes to the information.

How to submit your comments

Your feedback is important in helping us to produce accurate, high-quality information. You can add comments about this information in IBM Knowledge Center:

http://www.ibm.com/support/knowledgecenter/SSYSP8/sts_welcome.html

To contact the IBM Spectrum Scale development organization, send your comments to the following email address:

scale@us.ibm.com

Deploying the Elastic Storage Server - for experienced users

This topic includes a concise set of deployment instructions for those who are familiar with Elastic Storage Server (ESS) systems.

In these instructions:

- All version numbers shown are examples. The version depends on the release and edition that is being deployed.
- Node names `ems1`, `gssio1`, and `gssio2` are examples. Each environment could have its own unique naming conventions.

Follow these high-level steps:

1. Complete the prerequisite tasks.
2. Unpack the ESS install/upgrade software from FixCentral at http://www-933.ibm.com/support/fixcentral/swg/selectFixes?parent=Software%2Bdefined%2Bstorage&product=ibm/StorageSoftware/IBM+Spectrum+Scale+RAID&release=All&platform=All&function=all
3. Obtain the required Kernel Errata. For more information, see Appendix E, "Obtaining kernel for system upgrades," on page 37.
4. Obtain the systemd update and apply it on each ESS node.

   Important: Doing the systemd update is mandatory and the systemd update must be applied to each ESS node.
5. Complete one of the following tasks:
   a. Install the ESS system.
   b. Upgrade the ESS system.
6. To facilitate installation and deployment, ESS release 4.5 ships with the ESS Installation and Deployment Toolkit. The toolkit is a menu-driven interface that can perform some steps of install and deploy tasks. To start the Toolkit, issue the **`gssutils`** command at the command line. Refer to the **`gssutils`** man page for details.
   a. The Toolkit can be used to help and simplify some of the steps; however, this document and CLI should be used as the primary installation and deployment instructions and method.
   b. The steps for installing the management server software (provided in the section titled "Install the management server software " on page 2) must be done outside of the toolkit.
   c. Cleaning of the xCAT configuration and associated configurations (provided in step 2. of the section titled "Install the ESS system" on page 2) must be done outside of the toolkit.
   d. Review this guide to become familiar with the steps before using the toolkit.
7. Call home is supported for the servers in the IBM Elastic Storage™ Server. When properly enabled and configured, server platform events (power, cooling, processor, memory) will automatically be reported to IBM when they reach a service action required state.

   Note: Errors associated with devices and adapters within the servers, or any errors associated with the expansion I/O drawers and drives are not supported in this initial release.

   A Lab Based Services engagement is required to configure and initialize the call home application after installing or upgrading to ESS 4.5. Contact your IBM Sales representative to arrange this engagement.

Complete the prerequisite tasks

Complete these tasks before proceeding:

1. Ensure nodes are properly prepared for deployment.
   - EMS and IO Server node network requirements are met with correct `/etc/hosts` entries in EMS node. Review and address the items described in Table 3 on page 23
   - HMC is properly configured for the EMS and I/O server nodes and partition names are correctly set
   - Nodes are powered up
2. Obtain a Red Hat Enterprise Linux 7.1 ISO image file (e.g., rhel-server-7.1-ppc64-dvd.iso) or DVD for 64-bit IBM Power Systems architecture. The ISO or DVD is used to upgrade EMS node as well as upgrade or deploy I/O server nodes.
3. Obtain the ESS software archive from FixCentral at http://www-933.ibm.com/support/fixcentral/swg/selectFixes?parent=Software%2Bdefined%2Bstorage&product=ibm/StorageSoftware/IBM+Spectrum+Scale+RAID&release=All&platform=All&function=all.
4. Review the list of known issues for the ESS version you are installing or to which you are upgrading. For more information, see Appendix A, "Known issues," on page 15.

Install the management server software

1. Unpack the ESS software archive:

   ```
   tar -zxvf gss_install-4.5.2_ppc64_advanced_20170317T125741Z.tgz
   ```

2. Check the MD5 checksum:

   ```
   md5sum -c gss_install-4.5.2_ppc64_advanced_20170317T125741Z.md5
   ```

3. Make sure the `/opt/ibm/gss/install` directory is clean:

   ```
   /bin/sh gss_install-gss_install-4.5.2_ppc64_advanced_20170317T125741Z --remove
   ```

4. Extract the ESS packages and accept the license as shown below. By default it will be extracted to:

   ```
   /opt/ibm/gss/install directory
   /bin/sh gss_install-4.5.2_ppc64_advanced_20170317T125741Z —-text-only
   ```

5. For install and deployment go to the section titled "Install the ESS system" To upgrade an existing ESS system, go to the section titled "Upgrade the ESS system" on page 8.

Install the ESS system

Follow these steps to perform a new installation of the ESS software on a management server node and I/O server nodes. Node hostnames **ems1**, **gssio1**, and **gssio2** are examples. Each environment could have its own unique naming conventions. For an xCAT command such as **updatenode**, use an xCAT hostname. For the IBM Spectrum Scale commands (those start with mm), use an IBM Spectrum Scale hostname. For example, **ems1** is an xCAT hostname (typically a hostname associated with the management interface) and **ems1-hs** is the corresponding IBM Spectrum Scale hostname (typically a hostname associated with the high speed interface).

1. Copy the **gssdeploy** script and **customize** it for your environment by editing it:

   ```
   cp /opt/ibm/gss/install/samples/gssdeploy /var/tmp
   chmod +x /var/tmp/gssdeploy
   ```

   `/var/tmp` is a sample directory name. You can specify a different directory name. The **gssdeploy** script uses a directory called `/tmp/gssdeploy`, so do *not* copy the script to `/tmp`. Copy the ISO (if ISO is used) into the directory defined in the customized section of the script. By default it is set to `/opt/ibm/gss/iso`.

2. Clean the current xCAT installation and associated configuration to remove any preexisting xCAT configuration:

   ```
   /var/tmp/gssdeploy -c
   ```

3. Update the ESS repositories on the management server node:

```
cd /opt/ibm/gss/install
installer/gssinstall -m manifest -u
```

4. Run the **gssdeploy** script:

```
/var/tmp/gssdeploy -x
```

If any step fails, address the issue by looking at the error message. Depending on the nature of the problem, you may need to restart from the last completed step, or you may need to start from the beginning. If you are starting from the beginning, restart from step 2 (gssdeploy -c) of this section.

5. Log out and then log back in to acquire the environment updates.
6. Set up the Kernel Errata repository. For information about obtaining and packaging the kernel update, see Appendix E, "Obtaining kernel for system upgrades," on page 37. For information about staging the kernel update for installation, see Appendix F, "Instructions for installing the ESS Red Hat Linux Errata Kernel Update," on page 39.
7. Update the management server node. Here **ems1** is the xCAT hostname. This step updates the node system profile, prepares OFED and **gplbin rpm** and installs the required rpms.

```
updatenode ems1 -P gss_updatenode
```

| Use **systemctl reboot** to reboot the management server node and run this step again as shown
| below. This additional step is required to account for changes in kernel, removal of drivers, etc.

```
updatenode ems1 -P gss_updatenode
```

8. Update OFED on the management server node:

```
updatenode ems1 -P gss_ofed
```

9. Update the IP RAID Adapter firmware on the management server node:

```
updatenode ems1 -P gss_ipraid
```

| 10. Use **systemctl reboot** to reboot the management server node.

Deploy the I/O server nodes

1. Before initiating the deployment of the I/O server nodes, verify that the attached storage enclosures are powered off.
2. Deploy on the I/O server nodes using the customized deploy script:

```
./gssdeploy -d
```

3. Run:

```
nodestat gss_ppc64
```

The installation is complete when **nodestat** displays `sshd` (it may take approximately 30 minutes) for all I/O server nodes. Here **gss_ppc64** is the xCAT nodegroup containing I/O Server nodes. To follow the progress of an I/O server install, you can tail the console log by doing

| `tailf /var/log/consoles/gssio1`

| where gssio1 is an example default I/O node name.

4. At the end of the deployment wait approximately five minutes and reboot the node:

```
xdsh gss_ppc64 systemctl reboot
```

| **Note:** Make sure the xCAT post-installation script is complete before rebooting the nodes. You can
| check xCAT post process running on the I/O server nodes as follows:

| `xdsh gss_ppc64 "ps -eaf | grep -v grep | grep xcatpost"`

If there are any processes still running, wait for them to complete.

5. Update the IP RAID Adapter firmware on the I/O server nodes:
   ```
   updatenode gss_ppc64 -P gss_ipraid
   ```
6. Once rebooted, verify the installation by running **gssinstallcheck:**
   ```
   gssinstallcheck -G ems1,gss_ppc64
   ```
   Check for any error with the following:
   a. Installed packages
   b. Linux kernel release
   c. OFED level
   d. IPR SAS FW
   e. ipraid RAID level
   f. ipraid RAID status
   g. IPR SAS queue depth
   h. System firmware
   i. System profile setting
   j. Host adapter driver
   k. Storage Adapters with 64-bit DMA setting

      Note: This item (Storage Adapters) will only be shown for I/O Server nodes.
   l. Network Adapters with 64-bit DMA setting

Ignore other errors that might be flagged by the **gssinstallcheck** script. They will go away after the remaining installation steps are completed.

Check the system hardware

After the I/O server nodes have been installed successfully, power on the attached enclosures. Wait approximately five to 10 minutes from power on for discovery to complete before moving on to the next step. Here is the list of key log files that should be reviewed for possible problem resolution during deployment.

- By default `/var/log/message` log from all IO Server nodes are directed to the message log in the EMS node.
- gssdeploy log is located at `/var/log/gss`
- xCAT log is located at `/var/log/xcat`
- Console outputs from the IO Server node during deployment are located at `/var/log/consoles`

1. Run **gssstoragequickcheck**:

   ```
   gssstoragequickcheck -G gss_ppc64
   ```
2. Run **gss_sashba script**:

   ```
   updatenode gss_ppc64 -P gss_sashba
   ```
3. Run **gssfindmissingdisks**:

   ```
   gssfindmissingdisks -G gss_ppc64
   ```

   If **gssfindmissingdisks** displays an error, run **mmgetpdisktopology** and pipe it to **topsummary** on each I/O server node to obtain more information about the error:

   ```
   mmgetpdisktopology | topsummary
   ```
4. Run **gsscheckdisks**:
   ```
   GSSENV=INSTALL gsscheckdisks -G gss_ppc64 --encl all --iotest a --write-enable
   ```

Attention: When run with **--iotest w** (write) or **--iotest a** (all), **gsscheckdisks** will perform write I/O to the disks attached through the JBOD. This will overwrite the disks and will result in the loss of any configuration or user data stored on the attached disks. **gsscheckdisks** should be run only during the installation of a building block to validate that read and write operations can be performed to the attached drives without any error. The **GSSENV** environment variable must be set to **INSTALL** to indicate that **gsscheckdisks** is being run during installation.

5. Check for any hardware serviceable events and address them as needed. To view the serviceable events, issue the following command:

```
gssinstallcheck -N ems1,gss_ppc64 --srv-events
```

Note: During initial deployment of the nodes, SRC BA15D001 might be logged as serviceable event by Partition Firmware. This is normal and must be cleared after the initial deployment. For more information, see Appendix A, "Known issues," on page 15.

Set up the high-speed network

Set up the high-speed network that will be used for the cluster data communication. Update your `/etc/hosts` file with high-speed network entries showing the high-speed IP address and corresponding host name. Copy the modified `/etc/hosts` to I/O Server nodes of the cluster. With the Ethernet high-speed network, you can use the **gssgennetworks** script to create a bonded Ethernet interface over active (up) high-speed network interfaces. See Appendix D, "Installation: reference," on page 27.

1. Update `/etc/hosts` file with high-speed hostname entries in the EMS node and copy the modified `/etc/hosts` file to the I/O Server nodes as follows:

```
xdcp gss_ppc64 /etc/hosts /etc/hosts
```

2. To see the current set of active (up) interfaces on all nodes, run:

```
gssgennetworks -N ems1,gss_ppc64 --suffix=-hs
```

where ems1 is the name of the EMS node and gss_ppc64 is the I/O server node group and -hs is the nodename suffix of the high-speed host name.

3. To create a bonded interface, in all nodes run:

```
gssgennetworks -N ems1,gss_ppc64 --suffix=-hs --create-bond
```

The script sets `miimon` to `100`, the bonding mode to `802.3ad` (LACP), and `xmit_hash_policy` to `layer2+3`. In some network configurations xmit_hash_policy of layer3+4 may provide a better load balancing of the traffic over the slave interfaces of the bond. You can change the xmit_hash_policy as follows once the bond is created.

```
nmcli c mod bond-bond0 +bond.option xmit_hash_policy=layer3+4
```

where bond-bond0 is the bonded interface.

The other bond options are left with the default values, including `lacp_rate` (the default is `slow`). For proper network operation, the Ethernet switch settings in the networking infrastructure must match the I/O server node interface bond settings. If the Ethernet switch and network infrastructure cannot support bonding mode 802.3ad (LACP), another bonding mode can be selected. For more information, see the **gssgennetworks** man page.

See Appendix D, "Installation: reference," on page 27 for information on how to set up networking.

Note: If using bonded IP over IB do the following:

Remove or comment out **CONNECTED_MODE=yes** statement, if it exists, from the corresponding slave-bond interface scripts located in `/etc/sysconfig/network-scripts` directory of the EMS and IO Server nodes. These scripts are created as part of the IP over IB bond creation. An example of the slave-bond interface with the modification is shown below.

```
TYPE=Infiniband         <= change from Ethernet to Infiniband
NAME=bond-slave-ib0     <= bond-slave-ib0 is the slave connection
UUID=86c0af63-4b6c-475c-a724-0fb074dc9092
```

```
DEVICE=ib0              <= slave interfaceONBOOT=yes
MASTER=bond0            <= master bond interface
SLAVE=yes
#CONNECTED_MODE=yes      <= do not add this line, comment out if this statement exists
NM_CONTROLLED=yes       <= add this line
```

Run:`nmcli c reload bond-bond0`

Here **bond-bond0** is the connection name of the bond.

4. Once the high-speed network is set up, stress test the high-speed network as follows:

   ```
   GSSENV=TEST gssnettest -N ems1,gss_ppc64 --suffix=-hs
   ```

   where ems1 is the EMS node name and gss_ppc64 is the I/O server node group and -hs is the nodename suffix of the high-speed host name. This test should only be run in a test environment as it can highly stress the high-speed network.

Create the cluster, recovery groups, and file system

1. Create the GPFS™ cluster:

   ```
   gssgencluster -C test01 -G gss_ppc64 --suffix=-hs --accept-license
   ```

   In this example, `test01` is used as the cluster name and `-hs` is used as the suffix of the hostname.
2. Create the recovery groups:

   ```
   gssgenclusterrgs -G gss_ppc64 --suffix=-hs
   ```
3. Create the vdisks, NSDs, and file system:

   ```
   gssgenvdisks --create-vdisk --create-nsds --create-filesystem --contact-node gssio1
   ```
4. Add the management server node to the cluster:

   ```
   gssaddnode -N ems1 --cluster-node gssio1 --suffix=-hs --accept-license --no-fw-update
   ```

   In this example, the management server hostname is `ems1` with a suffix of `-hs` (`ems1-hs`) in the high-speed network. The **--no-fw-update** option is used because the management server node does not contain a SAS adapter or attached drives.

Check the installed software and system health

1. Run **gssinstallcheck** on the management server:

   ```
   gssinstallcheck -N ems1
   ```
2. Run **gssinstallcheck** on the I/O server nodes:

   ```
   gssinstallcheck -G gss_ppc64
   ```
3. Shut down GPFS in all nodes and reboot all nodes.
   a. Shut down GPFS all nodes:

      ```
      mmshutdown -a
      ```
   b. Reboot all server nodes:

      ```
      xdsh gss_ppc64 "systemctl reboot"
      ```
   c. Reboot EMS node:

      ```
      systemctl reboot
      ```
4. After reboots, run:

   ```
   gssinstallcheck -G gss_ppc64 --phy-mapping
   ```

   Ensure that the phy mapping check is OK.
5. Restart GPFS in all nodes and wait for all nodes to become active:

   ```
   mmstartup -a
   ```
6. Mount the filesystem and perform a stress test. For example, run:

```
mmmount gpfs0 -a
gssstress /gpfs/gpfs0 gssio1 gssio2
```

In this example, **gssstress** is invoked on the management server node. It is run on I/O server nodes `gssio1` and `gssio2` with `/gpfs/gpfs0` as the target path. By default gssstress runs for 20 iterations and can be adjusted using the -i option (type **gssstress** and press Enter to see the available options).
During the IO stress test, check for network error by running from another console:

```
gssinstallcheck -N ems1,gss_ppc64 --net-errors
```

7. Perform a health check. Run:

```
gnrhealthcheck
/usr/lpp/mmfs/bin/mmhealth node show -N all --verbose
```

Address any issues that are identified.

8. Check for any open hardware serviceable events and address them as needed. The serviceable events can be viewed as follows:

```
gssinstallcheck -N ems1,gss_ppc64 --srv-events
```

Install the ESS GUI

Important: Complete all of the following steps carefully including the steps for configuring **mmperfmon** and restricting certain sensors to the management server node (EMS) only.

1. Generate performance collector on the management server node by running the following command. The management server node must be part of the ESS cluster and the node name must be the node name used in the cluster (e.g., ems1-hs).

```
mmperfmon config generate --collectors ems1-hs
```

2. Set up the nodes in the *ems nodeclass* and *gss_ppc64 nodeclass* for performance monitoring by running the following command.

```
mmchnode --perfmon -N ems,gss_ppc64
```

3. Capacity and fileset quota monitoring is not enabled in the GUI by default. You must correctly update the values and restrict collection to the management server node only.
   a. To modify the GPFS Disk Capacity collection interval, run the following command:

```
mmperfmon config update GPFSDiskCap.restrict=EMSNodeName
          GPFSDiskCap.period=PeriodInSeconds
```

   The recommended period is 86400 so that the collection is done once per day.
   b. To restrict GPFS Fileset Quota to run on the management server node only, run the following command:

```
mmperfmon config update GPFSFilesetQuota.restrict=EMSNodeName
```

   Here the *EMSNodeName* must be the name shown in the **mmlscluster** output.

   Note: To enable quota, the filesystem quota checking must be enabled. Refer **mmchfs -Q** and **mmcheckquota** commands in the *IBM Spectrum Scale: Command and Programming Reference*.

4. Verify that the values are set correctly in the performance monitoring configuration by running the **mmperfmon config show** command on the management server node. Make sure that `GPFSDiskCap.period` is properly set, and `GPFSFilesetQuota` and `GPFSDiskCap` are both restricted to the management server node only.

Note: If you are moving from manual configuration to auto configuration then all sensors are set to default. Make the necessary changes using the **mmperfmon** command to customize your environment accordingly. For information on how to configure various sensors using **mmperfmon**, see Manually installing IBM Spectrum Scale GUI.

5. Start the performance collector on the management server node:
   ```
   systemctl start pmcollector
   ```
6. Enable and start gpfsgui:
   ```
   systemctl enable gpfsgui.service
   systemctl start gpfsgui
   ```
7. To launch the ESS GUI in a browser, go to: `https://EssGuiNode` where ESSGuiNode is the hostname or IP address of the management server node for GUI access. To log in, type admin in the User Name field and your password in the Password field on the login page. The default password for admin is admin001. Walk through each panel and complete the GUI Setup Wizard.

This completes the installation task of the ESS system.

Upgrade the ESS system

Perform a hardware and software healthcheck and address any issues before starting the upgrade of the system. During the upgrade process if a step fails, it must be addressed before moving to the next step. Follow these steps to perform an upgrade of the ESS system.

Prepare the system for upgrade

1. Perform a health check. Run:
   ```
   gnrhealthcheck
   ```
   Address any issues that are identified.
2. Wait for any of these commands that are performing file system maintenance tasks to complete:
   mmadddisk
   mmapplypolicy
   mmcheckquota
   mmdeldisk
   mmfsck
   mmlssnapshot
   mmrestorefs
   mmrestripefile
   mmrestripefs
   mmrpldisk
3. It is recommended that you stop the creation and deletion of snapshots using **mmcrsnapshot** and **mmdelsnapshot** during the upgrade window.
4. Check for any hardware serviceable events from the HMC. If you are running ESS 4.0.3 or newer you can obtain serviceable events using **gssinstallcheck** as follows:
   ```
   gssinstallcheck -N ems1,gss_ppc64 --srv-events
   ```
   Address any hardware issues identified in the serviceable events.
5. Check the status of the local boot drive shown in the **ipraid RAID status** line output of the **gssinstallcheck**:
   ```
   gssinstallcheck -N ems1,gss_ppc64
   ```
   If the status does not show that it is optimized, get it serviced. The local drive state can also be obtained by running **iprconfig** from the command line and selecting option 1. Check for RAID 10 Array Status. It should state Optimized.

Upgrading from versions earlier than ESS 3.5

ESS versions earlier than 3.5 should first be upgraded to ESS 3.5.x prior to upgrading to ESS 4.5. Contact Customer Support for further details.

Upgrading from ESS 3.5.x, ESS 4.0.x, or ESS 4.5.x

Perform the following steps if you are upgrading from ESS 3.5.x, ESS 4.0.x, or ESS 4.5.x:

1. Perform cleanup and save a backup copy of the xCAT database:
   ```
   /opt/ibm/gss/install/samples/gssdeploy -c -r /var/tmp/xcatdb
   ```
2. Update ESS repositories on the management server node:
   ```
   cd /opt/ibm/gss/install
   installer/gssinstall -m manifest -u
   ```
3. Install tools and xCAT and restore the xCAT database:
   ```
   /var/tmp/gssdeploy -x -r /var/tmp/xcatdb
   ```
4. Go to the next section titled *Update the management server node*.

Update the management server node

1. On the management server node, stop GUI services:
   ```
   systemctl stop gpfsgui
   ```
2. Save collector configuration files in the EMS node for later usage:
   ```
   cp /opt/IBM/zimon/ZIMonCollector.cfg /var/tmp
   ```
3. Save sensor configuration files in EMS and IO Server nodes (ssh to I/O Server nodes and run the command locally) for later usage:
   ```
   cp /opt/IBM/zimon/ZIMonSensors.cfg /var/tmp
   ```
4. Shut down IBM Spectrum Scale on the management server node while making sure quorum is still maintained. Run:
   ```
   mmshutdown
   ```
5. If you have not installed the Kernel Errata stated in the http://www-01.ibm.com/support/docview.wss?uid=ssg1S1005719, obtain the required Kernel Errata stated in this link. Complete the steps in Appendix F, “Instructions for installing the ESS Red Hat Linux Errata Kernel Update,” on page 39. Otherwise, skip this step.
6. Update the management server node:
   ```
   updatenode ems1 -P gss_updatenode
   ```

   Use **systemctl reboot** to reboot the management server node and complete this step again as follows:
   ```
   updatenode ems1 -P gss_updatenode
   ```
7. Perform the following steps to upgrade to upgrade IBM Spectrum Scale RAID configuration parameters.

   If upgrading from ESS 3.5.x:
   ```
   /opt/ibm/gss/tools/samples/gssupg450.sh -b
   /opt/ibm/gss/tools/samples/gssupg450.sh -c
   /opt/ibm/gss/tools/samples/gssupg450.sh -p
   ```

   If upgrading from ESS 4.0.x:
   ```
   /opt/ibm/gss/tools/samples/gssupg450.sh -b
   /opt/ibm/gss/tools/samples/gssupg450.sh -c
   ```
8. Update OFED on the management server node:
   ```
   updatenode ems1 -P gss_ofed
   ```
9. Update IP RAID Adapter firmware on the management server node:
   ```
   updatenode ems1 -P gss_ipraid
   ```

10. If using bonded IP over IB, remove or comment out **CONNECTED_MODE=yes** statement, if it exists, from the corresponding slave-bond interface scripts located in the `/etc/sysconfig/network-scripts` directory of the ems1 node. An example of the slave-bond interface with the modification is shown below.

```
TYPE=Infiniband
NAME=bond-slave-ib0
UUID=86c0af63-4b6c-475c-a724-0fb074dc9092
DEVICE=ib0
ONBOOT=yes
MASTER=bond0
SLAVE=yes
#CONNECTED_MODE=yes       <= this line commented out
NM_CONTROLLED=yes
```

11. Use **`systemctl reboot`** to reboot the management server node.
12. Start IBM Spectrum Scale on the management server node:

```
mmstartup
```

13. Verify that IBM Spectrum Scale is in the active state:

```
mmgetstate
```

    Do not proceed if the system is not active.

Update the I/O server nodes

Repeat the following steps for each I/O server node, one node at a time.

1. Move the cluster and file system manager role to another node if the current node is a cluster manager or file system manager.
   a. To find the cluster and file system managers, run:

```
mmlsmgr
```

   b. To change the file system manager, run:

```
mmchmgr gpfs0 gssio2-hs
```

      In this example, `gssio2-hs` is the new file system manager of file system `gpfs0`.

   c. To change the cluster manager, run:

```
mmchmgr -c gssio2-hs
```

      In this example, `gssio2-hs` is the new cluster manager.
2. Move the recovery group in the current I/O server node to the peer I/O server node in the same building block.
   a. To list the recovery groups, run:

```
mmlsrecoverygroup
```

   b. To list the active server, primary server, and secondary server, run:

```
mmlsrecoverygroup rg_gssio1-hs -L | grep active -A2
```

   c. To move the recovery group from the current active I/O server node (`rg_gssio1-hs`) to the peer I/O server node (`gssio2-hs`) in the same building block, run:

```
mmchrecoverygroup rg_gssio1-hs --servers gssio2-hs,gssio1-hs
```

3. After confirming that the recovery group has been successfully moved to the peer I/O server node, unmount all GPFS filesystems if mounted, and shut down IBM Spectrum Scale on the current I/O server node while maintaining quorum:

```
mmunmountall -N CurrentIoServer-hs
mmshutdown -N CurrentIoServer-hs
```

4. Run **updatenode**:

   ```
   updatenode CurrentIoServer -P gss_updatenode
   ```

   Reboot the I/O server node and complete this step again if you are instructed to do so in the **updatenode** output. Reboot the I/O server node as follows :

   ```
   xdsh CurrentIoServer "systemctl reboot"
   ```

5. Update OFED.

   ```
   updatenode CurrentIoServer -P gss_ofed
   ```

6. Update IP RAID FW in the I/O Server node that is being upgraded.

   ```
   updatenode CurrentIoServer -P gss_ipraid
   ```

7. If using bonded IP over IB, remove or comment out **CONNECTED_MODE=yes** statement, if it exists, from the corresponding slave-bond interface scripts located in the `/etc/sysconfig/network-scripts` directory of the *CurrentIOServer* node. An example of the slave-bond interface with the modification is shown below.

   ```
   TYPE=Infiniband
   NAME=bond-slave-ib0
   UUID=86c0af63-4b6c-475c-a724-0fb074dc9092
   DEVICE=ib0
   ONBOOT=yes
   MASTER=bond0
   SLAVE=yes
   #CONNECTED_MODE=yes       <= this line commented out
   NM_CONTROLLED=yes
   ```

8. Reboot the I/O server node as follows:

   ```
   xdsh CurrentIoServer "systemctl reboot"
   ```

9. Update the SAS host adapter firmware on *CurrentIoServer*:

   ```
   CurrentIoServer$ mmchfirmware --type host-adapter
   ```

   Here CurrentIOServer is an IO Server node and the command is run on the IO Server node.

10. Update the node configuration:

    ```
    /opt/ibm/gss/tools/samples/gssupg450.sh -s CurrentIoServer-hs
    ```

    This command is run from the EMS node.

11. Run phy check and ensure that the phy mapping is OK:

    ```
    gssinstallcheck -N CurrentIoServer --phy-mapping
    ```

12. Start IBM Spectrum Scale on the I/O server node. Run:

    ```
    mmstartup -N CurrentIoServer-hs
    ```

    Once Spectrum Scale daemon is successfully started, move back the cluster manager and the file system manager if required. Move back the recovery group that was moved to the peer I/O Server node of the same building block in Step 2 above.

13. Wait until the management server node shows that it is active, using the following command:

    ```
    mmgetstate
    ```

14. Repeat steps 1 through 13 for the peer I/O Server node of the same building block.
15. Repeat all steps in this section for each additional building block.

Update the enclosure and drive firmware

1. To update the storage enclosure firmware, run the following command from one I/O Server node of each building block:

   ```
   CurrentIoServer$ mmchfirmware --type storage-enclosure
   ```

2. To update the drive firmware, run the following command from **each** I/O Server node of each building block:

```
CurrentIoServer$ mmchfirmware --type drive
```

The drive update can take some time to complete. You can update the drives more quickly by taking the system offline (shutting down IBM Spectrum Scale) and using the **--fast-offline** option.

Check the installed software and system health

1. Run **gssinstallcheck** on the management server:
   ```
   gssinstallcheck -N ems1
   ```
2. Run **gssinstallcheck** on the I/O server nodes:
   ```
   gssinstallcheck -G gss_ppc64
   ```
3. Perform a health check. Run:

   ```
   gnrhealthcheck
   /usr/lpp/mmfs/bin/mmhealth node show -N all --verbose
   ```
4. Check for any hardware serviceable events and address them as needed. To view the serviceable events, issue the following command:
   ```
   gssinstallcheck -N ems1,gss_ppc64 --srv-events
   ```
   Note that during the initial deployment of the nodes, SRC BA15D001 may be logged as a serviceable event by Partition Firmware. This is normal and should be cleared after the initial deployment. For more information, see Appendix A, “Known issues,” on page 15.

Upgrading GUI

Note: Some of these steps might fail if they are already implemented in previous versions of ESS. If you see any failures indicating **mmperfmon** has already been configured, ignore these failure messages and continue with the remaining steps.

Upgrading from ESS 3.0.x

ESS versions earlier than 3.5 should first be upgraded to ESS 3.5.x prior to upgrading to ESS 4.5. Contact Customer Support for further details.

Upgrading from ESS 3.5.x or ESS 4.0.x

Perform the following steps to upgrade from ESS 3.5.x:

1. Generate performance collector on the management server node by running the following command. The management server node must be part of the ESS cluster and the node name must be the node name used in the cluster (e.g., ems1-hs).

   ```
   mmperfmon config generate --collectors ems1-hs
   ```
2. Set up the nodes in the *ems nodeclass* and *gss_ppc64 nodeclass* for performance monitoring by running the following command.
   ```
   mmchnode --perfmon -N ems,gss_ppc64
   ```
3. Capacity and fileset quota monitoring is not enabled in the GUI by default. You must correctly update the values and restrict collection to the management server node only.
   a. To modify the GPFS Disk Capacity collection interval, run the following command:
      ```
      mmperfmon config update GPFSDiskCap.restrict=EMSNodeName
               GPFSDiskCap.period=PeriodInSeconds
      ```
      The recommended period is 86400 so that the collection is done once per day.
   b. To restrict GPFS Fileset Quota to run on the management server node only, run the following command:
      ```
      mmperfmon config update GPFSFilesetQuota.restrict=EMSNodeName
      ```
      Here the *EMSNodeName* must be the name shown in the **mmlscluster** output.

Note: To enable quota, the filesystem quota checking must be enabled. Refer **mmchfs -Q** and **mmcheckquota** commands in the *IBM Spectrum Scale: Command and Programming Reference*.

4. Verify that the values are set correctly in the performance monitoring configuration by running the **mmperfmon config show** command on the management server node. Make sure that `GPFSDiskCap.period` is properly set, and `GPFSFilesetQuota` and `GPFSDiskCap` are both restricted to the management server node only.

   Note: If you are moving from manual configuration to auto configuration then all sensors are set to default. Make the necessary changes using the **mmperfmon** command to customize your environment accordingly. For information on how to configure various sensors using **mmperfmon**, see Manually installing IBM Spectrum Scale GUI.

5. Start the performance collector on the management server node:

   ```
   systemctl start pmcollector
   ```

6. Enable and start gpfsgui:

   ```
   systemctl enable gpfsgui.service
   systemctl start gpfsgui
   ```

7. To launch the ESS GUI in a browser, go to: `https://EssGuiNode` where ESSGuiNode is the hostname or IP address of the management server node for GUI access. To log in, type admin in the User Name field and your password in the Password field on the login page. The default password for admin is admin001. Walk through each panel and complete the GUI Setup Wizard.

This completes the upgrade task of the ESS system. For information on applying optimized configuration settings to a set of client nodes or a node class, see "Adding IBM Spectrum Scale nodes to an ESS cluster" on page 34.

Appendix A. Known issues

This topic includes known issues for ESS.

ESS 4.5 issues

Table 2 includes information about known issues in ESS 4.5 and how to resolve these issues. Depending on which fix level you are installing, these might or might not apply to you.

Table 2. Known issues in ESS 4.5

Issue	Environment affected	Description	Resolution or action
1. gssinstallcheck might flag an error regarding page pool size in multi-building block situations if the physical memory sizes differ.	Software validation Type: Install or Upgrade Version: Advanced or Standard Affected nodes: I/O	**gssinstallcheck** is a tool introduced in ESS 3.5, that helps validate software, firmware, and configuration settings. If adding (or installing) building blocks of a different memory footprint installcheck will flag this as an error. As a best practice all of your I/O servers must have the same memory footprint, thus page pool value. Page pool is currently set at ~60% of physical memory per I/O node. Example from **gssinstallcheck**: `[ERROR] pagepool: found 142807662592 expected range 147028338278 - 179529339371`	1. Confirm each I/O node's individual memory footprint. From the EMS, run the following command against your I/O xCAT group: `xdsh gss_ppc64 "cat/ proc/meminfo | grep MemTotal"` **Note:** This value is in KB If the physical memory varies between servers and/or building blocks, consider adding memory and re-calculating page pool to ensure consistency. 2. Validate the page pool settings in IBM Spectrum Scale: `mmlsconfig | grep -A 1 pagepool` **Note:** This value is in MB. If the page pool value setting is not roughly ~60% of physical memory, then you should consider recalculating and setting an updated value. For information on how to update the pagepool value, see the IBM Spectrum Scale documentation.

Table 2. Known issues in ESS 4.5 (continued)

Issue	Environment affected	Description	Resolution or action
2. The **gssgennetworks** script requires high-speed host names to be derived from I/O server (xCAT) host names using suffix, prefix, or both.	High-speed network generation Type: Install Version: Advanced or Standard Affected nodes: I/O and EMS	**gssgennetworks** requires that the target host name provided in -N or -G option are reachable to create the high-speed network on the target node. If the xCAT node name does not contain the same base name as the high-speed name you might be affected by this issue. A typical deployment scenario is: `gssio1 // xCAT name` `gssio1-hs // high-speed` An issue scenario is: `gssio1 // xCAT name` `foo1abc-hs // high-speed name`	Create entries in the `/etc/hosts` with node names that are reachable over the management network such that the high-speed host names can be derived from it using some combination of suffix and/or prefix. For example, if the high-speed host names are `foo1abc-hs` and `goo1abc-hs`: 1. Add foo1 and goo1 to the `/etc/hosts` using management network address (reachable) in the EMS node only. 2. Issue: **gssgennetworks -N foo1,goo1 – suffix abc-hs --create-bond** 3. Remove the entries foo1 and goo1 from `/etc/hosts` file on the EMS node once the high-speed networks are created. Example of how to fix (`/etc/hosts`): **// Before** `<IP><Long Name><Short Name>` `192.168.40.21` `gssio1.gpfs.net` `gssio1` `192.168.40.22` `gssio2.gpfs.net` `gssio2` `X.X.X.X` `foo1abc-hs.gpfs.net` `foo1abc-hs` `X.X.X.Y` `goo1abc-hs.gpfs.net` `goo1abc-hs` **// Fix** `192.168.40.21` `gssio1.gpfs.net` `gssio1 foo1` `192.168.40.22` `gssio2.gpfs.net` `gssio2 goo1` `X.X.X.X` `foo1abc-hs.gpfs.net` `foo1abc-hs` `X.X.X.Y` `goo1abc-hs.gpfs.net` `goo1abc-hs` `gssgennetworks -N foo1,` `goo1 --suffix=abc-hs` `--create-bond`

Table 2. Known issues in ESS 4.5 (continued)

Issue	Environment affected	Description	Resolution or action
3. When upgrading to ESS 4.5.x the gss_ofed script might fail for nodes upgraded to Red Hat Enterprise Linux 7.2. ESS 4.5.2 is supported on Red Hat Enterprise Linux 7.2 only. Any accidental upgrades to 7.2 or higher will cause issues.	Cluster Network Type: Upgrade Version: Advanced or Standard Affected Nodes:I/O + EMS	Users who connect nodes to the Red Hat Network must only apply specific security updates. If all non-essential updates are applied (not advised), systems may not fully upgrade to Red Hat Enterprise Linux 7.2. When upgrading to ESS 4.5.x, users would encounter the following issue when running the gss_ofed script: `gssio1: gss_ofed [DEBUG]: Detected rhel7u2 ppc64. Disabling installing 32bit rpms...` `gssio1: gss_ofed [DEBUG]: Error: The current MLNX_OFED_LINUX is intended for rhel7.1` `gssio1: gss_ofed [ERROR]: Mellanox install failed RC: 172`	The workaround for this issue is to downgrade the Red Hat Enterprise Linux release version back to 7.1. For each node experiencing this problem run the following command: `yum -y downgrade redhat-release-server unbound-libs` Then validate the level again:`cat /etc/redhat-release` Red Hat Enterprise Linux Server release 7.1 (Maipo) You may now re-run the gss_ofed script to properly update the Mellanox firmware and driver levels.
4. During initial deployment of the nodes, SRC BA15D001 is logged by Partition	Platform Firmware Type: Install Version: Advanced or Standard Affected Nodes: I/O and EMS	During deployment of EMS and I/O server nodes, Partition Firmware may log Unrecoverable Error, Loss of Function and SRC BA15D001.	This error is triggered during scanning of network ports that are not populated. The scanning is done to determine xCAT management network connection topology between EMS and I/O server nodes where some network ports can remain unpopulated in a valid configuration. Therefore, this error can be ignored when generated during deployment.
5. The GUI might not display data under **Monitoring** > **Capacity** > **Pools** or **Files** > **Filesystem (capacity)**if upgrading from ESS 4.0.x to ESS 4.5.x.	GUI Type: Upgrade Version: Advanced or Standard Affected Nodes: EMS (GUI host)	In certain instances, the GUI may not show data after upgrading from ESS 4.0.x to ESS 4.5.x in the following panels: 1. **Monitoring** > **Capacity** > **Pools** 2. **Files** > **Filesystem (capacity)**	Try the following steps if you hit this scenario: 1. Check that your pmsensors and pmcollector are correctly configured and started. 2. Restart the pmsensors on all nodes and the pmcollector on the EMS node. 3. Restart the `gpfsgui` service on the EMS node. 4. Run the runtask for CAPACITY, FILESYSTEMS, and POOLS To view the specific list of GUI tasks, use the following command: `/usr/lpp/mmfs/gui/cli/ lstask -x` To execute a task, use the following command: `/usr/lpp/mmfs/gui/cli/ runtask <task>` **Note:** For information on configuring **mmperfmon**, see “Install the ESS GUI ” on page 7.

Table 2. Known issues in ESS 4.5 (continued)

Issue	Environment affected	Description	Resolution or action
6. Running **gssutils** over PuTTY might show horizontal lines as “qqq” and vertical lines as “xxx”.	ESS Install and Deployment toolkit Type: Install Version: Advanced or Standard Affected Nodes: EMS and I/O Server nodes	PuTTY translation default Remote Character set UTF-8 may not translate horizontal line and vertical character sets correctly.	1. On the PuTTY terminal, do the following: In Window -> Translation, change Remote character set from UTF-8 to ISO-8859-1:1998 (Latin-1, West Europe) (This should be first option after UTF-8). 2. Open session
7. The GUI might show ‘unknown’ hardware states for storage enclosures and Power 8 servers in the ESS building block. Part info and firmware levels under the ‘Hardware Details’ panel may also be missing.	GUI Type: Upgrade Version: Advanced or Standard Affected nodes: ALL	The ESS GUI (running on the EMS) might show ‘unknown’ under the Hardware panel for the ESS building block members. The ESS GUI might also be missing information under Part Info and Firmware version within the **Hardware Details** panel. .	To fix most of these issues, do the following: 1. Login to the EMS 2. Run the following in order: `/usr/lpp/mmfs/gui/cli/runtask RECOVERY_GROUP` `/usr/lpp/mmfs/gui/cli/runtask DISK_ENCLOSURES` `/usr/lpp/mmfs/gui/cli/runtask ENCLOSURE_FW` `/usr/lpp/mmfs/gui/cli/runtask CHECK_FIRMWARE` After running, the GUI should refresh with the issues resolved.
8. The GUI might display the long-waiters warning: `Spectrum Scale long-waiters monitoring returned unknown result`	GUI Type: Upgrade Version: Advanced or Standard Affected nodes: ALL	Upon new installs (or upgrades) to ESS 4.5.x, the GUI might show an error due to a bad return code from **mmhealth** in its querying of long-waiters information. `/usr/lpp/mmfs/bin/mmdiag --deadlock Failed to connect to file system daemon: No such process RC=50`	There is no current fix but it is advised to verify on the command line that no long-waiters exist. If the system is free from this symptom, mark the event as read on the GUI by clicking under the Action column. Doing so will clear the event.
9. Creating small file systems in the GUI (below 16G) will result in incorrect sizes	GUI Type: Install Version: Advanced or Standard Affected nodes: ALL	When creating file systems in the GUI smaller than 16GB (usually done to create CES_ROOT for protocol nodes) the size will come out larger than expected.**gssgenvdisks** also has this problem for those used to the command-line method of creating file systems.	There is currently no resolution. The smallest size you might be able to create is 16GB. Experienced users may consider creating a customer `vdisk.stanza` file for specific sizes you require.

Table 2. Known issues in ESS 4.5 (continued)

Issue	Environment affected	Description	Resolution or action
10. Creating file systems in the GUI might immediately result in lack of capacity data	GUI Type: Install Version: Advanced or Standard Affected nodes: ALL	When creating file systems in the GUI you might not immediately see capacity data show up.	You may wait up to 24 hours for the capacity data to display or simply visit the command line which should accurately show the file system size.
11. Fileset information might not show in the GUI (under **Monitoring** > **Filesets**)	GUI Type: Install Version: Advanced or Standard Affected nodes: ALL	You might not see fileset information displayed under the **Monitoring** tab in the ESS GUI.	There is no workaround currently for this problem. An alternative method to view fileset information is to use the **mmlsfileset** command.

Appendix B. Networking requirements for ESS

This topic provides instructions for installing the Red Hat Linux Errata Kernel Update for ESS.

Networking requirements

The following networks are required:

- **Service network**

  This network connects the flexible service processor (FSP) on the management server and I/O server nodes with the HMC, as shown in yellow in Figure 1 on the following page.

- **Management and provisioning network**

  This network connects the management server to the I/O server nodes and HMCs, as shown as blue in in Figure 1 on the following page. The management server runs DHCP on the management and provisioning network. If a management server is not included in the solution order, a customer-supplied management server is used.

- **Clustering network**

  This high-speed network is used for clustering and client node access. It can be a 10 Gigabit Ethernet (GbE), 40 GbE, or InfiniBand network. It might not be included in the solution order.

- **External and campus management network**

  This public network is used for external and campus management of the management server, the HMC, or both.

Figure 1, Network Topology, is a high-level logical view of the management and provisioning network and the service network for an ESS building block.

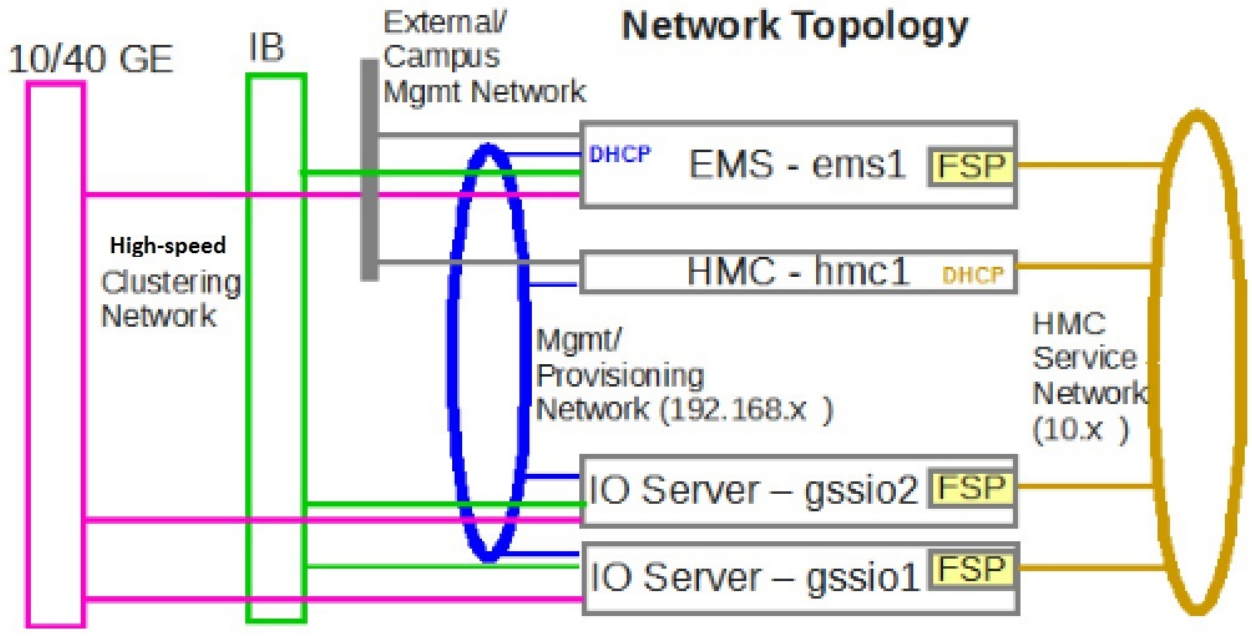

Figure 1. The management and provisioning network and the service network: a logical view

The management and provisioning network and the service network must run as two non-overlapping networks implemented as two separate physical networks or two separate virtual local-area networks (VLANs).

The HMC, the management server, and the switches (1 GbE switches and high-speed switches) might not be included in a solution order in which an existing or customer-supplied HMC or management server is used. Perform any advance planning tasks that might be needed to access and use these solution components.

Appendix C. Pre-installation tasks for ESS

This topic provides the pre-installation tasks required for ESS.

Table 3. Pre-installation tasks

ESS component	Description	Required actions	System settings
1. Service network	This private network connects the HMC with the management server's FSP and the I/O server nodes. The service network must not be seen by the OS running on the node being managed (that is, the management server or the I/O server node). The HMC uses this network to discover the management server and the I/O server nodes and perform such hardware management tasks as creating and managing logical partitions, allocating resources, controlling power, and rebooting.	Perform any advance planning tasks that might be needed to access and use the HMC if it is not part of the solution order and a customer-supplied HMC will be used. Set up this network if it has not been set up already.	Set the HMC to be the DHCP server for the service network.
2. Management and provisioning network	This network connects the management server node with the HMC and the I/O server nodes. It typically runs over 1Gb. • This network is visible to the OS that is running on the nodes. • The management server uses this network to communicate with the HMC and to discover the I/O server nodes. • The management server will be the DHCP server on this network. There cannot be any other DHCP server on this network. • This network is also used to provision the node and therefore deploy and install the OS on the I/O server nodes.	Perform any advance planning tasks that might be needed to access and use the management server if it is not part of the solution order and a customer-supplied management server will be used. Set up this network if it has not been set up already.	
3. Clustering network	This network is for high-performance data access. In most cases, this network is also part of the clustering network. It is typically composed of 10GbE, 40GbE, or InfiniBand networking components.	Set up this network if it has not been set up already.	
4. Management network domain	The management server uses this domain for the proper resolution of hostnames.	Set the domain name using *lowercase* characters. Do *not* use any uppercase characters.	Example: `gpfs.net`

Table 3. Pre-installation tasks (continued)

ESS component	Description	Required actions	System settings
5. HMC node (IP address and hostname)	The IP address of the HMC node on the management network has a console name, which is the hostname and a domain name. • This IP address must be configured and the link to the network interface must be up. • The management server must be able to reach the HMC using this address.	Set the fully-qualified domain name (FQDN) and the hostname using *lowercase* characters. Do *not* use any uppercase characters. Do *not* use a suffix of `-en`*x*, where *x* is any character. Do *not* use an _ (underscore) in the hostname.	Example: IP address: `192.168.45.9` Hostname: `hmc1` FQDN: `hmc1.gpfs.net`
6. Management server node (IP address)	The IP address of the management server node has an FQDN and a hostname. • This IP address must be configured and the link to the network interface must be up. • The management network must be reachable from this IP address.	Set the FQDN and hostname using *lowercase* characters. Do *not* use any uppercase characters. Do *not* use a suffix of `-en`*x*, where *x* is any character. Do *not* use an _ (underscore) in the hostname.	Example: IP address: `192.168.45.10` Hostname: `ems1` FQDN: `ems1.gpfs.net`
7. I/O server nodes (IP addresses)	The IP addresses of the I/O server nodes have FQDNs and hostnames. • These addresses are assigned to the I/O server nodes during node deployment. • The I/O server nodes must be able to reach the management network using this address.	Set the FQDN and hostname using *lowercase* characters. These names must match the name of the partition created for these nodes using the HMC. Do *not* use any uppercase characters. Do *not* use a suffix of `-en`*x*, where *x* is any character. Do *not* use an _ (underscore) in the host name.	Example: I/O server 1: IP address: `192.168.45.11` Hostname: `gssio1` FQDN: `gssio1.gpfs.net` I/O server 2: IP address: `192.168.45.12` Hostname: `gssio2` FQDN: `gssio2.gpfs.net`
8. Management server node (management network interface)	The management network interface of the management server node must have the IP address that you set in item 6 assigned to it. This interface must have only one IP address assigned.	To obtain this address, run: `ip addr`	Example: `enP7p128s0f0`
9. HMC (**hscroot** password)		Set the password for the **hscroot** user ID.	Example: `abc123` This is the default password.
10. I/O servers (user IDs and passwords)	The user IDs and passwords of the I/O servers are assigned during deployment.		Example: User ID: `root` Password: `cluster` (this is the default password)

Table 3. Pre-installation tasks (continued)

ESS component	Description	Required actions	System settings
11. Clustering network (hostname prefix or suffix)	This high-speed network is implemented on a 10Gb Ethernet, 40Gb Ethernet or InfiniBand network.	Set a hostname for this network. It is customary to use hostnames for the high-speed network that use the prefix and suffix of the actual hostname. Do *not* use a suffix of `-en`*x*, where *x* is any character.	Examples: Suffixes: `-bond0`, `-ib`, `-10G`, `-40G` Hostnames with a suffix: `gssio1-ib`, `gssio2-ib`
12. High-speed cluster network (IP address)	The IP addresses of the management server nodes and I/O server nodes on the high-speed cluster network have FQDNs and hostnames. In the example, `172.10.0.11` is the IP address that the GPFS daemon uses for clustering. The corresponding FQDN and hostname are `gssio1-ib` and `gssio1-ib.data.net`, respectively.	Set the FQDNs and hostnames. Do *not* make changes in the `/etc/hosts` file for the high-speed network until the deployment is complete. Do *not* create or enable the high-speed network interface until the deployment is complete.	Example: Management server: IP address: `172.10.0.10` Hostname: `ems1-ib` FQDN: `ems1-ib.gpfs.net` I/O server 1: IP address: `172.10.0.11` Hostname: `gssio1-ib` FQDN: `gssio1-ib.data.net` I/O server 2: IP address: `172.10.0.12` Hostname: `gssio2-ib` FQDN: `gssio2-ib.data.net`
13. Red Hat Enterprise Linux 7.1	The Red Hat Enterprise Linux 7.1 DVD or ISO file is used to create a temporary repository for the xCAT installation. xCAT uses it to create a Red Hat Enterprise Linux repository on the management server node.	Obtain this DVD or ISO file and download. For more information, see the Red Hat Enterprise Linux website: **http://access.redhat.com/products/red-hat-enterprise-linux/**	Example: `RHEL-7.1-20150219.1-Server-ppc64-dvd1.iso`

Table 3. Pre-installation tasks (continued)

ESS component	Description	Required actions	System settings
14. Management network switch	The switch that implements the management network must allow the Bootstrap Protocol (BOOTP) to go through.	Obtain the IP address and access credentials (user ID and password) of this switch. Some switches generate many Spanning Tree Protocol (STP) messages, which interfere with the network boot process. You need to disable STP to mitigate this.	
15. Target file system	You need to provide information about the target file system that is created using storage in the ESS building blocks.	Set the target file system name, the mount point, the block size, the number of data NSDs, and the number of metadata NSDs.	Example: `Block size = 8M, #datansd=4, #metadatansd=2`

Appendix D. Installation: reference

This topic provides information on creating a bonded interface with Ethernet, adding IBM Spectrum Scale nodes to an ESS cluster, and node name considerations.

bonded interface

A bonded interface with Ethernet

Starting with ESS 3.5, you can use a script to help you quickly create a bonded interface with Ethernet. See the man page for the **gssgennetworks** command for more information. Otherwise, complete the following steps.

Connect the network cables to the corresponding switch. Check that the links are up at the device level. To create a bonding, add connections for the master, add connections for the slave, bring up the connection for the slaves, and then bring up the connection for the master (bond). Run:

```
ibdev2netdev
```

The system displays output similar to this:

```
[root@gssio2 ~]# ibdev2netdev

mlx4_0 port 1 ==> enp1s0 (Up)
mlx4_0 port 2 ==> enp1s0d1 (Up)
mlx5_0 port 1 ==> ib0 (Down)
mlx5_0 port 2 ==> ib1 (Down)
mlx5_1 port 1 ==> ib2 (Down)
mlx5_1 port 2 ==> ib3 (Down)
```

This example shows two 10GbE network ports that are up and are connected to the switch properly. Now you will create a bond with this two ports.

Check the connection and make sure there are no connections defined for these ports. You can do this using network manager connection and device commands.

To check the connection, run:

```
nmcli -p c
```

The system displays output similar to this:

```
[root@gssio2 ~]# nmcli -p c

================================================================================
                          NetworkManager connection profiles
================================================================================
NAME             UUID                                  TYPE            DEVICE
--------------------------------------------------------------------------------
enp1s0d1         6d459dc7-db53-43d4-9236-8257ee900aae  802-3-ethernet  --
enP7p128s0f2     72b6533e-6eaa-4763-98fa-0b4ed372e377  802-3-ethernet  --
enP7p128s0f3     1b0a97e7-1b90-4d26-89cf-8f4fc8e5a00e  802-3-ethernet  --
enP7p128s0f1     5dffee0e-b0b6-4472-864e-acc2dc0cc043  802-3-ethernet  --
enp1s0           060d342f-3388-4e9f-91bb-13c3aa30847f  802-3-ethernet  --
GSS enP7p128s0f0 5f755525-2340-7e18-ef9d-0d4bfdba4c30  802-3-ethernet  enP7p128s0f0
```

To check the device, run:

```
nmcli -p d
```

The system displays output similar to this:

```
[root@gssio2 ~]# nmcli -p d

============================================================
                    Status of devices
============================================================
DEVICE        TYPE        STATE         CONNECTION
------------------------------------------------------------
enP7p128s0f0  ethernet    connected     GSS enP7p128s0f0
enP7p128s0f1  ethernet    disconnected  --
enP7p128s0f2  ethernet    disconnected  --
enP7p128s0f3  ethernet    disconnected  --
enp1s0        ethernet    disconnected  --
enp1s0d1      ethernet    disconnected  --
ib0           infiniband  disconnected  --
ib1           infiniband  disconnected  --
ib2           infiniband  disconnected  --
ib3           infiniband  disconnected  --
lo            loopback    unmanaged     --
```

As you can see, there is no connection defined for the devices and the device state is down. Add a connection for the bond `bond0`. In this case, specify `802.3ad` for the Link Aggregation Control Protocol (LACP) and an IPv4 address of `172.16.45.22/24`. For the bonding parameters, specify a `miimon` value of 100 milliseconds (msec).

```
[root@gssio2 ~]# nmcli c add type bond ifname bond0 miimon 100 mode 802.3ad ip4 172.16.45.22/24

Connection 'bond-bond0' (c929117b-6d92-488d-8bcb-d98e7e0c8b91) successfully added.
```

Note that by default, `xmit_hash_policy` is set to `layer2`. For optimal performance, you might want to set it to `layer3+4`, as follows:

```
nmcli c mod bond-bond0 +bond.option xmit_hash_policy=layer3+4
```

To view the connection properties, run:

```
nmcli c show bond-bond0
```

Add connections for the slaves:

```
[root@gssio2 ~]# nmcli c add type bond-slave ifname enp1s0 master bond0

Connection 'bond-slave-enp1s0' (d9e21d55-86ea-4551-9371-1fc24d674751) successfully added.

[root@gssio2 ~]# nmcli c add type bond-slave ifname enp1s0d1 master bond0

Connection 'bond-slave-enp1s0d1' (8432645a-5ddc-44fe-b5fb-2884031c790c) successfully added.
```

Bring the connection up for the slaves:

```
[root@gssio2 ~]# nmcli c up bond-slave-enp1s0d1

Connection successfully activated (D-Bus active path:
/org/freedesktop/NetworkManager/ActiveConnection/4)

[root@gssio2 ~]# nmcli c up bond-slave-enp1s0

Connection successfully activated (D-Bus active path:
/org/freedesktop/NetworkManager/ActiveConnection/6)
```

Bring the connection up for bond-bond0:

```
[root@gssio2 ~]# nmcli c up bond-bond0

Connection successfully activated (D-Bus active path:
/org/freedesktop/NetworkManager/ActiveConnection/7)
```

Finally, make sure the appropriate bond devices have been created:

```
[root@gssio2 ~]# cat /proc/net/bonding/bond0

Ethernet Channel Bonding Driver: v3.7.1 (April 27, 2011)

Bonding Mode: IEEE 802.3ad Dynamic link aggregation
Transmit Hash Policy: layer2 (0)
MII Status: up
MII Polling Interval (ms): 100
Up Delay (ms): 0
Down Delay (ms): 0

802.3ad info
LACP rate: slow
Min links: 0
Aggregator selection policy (ad_select): stable
Active Aggregator Info:
        Aggregator ID: 1
        Number of ports: 1
        Actor Key: 33
        Partner Key: 1
        Partner Mac Address: 00:00:00:00:00:00

Slave Interface: enp1s0
MII Status: up
Speed: 10000 Mbps
Duplex: full
Link Failure Count: 0
Permanent HW addr: f4:52:14:df:af:74
Aggregator ID: 1
Slave queue ID: 0

Slave Interface: enp1s0d1
MII Status: up
Speed: 10000 Mbps
Duplex: full
Link Failure Count: 0
Permanent HW addr: f4:52:14:df:af:75
Aggregator ID: 2
Slave queue ID: 0
```

Changing the MTU

To change the maximum transmission unit (MTU), follow these steps:

1. Create a file, copy the following script into it, and save the file in the `/etc/NetworkManager/dispatcher.d` directory of the nodes where bonding is run If the executable (**x**) bit gets reset, use **chmod +x** to make the file executable. The `/opt/ibm/gss/tools/samples` directory includes the **mtuset** script for your use.

```
#!/bin/sh
INTERFACE_NAME_REGEX="^bond?"
if [[ $CONNECTION_ID =~ $INTERFACE_NAME_REGEX ]]; then
        if [[ $2 == up ]]; then
                MTU=$(awk -F "=" '($1 ~ "^MTU") {print $NF}' /etc/sysconfig/network-scripts/
ifcfg-$CONNECTION_ID)
                if [[ $MTU > 0 ]] && [[ $MTU != 1500 ]]; then
                        logger -s "Setting MTU of $CONNECTION_ID to $MTU..."
                        if /usr/sbin/ip link set dev $1 mtu $MTU ; then
                                logger "Successfully set MTU of $CONNECTION_ID to $MTU"
                        else
                                logger "Failed to set MTU of $CONNECTION_ID to $MTU"
                        fi
                fi
        fi
fi
```

See **https://access.redhat.com/solutions/1309583** for more information.

2. Add the MTU parameter value to the bond's interface configuration file. To set an MTU of 9000, specify:

   ```
   MTU=9000
   ```

   For example, add **MTU=9000** to `ifcfg-bond-bond0`, `ifcfg-bond-slave-xxxx`, and `ifcfg-bond-slave-yyyy`. The script shown in the previous step checks for the MTU setting and uses **ip link set** to set them appropriately. The script assumes that the bond connection starts with `bond?-xxxx`. Make changes in the scripts as needed.

3. To enable the network manager dispatch service in each node, run these commands:

   ```
   [root@gssio2 network-scripts]# systemctl enable NetworkManager-dispatcher.service
   [root@gssio2 network-scripts]# systemctl start NetworkManager-dispatcher.service
   ```

4. To restart networking, run:

   ```
   nmcli networking off; modprobe -r bonding ; nmcli networking on
   ```

   While restarting networking, you could lose the connection to the I/O server nodes. Use **rcons** to establish the console connection, if needed.

   a. Open a console to each node. For example, run:

      ```
      rcons gssio1
      ```

      If **rcons** does not open, the console server is probably not running. To restart it at the management server node, run:

      ```
      makeconservercf NodeName
      ```

      or

      ```
      makeconservercf NodeGroup
      ```

      Log in to the console. The default user ID is **root** and the default password is **cluster**.

   b. To disconnect from the console server, press **<Ctrl-e> c .** (period).

Bonding with InfiniBand

Connect the InfiniBand cables to the corresponding switch. Check that the links are up at the device level. To create a bonding, add connections for the master and for the slave. You will have to modify the network script file and reload the connections in Network Manager. After the connections are reloaded, bonding should be available. When the system is restarted or rebooted, it could take some time (more than five minutes) before the bonding interface is ready. Check the device status on each node to make sure all of the links are up. Run:

```
ibdev2netdev
```

The system displays output similar to this:

```
[root@gssio2 ~]# ibdev2netdev

mlx5_0 port 1 ==> ib0 (Up)
mlx5_0 port 2 ==> ib1 (Up)
mlx5_1 port 1 ==> ib2 (Up)
mlx5_1 port 2 ==> ib3 (Up)
mlx5_2 port 1 ==> ib4 (Up)
mlx5_2 port 2 ==> ib5 (Up)
```

You can also use **ibstat**.

Check the connection using `nmcli c` and make sure there is no existing bond already defined in these interfaces. Add the bond connection first. In this example, active-backup mode is selected. In IP over InfiniBand (IPoIB), only active-backup bond is supported. Run:

```
nmcli c add type bond ifname bond0 mode
```

The system displays output similar to this:

```
[root@gssio2 network-scripts]# nmcli c add type bond ifname bond0 mode

active-backup ip4 172.16.45.22/24
Connection 'bond-bond0' (66f182d1-d0da-42cf-b4c9-336d5266bbe7) successfully
added.
```

Add the slave connections as follows. In this example, `ib0` and `ib1` are the slave devices. Make appropriate changes as needed. First, run:

```
nmcli c add type bond-slave ifname ib0 master bond0
```

The system displays output similar to this:

```
[root@gssio2 network-scripts]#  nmcli c add type bond-slave ifname ib0 master bond0

Connection 'bond-slave-ib0' (86c0af63-4b6c-475c-a724-0fb074dc9092) successfully added.
```

Next, run:

```
nmcli c add type bond-slave ifname ib1 master bond0
```

The system displays output similar to this:

```

[root@gssio2 network-scripts]#  nmcli c add type bond-slave ifname ib1 master bond0

Connection 'bond-slave-ib1' (1d0cb5c3-268d-487c-9e40-7c0cf268150f) successfully added.
```

To check the connections, run:

```
nmcli c
```

The system displays output similar to this:

```
[root@gssio2 network-scripts]# nmcli c

NAME              UUID                                   TYPE            DEVICE
GSS enP7p128s0f0  5f755525-2340-7e18-ef9d-0d4bfdba4c30  802-3-ethernet  enP7p128s0f0
bond-slave-ib1    1d0cb5c3-268d-487c-9e40-7c0cf268150f  802-3-ethernet  --
bond-slave-ib0    86c0af63-4b6c-475c-a724-0fb074dc9092  802-3-ethernet  --
bond-bond0        66f182d1-d0da-42cf-b4c9-336d5266bbe7  bond            bond0
enP7p128s0f1      2eb8617f-5c7d-4d68-a7fe-88a030fdb28b  802-3-ethernet  --
enP7p128s0f3      7dea32aa-caa1-4016-9414-a47c62de27e9  802-3-ethernet  --
enP7p128s0f2      4416229e-2233-414f-b3ad-929c54c15f27  802-3-ethernet  --
```

You can see that the slave connections are created, but there are no devices for these connections.

To check the devices, run:

```
nmcli d
```

The system displays output similar to this:

```
[root@gssio2 network-scripts]# nmcli d

DEVICE        TYPE        STATE         CONNECTION
bond0         bond        connected     bond-bond0
enP7p128s0f0  ethernet    connected     GSS enP7p128s0f0
enP7p128s0f1  ethernet    disconnected  --
enP7p128s0f2  ethernet    disconnected  --
enP7p128s0f3  ethernet    disconnected  --
ib0           infiniband  disconnected  --
ib1           infiniband  disconnected  --
ib2           infiniband  disconnected  --
```

```
ib3           infiniband  disconnected  --
ib4           infiniband  disconnected  --
ib5           infiniband  disconnected  --
lo            loopback    unmanaged     --
```

The devices `ib0` and `ib1` are disconnected (this is the view from Network Manager).

Check `/etc/sysconfig/network-scripts` directory for the network script for each of the connections that were just created.

```
-rw-r--r--  1 root root   354 Jan 19 04:12 ifcfg-bond-bond0
-rw-r--r--  1 root root   121 Jan 19 04:12 ifcfg-bond-slave-ib0
-rw-r--r--  1 root root   121 Jan 19 04:12 ifcfg-bond-slave-ib1
```

You need to make some changes to the slave connection scripts (`ifcfg-bond-slave-ib0` and `ifcfg-bond-slave-ib1`). In most cases, the master bond interface script remains unchanged:

```
cat ifcfg-bond-bond0

DEVICE=bond0
BONDING_OPTS=mode=active-backup
TYPE=Bond
BONDING_MASTER=yes
BOOTPROTO=none
IPADDR0=172.16.45.22
PREFIX0=24
GATEWAY0=172.6.45.20
DEFROUTE=yes
IPV4_FAILURE_FATAL=no
IPV6INIT=yes
IPV6_AUTOCONF=yes
IPV6_DEFROUTE=yes
IPV6_PEERDNS=yes
IPV6_PEERROUTES=yes
IPV6_FAILURE_FATAL=no
NAME=bond-bond0
UUID=66f182d1-d0da-42cf-b4c9-336d5266bbe7
ONBOOT=yes
```

Modify the first slave-bond interface script as indicated in bold typeface:

```
TYPE=Infiniband                 <= change from Ethernet to Infiniband
NAME=bond-slave-ib0
UUID=86c0af63-4b6c-475c-a724-0fb074dc9092
DEVICE=ib0
ONBOOT=yes
MASTER=bond0
SLAVE=yes
NM_CONTROLLED=yes               <= add this line
```

Modify the second slave-bond interface script as indicated in bold typeface:

```
TYPE=Infiniband                 <= change from Ethernet to Infiniband
NAME=bond-slave-ib1
UUID=1d0cb5c3-268d-487c-9e40-7c0cf268150f
DEVICE=ib1
ONBOOT=yes
MASTER=bond0
SLAVE=yes
NM_CONTROLLED=yes               <= add this line
```

Now reload the connections:

```
[root@gssio2 network-scripts]# nmcli c reload
```

To check the connections, run:

```
nmcli c
```

The system displays output similar to this:

```
[root@gssio2 network-scripts]# nmcli c

NAME              UUID                                  TYPE            DEVICE
GSS enP7p128s0f0  5f755525-2340-7e18-ef9d-0d4bfdba4c30  802-3-ethernet  enP7p128s0f0
bond-slave-ib1    1d0cb5c3-268d-487c-9e40-7c0cf268150f  infiniband      ib1
bond-slave-ib0    86c0af63-4b6c-475c-a724-0fb074dc9092  infiniband      ib0
bond-bond0        66f182d1-d0da-42cf-b4c9-336d5266bbe7  bond            bond0
enP7p128s0f1      2eb8617f-5c7d-4d68-a7fe-88a030fdb28b  802-3-ethernet  --
enP7p128s0f3      7dea32aa-caa1-4016-9414-a47c62de27e9  802-3-ethernet  --
enP7p128s0f2      4416229e-2233-414f-b3ad-929c54c15f27  802-3-ethernet  --
```

Now you can see that the bond slave connections have devices assigned to them.

To check the devices, run:

```
nmcli d
```

The system displays output similar to this:

```
[root@gssio2 network-scripts]# nmcli d

DEVICE        TYPE        STATE         CONNECTION
bond0         bond        connected     bond-bond0
enP7p128s0f0  ethernet    connected     GSS enP7p128s0f0
ib0           infiniband  connected     bond-slave-ib0
ib1           infiniband  connected     bond-slave-ib1
enP7p128s0f1  ethernet    disconnected  --
enP7p128s0f2  ethernet    disconnected  --
enP7p128s0f3  ethernet    disconnected  --
ib2           infiniband  disconnected  --
ib3           infiniband  disconnected  --
ib4           infiniband  disconnected  --
ib5           infiniband  disconnected  --
lo            loopback    unmanaged     --
```

This shows that devices `ib0` (connection name: `bond-slave-ib0`) and `ib1` (connection name: `bond-slave-ib1`) are now connected.

To check the `bond0` state in the `proc` file system, run:

```
cat /proc/net/bonding/bond0
```

The system displays output similar to this:

```
[root@gssio2 network-scripts]# cat /proc/net/bonding/bond0

Ethernet Channel Bonding Driver: v3.7.1 (April 27, 2011)

Bonding Mode: fault-tolerance (active-backup) (fail_over_mac active)
Primary Slave: None
Currently Active Slave: ib0
MII Status: up
MII Polling Interval (ms): 100
Up Delay (ms): 0
Down Delay (ms): 0

Slave Interface: ib0
MII Status: up
Speed: 40000 Mbps
Duplex: full
Link Failure Count: 0
Permanent HW addr: a0:00:00:27:fe:80
Slave queue ID: 0

Slave Interface: ib1
```

```
MII Status: up
Speed: 40000 Mbps
Duplex: full
Link Failure Count: 0
Permanent HW addr: a0:00:00:29:fe:80
Slave queue ID: 0
```

To ping the other node on the same bonded network, run:

```
ping 172.16.45.22
```

The system displays output similar to this:

```
[root@gssio1 ~]# ping 172.16.45.22

PING 172.16.45.22 (172.16.45.22) 56(84) bytes of data.
64 bytes from 172.16.45.22: icmp_seq=1 ttl=64 time=8.52 ms
64 bytes from 172.16.45.22: icmp_seq=2 ttl=64 time=0.059 ms
64 bytes from 172.16.45.22: icmp_seq=3 ttl=64 time=0.055 ms
64 bytes from 172.16.45.22: icmp_seq=4 ttl=64 time=0.042 ms
64 bytes from 172.16.45.22: icmp_seq=5 ttl=64 time=0.043 ms
64 bytes from 172.16.45.22: icmp_seq=6 ttl=64 time=0.040 ms
```

Adding IBM Spectrum Scale nodes to an ESS cluster

IBM Spectrum Scale node configuration is optimized for running Spectrum Scale RAID functions.

1. ESS cluster node configuration is optimized for running Spectrum Scale RAID functions. Protocols, other gateways, or any other non-ESS services must not be run on ESS management server nodes or I/O server nodes. In a cluster with high I/O load, avoid using ESS nodes as cluster manager or filesystem manager. For optimal performance the NSD client nodes accessing ESS nodes should be properly configured. ESS ships with **gssClientConfig.sh script** located in `/usr/lpp/mmfs/samples/gss/` directory. Use this script to configure the client as follows:

   ```
   /usr/lpp/mmfs/samples/gss/gssClientConfig.sh <Comma Separated list of
   client nodes or nodeclass>
   ```

   You can run the following to see configuration parameter settings without setting them:

   ```
   /usr/lpp/mmfs/samples/gss/gssClientConfig.sh -D
   ```

2. When IBM Spectrum Scale nodes deployed with protocols are added to the ESS cluster, quorum, cluster manager, and filesystem manager functions should be moved from the ESS to the protocol nodes after adding protocol nodes to the cluster.

For information about adding an IBM Spectrum Scale protocol node to an ESS cluster, see:

- Preparing a cluster that contains ESS for adding protocols
- Spectrum Scale Protocols Quick Overview

Node name considerations

Carefully select the hostname, suffix, and prefix of the management server and I/O server so that the hostname used in the high-speed network and by the ESS cluster can be generated from the suffix or prefix.

High-speed host names

Example 1:

```
a-bcd-edf-1
a-bcd-edf-2
a-bcd-edf-3
a-bcd-edf-4
```

Here, `a-bcd-` is the prefix and `edf-1`, `edf-2`, `edf-3`, and `edf-4` are the xCAT names of the nodes.

Example 2:

```
1-a-bcd-edf
2-b-bcd-edf
3-c-bcd-edf
4-d_bcd_edf
```

Here, `-edf` is the suffix and `1-a-bcd`, `2-a-bcd`, `3-a-bcd`, and `4-a-bcd` are the xCAT names of the nodes.

If possible, avoid using high-speed node names with variations at the beginning and the end, such as:

```
A-a-bcd-edf-1
B-b-bdc-edf-2
C-c-bcd-edf-3
D-d-bcd-edf-4
```

In such cases, use the **-N** option and specify the node list with the **gssgencluster** and **gssgenclusterrgs** commands. The node names must be reachable from the management server node. xCAT requires that the target nodes be part of a node group and a warning might be issued if the hostname is not defined as an xCAT object.

Example:

1. The xCAT hostnames are **gssio1**, **gssio2**, **gssio3**, and **gssio4**.
2. The high-speed hostnames are **A-test1**, **B-test2**, **C-test3**, **D-test4**. These hostnames are reachable from the management server node. They are not defined in xCAT.

Run:

```
gssgencluster -C test01 -N A-test1,B-test2,C-test3,D-test4
```

Appendix E. Obtaining kernel for system upgrades

For new system installation, the kernel is shipped with the system. However, for upgrades, you need to obtain and package the kernel update, and then follow the kernel update installation procedure.

Use the following steps on the EMS node during an upgrade to obtain and package the kernel update.

1. Clear the version locks.

```
yum versionlock clear
```

2. Connect the management server node to the Red Hat Network.

```
subscription-manager register --username=<X> --password=<Y>
subscription-manager list --available   // list pools
subscription-manager attach --pool=<X>
OR
subscription-manager attach --auto
```

3. Create a directory for the kernel update package.

```
mkdir -p /tmp/kernel/RHSA-2017-0403/
```

4. List all repositories and enable the repositories that are disabled, as required.

```
yum repolist all
yum-config-manager --enable rhel*

OR
subscription-manager config --rhsm.manage_repos=1
```

5. Download the kernel update package.

```
yum update *229.49.1* --downloadonly --downloaddir=/tmp/kernel/RHSA-2017-0403/
yum update dracut*033-241.el7_1.5* --downloadonly --downloaddir=/tmp/kernel/RHSA-2017-0403/
```

6. Package the directory.

```
cd /tmp/kernel ; tar -zcvf RHSA-2017-0403.tgz  RHSA-2017-0403
```

7. Disable the Red Hat Network connections.

```
subscription-manager config --rhsm.manage_repos=0
yum clean all
```

Continue with the kernel update installation steps for `RHSA-2017-xxxx.tgz`. For more information, see Appendix F, “Instructions for installing the ESS Red Hat Linux Errata Kernel Update,” on page 39.

Appendix F. Instructions for installing the ESS Red Hat Linux Errata Kernel Update

This topic provides instructions for installing the Red Hat Linux Errata Kernel Update for ESS.

Perform the following steps to prepare for installation of the ESS Red Hat Linux Errata Kernel Update.

1. Obtain the required Kernel Errata here: http://www-01.ibm.com/support/docview.wss?uid=ssg1S1005719.

   The following packages are provided in the `RHSA-2017-0403.tgz`:

   ```
   dracut-033-241.el7_1.5.ppc64.rpm
   dracut-network-033-241.el7_1.5.ppc64.rpm
   python-perf-3.10.0-229.49.1.el7.ppc64.rpm
   kernel-debug-3.10.0-229.49.1.el7.ppc64.rpm
   kernel-tools-libs-devel-3.10.0-229.49.1.el7.ppc64.rpm
   perf-3.10.0-229.49.1.el7.ppc64.rpm
   kernel-debug-devel-3.10.0-229.49.1.el7.ppc64.rpm
   kernel-devel-3.10.0-229.49.1.el7.ppc64.rpm
   kernel-debuginfo-common-ppc64-3.10.0-229.49.1.el7.ppc64.rpm
   perf-debuginfo-3.10.0-229.49.1.el7.ppc64.rpm
   kernel-abi-whitelists-3.10.0-229.49.1.el7.noarch.rpm
   python-perf-debuginfo-3.10.0-229.49.1.el7.ppc64.rpm
   kernel-tools-3.10.0-229.49.1.el7.ppc64.rpm
   kernel-3.10.0-229.49.1.el7.src.rpm
   kernel-doc-3.10.0-229.49.1.el7.noarch.rpm
   kernel-tools-libs-3.10.0-229.49.1.el7.ppc64.rpm
   kernel-bootwrapper-3.10.0-229.49.1.el7.ppc64.rpm
   kernel-debug-debuginfo-3.10.0-229.49.1.el7.ppc64.rpm
   kernel-3.10.0-229.49.1.el7.ppc64.rpm
   dracut-config-rescue-033-241.el7_1.5.ppc64.rpm
   kernel-debuginfo-3.10.0-229.49.1.el7.ppc64.rpm
   kernel-headers-3.10.0-229.49.1.el7.ppc64.rpm
   kernel-tools-debuginfo-3.10.0-229.49.1.el7.ppc64.rpm
   ```

2. Copy RPMs and setup repository:

   a. Unpack the Errata Kernel archive on the management server node:

   ```
   $ cd /var/tmp
   $ tar -zxvf RHSA-2017-0403.tgz
   ```

   b. Set up the management server node Kernel repository:

   ```
   $ /opt/ibm/gss/xcat/bin/gssxcatconfig -k
   ```

   c. Remove old RPMs from the kernel repository directory:

```
$ cd /install/gss/otherpkgs/rhels7/ppc64/kernel/
$ rm -f *.rpm
```

d. Copy the Errata Kernel RPMs into the repository directory:

```
$ cd /install/gss/otherpkgs/rhels7/ppc64/kernel/
$ cp /var/tmp/RHSA-2017-0403/*.rpm .
```

e. Update the Kernel repository information:

```
$ cd /install/gss/otherpkgs/rhels7/ppc64/kernel/
$ createrepo .
```

3. Return to the corresponding install or upgrade task.

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing IBM Corporation North Castle Drive Armonk, NY 10504-1785 U.S.A.

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing Legal and Intellectual Property Law IBM Japan Ltd. 19-21,

Nihonbashi-Hakozakicho, Chuo-ku Tokyo 103-8510, Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Corporation
Dept. 30ZA/Building 707
Mail Station P300

2455 South Road,
Poughkeepsie, NY 12601-5400
U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment or a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at “Copyright and trademark information” at www.ibm.com/legal/copytrade.shtml.

Intel is a trademark of Intel Corporation or its subsidiaries in the United States and other countries.

Java™ and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, and Windows NT are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Glossary

This glossary provides terms and definitions for the ESS solution.

The following cross-references are used in this glossary:

- *See* refers you from a non-preferred term to the preferred term or from an abbreviation to the spelled-out form.
- *See also* refers you to a related or contrasting term.

For other terms and definitions, see the IBM Terminology website (opens in new window):

http://www.ibm.com/software/globalization/terminology

B

building block
A pair of servers with shared disk enclosures attached.

BOOTP
See *Bootstrap Protocol (BOOTP)*.

Bootstrap Protocol (BOOTP)
A computer networking protocol thst is used in IP networks to automatically assign an IP address to network devices from a configuration server.

C

CEC See *central processor complex (CPC)*.

central electronic complex (CEC)
See *central processor complex (CPC)*.

central processor complex (CPC)
A physical collection of hardware that consists of channels, timers, main storage, and one or more central processors.

cluster
A loosely-coupled collection of independent systems, or *nodes*, organized into a network for the purpose of sharing resources and communicating with each other. See also *GPFS cluster*.

cluster manager
The node that monitors node status using disk leases, detects failures, drives recovery, and selects file system managers. The cluster manager is the node with the lowest node number among the quorum nodes that are operating at a particular time.

compute node
A node with a mounted GPFS file system that is used specifically to run a customer job. ESS disks are not directly visible from and are not managed by this type of node.

CPC See *central processor complex (CPC)*.

D

DA See *declustered array (DA)*.

datagram
A basic transfer unit associated with a packet-switched network.

DCM See *drawer control module (DCM)*.

declustered array (DA)
A disjoint subset of the pdisks in a recovery group.

dependent fileset
A fileset that shares the inode space of an existing independent fileset.

DFM See *direct FSP management (DFM)*.

DHCP See *Dynamic Host Configuration Protocol (DHCP)*.

direct FSP management (DFM)
The ability of the xCAT software to communicate directly with the Power Systems server's service processor without the use of the HMC for management.

drawer control module (DCM)
Essentially, a SAS expander on a storage enclosure drawer.

Dynamic Host Configuration Protocol (DHCP)
A standardized network protocol that is used on IP networks to dynamically distribute such network configuration parameters as IP addresses for interfaces and services.

E

Elastic Storage Server (ESS)
A high-performance, GPFS NSD solution

made up of one or more building blocks that runs on IBM Power Systems servers. The ESS software runs on ESS nodes - management server nodes and I/O server nodes.

encryption key
A mathematical value that allows components to verify that they are in communication with the expected server. Encryption keys are based on a public or private key pair that is created during the installation process. See also *file encryption key (FEK)*, *master encryption key (MEK)*.

ESS See *Elastic Storage Server (ESS)*.

environmental service module (ESM)
Essentially, a SAS expander that attaches to the storage enclosure drives. In the case of multiple drawers in a storage enclosure, the ESM attaches to drawer control modules.

ESM See *environmental service module (ESM)*.

Extreme Cluster/Cloud Administration Toolkit (xCAT)
Scalable, open-source cluster management software. The management infrastructure of ESS is deployed by xCAT.

F

failback
Cluster recovery from failover following repair. See also *failover*.

failover
(1) The assumption of file system duties by another node when a node fails. (2) The process of transferring all control of the ESS to a single cluster in the ESS when the other clusters in the ESS fails. See also *cluster*. (3) The routing of all transactions to a second controller when the first controller fails. See also *cluster*.

failure group
A collection of disks that share common access paths or adapter connection, and could all become unavailable through a single hardware failure.

FEK See *file encryption key (FEK)*.

file encryption key (FEK)
A key used to encrypt sectors of an individual file. See also *encryption key*.

file system
The methods and data structures used to control how data is stored and retrieved.

file system descriptor
A data structure containing key information about a file system. This information includes the disks assigned to the file system (*stripe group*), the current state of the file system, and pointers to key files such as quota files and log files.

file system descriptor quorum
The number of disks needed in order to write the file system descriptor correctly.

file system manager
The provider of services for all the nodes using a single file system. A file system manager processes changes to the state or description of the file system, controls the regions of disks that are allocated to each node, and controls token management and quota management.

fileset A hierarchical grouping of files managed as a unit for balancing workload across a cluster. See also *dependent fileset*, *independent fileset*.

fileset snapshot
A snapshot of an independent fileset plus all dependent filesets.

flexible service processor (FSP)
Firmware that provices diagnosis, initialization, configuration, runtime error detection, and correction. Connects to the HMC.

FQDN
See *fully-qualified domain name (FQDN)*.

FSP See *flexible service processor (FSP)*.

fully-qualified domain name (FQDN)
The complete domain name for a specific computer, or host, on the Internet. The FQDN consists of two parts: the hostname and the domain name.

G

GPFS cluster
A cluster of nodes defined as being available for use by GPFS file systems.

GPFS portability layer
The interface module that each

installation must build for its specific hardware platform and Linux distribution.

GPFS Storage Server (GSS)
A high-performance, GPFS NSD solution made up of one or more building blocks that runs on System x servers.

GSS See *GPFS Storage Server (GSS)*.

H

Hardware Management Console (HMC)
Standard interface for configuring and operating partitioned (LPAR) and SMP systems.

HMC See *Hardware Management Console (HMC)*.

I

IBM Security Key Lifecycle Manager (ISKLM)
For GPFS encryption, the ISKLM is used as an RKM server to store MEKs.

independent fileset
A fileset that has its own inode space.

indirect block
A block that contains pointers to other blocks.

inode The internal structure that describes the individual files in the file system. There is one inode for each file.

inode space
A collection of inode number ranges reserved for an independent fileset, which enables more efficient per-fileset functions.

Internet Protocol (IP)
The primary communication protocol for relaying datagrams across network boundaries. Its routing function enables internetworking and essentially establishes the Internet.

I/O server node
An ESS node that is attached to the ESS storage enclosures. It is the NSD server for the GPFS cluster.

IP See *Internet Protocol (IP)*.

IP over InfiniBand (IPoIB)
Provides an IP network emulation layer on top of InfiniBand RDMA networks, which allows existing applications to run over InfiniBand networks unmodified.

IPoIB See *IP over InfiniBand (IPoIB)*.

ISKLM
See *IBM Security Key Lifecycle Manager (ISKLM)*.

J

JBOD array
The total collection of disks and enclosures over which a recovery group pair is defined.

K

kernel The part of an operating system that contains programs for such tasks as input/output, management and control of hardware, and the scheduling of user tasks.

L

LACP See *Link Aggregation Control Protocol (LACP)*.

Link Aggregation Control Protocol (LACP)
Provides a way to control the bundling of several physical ports together to form a single logical channel.

logical partition (LPAR)
A subset of a server's hardware resources virtualized as a separate computer, each with its own operating system. See also *node*.

LPAR See *logical partition (LPAR)*.

M

management network
A network that is primarily responsible for booting and installing the designated server and compute nodes from the management server.

management server (MS)
An ESS node that hosts the ESS GUI and xCAT and is not connected to storage. It can be part of a GPFS cluster. From a system management perspective, it is the central coordinator of the cluster. It also serves as a client node in an ESS building block.

master encryption key (MEK)
A key that is used to encrypt other keys. See also *encryption key*.

maximum transmission unit (MTU)
The largest packet or frame, specified in octets (eight-bit bytes), that can be sent in a packet- or frame-based network, such as the Internet. The TCP uses the MTU to determine the maximum size of each packet in any transmission.

MEK See *master encryption key (MEK)*.

metadata
A data structure that contains access information about file data. Such structures include inodes, indirect blocks, and directories. These data structures are not accessible to user applications.

MS See *management server (MS)*.

MTU See *maximum transmission unit (MTU)*.

N

Network File System (NFS)
A protocol (developed by Sun Microsystems, Incorporated) that allows any host in a network to gain access to another host or netgroup and their file directories.

Network Shared Disk (NSD)
A component for cluster-wide disk naming and access.

NSD volume ID
A unique 16-digit hexadecimal number that is used to identify and access all NSDs.

node An individual operating-system image within a cluster. Depending on the way in which the computer system is partitioned, it can contain one or more nodes. In a Power Systems environment, synonymous with *logical partition*.

node descriptor
A definition that indicates how IBM Spectrum Scale uses a node. Possible functions include: manager node, client node, quorum node, and non-quorum node.

node number
A number that is generated and maintained by IBM Spectrum Scale as the cluster is created, and as nodes are added to or deleted from the cluster.

node quorum
The minimum number of nodes that must be running in order for the daemon to start.

node quorum with tiebreaker disks
A form of quorum that allows IBM Spectrum Scale to run with as little as one quorum node available, as long as there is access to a majority of the quorum disks.

non-quorum node
A node in a cluster that is not counted for the purposes of quorum determination.

O

OFED See *OpenFabrics Enterprise Distribution (OFED)*.

OpenFabrics Enterprise Distribution (OFED)
An open-source software stack includes software drivers, core kernel code, middleware, and user-level interfaces.

P

pdisk A physical disk.

PortFast
A Cisco network function that can be configured to resolve any problems that could be caused by the amount of time STP takes to transition ports to the Forwarding state.

R

RAID See *redundant array of independent disks (RAID)*.

RDMA
See *remote direct memory access (RDMA)*.

redundant array of independent disks (RAID)
A collection of two or more disk physical drives that present to the host an image of one or more logical disk drives. In the event of a single physical device failure, the data can be read or regenerated from the other disk drives in the array due to data redundancy.

recovery
The process of restoring access to file system data when a failure has occurred. Recovery can involve reconstructing data or providing alternative routing through a different server.

recovery group (RG)
A collection of disks that is set up by IBM Spectrum Scale RAID, in which each disk is connected physically to two servers: a primary server and a backup server.

remote direct memory access (RDMA)
A direct memory access from the memory of one computer into that of another without involving either one's operating system. This permits high-throughput, low-latency networking, which is especially useful in massively-parallel computer clusters.

RGD See *recovery group data (RGD)*.

remote key management server (RKM server)
A server that is used to store master encryption keys.

RG See *recovery group (RG)*.

recovery group data (RGD)
Data that is associated with a recovery group.

RKM server
See *remote key management server (RKM server)*.

S

SAS See *Serial Attached SCSI (SAS)*.

secure shell (SSH)
A cryptographic (encrypted) network protocol for initiating text-based shell sessions securely on remote computers.

Serial Attached SCSI (SAS)
A point-to-point serial protocol that moves data to and from such computer storage devices as hard drives and tape drives.

service network
A private network that is dedicated to managing POWER8 servers. Provides Ethernet-based connectivity among the FSP, CPC, HMC, and management server.

SMP See *symmetric multiprocessing (SMP)*.

Spanning Tree Protocol (STP)
A network protocol that ensures a loop-free topology for any bridged Ethernet local-area network. The basic function of STP is to prevent bridge loops and the broadcast radiation that results from them.

SSH See *secure shell (SSH)*.

STP See *Spanning Tree Protocol (STP)*.

symmetric multiprocessing (SMP)
A computer architecture that provides fast performance by making multiple processors available to complete individual processes simultaneously.

T

TCP See *Transmission Control Protocol (TCP)*.

Transmission Control Protocol (TCP)
A core protocol of the Internet Protocol Suite that provides reliable, ordered, and error-checked delivery of a stream of octets between applications running on hosts communicating over an IP network.

V

VCD See *vdisk configuration data (VCD)*.

vdisk A virtual disk.

vdisk configuration data (VCD)
Configuration data that is associated with a virtual disk.

X

xCAT See *Extreme Cluster/Cloud Administration Toolkit*.

Printed in USA

SC27-8580-03