Elastic Storage Server
Version 4.0

# *Quick Deployment Guide*

SC27-8580-01

Elastic Storage Server
Version 4.0

# *Quick Deployment Guide*

SC27-8580-01

**Note**
Before using this information and the product it supports, read the information in "Notices" on page 21.

This edition applies to version 4.0.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 4.0.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 4.0.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_4.0.0/sts40_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.<br><br>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.<br><br>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.<br><br>*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:<br>`mkcondition -r IBM.FileSystem -e "PercentTotUsed > 90" \`<br>`-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*.<br><br>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: http://www-01.ibm.com/support/docview.wss?uid=ssg1S1005719.
4. Complete one of the following tasks:
   a. Install the ESS system.
   b. Upgrade the ESS system.

## 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
   - HMC is properly configured for the EMS and IO 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 IO 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.

   See Appendix A, “Known issues,” on page 13 for more information.

## Install the management server software

1. Unpack the ESS software archive:

```
tar -zxvf gss_install-4.0.5_ppc64_advanced_20160729T004528Z.tgz
```

2. Check the MD5 checksum:

```
md5sum -c gss_install-4.0.5_ppc64_advanced_20160729T004528Z.md5
```

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

```
/bin/sh gss_install-4.0.5_ppc64_advanced_20160729T004528Z --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-4.0.5_ppc64_advanced_20160729T004528Z –text-only
```

5. For install and deployment go to the section titled **Install the ESS system** on page 2. To upgrade an existing ESS system, go to the section titled **Upgrade the ESS system** on page 6.

## 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. Obtain the required Kernel Errata stated in http://www-01.ibm.com/support/docview.wss?uid=ssg1S1005719 and complete steps 1 to 5 stated in the Section Install the Kernel Update. Also see Appendix B, “Instructions for installing the ESS Red Hat Linux Errata Kernel Update,” on page 19
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
   ```
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 may 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 **gss_iprraid**:
   ```
   updatenode gss_ppc64 -P gss_iprraid
   ```
2. Run **gssstoragequickcheck**:
   ```
   gssstoragequickcheck -G gss_ppc64
   ```
3. Run **gss_sashba script**:
   ```
   updatenode gss_ppc64 -P gss_sashba
   ```
4. 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
   ```
5. 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.
6. 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 13.

## 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 “ESS 4.0.5 issues” on page 13

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 `layer3+4`. In some network configurations xmit_hash_policy of layer2+3 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=layer2+3
```

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 *Deploying the Elastic Storage Server* for:

- more information about **gssgennetworks**
- information about creating a bonded InfiniBand network.

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

Remove or comment out any **CONNECTED_MODE=yes** statement 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
   ```

   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
   ```
   Note that during initial deployment of the nodes, SRC BA15D001 may be logged as 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 13.

## Install the ESS GUI

1. Generate performance collector in the EMS node. The EMS node must be part of the ESS cluster and node name must be the node name used in the cluster (e.g., esm1-hs). Run:

   ```
   mmperfmon config generate --collectors ems1-hs
   ```
2. Set up the nodes in the *ems nodeclass* and *gss_ppc64 nodeclass* for performance monitoring. Run:
   ```
   mmchnode --perfmon -N ems,gss_ppc64
   ```

3. Capacity and filesetquota monitoring is not enabled in the GUI by default.
   a. To enable capacity and filesetquota run the following command:

      ```
      mmperfmon config update GPFSDiskCap.restrict=<EMS node
                name> GPFSDiskCap.period=<period in sec>
      ```

      Here the EMS node name must be the name shown in the mmlscluster output.
   b. Verify that the GPFSDiskCap.period is set correctly in the /opt/IBM/zimon/ZIMonSensors.cfg. If it is not set correctly, verify that the EMS node is correctly provided. The following example shows what it should look like when period is set to 86400 sec (one day) and the task only runs in the ems-hs (mmlscluster output showing ems1-hs.gpfs.net) node.

      ```
      {
                  name = "GPFSDiskCap"
                  period = 86400
                  restroct = "ems1-hs.gpfs.net"
               }
      ```

      **Note:** To enable quota the filesystem quota checking must be enabled. Refer **mmchfs -Q** and **mmcheckquota** commands in the *IBM Spectrum Scale Administration and Programming Reference Guide.*
4. Start sensors in the EMS node and I/O server nodes:

   ```
   systemctl start pmsensors
   xdsh gss_ppc64 "systemctl start pmsensors"
   ```
5. Start performance collector in the EMS 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 EMS 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 Install 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 3.0

ESS versions 2.5.x should first be upgraded to ESS 3.5.2 prior to upgrading to ESS 4.0. Contact Customer Support for further details.

## Upgrading from ESS 3.x.y and ESS 4.0.x

Perform the following steps if you are upgrading from ESS 3.x.y and ESS 4.0.x:

1. Update ESS repositories on the management server node:

   ```
   cd /opt/ibm/gss/install
   installer/gssinstall -m manifest -u
   ```
2. 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 /tmp
   ```
3. Save sensor configuration files in EMS and IO Server nodes (ssh to IO Server nodes and run the command locally) for later usage:

   ```
   cp /opt/IBM/zimon/ZIMonSensors.cfg /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 steps 1 to 5 in the section stated in the Section Install the Kernel Update. Also see Appendix B, “Instructions for installing the ESS Red Hat Linux Errata Kernel Update,” on page 19. 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. If upgrading from ESS 3.x.y perform the following steps to upgrade IBM Spectrum Scale RAID configuration parameters.

   Skip these steps if upgrading from ESS 4.0.x.

   ```
   /opt/ibm/gss/tools/samples/gssupg400.sh -b
   /opt/ibm/gss/tools/samples/gssupg400.sh -c
   /opt/ibm/gss/tools/samples/gssupg400.sh -p
   ```

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 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, shut down IBM Spectrum Scale on the current I/O server node while maintaining quorum:

```
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 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/gssupg400.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 one 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
   ```

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 13.

## Upgrading GUI

Upgrading from ESS 3.0.x

Perform the following steps to upgrade from ESS 3.0.x:

1. Restore the sensor configuration file (that was saved in step 2 of the Update the management server node section) in each node (ssh to EMS and IO Server nodes and run the command locally):

   ```
   cp /tmp/ZIMonSensors.cfg /opt/IBM/zimon
   ```

2. Restore the collector configuration file that was saved in step 2 of the Update management server node section:

   ```
   cp /tmp/ZIMonCollector.cfg /opt/IBM/zimon
   ```

3. Start the performance data collector service in the EMS node and sensor for performance data collection in all nodes:

   ```
   systemctl start pmcollector
   systemctl start pmsensors
   xdsh gss_ppc64 "systemctl restart pmsensors"
   ```

4. Enable and start the GUI:

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

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 in the EMS node. The EMS 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). Run:

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

2. Set up the nodes in the *ems nodeclass* and *gss_ppc64 nodeclass* for performance monitoring. Run:

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

3. Capacity and filesetquota monitoring is not enabled in the GUI by default.
   a. To enable capacity and filesetquota run the following command:

      ```
      mmperfmon config update GPFSDiskCap.restrict=<EMS node
      name> GPFSDiskCap.period=<period in sec>
      ```

      Here the EMS node name must be the name shown in the mmlscluster output.

   b. Verify that the GPFSDiskCap.period is set correctly in the /opt/IBM/zimon/ZIMonSensors.cfg. If it is not set correctly, verify that the EMS node is corectly provided. The following example shows what it should look like when period is set to 86400 sec (one day) and the task only runs in the ems1-hs (mmlscluster output showing ems1-hs.gpfs.net) node.

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

      ```
      {
                  name = "GPFSDiskCap"
                  period = 86400
                  restroct = "ems1-hs.gpfs.net"
              }
      ```

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

4. Start sensors in the EMS node and I/O server nodes:

   ```
   systemctl start pmsensors
   systemctl start pmcollector
   xdsh gss_ppc64 "systemctl start pmsensors"
   ```

5. Enable and start the GUI:

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

This completes the upgrade task of the ESS system.

# Appendix A. Known issues

This topic includes known issues for ESS.

## ESS 4.0.5 issues

Table 2 includes information about known issues in ESS 4.0.5 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 4.0.5*

| Issue | Environment affected | Description | Resolution or action |
|---|---|---|---|
| **1.** After upgrading to to ESS 4.0.5 (from pre-ESS 3.0.0), the Java™ 1.6 rpm still exists which may fail some security scans. | Cluster software upgrade<br><br>Type: Upgrade<br><br>IBM Spectrum ScaleVersion: Advanced or Standard<br><br>Affected nodes: I/O IO + EMS | As of 3.0.0 a separate java rpm is no longer required for GUI operation (java is packaged within the GUI rpm). The java rpm remains on the system and may cause some security scans to fail. The workaround is to simply remove the rpm via yum.<br><br>The rpm in question is:<br><br>java-1.6.0-ibm-1.6.0.16.7-1jpp.ppc64 | To work around this issue, do the following:<br>1. Run:<br>`xdsh ems1,gss_ppc64 "yum -y remove java-1.6*"`<br>2. Confirm the removal by running this command:<br>`xdsh ems1,gss_ppc64 "rpm -qa \| grep -i java-1.6"`<br>No output should be returned. |

*Table 2. Known issues in ESS 4.0.5 (continued)*

| Issue | Environment affected | Description | Resolution or action |
|---|---|---|---|
| **2. gssinstallcheck** may flag an error regarding page pool size in multi-building block situations if the physical memory sizes differ. | Software validation<br><br>Type: Install or Upgrade<br><br>IBM Spectrum Scale Version: Advanced or Standard<br><br>Affected nodes: I/O | **gssinstallcheck** is a new tool introduced in ESS 3.5, that helps validate software, firmware, and configuration settings.<br><br>If adding (or installing) building blocks of a different memory footprint installcheck will flag this as an error. Best practice states that your IO servers should all have the same memory footprint, thus pagepool value. Page pool is currently set at ~60% of physical memory per IO node.<br><br>Example from **gssinstallcheck**:<br>`[ERROR] pagepool: found 142807662592 expected range 147028338278 - 179529339371` | 1. Confirm each IO node's individual memory footprint. From the EMS, run the following command against your IO xCAT group:<br>`xdsh gss_ppc64 "cat/ proc/meminfo \| grep MemTotal"`<br><br>**Note:** This value is in KB<br>If the physical memory varies between servers and/or building blocks, consider adding mermory and re-calculating pagepool to ensure consistency.<br>2. Validate the pagepool settings in Spectrum Scale:<br>`mmlsconfig \| grep -A 1 pagepool`<br><br>**Note:** This value is in MB.<br>If the pagepool value setting is not roughly ~60% of physical memory, then you should consider recalculating and setting an updated value.<br>Please refer to the Spectrum Scale documentation on how to update the pagepool value.<br>http://www-01.ibm.com/support/knowledgecenter/SSFKCN/gpfs_welcome.html |
| **3.** New Disk Prep Script fails when copying to IO nodes. Issue will be called out in gssinstallcheck. This issue is seen when upgrading from ESS 3.0.X only | Software validation<br><br>Type: Upgrade<br><br>Version: Advanced or Standard<br><br>Affected Nodes:IO | When a user upgrades from ESS 3.0.X, the command **gssUpg400.sh -p** fails to copy the New Disk Prep Script to the IO nodes.<br><br>Users will see “End of Task” vs “Copying prepare new disk script".<br><br>Further validation of the issue can be found when running **gssinstallcheck**. A user would see:<br>`[ERROR] New disk prep script not found.` | To work around this issue, run the following from the EMS:<br>`/usr/lpp/mmfs/bin/ mmdsh -N gss_ppc64 "cp /usr/lpp/mmfs/ samples/vdisk/ tspreparenewpdiskforuse /usr/lpp/mmfs/bin/ tspreparenewpdiskforuse`<br><br>To validate, do the following:<br>`xdsh gss_ppc64 "ls -l /usr/lpp/mmfs/bin/ tspreparenewpdiskforuse"`<br><br>Finally, re-run **gssinstallcheck** to validate that the error has been resolved. |

*Table 2. Known issues in ESS 4.0.5 (continued)*

| Issue | Environment affected | Description | Resolution or action |
|---|---|---|---|
| **4.** When upgrading to ESS 4.0.5 the gss_ofed script may fail for nodes upgraded to Red Hat 7.2. | Cluster Network<br><br>Type: Upgrade<br><br>Version: Advanced or Standard<br><br>Affected Nodes:IO + EMS | Users who connect nodes to the Red Hat Network should only apply specific security updates. If all non-essential updates are applied (not advised), systems may not fully upgrade to Red Hat 7.2. When upgrading to ESS 4.0.5, users would encounter the following issue when running the gss_ofed script:<br>`gssio1: gss_ofed [DEBUG]: Detected rhel7u2 ppc64. Disabling installing 32bit rpms...`<br>`gssio1: gss_ofed [DEBUG]: Error: The current MLNX_OFED_LINUX is intended for rhel7.1`<br>`gssio1: gss_ofed [ERROR]: Mellanox install failed RC: 172` | The workaround for this issue is to downgrade the redhat-release version back to 7.1. For each node experiencing this problem run the following command:<br>`yum -y downgrade redhat-release-server unbound-libs`<br><br>Then validate the level again:`cat /etc/redhat-release`<br><br>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. |

*Table 2. Known issues in ESS 4.0.5 (continued)*

| Issue | Environment affected | Description | Resolution or action |
|---|---|---|---|
| **5.** gssgennetworks script requires high-speed hostnames to be derived from IO Server (xCAT) hostnames using suffix, prefix, or both. | High-speed network generation<br><br>Type: Install<br><br>Version: Advanced or Standard | Gssgennetworks requires the target hostname 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<br><br>high-speed name you may be affected by this issue.<br><br>A typical deployment scenario is:<br><br>`gssio1 // xCAT name`<br><br>`gssio1-hs // high-speed`<br><br>An Issue scenario is:<br><br>`gssio1 // xCAT name`<br><br>`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 hostnames can be derived from it using some combination of suffix and/or prefix.<br><br>For example if the high-speed hostnames are foo1abc-hs, goo1abc-hs :<br><br>a. Add foo1 and goo1 to the /etc/hosts using management network address (reachable) in the EMS node only.<br><br>b. Use:<br><br>gssgennetworks -N foo1,goo1 –suffix abc-hs --create-bond<br><br>c. Remove the entries foo1 and goo1 from /etc/hosts file on the EMS node once the high-speed networks are created.<br><br>Example of how to fix (/etc/hosts):<br>`// Before`<br>`<IP><Long Name><Short Name>`<br>`192.168.40.21`<br>`gssio1.gpfs.net`<br>`gssio1`<br>`192.168.40.22`<br>`gssio2.gpfs.net`<br>`gssio2`<br>`X.X.X.X`<br>`foo1abc-hs.gpfs.net`<br>`foo1abc-hs`<br>`X.X.X.Y`<br>`goo1abc-hs.gpfs.net`<br>`goo1abc-hs`<br>`// Fix`<br>`192.168.40.21`<br>`gssio1.gpfs.net`<br>`gssio1 foo1`<br>`192.168.40.22`<br>`gssio2.gpfs.net`<br>`gssio2 goo1`<br>`X.X.X.X`<br>`foo1abc-hs.gpfs.net`<br>`foo1abc-hs`<br>`X.X.X.Y`<br>`goo1abc-hs.gpfs.net`<br>`goo1abc-hs`<br>`gssgennetworks -N foo1,`<br>`goo1 --suffix=abc-hs`<br>`--create-bond` |

*Table 2. Known issues in ESS 4.0.5 (continued)*

| Issue | Environment affected | Description | Resolution or action |
|---|---|---|---|
| **6.** During the GUI wizard setup, you may see an erroneous error during the firmware verification. | GUI<br><br>Type: Install<br><br>Version:<br><br>Advanced or Standard<br><br>Affected Nodes: EMS (GUI host) | Upon a new installation of an ESS 4.0.5 system users may receive an error during the GUI wizard setup stating that the ESM firmware is unknown (Firmware levels are not up to date).<br><br>Verify firmware is only run if the user checks the box during the wizard setup. This error is deemed invalid and should not prevent the successful completion of the ESS install or GUI wizard. | There is no resolution at this time. Please take the following steps though to assure a healthy system:<br>• Make sure gnrhealthcheck is clean of any problems.<br>• Make sure gssinstallcheck is run and all firmware levels are up to date and valid.<br><br>After the GUI wizard is complete make sure there are no events requiring action. |
| **7.** During the GUI wizard setup, you may see a failure during the 'Verify Installation' system setup screen regarding Performance Monitoring. | GUI<br><br>Type: Install<br><br>Version:<br><br>Advanced or Standard<br><br>Affected Nodes: EMS (GUI host) | The performance monitoring check of 'Verify Installation' may fail in the GUI wizard. You will see an error like:<br><br>"Did not get performance data from any node..."<br><br>Or "Could not get performance counters from node 'X'" | The workaround is simply to press the 'Verify Installation again' button. You may receive the error 2 or 3 times before a successful green checkmark is displayed next to Performance Monitoring. If after 3 times the Performance Monitoring check continues to fail, please check that your pmsensors and pmcollector are correctly configured and started. |
| **8.** During initial deployment of the nodes, SRC BA15D001 is logged by Partition Firmware | Platform Firmware<br><br>Type: Install<br><br>Version:<br><br>Advanced or Standard<br><br>Affected Nodes: IO + EMS | During deployment of EMS and IO 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 IO Server nodes where some network ports can remain unpopulated in a valid configuration. Therefore, this error can be ignored when generated during deployment. |

# Appendix B. 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 `RHBA-2016-1183_for_ESS.tgz`:

   ```
   dracut-033-241.el7_1.5.ppc64.rpm
   dracut-config-rescue-033-241.el7_1.5.ppc64.rpm
   dracut-network-033-241.el7_1.5.ppc64.rpm
   kernel-3.10.0-229.34.1.el7.ppc64.rpm
   kernel-abi-whitelists-3.10.0-229.34.1.el7.noarch.rpm
   kernel-bootwrapper-3.10.0-229.34.1.el7.ppc64.rpm
   kernel-debug-devel-3.10.0-229.34.1.el7.ppc64.rpm
   kernel-devel-3.10.0-229.34.1.el7.ppc64.rpm
   kernel-doc-3.10.0-229.34.1.el7.noarch.rpm
   kernel-headers-3.10.0-229.34.1.el7.ppc64.rpm
   kernel-tools-3.10.0-229.34.1.el7.ppc64.rpm
   kernel-tools-libs-3.10.0-229.34.1.el7.ppc64.rpm
   perf-3.10.0-229.34.1.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 RHBA-2016-1183_for_ESS.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/RHBA-2016-1183_for_ESS/*.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

SC27-8580-01