

Elastic Storage Server
Version 5.3.2.1

*Protocol Nodes Quick Deployment
Guide*



Elastic Storage Server
Version 5.3.2.1

*Protocol Nodes Quick Deployment
Guide*



Note

Before using this information and the product it supports, read the information in “Notices” on page 35.

- | This edition applies to version 5.3.2.1 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 ix. 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.

© **Copyright IBM Corporation 2018, 2019.**

US Government Users Restricted Rights – Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

Tables	v	Upgrading ESS Red Hat software and system firmware	29
About this information	vii	Performing prechecks to prepare for IBM Spectrum Scale upgrade	30
Who should read this information.	vii	Upgrading IBM Spectrum Scale on protocol nodes by using the installation toolkit.	32
Related information	vii		
Conventions used in this information	viii		
How to submit your comments	ix		
Change history	xi		
Chapter 1. CES and protocol nodes support in ESS.	1	Appendix. 5148-22L protocol node diagrams.	33
Chapter 2. Configuration 1: 5148-22L protocol nodes ordered and racked with a new 5148 ESS (PPC64LE)	5	Notices	35
Chapter 3. Configuration 2: 5148-22L protocol nodes ordered standalone and added to an existing 5148 ESS (PPC64LE)	13	Trademarks	36
Chapter 4. GPFS configuration parameters for protocol nodes	23	Glossary	39
Chapter 5. OS tuning for RHEL 7.5 PPC64LE protocol nodes.	25		
Chapter 6. Upgrading a cluster containing ESS and protocol nodes	27		
Planning upgrade in a cluster containing ESS and protocol nodes	27		

Tables

1.	Conventions	viii
----	-----------------------	------

About this information

This information guides you in installing, or upgrading to, version 5.3.2.x of the Elastic Storage Server (ESS).

Who should read this information

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

Related information

ESS information

The ESS 5.3.2.x library consists of following information units. You can access these publications on IBM® Knowledge Center or IBM Publications Center.

- *Elastic Storage Server: Quick Deployment Guide*, SC27-9526
- | • *Elastic Storage Server: Protocol Nodes Quick Deployment Guide*, SC27-9527
- *Elastic Storage Server: Problem Determination Guide*, SC27-9529
- *Elastic Storage Server: Command Reference*, SC27-9530
- *IBM Spectrum Scale RAID: Administration*, SC27-9528
- *IBM ESS Expansion: Quick Installation Guide (Model 084)*, SC27-4627
- *IBM ESS Expansion: Installation and User Guide (Model 084)*, SC27-4628
- *IBM ESS Expansion: Hot Swap Side Card - Quick Installation Guide (Model 084)*, GC27-9210
- | • *IBM ESS Expansion: Hardware Installation and Maintenance Guide (Model 106)*, SC27-9211
- | • *IBM ESS Expansion: Overview of CMA and Rail Kit Hardware Fasteners (Model 106)*, SC27-9296
- *Installing the Model 024, ESLL, or ESLS storage enclosure*, GI11-9921
- *Removing and replacing parts in the 5147-024, ESLL, and ESLS storage enclosure*
- *Disk drives or solid-state drives for the 5147-024, ESLL, or ESLS storage enclosure*
- For information about the DCS3700 storage enclosure, see:
 - *System Storage® DCS3700 Quick Start Guide*, GA32-0960-04:
 - *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=s8g1S7004920>
- For information about 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

For more information, see IBM Knowledge Center:

http://www-01.ibm.com/support/knowledgecenter/SSYSP8_5.3.2/sts532_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/gnrfaq.html>

Other related information

For information about:

- IBM Spectrum Scale, see:
http://www.ibm.com/support/knowledgecenter/STXKQY/ibmspectrumscale_welcome.html
- IBM Spectrum Scale call home, see Understanding call home.
- Installing IBM Spectrum Scale and CES protocols with the installation toolkit, see Installing IBM Spectrum Scale on Linux nodes with the installation toolkit.
- IBM POWER8® servers, see IBM Knowledge Center:
<http://www.ibm.com/support/knowledgecenter/POWER8/p8hdx/POWER8welcome.htm>
- Extreme Cluster/Cloud Administration Toolkit (xCAT), go to the xCAT website :
<http://xcat.org/>
 - xCAT 2.14.3 Release Notes
- Mellanox OFED Release Notes (4.4), go to https://www.mellanox.com/related-docs/prod_software/Mellanox_OFED_Linux_Release_Notes_4_4-1_0_0_0.pdf
- IBM Electronic Service Agent (ESA) documentation, go to <https://www-01.ibm.com/support/esa/>.
- Drive call home, go to Drive call home in 5146 and 5148 systems.

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.
<u>bold underlined</u>	<u>bold underlined</u> 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.
<i>italic</i>	<i>Italic</i> words or characters represent variable values that you must supply. <i>Italics</i> 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 <i>Enter</i> .
\	In command examples, a backslash indicates that the command or coding example continues on the next line. For example: <pre>mkcondition -r IBM.FileSystem -e "PercentTotUsed > 90" \ -E "PercentTotUsed < 85" -m p "FileSystem space used"</pre>
{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>.

Table 1. Conventions (continued)

Convention	Usage
<i>item...</i>	Ellipses indicate that you can repeat the preceding item one or more times.
	In <i>synopsis</i> statements, vertical lines separate a list of choices. In other words, a vertical line means <i>Or</i> . 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

Change history

Version	PDF form number	Summary
2	SC27-9527-01	Initial version for ESS 5.3.2.1
1	SC27-9527-00	Initial version for ESS 5.3.2

Chapter 1. CES and protocol nodes support in ESS

- “Overview of CES and protocol nodes”
- “Supported protocol node configurations” on page 2
- “5148-22L protocol node hardware” on page 2
- “5148-22L protocol node software” on page 2
- “Customer supplied protocol node hardware recommendations” on page 4
- “Customer supplied protocol node software management recommendations” on page 4

Overview of CES and protocol nodes

Cluster Export Services (CES) provides highly available file and object services to an IBM Spectrum Scale cluster by using Network File System (NFS), Object, or Server Message Block (SMB) protocols. Because CES has specific hardware and software requirements, the code must be installed on nodes designated to run the CES software stack. These nodes are called protocol nodes.

Protocol nodes can be added to an IBM Spectrum Scale cluster containing an ESS building block. They can also exist in non-ESS IBM Spectrum Scale clusters. SMB and NFS functions on protocol nodes can exist in clusters in which their storage is remotely mounted.

For more information, see the following resources.

- IBM Spectrum Scale FAQ
 - Covers all IBM Spectrum Scale levels and is frequently updated
 - Contains minimum requirements for hardware and software
 - Contains code level support statements
- Protocols Quick Overview for IBM Spectrum Scale
 - General flow of the IBM Spectrum Scale installation toolkit along with examples
 - ESS specific examples are on page 2
- IBM Spectrum Scale CES protocols documentation:
 - Protocols support overview
 - Planning for protocols
 - Best practices for SMB
 - Installation and deployment of protocols
 - Upgrade of IBM Spectrum Scale components including protocols
 - Configuration of CES and protocols
 - Managing protocol services
 - Managing protocol user authentication
 - Managing SMB and NFS data exports
 - Managing object storage
 - Monitoring IBM Spectrum Scale components including protocols
 - Troubleshooting
 - Configuration changes required on protocol nodes to collect core dump data
 - mmsmb command
 - mmnfs command
 - mmobj command

Supported protocol node configurations

The following protocol node configurations that are currently supported.

- Chapter 2, “Configuration 1: 5148-22L protocol nodes ordered and racked with a new 5148 ESS (PPC64LE),” on page 5
- Chapter 3, “Configuration 2: 5148-22L protocol nodes ordered standalone and added to an existing 5148 ESS (PPC64LE),” on page 13

5148-22L protocol node hardware

With the ESS 5.3.1.1 release, a protocol node feature code is introduced. This protocol node feature code allows the purchase of Power8 nodes with a very specific hardware configuration, tested and tuned by IBM for providing CES services. The machine type and model (MTM) for protocol nodes is 5148-22L and it ships with the following hardware configuration.

- 8247-22L Power8 model
- 2 x 10core 3.34 Ghz Power8 Processors
- 128 GB or greater memory
- Two 600 GB 10k RPM SAS HDDs in RAID10 mirror using IPRAID adapter
- 1 GbE 4port network adapter in slot C12
- ¹ Three x 16 or x8 network adapters in slots C5, C6, C7
- ¹ Four x 8 network adapters in slots C2, C3®, C10, C11

¹ It is recommended to plan for the GPFS™ admin or daemon network to use separate network adapter(s) from the Cluster Export Services.

For more information, see “5148-22L protocol node diagrams,” on page 33.

5148-22L protocol node software

With the ESS 5.3.1.1 release or later, the ESS EMS node contains the capability to manage certain aspects of a 5148-22L protocol node. This allows the protocol node(s) and ESS building block to have OS, driver, and firmware levels kept in synchronization. To understand which toolsets manage each aspect of the protocol node and the supported code levels, see the following table.

5148-22L protocol node components	Supported level	Managed by
Power8 FW	01SV860_165 (FW860.51) (in OPAL mode)	User (using update_flash)
OS	RHEL7.5 PPC64LE	EMS node (using gssdeploy)
kernel	3.10.0-862.25.3.el7	
systemd	219-57.el7_5.3	EMS node (using gssdeploy)
network manager	1.12.0-8.el7_6	EMS node (using gssdeploy)
Mellanox OFED (Firmware and driver)	Driver: MLNX_OFED_LINUX-4.4-2.0.8.1	EMS node (using gssdeploy)
Power8 IPR (Firmware and driver)	19511000	EMS node (using gssdeploy)
sysctl, tuned, udev rules	Chapter 5, “OS tuning for RHEL 7.5 PPC64LE protocol nodes,” on page 25	EMS node (using gssdeploy)
gpfs.gss.tools	5.3.2.1	EMS node (using gssdeploy)

Red Hat Enterprise Linux OS syslog	---	User managed for debug (contained on each Protocol node)
SSH key setup	---	EMS node (using gssdeploy) or User
/etc/hosts	---	EMS node (using gssdeploy) or User
Repositories	---	EMS node (using gssdeploy) (persistent) IBM Spectrum Scale installation toolkit (only while running)
/etc/resolv.conf	---	User
GPFS network configuration	---	User (using gssgennetwork)
CES base network configuration	---	User
CES shared root creation	---	EMS node (using gssgenvdisks --crcefs)
IBM Spectrum Scale code (RPM install, node add, license)	5.0.2.1 or later Note: It is recommended to keep IBM Spectrum Scale at the same level across all nodes, but a higher level of IBM Spectrum Scale can be used for protocol nodes, if desired. For more information, see IBM Spectrum Scale supported upgrade paths.	IBM Spectrum Scale installation toolkit (installation phase)
IBM Spectrum Scale CES code (rpm install, CES IPs, protocol enablement, license)		IBM Spectrum Scale installation toolkit (deployment phase)
CES protocol performance monitoring sensors		IBM Spectrum Scale installation toolkit (deployment phase)
CES authentication configuration		User (using mmuserauth) or IBM Spectrum Scale installation toolkit
Call home config		EMS (via gsscallhomeconf)
GUI integration of protocol nodes		EMS node (automatically detects CES nodes) User (using GUI rack location adjustment)
GPFS configuration parameters	Chapter 4, "GPFS configuration parameters for protocol nodes," on page 23	User (using mmchconfig) User (using GUI)
Health checks	---	EMS (using gssinstallcheck) EMS (using gssstoragequickcheck) IBM Spectrum Scale installation toolkit (prechecks and postchecks) User (using mmhealth and GUI)

Customer supplied protocol node hardware recommendations

Non-5148-22L protocol nodes can be based upon PPC64LE, PPC64BE, or x86_64 architectures and they are suggested to adhere to the guidance in IBM Spectrum Scale FAQ. All hardware, cabling or connections, and power sequencing is owned by the customer.

Support for IBM Spectrum Scale on the chosen hardware is handled by the general IBM Spectrum Scale support and not by ESS Solution Support.

Customer supplied protocol node software management recommendations

Only 5148-22L protocol nodes are supported for management by the EMS node. If a protocol node is not ordered with this exact model and type from IBM manufacturing, it cannot be added to the EMS xCAT server. While non-5148-22L protocol nodes of type PPC64LE, PPC64BE, or x86_64, might be joined to an ESS cluster using the IBM Spectrum Scale installation toolkit or the `mmaddnode` command, they must be owned, installed, deployed, upgraded, and managed by the customer. The ESS toolsets cannot be used for the management of OS, kernel, network manager, systemd, OFED, or firmware on non-5148-22L nodes. Using the ESS toolsets, including the EMS xCAT server, to manage customer supplied non-5148-22L protocol nodes is not supported.

The IBM Spectrum Scale code on customer supplied non-5148-22L protocol nodes is also managed separately from the ESS toolsets. The IBM Spectrum Scale installation toolkit can be used in this configuration to install or deploy and upgrade the IBM Spectrum Scale code on any customer supplied non-5148-22L protocol nodes. For more information about the installation toolkit, see IBM Spectrum Scale installation toolkit.

For latest information on supported levels, see IBM Spectrum Scale FAQ.

Chapter 2. Configuration 1: 5148-22L protocol nodes ordered and racked with a new 5148 ESS (PPC64LE)

In this configuration, both a new 5148 ESS and new 5148-22L protocol nodes are ordered and racked together. The EMS node, I/O server nodes, and protocol nodes have OS, kernel, systemd, network manager, firmware, and OFED, kept in synchronization as xCAT running on the EMS is used to manage and coordinate these levels. It is recommended to match IBM Spectrum Scale code levels between the ESS and protocol nodes, but this is not mandatory.

Note: All protocol nodes in a cluster must be at the same code level.

Overall flow

- “A) Starting point and what to expect upon initial power-on”
- “B) Power on the protocol nodes and perform health checks” on page 6
- “C) Decide which adapter(s) to use for the GPFS network(s) vs CES protocol network(s)” on page 6
- “D) Set up the high-speed network for GPFS” on page 6
- “E) Configure network adapters to be used for CES protocols” on page 7
- “F) Create a CES shared root file system for use with protocol nodes” on page 8
- “G) Locate the IBM Spectrum Scale protocols package on a protocol node in /root” on page 8
- “H) Extract the IBM Spectrum Scale protocols package” on page 8
- “I) Configure the IBM Spectrum Scale installation toolkit” on page 8
- “J) Installation phase of IBM Spectrum Scale installation toolkit” on page 9
- “K) Deployment phase of IBM Spectrum Scale installation toolkit” on page 10
- “L) Tune the protocol nodes as desired” on page 10
- “M) GUI configuration” on page 10
- “N) Call home setup” on page 11
- “O) Movement of quorum or management function to protocol nodes and off EMS or I/O nodes” on page 11

A) Starting point and what to expect upon initial power-on

- This configuration assumes Fusion mode has been run, thus retaining the shipped xCAT configuration of EMS, I/O server nodes, and protocol nodes. If Fusion mode has not been run and the ESS system has been reinstalled, then start with Configuration 2.
- The EMS and I/O server nodes are installed, high speed network has been created, and GPFS is active and healthy.
- The GUI has not yet been started nor configured. This is done following the protocol node configuration.
- Hardware and software call home is not yet configured. This is done after protocols are deployed.
- Protocol nodes exist in XCAT as prt01, prt0N objects.
Run the following command to verify.

```
# lsdef
```
- Protocol nodes are part of the ces_ppc64 xCAT group.
Run the following command to verify.

```
# lsdef -t group ces_ppc64
```
- An xCAT OS image specific for CES exists (rhels7.5-ppc64le-install-ces).
Run the following command to verify.

```
# lsdef -t osimage
```

- Protocol nodes come with authentication prerequisites pre-installed upon them (includes sssd, ypbind, openldap-clients, krb5-workstation).
- A deploy template exists, already customized for the protocol nodes.
Check for `/var/tmp/gssdeployces.cfg` on the EMS node.
- Protocol nodes already have RHEL 7.5 OS, kernel, network manager, systemd, OFED, iprraid, and firmware loaded at the same levels as the EMS and I/O server nodes. This is verified in this step.
- Protocol nodes do not have any GPFS RPMS installed except `gpfs.gss.tools`.
- Protocol nodes are already cabled to the internal ESS switch for use with the xCAT network and the FSP network. For more information, see Figure 2 on page 33.

B) Power on the protocol nodes and perform health checks

Important: Before proceeding, power on the protocol nodes, if they have not been powered on yet.

- Run `gssstoragequickcheck -G ces_ppc64` to verify network adapter types, slots, and machine type model of the protocol nodes.
- Run `gssinstallcheck -G ces_ppc64` to verify code and firmware levels on the protocol nodes.

C) Decide which adapter(s) to use for the GPFS network(s) vs CES protocol network(s)

It is recommended to plan for separation of the GPFS and CES networks, both by subnet and by card.

Note: Before proceeding, protocol nodes must be cabled up to the GPFS cluster network and to the CES network.

D) Set up the high-speed network for GPFS

Customer networking requirements are site-specific. The use of bonding to increase fault-tolerance and performance is advised but guidelines for doing this have not been provided in this document. Consult with your local network administrator before proceeding further. Before creating network bonds, carefully read *ESS networking considerations* in *Elastic Storage Server: Quick Deployment Guide*.

Make sure that the protocol nodes high speed network IPs and host names are present in `/etc/hosts` on all nodes.

Here is an example excerpt from `/etc/hosts`, showing the `-hs` suffix IPs and host names to be used for the GPFS cluster configuration.

```
172.31.250.3 ems1-hs.gpfs.net ems1-hs
172.31.250.1 gssio1-hs.gpfs.net gssio1-hs
172.31.250.2 gssio2-hs.gpfs.net gssio2-hs
172.31.250.11 prt01-hs.gpfs.net prt01-hs
172.31.250.12 prt02-hs.gpfs.net prt02-hs
172.31.250.13 prt03-hs.gpfs.net prt03-hs
```

Note:

- All nodes must be able to resolve all IPs, FQDNs, and host names, and ssh-keys must work.
- If the `/etc/hosts` file is already set up on the EMS node, copy it to the protocol node(s) first and then modify it. Each protocol node must have the same `/etc/hosts` file.

To set up bond over IB, run the following command.

```
gssgennetworks -G ces_ppc64 --create --ipoib --suffix=-hs --mtu 4092
```

In this example, MTU is set to 4092. The default MTU is 2048 (2K) and the **gssgennetworks** command supports 2048 (2K) and 4092 (4K) MTU. Consult your network administrator for the proper MTU setting.

To set up bond over Ethernet, run the following command.

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

For information on Infiniband issue with multiple fabrics, see the *Infiniband with multiple fabric* section in the *ESS networking considerations* topic of *Elastic Storage Server: Quick Deployment Guide*.

E) Configure network adapters to be used for CES protocols

Before deploying protocols, it is important to understand the customer network and protocol access requirements. CES protocols use a pool of CES IPs which float between nodes, providing redundancy in the case of node failure or degradation. The CES IPs are assigned and aliased by IBM Spectrum Scale to an adapter on each protocol node that has a matching predefined route and subnet. It is important that each protocol node has a base network adapter or bonded group of network adapters or ports with an established IP and routing so that CES IPs can be assigned by IBM Spectrum Scale code during protocol deployment.

Note: CES IPs are never assigned using **ifcfg** or **nmcli** commands. This is handled by the IBM Spectrum Scale code.

The following must be taken into account when planning this network:

- Bandwidth requirements per protocol node (how many ports per bond, bonding mode, and adapter speed)
- Redundancy of each protocol node, if needed. This determines the bonding mode used.
- Authentication domain and DNS. This determines the subnet(s) required for each protocol node.
- Are VLAN tags needed?
- Set aside 1 IP per protocol node, per desired CES subnet. You will be using these when configuring the CES base adapter(s) on each protocol node. These IPs must be setup for forward and reverse DNS lookup.
- Set aside a pool of CES IPs for later use. These IPs must be in DNS and be setup for both forward and reverse DNS lookup. You will not be assigning these IPs to network adapters on protocol nodes.
- Prepare to configure each protocol node to point to the authentication domain or DNS. You need to do this manually using **ifcfg** or **nmcli** commands and by verifying `/etc/resolv.conf` after the settings have taken effect. When deployed from an EMS node, each protocol node might already have a default domain of `gpfs.net` present in `/etc/resolv.conf` and the `ifcfg` files. This default domain can be removed so that it does not interfere with the authentication setup and DNS for protocols.

Proceed with either configuring the CES protocol adapters manually using **ifcfg** or **nmcli** commands or by using **gssgennetworks**. The **gssgennetworks** command cannot be used if your CES protocol network requires VLAN tags nor does it set up additional domain or DNS servers.

When the network is configured on each protocol nodes, verify it using the **mmnetverify** command with these actions:

- Make sure all protocol nodes can ping each other's base CES network by IP, host name, and FQDN.
- Make sure all protocol nodes can ping the authentication server by IP, host name, and FQDN.
- Make sure the authentication server can ping each protocol node base CES network by IP, host name, and FQDN
- Spot check the desired NFS, SMB, or OBJ clients, external to the GPFS cluster and verify that they can ping each protocol node base CES network by IP, hostname, and FQDN
- Even though the CES IP pool is not yet set up, because protocols are not deployed, double check that each protocol node can resolve each CES IP or host name using **nslookup**.

For an example showing how CES IP aliasing relies upon an established base adapter with proper subnets or routing, see CES IP aliasing to network adapters on protocol nodes.

For an example of CES-IP configuration that can be performed after deployment of protocols, see Configuring CES protocol service IP addresses.

F) Create a CES shared root file system for use with protocol nodes

CES protocols require a shared file system to store configuration and state data. This file system is called CES shared root and it is a replicated file system that is recommended to be of size between 4GB and 10GB. The following ESS command automatically creates this file system with the recommended size and mirroring.

- Run the following command from the EMS node.

```
# gssgenvdisks --create-vdisk --create-nsds --create-filesystem --crcesfs  
# mmmount cesSharedRoot -N ems1-hs
```

A file system named `cesSharedRoot` with a mount path of `/gpfs/cesSharedRoot` is created and mounted. Later in these steps, the IBM Spectrum Scale installation toolkit is pointed to this file system to use when deploying protocols.

G) Locate the IBM Spectrum Scale protocols package on a protocol node in /root

An example package name is: `Spectrum_Scale_Protocols_Data_Management-5.0.2.1-ppc64LE-Linuxinstall`

Each protocol node ships with an IBM Spectrum Scale protocols package in `/root`. The version and the license of this package matches with the ESS version that the protocol node was ordered with.

- If the package is of the desired version and license, proceed with extraction.
- If a different level is desired, proceed to IBM FixCentral to download and replace this version.

If replacing this version, the following rules apply:

- The IBM Spectrum Scale version must be 5.0.2.1 or later.
- The CPU architecture must be PPC64LE.
- The package must be a protocols package (The title and the file name must specifically contain `Protocols`).

H) Extract the IBM Spectrum Scale protocols package

- Enter the following at the command prompt: `/root/Spectrum_Scale_Protocols_Data_Management-5.0.2.1-ppc64LE-Linux-install`
- By default, the package is extracted to `/usr/lpp/mmfs/5.0.2.1/`.

I) Configure the IBM Spectrum Scale installation toolkit

1. Change directory to the installation toolkit directory:

```
cd /usr/lpp/mmfs/5.0.2.1/installer
```

View the installation toolkit usage help as follows.

```
/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale -h
```

2. Set up the installation toolkit by specifying which local IP to use for communicating with the rest of the nodes. Preferably, this IP should be the same IP used for the GPFS network. Set the toolkit to the ESS mode.

```
/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale setup -s IP_Address -st ess
```

3. Populate the installation toolkit configuration file with the current cluster configuration by pointing it to the EMS node.

```
/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale config populate -N EMSNode
```

There are limits to the config populate functionality. If it does not work, simply add the EMS node to the installation toolkit and continue.

View the current cluster configuration as follows.

```
/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale node list
/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale config gpfs
```

Note: ESS I/O nodes do not get listed in the installation toolkit node list.

4. Configure the details of the protocol nodes to be added to the cluster.

```
./spectrumscale node add ems1-hs.gpfs.net -e -a -g    ## No need to perform this step
                                                    ## if the config populate ran without error
./spectrumscale node add prt01-hs.gpfs.net -p        ## Designate a protocol node
./spectrumscale node add prt02-hs.gpfs.net -p        ## Designate a protocol node
./spectrumscale node add prt03-hs.gpfs.net -p        ## Designate a protocol node
./spectrumscale node add client01-hs.gpfs.net        ## Example of a non-protocol client node, if desired
./spectrumscale node add nsd01-hs.gpfs.net           ## Example of a non-ESS nsd node, if desired
./spectrumscale enable smb                          ## If you'd like to enable and use the SMB protocol
                                                    ## (it will be installed regardless)
./spectrumscale enable nfs                          ## If you'd like to enable and use the NFS protocol
                                                    ## (it will be installed regardless)
./spectrumscale enable object                       ## If you'd like to enable and use the Object protocol
                                                    ## (it will be installed regardless)
./spectrumscale config protocols -e CESIP1,CESIP2,CESIP3  ## Input the CES IPs set aside from step (G) of
                                                    ## this procedure. Toolkit assigns IPs listed.
./spectrumscale config protocols -f cesSharedRoot -m /gpfs/cesSharedRoot  ## FS name and mount point for
                                                    ## CES shared root, previously
                                                    ## setup during step (D)

./spectrumscale config object -e <endpoint IP or hostname>  ## This address should be an RRDNS or similar address
                                                    ## that resolves to the pool of CES IP addresses.
./spectrumscale config object -o Object_Fileset            ## This fileset will be created during deploy
./spectrumscale config object -f ObjectFS -m /gpfs/ObjectFS  ## This must point to an existing FS
                                                    ## create the FS on EMS if it doesn't already exist
./spectrumscale config object -au admin -ap -dp           ## Usernames and passwords for Object

./spectrumscale config perfmon -r on                     ## Turn on performance sensors for the protocol nodes.
                                                    ## EMS GUI picks up sensor data once protocols are deployed
./spectrumscale node list                                ## Lists out the node config (ESS IO nodes never show up here)
./spectrumscale config protocols                         ## Shows the protocol config
```

For more information, see IBM Spectrum Scale installation toolkit.

J) Installation phase of IBM Spectrum Scale installation toolkit

1. Run the installation toolkit installation precheck.

```
./spectrumscale install --precheck
```

2. Run the installation toolkit installation procedure.

```
./spectrumscale install
```

Installation toolkit performs the following actions and it can be re-run in the future to:

- Install GPFS, call home, performance monitoring, license RPMs on each node specified to the installation toolkit. The EMS and I/O server nodes are not acted upon by the installation toolkit.
- Add nodes to the cluster (protocol, client, NSD).
- Add non-ESS NSDs, if desired.
- Start GPFS and mount all file systems on the newly added nodes.
- Configure performance monitoring sensors.
- Set client or server licenses.

GPFS configuration parameters such as **pagepool**, **maxFilesToCache**, **verbsPorts** need to be set up manually. You can do this after completing the installation phase or after completing the deployment

phase. For more information about these parameters, see Chapter 4, “GPFS configuration parameters for protocol nodes,” on page 23.

K) Deployment phase of IBM Spectrum Scale installation toolkit

1. Run the installation toolkit deployment precheck.
`./spectrumscale deploy --precheck`
2. Run the installation toolkit installation procedure.
`./spectrumscale deploy`

Installation toolkit performs the following actions during deployment and it can be re-run in the future to:

- Install SMB, NFS, and object RPMs on each protocol node specified to the installation toolkit.
- Enable one or more protocols.
- Assign CES IPs. IBM Spectrum Scale code aliases these IPs to the CES base network adapter configured during step G.
- Enable authentication for file or object.
- Create additional file systems using non-ESS NSD nodes. You must run installation first to add more non-ESS NSDs.
- Add additional protocol nodes. You must run installation first to add more nodes and then run deployment for the protocol specific piece.

L) Tune the protocol nodes as desired

Protocol nodes should already be tuned with the same tuned and sysctl settings, and udev rules as the I/O server nodes. For more information, see Chapter 5, “OS tuning for RHEL 7.5 PPC64LE protocol nodes,” on page 25.

At this point, the main tuning settings to be aware of includes:

- **RDMA.** If IB RDMA is in use (check using `mm1sconfig verbsRDMA`), issue `mm1sconfig` and verify that the `verbPorts` parameter refers to the correct ports on each protocol node.
- **pagepool.** Use `mm1sconfig` to view the `pagepool` settings of the EMS and I/O server nodes. The protocol nodes do not have `pagepool` defined at this point. Define `pagepool` using `mmchconfig pagepool=XX -N cesNodes` command.
Where XX is typically 25% the system memory for protocol nodes in a cluster containing ESS. For more information, see Chapter 4, “GPFS configuration parameters for protocol nodes,” on page 23.
- **maxFilesToCache.** Use `mm1sconfig` to view the `maxFilesToCache` settings of the EMS and I/O server nodes. The protocol nodes do not have `maxFilesToCache` defined at this point. Define `maxFilesToCache` using `mmchconfig maxFilesToCache=XX -N cesNodes` command.
Where XX is typically 2M for protocol nodes in a cluster containing ESS. For more information, see Chapter 4, “GPFS configuration parameters for protocol nodes,” on page 23.

M) GUI configuration

Configure the GUI as follows.

1. Enable and start gpfs GUI:
`systemctl enable gpfs GUI.service`
`systemctl start gpfs GUI`

Note: If the GUI was started before protocol nodes were installed, then restart the GUI with `systemctl restart gpfs GUI` command.

2. To launch the ESS GUI in a browser, go to: `https://EssGuiNode` where `ESSGuiNode` is the host name 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.

During GUI configuration, you are allowed to specify rack locations of all components, including the protocol nodes. After the GUI configuration has completed, protocol nodes are a part of the GUI. Protocols (NFS, SMB) and Object panels show up in the GUI depending on which protocols were enabled during the installation toolkit deployment phase. Performance sensors for protocols are available on the **Monitoring > Statistics** page.

N) Call home setup

Now that protocol nodes are deployed, call home needs to be configured.

1. Check if call home has already been configured and if so, record the settings. Reconfiguring call home might require entering the settings again.
 - Check the hardware call home settings on the EMS node.
`gsscallhomeconf --show -E ems1`
 - Check the software call home setup.
`mmcallhome info list`
2. Set up call home from the EMS node.
`gsscallhomeconf -N ems1,gss_ppc64,ces_ppc64 --suffix=-hs -E ems1 --register all --crvpd`

For more information, see call home documentation in *Elastic Storage Server: Problem Determination Guide*.

O) Movement of quorum or management function to protocol nodes and off EMS or I/O nodes

Quorum and management functions can be resource intensive. In an ESS cluster that also has extra nodes, such as protocols, within the same cluster, it is recommended to move these functions to the protocol nodes. For information on changing node designation, see `mmchnode` command.

Chapter 3. Configuration 2: 5148-22L protocol nodes ordered standalone and added to an existing 5148 ESS (PPC64LE)

In this configuration, protocol nodes are ordered for attachment to an existing previously installed ESS. The EMS node, I/O server nodes, and protocol nodes have OS, kernel, systemd, network manager, firmware, and OFED, kept in synchronization as xCAT running on the EMS is used to manage and coordinate these levels. It is recommended to match IBM Spectrum Scale levels between the ESS and protocol nodes, but this is not mandatory.

Note: All protocol nodes in a cluster must be at the same code level.

Overall flow

- “A) Starting point and what to expect”
- “B) Protocol node OS deployment” on page 14
- “C) Decide which adapter(s) to use for the GPFS network(s) vs CES protocol network(s)” on page 16
- “D) Configure network adapters to be used for GPFS” on page 16
- “E) Configure network adapters to be used for CES protocols” on page 16
- “F) Create a CES shared root file system for use with protocol nodes” on page 17
- “G) Download the IBM Spectrum Scale protocols package (version 5.0.2.1 or later) on one of the protocol nodes ” on page 18
- “H) Extract the IBM Spectrum Scale protocols package” on page 18
- “I) Configure the IBM Spectrum Scale installation toolkit” on page 18
- “J) Installation phase of IBM Spectrum Scale installation toolkit” on page 19
- “K) Deployment phase of IBM Spectrum Scale installation toolkit” on page 20
- “L) Tune the protocol nodes as desired” on page 20
- “M) GUI configuration” on page 20
- “N) Call home setup” on page 21
- “O) Movement of quorum or management function to protocol nodes and off EMS or I/O nodes” on page 21

A) Starting point and what to expect

- An ESS is already installed and running (EMS and 2 I/O server nodes + possibly additional nodes)
- A cluster has already been created.
Run **mm1scluster** to check.
- A GUI is active on the EMS node and it has been logged into.
Run **systemctl status gpfsGUI** to check.
- The Performance Monitoring collector is configured and running on the EMS node.
Run **systemctl status pmcollector** and **mmperfmon config show** to check.
- Protocol nodes may or may not already exist on this cluster.
- The ESS is at code level 5.3.2 or later.
Run **/opt/ibm/gss/install/rhel7/ppc64le/installer/gssinstall -V** to check.
- An xCAT OS image specific for CES exists (rhel7.5-ppc64le-install-ces).
Run the following command to verify.

```
# lsdef -t osimage
```
- Newly ordered protocol nodes come with authentication prerequisites pre-installed upon them (includes sssd, ypbind, openldap-clients, krb5-workstation).

- A default deploy template exists for the protocol nodes.
Check for `/var/tmp/gssdeployces.cfg.default` on the EMS node
- New standalone protocol node orders arrives with OS, kernel, OFED, iprraid, and firmware pre-loaded.
This is verified in step B7.
- New standalone protocol node orders arrive with an IBM Spectrum Scale Protocols code package in `/root`. This is verified in steps G and H.
- New standalone protocol nodes do have not have any GPFS RPMs installed on them
- Hardware and software call home may already be configured on the existing ESS system. Call home is reconfigured after deployment of the protocol nodes.

Important: Before proceeding:

- Protocol nodes must be cabled up to the ESS switch for use with the xCAT network and the FSP network. For more information, see Figure 2 on page 33.
- Protocol nodes can be in the powered off state at this point.

B) Protocol node OS deployment

1. Verify that the protocol nodes are cabled up to the ESS xCAT network and FSP network.
2. On the EMS node, find the `/var/tmp/gssdeployces.cfg.default` file and copy to `/var/tmp/gssdeployces.cfg`.

The default CES template is pre-filled so that the only fields needing customization to match the current cluster are:

- **DEPLOYMENT_TYPE:** If these are your first protocol nodes, the type must be CES. If you are adding more protocol nodes to an ESS system that already has protocol nodes, use the type `ADD_CES`. Read the tips in the `gssdeployces.cfg` file carefully because using the incorrect deployment type and filling out the configuration file incorrectly could result in rebooting or reloading of any existing protocol nodes that might be a part of the cluster. Read all on-screen warnings.
- **EMS_HOSTNAME**
- **EMS_MGTNETINTERFACE**
- **SERVERS_UID**
- **SERVERS_PASSWD**
- **SERVERS_SERIAL:** Change the serial numbers to match each protocol node being added.
- **SERVERS_NODES:** Separate each desired protocol node name with a space.

3. Configure `/etc/hosts` on the EMS node to list the protocol nodes.

Note: This `/etc/hosts` helps during network setup in this step.

Here is an example of the IP, FQDN, and hostname configured for EMS, IO, and 3 protocol nodes.

- The EMS node is 192.168.45.20.
- The I/O server nodes are 192.168.45.21 → 192.168.45.30.
- The protocol nodes are 192.168.45.31 → X.

```
192.168.45.20 ems1.gpfs.net ems1
192.168.45.21 gssio1.gpfs.net gssio1
192.168.45.22 gssio2.gpfs.net gssio2
192.168.45.31 prt01.gpfs.net prt01
192.168.45.32 prt02.gpfs.net prt02
192.168.45.33 prt03.gpfs.net prt03
172.31.250.3 ems1-hs.gpfs.net ems1-hs
172.31.250.1 gssio1-hs.gpfs.net gssio1-hs
172.31.250.2 gssio2-hs.gpfs.net gssio2-hs
172.31.250.11 prt01-hs.gpfs.net prt01-hs
172.31.250.12 prt02-hs.gpfs.net prt02-hs
172.31.250.13 prt03-hs.gpfs.net prt03-hs
```

Note: If the `/etc/hosts` file is already set up on the EMS node, copy it to the protocol node(s) first and then modify it. Each protocol node must have the same `/etc/hosts` file.

4. Detect and add the protocol node objects to xCAT as follows.

```
/var/tmp/gssdeploy -o /var/tmp/gssdeployces.cfg
```

Proceed through all steps. Protocol nodes should be listed in xCAT afterwards.

```
# lsdef
ems1 (node)
gssio1 (node)
gssio2 (node)
prt01 (node)
prt02 (node)
prt03 (node)
```

5. Deploy the OS, kernel, systemd, netmgr, OFED, and IPR as follows.

This is a decision point with two options depending upon the requirement.

- **Option 1:** All standalone protocol nodes come preinstalled with OS, kernel, systemd, netmgr, OFED, and IPR. Now that the protocol nodes are discovered by xCAT, they can be set to boot from their hard drives, without reinstalling anything. If these preloaded levels are sufficient, then proceed with these steps. This option is quicker than option 2.
 - a. Power off all protocol nodes.

```
rpower ProtocolNodesList off
```
 - b. Set the protocol node(s) to HD boot from the EMS node.

```
rsetboot ProtocolNodesList hd
```
 - c. Power on the protocol nodes.

```
rpower ProtocolNodesList on
```
- **Option 2:** It is also possible to completely wipe and reload the protocol nodes if desired.

Remember: If you already have existing and active protocol nodes in the cluster, you must be very careful about which xCAT group is used and whether your `gssdeployces.cfg` file has a `DEPLOYMENT_TYPE` of `CES` or `ADD_CES`. Read the tips in the configuration file carefully, and read all on-screen warnings.

Run the following commands to proceed.

- a. Reload the protocol nodes with the same levels of OS, kernel, systemd, netmgr, OFED, and IPR existing on the EMS / IO nodes.

```
/var/tmp/gssdeploy -d /var/tmp/gssdeployces.cfg
```
- b. Proceed through all steps of the **gssdeploy** command. Protocol node installation progress can be watched using **rcons Node**.

Note: Unlike on I/O server nodes, this step does not install any GPFS RPMs on protocol nodes except `gpfs.gss.tools`.

6. If `ADD_CES` was used to add protocol nodes to a cluster that had existing or active protocol nodes, they would have been added to an xCAT group other than `ces_ppc64`. These protocol nodes must be moved to the `ces_ppc64` group using these steps, run from the EMS node.
 - a. Check to see which xCAT group was used for adding protocol nodes. Replace the configuration file name with the one used for **gssdeploy -o** and **-d**.

```
# cat /var/tmp/gssdeployces.cfg | grep GSS_GROUP
```
 - b. Move the added protocol nodes to the `ces_ppc64` group.

```
# chdef GroupNameUsed groups=all,ces_ppc64
```
7. Once deployed, run **gssinstallcheck** or **gssstoragequickcheck** to verify the nodes are in a healthy state.
 - a. Run **gssstoragequickcheck -G ces_ppc64** to verify network adapter types, slots, and machine type model of the protocol nodes.

- b. Run `gssinstallcheck -G ces_ppc64` to verify code and firmware levels on the protocol nodes.

C) Decide which adapter(s) to use for the GPFS network(s) vs CES protocol network(s)

It is recommended to plan for separation of the GPFS and CES networks, both by subnet and by card.

If adding protocol nodes to an existing or active protocol setup, in most cases it is recommended to match configurations of both the GPFS network and CES protocol networks to the existing protocol nodes. If planning a stretch cluster, or configurations in which not all protocol nodes see the same CES network, refer to IBM Spectrum Scale Knowledge Center.

Note: Before proceeding, protocol nodes must be cabled up to the GPFS cluster network and to the CES network.

D) Configure network adapters to be used for GPFS

Customer networking requirements are site-specific. The use of bonding to increase fault-tolerance and performance is advised but guidelines for doing this have not been provided in this document. Consult with your local network administrator before proceeding further. Before creating network bonds, carefully read *ESS networking considerations* in *Elastic Storage Server: Quick Deployment Guide*.

Make sure that the protocol nodes high speed network IPs and host names are present in `/etc/hosts` on all nodes.

Here is an example excerpt from `/etc/hosts`, showing the `-hs` suffix IPs and host names to be used for the GPFS cluster configuration.

```
172.31.250.3 ems1-hs.gpfs.net ems1-hs
172.31.250.1 gssio1-hs.gpfs.net gssio1-hs
172.31.250.2 gssio2-hs.gpfs.net gssio2-hs
172.31.250.11 prt01-hs.gpfs.net prt01-hs
172.31.250.12 prt02-hs.gpfs.net prt02-hs
172.31.250.13 prt03-hs.gpfs.net prt03-hs
```

Note:

- All nodes must be able to resolve all IPs, FQDNs, and host names, and ssh-keys must work.
- If the `/etc/hosts` file is already set up on the EMS node, copy it to the protocol node(s) first and then modify it. Each protocol node must have the same `/etc/hosts` file.

To set up bond over IB, run the following command.

```
gssgennetworks -G ces_ppc64 --create --ipoib --suffix=-hs --mtu 4092
```

In this example, MTU is set to 4092. The default MTU is 2048 (2K) and the `gssgennetworks` command supports 2048 (2K) and 4092 (4K) MTU. Consult your network administrator for the proper MTU setting.

To set up bond over Ethernet, run the following command.

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

For information on Infiniband issue with multiple fabrics, see the *Infiniband with multiple fabric* section in the *ESS networking considerations* topic of *Elastic Storage Server: Quick Deployment Guide*.

E) Configure network adapters to be used for CES protocols

Before deploying protocols, it is important to understand the customer network and protocol access requirements. CES protocols use a pool of CES IPs which float between nodes, providing redundancy in the case of node failure or degradation. The CES IPs are assigned and aliased by IBM Spectrum Scale to

an adapter on each protocol node that has a matching predefined route and subnet. It is important that each protocol node has a base network adapter or bonded group of network adapters or ports with an established IP and routing so that CES IPs can be assigned by IBM Spectrum Scale code during protocol deployment.

Note: CES IPs are never assigned using **ifcfg** or **nmcli** commands. This is handled by the IBM Spectrum Scale code.

The following must be taken into account when planning this network:

- Bandwidth requirements per protocol node (how many ports per bond, bonding mode, and adapter speed)
- Redundancy of each protocol node, if needed. This determines the bonding mode used.
- Authentication domain and DNS. This determines the subnet(s) required for each protocol node.
- Are VLAN tags needed?
- Set aside 1 IP per protocol node, per desired CES subnet. You will be using these when configuring the CES base adapter(s) on each protocol node. These IPs must be setup for forward and reverse DNS lookup.
- Set aside a pool of CES IPs for later use. These IPs must be in DNS and be setup for both forward and reverse DNS lookup. You will not be assigning these IPs to network adapters on protocol nodes.
- Prepare to configure each protocol node to point to the authentication domain or DNS. You need to do this manually using **ifcfg** or **nmcli** commands and by verifying `/etc/resolv.conf` after the settings have taken effect. When deployed from an EMS node, each protocol node might already have a default domain of `gpfs.net` present in `/etc/resolv.conf` and the `ifcfg` files. This default domain can be removed so that it does not interfere with the authentication setup and DNS for protocols.

Proceed with either configuring the CES protocol adapters manually using **ifcfg** or **nmcli** commands or by using **gssgennetworks**. The **gssgennetworks** command cannot be used if your CES protocol network requires VLAN tags nor does it set up additional domain or DNS servers.

When the network is configured on each protocol nodes, verify it using the **mmnetverify** command with these actions:

- Make sure all protocol nodes can ping each other's base CES network by IP, host name, and FQDN.
- Make sure all protocol nodes can ping the authentication server by IP, host name, and FQDN.
- Make sure the authentication server can ping each protocol node base CES network by IP, host name, and FQDN
- Spot check the desired NFS, SMB, or OBJ clients, external to the GPFS cluster and verify that they can ping each protocol node base CES network by IP, hostname, and FQDN
- Even though the CES IP pool is not yet set up, because protocols are not deployed, double check that each protocol node can resolve each CES IP or host name using **nslookup**.

For an example showing how CES IP aliasing relies upon an established base adapter with proper subnets or routing, see CES IP aliasing to network adapters on protocol nodes.

For an example of CES-IP configuration that can be performed after deployment of protocols, see Configuring CES protocol service IP addresses.

F) Create a CES shared root file system for use with protocol nodes

If you already have an existing or active protocol setup and `cesSharedRoot` should already exist. Skip this step.

CES protocols require a shared file system to store configuration and state data. This file system is called CES shared root and it is a replicated file system that is recommended to be of size between 4GB and 10GB. The following ESS command automatically creates this file system with the recommended size and mirroring.

- Run the following command from the EMS node.

```
# gssgenvdisks --create-vdisk --create-nsds --create-filesystem --crcefs  
# mmmount cesSharedRoot -N ems1-hs
```

A file system named cesSharedRoot with a mount path of /gpfs/cesSharedRoot is created and mounted. Later in these steps, the IBM Spectrum Scale installation toolkit is pointed to this file system to use when deploying protocols.

G) Download the IBM Spectrum Scale protocols package (version 5.0.2.1 or later) on one of the protocol nodes

An example package name is: Spectrum_Scale_Protocols_Data_Management-5.0.2.1-ppc64LE-Linuxinstall

Each protocol node ships with an IBM Spectrum Scale protocols package in /root. The version and the license of this package matches with the ESS version that the protocol node was ordered with.

- If the package is of the desired version and license, proceed with extraction.
- If a different level is desired, proceed to IBM FixCentral to download and replace this version.

If replacing this version, the following rules apply:

- The IBM Spectrum Scale version must be 5.0.2.1 or later.
- The CPU architecture must be PPC64LE.
- The package must be a protocols package (The title and the file name must specifically contain Protocols).

Note: If Option 2 was specified in step B5 and the protocol node was reloaded, then there will be no Spectrum Scale protocols package in /root. It will need to be downloaded.

H) Extract the IBM Spectrum Scale protocols package

- Enter the following at the command prompt: /root/Spectrum_Scale_Protocols_Data_Management-5.0.2.1-ppc64LE-Linux-install
- By default, the package is extracted to /usr/lpp/mmfs/5.0.2.1/.

I) Configure the IBM Spectrum Scale installation toolkit

1. Change directory to the installation toolkit directory:

```
cd /usr/lpp/mmfs/5.0.2.1/installer
```

View the installation toolkit usage help as follows.

```
/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale -h
```

2. Set up the installation toolkit by specifying which local IP to use for communicating with the rest of the nodes. Preferably, this IP must be the same IP used for the GPFS network. Set the toolkit to the ESS mode.

```
/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale setup -s IP_Address -st ess
```

3. Populate the installation toolkit configuration file with the current cluster configuration by pointing it to the EMS node.

```
/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale config populate -N EMSNode
```

There are limits to the config populate functionality. If it does not work, simply add the EMS node to the installation toolkit and continue.

View the current cluster configuration as follows.

```
/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale node list
/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale config gpfs
```

Note: ESS I/O nodes do not get listed in the installation toolkit node list.

4. Configure the details of the protocol nodes to be added to the cluster. If adding protocol nodes to an existing or active protocol setup, make sure to add all existing protocol nodes and configuration details. Note that a successful config populate operation from step 3 would have already performed this action.

```
./spectrumscale node add ems1-hs.gpfs.net -e -a -g    ## No need to perform this step
                                                    ## if the config populate ran without error
./spectrumscale node add prt01-hs.gpfs.net -p        ## Add a protocol node
./spectrumscale node add prt02-hs.gpfs.net -p        ## Add a protocol node
./spectrumscale node add client01-hs.gpfs.net        ## Example of a non-protocol client node, if desired
./spectrumscale node add nsd01-hs.gpfs.net           ## Example of a non-ESS nsd node, if desired
./spectrumscale enable smb                          ## If you'd like to enable and use the SMB protocol
                                                    ## (it will be installed regardless)
./spectrumscale enable nfs                          ## If you'd like to enable and use the NFS protocol
                                                    ## (it will be installed regardless)
./spectrumscale enable object                       ## If you'd like to enable and use the Object protocol
                                                    ## (it will be installed regardless)
./spectrumscale config protocols -e CESIP1,CESIP2,CESIP3  ## Input the CES IPs set aside from step (G) of
                                                    ## this procedure. Toolkit assigns IPs listed.
./spectrumscale config protocols -f cesSharedRoot -m /gpfs/cesSharedRoot  ## FS name and mount point for
                                                    ## CES shared root, previously
                                                    ## setup during step (D)
./spectrumscale config object -e <endpoint IP or hostname>  ## This address should be an RRDNS or similar address
                                                    ## that resolves to the pool of CES IP addresses.
./spectrumscale config object -o Object_Fileset           ## This fileset will be created during deploy
./spectrumscale config object -f ObjectFS -m /gpfs/ObjectFS  ## This must point to an existing FS
                                                    ## create the FS on EMS if it doesn't already exist
./spectrumscale config object -au admin -ap -dp          ## Usernames and passwords for Object
./spectrumscale config perfmon -r on                   ## Turn on performance sensors for the protocol nodes.
                                                    ## EMS GUI picks up sensor data once protocols are deployed
./spectrumscale node list                             ## Lists out the node config (ESS IO nodes never show up here)
./spectrumscale config protocols                      ## Shows the protocol config
```

For more information, see IBM Spectrum Scale installation toolkit.

J) Installation phase of IBM Spectrum Scale installation toolkit

1. Run the installation toolkit installation precheck.

```
./spectrumscale install --precheck
```
2. Run the installation toolkit installation procedure.

```
./spectrumscale install
```

Installation toolkit performs the following actions and it can be re-run in the future to:

- Install GPFS, call home, performance monitoring, license RPMs on each node specified to the installation toolkit. The EMS and I/O server nodes are not acted upon by the installation toolkit.
- Add nodes to the cluster (protocol, client, NSD).
- Add non-ESS NSDs, if desired.
- Start GPFS and mount all file systems on the newly added nodes.
- Configure performance monitoring sensors.
- Set client or server licenses.

GPFS configuration parameters such as **pagepool**, **maxFilesToCache**, **verbsPorts** need to be set up manually. You can do this after completing the installation phase or after completing the deployment phase. For more information about these parameters, see Chapter 4, “GPFS configuration parameters for protocol nodes,” on page 23.

K) Deployment phase of IBM Spectrum Scale installation toolkit

1. Run the installation toolkit deployment precheck.
`./spectrumscale deploy --precheck`
2. Run the installation toolkit installation procedure.
`./spectrumscale deploy`

Installation toolkit performs the following actions during deployment and it can be re-run in the future to:

- Install SMB, NFS, and object RPMs on each protocol node specified to the installation toolkit.
- Enable one or more protocols.
- Assign CES IPs. IBM Spectrum Scale code aliases these IPs to the CES base network adapter configured during step E.
- Enable authentication for file or object.
- Create additional file systems using non-ESS NSD nodes. You must run installation first to add more non-ESS NSDs.
- Add additional protocol nodes. You must run installation first to add more nodes and then run deployment for the protocol specific piece.

L) Tune the protocol nodes as desired

Protocol nodes should already be tuned with the same tuned and sysctl settings, and udev rules as the I/O server nodes. For more information, see Chapter 5, “OS tuning for RHEL 7.5 PPC64LE protocol nodes,” on page 25.

At this point, the main tuning settings to be aware of includes:

- **RDMA.** If IB RDMA is in use (check using `mmisconfig verbsRDMA`), issue `mmisconfig` and verify that the `verbPorts` parameter refers to the correct ports on each protocol node.
- **pagepool.** Use `mmisconfig` for the `pagepool` settings of the EMS and I/O server nodes. The protocol nodes do not have `pagepool` defined at this point. Define `pagepool` using `mmchconfig pagepool=XX -N cesNodes` command.
Where XX is typically 25% to 50% of the system memory. For more information, see Chapter 4, “GPFS configuration parameters for protocol nodes,” on page 23.
- **maxFilesToCache.** Use `mmisconfig` to view the `maxFilesToCache` settings of the EMS and I/O server nodes. The protocol nodes do not have `maxFilesToCache` defined at this point. Define `maxFilesToCache` using `mmchconfig maxFilesToCache=XX -N cesNodes` command.

Where XX is typically 2M for protocol nodes in a cluster containing ESS. For more information, see Chapter 4, “GPFS configuration parameters for protocol nodes,” on page 23.

M) GUI configuration

1. Open the existing EMS GUI in a web browser using the URL `https://EssGuiNode` where *ESSGuiNode* is the host name or IP address of the management server node. If the GUI is not yet set up, perform step 8 of the *Install the ESS GUI* procedure in the *Installing Elastic Storage Server* topic of *Elastic Storage Server: Quick Deployment Guide*.
2. Software monitoring of protocol nodes occurs automatically by the GUI and all newly added protocol nodes should now exist in the HOME page **Nodes** section. Hardware monitoring of protocol nodes must be configured within the GUI panels as follows.
 - a. Click **Monitoring > Hardware > Edit Rack Components**.
The Edit Rack Components wizard is displayed.
 - b. On the Welcome screen, select **Yes, discover new servers and enclosures first** and then click **Next**.
This might take a few minutes. After the detection is complete, click **OK**.
 - c. Click **Next** through the screens allowing edits of various rack components.

The Other Servers screen shows the protocol nodes.

- d. On the Other Servers screen, select a rack and location for each protocol node and then click **Next**.
- e. Click **Finish** on the Summary screen.

The protocol nodes are now fully configured for hardware monitoring in the GUI.

N) Call home setup

Now that protocol nodes are deployed, call home needs to be configured.

1. Check if call home has already been configured and if so, record the settings. Reconfiguring call home might require entering the settings again.
 - Check the hardware call home settings on the EMS node.
`gsscallhomeconf --show -E ems1`
 - Check the software call home setup.
`mmcallhome info list`
2. Set up call home from the EMS node.
`gsscallhomeconf -N ems1,gss_ppc64,ces_ppc64 --suffix=-hs -E ems1 --register all --crvpd`

For more information, see call home documentation in *Elastic Storage Server: Problem Determination Guide*.

O) Movement of quorum or management function to protocol nodes and off EMS or I/O nodes

Quorum and management functions can be resource intensive. In an ESS cluster that also has extra nodes, such as protocols, within the same cluster, it is recommended to move these functions to the protocol nodes. For information on changing node designation, see `mmchnode` command.

Chapter 4. GPFS configuration parameters for protocol nodes

For protocol nodes in a cluster containing ESS, the following values are recommended for a few GPFS configuration parameters. For GPFS configuration parameters that are not listed, default values are used.

Note: Each cluster, network environment, and planned workload is different and therefore these values might need to be altered to best fit the respective environment.

Use the **mmchconfig** command to change these parameters after a cluster is created. For example, to change the **maxFilesToCache** parameter to 2M on all protocol nodes in the cluster, issue the following command:

```
mmchconfig maxFilesToCache=2M -N cesNodes
```

cesNodes is the node class that includes all protocol nodes in the cluster.

For changing some of these parameters, GPFS needs to be stopped and restarted before the new values can take effect. For more information, see [Changing the GPFS cluster configuration data](#).

For protocol nodes in a cluster containing ESS, these values are recommended for cluster parameters.

Cluster configuration parameter	Recommended value
autoload	yes
maxblocksize	16M
numaMemoryInterleave	yes
enforceFilesetQuotaOnRoot	yes
ignorePrefetchLUNCount	yes
workerThreads	1024
maxFilesToCache	2M
pagepool	32G Note: Assuming that modern servers such as 5148-22L have greater than or equal to 128G of memory, set pagepool to at least 25% of that value.
nsdClientCksumTypeLocal	ck64
nsdClientCksumTypeRemote	ck64
maxMBpS	10000 Note: This value can be adjusted to match the bandwidth between the protocol node and ESS.

For protocol nodes with Infiniband connectivity in a cluster containing ESS, these values are recommended.

Configuration parameter for protocol nodes with Infiniband	Recommended value
verbsRdma	enable
verbsRdmaSend	yes
verbsPorts	<i>Active_Infiniband_ports</i>

File systems being accessed by protocol nodes are recommended to have the following parameter. This parameter can be set at file system creation using GUI, **mmcrfs**, or **gssgenvdisks**. The parameter can be changed for existing file systems using **mmchfs**.

File system related configuration parameter	Recommended value
ACLSemantics	nfs4 Note: This is mandatory for SMB access to a file system.

Related reference:

Chapter 5, “OS tuning for RHEL 7.5 PPC64LE protocol nodes,” on page 25

The following operating system tuning is applicable for Red Hat Enterprise Linux 7.5 PPC64LE protocol nodes.

Chapter 5. OS tuning for RHEL 7.5 PPC64LE protocol nodes

The following operating system tuning is applicable for Red Hat Enterprise Linux 7.5 PPC64LE protocol nodes.

These tuning settings are automatically applied to 5148-22L protocol nodes on deployment. For non-5148-22L configurations, it is up to the customers to manually apply these tuning settings to the OS. These recommended tuning settings are specifically tested for 5148-22L PPC64LE configurations. Non-5148-22L environments might need different tuning settings.

tuned.conf settings for RHEL 7.5 PPC64LE protocol nodes

Note: The tuned.conf file contains the sysctl settings as well.

```
[main]
include=throughput-performance

[cpu]
force_latency=1
governor=performance
energy_perf_bias=performance
min_perf_pct=100

[script]
script=script.sh

[sysctl]
# ktune sysctl settings for rhel servers, maximizing i/o throughput
#
# Minimal preemption granularity for CPU-bound tasks:
# (default: 1 msec# (1 + ilog(ncpus)), units: nanoseconds)
kernel.sched_min_granularity_ns=100000000

# If a workload mostly uses anonymous memory and it hits this limit, the entire
# working set is buffered for I/O, and any more write buffering would require
# swapping, so it's time to throttle writes until I/O can catch up. Workloads
# that mostly use file mappings may be able to use even higher values.
#
# The generator of dirty data starts writeback at this percentage (system default
# is 20%)
vm.dirty_ratio=10

# Start background writeback (via writeback threads) at this percentage (system
# default is 10%)
vm.dirty_background_ratio=3

# The swappiness parameter controls the tendency of the kernel to move
# processes out of physical memory and onto the swap disk.
# 0 tells the kernel to avoid swapping processes out of physical memory
# for as long as possible
# 100 tells the kernel to aggressively swap processes out of physical memory
# and move them to swap cache
vm.swappiness=0

# The total time the scheduler will consider a migrated process
# "cache hot" and thus less likely to be re-migrated
# (system default is 500000, i.e. 0.5 ms)
kernel.sched_migration_cost_ns=5000000

#number of max socket connections , should be >size of cluster
net.core.somaxconn = 10000
```

```

# Sets the maximum number of packets allowed to queue when
# a particular interface receives packets faster than the kernel can process them
net.core.netdev_max_backlog = 250000

# allow more ports to be used
net.ipv4.ip_local_port_range = 2000 65535
net.ipv4.tcp_rfc1337 = 1
net.ipv4.tcp_max_tw_buckets = 1440000

net.ipv4.tcp_mtu_probing=1
net.ipv4.tcp_window_scaling=1
net.ipv4.tcp_adv_win_scale=1
net.ipv4.tcp_low_latency=1
net.ipv4.tcp_max_syn_backlog=4096
net.ipv4.tcp_fin_timeout=10

net.core.rmem_max=4194304
net.core.wmem_max=4194304
net.core.rmem_default=4194304
net.core.wmem_default=4194304
net.core.optmem_max=4194304
net.ipv4.tcp_rmem=4096 87380 4194304
net.ipv4.tcp_wmem=4096 65536 4194304

# make sure there is enough free space to prevent OOM until no real free memory left
vm.min_free_kbytes = 512000

# allow magic keys for debugging on console
kernel.sysrq = 1
kernel.shmmax = 137438953472

# don't panic at oops so we can finish writing dumps and traces.
kernel.panic_on_oops = 0

# Disable IPv6
net.ipv6.conf.all.disable_ipv6=1
net.ipv6.conf.default.disable_ipv6=1

```

udev rules for RHEL 7.5 PPC64LE protocol nodes

```

ACTION=="add|change", SUBSYSTEM=="block", \
KERNEL=="sd*[^0-9]", PROGRAM="/usr/bin/lsblk -rno FSTYPE,MOUNTPOINT,NAME /dev/%k", \
RESULT=="* /boot ", ATTR{queue/nr_requests}="128", ATTR{device/queue_depth}="64"
ACTION=="add|change", SUBSYSTEM=="block", \
KERNEL=="sd*[^0-9]", PROGRAM="/usr/bin/lsblk -rno FSTYPE,MOUNTPOINT,NAME /dev/%k", \
RESULT!=" /boot *", ATTR{queue/scheduler}="deadline", \
ATTR{queue/nr_requests}="256", ATTR{device/queue_depth}="31", \
ATTR{queue/max_sectors_kb}="8192", ATTR{queue/read_ahead_kb}="0", ATTR{queue/rq_affinity}="2"

```

Related reference:

Chapter 4, “GPFS configuration parameters for protocol nodes,” on page 23

For protocol nodes in a cluster containing ESS, the following values are recommended for a few GPFS configuration parameters. For GPFS configuration parameters that are not listed, default values are used.

Chapter 6. Upgrading a cluster containing ESS and protocol nodes

The procedure for upgrading a cluster containing ESS and protocol nodes comprises several phases. Although the protocol node upgrade procedure is detailed here, the same procedure can be tweaked and used for client and NSD nodes as well.

1. “Planning upgrade in a cluster containing ESS and protocol nodes”
2. “Upgrading ESS Red Hat software and system firmware” on page 29
3. “Performing prechecks to prepare for IBM Spectrum Scale upgrade” on page 30
4. “Upgrading IBM Spectrum Scale on protocol nodes by using the installation toolkit” on page 32

Planning upgrade in a cluster containing ESS and protocol nodes

Before scheduling an upgrade of a cluster containing ESS and protocol nodes, planning discussions must take place to know the current cluster configuration and to understand which functions might face an outage.

The planning phase comprises the following steps.

1. Note all products and functions currently installed in the cluster that is being upgraded.

Important: Any function that is actively accessing files on a specific node that is undergoing upgrade might prevent a file system from properly unmounting and thus prevent IBM Spectrum Scale from unloading kernel modules, which is required for RPM updates. If this occurs, the upgrade can be resumed after doing the following steps:

- a. Reboot the node that could not unload kernel modules properly.
- b. Verify that SMB on this node is not at a level differing from SMB on other nodes within the cluster. Resolve this issue by either upgrading or downgrading the SMB version on this node to match other nodes within the cluster.
- c. Bring the node online and resume the upgrade.

The following list contains the functions and products that must be considered for upgrade depending on your environment. Upgrade considerations for some of these functions and products are also listed.

- SMB: Requires quiescing all I/O for the duration of the upgrade. Due to the SMB clustering functionality, differing SMB levels cannot co-exist within a cluster at the same time. This requires a full outage of SMB during the upgrade.
- NFS: Recommended to quiesce all I/O for the duration of the upgrade. NFS experiences I/O pauses, and depending upon the client, mounts and disconnects during the upgrade.
- Object: Recommended to quiesce all I/O for the duration of the upgrade. Object service will be down or interrupted at multiple times during the upgrade process. Clients might experience errors or they might be unable to connect during this time. They should retry as appropriate.
- CES Groups: Follow SMB, NFS, and Object advice for quiescing I/O for the duration of the upgrade.
- TCT: Requires the cloud gateway service to be stopped on all TCT nodes prior to the upgrade by using the following command: `mmcloudgateway service stop -N Node | NodeClass`.
- AFM: Active AFM file transfers might hold open the file system.
- ILM: Recommended to quiesce or pause all ILM policies that might be set to trigger during an upgrade window.

- Restripe: Recommended to stop any **mmrestripefs** process prior to an upgrade. For a list of commands that might perform file system maintenance tasks, see the *Prepare the system for upgrade* section in the *Upgrading Elastic Storage Server* topic of *Elastic Storage Server: Quick Deployment Guide*.
- Snapshot creation or deletion: Recommended to stop or pause any policies that might create or delete snapshots during an upgrade window. For a list of commands that might perform file system maintenance tasks, see the *Prepare the system for upgrade* section in the *Upgrading Elastic Storage Server* topic of *Elastic Storage Server: Quick Deployment Guide*.
- IBM Spectrum Protect™ : All **mmbackup** operations must be quiesced prior to the upgrade.
- DMAPI flag for file systems: If the **cesSharedRoot** file system is DMAPI enabled, all HSM services must be stopped prior to upgrade by using **dsmmigfs stop** and **systemctl stop hsm**.

Furthermore, the installation toolkit upgrade process might fail while attempting to remount **cesSharedRoot**. This is because HSM processes must be restarted for the file system to mount. Perform this manually if the installation toolkit fails:

- Start the HSM service: **systemctl start hsm.service**
 - Start the HSM daemons: **dsmmigfs start**
 - Mount the **cesSharedRoot** file system: **mmmount cesSharedRoot -N cesNodes**
 - Restart the installation toolkit upgrade process.
- IBM Spectrum Archive™ EE: If IBM Spectrum Archive is enabled, do the following steps to upgrade.

Note: For latest information about IBM Spectrum Archive EE commands, refer to *IBM Spectrum Archive EE documentation on IBM Knowledge Center*.

- Stop IBM Spectrum Archive (LTFS) by issuing the following command on all IBM Spectrum Archive EE nodes.
`ltfsee stop`
- Unmount the media by issuing the following command on all IBM Spectrum Archive EE nodes.
`umount /ltfs`
- Deactivate failover operations by issuing the following command on all IBM Spectrum Archive EE nodes.
`dsmmigfs disablefailover`
- Stop the HSM daemons by issuing the following command on all IBM Spectrum Archive EE nodes.
`dsmmigfs stop`
- Stop the HSM service by issuing the following command on all IBM Spectrum Archive EE nodes.
`systemctl stop hsm.service`
- Upgrade using the installation toolkit. For more information, see “Upgrading IBM Spectrum Scale on protocol nodes by using the installation toolkit” on page 32.
- Upgrade IBM Spectrum Archive EE, if needed.
- Start the HSM service by issuing the following command on all IBM Spectrum Archive EE nodes.
`systemctl start hsm.service`
- Start the HSM daemons by issuing the following command on all IBM Spectrum Archive EE nodes.
`dsmmigfs start`
- Activate failover operations by issuing the following command on all IBM Spectrum Archive EE nodes.
`dsmmigfs enablefailover`
- Mount the media by issuing the following command on all IBM Spectrum Archive EE nodes.

```
ltfs -o devname=DEVICE /ltfs
```

1. Start IBM Spectrum Archive (LTFS) by issuing the following command on all IBM Spectrum Archive EE nodes.

```
ltfsee start
```

For information on how to manage a DMAPI enabled cesSharedRoot file system, see the entry *DMAPI flag for file systems* in this list.

- Encryption
 - cNFS
 - GUI - How many and which nodes?
 - Performance monitoring collectors - How many and where are they located?
 - Performance monitoring sensors - Are they installed on all nodes?
 - Which nodes run more than one of these functions?
2. Understand the source version and the number of hops needed to move to the target code version across all nodes and functions.
 - For information on IBM Spectrum Scale upgrade paths, see Supported online upgrade paths.
 - For information on ESS upgrade paths, see the *Supported upgrade paths* section in the *Installation and upgrade related information and checklists* topic of *Elastic Storage Server: Quick Deployment Guide*.
 3. Understand if the IBM Spectrum Scale installation toolkit can be used on the protocol nodes and also understand how the installation toolkit performs the upgrade. For information about installation toolkit limitations, see Limitations of the installation toolkit.

Note: This instruction set assumes that the installation toolkit is being used for protocol nodes.

4. Set expectations for functional currency and outages. For more information, see IBM Spectrum Scale FAQ.
5. Obtain the necessary packages.
6. Decide the upgrade sequence.
7. Decide whether operating system, driver, or firmware updates are needed on protocol nodes. This includes OFED, Power® firmware, x86 firmware. When making this decision, be aware that tools normally used within ESS might not be available to assist with automating these efforts outside of the ESS nodes.

Upgrading ESS Red Hat software and system firmware

Upgrading 5148-22L protocol nodes consists of two major tasks. The first task, which is described in this topic, is to upgrade the OS, OFED, kernel errata, systemd, and network manager on the 5148-22L protocol nodes.

This involves picking a protocol node to start with, shutting down GPFS on the node, upgrading, bringing it back online, and proceeding to the next protocol node that needs to be upgraded. This procedure allows concurrency and keeps the cluster and protocol services available during the upgrade. Once all 5148-22L protocol nodes have had their OS, OFED, kernel errata, systemd, and network manager software upgraded, they are back online and ready for the 2nd major task, described in “Performing prechecks to prepare for IBM Spectrum Scale upgrade” on page 30 and “Upgrading IBM Spectrum Scale on protocol nodes by using the installation toolkit” on page 32.

While upgrading a cluster containing ESS and protocol nodes, upgrade OS, OFED, kernel errata, systemd, and network manager on protocol nodes as follows.

1. Perform precheck to detect any errors and address them before proceeding further.

```
/opt/ibm/gss/tools/samples/gssprecheck -N Current_Protocol_Node --upgrade --file /var/tmp/gssdeployces.cfg
```
2. Shut down GPFS on the current protocol node that is ready for Red Hat software and system firmware upgrade.

```
mmshutdown -N CurrentProtocolServer-hs
mmgetstate -N CurrentProtocolServer-hs
```

Note: If the shutdown indicates that file systems were not unmounted, repeat the shutdown process on this node. If shutdown errors still occur, it might be necessary to reboot the node prior to continuing.

3. Issue the following command from the management server node to update the current protocol node.

```
updatenode Current_Protocol_Node -P gss_updatenode
```

Use **systemctl reboot** to reboot the current protocol node and run **updatenode** again as follows.

```
updatenode Current_Protocol_Node -P gss_updatenode
```

This additional step rebuilds OFED for the new kernel.

Note: You can use the **-V** option with the **updatenode** command for a more verbose output on the screen for a better understanding of failures, if any.

4. Update OFED on the current protocol node.

```
updatenode Current_Protocol_Node -P gss_ofed
```

5. Update IP RAID Adapter on the current protocol node.

```
updatenode Current_Protocol_Node -V -P gss_cesipraid
```

6. Use **systemctl reboot** to reboot the current protocol node.

7. If you are planning to upgrade protocol packages later, run the **mmbuildgpl** command on the protocol node before starting up GPFS on the protocol node. Or, after the upgrade of protocol packages, make sure that you use the IBM Spectrum Scale installation toolkit to upgrade core GPFS.

Note: Use the **mmbuildgpl** command to create the needed kernel extension for your kernel or copy the binaries from another node with the identical environment.

8. Update the system firmware. For more information, see *Updating the system firmware* in *Elastic Storage Server: Quick Deployment Guide*.

9. Start GPFS on the current protocol node.

A successful startup plus the remaining steps in this procedure help in verifying a successful upgrade of OS, OFED, kernel errata, systemd, and network manager prior to moving on to the IBM Spectrum Scale code upgrade procedures documented in this guide. The installation toolkit instructions used in the remaining topics require that all protocol nodes be in the active state.

```
mmstartup
```

10. Verify that GPFS is in the active state on the current protocol node.

```
mmgetstate
```

11. Ensure that the protocol node is fully updated and active.

```
gssinstallcheck -N Current_Protocol_Node
```

Performing prechecks to prepare for IBM Spectrum Scale upgrade

The precheck phase assists with the planning phase and it can be done on a cluster without any harm. It might be useful to run through the precheck steps the day before an upgrade is scheduled, or earlier, to guard against any unexpected situations that might lead to an upgrade failure.

1. Identify the protocol nodes to be upgraded.
2. Verify and, if needed, configure passwordless SSH on all protocol nodes and ESS nodes (EMS and I/O server nodes).

Important: Verify the following combinations.

- All nodes must be able to SSH to themselves and all other nodes using the IP
- All nodes must be able to SSH to themselves and all other nodes using the host name

- All nodes must be able to SSH to themselves and all other nodes using the FQDN
You can use the **mmnetverify** command to check for passwordless SSH access among nodes.
- 3. Verify that the contents of the `/etc/hosts` file on each protocol nodes are in the following format:
`<IP> <FQDN> <alias>`
- 4. Verify that firewall ports required for the necessary functions are open. For more information, see *Securing the IBM Spectrum Scale system using firewall*.
- 5. Download the new IBM Spectrum Scale self-extracting package using the sub-steps and then place it on one of the existing protocol nodes that you plan to designate as the installer node.
 - a. Go to the IBM Spectrum Scale page on Fix Central, select the new **spectrumscale** package, matching the licensed edition of the ESS (use **mm1slicense** or compare to the ESS code package name itself), and then click **Continue**.
 - b. Choose the download option **Download using Download Director** to download the new **spectrumscale** package and place it in the wanted location on the install node.
- 6. Extract the new IBM Spectrum Scale self-extracting package by using the package name (for example, `/tmp/Spectrum_Scale_Data_Management-5.0.2.1_ppc64LE-Linux_install`).
This creates a new directory structure (`/usr/lpp/mmfs/5.0.2.1/`).
- 7. Specify the installer node IP address and the setup type `ess`.
`/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale setup -s IPAddress -st ess`
- 8. Ensure that the configuration in the cluster definition file mirrors the current cluster configuration by using one of the following methods (a) or (b).
 - a. Populate the cluster definition file with the current cluster state with the EMS node specified.
`/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale config populate -N EMSNode`
This is the preferred and the easiest method. If this single step is successful, proceed to the upgrade precheck step.
 - b. If the **config populate** operation does not work in your environment, use one of the following options.
 - 1) If the installation toolkit was previously used, copy the old cluster definition file to the new code location.
For example:
`cp -p /usr/lpp/mmfs/4.2.3.10/installer/configuration/clusterdefinition.txt \`
`/usr/lpp/mmfs/5.0.2.1/installer/configuration`
 - 2) If you do not have the old cluster definition file or if the installation toolkit was not used previously in the cluster, manually enter the cluster information in the cluster definition file by referring to the following examples.
 - Specify the EMS node and also designate it as an admin node and a GUI node.
`/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale node add -e -a -g EMSNode`
 - Specify any protocol or non-protocol nodes on which you plan to use the installation toolkit.
For example, specify protocol nodes as follows:
`/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale node add node1.example.com -p`
`/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale node add node2.example.com -p`
`/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale node add node3.example.com -p`
For example, specify client node as follows:
`/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale node add node4.example.com`
- Enable SMB, NFS, and object, if required.
`/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale enable smb`
`/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale enable nfs`
`/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale enable object`
- Specify the existing CES IPs that you can obtain using **mmces address list**.
`/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale config protocols -e CESIP1,CESIP2,CESIP3`

- Specify the existing CES shared root file system.
`/usr/lpp/mmfs/5.0.2.1/installer/spectrumscale config protocols -f cesSharedRoot -m /gpfs/cesSharedRoot`
- Configure object-related settings.
`./spectrumscale config object -e EndpointIP or HostName`
`./spectrumscale config object -o Object_Fileset`
`./spectrumscale config object -f ObjectFS -m /gpfs/ObjectFS`
`./spectrumscale config object -au admin -ap -dp`
- Enable performance monitoring reconfiguration to ensure that sensors are also upgraded during the upgrade.
`./spectrumscale config perfmon -r on`

When specifying the cluster configuration information, the following considerations apply.

- If possible, always set up the installation toolkit in the ess mode with the `-st ess` flag and add the EMS node as the only admin node.
- Do not specify the I/O server nodes from the ESS system.
- It is not required to specify NSD or file system information.

The installation toolkit can be used for tasks other than upgrade, such as adding new protocols and protocol nodes. If you are planning to do this in the future, you need to expand the preceding example to input configuration details necessary for each future action. For more information, see Protocols Quick Overview Guide.

9. Run the installation toolkit upgrade precheck.

```
./spectrumscale upgrade precheck
```

A successful precheck implies that you are ready for using the installation toolkit to perform the upgrade.

10. Make all possible attempts to quiesce all I/O for Object, SMB, and NFS prior to the upgrade. For information on upgrade considerations for these and other functions, see “Planning upgrade in a cluster containing ESS and protocol nodes” on page 27.

Upgrading IBM Spectrum Scale on protocol nodes by using the installation toolkit

Use these steps to upgrade protocol nodes by using the installation toolkit.

This phase of the upgrading a cluster containing ESS and protocol nodes procedure is dependent on the successful completion of the planning and precheck phases.

1. Run the installation toolkit upgrade precheck.

Note: If a precheck has not occurred on the same day as the upgrade, make sure to run it again before proceeding with the full upgrade.

```
./spectrumscale upgrade precheck
```

If the precheck is successful, proceed to the next step.

Attention: Make all possible attempts to quiesce all I/O for Object, SMB, and NFS prior to the upgrade. For information on upgrade considerations for these and other functions, see “Planning upgrade in a cluster containing ESS and protocol nodes” on page 27.

2. Run the installation toolkit upgrade procedure.

```
./spectrumscale upgrade run
```

When this procedure is done, components including base GPFS and protocols will have been upgraded on all protocol nodes that were specified to the installation toolkit. This step does not need to be repeated on each node unless only a subset of nodes were specified to the installation toolkit.

Appendix. 5148-22L protocol node diagrams

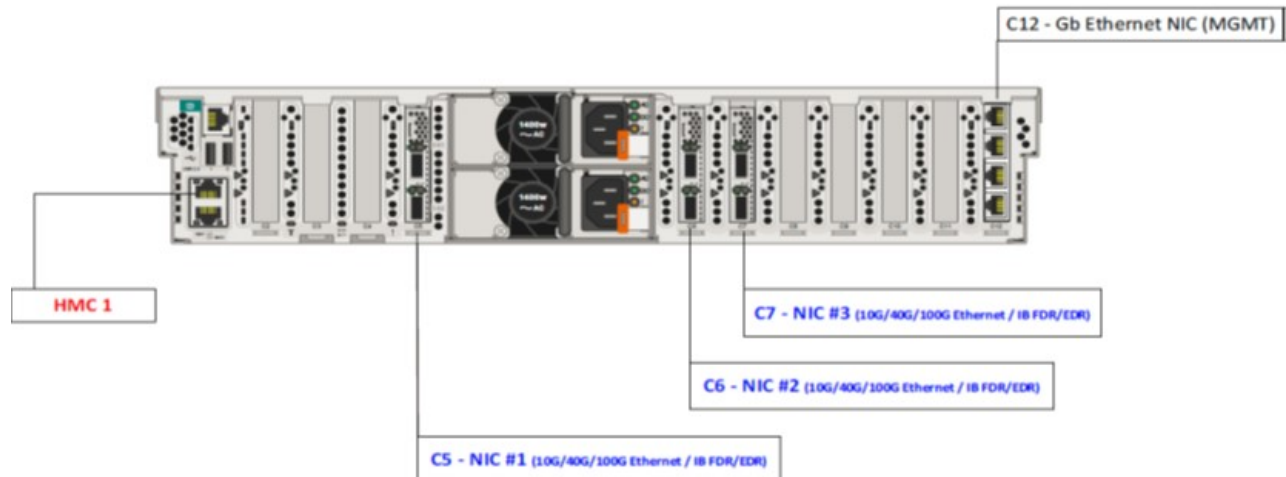


Figure 1. Card placement diagram

Note:

- In this release, there must be one 10G Ethernet card in slot C6 or C7.
- The system must have one network adapter option that matches the high speed network option of the ESS configuration.

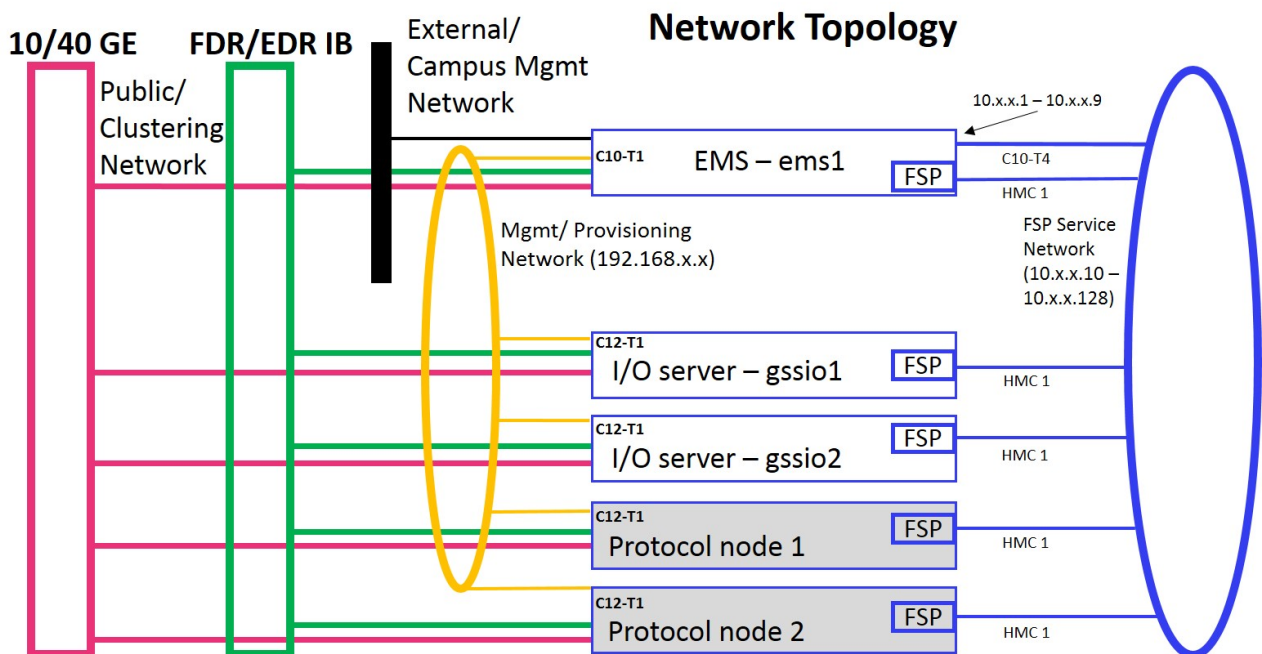


Figure 2. ESS (PPC64LE) with protocol nodes cabling diagram

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)):

<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 that 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.

ESS Management Server (EMS)

An xCAT server is required to discover the I/O server nodes (working with the HMC), provision the operating system (OS) on the I/O server nodes, and deploy the ESS software on the management node and I/O server nodes. One management server is required for each ESS system composed of one or more building blocks.

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 provides 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 must 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

SC27-9527-01

