

Elastic Storage Server
5.3.7

Problem Determination Guide



Note

Before using this information and the product it supports, read the information in [“Notices” on page 113](#).

This edition applies to Version 5.3.7 of the Elastic Storage Server (ESS) for Power®, to Version 5 release 0 modification 5 of the following product editions, and to all subsequent releases and modifications until otherwise indicated in new editions:

- IBM Spectrum® Scale Data Management Edition for IBM® ESS (product number 5765-DME)
- IBM Spectrum Scale Data Access Edition for IBM ESS (product number 5765-DAE)

Significant changes or additions to the text and illustrations are indicated by a vertical line (|) to the left of the change.

IBM welcomes your comments; see the topic [“How to submit your comments” on page xiv](#). 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 International Business Machines Corporation 2015, 2021.**

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

Contents

Tables.....	V
About this information.....	vii
Related information.....	xii
Conventions used in this information.....	xiii
How to submit your comments.....	xiv
Chapter 1. Re-creating the NVR partitions.....	1
Chapter 2. Re-creating NVRAM pdisks.....	3
Chapter 3. Replacing the logtip backup solid state drive	5
Chapter 4. Steps to restore an I/O node.....	9
Chapter 5. Best practices for troubleshooting.....	15
How to get started with troubleshooting.....	15
Back up your data.....	16
Resolve events in a timely manner.....	17
Keep your software up to date.....	17
Subscribe to the support notification.....	17
Know your IBM warranty and maintenance agreement details.....	17
Know how to report a problem.....	17
Chapter 6. Limitations.....	19
Limit updates to Red Hat Enterprise Linux (ESS 5.3).....	19
Chapter 7. Collecting information about an issue.....	21
Chapter 8. GUI Issues.....	23
Issue with loading GUI	23
Chapter 9. Recovery Group Issues.....	25
Chapter 10. Contacting IBM.....	27
Information to collect before contacting the IBM Support Center.....	27
How to contact the IBM Support Center.....	29
Chapter 11. Maintenance procedures.....	31
Updating the firmware for host adapters, enclosures, and drives.....	31
Disk diagnosis.....	32
Background tasks.....	33
Server failover in ESS.....	33
Data checksums.....	34
Disk replacement.....	34
Replacing failed disks in an ESS recovery group: a sample scenario.....	34
Using the mmchcarrier command to fix issues caused by improper disk removal.....	39
Commandless disk replacement.....	40

Replacing failed ESS storage enclosure components: a sample scenario.....	42
Replacing a failed ESS storage drawer: a sample scenario.....	42
Replacing a failed ESS storage enclosure: a sample scenario.....	51
Replacing failed disks in a Power 775 Disk Enclosure recovery group: a sample scenario	58
Other hardware service.....	63
Directed maintenance procedures available in the GUI.....	63
Replace disks.....	64
Update enclosure firmware.....	65
Update drive firmware.....	65
Update host-adapter firmware.....	66
Start NSD.....	66
Start GPFS daemon.....	66
Increase filespace.....	67
Synchronize node clocks.....	67
Start performance monitoring collector service.....	67
Start performance monitoring sensor service.....	68
Activate AFM performance monitoring sensors.....	68
Activate NFS performance monitoring sensors.....	69
Activate SMB performance monitoring sensors.....	69
Configure NFS sensors.....	69
Configure SMB sensors.....	70
Mount file system if it must be mounted.....	71
Start the GUI service on the remote nodes.....	71
Chapter 12. References.....	73
Events.....	73
Array events.....	73
Enclosure events.....	74
Virtual disk events.....	77
Physical disk events.....	78
Recovery group events.....	80
Server events.....	81
Canister events.....	86
Messages.....	89
Message severity tags.....	89
IBM Spectrum Scale RAID messages.....	91
Notices.....	113
Trademarks.....	114
Terms and conditions for product documentation.....	114
Glossary.....	117
Index.....	125

Tables

1. Conventions.....	xiii
2. IBM websites for help, services, and information.....	18
3. Background tasks.....	33
4. ESS fault tolerance for drawer/enclosure.....	43
5. ESS fault tolerance for drawer/enclosure.....	52
6. DMPs.....	64
7. NFS sensor configuration example.....	70
8. SMB sensor configuration example.....	70
9. Events for the Array component.....	73
10. Events for the Enclosure component.....	74
11. Events for the virtual disk component.....	77
12. Events for the physical disk component.....	78
13. Events for the Recovery group component.....	80
14. Server events.....	81
15. Events for the Canister component.....	86
16. IBM Spectrum Scale message severity tags ordered by priority.....	90
17. ESS GUI message severity tags ordered by priority.....	90

About this information

This information guides you in monitoring and troubleshooting the Elastic Storage Server (ESS) Version 5.x for Power and all subsequent modifications and fixes for this release.

Which Elastic Storage System 3000 information unit provides the information you need?

The Elastic Storage System 3000 consists of the information units listed in [#unique_5/unique_5_Connect_42_table_ess3k](#) on page vii.

Information unit	Type of information	Intended users
Hardware Planning and Installation Guide	This unit provides the following information: <ul style="list-style-type: none">• Notices• Technical overview• Planning for hardware• Installing• Troubleshooting (SSR tasks)• Planning worksheets (customer task)• Cabling Guide (SSR tasks)	System administrators and IBM support team
Quick Deployment Guide	This unit provides the following information: <ul style="list-style-type: none">• ESS 3000 contents• ESS 3000 example setup flow• ESS 3000 best practices• ESS 3000 deployment scenarios• ESS 3000 common setup instructions• ESS 3000 initial setup instructions• ESS 3000 upgrade instructions• ESS 3000 known issues• ESS 3000: EMS upgrade scenarios	System administrators, analysts, installers, planners, and programmers of IBM Spectrum Scale clusters who are very experienced with the operating systems on which each IBM Spectrum Scale cluster is based
Service Guide	This unit provides the following information: <ul style="list-style-type: none">• Events• Servicing (Customer tasks)• Servicing (service personnel only)• Part Listings	System administrators and IBM support team

Information unit	Type of information	Intended users
ESS 3000 Problem Determination Guide	<p>This unit provides the following information:</p> <ul style="list-style-type: none"> • Call home in systems to resolve events raised for hardware issue • Software call home • Replacing an ESS 3000 canister server • Best practices for troubleshooting • Collecting information about an issue • ESS 3000 deployment troubleshooting: Helpful podman, Ansible, and log information • GUI Issues • Recovery Group Issues • Contacting IBM • Maintenance procedures • References 	System administrators and IBM support team
Command Reference Guide	<p>This unit provides the following information:</p> <ul style="list-style-type: none"> • ESS 3000 commands • ESS 3000 scripts 	System administrators and IBM support team

Information unit	Type of information	Intended users
IBM Spectrum Scale RAID: Administration	<p>This unit provides the following information:</p> <ul style="list-style-type: none"> • Summary of changes • Introducing IBM Spectrum Scale RAID • Administering IBM Spectrum Scale RAID • Managing IBM Spectrum Scale RAID with the mmvdisk command • Managing IBM Spectrum Scale RAID • Managing TRIM support for storage space reclamation • Configuring components on the Elastic Storage Server • Monitoring IBM Spectrum Scale RAID • Checking the health of an ESS configuration: a sample scenario • Setting up IBM Spectrum Scale RAID on the Elastic Storage Server • IBM Spectrum Scale RAID management API • Best-practice recommendations for IBM Spectrum Scale RAID • IBM Spectrum Scale RAID commands • IBM Spectrum Scale RAID scripts • Setting up GNR on the Power 775 Disk Enclosure • Accessibility features for IBM Spectrum Scale RAID • Notices • Glossary 	<ul style="list-style-type: none"> • System administrators of IBM Spectrum Scale systems • Application programmers who are experienced with IBM Spectrum Scale systems and familiar with the terminology and concepts in the XDSM standard

Which Elastic Storage System 5000 information unit provides the information you need?

The Elastic Storage System 5000 consists of the information units listed in [#unique_5/unique_5_Connect_42_table_5000guides](#) on page x.

Information unit	Type of information	Intended users
ESS 5000 Hardware Guide	<p>This unit provides the following information:</p> <ul style="list-style-type: none"> • Notices • Edition notice • About this guide • Notices • System overview • Installing • Troubleshooting (SSR tasks) • Planning worksheets (customer task) • 5105 22E: Reference information • Accessibility features for the system • Glossary 	System administrators and IBM support team
ESS 5000 Quick Deployment Guide	<p>This unit provides the following information:</p> <ul style="list-style-type: none"> • Notices • Edition notice • About this information • ESS 5000 Software deployment preparation • ESS 5000 Common installation instructions • ESS 5000 Re-installation and cleanup instructions • How to set up chronyd (time server) • ESS 5000 protocol node deployment by using the IBM Spectrum Scale installation toolkit • Sample scenario: ESS 3000 and ESS 5000 mixed cluster and file system • Accessibility features for the system • Notices • Glossary 	System administrators, analysts, installers, planners, and programmers of IBM Spectrum Scale clusters who are very experienced with the operating systems on which each IBM Spectrum Scale cluster is based
Model 092 storage enclosures	<p>This unit provides the following information:</p> <ul style="list-style-type: none"> • Notices • Weight considerations: Model 092 expansion enclosure • System overview • Planning for hardware • Installing Model 092 expansion enclosure • Servicing • Planning worksheets (customer task) 	System administrators and IBM support team

Information unit	Type of information	Intended users
Model 106 storage enclosures	<p>This unit provides the following information:</p> <ul style="list-style-type: none"> • Hardware installation and maintenance information - Model 106 • CMA and rail kit hardware fasteners overview - Model 106 	
ESS 5000 Problem Determination Guide	<p>This unit provides the following information:</p> <ul style="list-style-type: none"> • Call home in 5105-22E systems to resolve events raised for hardware issue • Software call home • Re-creating NVRAM pdisks • Replacing the logtip backup solid state drive • Best practices for troubleshooting • Collecting information about an issue • ESS 5000 deployment troubleshooting: Helpful podman, Ansible, and log information • GUI Issues • Recovery Group Issues • Contacting IBM • Maintenance procedures • References 	System administrators and IBM support team
Command Reference Guide	<p>This unit provides the following information:</p> <ul style="list-style-type: none"> • ESS 5000 commands • ESS 5000 scripts 	System administrators and IBM support team

Information unit	Type of information	Intended users
IBM Spectrum Scale RAID: Administration	<p>This unit provides the following information:</p> <ul style="list-style-type: none"> • Summary of changes • Introducing IBM Spectrum Scale RAID • Administering IBM Spectrum Scale RAID • Managing IBM Spectrum Scale RAID with the mmvdisk command • Managing IBM Spectrum Scale RAID • Managing TRIM support for storage space reclamation • Configuring components on the Elastic Storage Server • Monitoring IBM Spectrum Scale RAID • Checking the health of an ESS configuration: a sample scenario • Setting up IBM Spectrum Scale RAID on the Elastic Storage Server • IBM Spectrum Scale RAID management API • Best-practice recommendations for IBM Spectrum Scale RAID • IBM Spectrum Scale RAID commands • IBM Spectrum Scale RAID scripts • Setting up GNR on the Power 775 Disk Enclosure • Accessibility features for IBM Spectrum Scale RAID • Notices • Glossary 	<ul style="list-style-type: none"> • System administrators of IBM Spectrum Scale systems • Application programmers who are experienced with IBM Spectrum Scale systems and familiar with the terminology and concepts in the XDSM standard

Related information

ESS information

The ESS 5.3.7.x library consists of following information units. You can access these publications on [IBM Documentation](#) or [IBM Publications Center](#).

- *Elastic Storage Server: Quick Deployment Guide*, SC28-3189
- *Elastic Storage Server: Protocol Nodes Quick Deployment Guide*, SC28-3190
- *Elastic Storage Server: Problem Determination Guide*, GC28-3191
- *Elastic Storage Server: Command Reference*, SC28-3192
- *IBM Spectrum Scale RAID: Administration*, SC28-3142
- *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
- For information about the IBM Power Systems EXP24S I/O Drawer (FC 5887), see [IBM Documentation](#).

For the latest support information about IBM Spectrum Scale RAID, see the IBM Spectrum Scale RAID FAQ in [IBM Documentation](#).

Other related information

For information about:

- IBM Spectrum Scale, see [IBM Documentation](#).
- 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](#).
- Detailed information about the IBM Spectrum Scale installation toolkit, see [Using the installation toolkit to perform installation tasks: Explanations and examples](#).
- CES HDFS, see [Adding CES HDFS nodes into the centralized file system](#).
- Installation toolkit ESS support, see [ESS awareness with the installation toolkit](#).
- IBM POWER8® servers, see [IBM Documentation](#).
- Extreme Cluster/Cloud Administration Toolkit (xCAT), go to the [xCAT website](#).
 - [xCAT 2.16.1 Release Notes®](#)
- Mellanox OFED Release Notes (4.9), go to <https://docs.mellanox.com/display/OFEDv490170/Release%20Notes>
- IBM Electronic Service Agent (ESA) documentation, go to <https://www-01.ibm.com/support/esa/>.
- Drive call home, see *Configuring call home* in *ESS Quick Deployment Guide*.

Conventions used in this information

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

Table 1. Conventions

Convention	Usage
bold	<p>Bold words or characters represent system elements that you must use literally, such as commands, flags, values, and selected menu options.</p> <p>Depending on the context, bold typeface sometimes represents path names, directories, or file names.</p>
<u>bold underlined</u>	<u>bold underlined</u> keywords are defaults. These take effect if you do not specify a different keyword.

Table 1. Conventions (continued)

Convention	Usage
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>.
item...	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 Documentation](#).

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

scale@us.ibm.com

Chapter 1. Re-creating the NVR partitions

The Non-Volatile Random-Access Memory (NVRAM) physically resides within the IPR-Raid adapter that is installed on the EMS, and each of the I/O nodes. The NVR partitions are created on the local sda drive that is installed on the ESS I/O nodes to hold data for the log tip pdisks.

Although a total of 6 partitions are created, only 2 are actually used per I/O node, one for each NVR pdisk. In some cases the NVRAM partitions might need to be recreated. For example, an hardware/OS failure would be one such case.

Before re-creating the NVR partitions, list all the existing partitions for sda. To list all partitions for sda, run the following command:

```
parted /dev/sda unit KiB print
```

This command will give a similar output:

```
Model: IBM IPR-10 749FFB00 (scsi)
Disk /dev/sda: 557727744kiB
Sector size (logical/physical): 512B/4096B
Partition Table: msdos
Disk Flags:

Number  Start      End          Size         Type         File system  Flags
 1      1024kiB    9216kiB      8192kiB      primary      boot, prep
 2      9216kiB   521216kiB    512000kiB    primary      xfs
 3      521216kiB 176649216kiB 176128000kiB primary      xfs
 4      176649216kiB 557727744kiB 381078528kiB extended
 5      176651264kiB 279051264kiB 102400000kiB logical      xfs
 6      279052288kiB 381452288kiB 102400000kiB logical      xfs
 7      381453312kiB 483853312kiB 102400000kiB logical      xfs
 8      483854336kiB 535054336kiB 51200000kiB logical      xfs
 9      535055360kiB 543247360kiB 8192000kiB   logical      linux-swap(v1)
```

For optimal alignment, each partition must be exactly 2048000 KiB in size, and must be 1024 KiB apart from each other.

In the sample output, the last end size pertains to Partition # 9, and has a value of 543247360 KiB.

To get the NVR partition's new start value, add 1024 KiB to the last end size value, and add 2048000 KiB to the start value to determine the new end as shown:

1. NVR Partition 1 new start value = Last end size value + 1024 KiB = 543247360 KiB + 1024 KiB = 543248384 KiB
2. NVR Partition 1 new end = NVR Partition 1 new start value + 2048000 KiB = 543248384 KiB + 2048000 KiB = 545296384 KiB

To create the first NVR partition, run the following command:

```
parted /dev/sda mkpart logical 543248384KiB 545296384KiB
```

To get the new start for the second partition, you need to add 1024 KiB to the end size value of partition 1. Repeat the steps to calculate the start and end positions for the second partition as shown:

1. NVR Partition 2 new start = NVR Partition 1 end value + 1024 KiB = 545296384 KiB + 1024 KiB = 545297408 KiB
2. NVR Partition 2 new end = NVR Partition 2 new start value + 2048000 KiB = 545297408 KiB + 2048000 KiB = 547345408 KiB

Repeat the above steps four times to create a total of six partitions. When complete, the partitions list for sda will look similar to the following:

```
[root@ems1 ~]# parted /dev/sda unit KiB print
Model: IBM IPR
```

10 749FFB00 (scsi)

Disk /dev/sda: 557727744kiB

Sector size (logical/physical): 512B/4096B

Partition Table: msdos

Disk Flags:

Number	Start	End	Size	Type	File system	Flags
1	1024kiB	9216kiB	8192kiB	primary		boot, prep
2	9216kiB	521216kiB	512000kiB	primary	xfs	
3	521216kiB	176649216kiB	176128000kiB	primary	xfs	
4	176649216kiB	557727744kiB	381078528kiB	extended		
5	176651264kiB	279051264kiB	102400000kiB	logical	xfs	
6	279052288kiB	381452288kiB	102400000kiB	logical	xfs	
7	381453312kiB	483853312kiB	102400000kiB	logical	xfs	
8	483854336kiB	535054336kiB	51200000kiB	logical	xfs	
9	535055360kiB	543247360kiB	8192000kiB	logical	linux-swap(v1)	
10	543248384kiB	545296384kiB	2048001kiB	logical	xfs	
11	545297408kiB	547345408kiB	2048001kiB	logical	xfs	
12	547346432kiB	549394432kiB	2048001kiB	logical	xfs	
13	549395456kiB	551443456kiB	2048001kiB	logical	xfs	
14	551444480kiB	553492480kiB	2048001kiB	logical	xfs	
15	553493504kiB	555541504kiB	2048001kiB	logical	xfs	

Chapter 2. Re-creating NVRAM pdisks

NVRAM pdisks are used to store the log tip data, which is eventually migrated to the log home vdisk. Although ESS can continue to function without NVRAM pdisks, the performance is impacted without their presence. Therefore, it is important to ensure that the NVRAM pdisks are functioning at all times.

The NVRAM pdisks may stop functioning and go into a missing state. This could be due to hardware failure of the IPR card, or corrupt or missing NVR OS partition caused by an OS failure. To fix this problem, the NVRAM pdisks must be recreated.

You can find the pdisks that are in a missing state by running the **mmmlsrecoverygroup** command as shown:

```
mmmlsrecoverygroup rg_gssio1 -L --pdisk | grep NVR
NVR          no          1          2          0,0          1          3632 MiB          14 days          inactive          0%          low
n1s01        0, 0          NVR          1816 MiB          missing
n2s01        0, 0          NVR          1816 MiB          missing

mmmlsrecoverygroup rg_gssio2 -L --pdisk | grep NVR
NVR          no          1          2          0,0          1          3632 MiB          14 days          inactive          0%          low
n1s02        0, 0          NVR          1816 MiB          missing
n2s02        0, 0          NVR          1816 MiB          missing
```

Before recreating the pdisks, ensure that all six NVRAM partitions exist on the sda by using the following command:

```
parted /dev/sda unit KiB print

Model: IBM IPR
-
10 749FFB00 (scsi)
Disk /dev/sda: 557727744kiB
Sector size (logical/physical): 512B/4096B
Partition Table: msdos
Disk Flags:
Number  Start      End          Size         Type         File system  Flags
1       1024kiB    9216kiB      8192kiB      primary      xfs          boot, prep
2       9216kiB   521216kiB   512000kiB    primary      xfs
3       521216kiB 176649216kiB 176128000kiB primary      xfs
4       176649216kiB 557727744kiB 381078528kiB extended
5       176651264kiB 279051264kiB 102400000kiB logical      xfs
6       279052288kiB 381452288kiB 102400000kiB logical      xfs
7       381453312kiB 483853312kiB 102400000kiB logical      xfs
8       483854336kiB 535054336kiB 51200000kiB logical      xfs
9       535055360kiB 543247360kiB 8192000kiB logical      linux-swaps(v1)
10      543248384kiB 545296384kiB 2048001kiB logical      xfse*/
11      545297408kiB 547345408kiB 2048001kiB logical      xfs
12      547346432kiB 549394432kiB 2048001kiB logical      xfs
13      549395456kiB 551443456kiB 2048001kiB logical      xfs
14      551444480kiB 553492480kiB 2048001kiB logical      xfs
15      553493504kiB 555541504kiB 2048001kiB logical      xfs
```

Note: In case the partitions are not present, you must recreate the 6 NVR partitions. For more information, see ["Re-creating the NVR partitions"](#).

After you have verified the 6 NVR partitions, create a stanza file for each of the NVRAM devices that are missing, and save it.

```
gssio1.stanza:
%pdisk: pdiskName=n1s01 device=//gssio1/dev/sda10 da=NVR rotationRate=NVRAM
%pdisk: pdiskName=n2s01 device=//gssio2/dev/sda10 da=NVR rotationRate=NVRAM
```

Run the **mmaddpdisk** command using the stanza file that was created to replace the missing pdisks.

```
mmaddpdisk rg_gssio1 -F gssio1.stanza --replace
```

The following pdisks will be formatted on the node `gssio.ess.com`:

- `//gssio1/dev/sda10`
- `//gssio2/dev/sda10`

Run the **mm`lsrecoverygroup`** command to confirm the current state of the pdisks.

```
mmlsrecoverygroup rg_gssio1 -L --pdisk | grep NVR
n1s01              1, 1      NVR          1816 MiB    ok
n2s01              1, 1      NVR          1816 MiB    ok
```

Run the **mm`addpdisk`** command to recreate the other missing NVRAM pdisks.

Chapter 3. Replacing the logtip backup solid state drive

The solid state drive (SSD) is used to hold the *logtip backup vdisk*. In a normal operation, when both the copies of the primary logtip vdisk on the NVRAM are available, the logtip backup vdisk is not used during the write. However, when a write to the logtip is not able to make copies on every replica of the vdisk, GNR also writes the logtip data to the logtip backup vdisk.

Ensure that the following checks are done before you begin to replace the SSD DA:

1. Ensure that the system is healthy by running the **gssprecheck**, **gnrhealthcheck**, **gssintallcheck**, and **mmhealth node show** commands.
2. Ensure that the hardware does not have any issues.
3. Stop or suspend any long running tasks.
4. Find a maintenance window with a light load. For example, weekends or off-hour shifts, etc.

Follow these steps to replace logtip backup devices:

1. Obtain the existing logTip and logTipBackup vdisks details, SSD DA details and the recovery group information.

Note: Ensure that you have the details of the SSD device names and recovery groups before and after SSD replacements. This information is required later for comparison.

- a. Run the **mmlsvdisk** command to get the logTip and logTipBackup vdisks details. The command gives an output similar to the following:

```
[root@gssio1 home]# mmlsvdisk --recovery-group rg_gssio1-te0 --declustered-array NVR
vdisk:
name = "rg_gssio1_te0_logtip"
raidCode = "2WayReplication"
recoveryGroup = "rg_gssio1-te0"
declusteredArray = "NVR"
logGroup = "root"
blockSizeInKib = 2048
size = "48 MiB"
rawSize = 100663296
checksumGranularity = 4096
state = "ok"
remarks = "logTip"
```

```
[root@gssio1 home]# mmlsvdisk --recovery-group rg_gssio2-te0 --declustered-array NVR
vdisk:
name = "rg_gssio2_te0_logtip"
raidCode = "2WayReplication"
recoveryGroup = "rg_gssio2-te0"
declusteredArray = "NVR"
logGroup = "root"
blockSizeInKib = 2048
size = "48 MiB"
rawSize = 100663296
checksumGranularity = 4096
state = "ok"
remarks = "logTip"
```

```
[root@gssio1 home]# mmlsvdisk --recovery-group rg_gssio1-te0 --declustered-array SSD
vdisk:
name = "rg_gssio1_te0_logtipbackup"
raidCode = "Unreplicated"
recoveryGroup = "rg_gssio1-te0"
declusteredArray = "SSD"
logGroup = "root"
blockSizeInKib = 2048 size = "48 MiB"
rawSize = 268435456
checksumGranularity = 4096
state = "ok"
remarks = "logTipBackup"
```

```
[root@gssio1 home]# mmlsvdisk --recovery-group rg_gssio2-te0 --declustered-array SSD
vdisk:
name = "rg_gssio2_te0_logtipbackup"
raidCode = "Unreplicated"
```

```
recoveryGroup = "rg_gssio2-te0"
declusteredArray = "SSD"
logGroup = "root"
blockSizeInKib = 2048
size = "48 MiB"
rawSize = 268435456
checksumGranularity = 4096
state = "ok"
remarks = "logTipBackup"
```

- b. Run the **mmlspdisk** command to get the SSD DA details. The command gives an output similar to the following:

```
[root@gssio1 tmp]# mmlspdisk all | grep ssd -A1 name = "e1s01ssd"
device = "///gssio2-te0/dev/sdad(notEnabled),///gssio2-te0/dev/
sdbv(notEnabled),///gssio1-te0/dev/sdag,///gssio1-te0/dev/sdbv"
--
name = "e1s24ssd"
device = "///gssio2-te0/dev/sdaw,///gssio2-te0/dev/sdcs,///gssio1-te0/
dev/sdw(notEnabled),///gssio1-te0/dev/sdcs(notEnabled)"
```

- c. Run the **mmlsrecoverygroup** command to list the information about the recovery groups. The command must be run against all recovery groups. The command gives an output similar to the following:

```
[root@gssio1 ~]# mmlsrecoverygroup rg_gssio1-hs -L
```

recovery group	declustered arrays	vdisks	pdisks	current format version	allowable format version
rg_gssio1-hs	3	11 1	77	4.2.2.0	4.2.2.0

declustered array	needs service	vdisks	pdisks	spares	replace threshold	free space	scrub duration	background activity task progress priority
NVR	no	1	2	0,0	1	3632 MiB	14 days	scrub 4% low
DA1	no	9	174	6,89	2	8192 MiB	14 days	scrub 42% low
SSD	no	1	1	0,0	1	372 GiB	14 days	scrub 4% low

vdisk	RAID code	declustered array	vdisk size	block size	checksum size granularity	state	remarks
rg_gssio1_hs_LOGTIP	2WayReplication	NVR	48 MiB	2 MiB	4096	ok	logTip
rg_gssio1_hs_LOGTIPBACKUP	Unreplicated	SSD	48 MiB	2 MiB	4096	ok	logTipBackup
rg_gssio1_hs_loghome	4WayReplication	DA1	40 GiB	2 MiB	4096	ok	log
rg_gssio1_hs_MetaData_1M_3W_1	3WayReplication	DA1	7355 GiB	1 MiB	32 KiB	ok	
rg_gssio1_hs_MetaData_1M_3W_2	3WayReplication	DA1	7355 GiB	1 MiB	32 KiB	ok	
rg_gssio1_hs_MetaData_1M_3W_3	3WayReplication	DA1	7355 GiB	1 MiB	32 KiB	ok	
rg_gssio1_hs_Data_8M_2p_1	8+2p	DA1	143 TiB	8 MiB	32 KiB	ok	
rg_gssio1_hs_Data_8M_2p_2	8+2p	DA1	143 TiB	8 MiB	32 KiB	ok	
rg_gssio1_hs_Data_8M_2p_3	8+2p	DA1	143 TiB	8 MiB	32 KiB	ok	
rg_gssio1_hs_ces_system_0	3WayReplication	DA1	23 GiB	1 MiB	32 KiB	ok	
rg_gssio1_hs_ces_data_0	8+2p	DA1	3952 GiB	1 MiB	32 KiB	ok	

config data	declustered array	spare space	remarks
rebuild space	DA1	93 pdisk	

config data	disk group fault tolerance	remarks
rg descriptor 2	enclosure limiting fault tolerance	system index 2 enclosure limited by rg descriptor

vdisk	disk group fault tolerance	remarks
rg_gssio1_hs_LOGTIP	1 pdisk	
rg_gssio1_hs_LOGTIPBACKUP	0 pdisk	
rg_gssio1_hs_loghome	2 enclosure limited by rg descriptor	
rg_gssio1_hs_MetaData_1M_3W_1	2 enclosure rg_gssio1_hs_MetaData_1M_3W_2	2 enclosure
rg_gssio1_hs_MetaData_1M_3W_3	2 enclosure rg_gssio1_hs_Data_8M_2p_1	1 enclosure
rg_gssio1_hs_Data_8M_2p_2	1 enclosure rg_gssio1_hs_Data_8M_2p_3	1 enclosure
rg_gssio1_hs_ces_system_0	2 enclosure rg_gssio1_hs_ces_data_0	1 enclosure

active recovery group server	servers
gssio1-hs.gpfs.net	gssio1-hs.gpfs.net,gssio2-hs.gpfs.net

2. Run the **mmdelvdisk** command to delete the logTip and logTipBackup vdisk. The command gives an output similar to the following:

```
[root@gssio1 home]# mmdelvdisk rg_gssio1_te0_logtip
mmdelvdisk: [I] Processing vdisk rg_gssio1_te0_logtip
mmdelvdisk: Propagating the cluster configuration data to all affected nodes. This is an asynchronous process.

[root@gssio1 home]# mmdelvdisk rg_gssio2_te0_logtip
mmdelvdisk: [I] Processing vdisk rg_gssio2_te0_logtip
mmdelvdisk: Propagating the cluster configuration data to all affected nodes. This is an asynchronous
```

```
process.

[root@gssio1 home]# mmdelvdisk rg_gssio1_te0_logtipbackup
mmdelvdisk: [I] Processing vdisk rg_gssio1_te0_logtipbackup
mmdelvdisk: Propagating the cluster configuration data to all affected nodes. This is an asynchronous
process.

[root@gssio1 home]# mmdelvdisk rg_gssio2_te0_logtipbackup
mmdelvdisk: [I] Processing vdisk rg_gssio2_te0_logtipbackup
mmdelvdisk: Propagating the cluster configuration data to all affected nodes. This is an asynchronous
process.
```

3. Run the **mmdelpdisk** command to delete the SSD DAs from both the recovery groups . The command gives an output similar to the following:

```
[root@gssio1 home]# mmdelpdisk rg_gssio1-te0 --declustered-array SSD
[root@gssio1 home]# mmdelpdisk rg_gssio2-te0 --declustered-array SSD
```

4. Swap the SSD drives.

Ensure that you have the correct FRUs in hand while you swap the drives.

The SSD device names might change. To ensure that you have the correct device names for the stanzas, use following procedure:

- a. Run the **mmgetpdisktopology** command to collect the topology information about the servers. The command gives an output similar to the following:

```
[root@gssio1 home]# mmgetpdisktopology > left_server.top
[root@gssio2 home]# mmgetpdisktopology > right_server.top
```

- b. Run the **mkrginput** command on any one of the servers to get the topology files and recovery group stanza information. The recovery group stanza file contains the correct device mapping and other details. The command gives an output similar to the following:

```
[root@gssio1 home]# mkrginput left_server.top right_server.top
GMR server disk topology: ESS GS2 HDD
GMR recovery group layout: GS GSS/ESS RG Number of recovery groups: 2
Created stanza file: G51704ML.stanza
Created stanza file: G51704MR.stanza
Stanza file contains '%vdisk:' lines with log vdisk attributes.
```

- c. Retrieve the SSD DA, pdisk, and the required vdisk definitions from the stanza file.

```
[root@gssio1 home]# grep -e SSD -e LOGTIP G51704ML.stanza > ssd_daL.conf
[root@gssio1 home]# grep -e SSD -e LOGTIP G51704MR.stanza > ssd_daR.conf

[root@gssio1 home]# cat ssd_daL.conf
%da: daName=SSD replaceThreshold=1 spares=0 VCDspares=0 canHoldVCD=no nspdEnable=no auLogSize=120m
%pdisk: pdiskName=e1s01ssd device=/gssio1-te0.gpfs.net/dev/sdag da=SSD
%#vdisk: vdiskName={rg}LOGTIP rg={rg} da=NVR diskUsage=vdiskLogTip
checksumGranularity=4096 raidCode=2WayReplication blocksize=2m size=48m
%#vdisk: vdiskName={rg}LOGTIPBACKUP rg={rg} da=SSD diskUsage=vdiskLogTipBackup
checksumGranularity=4096 raidCode=Unreplicated blocksize=2m size=48m

[root@gssio1 home]# cat ssd_daR.conf
%da: daName=SSD replaceThreshold=1 spares=0 VCDspares=0 canHoldVCD=no nspdEnable=no auLogSize=120m
%pdisk: pdiskName=e1s24ssd device=/gssio1-te0.gpfs.net/dev/sdcs da=SSD
%#vdisk: vdiskName={rg}LOGTIP rg={rg} da=NVR diskUsage=vdiskLogTip
checksumGranularity=4096 raidCode=2WayReplication blocksize=2m size=48m
%#vdisk: vdiskName={rg}LOGTIPBACKUP rg={rg} da=SSD

diskUsage=vdiskLogTipBackup checksumGranularity=4096 raidCode=Unreplicated blocksize=2m size=48m
```

- d. In the newly created .conf file, uncomment the %vdisk line and replace the vdiskName and recovery group information with the corresponding names saved in [step 1](#). After the modification the stanza configuration files gives an output similar to the following:

```
[root@gssio1 home]# cat ssd_daL.conf
%da: daName=SSD replaceThreshold=1 spares=0 VCDspares=0 canHoldVCD=no nspdEnable=no auLogSize=120m
%pdisk: pdiskName=e1s01ssd device=/gssio1-te0.gpfs.net/dev/sdag da=SSD
%vdisk: vdiskName=rg_gssio1_te0_logtip rg=rg_gssio1-te0
da=NVR diskUsage=vdiskLogTip checksumGranularity=4096 raidCode=2WayReplication blocksize=2m size=48m
%vdisk: vdiskName=rg_gssio1_te0_logtipbackup rg=rg_gssio1-te0
```

```
da=SSD diskUsage=vdiskLogTipBackup checksumGranularity=4096 raidCode=Unreplicated blocksize=2m size=48m

[root@gssio1 home]# cat ssd_daR.conf
%da: daName=SSD replaceThreshold=1 spares=0 VCDSpares=0 canHoldVCD=no nspdEnable=no auLogSize=120m
%pdisk: pdiskName=e1s24ssd device=//gssio1-te0.gpfs.net/dev/sdcs da=SSD
%vdisk: vdiskName=rg_gssio2_te0_logtip rg=rg_gssio2-te0
da=NVR diskUsage=vdiskLogTip checksumGranularity=4096 raidCode=2WayReplication blocksize=2m size=48m
%vdisk: vdiskName=rg_gssio2_te0_logtipbackup rg=rg_gssio2-te0
da=SSD diskUsage=vdiskLogTipBackup checksumGranularity=4096 raidCode=Unreplicated blocksize=2m size=48m
```

5. Add the SSD DAs and create the logTip and logTipBackup vdisks.

- a. Run the **mmaddpdisk** command to add the SSD DAs for each recovery group. The command gives an output similar to the following:

```
[root@gssio1 home]# mmaddpdisk rg_gssio1-te0 -F ssd_daL.conf

The following pdisks will be formatted on node gssio1.gpfs.net:
//gssio2-te0/dev/sdad, //gssio2-te0/dev/sdbv, //gssio1-te0/dev/
sdag, //gssio1-te0/dev/sdbv

[root@gssio1 home]# mmaddpdisk rg_gssio2-te0 -F ssd_daR.conf

The following pdisks will be formatted on node gssio2.gpfs.net:
//gssio2-te0/dev/sdaw, //gssio2-te0/dev/sdcs, //gssio1-te0/dev/sdw, //
gssio1-te0/dev/sdcs
```

- b. Run the **mmcrvdisk** command to add the logTip and logTipBack vdisk for each recovery group. The command gives an output similar to the following:

```
[root@gssio1 home]# mmcrvdisk -F ssd_daL.conf
mmcrvdisk: [I] Processing vdisk rg_gssio1_te0_logtip
mmcrvdisk: [I] Processing vdisk rg_gssio1_te0_logtipbackup
mmcrvdisk: Propagating the cluster configuration data to all affected nodes.
This is an asynchronous process.

[root@gssio1 home]# mmcrvdisk -F ssd_daR.conf
mmcrvdisk: [I] Processing vdisk rg_gssio2_te0_logtip
mmcrvdisk: [I] Processing vdisk rg_gssio2_te0_logtipbackup
mmcrvdisk: Propagating the cluster configuration data to all affected nodes.
This is an asynchronous process.
```

6. Verify and perform health checks.

Check the logtip and logtipbackup vdisks. The information in them must match the data that was saved in step 1. Post-replacement recovery group information also must match the data that was saved in step 1. Perform a final health check by running the **gssprecheck**, **gnrhealthcheck**, **gssintallcheck**, and **mmhealth node show** commands.

Chapter 4. Steps to restore an I/O node

If an I/O node fails due to a hardware or OS problem, and the OS is no longer accessible, you must restore the node by using the existing configuration settings that are stored in xCAT, which typically stored on the EMS node.

This process restores the OS image and the required ESS software, drivers, and firmware.

Note: For the following steps, assume that the gssio1 node is the node that is being restored.

1. Disable the GPFS auto load by using the mmchconfig command.

Note: This prevents GPFS from restarting automatically upon reboot.

```
[ems1]# mmlsconfig autoload
autoload yes
[ems1]# mmchconfig autoload=no
[ems1]# mmlsconfig autoload
autoload no
```

2. List the recovery groups by using the mmlsrecoverygroup command to verify that the replacement node is not an active recovery group server currently.

```
[ems1]# mmlsrecoverygroup
recovery group      vdisks      vdisks  servers
-----
rg_gssio1           3          18      gssio1,gssio2
rg_gssio2           3          18      gssio2,gssio1
```

List the current active recovery group server for each recovery group.

```
[ems1]# mmlsrecoverygroup rg_gssio1 -L | grep "active recovery" -A2
active recovery group server      servers
-----
gssio1                           gssio1,gssio2

[ems1]# mmlsrecoverygroup rg_gssio2 -L | grep "active recovery" -A2
active recovery group server      servers
-----
gssio2                           gssio2,gssio1
```

Note: When you restore gssio1, the primary and active recovery group server for rg_gssio1 must be gssio2. If the server is not set to gssio2, you must run the **mmchrecoverygroup** command. If the recovery group is under the mmvdisk control, you must run the **mmvdisk** command to change the server.

```
[ems1]# mmchrecoverygroup <RG> --servers <NEW PRIMARY NODE>,<OLD PRIMARY NODE> -v no
[ems1]# mmchrecoverygroup <RG> --active <NEW PRIMARY NODE>
or
[ems1]# mmvdisk rg change --rg <RG> --primary <NEW PRIMARY NODE> --backup <OLD PRIMARY
NODE> -v no
[ems1]# mmvdisk rg change --rg <RG> --active <NEW PRIMARY NODE>

[root@gssio1 ~]# mmchrecoverygroup rg_gssio1 --servers gssio2,gssio1 -v no
[root@gssio1 ~]# mmchrecoverygroup rg_gssio1 --active gssio2
[ems1]# mmlsrecoverygroup rg_gssio1 -L | grep "active recovery" -A2
active recovery group server      servers
-----
gssio2                           gssio2,gssio1

[ems1]# mmlsrecoverygroup rg_gssio2 -L | grep "active recovery" -A2
active recovery group server      servers
-----
gssio2                           gssio2,gssio1
```

3. Create a backup of the replacement node's network file.

```
[ems]# rm -rf /tmp/replacement_node_network_backup
[ems]# mkdir /tmp/replacement_node_network_backup
[ems]# scp <REPLACEMENT NODE>:/etc/sysconfig/network-scripts/ifcfg-*
/tmp/replacement_node_network_backup/
[ems]# scp gssio2:/etc/sysconfig/network-scripts/ifcfg-*
/tmp/replacement_node_network_backup/
```

Note: This is an optional step, and can be taken only when the replacement node can be accessed.

4. Check for the RHEL images available for installation on the EMS node.

The RHEL image is needed to re-image the node that is being restored. The OS image must be located on the EMS node under the following directory:

```
[ems]# ls /tftpboot/xcat/osimage/
rhels7.3-ppc64-install-gss
```

5. Configure the replacement node's boot state to Install for the specified OS image.

```
[ems]# nodeset <REPLACEMENT NODE> osimage=<OS_ISO_image>
[root@ems1 ~]# nodeset gssio2 osimage=rhels7.3-ppc64-install-gss
gssio2: install rhels7.3-ppc64-gss
```

6. Ensure that the remote console is properly configured on the EMS node.

```
[ems]# makeconservercf <REPLACEMENT NODE>
[root@ems1 ~]# makeconservercf gssio2
```

7. Reboot the replaced node to initiate the installation process.

```
[ems]# rnetboot <REPLACEMENT NODE> -V
[root@ems1 ~]# rnetboot gssio2 -V
lpar_netboot Status: List only ent adapters
lpar_netboot Status: -v (verbose debug) flag detected
lpar_netboot Status: -i (force immediate shutdown) flag detected
lpar_netboot Status: -d (debug) flag detected
node:gssio2
Node is gssio2
...
# Network boot proceeding - matched BOOTP, exiting.
# Finished.
sending commands ~. to expect
gssio2: Success
```

Monitor the progress of the installation, and wait for the xcatpost/yum/etc script to finish.

```
[ems]# watch "nodestat <REPLACEMENT NODE>; echo; tail /var/log/containers/<REPLACEMENT NODE>"
[root@ems1 ~]# watch "nodestat gssio2; echo; tail /var/log/containers/gssio2"
gssio2: noping
...
gssio2: install rhels7.3-ppc64-gss
...
gssio2: sshd
```

```
[ems]# watch -n .5 "ssh <REPLACEMENT NODE> 'ps -eaf | grep -v grep' |
egrep 'xcatpost|yum|rpm|vdp'"
[root@ems1 ~]# watch -n .5 "ssh gssio2 'ps -eaf | grep -v grep' |
egrep 'xcatpost|yum|rpm|vdp'"
```

Note: Depending on what needs to be updated, the node might reboot more than once. You must wait until there is no process output before taking the next step.

8. Verify that the upgrade files are copied to the I/O node sync directory, /install/gss/sync/ppc64/.

```
[ems]# ssh <REPLACEMENT NODE> "ls /install/gss/sync/ppc64/"
[root@ems1]# ssh gssio2 "ls /install/gss/sync/ppc64/"
gssio2: mofed
```

Wait for the directory to sync. After the mofed directory is created, you can take the next step.

9. Copy the host files from the healthy node to the replacement node.

```
[ems]# scp /etc/hosts <REPLACEMENT NODE>:/etc/  
[root@ems1 mofed]# scp /etc/hosts gssio2:/etc/
```

10. Configure the network on the replacement node.

If you created a backup of the network files previously, you can copy them over to the node, and restart the node. Verify that the names of the devices are consistent with the names in the backup file before you replace the files.

You can also apply the Red Hat® updates not included in the xCAT image, if necessary.

11. Rebuild the GPFS kernel extensions on the replacement node.

If the kernel patches were applied, it might be necessary to rebuild the GPFS portability layer by running the **mmbuildgpl** command.

```
[ems]# ssh <REPLACEMENT NODE> "/usr/lpp/mmfs/bin/mmbuildgpl"  
[root@ems1 ~]# ssh gssio2 "/usr/lpp/mmfs/bin/mmbuildgpl"  
-----  
mmbuildgpl: Building GPL module begins at Wed Nov  8 17:18:21 EST 2017.  
-----  
Verifying Kernel Header...  
kernel version = 31000514 (3.10.0-514.28.1.el7.ppc64, 3.10.0-514.28.1)  
module include dir = /lib/modules/3.10.0-514.28.1.el7.ppc64/build/include  
module build dir  = /lib/modules/3.10.0-514.28.1.el7.ppc64/build  
kernel source dir  = /usr/src/linux-3.10.0-514.28.1.el7.ppc64/include  
Found valid kernel header file under /usr/src/kernels/3.10.0-514.28.1.el7.ppc64/include  
Verifying Compiler...  
make is present at /bin/make  
cpp is present at /bin/cpp  
gcc is present at /bin/gcc  
g++ is present at /bin/g++  
ld is present at /bin/ld  
Verifying Additional System Headers...  
Verifying kernel-headers is installed ...  
Command: /bin/rpm -q kernel-headers  
The required package kernel-headers is installed  
make World ...  
make InstallImages ...  
-----  
mmbuildgpl: Building GPL module completed successfully at Wed Nov  8 17:18:39 EST 2017.
```

12. Restore the GPFS configuration from an existing healthy node in the cluster.

```
[ems]# ssh <REPLACEMENT NODE> "/usr/lpp/mmfs/bin/mmsdrrestore -p <GOOD NODE>"  
[root@ems ~]# ssh gssio2 "/usr/lpp/mmfs/bin/mmsdrrestore -p ems1"  
mmsdrrestore: Processing node gssio1  
mmsdrrestore: Node gssio1 successfully restored.
```

Note: This code is run on the replacement node, and the **-p** option is applied to an existing healthy node.

13. Start GPFS on the recovered node, and enable the GPFS auto load.

- a. Before you start GPFS, verify that the replacement node is still in the DOWN state.

```
[ems]# mmgetstate -al  
Node number  Node name  Quorum  Nodes up  Total nodes  GPFS state  Remarks  
-----  
1            gssio1    2        2         5           active     quorum node  
2            gssio2    0        0         5           down       quorum node  
3            ems1      2        2         5           active     quorum node  
4            gsscomp1  2        2         5           active  
5            gsscomp   2        2         5           active
```

- b. Start GPFS on the replacement node.

```
[ems]# mmstartup -N <REPLACEMENT NODE>  
mmstartup: Starting GPFS ...
```

- c. Verify that the replacement node is active.

```
[ems]# mmgetstate -al  
Node number  Node name  Quorum  Nodes up  Total nodes  GPFS state  Remarks  
-----  
1            gssio1    2        3         5           active     quorum node
```

2	gssio2	2	3	5	active	quorum node
3	ems1	2	3	5	active	quorum node
4	gsscomp1	2	3	5	active	
5	gsscomp2	2	3	5	active	

d. Ensure that all the file systems are mounted on the replacement node.

```
[ems]# mmmount all -N <REPLACEMENT NODE>
[ems]# mmlsmount all -L
```

e. Re-enable the GPFS auto load.

```
[ems]# mmlsconfig autoloader
autoloader no

[ems]# mmchconfig autoloader=yes
mmchconfig: Command successfully completed

[ems]# mmlsconfig autoloader
autoloader yes
```

14. The primary and active recovery group server for rg_gssio1 must be gssio1. If the server is not set to gssio1, you must run the **mmchrecoverygroup** command to set the server. If the recovery group is under the mmvdisk control, you must run the **mmvdisk** command to change the server.

```
[ems1]# mmchrecoverygroup <RG> --servers <NEW PRIMARY NODE>,<OLD PRIMARY NODE> -v no
[ems1]# mmchrecoverygroup <RG> --active <NEW PRIMARY NODE>
or
[ems1]# mmvdisk rg change --rg <RG> --primary <NEW PRIMARY NODE> --backup <OLD PRIMARY
NODE> -v no
[ems1]# mmvdisk rg change --rg <RG> --active <NEW PRIMARY NODE>

[root@gssio1 ~]# mmchrecoverygroup rg_gssio1 --servers gssio1,gssio2 -v no
[root@gssio1 ~]# mmchrecoverygroup rg_gssio1 --active gssio1
[ems1]# mmlsrecoverygroup rg_gssio1 -L | grep "active recovery" -A2
active recovery group server          servers
-----
gssio1                                gssio1,gssio2

[ems1]# mmlsrecoverygroup rg_gssio2 -L | grep "active recovery" -A2
active recovery group server          servers
-----
gssio2                                gssio2,gssio1
```

15. Verify that the NVRAM partition exists, and ensure that the following conditions are met:

- There must be 11 partitions.
- Partitions 6 through 11 must be 2 GB.
- Partitions 6 through 9 are marked as xfs for file system.
- Partitions 10 and 11 must not have a file system that is associated with it.
- After re-imaging, the node that was re-imaged will have an xfs file system as shown:

```
[ems]# ssh gssio1 "lsblk | egrep 'NAME|sda[0-9]'"
NAME        MAJ:MIN RM  SIZE RO TYPE MOUNTPOINT
--sda1       8:1    0    8M  0 part
--sda2       8:2    0 500M  0 part /boot
--sda3       8:3    0 246.1G 0 part /
--sda4       8:4    0    1K  0 part
--sda5       8:5    0   3.9G 0 part [SWAP]
--sda6       8:6    0    2G  0 part
--sda7       8:7    0    2G  0 part
--sda8       8:8    0    2G  0 part
--sda9       8:9    0    2G  0 part
--sda10      8:10   0    2G  0 part
--sda11      8:11   0    2G  0 part

[ems1]# ssh gssio1 "parted /dev/sda -l | egrep 'boot, prep' -B 1 -A 10"
Number      Start      End        Size      Type      File system  Flags
--
1            1049kB     9437kB     8389kB     primary
2            9437kB     534MB      524MB      primary    xfs
3            534MB      265GB      264GB      primary    xfs
4            265GB      284GB      18.9GB     extended
5            265GB      269GB      4194MB     logical    linux-swap(v1)
```

```

6      269GB  271GB  2097MB  logical  xfs
7      271GB  273GB  2097MB  logical  xfs
8      273GB  275GB  2097MB  logical  xfs
9      275GB  277GB  2097MB  logical  xfs
10     277GB  279GB  2097MB  logical
11     279GB  282GB  2097MB  logical

[ems1]# ssh gssio2 "lsblk | egrep 'NAME|sda[0-9]'"
NAME        MAJ:MIN RM   SIZE RO TYPE MOUNTPOINT
├─sda1       8:1    0     8M  0 part
├─sda2       8:2    0   500M  0 part /boot
├─sda3       8:3    0  246.1G  0 part /
├─sda4       8:4    0     1K  0 part
├─sda5       8:5    0   3.9G  0 part [SWAP]
├─sda6       8:6    0     2G  0 part
├─sda7       8:7    0     2G  0 part
├─sda8       8:8    0     2G  0 part
├─sda9       8:9    0     2G  0 part
├─sda10      8:10   0     2G  0 part
└─sda11      8:11   0     2G  0 part

[ems1]# ssh gssio2 "parted /dev/sda -l | egrep 'boot, prep' -B 1 -A 10"
Number      Start      End      Size      Type      File system      Flags
1            1049kB    9437kB   8389kB    primary   boot, prep
2            9437kB    534MB    524MB     primary   xfs
3            534MB     265GB    264GB     primary   xfs
4            265GB     284GB    18.9GB    extended
5            265GB     269GB    4194MB    logical   linux-swap(v1)
6            269GB     271GB    2097MB    logical   xfs
7            271GB     273GB    2097MB    logical   xfs
8            273GB     275GB    2097MB    logical   xfs
9            275GB     277GB    2097MB    logical   xfs
10           277GB     279GB    2097MB    logical   xfs
11           279GB     282GB    2097MB    logical   xfs

```

If the partitions do not exist, you need to create them. For more information, see [Chapter 1, “Re-creating the NVR partitions,”](#) on page 1

16. View the current NVR device status.

```

[ems1]# mmlsrecoverygroup rg_gssio1 -L --pdisk | egrep "n[0-9]s[0-9]"
n1s01      1, 1      NVR      1816 MiB  ok
n2s01      0, 0      NVR      1816 MiB  missing

[ems1]# mmlsrecoverygroup rg_gssio2 -L --pdisk | egrep "n[0-9]s[0-9]"
n1s02      1, 1      NVR      1816 MiB  ok
n2s02      0, 0      NVR      1816 MiB  missing

```

Note: The missing NVR devices must be re-created or replaced. For more information, see [Chapter 2, “Re-creating NVRAM pdisks,”](#) on page 3.

Chapter 5. Best practices for troubleshooting

Following certain best practices makes the troubleshooting process easier.

For information on IBM Spectrum Scale issues and their resolution, see the *Troubleshooting* section in the IBM Spectrum Scale documentation.

How to get started with troubleshooting

Troubleshooting the issues that are reported in the system is easier when you follow the process step-by-step.

When you experience some issues with the system, go through the following steps to get started with the troubleshooting:

1. Check the events that are reported in various nodes of the cluster by using the **mmhealth cluster show** and **mmhealth node show** commands.
2. Check the user action corresponding to the active events and take the appropriate action. For more information on the events and corresponding user action, see “Events” on page 73.
3. Check for events that happened before the event you are trying to investigate. They might give you an idea about the root cause of problems. For example, if you see an event `nfs_in_grace` and `node_resumed` a minute before you get an idea about the root cause why NFS entered the grace period, it means that the node resumed after a suspend.
4. Collect the details of the issues through logs, dumps, and traces. You can use various CLI commands and the **Settings > Diagnostic Data** GUI page to collect the details of the issues reported in the system.
5. Based on the type of issue, browse through the various topics that are listed in the troubleshooting section and try to resolve the issue.
6. If you cannot resolve the issue by yourself, contact IBM Support.

Example command output

```
# mmhealth cluster show
Component      Total      Failed      Degraded      Healthy      Other
-----
NODE            3           0           2             1           0
GPFS            3           0           1             2           0
NETWORK         3           0           0             3           0
FILESYSTEM      1           0           0             1           0
DISK            2           0           0             2           0
GUI             1           0           0             1           0
NATIVE_RAID     2           2           0             0           0
PERFMON         3           0           0             3           0
THRESHOLD       3           0           0             3           0

# mmhealth node show
Node name:      ems2-te0.example.net
Node status:    HEALTHY
Status Change:  4 days ago
Component      Status      Status Change      Reasons
-----
GPFS            HEALTHY     4 days ago         -
NETWORK         HEALTHY     4 days ago         -
FILESYSTEM      HEALTHY     4 days ago         -
GUI             HEALTHY     4 days ago         -
PERFMON         HEALTHY     4 days ago         -
THRESHOLD       HEALTHY     4 days ago         -

# gnrihealthcheck
#####
# gnrihealthcheck invoked: Wed Mar 17 10:33:16 UTC 2021
#####
# Beginning topology checks.
#####
Topology checks successful.
```

```
#####
# Beginning enclosure checks.
#####
Enclosure checks successful.
#####
# Beginning recovery group checks.
#####
Recovery group checks successful.
#####
# Beginning pdisk checks.
#####
Pdisk checks successful.
#####
# Beginning IBM Power RAID checks.
#####
IBM Power RAID checks successful.
#####
# Beginning the NVMe Controller checks.
#####
The NVMe Controller checks are successful.
#####
# Beginning SSD endurance checks
#####
The SSD endurance checks are successful.
```

Back up your data

You need to back up data regularly to avoid data loss. It is also recommended to take backups before you start troubleshooting. IBM Spectrum Scale provides various options to create data backups.

Follow the guidelines in the following sections to avoid any issues while creating backup:

- *GPFS(tm) backup data in IBM Spectrum Scale: Concepts, Planning, and Installation Guide*
- *Backup considerations for using IBM Spectrum Protect in IBM Spectrum Scale: Concepts, Planning, and Installation Guide*
- *Configuration reference for using IBM Spectrum Protect with IBM Spectrum Scale(tm) in IBM Spectrum Scale: Administration Guide*
- *Protecting data in a file system using backup in IBM Spectrum Scale: Administration Guide*
- *Backup procedure with SOBAR in IBM Spectrum Scale: Administration Guide*

The following best practices help you to troubleshoot the issues that might arise in the data backup process:

1. Enable the most useful messages in the **mmbackup** command by setting the **MMBACKUP_PROGRESS_CONTENT** and **MMBACKUP_PROGRESS_INTERVAL** environment variables in the command environment prior to issuing the **mmbackup** command. Setting **MMBACKUP_PROGRESS_CONTENT=7** provides the most useful messages. For more information on these variables, see the *mmbackup* command in the *IBM Spectrum Scale: Command and Programming Reference*.
2. If the **mmbackup** process is failing regularly, enable debug options in the backup process:

Use the **DEBUGmmbackup** environment variable or the **-d** option that is available in the **mmbackup** command to enable debugging features. This variable controls what debugging features are enabled. It is interpreted as a bitmask with the following bit meanings:

0x001

Specifies that basic debug messages are printed to STDOUT. There are multiple components that comprise **mmbackup**, so the debug message prefixes can vary. Some examples include:

```
mmbackup:mbackup.sh
DEBUGtsbackup33:
```

0x002

Specifies that temporary files are to be preserved for later analysis.

0x004

Specifies that all **dsmc** command output is to be mirrored to STDOUT.

The **-d** option in the **mmbackup** command line is equivalent to **DEBUGmmbackup=1**.

3. To troubleshoot problems with backup subtask execution, enable debugging in the **tsbuhelper** program.

Use the **DEBUGtsbuhelper** environment variable to enable debugging features in the **mmbackup** helper program **tsbuhelper**.

Resolve events in a timely manner

Resolving the issues in a timely manner helps to get attention on the new and most critical events. If there are a number of unfixed alerts, fixing any one event might become more difficult because of the effects of the other events. You can use either the CLI or the GUI to view the list of issues that are reported in the system.

You can use the **mmhealth node eventlog** to list the events that are reported in the system.

The **Monitoring > Events** GUI page lists all events reported in the system. You can also mark certain events as read to change the status of the event in the events view. The status icons become gray in case an error or warning is fixed or if it is marked as read. Some issues can be resolved by running a fix procedure. Use the action **Run Fix Procedure** to do so. The **Events** page provides a recommendation for which fix procedure to run next.

Keep your software up to date

Check for new code releases and update your code on a regular basis.

This can be done by checking the IBM support website to see if new code releases are available: . The release notes provide information about new functions in a release plus any issues that are resolved with the new release. Update your code regularly if the release notes indicate a potential issue.

Note: If a critical problem is detected on the field, IBM may send a flash, advising the user to contact IBM for an efix. The efix when applied might resolve the issue.

Subscribe to the support notification

Subscribe to support notifications so that you are aware of best practices and issues that might affect your system.

Subscribe to support notifications by visiting the IBM support page on the following IBM website: <http://www.ibm.com/support/mynotifications>.

By subscribing, you are informed of new and updated support site information, such as publications, hints and tips, technical notes, product flashes (alerts), and downloads.

Know your IBM warranty and maintenance agreement details

If you have a warranty or maintenance agreement with IBM, know the details that must be supplied when you call for support.

For more information on the IBM Warranty and maintenance details, see [Warranties, licenses and maintenance](#).

Know how to report a problem

If you need help, service, technical assistance, or want more information about IBM products, then you can find a wide variety of sources available from IBM to assist you.

IBM maintains pages on the web where you can get information about IBM products and fee services, product implementation and usage assistance, break and fix service support, and the latest technical information. The following table provides the URLs of the IBM websites where you can find the support information.

Table 2. IBM websites for help, services, and information

Website	Address
IBM home page	http://www.ibm.com
Directory of worldwide contacts	http://www.ibm.com/planetwide
Support for ESS	IBM Elastic Storage® Server support website
Support for IBM System Storage and IBM Total Storage products	http://www.ibm.com/support/entry/portal/product/system_storage/

Note: Available services, telephone numbers, and web links are subject to change without notice.

Before you call

Make sure that you have taken steps to try to solve the problem yourself before you call. Some suggestions for resolving the problem before calling IBM Support include:

- Check all hardware for issues beforehand.
- Use the troubleshooting information in your system documentation. The troubleshooting section of the IBM Documentation contains procedures to help you diagnose problems.

To check for technical information, hints, tips, and new device drivers or to submit a request for information, go to the [IBM Elastic Storage Server support website](#).

Using the documentation

Information about your IBM storage system is available in the documentation that comes with the product. That documentation includes printed documents, online documents, readme files, and help files in addition to the IBM Documentation.

Chapter 6. Limitations

Read this section to learn about product limitations.

Limit updates to Red Hat Enterprise Linux (ESS 5.3)

Limit errata updates to RHEL to security updates and updates requested by IBM Service.

The required Red Hat components are:

- RHEL 7.3 ISO (PPC64BE and PPC64LE)
- Network manager version : 1.8.0-11.el7_4
- Systemd version: 219-42.el7_4.10
- Kernel version: 3.10.0-514.44

ESS 5.3 supports Red Hat Enterprise Linux 7.3 (3.10.0-514.44.1 ppc64BE and LE). It is highly recommended that you install only the following types of updates to RHEL:

- Security updates.
- Errata updates that are requested by IBM Service.

Chapter 7. Collecting information about an issue

To begin the troubleshooting process, collect information about the issue that the system is reporting.

From the EMS, issue the following command:

```
gsssnap -i -g -N <IO node1>,<IO node 2>,...,<IO node X>
```

The system will return a **gpfs.snap**, an **gssinstallcheck**, and the data from each node.

Chapter 8. GUI Issues

When troubleshooting GUI issues, it is recommended to view the logs that are located under `/var/log/cnlog/mgtsrv`. By default, the GUI is installed on the EMS node. It is possible that the customer installed it in another node. In such cases, the GUI logs are stored in the node where the GUI is installed.

The following logs can be viewed to troubleshoot the GUI issues:

mgtsrv-system-log

Logs everything that runs in background processes such as refresh tasks. This is the most important log for GUI.

mgtsrv-trace-log

Logs everything that is directly triggered by the GUI user. For example, starting an action, clicking a button, executing a GUI CLI command, etc.

wlp-messages.log

This log covers the underlying Websphere Liberty. The log is mostly relevant during the startup phase.

gpfsgui_trc.log

Logs the issues related to incoming requests from the browser. Users must check this log if the GUI displays the error message:

```
'Server was unable to process request.'
```

Issue with loading GUI

If there are problems in loading the GUI, you can reconfigure the GUI to see if that resolves the problem.

Follow these steps to reconfigure the GUI:

1. Run the following command to force the GUI to launch the wizard after the next login:

```
/usr/lpp/mmfs/gui/cli/debug enablewizard  
systemctl restart gpfsgui
```

2. Run the following command to force the GUI to no longer display the wizard after login:

```
/usr/lpp/mmfs/gui/cli/debug disablewizard  
systemctl restart gpfsgui
```

3. If the problem persists, reinstall the GUI RPM that can be found on the EMS node using the following command:

```
yum -Uvh /opt/ibm/gss/install/rhel7/<arch>/gui/gpfs.gui*
```

4. If there is a possibility that the GUI database has become corrupt or has inconsistencies that are preventing the GUI from loading properly, take the following steps.



CAUTION: This should be done as a last resort since the GUI configuration settings will be lost after you execute the following steps:

- a. Stop the GUI service.

```
systemctl stop gpfsgui
```

- b. Drop the GUI schema from the postgres database.

```
psql postgres postgres -c "DROP SCHEMA FSCC CASCADE"
```

c. Start the GUI service.

```
systemctl start gpfsGui
```

Chapter 9. Recovery Group Issues

Use the **mmlsrucoverygroup** command to check which recovery groups are available:

```
mmlsrucoverygroup
```

The command will give output similar to the following:

recovery group	vdisks	declustered arrays with vdisks	servers
rg_rchgss1-hs	3	5	rchgss1-hs.gpfs.rchland.ibm.com, rchgss2-hs.gpfs.rchland.ibm.com
rg_rchgss2-hs	3	5	rchgss2-hs.gpfs.rchland.ibm.com, rchgss1-hs.gpfs.rchland.ibm.com

List the active recovery group and their primary servers using the following command :

```
mmlsrucoverygroup rg_rchgss1-hs -L | grep active -A2
active recovery group server          servers
-----
rchgss1-hs.gpfs.rchland.ibm.com      rchgss1-hs.gpfs.rchland.ibm.com, rchgss2-hs.gpfs.rchland.ibm.com

mmlsrucoverygroup rg_rchgss2-hs -L | grep active -A2
active recovery group server          servers
-----
rchgss2-hs.gpfs.rchland.ibm.com      rchgss2-hs.gpfs.rchland.ibm.com, rchgss1-hs.gpfs.rchland.ibm.com
```

Each of the recovery groups must be served by its own server. If the server is unavailable due to maintenance or other issues, the recovery group must be served by an available server. After a failure or maintenance event, when the recovery group's primary server becomes active again, it must automatically begin serving its recovery group. You will find the following information under the `/var/adm/ras/mmfs.log.latest` file under in the recovery group server:

- Now serving recovery group rg_rchgss1-hs.
- Reason for takeover of rg_rchgss1-hs: 'primary server became ready'.

If the recovery group is not being served by its respective server, examine the gpfs log on that server for errors that might prevent the server from serving the recovery group. If there are no issues, you can manually activate the recovery group. For example, to allow `rchgss1-hs.gpfs.rchland.ibm.com` to serve the `rg_rchgss1-hs` RG, execute:

```
mmchrecoverygroup rg_rchgss1-hs --active rchgss1-hs.gpfs.rchland.ibm.com
```

If the recovery group is **mmvdisk** controlled, execute:

```
mmvdisk rg change --rg rg_rchgss1-hs --active rchgss1-hs.gpfs.rchland.ibm.com
```

In certain situations, if an ESS server node experiences a disk failure, the disks might be marked down, and does not automatically start. This can prevent the recovery group from becoming active. For more information on troubleshooting disk problems, see the *Disk Issues* section in the [IBM Spectrum Scale documentation](#).

Before troubleshooting further, ensure that GPFS is in the active state for the node in question by running the **mmgetstate** command:

```
mmgetstate -a
```

The command gives an output similar to the following:

Node number	Node name	GPFS state
-----	-----	-----

1	rchgss1-hs	active
2	rchgss2-hs	active
3	rchems1	active

Execute the **mm1sdisk** command to check the status of the disks. The -e option will only display disks with errors.

```
mm1sdisk gpfs0 -e
```

The command gives an output similar to the following:

disk name	driver type	sector size	failure group	holds metadata	holds data	status	availability	storage pool
rg_rchgss1_hs_MetaData_1M_3W_1	nsd	512	30	Yes	No	to be emptied	up	system



Attention: Due to an earlier configuration change the file system might contain data that is at risk of being lost.

In the previous example, the disk is in the suspended state, hence the to be emptied status. Other disks might be in the non-ready state or might be unavailable, so this prevents the disks from being used by GPFS or ESS.

disk name	driver type	sector size	failure group	holds metadata	holds data	status	availability	storage disk id pool	remarks
rg_rchgss1_hs_MetaData_1M_3W_1	nsd	512	30	Yes	No	ready	up	1	system
rg_rchgss1_hs_Data_16M_2p_1	nsd	512	30	No	Yes	ready	up	2	data desc
rg_rchgss2_hs_MetaData_1M_3W_1	nsd	512	30	Yes	No	ready	up	3	system
desc									
rg_rchgss2_hs_Data_16M_2p_1	nsd	512	30	No	Yes	ready	up	4	data desc

You can try to manually start the disks by running the **mmchdisk** command.

```
mmchdisk gpfs0 start -d rg_rchgss1_hs_MetaData_1M_3W_1
mmnsddiscover: Attempting to rediscover the disks. This may take a while ...
mmnsddiscover: Finished.
rchgss1-hs.gpfs.rchland.ibm.com: Rediscovered nsd server access to
rg_rchgss1_hs_MetaData_1M_3W_1
```

If multiple disks are down, you can run the command:

```
mmchdisk gpfs0 start -a
```

Note: Depending on the number of disks that are down and their size, the **mmnsddiscover** command might take a while to complete.

Chapter 10. Contacting IBM

Specific information about a problem such as: symptoms, traces, error logs, GPFS logs, and file system status is vital to IBM in order to resolve an IBM Spectrum Scale RAID problem.

Obtain this information as quickly as you can after a problem is detected, so that error logs will not wrap and system parameters that are always changing will be captured as close to the point of failure as possible. When a serious problem is detected, collect this information and then call IBM.

Information to collect before contacting the IBM Support Center

For effective communication with the IBM Support Center to help with problem diagnosis, you need to collect certain information.

Information to collect for all problems related to IBM Spectrum Scale RAID

Regardless of the problem encountered with IBM Spectrum Scale RAID, the following data should be available when you contact the IBM Support Center:

1. A description of the problem.
2. Output of the failing application, command, and so forth.

To collect the **gpfs.snap** data and the ESS tool logs, issue the following from the EMS:

```
gsssnap -g -i -n <IO node1>, <IOnode2>,... <ioNodeX>
```

3. A tar file generated by the **gpfs.snap** command that contains data from the nodes in the cluster. In large clusters, the **gpfs.snap** command can collect data from certain nodes (for example, the affected nodes, NSD servers, or manager nodes) using the **-N** option.

For more information about gathering data using the **gpfs.snap** command, see the *IBM Spectrum Scale: Problem Determination Guide*.

If the **gpfs.snap** command cannot be run, collect these items:

- a. Any error log entries that are related to the event:
 - On a Linux node, create a tar file of all the entries in the `/var/log/messages` file from all nodes in the cluster or the nodes that experienced the failure. For example, issue the following command to create a tar file that includes all nodes in the cluster:

```
mmdsh -v -N all "cat /var/log/messages" > all.messages
```

- On an AIX® node, issue this command:

```
errpt -a
```

For more information about the operating system error log facility, see the *IBM Spectrum Scale: Problem Determination Guide*.

- b. A master GPFS log file that is merged and chronologically sorted for the date of the failure. See the *IBM Spectrum Scale: Problem Determination Guide* for information about creating a master GPFS log file.
- c. If the cluster was configured to store dumps, collect any internal GPFS dumps written to that directory relating to the time of the failure. The default directory is `/tmp/mmfs`.
- d. On a failing Linux node, gather the installed software packages and the versions of each package by issuing this command:

```
rpm -qa
```

- e. On a failing AIX node, gather the name, most recent level, state, and description of all installed software packages by issuing this command:

```
lslpp -l
```

- f. For the file system attributes for all of the failing file systems, issue:

```
mmlsfs Device
```

- g. For the current configuration and state of the disks for all of the failing file systems, issue:

```
mmlsdisk Device
```

- h. A copy of file `/var/mmfs/gen/mmsdrfs` from the primary cluster configuration server.

4. If you are experiencing one of the following problems, see the appropriate section before contacting the IBM Support Center:

- For delay and deadlock issues, see [“Additional information to collect for delays and deadlocks” on page 28](#).
- For file system corruption or MMFS_FSSTRUCT errors, see [“Additional information to collect for file system corruption or MMFS_FSSTRUCT errors” on page 28](#).
- For GPFS daemon crashes, see [“Additional information to collect for GPFS daemon crashes” on page 29](#).

Additional information to collect for delays and deadlocks

When a delay or deadlock situation is suspected, the IBM Support Center will need additional information to assist with problem diagnosis. If you have not done so already, make sure you have the following information available before contacting the IBM Support Center:

1. Everything that is listed in [“Information to collect for all problems related to IBM Spectrum Scale RAID” on page 27](#).
2. The deadlock debug data collected automatically.
3. If the cluster size is relatively small and the `maxFilesToCache` setting is not high (less than 10,000), issue the following command:

```
gpfs.snap --deadlock
```

If the cluster size is large or the `maxFilesToCache` setting is high (greater than 1M), issue the following command:

```
gpfs.snap --deadlock --quick
```

For more information about the `--deadlock` and `--quick` options, see the *IBM Spectrum Scale: Problem Determination Guide*.

Additional information to collect for file system corruption or MMFS_FSSTRUCT errors

When file system corruption or MMFS_FSSTRUCT errors are encountered, the IBM Support Center will need additional information to assist with problem diagnosis. If you have not done so already, make sure you have the following information available before contacting the IBM Support Center:

1. Everything that is listed in [“Information to collect for all problems related to IBM Spectrum Scale RAID” on page 27](#).
2. Unmount the file system everywhere, then run `mmfsck -n` in offline mode and redirect it to an output file.

The IBM Support Center will determine when and if you should run the `mmfsck -y` command.

Additional information to collect for GPFS daemon crashes

When the GPFS daemon is repeatedly crashing, the IBM Support Center will need additional information to assist with problem diagnosis. If you have not done so already, make sure you have the following information available before contacting the IBM Support Center:

1. Everything that is listed in [“Information to collect for all problems related to IBM Spectrum Scale RAID”](#) on page 27.
2. Make sure the `/tmp/mmfs` directory exists on all nodes. If this directory does not exist, the GPFS daemon will not generate internal dumps.
3. Set the traces on this cluster and *all* clusters that mount any file system from this cluster:

```
mmtracectl --set --trace=def --trace-recycle=global
```

4. Start the trace facility by issuing:

```
mmtracectl --start
```

5. Recreate the problem if possible or wait for the assert to be triggered again.
6. Once the assert is encountered on the node, turn off the trace facility by issuing:

```
mmtracectl --off
```

If traces were started on multiple clusters, `mmtracectl --off` should be issued immediately on all clusters.

7. Collect `gpfs.snap` output:

```
gpfs.snap
```

How to contact the IBM Support Center

IBM support is available for various types of IBM hardware and software problems that IBM Spectrum Scale customers might encounter.

These problems include the following:

- IBM hardware failure
- Node halt or crash not related to a hardware failure
- Node hang or response problems
- Failure in other software supplied by IBM

If you have an IBM Software Maintenance service contract

If you have an IBM Software Maintenance service contract, contact IBM Support as follows:

Your location	Method of contacting IBM Support
In the United States	Call 1-800-IBM-SERV for support.
Outside the United States	Contact your local IBM Support Center or see the Directory of worldwide contacts (www.ibm.com/planetwide) .

When you contact IBM Support, the following will occur:

1. You will be asked for the information you collected in [“Information to collect before contacting the IBM Support Center”](#) on page 27.
2. You will be given a time period during which an IBM representative will return your call. Be sure that the person you identified as your contact can be reached at the phone number you provided in the PMR.

3. An online Problem Management Record (PMR) will be created to track the problem you are reporting, and you will be advised to record the PMR number for future reference.
4. You might be requested to send data related to the problem you are reporting, using the PMR number to identify it.
5. Should you need to make subsequent calls to discuss the problem, you will also use the PMR number to identify the problem.

If you do not have an IBM Software Maintenance service contract

If you do not have an IBM Software Maintenance service contract, contact your IBM sales representative to find out how to proceed. Be prepared to provide the information you collected in [“Information to collect before contacting the IBM Support Center”](#) on page 27.

For failures in non-IBM software, follow the problem-reporting procedures provided with that product.

Chapter 11. Maintenance procedures

Very large disk systems, with thousands or tens of thousands of disks and servers, will likely experience a variety of failures during normal operation.

To maintain system productivity, the vast majority of these failures must be handled automatically without loss of data, without temporary loss of access to the data, and with minimal impact on the performance of the system. Some failures require human intervention, such as replacing failed components with spare parts or correcting faults that cannot be corrected by automated processes.

You can also use the ESS GUI to perform various maintenance tasks. The ESS GUI lists various maintenance-related events in its event log in the **Monitoring > Events** page. You can set up email alerts to get notified when such events are reported in the system. You can resolve these events or contact the IBM Support Center for help as needed. The ESS GUI includes various maintenance procedures to guide you through the fix process.

Updating the firmware for host adapters, enclosures, and drives

The following section describes the procedure to update the firmware for host adapters, enclosures, and drives.

After you create a GPFS cluster, install the most current firmware for host adapters, enclosures, and drives only if instructed to do so by IBM support.

You can update the firmware either manually or with the help of directed maintenance procedures (DMP) that are available in the GUI. The ESS GUI lists events in its event log in the **Monitoring > Events** page if the host adapter, enclosure, or drive firmware is not up-to-date, compared to the firmware packages on the servers that are currently available. Select **Action > Run Fix Procedure** for the firmware-related event to start the corresponding DMP in the GUI. For more information on the available DMPs, see *Directed maintenance procedures* in the *Elastic Storage Server: Problem Determination Guide*.

The most current firmware is packaged as the `gpfs.gss.firmware` RPM. You can find the most current firmware on [Fix Central](#).

1. Sign in with your IBM ID and password.
2. On the **Find product** tab:
 - a. In the **Product selector** field, type: IBM Spectrum Scale RAID, and click the right arrow.
 - b. On the **Installed Version** menu, select: 5.0.0
 - c. On the **Platform** menu, select: Linux 64-bit,pSeries.
 - d. Click **Continue**.
3. On the **Select fixes** page, select the most current fix pack.
4. Click **Continue**.
5. On the **Download options** page, select your preferred downloading method. Make sure the check box to the left of Include prerequisites and co-requisite fixes (you can select the ones you need later) has a check mark in it.
6. Click **Continue** to go to the **Continue** page and download the fix pack files.

The `gpfs.gss.firmware` RPM needs to be installed on all ESS server nodes. It contains the most current updates of the following types of supported firmware for an ESS configuration:

- Host adapter firmware
- Enclosure firmware
- Drive firmware
- Firmware loading tools

For command syntax and examples, see the *mmchfirmware* command in *IBM Spectrum Scale RAID: Administration*.

Disk diagnosis

For information about disk hospital, see *Disk hospital* in *IBM Spectrum Scale RAID: Administration*.

When an individual disk I/O operation (read or write) encounters an error, IBM Spectrum Scale RAID completes the NSD client request by reconstructing the data (for a read) or by marking the unwritten data as stale and relying on successfully written parity or replica strips (for a write), and starts the disk hospital to diagnose the disk. While the disk hospital is diagnosing, the affected disk will not be used for serving NSD client requests.

Similarly, if an I/O operation does not complete in a reasonable time period, it is timed out, and the client request is treated just like an I/O error. Again, the disk hospital will diagnose what went wrong. If the timed-out operation is a disk write, the disk remains temporarily unusable until a pending timed-out write (PTOW) completes.

The disk hospital then determines the exact nature of the problem. If the cause of the error was an actual media error on the disk, the disk hospital marks the offending area on disk as temporarily unusable, and overwrites it with the reconstructed data. This cures the media error on a typical HDD by relocating the data to spare sectors reserved within that HDD.

If the disk reports that it can no longer write data, the disk is marked as `readonly`. This can happen when no spare sectors are available for relocating in HDDs, or the flash memory write endurance in SSDs was reached. Similarly, if a disk reports that it cannot function at all, for example not spin up, the disk hospital marks the disk as dead.

The disk hospital also maintains various forms of *disk badness*, which measure accumulated errors from the disk, and of *relative performance*, which compare the performance of this disk to other disks in the same declustered array. If the badness level is high, the disk can be marked dead. For less severe cases, the disk can be marked `failing`.

Finally, the IBM Spectrum Scale RAID server might lose communication with a disk. This can either be caused by an actual failure of an individual disk, or by a fault in the disk interconnect network. In this case, the disk is marked as `missing`. If the relative performance of a disk falls below a particular threshold, the disk is declared as `slow` in the `pdisk` state, and the disk is prepared for replacement. To check the current value, run the **`mmlsconfig nsdRAIDDiskPerformanceMinLimitPct`** command.

If a disk would have to be marked dead, `missing`, or `readonly`, and the problem affects individual disks only (not a large set of disks), the disk hospital tries to recover the disk. If the disk reports that it is not started, the disk hospital attempts to start the disk. If nothing else helps, the disk hospital power-cycles the disk (assuming the JBOD hardware supports that), and then waits for the disk to return online.

Before actually reporting an individual disk as `missing`, the disk hospital starts a search for that disk by polling all disk interfaces to locate the disk. Only after that fast poll fails is the disk actually declared `missing`.

If a large set of disks has faults, the IBM Spectrum Scale RAID server can continue to serve read and write requests, provided that the number of failed disks does not exceed the fault tolerance of either the RAID code for the vdisk or the IBM Spectrum Scale RAID vdisk configuration data. When any disk fails, the server begins rebuilding its data onto spare space. If the failure is not considered *critical*, rebuilding is throttled when user workload is present. This ensures that the performance impact to user workload is minimal. A failure might be considered critical if a vdisk has no remaining redundancy information, for example three disk faults for 4-way replication and $8 + 3p$ or two disk faults for 3-way replication and $8 + 2p$. During a critical failure, critical rebuilding will run as fast as possible because the vdisk is in imminent danger of data loss, even if that impacts the user workload. Because the data is declustered, or spread out over many disks, and all disks in the declustered array participate in rebuilding, a vdisk will remain in critical rebuild only for short periods of time (several minutes for a typical system). A double or triple fault is extremely rare, so the performance impact of critical rebuild is minimized.

In a multiple fault scenario, the server might not have enough disks to fulfill a request. More specifically, such a scenario occurs if the number of unavailable disks exceeds the fault tolerance of the RAID code. If some of the disks are only temporarily unavailable, and are expected back online soon, the server will stall the client I/O and wait for the disk to return to service. Disks can be temporarily unavailable for any of the following reasons:

- The disk hospital is diagnosing an I/O error.
- A timed-out write operation is pending.
- A user intentionally suspended the disk, perhaps it is on a carrier with another failed disk that has been removed for service.

If too many disks become unavailable for the primary server to proceed, it will fail over. In other words, the whole recovery group is moved to the backup server. If the disks are not reachable from the backup server either, then all vdisks in that recovery group become unavailable until connectivity is restored.

A vdisk will suffer data loss when the number of permanently failed disks exceeds the vdisk fault tolerance. This data loss is reported to NSD clients when the data is accessed.

Background tasks

While IBM Spectrum Scale RAID primarily performs NSD client read and write operations in the foreground, it also performs several long-running maintenance tasks in the background, which are referred to as *background tasks*.

The background task that is currently in progress for each declustered array is reported in the long-form output of the **mmlsrecoverygroup** command. [Table 3 on page 33](#) describes the long-running background tasks.

Table 3. Background tasks

Task	Description
repair-RGD/VCD	Repairing the internal recovery group data and vdisk configuration data from the failed disk onto the other disks in the declustered array.
rebuild-critical	Rebuilding virtual tracks that cannot tolerate any more disk failures.
rebuild-1r	Rebuilding virtual tracks that can tolerate only one more disk failure.
rebuild-2r	Rebuilding virtual tracks that can tolerate two more disk failures.
rebuild-offline	Rebuilding virtual tracks where failures exceeded the fault tolerance.
rebalance	Rebalancing the spare space in the declustered array for either a missing pdisk that was discovered again, or a new pdisk that was added to an existing array.
scrub	Scrubbing vdisks to detect any silent disk corruption or latent sector errors by reading the entire virtual track, performing checksum verification, and performing consistency checks of the data and its redundancy information. Any correctable errors found are fixed.

Server failover

If the primary IBM Spectrum Scale RAID server loses connectivity to a sufficient number of disks, the recovery group attempts to fail over to the backup server.

If the backup server is also unable to connect, the recovery group becomes unavailable until connectivity is restored. If the backup server had taken over, it will relinquish the recovery group to the primary server when it becomes available again.

Data checksums

IBM Spectrum Scale RAID stores checksums of the data and redundancy information on all disks for each vdisk. Whenever data is read from disk or received from an NSD client, checksums are verified. If the checksum verification on a data transfer to or from an NSD client fails, the data is retransmitted. If the checksum verification fails for data read from disk, the error is treated similarly to a media error:

- The data is reconstructed from redundant data on other disks.
- The data on disk is rewritten with reconstructed good data.
- The disk badness is adjusted to reflect the silent read error.

Disk replacement

You can use the ESS GUI for detecting failed disks and for disk replacement.

When one disk fails, the system rebuilds the data that was on the failed disk onto spare space and continues to operate normally. However, the performance is slightly reduced because the same workload is shared among fewer disks. With the default setting of two spare disks for each large declustered array, failure of a single disk would typically not be a sufficient reason for maintenance.

The system might continue to operate in the presence of several disk failures even if there is no remaining spare space as long as the number of failures is less than the configured vdisk fault tolerance. The next disk failure would make the system unable to maintain the redundancy that the user requested during vdisk creation. A service request is sent to a maintenance management application that requests replacement of the failed disks, and specifies the disk FRU numbers and locations.

Call home for disk maintenance is requested when the number of failed disks in a declustered array reaches the disk replacement threshold. By default, the replace threshold is one if the number of data spares is zero or one, or two if the number of spares is two or greater. The maximum value is one more than the number of spares.

Disk maintenance is done by using the **mmvdisk pdisk replace** with the **--prepare** option for mmvdisk-managed recovery groups, or by using the **mmchcarrier** command with the **--release** option for non-mmvdisk recovery groups, which:

- Suspends any functioning disks on the carrier if the multi-disk carrier is shared with the disk that is being replaced.
- If possible, powers down the disk to be replaced or all of the disks on that carrier.
- Turns on indicators on the disk enclosure and disk or carrier to help locate and identify the disk that needs to be replaced.
- If necessary, unlocks the carrier for disk replacement.

After the disk is replaced and the carrier is reinserted, **mmvdisk pdisk replace** for mmvdisk-managed recovery groups or the **mmchcarrier** command with the **--replace** option for non-mmvdisk recovery groups powers on the disks.

You can replace the disk either manually or with the help of directed maintenance procedures (DMP) that are available in the GUI. The ESS GUI lists events in its event log in the **Monitoring > Events** page if a disk failure is reported in the system. Select the *gnr_pdisk_replaceable* event from the list of events, and then select **Action > Run Fix Procedure** from the menu to start the `replace disk` DMP in the GUI. For more information, see [“Replace disks” on page 64](#).

Replacing failed disks in an ESS recovery group: a sample scenario

The scenario presented here shows how to detect and replace failed disks in a recovery group built on an ESS building block.

Detecting failed disks in your ESS enclosure

Assume a GL4 building block on which the following two recovery groups are defined:

- BB1RGL, containing the disks in the left side of each drawer
- BB1RGR, containing the disks in the right side of each drawer

Each recovery group contains the following:

- One log declustered array (LOG)
- Two data declustered arrays (DA1, DA2)

The data declustered arrays are defined according to GL4 best practices as follows:

- 58 pdisks per data declustered array
- Default disk replacement threshold value set to 2

The replacement threshold of 2 means that IBM Spectrum Scale RAID only requires disk replacement when two or more disks fail in the declustered array; otherwise, rebuilding onto spare space or reconstruction from redundancy is used to supply affected data. This configuration can be seen in the output of `mmfsrecoverygroup` for the recovery groups, which are shown here for BB1RGL:

```
# mmfsrecoverygroup BB1RGL -L
```

recovery group	declustered arrays	vdisks	pdisks	format version
BB1RGL	4	8	119	4.1.0.1

declustered array	needs service	vdisks	pdisks	spares	replace threshold	free space	scrub duration	background activity task	progress	priority
SSD	no	1	1	0,0	1	186 GiB	14 days	scrub	8%	low
NVR	no	1	2	0,0	1	3648 MiB	14 days	scrub	8%	low
DA1	no	3	58	2,31	2	50 TiB	14 days	scrub	7%	low
DA2	no	3	58	2,31	2	50 TiB	14 days	scrub	7%	low

vdisk	RAID code	declustered array	vdisk size	block size	checksum granularity	state	remarks
ltip_BB1RGL	2WayReplication	NVR	48 MiB	2 MiB	512	ok	logTip
ltbackup_BB1RGL	Unreplicated	SSD	48 MiB	2 MiB	512	ok	logTipBackup
lhome_BB1RGL	4WayReplication	DA1	20 GiB	2 MiB	512	ok	log
reserved1_BB1RGL	4WayReplication	DA2	20 GiB	2 MiB	512	ok	logReserved
BB1RGLMETA1	4WayReplication	DA1	750 GiB	1 MiB	32 KiB	ok	
BB1RGLDATA1	8+3p	DA1	35 TiB	16 MiB	32 KiB	ok	
BB1RGLMETA2	4WayReplication	DA2	750 GiB	1 MiB	32 KiB	ok	
BB1RGLDATA2	8+3p	DA2	35 TiB	16 MiB	32 KiB	ok	

config data	declustered array	VCD spares	actual rebuild spare space	remarks
rebuild space	DA1	31	35 pdisk	
rebuild space	DA2	31	35 pdisk	

config data	max disk group fault tolerance	actual disk group fault tolerance	remarks
rg descriptor	1 enclosure + 1 drawer	1 enclosure + 1 drawer	limiting fault tolerance
system index	2 enclosure	1 enclosure + 1 drawer	limited by rg descriptor

vdisk	max disk group fault tolerance	actual disk group fault tolerance	remarks
ltip_BB1RGL	1 pdisk	1 pdisk	
ltbackup_BB1RGL	0 pdisk	0 pdisk	
lhome_BB1RGL	3 enclosure	1 enclosure + 1 drawer	limited by rg descriptor
reserved1_BB1RGL	3 enclosure	1 enclosure + 1 drawer	limited by rg descriptor
BB1RGLMETA1	3 enclosure	1 enclosure + 1 drawer	limited by rg descriptor
BB1RGLDATA1	1 enclosure	1 enclosure	
BB1RGLMETA2	3 enclosure	1 enclosure + 1 drawer	limited by rg descriptor
BB1RGLDATA2	1 enclosure	1 enclosure	

active recovery group server	servers
c45f01n01-ib0.gpfs.net	c45f01n01-ib0.gpfs.net, c45f01n02-ib0.gpfs.net

The indication that disk replacement is called for in this recovery group is the value of yes in the needs service column for declustered array DA1.

The fact that DA1 is undergoing rebuild of its IBM Spectrum Scale RAID tracks that can tolerate one strip failure is by itself not an indication that disk replacement is required; it merely indicates that data from a failed disk is being rebuilt onto spare space. Only if the replacement threshold has been met will disks be marked for replacement and the declustered array marked as needing service.

IBM Spectrum Scale RAID provides several indications that disk replacement is required:

- Entries in the Linux syslog
- The `pdReplacePdisk` callback, which can be configured to run an administrator-supplied script at the moment a pdisk is marked for replacement
- The output from the following commands, which may be performed from the command line on any IBM Spectrum Scale RAID cluster node (see the examples that follow):
 1. `mmlsrecoverygroup` with the `-L` flag shows `yes` in the `needs service` column
 2. `mmlsrecoverygroup` with the `-L` and `--pdisk` flags; this shows the states of all pdisks, which may be examined for the `replace pdisk` state
 3. `mmlspdisk` with the `--replace` flag, which lists only those pdisks that are marked for replacement

Note: Because the output of `mmlsrecoverygroup -L --pdisk` is long, this example shows only some of the pdisks (but includes those marked for replacement).

```
# mmlsrecoverygroup BB1RGL -L --pdisk
```

recovery group	declustered arrays	vdisks	pdisks									
BB1RGL	3	5	119									
declustered array priority	needs service	vdisks	pdisks	spares	replace threshold	free space	scrub duration	background task	activity progress			
LOG	no	1	3	0	1	534 GiB	14 days	scrub		1%	low	
DA1	yes	2	58	2	2	0 B	14 days	rebuild-1r		4%	low	
DA2	no	2	58	2	2	1024 MiB	14 days	scrub		27%	low	
pdisk	n. active, total paths	declustered array	free space	user condition	state, remarks							
[...]												
e1d4s06	2, 4	DA1	62 GiB	normal	ok							
e1d5s01	0, 0	DA1	70 GiB	replaceable	slow/noPath/systemDrain/noRGD/noVCD/replace							
e1d5s02	2, 4	DA1	64 GiB	normal	ok							
e1d5s03	2, 4	DA1	63 GiB	normal	ok							
e1d5s04	0, 0	DA1	64 GiB	replaceable	failing/noPath/systemDrain/noRGD/noVCD/replace							
e1d5s05	2, 4	DA1	63 GiB	normal	ok							
[...]												

The preceding output shows that the following pdisks are marked for replacement:

- e1d5s01 in DA1
- e1d5s04 in DA1

The naming convention used during recovery group creation indicates that these disks are in Enclosure 1 Drawer 5 Slot 1 and Enclosure 1 Drawer 5 Slot 4. To confirm the physical locations of the failed disks, use the `mmlspdisk` command to list information about the pdisks in declustered array DA1 of recovery group BB1RGL that are marked for replacement:

```
# mmlspdisk BB1RGL --declustered-array DA1 --replace
pdisk:
  replacementPriority = 0.98
  name = "e1d5s01"
  device = ""
  recoveryGroup = "BB1RGL"
  declusteredArray = "DA1"
  state = "slow/noPath/systemDrain/noRGD/noVCD/replace"
  .
  .
```

```
.
pdisk:
  replacementPriority = 0.98
  name = "e1d5s04"
  device = ""
  recoveryGroup = "BB1RGL"
  declusteredArray = "DA1"
  state = "failing/noPath/systemDrain/noRGD/noVCD/replace"
.
.
.
```

The physical locations of the failed disks are confirmed to be consistent with the pdisk naming convention and with the IBM Spectrum Scale RAID component database:

Disk	Location	User Location
pdisk e1d5s01	SV21314035-5-1	Rack BB1 U01-04, Enclosure BB1ENC1 Drawer 5 Slot 1
pdisk e1d5s04	SV21314035-5-4	Rack BB1 U01-04, Enclosure BB1ENC1 Drawer 5 Slot 4

This shows how the component database provides an easier-to-use location reference for the affected physical disks. The pdisk name e1d5s01 means "Enclosure 1 Drawer 5 Slot 1." Additionally, the location provides the serial number of enclosure 1, SV21314035, with the drawer and slot number. But the user location that has been defined in the component database can be used to precisely locate the disk in an equipment rack and a named disk enclosure: This is the disk enclosure that is labeled "BB1ENC1," found in compartments U01 - U04 of the rack labeled "BB1," and the disk is in drawer 5, slot 1 of that enclosure.

The relationship between the enclosure serial number and the user location can be seen with the `mmfscmp` command:

```
# mmfscmp --serial-number SV21314035

Storage Enclosure Components

Comp ID  Part Number  Serial Number  Name      Display ID
-----  -
      2  1818-80E      SV21314035    BB1ENC1
```

Replacing failed disks in a GL4 recovery group

Note: In this example, it is assumed that two new disks with the appropriate Field Replaceable Unit (FRU) code, as indicated by the fru attribute (90Y8597 in this case), have been obtained as replacements for the failed pdisks e1d5s01 and e1d5s04.

Replacing each disk is a three-step process:

1. Using the `mmchcarrier` command with the `--release` flag to inform IBM Spectrum Scale to locate the disk, suspend it, and allow it to be removed.
2. Locating and removing the failed disk and replacing it with a new one.
3. Using the `mmchcarrier` command with the `--replace` flag to begin use of the new disk.

IBM Spectrum Scale RAID assigns a priority to pdisk replacement. Disks with smaller values for the `replacementPriority` attribute should be replaced first. In this example, the only failed disks are in DA1 and both have the same `replacementPriority`.

Disk e1d5s01 is chosen to be replaced first.

1. To release pdisk e1d5s01 in recovery group BB1RGL:

```
# mmchcarrier BB1RGL --release --pdisk e1d5s01
[I] Suspending pdisk e1d5s01 of RG BB1RGL in location SV21314035-5-1.
[I] Location SV21314035-5-1 is Rack BB1 U01-04, Enclosure BB1ENC1 Drawer 5 Slot 1.
[I] Carrier released.

- Remove carrier.
```

- Replace disk in location SV21314035-5-1 with FRU 90Y8597.
- Reinsert carrier.
- Issue the following command:

```
mmchcarrier BB1RGL --replace --pdisk 'e1d5s01'
```

IBM Spectrum Scale RAID issues instructions as to the physical actions that must be taken, and repeats the user-defined location to help find the disk.

2. To allow the enclosure BB1ENC1 with serial number SV21314035 to be located and identified, IBM Spectrum Scale RAID will turn on the enclosure's amber "service required" LED. The enclosure's bezel should be removed. This will reveal that the amber "service required" and blue "service allowed" LEDs for drawer 5 have been turned on.

Drawer 5 should then be unlatched and pulled open. The disk in slot 1 will be seen to have its amber and blue LEDs turned on.

Unlatch and pull up the handle for the identified disk in slot 1. Lift out the failed disk and set it aside. The drive LEDs turn off when the slot is empty. A new disk with FRU 90Y8597 should be lowered in place and have its handle pushed down and latched.

Since the second disk replacement in this example is also in drawer 5 of the same enclosure, leave the drawer open and the enclosure bezel off. If the next replacement were in a different drawer, the drawer would be closed; and if the next replacement were in a different enclosure, the enclosure bezel would be replaced.

3. To finish the replacement of pdisk e1d5s01:

```
# mmchcarrier BB1RGL --replace --pdisk e1d5s01
[I] The following pdisks will be formatted on node server1:
    /dev/sdmi
[I] Pdisk e1d5s01 of RG BB1RGL successfully replaced.
[I] Resuming pdisk e1d5s01#026 of RG BB1RGL.
[I] Carrier resumed.
```

When the `mmchcarrier --replace` command returns successfully, IBM Spectrum Scale RAID begins rebuilding and rebalancing IBM Spectrum Scale RAID strips onto the new disk, which assumes the pdisk name `e1d5s01`. The failed pdisk may remain in a temporary form (indicated here by the name `e1d5s01#026`) until all data from it rebuilds, at which point it is finally deleted. Notice that only one block device name is mentioned as being formatted as a pdisk; the second path will be discovered in the background.

Disk `e1d5s04` is still marked for replacement, and `DA1` of `BBRGL` will still need service. This is because the IBM Spectrum Scale RAID replacement policy expects all failed disks in the declustered array to be replaced after the replacement threshold is reached.

Pdisk `e1d5s04` is then replaced following the same process.

1. Release pdisk `e1d5s04` in recovery group `BB1RGL`:

```
# mmchcarrier BB1RGL --release --pdisk e1d5s04
[I] Suspending pdisk e1d5s04 of RG BB1RGL in location SV21314035-5-4.
[I] Location SV21314035-5-4 is Rack BB1 U01-04, Enclosure BB1ENC1 Drawer 5 Slot 4.
[I] Carrier released.

- Remove carrier.
- Replace disk in location SV21314035-5-4 with FRU 90Y8597.
- Reinsert carrier.
- Issue the following command:

mmchcarrier BB1RGL --replace --pdisk 'e1d5s04'
```

2. Find the enclosure and drawer, unlatch and remove the disk in slot 4, place a new disk in slot 4, push in the drawer, and replace the enclosure bezel.
3. To finish the replacement of pdisk `e1d5s04`:

```
# mmchcarrier BB1RGL --replace --pdisk e1d5s04
[I] The following pdisks will be formatted on node server1:
```

```

/dev/sdfd
[I] Pdisk e1d5s04 of RG BB1RGL successfully replaced.
[I] Resuming pdisk e1d5s04#029 of RG BB1RGL.
[I] Carrier resumed.

```

The disk replacements can be confirmed with `mmlsrecoverygroup -L --pdisk`:

```
# mmlsrecoverygroup BB1RGL -L --pdisk
```

recovery group	declustered arrays	vdisks	pdisks
BB1RGL	3	5	121

declustered array	needs service	vdisks	pdisks	spares	replace threshold	free space	scrub duration	background task	activity progress	priority
LOG	no	1	3	0	1	534 GiB	14 days	scrub	1%	low
DA1	no	2	60	2	2	3647 GiB	14 days	rebuild-1r	4%	low
DA2	no	2	58	2	2	1024 MiB	14 days	scrub	27%	low

pdisk	n. active, total paths	declustered array	free space	user condition	state, remarks
[...]					
e1d4s06	2, 4	DA1	62 GiB	normal	ok
e1d5s01	2, 4	DA1	1843 GiB	normal	ok
e1d5s01#026	0, 0	DA1	70 GiB	draining	slow/noPath/systemDrain/adminDrain/noRGD/noVCD
e1d5s02	2, 4	DA1	64 GiB	normal	ok
e1d5s03	2, 4	DA1	63 GiB	normal	ok
e1d5s04	2, 4	DA1	1853 GiB	normal	ok
e1d5s04#029	0, 0	DA1	64 GiB	draining	failing/noPath/systemDrain/adminDrain/noRGD/noVCD
e1d5s05	2, 4	DA1	62 GiB	normal	ok
[...]					

Notice that the temporary pdisks (e1d5s01#026 and e1d5s04#029) representing the now-removed physical disks are counted toward the total number of pdisks in the recovery group BB1RGL and the declustered array DA1. They exist until IBM Spectrum Scale RAID rebuild completes the reconstruction of the data that they carried onto other disks (including their replacements). When rebuild completes, the temporary pdisks disappear, and the number of disks in DA1 will once again be 58, and the number of disks in BBRGL will once again be 119.

Using the `mmchcarrier` command to fix issues caused by improper disk removal

Pdisks are identified by descriptors written onto the disks, not by their physical locations. If a pdisk is moved to a different enclosure slot, the system still correctly identifies the pdisk and continues to use it. In general, the system cannot prevent an operator from swapping disks between slots. Continuing to use a disk found in an unexpected location avoids risk of data unavailability.

The location code associated with a pdisk reflects the enclosure slot where the pdisk was last seen. Thus, if a pdisk is moved to a different slot, the system automatically updates the location code to reflect where it currently is.

There are only two ways a location code can be empty:

- The location is unknown since the time of installation.
- The pdisk was removed; another pdisk from the same GNR recovery group pair was inserted into the slot, and the new pdisk took over the location.

Devices such as logtip disks might not have location codes and can fall into the first case. But devices in external enclosures that automatically detect the location are not likely to have been forever blank. Blank location codes on these disks, therefore, suggest that disks have been pulled out and other disks from the same recovery group pair have been placed into their slots.

The user location code comes from a table in the **mmcomp** database that maps location code to user location code. A blank user location might indicate a blank location code as mentioned above, or it may indicate a missing row in the table. Verify that the regular location code is also blank.

Test case of issues caused due to improper disk removal

Consider a situation where the pdisk e2s046 has failed. The admin runs the **mmchcarrier rg_alpine-nsds4b2-bond1--release --pdisk e2s046** command, and removes the bad drive. The system is now expecting a new disk to be inserted. However, instead of inserting a new disk, the admin pulls pdisk e2s047 from one slot over, inserts it into slot 46, then runs the **mmchcarrier rg_alpine-nsds4b2-bond1--replace --pdisk e2s046** command. The replace command detects what happened and fails, and displays the following error message:

```
[E] Pdisk e2s047 of recovery group rg_alpine-nsds4b2-bond1 in location
78T02D4-46 cannot be used as a replacement for pdisk e2s046 of recovery group
rg_alpine-nsds4b2-bond1.
```

But because e2s047 now occupies the slot, it has taken on the location code 78T02D4-46, clearing it from pdisk e2s046. The system no longer knows the location e2s046; it just knows that the location is not 78T02D4-46. Even, if at this point, the admin realizes the mistake and moves e2s047 back into slot 47, e2s047's location is updated to slot 47, but e2s046's location remains blank.

Solution

You can put the disks back into the right slot and solve this issue in case the following criteria are met:

- You have all the drives.
- All the drives are functional and the system can read the descriptors from them.
- You have not used dd or other tools to clear the descriptors.

When the system discovers the disks, it automatically updates the location codes. After the location codes are updated, replace any bad disks using the **mmchcarrier** command. In order to pull a drive that is in the wrong slot, use the **mmchpdisk RNAME --pdisk PDNAME --suspend** command to quiesce the disk before pulling it. Run the **mmchpdisk RNAME --pdisk PDNAME --resume** command after reinserting the disk. Suspending the disk before pulling it avoids unnecessary I/O errors and the risk of causing a recovery group resign.

If some of the disks are no longer available or the descriptors are unreadable, then you can use the `replace-at-location` script to replace them. This script is found in `/usr/lpp/mmfs/samples/vdisk` as shown:

1. Insert a new, blank disk into the empty slot 46 where the bad e2s046 drive was.
2. Run `replace-at-location rg_alpine-nsds4b2-bond1 e2s046 78T02D4-46`.

Commandless disk replacement

Commandless disk replacement automates the process of replacing a failed or bad drive with a new drive.

The commandless disk replacement feature helps in automating the process of replacing a failed drive with a new drive. The disk hospital begins moving data off that drive whenever the drive fails in preparation for replacement. When data is drained off the drive, or if the drive is undrainable or dead, the disk hospital marks the drive as `replaceable`. Then, the commandless disk replacement component runs the `prepare for replacement` operation on the drive. As a part of the prepare operation, the LED on the drive for replacement is turned on, indicating that it is ready for removal.

If the prepare operation fails, then use commands like the **mmvdisk pdisk replace** for mmvdisk-managed recovery groups or **mmchcarrier** for non-mmvdisk recovery groups to replace the bad drive manually. For more information, see [Disk Replacement](#).

Identify the bad drive that needs to be replaced and verify that the `replace` LED is turned on. Next, remove the bad drive, and insert a new drive. The new drive is identified by commandless disk replacement within minutes, as defined by the configuration keyword

`nsdRAIDDiskDiscoveryInterval`. Commandless disk replacement then runs the `replace` operation. This operation makes the new drive ready for use. If it runs successfully without any errors, then no further action is required.

If the `replace` operation on a new drive fails for any reason, error messages, indicating the failure, are logged in the `mmfs.log` file. You can then replace the drive manually by following the manual disk replacement procedure. For more information, see [Disk Replacement](#).

Replacing bad drives with new drives by using commandless disk replacement

The following section describes the procedure to replace bad drives with new drives by using commandless disk replacement.

You can replace bad drives with good drives by using the following commandless disk replacement process.

1. To enable commandless disk replacement, use the **`mmchconfig enableAutomaticDiskReplacement=yes -i`** command.
Note: This step needs to be performed only once.
2. Wait for any drive to fail and become replaceable.
3. List out all failed drives by using the **`mmvdisk pdisk list --recovery-group all --replace`** command.
4. When a drive fails and is ready to be replaced, the `release` operation runs automatically.
5. When the drive becomes replaceable, two things happen:
 - The replace LED is turned on.
 - The following message is logged in the `mmfs.log` file:

```
[I] Automatic Pdisk Release for pdisk:e1s19 in RG:BB01L succeeded
```

Note:

- If the automatic release on a bad drive fails, then an error message is logged in the `mmfs.log` file. Also, the replace LED is not turned on.

The following message is logged in the `mmfs.log` file:

```
[E] Automatic Pdisk Release for pdisk:e1s19 in RG:BB01L failed with err:<Error code>
```

If the automatic `release` fails, then follow the manual procedure to replace the failed drive. For more information, see [Disk Replacement](#).

6. When the failed drive is identified and its replace LED is turned on, remove the failed drive and insert the new drive in the same slot.
7. When the new drive is successfully accepted, data is rebalanced onto the new drive automatically. On the successful acceptance of the new drive, the following events occur:
 - The replace LED is turned off.
 - The following message is logged to `mmfs.log` file:

```
[I] Automatic Pdisk Replace for pdisk:e1s19 in RG:BB01L succeeded
```

If the automatic `replace` on the newly replaced drive fails, the following error message is logged in `mmfs.log` file, and the replace LED is not turned off.

```
[E] Automatic Pdisk Replace for pdisk:e1s19 in RG:BB01L failed with err:<Error code>
```

If the new drive is not accepted successfully, then follow the manual procedure to replace the failed drive. For more information, see [Disk Replacement](#).

Note:

You might need to wait for the maximum time, in minutes, defined by the configuration keyword `nsdRAIDDiskDiscoveryInterval` before the replacement drive is discovered and the replace LED is turned off.

The command **`mmvdisk pdisk list --recovery-group all --replace`** can be used to verify that the replaced drive is no longer in the list.

Replacing failed ESS storage enclosure components: a sample scenario

The scenario presented here shows how to detect and replace failed storage enclosure components in an ESS building block.

Detecting failed storage enclosure components

The `mmIsenclosure` command can be used to show you which enclosures need service along with the specific component. A best practice is to run this command every day to check for failures.

```
# mmIsenclosure all -L --not-ok
```

serial number	needs service	nodes
SV21313971	yes	c45f02n01-ib0.gpfs.net

component type	serial number	component id	failed	value	unit	properties
fan	SV21313971	1_BOT_LEFT	yes		RPM	FAILED

This indicates that enclosure SV21313971 has a failed fan.

When you are ready to replace the failed component, use the **`mmchenclosure`** command to identify whether it is safe to complete the repair action, or whether IBM Spectrum Scale needs to be shut down first:

```
# mmchenclosure SV21313971 --component fan --component-id 1_BOT_LEFT
mmenclosure: Proceed with the replace operation.
```

The fan can now be replaced.

Special note about detecting failed enclosure components

In the following example, only the enclosure itself is being called out as having failed; the specific component that has actually failed is not identified. This typically means that there are drive "Service Action Required (Fault)" LEDs that have been turned on in the drawers. In such a situation, the `mmIsdisk all --not-ok` command can be used to check for dead or failing disks.

```
mmIsenclosure all -L --not-ok
```

serial number	needs service	nodes
SV13306129	yes	c45f01n01-ib0.gpfs.net

component type	serial number	component id	failed	value	unit	properties
enclosure	SV13306129	ONLY	yes			NOT_IDENTIFYING,FAILED

Replacing a failed ESS storage drawer: a sample scenario

Prerequisite information:

- IBM Spectrum Scale 4.1.1 PTF8 or 4.2.1 PTF1 is a prerequisite for this procedure to work. If you are not at one of these levels or higher, contact IBM.
- This procedure is intended to be done as a partnership between the storage administrator and a hardware service representative. The storage administrator is expected to understand the IBM Spectrum Scale RAID concepts and the locations of the storage enclosures. The storage administrator is responsible for all the steps except those in which the hardware is actually being worked on.
- The pdisks in a drawer span two recovery groups; therefore, it is very important that you examine the pdisks and the fault tolerance of the vdisks in both recovery groups when going through these steps.
- An underlying principle is that drawer replacement should never deliberately put any vdisk into critical state. When vdisks are in critical state, there is no redundancy and the next single sector or IO error can cause unavailability or data loss. If drawer replacement is not possible without making the system critical, then the ESS has to be shut down before the drawer is removed. An example of drawer replacement will follow these instructions.

Replacing a failed ESS storage drawer requires the following steps:

1. If IBM Spectrum Scale is shut down: perform drawer replacement as soon as possible. Perform steps 4b and 4c and then restart IBM Spectrum Scale.
2. Examine the states of the pdisks in the affected drawer. If all the pdisk states are missing, dead, or replace, then go to step 4b to perform drawer replacement as soon as possible without going through any of the other steps in this procedure.

Assuming that you know the enclosure number and drawer number and are using standard pdisk naming conventions, you could use the following commands to display the pdisks and their states:

```
mmlsrecoverygroup LeftRecoveryGroupName -L --pdisk | grep
e{EnclosureNumber}d{DrawerNumber}
mmlsrecoverygroup RightRecoveryGroupName -L --pdisk | grep
e{EnclosureNumber}d{DrawerNumber}
```

3. Determine whether online replacement is possible.
 - a. Consult the following table to see if drawer replacement is theoretically possible for this configuration. The only required input at this step is the ESS model.

The table shows some of the possible ESS system configurations, as well as the configuration parameters for the systems. If the table indicates that online replacement is impossible, IBM Spectrum Scale will need to be shut down (on at least the two I/O servers involved) and you should go back to step 1. The fault tolerance notation uses E for enclosure, D for drawer, and P for pdisk.

Additional background information on interpreting the fault tolerance values:

- For many of the systems, 1E is reported as the fault tolerance; however, this does not mean that failure of x arbitrary drawers or y arbitrary pdisks can be tolerated. It means that the failure of all the entities in one entire enclosure can be tolerated.
- A fault tolerance of 1E+1D or 2D implies that the failure of two arbitrary drawers can be tolerated.

Table 4. ESS fault tolerance for drawer/enclosure											
Hardware type (model name...)			DA configuration			Fault tolerance					Is online replacement possible?
IBM ESS	Enclosure type	# Encl.	# Data DA per RG	Disks per DA	# Spares	RG desc	Mirrored vdisk		Parity vdisk		
GS1	2U-24	1	1	12	1	4P	3Way	2P	8+2p	2P	No drawers, enclosure impossible
							4Way	3P	8+3p	3P	No drawers, enclosure impossible

Table 4. ESS fault tolerance for drawer/enclosure (continued)

Hardware type (model name...)			DA configuration			Fault tolerance					Is online replacement possible?
GS2	2U-24	2	1	24	2	4P	3Way	2P	8+2p	2P	No drawers, enclosure impossible
							4Way	3P	8+3p	3P	No drawers, enclosure impossible
GS4	2U-24	4	1	48	2	1E+1P	3Way	1E+1P	8+2p	2P	No drawers, enclosure impossible
							4Way	1E+1P	8+3p	1E	No drawers, enclosure impossible
GS6	2U-24	6	1	72	2	1E+1P	3Way	1E+1P	8+2p	1E	No drawers, enclosure impossible
							4Way	1E+1P	8+3p	1E+1P	No drawers, enclosure possible.
GL2	4U-60 (5d)	2	1	58	2	4D	3Way	2D	8+2p	2D	Drawer possible, enclosure impossible.
							4Way	3D	8+3p	1D+1P	Drawer possible, enclosure impossible
GL4	4U-60 (5d)	4	2	58	2	1E+1D	3Way	1E+1D	8+2p	2D	Drawer possible, enclosure impossible
							4Way	1E+1D	8+3p	1E	Drawer possible, enclosure impossible
GL4	4U-60 (5d)	4	1	116	4	1E+1D	3Way	1E+1D	8+2p	2D	Drawer possible, enclosure impossible
							4Way	1E+1D	8+3p	1E	Drawer possible, enclosure impossible
GL6	4U-60 (5d)	6	3	58	2	1E+1D	3Way	1E+1D	8+2p	1E	Drawer possible, enclosure impossible
							4Way	1E+1D	8+3p	1E+1D	Drawer possible, enclosure possible.
GL6	4U-60 (5d)	6	1	174	6	1E+1D	3Way	1E+1D	8+2p	1E	Drawer possible, enclosure impossible
							4Way	1E+1D	8+3p	1E+1D	Drawer possible, enclosure possible.

Table 4. ESS fault tolerance for drawer/enclosure (continued)											
Hardware type (model name...)			DA configuration			Fault tolerance					Is online replacement possible?
GLS1	5147-084 (2d)	1	1	41	2	4P	3Way	2P	8+2p	2P	Drawer impossible, enclosure impossible
							4Way	3P	8+3p	3P	Drawer impossible, enclosure impossible.
GL2S	5147-084 (2d)	2	1	83	2	1D+1P	3Way	1D+1P			Drawer possible, enclosure impossible
									8+2p	2P	Drawer impossible, enclosure impossible
							4Way	1D+1P			Drawer possible, enclosure impossible.
									8+3p	1D	Drawer impossible, enclosure impossible.
GL4S	5147-084 (2d)	4	1	167	4	1E+1P	3Way	1E+1P			Drawer possible, enclosure possible
									8+2p	1D	Drawer impossible, enclosure impossible
							4Way	1E+1P			Drawer possible, enclosure possible.
									8+3p	1E	Drawer possible, enclosure impossible.
GL6S	5147-084 (2d)	6	1	251	6	2E	3Way	2E			Drawer possible, enclosure possible
									8+2p	1E	Drawer possible, enclosure impossible
							4Way	2E			Drawer possible, enclosure possible.
									8+3p	1E+1D	Drawer possible, enclosure possible.

For information about the fault tolerance of GLxC models, see [“ESS fault tolerance for GLxC” on page 50.](#)

- b. Determine the actual disk group fault tolerance of the vdisks in both recovery groups using the **mmlsrecoverygroup RecoveryGroupName -L** command. The rg descriptor and all the vdisks must be able to tolerate the loss of the item being replaced plus one other item. This is necessary

because the disk group fault tolerance code uses a definition of "tolerance" that includes the system running in `critical` mode. But since putting the system into `critical` is not advised, one other item is required. For example, all the following would be a valid fault tolerance to continue with drawer replacement: 1E+1D, 1D+1P, and 2D.

- c. Compare the actual disk group fault tolerance with the disk group fault tolerance listed in [Table 4 on page 43](#). If the system is using a mix of 2-fault-tolerant and 3-fault-tolerant vdisks, the comparisons must be done with the weaker (2-fault-tolerant) values. If the fault tolerance can tolerate at least the item being replaced plus one other item, then replacement can proceed. Go to step 4.

4. Drawer Replacement procedure.

- a. Quiesce the pdisks.

Choose one of the following methods to suspend all the pdisks in the drawer.

- Using the `chdrawer` sample script:

```
/usr/lpp/mmfs/samples/vdisk/chdrawer EnclosureSerialNumber DrawerNumber --release
```

- Manually using the **mmchpdisk** command:

```
for slotNumber in 01 02 03 04 05 06 ; do mmchpdisk LeftRecoveryGroupName --pdisk \
e{EnclosureNumber}d{DrawerNumber}s{slotNumber} --suspend ; done

for slotNumber in 07 08 09 10 11 12 ; do mmchpdisk RightRecoveryGroupName --pdisk \
e{EnclosureNumber}d{DrawerNumber}s{slotNumber} --suspend ; done
```

Verify that the pdisks were suspended using the **mmlsrucoverygroup** command as shown in step 2.

- b. Remove the drives; make sure to record the location of the drives and label them. You will need to replace them in the corresponding slots of the new drawer later.
- c. Replace the drawer following standard hardware procedures.
- d. Replace the drives in the corresponding slots of the new drawer.
- e. Resume the pdisks.

Choose one of the following methods to resume all the pdisks in the drawer.

- Using the `chdrawer` sample script:

```
/usr/lpp/mmfs/samples/vdisk/chdrawer EnclosureSerialNumber DrawerNumber --replace
```

- Manually using the **mmchpdisk** command:

```
for slotNumber in 01 02 03 04 05 06 ; do mmchpdisk LeftRecoveryGroupName --pdisk \
e{EnclosureNumber}d{DrawerNumber}s{slotNumber} --resume ; done

for slotNumber in 07 08 09 10 11 12 ; do mmchpdisk RightRecoveryGroupName --pdisk \
e{EnclosureNumber}d{DrawerNumber}s{slotNumber} --resume ; done
```

You can verify that the pdisks are no longer suspended using the **mmlsrucoverygroup** command as shown in step 2.

5. Verify that the drawer has been successfully replaced.

Examine the states of the pdisks in the affected drawers. All the pdisk states should be ok and the second column of the output should all be "2" indicating that 2 paths are being seen. Assuming that you know the enclosure number and drawer number and are using standard pdisk naming conventions, you could use the following commands to display the pdisks and their states:

```
mmlsrucoverygroup LeftRecoveryGroupName -L --pdisk | grep
e{EnclosureNumber}d{DrawerNumber}
```

```
mmlsrecoverygroup RightRecoveryGroupName -L --pdisk | grep
e{EnclosureNumber}d{DrawerNumber}
```

Example

The system is a GL4 with vdisks that have 4way mirroring and 8+3p RAID codes. Assume that the drawer that contains pdisk e2d3s01 needs to be replaced because one of the drawer control modules has failed (so that you only see one path to the drives instead of 2). This means that you are trying to replace drawer 3 in enclosure 2. Assume that the drawer spans recovery groups rgL and rgR.

Determine the enclosure serial number:

```
> mmlspdisk rgL --pdisk e2d3s01 | grep -w location
location = "SV21106537-3-1"
```

Examine the states of the pdisks and find that they are all ok.

```
> mmlsrecoverygroup rgL -L --pdisk | grep e2d3
e2d3s01      1, 2      DA1      1862 GiB  normal    ok
e2d3s02      1, 2      DA1      1862 GiB  normal    ok
e2d3s03      1, 2      DA1      1862 GiB  normal    ok
e2d3s04      1, 2      DA1      1862 GiB  normal    ok
e2d3s05      1, 2      DA1      1862 GiB  normal    ok
e2d3s06      1, 2      DA1      1862 GiB  normal    ok

> mmlsrecoverygroup rgR -L --pdisk | grep e2d3
e2d3s07      1, 2      DA1      1862 GiB  normal    ok
e2d3s08      1, 2      DA1      1862 GiB  normal    ok
e2d3s09      1, 2      DA1      1862 GiB  normal    ok
e2d3s10      1, 2      DA1      1862 GiB  normal    ok
e2d3s11      1, 2      DA1      1862 GiB  normal    ok
e2d3s12      1, 2      DA1      1862 GiB  normal    ok
```

Determine whether online replacement is theoretically possible by consulting Table 4 on page 43.

The system is ESS GL4, so according to the last column drawer replacement is theoretically possible.

Determine the actual disk group fault tolerance of the vdisks in both recovery groups.

```
> mmlsrecoverygroup rgL -L
```

recovery group		declustered arrays		vdisks	pdisks	format version				
rgL		4		5	119	4.2.0.1				
declustered array priority	needs service	vdisks	pdisks	spares	replace threshold	free space	scrub duration	background activity		
								task	progress	
SSD	no	1	1	0,0	1	186 GiB	14 days	scrub	8%	low
NVR	no	1	2	0,0	1	3632 MiB	14 days	scrub	8%	low
DA1	no	3	58	2,31	2	16 GiB	14 days	scrub	5%	low
DA2	no	0	58	2,31	2	152 TiB	14 days	inactive	0%	low
vdisk	RAID code		declustered array		vdisk size	block size	checksum granularity	state	remarks	
logtip_rgL	2WayReplication		NVR		48 MiB	2 MiB	4096	ok	logTip	
logtipbackup_rgL	Unreplicated		SSD		48 MiB	2 MiB	4096	ok		
logTipBackup										
loghome_rgL	4WayReplication		DA1		20 GiB	2 MiB	4096	ok	log	
md_DA1_rgL	4WayReplication		DA1		101 GiB	512 KiB	32 KiB	ok		
da_DA1_rgL	8+3p		DA1		110 TiB	8 MiB	32 KiB	ok		
config data	declustered array		VCD spares		actual rebuild spare space		remarks			
rebuild space	DA1		31		35 pdisk					
rebuild space	DA2		31		36 pdisk					
config data	max disk group fault tolerance				actual disk group fault tolerance		remarks			

rg descriptor	1 enclosure + 1 drawer	1 enclosure + 1 drawer	limiting fault
tolerance			
system index	2 enclosure	1 enclosure + 1 drawer	limited by rg
descriptor			
vdisk	max disk group fault tolerance	actual disk group fault tolerance	remarks
-----	-----	-----	-----
logtip_rgL	1 pdisk	1 pdisk	
logtipbackup_rgL	0 pdisk	0 pdisk	
loghome_rgL	3 enclosure	1 enclosure + 1 drawer	limited by rg
descriptor			
md_DA1_rgL	3 enclosure	1 enclosure + 1 drawer	limited by rg
descriptor			
da_DA1_rgL	1 enclosure	1 enclosure	
active recovery group server		servers	
-----	-----	-----	-----
c55f05n01-te0.gpfs.net		c55f05n01-te0.gpfs.net,c55f05n02-te0.gpfs.net	
.			
.			
.			

```
> mmlsrecoverygroup rgR -L
```

recovery group	declustered arrays	vdisks	pdisks	format version														
-----	-----	-----	-----	-----														
rgR	4	5	119	4.2.0.1														
declustered array priority	needs service	vdisks	pdisks	spares	replace threshold	free space	scrub duration	background activity task	progress									
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----								
SSD	no	1	1	0,0	1	186 GiB	14 days	scrub	8%	low								
NVR	no	1	2	0,0	1	3632 MiB	14 days	scrub	8%	low								
DA1	no	3	58	2,31	2	16 GiB	14 days	scrub	5%	low								
DA2	no	0	58	2,31	2	152 TiB	14 days	inactive	0%	low								
vdisk	RAID code	declustered array	vdisk size	block size	checksum granularity	state	remarks											
-----	-----	-----	-----	-----	-----	-----	-----											
logtip_rgR	2WayReplication	NVR	48 MiB	2 MiB	4096	ok	logTip											
logtipbackup_rgR	Unreplicated	SSD	48 MiB	2 MiB	4096	ok												
logTipBackup																		
loghome_rgR	4WayReplication	DA1	20 GiB	2 MiB	4096	ok	log											
md_DA1_rgR	4WayReplication	DA1	101 GiB	512 KiB	32 KiB	ok												
da_DA1_rgR	8+3p	DA1	110 TiB	8 MiB	32 KiB	ok												
config data	declustered array	VCD spares	actual rebuild	spare space	remarks													
-----	-----	-----	-----	-----	-----													
rebuild space	DA1	31	35	pdisk														
rebuild space	DA2	31	36	pdisk														
config data	max disk group fault tolerance	actual disk group fault tolerance	remarks															
-----	-----	-----	-----															
rg descriptor	1 enclosure + 1 drawer	1 enclosure + 1 drawer	limiting fault															
tolerance																		
system index	2 enclosure	1 enclosure + 1 drawer	limited by rg															
descriptor																		
vdisk	max disk group fault tolerance	actual disk group fault tolerance	remarks															
-----	-----	-----	-----															
logtip_rgR	1 pdisk	1 pdisk																
logtipbackup_rgR	0 pdisk	0 pdisk																
loghome_rgR	3 enclosure	1 enclosure + 1 drawer	limited by rg															
descriptor																		
md_DA1_rgR	3 enclosure	1 enclosure + 1 drawer	limited by rg															
descriptor																		
da_DA1_rgR	1 enclosure	1 enclosure																
active recovery group server		servers																
-----	-----	-----	-----															
c55f05n02-te0.gpfs.net		c55f05n02-te0.gpfs.net,c55f05n01-te0.gpfs.net																

The rg descriptor has an actual fault tolerance of 1 enclosure + 1 drawer (1E+1D). The data vdisks have a RAID code of 8+3P and an actual fault tolerance of 1 enclosure (1E). The metadata vdisks have a RAID code of 4WayReplication and an actual fault tolerance of 1 enclosure + 1 drawer (1E+1D).

Compare the actual disk group fault tolerance with the disk group fault tolerance listed in Table 4 on page 43.

The actual values match the table values exactly. Therefore, drawer replacement can proceed.

Quiesce the pdisks.

Choose one of the following methods to suspend all the pdisks in the drawer.

- Using the chdrawer sample script:

```
/usr/lpp/mmfs/samples/vdisk/chdrawer SV21106537 3 --release
```

- Manually using the **mmchpdisk** command:

```
for slotNumber in 01 02 03 04 05 06 ; do mmchpdisk rgl --pdisk e2d3s$slotNumber --
suspend ; done
for slotNumber in 07 08 09 10 11 12 ; do mmchpdisk rgR --pdisk e2d3s$slotNumber --
suspend ; done
```

Verify the states of the pdisks and find that they are all suspended.

```
> mmlsrecoverygroup rgl -L --pdisk | grep e2d3
e2d3s01      0, 2      DA1      1862 GiB  normal  ok/suspended
e2d3s02      0, 2      DA1      1862 GiB  normal  ok/suspended
e2d3s03      0, 2      DA1      1862 GiB  normal  ok/suspended
e2d3s04      0, 2      DA1      1862 GiB  normal  ok/suspended
e2d3s05      0, 2      DA1      1862 GiB  normal  ok/suspended
e2d3s06      0, 2      DA1      1862 GiB  normal  ok/suspended
> mmlsrecoverygroup rgR -L --pdisk | grep e2d3
e2d3s07      0, 2      DA1      1862 GiB  normal  ok/suspended
e2d3s08      0, 2      DA1      1862 GiB  normal  ok/suspended
e2d3s09      0, 2      DA1      1862 GiB  normal  ok/suspended
e2d3s10      0, 2      DA1      1862 GiB  normal  ok/suspended
e2d3s11      0, 2      DA1      1862 GiB  normal  ok/suspended
e2d3s12      0, 2      DA1      1862 GiB  normal  ok/suspended
```

Remove the drives; make sure to record the location of the drives and label them. You will need to replace them in the corresponding slots of the new drawer later.

Replace the drawer following standard hardware procedures.

Replace the drives in the corresponding slots of the new drawer.

Resume the pdisks.

- Using the chdrawer sample script:

```
/usr/lpp/mmfs/samples/vdisk/chdrawer EnclosureSerialNumber DrawerNumber --replace
```

- Manually using the **mmchpdisk** command:

```
for slotNumber in 01 02 03 04 05 06 ; do mmchpdisk rgl --pdisk e2d3s$slotNumber --
resume ; done
for slotNumber in 07 08 09 10 11 12 ; do mmchpdisk rgR --pdisk e2d3s$slotNumber --
resume ; done
```

Verify that all the pdisks have been resumed.

```
> mmlsrecoverygroup rgl -L --pdisk | grep e2d3
e2d3s01      2, 2      DA1      1862 GiB  normal  ok
e2d3s02      2, 2      DA1      1862 GiB  normal  ok
e2d3s03      2, 2      DA1      1862 GiB  normal  ok
e2d3s04      2, 2      DA1      1862 GiB  normal  ok
e2d3s05      2, 2      DA1      1862 GiB  normal  ok
e2d3s06      2, 2      DA1      1862 GiB  normal  ok
> mmlsrecoverygroup rgR -L --pdisk | grep e2d3
e2d3s07      2, 2      DA1      1862 GiB  normal  ok
e2d3s08      2, 2      DA1      1862 GiB  normal  ok
e2d3s09      2, 2      DA1      1862 GiB  normal  ok
e2d3s10      2, 2      DA1      1862 GiB  normal  ok
```

e2d3s11	2, 2	DA1	1862 GiB	normal	ok
e2d3s12	2, 2	DA1	1862 GiB	normal	ok

ESS fault tolerance for GLxC

The following section describes ESS fault tolerance with erasure code settings for the 4u106 enclosure models, GLxC.

The following are applicable for the 4u106 enclosure models, GLxC.

- There is no configuration that allows a single online enclosure replacement using vdisks with 8+2p.
- There is no configuration that allows for multiple online enclosure replacements.
- The GLxC models (4u106) have no drawers. These are single slider enclosures.

Hardware Type			DA Configuration			Fault Tolerance					Is online replaceme nt possible?
IBM ESS	Enclosure Type	# Encl.	#Data DA per RG	Disks per RG	# data Spares	RG Desc	Mirrored Vdisk		Parity vdisk		
GL1C	4U-106	1	1	52	2	4P	3Way	2P	8+2p	2P	Enclosure Impossible
							3Way	2P	4+2p	2P	Enclosure Impossible
							4Way	3P	8+3p	3P	Enclosure Impossible
							4Way	3P	4+3p	3P	Enclosure Impossible
GL2C	4U-106	2	1	105	2	4P	3Way	2P	8+2p	2P	Enclosure impossible
							3Way	2P	4+2p	2P	Enclosure Impossible
							4Way	3P	8+3p	3P	Enclosure Impossible
							4Way	3P	4+3p	3P	Enclosure Impossible
GL3C	4U-106	3	1	158	3	1E + 1P	3Way	1E + 1P	8+2p	2P	Enclosure Impossible
							3Way	1E + 1P	4+2p	1E	Enclosure Impossible
							4Way	1E + 1P	8+3p	3P	Enclosure Impossible
							4Way	1E + 1P	4+3p	1E	Enclosure Impossible

GL4C	4U-106	4	1	211	4	1E + 1P	3Way	1E + 1P	8+2p	2P	Enclosure Impossible
							3Way	1E + 1P	4+2p	1E	Enclosure Impossible
							4Way	1E + 1P	8+3p	1E	Enclosure Impossible
							4Way	1E + 1P	4+3p	1E + 1P	Enclosure possible
GL5C	4U-106	5	1	264	5	2E	3Way	2E	8+2p	1E	Enclosure Impossible
							3Way	2E	4+2p	1E	Enclosure Impossible
							4Way	2E	8+3p	1E	Enclosure Impossible
							4Way	2E	4+3p	1E + 1P	Enclosure possible
GL6C	4U-106	6	1	317	6	2E	3Way	2E	8+2p	1E	Enclosure Impossible
							3Way	2E	4+2p	2E	Enclosure possible
							4Way	2E	8+3p	1E + 1P	Enclosure possible
							4Way	2E	4+3p	2E	Enclosure possible
GL7C	4U-106	7	1	370	7	2E	3Way	2E	8+2p	1E	Enclosure Impossible
							3Way	2E	4+2p	2E	Enclosure possible
							4Way	2E	8+3p	1E + 1P	Enclosure possible
							4Way	2E	4+3p	2E	Enclosure possible
GL8C	4U-106	8	1	423	8	2E	3Way	2E	8+2p	1E	Enclosure Impossible
							3Way	2E	4+2p	2E	Enclosure possible
							4Way	2E	8+3p	1E + 1P	Enclosure possible
							4Way	2E	4+3p	2E	Enclosure possible

Replacing a failed ESS storage enclosure: a sample scenario

Enclosure replacement should be rare. Online replacement of an enclosure is only possible on a GL6 and GS6.

Prerequisite information:

- IBM Spectrum Scale 4.1.1 PTF8 or 4.2.1 PTF1 is a prerequisite for this procedure to work. If you are not at one of these levels or higher, contact IBM.
- This procedure is intended to be done as a partnership between the storage administrator and a hardware service representative. The storage administrator is expected to understand the IBM

Spectrum Scale RAID concepts and the locations of the storage enclosures. The storage administrator is responsible for all the steps except those in which the hardware is actually being worked on.

- The pdisks in a drawer span two recovery groups; therefore, it is very important that you examine the pdisks and the fault tolerance of the vdisks in both recovery groups when going through these steps.
 - An underlying principle is that enclosure replacement should never deliberately put any vdisk into critical state. When vdisks are in critical state, there is no redundancy and the next single sector or IO error can cause unavailability or data loss. If drawer replacement is not possible without making the system critical, then the ESS has to be shut down before the drawer is removed. An example of drawer replacement will follow these instructions.
1. If IBM Spectrum Scale is shut down: perform the enclosure replacement as soon as possible. Perform steps 4b through 4h and then restart IBM Spectrum Scale.
 2. Examine the states of the pdisks in the affected enclosure. If all the pdisk states are missing, dead, or replace, then go to step 4b to perform drawer replacement as soon as possible without going through any of the other steps in this procedure.

Assuming that you know the enclosure number and are using standard pdisk naming conventions, you could use the following commands to display the pdisks and their states:

```
mmlsrecoverygroup LeftRecoveryGroupName -L --pdisk | grep e{EnclosureNumber}
mmlsrecoverygroup RightRecoveryGroupName -L --pdisk | grep e{EnclosureNumber}
```

3. Determine whether online replacement is possible.
 - a. Consult the following table to see if enclosure replacement is theoretically possible for this configuration. The only required input at this step is the ESS model. The table shows each possible ESS system as well as the configuration parameters for the systems. If the table indicates that online replacement is impossible, IBM Spectrum Scale will need to be shut down (on at least the two I/O servers involved) and you should go back to step 1. The fault tolerance notation uses E for enclosure, D for drawer, and P for pdisk.

Additional background information on interpreting the fault tolerance values:

- For many of the systems, 1E is reported as the fault tolerance; however, this does not mean that failure of x arbitrary drawers or y arbitrary pdisks can be tolerated. It means that the failure of all the entities in one entire enclosure can be tolerated.
- A fault tolerance of 1E+1D or 2D implies that the failure of two arbitrary drawers can be tolerated.

Hardware type (model name...)			DA configuration			Fault tolerance				Is online replacement possible?
IBM ESS	Enclosure type	# Encl.	# Data DA per RG	Disks per DA	# Spares	RG desc	Mirrored vdisk		Parity vdisk	
GS1	2U-24	1	1	12	1	4P	3Way	2P	8+2p	2P
							4Way	3P	8+3p	3P
GS2	2U-24	2	1	24	2	4P	3Way	2P	8+2p	2P
							4Way	3P	8+3p	3P
GS4	2U-24	4	1	48	2	1E+1P	3Way	1E+1P	8+2p	2P
							4Way	1E+1P	8+3p	1E

Table 5. ESS fault tolerance for drawer/enclosure (continued)											
Hardware type (model name...)			DA configuration			Fault tolerance					Is online replacement possible?
GS6	2U-24	6	1	72	2	1E+1P	3Way	1E+1P	8+2p	1E	No drawers, enclosure impossible
							4Way	1E+1P	8+3p	1E+1P	No drawers, enclosure possible.
GL2	4U-60 (5d)	2	1	58	2	4D	3Way	2D	8+2p	2D	Drawer possible, enclosure impossible.
							4Way	3D	8+3p	1D+1P	Drawer possible, enclosure impossible
GL4	4U-60 (5d)	4	2	58	2	1E+1D	3Way	1E+1D	8+2p	2D	Drawer possible, enclosure impossible
							4Way	1E+1D	8+3p	1E	Drawer possible, enclosure impossible
GL4	4U-60 (5d)	4	1	116	4	1E+1D	3Way	1E+1D	8+2p	2D	Drawer possible, enclosure impossible
							4Way	1E+1D	8+3p	1E	Drawer possible, enclosure impossible
GL6	4U-60 (5d)	6	3	58	2	1E+1D	3Way	1E+1D	8+2p	1E	Drawer possible, enclosure impossible
							4Way	1E+1D	8+3p	1E+1D	Drawer possible, enclosure possible.
GL6	4U-60 (5d)	6	1	174	6	1E+1D	3Way	1E+1D	8+2p	1E	Drawer possible, enclosure impossible
							4Way	1E+1D	8+3p	1E+1D	Drawer possible, enclosure possible.

- b. Determine the actual disk group fault tolerance of the vdisks in both recovery groups using the **mmfssrecoverygroup RecoveryGroupName -L** command. The rg descriptor and all the vdisks must be able to tolerate the loss of the item being replaced plus one other item. This is necessary because the disk group fault tolerance code uses a definition of "tolerance" that includes the system running in **critical** mode. But since putting the system into **critical** is not advised, one other item is required. For example, all the following would be a valid fault tolerance to continue with enclosure replacement: 1E+1D and 1E+1P.
 - c. Compare the actual disk group fault tolerance with the disk group fault tolerance listed in [Table 5 on page 52](#). If the system is using a mix of 2-fault-tolerant and 3-fault-tolerant vdisks, the comparisons must be done with the weaker (2-fault-tolerant) values. If the fault tolerance can tolerate at least the item being replaced plus one other item, then replacement can proceed. Go to step 4.
4. Enclosure Replacement procedure.
 - a. Quiesce the pdisks.

For GL systems, issue the following commands for each drawer.

```
for slotNumber in 01 02 03 04 05 06 ; do mmchpdisk LeftRecoveryGroupName --pdisk
e{EnclosureNumber}d{DrawerNumber}s{slotNumber} --suspend ; done

for slotNumber in 07 08 09 10 11 12 ; do mmchpdisk RightRecoveryGroupName --pdisk
e{EnclosureNumber}d{DrawerNumber}s{slotNumber} --suspend ; done
```

For GS systems, issue:

```
for slotNumber in 01 02 03 04 05 06 07 08 09 10 11 12 ; do mmchpdisk LeftRecoveryGroupName --pdisk
e{EnclosureNumber}s{$slotNumber} --suspend ; done
for slotNumber in 13 14 15 16 17 18 19 20 21 22 23 24 ; do mmchpdisk RightRecoveryGroupName --pdisk
e{EnclosureNumber}s{$slotNumber} --suspend ; done
```

Verify that the pdisks were suspended using the **mmlsrecoverygroup** command as shown in step 2.

- b. Remove the drives; make sure to record the location of the drives and label them. You will need to replace them in the corresponding slots of the new enclosure later.
- c. Replace the enclosure following standard hardware procedures.
 - Remove the SAS connections in the rear of the enclosure.
 - Remove the enclosure.
 - Install the new enclosure.
- d. Replace the drives in the corresponding slots of the new enclosure.
- e. Connect the SAS connections in the rear of the new enclosure.
- f. Power up the enclosure.
- g. Verify the SAS topology on the servers to ensure that all drives from the new storage enclosure are present.
- h. Update the necessary firmware on the new storage enclosure as needed.
- i. Resume the pdisks.

For GL systems, issue:

```
for slotNumber in 01 02 03 04 05 06 ; do mmchpdisk LeftRecoveryGroupName --pdisk
e{EnclosureNumber}d{DrawerNumber}s{$slotNumber} --resume ; done
for slotNumber in 07 08 09 10 11 12 ; do mmchpdisk RightRecoveryGroupName --pdisk
e{EnclosureNumber}d{DrawerNumber}s{$slotNumber} --resume ; done
```

For GS systems, issue:

```
for slotNumber in 01 02 03 04 05 06 07 08 09 10 11 12 ; do mmchpdisk LeftRecoveryGroupName --pdisk
e{EnclosureNumber}s{$slotNumber} --resume ; done
for slotNumber in 13 14 15 16 17 18 19 20 21 22 23 24 ; do mmchpdisk RightRecoveryGroupName --pdisk
e{EnclosureNumber}s{$slotNumber} --resume ; done
```

Verify that the pdisks were resumed using the **mmlsrecoverygroup** command as shown in step 2.

Example

The system is a GL6 with vdisks that have 4way mirroring and 8+3p RAID codes. Assume that the enclosure that contains pdisk e2d3s01 needs to be replaced. This means that you are trying to replace enclosure 2.

Assume that the enclosure spans recovery groups rgL and rgR.

Determine the enclosure serial number:

```
> mmlspdisk rgL --pdisk e2d3s01 | grep -w location
location = "SV21106537-3-1"
```

Examine the states of the pdisks and find that they are all ok instead of missing. (Given that you have a failed enclosure, all the drives would not likely be in an ok state, but this is just an example.)

```
> mmlsrecoverygroup rgL -L --pdisk | grep e2
```

e2d1s01	2,	4	DA1	96	GiB	normal	ok
e2d1s02	2,	4	DA1	96	GiB	normal	ok
e2d1s04	2,	4	DA1	96	GiB	normal	ok
e2d1s05	2,	4	DA2	2792	GiB	normal	ok/noData
e2d1s06	2,	4	DA2	2792	GiB	normal	ok/noData
e2d2s01	2,	4	DA1	96	GiB	normal	ok
e2d2s02	2,	4	DA1	98	GiB	normal	ok
e2d2s03	2,	4	DA1	96	GiB	normal	ok
e2d2s04	2,	4	DA2	2792	GiB	normal	ok/noData
e2d2s05	2,	4	DA2	2792	GiB	normal	ok/noData
e2d2s06	2,	4	DA2	2792	GiB	normal	ok/noData
e2d3s01	2,	4	DA1	96	GiB	normal	ok
e2d3s02	2,	4	DA1	94	GiB	normal	ok
e2d3s03	2,	4	DA1	96	GiB	normal	ok
e2d3s04	2,	4	DA2	2792	GiB	normal	ok/noData
e2d3s05	2,	4	DA2	2792	GiB	normal	ok/noData
e2d3s06	2,	4	DA2	2792	GiB	normal	ok/noData
e2d4s01	2,	4	DA1	96	GiB	normal	ok
e2d4s02	2,	4	DA1	96	GiB	normal	ok
e2d4s03	2,	4	DA1	96	GiB	normal	ok
e2d4s04	2,	4	DA2	2792	GiB	normal	ok/noData
e2d4s05	2,	4	DA2	2792	GiB	normal	ok/noData
e2d4s06	2,	4	DA2	2792	GiB	normal	ok/noData
e2d5s01	2,	4	DA1	96	GiB	normal	ok
e2d5s02	2,	4	DA1	96	GiB	normal	ok
e2d5s03	2,	4	DA1	96	GiB	normal	ok
e2d5s04	2,	4	DA2	2792	GiB	normal	ok/noData
e2d5s05	2,	4	DA2	2792	GiB	normal	ok/noData
e2d5s06	2,	4	DA2	2792	GiB	normal	ok/noData

```
> mmlsrecoverygroup rgR -L --pdisk | grep e2
```

e2d1s07	2,	4	DA1	96	GiB	normal	ok
e2d1s08	2,	4	DA1	94	GiB	normal	ok
e2d1s09	2,	4	DA1	96	GiB	normal	ok
e2d1s10	2,	4	DA2	2792	GiB	normal	ok/noData
e2d1s11	2,	4	DA2	2792	GiB	normal	ok/noData
e2d1s12	2,	4	DA2	2792	GiB	normal	ok/noData
e2d2s07	2,	4	DA1	96	GiB	normal	ok
e2d2s08	2,	4	DA1	96	GiB	normal	ok
e2d2s09	2,	4	DA1	94	GiB	normal	ok
e2d2s10	2,	4	DA2	2792	GiB	normal	ok/noData
e2d2s11	2,	4	DA2	2792	GiB	normal	ok/noData
e2d2s12	2,	4	DA2	2792	GiB	normal	ok/noData
e2d3s07	2,	4	DA1	94	GiB	normal	ok
e2d3s08	2,	4	DA1	96	GiB	normal	ok
e2d3s09	2,	4	DA1	96	GiB	normal	ok
e2d3s10	2,	4	DA2	2792	GiB	normal	ok/noData
e2d3s11	2,	4	DA2	2792	GiB	normal	ok/noData
e2d3s12	2,	4	DA2	2792	GiB	normal	ok/noData
e2d4s07	2,	4	DA1	96	GiB	normal	ok
e2d4s08	2,	4	DA1	94	GiB	normal	ok
e2d4s09	2,	4	DA1	96	GiB	normal	ok
e2d4s10	2,	4	DA2	2792	GiB	normal	ok/noData
e2d4s11	2,	4	DA2	2792	GiB	normal	ok/noData
e2d4s12	2,	4	DA2	2792	GiB	normal	ok/noData
e2d5s07	2,	4	DA2	2792	GiB	normal	ok/noData
e2d5s08	2,	4	DA1	108	GiB	normal	ok
e2d5s09	2,	4	DA1	108	GiB	normal	ok
e2d5s10	2,	4	DA2	2792	GiB	normal	ok/noData
e2d5s11	2,	4	DA2	2792	GiB	normal	ok/noData

Determine whether online replacement is theoretically possible by consulting [Table 5](#) on page 52.

The system is ESS GL6, so according to the last column enclosure replacement is theoretically possible.

Determine the actual disk group fault tolerance of the vdisks in both recovery groups.

```
> mmlsrecoverygroup rgL -L
```

recovery_group	declustered arrays	vdisks	pdisks	format version
rgL	4	5	177	4.2.0.1

declustered array	needs service	vdisks	pdisks	spares	replace threshold	free space	scrub duration	background task	activity progress	priority	
SSD	no	1	1	0,0	1	186 GiB	14 days	scrub	8%	low	
NVR	no	1	2	0,0	1	3632 MiB	14 days	scrub	8%	low	
DA1	no	3	174	2,31	2	16 GiB	14 days	scrub	5%	low	
vdisk	RAID code		declustered array	vdisk size	block size	checksum granularity	state	remarks			
logtip_rgL	2WayReplication		NVR	48 MiB	2 MiB	4096	ok	logTip			
logtipbackup_rgL	Unreplicated		SSD	48 MiB	2 MiB	4096	ok	logTipBackup			
loghome_rgL	4WayReplication		DA1	20 GiB	2 MiB	4096	ok	log			
md_DA1_rgL	4WayReplication		DA1	101 GiB	512 KiB	32 KiB	ok				
da_DA1_rgL	8+3p		DA1	110 TiB	8 MiB	32 KiB	ok				
config data	declustered array	VCD spares	actual rebuild spare space				remarks				
rebuild space	DA1	31	35 pdisk								
config data	max disk group fault tolerance			actual disk group fault tolerance				remarks			
rg descriptor	1 enclosure + 1 drawer			1 enclosure + 1 drawer				limiting fault tolerance			
system index	2 enclosure			1 enclosure + 1 drawer				limited by rg descriptor			
vdisk	max disk group fault tolerance			actual disk group fault tolerance				remarks			
logtip_rgL	1 pdisk			1 pdisk							
logtipbackup_rgL	0 pdisk			0 pdisk							
loghome_rgL	3 enclosure			1 enclosure + 1 drawer				limited by rg descriptor			
md_DA1_rgL	3 enclosure			1 enclosure + 1 drawer				limited by rg descriptor			
da_DA1_rgL	1 enclosure + 1 drawer			1 enclosure + 1 drawer							
active recovery group server				servers							
c55f05n01-te0.gpfs.net				c55f05n01-te0.gpfs.net,c55f05n02-te0.gpfs.net							
.											
.											
.											

```
> mmlsrecoverygroup rgR -L
```

recovery group		declustered arrays		vdisks	pdisks	format version				
rgR		4		5	177	4.2.0.1				
declustered array	needs service	vdisks	pdisks	spares	replace threshold	free space	scrub duration	task	background activity	priority
SSD	no	1	1	0,0	1	186 GiB	14 days	scrub	8%	low
NVR	no	1	2	0,0	1	3632 MiB	14 days	scrub	8%	low
DA1	no	3	174	2,31	2	16 GiB	14 days	scrub	5%	low
vdisk	RAID code			declustered array	vdisk size	block size	checksum granularity	state	remarks	
logtip_rgR	2WayReplication			NVR	48 MiB	2 MiB	4096	ok	logTip	
logtipbackup_rgR	Unreplicated			SSD	48 MiB	2 MiB	4096	ok	logTipBackup	
loghome_rgR	4WayReplication			DA1	20 GiB	2 MiB	4096	ok	log	
md_DA1_rgR	4WayReplication			DA1	101 GiB	512 KiB	32 KiB	ok		
da_DA1_rgR	8+3p			DA1	110 TiB	8 MiB	32 KiB	ok		
config data	declustered array		VCD spares	actual rebuild spare space			remarks			
rebuild space	DA1		31	35 pdisk						
config data	max disk group fault tolerance			actual disk group fault tolerance			remarks			
rg descriptor	1 enclosure + 1 drawer			1 enclosure + 1 drawer			limiting fault tolerance			
system index	2 enclosure			1 enclosure + 1 drawer			limited by rg descriptor			
vdisk	max disk group fault tolerance			actual disk group fault tolerance			remarks			
logtip_rgR	1 pdisk			1 pdisk						
logtipbackup_rgR	0 pdisk			0 pdisk						
loghome_rgR	3 enclosure			1 enclosure + 1 drawer			limited by rg descriptor			
md_DA1_rgR	3 enclosure			1 enclosure + 1 drawer			limited by rg descriptor			
da_DA1_rgR	1 enclosure + 1 drawer			1 enclosure + 1 drawer						
active recovery group server				servers						
c55f05n02-te0.gpfs.net				c55f05n02-te0.gpfs.net,c55f05n01-te0.gpfs.net						

The rg descriptor has an actual fault tolerance of 1 enclosure + 1 drawer (1E+1D). The data vdisks have a RAID code of 8+3P and an actual fault tolerance of 1 enclosure (1E). The metadata vdisks have a RAID code of 4WayReplication and an actual fault tolerance of 1 enclosure + 1 drawer (1E+1D).

Compare the actual disk group fault tolerance with the disk group fault tolerance listed in [Table 5 on page 52](#).

The actual values match the table values exactly. Therefore, enclosure replacement can proceed.

Quiesce the pdisks.

```
for slotNumber in 01 02 03 04 05 06 ; do mmchpdisk rgl --pdisk e2d1s$slotNumber --suspend ; done
for slotNumber in 07 08 09 10 11 12 ; do mmchpdisk rgR --pdisk e2d1s$slotNumber --suspend ; done

for slotNumber in 01 02 03 04 05 06 ; do mmchpdisk rgl --pdisk e2d2s$slotNumber --suspend ; done
for slotNumber in 07 08 09 10 11 12 ; do mmchpdisk rgR --pdisk e2d2s$slotNumber --suspend ; done

for slotNumber in 01 02 03 04 05 06 ; do mmchpdisk rgl --pdisk e2d3s$slotNumber --suspend ; done
for slotNumber in 07 08 09 10 11 12 ; do mmchpdisk rgR --pdisk e2d3s$slotNumber --suspend ; done

for slotNumber in 01 02 03 04 05 06 ; do mmchpdisk rgl --pdisk e2d4s$slotNumber --suspend ; done
for slotNumber in 07 08 09 10 11 12 ; do mmchpdisk rgR --pdisk e2d4s$slotNumber --suspend ; done

for slotNumber in 01 02 03 04 05 06 ; do mmchpdisk rgl --pdisk e2d5s$slotNumber --suspend ; done
for slotNumber in 07 08 09 10 11 12 ; do mmchpdisk rgR --pdisk e2d5s$slotNumber --suspend ; done
```

Verify the pdisks were suspended using the `mm1srecoverygroup` command. You should see suspended as part of the pdisk state.

```
> mm1srecoverygroup rgl -L --pdisk | grep e2d
e2d1s01      0, 4      DA1      96 GiB  normal  ok/suspended
e2d1s02      0, 4      DA1      96 GiB  normal  ok/suspended
e2d1s04      0, 4      DA1      96 GiB  normal  ok/suspended
e2d1s05      0, 4      DA2     2792 GiB normal  ok/suspended
e2d1s06      0, 4      DA2     2792 GiB normal  ok/suspended
e2d2s01      0, 4      DA1      96 GiB  normal  ok/suspended
.
.
.

> mm1srecoverygroup rgR -L --pdisk | grep e2d
e2d1s07      0, 4      DA1      96 GiB  normal  ok/suspended
e2d1s08      0, 4      DA1      94 GiB  normal  ok/suspended
e2d1s09      0, 4      DA1      96 GiB  normal  ok/suspended
e2d1s10      0, 4      DA2     2792 GiB normal  ok/suspended
e2d1s11      0, 4      DA2     2792 GiB normal  ok/suspended
e2d1s12      0, 4      DA2     2792 GiB normal  ok/suspended
e2d2s07      0, 4      DA1      96 GiB  normal  ok/suspended
e2d2s08      0, 4      DA1      96 GiB  normal  ok/suspended
.
.
.
```

Remove the drives; make sure to record the location of the drives and label them. You will need to replace them in the corresponding drawer slots of the new enclosure later.

Replace the enclosure following standard hardware procedures.

- Remove the SAS connections in the rear of the enclosure.
- Remove the enclosure.
- Install the new enclosure.

Replace the drives in the corresponding drawer slots of the new enclosure.

Connect the SAS connections in the rear of the new enclosure.

Power up the enclosure.

Verify the SAS topology on the servers to ensure that all drives from the new storage enclosure are present.

Update the necessary firmware on the new storage enclosure as needed.

Resume the pdisks.

```
for slotNumber in 01 02 03 04 05 06 ; do mmchpdisk rgl --pdisk e2d1s$slotNumber --resume ; done
for slotNumber in 07 08 09 10 11 12 ; do mmchpdisk rgR --pdisk e2d1s$slotNumber --resume ; done
for slotNumber in 01 02 03 04 05 06 ; do mmchpdisk rgl --pdisk e2d2s$slotNumber --resume ; done
for slotNumber in 07 08 09 10 11 12 ; do mmchpdisk rgR --pdisk e2d2s$slotNumber --resume ; done
for slotNumber in 01 02 03 04 05 06 ; do mmchpdisk rgl --pdisk e2d3s$slotNumber --resume ; done
for slotNumber in 07 08 09 10 11 12 ; do mmchpdisk rgR --pdisk e2d3s$slotNumber --resume ; done
for slotNumber in 01 02 03 04 05 06 ; do mmchpdisk rgl --pdisk e2d4s$slotNumber --resume ; done
for slotNumber in 07 08 09 10 11 12 ; do mmchpdisk rgR --pdisk e2d4s$slotNumber --resume ; done
```

```
for slotNumber in 01 02 03 04 05 06 ; do mmchpdisk rgl --pdisk e2d5s$slotNumber --resume ; done
for slotNumber in 07 08 09 10 11 12 ; do mmchpdisk rgR --pdisk e2d5s$slotNumber --resume ; done
```

Verify that the pdisks were resumed by using the `mmlsrecoverygroup` command.

```
> mmlsrecoverygroup rgl -L --pdisk | grep e2

e2d1s01          2, 4      DA1          96 GiB normal      ok
e2d1s02          2, 4      DA1          96 GiB normal      ok
e2d1s04          2, 4      DA1          96 GiB normal      ok
e2d1s05          2, 4      DA2        2792 GiB normal      ok/noData
e2d1s06          2, 4      DA2        2792 GiB normal      ok/noData
e2d2s01          2, 4      DA1          96 GiB normal      ok
.
.
.

> mmlsrecoverygroup rgR -L --pdisk | grep e2

e2d1s07          2, 4      DA1          96 GiB normal      ok
e2d1s08          2, 4      DA1          94 GiB normal      ok
e2d1s09          2, 4      DA1          96 GiB normal      ok
e2d1s10          2, 4      DA2        2792 GiB normal      ok/noData
e2d1s11          2, 4      DA2        2792 GiB normal      ok/noData
e2d1s12          2, 4      DA2        2792 GiB normal      ok/noData
e2d2s07          2, 4      DA1          96 GiB normal      ok
e2d2s08          2, 4      DA1          96 GiB normal      ok
.
.
.
```

Replacing failed disks in a Power 775 Disk Enclosure recovery group: a sample scenario

The scenario presented here shows how to detect and replace failed disks in a recovery group built on a Power 775 Disk Enclosure.

Detecting failed disks in your enclosure

Assume a fully-populated Power 775 Disk Enclosure (serial number 000DE37) on which the following two recovery groups are defined:

- 000DE37TOP containing the disks in the top set of carriers
- 000DE37BOT containing the disks in the bottom set of carriers

Each recovery group contains the following:

- one log declustered array (LOG)
- four data declustered arrays (DA1, DA2, DA3, DA4)

The data declustered arrays are defined according to Power 775 Disk Enclosure best practice as follows:

- 47 pdisks per data declustered array
- each member pdisk from the same carrier slot
- default disk replacement threshold value set to 2

The replacement threshold of 2 means that GNR will only require disk replacement when two or more disks have failed in the declustered array; otherwise, rebuilding onto spare space or reconstruction from redundancy will be used to supply affected data.

This configuration can be seen in the output of `mmlsrecoverygroup` for the recovery groups, shown here for 000DE37TOP:

```
# mmlsrecoverygroup 000DE37TOP -L

recovery group      declustered   vdisks   pdisks
-----

```


000DE37TOP		5	9	192					
declustered activity		needs			replace			scrub	background
array progress	service priority	vdisks	pdisks	spares	threshold	free space	duration	task	
-----		-----	-----	-----	-----	-----	-----		
DA1	no	2	47	2	2	3072 MiB	14 days	scrub	
63% low									
DA2	no	2	47	2	2	3072 MiB	14 days	scrub	
19% low									
DA3	yes	2	47	2	2	0 B	14 days	rebuild-2r	
48% low									
DA4	no	2	47	2	2	3072 MiB	14 days	scrub	
33% low									
LOG	no	1	4	1	1	546 GiB	14 days	scrub	
87% low									
vdisk		RAID code		declustered array		vdisk size		remarks	
-----		-----		-----		-----		-----	
000DE37TOPLOG		3WayReplication		LOG		4144 MiB		log	
000DE37TOPDA1META		4WayReplication		DA1		250 GiB			
000DE37TOPDA1DATA		8+3p		DA1		17 TiB			
000DE37TOPDA2META		4WayReplication		DA2		250 GiB			
000DE37TOPDA2DATA		8+3p		DA2		17 TiB			
000DE37TOPDA3META		4WayReplication		DA3		250 GiB			
000DE37TOPDA3DATA		8+3p		DA3		17 TiB			
000DE37TOPDA4META		4WayReplication		DA4		250 GiB			
000DE37TOPDA4DATA		8+3p		DA4		17 TiB			
active recovery group server					servers				
-----					-----				
server1					server1,server2				

The indication that disk replacement is called for in this recovery group is the value of yes in the needs service column for declustered array DA3.

The fact that DA3 (the declustered array on the disks in carrier slot 3) is undergoing rebuild of its RAID tracks that can tolerate two strip failures is by itself not an indication that disk replacement is required; it merely indicates that data from a failed disk is being rebuilt onto spare space. Only if the replacement threshold has been met will disks be marked for replacement and the declustered array marked as needing service.

GNR provides several indications that disk replacement is required:

- entries in the AIX error report or the Linux syslog
- the pdReplacePdisk callback, which can be configured to run an administrator-supplied script at the moment a pdisk is marked for replacement
- the POWER7® cluster event notification TEAL agent, which can be configured to send disk replacement notices when they occur to the POWER7 cluster EMS
- the output from the following commands, which may be performed from the command line on any GPFS cluster node (see the examples that follow):
 1. mmlsrecoverygroup with the -L flag shows yes in the needs service column
 2. mmlsrecoverygroup with the -L and --pdisk flags; this shows the states of all pdisks, which may be examined for the replace pdisk state
 3. mmlspdisk with the --replace flag, which lists only those pdisks that are marked for replacement

Note: Because the output of mmlsrecoverygroup -L --pdisk for a fully-populated disk enclosure is very long, this example shows only some of the pdisks (but includes those marked for replacement).

```
# mmlsrecoverygroup 000DE37TOP -L --pdisk
```

recovery group	declustered arrays	vdisks	pdisks
-----	-----	-----	-----
000DE37TOP	5	9	192

declustered array	needs service	vdisk	pdisk	spares	replace threshold	free space	scrub duration	background activity		
								task	progress	priority
DA1	no	2	47	2	2	3072 MiB	14 days	scrub	63%	low
DA2	no	2	47	2	2	3072 MiB	14 days	scrub	19%	low
DA3	yes	2	47	2	2	0 B	14 days	rebuild-2r	68%	low
DA4	no	2	47	2	2	3072 MiB	14 days	scrub	34%	low
LOG	no	1	4	1	1	546 GiB	14 days	scrub	87%	low

pdisk	n. active, total paths	declustered array	free space	user condition	state, remarks
[...]					
c014d1	2, 4	DA1	62 GiB	normal	ok
c014d2	2, 4	DA2	279 GiB	normal	ok
c014d3	0, 0	DA3	279 GiB	replaceable	dead/systemDrain/noRGD/noVCD/
replace					
c014d4	2, 4	DA4	12 GiB	normal	ok
[...]					
c018d1	2, 4	DA1	24 GiB	normal	ok
c018d2	2, 4	DA2	24 GiB	normal	ok
c018d3	2, 4	DA3	558 GiB	replaceable	dead/systemDrain/noRGD/noVCD/
noData/replace					
c018d4	2, 4	DA4	12 GiB	normal	ok
[...]					

The preceding output shows that the following pdisks are marked for replacement:

- c014d3 in DA3
- c018d3 in DA3

The naming convention used during recovery group creation indicates that these are the disks in slot 3 of carriers 14 and 18. To confirm the physical locations of the failed disks, use the `mm1spdisk` command to list information about those pdisks in declustered array DA3 of recovery group 000DE37TOP that are marked for replacement:

```
# mm1spdisk 000DE37TOP --declustered-array DA3 --replace
pdisk:
  replacementPriority = 1.00
  name = "c014d3"
  device = "/dev/rhdisk158,/dev/rhdisk62"
  recoveryGroup = "000DE37TOP"
  declusteredArray = "DA3"
  state = "dead/systemDrain/noRGD/noVCD/replace"
  .
  .
  .

pdisk:
  replacementPriority = 1.00
  name = "c018d3"
  device = "/dev/rhdisk630,/dev/rhdisk726"
  recoveryGroup = "000DE37TOP"
  declusteredArray = "DA3"
  state = "dead/systemDrain/noRGD/noVCD/noData/replace"
  .
  .
  .
```

The preceding location code attributes confirm the pdisk naming convention:

Disk	Location code	Interpretation
pdisk c014d3	78AD.001.000DE37-C14-D3	Disk 3 in carrier 14 in the disk enclosure identified by enclosure type 78AD.001 and serial number 000DE37
pdisk c018d3	78AD.001.000DE37-C18-D3	Disk 3 in carrier 18 in the disk enclosure identified by enclosure type 78AD.001 and serial number 000DE37

Replacing the failed disks in a Power 775 Disk Enclosure recovery group

Note: In this example, it is assumed that two new disks with the appropriate Field Replaceable Unit (FRU) code, as indicated by the fru attribute (74Y4936 in this case), have been obtained as replacements for the failed pdisks c014d3 and c018d3.

Replacing each disk is a three-step process:

1. Using the `mmchcarrier` command with the `--release` flag to suspend use of the other disks in the carrier and to release the carrier.
2. Removing the carrier and replacing the failed disk within with a new one.
3. Using the `mmchcarrier` command with the `--replace` flag to resume use of the suspended disks and to begin use of the new disk.

GNR assigns a priority to pdisk replacement. Disks with smaller values for the `replacementPriority` attribute should be replaced first. In this example, the only failed disks are in DA3 and both have the same `replacementPriority`.

Disk c014d3 is chosen to be replaced first.

1. To release carrier 14 in disk enclosure 000DE37:

```
# mmchcarrier 000DE37TOP --release --pdisk c014d3
[I] Suspending pdisk c014d1 of RG 000DE37TOP in location 78AD.001.000DE37-C14-D1.
[I] Suspending pdisk c014d2 of RG 000DE37TOP in location 78AD.001.000DE37-C14-D2.
[I] Suspending pdisk c014d3 of RG 000DE37TOP in location 78AD.001.000DE37-C14-D3.
[I] Suspending pdisk c014d4 of RG 000DE37TOP in location 78AD.001.000DE37-C14-D4.
[I] Carrier released.

- Remove carrier.
- Replace disk in location 78AD.001.000DE37-C14-D3 with FRU 74Y4936.
- Reinsert carrier.
- Issue the following command:

    mmchcarrier 000DE37TOP --replace --pdisk 'c014d3'

Repair timer is running. Perform the above within 5 minutes
to avoid pdisks being reported as missing.
```

GNR issues instructions as to the physical actions that must be taken. Note that disks may be suspended only so long before they are declared missing; therefore the mechanical process of physically performing disk replacement must be accomplished promptly.

Use of the other three disks in carrier 14 has been suspended, and carrier 14 is unlocked. The identify lights for carrier 14 and for disk 3 are on.

2. Carrier 14 should be unlatched and removed. The failed disk 3, as indicated by the internal identify light, should be removed, and the new disk with FRU 74Y4936 should be inserted in its place. Carrier 14 should then be reinserted and the latch closed.
3. To finish the replacement of pdisk c014d3:

```
# mmchcarrier 000DE37TOP --replace --pdisk c014d3
[I] The following pdisks will be formatted on node server1:
    /dev/rhdisk354
[I] Pdisk c014d3 of RG 000DE37TOP successfully replaced.
[I] Resuming pdisk c014d1 of RG 000DE37TOP.
[I] Resuming pdisk c014d2 of RG 000DE37TOP.
[I] Resuming pdisk c014d3#162 of RG 000DE37TOP.
[I] Resuming pdisk c014d4 of RG 000DE37TOP.
[I] Carrier resumed.
```

When the `mmchcarrier --replace` command returns successfully, GNR has resumed use of the other 3 disks. The failed pdisk may remain in a temporary form (indicated here by the name `c014d3#162`) until all data from it has been rebuilt, at which point it is finally deleted. The new replacement disk, which has assumed the name `c014d3`, will have RAID tracks rebuilt and rebalanced onto it. Notice that only one block device name is mentioned as being formatted as a pdisk; the second path will be discovered in the background.

This can be confirmed with `mmfsrecoverygroup -L --pdisk:`

```
# mmfsrecoverygroup 000DE37TOP -L --pdisk
```

recovery group	declustered arrays	vdisk	pdisk			
000DE37TOP	5	9	193			
declustered array	needs service	vdisk	pdisk	spares	replace threshold	free space
DA1	no	2	47	2	2	3072 MiB
DA2	no	2	47	2	2	3072 MiB
DA3	yes	2	48	2	2	0 B
DA4	no	2	47	2	2	3072 MiB
LOG	no	1	4	1	1	546 GiB
duration	task	progress	priority			
14 days	scrub	63%	low			
14 days	scrub	19%	low			
14 days	rebuild-2r	89%	low			
14 days	scrub	34%	low			
14 days	scrub	87%	low			
pdisk	n. active, total paths	declustered array	free space	user condition	state, remarks	
[...]						
c014d1	2, 4	DA1	23 GiB	normal	ok	
c014d2	2, 4	DA2	23 GiB	normal	ok	
c014d3	2, 4	DA3	550 GiB	normal	ok	
c014d3#162	0, 0	DA3	543 GiB	replaceable	dead/adminDrain/noRGD/noVCD/noPath	
c014d4	2, 4	DA4	23 GiB	normal	ok	
[...]						
c018d1	2, 4	DA1	24 GiB	normal	ok	
c018d2	2, 4	DA2	24 GiB	normal	ok	
c018d3	0, 0	DA3	558 GiB	replaceable	dead/systemDrain/noRGD/noVCD/	
noData/replace						
c018d4	2, 4	DA4	23 GiB	normal	ok	
[...]						

Notice that the temporary pdisk `c014d3#162` is counted in the total number of pdisks in declustered array DA3 and in the recovery group, until it is finally drained and deleted.

Notice also that pdisk `c018d3` is still marked for replacement, and that DA3 still needs service. This is because GNR replacement policy expects all failed disks in the declustered array to be replaced once the replacement threshold is reached. The `replace` state on a pdisk is not removed when the total number of failed disks goes under the threshold.

Pdisk `c018d3` is replaced following the same process.

1. Release carrier 18 in disk enclosure 000DE37:

```
# mmhcarrier 000DE37TOP --release --pdisk c018d3
[I] Suspending pdisk c018d1 of RG 000DE37TOP in location 78AD.001.000DE37-C18-D1.
[I] Suspending pdisk c018d2 of RG 000DE37TOP in location 78AD.001.000DE37-C18-D2.
[I] Suspending pdisk c018d3 of RG 000DE37TOP in location 78AD.001.000DE37-C18-D3.
[I] Suspending pdisk c018d4 of RG 000DE37TOP in location 78AD.001.000DE37-C18-D4.
[I] Carrier released.

- Remove carrier.
- Replace disk in location 78AD.001.000DE37-C18-D3 with FRU 74Y4936.
- Reinsert carrier.
- Issue the following command:

    mmhcarrier 000DE37TOP --replace --pdisk 'c018d3'

Repair timer is running. Perform the above within 5 minutes
to avoid pdisks being reported as missing.
```

2. Unlatch and remove carrier 18, remove and replace failed disk 3, reinsert carrier 18, and close the latch.

3. To finish the replacement of pdisk `c018d3`:

```
# mmhcarrier 000DE37TOP --replace --pdisk c018d3

[I] The following pdisks will be formatted on node server1:
/dev/rhdisk674
[I] Pdisk c018d3 of RG 000DE37TOP successfully replaced.
[I] Resuming pdisk c018d1 of RG 000DE37TOP.
```

```
[I] Resuming pdisk c018d2 of RG 000DE37TOP.
[I] Resuming pdisk c018d3#166 of RG 000DE37TOP.
[I] Resuming pdisk c018d4 of RG 000DE37TOP.
[I] Carrier resumed.
```

Running `mmfsrecoverygroup` again will confirm the second replacement:

```
# mmfsrecoverygroup 000DE37TOP -L --pdisk
```

recovery group	declustered arrays	vdisks	pdisks								
000DE37TOP	5	9	192								
declustered array	needs service	vdisks	pdisks	spares	replace threshold	free space	scrub duration	background activity			
								task	progress	priority	
DA1	no	2	47	2	2	3072 MiB	14 days	scrub	64%	low	
DA2	no	2	47	2	2	3072 MiB	14 days	scrub	22%	low	
DA3	no	2	47	2	2	2048 MiB	14 days	rebalance	12%	low	
DA4	no	2	47	2	2	3072 MiB	14 days	scrub	36%	low	
LOG	no	1	4	1	1	546 GiB	14 days	scrub	89%	low	

pdisk	n. active, total paths	declustered array	free space	user condition	state, remarks
[...]					
c014d1	2, 4	DA1	23 GiB	normal	ok
c014d2	2, 4	DA2	23 GiB	normal	ok
c014d3	2, 4	DA3	271 GiB	normal	ok
c014d4	2, 4	DA4	23 GiB	normal	ok
[...]					
c018d1	2, 4	DA1	24 GiB	normal	ok
c018d2	2, 4	DA2	24 GiB	normal	ok
c018d3	2, 4	DA3	542 GiB	normal	ok
c018d4	2, 4	DA4	23 GiB	normal	ok
[...]					

Notice that both temporary pdisks have been deleted. This is because `c014d3#162` has finished draining, and because pdisk `c018d3#166` had, before it was replaced, already been completely drained (as evidenced by the `noData` flag). Declustered array `DA3` no longer needs service and once again contains 47 pdisks, and the recovery group once again contains 192 pdisks.

Other hardware service

While IBM Spectrum Scale RAID can easily tolerate a single disk fault with no significant impact, and failures of up to three disks with various levels of impact on performance and data availability, it still relies on a majority of all the disks functioning properly and reachable from the server. If a major equipment malfunction prevents both the primary and backup server from accessing more than that number of disks, or if those disks are destroyed, all vdisks in the recovery group become either unavailable or suffer permanent data loss. As IBM Spectrum Scale RAID cannot recover from such catastrophic problems, it also does not attempt to diagnose them or organize their maintenance.

In the case that a IBM Spectrum Scale RAID server becomes permanently disabled, a manual failover procedure exists that requires recabling to an alternative server. For more information, see the `mmchrecoverygroup` command in the *IBM Spectrum Scale RAID: Administration*. If both the primary and backup IBM Spectrum Scale RAID servers for a recovery group fail, the recovery group is unavailable until one of the servers is repaired.

Directed maintenance procedures available in the GUI

The directed maintenance procedures (DMPs) assist you to repair a problem when you select the action **Run fix procedure** on a selected event from the **Monitoring > Events** page. DMPs are present for only a few events reported in the system.

The following table provides details of the available DMPs and the corresponding events.

Table 6. DMPs	
DMP	Event ID
Replace disks	gnr_pdisk_replaceable
Update enclosure firmware	enclosure_firmware_wrong
Update drive firmware	drive_firmware_wrong
Update host-adapter firmware	adapter_firmware_wrong
Start NSD	disk_down
Start GPFS daemon	gpfs_down
Increase fileset space	inode_error_high and inode_warn_high
Start performance monitoring collector service	pmcollector_down
Start performance monitoring sensor service	pmsensors_down
Activate AFM performance monitoring sensors	afm_sensors_inactive
Activate NFS performance monitoring sensors	nfs_sensors_inactive
Activate SMB performance monitoring sensors	smb_sensors_inactive
Configure NFS sensor	nfs_sensors_not_configured
Configure SMB sensor	smb_sensors_not_configured
Mount file systems	unmounted_fs_check
Start GUI service on remote node	gui_down
Repair a failed GUI refresh task	gui_refresh_task_failed

Replace disks

The replace disks DMP assists you to replace the disks.

The following are the corresponding event details and proposed solution:

- **Event name:** gnr_pdisk_replaceable
- **Problem:** The state of a physical disk is changed to “replaceable”.
- **Solution:** Replace the disk.

The ESS GUI detects if a disk is broken and whether it needs to be replaced. In this case, launch this DMP to get support to replace the broken disks. You can use this DMP either to replace one disk or multiple disks.

The DMP automatically launches in corresponding mode depending on situation. You can launch this DMP from the pages in the GUI and follow the wizard to release one or more disks:

- **Monitoring > Hardware** page: Select **Replace Broken Disks** from the **Actions** menu.
- **Monitoring > Hardware** page: Select the broken disk to be replaced in an enclosure and then select **Replace** from the **Actions** menu.
- **Monitoring > Events** page: Select the *gnr_pdisk_replaceable* event from the event listing and then select **Run Fix Procedure** from the **Actions** menu.
- **Storage > Physical Disks** page: Select **Replace Broken Disks** from the **Actions** menu.
- **Storage > Physical Disks** page: Select the disk to be replaced and then select **Replace Disk** from the **Actions** menu.

The system uses the following command on an *mmvdisk-enabled* environment to release and replace the disk:

```
mmvdisk pdisk replace [--prepare | --cancel] --recovery-group DiskRecoveryGroup --pdisk DiskName
```

For the systems with ESS version 5.3.0 or earlier, the system issues the **mmchcarrier** command to replace disks as given in the following format:

```
/usr/lpp/mmfs/bin/mmchcarrier <<Disk_RecoveryGroup>>  
--replace|--release|--resume --pdisk <<Disk_Name>> [--force-release]
```

For example: `/usr/lpp/mmfs/bin/mmchcarrier G1 --replace --pdisk G1FSP11`

Update enclosure firmware

The update enclosure firmware DMP assists to update the enclosure firmware to the latest level.

The following are the corresponding event details and the proposed solution:

- **Event name:** enclosure_firmware_wrong
- **Problem:** The reported firmware level of the environmental service module is not compliant with the recommendation.
- **Solution:** Update the firmware.

If more than one enclosure is not running the newest version of the firmware, the system prompts to update the firmware. The system issues the **mmchfirmware** command to update firmware as given in the following format:

```
mmchfirmware --esms <<ESM_Name>> --cluster  
  <<Cluster_Id>>- for all the enclosures :    mmchfirmware --esms --cluster  
  <<Cluster_Id>>
```

For example, for a single enclosure:

```
mmchfirmware --esms 181880E-SV20706999_ESM_B -cluster 1857390657572243170
```

For all enclosures:

```
mmchfirmware --esms -cluster 1857390657572243170
```

Update drive firmware

The update drive firmware DMP assists to update the drive firmware to the latest level so that the physical disk becomes compliant.

The following are the corresponding event details and the proposed solution:

- **Event name:** drive_firmware_wrong
- **Problem:** The reported firmware level of the physical disk is not compliant with the recommendation.
- **Solution:** Update the firmware.

If more than one disk is not running the newest version of the firmware, the system prompts to update the firmware. The system issues the **chfirmware** command to update firmware as given in the following format:

For single disk:

```
chfirmware --pdisks <<entity_name>> --cluster <<Cluster_Id>>
```

For example:

```
chfirmware --pdisks <<ENC123001/DRV-2>> --cluster 1857390657572243170
```

For all disks:

```
chfirmware --pdisks --cluster <<Cluster_Id>>
```

For example:

```
chfirmware --pdisks -cluster 1857390657572243170
```

Update host-adapter firmware

The Update host-adapter firmware DMP assists to update the host-adapter firmware to the latest level.

The following are the corresponding event details and the proposed solution:

- **Event name:** adapter_firmware_wrong
- **Problem:** The reported firmware level of the host adapter is not compliant with the recommendation.
- **Solution:** Update the firmware.

If more than one host-adapter is not running the newest version of the firmware, the system prompts to update the firmware. The system issues the **chfirmware** command to update firmware as given in the following format:

For single disk:

```
chfirmware --hostadapter <<Host_Adapter_Name>> --cluster <<Cluster_Id>>
```

For example:

```
chfirmware --hostadapter <<c45f02n04_HBA_2>> --cluster 1857390657572243170
```

For all disks:

```
chfirmware --hostadapter --cluster <<Cluster_Id>>
```

For example:

```
chfirmware --pdisks -cluster 1857390657572243170
```

Start NSD

The Start NSD DMP assists to start NSDs that are not working.

The following are the corresponding event details and the proposed solution:

- **Event ID:** disk_down
- **Problem:** The availability of an NSD is changed to “down”.
- **Solution:** Recover the NSD.

The DMP provides the option to start the NSDs that are not functioning. If multiple NSDs are down, you can select whether to recover only one NSD or all of them.

The system issues the **mmchdisk** command to recover NSDs as given in the following format:

```
/usr/lpp/mmfs/bin/mmchdisk <device> start -d <disk description>
```

For example: `/usr/lpp/mmfs/bin/mmchdisk r1_FS start -d G1_r1_FS_data_0`

Start GPFS daemon

When the GPFS daemon is down, GPFS functions do not work properly on the node.

The following are the corresponding event details and the proposed solution:

- **Event ID:** gpfs_down

- **Problem:** The GPFS daemon is down. GPFS is not operational on node.
- **Solution:** Start GPFS daemon.

The system issues the **mmstartup -N** command to restart GPFS daemon as given in the following format:

```
/usr/lpp/mmfs/bin/mmstartup -N <Node>
```

For example: `usr/lpp/mmfs/bin/mmstartup -N gss-05.localnet.com`

Increase fileset space

The system needs inodes to allow I/O on a fileset. If the inodes allocated to the fileset are exhausted, you need to either increase the number of maximum inodes or delete the existing data to free up space.

The procedure helps to increase the maximum number of inodes by a percentage of the already allocated inodes. The following are the corresponding event details and the proposed solution:

- **Event ID:** inode_error_high and inode_warn_high
- **Problem:** The inode usage in the fileset reached an exhausted level.
- **Solution:** increase the maximum number of inodes.

The system issues the **mmchfileset** command to recover NSDs as given in the following format:

```
/usr/lpp/mmfs/bin/mmchfileset <Device> <Fileset> --inode-limit <inodesMaxNumber>
```

For example: `/usr/lpp/mmfs/bin/mmchfileset r1_FS testFileset --inode-limit 2048`

Synchronize node clocks

The time must be in sync with the time set on the GUI node. If the time is not in sync, the data that is displayed in the GUI might be wrong or it does not even display the details. For example, the GUI does not display the performance data if time is not in sync.

The procedure assists to fix timing issue on a single node or on all nodes that are out of sync. The following are the corresponding event details and the proposed solution:

- **Event ID:** time_not_in_sync
- **Limitation:** This DMP is not available in sudo wrapper clusters. In a sudo wrapper cluster, the user name is different from 'root'. The system detects the user name by finding the parameter `GPFS_USER=<user name>`, which is available in the file `/usr/lpp/mmfs/gui/conf/gpfsgui.properties`.
- **Problem:** The time on the node is not synchronous with the time on the GUI node. It differs more than 1 minute.
- **Solution:** Synchronize the time with the time on the GUI node.

The system issues the **sync_node_time** command as given in the following format to synchronize the time in the nodes:

```
/usr/lpp/mmfs/gui/bin/sync_node_time <nodeName>
```

For example: `/usr/lpp/mmfs/gui/bin/sync_node_time c55f06n04.gpfs.net`

Start performance monitoring collector service

The collector services on the GUI node must be functioning properly to display the performance data in the IBM Spectrum Scale management GUI.

The following are the corresponding event details and the proposed solution:

- **Event ID:** pmcollector_down

- **Limitation:** This DMP is not available in sudo wrapper clusters when a remote *pmcollector* service is used by the GUI. A remote *pmcollector* service is detected in case a different value than localhost is specified in the ZIMonAddress in file, which is located at: /usr/lpp/mmfs/gui/conf/gpfsgui.properties. In a sudo wrapper cluster, the user name is different from 'root'. The system detects the user name by finding the parameter GPFS_USER=<user name>, which is available in the file /usr/lpp/mmfs/gui/conf/gpfsgui.properties.
- **Problem:** The performance monitoring collector service *pmcollector* is in inactive state.
- **Solution:** Issue the **systemctl status pmcollector** to check the status of the collector. If *pmcollector* service is inactive, issue **systemctl start pmcollector**.

The system restarts the performance monitoring services by issuing the **systemctl restart pmcollector** command.

The performance monitoring collector service might be on some other node of the current cluster. In this case, the DMP first connects to that node, then restarts the performance monitoring collector service.

```
ssh <nodeAddress> systemctl restart pmcollector
```

For example: `ssh 10.0.100.21 systemctl restart pmcollector`

In a sudo wrapper cluster, when collector on remote node is down, the DMP does not restart the collector services by itself. You need to do it manually.

Start performance monitoring sensor service

You need to start the sensor service to get the performance details in the collectors. If sensors and collectors are not started, the GUI and CLI do not display the performance data in the IBM Spectrum Scale management GUI.

The following are the corresponding event details and the proposed solution:

- **Event ID:** pmsensors_down
- **Limitation:** This DMP is not available in sudo wrapper clusters. In a sudo wrapper cluster, the user name is different from 'root'. The system detects the user name by finding the parameter GPFS_USER=<user name>, which is available in the file /usr/lpp/mmfs/gui/conf/gpfsgui.properties.
- **Problem:** The performance monitoring sensor service *pmsensor* is not sending any data. The service might be down or the difference between the time of the node and the node hosting the performance monitoring collector service *pmcollector* is more than 15 minutes.
- **Solution:** Issue **systemctl status pmsensors** to verify the status of the sensor service. If *pmsensor* service is inactive, issue **systemctl start pmsensors**.

The system restarts the sensors by issuing **systemctl restart pmsensors** command.

For example: `ssh gss-15.localnet.com systemctl restart pmsensors`

Activate AFM performance monitoring sensors

The activate SMB performance monitoring sensors DMP assists to activate the inactive SMB sensors.

The following are the corresponding event details and the proposed solution:

- **Event ID:** afm_sensors_inactive
- **Problem:** The AFM performance cannot be monitored because one or more of the performance sensors GPFSAFMFS, GPFSAFMFSET, and GPFSAFM are offline.
- **Solution:** Activate the AFM sensors.

The DMP provides the option to activate the AFM monitoring sensor and select a data collection interval that defines how frequently the sensors must collect data. It is recommended to select a value that is greater than or equal to 10 as the data collection frequency to reduce the impact on the system performance.

The system issues the **mmperfmon** command to activate AFM sensors as given in the following format:

```
/usr/lpp/mmfs/bin/mmperfmon config update <<sensor_name>>.restrict=<<afm_gateway_nodes>>  
/usr/lpp/mmfs/bin/mmperfmon config update <<sensor_name>>.period=<<seconds>>
```

For example:

```
/usr/lpp/mmfs/bin/mmperfmon config update GPFSAFM.restrict=gss-41  
/usr/lpp/mmfs/bin/mmperfmon config update GPFSAFM.period=30
```

Activate NFS performance monitoring sensors

The activate NFS performance monitoring sensors DMP assists to activate the inactive NFS sensors.

The following are the corresponding event details and the proposed solution:

- **Event ID:** nfs_sensors_inactive
- **Problem:** The NFS performance cannot be monitored because the performance monitoring sensor NFSIO is inactive.
- **Solution:** Activate the SMB sensors.

The DMP provides the option to activate the NFS monitoring sensor and select a data collection interval that defines how frequently the sensors must collect data. It is recommended to select a value that is greater than or equal to 10 as the data collection frequency to reduce the impact on the system performance.

The system issues the **mmperfmon** command to activate the sensors as given in the following format:

```
/usr/lpp/mmfs/bin/mmperfmon config update NFSIO.restrict=cesNodes NFSIO.period=<<seconds>>
```

For example: `/usr/lpp/mmfs/bin/mmperfmon config update NFSIO.restrict=cesNodes NFSIO.period=10`

Activate SMB performance monitoring sensors

The activate SMB performance monitoring sensors DMP assists to activate the inactive SMB sensors.

The following are the corresponding event details and the proposed solution:

- **Event ID:** smb_sensors_inactive
- **Problem:** The SMB performance cannot be monitored because either one or both the SMBStats and SMBGlobalStats sensors are inactive.
- **Solution:** Activate the SMB sensors.

The DMP provides the option to activate the SMB monitoring sensor and select a data collection interval that defines how frequently the sensors must collect data. It is recommended to select a value that is greater than or equal to 10 as the data collection frequency to reduce the impact on the system performance.

The system issues the **mmperfmon** command to activate the sensors as given in the following format:

```
/usr/lpp/mmfs/bin/mmperfmon config update SMBStats.restrict=cesNodes SMBStats.period=<<seconds>>
```

For example: `/usr/lpp/mmfs/bin/mmperfmon config update SMBStats.restrict=cesNodes SMBStats.period=10`

Configure NFS sensors

The configure NFS sensor DMP assists you to configure NFS sensors.

The following are the details of the corresponding event:

- **Event ID:** nfs_sensors_not_configured

- **Problem:** The configuration details of the NFS sensor is not available in the sensor configuration.
- **Solution:** The sensor configuration is stored in a temporary file that is located at: `/var/lib/mmfs/gui/tmp/sensorDMP.txt`. The DMP provides options to enter the following details in the `sensorDMP.txt` file and later add them to the configuration by using the **mmperfmon config add** command.

Table 7. NFS sensor configuration example			
Sensor	Restrict to nodes	Intervals	Contents of the sensorDMP.txt file
NFSIO	Node class - cesNodes	1, 5, 10, 15, 30 Default value is 10.	<pre>sensors={ name = "sensorName" period = period proxyCmd = "/opt/IBM/zimon/ GaneshaProxy" restrict = "cesNodes" type = "Generic" }</pre>

Only users with *ProtocolAdministrator*, *SystemAdministrator*, *SecurityAdministrator*, and *Administrator* roles can use this DMP to configure NFS sensor.

After you complete the steps in the DMP, refresh the configuration by issuing the following command:

```
/usr/lpp/mmfs/bin/mmhealth node show nfs --refresh -N cesNodes
```

Issue the **mmperfmon config show** command to verify whether the NFS sensor is configured properly.

Configure SMB sensors

The configure SMB sensor DMP assists you to configure SMB sensors.

The following are the details of the corresponding event:

- **Event ID:** `smb_sensors_not_configured`
- **Problem:** The configuration details of the SMB sensor is not available in the sensor configuration.
- **Solution:** The sensor configuration is stored in a temporary file that is located at: `/var/lib/mmfs/gui/tmp/sensorDMP.txt`. The DMP provides options to enter the following details in the `sensorDMP.txt` file and later add them to the configuration by using the **mmperfmon config add** command.

Table 8. SMB sensor configuration example			
Sensor	Restrict to nodes	Intervals	Contents of the sensorDMP.txt file
SMBStats SMBGlobalStats	Node class - cesNodes	1, 5, 10, 15, 30 Default value is 10.	<pre>sensors={ name = "sensorName" period = period restrict = "cesNodes" type = "Generic" }</pre>

Only users with *ProtocolAdministrator*, *SystemAdministrator*, *SecurityAdministrator*, and *Administrator* roles can use this DMP to configure SMB sensor.

After you complete the steps in the DMP, refresh the configuration by issuing the following command:

```
/usr/lpp/mmfs/bin/mmhealth node show SMB --refresh -N cesNodes
```

Issue the **mmperfmon config show** command to verify whether the SMB sensor is configured properly.

Mount file system if it must be mounted

The mount file system DMP assists you to mount the file systems that must be mounted.

The following are the details of the corresponding event:

- **Event ID:** `unmounted_fs_check`
- **Problem:** A file system is assumed to be mounted all the time because it is configured to mount automatically, but the file system is currently not mounted on all nodes.
- **Solution:** Mount the file system on the node where it is not mounted.

Only users with *ProtocolAdministrator*, *SystemAdministrator*, *SecurityAdministrator*, and *Administrator* roles can use this DMP to mount the file systems on the required nodes.

If there is more than one instance of *unmounted_fs_check* event for the file system, you can choose whether to mount the file system on all nodes where it is not mounted but supposed to be mounted.

The DMP issues the following command for mounting the file system on one node:

```
mmmount Filesystem -N Node
```

The DMP issues the following command for mounting the file system on several nodes if automatic mount is not included:

```
mmmount Filesystem -N all
```

The DMP issues the following command for mounting the file system on certain nodes if automatic mount is not included in those nodes:

```
mmmount Filesystem -N Nodes (comma-separated list)
```

Note: Nodes where the file `/var/mmfs/etc/ignoreStartupMount.filesystem` or `/var/mmfs/etc/ignoreStartupMount` exists are excluded from automatic mount of this file system.

After running the **mmmount** command, the DMP waits until the *unmounted_fs_check* event disappear from the event list. If the *unmounted_fs_check* event does not get removed from the event list after 120 seconds, a warning message is displayed.

Start the GUI service on the remote nodes

You can start the GUI service on the remote nodes by using this DMP.

The following are the details of the corresponding event:

- **Event ID:** `gui_down`
- **Problem:** A GUI service is supposed to be running but it is down.
- **Solution:** Start the GUI service.
- **Limitation:** This DMP can only be used if GUI service is down on the remote nodes.

Only users with *ProtocolAdministrator*, *SystemAdministrator*, *SecurityAdministrator*, and *Administrator* roles can use this DMP to mount the file systems on the required nodes.

The DMP issues the **systemctl restart gpfsGui** command to start the GUI service on the remote node.

After running the **mmmount** command, the DMP waits until the *gui_down* event disappears from the event list. If the *gui_down* event does not get removed from the event list after 120 seconds, a warning message is displayed.

Chapter 12. References

The IBM Elastic Storage Server displays a warning or error message when it encounters an issue that needs user attention. The message severity tags indicate the severity of the issue

Events

The recorded events are stored in the local database on each node. The user can get a list of recorded events by using the **mmhealth node eventlog** command. Users can use the **mmhealth node show** or **mmhealth cluster show** commands to display the active events in the node and cluster respectively.

For more details, see the *mmces* and *mmhealth* commands in the *IBM Spectrum Scale: Command and Programming Reference* in [IBM Spectrum Scale documentation](#). The recorded events can also be displayed through the GUI.

When you upgrade to IBM Spectrum Scale 5.0.5.3 or a later version, the nodes where no `sqlite3` package is installed have their RAS event logs converted to a new database format to prevent known issues. The old RAS event log is emptied automatically. You can verify that the event log is emptied either by using the **mmhealth node eventlog** command or in the IBM Spectrum Scale GUI.

Note: The event logs are updated only the first time IBM Spectrum Scale is upgraded to version 5.0.5.3 or higher.

The following sections list the RAS events that are applicable to various components of the IBM Spectrum Scale system:

Array events

The following table lists the events that are created for the *Array* component.

Table 9. Events for the Array component						
Event	Event Type	Severity	Message	Description	Cause	User Action
gnr_array_found	INFO_ADD_ENTITY	INFO	GNR declustered array {0} was found.	A GNR declustered array listed in the IBM Spectrum Scale configuration was detected.		N/A
gnr_array_needservice	STATE_CHANGE	WARNING	GNR declustered array {0} needs service.	The declustered array state needs service.	N/A	N/A
gnr_array_ok	STATE_CHANGE	INFO	GNR declustered array {0} is ok.	The declustered array state is ok.	N/A	N/A
gnr_array_unknown	STATE_CHANGE	WARNING	GNR declustered array {0} is in unknown state.	The declustered array state is unknown.	N/A	N/A
gnr_array_vanished	INFO_DELETE_ENTITY	INFO	GNR declustered array {0} has vanished.	A GNR declustered array listed in the IBM Spectrum Scale configuration was not detected.	A GNR declustered array, listed in the IBM Spectrum Scale configuration as mounted before, is not found. This could be a valid situation	Run the mmfsrecoverygroup command to verify that all the expected GNR declustered arrays exist.

Enclosure events

The following table lists the events that are created for the *Enclosure* component.

Table 10. Events for the Enclosure component						
Event	Event Type	Severity	Message	Description	Cause	User Action
adapter_bios_notavail	STATE_CHANGE	WARNIN G	The bios level of adapter {0} is not available.	The bios level of the adapter is not available.	N/A	Check the installed BIOS level using the mmIsfirmware command.
adapter_bios_ok	STATE_CHANGE	INFO	The BIOS level of adapter {0} is correct.	The BIOS level of the adapter is correct.	N/A	N/A
adapter_bios_wrong	STATE_CHANGE	WARNIN G	The bios level of adapter {0} is wrong.	The bios level of the adapter is wrong.	N/A	Check the installed BIOS level using the mmIsfirmware command.
adapter_firmware_notavail	STATE_CHANGE	WARNIN G	The firmware level of adapter {0} is not available.	The firmware level of the adapter is not available.	N/A	Check the installed BIOS level using the mmIsfirmware command.
adapter_firmware_ok	STATE_CHANGE	INFO	The firmware level of adapter {0} is correct.	The firmware level of the adapter is correct.	N/A	N/A
adapter_firmware_wrong	STATE_CHANGE	WARNIN G	The firmware level of adapter {0} is wrong.	The firmware level of the adapter is wrong.	N/A	Check the installed BIOS level using the mmIsfirmware command.
current_failed	STATE_CHANGE	ERROR	currentSensor {0} failed.	The currentSensor state is failed.	N/A	N/A
current_ok	STATE_CHANGE	INFO	currentSensor {0} is ok.	The currentSensor state is ok.	N/A	N/A
current_warn	STATE_CHANGE	WARNIN G	currentSensor {0} is degraded.	The currentSensor state is degraded.	N/A	N/A
dcm_drawer_open	STATE_CHANGE	WARNIN G	DCM {0} drawer is open.	The DCM drawer is open.	N/A	N/A
dcm_failed	STATE_CHANGE	WARNIN G	DCM {0} is failed.	The DCM state is failed.	N/A	N/A
dcm_not_available	STATE_CHANGE	WARNIN G	DCM {0} is not available.	The DCM is not installed or not responding.	N/A	N/A
dcm_ok	STATE_CHANGE	INFO	DCM {id[1]} is ok.	The DCM state is ok.	N/A	N/A
drawer_failed	STATE_CHANGE	ERROR	drawer {0} is failed.	The drawer state is failed.	N/A	N/A
drawer_ok	STATE_CHANGE	INFO	drawer {0} is ok.	The drawer state is ok.	N/A	N/A
drive_firmware_notavail	STATE_CHANGE	WARNIN G	The firmware level of drive {0} is not available.	The firmware level of the drive is not available.	N/A	Check the installed firmware level using the mmIsfirmware command.
drive_firmware_ok	STATE_CHANGE	INFO	The firmware level of drive {0} is correct.	The firmware level of the drive is correct.	N/A	N/A
drive_firmware_wrong	STATE_CHANGE	WARNIN G	The firmware level of drive {0} is wrong.	The firmware level of the drive is wrong.	N/A	Check the installed firmware level using the mmIsfirmware command.
enclosure_data	STATE_CHANGE	INFO	Enclosure data found.	Successfully queried the enclosure details.	The mmIsenclosure all -l -y command reports enclosure data.	N/A

Table 10. Events for the Enclosure component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
enclosure_firmware_notavail	STATE_CHANGE	WARNIN G	The firmware level of enclosure {0} is not available.	The firmware level of the enclosure is not available.	N/A	Check the installed firmware level using the mmlsfirmware command.
enclosure_firmware_ok	STATE_CHANGE	INFO	The firmware level of enclosure {0} is correct.	The firmware level of the enclosure is correct.	N/A	N/A
enclosure_firmware_unknown	STATE_CHANGE	WARNIN G	The firmware level of enclosure {0} is unknown.	The SAS card is unable to read enclosure firmware.	The SAS card does not report the enclosure firmware.	Check the SAS connectivity from node to enclosure. Use the mmlsrecoverygroup rg_name -L --pdisk command to verify if all the paths to pdisk are available. Check the SAS connectivity using a combination of the mmgetpdisktopology and the topsummary command. If there is an issue with the SAS HBA or SAS Cable, reboot the node to see if this resolves the issue. If not contact your IBM representative.
enclosure_firmware_wrong	STATE_CHANGE	WARNIN G	The firmware level of enclosure {0} is wrong.	The firmware level of the enclosure is wrong.	N/A	Check the installed firmware level using mmlsfirmware command.
enclosure_found	INFO_ADD_ENTITY	INFO	Enclosure {0} was found.	A GNR enclosure listed in the IBM Spectrum Scale configuration was detected.	N/A	N/A
enclosure_needservice	STATE_CHANGE	WARNIN G	Enclosure {0} needs service.	The enclosure needs service.	N/A	N/A
enclosure_ok	STATE_CHANGE	INFO	Enclosure {0} is ok.	The enclosure state is ok.	N/A	N/A
enclosure_unknown	STATE_CHANGE	WARNIN G	Enclosure state {0} is unknown.	The enclosure state is unknown.	N/A	N/A
enclosure_vanished	INFO_DELETE_ENTITY	INFO	Enclosure {0} has vanished.	A GNR enclosure listed in the IBM Spectrum Scale configuration was not detected.	A GNR enclosure, listed in the IBM Spectrum Scale configuration as mounted before, is not found. This could be a valid situation.	Run the mmlsenclosure command to verify that all expected enclosures exist.
esm_absent	STATE_CHANGE	WARNIN G	ESM {0} is absent.	The ESM state is not installed .	N/A	N/A
esm_failed	STATE_CHANGE	WARNIN G	ESM {0} is failed.	The ESM state is failed.	N/A	N/A
esm_ok	STATE_CHANGE	INFO	ESM {0} is ok.	The ESM state is ok.	N/A	N/A
expander_absent	STATE_CHANGE	WARNIN G	expander {0} is absent.	The expander is absent.	N/A	N/A
expander_failed	STATE_CHANGE	ERROR	expander {0} is failed.	The expander state is failed.	N/A	N/A
expander_ok	STATE_CHANGE	INFO	expander {0} is ok.	The expander state is ok.	N/A	N/A

Table 10. Events for the Enclosure component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
fan_failed	STATE_CHANGE	WARNIN G	Fan {0} is failed.	The fan state is failed.	N/A	N/A
fan_ok	STATE_CHANGE	INFO	Fan {0} is ok.	The fan state is ok.	N/A	N/A
fan_speed_high	STATE_CHANGE	WARNIN G	Fan {0} speed is too high	The fan speed is out of the tolerance range	N/A	Check the enclosure cooling module LEDs for fan faults.
fan_speed_low	STATE_CHANGE	WARNIN G	Fan {0} speed is too low	The fan speed is out of the tolerance range	N/A	Check the enclosure cooling module LEDs for fan faults.
no_enclosure_data	STATE_CHANGE	WARNIN G	Enclosure data and state information cannot be queried.	Cannot query the enclosure details. State reporting for all enclosures and canisters will be incorrect.	The mm1senclosure all -L -Y command fails to report any enclosure data.	Run the mm1senclosure command to check for errors. Use the 1smod command to verify that the pemsmo d is loaded.
power_high_current	STATE_CHANGE	WARNIN G	Power supply {0} reports high current.	The DC power supply current is greater than the threshold.	N/A	N/A
power_high_voltage	STATE_CHANGE	WARNIN G	Power supply {0} reports high voltage.	The DC power supply voltage is greater than the threshold.	N/A	N/A
power_no_power	STATE_CHANGE	WARNIN G	Power supply {0} has no power.	Power supply has no input AC power. The power supply may be turned off or disconnected from the AC supply.	N/A	N/A
power_supply_absent	STATE_CHANGE	WARNIN G	Power supply {0} is missing.	The power supply is missing	N/A	N/A
power_supply_failed	STATE_CHANGE	WARNIN G	Power supply {0} is failed.	The power supply state is failed.	N/A	N/A
power_supply_off	STATE_CHANGE	WARNIN G	Power supply {0} is off.	The power supply is not providing power.	N/A	N/A
power_supply_ok	STATE_CHANGE	INFO	Power supply {0} is ok.	The power supply state is ok.	N/A	N/A
power_switched_off	STATE_CHANGE	WARNIN G	Power supply {0} is switched off.	The requested on bit is off, indicating that the power supply has not been manually turned on or been requested to turn on by setting the requested on bit.	N/A	N/A
sideplane_failed	STATE_CHANGE	ERROR	sideplane {0} failed.	The sideplane state is failed.	N/A	N/A
sideplane_ok	STATE_CHANGE	INFO	sideplane {0} is ok.	The sideplane state is ok.	N/A	N/A
temp_bus_failed	STATE_CHANGE	WARNIN G	Temperature sensor {0} I2C bus is failed.	The temperature sensor I2C bus has failed.	N/A	N/A
temp_high_critical	STATE_CHANGE	WARNIN G	Temperature sensor {0} measured a high temperature value.	The temperature has exceeded the actual high critical threshold value for at least one sensor.	N/A	N/A

Table 10. Events for the Enclosure component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
temp_high_warn	STATE_CHANGE	WARNIN G	Temperature sensor {0} measured a high temperature value.	The temperature has exceeded the actual high warning threshold value for at least one sensor.	N/A	N/A
temp_low_critical	STATE_CHANGE	WARNIN G	Temperature sensor {0} measured a low temperature value.	The temperature has fallen below the actual low critical threshold value for at least one sensor.	N/A	N/A
temp_low_warn	STATE_CHANGE	WARNIN G	Temperature sensor {0} measured a low temperature value.	The temperature has fallen below the actual low warning threshold value for at least one sensor.	N/A	N/A
temp_sensor_failed	STATE_CHANGE	WARNIN G	Temperature sensor {0} is failed.	The temperature sensor state is failed.	N/A	N/A
temp_sensor_ok	STATE_CHANGE	INFO	Temperature sensor {0} is ok.	The temperature sensor state is ok.	N/A	N/A
voltage_bus_failed	STATE_CHANGE	WARNIN G	Voltage sensor {0} I2C bus is failed.	The voltage sensor I2C bus has failed.	N/A	N/A
voltage_high_critical	STATE_CHANGE	WARNIN G	Voltage sensor {0} measured a high voltage value.	The voltage has exceeded the actual high critical threshold value for at least one sensor.	N/A	N/A
voltage_high_warn	STATE_CHANGE	WARNIN G	Voltage sensor {0} measured a high voltage value.	The voltage has exceeded the actual high warning threshold value for at least one sensor.	N/A	N/A
voltage_low_critical	STATE_CHANGE	WARNIN G	Voltage sensor {0} measured a low voltage value.	The voltage has fallen below the actual low critical threshold value for at least one sensor.	N/A	N/A
voltage_low_warn	STATE_CHANGE	WARNIN G	Voltage sensor {0} measured a low voltage value.	The voltage has fallen below the actual low warning threshold value for at least one sensor.	N/A	N/A
voltage_sensor_failed	STATE_CHANGE	WARNIN G	Voltage sensor {0} is failed.	The voltage sensor state is failed.	N/A	N/A
voltage_sensor_ok	STATE_CHANGE	INFO	Voltage sensor {0} is ok.	The voltage sensor state is ok.	N/A	N/A

Virtual disk events

The following table lists the events that are created for the *Virtual disk* component.

Table 11. Events for the virtual disk component

Event	Event Type	Severity	Message	Description	Cause	User Action
gnr_vdisk_critical	STATE_CHANGE	ERROR	GNR vdisk {0} is critical degraded.	The vdisk state is critical degraded.	N/A	N/A
gnr_vdisk_degraded	STATE_CHANGE	WARNING	GNR vdisk {0} is degraded.	The vdisk state is degraded.	N/A	N/A
gnr_vdisk_found	INFO_ADD_ENTIT Y	INFO	GNR vdisk {0} was found.	A GNR vdisk listed in the IBM Spectrum Scale configuration was detected.	N/A	N/A

Table 11. Events for the virtual disk component (continued)						
Event	Event Type	Severity	Message	Description	Cause	User Action
gnr_vdisk_offline	STATE_CHANGE	ERROR	GNR vdisk {0} is offline.	The vdisk state is offline.	N/A	N/A
gnr_vdisk_ok	STATE_CHANGE	INFO	GNR vdisk {0} is ok.	The vdisk state is ok.	N/A	N/A
gnr_vdisk_unknown	STATE_CHANGE	WARNING	GNR vdisk {0} is unknown.	The vdisk state is unknown.	N/A	N/A
gnr_vdisk_vanished	INFO_DELETE_ENTITY	INFO	GNR vdisk {0} has vanished.	A GNR vdisk listed in the IBM Spectrum Scale configuration was not detected.	A GNR vdisk, listed in the IBM Spectrum Scale configuration as mounted before, is not found. This could be a valid situation.	Run the mmlsvdisk command to verify that all expected GNR vdisk exist.

Physical disk events

The following table lists the events that are created for the *Physical disk* component.

Table 12. Events for the physical disk component						
Event	Event Type	Severity	Message	Description	Cause	User Action
gnr_nvram_degraded	STATE_CHANGE	WARNING	The NVDIMM of the pdisk {0} is degraded.	The NVRAM drive of the disk is in degraded state.	The tslsnvramstatus command shows degraded state for the NVRAM drive of the disk.	N/A
gnr_nvram_disarmed	STATE_CHANGE	ERROR	The NVDIMM of the pdisk {0} is disarmed.	NVDIMM is unable to preserve future content.	The tslsnvramstatus command reports disarmed failure condition for the NVRAM drive of the disk.	Identify the NVDIMM cards or BPM, which encountered the errors from FSP log or call home data, and replace the faulty NVDIMM cards, BPM or both as soon as possible.
gnr_nvram_erased	STATE_CHANGE	ERROR	The NVDIMM of the pdisk {0} reports erased image.	Image erased. The NVDIMM contents not persisted.	The tslsnvramstatus command reports erased failure condition for the NVRAM drive of the disk.	Verify that any NVDIMM cards, BPM encountered any errors from FSP log or call home data. If any then replace the faulty NVDIMM cards, BPM or both as soon as possible. If no errors are found, errors are found then try to add the drive back to RG.
gnr_nvram_error	STATE_CHANGE	ERROR	The NVDIMM of the pdisk {0} is failed.	The NVRAM drive of the disk is in error state.	The tslsnvramstatus command shows failed state for the NVRAM drive of the disk.	N/A

Table 12. Events for the physical disk component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
gnr_nvram_ok	STATE_CHANGE	INFO	The NVDIMM of the pdisk {0} is normal.	The NVRAM drive of the disk is in normal state.	N/A	N/A
gnr_nvram_persist_error	STATE_CHANGE	ERROR	The NVDIMM of the pdisk {0} might not persist.	NVDIMM failed to save or restore the memory contents.	The tslnvramstatus command reports FailToPersist failure condition for the NVRAM drive of the disk.	Identify the NVDIMM cards or BPM, which encountered the errors from FSP log or call home data, and replace the faulty NVDIMM cards, BPM or both as soon as possible.
gnr_nvram_unhealthy	STATE_CHANGE	WARNING	The NVDIMM of the pdisk {0} is unhealthy.	Error is detected but save or restore might still work for the NVRAM drive of the disk.	The tslnvramstatus command reports unhealthy failure condition for the NVRAM drive of the disk.	Identify the NVDIMM cards or BPM, which encountered the errors from FSP log or call home data, and replace the faulty NVDIMM cards, BPM or both.
gnr_pdisk_degraded	WARNING	WARNING	GNR pdisk {0} is degraded.	The pdisk state is degraded.	N/A	N/A
gnr_pdisk_diagnosing	INFO	WARNING	GNR pdisk {0} is diagnosing.	The pdisk state is diagnosing.	N/A	N/A
gnr_pdisk_draining	STATE_CHANGE	ERROR	GNR pdisk {0} is draining.	The pdisk state is draining.	N/A	N/A
gnr_pdisk_disks	STATE_CHANGE	INFO	Pdisks found on this node.	Pdisks found		N/A
gnr_pdisk_found	INFO_ADD_ENTITY	INFO	GNR pdisk {0} was found.	A GNR pdisk listed in the IBM Spectrum Scale configuration was detected.	N/A	N/A
gnr_pdisk_maintenance	STATE_CHANGE	WARNING	GNR pdisk {0} is in maintenance.	The GNR pdisk is in maintenance because the state is either suspended, serviceDrain, pathMaintenance, or deleting. This might be caused by some administration commands like mmdeildisk .	The mmispdisk command displays maintenance user condition for the disk.	Complete the maintenance action. Contact IBM support if you are not sure how to solve this problem.
gnr_pdisk_missing	STATE_CHANGE	WARNING	GNR pdisk {0} is missing.	The pdisk state is missing.	N/A	N/A
gnr_pdisk_needanalysis	STATE_CHANGE	ERROR	GNR pdisk {0} needs analysis.	The GNR pdisk has a problem that has to be analyzed and solved by an expert.	The mmispdisk command displays attention user condition for the disk.	Contact IBM support if you are not sure how to solve this problem.
gnr_pdisk_nodisks	STATE_CHANGE	INFO	No pdisks found on this node.	No pdisks found, but some pdisks are expected on recovery group nodes.	The mmvdisk pdisk list command returned no pdisks.	Run the mmvdisk pdisk list command to verify if this is correct.
gnr_pdisk_ok	STATE_CHANGE	INFO	GNR pdisk {0} is ok.	The pdisk state is ok.	N/A	N/A
gnr_pdisk_replaceable	STATE_CHANGE	ERROR	GNR pdisk {0} is replaceable.	The pdisk state is replaceable.	N/A	N/A
gnr_pdisk_sedlocked	STATE_CHANGE	ERROR	GNR pdisk {0} is locked (Self-encrypting drive).	A self-encrypting drive which has encryption enabled is locked. GNR does not have access to any data on the drive.	The mmispdisk command shows that the pdisk state contains sedLocked.	The drive must be unlocked to be used by GNR.

Table 12. Events for the physical disk component (continued)						
Event	Event Type	Severity	Message	Description	Cause	User Action
gnr_pdisk_unknown	STATE_CHANGE	WARNING	GNR pdisks are in unknown state.	The pdisk state is unknown.	N/A	N/A
gnr_pdisk_vanished	INFO_DELETE_ENTITY	INFO	GNR pdisk {0} has vanished.	A GNR pdisk listed in the IBM Spectrum Scale configuration was not detected.	A GNR pdisk, listed in the IBM Spectrum Scale configuration as mounted before, is not found. This could be a valid situation.	Run the mmlspdisk command to verify that all expected GNR pdisk exist.
gnr_pdisk_vwce	STATE_CHANGE	ERROR	GNR pdisk {0} has volatile write cache enabled.	Volatile write cache is enabled on the drive. Already committed writes could be lost in case of power loss. GNR will read-only from this disk.	The mmlspdisk command shows that the pdisk state contains VWCE.	Check why the volatile write cache is enabled (e.g. new drive added with wrong default, wrong UDEV rules) and fix the modes using the sg_wr_modes command.
ssd_endurance_ok	STATE_CHANGE	INFO	The ssdEndurance Percentage of GNR pdisk {0} is ok.	The ssdEndurance Percentage value is ok.	N/A	N/A
ssd_endurance_warn	STATE_CHANGE	WARNING	The ssdEndurance Percentage of GNR pdisk {0} is on a warning value.	The ssdEndurance Percentage value is warning.	The ssdEndurance Percentage value of the pdisk is between 95 and 100.	SSDs have a finite lifetime based on the number of drive writes per day. The ssd-endurance-percentage values actually reported will be a number between 0 and 255. This value indicates the percentage of life that is used by the drive. The value 0 indicates that full life remains, and 100 indicates that the drive is at or past its end of life. The drive must be replaced when the value exceeds 100", "state": "DEGRADED" }.

Recovery group events

The following table lists the events that are created for the *Recovery group* component.

Table 13. Events for the Recovery group component						
Event	Event Type	Severity	Message	Description	Cause	User Action
gnr_rg_failed	STATE_CHANGE	ERROR	GNR recoverygroup {0} is not active.	The recovery group is not active.	N/A	N/A
gnr_rg_found	INFO_ADD_ENTITY	INFO	GNR recovery group {0} was found.	A GNR recovery group listed in the IBM Spectrum Scale configuration was detected.	N/A	N/A
gnr_rg_ok	STATE_CHANGE	INFO	GNR recoverygroup {0} is ok.	The recovery group is ok.	N/A	N/A

Table 13. Events for the Recovery group component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
gnr_rg_vanished	INFO_DELETE_ENTITY	INFO	GNR recovery group {0} has vanished.	A GNR recovery group listed in the IBM Spectrum Scale configuration was not detected.	A GNR recovery group, listed in the IBM Spectrum Scale configuration as mounted before, is not found. This could be a valid situation.	Run the mmfsrecoverygroup command to verify that all expected GNR recovery group exist.

Server events

The following table lists the events that are created for the *Server* component.

Server events

Table 14. Server events

Event	Event Type	Severity	Message	Description	Cause	User Action
cpu_peci_ok	STATE_CHANGE	INFO	PECI state of CPU {0} is ok.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
cpu_peci_failed	STATE_CHANGE	ERROR	PECI state of CPU {0} failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
cpu_qpi_link_ok	STATE_CHANGE	INFO	QPI Link of CPU {0} is ok.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
cpu_qpi_link_failed	STATE_CHANGE	ERROR	QPI Link of CPU {0} is failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
cpu_temperature_ok	STATE_CHANGE	ERROR	QPI Link of CPU {0} is failed.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
cpu_temperature_ok	STATE_CHANGE	INFO	CPU {0} temperature is normal ({1}).	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.

Table 14. Server events (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
cpu_temperature_failed	STATE_CHANGE			The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_power_supply_temp_ok	STATE_CHANGE	INFO	Temperature of Power Supply {0} is ok. ({1})	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_power_supply_temp_failed	STATE_CHANGE	ERROR	Temperature of Power Supply {0} is too high. ({1})	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_power_supply_oc_line_12V_ok	STATE_CHANGE	INFO	OC Line 12V of Power Supply {0} is ok.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_power_supply_oc_line_12V_failed	STATE_CHANGE	ERROR	OC Line 12V of Power Supply {0} failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_power_supply_ov_line_12V_ok	STATE_CHANGE	INFO	OV Line 12V of Power Supply {0} is ok.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_power_supply_ov_line_12V_failed	STATE_CHANGE	ERROR	OV Line 12V of Power Supply {0} failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_power_supply_uv_line_12V_ok	STATE_CHANGE	INFO	UV Line 12V of Power Supply {0} is ok.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_power_supply_uv_line_12V_failed	STATE_CHANGE	ERROR	UV Line 12V of Power Supply {0} failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_power_supply_aux_line_12V_ok	STATE_CHANGE	INFO	AUX Line 12V of Power Supply {0} is ok.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_power_supply_aux_line_12V_failed	STATE_CHANGE	ERROR	AUX Line 12V of Power Supply {0} failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_power_supply_fan_ok	STATE_CHANGE	INFO	Fan of Power Supply {0} is ok.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_power_supply_fan_failed	STATE_CHANGE	ERROR	Fan of Power Supply {0} failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.

Table 14. Server events (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
server_power_supply_voltage_ok	STATE_CHANGE	INFO	Voltage of Power Supply {0} is ok.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_power_supply_voltage_failed	STATE_CHANGE	ERROR	Voltage of Power Supply {0} is not ok.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_power_supply_ok	STATE_CHANGE	INFO	Power Supply {0} is ok.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_power_supply_failed	STATE_CHANGE	ERROR	Power Supply {0} failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
pci_riser_temp_ok	STATE_CHANGE	INFO	The temperature of PCI Riser {0} is ok. ({1})	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
pci_riser_temp_failed	STATE_CHANGE	ERROR	The temperature of PCI Riser {0} is too high. ({1})	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_fan_ok	STATE_CHANGE	INFO	Fan {0} is ok. ({1})	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_fan_failed	STATE_CHANGE	ERROR	Fan {0} failed. ({1})	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
dimmem_ok	STATE_CHANGE	INFO	DIMM {0} is ok.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
dimmem_failed	STATE_CHANGE	ERROR	DIMM {0} failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
pci_ok	STATE_CHANGE	INFO	PCI {0} is ok.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
pci_failed	STATE_CHANGE	ERROR	PCI {0} failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
fan_zone_ok	STATE_CHANGE	INFO	Fan Zone {0} is ok.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.

Table 14. Server events (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
fan_zone_failed	STATE_CHANGE	ERROR	Fan Zone {0} failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
drive_ok	STATE_CHANGE	INFO	Drive {0} is ok.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
drive_failed	STATE_CHANGE	ERROR	Drive {0} failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
dasd_backplane_ok	STATE_CHANGE	INFO	DASD Backplane {0} is ok.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
dasd_backplane_failed	STATE_CHANGE	ERROR	DASD Backplane {0} failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_cpu_ok	STATE_CHANGE	INFO	All CPUs of server {0} are fully available.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_cpu_failed	STATE_CHANGE	ERROR	At least one CPU of server {0} failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_dimm_ok	STATE_CHANGE	INFO	All DIMMs of server {0} are fully available.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_dimm_failed	STATE_CHANGE	ERROR	At least one DIMM of server {0} failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_pci_ok	STATE_CHANGE	INFO	All PCIs of server {0} are fully available.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_pci_failed	STATE_CHANGE	ERROR	At least one PCI of server {0} failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_ps_conf_ok	STATE_CHANGE	INFO	All Power Supply Configurations of server {0} are ok.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_ps_conf_failed	STATE_CHANGE	ERROR	At least one Power Supply Configuration of server {0} is not ok.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.

Table 14. Server events (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
server_ps_heavyload_ok	STATE_CHANGE	INFO	No Power Supplies of server {0} are under heavy load.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_ps_heavyload_failed	STATE_CHANGE	ERROR	At least one Power Supply of server {0} is under heavy load.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_ps_resource_ok	STATE_CHANGE	INFO	Power Supply resources of server {0} are ok.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_ps_resource_failed	STATE_CHANGE	ERROR	At least one Power Supply of server {0} has insufficient resources.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_ps_unit_ok	STATE_CHANGE	INFO	All Power Supply units of server {0} are fully available.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_ps_unit_failed	STATE_CHANGE	ERROR	At least one Power Supply unit of server {0} failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_ps_ambient_ok	STATE_CHANGE	INFO	Power Supply ambient of server {0} is ok.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_ps_ambient_failed	STATE_CHANGE	ERROR	At least one Power Supply ambient of server {0} is not okay.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_boot_status_ok	STATE_CHANGE	INFO	The boot status of server {0} is normal.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_boot_status_failed	STATE_CHANGE	ERROR	System Boot failed on server {0}.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_planar_ok	STATE_CHANGE	INFO	Planar state of server {0} is healthy, the voltage is normal ({1}).	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_planar_failed	STATE_CHANGE	ERROR	Planar state of server {0} is unhealthy, the voltage is too low or too high ({1}).	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_sys_board_ok	STATE_CHANGE	INFO	The system board of server {0} is healthy.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.

Table 14. Server events (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
server_sys_board_failed	STATE_CHANGE	ERROR	The system board of server {0} failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_system_event_log_ok	STATE_CHANGE	INFO	The system event log of server {0} operates normally.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_system_event_log_full	STATE_CHANGE	ERROR	The system event log of server {0} is full.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
server_ok	STATE_CHANGE	INFO	The server {0} is healthy.	The GUI checks the hardware state using xCAT.	The hardware part is ok.	None.
server_failed	STATE_CHANGE	ERROR	The server {0} failed.	The GUI checks the hardware state using xCAT.	The hardware part failed.	None.
hmc_event	STATE_CHANGE	INFO	HMC Event: {1}	The GUI collects events raised by the HMC.	An event from the HMC arrived.	None.

Canister events

The following table lists the events that are created for the *Canister* component.

Table 15. Events for the Canister component

Event	Event Type	Severity	Message	Description	Cause	User Action
bootdrive_mirror_degraded	STATE_CHANGE	WARNING	The bootdrive's mirroring is degraded.	The bootdrive's mirroring is degraded.	The tsplatformstat -a command returns a DEGRADED value for at least one partition.	N/A
bootdrive_installed	STATE_CHANGE	INFO	The bootdrive attached to port {0} is available.	The bootdrive is available.	The tsplatformstat -a command returns the bootdrives as expected.	N/A
bootdrive_mirror_failed	STATE_CHANGE	ERROR	The bootdrive's mirroring is failed.	The bootdrive's mirroring is failed.	The tsplatformstat -a command returns a FAILED value for at least one partition.	N/A
bootdrive_mirror_ok	STATE_CHANGE	INFO	The bootdrive's mirroring is OK.	The bootdrive's mirroring is OK.	The tsplatformstat -a command returns optimal for all partitions.	N/A
bootdrive_mirror_unconfigured	STATE_CHANGE	WARNING	The bootdrive's mirroring is unconfigured.	The bootdrive's mirroring is unconfigured.	The tsplatformstat -a command returns unconfigured for mirroring.	N/A

Table 15. Events for the Canister component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
bootdrive_missing	STATE_CHANGE	ERROR	The bootdrive on port {0} is missing or dead.	One bootdrive is missing or dead. Redundancy is not given anymore.	The tsplatformstat -a command returns only one instead of two bootdrives. Two drives are expected to ensure redundancy.	Inspect that the drive is correctly installed on the referenced port. Else insert or replace the drive.
bootdrive_smart_failed	STATE_CHANGE	ERROR	The smart assessment of bootdrive {0} attached to port {1} does not return OK.	The bootdrive's smart assessment does not return OK.	The tsplatformstat -a command does not return a PASSED value in the selfAssessment field for the bootdrive.	Verify the smart status of the bootdrive using tsplatformstat command or smartctl .
bootdrive_smart_ok	STATE_CHANGE	INFO	The smart assessment of bootdrive {0} attached to port {1} returns OK.	The bootdrive's smart assessment returns OK.	The tsplatformstat -a command returns a PASSED in the selfAssessment field for the bootdrive.	N/A
can_fan_failed	STATE_CHANGE	WARNING	Fan {0} is failed.	The fan state is failed.	The mmlsenclosure command reports the fan as failed.	Check the fan status by using the mmlsenclosure command. Replace the fan module in the canister.
can_fan_ok	STATE_CHANGE	INFO	Fan {0} is OK.	The fan state is OK.	The mmlsenclosure command reports the fan as working.	N/A
can_temp_bus_failed	STATE_CHANGE	WARNING	Temperature sensor {0} I2C bus is failed.	The temperature sensor I2C bus failed.	The mmlsenclosure command reports the temperature sensor with a failure.	Check the temperature status by using the mmlsenclosure command.
can_temp_high_critical	STATE_CHANGE	WARNING	Temperature sensor {0} measured a high temperature value.	The temperature exceeded the actual high critical threshold value for at least one sensor.	The mmlsenclosure command reports the temperature sensor with a failure.	Check the temperature status by using the mmlsenclosure command.
can_temp_high_warn	STATE_CHANGE	WARNING	Temperature sensor {0} measured a high temperature value.	The temperature exceeded the actual high warning threshold value for at least one sensor.	The mmlsenclosure command reports the temperature sensor with a failure.	Check the temperature status by using the mmlsenclosure command.
can_temp_low_critical	STATE_CHANGE	WARNING	Temperature sensor {0} measured a low temperature value.	The temperature has fallen below the actual low critical threshold value for at least one sensor.	The mmlsenclosure command reports the temperature sensor with a failure.	Check the temperature status by using the mmlsenclosure command.
can_temp_low_warn	STATE_CHANGE	WARNING	Temperature sensor {0} measured a low temperature value.	The temperature has fallen below the actual low warning threshold value for at least one sensor.	The mmlsenclosure command reports the temperature sensor with a failure.	Check the temperature status by using the mmlsenclosure command.
can_temp_sensor_failed	STATE_CHANGE	WARNING	Temperature sensor {0} is failed.	The temperature sensor state is failed.	The mmlsenclosure command reports the temperature sensor with a failure.	Check the temperature status by using the mmlsenclosure command. Replace the canister.
can_temp_sensor_ok	STATE_CHANGE	INFO	Temperature sensor {0} is OK.	The temperature sensor state is OK.	N/A	N/A

Table 15. Events for the Canister component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
canister_failed	STATE_CHANGE	ERROR	Canister {0} is failed.	The canister is reporting a failed hardware state. This might be caused by a failure of an underlying component. For example, the fan.	The mmIsenclosure command reports the canister as failed.	Check for detailed error events of canister components by using the mmhealth command. Inspect the output of mmIsenclosure all -L command for the referenced canister.
canister_ok	STATE_CHANGE	INFO	Canister {0} is OK.	The canister state is OK.	The mmIsenclosure command reports the canister as failed.	N/A
cpu_inspection_failed	STATE_CHANGE	ERROR	The inspection of the CPU slots found a mismatch	Number of populated CPU slots, number of enabled CPUs, number of CPU cores, number of CPU threads or CPU speed is not as expected.	The /opt/ibm/gss/tools/bin/ess3kplt command returned an InspectionPassed unequal to True value.	Check for specific events related to CPUs by using the mmhealth command. Inspect the output of the ess3kplt command for details.
cpu_inspection_passed	STATE_CHANGE	INFO	The CPUs of the canister are OK.	The CPU speed and number of populated CPU slots is as expected.	The /opt/ibm/gss/tools/bin/ess3kplt command returned an InspectionPassed equal to True value.	N/A
cpu_speed_ok	STATE_CHANGE	INFO	The CPU speed is OK.	The speed of all CPUs is as expected.	The /opt/ibm/gss/tools/bin/ess3kplt command returned no speed errors.	N/A
cpu_speed_wrong	STATE_CHANGE	ERROR	One or more CPUs have an unsupported speed.	The speed of one or more CPUs is not as expected. This configuration is not supported.	The /opt/ibm/gss/tools/bin/ess3kplt command returned one or more speed errors.	Inspect the output of the ess3kplt command to see which CPUs have an unsupported speed.
dimm_inspection_failed	STATE_CHANGE	ERROR	The inspection of the memory dimm slots found a failure.	The capacity, speed, or number of populated dimm slots is not as expected.	The /opt/ibm/gss/tools/bin/ess3kplt command returned an InspectionPassed unequal to True value.	Check for specific events related to dimms by using the mmhealth command. Inspect the output of the ess3kplt command for details.
dimm_inspection_passed	STATE_CHANGE	INFO	The memory dimms of the canister is OK.	The capacity, speed, and number of populated dimm slots is as expected.	The /opt/ibm/gss/tools/bin/ess3kplt command returned an InspectionPassed equal to True value.	N/A
dimm_size_ok	STATE_CHANGE	INFO	All installed memory dimms have the expected capacity.	The capacity of all populated memory dimm slots is as expected.	The /opt/ibm/gss/tools/bin/ess3kplt command returned no capacity errors.	N/A
dimm_size_wrong	STATE_CHANGE	ERROR	One or more memory dimm modules have an unsupported capacity.	The capacity of one or more memory dimm slots is not as expected. This configuration is not supported.	The /opt/ibm/gss/tools/bin/ess3kplt command returned some capacity errors.	Inspect the output of the ess3kplt command to see which memory dimm slots have an unsupported capacity and replace those dimm modules.
dimm_speed_ok	STATE_CHANGE	INFO	All installed memory dimms have the expected speed.	The speed of all populated memory dimm slots is as expected.	The /opt/ibm/gss/tools/bin/ess3kplt command returned no speed errors.	N/A

Table 15. Events for the Canister component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
dimm_speed_wrong	STATE_CHANGE	ERROR	One or more memory dimm modules have an unsupported speed.	The speed of one or more memory dimm slots is not as expected. This configuration is not supported.	The <code>/opt/ibm/gss/tools/bin/ess3kplt</code> command returned some speed errors.	Inspect the output of the <code>ess3kplt</code> command to see which memory dimm slots have an unsupported speed and replace those dimm modules.
pair_canister_missing	STATE_CHANGE	WARNING	Pair canister {0} is missing or dead.	Could not get the state of the pair canister. It might be missing or dead.	The <code>mm1senclosure</code> command reports only one canister instead of two.	Check for detailed error events of the referenced canister node by using the <code>mmhealth</code> command. Inspect the output of the <code>mm1senclosure all -L</code> command for the referenced canister.
pair_canister_visible	STATE_CHANGE	INFO	Pair canister {0} is visible.	Successfully get the state of the pair canister.	The <code>mm1senclosure</code> command reports both canisters.	N/A

Messages

This topic contains explanations for IBM Spectrum Scale RAID and ESS GUI messages.

For information about IBM Spectrum Scale messages, see the *IBM Spectrum Scale: Problem Determination Guide*.

Message severity tags

IBM Spectrum Scale and ESS GUI messages include message severity tags.

A severity tag is a one-character alphabetic code (**A** through **Z**).

For IBM Spectrum Scale messages, the severity tag is optionally followed by a colon (:) and a number, and surrounded by an opening and closing bracket (**[]**). For example:

```
[E] or [E:nnn]
```

If more than one substring within a message matches this pattern (for example, **[A]** or **[A:nnn]**), the severity tag is the first such matching string.

When the severity tag includes a numeric code (*nnn*), this is an error code associated with the message. If this were the only problem encountered by the command, the command return code would be *nnn*.

If a message does not have a severity tag, the message does not conform to this specification. You can determine the message severity by examining the text or any supplemental information provided in the message catalog, or by contacting the IBM Support Center.

Each message severity tag has an assigned priority.

For IBM Spectrum Scale messages, this priority can be used to filter the messages that are sent to the error log on Linux. Filtering is controlled with the `mmchconfig` attribute `systemLogLevel`. The default for `systemLogLevel` is `error`, which means that IBM Spectrum Scale will send all error **[E]**, critical **[X]**, and alert **[A]** messages to the error log. The values allowed for `systemLogLevel` are: `alert`, `critical`, `error`, `warning`, `notice`, `configuration`, `informational`, `detail`, or `debug`. Additionally, the value `none` can be specified so no messages are sent to the error log.

For IBM Spectrum Scale messages, alert **[A]** messages have the highest priority and debug **[B]** messages have the lowest priority. If the `systemLogLevel` default of `error` is changed, only messages with the specified severity and all those with a higher priority are sent to the error log.

The following table lists the IBM Spectrum Scale message severity tags in order of priority:

Table 16. IBM Spectrum Scale message severity tags ordered by priority

Severity tag	Type of message (systemLogLevel attribute)	Meaning
A	alert	Indicates a problem where action must be taken immediately. Notify the appropriate person to correct the problem.
X	critical	Indicates a critical condition that should be corrected immediately. The system discovered an internal inconsistency of some kind. Command execution might be halted or the system might attempt to continue despite the inconsistency. Report these errors to IBM.
E	error	Indicates an error condition. Command execution might or might not continue, but this error was likely caused by a persistent condition and will remain until corrected by some other program or administrative action. For example, a command operating on a single file or other GPFS object might terminate upon encountering any condition of severity E . As another example, a command operating on a list of files, finding that one of the files has permission bits set that disallow the operation, might continue to operate on all other files within the specified list of files.
W	warning	Indicates a problem, but command execution continues. The problem can be a transient inconsistency. It can be that the command has skipped some operations on some objects, or is reporting an irregularity that could be of interest. For example, if a multipass command operating on many files discovers during its second pass that a file that was present during the first pass is no longer present, the file might have been removed by another command or program.
N	notice	Indicates a normal but significant condition. These events are unusual, but are not error conditions, and could be summarized in an email to developers or administrators for spotting potential problems. No immediate action is required.
C	configuration	Indicates a configuration change; such as, creating a file system or removing a node from the cluster.
I	informational	Indicates normal operation. This message by itself indicates that nothing is wrong; no action is required.
D	detail	Indicates verbose operational messages; no is action required.
B	debug	Indicates debug-level messages that are useful to application developers for debugging purposes. This information is not useful during operations.

For ESS GUI messages, error messages (**(E)**) have the highest priority and informational messages (**(I)**) have the lowest priority.

The following table lists the ESS GUI message severity tags in order of priority:

Table 17. ESS GUI message severity tags ordered by priority

Severity tag	Type of message	Meaning
E	Error	Indicates a critical condition that should be corrected immediately. The system discovered an internal inconsistency of some kind. Command execution might be halted or the system might attempt to continue despite the inconsistency. Report these errors to IBM.

Table 17. ESS GUI message severity tags ordered by priority (continued)

Severity tag	Type of message	Meaning
W	warning	Indicates a problem, but command execution continues. The problem can be a transient inconsistency. It can be that the command has skipped some operations on some objects, or is reporting an irregularity that could be of interest. For example, if a multipass command operating on many files discovers during its second pass that a file that was present during the first pass is no longer present, the file might have been removed by another command or program.
I	informational	Indicates normal operation. This message by itself indicates that nothing is wrong; no action is required.

IBM Spectrum Scale RAID messages

This section lists the IBM Spectrum Scale RAID messages.

For information about the severity designations of these messages, see [“Message severity tags” on page 89](#).

6027-1850 [E] NSD-RAID services are not configured on node *nodeName*. Check the nsdRAIDTracks and nsdRAIDBufferPoolSizePct configuration attributes.

Explanation:

A IBM Spectrum Scale RAID command is being executed, but NSD-RAID services are not initialized either because the specified attributes have not been set or had invalid values.

User response:

Correct the attributes and restart the GPFS daemon.

6027-1851 [A] Cannot configure NSD-RAID services. The nsdRAIDBufferPoolSizePct of the pagepool must result in at least 128MiB of space.

Explanation:

The GPFS daemon is starting and cannot initialize the NSD-RAID services because of the memory consideration specified.

User response:

Correct the nsdRAIDBufferPoolSizePct attribute and restart the GPFS daemon.

6027-1852 [A] Cannot configure NSD-RAID services. nsdRAIDTracks is too large, the maximum on this node is *value*.

Explanation:

The GPFS daemon is starting and cannot initialize the NSD-RAID services because the nsdRAIDTracks attribute is too large.

User response:

Correct the nsdRAIDTracks attribute and restart the GPFS daemon.

6027-1853 [E] Recovery group *recoveryGroupName* does not exist or is not active.

Explanation:

A command was issued to a RAID recovery group that does not exist, or is not in the active state.

User response:

Retry the command with a valid RAID recovery group name or wait for the recovery group to become active.

6027-1854 [E] Cannot find declustered array *arrayName* in recovery group *recoveryGroupName*.

Explanation:

The specified declustered array name was not found in the RAID recovery group.

User response:

Specify a valid declustered array name within the RAID recovery group.

6027-1855 [E] Cannot find pdisk *pdiskName* in recovery group *recoveryGroupName*.

Explanation:

The specified pdisk was not found.

User response:

Retry the command with a valid pdisk name.

6027-1856 [E] Vdisk *vdiskName* not found.

Explanation:

The specified vdisk was not found.

User response:

Retry the command with a valid vdisk name.

6027-1857 [E] **A recovery group must contain between *number* and *number* pdisks.**

Explanation:

The number of pdisks specified is not valid.

User response:

Correct the input and retry the command.

6027-1858 [E] **Cannot create declustered array *arrayName*; there can be at most *number* declustered arrays in a recovery group.**

Explanation:

The number of declustered arrays allowed in a recovery group has been exceeded.

User response:

Reduce the number of declustered arrays in the input file and retry the command.

6027-1859 [E] **Sector size of pdisk *pdiskName* is invalid.**

Explanation:

All pdisks in a recovery group must have the same physical sector size.

User response:

Correct the input file to use a different disk and retry the command.

6027-1860 [E] **Pdisk *pdiskName* must have a capacity of at least *number* bytes.**

Explanation:

The pdisk must be at least as large as the indicated minimum size in order to be added to this declustered array.

User response:

Correct the input file and retry the command.

6027-1861 [W] **Size of pdisk *pdiskName* is too large for declustered array *arrayName*. Only *number* of *number* bytes of that capacity will be used.**

Explanation:

For optimal utilization of space, pdisks added to this declustered array should be no larger than the indicated maximum size. Only the indicated portion of the total capacity of the pdisk will be available for use.

User response:

Consider creating a new declustered array consisting of all larger pdisks.

6027-1862 [E] **Cannot add pdisk *pdiskName* to declustered array *arrayName*;**

there can be at most *number* pdisks in a declustered array.

Explanation:

The maximum number of pdisks that can be added to a declustered array was exceeded.

User response:

None.

6027-1863 [E] **Pdisk sizes within a declustered array cannot vary by more than *number*.**

Explanation:

The disk sizes within each declustered array must be nearly the same.

User response:

Create separate declustered arrays for each disk size.

6027-1864 [E] **[E] At least one declustered array must contain *number* + vdisk configuration data spares or more pdisks and be eligible to hold vdisk configuration data.**

Explanation:

When creating a new RAID recovery group, at least one of the declustered arrays in the recovery group must contain at least 2T+1 pdisks, where T is the maximum number of disk failures that can be tolerated within a declustered array. This is necessary in order to store the on-disk vdisk configuration data safely. This declustered array cannot have canHoldVCD set to no.

User response:

Supply at least the indicated number of pdisks in at least one declustered array of the recovery group, or do not specify canHoldVCD=no for that declustered array.

6027-1866 [E] **Disk descriptor for *diskName* refers to an existing NSD.**

Explanation:

A disk being added to a recovery group appears to already be in-use as an NSD disk.

User response:

Carefully check the disks given to tscrrecgroup, tsaddpdisk or tschcarrier. If you are certain the disk is not actually in-use, override the check by specifying the -v no option.

6027-1867 [E] **Disk descriptor for *diskName* refers to an existing pdisk.**

Explanation:

A disk being added to a recovery group appears to already be in-use as a pdisk.

User response:

Carefully check the disks given to tscrrecgroup, tsaddpdisk or tschcarrier. If you are certain

the disk is not actually in-use, override the check by specifying the -v no option.

6027-1869 [E] Error updating the recovery group descriptor.

Explanation:

Error occurred updating the RAID recovery group descriptor.

User response:

Retry the command.

6027-1870 [E] Recovery group name *name* is already in use.

Explanation:

The recovery group name already exists.

User response:

Choose a new recovery group name using the characters a-z, A-Z, 0-9, and underscore, at most 63 characters in length.

6027-1871 [E] There is only enough free space to allocate *number* spare(s) in declustered array *arrayName*.

Explanation:

Too many spares were specified.

User response:

Retry the command with a valid number of spares.

6027-1872 [E] Recovery group still contains vdisks.

Explanation:

RAID recovery groups that still contain vdisks cannot be deleted.

User response:

Delete any vdisks remaining in this RAID recovery group using the `tsdelvdisk` command before retrying this command.

6027-1873 [E] Pdisk creation failed for pdisk *pdiskName*: *err=errorNum*.

Explanation:

Pdisk creation failed because of the specified error.

User response:

None.

6027-1874 [E] Error adding pdisk to a recovery group.

Explanation:

`tsaddpdisk` failed to add new pdisks to a recovery group.

User response:

Check the list of pdisks in the -d or -F parameter of `tsaddpdisk`.

6027-1875 [E] Cannot delete the only declustered array.

Explanation:

Cannot delete the only remaining declustered array from a recovery group.

User response:

Instead, delete the entire recovery group.

6027-1876 [E] Cannot remove declustered array *arrayName* because it is the only remaining declustered array with at least *number* pdisks eligible to hold vdisk configuration data.

Explanation:

The command failed to remove a declustered array because no other declustered array in the recovery group has sufficient pdisks to store the on-disk recovery group descriptor at the required fault tolerance level.

User response:

Add pdisks to another declustered array in this recovery group before removing this one.

6027-1877 [E] Cannot remove declustered array *arrayName* because the array still contains vdisks.

Explanation:

Declassified arrays that still contain vdisks cannot be deleted.

User response:

Delete any vdisks remaining in this declustered array using the `tsdelvdisk` command before retrying this command.

6027-1878 [E] Cannot remove pdisk *pdiskName* because it is the last remaining pdisk in declustered array *arrayName*. Remove the declustered array instead.

Explanation:

The `tsdelvdisk` command can be used either to delete individual pdisks from a declustered array, or to delete a full declustered array from a recovery group. You cannot, however, delete a declustered array by deleting all of its pdisks -- at least one must remain.

User response:

Delete the declustered array instead of removing all of its pdisks.

6027-1879 [E] Cannot remove pdisk *pdiskName* because *arrayName* is the only remaining declustered array with at least *number* pdisks.

Explanation:

The command failed to remove a pdisk from a declustered array because no other declustered array in the recovery group has sufficient pdisks to store the on-disk recovery group descriptor at the required fault tolerance level.

User response:

Add pdisks to another declustered array in this recovery group before removing pdisks from this one.

6027-1880 [E] Cannot remove pdisk *pdiskName* because the number of pdisks in declustered array *arrayName* would fall below the code width of one or more of its vdisks.

Explanation:

The number of pdisks in a declustered array must be at least the maximum code width of any vdisk in the declustered array.

User response:

Either add pdisks or remove vdisks from the declustered array.

6027-1881 [E] Cannot remove pdisk *pdiskName* because of insufficient free space in declustered array *arrayName*.

Explanation:

The `tsdelpdisk` command could not delete a pdisk because there was not enough free space in the declustered array.

User response:

Either add pdisks or remove vdisks from the declustered array.

6027-1882 [E] Cannot remove pdisk *pdiskName*; unable to drain the data from the pdisk.

Explanation:

Pdisk deletion failed because the system could not find enough free space on other pdisks to drain all of the data from the disk.

User response:

Either add pdisks or remove vdisks from the declustered array.

6027-1883 [E] Pdisk *pdiskName* deletion failed: process interrupted.

Explanation:

Pdisk deletion failed because the deletion process was interrupted. This is most likely because of the recovery group failing over to a different server.

User response:

Retry the command.

6027-1884 [E] Missing or invalid vdisk name.

Explanation:

No vdisk name was given on the `tscrvdisk` command.

User response:

Specify a vdisk name using the characters a-z, A-Z, 0-9, and underscore of at most 63 characters in length.

6027-1885 [E] Vdisk block size must be a power of 2.

Explanation:

The `-B` or `--blockSize` parameter of `tscrvdisk` must be a power of 2.

User response:

Reissue the `tscrvdisk` command with a correct value for block size.

6027-1886 [E] Vdisk block size cannot exceed *maxBlockSize* (*number*).

Explanation:

The virtual block size of a vdisk cannot be larger than the value of the `maxblocksize` configuration attribute of the IBM Spectrum Scale `mmchconfig` command.

User response:

Use a smaller vdisk virtual block size, or increase the value of `maxBlockSize` using `mmchconfig maxblocksize=newSize`.

6027-1887 [E] Vdisk block size must be between *number* and *number* for the specified code.

Explanation:

An invalid vdisk block size was specified. The message lists the allowable range of block sizes.

User response:

Use a vdisk virtual block size within the range shown, or use a different vdisk RAID code.

6027-1888 [E] Recovery group already contains *number* vdisks.

Explanation:

The RAID recovery group already contains the maximum number of vdisks.

User response:

Create vdisks in another RAID recovery group, or delete one or more of the vdisks in the current RAID recovery group before retrying the `tscrvdisk` command.

6027-1889 [E] Vdisk name *vdiskName* is already in use.

Explanation:

The vdisk name given on the `tscrvdisk` command already exists.

User response:

Choose a new vdisk name less than 64 characters using the characters a-z, A-Z, 0-9, and underscore.

6027-1890 [E] A recovery group may only contain one log home vdisk.

Explanation:

A log vdisk already exists in the recovery group.

User response:

None.

6027-1891 [E] Cannot create vdisk before the log home vdisk is created.

Explanation:

The log vdisk must be the first vdisk created in a recovery group.

User response:

Retry the command after creating the log home vdisk.

6027-1892 [E] Log vdisks must use replication.

Explanation:

The log vdisk must use a RAID code that uses replication.

User response:

Retry the command with a valid RAID code.

6027-1893 [E] The declustered array must contain at least as many non-spare pdisks as the width of the code.

Explanation:

The RAID code specified requires a minimum number of disks larger than the size of the declustered array that was given.

User response:

Place the vdisk in a wider declustered array or use a narrower code.

6027-1894 [E] There is not enough space in the declustered array to create additional vdisks.

Explanation:

There is insufficient space in the declustered array to create even a minimum size vdisk with the given RAID code.

User response:

Add additional pdisks to the declustered array, reduce the number of spares or use a different RAID code.

6027-1895 [E] Unable to create vdisk *vdiskName* because there are too many failed pdisks in declustered array *declusteredArrayName*.

Explanation:

Cannot create the specified vdisk, because there are too many failed pdisks in the array.

User response:

Replace failed pdisks in the declustered array and allow time for rebalance operations to more evenly distribute the space.

6027-1896 [E] Insufficient memory for vdisk metadata.

Explanation:

There was not enough pinned memory for IBM Spectrum Scale to hold all of the metadata necessary to describe a vdisk.

User response:

Increase the size of the GPFS page pool.

6027-1897 [E] Error formatting vdisk.

Explanation:

An error occurred formatting the vdisk.

User response:

None.

6027-1898 [E] The log home vdisk cannot be destroyed if there are other vdisks.

Explanation:

The log home vdisk of a recovery group cannot be destroyed if vdisks other than the log tip vdisk still exist within the recovery group.

User response:

Remove the user vdisks and then retry the command.

6027-1899 [E] Vdisk *vdiskName* is still in use.

Explanation:

The vdisk named on the `tsdelvdisk` command is being used as an NSD disk.

User response:

Remove the vdisk with the `mmdeletnsd` command before attempting to delete it.

6027-3000 [E] No disk enclosures were found on the target node.

Explanation:

IBM Spectrum Scale is unable to communicate with any disk enclosures on the node serving the specified pdisks. This might be because there are no disk enclosures attached to the node, or it might indicate a problem in communicating with the disk enclosures. While the problem persists, disk maintenance with the `mmchcarrier` command is not available.

User response:

Check disk enclosure connections and run the command again. Use `mmaddpdisk --replace` as an alternative method of replacing failed disks.

6027-3001 [E] Location of pdisk *pdiskName* of recovery group *recoveryGroupName* is not known.

Explanation:

IBM Spectrum Scale is unable to find the location of the given pdisk.

User response:

Check the disk enclosure hardware.

6027-3002 [E] Disk location code *locationCode* is not known.

Explanation:

A disk location code specified on the command line was not found.

User response:

Check the disk location code.

6027-3003 [E] Disk location code *locationCode* was specified more than once.

Explanation:

The same disk location code was specified more than once in the `tschcarrier` command.

User response:

Check the command usage and run again.

6027-3004 [E] Disk location codes *locationCode* and *locationCode* are not in the same disk carrier.

Explanation:

The `tschcarrier` command cannot be used to operate on more than one disk carrier at a time.

User response:

Check the command usage and rerun.

6027-3005 [W] Pdisk in location *locationCode* is controlled by recovery group *recoveryGroupName*.

Explanation:

The `tschcarrier` command detected that a pdisk in the indicated location is controlled by a different recovery group than the one specified.

User response:

Check the disk location code and recovery group name.

6027-3006 [W] Pdisk in location *locationCode* is controlled by recovery group *id idNumber*.

Explanation:

The `tschcarrier` command detected that a pdisk in the indicated location is controlled by a different recovery group than the one specified.

User response:

Check the disk location code and recovery group name.

6027-3007 [E] Carrier contains pdisks from more than one recovery group.

Explanation:

The `tschcarrier` command detected that a disk carrier contains pdisks controlled by more than one recovery group.

User response:

Use the `tschpdisk` command to bring the pdisks in each of the other recovery groups offline and then rerun the command using the `--force-RG` flag.

6027-3008 [E] Incorrect recovery group given for location.

Explanation:

The `mmchcarrier` command detected that the specified recovery group name given does not match that of the pdisk in the specified location.

User response:

Check the disk location code and recovery group name. If you are sure that the disks in the carrier are not being used by other recovery groups, it is possible to override the check using the `--force-RG` flag. Use this flag with caution as it can cause disk errors and potential data loss in other recovery groups.

6027-3009 [E] Pdisk *pdiskName* of recovery group *recoveryGroupName* is not currently scheduled for replacement.

Explanation:

A pdisk specified in a `tschcarrier` or `tsaddpdisk` command is not currently scheduled for replacement.

User response:

Make sure the correct disk location code or pdisk name was given. For the `mmchcarrier` command, the `--force-release` option can be used to override the check.

6027-3010 [E] Command interrupted.

Explanation:

The `mmchcarrier` command was interrupted by a conflicting operation, for example the `mmchpdisk --resume` command on the same pdisk.

User response:

Run the `mmchcarrier` command again.

6027-3011 [W] Disk location *locationCode* failed to power off.

Explanation:

The `mmchcarrier` command detected an error when trying to power off a disk.

User response:

Check the disk enclosure hardware. If the disk carrier has a lock and does not unlock, try running the command again or use the manual carrier release.

6027-3012 [E] Cannot find a pdisk in location *locationCode*.

Explanation:

The `tschcarrier` command cannot find a pdisk to replace in the given location.

User response:

Check the disk location code.

6027-3013 [W] Disk location *locationCode* failed to power on.

Explanation:

The `mmchcarrier` command detected an error when trying to power on a disk.

User response:

Make sure the disk is firmly seated and run the command again.

6027-3014 [E] Pdisk *pdiskName* of recovery group *recoveryGroupName* was expected to be replaced with a new disk; instead, it was moved from location *locationCode* to location *locationCode*.

Explanation:

The `mmchcarrier` command expected a pdisk to be removed and replaced with a new disk. But instead of being replaced, the old pdisk was moved into a different location.

User response:

Repeat the disk replacement procedure.

6027-3015 [E] Pdisk *pdiskName* of recovery group *recoveryGroupName* in location *locationCode* cannot be used as a replacement for pdisk *pdiskName* of recovery group *recoveryGroupName*.

Explanation:

The `tschcarrier` command expected a pdisk to be removed and replaced with a new disk. But instead of finding a new disk, the `mmchcarrier` command found that another pdisk was moved to the replacement location.

User response:

Repeat the disk replacement procedure, making sure to replace the failed pdisk with a new disk.

6027-3016 [E] Replacement disk in location *locationCode* has an incorrect type *fruCode*; expected type code is *fruCode*.

Explanation:

The replacement disk has a different field replaceable unit type code than that of the original disk.

User response:

Replace the pdisk with a disk of the same part number. If you are certain the new disk is a valid substitute,

override this check by running the command again with the `--force-fru` option.

6027-3017 [E] Error formatting replacement disk *diskName*.

Explanation:

An error occurred when trying to format a replacement pdisk.

User response:

Check the replacement disk.

6027-3018 [E] A replacement for pdisk *pdiskName* of recovery group *recoveryGroupName* was not found in location *locationCode*.

Explanation:

The `tschcarrier` command expected a pdisk to be removed and replaced with a new disk, but no replacement disk was found.

User response:

Make sure a replacement disk was inserted into the correct slot.

6027-3019 [E] Pdisk *pdiskName* of recovery group *recoveryGroupName* in location *locationCode* was not replaced.

Explanation:

The `tschcarrier` command expected a pdisk to be removed and replaced with a new disk, but the original pdisk was still found in the replacement location.

User response:

Repeat the disk replacement, making sure to replace the pdisk with a new disk.

6027-3020 [E] Invalid state change, *stateChangeName*, for pdisk *pdiskName*.

Explanation:

The `tschpdisk` command received an state change request that is not permitted.

User response:

Correct the input and reissue the command.

6027-3021 [E] Unable to change identify state to *identifyState* for pdisk *pdiskName*: *err=errorNum*.

Explanation:

The `tschpdisk` command failed on an identify request.

User response:

Check the disk enclosure hardware.

6027-3022 [E] Unable to create vdisk layout.

Explanation:

The `tscrvdisk` command could not create the necessary layout for the specified vdisk.

User response:

Change the vdisk arguments and retry the command.

6027-3023 [E] Error initializing vdisk.

Explanation:

The `tscrvdisk` command could not initialize the vdisk.

User response:

Retry the command.

6027-3024 [E] Error retrieving recovery group *recoveryGroupName* event log.

Explanation:

Because of an error, the `tslsrecoverygroupevents` command was unable to retrieve the full event log.

User response:

None.

6027-3025 [E] Device *deviceName* does not exist or is not active on this node.

Explanation:

The specified device was not found on this node.

User response:

None.

6027-3026 [E] Recovery group *recoveryGroupName* does not have an active log home vdisk.

Explanation:

The indicated recovery group does not have an active log vdisk. This may be because the log home vdisk has not yet been created, because a previously existing log home vdisk has been deleted, or because the server is in the process of recovery.

User response:

Create a log home vdisk if none exists. Retry the command.

6027-3027 [E] Cannot configure NSD-RAID services on this node.

Explanation:

NSD-RAID services are not supported on this operating system or node hardware.

User response:

Configure a supported node type as the NSD RAID server and restart the GPFS daemon.

6027-3028 [E] There is not enough space in declustered array *declusteredArrayName* for the requested vdisk size. The

maximum possible size for this vdisk is *size*.

Explanation:

There is not enough space in the declustered array for the requested vdisk size.

User response:

Create a smaller vdisk, remove existing vdisks or add additional pdisks to the declustered array.

6027-3029 [E] There must be at least *number* non-spare pdisks in declustered array *declusteredArrayName* to avoid falling below the code width of vdisk *vdiskName*.

Explanation:

A change of spares operation failed because the resulting number of non-spare pdisks would fall below the code width of the indicated vdisk.

User response:

Add additional pdisks to the declustered array.

6027-3030 [E] There must be at least *number* non-spare pdisks in declustered array *declusteredArrayName* for configuration data replicas.

Explanation:

A delete pdisk or change of spares operation failed because the resulting number of non-spare pdisks would fall below the number required to hold configuration data for the declustered array.

User response:

Add additional pdisks to the declustered array. If replacing a pdisk, use `mmchcarrier` or `mmaddpdisk --replace`.

6027-3031 [E] There is not enough available configuration data space in declustered array *declusteredArrayName* to complete this operation.

Explanation:

Creating a vdisk, deleting a pdisk, or changing the number of spares failed because there is not enough available space in the declustered array for configuration data.

User response:

Replace any failed pdisks in the declustered array and allow time for rebalance operations to more evenly distribute the available space. Add pdisks to the declustered array.

6027-3032 [E] Temporarily unable to create vdisk *vdiskName* because more time is required to rebalance the available space in declustered array *declusteredArrayName*.

Explanation:

Cannot create the specified vdisk until rebuild and rebalance processes are able to more evenly distribute the available space.

User response:

Replace any failed pdisks in the recovery group, allow time for rebuild and rebalance processes to more evenly distribute the spare space within the array, and retry the command.

6027-3034 [E] The input pdisk name (*pdiskName*) did not match the pdisk name found on disk (*pdiskName*).

Explanation:

Cannot add the specified pdisk, because the input *pdiskName* did not match the *pdiskName* that was written on the disk.

User response:

Verify the input file and retry the command.

6027-3035 [A] Cannot configure NSD-RAID services. maxblocksize must be at least *value*.

Explanation:

The GPFS daemon is starting and cannot initialize the NSD-RAID services because the maxblocksize attribute is too small.

User response:

Correct the maxblocksize attribute and restart the GPFS daemon.

6027-3036 [E] Partition size must be a power of 2.

Explanation:

The partitionSize parameter of some declustered array was invalid.

User response:

Correct the partitionSize parameter and reissue the command.

6027-3037 [E] Partition size must be between *number* and *number*.

Explanation:

The partitionSize parameter of some declustered array was invalid.

User response:

Correct the partitionSize parameter to a power of 2 within the specified range and reissue the command.

6027-3038 [E] AU log too small; must be at least *number* bytes.

Explanation:

The auLogSize parameter of a new declustered array was invalid.

User response:

Increase the auLogSize parameter and reissue the command.

6027-3039 [E] A vdisk with disk usage vdiskLogTip must be the first vdisk created in a recovery group.

Explanation:

The --logTip disk usage was specified for a vdisk other than the first one created in a recovery group.

User response:

Retry the command with a different disk usage.

6027-3040 [E] Declustered array configuration data does not fit.

Explanation:

There is not enough space in the pdisks of a new declustered array to hold the AU log area using the current partition size.

User response:

Increase the partitionSize parameter or decrease the auLogSize parameter and reissue the command.

6027-3041 [E] Declustered array attributes cannot be changed.

Explanation:

The partitionSize, auLogSize, and canHoldVCD attributes of a declustered array cannot be changed after the declustered array has been created. They may only be set by a command that creates the declustered array.

User response:

Remove the partitionSize, auLogSize, and canHoldVCD attributes from the input file of the mmaddpdisk command and reissue the command.

6027-3042 [E] The log tip vdisk cannot be destroyed if there are other vdisks.

Explanation:

In recovery groups with versions prior to 3.5.0.11, the log tip vdisk cannot be destroyed if other vdisks still exist within the recovery group.

User response:

Remove the user vdisks or upgrade the version of the recovery group with mmchrecoverygroup --version, then retry the command to remove the log tip vdisk.

6027-3043 [E] Log vdisks cannot have multiple use specifications.

Explanation:

A vdisk can have usage vdiskLog, vdiskLogTip, or vdiskLogReserved, but not more than one.

User response:

Retry the command with only one of the --log, --logTip, or --logReserved attributes.

6027-3044 [E] Unable to determine resource requirements for all the recovery groups served by node *value*: to override this check reissue the command with the -v no flag.

Explanation:

A recovery group or vdisk is being created, but IBM Spectrum Scale can not determine if there are enough non-stealable buffer resources to allow the node to successfully serve all the recovery groups at the same time once the new object is created.

User response:

You can override this check by reissuing the command with the -v flag.

6027-3045 [W] Buffer request exceeds the non-stealable buffer limit. Check the configuration attributes of the recovery group servers: pagepool, nsdRAIDBufferPoolSizePct, nsdRAIDNonStealableBufPct.

Explanation:

The limit of non-stealable buffers has been exceeded. This is probably because the system is not configured correctly.

User response

Check the settings of the pagepool, nsdRAIDBufferPoolSizePct, and nsdRAIDNonStealableBufPct attributes and make sure the server has enough real memory to support the configured values.

Use the mmchconfig command to correct the configuration.

6027-3046 [E] The nonStealable buffer limit may be too low on server *serverName* or the pagepool is too small. Check the configuration attributes of the recovery group servers: pagepool, nsdRAIDBufferPoolSizePct, nsdRAIDNonStealableBufPct.

Explanation:

The limit of non-stealable buffers is too low on the specified recovery group server. This is probably because the system is not configured correctly.

User response

Check the settings of the pagepool, nsdRAIDBufferPoolSizePct, and nsdRAIDNonStealableBufPct attributes and make sure the server has sufficient real memory to support the configured values. The specified configuration variables should be the same for the recovery group servers.

Use the mmchconfig command to correct the configuration.

6027-3047 [E] Location of pdisk *pdiskName* is not known.

Explanation:

IBM Spectrum Scale is unable to find the location of the given pdisk.

User response:

Check the disk enclosure hardware.

6027-3048 [E] Pdisk *pdiskName* is not currently scheduled for replacement.

Explanation:

A pdisk specified in a tschcarrier or tsaddpdisk command is not currently scheduled for replacement.

User response:

Make sure the correct disk location code or pdisk name was given. For the tschcarrier command, the --force-release option can be used to override the check.

6027-3049 [E] The minimum size for vdisk *vdiskName* is *number*.

Explanation:

The vdisk size was too small.

User response:

Increase the size of the vdisk and retry the command.

6027-3050 [E] There are already *number* suspended pdisks in declustered array *arrayName*. You must resume pdisks in the array before suspending more.

Explanation:

The number of suspended pdisks in the declustered array has reached the maximum limit. Allowing more pdisks to be suspended in the array would put data availability at risk.

User response:

Resume one more suspended pdisks in the array by using the mmchcarrier or mmchpdisk commands then retry the command.

6027-3051 [E] Checksum granularity must be *number* or *number*.

Explanation:

The only allowable values for the checksumGranularity attribute of a data vdisk are 8K and 32K.

User response:

Change the checksumGranularity attribute of the vdisk, then retry the command.

6027-3052 [E] Checksum granularity cannot be specified for log vdisks.

Explanation:

The checksumGranularity attribute cannot be applied to a log vdisk.

User response:

Remove the checksumGranularity attribute of the log vdisk, then retry the command.

6027-3053 [E] Vdisk block size must be between *number* and *number* for the specified code when checksum granularity *number* is used.

Explanation:

An invalid vdisk block size was specified. The message lists the allowable range of block sizes.

User response:

Use a vdisk virtual block size within the range shown, or use a different vdisk RAID code, or use a different checksum granularity.

6027-3054 [W] Disk in location *locationCode* failed to come online.

Explanation:

The mmchcarrier command detected an error when trying to bring a disk back online.

User response:

Make sure the disk is firmly seated and run the command again. Check the operating system error log.

6027-3055 [E] The fault tolerance of the code cannot be greater than the fault tolerance of the internal configuration data.

Explanation:

The RAID code specified for a new vdisk is more fault-tolerant than the configuration data that will describe the vdisk.

User response:

Use a code with a smaller fault tolerance.

6027-3056 [E] Long and short term event log size and fast write log percentage are only applicable to log home vdisk.

Explanation:

The longTermEventLogSize, shortTermEventLogSize, and fastWriteLogPct options are only applicable to log home vdisk.

User response:

Remove any of these options and retry vdisk creation.

6027-3057 [E] Disk enclosure is no longer reporting information on location *locationCode*.

Explanation:

The disk enclosure reported an error when IBM Spectrum Scale tried to obtain updated status on the disk location.

User response:

Try running the command again. Make sure that the disk enclosure firmware is current. Check for improperly-seated connectors within the disk enclosure.

6027-3058 [A] GSS license failure - IBM Spectrum Scale RAID services will not be configured on this node.

Explanation:

The Elastic Storage Server has not been installed validly. Therefore, IBM Spectrum Scale RAID services will not be configured.

User response:

Install a licensed copy of the base IBM Spectrum Scale code and restart the GPFS daemon.

6027-3059 [E] The serviceDrain state is only permitted when all nodes in the cluster are running daemon version *version* or higher.

Explanation:

The mmchpdisk command option --begin-service-drain was issued, but there are backlevel nodes in the cluster that do not support this action.

User response:

Upgrade the nodes in the cluster to at least the specified version and run the command again.

6027-3060 [E] Block sizes of all log vdisks must be the same.

Explanation:

The block sizes of the log tip vdisk, the log tip backup vdisk, and the log home vdisk must all be the same.

User response:

Try running the command again after adjusting the block sizes of the log vdisks.

6027-3061 [E] Cannot delete path *pathName* because there would be no other working paths to pdisk *pdiskName* of RG *recoveryGroupName*.

Explanation:

When the -v yes option is specified on the --delete-paths subcommand of the tschregroup command, it is not allowed to delete the last working path to a pdisk.

User response:

Try running the command again after repairing other broken paths for the named pdisk, or reduce the list of paths being deleted, or run the command with -v no.

6027-3062 [E] Recovery group version *version* is not compatible with the current recovery group version.

Explanation:

The recovery group version specified with the `--version` option does not support all of the features currently supported by the recovery group.

User response:

Run the command with a new value for `--version`. The allowable values will be listed following this message.

6027-3063 [E] Unknown recovery group version *version*.

Explanation:

The recovery group version named by the argument of the `--version` option was not recognized.

User response:

Run the command with a new value for `--version`. The allowable values will be listed following this message.

6027-3064 [I] Allowable recovery group versions are:

Explanation:

Informational message listing allowable recovery group versions.

User response:

Run the command with one of the recovery group versions listed.

6027-3065 [E] The maximum size of a log tip vdisk is *size*.

Explanation:

Running `mmcrvdisk` for a log tip vdisk failed because the size is too large.

User response:

Correct the size parameter and run the command again.

6027-3066 [E] A recovery group may only contain one log tip vdisk.

Explanation:

A log tip vdisk already exists in the recovery group.

User response:

None.

6027-3067 [E] Log tip backup vdisks not supported by this recovery group version.

Explanation:

Vdisks with usage type `vdiskLogTipBackup` are not supported by all recovery group versions.

User response:

Upgrade the recovery group to a later version using the `--version` option of `mmchrecoverygroup`.

6027-3068 [E] The sizes of the log tip vdisk and the log tip backup vdisk must be the same.

Explanation:

The log tip vdisk must be the same size as the log tip backup vdisk.

User response:

Adjust the vdisk sizes and retry the `mmcrvdisk` command.

6027-3069 [E] Log vdisks cannot use code *codeName*.

Explanation:

Log vdisks must use a RAID code that uses replication, or be unreplicated. They cannot use parity-based codes such as 8+2P.

User response:

Retry the command with a valid RAID code.

6027-3070 [E] Log vdisk *vdiskName* cannot appear in the same declustered array as log vdisk *vdiskName*.

Explanation:

No two log vdisks may appear in the same declustered array.

User response:

Specify a different declustered array for the new log vdisk and retry the command.

6027-3071 [E] Device not found: *deviceName*.

Explanation:

A device name given in an `mmcrrecoverygroup` or `mmaddpdisk` command was not found.

User response:

Check the device name.

6027-3072 [E] Invalid device name: *deviceName*.

Explanation:

A device name given in an `mmcrrecoverygroup` or `mmaddpdisk` command is invalid.

User response:

Check the device name.

6027-3073 [E] Error formatting pdisk *pdiskName* on device *diskName*.

Explanation:

An error occurred when trying to format a new pdisk.

User response:

Check that the disk is working properly.

6027-3074 [E] Node *nodeName* not found in cluster configuration.

Explanation:

A node name specified in a command does not exist in the cluster configuration.

User response:

Check the command arguments.

6027-3075 [E] The --servers list must contain the current node, *nodeName*.

Explanation:

The --servers list of a tscrrecgroup command does not list the server on which the command is being run.

User response:

Check the --servers list. Make sure the tscrrecgroup command is run on a server that will actually server the recovery group.

6027-3076 [E] Remote pdisks are not supported by this recovery group version.

Explanation:

Pdisks that are not directly attached are not supported by all recovery group versions.

User response:

Upgrade the recovery group to a later version using the --version option of mmchrecoverygroup.

6027-3077 [E] There must be at least *number* pdisks in recovery group *recoveryGroupName* for configuration data replicas.

Explanation:

A change of pdisks failed because the resulting number of pdisks would fall below the needed replication factor for the recovery group descriptor.

User response:

Do not attempt to delete more pdisks.

6027-3078 [E] Replacement threshold for declustered array *declusteredArrayName* of recovery group *recoveryGroupName* cannot exceed *number*.

Explanation:

The replacement threshold cannot be larger than the maximum number of pdisks in a declustered array. The maximum number of pdisks in a declustered array depends on the version number of the recovery group. The current limit is given in this message.

User response:

Use a smaller replacement threshold or upgrade the recovery group version.

6027-3079 [E] Number of spares for declustered array *declusteredArrayName* of recovery group *recoveryGroupName* cannot exceed *number*.

Explanation:

The number of spares cannot be larger than the maximum number of pdisks in a declustered array. The maximum number of pdisks in a declustered array depends on the version number of the recovery group. The current limit is given in this message.

User response:

Use a smaller number of spares or upgrade the recovery group version.

6027-3080 [E] Cannot remove pdisk *pdiskName* because declustered array *declusteredArrayName* would have fewer disks than its replacement threshold.

Explanation:

The replacement threshold for a declustered array must not be larger than the number of pdisks in the declustered array.

User response:

Reduce the replacement threshold for the declustered array, then retry the mmdelpdisk command.

6027-3084 [E] VCD spares feature must be enabled before being changed. Upgrade recovery group version to at least *version* to enable it.

Explanation:

The vdisk configuration data (VCD) spares feature is not supported in the current recovery group version.

User response:

Apply the recovery group version that is recommended in the error message and retry the command.

6027-3085 [E] The number of VCD spares must be greater than or equal to the number of spares in declustered array *declusteredArrayName*.

Explanation:

Too many spares or too few vdisk configuration data (VCD) spares were specified.

User response:

Retry the command with a smaller number of spares or a larger number of VCD spares.

6027-3086 [E] There is only enough free space to allocate *n* VCD spare(s) in declustered array *declusteredArrayName*.

Explanation:

Too many vdisk configuration data (VCD) spares were specified.

User response:

Retry the command with a smaller number of VCD spares.

6027-3087 [E] Specifying Pdisk rotation rate not supported by this recovery group version.

Explanation:

Specifying the Pdisk rotation rate is not supported by all recovery group versions.

User response:

Upgrade the recovery group to a later version using the `--version` option of the `mmchrecoverygroup` command. Or, don't specify a rotation rate.

6027-3088 [E] Specifying Pdisk expected number of paths not supported by this recovery group version.

Explanation:

Specifying the expected number of active or total pdisk paths is not supported by all recovery group versions.

User response:

Upgrade the recovery group to a later version using the `--version` option of the `mmchrecoverygroup` command. Or, don't specify the expected number of paths.

6027-3089 [E] Pdisk *pdiskName* location *locationCode* is already in use.

Explanation:

The pdisk location that was specified in the command conflicts with another pdisk that is already in that location. No two pdisks can be in the same location.

User response:

Specify a unique location for this pdisk.

6027-3090 [E] Enclosure control command failed for pdisk *pdiskName* of RG *recoveryGroupName* in location *locationCode*: *err errorNum*. Examine mmfs log for `tsctlencslot`, `tsonosdisk` and `tsoffosdisk` errors.

Explanation:

A command used to control a disk enclosure slot failed.

User response:

Examine the mmfs log files for more specific error messages from the `tsctlencslot`, `tsonosdisk`, and `tsoffosdisk` commands.

6027-3091 [W] A command to control the disk enclosure failed with error code *errorNum*. As a result, enclosure indicator lights may not have changed to the correct states. Examine the mmfs log on nodes attached to the disk

enclosure for messages from the `tsctlencslot`, `tsonosdisk`, and `tsoffosdisk` commands for more detailed information.

Explanation:

A command used to control disk enclosure lights and carrier locks failed. This is not a fatal error.

User response:

Examine the mmfs log files on nodes attached to the disk enclosure for error messages from the `tsctlencslot`, `tsonosdisk`, and `tsoffosdisk` commands for more detailed information. If the carrier failed to unlock, either retry the command or use the manual override.

6027-3092 [I] Recovery group *recoveryGroupName* assignment delay *delaySeconds* seconds for safe recovery.

Explanation:

The recovery group must wait before meta-data recovery. Prior disk lease for the failing manager must first expire.

User response:

None.

6027-3093 [E] Checksum granularity must be *number* or *number* for log vdisks.

Explanation:

The only allowable values for the `checksumGranularity` attribute of a log vdisk are 512 and 4K.

User response:

Change the `checksumGranularity` attribute of the vdisk, then retry the command.

6027-3094 [E] Due to the attributes of other log vdisks, the checksum granularity of this vdisk must be *number*.

Explanation:

The checksum granularities of the log tip vdisk, the log tip backup vdisk, and the log home vdisk must all be the same.

User response:

Change the `checksumGranularity` attribute of the new log vdisk to the indicated value, then retry the command.

6027-3095 [E] The specified declustered array name (*declusteredArrayName*) for the new pdisk *pdiskName* must be *declusteredArrayName*.

Explanation:

When replacing an existing pdisk with a new pdisk, the declustered array name for the new pdisk must match the declustered array name for the existing pdisk.

User response:

Change the specified declustered array name to the indicated value, then run the command again.

6027-3096 [E] Internal error encountered in NSD-RAID command: *err=errorNum*.

Explanation:

An unexpected GPFS NSD-RAID internal error occurred.

User response:

Contact the IBM Support Center.

6027-3097 [E] Missing or invalid pdisk name (*pdiskName*).

Explanation:

A pdisk name specified in an **mmcrrecoverygroup** or **mmaddpdisk** command is not valid.

User response:

Specify a pdisk name that is 63 characters or less. Valid characters are: a to z, A to Z, 0 to 9, and underscore (_).

6027-3098 [E] Pdisk name *pdiskName* is already in use in recovery group *recoveryGroupName*.

Explanation:

The pdisk name already exists in the specified recovery group.

User response:

Choose a pdisk name that is not already in use.

6027-3099 [E] Device with path(s) *pathName* is specified for both new pdisks *pdiskName* and *pdiskName*.

Explanation:

The same device is specified for more than one pdisk in the stanza file. The device can have multiple paths, which are shown in the error message.

User response:

Specify different devices for different new pdisks, respectively, and run the command again.

6027-3800 [E] Device with path(s) *pathName* for new pdisk *pdiskName* is already in use by pdisk *pdiskName* of recovery group *recoveryGroupName*.

Explanation:

The device specified for a new pdisk is already being used by an existing pdisk. The device can have multiple paths, which are shown in the error message.

User response:

Specify an unused device for the pdisk and run the command again.

6027-3801 [E] [E] The checksum granularity for log vdisks in declustered array *declusteredArrayName* of RG *recoveryGroupName* must be at least *number* bytes.

Explanation:

Use a checksum granularity that is not smaller than the minimum value given. You can use the **mmlspdisk** command to view the logical block sizes of the pdisks in this array to identify which pdisks are driving the limit.

User response:

Change the **checksumGranularity** attribute of the new log vdisk to the indicated value, and then retry the command.

6027-3802 [E] [E] Pdisk *pdiskName* of RG *recoveryGroupName* has a logical block size of *number* bytes; the maximum logical block size for pdisks in declustered array *declusteredArrayName* cannot exceed the log checksum granularity of *number* bytes.

Explanation:

Logical block size of pdisks added to this declustered array must not be larger than any log vdisk's checksum granularity.

User response:

Use pdisks with equal or smaller logical block size than the log vdisk's checksum granularity.

6027-3803 [E] [E] NSD format version 2 feature must be enabled before being changed. Upgrade recovery group version to at least *recoveryGroupVersion* to enable it.

Explanation:

NSD format version 2 feature is not supported in current recovery group version.

User response:

Apply the recovery group version recommended in the error message and retry the command.

6027-3804 [W] Skipping upgrade of pdisk *pdiskName* because the disk capacity of *number* bytes is less than the *number* bytes required for the new format.

Explanation:

The existing format of the indicated pdisk is not compatible with NSD V2 descriptors.

User response:

A complete format of the declustered array is required in order to upgrade to NSD V2.

6027-3805 [E] NSD format version 2 feature is not supported by the current recovery group version. A recovery group version of at least *rgVersion* is required for this feature.

Explanation:

NSD format version 2 feature is not supported in the current recovery group version.

User response:

Apply the recovery group version recommended in the error message and retry the command.

6027-3806 [E] The device given for pdisk *pdiskName* has a logical block size of *logicalBlockSize* bytes, which is not supported by the recovery group version.

Explanation:

The current recovery group version does not support disk drives with the indicated logical block size.

User response:

Use a different disk device or upgrade the recovery group version and retry the command.

6027-3807 [E] NSD version 1 specified for pdisk *pdiskName* requires a disk with a logical block size of 512 bytes. The supplied disk has a block size of *logicalBlockSize* bytes. For this disk, you must use at least NSD version 2.

Explanation:

Requested logical block size is not supported by NSD format version 1.

User response:

Correct the input file to use a different disk or specify a higher NSD format version.

6027-3808 [E] Pdisk *pdiskName* must have a capacity of at least *number* bytes for NSD version 2.

Explanation:

The pdisk must be at least as large as the indicated minimum size in order to be added to the declustered array.

User response:

Correct the input file and retry the command.

6027-3809 [I] Pdisk *pdiskName* can be added as NSD version 1.

Explanation:

The pdisk has enough space to be configured as NSD version 1.

User response:

Specify NSD version 1 for this disk.

6027-3810 [W] [W] Skipping the upgrade of pdisk *pdiskName* because no I/O paths are currently available.

Explanation:

There is no I/O path available to the indicated pdisk.

User response:

Try running the command again after repairing the broken I/O path to the specified pdisk.

6027-3811 [E] Unable to *action* vdisk MDI.

Explanation:

The **tscrvdisk** command could not create or write the necessary vdisk MDI.

User response:

Retry the command.

6027-3812 [I] Log group *logGroupName* assignment delay *delaySeconds* seconds for safe recovery.

Explanation:

The recovery group configuration manager must wait. Prior disk lease for the failing manager must expire before assigning a new worker to the log group.

User response:

None.

6027-3813 [A] Recovery group *recoveryGroupName* could not be served by node *nodeName*.

Explanation:

The recovery group configuration manager could not perform a node assignment to manage the recovery group.

User response:

Check whether there are sufficient nodes and whether errors are recorded in the recovery group event log.

6027-3814 [A] Log group *logGroupName* could not be served by node *nodeName*.

Explanation:

The recovery group configuration manager could not perform a node assignment to manage the log group.

User response:

Check whether there are sufficient nodes and whether errors are recorded in the recovery group event log.

6027-3815 [E] Erasure code not supported by this recovery group version.

Explanation:

Vdisks with 4+2P and 4+3P erasure codes are not supported by all recovery group versions.

User response:

Upgrade the recovery group to a later version using the **--version** option of the **mmchrecoverygroup** command.

6027-3816 [E] Invalid declustered array name (*declusteredArrayName*).

Explanation:

A declustered array name given in the **mmcrrecoverygroup** or **mmaddpdisk** command is invalid.

User response:

Use only the characters a-z, A-Z, 0-9, and underscore to specify a declustered array name and you can specify up to 63 characters.

6027-3817 [E] Invalid log group name (*logGroupName*).

Explanation:

A log group name given in the **mmcrrecoverygroup** or **mmaddpdisk** command is invalid.

User response:

Use only the characters a-z, A-Z, 0-9, and underscore to specify a declustered array name and you can specify up to 63 characters.

6027-3818 [E] Cannot create log group *logGroupName*; there can be at most *number* log groups in a recovery group.

Explanation:

The number of log groups allowed in a recovery group has been exceeded.

User response:

Reduce the number of log groups in the input file and retry the command.

6027-3819 [I] Recovery group *recoveryGroupName* delay *delaySeconds* seconds for assignment.

Explanation:

The recovery group configuration manager must wait before assigning a new manager to the recovery group.

User response:

None.

6027-3820 [E] Specifying *canHoldVCD* not supported by this recovery group version.

Explanation:

The ability to override the default decision of whether a declustered array is allowed to hold vdisk configuration data is not supported by all recovery group versions.

User response:

Upgrade the recovery group to a later version using the **--version** option of the **mmchrecoverygroup** command.

6027-3821 [E] Cannot set *canHoldVCD=yes* for small declustered arrays.

Explanation:

Declustered arrays with less than 9+vcdSpares disks cannot hold vdisk configuration data.

User response:

Add more disks to the declustered array or do not specify *canHoldVCD=yes*.

6027-3822 [I] Recovery group *recoveryGroupName* working index delay *delaySeconds* seconds for safe recovery.

Explanation:

Prior disk lease for the workers must expire before recovering the working index metadata.

User response:

None.

6027-3823 [E] Unknown node *nodeName* in the recovery group configuration.

Explanation:

A node name does not exist in the recovery group configuration manager.

User response:

Check for damage to the **mmhdrfs** file.

6027-3824 [E] The defined server *serverName* for recovery group *recoveryGroupName* could not be resolved.

Explanation:

The host name of recovery group server could not be resolved by `gethostbyname()`.

User response:

Fix host name resolution.

6027-3825 [E] The defined server *serverName* for node class *nodeClassName* could not be resolved.

Explanation:

The host name of recovery group server could not be resolved by `gethostbyname()`.

User response:

Fix host name resolution.

6027-3826 [A] Error reading volume identifier for recovery group *recoveryGroupName* from configuration file.

Explanation:

The volume identifier for the named recovery group could not be read from the **mmsdrfs** file. This should never occur.

User response:

Check for damage to the **mmsdrfs** file.

6027-3827 [A] Error reading volume identifier for vdisk *vdiskName* from configuration file.

Explanation:

The volume identifier for the named vdisk could not be read from the **mmsdrfs** file. This should never occur.

User response:

Check for damage to the **mmsdrfs** file.

6027-3828 [E] Vdisk *vdiskName* could not be associated with its recovery group *recoveryGroupName* and will be ignored.

Explanation:

The named vdisk cannot be associated with its recovery group.

User response:

Check for damage to the **mmsdrfs** file.

6027-3829 [E] A server list must be provided.

Explanation:

No server list is specified.

User response:

Specify a list of valid servers.

6027-3830 [E] Too many servers specified.

Explanation:

An input node list has too many nodes specified.

User response:

Verify the list of nodes and shorten the list to the supported number.

6027-3831 [E] A vdisk name must be provided.

Explanation:

A vdisk name is not specified.

User response:

Specify a vdisk name.

6027-3832 [E] A recovery group name must be provided.

Explanation:

A recovery group name is not specified.

User response:

Specify a recovery group name.

6027-3833 [E] Recovery group *recoveryGroupName* does not have an active root log group.

Explanation:

The root log group must be active before the operation is permitted.

User response:

Retry the command after the recovery group becomes fully active.

6027-3836 [I] Cannot retrieve MSID for device: *devFileName*.

Explanation:

Command usage message for **tsgetmsid**.

User response:

None.

6027-3837 [E] Error creating worker vdisk.

Explanation:

The **tscrvdisk** command could not initialize the vdisk at the worker node.

User response:

Retry the command.

6027-3838 [E] Unable to write new vdisk MDI.

Explanation:

The **tscrvdisk** command could not write the necessary vdisk MDI.

User response:

Retry the command.

6027-3839 [E] Unable to write update vdisk MDI.

Explanation:

The **tscrvdisk** command could not write the necessary vdisk MDI.

User response:

Retry the command.

6027-3840 [E] Unable to delete worker vdisk *vdiskName* err=*errorNum*.

Explanation:

The specified vdisk worker object could not be deleted.

User response:

Retry the command with a valid vdisk name.

6027-3841 [E] Unable to create new vdisk MDI.

Explanation:

The **tscrvdisk** command could not create the necessary vdisk MDI.

User response:

Retry the command.

6027-3843 [E] Error returned from node *nodeName* when preparing new pdisk *pdiskName* of RG *recoveryGroupName* for use: err *errorNum*

Explanation:

The system received an error from the given node when trying to prepare a new pdisk for use.

User response:

Retry the command.

6027-3844 [E] Unable to prepare new pdisk *pdiskName* of RG *recoveryGroupName* for use: exit status *exitStatus*.

Explanation:

The system received an error from the **tspreparenewpdiskforuse** script when trying to prepare a new pdisk for use.

User response:

Check the new disk and retry the command.

6027-3845 [E] Unrecognized pdisk state: *pdiskState*.

Explanation:

The given pdisk state name is invalid.

User response:

Use a valid pdisk state name.

6027-3846 [E] Pdisk state change *pdiskState* is not permitted.

Explanation:

An attempt was made to use the **mmchpdisk** command either to change an internal pdisk state, or to create an invalid combination of states.

User response:

Some internal pdisk state flags can be set indirectly by running other commands. For example, the *deleting* state can be set by using the **mmde1pdisk** command.

6027-3847 [E] [E] The *serviceDrain* state feature must be enabled to use this command. Upgrade the recovery group version to at least *version* to enable it.

Explanation:

The **mmchpdisk** command option **--begin-service-drain** was issued, but there are back-level nodes in the cluster that do not support this action.

User response:

Upgrade the nodes in the cluster to at least the specified version and run the command again.

6027-3848 [E] The simulated dead and failing state feature must be enabled to use this command. Upgrade the recovery group version to at least *version* to enable it.

Explanation:

The **mmchpdisk** command option **--begin-**

service-drain was issued, but there are back-level nodes in the cluster that do not support this action.

User response:

Upgrade the nodes in the cluster to at least the specified version and run the command again.

6027-3849 [E] The pdisk *pdiskName* of recovery group *recoveryGroupName* could not be revived. Pdisk state is *pdiskState*.

Explanation:

An **mmchpdisk --revive** command was unable to bring a pdisk back online.

User response:

If the state is missing, restore connectivity to the disk. If the disk is in failed state replace the pdisk. A pdisk with the status dead, readOnly, failing, or slot is considered as failed.

6027-3850 [E] Location *locationCode* contains multiple disk devices. You cannot use this command to replace disks in the specific location.

Explanation:

The **mmvdisk pdisk replace** command or the **mmchcarrier** command was given a location that contains multiple disk devices. An example of a location with multiple disk devices is the situation where the operating system (OS) root disk and log tip devices share the same underlying storage.

User response:

If the problem PDisk is one of the log tip devices and it shares storage with other log tip devices or the OS root, first make sure that the device has failed. That is, it is in "dead", "readOnly" or "failing" state as opposed to being temporarily inaccessible because node is down. If the device is really down, delete the log tip VDisk and declustered array from the recovery group, then replace the failed hardware. Finally, re-create the log tip DA and VDisk. Refer to the product documentation for more detailed instructions.

6027-3851 [E] Command interrupted by recovery group *recoveryGroupName* failover.

Explanation:

A recovery group command failed because the recovery group stopped serving, probably because it failed over to another node.

User response:

Run the command again.

6027-3852 [A] Cannot configure NSD-RAID services. The *nsdRAIDBufferPoolSizePct* attribute of the pagepool

must result in at least
nsdRAIDMasterBufferPoolSize
(*number*) bytes + 128 MiB of space.

Explanation:

The GPFS daemon is starting and cannot initialize the NSD-RAID services because of the memory consideration specified.

User response:

Correct the *nsdRAIDBufferPoolSizePct* attribute of the pagepool and restart the GPFS daemon.

6027-3853 [W] Buffer request (*name*) exceeds the master reserved buffer limit (*number*). Check the configuration attributes of the recovery group servers: *nsdRAIDMasterBufferPoolSize*.

Explanation:

The limit of master reserved buffers is exceeded. This is probably because of an improperly configured system. Check the setting of the *nsdRAIDMasterBufferPoolSize* parameter, and whether the server has sufficient memory to support the configured value.

User response:

Use the *mmchconfig* command to correct the configuration.

6027-3854 [E] Recovery group configuration manager takeover failed: scheduled *scheduled stopping stopping*

Explanation:

The recovery group configuration manager takeover schedule failed.

User response:

Contact the IBM Support.

6027-3855 [E] *rgcmRefreshConfig* error. Duplicated NID *nsdID (vdiskName)* found in *recoveryGroupName*.

Explanation:

Duplicated ID found by RGCM during initialization.

User response:

Contact the IBM Support.

6027-3856 [E] Recovery group configuration manager takeover failed: *err errMsg*

Explanation:

The recovery group configuration manager takeover failed with error.

User response:

Contact the IBM Support.

6027-3857 [E] Log group *logGroupName* of recovery group *recoveryGroupName* could not be served.

Explanation:

The recovery group configuration manager could not perform a node assignment to manage the log group.

User response:

Check whether there are sufficient nodes and whether errors are recorded in the recovery group event log.

6027-3858 [E] Recovery group configuration manager failed to start. *err errMsg*

Explanation:

Recovery group configuration manager final takeover failed.

User response:

Contact IBM Support.

6027-3859 [E] Trim to device not supported by this recovery group version. Upgrade the recovery group version to at least *version* to enable it.

Explanation:

The ability to enable trim to device is not supported by the current recovery group version.

User response:

Upgrade the recovery group to a later version by using the *--version* option of the *mmchrecoverygroup* command.

6027-3860 [E] Recovery group descriptor for PDisk *pdiskName* of recovery group *recoveryGroupName* could not be written because volatile write caching is enabled on this drive.

Explanation:

GPFS Native RAID refused to write a recovery group descriptor to a drive because it detected that volatile write caching was enabled. This error can occur when creating a recovery group, adding a new PDisk to an existing recovery group, or when replacing a drive.

User response:

Disable volatile write caching and related settings to comply with supported configurations.

6027-3861 [E] Recovery group descriptor for PDisk *pdiskName* of recovery group *recoveryGroupName* not be written *err=errMsg*.

Explanation:

GPFS Native RAID refused to write recovery group descriptor to a drive due to an internal error. This error can occur when creating a recovery group, adding a new PDisk to an existing recovery group, or when replacing a drive.

User response:

Contact IBM Support.

6027-3862 [E] Trim to declustered array *arrayName* of recovery group *recoveryGroupName* is not supported for hardware type *hardwareType*

Explanation:

Trim to device will not be enabled for a declustered array that contains drives of an unsupported hardware type.

User response:

Review hardware documentation for device trim capability or GNR trim documentation for a list of supported configurations.

6027-3863 [E] Recovery group *recoveryGroupName* stops serving after exceeding retry limit *nsdRAIDMaxRecoveryRetries*.

Explanation:

The recovery group could not start due to retry failure exceeding retry limit.

User response:

Check disk storage connection and run mmvdisk to restart the recovery group.

6027-3864 [E] [E] Slot location is missing from pdisk *pdiskName* device(s) *deviceName* of declustered array *arrayName* in recovery group *recoveryGroupName* with hardware type *hardwareType*.

Explanation:

Perform strict pdisk slot location checking to find empty slot location for this pdisk.

User response:

Review GNR documentation to make sure the disk drives are configured properly and the slot location mapping is set up correctly. Fix the problems for this pdisk and retry the command. Contact IBM support if it doesn't solve the problem.

6027-3865 [E] [E] *nFailures* empty slot *locationstring* found in recovery group *recoveryGroupName*. Make sure \ disk drives and slot location mapping are configured properly.

Explanation:

Perform strict pdisk slot location checking to find empty slot location for at least one pdisk. The pdisks that have problem are shown above this message.

User response:

Review GNR documentation to make sure the disk drives are configured properly and the slot location mapping is set up correctly. Fix the problems for all these pdisks and retry the command. Contact IBM support if it doesn't solve the problem.

6027-3866 [E] Log group *logGroupName* of recovery group *recoveryGroupName* not reachable. Verify recovery group health and try again.

Explanation:

Vdisk deletion fails if the log group is down. Log group and corresponding recovery group are shown in the message.

User response:

Verify the state of recovery group before performing vdisk deletion.

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](http://www.ibm.com/legal/copytrade.shtml)" 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.

The registered trademark Linux is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

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

Red Hat and Ansible® are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

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

Terms and conditions for product documentation

Permissions for the use of these publications are granted subject to the following terms and conditions.

Applicability

These terms and conditions are in addition to any terms of use for the IBM website.

Personal use

You may reproduce these publications for your personal, noncommercial use provided that all proprietary notices are preserved. You may not distribute, display or make derivative work of these publications, or any portion thereof, without the express consent of IBM.

Commercial use

You may reproduce, distribute and display these publications solely within your enterprise provided that all proprietary notices are preserved. You may not make derivative works of these publications, or reproduce, distribute or display these publications or any portion thereof outside your enterprise, without the express consent of IBM.

Rights

Except as expressly granted in this permission, no other permissions, licenses or rights are granted, either express or implied, to the publications or any information, data, software or other intellectual property contained therein.

IBM reserves the right to withdraw the permissions granted herein whenever, in its discretion, the use of the publications is detrimental to its interest or, as determined by IBM, the above instructions are not being properly followed.

You may not download, export or re-export this information except in full compliance with all applicable laws and regulations, including all United States export laws and regulations.

IBM MAKES NO GUARANTEE ABOUT THE CONTENT OF THESE PUBLICATIONS. THE PUBLICATIONS ARE PROVIDED "AS-IS" AND WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT, AND FITNESS FOR A PARTICULAR PURPOSE.

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](http://www.ibm.com/software/globalization/terminology) (opens in new window):

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

B

building block

A pair of servers with shared disk enclosures attached.

BOOTP

See Bootstrap Protocol (BOOTP).

Bootstrap Protocol (BOOTP)

A computer networking protocol 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 filesset

A filesset 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 filesset, which enables more efficient per-filesset 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*.

Index

Special Characters

/tmp/mmfs directory [27](#)

A

array, declustered
background tasks [33](#)

B

back up data [16](#)
background tasks [33](#)
best practices for troubleshooting [15](#), [19](#), [21](#)

C

checksum
data [34](#)
Command
mmchcarrier
usage [39](#)
commands
errpt [27](#)
gpfs.snap [27](#)
lslpp [28](#)
mmlsdisk [28](#)
mmlsfs [28](#)
rpm [27](#)
comments [xiv](#)
components of storage enclosures
replacing failed [42](#)
contacting IBM [29](#)

D

data checksum [34](#)
declustered array
background tasks [33](#)
diagnosis, disk [32](#)
directed maintenance procedure
activate AFM [68](#)
activate NFS [69](#)
activate SMB [69](#)
configure NFS sensors [69](#)
configure SMB sensors [70](#)
increase fileset space [67](#)
mount file system [71](#)
replace disks [64](#)
start gpfs daemon [66](#)
start NSD [66](#)
start performance monitoring collector service [67](#)
start performance monitoring sensor service [68](#)
start the GUI service [71](#)
synchronize node clocks [67](#)
update drive firmware [65](#)

directed maintenance procedure (*continued*)
update enclosure firmware [65](#)
update host-adapter firmware [66](#)
directories
/tmp/mmfs [27](#)
disks
diagnosis [32](#)
hardware service [63](#)
hospital [32](#)
maintaining [31](#)
replacement [34](#)
replacing failed [35](#), [58](#)
DMP
replace disks [64](#)
update drive firmware [65](#)
update enclosure firmware [65](#)
update host-adapter firmware [66](#)
documentation
on web [xii](#)
drive firmware
updating [31](#)

E

enclosure components
replacing failed [42](#)
enclosure firmware
updating [31](#)
errpt command [27](#)
events
Array events [73](#)
Canister events [86](#)
Enclosure events [74](#)
Physical disk events [78](#)
Recovery group events [80](#)
server events [81](#)
virtual disk events [77](#)

F

failed disks, replacing [35](#), [58](#)
failed enclosure components, replacing [42](#)
failover, server [33](#)
files
mmfs.log [27](#)
firmware
updating [31](#)

G

getting started with troubleshooting [15](#)
GPFS
events [73](#), [74](#), [77](#), [78](#), [80](#), [81](#), [86](#)
RAS events
Array events [73](#)
Canister events [86](#)

GPFS (*continued*)

RAS events (*continued*)

Enclosure events [74](#)

Physical disk events [78](#)

Recovery group events [80](#)

server events [81](#)

virtual disk events [77](#)

GPFS log [27](#)

gpfs.snap command [27](#)

GUI

directed maintenance procedure [63](#)

DMP [63](#)

logs [23](#)

logsIssues with loading GUI [23](#), [25](#)

H

hardware service [63](#)

hospital, disk [32](#)

host adapter firmware

updating [31](#)

I

I/O node failure

restore [9](#)

IBM Elastic Storage Server

best practices for troubleshooting [19](#), [21](#)

IBM Spectrum Scale

back up data [16](#)

best practices for troubleshooting [15](#), [19](#)

ESS [5](#), [39](#)

events [73](#), [74](#), [77](#), [78](#), [80](#), [81](#), [86](#)

RAS events [73](#), [74](#), [77](#), [78](#), [80](#), [81](#), [86](#)

troubleshooting

best practices [17](#)

getting started [15](#)

warranty and maintenance [17](#)

information overview [vii](#)

L

license inquiries [113](#)

lspp command [28](#)

M

maintenance

disks [31](#)

message severity tags [89](#)

mmfs.log [27](#)

mmlsdisk command [28](#)

mmlsfs command [28](#)

N

node

crash [29](#)

hang [29](#)

notices [113](#)

NVR Partitions [1](#)

NVRAM pdisks

recreate [3](#)

O

overview

of information [vii](#)

P

patent information [113](#)

PMR [29](#)

preface [vii](#)

problem determination

documentation [27](#)

reporting a problem to IBM [27](#)

Problem Management Record [29](#)

R

RAS events

Array events [73](#)

Canister events [86](#)

Enclosure events [74](#)

Physical disk events [78](#)

Recovery group events [80](#)

server events [81](#)

virtual disk events [77](#)

rebalance, background task [33](#)

rebuild-1r, background task [33](#)

rebuild-2r, background task [33](#)

rebuild-critical, background task [33](#)

rebuild-offline, background task [33](#)

recovery groups

server failover [33](#)

repair-RGD/VCD, background task [33](#)

Replace

bad drives commandless disk replacement [41](#)

Replace disk

commandless [40](#)

replace disks [64](#)

replacement, disk [34](#)

replacing failed disks [35](#), [58](#)

replacing failed storage enclosure components [42](#)

report problems [17](#)

reporting a problem to IBM [27](#)

resolve events [17](#)

resources

on web [xii](#)

Restore

I/O node [9](#)

rpm command [27](#)

S

scrub, background task [33](#)

sda

NVR Partitions [1](#)

server failover [33](#)

service

reporting a problem to IBM [27](#)

service, hardware [63](#)

severity tags

messages [89](#)

SSD

logtip backup [5](#)

submitting [xiv](#)
support notifications [17](#)

T

tasks, background [33](#)
the IBM Support Center [29](#)
trademarks [114](#)
troubleshooting
 best practices
 report problems [17](#)
 resolve events [17](#)
 support notifications [17](#)
 update software [17](#)
 getting started [15](#)
 improper disk removal [39](#)
 Replacing logtip backup [5](#)
 warranty and maintenance [17](#)
Troubleshooting
 commandless disk replacement [40](#)
 fault tolerance for GLxC [50](#)
 GUI [23](#)
 Recovery Groups [25](#)
 Replace bad drives [41](#)

U

update drive firmware [65](#)
update enclosure firmware [65](#)
update host-adapter firmware [66](#)

V

vdisks
 data checksum [34](#)

W

warranty and maintenance [17](#)
web
 documentation [xii](#)
 resources [xii](#)



Product Number: 5765-DME
5765-DAE

GC28-3191-01

