IBM Elastic Storage System 3000
Version 6.0.0

Problem Determination Guide

GC28-3121-01

Note

Before using this information and the product it supports, read the information in "Notices" on page 95.

This edition applies to version 6 release 0 modification 0 of the following product 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)

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

Contents

Tables

About this information

This information is intended as a guide for administering IBM Elastic Storage® System (ESS) 3000.

Who should read this information

This information is intended for administrators of IBM Elastic Storage System (ESS) 3000 systems that include IBM Spectrum Scale RAID.

Related information

Related information

For information about:

- IBM Spectrum Scale, see:

  http://www.ibm.com/support/knowledgecenter/STXKQY/ibmspectrumscale_welcome.html
- mmvdisk command, see mmvdisk documentation.

Conventions used in this information

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

Table 1. Conventions

Convention	Usage
bold	**Bold** words or characters represent system elements that you must use literally, such as commands, flags, values, and selected menu options. Depending on the context, **bold** typeface sometimes represents path names, directories, or file names.
bold underlined	bold underlined keywords are defaults. These take effect if you do not specify a different keyword.
`constant width`	Examples and information that the system displays appear in `constant-width` typeface. Depending on the context, `constant-width` typeface sometimes represents path names, directories, or file names.
italic	*Italic* words or characters represent variable values that you must supply. *Italics* are also used for information unit titles, for the first use of a glossary term, and for general emphasis in text.
<*key*>	Angle brackets (less-than and greater-than) enclose the name of a key on the keyboard. For example, `<Enter>` refers to the key on your terminal or workstation that is labeled with the word *Enter*.

Table 1. Conventions (continued)

Convention	**Usage**

\	In command examples, a backslash indicates that the command or coding example continues on the next line. For example:
	`mkcondition -r IBM.FileSystem -e "PercentTotUsed > 90" \`
	`-E "PercentTotUsed < 85" -m p "FileSystem space used"`

{*item*}	Braces enclose a list from which you must choose an item in format and syntax descriptions.
[*item*]	Brackets enclose optional items in format and syntax descriptions.
`<Ctrl-x>`	The notation `<Ctrl-x>` indicates a control character sequence. For example, `<Ctrl-c>` means that you hold down the control key while pressing `<c>`.
item...	Ellipses indicate that you can repeat the preceding item one or more times.
|	In *synopsis* statements, vertical lines separate a list of choices. In other words, a vertical line means *Or*. In the left margin of the document, vertical lines indicate technical changes to the information.

How to submit your comments

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

`scale@us.ibm.com`

Chapter 1. Call home in 5146 and 5148 systems to resolve events raised for hardware issue

ESS version 5.x can generate call home events when a physical drive needs to be replaced in an attached enclosures.

ESS version 5.x automatically opens an IBM Service Request with service data, such as the location and FRU number to carryout the service task. The drive call home feature is only supported for drives installed in 5887, DCS3700 (1818), 5147-024 and 5147-084 enclosures in the 5146 and 5148 systems.

Background and overview

ESS 4.5 introduced ESS Management Server and I/O Server HW call home capability in ESS 5146 systems, where hardware events are monitored by the HMC managing these servers.

When a serviceable event occurs on one of the monitored servers, the Hardware Management Console (HMC) generates a call home event. This feature is only available in the 5146 systems as the 5146 systems are managed by the HMC. This feature is not available in 5148 systems as the 5148 systems are not managed by the HMC.

ESS 5.X provides additional Call Home capabilities for the drives in the attached enclosures of ESS 5146 and ESS 5148 systems. The call home for drive events does not require HMC, and uses the Electronic Service Agent (ESA) running on the EMS node.

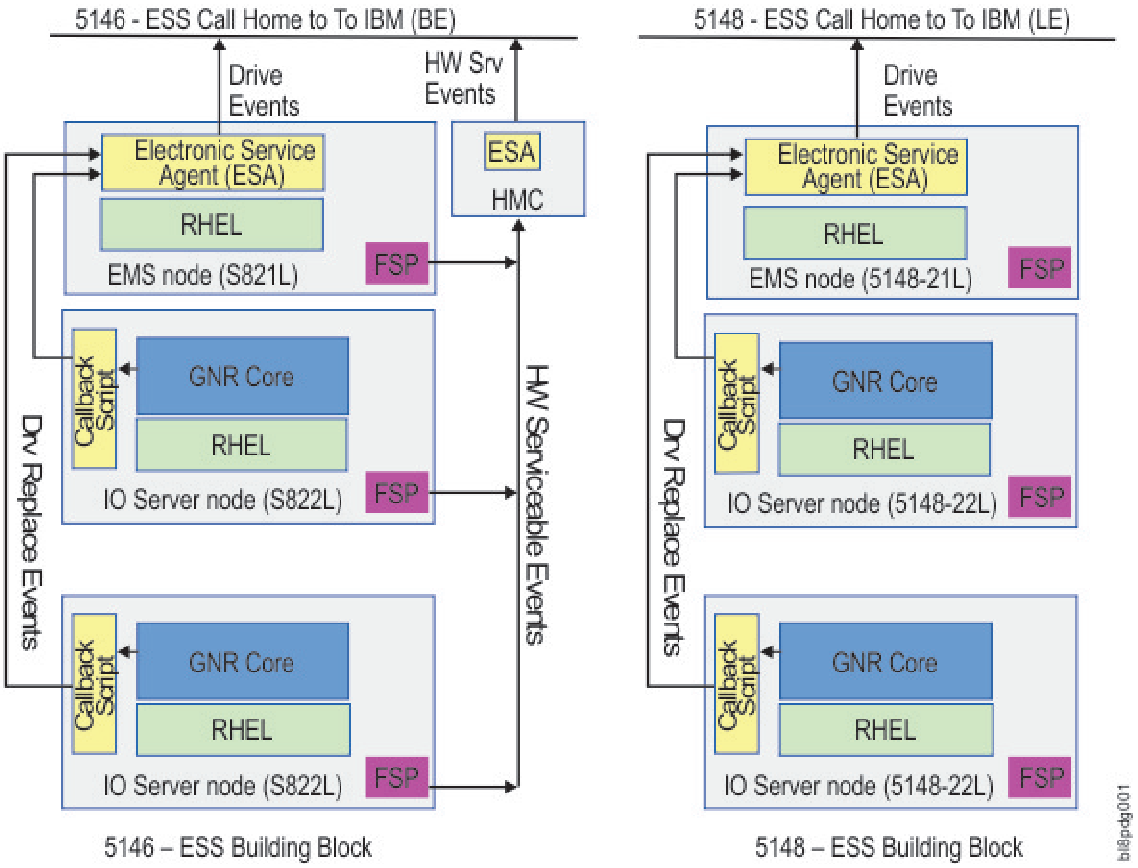


Figure 1. ESS Call Home block diagram

In ESS 5146 the HMC obtains the health status from the Flexible Service Process (FSP) of each server. When there is a serviceable event detected by the FSP, it is sent to the HMC, which initiates a call home event if needed. This function is not available in ESS 5148 systems.

The IBM Spectrum Scale RAID `pdisk` is an abstraction of a physical disk. A `pdisk` corresponds to exactly one physical disk, and belongs to exactly one de-clustered array within exactly one recovery group.

The attributes of a `pdisk` includes the following:

- The state of the `pdisk`
- The disk's unique worldwide name (WWN)
- The disk's field replaceable unit (FRU) code
- The disk's physical location code

When the pdisk state is `ok`, the `pdisk` is healthy and functioning normally. When the pdisk is in a `diagnosing` state, the IBM Spectrum Scale RAID disk hospital is performing a diagnosis task after an error has occurred.

The disk hospital is a key feature of the IBM Spectrum Scale RAID that asynchronously diagnoses errors and faults in the storage subsystem. When the pdisk is in a `missing` state, it indicates that the IBM Spectrum Scale RAID is unable to communicate with a disk. If a missing disk becomes reconnected and functions properly, its state changes back to ok. For a complete list of pdisk states and further information on pdisk configuration and administration, see IBM Spectrum Scale RAID Administration .

Any pdisk that is in the `dead`, `missing`, `failing` or `slow` state is known as a non-functioning pdisk. When the disk hospital concludes that a disk is no longer operating effectively and the number of non-functioning pdisks reaches or exceeds the replacement threshold of their de-clustered array, the disk hospital adds the replace flag to the pdisk state. The replace flag indicates the physical disk corresponding to the pdisk that must be replaced as soon as possible. When the pdisk state becomes `replace`, the drive replacement callback script is run.

The callback script communicates with the ESA over a REST API. The ESA is installed in the ESS Management Server (EMS), and initiates a call home task. The ESA is responsible for automatically opening a Service Request (PMR) with IBM support, and managing end-to-end life cycle of the problem.

Installing the IBM Electronic Service Agent

IBM Electronic Service Agent (ESA) for PowerLinux version 4.1 and later can monitor the ESS systems. It is installed in the ESS Management Server (EMS) during the installation of ESS version 5.X, or when upgrading to ESS 5.X.

The IBM Electronic Service Agent is installed when the **gssinstall** command is run. The **gssinstall** command can be used in one of the following ways depending on the system:

- For 5146 system:

  ```
  gssinstall_ppc64 -u
  ```

- For 5148 system:

  ```
  gssinstall_ppc64le -u
  ```

The rpm files for the esagent is found in the `/install/gss/otherpkgs/rhels7/<arch>/gss` directory.

Issue the following command to verify that the rpm for the esagent is installed:

```
rpm -qa | grep esagent
```

This gives an output similar to the following:

```
esagent.pLinux-4.5.2-1.noarch
```

If ESA is not installed, issue the following command:

```
cd /install/gss/otherpkgs/rhels7/<arch>/gss
rpm -ihv --nodeps esagent.pLinux-4.5.2-1.noarch.rpm
```

Login and activation

After the ESA is installed, the ESA portal can be reached by going to the following link:

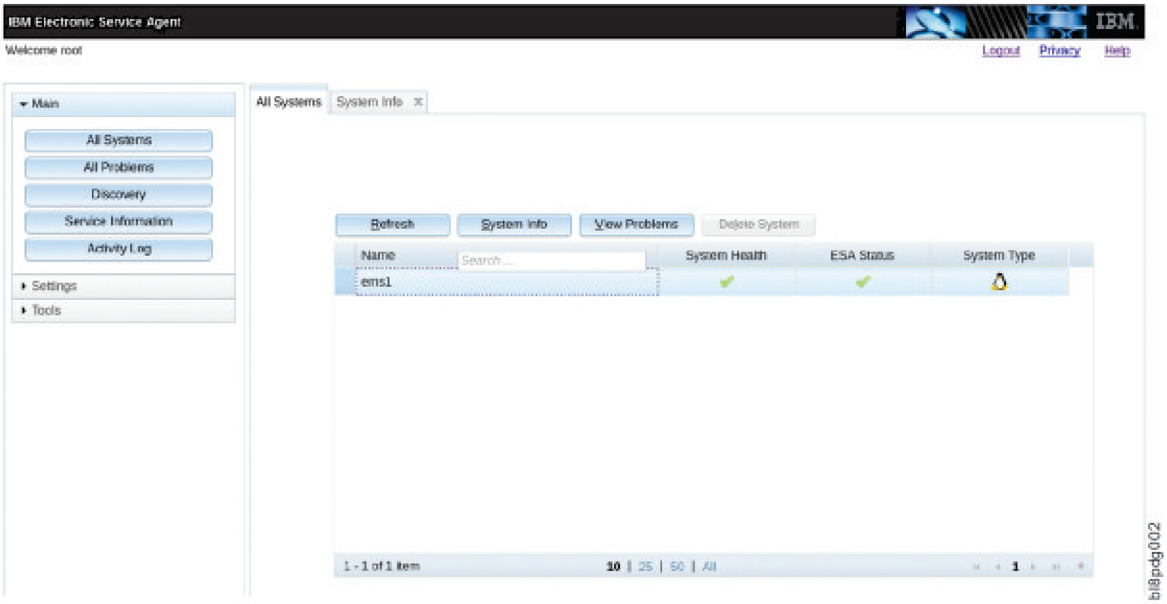
```
https://<EMS or ip>:5024/esa
```

For example:

```
https://192.168.45.20:5024/esa
```

The ESA uses port 5024 by default. It can be changed by using the ESA CLI if needed. For more information on ESA, see IBM Electronic Service Agent. On the Welcome page, log in to the IBM Electronic Service Agent GUI. If an untrusted site certificate warning is received, accept the certificate or click **Yes** to proceed to the IBM Electronic Service Agent GUI. You can get the context sensitive help by selecting the **Help** option located in the upper right corner.

After you have logged in, go to the **Main Activate ESA**, to run the activation wizard. The activation wizard requires valid contact, location and connectivity information.



Name	System Health	ESA Status	System Type
ems1	✓	✓	🐧

Figure 2. ESA portal after login

The All Systems menu option shows the node where ESA is installed. For example, ems1. The node where ESA is installed is shown as PrimarySystem in the **System Info**. The ESA Status is shown as **Online** only on the PrimarySystem node in the **System Info** tab.

Note: The ESA is not activated by default. In case it is not activated, you will get a message similar to the following:

```
[root@ems1 tmp]# gsscallhomeconf -E ems1 --show
IBM Electronic Service Agent (ESA) is not activated.
Activated ESA using /opt/ibm/esa/bin/activator -C and retry.
```

.

Electronic Service Agent configuration

Entities or systems that can generate events are called endpoints. The EMS, I/O Servers, and attached enclosures can be endpoints in ESS. Only enclosure endpoints can generate events, and the only event generated for call home is the disk replacement event. In the ESS 5146 systems, HMC can generate call home for certain node-related events.

In ESS, the ESA is only installed on the EMS, and automatically discovers the EMS as `PrimarySystem`. The EMS and I/O Servers have to be registered to ESA as endpoints. The **`gsscallhomeconf`** command is used to perform the registration task. The command also registers enclosures attached to the I/O servers by default.

The software call home is registered based on the customer information given while configuring the ESA agent. A software call home group `auto` is configured by default and the EMS node acts as the software call home server. The weekly and daily software call home data collection configuration is also activated by default.

The software call home uses the ESA network connection settings to upload the data to IBM. The ESA agent network setup must be complete and working for the software call home to work.

Note: You cannot configure the software call home without configuring the ESA. For more information, see Chapter 2, “Software call home,” on page 15.

```
usage: gsscallhomeconf [-h] ([-N NODE-LIST | -G NODE-GROUP] [--show] [--prefix PREFIX] [--suffix SUFFIX]
 -E ESA-AGENT [--register {node,all}] [--no-swcallhome] [--crvpd]
[--serial SOLN-SERIAL] [--model SOLN-MODEL] [--verbose]

optional arguments:
-h, --help                Show this help message and exit
-N NODE-LIST              Provide a list of nodes to configure.
-G NODE-GROUP             Provide name of node group.
--show                    Show callhome configuration details.
--prefix PREFIX           Provide hostname prefix. Use = between --prefix and value if the value starts with
-.
--suffix SUFFIX           Provide hostname suffix. Use = between --suffix and value if the value starts with
-.
-E ESA-AGENT              Provide nodename for esa agent node
--register {node,all}        Register endpoints(nodes, enclosure or all) with ESA.

--no-swcallhome              Do not configure software callhome while configuring hardware callhome
--crvpd                   Create vpd file.
--serial SOLN-SERIAL         Provide ESS solution serial number.
--model SOLN-MODEL        Provide ESS model.
--verbose                 Provide verbose output
```

A sample output is shown:

```
[root@ems1 ~]# gsscallhomeconf -E ems1 -N ems1,gss_ppc64 --suffix=-ib
2017-02-07T21:46:27.952187 Generating node list...
2017-02-07T21:46:29.108213 nodelist:    ems1 essio11 essio12
2017-02-07T21:46:29.108243 suffix used for endpoint hostname: -ib
End point ems1-ib registered successfully with systemid 802cd01aa0d3fc5137f006b7c9d95c26
End point essio11-ib registered successfully with systemid c7dba51e109c92857dda7540c94830d3
End point essio12-ib registered successfully with systemid 898fb33e04f5ea12f2f5c7ec0f8516d4
End point enclosure G5CT018 registered successfully with systemid
c14e80c240d92d51b8daae1d41e90f57
End point enclosure G5CT016 registered successfully with systemid
524e48d68ad875ffbeeec5f3c07e1acf
ESA configuration for ESS Callhome is complete.

Started configuring software callhome
Checking for ESA is activated or not before continuing.
Fetching customer detail from ESA.
Customer detail has been successfully fetched from ESA.
Setting software callhome customer detail.
Successfully set the customer detail for software callhome.
Enabled daily schedule for software callhome.
Enabled weekly schedule for software callhome.
Direct connection will be used for software calhome.
Successfully set the direct connection settings for software callhome.
Enabled software callhome capability.
Creating callhome automatic group
Created auto group for software call home and enabled it.
Software callhome configuration completed.
```

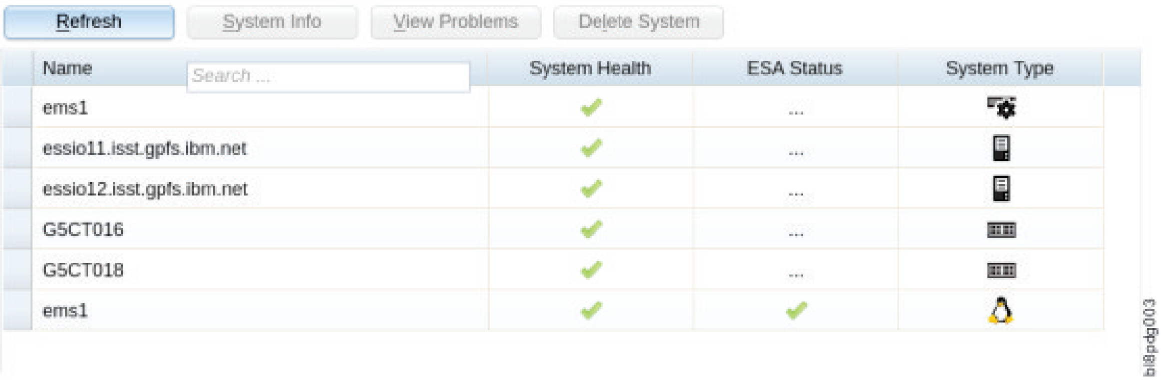
The **`gsscallhomeconf`** command logs the progress and error messages in the `/var/log/messages` file. There is a **`--verbose`** option that provides more details of the progress, as well error messages. The following example displays the type of information sent to the /var/log/messages file in the EMS by the **gsscallhomeconf** command.

```
[root@ems1 vpd]# grep ems1 /var/log/messages | grep gsscallhomeconf

Feb 8 01:37:39 ems1 gsscallhomeconf: [I] End point ems1-ib registered successfully with
 systemid 802cd01aa0d3fc5137f006b7c9d95c26
Feb 8 01:37:40 ems1 gsscallhomeconf: [I] End point essio11-ib registered successfully
 with systemid c7dba51e109c92857dda7540c94830d3
Feb 8 01:37:41 ems1 gsscallhomeconf: [I] End point essio12-ib registered successfully
 with systemid 898fb33e04f5ea12f2f5c7ec0f8516d4
Feb 8 01:43:04 ems1 gsscallhomeconf: [I] ESA configuration for ESS Callhome is complete.
```

⚠ **Attention:** The **gsscallhomeconf** command also configures the IBM Spectrum Scale call home setup. The IBM Spectrum Scale call home feature collects files, logs, traces, and details of certain system health events from the I/O and EMS nodes and services running on those nodes. These details are shared with the IBM support center for monitoring and problem determination. For more information on IBM Spectrum Scale call home, see the *Understanding call home* section in the IBM Knowledge Center.

The endpoints are visible in the ESA portal after registration, as shown in the following figure:



Name	System Health	ESA Status	System Type
ems1	✓	...	
essio11.isst.gpfs.ibm.net	✓	...	
essio12.isst.gpfs.ibm.net	✓	...	
G5CT016	✓	...	
G5CT018	✓	...	
ems1	✓	✓	

Figure 3. ESA portal after node registration

Name

Shows the name of the endpoints that are discovered or registered.

SystemHealth

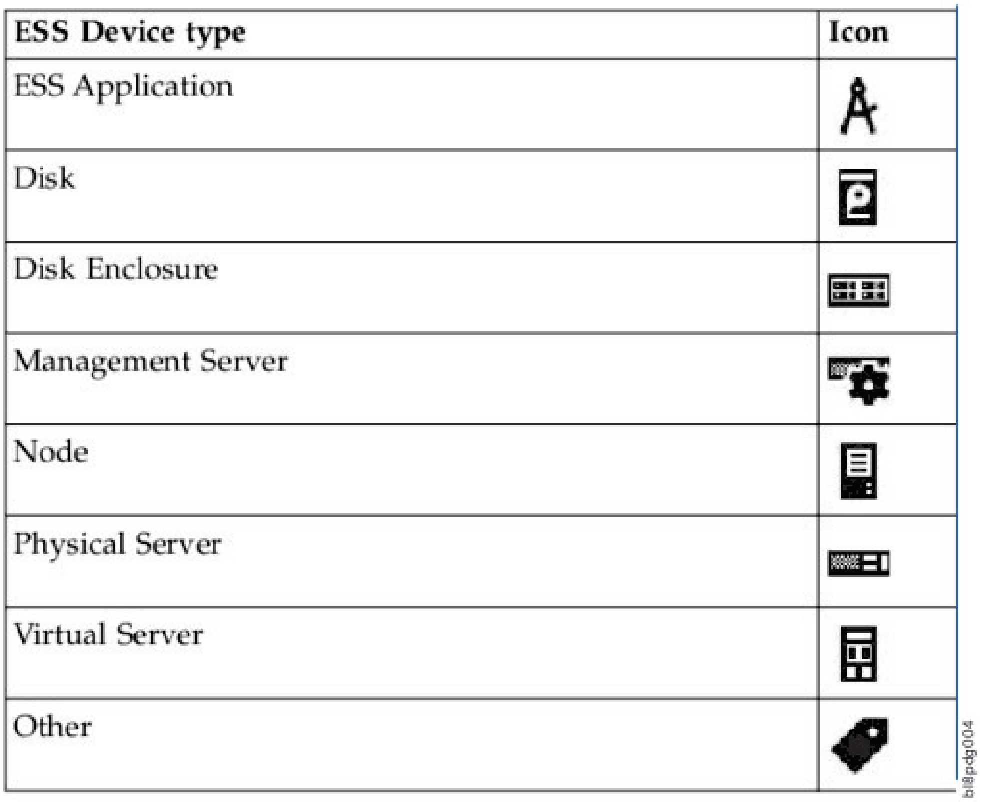
Shows the health of the discovered endpoints. A green icon (√) indicates that the discovered system is working fine. The red (X) icon indicates that the discovered endpoint has some problem.

ESAStatus

Shows that the endpoint is reachable. It is updated whenever there is a communication between the ESA and the endpoint.

SystemType

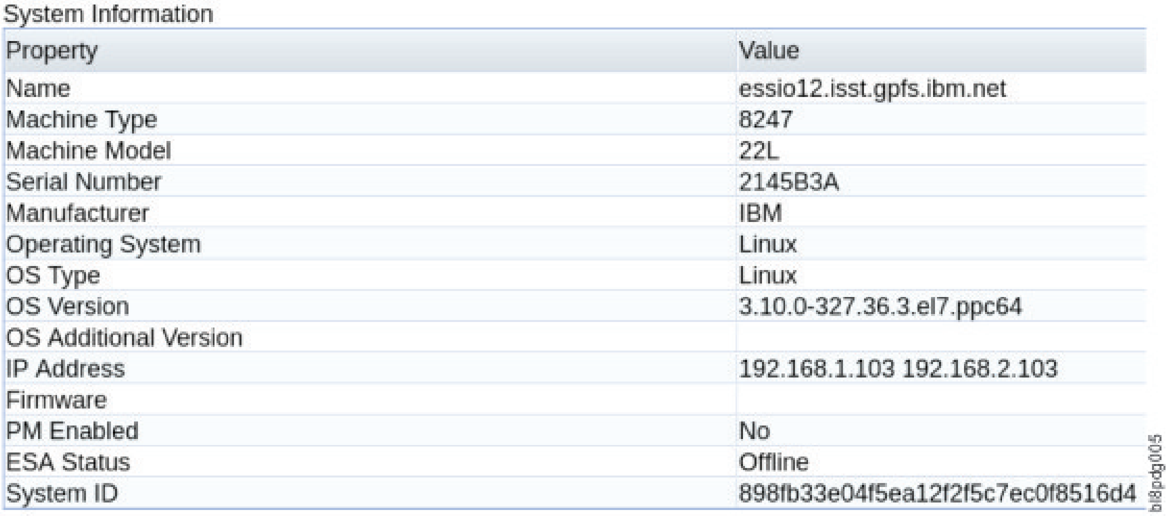
Shows the type of system being used. Following are the various ESS device types that the ESA supports.



ESS Device type	Icon
ESS Application	
Disk	
Disk Enclosure	
Management Server	
Node	
Physical Server	
Virtual Server	
Other	

Figure 4. List of icons showing various ESS device types

Detail information about the node can be obtained by selecting **System Information**. Here is an example of the system information:



System Information

Property	Value
Name	essio12.isst.gpfs.ibm.net
Machine Type	8247
Machine Model	22L
Serial Number	2145B3A
Manufacturer	IBM
Operating System	Linux
OS Type	Linux
OS Version	3.10.0-327.36.3.el7.ppc64
OS Additional Version	
IP Address	192.168.1.103 192.168.2.103
Firmware	
PM Enabled	No
ESA Status	Offline
System ID	898fb33e04f5ea12f2f5c7ec0f8516d4

Figure 5. System information details

When an endpoint is successfully registered, the ESA assigns a unique system identification (system id) to the endpoint. The system id can be viewed using the `--show` option.
For example:

```
[root@ems1 vpd]# gsscallhomeconf -E ems1 --show
System id and system name from ESA agent

{
"c14e80c240d92d51b8daae1d41e90f57": "G5CT018",
"c7dba51e109c92857dda7540c94830d3": "essio11-ib",
"898fb33e04f5ea12f2f5c7ec0f8516d4": "essio12-ib",
"802cd01aa0d3fc5137f006b7c9d95c26": "ems1-ib",
"524e48d68ad875ffbeeec5f3c07e1acf": "G5CT016"
}
```

When an event is generated by an endpoint, the node associated with the endpoint must provide the system id of the endpoint as part of the event. The ESA then assigns a unique event id for the event. The system id of the endpoints are stored in a file called `esaepinfo01.json` in the `/vpd` directory of the EMS and I/O servers that are registered. The following example displays a typical `esaepinfo01.json` file:

```
[root@ems1 vpd]# cat esaepinfo01.json
{
"encl": {
"G5CT016": "524e48d68ad875ffbeeec5f3c07e1acf",
"G5CT018": "c14e80c240d92d51b8daae1d41e90f57"
},
"esaagent": "ems1", "node": {
"ems1-ib": "802cd01aa0d3fc5137f006b7c9d95c26",
"essio11-ib": "c7dba51e109c92857dda7540c94830d3",
 "essio12-ib": "898fb33e04f5ea12f2f5c7ec0f8516d4"
}
```

In the ESS 5146, the **`gsscallhomeconf`** command requires the ESS solution vpd file that contains the IBM Machine Type and Model (MTM) and serial number information to be present. The vpd file is used by the ESA in the call home event. If the vpd file is absent, the **`gsscallhomeconf`** command fails, and displays an error message that the vpd file is missing. In this case, you can rerun the command with the `--crvpd` option, and provide the serial number and model number using the `--serial` and `--model` options. In ESS 5148, the vpd file is auto generated if not present.

The system vpd information is stored in the `essvpd01.json` file in the EMS `/vpd` directory. Here is an example of a vpd file:

```
[root@ems1 vpd]# cat essvpd01.json
{
"groupname": "ESSHMC", "model": "GS2",
 "serial": "219G17G", "system": "ESS", "type": "5146"
}
[root@ems1 vpd]# cat essvpd01.json
{
"groupname": "ESSHMC", "model": "GS2",
 "serial": "219G17G", "system": "ESS", "type": "5146"
}
```

To check if the ESA rpms are installed, run the following command:

```
rpm -qa | grep esagent
```

To check if the ESA is configured and activated, run the following command:

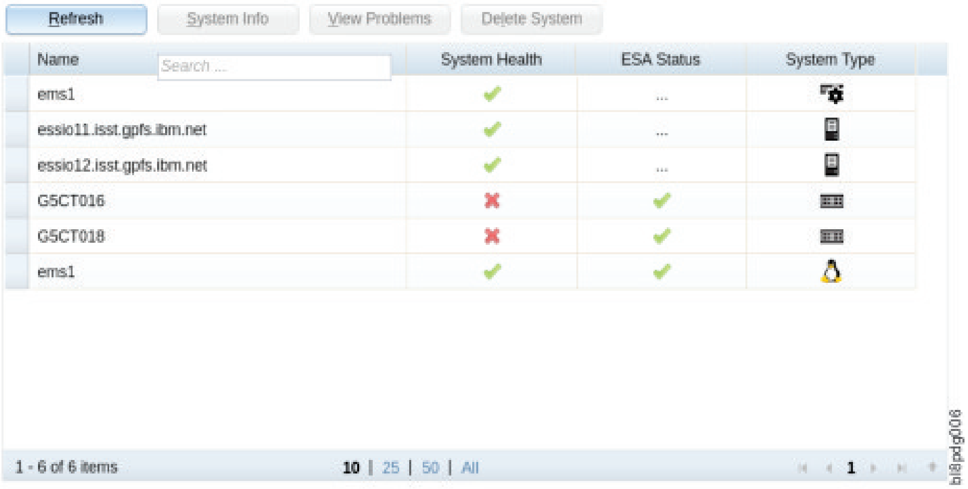
```
gsscallhomeconf -E ems1 --show
```

For more information on ESA configuration and activation, see "Login and activation" on page 3. For information on network connectivity and end-to-end setup, see "Test call home" on page 11.

Overview of a problem report

After the ESA is activated, and the endpoints for the nodes and enclosures are registered, they can send an event request to the ESA to initiate a call home.

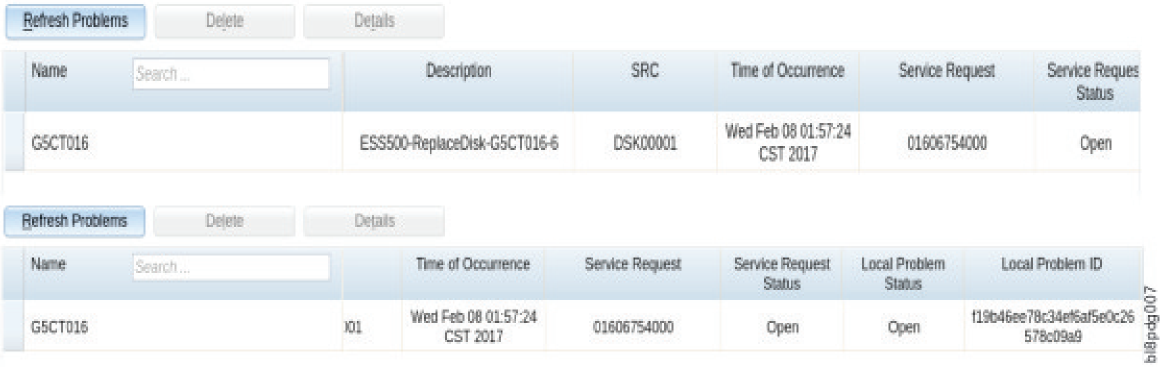
For example, when `replace` is added to a pdisk state, indicating that the corresponding physical drive needs to be replaced, an event request is sent to the ESA with the associated system id of the enclosure where the physical drive resides. Once the ESA receives the request it generates a call home event. Each server in the ESS is configured to enable callback for IBM Spectrum Scale RAID related events. These callbacks are configured during the cluster creation, and updated during the code upgrade. The ESA can filter out duplicate events when event requests are generated from different nodes for the same physical drive. The ESA returns an event identification value when the event is successfully processed. The ESA portal updates the status of the endpoints. The following figure shows the status of the enclosures when the enclosure contains one or more physical drives identified for replacement:



Name	System Health	ESA Status	System Type
ems1	✓	...	
essio11.isst.gpfs.ibm.net	✓	...	
essio12.isst.gpfs.ibm.net	✓	...	
G5CT016	✗	✓	
G5CT018	✗	✓	
ems1	✓	✓	

Figure 6. ESA portal showing enclosures with drive replacement events

The problem descriptions of the events can be seen by selecting the endpoint. You can select an endpoint by clicking the red X. The following figure shows an example of the problem description.



Name	Description	SRC	Time of Occurrence	Service Request	Service Reques Status
G5CT016	ESS500-ReplaceDisk-G5CT016-6	DSK00001	Wed Feb 08 01:57:24 CST 2017	01606754000	Open

Name		Time of Occurrence	Service Request	Service Request Status	Local Problem Status	Local Problem ID
G5CT016	)01	Wed Feb 08 01:57:24 CST 2017	01606754000	Open	Open	f19b46ee78c34ef6af5e0c26578c09a9

Figure 7. Problem Description

Name

It is the serial number of the enclosure containing the drive to be replaced.

Description

It is a short description of the problem. It shows ESS version or generation, service task name and location code. This field is used in the synopsis of the problem (PMR) report.

SRC

It is the Service Reference Code (SRC). An SRC identifies the system component area. For example, DSK XXXXX, that detected the error and additional codes describing the error condition. It is used by the support team to perform further problem analysis, and determine service tasks associated with the error code and event.

Time of Occurrence

It is the time when the event is reported to the ESA. The time is reported by the endpoints in the UTC time format, which ESA displays in local format.

Service request

It identifies the problem number (PMR number).

Service Request Status

It indicates reporting status of the problem. The status can be one of the following:

Open

No action is taken on the problem.

Pending

The system is in the process of reporting to the IBM support.

Failed

All attempts to report the problem information to the IBM support has failed. The ESA automatically retries several times to report the problem. The number of retries can be configured. Once failed, no further attempts are made.

Reported

The problem is successfully reported to the IBM support.

Closed

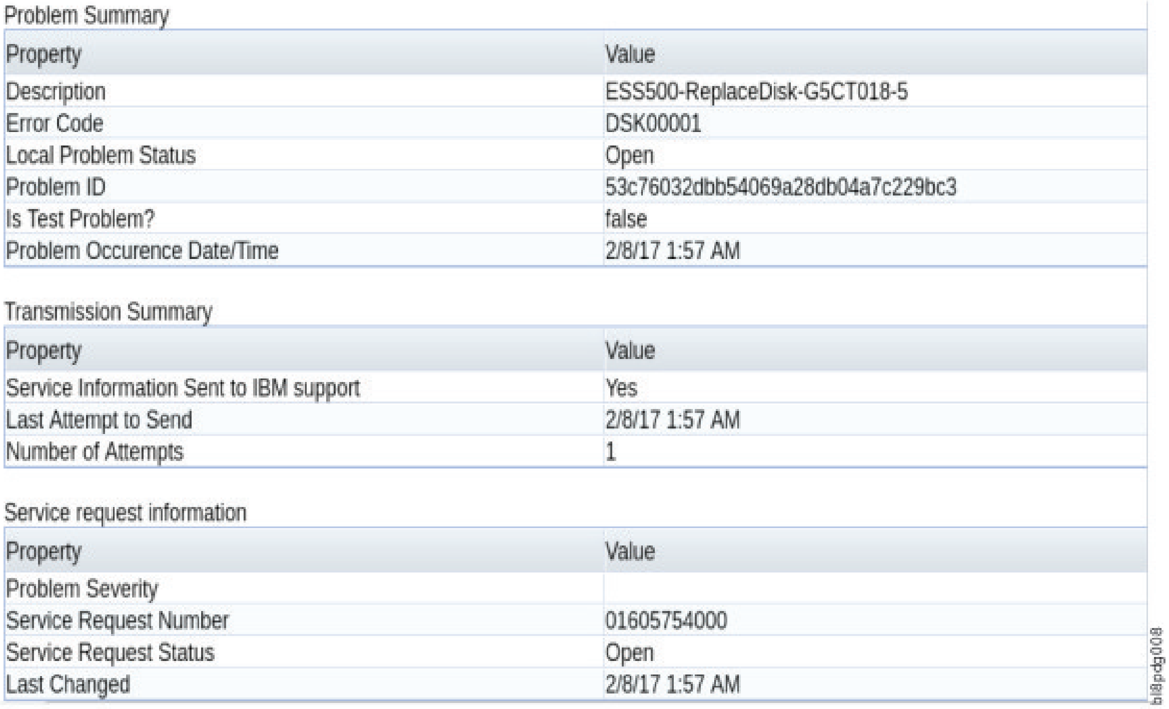
The problem is processed and closed.

Local Problem ID

It is the unique identification or event id that identifies a problem.

Problem details

Further details of a problem can be obtained by clicking the **Details** button. The following figure shows an example of a problem detail.



Problem Summary

Property	Value
Description	ESS500-ReplaceDisk-G5CT018-5
Error Code	DSK00001
Local Problem Status	Open
Problem ID	53c76032dbb54069a28db04a7c229bc3
Is Test Problem?	false
Problem Occurence Date/Time	2/8/17 1:57 AM

Transmission Summary

Property	Value
Service Information Sent to IBM support	Yes
Last Attempt to Send	2/8/17 1:57 AM
Number of Attempts	1

Service request information

Property	Value
Problem Severity	
Service Request Number	01605754000
Service Request Status	Open
Last Changed	2/8/17 1:57 AM

Figure 8. Example of a problem summary

If an event is successfully reported to the ESA, and an event ID is received from the ESA, the node reporting the event uploads additional support data to the ESA that are attached to the problem (PMR) for further analysis by the IBM support team.

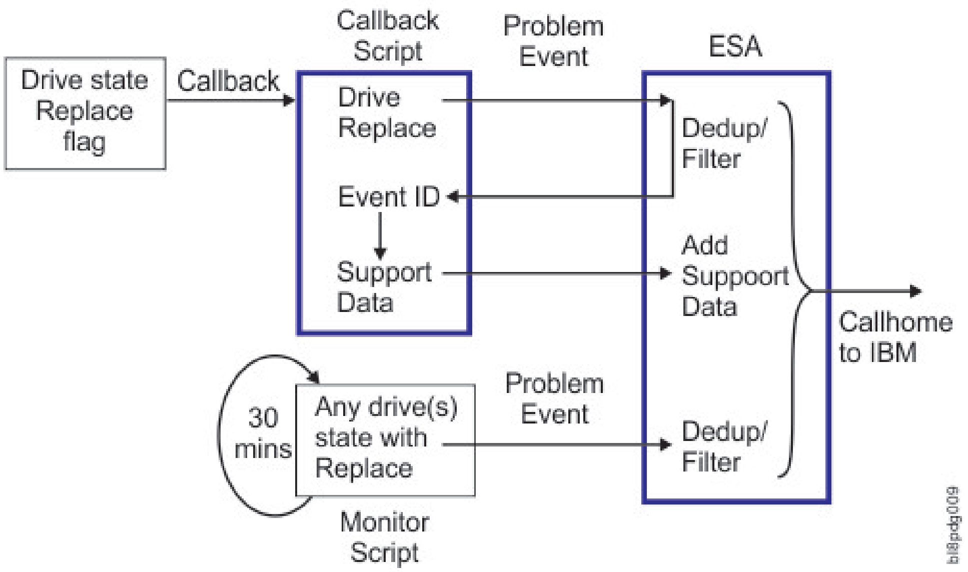


Figure 9. Call home event flow

The callback script logs information in the `/var/log/messages` file during the problem reporting episode. The following examples display the messages logged in the `/var/log/message` file generated by the `essio11` node:

- Callback script is invoked when the drive state changes to `replace`. The callback script sends an event to the ESA:

  ```
  Feb 8 01:57:24 essio11 gsscallhomeevent: [I] Event successfully sent
  for end point G5CT016, system.id 524e48d68ad875ffbeeec5f3c07e1acf,
   location G5CT016-6, fru 00LY195.
  ```

- The ESA responds by returning a unique event ID for the system ID in the json format.

  ```
  Feb 8 01:57:24 essio11 gsscallhomeevent:
  {#012 "status-details": "Received and ESA is processing",
  #012 "event.id": "f19b46ee78c34ef6af5e0c26578c09a9",
  #012 "system.id": "524e48d68ad875ffbeeec5f3c07e1acf",
  #012 "last-activity": "Received and ESA is processing"
  #012}
  ```

  Note: Here `#012` represents the new line feed `\n`.

- The callback script runs the **`ionodedatacol.sh`** script to collect the support data. It collects the `mmfs.log.latest,` file and the last 24 hours of the kernel messages in the journal into a .tgz file.

  ```
  Feb 8 01:58:15 essio11 gsscallhomeevent: [I] Callhome data collector
  /opt/ibm/gss/tools/samples/ionodechdatacol.sh finished

  Feb 8 01:58:15 essio11 gsscallhomeevent: [I] Data upload successful
  for end point 524e48d68ad875ffbeeec5f3c07e1acf
  and event.id f19b46ee78c34ef6af5e0c26578c09a9
  ```

Call home monitoring

A callback is a one-time event. Therefore, it is triggered when the disk state changes to `replace`. If the ESA misses the event , for example if the EMS is down for maintenance, the call home event is not generated by the ESA.

To mitigate this situation, the `callhomemon.sh` script is provided in the `/opt/ibm/gss/tools/samples` directory of the EMS. This script checks for pdisks that are in the `replace` state, and sends an event to the ESA to generate a call home event if there is no open PMR for the corresponding physical drive. This script can be run on a periodic interval. For example, every 30 minutes.

In the EMS, create a cronjob as follows:

1. Open crontab editor using the following command:

   ```
   crontab -e
   ```

2. Setup a periodic cronjob by adding the following line:

   ```
   */30 * * * * /opt/ibm/gss/tools/samples/callhomemon.sh
   ```

3. View the cronjob using the following command:

   ```
   crontab -l
    [root@ems1 deploy]# crontab -l
   */30 * * * * /opt/ibm/gss/tools/samples/callhomemon.sh
   ```

The call home monitoring protects against missing a call home due to the ESA missing a callback event. If a problem report is not already created, the call home monitoring ensures that a problem report is created.

Note: When the call home problem report is generated by the monitoring script, as opposed to being triggered by the callback, the problem support data is not automatically uploaded. In this scenario, the IBM support can request support data from the customer.

Upload data
The following support data is uploaded when the system displays a drive replace notification:

- The output of **`mmlspdisk`** command for the pdisk that is in replace state.
- Additional support data is provided only when the event is initiated as a response to a callback. The following information is supplied in a .tgz file as additional support data:
  - `mmfs.log.latest` from the node which generates the event.
  - Last 24 hours of the kernel messages (from journal) from the node which generates the event.

Note: If a PMR is created because of the periodic checking of the replaced drive state, for example, when the callback event is missed, additional support data is not provided.

Uninstalling and reinstalling the IBM Electronic Service Agent

The ESA is not removed when the **`gssdeploy -c`** command is run to clean up the system.

The ESA rpm files must be removed manually if needed. Issue the following command to remove the rpm files for the esagent:

```
yum remove esagent.pLinux-4.2.0-9.noarch
```

You can issue the following command to reinstall the rpm files for the esagent. The esagent requires the `gpfs.java` file to be installed. The `gpfs.java` file is automatically installed by the `gssinstall` and `gssdeploy` script. The dependencies may still not be resolved. In such case, use the `--nodeps` option to install it.

```
cd /install/gss/otherpkgs/rhels7/<arch>/gss
rpm -ivh --nodeps esagent.pLinux-4.5.2-1.noarch.rpm
```

Test call home

The configuration and setup for call home must be tested to ensure that the disk replace event can trigger a call home.

The test is composed of three steps:

- ESA connectivity to IBM - Check connectivity from ESA to IBM network. This might not be required if done during the activation.

```
/opt/ibm/esa/bin/verifyConnectivity -t
```

- ESA test Call Home - Test call home from the ESA portal. From the `All System` tab, check the system health of the endpoint, and it will show the button for generating Test Problem.
- ESS call home script setup to ensure that the callback script is setup correctly.

Verify that the periodic monitoring is setup.

```
crontab -l
 [root@ems1 deploy]# crontab -l
*/30 * * * * /opt/ibm/gss/tools/samples/callhomemon.sh
```

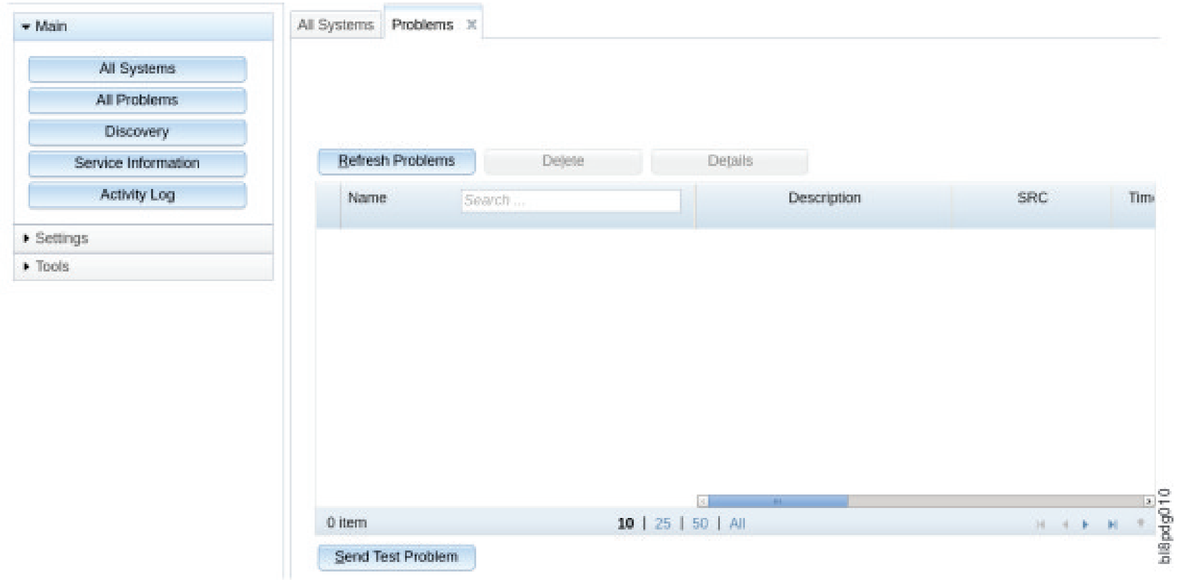


Figure 10. Sending a Test Problem

Callback Script Test

Verify that the system is healthy by issuing the **gnrhealthcheck** command. You must also verify that the active recovery group (RG) server is the primary recovery group server for all recovery groups. For more recovery group details, see the *IBM Spectrum Scale RAID: Administration* guide.

To test the callback script, select a pdisk from each enclosure alternating recovery groups. The purpose of the test call home events is to ensure that all the attached enclosures can generate call home events by using both the I/O servers in the building block.

For example, in a GS2 system with 5885 enclosure, one can select pdisks e1s02 (left RG) and e2s20 (right RG). You must find the corresponding recovery group and active server for these pdisks. Send a disk event to the ESA from the active recovery group server as shown in the following steps:.

Examples:

1. ssh to essio11

   ```
   gsscallhomeevent --event pdReplacePdisk
   --eventName "Test symptom generated by Electronic Service Agent"
   --rgName rg_essio11-ib --pdName e1s02
   ```

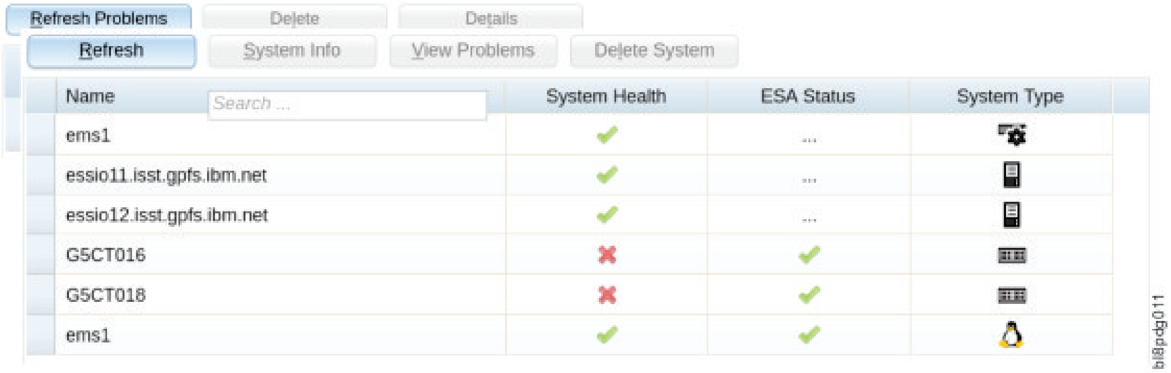
   Here the recovery group is *rg_essio11-ib*, and the active server is *essio11-ib*.
2. ssh to essio12

```
gsscallhomeevent --event pdReplacePdisk
 --eventName "Test symptom generated by Electronic Service Agent”
 --rgName rg_essio12-ib --pdName e2s20
```

Here the recovery group is *rg_essio12-ib*, and the active server is *essio12-ib*.

Note: Ensure that you state `Test symptom generated by Electronic Service Agent` in the **`--eventName`** option. Check in the ESA that the enclosure system health is showing the event. You might have to refresh the screen to make the event visible.

Select the event to see the details.



Name	System Health	ESA Status	System Type
ems1	✓	...	⚙
essio11.isst.gpfs.ibm.net	✓	...	▯
essio12.isst.gpfs.ibm.net	✓	...	▯
G5CT016	✗	✓	▤
G5CT018	✗	✓	▤
ems1	✓	✓	🐧

Figure 11. List of events

For DCS3700 enclosures, the pdisks to test call home can have the e1d1s1 and the e2d5s10 (e3d1s1, e4d5s10 etc.) alternating for recovery groups. For 5148-084 enclosures, the pdisks to test call home can have the e1d1s1 (or e1d1s1ssd) and the e2d2s14 (e3d1s1, e4d2s14 etc) alternating for the recovery groups.

Post setup activities

- Delete any test problems.
- If the system has a 4U enclosure (DCS3700) in the configuration, obtain the actual matching seven digit serial number, and keep it available if needed. The IBM support will need this serial number for handling the problem properly.

Chapter 2. Software call home

The software call home feature collects files, logs, traces, and details of certain system health events from different nodes and services in an IBM Spectrum Scale cluster.

These details are shared with the IBM® support center for monitoring and problem determination. For more information on call home, see Installing call home and Understanding call home.

Configuring hardware and software call home

You can configure call home (hardware and software) using the **gsscallhomeconf** command. You can use the --no-swcallhome option to set up just the call home hardware, and skip the software call home set up.

The call home hardware and software call home can be set up using the following command:

```
[root@ems1 ~]# gsscallhomeconf -E ems1 -N ems1,gss_ppc64 --suffix=-ib
```

The command gives an output similar to the following:

```
2017-02-07T21:46:27.952187 Generating node list...
2017-02-07T21:46:29.108213 nodelist: ems1 essio11 essio12
2017-02-07T21:46:29.108243 suffix used for endpoint hostname: -ib
End point ems1-ib registered successfully with systemid 802cd01aa0d3fc5137f006b7c9d95c26
End point essio11-ib registered successfully with systemid c7dba51e109c92857dda7540c94830d3
End point essio12-ib registered successfully with systemid 898fb33e04f5ea12f2f5c7ec0f8516d4
End point enclosure G5CT018 registered successfully with systemid
c14e80c240d92d51b8daae1d41e90f57
End point enclosure G5CT016 registered successfully with systemid
524e48d68ad875ffbeeec5f3c07e1acf
ESA configuration for ESS Callhome is complete.
Started configuring software callhome
Checking for ESA is activated or not before continuing.
Fetching customer detail from ESA.
Customer detail has been successfully fetched from ESA.
Setting software callhome customer detail.
Successfully set the customer detail for software callhome.
Enabled daily schedule for software callhome.
Enabled weekly schedule for software callhome.
Direct connection will be used for software calhome.
Successfully set the direct connection settings for software callhome.
Enabled software callhome capability.
Creating callhome automatic group
Created auto group for software call home and enabled it.
Software callhome configuration completed.
```

If you want to skip the software call home set up, use the following command:

```
[root@ems3 ~]# gsscallhomeconf -E ems3 -N ems3,gss_ppc64 --suffix=-te --register=all --no-swcallhome
```

The command gives an output similar to the following:

```
2017-01-23T05:34:42.005215 Generating node list...
2017-01-23T05:34:42.827295 nodelist:    ems3 essio31 essio32
2017-01-23T05:34:42.827347 suffix used for endpoint hostname: -te
End point ems3-te registered sucessfully with systemid 37e5c23f98090750226f400722645655
End point essio31-te registered sucessfully with systemid 35ae41e0388e08fd01378ae5c9a6ffef
End point essio32-te registered sucessfully with systemid 9ea632b549434d57baef7c999dbf9479
End point enclosure SV50321280 registered sucessfully with systemid 600755dc0aa2014526fe5945981b0e08
End point enclosure SV50918672 registered sucessfully with systemid 92aa6428102b44a4a1c9a293402b324c
ESA configuration for ESS Callhome is complete.
```

Important: If the software call home set up has been skipped, it can be reconfigured again. However, the user needs to reconfigure both the call home hardware and the software call home again.

Chapter 3. Replacing an ESS 3000 canister server

If one of the canister servers in ESS 3000 has a hardware failure, the ESS 3000 continues to operate on the remaining canister server.

The failed canister server must be replaced as soon as possible. Follow these steps to replace the failed canister:

1. Obtain the appropriate FRU for the canister.
2. Follow the steps that are given in the *Replacing a failed node canister* section in the ESS 3000 Service Guide.
3. Follow the software deployment procedure to reinstall the new canister by using the same hostname and network address as the failed canister, given in the ESS 3000 quick deployment.

   Important: The replacement canister must be installed with the same xCAT hostname and network addresses and configuration as the failed canister.
4. Perform an `essstoragequickcheck` to verify that the expected NVMe drive configuration is visible to the new canister. Do not perform any other disk checks as the drives contain user data, and are in use by the remaining canister.
5. Run the **`mmsdrrestore`** command to restore the GPFS cluster identity for the new canister.
6. Start GPFS on the new canister.

Chapter 4. Best practices for troubleshooting

Following certain best practices make the troubleshooting process easier.

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 59.
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 **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.

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. The 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 **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 *mmbackup command* in *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 CLI or 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 function 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, you 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 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 Knowledge Center 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 .

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 Knowledge Center.

Chapter 5. 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:

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

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

Chapter 6. ESS 3000 deployment troubleshooting: Helpful podman, Ansible, and log information

Creating CES shared root file system for protocol nodes

Use the following commands to create a CES shared root file system for protocol nodes.

```
mmvdisk vs define --vs vs_cesroot --rg gssio1rg,gssio2rg --code 8+2p --bs 4M --ss 20g --nsd-
usage dataAndMetadata --sp system
mmvdisk vs create --vs vs_cesroot
mmvdisk fs create --fs cesSharedRoot --vs vs_cesroot --mmcrfs -T /gpfs/cesSharedRoot
```

Adding additional ESS 3000 storage to existing file system

Before doing these steps, follow the steps in *ESS 3000 initial setup instructions* in *ESS 3000: Quick Deployment Guide*. Make sure that you update the /etc/hosts file with the new node names and IP addresses. Copy the updated /etc/hosts to all nodes before starting. Stop after creating the network bonds.

1. Add ESS 3000 nodes to the current file system.

   ```
   ess3krun -N NodesAlreadyinCluster cluster --add-3k \
   NewNode1,NewNode2 --suffix=Suffix
   ```

2. Configure the mmvdisk node class. A unique node class name is required for a new building block.

   ```
   mmvdisk server configure --nc ChosenNodeClassName --recycle one
   ```

3. Create the recovery group.

   ```
   mmvdisk rg create --rg ChosenRGName \
   --nc ChosenNodeClassName
   ```

4. Define the vdisk set.

   ```
   mmvdisk vs define --vs ChosenVdiskSetName --rg ChosenRGName --code RAIDCode \
   --bs BlockSize --ss SetSize --nsd-usage dataOnly --sp data
   ```

   Note: For this example command, it is assumed that you are adding data only vdisks to the existing file system. You might have a different use case, so adjust options accordingly.

   ```
   Example values (adjust to meet needs of existing filesystem):
   --code 8+2p
   ```

```
--bs 4M
--ss 80%
```

5. Create the vdisk set.

   ```
   mmvdisk vs create --vs ChosenVdiskSetName
   ```

6. Add the vdisk set to the file system.

   ```
   mmvdisk fs add --file-system FileSystem --vdisk-set ChosenVdiskSetName
   ```

   FileSystem is the name of the file system that you are adding the storage to.

7. Add the new nodes to performance monitoring.

   ```
   mmchnode --perfmon -N NewNode1,NewNode2
   ```

8. Fix the compDB.

   ```
   mmaddcompspec default --replace
   ```

9. Start or restart the GUI on the EMS node.

   ```
   systemctl restart gpfsgui
   ```

Adding ESS 3000 to an ESS for Power environment

Before adding ESS 3000 to an existing ESS for Power® environment, the existing ESS system must already be converted to mmvdisk.

Before doing these steps, follow the steps in *ESS 3000 initial setup instructions* in *ESS 3000: Quick Deployment Guide*. Make sure that you update the `/etc/hosts` file with the new node names and IP addresses. Copy the updated `/etc/hosts` to all nodes before starting. Stop after creating the network bonds.

1. Add ESS 3000 nodes to the existing ESS system by running the following command from one of the canister nodes.

   ```
   essaddnode -N ess3k4a,ess3k4b --suffix=-ib --accept-license --no-fw-update  \
   --cluster-node ems1-ib --nodetype ess3k
   ```

   For this example command, it is assumed that:

   - The new ESS 3000 system has two canister nodes called `ess3k4a` and `ess3k4b`.
   - You are adding the nodes over Infiniband. Although, this procedure also works with Ethernet.

2. Configure mmvdisk node class. A unique node class name is required for a new building block.

   ```
   mmvdisk server configure --nc ChosenNodeClassName --recycle one
   ```

3. Create the recovery group.

   ```
   mmvdisk rg create --rg ChosenRGName \
   --nc ChosenNodeClassName
   ```

4. Define the vdisk set.

   ```
   mmvdisk vs define --vs ChosenVdiskSetName --rg ChosenRGName --code RAIDCode \
   --bs BlockSize --ss SetSize --nsd-usage dataOnly --sp data
   ```

   Note: For this example command, it is assumed that you are adding data only vdisks to the existing file system. You might have a different use case, so adjust options accordingly.

   ```
   Example values (adjust to meet needs of existing filesystem):
   --code 8+2p
   --bs 4M
   --ss 80%
   ```

5. Create the vdisk set.

   ```
   mmvdisk vs create --vs ChosenVdiskSetName
   ```

6. Add the vdisk set to the file system.

   ```
   mmvdisk fs add --file-system FileSystem --vdisk-set ChosenVdiskSetName
   ```

   FileSystem is the name of the file system that you are adding the storage to.

7. Add the new nodes to performance monitoring.

   ```
   mmchnode --perfmon -N NewNode1,NewNode2
   ```

8. Fix the compDB.

   ```
   mmaddcompspec default --replace
   ```

9. Start or restart the GUI on the EMS node.

   ```
   systemctl restart gpfsgui
   ```

Cleaning up an existing mmvdisk environment

1. Unmount the file system:

   ```
   mmumount FileSystem -a
   ```

2. Delete the file system:

   ```
   mmdelfs FileSystem
   ```

   You can also delete the file system by using **mmvdisk** (including vdisk set and recovery group):

   ```
   mmvdisk filesystem delete --file-system FileSystem
   ```

   This command also deletes the vdisk set.

3. List the vdisk sets:

   ```
   mmvdisk vdiskset list
   ```

4. Delete the vdisk set for the deleted file system:

   ```
   mmvdisk vdiskset delete --vdisk-set VdiskSet
   ```

   This command also deletes the NSDs and data and metadata vdisk.

5. Undefine vdisk sets:

   ```
   mmvdisk vdiskset undefine --vdisk-set VdiskSet
   ```

6. List the recovery groups:

   ```
   mmvdisk recoverygroup list
   ```

7. Delete the recovery groups:

   ```
   mmvdisk recoverygroup delete --recovery-group RecoveryGroup
   ```

8. List the mmvdisk servers:

   ```
   mmvdisk server list
   ```

9. Unconfigure the servers:

   ```
   mmvdisk server unconfigure --node-class ServerNodeClass
   ```

10. Delete the node class:

```
mmvdisk nodeclass delete --node-class ServerNodeClass
```

Troubleshooting issues when running the container

If you are facing issues when running container with **`essmgr -r`**, you can try these steps.

1. Clean up the `cni` directory by removing this directory.

   ```
   /var/lib/cni/networks/podman
   ```

2. Re-create the bridge.

   ```
   ip link set dmgtbr down ; brctl delbr dmgtbr
   ./essmgr -n -c ./essmgr1.yml
   brctl show
   ```

Debugging deployment issues

When the **`ess3krun`** is used, it issues Ansible commands to the target. You can check the following logs to debug the progress of those commands.

- On the canister, run this command: **`grep -i ansible-command /var/log/messages`**

  Example output:

  ```
  Feb 28 17:21:59 fab3a ansible-command[7300]: Invoked with _raw_params=ofed_info -n warn=True
  _uses_shell=False stdin_add_newline=True
  strip_empty_ends=True argv=None chdir=None executable=None creates=None removes=None
  stdin=None
  Feb 28 17:27:01 fab3a ansible-command[4884]: Invoked with _raw_params=/xcatpost/
  ess_ofed.ess3k warn=True _uses_shell=False stdin_add_newline=True
  strip_empty_ends=True argv=None chdir=None executable=None creates=None removes=None
  stdin=None
  Feb 28 17:41:43 fab3a ansible-command[44520]: Invoked with _raw_params=/usr/lpp/mmfs/bin/
  mmlscluster warn=True _uses_shell=False stdin_add_newline=True
  strip_empty_ends=True argv=None chdir=None executable=None creates=None removes=None
  stdin=None
  Feb 28 17:41:44 fab3a ansible-command[44636]: Invoked with _uses_shell=True
  _raw_params=/usr/lpp/mmfs/bin/mmcommon showLocks | grep CCR warn=True stdin_add_newline=True
  strip_empty_ends=True argv=None chdir=None executable=None creates=None removes=None
  stdin=None
  Feb 28 17:46:47 fab3a ansible-command[5133]: Invoked with _raw_params=/usr/lpp/mmfs/bin/
  mmbuildgpl warn=True _uses_shell=False
  stdin_add_newline=True strip_empty_ends=True argv=None chdir=None executable=None
  creates=None removes=None stdin=None
  ```

- On the container, view the `ansible.log` file.

  ```
  /var/log/ansible.log
  ```

- The default log location for ESS 3000 commands is: `/var/log/ess/esslog_`*`timestamp`*`/`

  Use this location to debug details of the various python based commands running under Ansible control.

- To debug OS or package upgrades, you can view the DNF log.

  ```
  /var/log/dnf.log
  ```

- If you add `-v` to any **`ess3krun`** command, you can see the verbose output. This might be helpful, additional debug information.

Customizing file system parameters

If you want to customize the file system parameters from the defaults, do the following steps from within the container before running the **`ess3krun filesystem`** command:

1. Open the `/opt/ibm/ess/tools/ansible/vars.yml` file.

   ```
   vim /opt/ibm/ess/tools/ansible/vars.yml
   ```

2. Edit these values as required.

   ```
   Node_Class: "ess_x86_64_mmvdisk"
   Recovery_Group: "ess3k"
   Code: "8+2p"
   Block_Size: "4M"
   Size: "80%"
   Mount_Point: "/gpfs"
   ```

   Note: You must use a **`Size`** value of lower than or equal to 80%.

3. Save the file and quit.

   ```
   :wq
   ```

Configuring the GPFS pagepool size to the 60% target

Identify the node class name to use and list the current pagepool settings by running the following commands from either one of the canister nodes.

1. List the node classes and identify the node class name associated with the system going through MES.

   ```
   # mmvdisk nc list

   node class            recovery group
   --------------------  ---------------
   ess_x86_64_mmvdisk    ess3k
   ess_x86_64_mmvdisk_5  ess3k5
   gssio1_ibgssio2_ib    -
   ```

2. View the current pagepool configuration.

   ```
   # mmvdisk server list --nc ess_x86_64_mmvdisk_5 --config

   node
   number  server                            active   memory    pagepool  nsdRAIDTracks
   ------  --------------------------------  -------  --------  --------  -------------
       21  ess3k5a-ib.example.net               no        754 GiB    75 GiB         131072
       22  ess3k5b-ib.example.net               no        754 GiB    75 GiB         131072
   ```

   Here, the pagepool percentage is less than 25% of the physical memory.

   To change the pagepool percentage, GPFS must be running.

3. Restart GPFS.

   ```
   # mmstartup -N ess_x86_64_mmvdisk_5
   Wed Feb 19 16:37:02 EST 2020: mmstartup: Starting GPFS ...
   ```

4. Change the pagepool percentage to 60%.

   In these example steps, 60% of the 754 GiB physical memory is roughly 460 GiB.

   ```
   mmchconfig pagepool=460G -N ess_x86_64_mmvdisk_5
   ```

Turning on syslog redirection

Use these steps to redirect the `/var/log/messages` file on each canister node to the EMS node. Doing this allows you to access logs from a centralized location to debug any issues that might occur.

1. Log in to each canister node.
2. Edit the `/etc/rsyslogd.conf` file to add the IP address of the EMS node at the bottom of the file.

   For example:

```
*.* @192.168.20.1
```

Where 192.168.20.1 is the IP of the EMS node (bridge IP address).

3. Save the file and restart **rsyslogd**.

```
systemctl restart rsyslog
```

Restoring the backup files and SSH keys

Note:

- For the following command example, it is assumed that the backup location is `/home/backup/6001/xcatdb`.

```
/tmp/cems_restore.sh /home/backup/6001/xcatdb
cp -a /home/backup/6001/hostkeys /etc/xcat/hostkeys
```

Helpful podman commands

- List installed images:

```
podman images
```

- List containers:

```
podman ps -a
```

- Stop container:

```
podman stop  ContainerName
```

- Remove container:

```
podman rm ContainerName
```

- Remove image:

```
podman image rm ContainerName -f
```

- Re-create network bridge:

```
From within ESS3000 extracted directory run ./essmgr -n
```

- Re-run container:

```
From within ESS3000 extracted directory run ./essmgr -r
```

- Re-attach to running container:

```
podman attach ContainerName
```

- Start a container:

```
podman start ContainerName
```

- Exit container without stopping it:

```
Ctrl +p then Ctrl + q
```

- Enter container quietly:

```
podman exec -it ContainerName /bin/bash
```

Chapter 7. 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 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 which 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 is 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 8. Recovery Group Issues

An ESS 3000 recovery group has a different structure from the recovery groups in ESS version 5.3.5.

The recovery groups in ESS 5.3.5 are called `paired recovery groups` and always come in pairs, dividing ownership of the enclosure disks in half, with one recovery group primary to each of the two servers in the ESS building block. An ESS 3000 building block contains two canister servers and an NVMe enclosure, and configures as a single recovery group that is simultaneously active on both canister servers. An ESS 3000 recovery group is called a `shared recovery group` because the enclosure disks are shared by both the canister servers in the building block. The single shared recovery group structure is necessitated because the ESS 3000 can have as few as 12 disks, which is the smallest number of disks a recovery group can contain. having 12 disks allows for one equivalent spare and 11-wide 8+3P RAID codes. In contrast, ESS 5.3.5 building blocks always contain a minimum of 24 disks, which can therefore be divided into two paired recovery groups of at least 12 disks.

The following example displays a server pair of a representative ESS 5.3.5 building block, that is using the individual building block node class ESS:

```
 # mmvdisk server list --node-class ESS
 node
number  server                            active   remarks
------  --------------------------------  -------  -------
     1  server1.gpfs.net                  yes      serving ESSRG1
     2  server2.gpfs.net                  yes      serving ESSRG2
#
```

Server workload within the building block is balanced by each server that is serving one of the two paired recovery groups. The following example displays a canister server pair of a representative ESS 3000 building block, that is using the individual building block node class `ESS3000`:

```
 # mmvdisk server list --node-class ESS3000
 node
number  server                            active   remarks
------  --------------------------------  -------  -------
     3  canister1.gpfs.net                yes      serving ESS3000RG: LG002, LG004
     4  canister2.gpfs.net                yes      serving ESS3000RG: root, LG001, LG003
```

In the case of ESS 3000, each server is simultaneously serving the same single recovery group, `ESS3000RG`.The server workload within the building block is balanced by subdividing the single shared recovery group into the following log groups: `LG001`, `LG002`, `LG003`, `LG004`, and the lightweight root or master log group. The non-root log groups are called `user log groups`. Only the user log groups contain the file system vdisk NSDs.

All recovery groups in a cluster can be listed by using the **`mmvdisk recoverygroup`** list command:

```
# mmvdisk recoverygroup list
                                                            needs    user
recovery group  active   current or master server          service  vdisks  remarks
--------------  -------  --------------------------------  -------  ------  -------
ESS3000RG       yes      canister2.gpfs.net                no           16
ESSRG1          yes      server1.gpfs.net                  no            8
ESSRG2          yes      server2.gpfs.net                  no            8
```

The `needs service` column in all the IBM Spectrum Scale RAID commands is narrowly defined to mean whether a disk in the recovery group is called out for replacement. The **`mmvdisk recoverygroup list --not-ok`** command can be used to show other recovery group issues, including those involving log groups or servers:

```
# mmvdisk recoverygroup list --not-ok
recovery group  remarks
--------------  -------
ESS3000RG       server canister2.gpfs.net 'down'
#
```

If one server of an ESS 3000 shared recovery group is down, all the log groups must failover to the remaining server:

```
 # mmvdisk recoverygroup list --server --recovery-group ESS3000RG
 node
number  server                            active   remarks
------  --------------------------------  -------  -------
     3  canister1.gpfs.net                yes      serving ESS3000RG: root, LG001, LG002,
LG003, LG004
     4  canister2.gpfs.net                no       configured
```

When the down server is brought back up, the Recovery Group Configuration Manager (RGCM) process that is running on the cluster manager node assigns it two of the user log groups to rebalance the recovery group server workload.

Other than cases where there is a failover or while servers are rejoining a recovery group, RGCM must always keep two user log groups on each server. In the unlikely event that both servers are active but each server does not have two user log groups, you can shut down one of the servers and restart it. Shutting down the servers and restarting them causes the RGCM to redistribute the user log groups to the servers.

For example, consider a situation where the following allocation of log groups lasts for five or more minutes:

```
 # mmvdisk recoverygroup list --server --recovery-group ESS3000RG
 node
number  server                            active   remarks
------  --------------------------------  -------  -------
     3  canister1.gpfs.net                yes      serving ESS3000RG: root, LG001, LG002, LG003
     4  canister2.gpfs.net                yes      serving ESS3000RG: LG004
```

In such cases, shutting down `canister2` and starting it back up restores the log group workload balance in the building block within five or fewer minutes:

```
# mmshutdown -N canister2.gpfs.net
# mmstartup -N canister2.gpfs.net
# sleep 300
# mmvdisk recoverygroup list --server --recovery-group ESS3000RG
 node
number  server                            active   remarks
------  --------------------------------  -------  -------
     3  canister1.gpfs.net                yes      serving ESS3000RG: root, LG002, LG003
     4  canister2.gpfs.net                yes      serving ESS3000RG: LG001, LG003
```

Chapter 9. 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:

   ```
   esssnap -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. File system attributes for all of the failing file systems, issue:

```
mmlsfs Device
```

g. 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 36.
   - For file system corruption or MMFS_FSSTRUCT errors, see "Additional information to collect for file system corruption or MMFS_FSSTRUCT errors" on page 36.
   - For GPFS daemon crashes, see "Additional information to collect for GPFS daemon crashes" on page 37.

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 35.
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 35.
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 35.
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 may 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 35.
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 may 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 35.

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

Chapter 10. 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 3000 GUI to perform various maintenance tasks. The ESS 3000 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 3000 GUI includes various maintenance procedures to guide you through the fix process.

Updating the firmware for host adapters, enclosures, and drives

After you create a GPFS cluster, you can install the most current 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. Then, address issues that occur because ESS 3000 is not upgraded to a later version.

You can update the firmware either manually or with the help of directed maintenance procedures (DMP) that are available in the GUI. The ESS 3000 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 *Elastic Storage System: Problem Determination Guide*.

The most current firmware is packaged as the `gpfs.ess.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 Elastic Storage System(ESS)`, and click the right arrow.
   b. On the **Installed Version** menu, select: `6.0.0`
   c. On the **Platform** menu, select: `Linux 64-bit,x 86_64`
   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.ess.firmware` RPM needs to be installed on all ESS 3000 server nodes. It contains the most current updates of the following types of supported firmware for a ESS 3000 configuration:

- Host adapter firmware
- Enclosure firmware

- Drive firmware
- Firmware loading tools.

For command syntax and examples, see *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 the disk drops below 66% of the other disks for an extended period, the disk will be declared `slow`.

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 41 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

Each of the two canister servers of an ESS 3000 shared recovery group is capable of serving the entire recovery group if the other canister is not available. When only one canister server is available, all of the log groups are served by the remaining server. When an unavailable server becomes active again, it takes back two of the user log groups from the other server.

During a normal operation both the ESS 3000 servers are active, and each serves two of the user log groups:

```
# mmvdisk recoverygroup list --recovery-group ESS3000RG --server
 node
number  server                            active   remarks
------  -------------------------------   -------  -------
     3  canister1.gpfs.net                yes      serving ESS3000RG: LG001, LG003
     4  canister2.gpfs.net                yes      serving ESS3000RG: root, LG002, LG004
```

If canister2 fails or is shutdown, its two user log groups transparently switch to being served by canister1. The root log group also fails over if it is located on canister2. Application workload to the affected log groups is paused while the log groups are recovered on canister1, but are not otherwise affected.

When an ESS 3000 recovery group is operating with server failover, all the log groups are located on one server, and the recovery group is reported as `not OK`:

```
# mmvdisk recoverygroup list --recovery-group ESS3000RG --server
 node
number  server                            active   remarks
------  -------------------------------   -------  -------
     3  canister1.gpfs.net                yes      serving ESS3000RG: root, LG001, LG002,
LG003, LG004
     4  canister2.gpfs.net                no       configured
# mmvdisk rg list --not-ok
recovery group  remarks
--------------  -------
ESS3000RG       server ccanister2.gpfs.net 'down'
```

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 3000 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 continue 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.

When several disks fail, the system continues to operate even if there is no more spare space. 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 performed by using the **`mmvdisk pdisk replace`** command with the `--prepare` option for ESS 3000 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, the **`mmvdisk pdisk replace`** command powers on the replacement disk and integrates it into the ESS 3000 recovery group.

You can replace the disk either manually or with the help of directed maintenance procedures (DMP) that are available in the GUI. The ESS 3000 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 launch the `replace disk` DMP in the GUI. For more information, see *Replace disks* in *Elastic Storage System: Problem Determination Guide*.

Replacing failed disks in an ESS 3000 recovery group: a sample scenario

This scenario shows how to detect and replace failed disks in a recovery group that is built on an ESS 3000 building block.

Detecting failed disks in your ESS 3000 enclosure

The recovery group contains one declustered array DA1 containing log home and user data VDisk.

The data declustered array is defined as follows:

- 24 pdisks per data declustered array
- Default disk replacement threshold value set to two

The replacement threshold of two means that IBM Spectrum Scale RAID requires disk replacement only 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 **`mmvdisk recoverygroup list`** for the recovery groups, which are shown here for RG1:

```
# mmvdisk recoverygroup list --recovery-group rg1 --declustered-array --vdisk
declustered needs vdisks pdisks replace capacity
array service type user log total spare threshold total raw free raw background task
----------- ------- ---- ---- --- ----- ----- --------- --------- -------- ---------------
DA1 no NVMe 8 5 24 2 2 76 TiB 45 TiB scrub 14d (9%)
mmvdisk: Total capacity is the raw space before any vdisk set definitions.
mmvdisk: Free capacity is what remains for additional vdisk set definitions.
                       declustered array                                     block size and
vdisk                  and  log group   activity   capacity  RAID code       checksum   granularity remarks
------------------  ------- ----------- --------  --------  ---------------  ---------  ---------   -------
RG001LG001LOGHOME       DA1    LG001    normal    4096 MiB  4WayReplication   2 MiB      4096       log home
RG001LG002LOGHOME       DA1    LG002    normal    4096 MiB  4WayReplication   2 MiB      4096       log home
RG001LG003LOGHOME       DA1    LG003    normal    4096 MiB  4WayReplication   2 MiB      4096       log home
RG001LG004LOGHOME       DA1    LG004    normal    4096 MiB  4WayReplication   2 MiB      4096       log home
RG001ROOTLOGHOME        DA1    root     normal    4096 MiB  4WayReplication   2 MiB      4096       log home
RG001LG001VS001         DA1    LG001    normal    4235 GiB  8+3p              8 MiB      32 KiB
RG001LG001VS002         A1     LG001    normal     481 GiB  4WayReplication   1 MiB      8192
RG001LG002VS001         DA1    LG002    normal    4235 GiB  8+3p              8 MiB      32 KiB
RG001LG002VS002         DA1    LG002    normal     481 GiB  4WayReplication   1 MiB      8192
RG001LG003VS001         DA1    LG003    normal    4235 GiB  8+3p              8 MiB      32 KiB
RG001LG003VS002         DA1    LG003    normal     481 GiB  4WayReplication   1 MiB      8192
RG001LG004VS001         DA1    LG004    normal    4235 GiB  8+3p              8 MiB      32 KiB
RG001LG004VS002         DA1    LG004    normal     481 GiB  4WayReplication   1 MiB      8192
```

The indication that disk replacement is called for in this recovery group is the value of no 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. This just indicates that data from a failed disk is being rebuilt onto the spare space. Only if the replacement threshold is met, the disks are marked for replacement and the declustered array are flagged as needing service.

IBM Spectrum Scale RAID provides the following 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 can be run from the CLI on any IBM Spectrum Scale RAID cluster node. Consider the following example:
  1. **mmvdisk recoverygroup list --rg** with the --declusterd-array flag shows yes in the needs service column.
  2. **mmvdisk recoverygroup list --rg** and the --pdisk flags shows the states of all pdisks, which might be examined for the replace pdisk state.
  3. **mmvdisk pdisk list** with the --replace flag, which lists only those pdisks that are marked for replacement.

Note: Because the output of **mmvdisk recoverygroup list --rg rg1 --pdisk** is long, this example shows only some of the disks, but includes the disks that are marked for replacement:

```
# mmvdisk recoverygroup list --rg rg1 --pdisk
             declustered     paths                           AU
pdisk        array       active total capacity free space log size state
------------ ----------- ------ ----- -------- ---------- -------- -----
e1s01         DA1         2      2    3576 GiB  2264 GiB   256 MiB  ok
e1s02         DA1         0      0    3576 GiB  2334 GiB   256 MiB  simulatedDead/draining/
replace
e1s03         DA1         2      2    3576 GiB  2266 GiB   256 MiB  ok
e1s04         DA1         2      2    3576 GiB  2262 GiB   256 MiB  ok
e1s05         DA1         2      2    3576 GiB  2262 GiB   256 MiB  ok
e1s06         DA1         2      2    3576 GiB  2264 GiB   256 MiB  ok
e1s07         DA1         2      2    3576 GiB  2264 GiB   256 MiB  ok
e1s08         DA1         2      2    3576 GiB  2264 GiB   256 MiB  ok
e1s09         DA1         2      2    3576 GiB  2264 GiB   256 MiB  ok
e1s10         DA1         2      2    3576 GiB  2264 GiB   256 MiB  ok
e1s11         DA1         2      2    3576 GiB  2264 GiB   256 MiB  ok
e1s12         DA1         0      0    3576 GiB  2318 GiB   256 MiB  simulatedDead/draining/
replace
e1s13         DA1         2      2    3576 GiB  2264 GiB   256 MiB  ok
e1s14         DA1         2      2    3576 GiB  2264 GiB   256 MiB  ok
e1s15         DA1         2      2    3576 GiB  2264 GiB   256 MiB  ok
e1s16         DA1         2      2    3576 GiB  2266 GiB   256 MiB  ok
e1s17         DA1         2      2    3576 GiB  2264 GiB   256 MiB  ok
e1s18         DA1         2      2    3576 GiB  2262 GiB   256 MiB  ok
e1s19         DA1         2      2    3576 GiB  2264 GiB   256 MiB  ok
e1s20         DA1         2      2    3576 GiB  2264 GiB   256 MiB  ok
e1s21         DA1         2      2    3576 GiB  2266 GiB   256 MiB  ok
e1s22         DA1         2      2    3576 GiB  2264 GiB   256 MiB  ok
e1s23         DA1         2      2    3576 GiB  2266 GiB   256 MiB  ok
e1s24         DA1         2      2    3576 GiB  2262 GiB   256 MiB  ok
```

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

- e1s02 in DA1
- e1s12 in DA1

The naming convention that is used during recovery group creation indicates that these disks are in Enclosure 1 Slot 2 and Enclosure 1 Slot 12. To confirm the physical locations of the failed disks, use the **mmvdisk pdisk list** command to list information about the pdisks in declustered array DA1 of recovery group Brg1 that are marked for replacement:

```
# mmvdisk pdisk list --recovery-group rg1 --replace
recovery group pdisk priority FRU (type) location
-------------- ------------ -------- --------------- --------
rg1 e1s02 14.23 3.84TB NVMe G3 Enclosure 5141-AF8-78E00KW Drive 2
rg1 e1s12 14.23 3.84TB NVMe G3 Enclosure 5141-AF8-78E00KW Drive 12
mmvdisk: A lower priority value means a higher need for replacement.
```

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 e1s02 78E00KW-2 Slot 2
--------------------------------------------------------------------------------
pdisk e1s12 78E00KW-12 Slot 12
--------------------------------------------------------------------------------
```

This example shows how the component database provides an easier-to-use location reference for the affected physical disks. The pdisk name `e1s02` means `Enclosure 1 Slot 2`. Additionally, the location provides the serial number of enclosure 1, `78E00KW`, with the slot number. But the user location that is defined in the component database can be used to precisely locate the disk in an equipment rack and a named disk enclosure. There is no external enclosure for an ESS 3000 system. All of the NVMe devices are in the canisters.

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

```
mmlscomp --serial-number 78E00KW
Storage Enclosure Components
Comp ID Part Number Serial Number Name Display ID
------- ----------- ------------- ---------------- ----------
3 5141-AF8 78E00KW 5141-AF8-78E00KW
```

Replacing failed disks in a recovery group

Note: In this example, it is assumed that two new disks with the appropriate Field Replaceable Unit (FRU) code are obtained as replacements for the failed pdisks `e1s02` and `e1s12`. In this case, the FRU attribute of the FRU is `3.84TB NVMe G3`.

Replacing each disk is a three-step process:

1. Use the **`mmvdisk pdisk replace`** command with the `--prepare` flag to inform IBM Spectrum Scale to locate the disk, suspend it, and allow it to be removed.
2. Locate and remove the failed disk and replace it with a new one.
3. Use the **`mmvdisk pdisk replace`** command to use the new disk.

IBM Spectrum Scale RAID assigns a priority to the pdisk replacement. Disks with smaller values for the `replacementPriority` attribute must be replaced first. In this example, the only failed disks are in `DA1` and both have the same `replacementPriority` value. Disk `e1s02` is chosen to be replaced first.

1. Release the pdisk `e1s02` in recovery group `rg1` by using the following command:

   ```
   # mmvdisk pdisk replace --prepare --recovery-group rg1 --pdisk e1s02
   2.
   3.
   # mmvdisk pdisk replace --recovery-group rg1 --pdisk e1s02
   [I] The following pdisks will be formatted on node c202f06fs03a:
   /dev/nvme11n1
   [I] Pdisk e1s02 of RG rg1 successfully replaced.[I]
   Resuming pdisk e1s02#026 of RG rg1.
   [I] Carrier resumed.
   mmvdisk: Suspending pdisk e1s02 of RG rg1 in location 78E00KW-2.
   mmvdisk: Location 78E00KW-2 is Enclosure 5141-AF8-78E00KW Drive 2.
   mmv di sk: Carrier released.
   mmvdisk:
   mmvdisk: - Remove carrier.
   mmvdisk: - Replace disk in location 78E00KW-2 with type '3.84TB NVMe G3 '.
   mmvdisk: - Reinsert carrier.
   mmvdisk: - Issue the following command:
   mmvdisk:
   mmvdisk: mmvdisk pdisk replace --recovery-group rg1 --pdisk 'e1s02'
   ```

2. Unlatch and pull the handle for the failed disk in slot 2. Slide out the failed disk and set it aside.

   Note: The amber LED is turned on for the failed disk. In this example, the failed disk is the disk in slot 2. The drive LEDs turn off when the slot is empty.

3. Insert the new disk with FRU `3.84TB NVMe G3` in place, push its handle forward, and latch it.

4. Finish the replacement of pdisk e1s02, by using the following command:

```
# mmvdisk pdisk replace --recovery-group rg1 --pdisk e1s02
[I] The following pdisks will be formatted on node c202f06fs03a:
/dev/nvme11n1
[I] Pdisk e1s02 of RG rg1 successfully replaced.[I] Resuming pdisk e1s02#026 of RG rg1.
[I] Carrier resumed.
```

When the **mmvdisk pdisk replace** command returns successfully, IBM Spectrum Scale RAID begins rebuilding and re balancing the IBM Spectrum Scale RAID strips onto the new disk, which assumes the pdisk name e1s02. The failed pdisk might remain in a temporary form, until all data from it rebuilds, at which point it is deleted. The temporary form is indicated in this example by the name e1s02#026. Only one block device name is mentioned as being formatted as a pdisk; the second path is discovered in the background.

Disk e1s12 is still marked for replacement, and DA1 of rg1 still needs 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.

To replace pdisk e1s12 following these steps:

1. Release pdisk e1s12 in recovery group rg1:

```
# mmvdisk pdisk replace --prepare --recovery-group rg1 --pdisk e1s12
[I] Suspending pdisk e1s12 of RG rg1 in location 78E00KW-12. [I] Location 78E00KW-12 is
Enclosure 5141-AF8-78E00KW Drive 12 [I] Carrier released.


[II] Remove carrier.
[III] Replace disk in location 78E00KW-12 with type '3.84TB NVMe G3 '.
[IV] Reinsert carrier.
[V] Issue the following command:


mmvdisk pdisk replace --recovery-group rg1 --pdisk     'e1s12'
```

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. Finish the replacement of pdisk e1s12, run the following command:

```
# mmvdisk pdisk replace --recovery-group rg1 --pdisk e1s12
[I] The following pdisks will be formatted on node c202f06fs03a:
/dev/nvme0n1
[I] Pdisk e1s12 of RG rg1 successfully replaced.[I] Resuming pdisk e1s12#029 of RG rg1.
[I] Carrier resumed.
```

The disk replacements can be confirmed by using the **mmvdisk recoverygroup list --rg rg1 --pdisk** command:

```

# mmvdisk recoverygroup list --rg rg1 --pdisk    --declustered-array

declustered    needs        vdisks              pdisks    replace                capacity
   array    service    type    user log    total spare    threshold    total raw free raw
background
task
-----------    -------    ----    ---- ---    ----- -----    ---------    --------- --------
---------------
DA1               no    NVMe    4    5    24    2        2              76 TiB   786 GiB
scrub 14d
(0%)

mmvdisk: Total capacity is the raw space before any vdisk set definitions. mmvdisk: Free
capacity is what remains for additional vdisk set definitions.

       declustered        paths                                               AU
pdisk    array        active    total    capacity    free space    log size    state
------     ------       --------    -------      ----------     -----------     ----------
------
e1s01     DA1              2        2     3576 GiB       342 GiB    256 MiB     ok
```

```
e1s02     DA1              2        2     3576 GiB       342 GiB    256 MiB    ok
e1s02#026    DA1              0        0     3576 GiB       342 GiB    256 MiB    simulatedDead/
deleting/draining/01008.6c0
e1s03     DA1              2        2     3576 GiB       344 GiB    256 MiB    ok
e1s04     DA1              2        2     3576 GiB       340 GiB    256 MiB    ok
e1s05     DA1              2        2     3576 GiB       342 GiB    256 MiB    ok
e1s06     DA1              2        2     3576 GiB       344 GiB    256 MiB    ok
e1s07     DA1              2        2     3576 GiB       342 GiB    256 MiB    ok
e1s08     DA1              2        2     3576 GiB       340 GiB    256 MiB    ok
e1s09     DA1              2        2     3576 GiB       338 GiB    256 MiB    ok
e1s10     DA1              2        2     3576 GiB       344 GiB    256 MiB    ok
e1s11     DA1              2        2     3576 GiB       340 GiB    256 MiB    ok
e1s12     DA1              2        2     3576 GiB       340 GiB    256 MiB    ok
e1s12#029    DA1              0        0     3576 GiB       342 GiB    256 MiB    simulatedDead/
deleting/draining/01008.6c0
e1s13     DA1              2        2     3576 GiB       344 GiB    256 MiB    ok
e1s14     DA1              2        2     3576 GiB       342 GiB    256 MiB    ok
e1s15     DA1              2        2     3576 GiB       344 GiB    256 MiB    ok
e1s16     DA1              2        2     3576 GiB       344 GiB    256 MiB    ok
e1s17     DA1              2        2     3576 GiB       340 GiB    256 MiB    ok
e1s18     DA1              2        2     3576 GiB       342 GiB    256 MiB    ok
e1s19     DA1              2        2     3576 GiB       342 GiB    256 MiB    ok
e1s20     DA1              2        2     3576 GiB       344 GiB    256 MiB    ok
e1s21     DA1              2        2     3576 GiB       340 GiB    256 MiB    ok
e1s22     DA1              2        2     3576 GiB       344 GiB    256 MiB    ok
e1s23     DA1              2        2     3576 GiB       342 GiB    256 MiB    ok
e1s24     DA1              2        2     3576 GiB       342 GiB    256 MiB    ok
```

Physical disks are counted toward the total number of pdisks in the recovery group `rg1` 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 becomes 24 again.

Using the mmvdisk command to fix issues caused by improper disk removal

Pdisks are identified by the descriptors that are 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 that is found in an unexpected location avoids risk of data unavailability.

The location code that is 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 be blank forever. 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 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 `e1s03` from one slot over, inserts it into slot 2, then runs the **`mmvdisk pdisk replace --recovery-group rg1 --pdisk e1s02`** command. The replace command detects what happened and fails, and displays the following error message:

```
[E] Pdisk e1s03 of recovery group rg1 in location 78E00KW-2 cannot be used as a
replacement for pdisk e1s03 of recovery group rg1.
```

But because e1s03 now occupies the slot, it has taken on the location code 78E00KW-2, clearing it from pdisk e2s046. The system no longer knows the location e1s03; it just knows that the location is not 78E00KW-2. Even, if the admin realizes the mistake and moves e1s03 back into slot 3, e1s03's location is updated to slot 3, but e1s03'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.
- dd or other tools are not used 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 by using the **mmvdisk pdisk change** command. To pull a drive that is in the wrong slot, use the **mmvdisk pdisk change --recovery-group RGNAME --pdisk PDNAME --suspend** command to quiesce the disk before you pull it. Run the **mmvdisk pdisk change --recovery-group RGNAME --pdisk PDNAME --resume** command after you reinsert the disk. Suspending the disk before you pull 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/vdisk/samples` as shown:

1. Insert a new, blank disk into the empty slot 2 where the bad e1s02 drive was.
2. Run replace-at-location `rg1 e1s02 78E00KW-2`.

Other hardware service

Other hardware components of the ESS 3000 such as boot drives, fans, and power supplies can be serviced by IBM authorized service personnel only. IBM service support representatives and lab based services personnel can access service information through the Service Guide located in the IBM Knowledge Center.

Note: An IBM intranet connection is required.

The status of many ESS 3000 components can be examined by using the **mmlsenclosure** command.

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 4. DMPs

DMP	Event ID
Start NSD	disk_down
Start GPFS daemon	gpfs_down
Increase fileset space	inode_error_high and inode_warn_high
Synchronize Node Clocks	time_not_in_sync

Table 4. DMPs (continued)

DMP	Event ID
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 singe 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 singe 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 5. 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.
			`sensors={`
			`name = "sensorName"`
			`period = period`
			`proxyCmd = "/opt/IBM/zimon/`
			`GaneshaProxy"`
			`restrict = "cesNodes"`
			`type = "Generic"`
			`}`

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 6. 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.
			`sensors={`
			`     name = "sensorName"`
			`     period = period`
			`     restrict = "cesNodes"`
			`     type = "Generic"`
			`}`

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 time because it is configured to mount automatically but the file system is currently not mounted on all nodes.

- **Solution:** Mount the file system one 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 are 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 disappear 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.

Maintenance procedures for NVMe and PCIe issues

This section details the maintenance procedures for NVMe and PCIe issues.

NVMe drive listing is not verified

Follow these steps to verify that the expected number of NVMe drives are listed:

1. Run the **`nvme list`** Linux command to query NVMe drives.
2. Verify that the expected number of drives is reported.

NVMe drives are missing from one or both I/O nodes

Follow these steps if the NVMe listing is done, but the listing displays no drives:

1. Validate that the PERST service, `systemctl status ess3k_perst.service`, is enabled and has run after boot.
2. If the PERST service is not enabled or does not exist, then reinstall the `gpfs.ess.platform.ess3k` rpm.

PCIe initialization settings are not validated

Various PCIe-related settings like error-reporting settings are set by `ess3k_initpcie.service`. Follow these steps to validate that the PCIe initialization settings are enabled:

1. Validate that the `systemctl status ess3k_initpcie.service` service is enabled and has run after boot.
2. If the service is not enabled or does not exist, then reinstall the `gpfs.ess.platform.ess3k` rpm.

Unexpected kernel crashes due to PCIe or NVMe activities:

PCIe or NVMe activities like reset, power off, power on, and so on might cause unexpected kernel crash if the system is not set up correctly. If NVMe drives encounter PCIe fabric-related errors or resets, those events produce a fabric error interrupt, that must be handled by the PCIe fabric. However, if the fabric-handling infrastructure does not exist, it might result in a kernel crash and reboot. To prevent such issues, verify that the Linux native PCIe interrupt handler is enabled. For more information, see “Linux native PCIe interrupt handler validation and enablement” on page 57.

Downstream port containment (DPC) bits are not clearing

ESS 3000 I/O nodes are DPC-enabled to provide isolation and containment of the PCIe-related issues for the NVMe drive endpoints. When an NVMe drive is removed or powered off, the PCIe fabric handles the event by performing a DPC. If the NVMe drive is reinserted or the slot is powered back on, and the NVMe drive does not show up again, it might be because the Linux native PCIe interrupt handler is not enabled. For more information, see “Linux native PCIe interrupt handler validation and enablement” on page 57.

Linux native PCIe interrupt handler validation and enablement

For the ESS 3000 I/O nodes, the native PCIe interrupt handler is enabled during the manufacturing phase and validated during the deployment phase.

However, if for some reason the enablement was removed, this section helps determine how to validate and enable it again.

1. To validate the PCIe native error handler, run the following query:

   ```
   cat /proc/cmdline | grep pcie_ports=native
   ```

   If the query comes back empty, then the PCIe native error handler must be enabled:
2. To enable the PCIe native error handler, follow these steps:
   a. Open the `/etc/default/grub` file for editing.
   b. Find the `GRUB_CMDLINE_LINUX` line.
   c. Append the text `pcie_ports=native` to the end of the `GRUB_CMDLINE_LINUX` line as shown:

   ```
   GRUB_CMDLINE_LINUX="nvme.sgl_threshold=0 sshd=1 noht crashkernel=auto resume=UUID=f0cccb47-da43-404d-
   a8f3-578129d3b8f7 rd.md.uuid=53d2b2a3:0c7532dd:72ba276b:179d3b74
   rd.md.uuid=519c1d9a:68fa26be:755637c7:9db5d8e4 rhgb quiet pcie_ports=native"
   ```

   d. Save and close the file.
   e. Make a new configuration with the updated grub file by running the following command:

```
grub2-mkconfig -o /boot/efi/EFI/redhat/grub.cfg
```

f. Reboot the server node.

g. When the server is back up, validate that the handler is enabled by running the following query:

```
cat /proc/cmdline | grep pcie_ports=native
```

PCIe-related data collection and debug

This section details the PCIe-related data collection and debug processes that can be done live on a node.

You can get more information about the active issues on the ESS 3000 for both the NVMe drive availability and the PCIe-related issues. Follow these steps to determine the possible steps towards resolving these issues:

1. Run the following script to show the NVMe-related PCIe fabric:

   ```
   lspci -tv |sed -n '/ +-\[0000:\(85\|3a\)\]/,/8546/p'
   ```

2. Run the following script to show the PCIe device link status for the NVMe drives:

   ```
   for u in 87 3c; do for i in $(seq 0 11); do d=$(printf "%02x" $i); lspci -vvs $u:$d.0; done;
   done | grep -E "^[0-9a-f]|LnkSta:|Bus:" | sed "/^[0-9a-f]/{s/ .*//;N;s/, sec-
   latency.*//;N;s/, TrErr.*//;s/\n//g;}"
   ```

3. Run the following script to show the Downstream Port Containment (DPC) status for the NVMe drives:

   ```
   for u in 87 3c; do for i in $(seq 0 11); do d=$(printf "%02x" $i); echo -n "$u:$d.0: ";
   lw1="0x"$(setpci -s $u:$d.0 0x1b4.l); lw2="0x"$(setpci -s $u:$d.0 0x1b8.l); echo "$lw1
   $lw2";done; done
   ```

   Note: If DPC is enabled for a particular PCIe port, observe a nonzero value in the rightmost column.

Chapter 11. References

The IBM Elastic Storage System system 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.

The recorded events can also be displayed through GUI.

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 7. 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_needsservice	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 **`mmlsrecoverygroup`** 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 8. Events for the Enclosure component

Event	Event Type	Severity	Message	Description	Cause	User Action
adapter_bios_notavail	STATE_CHANGE	WARNING	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 **mmlsfirmware** 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	WARNING	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 **mmlsfirmware** command.
adapter_firmware_notavail	STATE_CHANGE	WARNING	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 **mmlsfirmware** 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	WARNING	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 **mmlsfirmware** 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	WARNING	currentSensor {0} is degraded.	The currentSensor state is degraded.	N/A	N/A
dcm_drawer_open	STATE_CHANGE	WARNING	DCM {0} drawer is open.	The DCM drawer is open.	N/A	N/A
dcm_failed	STATE_CHANGE	WARNING	DCM {0} is failed.	The DCM state is failed.	N/A	N/A
dcm_not_available	STATE_CHANGE	WARNING	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	WARNING	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 **mmlsfirmware** 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	WARNING	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 **mmlsfirmware** command.
enclosure_data	STATE_CHANGE	INFO	Enclosure data found.	Successfully queried the enclosure details.	The **mmlsenclosure all -L -Y** command reports enclosure data.	N/A

Table 8. Events for the Enclosure component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
enclosure_firmware_notavail	STATE_CHANGE	WARNING	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	WARNING	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	WARNING	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_needsservice	STATE_CHANGE	WARNING	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	WARNING	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	WARNING	ESM {0} is absent.	The ESM state is not installed .	N/A	N/A
esm_failed	STATE_CHANGE	WARNING	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	WARNING	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 8. Events for the Enclosure component (continued)						
Event	**Event Type**	**Severity**	**Message**	**Description**	**Cause**	**User Action**
fan_failed	STATE_CHANGE	WARNING	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	WARNING	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	WARNING	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	WARNING	Enclosure data and state information cannot be queried.	Cannot query the enclosure details. State reporting for all enclosures and canisters will be incorrect.	The **`mmlsenclosure all -L -Y`** command fails to report any enclosure data.	Run the **`mmlsenclosure`** command to check for errors. Use the **`lsmod`** command to verify that the pemsmod is loaded.
power_high_current	STATE_CHANGE	WARNING	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	WARNING	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	WARNING	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	WARNING	Power supply {0} is missing.	The power supply is missing	N/A	N/A
power_supply_failed	STATE_CHANGE	WARNING	Power supply {0} is failed.	The power supply state is failed.	N/A	N/A
power_supply_off	STATE_CHANGE	WARNING	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	WARNING	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	WARNING	Temperature sensor {0} I2C bus is failed.	The temperature sensor I2C bus has failed.	N/A	N/A
temp_high_critical	STATE_CHANGE	WARNING	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 8. Events for the Enclosure component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
temp_high_warn	STATE_CHANGE	WARNING	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	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.	N/A	N/A
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.	N/A	N/A
temp_sensor_failed	STATE_CHANGE	WARNING	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	WARNING	Voltage sensor {0} I2C bus is failed.	The voltage sensor I2C bus has failed.	N/A	N/A
voltage_high_critical	STATE_CHANGE	WARNING	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	WARNING	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	WARNING	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	WARNING	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	WARNING	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 9. 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_ENTITY	INFO	GNR vdisk {0} was found.	A GNR vdisk listed in the IBM Spectrum Scale configuration was detected.	N/A	N/A

Table 9. 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 10. Events for the physical disk component

Event	Event Type	Severity	Message	Description	Cause	User Action
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 **mmdeldisk**.	The **mmlspdisk** 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.

Table 10. Events for the physical disk component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
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 **mmlspdisk** 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 **mmlspdisk** command shows that the pdisk state contains `sedLocked`.	The drive must be unlocked to be used by GNR.
gnr_pdisk_unknown	STATE_CHANGE	WARNING	GNR pdisks are in unknown state.	The pdisk state is unknown.	N/A	N/A

Table 10. Events for the physical disk component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
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 `ssdEndurancePercentage` of GNR pdisk {0} is ok.	The `ssdEndurancePercentage` value is ok.	N/A	N/A
ssd_endurance_warn	STATE_CHANGE	WARNING	The `ssdEndurancePercentage` of GNR pdisk {0} is on a warning value.	The `ssdEndurancePercentage` value is warning.	The `ssdEndurancePercentage` 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 11. 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

Table 11. Events for the Recovery group component (continued)						
Event	**Event Type**	**Severity**	**Message**	**Description**	**Cause**	**User Action**
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
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 **mmlsrecoverygroup** 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.

Canister events

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

Table 12. Events for the Canister component						
Event	**Event Type**	**Severity**	**Message**	**Description**	**Cause**	**User Action**
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_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_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

Table 12. Events for the Canister component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
bootdrive_mirror _unconfigured	STATE_CHANG E	WARNING	The bootdrive's mirroring is unconfigured.	The bootdrive's mirroring is unconfigured.	The **tsplatformstat -a** command returns unconfigured for mirroring.	N/A
bootdrive_missing	STATE_CHANG E	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_CHANG E	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_CHANG E	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_CHANG E	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_CHANG E	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_CHANG E	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_CHANG E	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_CHANG E	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_CHANG E	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_CHANG E	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_CHANG E	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.

Table 12. Events for the Canister component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
can_temp_sensor_ok	STATE_CHANGE	INFO	Temperature sensor {0} is OK.	The temperature sensor state is OK.	N/A	N/A
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 **mmlsenclosure** command reports the canister as failed.	Check for detailed error events of canister components by using the **mmhealth** command. Inspect the output of **mmlsenclosure all -L** command for the referenced canister.
canister_ok	STATE_CHANGE	INFO	Canister {0} is OK.	The canister state is OK.	The **mmlsenclosure** 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.

Table 12. Events for the Canister component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
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
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 **/opt/ibm/gss/tools/bin/ess3kplt** command returned some speed errors.	Inspect the output of the **ess3kplt** 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 **mmlsenclosure** command reports only one canister instead of two.	Check for detailed error events of the referenced canister node by using the **mmhealth** command. Inspect the output of the **mmlsenclosure all -L** 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 **mmlsenclosure** command reports both canisters.	N/A

Messages

This topic contains explanations for IBM Spectrum Scale RAID and ESS 3000 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 3000 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 13. 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 3000 GUI messages, error messages (**(E)**) have the highest priority and informational messages (**I**) have the lowest priority.

The following table lists the ESS 3000 GUI message severity tags in order of priority:

Table 14. ESS 3000 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.
`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 70.

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:
Declustered 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 `tsdelpdisk` 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 `mmdelnsd` 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 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 System 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 `tschrecgroup` 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 tsctlenclslot, 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 **tsctlenclslot**, **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 `tsctlenclslot`, `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 **tsctlenclslot**, **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 **mmsdrfs** 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 **`mmdelpdisk`** 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 *errorNum*

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 *errorNum*

Explanation:
Recovery group configuration manager final takeover failed.

User response:
Contact IBM Support.

Accessibility features for IBM Spectrum Scale RAID

Accessibility features help users who have a disability, such as restricted mobility or limited vision, to use information technology products successfully.

Accessibility features

The following list includes the major accessibility features in IBM Spectrum Scale RAID:

- Keyboard-only operation
- Interfaces that are commonly used by screen readers
- Keys that are discernible by touch but do not activate just by touching them
- Industry-standard devices for ports and connectors
- The attachment of alternative input and output devices

IBM Knowledge Center, and its related publications, are accessibility-enabled. The accessibility features are described in IBM Knowledge Center (www.ibm.com/support/knowledgecenter).

Keyboard navigation

This product uses standard Microsoft Windows navigation keys.

IBM and accessibility

See the IBM Human Ability and Accessibility Center (www.ibm.com/able) for more information about the commitment that IBM has to accessibility.

Notices

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

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

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

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

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

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

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

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

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

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

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

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

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

IBM Corporation
Dept. 30ZA/Building 707
Mail Station P300
2455 South Road,
Poughkeepsie, NY 12601-5400
U.S.A.

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

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

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

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

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

COPYRIGHT LICENSE:

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

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

Trademarks

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

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

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

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

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

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

Glossary

This glossary provides terms and definitions for the ESS 3000 solution.

The following cross-references are used in this glossary:

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

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

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

B

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

BOOTP
See *Bootstrap Protocol (BOOTP)*.

Bootstrap Protocol (BOOTP)
A computer networking protocol 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 3000 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 System (ESS 3000)
A high-performance, GPFS NSD solution made up of one or more building blocks that runs on IBM Power Systems servers. The ESS 3000 software runs on ESS 3000 nodes - management server nodes and I/O server nodes.

ESS 3000 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 3000 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 3000
See *Elastic Storage System (ESS 3000)*.

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

ESM
See *environmental service module (ESM)*.

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

F

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

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

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

FEK
See *file encryption key (FEK)*.

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

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

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

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

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

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

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

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

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

FSP
See *flexible service processor (FSP)*.

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

G

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

GPFS portability layer
The interface module that each installation must build for its specific hardware platform and Linux distribution.

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

GSS
See *GPFS Storage Server (GSS)*.

H

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

HMC
See *Hardware Management Console (HMC)*.

I

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

independent fileset
A fileset that has its own inode space.

indirect block
A block that contains pointers to other blocks.

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

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

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

I/O server node
An ESS node that is attached to the ESS 3000 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 3000 node that hosts the ESS 3000 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 3000 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

Part Number:
Product Number: 5765-DME
5765-DAE

(1P) P/N:

GC28-3121-01