



Fujitsu Software Technical Computing Suite V4.0L20

FEFS User's Guide

J2UL-2554-01ENZ0(07)
March 2025

Preface

Purpose of This Manual

This manual describes FEFS, which is a shared file system developed by Fujitsu Limited.

The purpose of the manual is to enable system administrators to install, operate, and manage FEFS.

Intended Readers

This manual is intended for the system administrators who install, operate, and manage FEFS.

The manual assumes that readers are knowledgeable about Linux, storage general, and ETERNUS.

Organization of This Manual

This manual is organized as follows.

[Chapter 1 Overview](#)

This chapter provides an overview of FEFS and describes FEFS configurations.

[Chapter 2 Functions](#)

This chapter describes FEFS functions.

[Chapter 3 Method of Installation and Maintenance](#)

This chapter describes how to install and maintain FEFS.

[Chapter 4 Method of Operation](#)

This chapter describes how to operate FEFS.

[Appendix A Reference](#)

This appendix is the reference manual for FEFS system calls and commands.

[Appendix B Messages](#)

This appendix explains the messages output by FEFS.

[Appendix C Settings Required after FEFS Build](#)

This appendix describes the settings required after FEFS build.

[Appendix D Procedures to Recover the File System](#)

This appendix describes the procedure to recover the file system.

[Appendix E Procedures to Continue the Job Operation when File System Failure](#)

This appendix describes procedures to continue the job operation when the file system failure.

[Appendix F Materials Required for Troubleshooting](#)

This appendix describes the materials needed for troubleshooting.

[Glossary](#)

The glossary explains the main FEFS terms.

Notation Used in This Manual

Abbreviations

This manual uses the following abbreviations.

Formal name	Abbreviation
Windows(R) 8.1 Windows(R) 8.1 Pro Windows(R) 8.1 Enterprise	Windows 8.1 or Windows
Windows(R) 10 Home Windows(R) 10 Pro Windows(R) 10 Enterprise	Windows 10 or Windows
Microsoft(R) Office Excel(R) 2010 Microsoft(R) Office Excel(R) 2013 Microsoft(R) Office Excel(R) 2016	Excel
Red Hat(R) Enterprise Linux(R)	RHEL

Representation of units

The following table lists the prefixes used to represent units in this manual. Be careful about specifying them when displaying or entering commands.

Prefix	Value	Prefix	Value
K (kilo)	10^3	Ki (kibi)	2^{10}
M (mega)	10^6	Mi (mebi)	2^{20}
G (giga)	10^9	Gi (gibi)	2^{30}
T (tera)	10^{12}	Ti (tebi)	2^{40}
P (peta)	10^{15}	Pi (pebi)	2^{50}
E (exa)	10^{18}	Ei (exbi)	2^{60}

Model name notation

In this manual the computer with Fujitsu CPU A64FX is referred to as the "FX server" and FUJITSU server PRIMERGY as the "PRIMERGY server" (or simply "PRIMERGY").

Some part of the functions explained in this manual may vary depending on the support model. In the explanation of such functions, the support models are indicated with the abbreviations as follows.

[FX]: Function to support the FX server.

[PG]: Function to support the PRIMERGY server.

Prompts in command input examples

- # means that the command is executed with administrator privileges (super user).
- \$ means that the command is executed with privileges other than administrator privileges.

Symbols in this manual

This manual uses the following symbols.



The Note symbol indicates an item requiring special care. Be sure to read these items.



The See symbol indicates the written reference source of detailed information.

Information

The Information symbol indicates a reference note related to FEFS.

Export Controls

Exportation/release of this document may require necessary procedures in accordance with the regulations of your resident country and/or US export control laws.

Trademarks

Lustre is a registered trademark of Seagate Technology LLC in the United States.

Linux(R) is the registered trademark of Linus Torvalds in the U.S. and other countries.

Red Hat is a registered trademark of Red Hat, Inc. in the U.S. and other countries.

Microsoft, Windows, and Excel are registered trademarks or trademarks of Microsoft Corporation in the U.S. and/or other countries.

All other trademarks are the property of their respective owners.

Publishing Dates and Revisions

Revision	Manual code
March 2025, Version 1.7	J2UL-2554-01ENZ0(07)
September 2023, Version 1.6	J2UL-2554-01ENZ0(06)
September 2022, Version 1.5	J2UL-2554-01ENZ0(05)
March 2022, Version 1.4	J2UL-2554-01ENZ0(04)
November 2021, Version 1.3	J2UL-2554-01ENZ0(03)
August 2021, Version 1.2	J2UL-2554-01ENZ0(02)
June 2020, Version 1.1	J2UL-2554-01ENZ0(01)
February 2020, First version	J2UL-2554-01ENZ0(00)

Copyright

Copyright FUJITSU LIMITED 2020-2025

Update History

Changes	Location	Version
Added procedure to configuring periodic deletion of FEFS logs.	C.4	1.7
Added procedure to obtain the IP address of the management network for the FEFS client.	C.1 C.2	1.6
Also fixed errors.	-	
Fixed the notes of the project QUOTA function.	2.3.2	1.5
Fixed the description of the --setup option for the fefs_sync command.	A.2.1	1.4
Added a compute cluster definition so that the evict script runs even if the service state changes.	C.1 C.2	1.3
Added description of procedures to continue the job operation when file system failure. Also added references and messages for the fefs_deactivate command.	Appendix E A.2.18 B.2.16	

Changes	Location	Version
Fixed the procedure for adding file systems.	4.16	1.2
Added file transfer considerations for data management tools.	2.8	1.1
The FEFS design sheet supports the following versions of Windows:. Windows 8.1, 10	3.1.3 3.12.1	
Added description of how to specify "Step 1. LLIO SETTING section". Added descriptions for b and d in "Step 2".	3.1.3.2	
Fixed execution example of "Step 4 Checking the FEFS status".	3.1.6	
The note when applying the OS update package has been moved to "Deleting a Package" in "Chapter 3" in "Job Operation Software Administrator's Guide for Maintenance" and has been deleted.	3.13	
Fixed stripe pattern execution example.	4.3.1	
Fixed example of mounting FEFS on /mnt/efs in "Step 2 How to set limit values".	4.5.1	
Fixed execution example of "Step 7 Check the FEFS status".	4.14	
Added packages that must be installed on the client node that executes when using the data management tool.	4.20.3	
Added a node on which the fefsbackup.conf and fefsbackup_rsync.conf files must be set.	4.20.4	
Added the maximum value to the set value of the item " MULTI_PUT_MAX " in fefsbackup_rsync.conf. Fixed the setting needed of item "BACKUP_ROOT" to "Required".	4.20.4.1	
Fixed the description of the port settings, which are the advance preparation for executing data management tools. Also added description of password.	4.20.6	
Changed the following for the procedure for transferring files between file systems with the copy subcommand: - Changed the description in step 1. - Changed the display message in step 2. - Fixed the execution example in step 3. - Changed the directory path specified in the option -d in the examples in steps 1 through 4. Added or changed the following notes: - Improved a note about file attributes at the copy time. - Added a note about hard-link files. - Improved a note on symbolic link files. - Added a note about the path name in the operand path or the pathlist in -f option. - Added a note about file path considerations if the transfer process fails. - Added a note about simultaneous command execution.	4.20.8	
Fixed the display format of the display parameters "Backed-up" in the output when the request ID is specified.	4.20.9	
Fixed the description of display parameters "command" when run without arguments. Fixed the display format of the display parameters "Started" in the output when the request ID is specified.	4.20.11	
Fixed the synopsis of the fefs_mkfs command.	A.2.3	
Fixed the synopsis of the fefs_mount command.	A.2.4	
Added or changed the following for the fefsbackup command reference: - Fixed synopsis. - Added description of command execution privileges. - Changed the description of the refsbackup copy subcommand. - Fixed a note when specifying option -L. - Improved description of options -f and --file.	A.2.11	

Changes	Location	Version
- Added description of the options -u and --update. - Fixed description of option -v, --verbose. - Improved description of option --ignore_err.		
Added a system log message that is output when a file system inconsistency is detected.	B.1	
Fixed the corrective action of the message "error on LL_IOC_LMV_SETSTRIPE 'Path'(fd): Msg".	B.2.6	
Fixed the meaning and corrective action of the message "error: mkdir: missing stripe offset and count".	B.2.8	
Added the following fefsbackup command messages:: 0002, 0010, 1035, and 1036 Changed message 1019. Fixed the corrective action of message 0019, 1022, 1027.	B.2.9	
Fixed the description of the state (state) before incorporation in "Incorporating an MDT" and "Incorporating an OST "of "Step1 Checking the status before incorporating". Fixed the procedure for executing the lfs df command on all client nodes after the incorporation is complete.	D.4.2.3	
MGT was added as a target disk to detect disk error. Changed the target node that performs settings changes or resets to avoid FEFS server panic.	D.4.4	
Clarified that this is the procedure for backup files when the OST can be mounted. Added steps to incorporate and disconnect OST on the login node. Fixed backup procedures.	D.4.9	
Fixed an example of performing a full file system backup procedure.	D.4.10	
Also fixed errors and expressions.	-	

This manual shall not be copied without the permission of the publisher.
The information in this manual is subject to change without notice.

Contents

Chapter 1 Overview	1
1.1 FEFS Features	1
1.2 FEFS System Configuration	1
1.2.1 Hardware configuration	1
1.2.2 Server configurations	2
1.2.3 Network configuration	6
1.3 FEFS Software Configuration	6
1.4 Upper limits and Lower limits	7
1.5 Precautions	8
 Chapter 2 Functions	10
2.1 Striping Function	10
2.1.1 Round-robin and striping	10
2.1.2 Advantages of the striping function	10
2.1.3 OST_pool function	11
2.2 Multi-MDS Function	12
2.2.1 Remote directory	13
2.2.2 Striped directory	13
2.3 QUOTA Function	14
2.3.1 QUOTA function for user or group	15
2.3.2 Project QUOTA function	16
2.4 QoS Function	17
2.4.1 IO priority control function between clients	18
2.4.2 Fair share function between users	18
2.5 ACL (Access Control List) Function	19
2.6 Journaling Function	19
2.7 RAS Function/FEFS Status Check Function	19
2.7.1 Failover configuration	19
2.7.2 FEFS service monitoring	21
2.7.3 FEFS status check	21
2.7.4 Server failover (MGS/MDS/OSS)	22
2.7.5 LNet multirail function	22
2.7.6 LNet router	23
2.8 Data Management Tool (fefbackup command) [PG]	23
2.9 FEFS statistics visualization function (fefssv.ph script)	24
2.10 Function for Linkage with File Systems Other than FEFS	25
2.10.1 Publishing using NFS to external systems	26
2.10.2 Connecting Lustre [PG]	27
 Chapter 3 Method of Installation and Maintenance	28
3.1 Installation Flow	28
3.1.1 Designing the FEFS Configuration	29
3.1.2 Applying FEFS packages	33
3.1.3 Creating FEFS design sheets	34
3.1.3.1 Filling in a Node sheet	34
3.1.3.2 Filling in an LLIO sheet	37
3.1.3.3 Filling in a GFS sheet	39
3.1.3.4 Checking the input data	42
3.1.4 Creating configuration definition files for the FEFS setup tool	42
3.1.5 Deploying the configuration definition files for the FEFS setup tool	43
3.1.6 Building FEFS	43
3.1.7 Changing the file system permission	44
3.1.8 Settings required after build	44
3.1.9 Additional settings of the compute nodes	44
3.1.10 How to build FEFS on individual nodes	45
3.2 Settings to Enable the QoS Function	46

3.2.1 Enable the QoS Function.....	46
3.2.2 Settings of the QoS definition file.....	48
3.3 Settings to Enable the File Lock.....	52
3.4 Settings to Enable the ACL Function.....	52
3.5 Settings to Enable the user Extended Attribute.....	52
3.6 Settings to Use the Failover Function.....	52
3.7 Operation during maintenance.....	52
3.8 Rolling Update.....	53
3.9 Configuring the FEFS Statistics Visualization Function.....	55
3.10 Settings for NFS Publishing.....	55
3.11 Recovering a Node That Has Encountered an Error.....	55
3.12 How to Configure FEFS in External Networks.....	56
3.12.1 Creating FEFS design sheets for external networks.....	57
3.12.1.1 Filling in a NODE sheet.....	57
3.12.1.2 Filling in a NET sheet.....	58
3.12.1.3 Filling in a GFS sheet.....	58
3.12.1.4 Checking the input data.....	59
3.12.2 Creating configuration definition files for the FEFS setup tool.....	59
3.12.3 Deploying the configuration definition files for the FEFS setup tool.....	59
3.12.4 Configuring the FEFS in the external network.....	60
3.12.4.1 Setting the FEFS server and configuring the router.....	60
3.12.4.2 Setting the FEFS client.....	61
3.13 Precautions.....	61
Chapter 4 Method of Operation.....	63
4.1 Starting FEFS Servers and Clients.....	63
4.2 Stopping FEFS Servers and Clients.....	63
4.3 Configuring the Striping Function.....	63
4.3.1 How to configure stripes.....	63
4.3.2 How to confirm Stripe settings.....	65
4.3.3 How to configure an OST_pool.....	65
4.4 How to Use Multi-MDS.....	68
4.4.1 Creating a remote directory.....	68
4.4.2 Creating a striped directory.....	69
4.5 Configuring the QUOTA Function.....	70
4.5.1 QUOTA setting for users and groups.....	70
4.5.2 QUOTA settings for the projects.....	72
4.6 Configuring the QoS Function.....	75
4.6.1 Checking the QoS status on the FEFS client.....	75
4.6.2 Changing the QoS status on the FEFS client.....	75
4.6.3 Checking the QoS status on the MDS.....	76
4.6.4 Changing the QoS status on the MDS.....	77
4.7 Tuning the QoS Function.....	77
4.7.1 Analyzing the client node (metadata operation).....	78
4.7.2 Analyzing the client node (data operation).....	79
4.7.3 Analyzing the MDS.....	82
4.7.4 Analyzing the OSS.....	84
4.8 Repairing File System Inconsistencies.....	88
4.8.1 Stop the FEFS service.....	89
4.8.2 Repair on the MGS.....	89
4.8.3 Repair on the MDS.....	89
4.8.4 Repair on the OSS.....	89
4.8.5 Repair of FEFS.....	90
4.9 How to Configure the ACL.....	91
4.10 How to Set user Extended Attributes.....	91
4.11 Checking the FEFS Status.....	92
4.12 Failover.....	93

4.13 Adding an MDS.....	94
4.14 Adding an OSS.....	95
4.15 Adding Clients.....	96
4.16 Adding a File System.....	97
4.17 Deleting a File System.....	98
4.18 Adding Rack and BoB.....	99
4.19 Protecting Data in a Built File System.....	100
4.19.1 Procedure for protecting data in a file system.....	100
4.19.2 Procedure for canceling the protection of data in a file system.....	101
4.20 How to Use the Data Management Tool (fefsbackup Command) [PG].....	101
4.20.1 Subcommand overview.....	101
4.20.2 Transfer information management.....	102
4.20.3 Mandatory package.....	102
4.20.4 Config files.....	102
4.20.4.1 Config file details.....	102
4.20.5 Design where management information is stored.....	104
4.20.6 Advance preparation.....	104
4.20.7 Configuring the data management tool.....	105
4.20.8 copy subcommand.....	105
4.20.9 list subcommand.....	107
4.20.10 delete subcommand.....	108
4.20.11 status subcommand.....	109
4.21 JobStats Function.....	109
4.22 How to Use the FEFS Statistics Visualization Function (fefssv.ph Script).....	110
4.22.1 How to collect information.....	110
4.22.2 How to output information.....	111
4.22.3 Options and output information.....	115
4.23 Connecting Lustre [PG].....	117
4.23.1 Settings on Lustre server.....	117
4.23.2 Mounting an FEFS server from a Lustre client.....	117
4.23.3 Mounting a Lustre server from an FEFS client.....	117
Appendix A Reference.....	118
A.1 System Calls.....	118
A.2 Commands.....	120
A.2.1 fefs_sync.....	120
A.2.2 fefsconfig.....	122
A.2.3 fefs_mkfs.....	124
A.2.4 fefs_mount.....	124
A.2.5 fefssnap.....	125
A.2.6 lfs command.....	125
A.2.7 lctl command.....	132
A.2.8 fsck.ldiskfs.....	148
A.2.9 tuneefs.lustre.....	149
A.2.10 debugfs.ldiskfs.....	150
A.2.11 fefsbackup.....	150
A.2.12 fefs_ost2fid.....	153
A.2.13 find_file_ost.....	154
A.2.14 convert_fid2path.....	155
A.2.15 force_intr.....	155
A.2.16 evict_client.....	157
A.2.17 fefs_yaml2csv.....	157
A.2.18 fefs_deactivate.....	158
Appendix B Messages.....	160
B.1 Messages Output to System Log.....	160
B.2 Messages Output by Commands.....	171
B.2.1 fefs_sync.....	171

B.2.2 fefsconfig.....	174
B.2.3 fefs_mkfs.....	178
B.2.4 fefs_mount.....	180
B.2.5 fefssnap.....	182
B.2.6 lfs command.....	183
B.2.7 lctl command.....	196
B.2.8 fsck.ldiskfs.....	207
B.2.9 fefsbackup [PG].....	209
B.2.10 File identification tool common.....	216
B.2.11 find_file_ost.....	219
B.2.12 convert_fid2path.....	220
B.2.13 force_intr.....	220
B.2.14 evict_client.....	221
B.2.15 fefs_yaml2csv.....	221
B.2.16 fefs_deactivate.....	222
 Appendix C Settings Required after FEFS Build.....	224
C.1 Configuring the FEFS Script.....	224
C.2 Configuring the FEFS Script in an Environment Having Multiple System Management Nodes.....	227
C.3 Settings Required for Using ETERNUS.....	231
C.3.1 Enabling ETERNUS NRDY measures on the MDS.....	231
C.3.2 Procedure for configuring the OSS autoreboot script.....	231
C.4 Configuring periodic deletion of FEFS logs.....	235
 Appendix D Procedures to Recover the File System.....	237
D.1 Introduction.....	237
D.2 Impact.....	237
D.3 Failure Recovery Flow.....	237
D.3.1 Recovery flow when a defective block is detected.....	237
D.3.2 Recovery flow after a disk failure or file system crash.....	239
D.3.3 Trouble handling flow when both systems stop.....	242
D.4 Response Procedures.....	243
D.4.1 Checking the status for a block number where a defective block occurred.....	243
D.4.2 Disconnecting/incorporating an FEFS server.....	245
D.4.3 Preventing automatic failover.....	248
D.4.4 Changing the FEFS setting.....	248
D.4.5 Executing fsck.....	249
D.4.6 Executing lfsck.....	249
D.4.7 Partially rebuilding the file system.....	249
D.4.8 Creating a list of backup files.....	253
D.4.9 Backing up files.....	254
D.4.10 Rebuilding the file system.....	255
D.5 Impact on Access.....	256
 Appendix E Procedures to Continue the Job Operation when File System Failure.....	257
E.1 Disconnecting/Incorporating a File System.....	257
E.1.1 Disconnecting.....	257
E.1.2 Incorporating.....	257
E.2 Procedures to Delete Hanged Jobs due to the File System Failure.....	258
E.3 Procedures to Start Node in during a File System Failure.....	259
 Appendix F Materials Required for Troubleshooting.....	261
Glossary.....	262

Chapter 1 Overview

This chapter describes FEFS features, configurations, and specifications.

1.1 FEFS Features

FEFS is a large-scale and high-performance parallel distributed file system based on the technology of Lustre, open source file system. FEFS has following features:

- Large-scale
Supports clients of 100,000 nodes and 8EiB file system size.
- High performance
Enhances I/O performance by using the striping method and the round-robin method to distribute and store file data onto storage.
- Easy to use
Prevents performance impact from other users issuing numerous I/O requests through the IO priority control function between clients and the fair share function between users (QoS function), and other functions.
- High reliability
Contains the failover function for MGS (Management Server), MDS (Meta Data Server), and OSS (Object Storage Server).
- Expandability
Allows dynamic expansion of meta data domain/data storage domain.

1.2 FEFS System Configuration

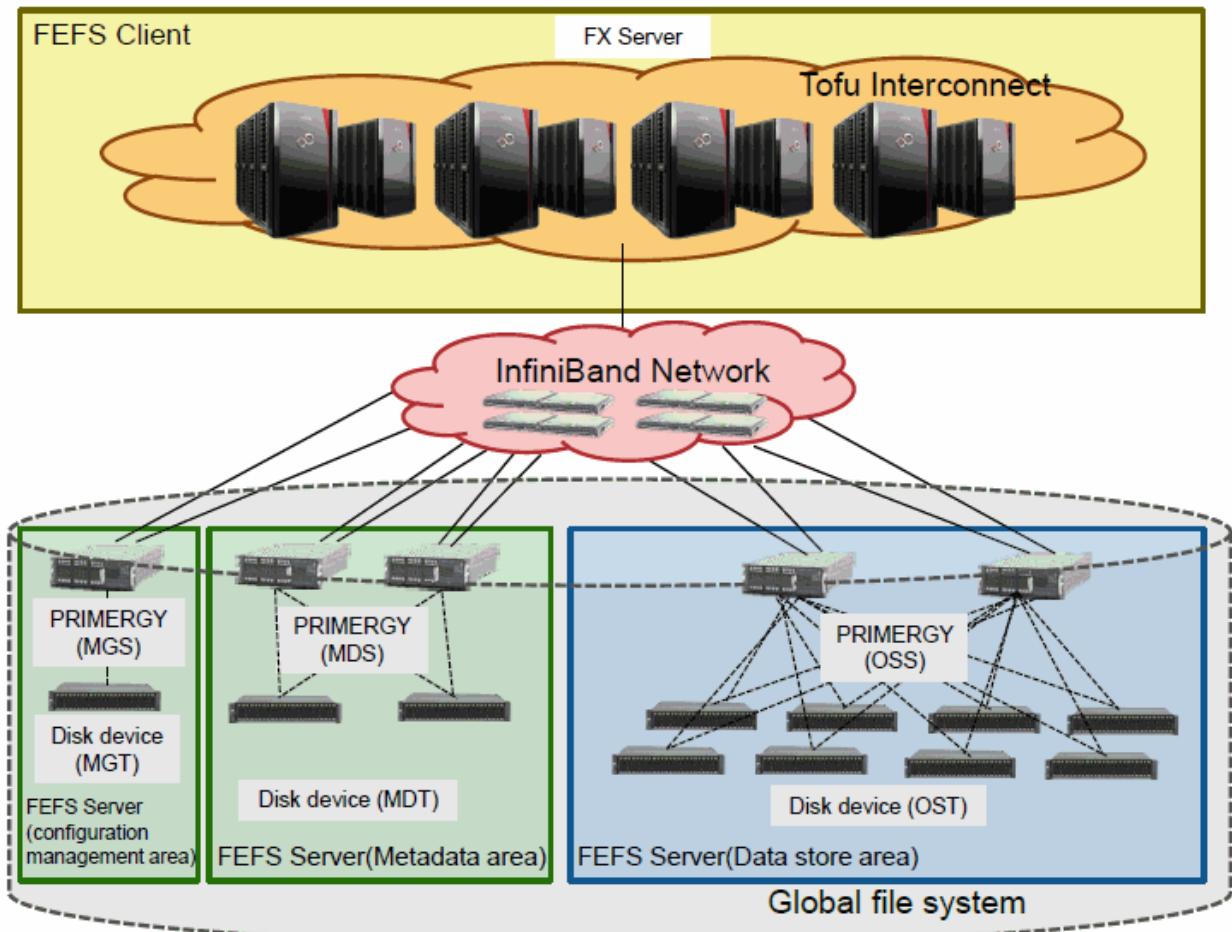
The system configuration of FEFS is shown as follows. It describes the client configuration, server configuration and network configuration of FEFS.

1.2.1 Hardware configuration

"[Figure 1.1 Hardware configuration using FX server](#)" shows the hardware configuration using Tofu interconnect.

In this configuration, the compute node uses FX server. Also, the networks to connect groups of compute nodes are configured by Tofu interconnect. The networks between server and client are configured by InfiniBand.

Figure 1.1 Hardware configuration using FX server



Node types that can be used as FEFS clients are as follows:

- Compute Node (CN)
- Compute Node and Storage I/O node (CN/SIO)
- Compute Node and Global I/O node (CN/GIO)
- Compute Node and Boot I/O node (CN/BIO)
- Login Node (LN)
- Compute Cluster Management node (CCM)
- Multiuse node

For details on the node types, see the *Job Operation Software Overview*.

1.2.2 Server configurations

FEFS consists of the servers or disk units listed below.

- Management server (referred to below as MGS):
The MGS is a server that manages MDS, MDT, OSS, and OST configurations.
- Management volume (management target, referred to below as MGT):
The MGT is a disk unit that stores the configuration information of file system.
- Metadata server (referred to below as MDS):
The MDS is a server that manages metadata (inode, stripe information, directory entry).

- Metadata logical volume (metadata target, referred to below as MDT):

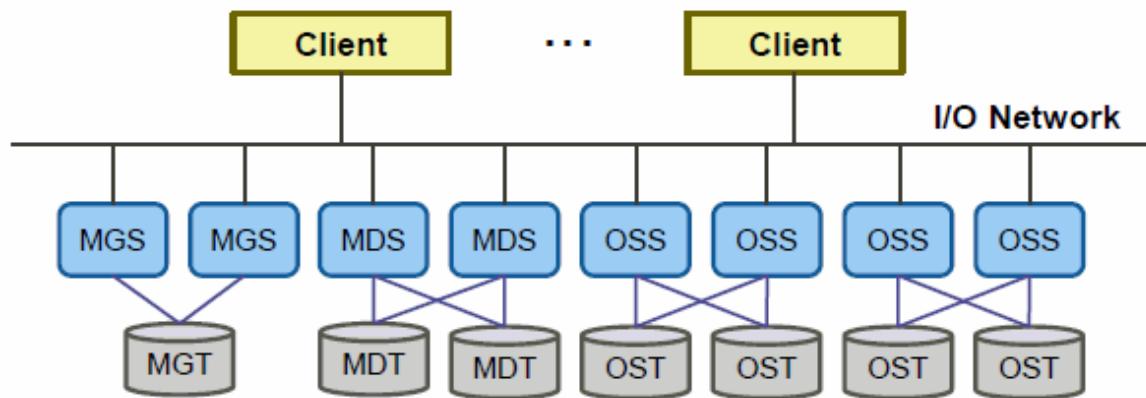
The MDT is a disk unit that stores metadata.
- Object storage server (referred to below as OSS):

The OSS is a server that manages the file data area.
- Object storage logical volume (object storage target, referred to below as OST):

The OST is a disk unit that stores the actual file data.

The following figures show examples of an FEFS server configuration.

Figure 1.2 Server configuration (when the MGS and MDS are built on different machines)



MGS/MGT configuration

MGS supports Active/Standby method redundant configuration. You can operate MGS and MDS on the same machine. However, we recommend configuring MGS and MDS on different machines to realize independent management. An independent MGT is required when MDS and MGS are made another server or the redundant configuration of MDS is assumed to be Active/Active method.

MGT volume configuration

We recommend using the MGT with a RAID1 (mirroring) configuration for reliability.

The volume size required for the MGT can be calculated by using the following expression:

<code>200MiB x number of file systems</code>
--

MDS/MDT configuration

FEFS achieves common file views and exclusive control through a common file tree and centralized management. FEFS shows space in the tree from the MDS to clients, and file open requests received from applications are managed in a centralized manner.

FEFS manages file system metadata with pairs of MDS and MDT. You can distribute the metadata access load by increasing the number of MDS and MDT pairs. For details, refer to "[2.2 Multi-MDS Function](#)".

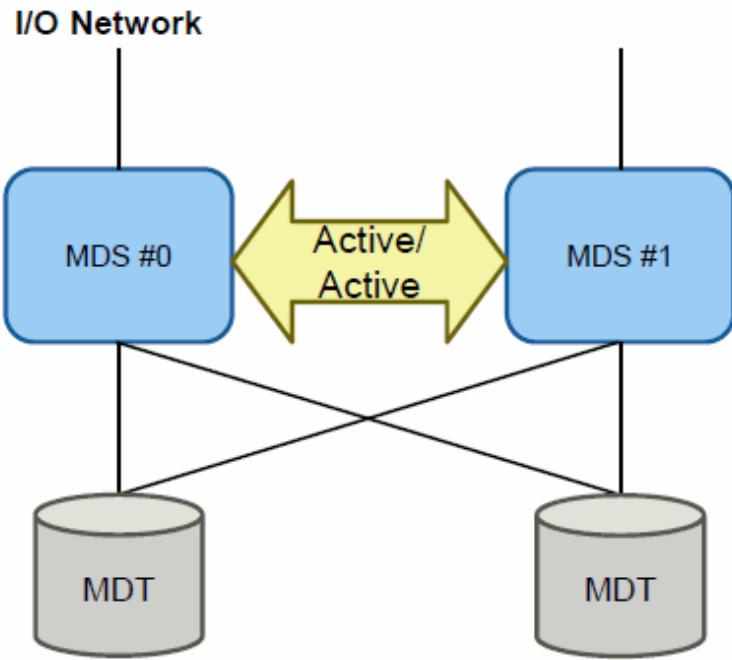
The MDS supports both Active/Standby method and Active/Active method redundant configuration, where operation can continue by switching to the standby node if the active node has failed.

Also, an independent MGT is required when the redundant configuration of MDS is assumed to be Active/Active method.

For details on redundant configuration and node switching, see "[2.7 RAS Function/FEFS Status Check Function](#)".

An example of the minimum MDS and MDT configuration is shown below.

Figure 1.3 MDS/MDT configuration



MDT volume configuration

We recommend using the MDT with a RAID10 (mirroring + striping) configuration for high-speed and reliability. The recommended minimum volume size of the MDT is 100 GiB.

You can calculate the required volume size for the MDT by using the following expression, based on the number of files (number of inodes) stored in the file system.

```
(Number of inodes + number of files on which an ACL or stripes are configured) x 4 KiB  
+ ACL area (*1)  
+ stripe area (*2)  
+ user extended attribute area (*3)
```

*1 A volume size calculated by using the following expression is required to configure an ACL(Access Control List):
Number of files or directories on which an ACL is configured x number of ACL entries per file or directory x 8 bytes

*2 A volume size calculated by using the following expression is required to configure stripes:
Number of files or directories on which stripes are configured x number of stripes per file or directory x 24 bytes

*3 A volume size calculated by using the following expression is required to configure the user extended attribute:
Number of files or directories on which the user extended attribute is configured x user extended attribute size (bytes) per file or directory



The maximum volume size of one MDT is 8 TiB - 2 KiB.

If the required volume size, calculated by the above formula, for the MDT exceeds 8 TiB - 2 KiB, consider the following configurations:

- Configuration with multiple file systems
- Multi-MDS configuration

For stable system operation, take measures for capacity monitoring during operation so that insufficient capacity does not occur.



For details on the ACL function, striping function, and user extended attributes, see the functional descriptions and methods of operation written elsewhere in this guide.

OSS/OST configuration

FEFS manages the actual data of a file system via the OSS/OST.

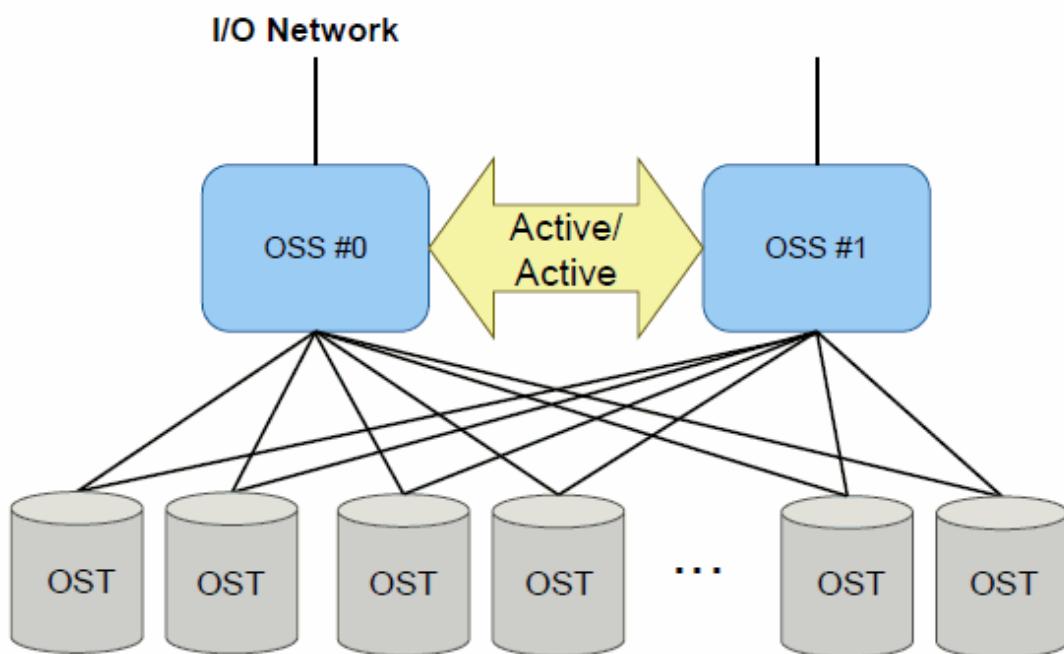
The OSS stores the client-written file data in the object storage logical volume (OST), and retrieves the client-read file data from the OST to transfer it to the client. Increasing the number of pairs of OSS and OST can increase the file system capacity and throughput performance.

The OSS supports the active/active redundant configuration, where operation can continue by switching to another node if a node has failed.

For details on the redundant configuration and node switchover, see "[2.7 RAS Function/FEFS Status Check Function](#)."

The following figure shows the minimum configuration of the OSS and OST.

Figure 1.4 OSS/OST configuration



OST volume configuration

The recommended OST configuration is RAID6 (double parity) because it satisfies both capacity and reliability requirements.

The recommended minimum volume size of the OST is 10 GiB.

The number of available inodes can be calculated as follows, depending on the volume size of the OST:

```
OST size (*1) / inode ratio (*2) = number of available inodes
*1 The volume size of a single OST. Note that this is not a total value across the entire system.
*2 The inode ratio is a disk size to be assigned to each inode at the time of OST formatting. The inode ratio varies depending on the OST size. The following table lists the ratios.
```

OST size	inode ratio
Less than 10 GiB	16 KiB
10 GiB to 1 TiB	68 KiB
1 TiB to 4 TiB	256 KiB
4 TiB to 16 TiB	512 KiB
16 TiB or larger	1 MiB

```

Calculation example 1: Number of available inodes when the size of the OST is 300 GiB
    300 GiB / 68 KiB = about 4.6 M inodes

Calculation example 2: Number of available inodes when the size of the OST is 2 TiB
    2 TiB / 256 KiB = about 8.3 M inodes

Calculation example 3: Number of available inodes when the size of each OST is 2 TiB in a file system
built with 10 OSTs
    2 TiB / 256 KiB * 10 = about 83 M inodes

```

If you want to specify an inode ratio irrespective of the size of the OST, specify the following for the OST OPTION of the MKFS OPTION in the GFS sheet of the FEFS design sheet:

```

inode ratio  the OST OPTION of the MKFS OPTION in the GFS sheet
For 16 KiB   --mkfsoptions="-i 16384"
For 68 KiB   --mkfsoptions="-i 69905"
For 256 KiB  --mkfsoptions="-i 262144"
For 512 KiB  --mkfsoptions="-i 524288"
For 1 MiB    --mkfsoptions="-i 1048576"

```

Note

Do not specify a value in inode ratio other than the above.

1.2.3 Network configuration

FEFS supports the following functions with dedicated network drivers.

- Tofu interconnect [FX]
 - For details on Tofu interconnect, see the *Job Operation Software Overview*.
- InfiniBand (OFED)
 - Delivers the performance of high-speed high-volume data transfer using RDMA (Remote Direct Memory Access) communication.
 - Realizes the following functions if multiple HCA cards are mounted:
 - High-bandwidth transfer via round-robin
 - Valid path selection and continuous file IO operation in the event of an InfiniBand failure

FEFS realizes the function to relay multiple networks mentioned above by communication networking referred to as LNet (Lustre Networking).

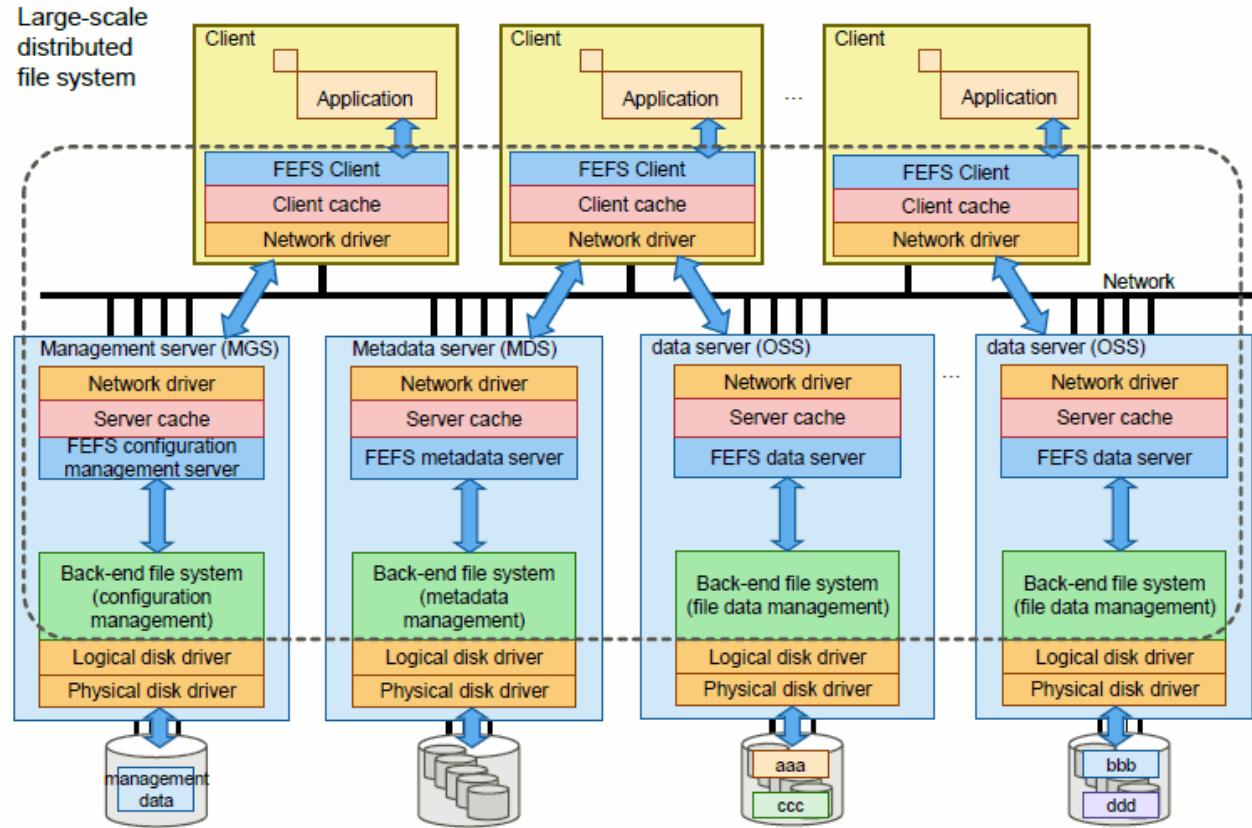
FEFS uses the following communication functions according to the purpose:

- The function for RDMA, which handles one-sided communication for the kernel, is used for high-speed data transfer.
- The system packet (Send/Receive) function is used for other communication.

1.3 FEFS Software Configuration.

"Figure 1.5 Software configuration" shows the FEFS software configuration.

Figure 1.5 Software configuration



1.4 Upper limits and Lower limits

The following table lists the upper limits and lower limits in FEFS.

Table 1.1 Upper limits and Lower limits in FEFS

Item	Specification
Maximum file system size	8 EiB
Maximum file size	62.5 PiB
Maximum number of files	16Ti
Maximum number of files or subdirectories in a single directory	10 Mi
Maximum number of meta volumes (MDTs)	4096
Maximum meta volume (MDT) size	8TiB - 2KiB
Minimum meta volume (MDT) size	32MiB
Maximum number of volumes (OSTs)	8150
Maximum volume (OST) size	2 PiB
Maximum number of clients	1 Mi
Maximum block size	4 KiB
Maximum number of stripes	4000
Minimum stripe size	64 KiB
Maximum stripe size	4194240KiB(4GiB-64KiB)
Maximum file name length	255

Item	Specification
Maximum path name length	4096 (includes the terminating null character)
Maximum number of files for QUOTA	8 Ei
Minimum number of files for QUOTA	1024 x number of MDT
Maximum disk size for QUOTA	8 EiB
Minimum disk size for QUOTA	1024 x number of OST
Maximum number of ACL entries	32
Maximum file system name length	8
Maximum OST pool name length	15

1.5 Precautions

This section describes points to consider when using FEFS.

Purpose of use

You can use the FEFS as application work areas and as areas under /home. You cannot use the FEFS as areas under the / (root directory), /var, or /usr directory, which is used by the operating system itself.

Mount point

Do not allow the same FEFS to be mounted at multiple mount points by the same FEFS client.

Support of multiple functions on the same node

This product enables the following combined configurations on the same node.

- MGS and MDS and OSS
- MGS and MDS
- This product does not support an operating server and a client on the same node.



If MDS and OSS are operated on the same node, the product does not support the following configurations.

- Multiple file systems configuration
- Server redundant configuration

Time synchronization

If the time differs between nodes, operational errors may occur. Synchronize the clocks on the nodes by using NTP, for example.

User authentication

If the UID/GID corresponding to a user name/group name is not consistent between the MDS and the client, the file system does not work correctly. For the UID/GID corresponding to each user/group using FEFS, make the setting the same throughout the MDS and the client.

If an EIDRM error (Identifier removed) error occurs during file access, user authentication may not have been successful. If you have attempted LDAP user authentication, check whether communication with the LDAP server is normal.

Notes on using FEFS with multiuse nodes

The multiuse nodes and multiuse clusters in Technical Computing Suite can be used for any purpose.

Multiuse nodes in a storage cluster cannot use the FEFS client functions.

For details on multiuse nodes and multiuse clusters, see the *Job Operation Software Overview*.

Access to the FEFS area

Access with a stat call such as ls -al to FEFS may take several tens of seconds before the processing ends.

Chapter 2 Functions

This chapter describes FEFS functions.

2.1 Striping Function

2.1.1 Round-robin and striping

FEFS distributes and stores file data onto the OST via the OSS.

You can select either of two methods as the distribution method.

- Round-robin method

The OST is selected by round-robin for each file, and then the file is stored there.

The maximum file size is the physical OST capacity.

This is the default distribution method.

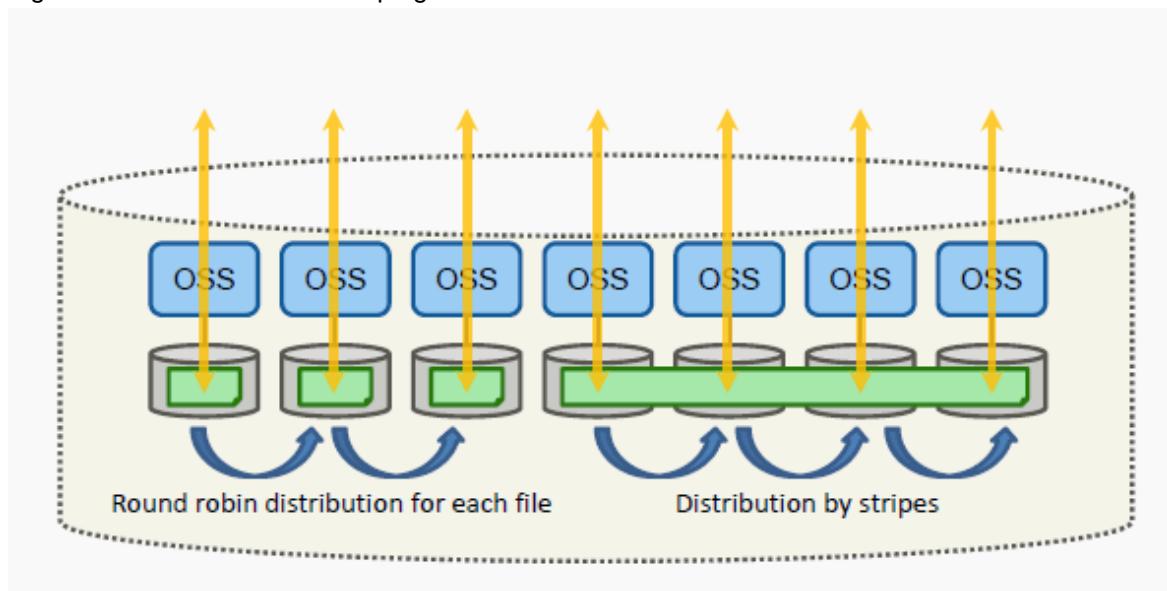
- Striping method

The data for a single file is distributed to and stored on OSTs.

The file size can exceed the physical one OST capacity.

To use this method, it must be set up. For details on how to set it, see "[4.3 Configuring the Striping Function](#)".

Figure 2.1 Round-robin and striping



2.1.2 Advantages of the striping function

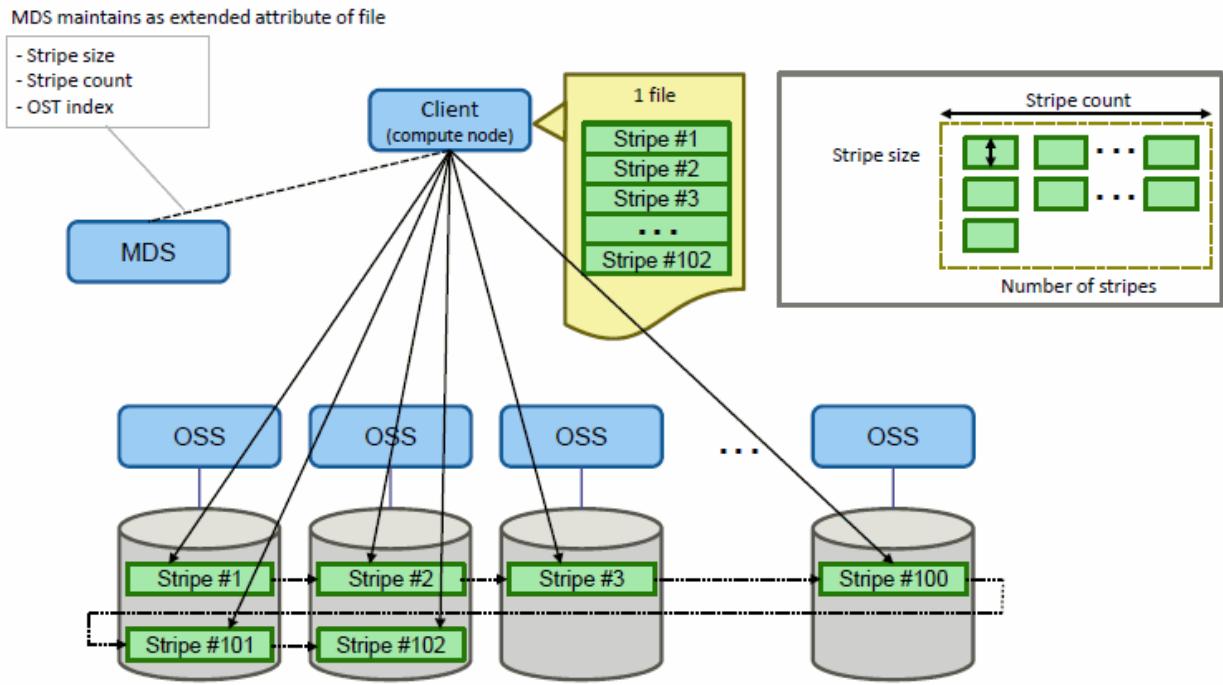
The striping function distributes and stores the data, in the specified size, for a single file from the beginning of the file onto multiple OSTs to distribute file access.

The advantages are as follows.

- A file of a size exceeding the physical OST capacity can be created.

- Distributing and storing the data for a single file onto multiple OSTs improves file access bandwidth.

Figure 2.2 Striping function



Note that to use the striping function, it must be set up.

You can specify a size for the data to be distributed (stripe size), OST range (stripe count), etc. in the setup.

For details, see "[4.3 Configuring the Striping Function.](#)"

2.1.3 OST_pool function

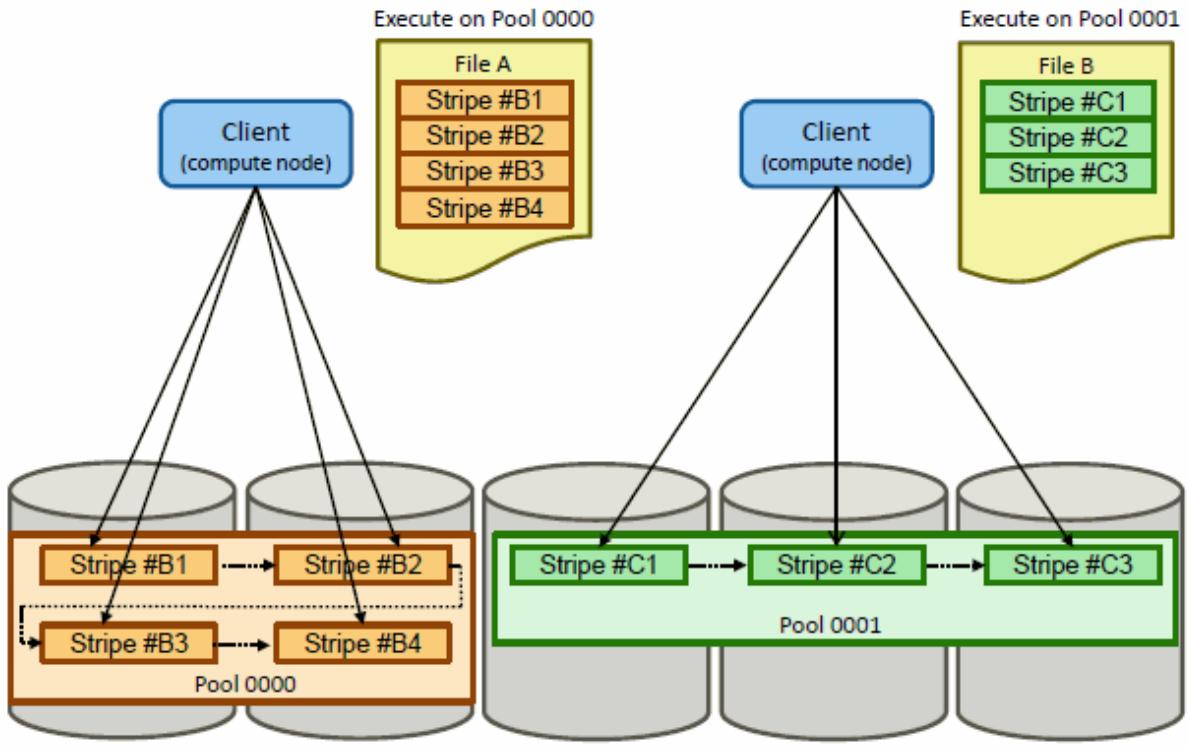
FEFS provides the OST_pool function, which groups and manages OSTs.

The OST_pool function groups multiple OSTs and defines them as one group.

You can specify the OST range allocated for file distribution by the striping function.

The following is the concept of OST_pool.

Figure 2.3 OST_pool function

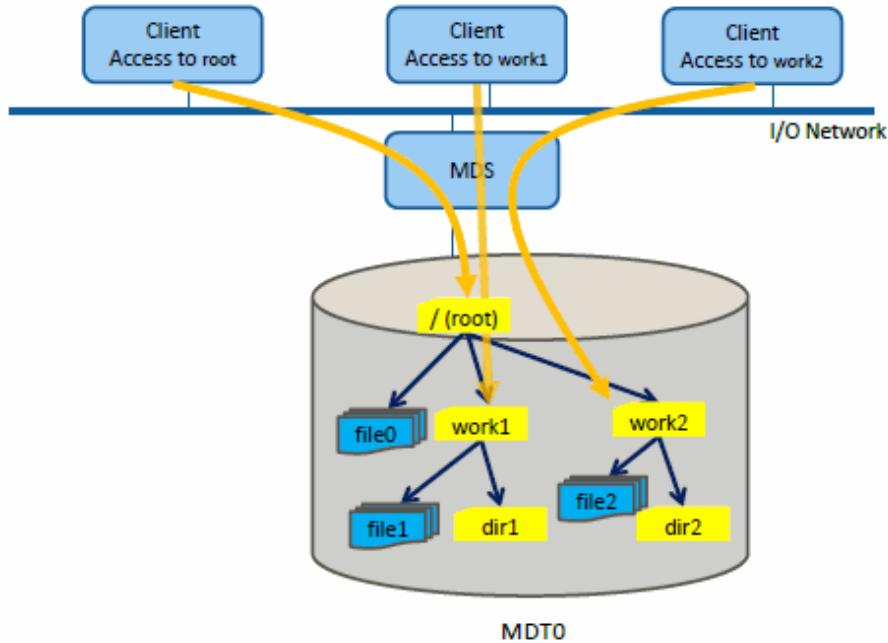


2.2 Multi-MDS Function

FEFS allows you to have a multi-MDS configuration where a single file system consists of multiple MDSs.

An example of a conventional single MDS configuration is shown below.

Figure 2.4 Example of single MDS configuration



In a single MDS configuration, metadata for the entire file system is managed by a single MDS/MDT pair, so metadata access becomes a performance bottleneck as the size of the file system increases. In a multi-MDS configuration, metadata can be managed by multiple MDS/MDTs.

MDT pairs to distribute the access load and improve overall system performance. In addition, the maximum number of files that can be managed can be expanded by increasing the MDT.

There are remote directory and striped directory in the method that uses the multi-MDS configuration.

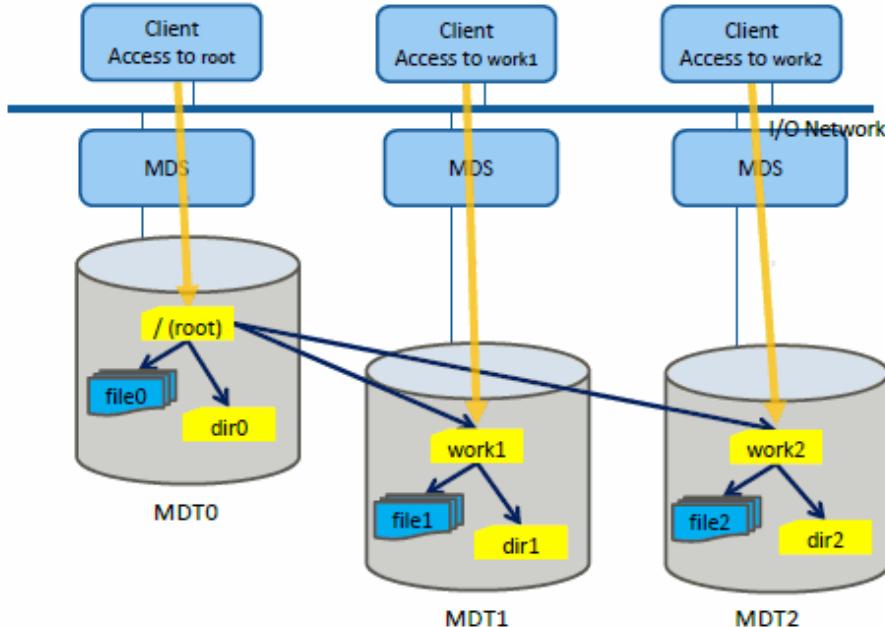
All metadata is stored only in MDT 0 unless you explicitly set up metadata distribution in either way.

2.2.1 Remote directory

Remote directory is a method of storing the metadata of a directory in a specific MDT that is different from the metadata of the parent directory. When a file is created under a remote directory, its metadata is stored in the same MDT as the remote directory.

The following is an example of a remote directory:

Figure 2.5 Remote directory



The root directory (/) is managed by MDS that mounts MDT0. By adding the MDS that mounts a certain MDT_n here, you can set so as to manage a certain subdirectory. MDT1 and MDT2 in the above figure manages part of the trees with /work1 and /work2 at the top respectively.

Remote directory can assign users and projects that require metadata access performance to specific MDTs and prevent the impact of other users on metadata access.



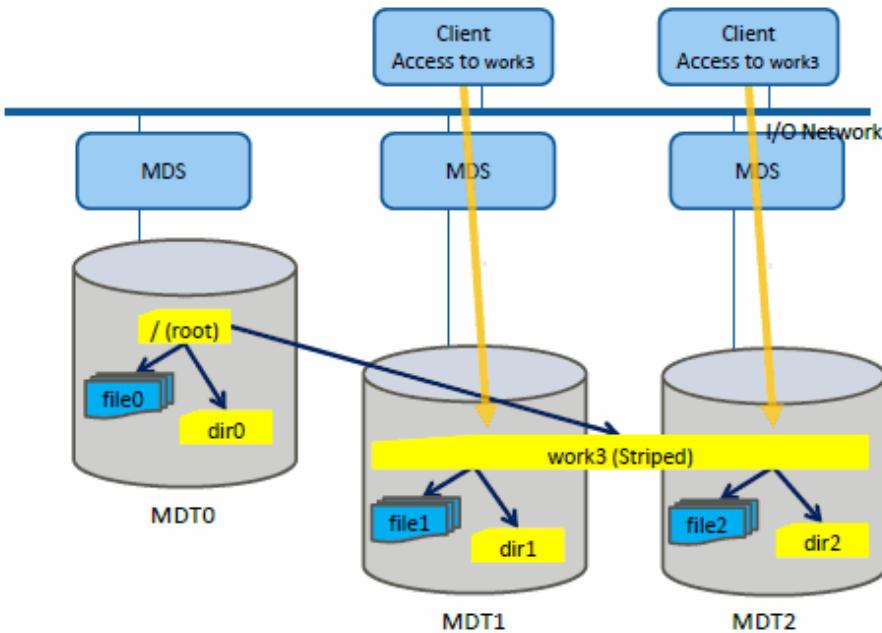
If the file under the directory allocated to a certain MDT is moved under the directory allocated to another MDT, inode is consumed both at the source MDT and the target MDT. For example, in the above figure, if file 1 under /work1 directory of MDT1 is moved under /work2 directory of MDT2, inode is consumed both at MDT1 and MDT2.

2.2.2 Striped directory

A striped directory is a method of storing metadata of a directory in a specified range of MDTs. When a file is created under the striped directory, its metadata is stored in one MDT within the specified range.

An example of a striped directory is shown below.

Figure 2.6 Striped directory



In the above example, the metadata of directories dir1 and dir2 and files file1 and file2 under /work3 are distributed and stored in MDT1 and MDT2, respectively.

You can specify the MDT range to be striped (stripe count) and the first MDT number (stripe index) when creating directory with the lfs mkdir command. You can allocate the metadata under the striping directory to MDT with round-robin.

Note

- When the file under the striped directory is moved, it may be allocated to another MDT. At that time, inode is consumed both at the source MDT and at the target MDT.

2.3 QUOTA Function

QUOTA is a function to limit the disk capacity and number of files that can be used. QUOTA can be configured by the system administrator for individual users, groups, or projects.

Note

- Cache and hard limit

The calculation of disk usage displayed by the lfs quota command is based on the number of blocks allocated to the OST. Sometimes, the value exceeds the hard limit.

For high-speed processing, FEFS temporarily caches write data in memory. Then, after a certain length of time has elapsed, FEFS writes the data to the OST. However, the data cached in memory does not consume any space on the OST, so writing the data may exceed the QUOTA limit.

Therefore, the executed lfs quota command may display a value exceeding the hard limit. However, it does not cause a problem in FEFS operation.

- Disk usage

The disk usage displayed by the lfs quota command is the total disk usage of each MDT and OST displayed by lfs quota -v. QUOTA limits only apply to OST disk usage, not MDT data limits. For this reason, the lfs quota command may display a value that exceeds the limit of MDT data usage, which is acceptable for FEFS.

- About "*" displayed when a soft limit is exceeded.

When a soft limit is exceeded by information displayed by the lfs quota command, "*" is added after the value.

Output example of lfs quota command

```
# lfs quota -u user1 /mnt/fefs
Disk quotas for user user1 (uid 1070):
Filesystem  kbytes  quota  limit  grace  files  quota  limit  grace
/mnt/fefs    118876     0      0     -    1505*   1500    3000  6d23h59m57s
```

"*" might not disappear even when it deletes a file from the above-mentioned and the number of inode becomes less than soft limit. There is no problem as operation of FEFS because the file creation is possible if the number of inode is less than soft limits even when moratorium (grace) is exceeded with "*" displayed.

- Handling sparse files

A sparse file is a file containing areas in the middle on which no data is written.

The areas without data in a sparse file are not included in the disk usage counted against QUOTA because they do not consume any blocks on the OST.

- An error of QUOTA limit

QUOTA limit for disk usage may differ from the set limit value. The following is Maximum value of the gap:

Write from CN [FX]

```
4MiB x <Number of OSTs> x <Number of clients>
```

Number of OSTs: Number of OSTs in the file system for which QUOTA limit for disk usage is to be set

Number of clients: Number of clients that are to have the file system for which QUOTA limit for disk usage is to be set mounted

Write from LN [PG]

```
512MiB x <Number of OSTs> x <Number of clients>
```

Number of OSTs: Number of OSTs in the file system for which QUOTA limit for disk usage is to be set

Number of clients: Number of clients that are to have the file system for which QUOTA limit for disk usage is to be set mounted

QUOTA limit for the number of inodes may deviate from the set limit value.

```
1024 x <Number of MDTs>
```

Number of MDTs: Number of MDTs in the file system for which QUOTA limit for disk usage is to be set

- Increasing the number of inodes by QUOTA by creating striped directories

Creating a striped directory increases the number of inodes by QUOTA (Stripe count + 1).

2.3.1 QUOTA function for user or group

Like ext3 and other file systems, FEFS can use the QUOTA function as follows:

- Settings of the number of files and disk usage
- Soft and hard limit settings

The following features are specific to the QUOTA function of FEFS:

- Special commands for QUOTA management (e.g., "lfs setquota" command. For details on commands, see "[A.2 Commands.](#)")
- The amount of each MDT of the disk use can be displayed for file system composed of two or more MDT. (-v option of "[lfs quota](#)" command)
- Capability (using the -v option of the "[lfs quota](#)" command) of displaying the disk usage of each OST in a file system configured with multiple OSTs

2.3.2 Project QUOTA function

In the FEFS, you can set QUOTA not only to users or groups, but also to projects. The "project" refers to a certain group of inodes specifying file and directory, to which the same "project ID" is assigned, respectively. Just like user name and group name, you can set and manage QUOTA for each project ID.

The project ID is usually assigned to a single directory. However, you can assign the same project ID to multiple directories and consolidate them for QUOTA management. In addition, you can also assign different project ID to each subdirectory in the hierarchical directory structure. Accordingly, this allows more flexible management compared to setting QUOTA directly to a certain directory.

The following shows the operation image and application scene of QUOTA function relating a certain directory to the project.

Figure 2.7 Image where QUOTA is operated

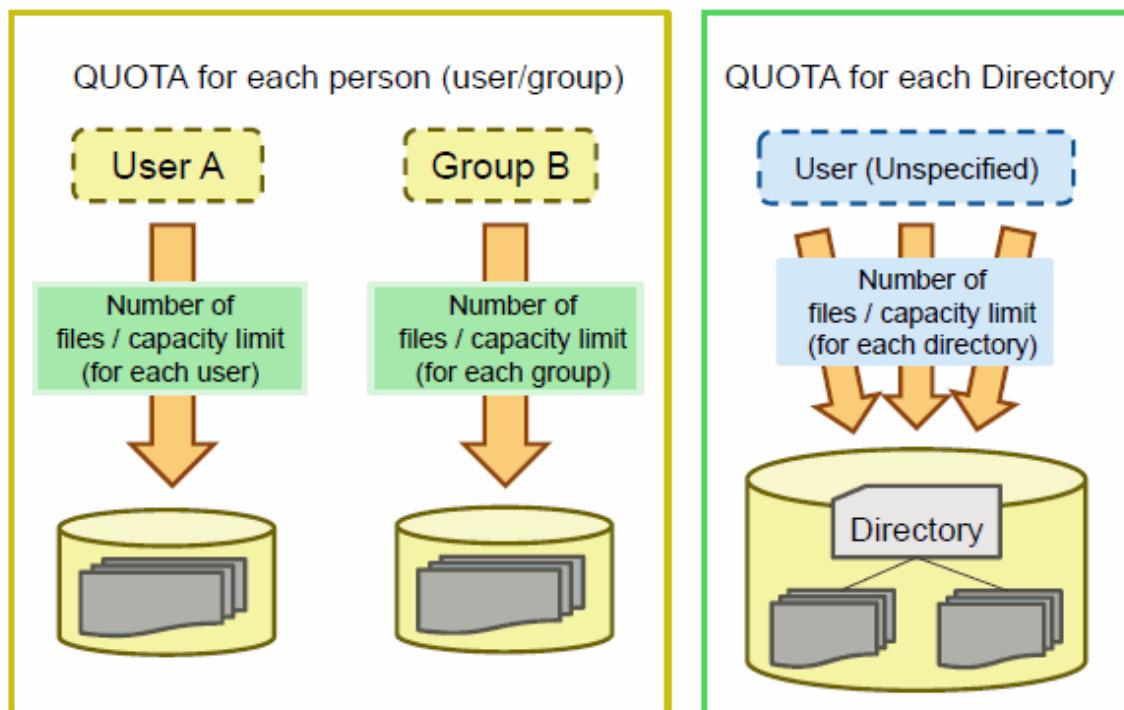
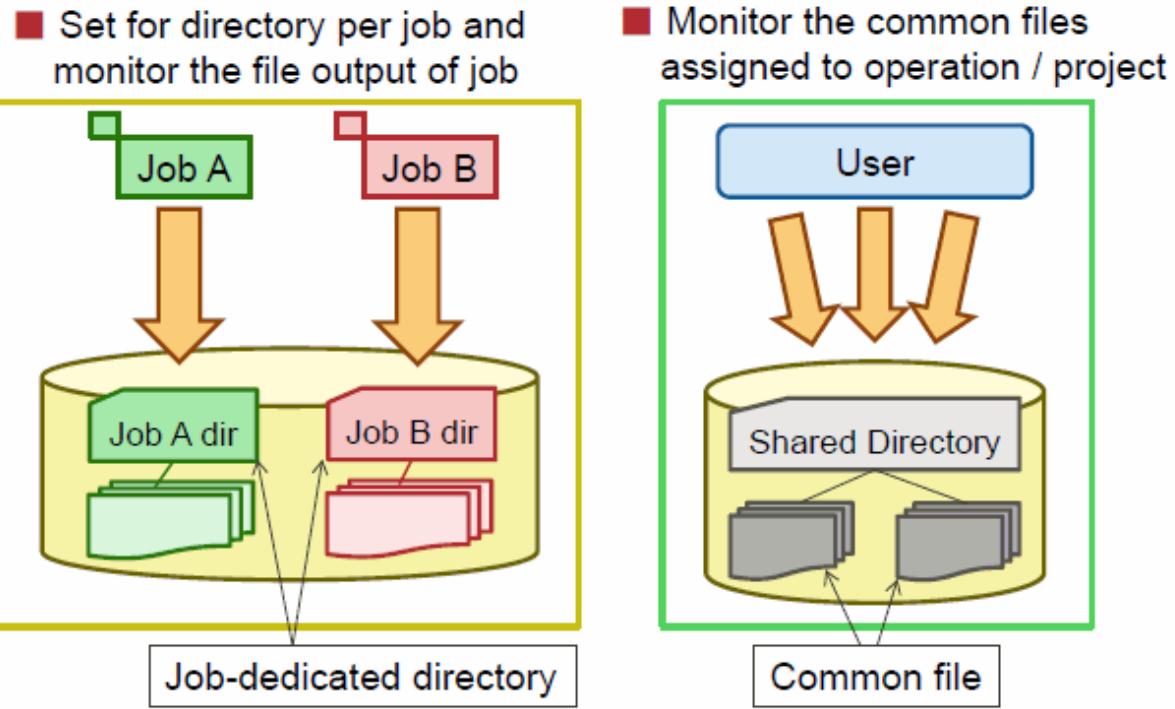


Figure 2.8 Scene where the Project QUOTA is applied



Note

- About rename and link System calls

When files are moved with the rename system call between directories that have different project QUOTA settings or directories, one of which has project QUOTA settings, no QUOTA exceeded error occurs even if the QUOTA limit of the destination project QUOTA is exceeded.

If directories are moved with the rename system call between directories that have different project QUOTA settings or directories, one of which has project QUOTA settings, -1 is returned (error number EXDEV).

If directories or files are created using the link system call between directories that have different project QUOTA settings or directories, one of which has project QUOTA settings, -1 is returned (error number EXDEV).
- Symbolic link

The disk capacity of a symbolic link should be the size used by an inode, not the linked entry size. For any existing symbolic link for a directory, entries in the linked directory cannot be counted as the disk usage and the number of inodes.
- The values that can be specified as project ID are from 1 to 4294967295.
- Manage the project ID appropriately so that a unique value is assigned to each project.

2.4 QoS Function

On a large-scale system used by many users, even a large amount of file IO by a particular user should not have an adverse effect on other users. Also, even with file access from a job on a compute node, user response on the login node should not be affected.

FEFS clears these hurdles by using a QoS (Quality of Service) function. The QoS function has the following features:

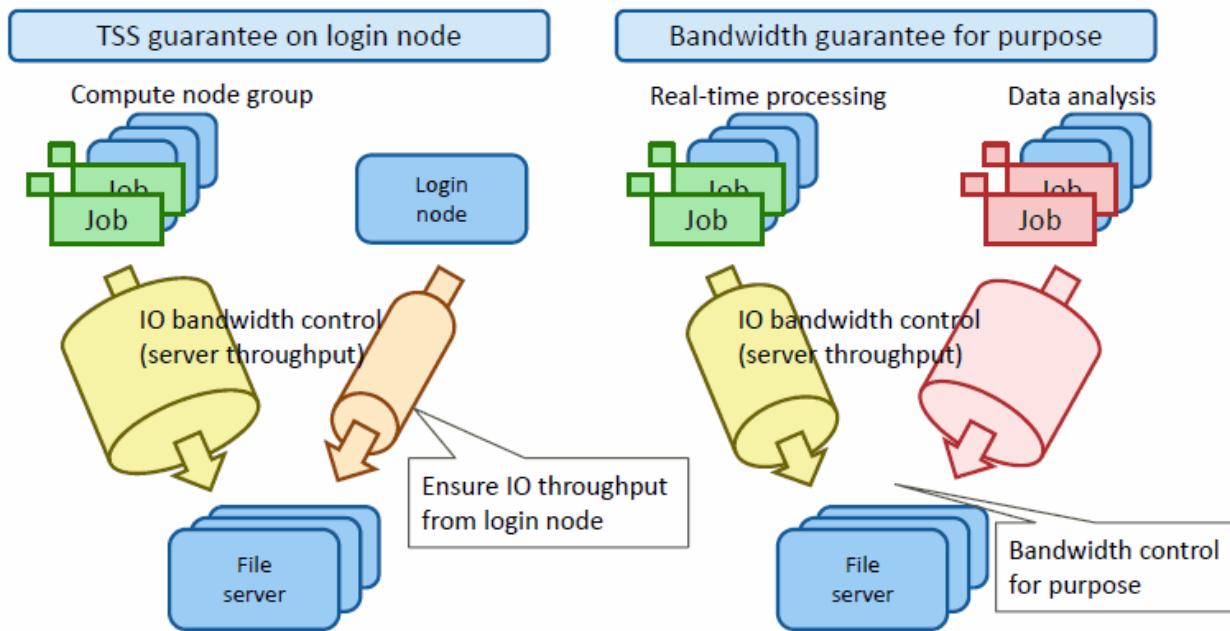
- IO priority control function between clients
- Fair share function between users

By default, the QoS function is disabled. For the procedure for enabling it, see "[3.2 Settings to Enable the QoS Function.](#)"

2.4.1 IO priority control function between clients

FEFS can restrict the IO throughput of a client group on a server. This function can prevent TSS response from deteriorating on a login node because of job IO. It can also guarantee the bandwidth suitable for cluster operation in a client group. "Figure 2.9 IO priority control function between clients" shows a conceptual image of IO priority control between clients.

Figure 2.9 IO priority control function between clients



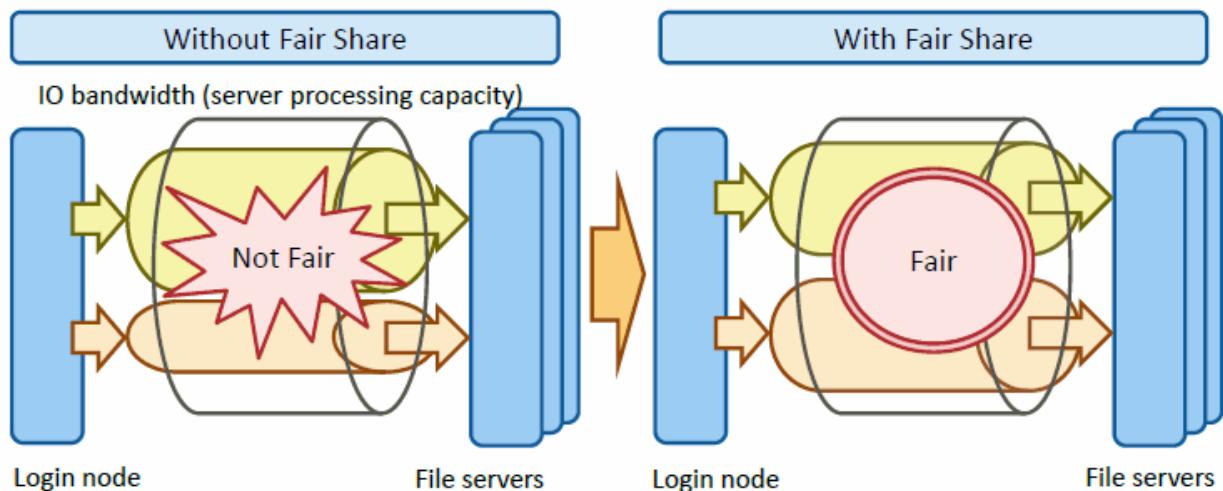
2.4.2 Fair share function between users

FEFS controls the number of IO requests that are issued and processed on each client and server, preventing a particular user from monopolizing IO resources.

The client restricts the IO requests that can be issued simultaneously by one user to an upper limit, thereby preventing a single user from issuing numerous IO requests and monopolizing IO bandwidth and server resources.

If the same user application issues IO requests simultaneously from multiple clients, such as on compute nodes, it can monopolize file server resources. Therefore, the server throughput capacity available to a single user is controlled on file servers to prevent a single user from monopolizing server resources through IO requests.

Figure 2.10 Fair share function between users



2.5 ACL (Access Control List) Function

The ACL function enables more flexible file access control than file access control with the conventional UNIX format (file owner/group/other user). With the ACL, you can set the access privileges of each user or group.

Like ext3 and other file systems, FEFS can use the ACL function as follows:

- setfacl command for setting the ACL
- getfacl command for displaying ACL information

The maximum number of ACL entries that can be set for a single file or directory is 8191.

By default, the ACL function is disabled. For the procedure for enabling the ACL function, see "[3.4 Settings to Enable the ACL Function.](#)"

2.6 Journaling Function

FEFS provides a journaling function to prevent conflicts that would otherwise occur in file system configuration information when the system is powered off or goes down. Separate journaling functions are provided for the MDT and OST.

The journaling function needs to periodically access the disk to perform the recording process for journal files. You can reduce this overhead by allocating the journal area on an external device. This is called the external journal function.

By default, the external journaling function is disabled. For details on how to set it, see "[3.1.3 Creating FEFS design sheets.](#)"



Note

For the journal, use the internal journal.

If you use an external journal, you need a free memory space for the size of the external journal to be set.

The performance of the MDS may deteriorate depending on the size set for the external journal. Therefore, when designing the external journal, contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

2.7 RAS Function/FEFS Status Check Function

This function provides the RAS function together with an FEFS status check function and a failover function.

FEFS enables continuous service as a file system, even at the time of a single point of failure, by duplicating hardware and controlling software to switch server and IO communication paths.

The failover function continues IO operations even if an error occurs on an FEFS server, thereby achieving continuous system operation.

2.7.1 Failover configuration

Failover can be classified by configuration as follows.

Active/Active configuration

In this configuration, both nodes of a failover pair act as the active system during normal operation.

"[Figure 2.11 Active/Active failover configuration](#)" shows the configuration and operational concept.

The FEFS service runs on both nodes and when a failover occurs, all services run only on the normal node.

Active/Standby configuration

In this configuration, normally, the nodes of a failover pair are completely separated as the active node and the standby node.

"[Figure 2.12 Active/Standby failover configuration](#)" shows the configuration and operational concept.

Normally, the active node provides the FEFS service, and the standby node is not used. Failover results in the standby node providing the FEFS service.

Figure 2.11 Active/Active failover configuration

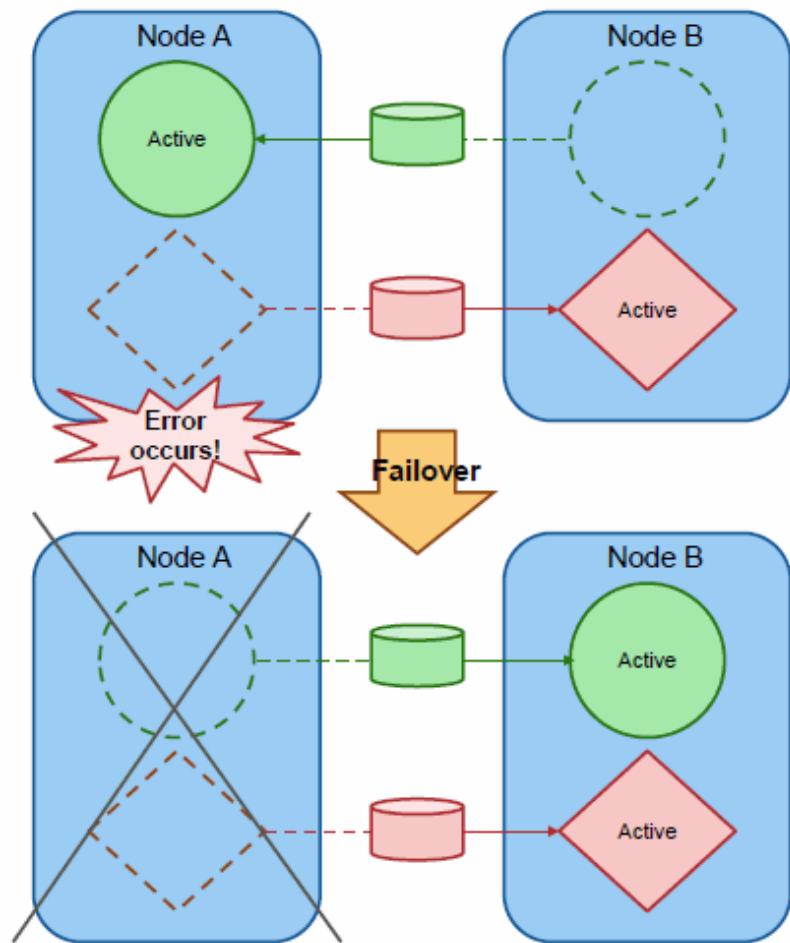
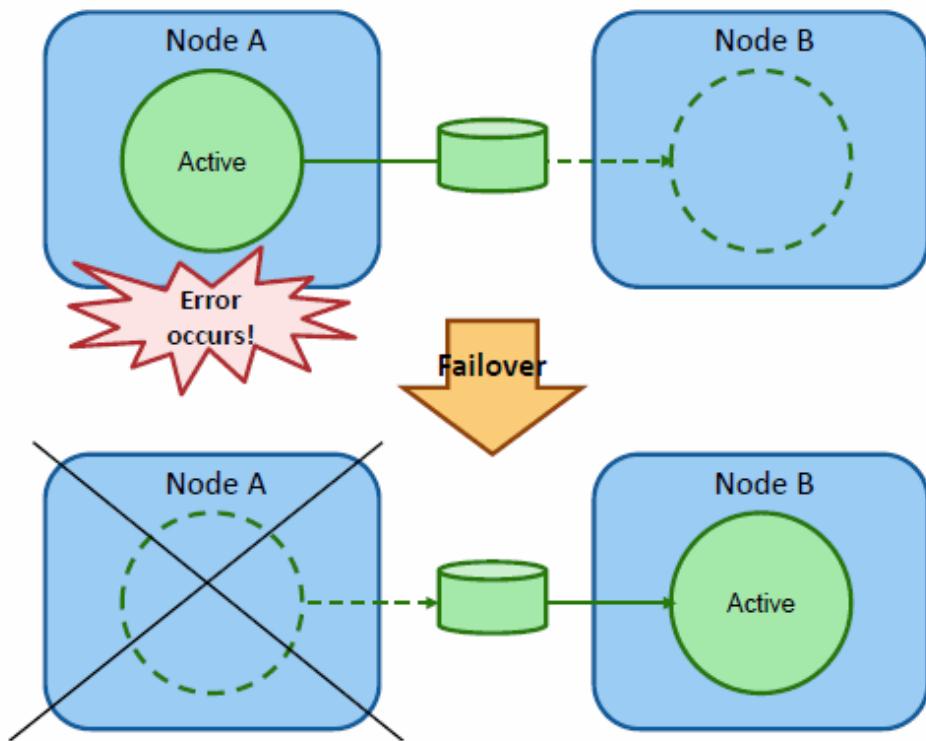


Figure 2.12 Active/Standby failover configuration





Note

- Although the operation can be executed with all services running on one side, it affects the performance.
The performance is not affected in an Active/Standby configuration.
- Before a redundant configuration can be used for MDSs, the MDSs and MDTs (MGTs) must be connected in a multipath configuration.
- During failover, job IO processing temporarily stops until the node switchover ends, and the job will continue after the IO processing resumes.

2.7.2 FEFS service monitoring

FEFS service monitoring is the function to monitor and report the FEFS related service status that can be used when cooperating with Job Operation Software. FEFS service monitoring daemon resides in each node and acquires and reports the FEFS related service status by the report from Job Operation Software.

Table 2.1 Monitoring target of FEFS service monitoring daemon (hardware)

Node type	FC	InfiniBand	SSD connected by NVMe *
MGS, MDS, OSS	o	o	-
CCM, LN, GIO[FX], Multiuse node	-	o	-
SIO[FX]	-	-	o
CN	-	-	-

* These devices are assumed when using LLIO. For details on LLIO, see the *LLIO User's Guide*.

o: Monitored

-: Not monitored

Table 2.2 Monitoring target of FEFS service monitoring daemon (software)

Node type	Mount		LLIO service	
	Storage	Global FS	Server	Client
MGS, MDS, OSS	o	-	-	-
CCM, LN	-	o	-	-
GIO[FX]	-	*	-	o
SIO[FX]	-	o	o	-
CN	-	*	-	o
Multiuse node	-	o	-	-

*: "-" In the environment where the LLIO service exists, "o" in the environment where the LLIO service does not exist.

o: Monitored

-: Not monitored

FEFS service monitoring function is divided into the following two services: FEFSSR service that monitors server function and relay function and FEFS service that monitors client function. For details, see "[4.11 Checking the FEFS Status](#)" and the *Job Operation Software Administrator's Guide for System Management*.

2.7.3 FEFS status check

You can check the FEFS service status with the pashowclst command.



For details on the state confirm method of service by the pashowlst command, see the following manual:

Job Operation Software Administrator's Guide for System Management

2.7.4 Server failover (MGS/MDS/OSS)

To use the failover function, two MGSs/MDSs/OSSs must be paired as a failover pair, and the same MGT/MDT/OST must be able to be mounted.

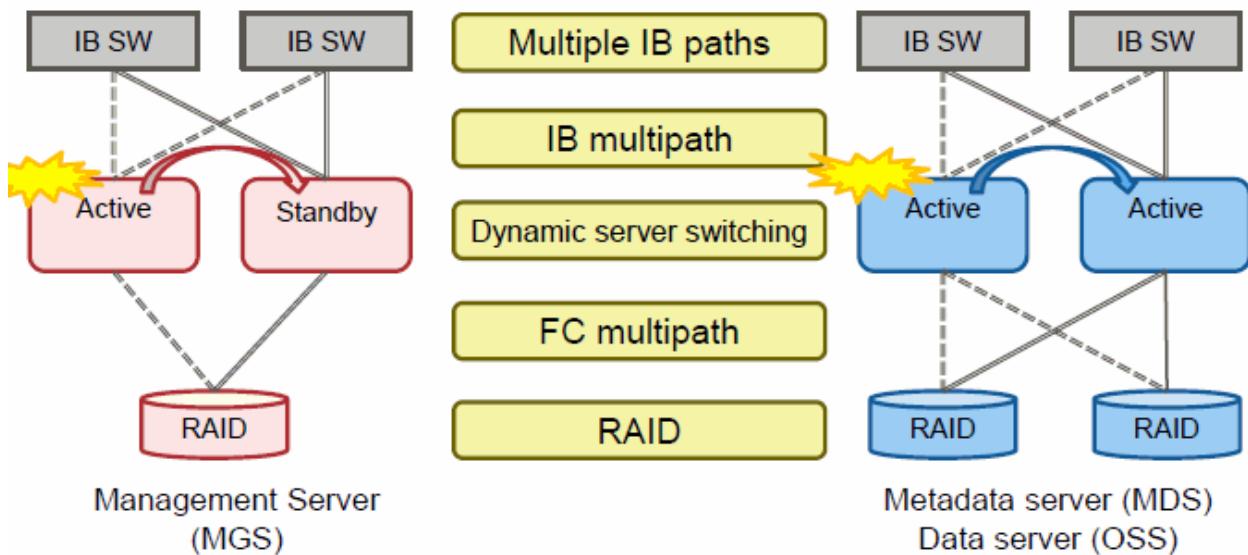
The MGTs/MDTs/OSTs shared by the failover pair are not mounted at the same time, but mounted as one of the two MDSs/OSSs.

With the MGT, MDT, and OST mounted on the other side of the failover pair, IO can continue when an error occurs.

The failover can take place only through linkage with the operational software.

When an error occurs in an FEFS related service, the FEFS service monitoring daemon resident in each node reports the error to the operational software, after which a failover automatically occurs.

Figure 2.13 Server failover (MGS/MDS/OSS)

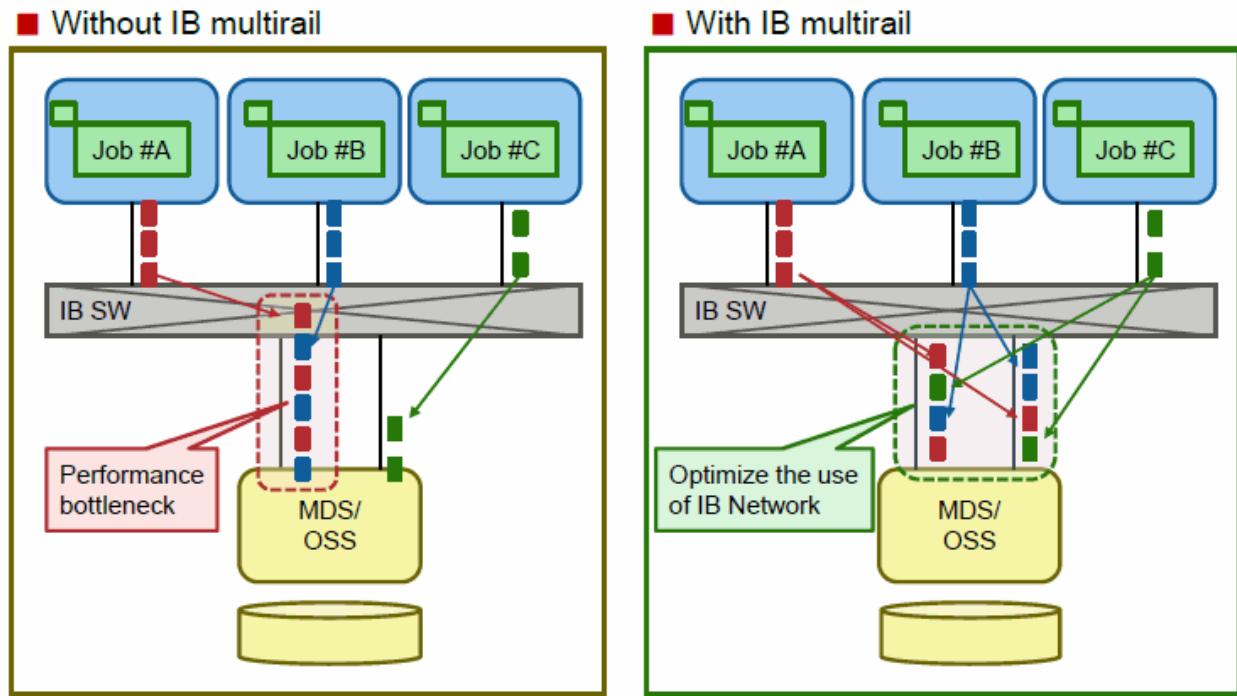


2.7.5 LNet multirail function

FEFS supports a communication function that uses multiple paths at the same time to deliver high bandwidth and high fault tolerance for multiple HCAs mounted on a server. This function is called the LNet multirail function.

If an InfiniBand communication path fails, the function automatically selects a valid path to continue each job.

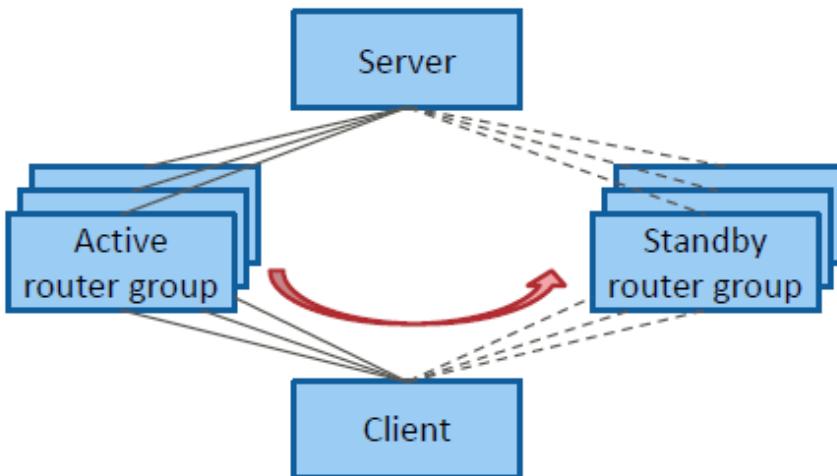
Figure 2.14 LNet multirail



2.7.6 LNet router

The LNet router has an active LNet router group and a standby LNet router group. When a fault occurs with the active LNet router group, the communication path is switched to the standby router group to continue operation.

Figure 2.15 Failover of LNet router



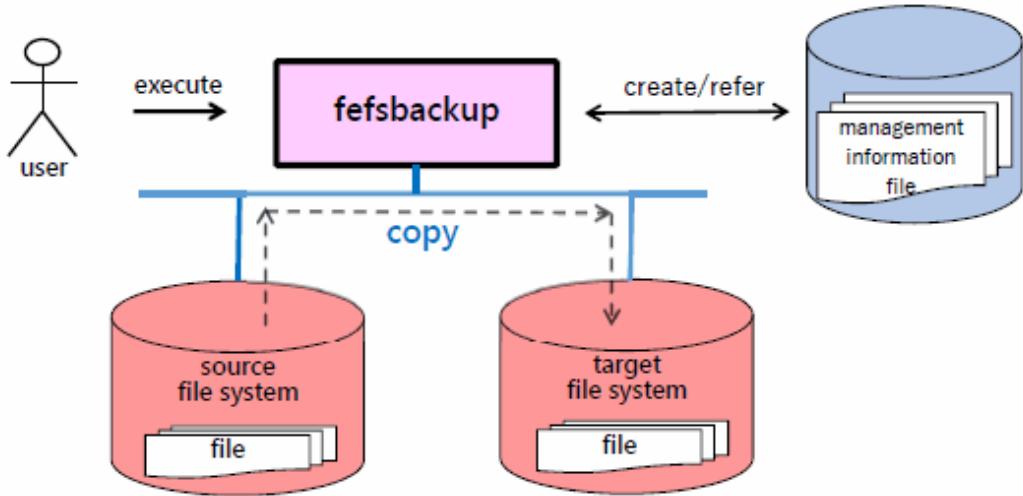
2.8 Data Management Tool (fefsb backup command) [PG]

FEFS provides a data management tool for fast data migration between file systems and for fast data evacuation in the event of a disk failure. The data management tool performs high-speed data transfer by taking the following measures.

- Parallel execution of multiple rsync.
- Multiple small-size files are consolidated in a tar archive on the node, which is then transferred.
- Only the difference file is transferred after the second time.

The following figure shows the operation of this function.

Figure 2.16 fefsbackup operation



For details on subcommands, see "[4.20 How to Use the Data Management Tool \(fefsbacup Command\) \[PG\]](#)".

Note

Data management tool does not support data migration from "K" or Technical Computing Suite V2.0/V3.0 system.

To use data management tool, rsync must be installed on both the server and the client.

FEFS is the only file system supported by data management tool.

Do not update, delete, or add files or directories while the data management tool is performing a file transfer.

2.9 FEFS statistics visualization function (fefssv.ph script)

The FEFS statistics visualization function uses a tool called the fefssv.ph script, which can acquire and output the frequency of executed FEFS requests, IO volume, etc. in the FEFS area by job on each server.

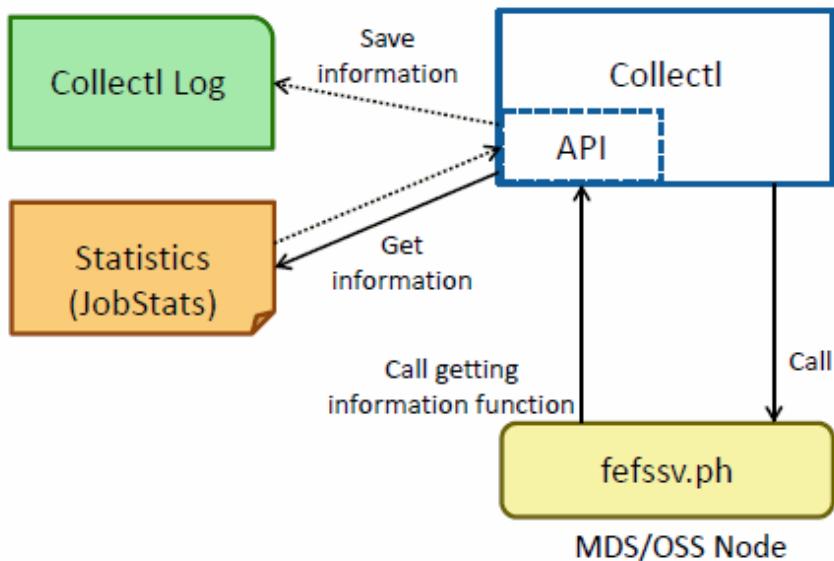
The purpose of this tool is to identify high-load jobs by finding out and aggregating various statistics about problems, such as server delay, occurring on each server.

The fefssv.ph script is called from collectl, which is a package widely used for monitoring purposes.

The JobStats function is used to collect information. For details on JobStats, see "[4.21 JobStats Function](#)."

The following figure shows the operation of this function.

Figure 2.17 fefsv.ph operation

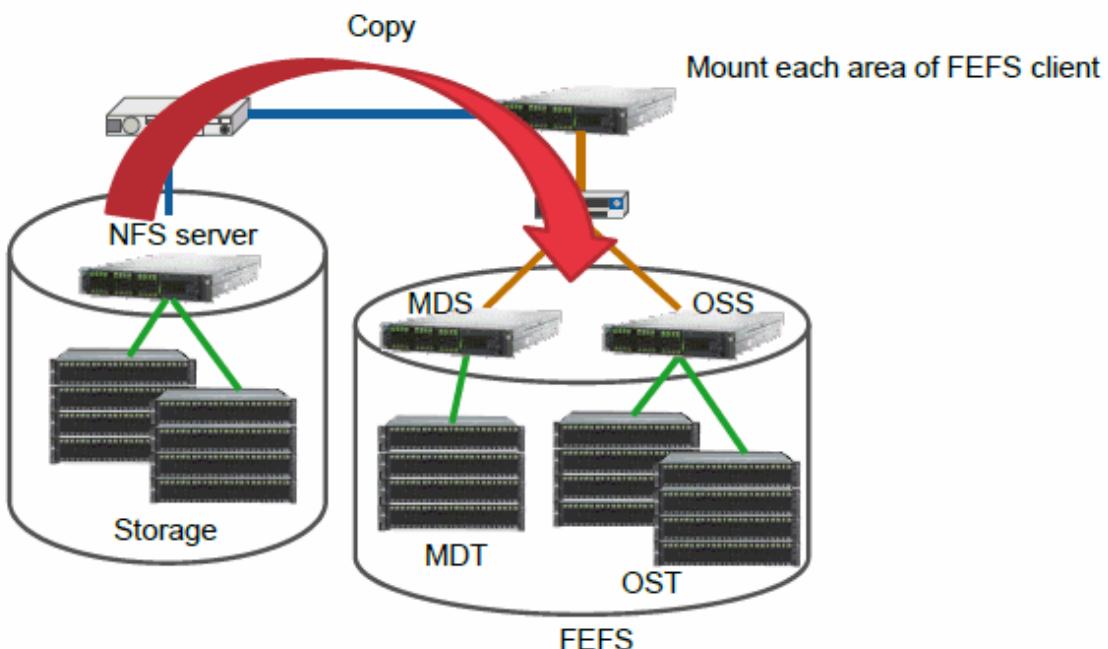


2.10 Function for Linkage with File Systems Other than FEFS

FEFS can work together with NFS and Lustre.

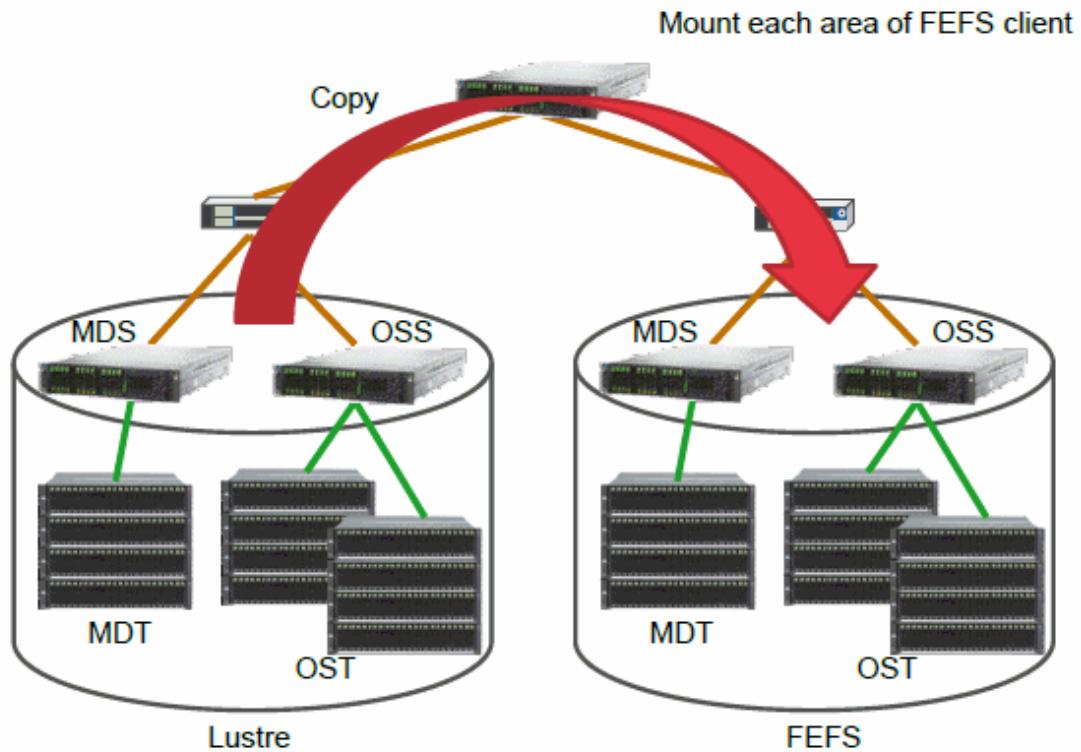
The following figures show examples of data migration from NFS to FEFS and from Lustre to FEFS.

Figure 2.18 Data migration from NFS to FEFS



NFS files are copied to FEFS via the network.

Figure 2.19 Data migration from Lustre to FEFS



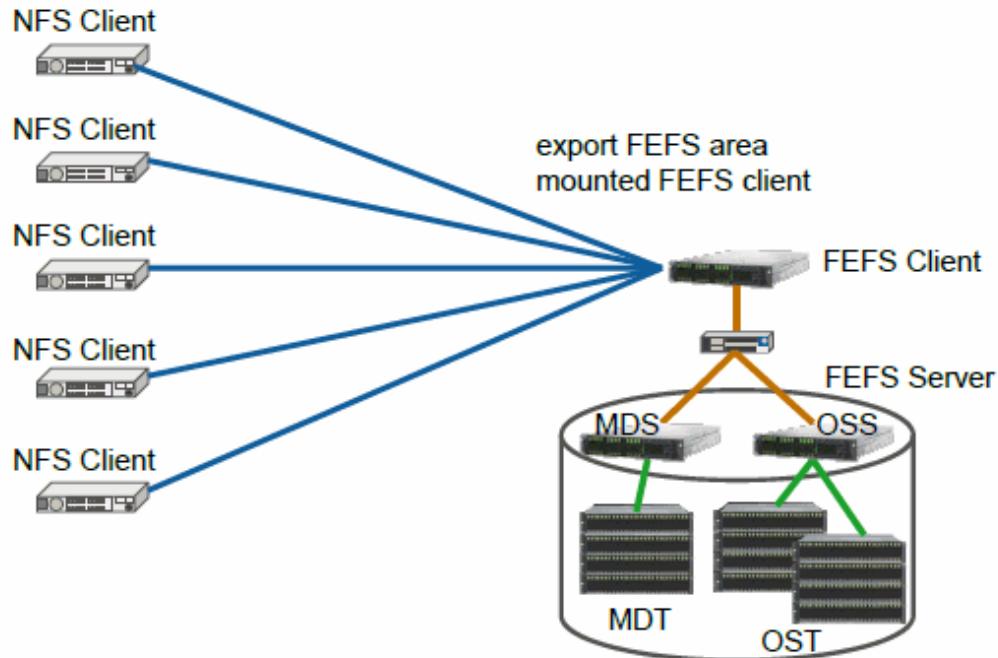
Lustre files are copied to FEFS via the network.

2.10.1 Publishing using NFS to external systems

FEFS mounted on the client node can be published in the external system by using NFS. To publish the FEFS file system provided by an FEFS server from the client using NFS, only one FEFS client can be the NFS server in order to ensure consistency and compatibility of the file data.

NFS version 3 and version 4 are supported for FEFS publishing using NFS.

Figure 2.20 Publishing using NFS



Information

FEFS satisfies the NFS requirements regarding the integrity and consistency of file data from NFS client nodes.

2.10.2 Connecting Lustre [PG]

The Lustre-compatible client provides functions for mounting a Lustre server from an FEFS client or mount an FEFS server from a Lustre client, and accessing files on the server.

The supported version for Lustre servers is 2.10.8.

Note

- An FEFS and a Lustre cannot be installed on the same node at the same time.
- Lustre server cannot be mounted from FEFS client of FX server.
- On a Lustre file system, the following functions cannot be used:
 - QoS function
 - LNet multirail function

Chapter 3 Method of Installation and Maintenance

This chapter describes how to install and maintain FEFS. The installation function of the system management installs software required for the system. The following explanations are provided assuming that the rpm package configuring FEFS (see the "Applying FEFS package") is configured in the appropriate node.

3.1 Installation Flow

Install FEFS in the following order.

1. Designing the FEFS configuration
2. Applying FEFS packages
3. Creating FEFS design sheets
4. Creating configuration definition files for the FEFS setup tool
5. Deploying the configuration definition files for the FEFS setup tool
6. Building FEFS
7. Changing the file system permission
8. Settings required after build

Note

Before you perform the step of "Building FEFS" and subsequent steps, you need to confirm that every nodes on which FEFS is installed, except for compute nodes, has started.

See

The following table lists node types specified during deployment.
For details on each node, see the *Job Operation Software Overview*.

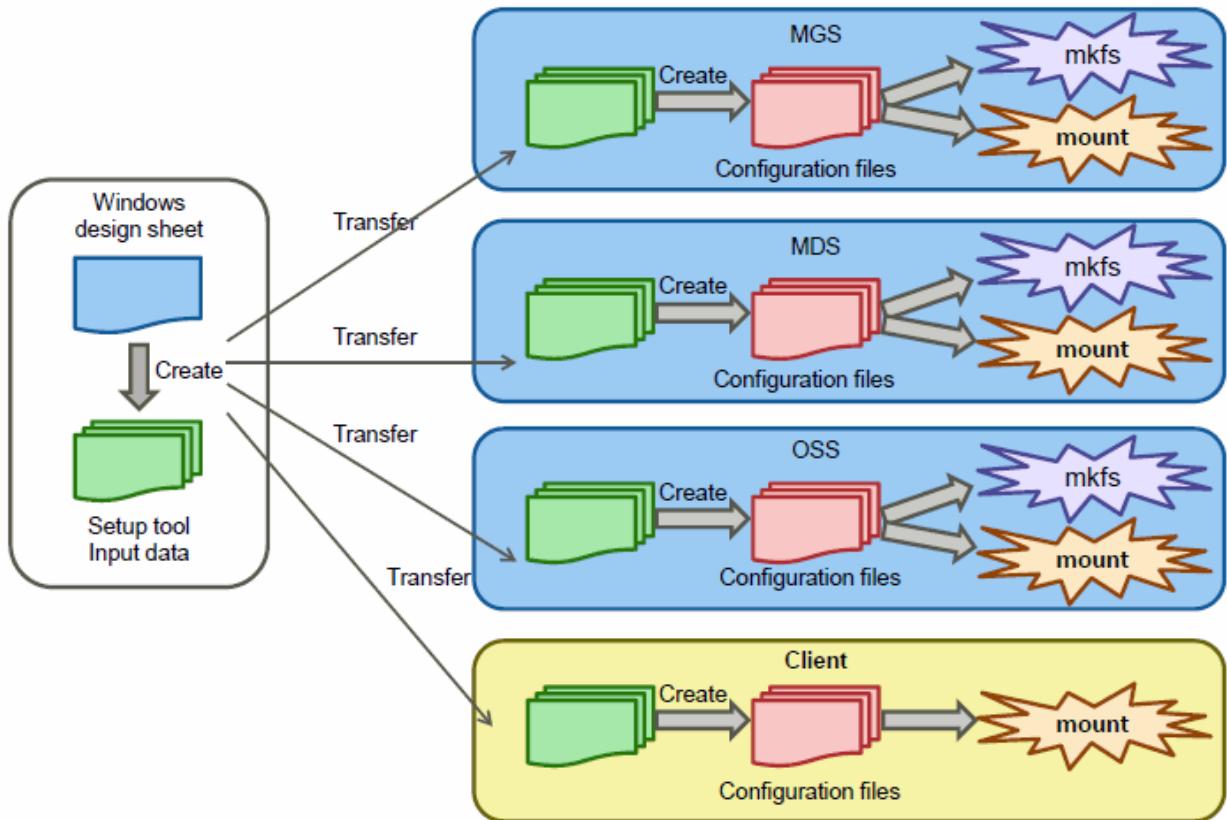
Table 3.1 Node Types

Type	Meaning
MGS	Management server
MDS	Meta data server
OSS	Object storage server
SMM	System management node
CCM	Compute cluster management node
LN	Login node
CN	Compute node [FX]
CN/BIO	Shared node of BIO and CN ("CN-BIO" in FEFS design sheets) [FX]
CN/GIO	Shared node of GIO and CN ("CN-GIO" in FEFS design sheets) [FX]
CN/SIO	Shared node of SIO and CN ("CN-SIO" in FEFS design sheets) [FX]

* For the handling of multiuse nodes, refer to "Note" in ["Figure 3.4 Example of entries in the NODE section"](#).

A conceptual image of the FEFS configuration definition files for the setup tool and the FEFS configuration files used in FEFS installation are shown as follows.

Figure 3.1 FEFS installation flow



3.1.1 Designing the FEFS Configuration

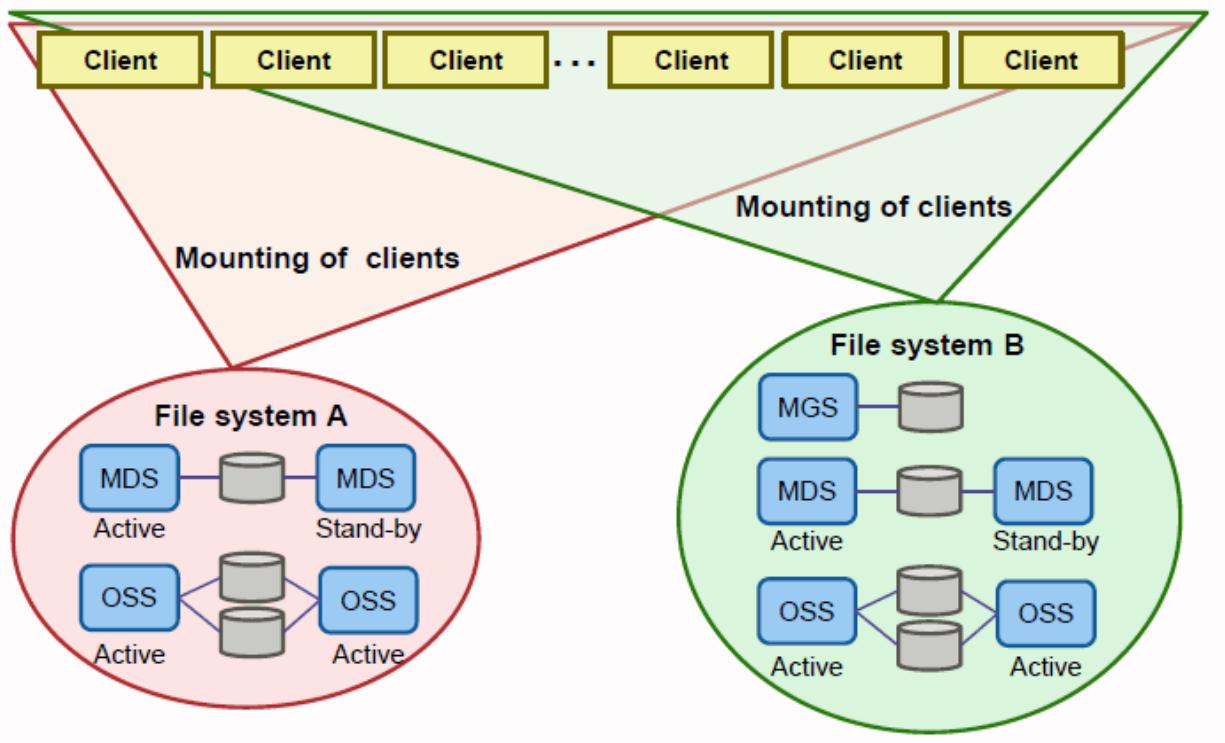
Determine the specifics of the FEFS configuration from the following aspects:

- Selecting functions and configurations
 - Whether the MDS and the OSS can use the external journaling function
 - Whether the QoS function can be used
 - Whether the ACL function can be used
 - Whether the QUOTA function can be used
- Determining configurations
 - Node configuration

Determine the MGS, MDS, OSS, and client configurations as shown in "[Figure 3.2 Determining the node configuration.](#)"

- Number of file systems
- Whether to make the MGS, MDS, and OSS redundant
- Number of clients
- Machine serving as both MGS and MDS

Figure 3.2 Determining the node configuration

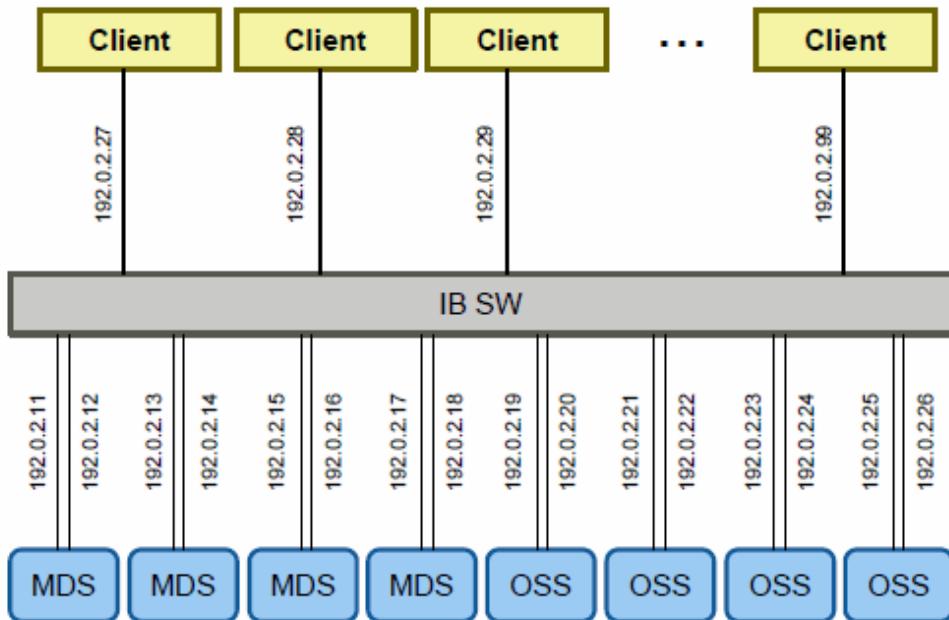


- Network configuration

Determine the InfiniBand configuration and IP addresses to be used as shown in "[Figure 3.3 Determining the network configuration.](#)"

- Number of InfiniBand connections to the MGS
- Number of InfiniBand connections to the MDS
- Number of InfiniBand connections to the OSS
- Number of InfiniBand connections to the client
- Number of InfiniBand connections to the CN/GIO
- IP addresses

Figure 3.3 Determining the network configuration



- Volume configuration

Determine the configuration of the volumes to be used.

- MDT-related volumes

MDT volume

MGT volume (MDS and MGS are another servers, or the redundant configuration of MDS to be Active/Active method)

- OST-related volumes

OST volume



See

When introducing LLIO, determine LLIO configuration as well. For details, see "3.1.1 Designing the LLIO configuration" of the *LLIO User's Guide*.



Information

You can identify volume information by the following method.

- Identify volume information

- ETERNUS multipath driver

Specify volumes by by-id name.

You can identify volume name by checking the symbolic link as follows on a node:

```
/dev/disk/by-id/scsi-3600000e00d000000002151900010000
```

- Device Mapper multipath

Specify volumes by /dev/mapper/<volumename>.

When using Device Mapper multipath, specify the WWID (World Wide Identifier) when the friendly name is invalid, and specify "mpathN" (mpatha, mpahb, etc.) when the friendly name is valid.

Other volume names are not supported.

You can identify volume name by executing the multipath -ll command.

- Friendly name is invalid:

```
# multipath -ll
3600000e00d000000001151a00000000 dm-0 FUJITSU,ETERNUS_DLX
size=818G features='1 queue_if_no_path' hwandler='0' wp=rw
|--- policy='round-robin 0' prio=10 status=enabled
| `-- 0:0:0:0 sda 8:0 active ready running
`--- policy='round-robin 0' prio=50 status=active
  `-- 2:0:0:0 sde 8:64 active ready running
```

- Friendly name is valid:

```
# multipath -ll
mpatha (3600000e00d000000001151a00000000) dm-0 FUJITSU,ETERNUS_DLX
size=398G features='1 queue_if_no_path' hwandler='0' wp=rw
|--- policy='round-robin 0' prio=10 status=enabled
| `-- 0:0:0:0 sda 8:0 active ready running
`--- policy='round-robin 0' prio=50 status=active
  `-- 2:0:0:0 sde 8:64 active ready running
```

- Identify storage device by volume name

For the method to identify storage device by volume name, see the manual of your storage device.

Example: Method to confirm device identification number on ETERNUS

- Confirm the device identification number and LUN (Logical Unit Number) on the node.

- ETERNUS multipath

```
/dev/disk/by-id/scsi-3600000e00d000000002151900010000
[021519:<device identification number>]
[0001:LUN]
```

- Device Mapper multipath

```
# multipath -ll
3600000e00d000000002151900010000 dm-1 FUJITSU,ETERNUS_DLX
[021519:<device identification number>]
[0001:LUN]
```

- Confirm the device identification number of ETERNUS.

Log in to ETERNUS with CLI (Command Line Interface), and get the device identification number from enclosure status.

```
CLI> show enclosure-status
Enclosure View
Name []
Model Upgrade Status [Not Possible]
Model Name [ET08E21B]
Serial Number [XXXXXXXXXX]
Device Identification Number [021519]
Status [Normal]
Cache Mode [Write Back Mode]
Remote Support [Not yet Set]
Operation Mode [Active]
CLI Connecting Controller Module [CM#0]
Firmware Version [V10L55-0000]
```

Controller Enclosure (2.5") [Undefined]
Drive Enclosure #1 (2.5") [Undefined]

.....

3.1.2 Applying FEFS packages

The package where FEFS is composed is as follows.

FEFS Server Packages

1. FJSVfefspogs-*x86_64.rpm
2. FJSVfefps-modules-*x86_64.rpm
3. FJSVfefps-osd-ldiskfs-modules-*x86_64.rpm
4. FJSVfefps-osd-ldiskfs-mount-*x86_64.rpm
5. FJSVfefps-*x86_64.rpm

(Note) "*" is replaced by the version number and the release name.

The following table lists the relation the packages and the nodes applied.

Table 3.2 FEFS server package and the node applied

Package name	Node type		
	MGS	MDS	OSS
FJSVfefspogs	o	o	o
FJSVfefps-modules	o	o	o
FJSVfefps-osd-ldiskfs-modules	o	o	o
FJSVfefps-osd-ldiskfs-mount	o	o	o
FJSVfefps	o	o	o

FEFS Client Packages [PG]

1. FJSVfefps-client-modules-*x86_64.rpm
2. FJSVfefps-client-*x86_64.rpm

(Note) "*" is replaced by the version number and the release name.

The following table lists the relation the packages and the nodes applied.

Table 3.3 FEFS client package and the node applied (PG client)

Package name	Node type				
	SMM	CCM	LN	CCS *1	Multiuse *2
FJSVfefps-client-modules	o	o	o	o	o
FJSVfefps-client	o	o	o	o	o

*1: Apply the package only when using a node serving as both CCS and CN.

*2: Apply the package only when using as an FEFS client.

FEFS Client Packages [FX]

1. FJSVfefps-client-modules-*aarch64.rpm
2. FJSVfefps-client-*aarch64.rpm

(Note) "*" is replaced by the version number and the release name.

The following table lists the relation the packages and the nodes applied.

Table 3.4 FEFS client package and the node applied (FX client)

Package name	Node type
	CN
FJSVfefefs-client-modules	o
FJSVfefefs-client	o



When introducing LLIO, determine LLIO configuration as well. For details, see "3.1.2 Applying the LLIO packages" of the *LLIO User's Guide*.

3.1.3 Creating FEFS design sheets

Create FEFS design sheets at your Windows station.

Use the FEFS design sheet templates included with the product. The file name of the FEFS design sheet is "FEFSDesignSheet.xlsx".

Before using them, enable the Excel macro function.

However, the entry items in red cells are the items requiring settings. Make sure to enter the values.



FEFS design sheet can be filled out in the following environment.

- Microsoft Windows 8.1, 10
- Microsoft Excel 2010, 2013, 2016

For an environment other than the above, consult with a Fujitsu system engineer (SE) or the Fujitsu Support Desk.

Create the following three types of sheets.

- NODE(1) sheet, NODE(2) sheet
These sheets define the configuration information independent of each file system such as network information.
Define the configuration information of the nodes using more than three InfiniBand in NODE(2) sheet.
- LLIO sheet
Define the configuration and device to use LLIO. LLIO is not used if not entered.
- GFS sheet

This sheet defines the configuration of each global file system. Create as many sheets as the number of file systems.

The following sections describe what is entered on each sheet.

3.1.3.1 Filling in a Node sheet

Set common information in each file system.

1. NODE section

List network information for all nodes relating to FEFS. If you use the installation function of Job Operation Software, import the node information to be inputted to the NODE (1) sheet and NODE (2) sheet from the information in the definition file of Job Operation Software.

For the import method, see "[Importing node information from the Job Operation Software](#)" below.

The following figure shows an example of entries in the NODE section.

Figure 3.4 Example of entries in the NODE section

HOSTNAME	NODETYPE	CLSTNAME	Primary Network			Secondary Network			Tofu Coord						Tofu IP ADDRESS	SYSTEM ID
			N/I	IP ADDRESS	PREFIX	N/I	IP ADDRESS	PREFIX	X	Y	Z	A	B	C		
mgs	MGS	storage	ib0	192.168.128.1	24	ib2	192.168.129.1	24								
mds1	MDS	storage	ib0	192.168.128.2	24	ib2	192.168.129.2	24								
mds2	MDS	storage	ib0	192.168.128.3	24	ib2	192.168.129.3	24								
mds3	MDS	storage	ib0	192.168.128.4	24	ib2	192.168.129.4	24								
mds4	MDS	storage	ib0	192.168.128.5	24	ib2	192.168.129.5	24								
oss1	OSS	storage	ib0	192.168.128.6	24	ib2	192.168.129.6	24								
oss2	OSS	storage	ib0	192.168.128.7	24	ib2	192.168.129.7	24								
oss3	OSS	storage	ib0	192.168.128.8	24	ib2	192.168.129.8	24								
oss4	OSS	storage	ib0	192.168.128.9	24	ib2	192.168.129.9	24								
oss5	OSS	storage	ib0	192.168.128.10	24	ib2	192.168.129.10	24								
oss6	OSS	storage	ib0	192.168.128.11	24	ib2	192.168.129.11	24								
oss7	OSS	storage	ib0	192.168.128.12	24	ib2	192.168.129.12	24								
oss8	OSS	storage	ib0	192.168.128.13	24	ib2	192.168.129.13	24								
ccm1	CCM	compute01	ib0	192.168.128.14	24	ib2	192.168.129.14	24								
ccm2	CCM	compute01	ib0	192.168.128.15	24	ib2	192.168.129.15	24								
login1	LN	compute01	ib0	192.168.128.16	24	ib2	192.168.129.16	24								
login2	LN	compute01	ib0	192.168.128.17	24	ib2	192.168.129.17	24								
client1	CN	compute01							0	0	0	0	1	0	10.0.0.1	
client2	CN	compute01							0	0	0	1	1	0	10.0.0.2	
client3	CN	compute01							0	0	0	1	1	1	10.0.0.3	

a. HOSTNAME

Specify the host name.

b. NODETYPE

Specify the node type. Select from the pull-down menu. For details on multiuse node, see the note shown below.

c. CLSTNAME

Define cluster name.

d. Primary Network, Secondary Network

e. Define InfiniBand information. Enter network interface, IP address and prefix.

f. Tofu Coord

Define Tofu coordinates (X, Y, Z, A, B, C).

g. Tofu IP ADDRESS

Define Tofu interconnect IP address.

h. SYSTEM ID

Use this ID when a node with the same Tofu coordinates exists. For details, see the note shown below.



For a node with one HCA, enter InfiniBand information in Primary Network.

For a node with a Tofu interconnect, enter the coordinates (X, Y, Z, A, B, C).

Nothing needs to be entered for the networks not used with FEFS.

Use either NODE (1) sheet or NODE (2) sheet depending on the number of Infiniband information. Use NODE (2) sheet only when there are three or more Infiniband information. For all other cases, use NODE (1) sheet. Add NODE (2) sheet by using an Excel macro.

Excel macro: [FEFS Design] > [Insert extra-IB node sheet]

NODETYPE of multiuse nodes

Technical Computing Suite has a node type called "multiuse node" (For details, see the following Job Operation Software manuals: *Overview* and *Setup Guide*). This node is not on the NODETYPE selection list.

For NODETYPE of multiuse node, set it with three to five characters, capital letters and numbers.

NODETYPE must be the same as the names defined in the design sheet of Job Operation Software.

For details on the naming rules, see the *Job Operation Software Setup Guide*.

However, you do not need to enter them when importing the node information in the system configuration design sheet of Job Operation Software.

Importing node information from the Job Operation Software

Import node information from the Job Operation Software to the NODE(1) sheet, and NODE(2) sheet.

1. Generating import information

Change the file format in order to import "node information definition file" and "node information definition file for FX server" into FEFS design sheet.

Execute the following command on the node to which the node information definition file and the node information definition file for FX server are placed.

```
# /sbin/fefs_yaml2csv <node information definition file> <output file>
# /sbin/fefs_yaml2csv <node information definition file for FX server> <output file>
```

The node information definition file and the node information definition file for FX server may be divided into multiple files. In such case, execute this command for each file as many times as the number of files.

For details on the "node information definition file" and the "node information definition file for FX server," see the *Job Operation Software for System Management*.

2. Selecting the files to import

Use an Excel macro to import generated importing information in FEFS design sheet.

Excel macro: [FEFS Design] > [Import installer data]

According to the dialogue output, import the files output in the section 1 above. When importing is completed, "Import Complete" appears.

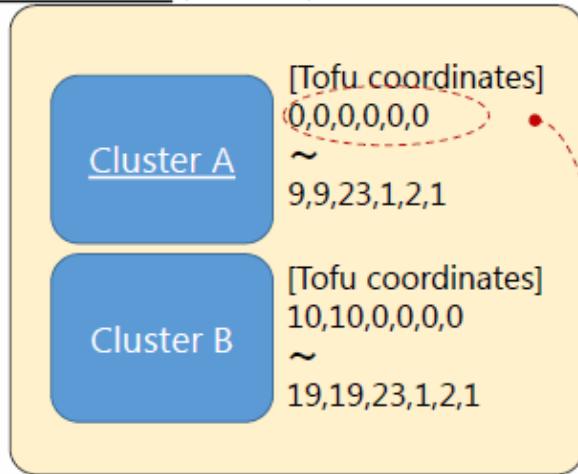
* When the node that does not use FEFS is contained, such as multiuse node, delete such node definition from the FEFS design sheet.

SYSTEM ID

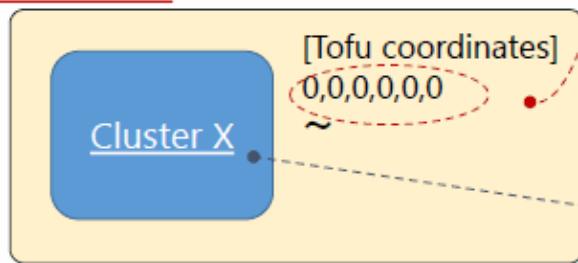
Normally, you do not need to enter the SYSTEM ID column of NODE (1) sheet. Use this ID when a node with the same Tofu coordinates exists. Also, you cannot use the same cluster name between different SYSTEM IDs.

Figure 3.5 Use of SYSTEM ID

SYSTEM ID = 1 (default)



SYSTEM ID = 2



Assign SYSTEM ID = 2 because

- there are nodes with the same Tofu coordinates

The name "Cluster A" "Cluster B"

- used with SYSTEM ID = 1 cannot be used.

3.1.3.2 Filling in an LLIO sheet

Set LLIO mount information and storage I/O node and first-layer storage device.

1. LLIO SETTING section

Figure 3.6 Example of entries in the LLIO SETTING section

■ LLIO SETTING	
FUNCTIONS	USE / UNUSE
Shared temporary	USE
Local temporary	USE
Global	USE

When using LLIO, specify USE for all items.

If you do not use LLIO, leave the initial value as UNUSE, and you do not need to set LLIO in the FEFS design sheet afterwards.

2. LLIO section

The following "Example of entries in the LLIO section" is provided along with an explanation.

Figure 3.7 Example of entries in the LLIO section

■ LLIO		
FUNCTIONS	MOUNT POINT	MOUNT OPTION
Shared temporary	/share	Configured
Local temporary	/local	Configured
Global	* Same as "MOUNT POINT [FEFS]" in GFS sheet *	Configured

- a. MOUNT POINT [Shared temporary]
Specify the mount point for shared temporary in the job.
- b. MOUNT OPTION [Shared temporary]
Specify the shared temporary mount option in a job. Normally, the option does not need to be changed.
- c. MOUNT POINT [Local temporary]
Specify the mount point for temporary in the node.
- d. MOUNT OPTION [Local temporary]
Specify the temporary mount option in a node. Normally, the option does not need to be changed.
- e. MOUNT OPTION [Global]
Specify the cache mount option for the second hierarchy storage in the job. Usually, this section is not necessary to change.
You cannot define the mount option without mount point. When it is defined properly, "Configured" is entered in the right of the entry column.

3. SIO section

The following "Example of entries in the SIO section" is provided along with an explanation.

Figure 3.8 Example of entries in the SIO section

SIO				
SIO HOSTNAME	SSD VOLUME	SSD SIZE(Mib)	MKFS OPTION	MOUNT OPTION
sio1	/dev/nvme0n1		819200 -f	pquota
sio2	/dev/nvme0n1			
sio3	/dev/nvme0n1			

- a. SIO HOSTNAME
Describe the storage I/O node host name.
- b. SSD VOLUME
Define the first-layer storage device path connected to that the storage I/O node.
- c. SSD SIZE (Mib)
Specify the device size used by the first-layer storage in units of Mib.
- d. MKFS OPTION
Specify the first-layer storage mkfs option. Usually, it is not necessary to change this section.
- e. MOUNT OPTION
Specify the mount option of first-layer storage device. Usually, it is not necessary to change this section.

4. MDT section

The following "Example of entries in the MDT section" is provided along with an explanation.

Figure 3.9 Example of entries in the MDT section

MDT	
NUMBER OF MDT	
0	

- a. NUMBER OF MDT
If you have dedicated MDTs for shared temporary use, specify the number of dedicated MDTs.
Specify 0 if you do not have a dedicated MDT for shared temporary. (Default 0)



For details on LLIO-related terms appearing in this section, see *LLIO User's Guide*.

3.1.3.3 Filling in a GFS sheet

Use this sheet when configuring a global file system. Use one sheet per file system for the settings. Accordingly, for the settings of multiple global file systems, add as many sheets for the settings as the number of additional file systems.

1. Adding a GFS sheet

Add a GFS sheet by using an Excel macro.

Excel macro: [FEFS Design] > [Insert global filesystem sheet]

2. FILESYSTEM section

The following example of entries in the FILESYSTEM section is provided along with an explanation.

Figure 3.10 Example of entries in the FILESYSTEM section

FILESYSTEM	
FSNAME	fefs01
MOUNT POINT [FEFS]	/fefs01

a. FSNAME

Set the file system name.



Note

The name must be unique in the system. Also, the file system name must consist of up to alphanumeric characters and eight characters.

b. MOUNT POINT

Specify the mount point for clients. Specify an absolute path.

3. MGS section

The following "Figure 3.11 Example of entries in the MGS section" is provided along with an explanation.

The MGS HOSTNAME (Active) and MGT VOLUME in this entry column must be filled in.

Figure 3.11 Example of entries in the MGS section

MGS						
MGS HOSTNAME (Active)	MGS HOSTNAME (Standby)	-	MGT VOLUME	-	MKFS OPTION	MOUNT OPTION
mgs		-	/dev/disk/by-id/scsi-3600000e00d110000001129ae00000000	-	--reformat --verbose	defaults,retry=6

a. MGS HOSTNAME (Active)

Specify the node name of the active MGS node where the MGT is to be mounted.



Note

For the MGS to double as an MDS, specify the same node name in MDS HOSTNAME(Active) in MDS section index0.

b. MGS HOSTNAME (Standby)

If the MGS is in an HA configuration, specify the node name of the corresponding standby MGS.



Note

For an MGS to double as the MDS, specify the same node name as MDS HOSTNAME (Standby) in MDS section index0.

c. MGT VOLUME

Specify the MGT volume.



Note

For an MGT to double as the MDT, specify the same volume name as the MGT in index0 in MDS section MDT VOLUME.

d. MKFS OPTION

Usually, this section is not necessary to change.

e. MOUNT OPTION

Usually, this section is not necessary to change.

4. MDS section

The following "Figure 3.12 Example of entries in the MDS section" is provided along with an explanation.

The MDS HOSTNAME(Active) and MDT VOLUME in index0 of this entry column must be filled in.

Figure 3.12 Example of entries in the MDS section

MDS							
MDS HOSTNAME (Active)	MDS HOSTNAME (Standby)	MDT INDEX	MDT VOLUME	JOURNAL VOLUME	MKFS OPTION	MOUNT OPTION	
mds1	mds2	0	/dev/disk/by-id/scsi-3600000e00d1100000011286800000000		-reformat--verbose	defaults,retry=6	
mds2	mds1	1	/dev/disk/by-id/scsi-3600000e00d110000001128680001000				
mds3	mds4	2	/dev/disk/by-id/scsi-3600000e00d110000001128680002000				
mds4	mds3	3	/dev/disk/by-id/scsi-3600000e00d110000001128680003000				

a. MDS HOSTNAME(Active)

Specify the node name of the active MDS where the MDT is mounted during normal operation.



For an MGS to double as the MDS, specify the same node name as MGS HOSTNAME(Active) in index0 in MDS HOSTNAME(Active).

b. MDS HOSTNAME(Standby)

If the MDS is in an HA configuration, specify the node name of the corresponding standby MDS.



For an MGS to double as the MDS, specify the same node name as MGS HOSTNAME(Standby) in index0 in MDS HOSTNAME(Standby).

c. MDT VOLUME

Specify the MDT volume.



For an MGT to double as the MDT, specify the same volume name as the MGT in index0 in MDT VOLUME.

d. JOURNAL VOLUME

To use an external journal, prepare and specify a volume for the journal.

e. MKFS OPTION

Usually, this section is not necessary to change.

f. MOUNT OPTION

Usually, this section is not necessary to change.

Add the options in the following cases.

- To enable the QoS function

Add the path of the QoS definition file. For details on settings to enable the QoS function, see "[3.2.1 Enable the QoS Function](#)".

- To enable the ACL function

Add the acl option. For details on settings to enable the ACL function, see "[Settings to Enable the ACL Function](#)".

5. OSS section

The following "Example of entries in the OSS section" is provided along with an explanation.

Figure 3.13 Example of entries in the OSS section

OSS							
OSS HOSTNAME (Active)	OSS HOSTNAME (Standby)	OST INDEX	OST VOLUME	JOURNAL VOLUME	MKFS OPTION	MOUNT OPTION	
oss1	oss2	0	0	1	-reformat -verbose	defaults,retry=6	
oss2	oss1	1	1	2			
oss3	oss4	2	2	3			
oss4	oss3	3	3				

a. OSS HOSTNAME(Active)

Specify the node name of the active OSS where the OST is mounted during normal operation.

b. OSS HOSTNAME(Standby)

If the OSS is in an HA configuration, specify the node name of the corresponding standby OSS.

c. OST VOLUME

Specify the OST volume.

d. JOURNAL VOLUME

To use an external journal, prepare and specify a volume for the journal.

6. CLIENT section

The following "Example of entries in the CLIENT section" is provided along with an explanation.

Figure 3.14 Example of entries in the CLIENT section

CLIENT	
MOUNT OPTION (CCM)	defaults,flock
MOUNT OPTION (LN)	defaults,flock
MOUNT OPTION (PG Client)	defaults,flock
MOUNT OPTION (FX Client)	defaults,flock

Specify the mount option of each node. Usually, this section is not necessary to change.

Add the options in the following cases.

- To enable the QoS function

Add the qos option. For details on settings to enable the QoS function, see "[3.2.1 Enable the QoS Function.](#)"

- To enable the user extended attribute

Add the user_xattr option. For details on settings to enable user extended attribute, see "[3.5 Settings to Enable the user Extended Attribute.](#)"

Rule for setting the client mount option of combined node and multiuse node

Set the option according to the rules below.

- CCS (when combined with client): Set to CLIENT OPTION (PG Client).

- Multiuse node: Set to CLIENT OPTION (LN).

7. FX CLIENT section

The following "Example of entries in the FX CLIENT section" is provided along with an explanation.

Figure 3.15 Example of entries in the FX CLIENT section

FX CLIENT		
SYSTEM	Tofu Coord	
ID	X	Y
	0	0
	0	1

a. SYSTEM ID

When multiple FX systems exist, input the system number. Input is not necessary in case of one system.

b. TOFU Coord

Input the list of TOFU coordinates (X, Y) of the main unit to which the client is mounted in FX.

8. PG CLIENT section

Specify the list of node names to which the client is mounted in PRIMERGY.

Figure 3.16 Example of entries in the PG CLIENT section

■ PG CLIENT
HOSTNAME
client1
client2
client3
login1
login2

3.1.3.4 Checking the input data

You can check the input data with an Excel macro. Confirm that the input data is correct and complete.

Excel macro: [FEFS Design] > [Check]

If the data is incorrect, the relevant locations are reported. In this case, correct the errors, and confirm the data again. When there is no problem, "OK:Check completed." appears.

3.1.4 Creating configuration definition files for the FEFS setup tool

You can create input data for setup with an Excel macro.

Excel macro: [FEFS Design] > [Create setup files]

Specify the output folder in the dialog box that appears. The macro creates configuration definition files for the FEFS setup tool in the specified folder.



When using LLIO, this operation also creates configuration definition files for LLIO.



You can create the FEFS design sheet from the configuration definition files for the FEFS setup tool (import function of configuration definition files for the FEFS setup tool).

Use the template for the FEFS design sheet contained in the product and execute the following Excel macro.

Excel macro: [FEFS Design] > [Import config files]

Specify the output folder (folder specified in Create config files) of configuration definition files for the FEFS setup tool according to the dialog box that appears. The macro imports configuration definition files for the FEFS setup tool in the FEFS design sheet.

Note the following during importing.

- The files are imported in the blank FEFS design sheet. You cannot add information to the FEFS design sheet with settings already input.
- When multiple file system settings exist, you can import the files in the FEFS design sheet in any order.

3.1.5 Deploying the configuration definition files for the FEFS setup tool

Deploy the configuration definition files for the FEFS setup tool in the following directory on active and standby system management node (referred to below as system management node) that has FEFS installed:

/etc/opt/FJSVfefs/config

3.1.6 Building FEFS

Use the fefs_sync command to build FEFS. For details on the fefs_sync command, see "[A.2.1 fefs_sync](#)."

Before you can perform the following operation, you need to ensure that all non-compute nodes where FEFS is installed are active.

In addition, before you can execute the following commands, you need to ensure that the Job Operation Software is installed.

Perform the following operation on the active system management node.



Information

For the procedure that recovers a node that has encountered an error during fefs_sync command execution, see "[3.11 Recovering a Node That Has Encountered an Error](#)."

To specify a multiuse cluster, specify the --compute option in the fefs_sync command.

1. Creating FEFS configuration files

Execute the following:

```
# fefs_sync --setup --storage=<cluster>[,...] --compute=<cluster>[,...]
```

Specify all clusters contained in the FEFS to be built.

--storage : Specify a storage cluster name.

--compute : Specify a compute cluster name or a multiuse cluster name.

2. Formatting MGT, MDTs and OSTs to initialize them

Format MGS, MDS, and OSS volumes.

Execute the following if formatting is required:

```
# fefs_sync --mkfs --storage=<cluster>[,...]
```

Specify all clusters contained in the FEFS to be built.

--storage : Specify a storage cluster name.

3. Starting FEFS service

Start FEFS service, and mount the MGS, MDS, OSS, CN, CCM, and LN.

Execute the following:

```
# fefs_sync --start --storage=<cluster>[,...] --compute=<cluster>[,...]
```

Specify all clusters contained in the FEFS to be built.

--storage : Specify a storage cluster name.

--compute : Specify a compute cluster name or a multiuse cluster name.



Note

In the configuration using LLIO, mounting with CN is not executed at the time the service is started. For details, refer to LLIO User's Guide.

4. Checking the FEFS status

Confirm that the FEFS service has normally started on the MGS, MDS, OSS, CN, CCM, and LN, using the pashowclst command.

Execute the following:

```
# pashowclst -v --nodetype MGS,MDS,OSS,CN,CCM,LN
```

If the FEFS status has changed to FEFSSR(o) and FEFS(o), the FEFS service has normally started.

3.1.7 Changing the file system permission

Set the permission for the mount point on one client.

Perform the work when the file system has been mounted.

(The default value is 755.)

3.1.8 Settings required after build

Scripts must be configured to prevent trouble in FEFS.

Use the procedures in "[Appendix C Settings Required after FEFS Build](#)."

3.1.9 Additional settings of the compute nodes

When building file system, even if some compute nodes could not be started because of hardware error, etc, you can build it later. Execute the following procedure on the active system management node.

1. Node isolation from operation

Make the setting of the target compute node disable.

Execute the following:

```
# paclstmgr -c <cluster> -n <nodeid> --disable
```

<cluster>: Specify the cluster that includes the node.

<nodeid>: Specify the node ID of the target node.

2. Starting the relevant node

Start the target compute node.

Execute the following:

```
# papwrctl -c <cluster> -n <nodeid> on
```

<cluster>: Specify the cluster that includes the node.

<nodeid>: Specify the node ID of the target node.

3. Creating FEFS configuration files

Create FEFS configuration files.

Execute the following:

```
# fefs_sync --setup --storage=<cluster>[,...] --compute=<cluster>[,...]
```

Specify all clusters contained in the FEFS to be built.

--storage : Specify a storage cluster name.

--compute : Specify a compute cluster name or a multiuse cluster name.

4. Restarting the relevant node

Restart the target compute node.

Execute the following:

```
# papwrctl -c <cluster> -n <nodeid> off  
# papwrctl -c <cluster> -n <nodeid> on
```

<cluster>: Specify the cluster that includes the node.
<nodeid>: Specify the node ID of the target node.

5. Node incorporation to operation

Make the setting of the target compute node enable.

Execute the following:

```
# paclstmgr -c <cluster> -n <nodeid> --enable
```

<cluster>: Specify the cluster that includes the node.
<nodeid>: Specify the node ID of the target node.

3.1.10 How to build FEFS on individual nodes

Perform the following procedure to build FEFS on individual nodes:

1. Creating FEFS design sheets

For details on performing this work, see "[3.1.3 Creating FEFS design sheets](#)."

2. Creating the configuration definition files for the FEFS setup tool

You can create input data for setup with an Excel macro.

Excel macro: [FEFS Design] > [Create setup files]

3. Deploying the configuration definition files for the FEFS setup tool

Deploy the configuration definition files for the FEFS setup tool in the following directory on the active and standby system management node:

/etc/opt/FJSVfefs/config

4. Creating FEFS configuration files

Execute the following on the system management node.

```
# fefs_sync --setup [--storage=<cluster> | --compute=<cluster>] --nodelist=<nodeidlist>
```

--storage : Specify the storage cluster name.
--compute : Specify the compute cluster name.
--nodelist : Specify a file that lists the node IDs of the target nodes.

5. Volume formatting

Perform this procedure if you are building FEFS on an MGS, MDS or OSS node.

Perform the following on the system management node if it is necessary to format the volume:

```
# fefs_sync --mkfs --storage=<cluster> --nodelist=<nodeidlist>
```

--storage : Specify the storage cluster name.
--nodelist : Specify a file that lists node IDs.

6. Starting FEFS

a. Starting the FEFS network

Execute the following on the system management node:

```
# fefs_sync --start [--storage=<cluster> | --compute=<cluster>] --nodelist=<nodeidlist>
```

--storage : Specify the storage cluster name.
--compute : Specify the compute cluster name.
--nodelist : Specify a file that lists node IDs.

b. Checking the FEFS status

Confirm that the FEFS service has started normally on the node using the pashowclst command.

```
# pashowclst -c <cluster> -n <nodeid>
```

Specify the cluster including the node and the node ID.

<cluster>: Specify the cluster that includes the node.

<nodeid>: Specify the node ID of the target node.

If the FEFS status has changed to FEFSSR(o) and FEFS(o), the FEFS service has normally started.



You can also specify a node with the --nodeid option. Refer to "[A.2.1 fefs_sync](#)" for details.

3.2 Settings to Enable the QoS Function

This section describes the setting to enable the QoS function.

3.2.1 Enable the QoS Function

The Qos function is enabled by setting up the QoS options in the FEFS design sheets on both the MDS and the FEFS client.

How to configure the MDS

Specify the QoS definition file path in the MOUNT OPTION field in the MDS section of the GFS sheet. This enables the QoS function. Refer to "[3.2.2 Settings of the QoS definition file](#)" for the method to create the QoS definition file.

Figure 3.17 Enabling the QoS function (MDS)

MDS						
MDS HOSTNAME (Active)	MDS HOSTNAME (Standby)	MDT INDEX	MDT VOLUME	JOURNAL VOLUME	MKFS OPTION	MOUNT OPTION
mds1	mds2	0	/dev/disk/by-id/scsi-3600000e00d1100000011286800000000		-reformat -verbose	defaults,retry=6,qosfile=/etc/opt/FJSVfefs/qosserver.conf
mds2	mds1	1	/dev/disk/by-id/scsi-3600000e00d1100000011286800001000			
mds3	mds4	2	/dev/disk/by-id/scsi-3600000e00d1100000011286800002000			
mds4	mds3	3	/dev/disk/by-id/scsi-3600000e00d1100000011286800003000			

For the procedure after creating the FEFS design sheet, see "[3.1.4 Creating configuration definition files for the FEFS setup tool](#)" and later.



Information

It is possible to enable the QoS function after MDS mounting as follows, however, we recommend the procedure for enabling during FEFS design sheet creation.

```
[MDS node]
# lctl qos on /etc/opt/FJSVfefs/qosserver.conf
```

For details on the lctl qos on command, see "[lctl qos](#)", which is a subcommand of the "[A.2.7 lctl command](#)".

How to configure an FEFS client

When creating FEFS design sheets, make settings in the CLIENT section of the GFS sheet. Specify the qos or qos_cache option in the MOUNT OPTION field to enable the QoS function. We recommend configuring the QoS function for clients only on clients (login nodes) that are shared by many users.

Figure 3.18 Enabling the QoS function (FEFS client)

CLIENT	
MOUNT OPTION (CCM)	defaults,flock
MOUNT OPTION (LN)	defaults,flock,qos,musermax=1,rd usermax=2,wrusermax=2
MOUNT OPTION (PG Client)	defaults,flock
MOUNT OPTION (FX Client)	defaults,flock

On the FEFS client, you can set the following qos options:

For the procedure after creating the FEFS design sheet, see "[3.1.4 Creating configuration definition files for the FEFS setup tool](#)" and later.

Table 3.5 qos options that can be set for FEFS clients

Option	Description
qos	Enables the QoS function (request control) on the client node. You can specify this option together with the qos_cache option.
qos_cache	Enables the QoS function (cache control) on the client node. You can specify this option together with the qos option.
noqos	Disables the QoS function on the client node (default).
musermax=<numerical-value>	Number of requests that one general user can issue simultaneously to the MDS. You can specify a numerical value in a range of 1 to 8. If omitted, the setting is 1. Specify this option together with the qos option.
rdusermax=<numerical-value>	Number of read requests that one general user can issue simultaneously to the OSS. You can specify a numerical value in a range of 1 to 16. If omitted, the setting is 2. Specify this option together with the qos option.
wrusermax=<numerical-value>	Number of write requests that one general user can issue simultaneously to the OSS. You can specify a numerical value in a range of 1 to 16. If omitted, the setting is 2. Specify this option together with the qos option.
mrootmax=<numerical-value>	Number of requests that a root user can issue simultaneously to the MDS. You can specify a numerical value in a range of 1 to 8. If omitted, the setting is 1. Specify this option together with the qos option.
rdrootmax=<numerical-value>	Number of read requests that a root user can issue simultaneously to the OSS. You can specify a numerical value in a range of 1 to 16. If omitted, the setting is 2. Specify this option together with the qos option.
wrrootmax=<numerical-value>	Number of write requests that a root user can issue simultaneously to the OSS. You can specify a numerical value in a range of 1 to 16. If omitted, the setting is 2. Specify this option together with the qos option.
mclientmax=<numerical-value>	Number of requests that can be issued simultaneously from within a client node to the MDS. You can specify a numerical value in a range of 1 to 16. If omitted, the setting is 4. Specify this option together with the qos option.
rdclientmax=<numerical-value>	Number of read requests that can be issued simultaneously from within a client node to the OSS. You can specify a numerical value in a range of 1 to 32. If omitted, the setting is 8. Specify this option together with the qos option.
wrclientmax=<numerical-value>	Number of write requests that can be issued simultaneously from within a client node to the OSS. You can specify a numerical value in a range of 1 to 32. If omitted, the setting is 8. Specify this option together with the qos option.
dpusermax=<numerical-value>	Percentage of the client cache on the client node that can be used by one general user. You can specify a numerical value in a range of 1 to 100. If omitted, the setting is 10. Specify this option together with the qos_cache option.

Option	Description
dpusermax=<numerical-value>	Percentage of the client cache on the client node that can be used by a root user. You can specify a numerical value in a range of 1 to 100. If omitted, the setting is 10. Specify this option together with the qos_cache option.

3.2.2 Settings of the QoS definition file

To use the QoS function, you need to create the QoS definition file /etc/opt/FJSVfef/qosserver.conf. This operation should be done by the administrator with root authority on the MDS. In case of multi MDS environment, execute on MDT0.



A sample QoS definition file is /etc/opt/FJSVfef/qosserver.conf.sample.

1. Creating the QoS definition file

Create the QoS definition file. The format of the QoS definition file is as follows.

```
MDS{
    item-name=setting-value
    item-name=setting-value
    (omitted)
OSS{
    item-name=setting-value
    item-name=setting-value
    (omitted)}
```



- The MDS and OSS sections must have these descriptions.
- Each line can consist of up to 1,024 characters including space characters and newline characters.
- A line that begins with the number sign (#) is a comment line. A comment cannot begin part way through a line.

An example of setting of the QoS definition file is as follows.

Example 1: Fair sharing between users

```
MDS{
    qos = on
# login node
    nodegrp1 = 30% 203.0.113.10, 203.0.113.20, 203.0.113.30
    usermax1 = 10%

# batch-job node
    nodegrp2 = 70% 192.0.2.[0-10], 198.51.100./*
    usermax2 = 20%
}
OSS{
    qos = same_mds
}
```

Example 2: Priority control of login nodes

```
MDS{
    qos = on
# login node
    nodegrp1 = 70% 203.0.113.10, 203.0.113.20, 203.0.113.30
```

```

# batch-job node
nodegrp2 = 30% 192.0.2.[0-10], 198.51.100. *
}
OSS{
qos = same_mds
}

```

The details of setting items are as follows.

Table 3.6 Items that can be written in the MDS section

Specification item	Description	Behavior when omitted
qos={on off }	on: The MDS performs QoS control. off: The MDS does not perform QoS control.	This is a required parameter and cannot be omitted.
nodegrp[1-10]=<numerical-value-1>%(<numerical-value-2>%)[ip-address-group]*1	Specify the maximum value (percentage) for the number of server threads assigned to a node group. Server threads up to the percentage specified by <numerical-value-1> are assigned. The range of <numerical-value-1> is 1 to 100. The total of <numerical-value-1> for nodegrp1 to nodegrp10 must be equal to or less than 100. If there are free server threads, server threads up to the percentage specified by <numerical-value-2> are assigned. The range of <numerical-value-2> is <numerical-value-1> to 100. If <numerical-value-2> is omitted, the same value as <numerical-value-1> is assumed specified. In this case, even if there are free server threads, server threads only up to the percentage specified by <numerical-value-1> are assigned. Specify the IO interconnect IP addresses of FEFS clients in [ip-address-group].	This is a required parameter, and at least one nodegrp specification is required.
usermax[1-10]=<numerical-value-1>%(<numerical-value-2>%)	Specify the maximum number (percentage) of server threads that can be assigned to a single general user in the node groups defined in the nodegrp specifications. The range of <numerical-value-1> is 1 to 100. If there are free server threads, server threads up to the percentage specified by <numerical-value-2> are assigned. The range of <numerical-value-2> is <numerical-value-1> to 100. If <numerical-value-2> is omitted, the same value as <numerical-value-1> is assumed specified. In this case, even if there are free server threads, server threads only up to the percentage specified by <numerical-value-1> are assigned.	100% is assumed specified.
rootmax[1-10]=<numerical-value-1>%(<numerical-value-2>%)	Specify the maximum number (percentage) of server threads that can be assigned to a root user in the node groups defined in the nodegrp specifications. The range of <numerical-value-1> is 1 to 100. If there are free server threads, server threads up to the percentage specified by <numerical-value-2> are assigned. The range of <numerical-value-2> is <numerical-value-1> to 100. If <numerical-value-2> is omitted, the same value as <numerical-value-1> is assumed specified. In this case, even if there are free server threads, server threads only up to the percentage specified by <numerical-value-1> are assigned.	100% is assumed specified.

*1 The ip-address-group format in nodegrp is as follows:

Format of ip-address-group of nodegrp

```
<ip-address-group> ::= <ip-range> { ,<ip-range>}
<ip-range> ::= <r-expr> ." <r-expr> ." <r-expr> ." <r-expr>
<r-expr> ::= <number> | "*" | "[" <r-list> "]"
<r-list> ::= <range> [ "," <r-list> ]
<range> ::= <number> [ "-" <number> [ "/" <stride> ] ]
<number> ::= "0-255"
<stride> ::= "1-255"
```

Examples of nodegrp definitions are as follows.

Example 1: IP addresses are specified one by one.

```
nodegrp1 = 30% 192.0.2.10, 192.0.2.27, 192.0.2.35
nodegrp2 = 70% 198.51.100.50, 198.51.100.55
```

Example 2: A range of IP addresses is specified.

```
nodegrp1 = 30% 192.0.2.[10-15]    -> Equivalent to 192.0.2.[10,11,12,13,14,15]
nodegrp2 = 70% 198.51.100.*       -> Equivalent to 198.51.100.[0-255]
```

Example 3: A range of IP addresses is specified with a constant interval.

```
nodegrp1 = 30% 192.0.2.[10-20/3]      -> Equivalent to 192.0.2.[10,13,16,19]
nodegrp2 = 70% 198.51.100.[50-100/10] -> Equivalent to 198.51.100.[50,60,70,80,90,100]
```



Note

If there is a repeated definition of an IP address in nodegrp1 to nodegrp10, the definition of the smaller nodegrp number has priority.

For example, for the following definitions, a request from 192.0.2.35 will be under QoS control of nodegrp1:

Example of wrong definitions

```
nodegrp1 = 30% 192.0.2.*
nodegrp2 = 70% 192.0.2.35
```

Write the following definitions for the correct QoS control of a request from 192.0.2.35:

Example of correct definitions

```
nodegrp1 = 70% 192.0.2.35
nodegrp2 = 30% 192.0.2.*
```

For the following definitions, a request from any FEFS client will be under QoS control of nodegrp1:

Example of wrong definitions

```
nodegrp1 = 70% *.*.*.*
nodegrp2 = 30% 192.0.2.*
```

Write the following definitions for the correct QoS control of a request from 192.0.2.*:

Example of correct definitions

```
nodegrp1 = 30% 192.0.2.*
nodegrp2 = 70% *.*.*.*
```

Table 3.7 Items that can be written in the OSS section

Specification item	Description	Behavior when omitted
qos={on off same_mds }	on: The OSS performs QoS control. off: The OSS does not perform QoS control. same_mds: The definition in the MDS section is used. (The definition in the OSS section is invalid.). However, for the parameter (load_limit_usec) that can only be specified in the OSS section, the definition in the OSS section is used.	This is a required parameter and cannot be omitted.
nodegrp[1-10]=<numerical-value-1>%(<numerical-value-2>%;) [ip-address-group] *1	Same as in the MDS section. See the description of the MDS section.	This is a required parameter, and at least one nodegrp specification is required.
usermax[1-10]=<numerical-value-1>%(<numerical-value-2>%)	Same as in the MDS section. See the description of the MDS section.	100% is assumed specified.
rootmax[1-10]=<numerical-value-1>%(<numerical-value-2>%)	Same as in the MDS section. See the description of the MDS section.	100% is assumed specified.
load_limit_usec=numerical-value	Specify the upper limit time in microseconds for one disk access to the OST. If the disk access time exceeds the value specified by this parameter, the result is as follows: QoS control will be exercised according to a value smaller than the specified usermax or rootmax value (percentage) for the number of server threads. You can specify a numerical value in a range of 0 to 100000000. If you specify 0, the number of threads will not be controlled according to the IO access time.	0 (invalid) is assumed to be specified.

*1 Describe ip address group of nodegrp referring to "Format of ip-address-group of nodegrp" above.

2. Checking the QoS definition file

Check the syntax of the created QoS definition file for errors. To check the QoS definition file, use lctl qos check command on the MDS.

Example 1: Where the syntax is correct

```
[MDS node]
# lctl qos check /etc/opt/FJSVfef/s qosserver.conf
QoS command was completed.
```

Example 2: Where the syntax is incorrect

```
[MDS node]
# lctl qos check /etc/opt/FJSVfef/s qosserver.conf
QoS config-file error. code=E_SEC_INVALID line=12
```

For details on the lctl qos check command, see "[lctl qos](#)", which is a subcommand of the "[A.2.7 lctl command](#)".

3. Enabling the QoS function

After confirming that the syntax in the QoS definition file is correct, enable the QoS function as shown in "[How to configure the MDS](#)" of "[3.2.1 Enable the QoS Function](#)".

3.3 Settings to Enable the File Lock

FEFS has the advisory lock function by the fcntl system call or the flock system call.

When you use this file lock, specify the flock option in the MOUNT OPTION field of the CLIENT section of the GFS sheet.

Figure 3.19 Specifying in the MOUNT OPTION

CLIENT	
MOUNT OPTION (CCM)	defaults,flock
MOUNT OPTION (LN)	defaults,flock
MOUNT OPTION (PG Client)	defaults,flock
MOUNT OPTION (FX Client)	defaults,flock

For the procedure after creating the FEFS design sheet, see "[3.1.4 Creating configuration definition files for the FEFS setup tool](#)" and later.

3.4 Settings to Enable the ACL Function

If you want to apply it to all file systems when creating FEFS design sheets, specify the acl option in the MOUNT OPTION filed of the MDS section of the GFS sheet to enable the ACL function.

Figure 3.20 Enabling the ACL function

MDS						
MDS HOSTNAME (Active)	MDS HOSTNAME (Standby)	MDT INDEX	MDT VOLUME	JOURNAL VOLUME	MKFS OPTION	MOUNT OPTION
mds1	mds2	0	/dev/disk/by-id/scsi-3600000e00d110000001128680000000		--reformat--verbose	defaults,retry=6,acl
mds2	mds1	1	/dev/disk/by-id/scsi-3600000e00d1100000011286800001000			
mds3	mds4	2	/dev/disk/by-id/scsi-3600000e00d1100000011286800002000			
mds4	mds3	3	/dev/disk/by-id/scsi-3600000e00d1100000011286800003000			

For the procedure after creating the FEFS design sheet, see "[3.1.4 Creating configuration definition files for the FEFS setup tool](#)" and later.

3.5 Settings to Enable the user Extended Attribute

The user extended attribute is information that is set and referred to by the system call setattr and the system call getxattr. The prefix indicating the namespace is "user." The user extended attribute is not available by default. If you want to use the user extended attribute, specify the user_xattr option in the MOUNT OPTION field of the CLIENT section of the GFS sheet when creating the FEFS design sheet.

Figure 3.21 Enabling the user extended attribute

CLIENT	
MOUNT OPTION (CCM)	defaults,flock,user_xattr
MOUNT OPTION (LN)	defaults,flock,user_xattr
MOUNT OPTION (PG Client)	defaults,flock,user_xattr
MOUNT OPTION (FX Client)	defaults,flock,user_xattr

For the procedure after creating the FEFS design sheet, see "[3.1.4 Creating configuration definition files for the FEFS setup tool](#)" and later.

3.6 Settings to Use the Failover Function

To use the MGS/MDS/OSS failover function, the following setting is required for preventing automatic restart at the node panic time.

For configuration instructions, refer to the documentation provided if you have installed Dump assist tool of the Fujitsu Linux Support Package.

If not, see the "Kernel Crash Dump Guide" published by Red Hat, Inc.

3.7 Operation during maintenance

If you want to change a module parameter after the FEFS has already been built, you need to once stop the FEFS and unload the module.

Stop the FEFS in the following sequence:

1. Advance preparation

Perform "Disconnecting from the operation" and "Transferring to the software maintenance mode." For details, see the "Preparation for software maintenance" in the *Job Operation Software Administrator's Guide for Maintenance*.

2. Stopping clients

Execute the following on every client.

```
# systemctl stop FJSVfef
```

3. Stopping OSSs

Execute the following on every OSS.

```
# systemctl stop FJSVfef
```

4. Stopping MDSs

Execute the following on every MDS.

```
# systemctl stop FJSVfef
```

5. Stopping MGSs

Execute the following on every MGS.

```
# systemctl stop FJSVfef
```

After unloading the module, change the FEFS parameters, and then restart the FEFS.

At this time, you do not need to reformat the file system if it is already formatted.

6. Incorporating into operation

Incorporate into operation the maintenance target that was disconnected from the operation at advance preparation. For details, see "Incorporating into operation after software maintenance" in the *Job Operation Software Administrator's Guide for Maintenance*.



Note

- In case of combined configuration of MGS and MDS, stop the node that mounts the MGT at the end.
- When the file system inconsistency occurs, you also need to stop FEFS to repair it. For details on repairing the file system, see "[4.8 Repairing File System Inconsistencies](#)".
- Stopping FEFS should be done after completing pasnap if it is being executed. If FEFS is stopped while executing pasnap, a node may panic.

3.8 Rolling Update

You can update the file system package without stopping the entire system.

For the package application, perform the following procedure.

For details on FEFS package type, see "[3.1.2 Applying FEFS packages](#)."

For details on LLIO package type, see the *LLIO User's Guide*.



Note

Check the availability and conditions of work in advance using the rpm -qi command for applied packages and the rpm -qpi command for packages to be applied. For details, see "Applying packages through rolling updates" in the *Job Operation Software Administrator's Guide for Maintenance*.

Advance preparation

Perform "Disconnecting from the operation" and "Transferring to the software maintenance mode." For details, see the "Preparation for software maintenance" in the *Job Operation Software Administrator's Guide for Maintenance*.

1. FEFS client package application

Perform the operation for each SIO group.

For details on the SIO group, see the *Job Operation Software Overview*.

- Stopping the FEFS service in the target range

Execute the following command in the system management node.

```
# fefs_sync --stop --compute=<cluster> --nodeid=<nodeid> --siogrp
```

--compute: Specify the compute cluster name.

--nodeid: The command is executed to the SIO group that includes <nodeid>.

- Maintenance and application operation

Perform package application in the range stopped in section a. above.

- Starting the FEFS service in the target range

Execute the following command in the system management node.

```
# fefs_sync --start --compute=<cluster> --nodeid=<nodeid> --siogrp
```

--compute: Specify the compute cluster name.

--nodeid: The command is executed to the SIO group that includes <nodeid>.



If the LLIO package is included in the target package, the LLIO package and the FEFS client package can be applied together as described above. To apply only the LLIO package, see "LLIO User's Guide".

2. FEFS server package application

When multiple file systems are configured, you can perform the maintenance for each file system.

- Unmounting the file system in the client

Unmount the file system subject to maintenance from the client.

Execute the following command in the system management node.

```
# fefs_sync --umount --compute=<cluster> --fsname=<fsname>
```

--compute: Specify the compute cluster name.

--fsname: Specify the name of the file system subject to maintenance.

- Stopping the maintenance target server

Stop the FEFS service in the maintenance target server.

Execute the following command in the system management node.

```
# fefs_sync --stop --storage=<cluster> --nodelist=<nodeidlist>
```

--storage: Specify the storage cluster name.

--nodelist: Specify the file listing the node IDs. List and specify the node ID of the maintenance target server.

- Maintenance and application operation

Perform package application in the range stopped in section b. above.

- Starting the maintenance target server

Start the FEFS service in the maintenance target server.

Execute the following command in the system management node.

```
# fefs_sync --start --storage=<cluster> --nodelist=<nodeidlist>
```

--storage: Specify the storage cluster name.

--nodelist: Specify the file listing the node IDs. List and specify the node ID of the maintenance target server.

- e. Mounting the file system in the client
 Remount the unmounted file system.
 Execute the following command in the system management node.

```
# fefs_sync --mount --compute=<cluster> --fsname=<fsname>
```

--compute: Specify the compute cluster name.
 --fsname: Specify the name of the maintenance target file system.

Incorporating into operation

Incorporate the maintenance target into the operation, which has been disconnected from the operation in the advance preparation. For details, see the "Incorporating into operation after software maintenance" in the *Job Operation Software Administrator's Guide for Maintenance*.

3.9 Configuring the FEFS Statistics Visualization Function

As described in "2.9 FEFS statistics visualization function (fefssv.ph script)," fefssv.ph is a script called from collectl when used. No particular setting is required because the fefssv.ph script is bundled with the FEFS package. You can use the fefssv.ph script by specifying it with the collectl option.



You will need to procure the collectl package separately and install it on the MDS and OSS.

If the following packages are installed, collectl log files are compressed and written out.

- perl-IO-Compress
- perl-Compress-Raw-Zlib
- perl-Compress-Raw-Bzip2

Only version 4.3.0 of collectl is supported.

3.10 Settings for NFS Publishing

To publish the FEFS, include the following options in the settings for NFS publishing in the /etc/exports file on the NFS server that has the FEFS.

Table 3.8 Required options in the settings for FEFS publishing

Option name	Remarks
fsid=num	Specify a non-zero 32-bit integer value in num. Always specify a unique value for each publishing point. We recommend setting a low value, which should be 1 or more.
no_root_squash	If no_root_squash has been used for export, an NFS client with root privileges can access resources in the file system. Therefore, adjust the NFS export settings to set a limit on the range of mountable clients so that mounting from an unintended node is not possible.

Also, the NFS server and the NFS client must have the same settings for account information, such as login user settings (e.g., user name, group name, uid, gid).

3.11 Recovering a Node That Has Encountered an Error

You can individually configure the node that has failed during the FEFS configuration or that has encountered an error. List the node IDs of the configuration target nodes in a file and specify to --nodelist option of the fefs_sync command.

```
# fefs_sync <operation> {--storage=<cluster> | --compute=<cluster>} --nodelist=<nodeidlist>
```

<operation>: Specify the operation options such as --start option and --stop option. For details, see "[A.2.1 fefs_sync](#)."
 <cluster>: Specify the cluster to which the failed node belongs.
 <nodeidlist>: Specify the file listing the node IDs.

When node failure occurs while building FEFS, the node IDs of the failed nodes are listed in the following file. To recover from the failure, save the following files and specify the saved files in the --nodelist option.

fefs_sync command name	File name
fefs_sync --setup	/var/opt/FJSVfefs/downnodeid_<cluster>_setup
fefs_sync --mkfs	/var/opt/FJSVfefs/downnodeid_<cluster>_mkfs
fefs_sync --mount	/var/opt/FJSVfefs/downnodeid_<cluster>_mount

<cluster> : Cluster to which failed nodes belong

If an error occurs in a node while building FEFS, the node IDs of the nodes with an error are listed in the following file. To recover from the error, save the following files and specify the saved files to fefs_sync --nodelist.

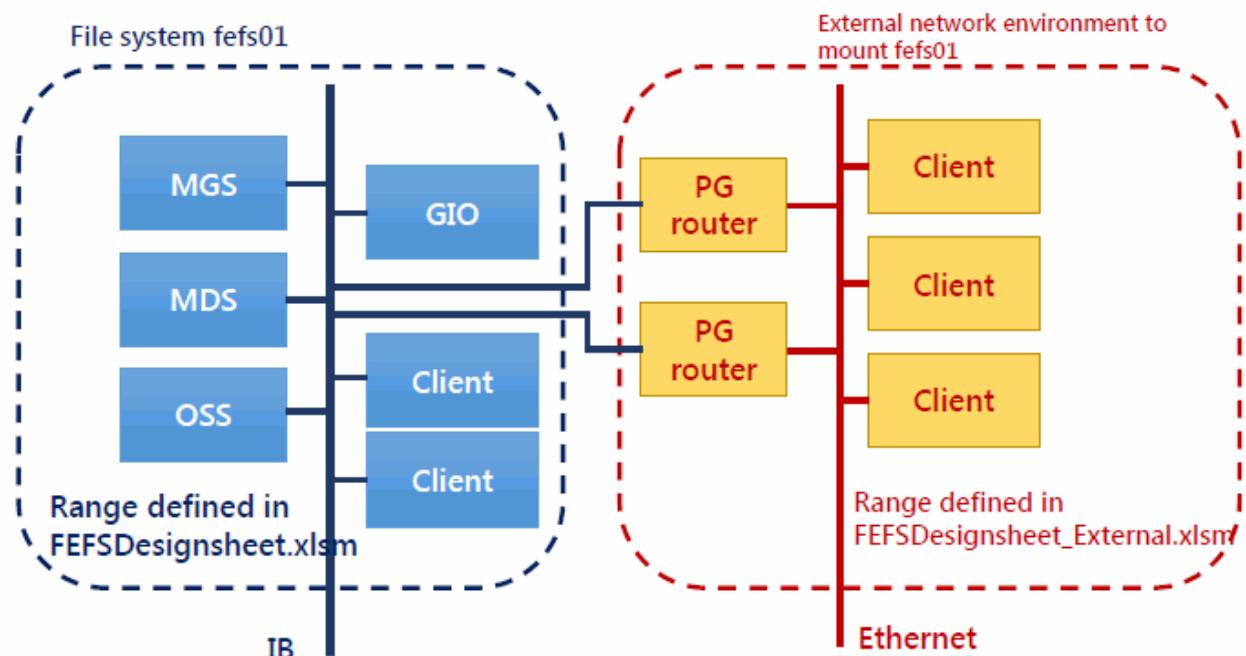
fefs_sync command name	File name
fefs_sync --setup	/var/opt/FJSVfefs/errornodeid_<cluster>_setup
fefs_sync --mkfs	/var/opt/FJSVfefs/errornodeid_<cluster>_mkfs
fefs_sync --mount	/var/opt/FJSVfefs/errornodeid_<cluster>_mount

<cluster> : Cluster to which nodes with an error belong

3.12 How to Configure FEFS in External Networks

You can use the FEFS from the node that is not managed with the Technical Computing Suite. A network that is not managed with the Technical Computing Suite is referred to as an external network. Ethernet is used for communication between FEFS and external networks. The following diagram shows the image of the system configuration that contains an external network.

Figure 3.22 Configuration image



The following section describes the settings required for using the FEFS from an external network client. Configure settings of the external network client and the router for external network connection (PG router).

To use this function, it is necessary to prepare a multiuse node as the router for external network connection.

3.12.1 Creating FEFS design sheets for external networks

Create the FEFS design sheet for an external network environment configuration (referred to below as "external FEFS design sheet") on the Windows terminal.

The template for the external FEFS design sheet is packaged in the product. The file name of the external FEFS design sheet is "FEFSDesignSheet_External.xlsx".

To start creating the external FEFS design sheet, enable the Excel macro function at first.



FEFS design sheet can be filled out in the following environment.

- Microsoft Windows 8.1, 10
- Microsoft Excel 2010, 2013, 2016

For an environment other than the above, consult with a Fujitsu system engineer (SE) or the Fujitsu Support Desk.

Create the following three kinds of sheets.

- NODE sheet
Define the configuration information independent of the file system such as network information.
- NET sheet
Define the router and client configuration.
- GFS sheet
Define the global file system configuration. Create as many as the number of file systems.

The following sections describe what to enter on each sheet.

3.12.1.1 Filling in a NODE sheet

Set common information in each file system.

1. List the network information for the target router and the client to configure in the external network. You do not need to define the node that is defined in the FEFS design sheet ("FEFSDesignSheet.xlsx").
2. The following shows an example of entries in the NODE sheet.

Figure 3.23 Example of entries in the NODE sheet

HOSTNAME	NODETYPE	CLSTNAME	IB(1)		IB(2)		Ethernet	
			N/I	IP ADDRESS	N/I	IP ADDRESS	N/I	IP ADDRESS
router01	MULTI	compute01	ib0	192.168.128.201	ib2	172.16.1.201		
router02	MULTI	compute01	ib0	192.168.128.202	ib2	172.16.1.202		
router03	MULTI	compute01	ib0	192.168.128.203			ens1f0	10.0.0.203
router04	MULTI	compute01	ib0	192.168.128.204			ens1f0	10.0.0.204
client001			ib0	172.16.1.101				
client002			ib0	172.16.1.102				
client003							ens1f0	10.0.0.103
client004							ens1f0	10.0.0.104

a. HOSTNAME

Define the host name.

b. NODETYPE

Define the node type.

Specify a multiuse node for the router node.

For details on multiuse nodes, see "[3.1.3.1 Filling in a Node sheet](#)".

The nodes outside the range of the Technical Computing Suite product are not necessary to enter.

c. CLSTNAME

Define the cluster name. The nodes outside the range of the Technical Computing Suite product are not necessary to enter.

d. IB(1), IB(2), Ethernet

Define the network interface and the IP address of the Infiniband or the Ethernet to use.



Note

You cannot use the LNet multirail function in the router.

For 1HCA node, enter the Infiniband information in IB(1).

In the node for the router, define information of both the network connected to the FEFS and the network connected to the client. You do not need to enter the network not used in the FEFS.

The node within the range of the Technical Computing Suite product needs to have the same name as defined in the design sheet for Job Operation Software. For the details on the naming rule, see the *Job Operation Software Setup Guide*.

3.12.1.2 Filling in a NET sheet

Define the router and client configuration.

The following shows an example of entries in the NET sheet.

Figure 3.24 Example of entries in the NET sheet

■ ROUTER				■ CLIENT	
HOSTNAME	LNET GROUP-ID	N/I		HOSTNAME	LNET GROUP-ID
		SERVER	CLIENT		
router01	3	ib0	ib2	client001	3
router02	3	ib0	ib2	client002	3
router03	8	ib0	ens1f0	client003	8
router04	8	ib0	ens1f0	client004	8

In the router section, define the node to be used as the router out of the nodes defined in the NODE sheet.

a. HOSTNAME

Define the node host name for the router.

b. LNET GROUP-ID

This is the identifier to define the network range. Define the different identifier for each network range. In addition, the identifier entered in the ROUTER section needs to be linked to the one entered in the CLIENT section.

The values that can be entered are from 1 to 65535. The values are not necessary to be serial or sequential.

c. N/I(SERVER,CLIENT)

Define the server network interface and the client network interface in the router node. Specify the network interface defined in the NODE sheet.

In the CLIENT section, define the node to be used as the client out of the nodes defined in the NODE sheet.

a. HOSTNAME

Define the client host name.

b. LNET GROUP-ID

This is the identifier to define the network range. For the client to be connected to the node defined in the ROUTER section, enter the same identifier as the router.

3.12.1.3 Filling in a GFS sheet

Set the global file system in the GFS sheet. Perform setting of one file system for one sheet. To set multiple file systems, add as many sheets as the number of file systems and set.

1. Adding the GFS sheet

Add the GFS sheet with an Excel macro.

Excel macro: [FEFS Design] > [Insert global filesystem sheet]

2. FILESYSTEM section

The following example of entries in the FILESYSTEM section is provided along with an explanation.

Figure 3.25 Example of entries in the FILESYSTEM section

FILESYSTEM		CLIENT	
FSNAME	fefs01	MOUNT OPTION	defaults,flock
MOUNT POINT	/fefs01		
PG CLIENT			
HOSTNAME			
client001			
client002			
client003			
client004			

a. FSNAME

Set the file system name.

The name is necessary to be same as the file system name defined in the FEFS design sheet ("FEFSDesignSheet.xlsx").

b. MOUNT POINT

Specify the client mount point. Specify the absolute path.

3. CLIENT section

Set the client mount option. Usually, this section is not necessary to change.

Add the option in the following cases.

- When enabling QoS function

Add the qos option. For the details on enabling QoS function, see "[3.2 Settings to Enable the QoS Function](#)."

- When enabling the user extended attribute

Add the user_xattr option. For details on settings to enable user extended attribute, see "[3.5 Settings to Enable the user Extended Attribute](#)."

4. PG CLIENT section

Define the client host name that mounts such file system.

3.12.1.4 Checking the input data

You can check the input data with an Excel macro below. Confirm that the input data is correct and complete.

Excel macro: [FEFS Design] > [Check]

If the data is incorrect, the relevant locations are reported. In this case, correct the errors, and confirm the data again. When there is no problem, "OK: Check completed." appears.

3.12.2 Creating configuration definition files for the FEFS setup tool

You can create input data for setup with an Excel macro below.

Excel macro: [FEFS Design] > [Create config files]

Specify the output folder in the dialog box that appears. The macro creates configuration definition files for the FEFS setup tool under the specified folder.

3.12.3 Deploying the configuration definition files for the FEFS setup tool

Deploy the configuration definition files for the FEFS setup tool in the following directory on active and standby system management node.

/etc/opt/FJSVfefs/config

3.12.4 Configuring the FEFS in the external network

Perform the configuration in two phases: the nodes within and outside the control of the Technical Computing Suite.

1. Setting the FEFS server and configuring the router (within the control of Technical Computing Suite)
2. Setting the client (outside the control of Technical Computing Suite)

As you cannot perform the configuration with the fefs_sync command in the client on the external network, perform the configuration individually on each node.

3.12.4.1 Setting the FEFS server and configuring the router

Use the fefs_sync command to set the FEFS server and configure the router. For details on the fefs_sync command, see "[A.2.1 fefs_sync](#)."

Before you execute the following commands, you need to ensure that Job Operation Software is installed.

Perform the following operation on the active system management node.



Note

If the FEFS building (building by using the FEFS design sheet "FEFSDesignSheet.xlsx") has not yet been completed, see "[3.1 Installation Flow](#)" and build FEFS in advance.



Information

For the procedure to recover a node that has encountered an error during fefs_sync command execution, see "[3.11 Recovering a Node That Has Encountered an Error](#)."

To specify multiuse cluster, specify the --compute option in the fefs_sync command.

1. Creating the FEFS setting file

Reflect the content set in "FEFSDesignSheet_External.xlsx" to the FEFS.

Execute the following command on the system management node node.

```
# fefs_sync --setup --storage=<cluster> --compute=<cluster>
```

Specify all clusters to build FEFS.

--storage : Specify the storage cluster name.

--compute : Specify the compute cluster name.

2. Starting the FEFS service

Start the FEFS service of the router node.

Execute the following command.

```
# fefs_sync --start --compute=<cluster> --nodelist=<nodeidlist>
```

Specify the cluster and the node ID of the router node.

--compute : Specify the compute cluster name.

--nodelist : Specify the file listing the node ID of the router node.

3. Checking the FEFS status

Confirm that the FEFS service is normally started in the router node with the pashowclst command.

Execute the following command.

```
# pashowclst -v --nodetype < node type of router node>
```

If the FEFS status transitions to FEFS(o), the FEFS service is normally started.

3.12.4.2 Setting the FEFS client

Perform settings of the FEFS client in the external network individually for each node. Perform the following command operations on all clients.

1. Creating the FEFS setting file

a. Deploying the FEFS configuration definition file

Distribute the configuration definition files for the FEFS setup tool on the system management node in the following directory of all client nodes.

```
/etc/opt/FJSVfefs/config
```

b. Creating the FEFS setting file

Execute the following command on the client node.

```
# fefsconfig --setup
```

2. Starting FEFS

a. Starting the FEFS service

Execute the following command on all client nodes.

```
# systemctl start FJSVfefs
```

b. Confirming the FEFS status

Confirm that the FEFS is mounted on all client nodes.

```
# mount
```

3.13 Precautions

The depletion of ARP cache

In a cluster of PRIMERGY node consisting of a large number of nodes, a communication from a single node to a large number of nodes may be generated. As a result, the number of ARP cache of the sender node is possibly insufficient as the default. When the ARP cache dries up, you fail in the address resolution, and FEFS cannot be accessed. To avoid such a situation, examine the tuning of the kernel parameter as follows.

- The nodes to be specified
MGS node, MDS node, and OSS node

- The kernel parameter to be specified

Specify the following kernel parameters concerning the number of ARP cache.

```
net.ipv4.neigh.default.gc_thresh3
```

- The value to be set to parameter

You can calculate the required number of ARP cache by using the following expression. Set a value larger than this value.

```
number of I/F(IP address) of receiver node x number of receiver node
```

When the value obtained above is larger than the default value(1024) of the kernel parameter net.ipv4.neigh.default.gc_thresh3 in RHEL, execute tuning. However, it is necessary to adjust it so as to avoid memory shortage.

The following is an example of calculating parameters.

```
When two Ethernet adapter are installed to the receiver node, the number of node is 1000, the number of ARP cache required for the sender node is given as follows.  
2 (Ethernet) x 1000 (number of receiver node) = 2000
```

To modify the kernel parameter, edit "/etc/sysctl.conf" file. For details, see the specifications of RHEL.

Environment variable LD_LIBRARY_PATH

If the common libraries required for program execution have not been placed in the current directory, do not include the current directory in the environment variable LD_LIBRARY_PATH.

When the current directory on FEFS includes a search path, unnecessary file access increases, resulting in a high access load on FEFS. The following examples include the current directory as a search target.

- (a) LD_LIBRARY_PATH=/usr/local/lib:
- (b) LD_LIBRARY_PATH=/usr/local/lib:..
- (c) LD_LIBRARY_PATH=/usr/local/lib:../usr/lib
- (d) LD_LIBRARY_PATH=:/usr/local/lib
- (e) LD_LIBRARY_PATH=/usr/local/lib::/usr/lib

As shown in (a), (d), and (e), the current directory may be a search target even in cases where the current directory (".") is not explicitly specified. For details, check the LD_LIBRARY_PATH specifications.

The number of accesses to the current directory varies depending on whether the current directory is set in the environment variable LD_LIBRARY_PATH. Examples are shown below.

Example 1: The current directory is not set in LD_LIBRARY_PATH

```
$ export LD_LIBRARY_PATH=
$ strace -e open dd if=/dev/zero of=./test.dat bs=4k count=1 |& grep ENOENT | wc -l
5
-> Open system calls that ended with ENOENT: 5
```

Example 2: The current directory is set in LD_LIBRARY_PATH

```
$ export LD_LIBRARY_PATH=.
$ strace -e open dd if=/dev/zero of=./test.dat bs=4k count=1 |& grep ENOENT | wc -l
17
-> Open system calls that ended with ENOENT: 17
```

In the above example, where the current directory is on FEFS, 12 (17 - 5) open system calls on FEFS ended with ENOENT. The above example shows execution on one node. In a case of execution on 100 compute nodes, the number of executed open system calls that end in an error with ENOENT is $12 \times 100 = 1200$.

Chapter 4 Method of Operation

This chapter describes how to operate FEFS.

4.1 Starting FEFS Servers and Clients

To start FEFS servers and clients, start nodes in the following order by node type.

FEFS is automatically mounted at the node start time.

1. MGS node, MDS node.
2. OSS node.
3. Client node.



Note

The run level cannot be changed while running the FEFS service. Change the run level after stopping the FEFS service, and start the FEFS service afterwards when you change the run level.



See

For details on the operation about starting nodes, see the following manual.

Job Operation Software Administrator's Guide for System Management

4.2 Stopping FEFS Servers and Clients

To stop FEFS servers and clients, stop nodes in the following order by node type.

1. Client node.
2. OSS node.
3. MGS node, MDS node.



Note

Stopping FEFS should be done after completing pasnap or fjsnap if it is being executed. If FEFS is stopped while executing pasnap or fjsnap, a node may panic.



See

For details on the operation about stopping nodes, see the following manual.

Job Operation Software Administrator's Guide for System Management

4.3 Configuring the Striping Function

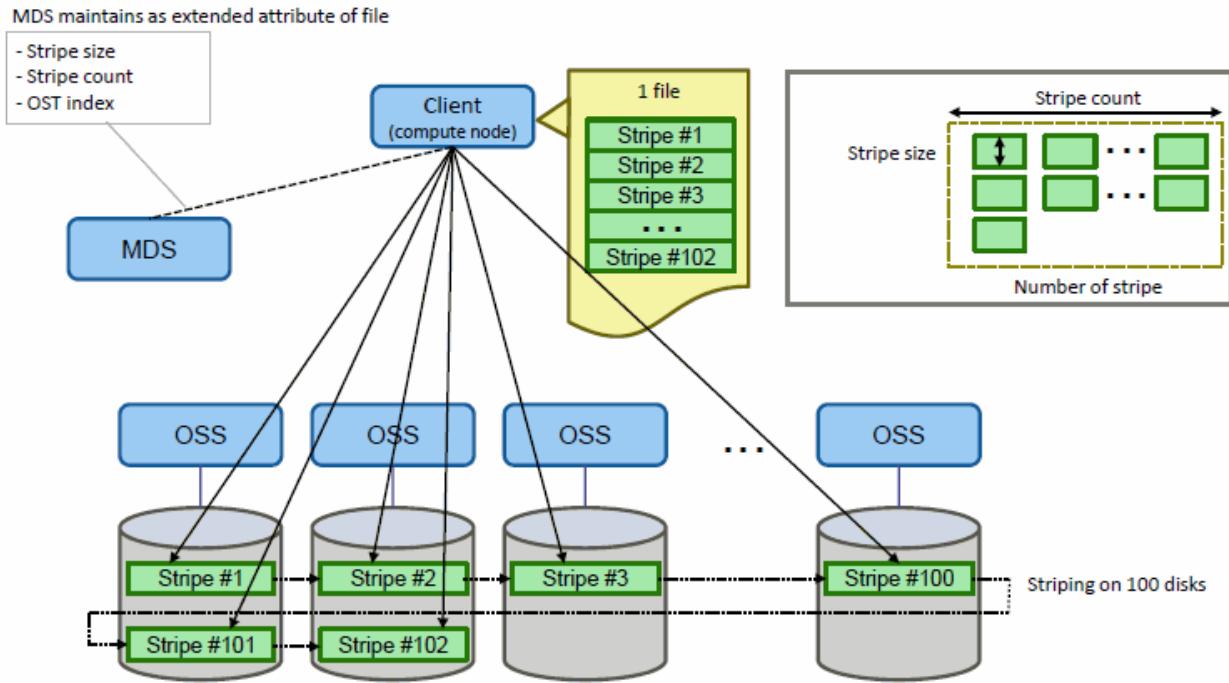
This section describes striping function settings.

4.3.1 How to configure stripes

Use the lfs setstripe command to configure stripes. Execute this command on a client node.

The following figure shows an example.

Figure 4.1 Example of stripes



With the lfs setstripe command, you can specify a size and range for the data to be distributed, and the number of units of data to be distributed.

- Stripe size

You can set a size for the data to be distributed.

- Stripe count

You can set an OST range for distributing data.

In "Figure 4.1 Example of stripes", stripes are allocated on 100 disks. This means the stripe count is 100. You can set a distribution range by specifying an OST index and an OST group named OST_pool.

For details on OST_pool settings, see "[4.3.3 How to configure an OST_pool](#)."

The lfs setstripe command creates a new file that has a stripe pattern. You can set the stripe pattern for an existing directory.

The following example sets a stripe pattern with a stripe size of 2 GiB and stripe count of 100 for the directory /fefs01. Stripes are allocated in an OST_pool named pool0001.

```
[Client node]
# lfs setstripe -S 2g -c 100 -p pool0001 /fefs01
```



- If the specified OST runs out of space, it may be set to another OST.
- The stripe size must be a multiple of 65536 bytes.



The OST index is available for reference in the information output by lfs df.

```
An error will occur if a file name or directory name containing a line feed exists [Client node]
# lfs df
  UUID          1K-blocks      Used   Available Use% Mounted on
gfefs0-MDT0000_UUID    189194456    130552   179298280   0% /feefs01[MDT:0]
gfefs0-OST0000_UUID    202702656    24792   192532828   0% /feefs01[OST:0]
gfefs0-OST0001_UUID    202702656    25304   192532312   0% /feefs01[OST:1]
gfefs0-OST0002_UUID    202702656    26328   192531288   0% /feefs01[OST:2]
gfefs0-OST0003_UUID    202702656    25816   192531804   0% /feefs01[OST:3]

filesystem summary:     810810624    102240   770128232   0% /feefs01
```

The rightmost number (blue character) on each line of the output results is the OST index number.



For details on the lfs setstripe command, see "[lfs setstripe](#)", which is a subcommand of the "[A.2.6 lfs command](#)".

4.3.2 How to confirm Stripe settings

Use the lfs getstripe command to confirm stripe settings. The command displays information on the stripe pattern of the specified file or directory. Execute this command on a client node.

The following example displays the stripe pattern of the directory /feefs01.

```
[Client node]
# lfs getstripe /feefs01
/feefs01
stripe_count:    100 stripe_size:    2147483648 stripe_offset: -1
```

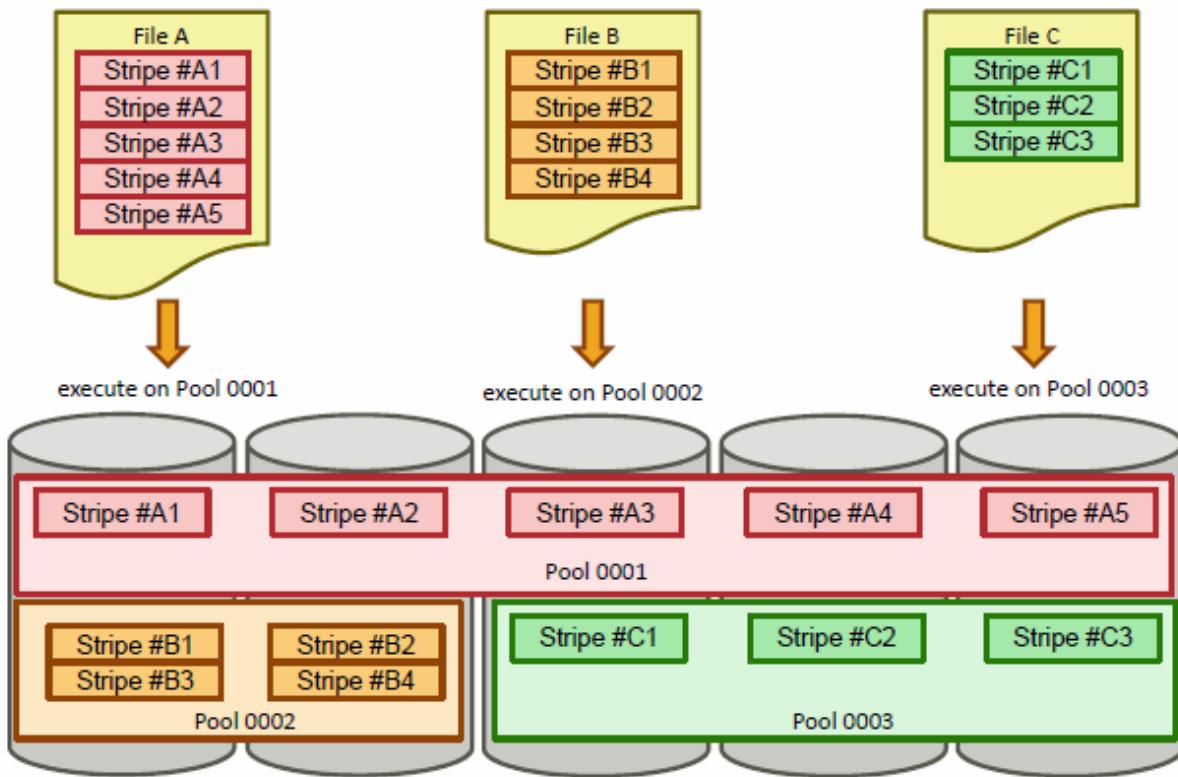


For details on the lfs getstripe command, see "[lfs getstripe](#)", which is a subcommand of the "[A.2.6 lfs command](#)".

4.3.3 How to configure an OST_pool

The OST_pool function is a function for grouping OSTs in advance. Use this function to specify a distribution range for the striping function.

Figure 4.2 OST_pool example



- Mounting clients on the MGS node (only if the MGS and MDS are different nodes)

If the MGS and MDS are different nodes, clients must be mounted on the MGS node before using the OST_pool function. From the MGS node, execute the command to mount clients on the MGS node.

The following command specifies 192.0.2.81 as the IB IP address of the MGS.

```
[MGS node]
# mount -t lustre 192.0.2.81@o2ib0:<fsname> /mnt/fefs
```

<fsname>: File system name

- Creating a new OST_pool

Use the lctl pool_new command to create a new OST_pool. Execute this command on an MDS node. However, if the MGS and MDS are different nodes, execute the command on the MGS node.

The following example creates an OST_pool named pool0001 on the file system fefs01.

```
[MGS/MDS node]
# lctl pool_new fefs01.pool0001
Pool fefs01.pool0001 created
```



For details on the lctl pool_new command, see "[lctl pool_new](#)", which is a subcommand of the "[A.2.7 lctl command](#)".

- Registering an OST in an OST_pool

Next, use the lctl pool_add command to register an OST. Execute this command on an MDS node. However, if the MGS and MDS are different nodes, execute the command on the MGS node.

The following example registers an OST in the OST_pool named pool0001.

```
[MGS/MDS node]
# lctl pool_add fefs01.pool0001 OST0000
OST TCS-OST0000_UUID added to pool fefs01.pool0001
```

The registered OST has an OST index ID of 0 in the OST_pool named pool0001.

OST names specified by this command corresponds to OST index IDs as follows.

OSTxxxx : OST index xxxx

When an OST index ID is 0, the OST name is OST0000. The same operation is repeated for all OSTs to be registered as members of pool0001.



For details on the lctl pool_add command, see "[lctl pool_add](#)", which is a subcommand of the "[A.2.7 lctl command](#)."

4. Displaying information on an OST_pool

Use the lctl pool_list command to display an OST_pool list and a list of the OSTs registered in an OST_pool. Execute this command on an MDS node. However, if the MGS and MDS are different nodes, execute the command on the MGS node.

The following example displays a list of the OST_pools registered at the mount point /fefs01.

```
[MGS/MDS node]
# lctl pool_list /fefs01
Pools from fefs01:
fefs01.pool0001
```

The following example displays a list of OSTs registered in the OST_pool named pool0001.

```
[MGS node]
# lctl pool_list fefs01.pool0001
```

The lfs pool_list command can also display the same information. Execute this command on a client node.

The following example displays a list of the OST_pools registered at the mount point /fefs01.

```
[Client node]
# lfs pool_list /fefs01
```

The following example displays a list of the OSTs registered in the OST_pool named pool0001.

```
[Client node]
# lfs pool_list fefs01.pool0001
```



For details on the lctl pool_list command, see "[lctl pool_list](#)", which is a subcommand of the "[A.2.7 lctl command](#)." Also For details on the lfs pool_list command, see "[lfs pool_list](#)", which is a subcommand of the "[A.2.6 lfs command](#)."

5. Deleting an OST from an OST_pool

Use the lctl pool_remove command to delete a registered OST entry from an OST_pool. Execute this command on an MDS node. However, if the MGS and MDS are different nodes, execute the command on the MGS node.

The following example deletes OST0000 registered in the OST_pool named pool0001.

```
[MGS/MDS node]
# lctl pool_remove fefs01.pool0001 OST0000
```



For details on the lctl pool_remove command, see "[lctl pool_remove](#)", which is a subcommand of the "[A.2.7 lctl command](#)."

6. Deleting an OST_pool

Use the lctl pool_destroy command to delete an OST_pool. Execute this command on an MDS node. However, if the MGS and MDS are different nodes, execute the command on the MGS node.

The following example deletes the OST_pool named pool0001.

```
[MGS/MDS node]
# lctl pool_destroy fefs01.pool0001
```



See

For details on the lctl pool_destroy command, see "[lctl pool_destroy](#)", which is a subcommand of the "[A.2.7 lctl command](#)".

7. Unmounting clients (only if the MGS and MDS are different nodes)

Unmount the clients mounted on the MGS node. From the MGS node, execute the command to unmount the clients from the MGS node.

```
[MGS node]
# umount /mnt/fefs
```



Note

Before executing a command to manipulate an OST_pool as described above, unmount all the mounted FEFS clients in the targeted FEFS server group.

4.4 How to Use Multi-MDS

4.4.1 Creating a remote directory

The following example displays how to specify MDT to create a directory in multi-MDS environment.

Assume that MDT1 and MDT 2 are index number 1 and index number 2 respectively.

```
[Client node]
# lfs mkdir -i 1 /mnt/fefs/usr
# lfs mkdir -i 2 /mnt/fefs/work
```

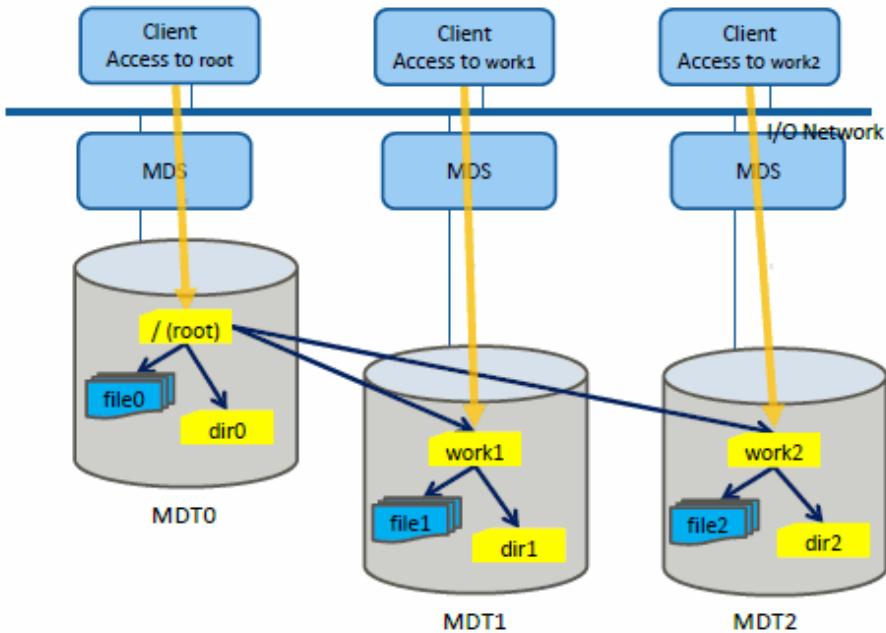
This command creates the directory /usr in MDT1 and the directory /work in MDT2.



Note

If a number larger than the MDT number is used as the specified value for -i, it is treated as if the MDT number was specified.

Figure 4.3 Creating directory in the specified MDT



4.4.2 Creating a striped directory

The following example displays how to create a stripe directory to distribute and deploy metadata in multiple MDTs in multi-MDS environment.

There are four target MDTs, assuming that they are MDT1, MDT2, MDT3, and MDT4 in the order of index numbers.

```
[Client node]
# lfs mkdir -c 4 -i 1 /mnt/fefs/dir1
```

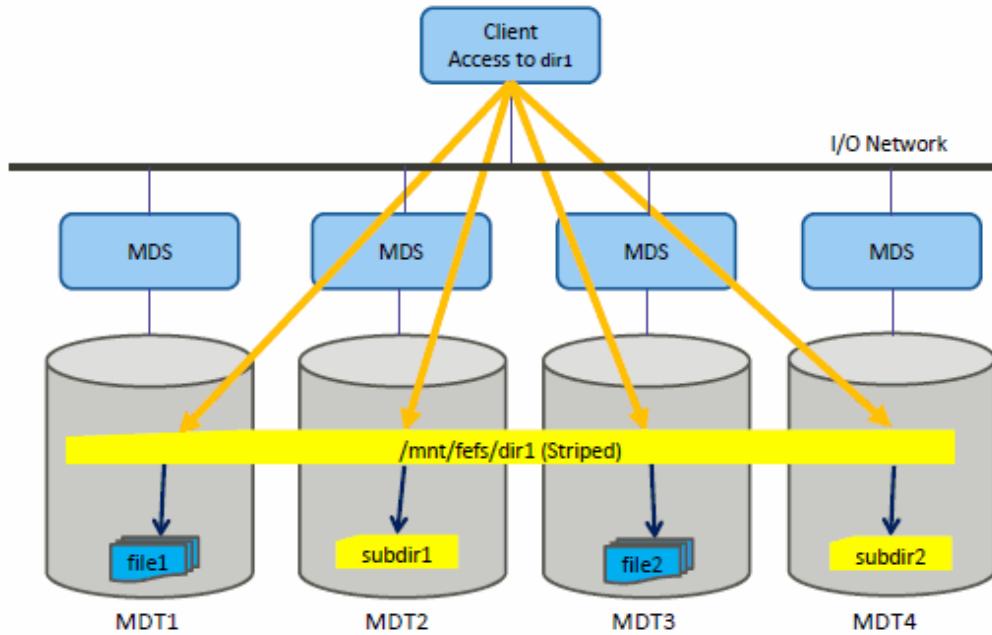
This command creates a striped directory in the four MDTs with MDT1 in the front.

After this, the files created under the directory are distributed and managed.



If a number larger than the number of MDTs is used as the specified value for -c or -i, it is treated as if the number of MDTs was specified.

Figure 4.4 Creating stripe directory



For details on the command, see "[lfs mkdir](#)," which is a subcommand of "[A.2.6 lfs command](#)."

Use the `lfs getdirstripe` command to confirm the directory striping information. The usage example is shown below.

```
[Client node]
$ lfs getdirstripe /mnt/fefs/dir1
lmv_stripe_count: 4 lmv_stripe_offset: 1 lmv_hash_type: fnv_1a_64
mdtidx      FID[seq:oid:ver]
1           [0x240000400:0x5:0x0]
2           [0x300000402:0x5:0x0]
3           [0x340000401:0x5:0x0]
0           [0x200000401:0x5:0x0]
```

The above means that the stripe count is 4 and the number of striped MDS starts with 1.



For details on the command, see "[lfs getdirstripe](#)," which is a subcommand of "[A.2.6 lfs command](#)."

4.5 Configuring the QUOTA Function

This section describes QUOTA function settings.

4.5.1 QUOTA setting for users and groups

This section describes how to configure the QUOTA function to files or directories. Unless otherwise noted, a user who has root privileges configures this function.

1. Preparation

Nothing is necessary for preparation.

2. How to set limit values

By enabling the QUOTA function, you can set limit values for each user or group. The following settings are an example of limit values for individual users and groups.

Table 4.1 Setting example of limit values for individual users

User name	Disk capacity	Number of files
user1	32 GiB	No limit
user2	No limit	10000
user3	64 GiB	5000

Grace times of the soft limit for users(blocks): 1000 seconds

Grace times of the soft limit for users(inodes): 3 days

Table 4.2 Setting example of limit values for individual groups

Group name	Disk capacity	Number of files
group1	512 GiB	No limit
group2	No limit	20000
group3	1 TiB	30000

Grace times of the soft limit for groups(blocks): 3 days

Grace times of the soft limit for groups(inodes): 1000 seconds



Note

The grace time for soft limit is set for all users or groups in common, not for individual users or groups.

As shown below, execute the lfs setquota command from an FEFS client to set the limit values. If you want unlimited disk space or files, specify "0" as the value.

This example assumes that FEFS is mounted at /mnt/fefs.

```
[Client node]
# lfs setquota -u user1 -B 33554432 /mnt/fefs          # user1 setting
# lfs setquota -u user2 -I 10000 /mnt/fefs            # user2 setting
# lfs setquota -u user3 -B 67108864 -I 5000 /mnt/fefs    # user3 setting
# lfs setquota -g group1 -B 536870912 /mnt/fefs        # group1 setting
# lfs setquota -g group2 -I 20000 /mnt/fefs            # group2 setting
# lfs setquota -g group3 -B 1073741824 -I 30000 /mnt/fefs # group3 setting
# lfs setquota -t -u -b 1000 -i 3d /mnt/fefs          # user soft limit grace times setting
# lfs setquota -t -g -b 3d -i 1000 /mnt/fefs           # group soft limit grace times setting
```

*Specify the disk capacity limit in KiB units.

Similarly, you can change limit values by executing the lfs setquota command.



See

For details on the command, see "[lfs setquota](#)", which is a subcommand of the "[A.2.6 lfs command](#)".



Note

- When you set again the value exceeding upper limit value while the QUOTA setting has reached the upper limit, you may not be able to create the file exceeding the first upper limit value. In such case, set 0 at first and then set the value again. This enables the set value.

- When there is 1MDT node in MDS or 1OST node in OSS, the QUOTA limit values are not enabled by default. To enable the value, execute the following command on the MGS node after confirming the server mount:

```
# lctl conf_param <fsname>.quota.ost=ugd
# lctl conf_param <fsname>.quota.mdt=ugd
```

<fsname>: File system name

3. How to confirm limit values

You can confirm limit values by executing the lfs quota command. This command is executed on the client node.

The following example shows how to confirm the limit for the user name: user1. In this case, specify the user name with the -u option of the lfs quota command.

```
[Client node]
# lfs quota -u user1 /mnt/fefs
Disk quotas for user user1 (uid 2000):
Filesystem kbytes quota limit grace files quota limit grace
/mnt/fefs 0 0 33554432 - 0 0 0 -
```

The following example shows how to confirm the limit for the your own user (user name: user1, group name to which you belong: group1). In this case, execute the lfs quota command as follows. No options are required.

```
[Client node]
$ lfs quota /mnt/fefs # Your user name is user1, and group name is group1
Disk quotas for user user1 (uid 2000):
Filesystem kbytes quota limit grace files quota limit grace
/mnt/fefs 0 0 33554432 - 0 0 0 -
Disk quotas for group group1 (gid 2000):
Filesystem kbytes quota limit grace files quota limit grace
/mnt/fefs 0 0 536870912 - 0 0 0 -
```

The following example shows how to confirm the grace times of the soft limit. In this case, specify the -t option to the lfs quota command as follows:

```
[Client node]
$ lfs quota -t -u /mnt/fefs # user soft limit grace times
Block grace time: 16m40s; Inode grace time: 3d
$ lfs quota -t -g /mnt/fefs # group soft limit grace times
Block grace time: 3d; Inode grace time: 16m40s
```



For details on the lfs quota command, see "[lfs quota](#)", which is a subcommand of the "[A.2.6 lfs command](#)".

4. Procedure for disabling the QUOTA function

Disable the QUOTA function by using the following procedure.

1. On the FEFS client, write down the limits set for all users and groups.
2. On the FEFS client, set each limit to 0 for all users and groups by using "[lfs setquota](#)."
To enable the QUOTA function again, set the values that you wrote down in step 1 by using "[lfs setquota](#)."

4.5.2 QUOTA settings for the projects

This section describes how to configure the Project QUOTA function. Unless otherwise noted, a user who has root privileges configures this function.

1. Preparation

Use the lfs project command to set the project ID and inheritance flags. If inheritance flags are set in the directory, the directory created underneath inherits the project ID and the inheritance flags of the parent directory.

The following example shows the procedure to set the project ID 1000 to the directory /mnt/fefs/dir1 which sets the project ID and the inheritance flags to all directories underneath.

```
[Client node]
# lfs project -srp 1000 /mnt/fefs/dir1 # set project ID 1000 and inheritance flags to all under
                                         # directory /mnt/fefs/dir1
# lfs project -d /mnt/fefs/dir1          # display the project setting of directory /mnt/fefs/dir1
1000 P /mnt/fefs/dir1
# lfs project -r /mnt/fefs/dir1          # display the project setting recursively under
                                         # directory /mnt/fefs/dir1
1000 P /mnt/fefs/dir1/subdir11          # directory /mnt/fefs/dir1
1000 P /mnt/fefs/dir1/file11
1000 P /mnt/fefs/dir1/subdir11/subdir21
1000 P /mnt/fefs/dir1/subdir11/file21
```

Note

- Without the -s option, the lfs project command does not set the project ID when creating the file directories under the directory afterwards and the project ID is not counted in the project QUOTA count of the parent directory.
- When you execute the lfs project command again to the file or directory to which the project ID has already been set, the command overwrites the project ID value.

See

For details on the lfs project command, see "[lfs project](#)," which is a subcommand of "[A.2.6 lfs command](#)."

2. How to set limit values

When you enable the project QUOTA function and relate the project to the directory, you can set limit values such as disk capacity and the number of files for each directory.

The following settings are an example of limit values for individual directories.

Table 4.3 Example of setting limit values for each project

Project ID	Directory	Disk capacity	Number of files
1000	/mnt/fefs/dir1	32 GiB	No limit
2000	/mnt/fefs/dir2	No limit	10000
3000	/mnt/fefs/dir3	64 GiB	5000

Grace times of the soft limit(blocks): 7200 seconds

Grace times of the soft limit(inodes): 1 day

Note

The grace time for the project QUOTA soft limit is set for each file system, not for each project.

As shown below, execute the lfs setquota command from an FEFS client to set the limit values.

This example assumes that FEFS is mounted at /mnt/fefs.

```
[Client node]
# mkdir /mnt/fefs/dir1
# lfs project -srp 1000 /mnt/fefs/dir1
# create directory /mnt/fefs/dir1
# set project ID 1000 to directory /mnt/
fefs/dir1
# lfs setquota -p 1000 -B 33554432 /mnt/fefs
# set QUOTA to project ID 1000
# mkdir /mnt/fefs/dir2
# lfs project -srp 2000 /mnt/fefs/dir2
# create directory /mnt/fefs/dir2
# set project ID 2000 to directory /mnt/
fefs/dir2
```

```

# lfs setquota -p 2000 -I 10000 /mnt/fefs          # set QUOTA to project ID 2000
# mkdir /mnt/fefs/dir3                           # create directory /mnt/fefs/dir3
# lfs project -srp 3000 /mnt/fefs/dir3          # set project ID 3000 to directory /mnt/
fefs/dir3
# lfs setquota -p 3000 -B 67108864 -I 5000 /mnt/fefs    # set QUOTA to project ID 3000
# lfs setquota -t -p -b 7200 -i 1d /mnt/fefs        # set grace times of project QUOTA soft
limit

```

*Specify the disk capacity limit in KiB units.

Similarly, you can change limit values too by executing the lfs setquota command.

Note

- When you set again the value exceeding upper limit value while the QUOTA setting has reached the upper limit, you may not be able to create the file exceeding the first upper limit value. In such case, set 0 at first and then set the value again. This enables the set value.
- When there is 1MDT node in MDS or 1OST node in OSS, the QUOTA limit values are not enabled by default. To enable the value, execute the following command on the MGS node after confirming the server mount:

```

# lctl conf_param <fsname>.quota.ost=ugd
# lctl conf_param <fsname>.quota.mdt=ugd

```

<fsname>: File system name

For details on the command, see "[lctl conf_param](#).", which is a subcommand of the "[A.2.7 lctl command](#)".

This command only needs to be executed once per file system, and there is no need to delete the set value.

See

For details on the command, see "[A.2.6 lfs command](#).", which is a subcommand of the "[A.2.6 lfs command](#)".

Information

We recommend the operation of creating empty directory and setting the project QUOTA and inheritance flags to transfer the files in that directory, instead of setting the entire project QUOTA in the existing directory. That is because the this operation is beneficial in terms of load to the system.

3. How to confirm limit values

You can confirm limit values of the project QUOTA function by executing the lfs quota command. Execute this command on the client node.

The following example shows how to confirm limit values of the project ID 1000. In this case, specify the directory path name with the -p option of the lfs quota command.

```

[Client node]
$ lfs quota -p 1000 /mnt/fefs
Disk quotas for prj 1000 (pid 1000):
  Filesystem  kbytes  quota  limit  grace   files  quota  limit  grace
  /mnt/fefs      4       0  33554432     -       1       0       0     -

```

The following example shows how to confirm grace times of the soft limit. In this case, specify the -t option to the lfs quota command.

```

[Client node]
$ lfs quota -t -p /mnt/fefs      # Project QUOTA soft limit grace times
Block grace time: 2h; Inode grace time: 1d

```



For details on the command, see "[lfs quota](#)", which is a subcommand of the "[A.2.6 lfs command](#)".

4. How to remove a project ID from a file

To disable the project QUOTA function, execute the lfs project-C command to clear the project ID. The following example shows that the project IDs under the directory /mnt/fefs/dir1 are all cleared.

```
[Client node]
# lfs project -C -r /mnt/fefs/dir1
```

4.6 Configuring the QoS Function

4.6.1 Checking the QoS status on the FEFS client

The following examples show how to confirm the QoS function options specified in the FEFS design sheet.

ex1) Confirming the QoS function

If qos is specified in the FEFS design sheet, 1 is displayed. If noqos is specified, 0 is displayed.

```
[Client node]
# cat /proc/fs/lustre/mdc/<fsname>*/qos
1
```

<fsname>: File system name specified in the FEFS design sheet

If qos_cache is specified in the FEFS design sheet, 1 is displayed. If noqos is specified, 0 is displayed.

```
[Client node]
# cat /proc/fs/lustre/mdc/<fsname>*/qos_cache
1
```

<fsname>: File system name specified in the FEFS design sheet

ex2) Confirming the numerical value of the mclientmax option

You can confirm the mclientmax option (the number of requests that can be issued concurrently within a client node to the MDS), which has a value of 4 in this example.

```
[Client node]
# cat /proc/fs/lustre/mdc/<fsname>*/mclientmax
4
```

4.6.2 Changing the QoS status on the FEFS client

To change the QoS function option specified for a mounted FEFS client, execute the following procedure.

1. Create the FEFS design sheets

Update the FEFS design sheets referring to "[How to configure an FEFS client](#)" of "[3.2.1 Enable the QoS Function](#)".

2. Create configuration definition files for the FEFS setup tool

Execute the procedure in "[3.1.4 Creating configuration definition files for the FEFS setup tool](#)".

3. Deploy the configuration definition files for the FEFS setup tool

Execute the procedure in "[3.1.5 Deploying the configuration definition files for the FEFS setup tool](#)".

4. Create FEFS configuration files

Execute the following:

```
[System management node]
# fefs_sync --setup --<cluster>[,...] --compute=<cluster>[,...]
```

Specify all clusters contained in the FEFS to be built.

- storage : Specify a storage cluster name.
- compute : Specify a compute cluster name or a multiuse cluster name.

5. System stop

Execute the following:

```
[System management node]
# pastop -c <cluster,cluster,...>
```

In the -c option, specify the names of all clusters to be stopped.



Note

- Stop the system in the order of compute cluster - storage cluster.
- If FEFS is running on a multiuse node, stop FEFS on that node before executing the pastop command.



See

For details on the pastop command, see the following manual.

Job Operation Software Administrator's Guide for System Management

6. System startup

Execute the following:

```
[System management node]
# pastart -c <cluster,cluster,...>
```

In the -c option, specify all the cluster name that you want to start.



Note

- Start the system in the order of storage cluster - compute cluster.
- If FEFS is used on a multiuse node, start FEFS on that node after executing the pastart command.



See

For details on the pastart command, see the following manual.

Job Operation Software Administrator's Guide for System Management

4.6.3 Checking the QoS status on the MDS

To check the status of the QoS function specified in the QoS definition file on the MDS, execute the lctl qos stat command.

When the QoS function is enabled (Enable), the QoS definition file path is displayed in the second line, the contents of the QoS definition file in the third and subsequent lines are displayed on the screen.

```
[MDS node]
# lctl qos stat
QoS is Enable.
#QoS file path = /etc/opt/FJSVfefs/qosserver.conf
MDS{
    qos = on
# login node
```

```

nodegrp1 = 30% 203.0.113.10, 203.0.113.20, 203.0.113.30
usermax1 = 10%
# batch-job node
nodegrp2 = 70% 192.0.2.[0-10], 198.51.100./*
usermax2 = 20%
}
OSS{
    qos = same_mds
}
QoS command was completed.

```



See

For details on the lctl qos stat command, see "[lctl qos](#)", which is a subcommand of the "[A.2.7 lctl command](#)".

4.6.4 Changing the QoS status on the MDS

To change the contents of the QoS definition file, use the lctl qos command. A user with root privileges performs this operation on the MDS.

The following describes the procedure.

1. Edit the QoS definition file
Edit the QoS definition file.

```
[MDS node]
# vi /etc/opt/FJSVfef/s/qosserver.conf
```

2. Check the syntax of the QoS definition file.
Check the syntax of the QoS definition file edited in step 1 with the lctl qos check command.

```
[MDS node]
# lctl qos check /etc/opt/FJSVfef/s/qosserver.conf
QoS command was completed.
```

3. Disable the QoS function.
Disable the QoS function with the lctl qos off command.

```
[MDS node]
# lctl qos off
QoS command was completed.
```

4. Enable the QoS function.
Enable the QoS function with the lctl qos on command.

```
[MDS node]
# lctl qos on /etc/opt/FJSVfef/s/qosserver.conf
QoS command was completed.
```



See

For details on the lctl qos check, lctl qos on, and lctl qos off command, see "[lctl qos](#)", which is a subcommand of the "[A.2.7 lctl command](#)".

4.7 Tuning the QoS Function

This section describes the work required to appropriately configure the QoS function.

If you encounter a phenomenon in which file access becomes slower than in normal operation, you should identify whether the cause is on the server side or on the client side.

Shown below is an example of such an identification.

Example

```
Phenomenon: The command response is slow on the login node.  
Identification method: From a different node (that is not a high-load node), execute the same  
command (such as ls) and check the response.  
If the response is normal on the different node -> It is more likely that the cause of the  
problem is on the client side.  
If the response is slow even on the different node -> It is more likely that the cause of the  
problem is on the server side.
```

If the cause seems to be on the client side, see the following sections:

- If file metadata operation (such as the ls or touch command) is slow: See "[4.7.1 Analyzing the client node \(metadata operation\)](#)"
- If file data operation (such as read/write system calls) is slow: See "[4.7.2 Analyzing the client node \(data operation\)](#)"

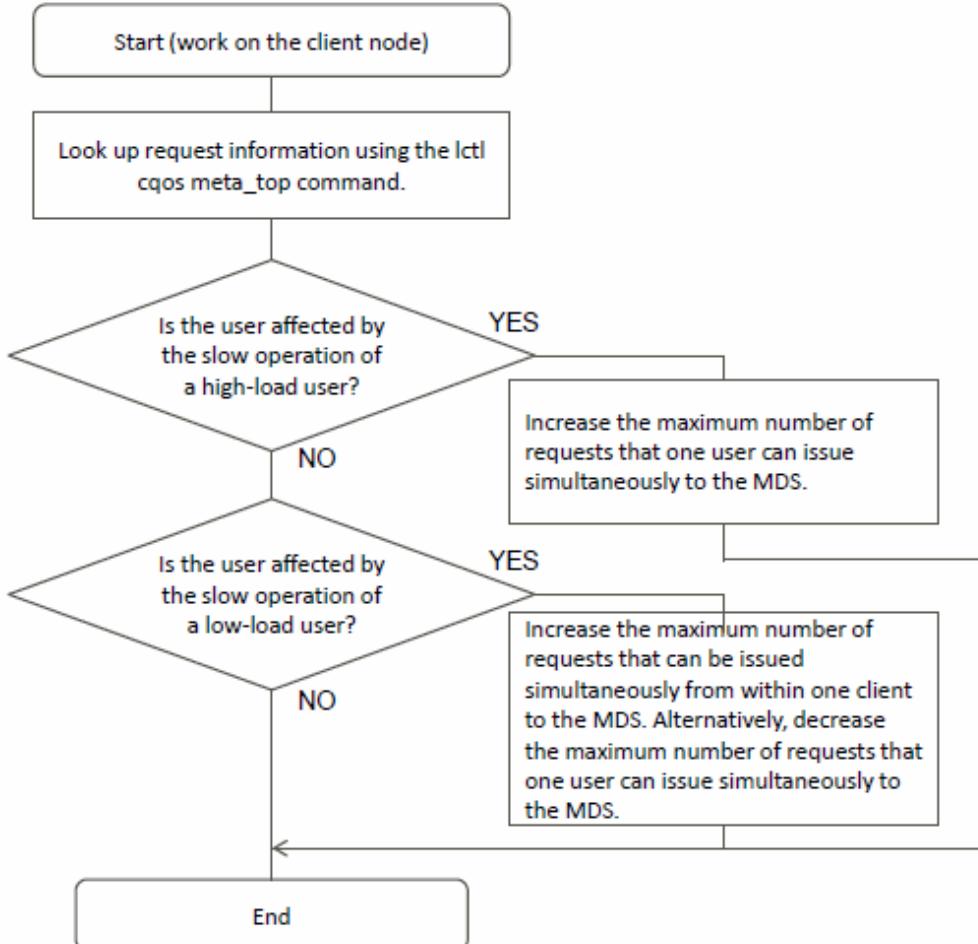
If the cause seems to be on the server side, see the following sections:

- If file metadata operation (such as the ls or touch command) is slow: See "[4.7.3 Analyzing the MDS](#)"
- If file data operation (such as read/write system calls) is slow: See "[4.7.4 Analyzing the OSS](#)"

4.7.1 Analyzing the client node (metadata operation)

If file metadata operation is slow, conduct investigation in the sequence indicated below:

Figure 4.5 Analyzing and tuning client node (metadata operations)



See the following for specific investigation examples:

1) Checking the maximum number of requests that one user can issue simultaneously

Example of execution on a client node

```
# lctl c qos meta_top /mnt/feefs/
mclientmax=4 mrootmax=1 musermax=1
<user info>
-----total_wait_cnt---- --own_wtime(usec)-- -other_wtime(usec)-
No. uid exec_cnt own other max avg max avg last_update
1 1053 10468 9325 0 82128 361 0 0 2013/09/04 16:20:17
2 1070 2595 0 0 0 0 0 0 2013/09/04 16:21:32
3 0 4 0 0 0 0 0 0 2013/09/04 16:21:37
CQoS command was completed.
```

* For details on each output item, see the description of "[lctl c qos](#)", which is a subcommand of the "[A.2.7 lctl command](#)."

The user with uid = 1053 holds an "own" value of 9325 under total_wait_cnt. This means that this user issued a larger number of requests than the maximum number (1 in the above example as indicated by musermax = 1) of requests that one user can issue simultaneously to the MDS. Users whose uid is not 1053 have an "own" value of 0, meaning that their request issuance did not exceed the maximum number of 1.

Suppose that the user who is experiencing slow file metadata operation is the user with uid = 1053. This user may have improved responses if a greater value is assigned to musermax to increase the maximum number of requests that one user can issue simultaneously to the MDS.

For details on how to change the musermax value, see "[4.6.2 Changing the QoS status on the FEFS client](#)."

2) Checking the maximum number of requests that can be issued simultaneously from within one client

Example of execution on a client node

```
# lctl c qos meta_top /mnt/feefs/
mclientmax=4 mrootmax=3 musermax=3
<user info>
-----total_wait_cnt---- --own_wtime(usec)-- -other_wtime(usec)-
No. uid exec_cnt own other max avg max avg last_update
1 1070 20518 2191 21982 87685 975 132039 1193 2013/09/04
16:30:39
2 1053 20266 2037 22215 70809 1269 153994 1066 2013/09/04
16:30:39
3 1071 2336 0 2613 0 0 33758 1122 2013/09/04 16:30:30
4 0 3 0 0 0 0 0 0 2013/09/04 16:30:41
CQoS command was completed.
```

The user with uid = 1071 holds an "own" value of 0 under total_wait_cnt. This means that this user did not issue a larger number of requests than the maximum number (3 in the above example as indicated by musermax = 3) of requests that one user can issue simultaneously to the MDS. The same user holds an "other" value of 2613. This means that this user was forced to wait due to the limitation on the maximum number (4 in the above example as indicated by mclientmax = 4) of requests that can be issued simultaneously from within one client to the MDS. The users with uid = 1070 and uid = 1053 have an "own" value of 1 or greater. It is highly probable that these two users issued so many requests simultaneously that requests from the user with uid = 1071 had to wait.

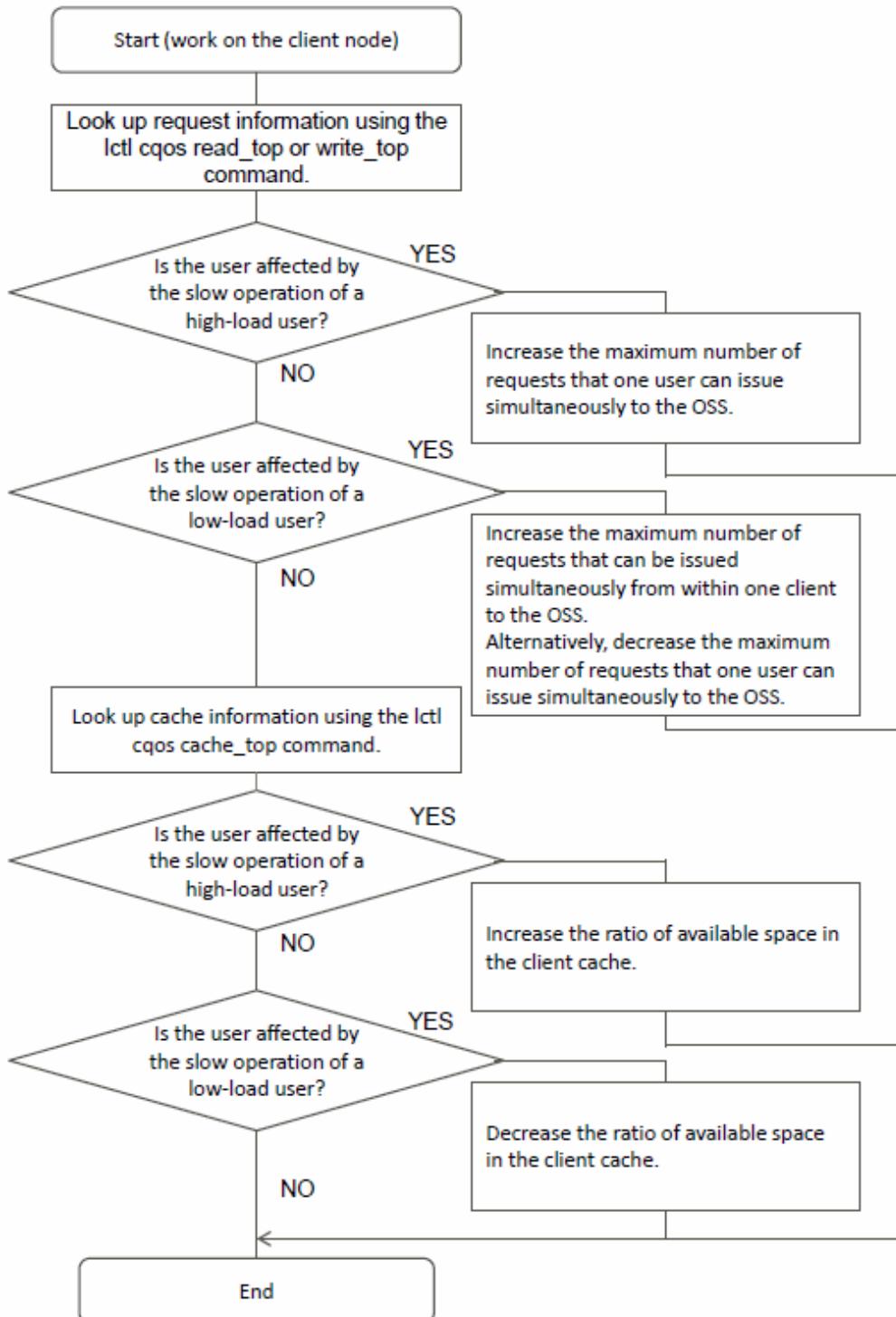
Suppose that the user who is experiencing slow file metadata operation is the user with uid = 1071. This user may have improved responses if a greater value is assigned to mclientmax to increase the maximum number of requests that can be issued simultaneously from within one client to the MDS. Another probable measure for improving this user's responses is to decrease the musermax value. This reduces the maximum number of requests that one user can issue simultaneously to the MDS, thus reducing the number of requests to be issued simultaneously by the users with uid = 1070 and uid = 1053.

For details on how to change the mclientmax or musermax value, see "[4.6.2 Changing the QoS status on the FEFS client](#)."

4.7.2 Analyzing the client node (data operation)

If file data operation is slow, conduct investigation in the sequence indicated below:

Figure 4.6 Analyzing and tuning client node (data operation)



See the following for specific investigation examples:

1) Checking the maximum number of requests that one user can issue simultaneously

Example of execution on a client node

```

# lctl cqos write_top /mnt/fefs/
wrclientmax=8 wrrootmax=2 wrusermax=2
<user info>
          -----total_wait_cnt-----  --own_wtime(usec)--  -other_wtime(usec)-
No.   uid      exec_cnt        own       other      max      avg      max      avg      last_update
  
```

1	1053	20021	5035	0	3808286	102252	0	0	2013/09/04
16:42:24									
2	1070	5017	0	0	0	0	0	0	2013/09/04 16:43:44
3	1071	2013	0	0	0	0	0	0	2013/09/04 16:44:13
CQoS command was completed.									

* The above is an example of output obtained when write_top is specified. To obtain information on read operations, specify read_top.

* For details on each output item, see the description of "[lctl c qos](#)", which is a subcommand of the "[A.2.7 lctl command](#)".

The user with uid = 1053 holds an "own" value of 5035 under total_wait_cnt. This means that this user issued a larger number of write requests than the maximum number (2 in the above example as indicated by wrusermax = 2) of write requests that one user can issue simultaneously to the OSS. Users whose uid is not 1053 have an "own" value of 0, meaning that their request issuance did not exceed the maximum number of 2.

Suppose that the user who is experiencing slow file data operation is the user with uid = 1053. This user may have improved responses if a greater value is assigned to wrusermax to increase the maximum number of write requests that one user can issue simultaneously to the OSS.

For details on how to change the wrusermax value, see "[4.6.2 Changing the QoS status on the FEFS client](#)".

2) Checking the maximum number of requests that can be issued simultaneously from within one client

Example of execution on a client node

# lctl c qos write_top /mnt/fefs/ wrclientmax=8 wrrootmax=6 wrusermax=6 <user info>									
No.	uid	exec_cnt	-----total_wait_cnt-----		--own_wtime(usec)--		-other_wtime(usec)-		
			own	other	max	avg	max	avg	last_update
1	1070	8029	55	3380	69994	21432	655177	29133	2013/09/04 18:57:11
2	1053	8027	276	3676	558694	29637	655604	29818	2013/09/04 18:57:11
3	1071	2006	0	499	0	0	558087	34493	2013/09/04 18:57:16 CQoS command was completed.

The user with uid = 1071 holds an "own" value of 0 under total_wait_cnt. This means that this user did not issue a larger number of write requests than the maximum number (6 in the above example as indicated by wrusermax = 6) of write requests that one user can issue simultaneously to the OSS. The same user holds an "other" value of 499. This means that this user was forced to wait due to the limitation on the maximum number (8 in the above example as indicated by wrclientmax = 8) of write requests that can be issued simultaneously from within one client to the OSS. The users with uid = 1070 and uid = 1053 have an "own" value of 1 or greater. It is highly probable that these two users issued so many write requests simultaneously that requests from the user with uid = 1071 had to wait.

Suppose that the user who is experiencing slow file data operation is the user with uid = 1071. This user may have improved responses if a greater value is assigned to wrclientmax to increase the maximum number of write requests that can be issued simultaneously from within one client to the OSS. Another probable measure for improving this user's responses is to decrease the wrusermax value. This reduces the maximum number of write requests that one user can issue simultaneously to the OSS, thus reducing the number of write requests to be issued simultaneously by users with uid = 1070 and uid = 1053.

For details on how to change the wrclientmax or wrusermax value, see "[4.6.2 Changing the QoS status on the FEFS client](#)".

3) Checking cache usage by each user

Example of execution on a client node

# lctl c qos cache_top /mnt/fefs/ dproutmax=10 dpusermax=10 <user info>									
No.	uid	write_page_cnt	-----total_wait_cnt-----		--own_wtime(usec)--		-other_wtime(usec)-		
			own	other	max	avg	max	avg	last_update
1	1053	8005	7149	0	752317	36295	0	0	2013/09/04 19:05:02
2	1071	1002	0	0	0	0	0	0	2013/09/04 19:05:21

```

3 1070      202      0      0      0      0      0      0 2013/09/04 19:05:17
CQoS command was completed.

```

The user with uid = 1053 holds an "own" value of 7149 under total_wait_cnt. This means that this user issued write requests beyond the maximum ratio (10% in the above example as indicated by dpusermax = 10) of the client cache space available to one user. Users whose uid is not 1053 have an "own" value of 0, meaning that their write request issuance did not beyond the maximum ratio of 10%.

Suppose that the user who is experiencing slow file data operation is the user with uid = 1053. This user may have improved responses if a greater value is assigned to dpusermax to increase the maximum ratio of the client cache space available to one user.

For details on how to change the dpusermax value, see "[4.6.2 Changing the QoS status on the FEFS client.](#)"

4) Checking cache usage by each client

Example of execution on a client node

```

# lctl cqos cache_top /mnt/fefs/
dprotover=70 dpusermax=70
<user info>
          -----total_wait_cnt-----  --own_wtime(usec)--  -other_wtime(usec)-
No.  uid write_page_cnt      own      other      max      avg      max      avg      last_update
 1   1070      128003      199      12592     10419     6340     584271    18072  2013/09/04
19:15:20
 2   1053      128003      497      13008     410463    18878     584403    18119  2013/09/04
19:15:24
 3   1071      501        0       247        0        0     133594    21781  2013/09/04 19:14:29
CQoS command was completed.

```

The user with uid = 1071 holds an "own" value of 0 under total_wait_cnt. This means that this user did not issue write requests beyond the maximum ratio (70% in the above example as indicated by dpusermax=70) of the client cache space available to one user. The same user holds an "other" value of 247. This means that this user was forced to wait due to client cache usage by other users. The users with uid = 1070 and uid = 1053 have an "own" value of 1 or greater. It is highly probable that these two users used the client cache space at so high a ratio that requests from the user with uid = 1071 had to wait.

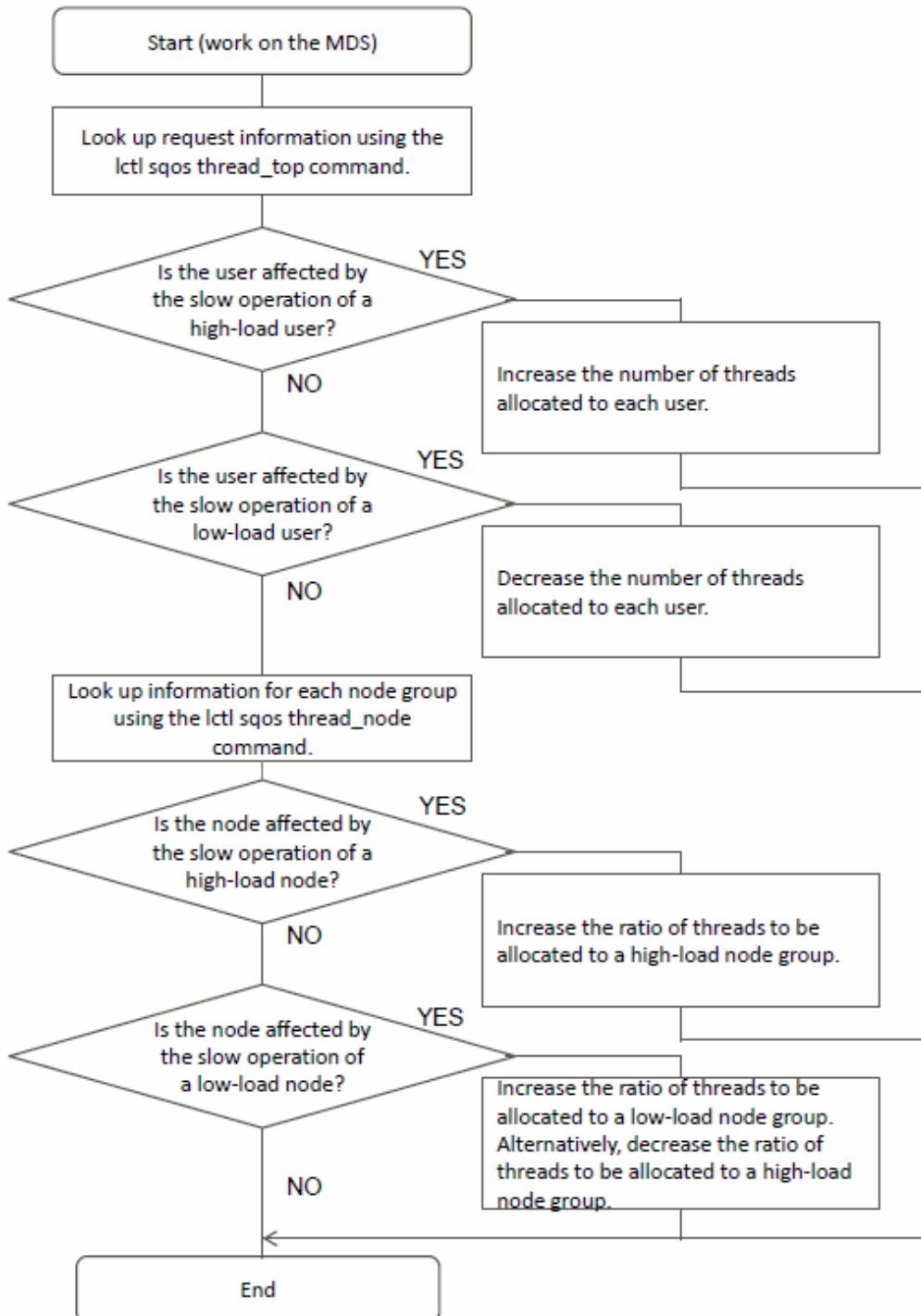
Suppose that the user who is experiencing slow file data operation is the user with uid = 1071. This user may have improved responses if dpusermax is decreased to reduce the maximum ratio of the client cache space available to one user, thus reducing the client cache usage by the users with uid = 1070 and uid = 1053.

For details on how to change the dpusermax value, see "[4.6.2 Changing the QoS status on the FEFS client.](#)"

4.7.3 Analyzing the MDS

If file metadata operation is slow, conduct investigation in the sequence indicated below:

Figure 4.7 Analyzing and tuning the MDS



See the following for specific investigation examples:

1) Checking individual users

If file metadata operation is slow for a specific user, execute the following command on the MDS:

Example of execution on the MDS

# lctl sqos thread_top												
nodegrp= 1		---thread---			---wait_req--		-wait_time(usec)-		-exec_time(usec)-			
No.	uid	exec_cnt	cur	max	lim	cur	max	max	avg	max	avg	last_update
1	1053	148203	0	12	12	0	133	112105	6684	105478	625	2013/08/12

18:11:54											
2	1070	2149	0	1	12	0	1	261	89	46171	619 2013/08/12
18:11:50											
3	1073	1801	0	1	12	0	1	351	92	52256	610 2013/08/12
18:11:53											
4	1071	1525	0	1	12	0	1	298	88	48222	599 2013/08/12
18:11:48											
5	1072	1250	0	1	12	0	1	401	101	51001	615 2013/08/12
18:11:51											

QoS command was completed.

* For details on each output item, see the description of "[lctl sqos](#)", which is a subcommand of the "[A.2.7 lctl command](#)."

The user with uid = 1053 holds a "max" value of 12 under "thread" which is the same as the "lim" value. This means that this user has ever used up 12 threads, which is the limit on the number of threads available to one user. Because the "cur" value under "thread" is 0, this user is not using any thread at present. The "max" value under wait_req is 133, meaning that the request queue for waiting for thread allocation has once reached a maximum of 133. Because the "cur" value under wait_req is 0, the request queue is empty at present.

As you can see, in this output example, the QoS function controls the number of threads executed by the user with uid = 1053 because this user is treated as a high-load user. On the other hand, the function does not control the number of threads executed by other users.

Suppose that the user who is experiencing slow file metadata operation is the high-load user (uid = 1053). This user may have improved responses if the ratio of threads allocated per user is increased. If the affected user is a low-load user (a user whose uid is not 1053), that user may have improved responses if the ratio of threads allocated per user is decreased to control processing for high-load users.

To change the ratio of thread allocation, you need to edit usermax in the QoS definition file. For details on how to edit the QoS definition file, see "[4.6.4 Changing the QoS status on the MDS](#)."

2) Checking individual node groups

If file metadata operation is slow on a specific node group, execute the following command on the MDS:

Example of execution on the MDS

# lctl sqos thread_node											
nodegrp	exec_cnt	---thread---			---wait_req---		-wait_time(usec)-		-exec_time(usec)-		
		cur	max	lim	cur	max	max	avg	max	avg	last_update
1	1218651	0	4	4	0	186	200384	5814	107811	343	2013/08/14 10:35:41
2	455	0	2	19	0	2	85	18	16491	258	2013/08/14 10:35:43

* For details on each output item, see the description of "[lctl sqos](#)", which is a subcommand of the "[A.2.7 lctl command](#)."

The node group with nodegrp = 1 holds a "max" value of 4 under "thread" which is the same as the "lim" value. This means that this node group has used up 4 threads which is the limit on the number of threads available to this node group. The node group with nodegrp = 2 has never used threads up to the limit because its "max" value under "thread" is less than its "lim" value.

As you can see, in this output example, the QoS function controls the number of threads executed by the node group with nodegrp = 1 because this node group is treated as a high-load node group. On the other hand, the function does not control the number of threads executed by the node group with nodegrp = 2.

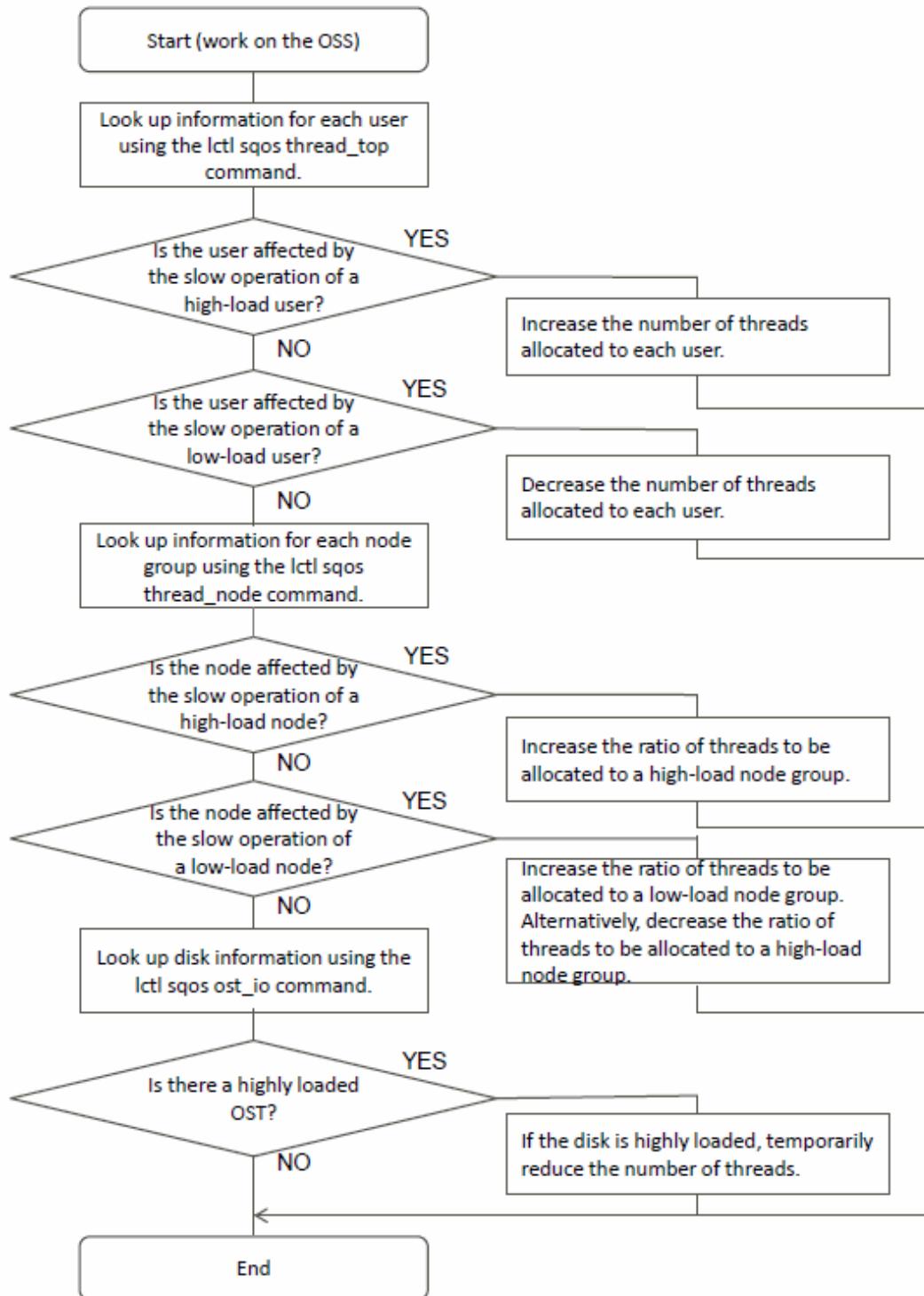
Suppose that the node group experiencing slow file metadata operation is a high-load node group (nodegrp = 1). This group may have improved responses if the ratio of threads allocated to the high-load node group is increased. If the node group experiencing slow file metadata operation is a low-load node group (nodegrp = 2), there are two possible measures to improve responses: One is to increase the ratio of threads allocated to the low-load node group and the other is to decrease the ratio of threads allocated to the high-load node group.

To change the ratio of thread allocation, you need to edit nodegrp in the QoS definition file. For details on how to edit the QoS definition file, see "[4.6.4 Changing the QoS status on the MDS](#)."

4.7.4 Analyzing the OSS

If file data operation is slow, conduct investigation in the sequence indicated below:

Figure 4.8 Analyzing and tuning the OSS



Specify the oss option for lctl sqos command when OSS combines MDS. For details on the lctl sqos command, see "lctl sqos", which is subcommand of the [A.2.7 lctl command](#).

See the following for specific investigation examples:

1) Checking individual users

If file data operation is slow for a specific user, execute the following command on the OSS:

Example of execution on the OSS

```
# lctl sqos thread_top
nodegrp= 1          ---thread---  ---wait_req--  -wait_time(usec)-  -exec_time(usec)-
No.  uid      exec_cnt   cur max lim    cur   max       max     avg      max   avg      last_update
  1  1053        9229      0  10  10     0     71     828997  567675   239225  84829  2013/08/14
11:11:16
  2  1070        1025      0   1  10     0     1       614     80    166067  49479  2013/08/14
11:10:54
  3  1072        980       0   1  10     0     1       522     83    178600  50100  2013/08/14
11:10:48
  4  1071        850       0   1  10     0     1       710     90    190520  51235  2013/08/14
11:10:55
  5  1073        715       0   1  10     0     1       620     79    150158  48155  2013/08/14
11:10:49

QoS command was completed.
```

* For details on each output item, see the description of "[lctl sqos](#)", which is a subcommand of the "[A.2.7 lctl command](#)."

The user with uid = 1053 holds a "max" value of 10 under "thread" which is the same as the "lim" value. This means that this user has used up 10 threads, which is the limit on the number of threads available to one user. Because the "cur" value under "thread" is 0, this user is not using any thread at present. The "max" value under wait_req is 71, meaning that the request queue for waiting for thread allocation has once reached a maximum of 71. Because the "cur" value under wait_req is 0, the request queue is empty at present.

As you can see, in this output example, the QoS function controls the number of threads executed by the user with uid = 1053 because this user is treated as a high-load user. On the other hand, the function does not control the number of threads executed by other users.

Suppose that the user who is experiencing slow file data operation is the high-load user (uid = 1053). This user may have improved responses if the ratio of threads allocated per user is increased. If the affected user is a low-load user (a user whose uid is not 1053), that user may have improved responses if the ratio of threads allocated per user is decreased to control processing for high-load users.

To change the ratio of thread allocation, you need to edit usermax in the QoS definition file. For details on how to edit the QoS definition file, see "[4.6.4 Changing the QoS status on the MDS](#)."

2) Checking individual node groups

If file data operation is slow on a specific node group, execute the following command on the OSS:

Example of execution on the OSS

```
# lctl sqos thread_node
          ---thread---  ---wait_req--  -wait_time(usec)-  -exec_time(usec)-
nodegrp    exec_cnt   cur max lim    cur   max       max     avg      max   avg      last_update
  1         9228      0   51  51     0     30     1059420  203963   1073840  422668  2013/08/14
11:57:23
  2         1025      0   9  204    0     9       1574     115    1111875  393811  2013/08/14
11:56:42

QoS command was completed.
```

* For details on each output item, see the description of "[lctl sqos](#)", which is a subcommand of the "[A.2.7 lctl command](#)."

The node group with nodegrp = 1 holds a "max" value of 51 under "thread" which is the same as the "lim" value. This means that this node group has ever used up 51 threads which is the limit on the number of threads available to this node group. The node group with nodegrp = 2 has never used threads up to the limit because its "max" value under "thread" is less than its "lim" value.

As you can see, in this output example, the QoS function controls the number of threads executed by the node group with nodegrp = 1 because this node group is treated as a high-load node group. On the other hand, the function does not control the number of threads executed by the node group with nodegrp = 2.

Suppose that the node group experiencing slow file data operation is a high-load node group (nodegrp = 1). This group may have improved responses if the ratio of threads allocated to the high-load node group is increased. If the node group experiencing slow file data operation

is a low-load node group (nodegrp = 2), there are two possible measures to improve responses: One is to increase the ratio of threads allocated to the low-load node group and the other is to decrease the ratio of threads allocated to the high-load node group.

To change the ratio of thread allocation, you need to edit nodegrp in the QoS definition file. For details on how to edit the QoS definition file, see "[4.6.4 Changing the QoS status on the MDS](#)."

3) Checking disk loads

If file data operation is slow, execute the following command on the OSS:

Example of execution on the OSS

```
# lctl sqos ost_io
                           --io_time(usec)--
ost_name      io_cnt      max     avg
fefs-OST0000   9228    1070053  416984
fefs-OST0001   1025    1111215  389789
QoS command was completed.
```

* For details on each output item, see the description of "[lctl sqos](#)", which is a subcommand of the "[A.2.7 lctl command](#)".

In the above output example, disk access target was two OSTs: fefs-OST0000 and fefs-OST0001. For fefs-OST0000, 9228 disk access requests were issued and the average time for one access was 416984 microseconds. For fefs-OST0001, 1025 disk access requests were issued and the average time for one access was 389789 microseconds.

In this example, because the time spent for one disk access is 0.3 to 0.4 seconds, it is probable that a disk conflict between threads has occurred. Such a situation is likely to cause a processing delay due to disk waits. To avoid this, specify a disk access time limit (load_limit_usec) in the QoS definition file to limit the number of threads that may simultaneously access the disk. For details on how to edit the QoS definition file, see "[4.6.4 Changing the QoS status on the MDS](#)".

How the result varies depending on whether load_limit_usec is specified or not is described below:

Example 1 of execution on the OSS (without specification for load_limit_usec)

```
# lctl sqos thread_top
nodegrp= 1      ---thread---  ---wait_req--  -wait_time(usec)-  -exec_time(usec)-
No.  uid      exec_cnt  cur  max lim    cur    max      max     avg      max     avg      last_update
1    1053      18447    0    81 128    0      61      6324    625    1086825  548842  2013/08/14
14:03:56
2    1070      513      0    1 128     0      1       444     75     1701725  203323  2013/08/14
14:03:57

QoS command was completed.

# lctl sqos ost_io
                           --io_time(usec)--
ost_name      io_cnt      max     avg
fefs-OST0000   18955    1695127  532015
QoS command was completed.
```

In the above output example, the user with uid = 1070 is a low-load user using only one thread, so that the thread allocation wait time (wait_time) of the user is short. However, the "avg" value under exec_time is 203323 microseconds, meaning that a processing delay has occurred due to disk waits. There are two methods to reduce the disk wait time for uid = 1070. One is to assign a smaller value to usermax in the QoS definition file to reduce the ratio of threads allocated per user. This can control the thread usage by the user with uid = 1053 who is a high-load user using as many as 81 threads. The other is to specify load_limit_usec in the QoS definition file to keep the number of threads below the ratio specified by usermax when the disk load is high.

Shown below is an output example that is obtained when load_limit_usec=30000 is specified in the QoS definition file.

Example 2 of execution on the OSS (with a specification of load_limit_usec=30000)

```
# lctl sqos thread_top
nodegrp= 1      ---thread---  ---wait_req--  -wait_time(usec)-  -exec_time(usec)-
No.  uid      exec_cnt  cur  max lim    cur    max      max     avg      max     avg      last_update
1    1053      18449    0    81 128    0      77      1083345  476457    929469  89972   2013/08/14
14:16:32
```

```

2 1070      513     0 1 128      0     1      296     95    173660   28032  2013/08/14
14:14:56

QoS command was completed.

# lctl sqos ost_io
                           --io_time(usec)--
ost_name          io_cnt      max      avg
fef0-OST0000      18955    917113   86815
QoS command was completed.

```

In the above example, due to the specification of load_limit_usec, the average (avg) io_time has decreased from 532015 microseconds to 86815 microseconds. As a consequence, the average (avg) exec_time for the low-load user with uid = 1070 decreased from 203323 microseconds to 28032 microseconds, indicating that the disk wait time for uid = 1070 has reduced.

In contrast, the average (avg) wait_time, which is the thread allocation wait time, for the high-load user with uid = 1053 has increased from 625 microseconds to 476457 microseconds. This is because the specification of load_limit_usec limits the maximum number of threads when the disk load is high.

4.8 Repairing File System Inconsistencies

Usually, no inconsistencies occur in the FEFS because FEFS has a journaling function.

However, a hardware error, etc. may cause a file system inconsistency.

In such cases, the file system inconsistency is detected at the MDT or OST mounting time.

Once the file system inconsistency is detected, you need to repair it by using the following procedure.

The repairing by the fsck.ldiskfs command can be repaired by executing only the volume where the inconsistencies of file system were detected.

- Repair the MGT by using the fsck.ldiskfs command.
- Repair the MDT by using the fsck.ldiskfs command.
- Repair the OST by using the fsck.ldiskfs command.
- Repair the FEFS by using the lctl lfsck_start command



For details on the fsck-related command, see "[A.2.8 fsck.ldiskfs](#).", "lctl lfsck_start", and "lctl lfsck_stop".

The following example shows how to repair a specific file system inconsistency.

Table 4.4 The example assumes the following environment.

Node Type	Host name	Use	Volume name
MGS	mgs1	MGT	/dev/disk/by-id/scsi-36003005700a5a69012f2037006a14054-part6
MDS	mds1	MDT	/dev/disk/by-id/scsi-36003005700adb66012c35b9708eb7a63-part6
	mds2	MDT	/dev/disk/by-id/scsi-36003005700adf5e012c35a3107cadfa1-part8
OSS	oss1	OST	/dev/disk/by-id/scsi-36003005700a3b0f012fd9c6830808043-part6
			/dev/disk/by-id/scsi-36003005700a3b0f012fd9c6830808043-part7
			/dev/disk/by-id/scsi-36003005700a3b0f012fd9c6830808043-part8
			/dev/disk/by-id/scsi-36003005700a3b0f012fd9c6830808043-part9
	oss2	OST	/dev/disk/by-id/scsi-36003005700a5388012be24d63761405f-part6
			/dev/disk/by-id/scsi-36003005700a5388012be24d63761405f-part7
			/dev/disk/by-id/scsi-36003005700a5388012be24d63761405f-part8

Node Type	Host name	Use	Volume name
			/dev/disk/by-id/scsi-36003005700a5388012be24d63761405f-part9

Information

Confirm the host name and each volume name in GFS sheet of the FEFS design sheet.

4.8.1 Stop the FEFS service

Before repairing the file system, you need to stop the FEFS service. For details on the procedure of stopping, see "[3.7 Operation during maintenance](#)".

4.8.2 Repair on the MGS

The MGT is repaired as follows on the MGS. For details on the fsck-related command, see "[A.2.8 fsck.ldiskfs](#)."

```
[mgs1 node]
# /opt/FJSVfefefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700a5a69012f2037006a14054-
part6
```

4.8.3 Repair on the MDS

The MDT is repaired as follows on the MDS. For details on the fsck-related command, see "[A.2.8 fsck.ldiskfs](#)."

```
[mds1 node]
# /opt/FJSVfefefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700adb66012c35b9708eb7a63-
part6
[mds2 node]
# /opt/FJSVfefefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700adf5e012c35a3107cadfa1-
part8
```

4.8.4 Repair on the OSS

The OST is repaired as follows on the OSS. For details on the fsck-related command, see "[A.2.8 fsck.ldiskfs](#)."

```
[oss1 node]
# /opt/FJSVfefefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700a3b0f012fd9c6830808043-
part6
# /opt/FJSVfefefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700a3b0f012fd9c6830808043-
part7
# /opt/FJSVfefefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700a3b0f012fd9c6830808043-
part8
# /opt/FJSVfefefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700a3b0f012fd9c6830808043-
part9
[oss2 node]
# /opt/FJSVfefefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700a5388012be24d63761405f-
part6
# /opt/FJSVfefefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700a5388012be24d63761405f-
part7
# /opt/FJSVfefefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700a5388012be24d63761405f-
part8
# /opt/FJSVfefefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700a5388012be24d63761405f-
part9
```

4.8.5 Repair of FEFS

Confirm the MDT-OST inconsistency and repair the FEFS. Execute the "lctl lfsck_start" command to repair the FEFS. The procedure is given below.

1) Mounting an MGT

Execute the following on mgs1.

```
# systemctl start FJSVfef
```

2) Mounting MDTs

Execute the following on mds1 and mds2.

```
# systemctl start FJSVfef
```

3) Mounting OSTs

Execute the following on oss1 and oss2.

```
# systemctl start FJSVfef
```

4) Confirming the client FEFS service

Confirm that the FEFS service is stopped in all clients.

When there is any client that is not stopped, execute the following command to end the service.

```
[Client node]
# systemctl stop FJSVfef
```

5) Repairing FEFS

Repair the FEFS on the MDS.

Confirm that MDTs and OSTs are mounted on all MDS and OSS and execute the following command on the MDS to which MDT0 is mounted.

When the file system repair starts, the lctl lfsck_start command recovers, but the repair procedure operates asynchronously.

```
[MDS node]
# lctl lfsck_start -M <fsname>-MDT0000 -A
Started LFSCK on the device fefs-MDT0000: scrub layout namespace
```

* <fsname>: Specify the FEFS file system name.

6) Confirming repair

To confirm that the file system repair is completed, execute the following command on each MDS and OSS node.

If "completed" appears in the status, the repair is complete.

```
[MDS node]
# lctl get_param osd-ldiskfs.<fsname>-MDT*.oi_scrub | grep -e status: -e =
osd-ldiskfs.<fsname>-MDT0000.oi_scrub=
status: completed                                # when file system repair is complete, "completed" appears.
# lctl get_param mdd.<fsname>-MDT*.lfsck_namespace | grep -e status: -e =
mdd.fefs-MDT0000.lfsck_namespace=
status: completed                                # when file system repair is complete, "completed" appears.
# lctl get_param mdd.<fsname>-MDT*.lfsck_layout | grep -e status: -e =
mdd.fefs-MDT0000.lfsck_layout=
status: completed                                # when file system repair is complete, "completed" appears.
```

```
[OSS node]
# lctl get_param osd-ldiskfs.<fsname>-OST*.oi_scrub | grep -e status: -e =
osd-ldiskfs.fefs-OST0000.oi_scrub=
status: completed # when file system repair is complete, "completed" appears.
# lctl get_param obdfilter.<fsname>-OST*.lfsck_layout | grep -e status: -e =
obdfilter.fefs-OST0000.lfsck_layout=
status: completed # when file system repair is complete, "completed" appears.
```

* <fsname>: Specify the FEFS file system name.

7) Mounting a client

Execute the following on every client.

```
# systemctl start FJSVfef
```

4.9 How to Configure the ACL

This section describes how to configure the ACL and how to acquire ACL information. Perform these operations on a client node.

- Configuring the ACL

To configure the ACL, execute the setfacl command with the -m option specified.

The following example grants the read, write, and execute permissions for the testfile file to user1.

```
[Client node]
# setfacl -m user:user1:rwx testfile
```

To delete the ACL settings, execute the setfacl command with the -x option specified.

The following example deletes the ACL configured in the above example.

```
[Client node]
# setfacl -x user:user1: testfile
```

- Acquiring ACL information

To acquire information on the configured ACL, use the getfacl command as shown below.

```
[Client node]
# getfacl testfile
```



For details on the setfacl command and getfacl command, see the reference manuals for the setfacl command and getfacl command.

4.10 How to Set user Extended Attributes

This section describes how to set the user extended attributes and obtain information of the user extended attributes. Perform these operations on the client node.

- Setting user extended attributes.

- To set the user extended attributes, specify the -n option in the setfattr command and execute.

If you also want to set the value of the extended attributes, specify the -v option together.

The following example shows the procedure to set value 0 at the user extended attribute test to the file testfile.

```
[Client node]
# setfattr -n user.test -v value0 testfile
```

- Obtaining information of the user extended attributes
- Execute the getfattr command to obtain the information of set user extended attribute.

```
[Client node]
# getfattr testfile          # display user extended attributes set in testfile
# file: testfile
user.test
# getfattr -n user.test testfile # display user extended attributes together with values
# file: testfile
user.test="value0"
```

- Deleting user extended attributes

Specify the -x option in the setfattr command and execute to delete the user extended attribute settings.

The following example shows the procedure to delete the user extended attributes set in the above.

```
[Client node]
# setfattr -x user.test testfile
```



Note

To use the setfattr command and getfattr command, the attr package must be installed separately on the client node to be executed.

For details on the setfattr command and the getfattr command, see the reference manual for the setfattr command and the getfattr command.

4.11 Checking the FEFS Status

This section describes how to check the FEFS status.

With the pashowclst command, you can check the FEFS service status of each node.

There are two types of status monitoring services, the FEFSSR service and the FEFS service.

The following table shows the status monitoring target nodes and status monitoring items for each service.

Table 4.5 Status monitoring target nodes and status monitoring items

Monitoring service name	Status monitoring item	Status monitoring target node
FEFSSR	FEFS server function	MGS node, MDS node, OSS node
	Global I/O node relay function	Global I/O node
	LLIO server function	Storage I/O node
FEFS	FEFS client function	Compute node, compute cluster management node, login node, multiuse node

ex) By using the -n option of pashowclst command, you can check the FEFS service status of the target node. An example of MDS is shown below.

```
[System management node]
# pashowclst -c clstname -n nodeid1
[ CLST: clstname ]
[ NODE: nodeid1 ]
NODE      NODETYPE      STATUS      REASON      PWR_STATUS      ARCH_STATUS      SRV_STATUS
nodeid1    MDS           Running     -          on             -              FEFSSR(o)
```

The following table shows the meaning of the service status output from the status monitoring target node.

Table 4.6 Status and meaning of the FEFSSR monitoring service in MGS, MDS and OSS node.

Service status	Meaning
o	All monitoring items are normal

Service status	Meaning
x	One or more monitoring items are abnormal
!	Fallback occurs in the multi-path driver or network
s	Initializing service
w	Failback is possible
*	Failover is performed and shifted to one side
f	Failover processing is performed

Table 4.7 Status and meaning of the FEFSSR monitoring service in global I/O node

Service status	Meaning
o	All monitoring items are normal
x	FEFS service is stopped or abnormal
!	Fallback or error occurs in the network
s	Initializing service
b	The global I/O node relay function remain unset

Table 4.8 Status and meaning of the FEFSSR monitoring service in storage I/O node

Service status	Meaning
o	All monitoring items are normal
x	One or more monitoring items are abnormal
!	Fallback or error occurs in network
s	Initializing service
b	LLIO clients remain unset

Table 4.9 Status and meaning of the FEFS monitoring service in compute node, compute cluster management node, login node and multiuse node.

Service status	Meaning
o	All monitoring items are normal
x	One or more monitoring items are abnormal
!	Fallback or error occurs in network
s	Initializing service
a	Error occurs in the FEFSSR monitoring service of SIO node or GIO node and the service is not available
b	FEFS or LLIO clients remain unset



For details on SRV_STATUS, see "Displaying the service operating status" in "Detailed display of operation status of the system" in the "Job Operation Software Administrator's Guide for System Management."

4.12 Failover

Failover can take place only through linkage with Job Operation Software.

When an error occurs in an FEFS related service, the FEFS service monitoring daemon resident in each node reports the error to Job Operation Software, after which a failover automatically occurs.

In this case, the node in error will be in the stopped state.

You can use the -d option of pashowclst command to check the node on which an error has occurred.

[System management node]						
# pashowclst -c clstname -v -d						
[CLST: clstname]						
[NODE: nodeid1]						
NODE	NODETYPE	STATUS	REASON	PWR_STATUS	ARCH_STATUS	SRV_STATUS
0xFFFFF0003	MGS	SoftError	SrvDown	on	-	FEFSSR(x)
0xFFFFF0004	MDS	SoftError	SrvDown	on	-	FEFSSR(x)
0xFFFFF0005	OSS	SoftError	SrvDown	on	-	FEFSSR(x)

clstname: cluster name

You can perform failback of the node with an error from the node where all the services are running, after the maintenance is completed.



Manual switching is also possible during maintenance.
For details on failover/failback, see the following manuals.

Job Operation Software Administrator's Guide for Maintenance

Job Operation Software Administrator's Guide for System Management

4.13 Adding an MDS

To incorporate a new MDS during operation, you need to update FEFS information before mounting the MDT. Use the procedure described below to dynamically add an MDS.



If you want to protect configured data, see "[4.19 Protecting Data in a Built File System](#)."

1. Update FEFS design sheets.

Add MDS information to the FEFS design sheets created during initial installation or to the FEFS design sheets used to protect data in a built file system.

For details on performing this work, see "[3.1.3 Creating FEFS design sheets](#)."

2. Create the configuration definition files for the FEFS setup tool

You can create input data for setup with an Excel macro.

Excel macro: [FEFS Design] > [Create setup files]

Specify the output folder in the dialog box that appears. The macro creates configuration definition files for the FEFS setup tool in the specified folder.

3. Deploy the configuration definition files for the FEFS setup tool

Deploy the configuration definition files for the FEFS setup tool in the following directory on the active and standby SMM:

/etc/opt/FJSVfefs/config

4. Create the FEFS configuration files.

Execute the following.

```
[System management node]
# fefs_sync --setup --storage=<cluster>[,...] --compute=<cluster>[,...]
```

Specify all clusters contained in the FEFS to be built.

--storage : Specify a storage cluster name.
--compute : Specify a compute cluster name or a multiuse cluster name.

5. Volume format

Execute the following.

```
[System management node]
# fefs_sync --mkfs --storage=<cluster> {--nodeid=<nodeid>|--nodelist=<nodeidlist>}
```

Specify the cluster and node ID of the MDS to be added.

--storage : Specify a storage cluster name.
--nodeid : Specify the node ID to be added.
--nodelist : Specify a file that lists the node IDs of the MDS to be added.

6. Mount FEFS

Execute the following.

```
[System management node]
# fefs_sync --mount --storage=<cluster> {--nodeid=<nodeid>|--nodelist=<nodeidlist>}
```

Specify the cluster and node ID of the MDS to be added.

--storage : Specify a storage cluster name.
--nodeid : Specify the node ID to be added.
--nodelist : Specify a file that lists the node IDs of the MDS to be added.

7. Check the FEFS status.

By executing the pashowclst command, confirm that the FEFS service has normally started on the added MDS.

```
[System management node]
# pashowclst -v --nodetype MDS
```

If the FEFSSR status has changed to FEFSSR(o), the FEFSSR service has normally started.



Note

Dynamic MDS deletion is not supported.

4.14 Adding an OSS

To incorporate a new OSS during operation, you need to update FEFS information before mounting the OST. Use the procedure described below to dynamically add an OSS.



See

If you want to protect a built file system, see "[4.19 Protecting Data in a Built File System](#)."

1. Update FEFS design sheets.

Add OSS information to the FEFS design sheets created during initial installation or to the FEFS design sheets used to protect data in a built file system.

For details on performing this work, see "[3.1.3 Creating FEFS design sheets](#)"

2. Create the configuration definition files for the FEFS setup tool

You can create input data for setup with an Excel macro.

Excel macro: [FEFS Design] > [Create setup files]

Specify the output folder in the dialog box that appears. The macro creates configuration definition files for the FEFS setup tool in the specified folder.

3. Deploy the configuration definition files for the FEFS setup tool.

Deploy the configuration definition files for the FEFS setup tool in the following directory on the SMM:

/etc/opt/FJSVfefs/config

4. Create the FEFS configuration files.

Execute the following.

```
[System management node]
# fefs_sync --setup --storage=<cluster>[ ,... ] --compute=<cluster>[ ,... ]
```

Specify all clusters contained in the FEFS to be built.

--storage : Specify a storage cluster name.

--compute : Specify a compute cluster name or a multiuse cluster name.

5. Volume format for the added OSS

Execute the following.

```
[System management node]
# fefs_sync --mkfs --storage=<cluster> --nodeid=<nodeid>
```

Specify the cluster name and the node ID of the added OSS.

--storage : Specify a storage cluster name.

--nodeid : Specify the node ID to be added.

6. Mount added OSS.

Execute the following.

```
[System management node]
# fefs_sync --mount --storage=<cluster> --nodeid=<nodeid>
```

Specify the cluster name and the node ID of the added OSS.

--storage : Specify a storage cluster name.

--nodeid : Specify the node ID to be added.

7. Check the FEFS status.

By executing the pashowclst command, confirm that the FEFS service has normally started on the added OSS.

```
[System management node]
# pashowclst -v --nodetype OSS
```

If the FEFS status has changed to FEFS(o), the FEFS service has normally started.

4.15 Adding Clients

If you want to add PG clients later, you can build in the following procedure.

1. Updating FEFS design sheets

Add PG clients information to the FEFS design sheets created during initial installation.

For details on performing this work, see "[3.1.3 Creating FEFS design sheets](#)."

2. Creating the configuration definition files for the FEFS setup tool

You can create input data for setup with an Excel macro.

Excel macro: [FEFS Design] > [Create setup files]

3. Deploying the configuration definition files for the FEFS setup tool

Deploy the configuration definition files for the FEFS setup tool in the following directory on the active and standby SMM:

/etc/opt/FJSVfefs/config

4. Creating FEFS Configuration files

Execute the following on the system management node.

```
# fefs_sync --setup --storage=<cluster>[,...] --compute=<cluster>[,...]
```

Specify all clusters on which to build the FEFS.

--storage : Specify the storage cluster name.

--compute : Specify the compute cluster name.

5. Starting the FEFS service

Execute the following on the system management node.

```
# fefs_sync --start --compute=<cluster> --nodelist=<nodeidlist>
```

--compute : Specify the compute cluster name.

--nodelist : Specify a file that lists the node IDs of the added client nodes.



See

You can specify the --nodeid option instead of --nodelist. Refer to "[A.2.1 fefs_sync](#)" for details.



Note

If you add an FEFS to a multiuse node, you must restart the service of the job operation software after setting up the FEFS as follows:

```
[System management node]
# paclstmgr -c <cluster> --service restart -n <nodeid>
```

<cluster> : Specify the cluster name.

<nodeid> : Specify the node ID of the multiuse node.

For details on the paclstmgr command, see the *Job Operation Software Setup Guide*.

4.16 Adding a File System

Add an FEFS global file system by using the following procedure.



See

If you want to protect a built file system, see "[4.19 Protecting Data in a Built File System](#)."

1. Update FEFS design sheets.

Add a GFS sheet to the FEFS design sheets created during initial installation or to the FEFS design sheets used to protect data in a built file system, and enter information for the added file system. Also, enter the MDS or OSS information to be added into the NODE sheet.

For details on performing this work, see "[3.1.3 Creating FEFS design sheets](#)"

2. Create the configuration definition files for the FEFS setup tool.

You can create input data for setup using an Excel macro.

Excel macro: [FEFS Design] > [Create setup files]

Specify the output folder in the dialog box that appears. The macro creates configuration definition files for the FEFS setup tool in the specified folder.

3. Deploy the configuration definition files for the FEFS setup tool.

Deploy the configuration definition files for the FEFS setup tool in the following directory on the SMM:

/etc/opt/FJSVfefs/config

4. Create the FEFS configuration files.

Execute the following.

```
[System management node]
# fefs_sync --setup --storage=<cluster>[,...] --compute=<cluster>[,...]
```

Specify all clusters contained in the FEFS to be built.

--storage : Specify a storage cluster name.

--compute : Specify a compute cluster name or a multiuse cluster name.

5. Volume format for the added MDT and OST

Execute the following.

```
[System management node]
# fefs_sync --mkfs --storage=<cluster> --nodelist=<nodeidlist> --fsname=<fsname>
```

Specify the cluster name and file system name of the added file system.

--storage : Specify a storage cluster name.

--nodelist : Specify a file that lists the node IDs of the MDS, OSS to be added.

--fsname : Specify a file system name.



Note

When adding the MDT and OST that are already formatted and used, you do not need to format them.

6. FEFS service stop

Execute the following.

```
[System management node]
# fefs_sync --stop --storage=<cluster> --compute=<cluster>
```

--storage : Specify the storage cluster name.

--compute : Specify the compute cluster name.

7. FEFS service startup

Execute the following.

```
[System management node]
# fefs_sync --start --storage=<cluster> --compute=<cluster>
```

--storage : Specify the storage cluster name.

--compute : Specify the compute cluster name.

8. Change the file system permission.

Set the permission at the mount point on a single arbitrary client.

Perform the work when the file system has been mounted.

(The default value is 755.)

4.17 Deleting a File System

Delete an FEFS global file system by using the following procedure.



Before deleting the file system, backup the necessary data.

1. Update FEFS design sheets.

Delete the GFS sheet of the file system to be deleted, from the FEFS design sheets.

2. Create the configuration definition files for the FEFS setup tool.

You can create input data for setup using an Excel macro.

Excel macro: [FEFS Design] > [Create setup files]

Specify the output folder in the dialog box that appears. The macro creates configuration definition files for the FEFS setup tool in the specified folder.

3. Deploy the configuration definition files for the FEFS setup tool.

Deploy the configuration definition files for the FEFS setup tool in the following directory on the SMM:

/etc/opt/FJSVfefs/config

4. Create the FEFS configuration files.

Execute the following.

```
[System management node]
# fefs_sync --setup --storage=<cluster> --compute=<cluster>
```

Specify all clusters contained in the FEFS to be built.

--storage : Specify a storage cluster name.

--compute : Specify a compute cluster name or a multiuse cluster name.

5. FEFS service stop

Execute the following.

```
[System management node]
# fefs_sync --stop --storage=<cluster> --compute=<cluster>
```

--storage : Specify the storage cluster name.

--compute : Specify the compute cluster name.

6. FEFS service startup

Execute the following.

```
[System management node]
# fefs_sync --start --storage=<cluster> --compute=<cluster>
```

--storage : Specify the storage cluster name.

--compute : Specify the compute cluster name.

4.18 Adding Rack and BoB

Add rack or BoB according to the following procedure.

1. Updating FEFS design sheets

Add node information to the FEFS design sheet created at the time of initial installation.

For setting method, see "[3.1.3 Creating FEFS design sheets](#)" and proceed with the operation.

2. Creating configuration definition files for the FEFS setup tool

You can create the input data for setup by using an Excel macro.

Excel macro: [FEFS Design] > [Create config files]

Specify the output folder in the dialog that appears. The macro creates configuration definition files for the FEFS setup tool in the specified folder.

3. Deploying configuration definition files for the FEFS setup tool.

Deploy the configuration definition files for the FEFS setup tool under the directory /etc/opt/FJSVfefs/config on the SMM.

4. Creating FEFS configuration files

Execute the following command.

```
[System management node]
# fefs_sync --setup --storage=<cluster> --compute=<cluster>
```

Specify all clusters to build FEFS.

--storage : Specify storage cluster name.

--compute : Specify compute cluster name.

5. Starting the FEFS service

Execute the following command.

```
[System management node]
# fefs_sync --start --compute=<cluster> --nodelist=<nodeidlist>
```

Specify the cluster and the node list of rack or BoB to add.

--compute : Specify the compute cluster name.

--nodelist : Specify the file listing the node IDs of the nodes to add.

6. Starting the FEFS service

Confirm that the FEFS service is started normally in the node mounted on the added rack or BoB.

Execute the following command.

```
[System management node]
# pashowclst -v
```

Confirm that the FEFS status transitioned to FEFSSR(o) and FEFS(o).

4.19 Protecting Data in a Built File System

When adding a file system, MDS, or OSS to an environment where a built file system exists, you need to format the volume to be added. Therefore, there is a risk of inadvertently destroying the built file system.

To prepare against this risk, protect the data in the built file system.

4.19.1 Procedure for protecting data in a file system

Protect data in a built file system by using the following procedure.

1. Updating FEFS design sheets

Delete the --reformat option that exists in MKFS OPTION of the GFS sheet corresponding to the file system containing the data to protect.

2. Creating configuration definition files for the FEFS setup tool

You can create input data for setup using an Excel macro.

Excel macro: [FEFS Design] > [Create setup files]

Specify the output folder in the dialog box that appears.

The macro creates configuration definition files for the FEFS setup tool in the specified folder.

3. Deploying the configuration definition files for the FEFS setup tool

Deploy the configuration definition files for the FEFS setup tool in the /etc/opt/FJSVfefc/config directory on the SMM.

4. Creating FEFS configuration files

Execute the following.

```
[System management node]
# fefs_sync --setup --storage=<cluster>[,...] --compute=<cluster>[,...]
```

Specify all clusters contained in the FEFS to be built.

--storage : Specify a storage cluster name.
--compute : Specify a compute cluster name or a multiuse cluster name.



If the disk device connected to the MDS to add has ever been formatted with Lustre or FEFS in the past, deleting the --reformat option causes an error.

To proceed with the setup without the --reformat option, you need to connect the disk device that is not formatted with Lustre or FEFS, to the MDS to add.

4.19.2 Procedure for canceling the protection of data in a file system

Cancel the protection of data in a built file system by using the following procedure.

1. Updating the FEFS design sheet

Set the --reformat option that exists in MKFS OPTION of the GFS sheet corresponding to the file system containing the data to cancel protection.

2. Creating configuration definition files for the FEFS setup tool

You can create input data for setup using an Excel macro.

Excel macro: [FEFS Design] > [Create setup files]

Specify the output folder in the dialog box that appears.

The macro creates configuration definition files for the FEFS setup tool in the specified folder.

3. Deploying the configuration definition files for the FEFS setup tool

Deploy the configuration definition files for the FEFS setup tool in the /etc/opt/FJSVfefsc/config directory on the SMM.

4. Creating FEFS configuration files

Execute the following.

```
[System management node]
# fefs_sync --setup --storage=<cluster>[,...] --compute=<cluster>[,...]
```

Specify all clusters contained in the FEFS to be built.

--storage : Specify a storage cluster name.

--compute : Specify a compute cluster name or a multiuse cluster name.

4.20 How to Use the Data Management Tool (fefsbacup Command) [PG]

This section describes how to use the data management tool.

4.20.1 Subcommand overview

This section provides a functional overview of the subcommands for the fefsbacup command. Unless otherwise noted, these subcommands are executed on an FEFS client. Also only the users who have root privileges can use these subcommands.

For details on the subcommands, see "[A.2.11 fefsbacup](#)."

Table 4.10 Overview of subcommand functions

Subcommand	Function
copy	Copies and transfers files from the source file system to the destination file system.
list	Displays lists of copied information.
delete	Deletes copied information.

Subcommand	Function
status	Displays the copy processes that are currently running or ended in an error.

4.20.2 Transfer information management

The data management tool creates a unique request ID for each user as management information during copying, and manages copied file information.

The tool automatically generates request IDs in the "date_host name of execution node_sequential number" format (example: 20180101_hostA_01 for the first execution time on hostA on 2018/01/01). However, if the -L option is specified, the specified character string is used as the request ID.

The sequential number has two digits, indicating one of the times that the tool was executed on the same day. To execute the data management tool 100 or more times in 1 day, specify a request ID with the -L option, and execute the command.

4.20.3 Mandatory package

When using the data management tool, install the following package on the client node to be executed.

```
tar
rsync
expect
openssh-clients
perl
perl-Data-Dumper
perl-Exporter
perl-File-Path
perl-File-Temp
perl-Getopt-Long
perl-PathTools
perl-TermReadKey
perl-Thread-Queue
perl-Time-HiRes
perl-Time-Local
perl-threads
perl-threads-shared
```

4.20.4 Config files

The following table lists the config files used with the data management tool.

There are no config files at the installation time, so they need to be created on the copy source node. For details, see "[4.20.6 Advance preparation](#)."

Table 4.11 Config files

Name	Description
/etc/opt/FJSVfefs/fefsbackup/config/fefsbackup.conf	Definition file of settings related to common processing independent of the environment
/etc/opt/FJSVfefs/fefsbackup/config/fefsbackup_rsync.conf	Definition file of settings related to processing for the file system environment

4.20.4.1 Config file details

This section provides an overview of each definition file.

Each setting item is specified in the following format.

```
<item name> = <setting value>
```

One single-byte space character is required before and after "=".

If an item name will not be used, delete the item name and its setting value.

fefsbackup.conf

This definition file has settings related to common processing independent of the environment. The following table lists its setting items.

Table 4.12 Overview of fefsbackup.conf setting items

Item name	Description
WORK_DIR	Root directory of the work directory (Absolute path)
WATCHDOG_INTERVAL	Live monitoring file update interval (seconds) A cycle of monitoring determines whether each subcommand is being executed. This is used by the status subcommand to output information.
USE_TAR	Flag indicating whether to use transferred tar archives
TAR_FILE_SIZE_MAX	Maximum size (bytes) for a file stored in a tar file *
TAR_TOTAL_SIZE_MAX	Total size (bytes) of files stored in 1 tar file
TAR_FILE_NUM_MAX	Number of files stored in 1 tar file *

* TAR_FILE_SIZE_MAX is that size threshold that determines whether to create a tar archive. The tool creates and transfers tar archives from files with a size of TAR_FILE_SIZE_MAX or smaller.

TAR_TOTAL_SIZE_MAX and TAR_FILE_NUM_MAX determine the maximum total size and number, respectively, of files contained in a single tar archive. If either value is exceeded, create and transfer another tar file.

Table 4.13 fefsbackup.conf setting values

Item name	Setting value	Setting needed?	Default value
WORK_DIR	Character string (up to 511 characters)	Required	None
WATCHDOG_INTERVAL	Integer between 1 and 86,400	Optional	60
USE_TAR	0 (Disable) 1 (Enable)	Optional	1
TAR_FILE_SIZE_MAX	Integer between 0 and 2^63	Optional	1048576
TAR_TOTAL_SIZE_MAX	Integer between 0 and 2^63	Optional	25000000000
TAR_FILE_NUM_MAX	Integer between 1 and 2^63	Optional	1000000

fefsbackup_rsync.conf

This definition file has settings related to processing for the file system. The following table lists its setting items.

Table 4.14 Overview of fefsbackup_rsync.conf setting items

Item name	Description
ARCHIVE_HOST	The IP address or host name of the node where the management information is stored. Specify a node managed by Technical Computing Suite.
BACKUP_ROOT	Directory path of the management information storage node, under which management information is stored.
RSYNC_COMMAND_LOCAL	Rsync path on the source node. *
RSYNC_COMMAND_REMOTE	Rsync path on the destination node. *
MULTI_PUT_MAX	Maximum number of parallel file transfers
SSH_PORT	Port number used by ssh

* Normally, execute the rsync command in the default path. However, you can specify a path for special use cases (for example, if you want to use different rsync versions, etc.).

Table 4.15 fefsbackup_rsync.conf setting values

Item name	Setting value	Setting needed?	Default value
ARCHIVE_HOST	IP address or host name	Required	None
BACKUP_ROOT	Character string (up to 1,023 characters)	Required	None
RSYNC_COMMAND_LOCAL	Character string (up to 1,023 characters)	Optional	None
RSYNC_COMMAND_REMOTE	Character string (up to 1,023 characters)	Optional	None
MULTI_PUT_MAX	Integer between 1 and 2^31 (Max: Number of CPUs of execution nodes x 2)	Optional	1
SSH_PORT	Integer between 0 and 65,536	Optional	22

4.20.5 Design where management information is stored

When copying a file using the data management tool, save the file information transferred to the management information storage node.

The disk space used for one management information on the storage node is as follows.

```
4096+((35+average file name length)*number of files)+(average directory path name length)
```

The maximum disk space used as a temporary space by a single instance of the copy process is as follows.

```
(2 MB * MULTI_PUT_MAX)+((35 + average file name length)* number of file + (35+directory path name length)* number of directory)
```

To execute the data management tool on multiple nodes, set the configuring WORK_DIR of config file on FEFS. By doing so, you can check the execution status with the status subcommand from another node, and you can retry the tool when an error occurs.

4.20.6 Advance preparation

This section lists the items that have to be prepared before execution of the FEFS data management tool.

- Configuring ports
Open the port used by ssh on the source, destination nodes, and management information storage nodes. To change the port number from the default value (22), change the config file (SSH_PORT).
- Creating a management directory
Create the storage directory (the directory specified for "BACKUP_ROOT" in the config file) in advance in the copy management information storage node (config file "ARCHIVE_HOST").
- Creating a work directory
Create a work directory (directory specified at "WORK_DIR" in a config file) in advance on the node where the data management tool is executed.
- Registering the host key of the management information storage node
Connect to the management information storage node beforehand to register the host key of the management information storage node with known_hosts for the ssh command. Also, if the host key of the management information storage node changes, the key must be reregistered.
- Password for a user with administrator privileges
The management information storage node should have the same password as the user with administrator privileges on the destination node.

4.20.7 Configuring the data management tool

1. Creating config files for the data management tool

Create the config files on the node that will be executing the data management tool.

Create the fefsbackup.conf and fefsbackup_rsync.conf config files to configure the data management tool.

Place the fefsbackup.conf and fefsbackup_rsync.conf config files at the following location.

```
/etc/opt/FJSVfefs/fefsbackup/config/fefsbackup.conf  
/etc/opt/FJSVfefs/fefsbackup/config/fefsbackup_rsync.conf
```



Neither fefsbackup.conf nor fefsbackup_rsync.conf exists at the FEFS installation time. Copy the sample files shown below and place them in the paths shown above.

Sample file fefsbackup.conf

```
/etc/opt/FJSVfefs/fefsbackup/config/sample/fefsbackup.conf.sample
```

Sample file fefsbackup_rsync.conf

```
/etc/opt/FJSVfefs/fefsbackup/config/sample/fefsbackup_rsync.conf.sample
```

2. Checking setting values

Check the set values.

The following example displays settings.

Example displaying the fefsbackup.conf config file

```
# cat /etc/opt/FJSVfefs/fefsbackup/config/fefsbackup.conf  
WORK_DIR  = /fefs/fefsbackup_workdir  
WATCHDOG_INTERVAL = 60  
USE_TAR = 1  
TAR_FILE_SIZE_MAX = 1048576
```

Example displaying the fefsbackup_rsync.conf config file

```
# cat /etc/opt/FJSVfefs/fefsbackup/config/fefsbackup_rsync.conf  
BACKUP_ROOT = /fefs/backup_dir  
MULTI_PUT_MAX = 4  
ARCHIVE_HOST = 10.0.0.18
```

4.20.8 copy subcommand

The copy subcommand transfers files between file systems. An example of execution is shown below.

1. Specify a path to transfer files between the same nodes.

```
[Client node]  
# /opt/FJSVfefs/bin/fefsbackup copy -a -x -d /fefs/backupdir /home/projectA  
password: ***** * Enter ssh password.  
Copying request 20180604_lhostA_01 is executing...  
Total 6 files (6291456 bytes) were copied from /home/projectA directory.  
Copying request 20180604_lhostA_01 is completed successfully.
```

2. Transfer the files and directories described in the pathlist specified in the -f option. Also specify a request ID for managing transferred information.

```
[Client node]  
# cat pathlist
```

```

/home/projectA/shell/test1.sh
/home/projectA/shell/test2.sh

# /opt/FJSVfefefs/bin/fefsbackup copy -L projectA_Backup -f pathlist -n rhostA -d /fefs/backupdir
password: ***** * Enter ssh password.
Copying request projectA_Backup is executing...
Total 2 files (2048 bytes) were copied from pathlist file.
Copying request projectA_Backup is completed successfully.

```

3. From an existing archive managed with the user-specified request ID, make a differential transfer under the specified path name. The information transferred by difference is managed by the specified request ID.

```

[Client node]
# /opt/FJSVfefefs/bin/fefsbackup copy -v -u 20180604_lhostA_01 -n rhostA \
-d /fefs/backupdir /home/projectA
password: ***** * Enter ssh password
Copying request 20180604_lhostA_01 is executing...

/home/projectA/shell
/home/projectA/shell/test1_new.sh
/home/projectA/shell/test2_new.sh
Total 2 files (2048 bytes) were copied from /home/projectA directory.
Copying request 20180604_lhostA_01 is completed successfully.

```

4. From an existing archive managed with the user-specified request ID, make a differential transfer of the files and directories described in the specified pathlist.

```

[Client node]
# cat pathlist
/home/projectA/shell/test1.sh
/home/projectA/shell/test2.sh

# /opt/FJSVfefefs/bin/fefsbackup copy -u 20180101_lhostA -f pathlist -n rhostA -d /fefs/backupdir
password: ***** * Enter ssh password
Copying request 20180101_lhostA is executing...
Total 2 files (2048 bytes) were copied from pathlist file..
Copying request 20180101_lhostA is completed successfully.

```

5. The transfer process may be interrupted by an error or the like, in which case you may be able to retry the process. If so, specify the request ID, and resume.

```

[Client node]
# /opt/FJSVfefefs/bin/fefsbackup copy -v -R 20180604_lhostA_01
password: ***** * Enter ssh password
Copying request 20180604_lhostA_01 is executing...
/home/projectA/shell/test1.sh : failed (No such file or directory)
/home/projectA/shell/test2.sh : failed (No such file or directory)
/home/projectA/shell/test3.sh
/home/projectA/shell/test4.sh
/home/projectA/shell/test5.sh
/home/projectA/shell/test6.sh
Total 4 files (3200826 bytes) were copied from /home/projectA directory.
Copying request 20180604_lhostA_01 is completed successfully.

```

Note

- Retry

If an error occurs, such as a file system access error, the data management tool ends with an error and the "Internal error has occurred" message. At this time, if the following message is output, you can resume processing from the point of the error after recovery. To do so, use the -R option to specify the request ID associated with the error, and try the copying again.

Resume the command after checking the internal error cause.

In this case, the command is executed with the same specified contents as when initially executed, except for the -v option. The contents will be the same regardless of what is specified at the retry time.

- File attributes (access rights, ownership, and time of last modification) at the copy time

The file attributes (access rights, ownership, and time of last modification) are copied from the attributes of the source file.
The following is not copied.

- Project ID
- Project ID inheritance flag
- Information about the striped directory
- Information about the stripes

- Hard link files

Hard link is not guaranteed to be copied.

Hard links may be copied if file size is less than or equal to the size specified for TAR_FILE_SIZE_MAX in fefsbackup.conf.

- Symbolic link files

Symbolic link files copy the linked file itself, not the referenced file.

- Special files, and device files

Special files such as named sockets, FIFOs, and device files are also copied.

- File name

An error will occur if a file name or directory name containing a line feed exists in the copy target.

- About the path name in the operand path or the pathlist in -f option

If the path name starts with a half-width symbol, use an absolute path or a parent directory that does not start with a half-width symbol.

- File path when a transfer process fails

When copying files fails, the path of the error message file may be printed without the leading "/" even if the path is specified as an absolute path in the format of the copy subcommand of fefsbackup.

The leading "/" is stripped from the pathname if the size of the file that failed to copy the file is less than or equal to the size specified for TAR_FILE_SIZE_MAX in fefsbackup.conf.

- Simultaneous command execution

Do not execute the copy subcommand of fefsbackup with specifying the same request ID with the -u option simultaneously.

4.20.9 list subcommand

The list subcommand outputs copy request ID information. An example of execution is shown below.

1. When executed without any arguments, the subcommand outputs a list of request IDs.

```
[Client node]
# /opt/FJSVfefsf/bin/fefsbackup list
password: **** * Enter ssh password
Request ID      Command  Files  Bytes  Backed-up          Src_Host  Dest_Host Target
20180606_hostA_01  copy    1K    1G    2018/06/06 18:34:49 shostA  dhostA
fefsfbackup_test_data
20180628_hostA_01  copy    1K    1G    2018/06/28 13:21:51 shostA  dhostA
(fefsbackup_test_data.list)
```

The following table lists details of display parameters for the subcommand executed without any arguments.

Table 4.16 Display parameter details

Parameter name	Description
Request ID	Request ID of the target
Command	Executed subcommand (copy)
Files	Number of copied files

Parameter name	Description
Bytes	Number of copied bytes
Backed-up	Instructed copy period Display format: YYYY/MM/DD hh:mm:ss
Src_Host	Execution node name
Dest_Host	Copy destination node name (IP address)
Target	Target path for copying The parentheses indicate the pathlist specified by -f at the respective subcommand execution time.

2. Specify a request ID to output a list of files managed with this request ID.

```
[Client node]
# /opt/FJSVfefefs/bin/fefsbackup list 20180709_hostA_01
password: **** * Enter ssh password
dirA/fileA
dirB/fileA
dirB/fileB
```

3. Add the -l option to output detailed information.

```
[Client node]
# /opt/FJSVfefefs/bin/fefsbackup list -l 20180709_hostA_01
password: **** * Enter ssh password
100664 usrA groupA 225 2018/07/06 15:55 dirA/fileA
100664 usrA groupA 8771 2018/07/02 10:36 dirB/fileA
100664 usrA groupA 4408 2018/07/03 10:01 dirB/fileB
```

The following table lists details of display parameters for times when a request ID and the -l option are specified.

Table 4.17 Display parameter details

Parameter name	Description
Access rights	Displays the access rights to the applicable file in the 1st column.
Owner	Displays the owner of the applicable file in the 2nd column.
Group	Displays the group of the applicable file in the 3rd column.
Size	Displays the size of the applicable file in the 4th column.
Update date/time	Displays the update date/time of the applicable file in the 5th column. The display format is YYYY/MM/DD hh:mm.
File name	Displays the name of the applicable file in the 6th column.

4.20.10 delete subcommand

The delete subcommand deletes the specified copy request IDs. An example of execution is shown below. The delete subcommand only deletes the management information and does not delete the copied files.

1. Delete the specified copy request IDs.

```
[Client node]
# /opt/FJSVfefefs/bin/fefsbackup delete 20180604_lhostA_01
password: **** * Enter ssh password
20180604_lhostA_01 will be deleted.
This request ID will be PERMANENTLY *LOST* and cannot be recovered.
Type "yes" to continue or "no" to abort. [no]
yes
Deleting request ID in backups...
Deleting request ID is completed successfully.
```

4.20.11 status subcommand

The status subcommand outputs a list of the request IDs for the copy commands that are currently running or ended in an error. If a request ID is specified, the subcommand outputs the execution information for that request ID. An example of execution is shown below.

- When executed without any arguments, the subcommand outputs a list of execution information for this user.

```
[Client node]
# /opt/FJSVfefefs/bin/fefsbackup status
password: **** * Enter ssh password
Request_id          Command  Status
20180514_hostA_01   copy     STOPPED
20180706_hostA_02   copy     RUNNING
```

The following table lists details of display parameters for the subcommand executed without any arguments.

Table 4.18 Display parameter details

Parameter name	Description
Request_id	Request ID
Command	Subcommand that is currently running or have ended in an error.
Status	Execution status. The output status is as follows: - RUNNING: Running - STOPPED: Ended due to an internal processing error such as system trouble

- Specify a request ID to output the execution information for that request ID.

```
[Client node]
# /opt/FJSVfefefs/bin/fefsbackup status 20180706_hostA_02
password: **** * Enter ssh password
RequestID : 20180706_hostA_02
Command   : copy
Status    : RUNNING
Files     : 234
Bytes     : 45379
Started   : 2018/07/06 16:17:05
UserID    : 0
```

The following table lists details of display parameters for the output when a request ID is specified.

Table 4.19 Display parameter details

Parameter name	Description
RequestID	Request ID of the target
Command	Running subcommand (copy)
Status	Execution status. The output status is as follows: - RUNNING: Running - STOPPED: Ended due to an internal processing error such as system trouble
Files	Number of completely transferred files
Bytes	Number of completely transferred bytes
Started	Execution start time. The display format is YYYY/MM/DD hh:mm:ss.
UserID	Executing user ID

4.21 JobStats Function

JobStats is a function that records the statistics of each job process. The JobStats function acquires the job ID from the environment variable of each job process that accesses FEFS, and saves statistics under each job ID. The saved statistics are accumulated on memory and

periodically cleared by the JobStats function. You can confirm the clear interval by executing the command indicated below. The units are indicated in seconds.

```
[MDS node]
# cat /proc/fs/lustre/mdt/<fsname>-MDT<xxxx>/job_cleanup_interval
```

<fsname> : name of the file system

<xxxx> : index number of the MDT

```
[OSS node]
# cat /proc/fs/lustre/obdfilter/<fsname>-OST<xxxx>/job_cleanup_interval
```

<fsname> : name of the file system

<xxxx> : index number of the OST



Note

- In an environment with multiple compute clusters, each compute cluster is numbered with an independent job ID. The function does not distinguish between any duplicated IDs among these independent job IDs, so their statistics would be displayed under the same job ID.
- The memory amount used by the JobStats function is as follows on the assumption that 10000 jobs are executed:"About 12 MB x the number of MDTs" in the case of an MDS or "about 5 MB x the number of OSTs" in the case of an OSS.
- If a process other than a job accesses FEFS on a compute node, collect statistical information using *process_name.uid* as the job ID.

4.22 How to Use the FEFS Statistics Visualization Function (fefssv.ph Script)

4.22.1 How to collect information

Describe the following in the collectl daemon startup option setting line (DaemonCommands) of the collectl configuration file /etc/collectl.conf.

```
[MDS node or OSS node]
$ collectl -f <output dir> [<collectl option>] --import /opt/FJSVfeffs/bin/fefssv.ph
```

<output dir>: Output directory name

<collectl option>: collectl command option

An execution example is shown below.

```
[MDS node or OSS node]
$ collectl -f /var/log/collectl -r00:00,7 -m -F60 -s+YZ -i10:60:300 --import /opt/FJSVfeffs/bin/
fefssv.ph
```

Execute the following command to start the collectl service automatically.

```
[MDS node or OSS node]
# systemctl enable collectl
# systemctl start collectl
```

To change the setting during operation, restart the collectl service as shown below.

```
[MDS node or OSS node]
# systemctl reload collectl
```

For details on the collectl command options, see the man manual of collectl(1).

4.22.2 How to output information

Specify a collectl log file as input, and specify what you want displayed with the fefssv.ph script in the --import option. The output is displayed on the screen.

```
[MDS node or OSS node]
$ collectl -p <data file> -s-all [<collectl option>] --import /opt/FJSVfeffs/bin/fefssv.ph,{mdt|ost}
$ collectl -p <data file> -s-all [<collectl option>] --import /opt/FJSVfeffs/bin/fefssv.ph,d,{mdt|ost}
[,fs=<FS name>]
$ collectl -p <data file> -s-all [<collectl option>] --import /opt/FJSVfeffs/bin/fefssv.ph,d,
{mdt=mdt_name|ost=ost_name}
$ collectl -p <data file> -s-all [<collectl option>] \
--import /opt/FJSVfeffs/bin/fefssv.ph,v,{mdt|ost}[,fs=<FS name>],[jobid=<job id>]
$ collectl -p <data file> -s-all [<collectl option>] \
--import /opt/FJSVfeffs/bin/fefssv.ph,v,{mdt=mdt_name|ost=ost_name}, [jobid=<job id>]
```

<data file>: Output file name

<collectl option>: collectl command option

<FS name>: File system name

<job ID>: Job ID

For details on other options of the fefssv.ph script, see "[4.22.3 Options and output information](#)."

Execution examples are shown below.

- Execution example with the MDS

```
$ collectl -p /var/log/collectl/rx200-001-20160325-000000.raw.gz -s-all -oDm --from 11:03:20-11:06:20
--import (*)
/opt/FJSVfeffs/bin/fefssv.ph,v,mdt=fefs-MDT0001

# Lustre Jobstats
#                               MDT_NAME      JOBID    open   close   mknod    link   unlink   mkdir
rmdir (*)
rename  setattr  getxattr  setxattr  statfs     sync  sameDir_rename crossDir_rename
20160325 11:03:20.001  fefs-MDT0001      17       0       0       0       0       0       0
0      0 (*)          0         0           0       0       0       0       0       0
20160325 11:03:20.001  fefs-MDT0001      18       4       3       0       0       0       0       1551
0      0 (*)          0         0           1       0       0       0       0       0
20160325 11:03:30.001  fefs-MDT0001      17       0       0       0       0       0       0       0
0      0 (*)          0         0           0       0       0       0       0       0
20160325 11:03:30.001  fefs-MDT0001      18       0       0       0       0       0       0       16620
0      0 (*)          0         0           0       0       0       0       0       0
20160325 11:03:40.001  fefs-MDT0001      17       0       0       0       0       0       0       0
0      0 (*)          0         0           0       0       0       0       0       0
20160325 11:03:40.001  fefs-MDT0001      18       0       0       0       0       0       0       0       21837
40000 (*)          0         0           0       0       0       0       0       0
0      40004          0         0           0       0       0       0       0       0
20160325 11:03:50.001  fefs-MDT0001      17       0       0       0       0       0       0       0
0      0 (*)          0         0           0       0       0       0       0       0
20160325 11:03:50.001  fefs-MDT0001      18       40000   40000     0       0       0       0       0
0      0 (*)          0         0           0       0       0       0       0       0
20160325 11:04:00.001  fefs-MDT0001      17       0       0       0       0       0       0       0
0      0 (*)          0         0           0       0       0       0       0       0
20160325 11:04:00.001  fefs-MDT0001      18       0       0       0       0       0       0       40000
5      4 (*)          0         0           0       0       0       0       0       0
0      34424          0         0           0       0       0       0       0       0
```

20160325	11:04:10.001	fefs-MDT0001	17	0	0	0	0	0	0
0	0 (*)								
0	0 0	0 0	0	0	0	0	0	0	0
20160325	11:04:10.001	fefs-MDT0001	18	0	0	0	0	0	40000
34852	(*)								
0	40012 0	0 0	0	0	0	0	0	0	0
20160325	11:04:20.001	fefs-MDT0001	17	0	0	0	0	0	0
0	0 (*)								
0	0 0	0 0	0	0	0	0	0	0	0
20160325	11:04:20.001	fefs-MDT0001	18	34137	34134	0	0	0	0
0	5148 (*)								
0	4 0	0 0	0	0	0	0	0	0	0
20160325	11:04:30.001	fefs-MDT0001	17	0	0	0	0	0	0
0	0 (*)								
0	0 0	0 0	0	0	0	0	0	0	0
20160325	11:04:30.001	fefs-MDT0001	18	5863	5866	0	0	0	40000
5	4 (*)								
0	40012 0	0 0	0	0	0	0	0	0	0
20160325	11:04:40.001	fefs-MDT0001	17	0	0	0	0	0	0
0	0 (*)								
0	0 0	0 0	0	0	0	0	0	0	0
20160325	11:04:40.001	fefs-MDT0001	18	0	0	0	0	0	40000
4	(*)								
0	40012 0	0 0	0	0	0	0	0	0	0
20160325	11:04:50.001	fefs-MDT0001	17	0	0	0	0	0	0
0	0 (*)								
0	0 0	0 0	0	0	0	0	0	0	0
20160325	11:04:50.001	fefs-MDT0001	18	12468	12464	0	0	0	0
0	39996 (*)								
0	4 0	0 0	0	0	0	0	0	0	0
20160325	11:05:00.001	fefs-MDT0001	17	0	0	0	0	0	0
0	0 (*)								
0	0 0	0 0	0	0	0	0	0	0	0
20160325	11:05:00.001	fefs-MDT0001	18	27532	27536	0	0	0	6443
0	0 (*)								
0	40004 0	0 0	0	0	0	0	0	0	0
20160325	11:05:10.001	fefs-MDT0001	17	0	0	0	0	0	0
0	0 (*)								
0	0 0	0 0	0	0	0	0	0	0	0
20160325	11:05:10.001	fefs-MDT0001	18	0	0	0	0	0	33557 40005
4	(*)								
0	9739 0	0 0	0	0	0	0	0	0	0
20160325	11:05:20.001	fefs-MDT0001	17	0	0	0	0	0	0
0	0 (*)								
0	0 0	0 0	0	0	0	0	0	0	0
20160325	11:05:20.001	fefs-MDT0001	18	0	0	0	0	0	0
0	40000 (*)								
0	30281 0	0 0	0	0	0	0	0	0	0
20160325	11:05:30.001	fefs-MDT0001	17	0	0	0	0	0	0
0	0 (*)								
0	0 0	0 0	0	0	0	0	0	0	0
20160325	11:05:30.001	fefs-MDT0001	18	40000	40000	0	0	0	0
0	0 (*)								
0	24632 0	0 0	0	0	0	0	0	0	0
20160325	11:05:40.001	fefs-MDT0001	17	0	0	0	0	0	0
0	0 (*)								
0	0 0	0 0	0	0	0	0	0	0	0
20160325	11:05:40.001	fefs-MDT0001	18	0	0	0	0	0	40000 3950
4	(*)								
0	15392 0	0 0	0	0	0	0	0	0	0
20160325	11:05:50.001	fefs-MDT0001	17	0	0	0	0	0	0
0	0 (*)								
0	0 0	0 0	0	0	0	0	0	0	0

20160325	11:05:50.001	fefs-MDT0001	18	0	0	0	0	0	29325
0	(*)								
0	0	0	0	0	0	0	0	0	0
20160325	11:06:00.001	fefs-MDT0001	17	0	0	0	0	0	0
0	0	(*)							
0	0	0	0	0	0	0	0	0	0
20160325	11:06:00.001	fefs-MDT0001	18	0	0	0	0	0	6730
40000	(*)								
0	40004	0	0	0	0	0	0	0	0
20160325	11:06:10.001	fefs-MDT0001	17	0	0	0	0	0	0
0	0	(*)							
0	0	0	0	0	0	0	0	0	0
20160325	11:06:10.001	fefs-MDT0001	18	40000	40000	0	0	0	0
0	0	(*)							
0	19208	0	0	0	0	0	0	0	0
20160325	11:06:20.001	fefs-MDT0001	17	0	0	0	0	0	0
0	0	(*)							
0	0	0	0	0	0	0	0	0	0
20160325	11:06:20.001	fefs-MDT0001	18	0	0	0	0	0	40000
0	4	(*)							
0	20804	0	0	0	0	0	0	0	0

Remarks: Due to space limitations, each (*) location in the example shown above is an inserted line break. The wrapped text is actually displayed as one line.

- Execution example with the OSS

\$ collectl -p /var/log/collectl/rx200-003-20160325-000000.raw.gz -s-all -oDm --from 11:18:10-11:19:50
--import (*)
/opt/FJSVfefsfes/bin/fefssv.ph,v,ost=fefs-OST0001
Lustre Jobstats
#
setattr (*)
punch sync destroy
create statfs get_info set_info quotactl
20160325 11:18:10.001 fefs-OST0001 17 0 0 0 0 0 0 0 0
0 0 (*)
0 0 0 0 0 0 0 0 0 0
20160325 11:18:10.001 fefs-OST0001 18 0 0 0 0 0 0 0 0
0 0 (*)
0 0 0 0 0 0 0 0 0 0
20160325 11:18:10.001 fefs-OST0001 19 0 0 0 0 747 782237696 0 0
0 0 (*)
0 0 0 0 0 0 0 0 0 0
20160325 11:18:20.001 fefs-OST0001 17 0 0 0 0 0 0 0 0
0 0 (*)
0 0 0 0 0 0 0 0 0 0
20160325 11:18:20.001 fefs-OST0001 18 0 0 0 0 0 0 0 0
0 0 (*)
0 0 0 0 0 0 0 0 0 0
20160325 11:18:20.001 fefs-OST0001 19 0 0 0 0 921 965738496 0 0
0 0 (*)
0 0 0 0 0 0 0 0 0 0
20160325 11:18:30.001 fefs-OST0001 17 0 0 0 0 0 0 0 0
0 0 (*)
0 0 0 0 0 0 0 0 0 0
20160325 11:18:30.001 fefs-OST0001 18 0 0 0 0 0 0 0 0
0 0 (*)
0 0 0 0 0 0 0 0 0 0
20160325 11:18:30.001 fefs-OST0001 19 0 0 0 0 887 930086912 0 0
0 0 (*)
0 0 0 0 0 0 0 0 0 0

20160325	11:18:40.001	fefs-OST0001	17	0	0	0	0	0
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:18:40.001	fefs-OST0001	18	0	0	0	0	0
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:18:40.001	fefs-OST0001	19	0	0	0	923	967835648
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:18:50.001	fefs-OST0001	17	0	0	0	0	0
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:18:50.001	fefs-OST0001	18	0	0	0	0	0
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:18:50.001	fefs-OST0001	19	0	0	0	911	955252736
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:19:00.001	fefs-OST0001	17	0	0	0	0	0
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:19:00.001	fefs-OST0001	18	0	0	0	0	0
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:19:00.001	fefs-OST0001	19	0	0	0	907	951058432
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:19:10.001	fefs-OST0001	17	0	0	0	0	0
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:19:10.001	fefs-OST0001	18	0	0	0	0	0
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:19:10.001	fefs-OST0001	19	0	0	0	939	984612864
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:19:20.001	fefs-OST0001	17	0	0	0	0	0
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:19:20.001	fefs-OST0001	18	0	0	0	0	0
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:19:20.001	fefs-OST0001	19	0	0	0	940	985661440
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:19:30.001	fefs-OST0001	17	0	0	0	0	0
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:19:30.001	fefs-OST0001	18	0	0	0	0	0
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:19:30.001	fefs-OST0001	19	0	0	0	885	927989760
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:19:40.001	fefs-OST0001	17	0	0	0	0	0
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:19:40.001	fefs-OST0001	18	0	0	0	0	0
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0
20160325	11:19:40.001	fefs-OST0001	19	7981	8368685056	115	120586240	
0	0 (*)							
0	0 0	0 0	0	0	0	0	0	0

20160325 11:19:50.001	fefs-OST0001	17	0	0	0	0	0
0 0 (*)	0 0	0 0	0 0	0 0	0 0	0 0	0 0
20160325 11:19:50.001	fefs-OST0001	18	0	0	0	0	0
0 0 (*)	0 0	0 0	0 0	0 0	0 0	0 0	0 0
20160325 11:19:50.001	fefs-OST0001	19	209	219152384	0	0	0
0 0 (*)	0 0	0 0	0 0	0 0	0 0	0 0	0 0

Remarks: Due to space limitations, each (*) location in the example shown above is an inserted line break. The wrapped text is actually displayed as one line.



For details on the collectl options, see the reference manual for `collectl(1)`.

4.22.3 Options and output information

The following table lists the `fefssv.ph` options that can be specified for this function.

Table 4.20 `fefssv.ph` options

Option	Description
v	Switches the output format. If this option is specified, the command outputs statistics for each job ID on each MDT (OST) as detailed output. You cannot use it together with the "d" option.
d	Switches the output format. If this option is specified, the command outputs statistics by volume. You cannot use it together with the "v" option.
mdt[=<MDT name>/<MDT name>/...]	Specifies the output information. If this option is specified, the command outputs MDT statistics. If an MDT name is specified, the statistics of only the specified MDT are output. You can specify multiple names by separating them with "/". Either this option or "ost" must be specified. (Volume is an optional specification.) You cannot use this option together with the "ost" option. When specifying a volume name, you need to also specify the "v" or "d" option.
ost[=<OST name>/<OST name>/...]	Specifies the output information. If this option is specified, the command outputs OST statistics. If an OST name is specified, the statistics of only the specified OST are output. You can specify multiple names by separating them with "/". Either this option or the "mdt" option must be specified. (Volume is an optional specification.) You cannot use this option together with the "mdt" option. When specifying a volume name, you need to also specify the "v" or "d" option.
fs=<FS name>[/<FS name>/...]	Specifies the output information. If this option is specified, the command outputs the statistics of only the specified file system. You can specify multiple names by separating them with "/". If the "mdt" or "ost" option specifies a volume and neither the "v" nor "d" option is specified, an error occurs.
jobid=<job ID>[/<job ID>/...]	Narrows down the output information. If this option is specified, the command outputs the statistics of only the specified job. You can specify multiple IDs by separating them with "/". The "v" option must be specified.

Output information

This function outputs the following information. The unit of aggregation for each item depends on the output pattern at the time of output.

Table 4.21 MDS output information

Item	Description
mdt_name	MDT name
jobid	Job ID
open	Number of times that open was executed within a unit time
close	Number of times that close was executed within a unit time
mknod	Number of times that mknod was executed within a unit time
link	Number of times that link was executed within a unit time
unlink	Number of times that unlink was executed within a unit time
mkdir	Number of times that mkdir was executed within a unit time
rmdir	Number of times that rmdir was executed within a unit time
rename	Number of times that rename was executed within a unit time
getattr	Number of times that getattr was executed within a unit time
setattr	Number of times that setattr was executed within a unit time
getxattr	Number of times that getxattr was executed within a unit time
setxattr	Number of times that setxattr was executed within a unit time
statfs	Number of times that statfs was executed within a unit time
sync	Number of times that sync was executed within a unit time
samedir_rename	Number of times that samedir_rename was executed within a unit time
crossdir_rename	Number of times that crossdir_rename was executed within a unit time

Table 4.22 OSS output information

Item	Description
ost_name	OST name
jobid	Job ID
read	Number of times that read was executed within a unit time
read_bytes	Quantity of read (bytes) within a unit time
write	Number of times that write was executed within a unit time
write_bytes	Quantity of write (bytes) within a unit time
getattr	Number of times that getattr was executed within a unit time
setattr	Number of times that setattr was executed within a unit time
punch	Number of times that punch was executed within a unit time
sync	Number of times that sync was executed within a unit time
destroy	Number of times that destroy was executed within a unit time
create	Number of times that create was executed within a unit time
statfs	Number of times that statfs was executed within a unit time
get_info	Number of times that get_info was executed within a unit time
set_info	Number of times that set_info was executed within a unit time

Item	Description
quotactl	Number of times that quotactl was executed within a unit time

4.23 Connecting Lustre [PG]

To connect FEFS to Lustre, you must perform the following setting work.

4.23.1 Settings on Lustre server

To enable communication between FEFS and Lustre, the receive port numbers for RPC requests and the number of RDMA fragments need to be lined up.

Add the following description to /etc/modprobe.conf at the Lustre server.

```
options lnet accept_port=30988
options ko2iblnd service=30987
options ko2iblnd map_on_demand=16
```

4.23.2 Mounting an FEFS server from a Lustre client

Mount the FEFS server from the Lustre client. The following example assumes that the IB for the FEFS server is at address 192.0.2.100.

```
# mount -t lustre 192.0.2.100@o2ib:/fsname1 /mountpoint
```

fsname1: File system name for the FEFS server to be mounted.

4.23.3 Mounting a Lustre server from an FEFS client

Mount the Lustre server from the FEFS client. The following example assumes that the IB for the Lustre server is at address 192.0.2.101..

```
# mount -t lustre 192.0.2.101@o2ib:/fsname2 /mountpoint
```

fsname2: File system name for the Lustre server to be mounted.



You cannot mount from an FX server.

Appendix A Reference

A.1 System Calls

The following "[Table A.1 Support of system calls](#)" lists the types of system calls supported by FEFS.

Yes: Supported; Limited: Supported only for advisory lock (Not supported for forced lock); VFS: Supported at VFS level; No: Not supported; No guarantee: Not guaranteed to operate properly

Table A.1 Support of system calls

System call	PG	FX
_llseek	Yes	Yes
access	Yes	Yes
bdflush	VFS	VFS
chdir	Yes	Yes
chmod	Yes	Yes
chown	Yes	Yes
chown32	Yes	Yes
chroot	Yes	Yes
close	Yes	Yes
creat	Yes	Yes
dup	VFS	VFS
dup2	VFS	VFS
execve	Yes	Yes
fchdir	Yes	Yes
fchmod	Yes	Yes
fchown	Yes	Yes
fchown32	Yes	Yes
fcntl	Limited	Limited
fcntl64	Limited	Limited
fdatasync	Yes	Yes
fgetxattr	Yes	Yes
flistxattr	Yes	Yes
flock	No	No
fremovexattr	Yes	Yes
fsetxattr	Yes	Yes
fstat	Yes	Yes
fstat64	Yes	Yes
fstatfs	Yes	Yes
fstatfs64	Yes	Yes
fsync	Yes	Yes
ftruncate	Yes	Yes
ftruncate64	Yes	Yes

System call	PG	FX
getdents	Yes	Yes
getdents64	Yes	Yes
getxattr	Yes	Yes
lchown	Yes	Yes
lchown32	Yes	Yes
lgetxattr	Yes	Yes
link	Yes	Yes
listxattr	Yes	Yes
llistxattr	Yes	Yes
lremovexattr	Yes	Yes
lseek	Yes	Yes
lsetxattr	Yes	Yes
lstat	Yes	Yes
lstat64	Yes	Yes
mkdir	Yes	Yes
mknod	Yes	Yes
mmap	Yes	Yes
mount	Yes	Yes
munmap	VFS	VFS
open	Yes	Yes
pipe	VFS	VFS
pivot_root	No	No
pread64	Yes	Yes
pwrite64	Yes	Yes
read	Yes	Yes
readdir	Yes	Yes
readlink	Yes	Yes
readv	Yes	Yes
removexattr	Yes	Yes
rename	Yes	Yes
rmdir	Yes	Yes
setrlimit	VFS	VFS
setxattr	Yes	Yes
stat	Yes	Yes
stat64	Yes	Yes
statfs	Yes	Yes
statfs64	Yes	Yes
swapon	No guarantee	No guarantee
swapoff	No guarantee	No guarantee

System call	PG	FX
symlink	Yes	Yes
sync	VFS	VFS
sysfs	Yes	No
truncate	Yes	Yes
truncate64	Yes	Yes
umount	Yes	Yes
umount2	Yes	Yes
unlink	Yes	Yes
utime	Yes	Yes
utimes	Yes	Yes
write	Yes	Yes
writev	Yes	Yes
FIEMAP ioctl	No	No
FIBMAP ioctl	No	No

A.2 Commands

A.2.1 fefs_sync

[NAME]

fefs_sync - Build or update FEFS

[SYNOPSIS]

```
/sbin/fefs_sync --setup --storage=<cluster>[,...] --compute=<cluster>[,...]
    [--directory=<directory>]
/sbin/fefs_sync --setup {--storage=<cluster>|--compute=<cluster>}
    [--nodetype=<nodetype>[,...] | --nodeid=<nodeid>[,...] |
     --nodelist=<nodelist>[--directory=<directory>]
/sbin/fefs_sync --mkfs --storage=<cluster>[,...] [--fsname=<fsname>]
/sbin/fefs_sync --mkfs --storage=<cluster>
    [--nodetype=<nodetype>[,...] | --nodeid=<nodeid>[,...] |
     --nodelist=<nodelist>[--fsname=<fsname>]
/sbin/fefs_sync --mount --storage=<cluster>[,...] --compute=<cluster>[,...] [--fsname=<fsname>]
/sbin/fefs_sync --mount {--storage=<cluster>|--compute=<cluster>}
    [--nodetype=<nodetype>[,...] | --nodeid=<nodeid>[,...] |
     --nodelist=<nodelist>[--fsname=<fsname>]
/sbin/fefs_sync --umount --storage=<cluster>[,...] --compute=<cluster>[,...] [--fsname=<fsname>]
/sbin/fefs_sync --umount {--storage=<cluster>|--compute=<cluster>}
    [--nodetype=<nodetype>[,...] | --nodeid=<nodeid>[,...] |
     --nodelist=<nodelist>[--fsname=<fsname>]
/sbin/fefs_sync --start --storage=<cluster>[,...] --compute=<cluster>[,...] [--llio]
/sbin/fefs_sync --start {--storage=<cluster>|--compute=<cluster>}
    {--nodeid=<nodeid>[,...] [--siogrp|--giogrp] | --nodelist=<nodelist>
     --nodegrp=<nodegid>[,...] | --bootgrp=<bootgid>[,...]}
    [--nodetype=<nodetype>[,...]] [--excludetype=<nodetype>[,...]]
    [--model=<model>[,...]] [--llio]
/sbin/fefs_sync --stop --storage=<cluster>[,...] --compute=<cluster>[,...] [--llio]
/sbin/fefs_sync --stop {--storage=<cluster>|--compute=<cluster>}
    {--nodeid=<nodeid>[,...] [--siogrp|--giogrp] | --nodelist=<nodelist>
     --nodegrp=<nodegid>[,...] | --bootgrp=<bootgid>[,...]}
    [--nodetype=<nodetype>[,...]] [--excludetype=<nodetype>[,...]]
```

```
[ --model=<model>[ , . . . ] [ --llio ]
/sbin/fefs_sync --router --storage=<cluster> --compute=<cluster> --hostname=<host>
--type={ disable|enable|info }
```

[DESCRIPTION]

The fefs_sync command builds or updates FEFS.

You can use this command from a system management node to execute a batch operation on a specified cluster to distribute FEFS configuration definition files to the servers and clients making up FEFS and formatting or mounting disk volumes on these servers and clients.

Execute this command on the system management node.

You need to specify both operation and common options.

Only the users who have administrator privileges can use this command.

[OPERATION OPTIONS]

--setup [--directory=<directory>]

This option creates the FEFS configuration files on the servers and clients making up FEFS.

If you want to change the directory of the distribution file, specify the --directory option. The default directory is /etc/opt/FJSVfefs/config.

--mkfs

This option executes mkfs on the FEFS server to format a disk volume.

--mount

This option mounts the disk volume on the FEFS server and mount the FEFS client on the FEFS client.

--umount

This option unmounts the disk volume on the FEFS server and mount the FEFS client on the FEFS client.

--start

This option starts the FEFS service.

--stop

This option stops the FEFS service.

--router --hostname=*host* --type=*ope*

This option changes the routing state.

In *host*, specify the node name of the router.

In *ope*, specify one of the following values:

disable: Disable

enable: Enable

info: Display the state



Note

--type=info is an option intended for displaying information for investigating an abnormality that occurred during node maintenance. Use this option when requested by a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[COMMON OPTIONS]

--storage=<cluster>

Specify the storage cluster name.

--compute=<cluster>

Specify the compute cluster name or the multiuse cluster name.

--nodeid=<nodeid> [--siogrp|--giogrp]

Specify the node ID. When --siogrp or --giogrp is specified, the command is executed for the SIO group or GIO group that includes the specified node ID.

--nodelist=<nodelist>

Specify a file that lists node IDs. Write one node ID per line in the file.

--fsname=<fsname>

Specify the file system name.

--nodetype=<nodetype>

Specify node type.

--excludetype=<nodetype>

Specify node type. If this option is specified, the command is not executed for the specified node type.

--nodegrp=<nodegid>

Specify node group ID.

--bootgrp=<bootgid>

Specify boot group ID.

--model=<model>

Specify the model. For <model>, you can specify the following abbreviations or any model defined by the user.

abbreviation	Model
PG	PRIMERGY Server
FT	FX Server

--llio

Specify this when narrowing the operation target to LLIO. When this option is specified, operation to FEFS is not performed.

[RETURN VALUES]

The following status is returned.

0: Normal end

1: Abnormal end

A.2.2 fefsconfig

[NAME]

fefsconfig - Generate an FEFS configuration file

[SYNOPSIS]

```
/usr/sbin/fefsconfig --setup
/usr/sbin/fefsconfig --make [MGS=<mgs address>]
                           [mount_option_mdt=<mount option>]
                           [mount_option_ost=<mount option>]
                           [mount_option_client=<mount option>]
                           [mount_point=<mount point>]
/usr/sbin/fefsconfig --mdtadd
/usr/sbin/fefsconfig --ostadd
/usr/sbin/fefsconfig --cleanup
```

[DESCRIPTION]

The fefsconfig command generates an FEFS configuration file.

The automatically generated FEFS configuration files are required for building an FEFS based on the configuration definition files for the FEFS setup tool.

This command must be executed on all the nodes configuring the FEFS.

Also, this command can create the configuration definition files for the FEFS setup tool.

Only the users who have administrator privileges can use this command.

[OPTIONS]

--setup

This option creates an FEFS configuration file.

--make [MGS=<mgs address>]
[mount_option_mdt=<mount option>]
[mount_option_ost=<mount option>]
[mount_option_client=<mount option>]
[mount_point=<mount point>]

This option creates configuration definition files for the FEFS setup tool.

The specifiable parameters are as follows:

MGS=<mgs address>

Specify the MGS address.

This parameter must be specified except for MGS (node).

mount_option_mdt=<mount option>

This parameter adds the option specified in "mount_option" as an MDT mount option.

The default value is "defaults,retry=6".

mount_option_ost=<mount option>

This parameter adds the option specified in "mount_option" as an OST mount option.

The default value is "defaults,retry=6".

mount_option_client=<mount option>

This parameter adds the option specified in "mount_option" as a client mount option.

The default value is "defaults,flock".

mount_point=<mount point>

This parameter specifies an FEFS mount point.

If this parameter is not specified, "/mnt/fefs" is assumed as the mount point.

--mdtadd

This option adds information on the MDT to be added, to the configuration definition file for the FEFS setup tool.

--ostadd

This option adds information on the OST to be added, to the configuration definition file for the FEFS setup tool.

--cleanup

This option deletes FEFS configuration files.

This command must be executed on the node which does not require FEFS.

[RETURN VALUES]

The following status is returned.

0: Normal end

1: Abnormal end

A.2.3 fefs_mkfs

[NAME]

fefs_mkfs - Format a volume used on FEFS

[SYNOPSIS]

```
/sbin/fefs_mkfs { -a | -f fsname | volume }
```

[DESCRIPTION]

The fefs_mkfs command formats a volume used on FEFS.

Only users who have administrator privileges on the MGS, MDS and OSS can use this command.

[OPTIONS]

-a

This option formats all volumes used by FEFS on the node where this command is executed.

You cannot use it together with another option.

-f *fsname*

This option formats the volume used by the specified file system name (*fsname*) on the node where this command is executed.

You cannot use it together with another option.

volume

This option formats the specified volume on the node where this command is executed.

You cannot use it together with another option.

[RETURN VALUES]

0: Formatting successful

Other than 0: Formatting failed

A.2.4 fefs_mount

[NAME]

fefs_mount - Mount a mount point used with FEFS

[SYNOPSIS]

```
/sbin/fefs_mount { -a | -f fsname | mountpoint | volume }
```

[DESCRIPTION]

The fefs_mount command mounts a mount point used with FEFS.

Only the users who have administrator privileges can use this command.

[OPTIONS]

-a

This option mounts all the mount points used by FEFS on the node where this command is executed.

You cannot use it together with another option.

-f *fsname*

This option mounts the mount point used by the specified file system name (*fsname*) on the node where this command is executed.

You cannot use it together with another option.

mountpoint

This option mounts the specified mount point on the node where this command is executed.

You cannot use it together with another option.

volume

This option mounts the specified volume on the node where this command is executed.

You cannot use it together with another option.

[RETURN VALUES]

The following status is returned.

0: Normal

Other than 0: Abnormal

A.2.5 **fefssnap**

[NAME]

`fefssnap` - collect materials necessary for investigation of FEFS

[SYNOPSIS]

```
/usr/sbin/fefssnap -d <outputdir>
/usr/sbin/fefssnap --help
```

[DESCRIPTION]

The `fefssnap` command collects materials necessary for investigation of FEFS.

The collected data are compressed in the form of tar + gzip, are stored in the directory specified by the following names.

`fefssnap_<timestamp>.tgz`

* timestamp: `yyyymmddHHMMSS`

Only the users who have administrator privileges can use this command.

[OPTIONS]

`-d <outputdir>`

This option specifies the directory in which to store the collected materials.

When collecting materials, this is a required option.

`--help`

This option displays the usage, and terminates.

[RETURN VALUES]

The following status is returned.

0: Normal

1: Abnormal

A.2.6 **lfs command**

`lfs` is an FEFS utility command with the following subcommands.

lfs df

[NAME]

`lfs df` - Display the current usage of disk capacity

[SYNOPSIS]

```
/usr/bin/lfs df [-ih] [mount_point]
```

[DESCRIPTION]

By default, the usage information for each FEFS file system currently mounted on that node, or the file system that contains the path if a path is specified, is displayed. It displays the current usage and total for each MDT and OST separately, and also displays a summary for each file system that matches the df (1) output for each file system.

[OPTIONS]

-i

This option displays the usage of an inode.

-h

This option makes the display easy to read.

mount_point

This option specifies the mount point of the file system.

lfs find

[NAME]

lfs find - Search files and directories

[SYNOPSIS]

```
/usr/bin/lfs find <dirname | filename>
      [[ !] {--name|-n} <pattern>]
      [--obd <uuid>]
      [--print0]
```

[DESCRIPTION]

The lfs find command searches files and directories to find those that match the given parameters.

It specifies a directory in <dirname> and a file in <filename> for the searches.

[OPTIONS]

[!] --name, -n <pattern>

Files and directories are searched to find those that match the name in this option.

If ! is attached to the option, files and directories are searched to exclude those that match.

--obd <uuid>

Searches for files on the OST with UUID specified in <uuid>.

--print0

Prints full file name, followed by a newline or NULL character correspondingly.

lfs project

[NAME]

lfs project - This command changes or displays the project attribute of the specified file or directory.

[SYNOPSIS]

```
/usr/bin/lfs project [-d|-r] <file|directory...>
/usr/bin/lfs project -s [-p ID] [-r] <file|directory...>
/usr/bin/lfs project -c [-d|-r [-p ID]] <file|directory...>
/usr/bin/lfs project -C [-r|-k] <file|directory...>
```

[DESCRIPTION]

Operate the inheritance flag to display whether to inherit the project ID and the ID to the file or directory.
Only the users who have administrator privileges can use this command.

lfs project [-d|-r] <file|directory...>

This command displays the project ID and the inheritance flag of the file or directory.

-d

This option displays the directory's own project ID and inheritance flags.

-r

This option displays the project ID and the inheritance flag of all subdirectories recursively.

Example

```
# lfs project -d dir1
  0 - dir1
# lfs project -r dir1
  0 - dir1/file1
1000 P dir1/dir2
1000 P dir1/dir2/file2
1000 P dir1/dir2/dir3
1000 P dir1/dir2/dir3/file3
```

The first number in the output represents the project ID. P indicates that the inheritance flag is set. - indicates that the inheritance flag is not set.

lfs project -s [-p ID] [-r] <file|directory...>

This command sets the project ID and the inheritance flag to file or directory. Afterwards, new files and subdirectories to be created under the specified directory inherit the project ID and inheritance flag from the parent directory.



If the -r option is not specified, the project ID and inheritance flag are set for the specified directory and the directories or files immediately below it.

For example, if the lfs project command is executed for the directory /dir1 without the -r option with the -s option, the same project ID and inheritance flag as /dir1 are set for the files directly under /dir1 and the directory /dir1/dir2. However, files under /dir1 /dir2, directories /dir1 /dir2 /dir3, and deeper file directories are not affected.

-p ID

This option specifies the project ID called *ID* to the specified file or directory.

Possible values to be specified as project ID are from 1 to 4294967295.



This command uses the above range from 1 to 4294967295 cyclically. Accordingly, when you specify 4294967296 and execute this command, the command sets 0 and does not enable the project ID. When you specify 4294967297, the command sets the project ID to 1, and then it sets 2 when you specify 4294967298 and 3 when you specify 4294967299. In addition, when you specify -1, the command sets 4294967295.

-r

This option sets an inheritance flag recursively to a subdirectory.

When you specify the -p option, it sets the specified project ID to all subdirectories.

lfs project -c [-d|-r [-p ID]] <file|directory...>

This command checks the project ID and the inheritance flag of the file and the directory and outputs abnormal values.

-p ID

This option checks whether or not to be different from the specified project ID. Without the -p option specified, it checks to the project ID on the specified directory. The following is an example of output when it is different from the specified project ID.

```
dir1/dir2/file1 - project identifier is not set (inode=6000, tree=6001)
```

-d

This option checks the project ID and the inheritance flag of the directory. The following is an example output when the inheritance flag is not set.

```
dir1/dir2/file1 - project inheritance flag is not set
```

-r

This option checks recursively the project ID and the inheritance flag to the subdirectory.

lfs project -C [-r|-k] <file|directory...>

This command clears the project ID and the inheritance flag set to the file or the directory. It clears the project ID and inheritance flags of the specified directory and the directories or files immediately below it. Project IDs and inheritance flags of directories or files deeper than that are not cleared.

-r

This option clears recursively the project ID and the inheritance flag to the specified directory and subdirectory.

-k

This option clears inheritance flags only.

lfs quota

[NAME]

lfs quota - Display the current disk usage and the usage limit, or the soft limit grace times of the QUOTA

[SYNOPSIS]

```
/usr/bin/lfs quota [-v] [-u uname| -g gname| -p pid] <mount_point>
/usr/bin/lfs quota -t {-u|-g|-p} <mount_point>
```

[DESCRIPTION]

The lfs quota command displays the current disk usage and the usage limit.

Specifies the user name *uname* if the option is "-u", the group name *gname* if the option is "-g", and the project ID *pid* to be displayed if the option is "-p". If none of the options "-u", "-g", or "-p" is specified, the usage status and usage limit of the user who executed the lfs command and the disk of the group to which the user belongs are displayed.

If you execute with "-v", the command displays the current disk usage and the usage limit of each MDTs/OSTs.

If you execute with "-t", the command displays the grace times of the QUOTA soft limit.

<mount_point> specifies the mount point of the file system.

[OPTIONS]

-u

This option displays the current usage and the grace times of soft limit of QUOTA for each user.

-g

This option displays the current usage and the grace times of soft limit of QUOTA for each group.

-p

This option displays the current usage and the grace times of soft limit of the project QUOTA.

lfs setquota

[NAME]

lfs setquota - Set QUOTA

[SYNOPSIS]

```
/usr/bin/lfs setquota {-u|-g|-p|} <name>
    [--block-softlimit <block-softlimit>]
    [--block-hardlimit <block-hardlimit>]
    [--inode-softlimit <inode-softlimit>]
    [--inode-hardlimit <inode-hardlimit>]
    <mount_point>
/usr/bin/lfs setquota {-u|-g|-p|} <name>
    [-b <block-softlimit>]
    [-B <block-hardlimit>]
    [-i <inode-softlimit>]
    [-I <inode-hardlimit>]
    <mount_point>
/usr/bin/lfs setquota -t
    {-u|-g|-p|}
    [--block-grace <block-grace>]
    [--inode-grace <inode-grace>]
    <mount_point>
/usr/bin/lfs setquota -t
    {-u|-g|-p|}
    [-b <block-grace>]
    [-i <inode-grace>]
    <mount_point>
```

[DESCRIPTION]

The lfs setquota command sets QUOTA.

<name> specifies a user name when "-u" is specified, a group name when "-g" is specified, and a target project ID for this operation when "-p" is specified, in the respective options.

<block-softlimit> sets the soft limit for the number of blocks used. It must be specified in KiB.

<block-hardlimit> sets the hard limit for the number of blocks used. It must be specified in KiB.

<inode-softlimit> sets the soft limit for the number of inodes.

<inode-hardlimit> sets the hard limit for the number of inodes.

If you execute with "-t", the command sets the grace times of the QUOTA soft limit.

<block-grace> sets the grace times of the soft limit(blocks). The grace times are specified in seconds, or in the format XXwXXdXXhXXmXXs. 1w4d means "1 week and 4 days". In addition, You cannot set the value more than 18446744073709551614 seconds in the grace times. The default value of the grace times is 7 days.

<inode-grace> sets the grace times of the soft limit(inodes) The grace times are specified in seconds, or in the format XXwXXdXXhXXmXXs. 1w4d means "1 week and 4 days". In addition, You cannot set the value more than 18446744073709551614 seconds in the grace times. The default value of the grace times is 7 days.

If you specify either <block-grace> or <inode-grace> option, grace times which you do not specify are unchanged.

<mount_point> specifies the mount point of the file system.

Root privileges are needed to execute this command.

[OPTIONS]

-u

This option sets QUOTA for a user.

-g

This option sets QUOTA for a group.

-p

This option sets Project QUOTA.

lfs setstripe

[NAME]

lfs setstripe - Set the stripe pattern

[SYNOPSIS]

```
/usr/bin/lfs setstripe [{--stripe-size|-S} stripe_size] [{--stripe-count|-c} stripe_count]
[{--stripe-index|-i} start_ost] [{--pool|-p} pool_name] <dirname|filename>
```

[DESCRIPTION]

The lfs setstripe command creates a new file that has the stripe pattern or sets the stripe pattern of an existing directory. It specifies the stripe pattern of the directory specified in *<dirname>*. Alternatively, it specifies the stripe pattern of the file specified in *<filename>*.

[OPTIONS]

--stripe-size,-S

This option sets the stripe size.

You can set a size in units of KiB, MiB, or GiB by using -S #k, -S #m, or -S #g, respectively.

--stripe-count,-c

This option sets the stripe count. If -1 is set, the writing is performed for every OST.

--stripe-index,-i

This option specifies the OST from which file writing starts. If -1 is set, file writing starts from a randomly selected OST.

--pool,-p

Use this option to specify the stripe pattern of OST_pool.

lfs getstripe

[NAME]

lfs getstripe - Display stripe pattern information

[SYNOPSIS]

```
/usr/bin/lfs getstripe [--mdt-index|-M] <dirname|filename> ...
```

[DESCRIPTION]

The lfs getstripe command displays information on the stripe pattern of the specified file or directory. It displays the stripe pattern of the directory specified in *<dirname>*. Alternatively, it displays the stripe pattern of the file specified in *<filename>*.

[OPTIONS]

--mdt-index,-M

This option displays the MDT index number for the specified file or directory.

The index number of the MDT that manages the specified file or directory is displayed.

```
[Client]
$ lfs getstripe -M something
1
```

lfs getdirstripe

[NAME]

lfs getdirstripe - This command lists the stripe pattern of the directory.

[SYNOPSIS]

```
/usr/bin/lfs getdirstripe [--mdt-count|-c] [--mdt-index|-i] [--recursive|-r] [{--obd|-O} <uuid>]
<dir>...
```

[DESCRIPTION]

This command acquires the stripe pattern information of the directory specified with *<dir>*. You can specify *<dir>* more than once.

[OPERATION OPTIONS]

-c, --mdt-count

This option displays only the stripe count of the directory.

-i, --mdt-index

This option displays only the stripe index of the directory.

-r, --recursive

This option recursively tracks back all subdirectories in the specified directory and lists the striping information.

-O, --obd <uuid>

This option displays only the stripe information of the directory on the MDT to which UUID is specified with *<uuid>*.

lfs mkdir

[NAME]

lfs mkdir - Create a directory on an MDT

[SYNOPSIS]

```
/usr/bin/lfs mkdir [ {-c | --count} <stripe_count>] [ {-i | --index} <mdt_idx>] <dir>...
```

[DESCRIPTION]

This command creates the directory specified with *<dir>* on a MDT.

Execute this command on the client node.

Only the users who have administrator privileges can use this command.

To delete the created directory, it is possible to delete by executing the rmdir command (/bin/rmdir) just like an ordinary directory.

[OPTIONS]

-i,--index <mdt_index>

This option creates a directory on an MDT starting with the number *<mdt_index>*. Unless the -c option is specified, this option is required.

-c,--count <stripe_count>

This option creates the stripe directory of which the stripe count is *<stripe_count>*. Unless the -i option is specified, this option is required.

lfs pool_list

[NAME]

lfs pool_list - Display either the OST_pool list or the OSTs in OST_pool.

[SYNOPSIS]

```
/usr/bin/lfs pool_list <fsname>[.<poolname>] | <mount_point>
```

[DESCRIPTION]

The lfs pool_list command displays a list of the OST_pool defined by *<fsname>*.

It displays a list of the OSTs included in the pool defined by *<fsname>[.<poolname>]*.

<mount_point> specifies the mount point of the file system.

OST_pool is a function for bundling multiple specified OSTs into a group and assigning files and directories to the OSTs in the group.

lfs fid2path

[NAME]

lfs fid2path - FID Output the file path name corresponding to FID

[SYNOPSIS]

```
/usr/bin/lfs fid2path [--link <linkno>] <fsname|rootpath> <fid> ...
```

[DESCRIPTION]

Print out the pathname(s) for the specified *<fid(s)>* from the filesystem mounted at *<rootpath>* or named *<fsname>*. If a file has multiple hard links, then all of the pathnames for that file are printed, unless --link limits the printing to only the specified link number (starting at 0, in no particular order). If multiple fids are specified, but only a single pathname is needed for each file, use --link 0.

Root privileges are needed to execute this command.

lfs help

[NAME]

lfs help - Display the lfs help.

[SYNOPSIS]

```
/usr/bin/lfs help [command]
```

[DESCRIPTION]

This command displays the lfs help.

[OPTIONS]

[command]

This option specifies the command to display about helpful information.

lfs --list-commands

[NAME]

lfs --list-commands - Displays a list of subcommands

[SYNOPSIS]

```
/usr/bin/lfs --list-commands
```

[DESCRIPTION]

This command displays a list of subcommands.

A.2.7 lctl command

lctl is an FEFS command for low-level configuration control with the following subcommands.

lctl device_list

[NAME]

lctl device_list - Display file system configuration information

[SYNOPSIS]

```
/usr/sbin/lctl device_list [-t]
```

[DESCRIPTION]

The lctl device_list command displays file system configuration information.

Root privileges are needed to execute this command.

[OPTIONS]

-t

If run on a client node, it prints with an NID.

lctl list_nids

[NAME]

lctl list_nids - Display valid NIDs

[SYNOPSIS]

```
/usr/sbin/lctl list_nids
```

[DESCRIPTION]

The command displays all the NIDs of the servers available for communication from the relevant node. The following shows a display example.

```
[MDS node, OSS node, Client node]
# lctl list_nids
192.0.2.1@o2tofu
192.0.2.1@o2tofu2
192.0.2.1@o2tofu514
198.51.100.23@o2ib
203.0.113.216@tcp
```

Root privileges are needed to execute this command.

lctl qos

[NAME]

lctl qos - Command used for QoS function-related settings

[SYNOPSIS]

```
/usr/sbin/lctl qos < on [filepath] | off | stat | check [filepath] >
```

[DESCRIPTION]

The lctl qos command sets the QoS function. This command must be executed on the MDS. Also, if your environment is multi-MDS, execute this command on the MDT0.

Root privileges are needed to execute this command.

[ARGUMENTS]

on [filepath]

This argument starts QoS control with the specified QoS definition file.

filepath specifies an absolute path. The default *filepath* is /etc/opt/FJSVfefs/qosserver.conf.

off

This argument terminates QoS control.

stat

This argument displays the QoS status on the first line.

"QoS is Enable." The QoS function is enabled.

"QoS is Disable." The QoS function is disabled.

If the QoS function is enabled, this argument displays the QoS definition file path name on the second line and the QoS definition file contents on the third and subsequent lines.

check [filepath]

This argument checks the syntax of the specified QoS definition file.

filepath specifies an absolute path. The default filepath is /etc/opt/FJSVfefs/qosserver.conf.

lctl sqos

[NAME]

lctl sqos - displays information about the QoS function of servers.

[SYNOPSIS]

```
/usr/sbin/lctl sqos [oss] <thread_top | thread_all | thread_user <user-id> | thread_node | ost_io | clear>
```

[DESCRIPTION]

The command displays information about the QoS function of servers. If you want to display information of the MDS, execute this command on the MDS. Also, if your environment is multi-MDS, execute this command for each MDS. If you want to display information of the OSS, execute this command on the OSS.

The displayed information is the total of from the time you enable QoS function. Information about the user who made last file access more than 30 days ago is not be displayed. Also, If the number of users more than 1000 people access the file, information of up to 1000 people will be displayed.

When the QoS function on the server is invalid, this command is not available.

Root privileges are needed to execute this command.

[OPTIONS]

oss

When both the MDS and the OSS are on the same node, if you want to display information about the OSS, specify this option.

[ARGUMENTS]

thread_top

This argument displays up to 10 people in decreasing order, the information of the users number of requests from clients.

thread_all

This argument displays the information of all users. Display order is out of order.

thread_user <user-id>

This argument displays the information of the specified user.

thread_node

This argument displays the information of each node group defined in the QoS definition file.

ost_io

This argument displays the access status of each OST.

This argument can be executed only when you run this command on the OSS.

clear

This argument clears statistics of the QoS function on the server

You can also clear the statistics by invalidating (lctl qos off command) on the server.

Display example of thread_top

Execution example on the MDS

# lctl sqos thread_top												
nodegrp= 1		---thread---			---wait_req--		-wait_time(usec)-		-exec_time(usec)-			
No.	uid	exec_cnt	cur	max	lim	cur	max	max	avg	max	avg	last_update
09:34:49		1079	1446260	0	2	12	0	2	638	18	48516	77 2013/08/14
09:34:49		1078	766438	0	2	12	0	2	361	20	38156	102 2013/08/14

3	1071	187964	0	3	12	0	2	348	20	40080	66	2013/08/14
09:34:49												
4	1072	168619	0	2	12	0	2	485	20	37456	54	2013/08/14
09:34:49												
5	1073	129596	0	2	12	0	1	305	20	30598	53	2013/08/14
09:34:49												
6	1074	125128	0	4	12	0	2	389	20	4843202	1495	2013/08/14
09:34:49												
7	0	41672	0	1	12	0	1	227	19	34176	122	2013/08/14 09:34:49
09:34:49												
8	1076	1176	0	1	12	0	1	242	21	16362	124	2013/08/14
09:34:49												
9	1075	36	0	1	12	0	1	117	42	16614	598	2013/08/14 09:34:08
10	1077	12	0	2	12	0	1	156	68	1858142	156280	2013/08/14
09:34:33												
nodegrp= 2			---thread---	---	wait_req--		wait_time(usec)-		exec_time(usec)-			
No.	uid	exec_cnt	cur	max	lim	cur	max	max	avg	max	avg	last_update
1	1070	642	0	2	12	0	2	186	20	2948302	18338	2013/08/14
09:35:06												
2	1053	420	0	1	12	0	1	150	20	573	58	2013/08/14 09:34:48
QoS command was completed.												

Execution example on the OSS

# lctl sqos thread_top												
nodegrp= 1			---thread---	---	wait_req--		wait_time(usec)-		exec_time(usec)-			
No.	uid	exec_cnt	cur	max	lim	cur	max	max	avg	max	avg	last_update
1	1076	56332	0	15	128	0	10	2004	102	7930509	47401	2013/08/14
09:35:30												
2	1075	4202	0	100	128	0	55	16627	500	7809997	426544	2013/08/14
09:35:34												
3	1077	2123	0	8	128	0	6	1683	214	7809594	426322	2013/08/14
09:34:49												
4	1071	907	0	9	128	0	3	850	89	7015763	176352	2013/08/14
09:35:34												
5	0	813	0	4	128	0	3	1007	91	4098120	190510	2013/08/14
09:35:14												
6	1074	554	0	5	128	0	4	894	109	6757076	179418	2013/08/14
09:35:01												
7	1073	184	0	7	128	0	6	1344	118	1331260	72051	2013/08/14
09:34:48												
8	1072	105	0	3	128	0	2	634	94	2371996	81728	2013/08/14
09:34:37												
9	1079	2	0	1	128	0	1	131	123	220	191	2013/08/14 09:30:05
10	1078	1	0	1	128	0	1	106	106	97	97	2013/08/14 09:30:04
nodegrp= 2			---thread---	---	wait_req--		wait_time(usec)-		exec_time(usec)-			
No.	uid	exec_cnt	cur	max	lim	cur	max	max	avg	max	avg	last_update
1	1070	21707	0	38	128	0	30	2288	211	7931280	329789	2013/08/14
09:35:14												
2	1053	5845	0	20	128	0	9	1906	157	8461552	355542	2013/08/14
09:35:19												
QoS command was completed.												

The description of each item is the following.

Item name	Description
nodegrp=	Number that corresponds to the node group defined in nodegrp[1-10] of the QoS definition file.

Item name	Description
	If the number is "undef", there is access from the node that has not been defined in the QoS definition file. Check the IP address defined in nodegrp[1-10] of the QoS definition file.
No.	Displays up to 10 people in descending order of exec_cnt
uid	User ID
exec_cnt	Number of the request from the client executed by the server thread.
thread	Number of the server thread that are running. cur : current value above max : maximum value above lim : the upper limit of the number of the executable server thread. This value is determined by the maximum value of the number of server threads(percentage) specified in the QoS definition file.
wait_req	Number of the request waiting for running of the server thread, cur : current value above max : maximum value above
wait_time(usec)	Latency time from the arrival of the request from the client at the server to beginning execution by the server thread. The units are indicated in microseconds. max : maximum value above avg : average value above
exec_time(usec)	Execution time of when the request from the client is executed by the server thread. The units are indicated in microseconds. max : maximum value above avg : average value above
last_update	Last date and time that you run the server thread.

Display example of thread_all

Execution example on the MDS

```
# lctl sqos thread_all
nodegrp= 1           ---thread---  ---wait_req--  -wait_time(usec)-  -exec_time(usec)-
      uid    exec_cnt  cur max lim   cur   max     max   avg   max   avg   last_update
      0       41672    0   1  12    0     1     227    19    34176   122  2013/08/14 09:34:49
      1076      1176    0   1  12    0     1     242    21    16362   124  2013/08/14
09:34:49
      1073      129596   0   2  12    0     1     305    20    30598    53  2013/08/14
09:34:49
      1078      766438   0   2  12    0     2     361    20    38156   102  2013/08/14
09:34:49
      1075       36     0   1  12    0     1     117    42    16614    598  2013/08/14 09:34:08
      1072     168619    0   2  12    0     2     485    20    37456    54  2013/08/14
09:34:49
      1077       12     0   2  12    0     1     156    68    1858142  156280  2013/08/14
09:34:33
      1074     125128    0   4  12    0     2     389    20    4843202   1495  2013/08/14
09:34:49
      1079     1446260   0   2  12    0     2     638    18    48516    77  2013/08/14
09:34:49
      1071     187964   0   3  12    0     2     348    20    40080    66  2013/08/14
09:34:49
```

```

nodegrp= 2          ---thread---  ---wait_req-- -wait_time(usec)- -exec_time(usec)-
    uid      exec_cnt  cur max lim   cur   max     max   avg   max   avg   last_update
    1070        642     0   2   12     0     2     186    20   2948302   18338  2013/08/14
09:35:06
    1053        420     0   1   12     0     1     150    20     573    58  2013/08/14 09:34:48

QoS command was completed.

```

*Description of each item is the same as lctl sqos thread_top described above.

Execution example on the OSS

```

# lctl sqos thread_all
nodegrp= 1          ---thread---  ---wait_req-- -wait_time(usec)- -exec_time(usec)-
    uid      exec_cnt  cur max lim   cur   max     max   avg   max   avg   last_update
    0        813     0   4   128     0     3     1007    91   4098120   190510  2013/08/14
09:35:14
    1076      56332    0  15  128     0    10     2004   102   7930509   47401   2013/08/14
09:35:30
    1073       184     0   7   128     0     6     1344   118   1331260   72051   2013/08/14
09:34:48
    1078       105     0   1   128     0     1     106    106     97    97  2013/08/14 09:30:04
    1075      4202     0 100  128     0    55     16627   500   7809997   426544  2013/08/14
09:35:34
    1072       105     0   3   128     0     2     634    94   2371996   81728   2013/08/14
09:34:37
    1077      2123     0   8   128     0     6     1683   214   7809594   426322  2013/08/14
09:34:49
    1074       554     0   5   128     0     4     894   109   6757076   179418   2013/08/14
09:35:01
    1079       205     0   1   128     0     1     131    123     220    191  2013/08/14 09:30:05
    1071      907     0   9   128     0     3     850    89   7015763   176352   2013/08/14
09:35:34

nodegrp= 2          ---thread---  ---wait_req-- -wait_time(usec)- -exec_time(usec)-
    uid      exec_cnt  cur max lim   cur   max     max   avg   max   avg   last_update
    1070      21707    0   38  128     0    30     2288   211   7931280   329789  2013/08/14
09:35:14
    1053       5845    0   20  128     0     9     1906   157   8461552   355542  2013/08/14
09:35:19

QoS command was completed.

```

*Description of each item is the same as lctl sqos thread_top described above.

Display example of thread_user

Execution example on the MDS

```

# lctl sqos thread_user 1076
uid(name)= 1076(fefs_guest06)
          ---thread---  ---wait_req-- -wait_time(usec)- -exec_time(usec)-
nodegrp      exec_cnt  cur max lim   cur   max     max   avg   max   avg   last_update
    1         1176     0   1   12     0     1     242    21   16362   124  2013/08/14 09:34:49
    2           0     0   0   12     0     0     0     0     0     0     0
QoS command was completed.

```

*Description of each item is the same as lctl sqos thread_top described above.

Execution example on the OSS

```

# lctl sqos thread_user 1076
uid(name)= 1076(fefs_guest06)
          ---thread---  ---wait_req-- -wait_time(usec)- -exec_time(usec)-
nodegrp      exec_cnt  cur max lim   cur   max     max   avg   max   avg   last_update
    1         56332    0  15  128     0    10     2004   102   7930509   47401  2013/08/14

```

```

09:35:30
      0   0   0 128      0   0      0   0      0   0
QoS command was completed.

```

*Description of each item is the same as lctl sqos thread_top described above.

Display example of thread_node

Execution example on the MDS

```

# lctl sqos thread_node
                                ---thread---  ---wait_req--  -wait_time(usec)-  -exec_time(usec)-
nodegrp      exec_cnt  cur max lim    cur   max       max     avg      max   avg      last_update
  1          2866901    0   4   12      0     2       638     19     4843202   144  2013/08/14
09:34:49
  2            1062     0   2   12      0     2       186     20     2948302  11108  2013/08/14
09:35:06
QoS command was completed.

```

*Description of each item is the same as lctl sqos thread_top described above.

Execution example on the OSS

```

# lctl sqos thread_node
                                ---thread---  ---wait_req--  -wait_time(usec)-  -exec_time(usec)-
nodegrp      exec_cnt  cur max lim    cur   max       max     avg      max   avg      last_update
  1          65223     0 100 128      0     55      16627    131     7930509   88982  2013/08/14
09:35:34
  2          27552     0   38 128      0     30      2288     199     8461552  335253  2013/08/14
09:35:19
QoS command was completed.

```

*Description of each item is the same as lctl sqos thread_top described above.

Display example of ost_io

Execution example on the OSS

```

# lctl sqos ost_io
                                --io_time(usec)--
ost_name      io_cnt        max     avg
  fefs-OST0000    23969    8461373  446831
  fefs-OST0001    17691    3028350  234968
QoS command was completed.

```

*The description of each item is the following.

Item name	Description
ost_name	Name of OST that accesses the disk.
io_cnt	Execution frequency of the disk access.
io_time(usec)	Disk access time per once. The units are indicated in microseconds. max : maximum value above avg : average value above

lctl cqos

[NAME]

lctl cqos - displays information about the QoS function of clients.

[SYNOPSIS]

```
/usr/sbin/lctl cqos < meta_top | read_top | write_top |
      meta_all | read_all | write_all |
```

```
meta_user <user-id> | read_user <user-id> | write_user <user-id> |
cache_top | cache_all | cache_user <user-id> | clear >
< mount-point >
```

[DESCRIPTION]

The command displays information about the QoS function of clients. Execute this command on the client node.

The displayed information is the total of from the time you enable QoS function. Information about the user who made last file access more than 30 days ago is not be displayed. Also, If the number of users more than 1000 people access the file, information of up to 1000 people will be displayed.

When the QoS function on the clients is invalid, this command is not available.

Root privileges are needed to execute this command.

[ARGUMENTS]

meta_top

This argument displays up to 10 people in decreasing order, the information of the users who often run the number of meta-operation.

read_top

This argument displays up to 10 people in decreasing order, the information of the users who often run the number of read.

write_top

This argument displays up to 10 people in decreasing order, the information of the users who often run the number of write.

meta_all

This argument displays the information of all the users who run meta-operation. Display order is in no particular order.

read_all

This argument displays the information of all the users who run read. Display order is in no particular order.

write_all

This argument displays the information of all the users who run write. Display order is in no particular order.

meta_user <user-id>

This argument displays the information of the specified user's meta-operation.

read_user <user-id>

This argument displays the information of the specified user's read.

write_user <user-id>

This argument displays the information of the specified user's write.

cache_top

This argument displays up to 10 people in decreasing order, the information of the users number of using write cache.

cache_all

This argument displays the information of all the users who use write cache. Display order is in no particular order.

cache_user <user-id>

This argument displays the information of the specified user's using write cache.

clear

This argument clears statistics of the QoS function on the client(request control, cache control).

mount-point

This argument specifies FEFS mount point.

Display example of meta_top

```
# lctl c qos meta_top /mnt/fefs/
mclientmax=4 mrootmax=1 musermax=1
```

<user info>									
No.	uid	exec_cnt	-----total_wait_cnt----		--own_wtime(usec)--		-other_wtime(usec)-		
			own	other	max	avg	max	avg	last_update
1	1079	1196513	95267	655	53656	174	5536	118	2013/09/04
2	1078	494573	50474	717	43589	192	31710	311	2013/09/04
3	1071	146154	2459	525	41067	338	50073	298	2013/09/04
4	1072	112171	12	503	9534	1069	33345	308	2013/09/04
5	1073	98563	9	500	1016	248	33034	321	2013/09/04
6	1074	38232	19	356	1645	383	30111	210	2013/09/04
7	0	24673	0	238	0	0	21814	221	2013/09/04 19:33:26
8	1076	519	0	7	0	0	185	77	2013/09/04 19:33:16
9	1053	341	27	10	460	121	211	62	2013/09/04
10	1075	116	0	0	0	0	0	0	2013/09/04 19:32:35
CQoS command was completed.									

*The description of each item is the following.

Item name	Description
mclientmax=4 mrootmax=1 musermax=1	Value of qos option specified in FEFS design sheet. For the meaning of each option, see " How to configure an FEFS client ".
No.	Displays up to 10 people in descending order of exec_cnt
uid	User ID
exec_cnt	Number of executions of the request
total_wait_cnt	Frequency in which it waits for execution by the QoS function by the time the request is executed. own : frequency in which it meets the execution of request for that it reached the upper limit of the concurrent number of requests per one user. other : frequency in which it meets the execution of request for that it reached the upper limit of the concurrent number of requests in the client node.
own_wtime(usec)	Time to have waited for the execution of request when it reached the upper limit of the concurrent number of requests per one user. The units are indicated in microseconds. max : maximum value above avg : average value above
other_wtime(usec)	Time to have waited for the execution of request when it reached the upper limit of the concurrent number of requests in the client node. The units are indicated in microseconds. max : maximum value above avg : average value above
last_update	Last date and time that you run the request.

Display example of read_top

```
# lctl cqos read_top /mnt/fefs/
rdclientmax=8 rdrootmax=2 rdusermax=2
<user info>
-----total_wait_cnt---- --own_wtime(usec)-- -other_wtime(usec)-

```

No.	uid	exec_cnt	own	other	max	avg	max	avg	last_update
1	1073	926011	0	0	0	0	0	0	2013/09/04 19:33:19
2	1072	713838	14	0	592974	42473	0	0	2013/09/04
19:33:19									
3	1071	452718	116	0	367087	8943	0	0	2013/09/04
19:33:19									
4	0	43804	0	0	0	0	0	0	2013/09/04 19:33:18
5	1074	23914	0	0	0	0	0	0	2013/09/04 19:33:19
6	1076	1610	0	0	0	0	0	0	2013/09/04 19:33:16
7	1053	34	5	0	1251163	884525	0	0	2013/09/04
19:33:16									
8	1075	29	0	0	0	0	0	0	2013/09/04 19:32:33
9	1078	25	0	0	0	0	0	0	2013/09/04 19:31:35
10	1079	25	0	0	0	0	0	0	2013/09/04 19:31:35
CQoS command was completed.									

*Description of each item is the same as lctl cqos meta_top described above.

Display example of write_top

# lctl cqos write_top /mnt/fefs/ wrclientmax=8 wrrootmax=2 wrusermax=2										
<user info>										
No.	uid	exec_cnt	own	other	-----total_wait_cnt-----	--own_wtime(usec)--	-other_wtime(usec)-	max	avg	last_update
1	1076	15830	0	59	0	0	104642	2636	2013/09/04	
19:33:18										
2	1075	12654	0	58	0	0	1030596	39467	2013/09/04	
19:33:18										
3	1074	6606	17	28	1742873	500635	478237	57307	2013/09/04	
19:33:19										
4	1071	3410	0	31	0	0	631005	74418	2013/09/04	
19:33:19										
5	0	2776	0	21	0	0	1694551	247850	2013/09/04	
19:33:18										
6	1053	2445	1959	27	2715973	65176	194776	15607	2013/09/04	
19:33:19										
7	1073	384	0	18	0	0	1030038	90040	2013/09/04	
19:33:05										
8	1072	271	0	1	0	0	110	110	2013/09/04 19:33:02	
9	1077	40	0	0	0	0	0	0	2013/09/04 19:33:17	
10	1078	1	0	0	0	0	0	0	2013/09/04 19:35:32	
CQoS command was completed.										

*Description of each item is the same as lctl cqos meta_top described above.

Display example of meta_all

# lctl cqos meta_all /mnt/fefs/ mclientmax=4 mrootmax=1 musermax=1									
<user info>									
uid	exec_cnt	own	other	-----total_wait_cnt-----	--own_wtime(usec)--	-other_wtime(usec)-	max	avg	last_update
0	24677	0	238	0	0	21814	221	2013/09/04 19:37:27	
1076	519	0	7	0	0	185	77	2013/09/04 19:33:16	
1073	98563	9	500	1016	248	33034	321	2013/09/04 19:33:19	
1078	494590	50474	717	43589	192	31710	311	2013/09/04 19:35:32	
1075	116	0	0	0	0	0	0	2013/09/04 19:32:35	
1072	112171	12	503	9534	1069	33345	308	2013/09/04 19:33:19	
1077	43	0	0	0	0	0	0	2013/09/04 19:31:34	
1074	38232	19	356	1645	383	30111	210	2013/09/04 19:33:19	
1053	341	27	10	460	121	211	62	2013/09/04 19:33:16	
1079	1196513	95267	655	53656	174	5536	118	2013/09/04 19:33:19	

```

1071      146154      2459      525      41067      338      50073      298  2013/09/04 19:33:19
CQoS command was completed.

```

*Description of each item is the same as lctl cqos meta_top described above.

Display example of read_all

```

# lctl cqos read_all /mnt/fefs/
rdclientmax=8 rdrootmax=2 rdusermax=2
<user info>
          -----total_wait_cnt----  --own_wtime(usec)--  -other_wtime(usec)-
uid    exec_cnt      own      other      max      avg      max      avg      last_update
  0        43804       0       0       0       0       0       0  2013/09/04 19:33:18
1076      1610       0       0       0       0       0       0  2013/09/04 19:33:16
1073     926011       0       0       0       0       0       0  2013/09/04 19:33:19
1078       25       0       0       0       0       0       0  2013/09/04 19:31:35
1075       29       0       0       0       0       0       0  2013/09/04 19:32:33
1072    713838       14       0   592974   42473       0       0  2013/09/04 19:33:19
1077       16       0       0       0       0       0       0  2013/09/04 19:31:36
1074    23914       0       0       0       0       0       0  2013/09/04 19:33:19
1053       34       5       0  1251163  884525       0       0  2013/09/04 19:33:16
1079       25       0       0       0       0       0       0  2013/09/04 19:31:35
1071    452718      116       0  367087   8943       0       0  2013/09/04 19:33:19
CQoS command was completed.

```

*Description of each item is the same as lctl cqos meta_top described above.

Display example of write_all

```

# lctl cqos write_all /mnt/fefs/
wrclientmax=8 wrrootmax=2 wrusermax=2
<user info>
          -----total_wait_cnt----  --own_wtime(usec)--  -other_wtime(usec)-
uid    exec_cnt      own      other      max      avg      max      avg      last_update
  0        2776       0      21       0       0  1694551  247850  2013/09/04 19:33:18
1076     15830       0      59       0       0  104642   2636  2013/09/04 19:33:18
1073      384       0      18       0       0  1030038  90040  2013/09/04 19:33:05
1078       1       0       0       0       0       0       0  2013/09/04 19:35:32
1075     12654       0      58       0       0  1030596  39467  2013/09/04 19:33:18
1072      271       0       1       0       0      110    110  2013/09/04 19:33:02
1077       40       0       0       0       0       0       0  2013/09/04 19:33:17
1074     6606       17      28  1742873  500635  478237  57307  2013/09/04 19:33:19
1053     2445      1959       27  2715973  65176  194776  15607  2013/09/04
19:33:19
  1071     3410       0      31       0       0  631005  74418  2013/09/04 19:33:19
CQoS command was completed.

```

*Description of each item is the same as lctl cqos meta_top described above.

Display example of meta_user

```

# lctl cqos meta_user 1076 /mnt/fefs/
uid(name) = 1076(fefs_guest06)
mclientmax=4 mrootmax=1 musermax=1
<user info>
          -----total_wait_cnt----  --own_wtime(usec)--  -other_wtime(usec)-
uid    exec_cnt      own      other      max      avg      max      avg      last_update
1076      519       0       7       0       0      185     77  2013/09/04 19:33:16
CQoS command was completed.

```

*Description of each item is the same as lctl cqos meta_top described above.

Display example of read_user

```

# lctl cqos read_user 1076 /mnt/fefs/
uid(name) = 1076(fefs_guest06)

```

```

rdclientmax=8 rdrootmax=2 rdusermax=2
<user info>
    -----total_wait_cnt---- --own_wtime(usec)-- -other_wtime(usec)-
    uid      exec_cnt      own     other      max      avg      max      avg      last_update
    1076        1610         0       0        0       0        0       0   2013/09/04 19:33:16
CQoS command was completed.

```

*Description of each item is the same as lctl cqos meta_top described above.

Display example of write_user

```

# lctl cqos write_user 1076 /mnt/fefs/
uid(name) = 1076(fefs_guest06)
wrclientmax=8 wrrootmax=2 wrusermax=2
<user info>
    -----total_wait_cnt---- --own_wtime(usec)-- -other_wtime(usec)-
    uid      exec_cnt      own     other      max      avg      max      avg      last_update
    1076        15830        0      59        0       0    104642     2636  2013/09/04 19:33:18
CQoS command was completed.

```

*Description of each item is the same as lctl cqos meta_top described above.

Display example of cache_top

```

# lctl cqos cache_top /mnt/fefs/
dprotoxmax=10 dpusermax=10
<user info>
    -----total_wait_cnt---- --own_wtime(usec)-- -other_wtime(usec)-
No.  uid write_page_cnt      own     other      max      avg      max      avg      last_update
  1  1076      309590      774       0    1618578  131839       0       0  2013/09/04
19:33:19
  2  1075      202432      591       0    2244578  170829       0       0  2013/09/04
19:33:19
  3  1053      118420      171       0    1439234  142419       0       0  2013/09/04
19:33:19
  4  1077      81922       253       0    1495088  163892       0       0  2013/09/04
19:33:19
  5  1071      12411        18       0    1380679  368360       0       0  2013/09/04
19:33:12
  6  1073      4608          3       0    648024  299229       0       0  2013/09/04
19:33:05
  7  1072      2508          0       0        0       0       0       0  2013/09/04 19:33:02
  8  0          1501          0       0        0       0       0       0  2013/09/04 19:33:18
  9  1074      1442          0       0        0       0       0       0  2013/09/04 19:33:18
10  1078      245           0       0        0       0       0       0  2013/09/04 19:35:32
CQoS command was completed.

```

*The description of each item is the following.

Item name	Description
dprotoxmax=10 dpusermax=10	Value of qos option specified in FEFS design sheet. For the meaning of each option, see " How to configure an FEFS client ".
No.	Displays up to 10 people in descending order of write_page_cnt.
uid	User ID
write_page_cnt	Number of pages written to the client cache.
total_wait_cnt	Frequency in waiting for the execution by the QoS function by writing on the client cache. own : frequency in waiting for the writing for that it reached the upper limit of the ratio per one user.

Item name	Description
	other : frequency in waiting for the writing for that it reached the upper limit of the client cache available on the client node.
own_wait_time(usec)	Time to have waited for the writing when it reached the upper limit of the ratio per one user. The units are indicated in microseconds. max : maximum value above avg : average value above
other_wait_time(usec)	Time to have waited for the writing when it reached the upper limit of the client cache available on the client node. max : maximum value above avg : average value above
last_update	Last date and time that you wrote to the client cache.

Display example of cache_all

```
# lctl cqos cache_all /mnt/fefs/
dproto=10 dpusermax=10
<user info>
      -----total_wait_cnt----  --own_wtime(usec)--  -other_wtime(usec)-
uid write_page_cnt      own      other      max      avg      max      avg      last_update
  0       1501        0        0        0        0        0        0  2013/09/04 19:33:18
1076     309590      774        0    1618578    131839        0        0  2013/09/04 19:33:19
1073     4608         3        0    648024    299229        0        0  2013/09/04 19:33:05
1078     245          0        0        0        0        0        0  2013/09/04 19:35:32
1075     202432      591        0    2244578    170829        0        0  2013/09/04 19:33:19
1072     2508          0        0        0        0        0        0  2013/09/04 19:33:02
1077     81922        253        0    1495088    163892        0        0  2013/09/04 19:33:19
1074     1442          0        0        0        0        0        0  2013/09/04 19:33:18
1053     118420        171        0    1439234    142419        0        0  2013/09/04 19:33:19
1071     12411         18        0    1380679    368360        0        0  2013/09/04 19:33:12
CQoS command was completed.
```

*Description of each item is the same as lctl cqos cache_top described above.

Display example of caches_user

```
# lctl cqos cache_user 1076 /mnt/fefs/
uid(name) = 1076(fefs_guest06)
dproto=10 dpusermax=10
<user info>
      -----total_wait_cnt----  --own_wtime(usec)--  -other_wtime(usec)-
uid write_page_cnt      own      other      max      avg      max      avg      last_update
1076     309590      774        0    1618578    131839        0        0  2013/09/04 19:33:19
CQoS command was completed.
```

*Description of each item is the same as lctl cqos cache_top described above.

lctl pool_list

[NAME]

lctl pool_list - Display a list of the OSTs included in a pool and an OST_pool list

[SYNOPSIS]

```
/usr/sbin/lctl pool_list <fsname>[ .<poolname> ] | <mount_point>
```

[DESCRIPTION]

The lctl pool_list command displays a list of the OST_pool defined by <fsname>.

It displays a list of the OSTs included in the pool defined by <fsname>[.<poolname>].

<mount_point> specifies the mount point of the file system.

OST_pool is a function for bundling multiple specified OSTs into a group and assigning files and directories to the OSTs in the group.

lctl pool_new

[NAME]

lctl pool_new - Create OST_pool

[SYNOPSIS]

```
/usr/sbin/lctl pool_new <fsname>.<poolname>
```

[DESCRIPTION]

The lctl pool_new command creates the new OST_pool defined by <fsname>.<poolname>.

Root privileges are needed to execute this command.

lctl pool_destroy

[NAME]

lctl pool_destroy - Delete OST_pool

[SYNOPSIS]

```
/usr/sbin/lctl pool_destroy <fsname>.<poolname>
```

[DESCRIPTION]

The lctl pool_destroy command deletes the OST_pool defined by <fsname>.<poolname>.

Root privileges are needed to execute this command.

lctl pool_add

[NAME]

lctl pool_add - Add the OST to OST_pool

[SYNOPSIS]

```
/usr/sbin/lctl pool_add <fsname>.<poolname> <ostname indexed list>
```

[DESCRIPTION]

The command adds the OST defined by <ostname indexed list> to the OST_pool defined by <fsname>.<poolname>.

Root privileges are needed to execute this command.

lctl pool_remove

[NAME]

lctl pool_remove - Remove the OST from OST_pool

[SYNOPSIS]

```
/usr/sbin/lctl pool_remove <fsname>.<poolname> <ostname indexed list>
```

[DESCRIPTION]

The lctl pool_remove command removes the OST defined by <ostname indexed list> from the OST_pool defined by <fsname>.<poolname>.

Root privileges are needed to execute this command.

lctl ping

[NAME]

lctl ping - Check communication through the LNet connection.

[SYNOPSIS]

```
/usr/sbin/lctl ping <nid>
```

[DESCRIPTION]

The lctl ping command checks communication through the LNet connection specified by *<nid>*.

Root privileges are needed to execute this command.

lctl set_param

[NAME]

lctl set_param - Set the value of an FEFS parameter.

[SYNOPSIS]

```
/usr/sbin/lctl set_param <parameter>=<value> ...>
```

[DESCRIPTION]

Set *<value>* in *<parameter>*.

Root privileges are needed to execute this command.

lctl get_param

[NAME]

lctl get_param - Get the value of an FEFS parameter.

[SYNOPSIS]

```
/usr/sbin/lctl get_param <parameter> ...>
```

[DESCRIPTION]

Print the value specified in *<parameter>*.

lctl conf_param

[NAME]

lctl conf_param - This command sets parameter to device.

[SYNOPSIS]

```
/usr/sbin/lctl conf_param [-d] <device|fsname>.<parameter>=<value>
```

[DESCRIPTION]

This command sets a permanent parameter to a certain device through MGS.

This command sets *<value>* to *<parameter>* of the device or file system specified with *<device|fsname>*.

This command needs to be executed with the root privileges on the MGS node.

[OPTIONS]

-d *<device | fsname>.<parameter>*

This option deletes parameter setting (uses default value at the time of next restarting). Even if null is set to *<value>*, the option can delete parameter setting.

lctl lfsck_start

[NAME]

lctl lfsck_start - Starts lfsck on the specified MDT device.

[SYNOPSIS]

```
/usr/sbin/lctl lfsck_start -M <fsname>-MDT0000 -A  
/usr/sbin/lctl lfsck_start -h
```

[DESCRIPTION]

This command checks for MDT inconsistencies and repair the file system.

Execute this command only on the MDS node that mounts MDT0.

Only the users who have root privileges can use this command.

For <fsname>, specify the FEFS file system name.

[OPTIONS]

-M <fsname>-MDT0000

Specifies the MDT device that is the target of the command.

-A

Starts this command on all available devices in the system.

-h

This option displays help information.

lctl lfsck_stop

[NAME]

lctl lfsck_stop - This command stops lfsck that is running.

[SYNOPSIS]

```
/usr/sbin/lctl lfsck_stop -M <fsname>-MDT0000 -A [option]
```

[DESCRIPTION]

This command stops lfsck that is running.

Execute this command on the MDS node that mounts MDT0.

Only the users who have root privileges can use this command.

For <fsname>, specify the FEFS file system name.

[OPTIONS]

-M <fsname>-MDT0000

Specifies the MDT device that is the target of the command.

-A

Stops this command on all available devices in the system.

-h

This option displays help information.

lctl help

[NAME]

lctl help - Display the lctl help.

[SYNOPSIS]

```
/usr/sbin/lctl help [command]
```

[DESCRIPTION]

This command displays the lctl help.

[OPTIONS]

[command]

This option specifies the command to display about helpful information.

lctl --list-commands

[NAME]

lctl --list-commands - Displays a list of subcommands

[SYNOPSIS]

```
/usr/bin/lctl --list-commands
```

[DESCRIPTION]

This command displays a list of subcommands.

A.2.8 fsck.ldiskfs

[NAME]

fsck.ldiskfs - Check and repair an FEFS device

[SYNOPSIS]

```
/opt/FJSVfefsprots/sbin/fsck.ldiskfs [-pnyfv] [-b superblock] [-B blocksize]
[-j external_journal] device
```

[DESCRIPTION]

This command is executed on the MGS/MDS/OSS node, and checks and repairs MGT/MDT/OST.

Only the users who have root privileges can use this command.

[OPTIONS]

-p

This option automatically repairs the file system without any questions.

-n

This option makes no changes to the file system.

-y

This option assumes that "yes" is the answer for all questions at the repair time.

-f

This option forces a check even if the file system is clean.

-v

This option produces more verbose output.

-b *superblock*

This option uses superblock as a super block instead of the normal super block.

-B *blocksize*

This option forces a search for super blocks of the specified block size.

-j external_journal

This option specifies a path name with an external journal.

[ARGUMENTS]

device

This argument specifies the device to be checked.

[EXIT CODE]

The exit code returned by fsck.ldiskfs is the sum of the following conditions:

0: No errors

1: File system errors corrected

2: File system errors corrected, system should be rebooted

4: File system errors left uncorrected

8: Operational error

16: Usage or syntax error

32: Fsck.ldiskfs canceled by user request

128: Shared library error

A.2.9 tunefs.lustre

[NAME]

tunefs.lustre - modify the Lustre configuration information on a disk

[SYNOPSIS]

/usr/sbin/tunefs.lustre [option] *device*

[DESCRIPTION]

tunefs.lustre is used to modify the configuration information on a Lustre target disk.

This does not reformat the disk or erase the target information, but modifying the configuration information can result in an unusable filesystem.

Changes made here will affect a filesystem only when the target is next mounted.

Only the users who have root privileges can use this command.

[OPTIONS]

--writeconf

Erase the configuration logs for the filesystem that this server is part of, and regenerate them. This is very dangerous.

All targets (OSTs/MDTs) must then be restarted to regenerate the logs.

Order of operations is:

1. Unmount all clients of this filesystem.
2. Unmount MDT and all OSTs of this filesystem.
3. Run tunefs.lustre --writeconf <*device*> on every server.
4. Mount MDT and OSTs.
5. Mount clients

[ARGUMENTS]

device

Specify the target device.

A.2.10 debugfs.ldiskfs

[NAME]

debugfs.ldiskfs - ldiskfs file system debugger

[SYNOPSIS]

```
/opt/FJSVfefefsprogs/sbin/debugfs.ldiskfs [-R request] [device]
```

[DESCRIPTION]

The debugfs.ldiskfs program is an interactive file system debugger.

device is a device or a file.

Only users with root privileges can execute this command on MGS/MDS/OSS.

[OPTIONS]

-R request

Causes debugfs.ldiskfs to execute the single command request, and then exit.

icheck

[SYNOPSIS]

```
icheck block ...
```

[FUNCTION]

Print a listing of the inodes which use the one or more blocks specified on the command line.

testb

[SYNOPSIS]

```
testb block [count]
```

[FUNCTION]

Test if the block number block is marked as allocated in the block bitmap. If the optional argument count is present, then count blocks starting at block number block will be tested.

quit

[FUNCTION]

Quit debugfs.ldiskfs

A.2.11 fefsbackup

[NAME]

feefsbackup - Data management tool.

[SYNOPSIS]

```
feefsbackup [--version] [-h] [subcommand]
```

[DESCRIPTION]

The main commands of the data management tool have the following subcommands as functions.

Only the users who have administrator privileges can use this command.

[OPTIONS]

--version

This option displays the version of the data management tool.

-h, --help

This option displays help information.

[SUBCOMMAND]

fefsb backup copy

[SYNOPSIS]

```
/opt/FJSVfeffs/bin/fefsb backup copy [-h]
/opt/FJSVfeffs/bin/fefsb backup copy [-L request_id] [-v] [option] -d destdir [-n node] {path
[path...] | -f pathlist}
/opt/FJSVfeffs/bin/fefsb backup copy [-v] -u request_id [option] -d destdir [-n node] {path
[path...] | -f pathlist}
/opt/FJSVfeffs/bin/fefsb backup copy [-v] -R request_id
```

[DESCRIPTION]

The file attributes (access rights, ownership, and time of last modification) are copied from the attributes of the source file.
The following is not copied.

- ACL and user extended attributes (Specify an option if you want to copy.)
- Hard links (there are cases when hard links not copy).
- Project ID
- Project ID inheritance flag
- Information about the striped directory
- Information about the stripes

If the file already exists on the copy destination directory, overwrite it.

If the path includes the ".." directory, copy the files under the path name specified after "..".

If it fails to copy the file, it prints the failed file path name and error information to standard error and exits with an error.

[OPTIONS]

-h, --help

This option displays help messages.

-f, --file pathlist

Handles files and directory paths listed in the specified *pathlist* as copy targets. Describe one file path per line in *pathlist*.
Moreover, this option cannot be used when an individual path has already been specified since an error will occur.

-L, --label request_id

Specify the request ID for copy management. Specify a unique ID for the request ID.



The request ID conditions that can be specified using the -L option are as follows.

Available characters: Half-width alphanumeric characters (a-z, A-Z, 0-9), half-width symbols (-_)

Number of characters: 1 to 32 characters

Format: Do not start with half-width symbols (-_)

-R, --resume request_id

If the copy process interrupted due to an error, etc. can be executed again, restart the process from the point of error.

-u, --update request_id

Specify the copied request ID that is the target of the difference copy. If the time of last modification, the ttime of last status change, or size of the file managed by the specified request ID is different from the copy source file system, transfer the file to the copy destination file system.

-v, --verbose

Display the transferred files and directories.

-a, --acl

Save the ACL. If not specified, it will not be saved.

-x, --xattr

Save extended attributes. If not specified, it will not be saved.

-n, --node *node*

Specify the copy destination node. If not specified, transfer within the node. Specify the destination host name or IP address.

-d, --destdir *destdir*

Specify the copy destination directory. This option is required. If not specified, an error will occur.

The destination directory must be an absolute path.

--ignore_err

If a file fails to copy, continue copying except for the failed file. The default is to fail if the file fails to copy.

[EXIT CODE]

0: Normal end

1: Copy failed with an internal error

2: Incorrect usage

fefsbackup list

[SYNOPSIS]

```
/opt/FJSVfeffs/bin/fefsbackup list [-h]
/opt/FJSVfeffs/bin/fefsbackup list [-l] request_id [path ...]
/opt/FJSVfeffs/bin/fefsbackup list [-l] [-f pathlist] request_id
```

[DESCRIPTION]

This command outputs copy request ID information. If no request ID is specified, a list of request IDs is output. When request ID of copy is specified, a list of files managed by the request ID is output.

If path or pathlist of -f option is specified in the argument, only the specified files are output.

If a directory is specified, a list of files under that directory is output.

[OPTIONS]

-h, --help

This option displays help information.

-l, --list

When specified, this option outputs detailed information on "ls -l" equivalent.

-f, --file *pathlist*

This option outputs only the information containing the file and directory path described in the specified *pathlist*. However, when this option is specified, error occurs in case that the path is specified individually.

[EXIT CODE]

0: Normal end

1: Command execution failed with an internal error

2: Incorrect usage

fefsbackup delete

[SYNOPSIS]

```
/opt/FJSVfeffs/bin/fefsbackup delete -h
/opt/FJSVfeffs/bin/fefsbackup delete [-F] request_id
```

[DESCRIPTION]

This command deletes the request ID of the specified copy. Only deletes the management information of the request ID. It does not delete copied files.

[OPTIONS]

-h, --help

This option displays help information.

-F, --force

This option assumes that "yes" is entered to all inquiry messages to user.

[EXIT CODE]

0: Normal end

1: Command execution failed with an internal error

2: Incorrect usage

fefsbackup status

[SYNOPSIS]

```
/opt/FJSVfefsf/bin/fefsbackup status -h  
/opt/FJSVfefsf/bin/fefsbackup status  
/opt/FJSVfefsf/bin/fefsbackup status request_id
```

[DESCRIPTION]

This command outputs a list of request IDs that are being executed and that ended with an error. If a request ID is specified, detailed information of the request ID is output.

[OPTIONS]

-h, --help

This option displays help information.

[EXIT CODE]

0: Normal end

1: Command execution failed with an internal error

2: Incorrect usage

A.2.12 fefs_ost2fid

[NAME]

fefs_ost2fid - show a FID corresponding to its inode number.

[SYNOPSIS]

```
/opt/FJSVfefsfprogs/sbin/fefs_ost2fid <OST device> <ino ...>
```

[DESCRIPTION]

fefs_ost2fid is used to show FID using inode number <ino>. <device> is an OST device path. fefs_ost2fid is available to only Administrators on OSS node.

Sample Output

```
# /opt/FJSVfefsfprogs/sbin/fefs_ost2fid /dev/sdb 13024 14254  
13034: [0xf:0x2367de45:0x4000]  
14254: [0x22:0x2367de46:0x4000]
```

[ARGUMENTS]

<OST device>

an OST device path

<ino>

inode number. Separate multiple entries with " " (Space). You can specify up to 1024 inode numbers.

[Command ends normally]

Output FID to standard output and return 0.

[Error message at command abend]

If there is an inode number for which an FID cannot be obtained, "No such file or directory" is output in the FID column for that inode number, and the command ends normally.

If an error is detected during the search, such as an invalid argument, a message is output to the standard error and the search continues.

The following is an example of failure in reading the FID with inode number 13648.

Sample Output

```
# /opt/FJSVfefefsprogs/sbin/fefs_ost2fid /dev/sdb 13648
Invalid EA entry in inode (13648)      # Output destination is standard error
```

If the entire command fails, it prints an error message to standard error and returns a return value of 1.

[RETURN VALUES]

0: Normal end

1: Abnormal end

A.2.13 find_file_ost

[NAME]

find_file_ost - show files located on OST.

[SYNOPSIS]

```
/opt/FJSVfefefsprogs/sbin/find_file_ost -o <filename> -d <device> <OST_index ...>
```

[DESCRIPTION]

find_file_ost is used to show file paths located on OST of *<ost_index>*. *<ost_index>* is an OST index corresponding to its OST device. *<device>* is a MDT device path.

find_file_ost is available to only Administrators on MDS node.

[OPTIONS]

-o *<filename>*

print results to *<filename>.out* if succeeded and *<filename>.err* if failed.

-d *<device>*

inspect the specified MDT device.

<OST_index>

To identify the target OST, use a number starting with 0 called "OST index". Refer to the OSS section of the FEFS design sheet "GFS sheet" and specify the value of "OST INDEX" corresponding to the failed device (OST VOLUME). Separate multiple numbers with " " (space). Up to 1024 OST indexes can be specified.

[Error message at command abend]

If an error is detected during the search, such as an invalid argument, a message is output to the standard error and the search continues.

If an error occurs in the middle of a command, it prints an error message to standard error and returns a return value of 1.

Sample Output

```
/opt/FJSVfefefsprogs/sbin/find_file_ost: Permission denied
```

[Handling command abends]

Since the parent directory list of the file that failed to be searched is output to "<outfile>.err", execute the convert_fid2path command on this file, and the files under the output path are targeted for recovery.

[RETURN VALUES]

- 0: Normal end
- 1: Abnormal end

A.2.14 convert_fid2path

[NAME]

convert_fid2path - lists file paths which have objects on specific OSTs.

[SYNOPSIS]

```
/opt/FJSVfefs/sbin/convert_fid2path -o <outfile> -m <mount_point> <infile> ...>
```

[DESCRIPTION]

convert_fid2path is used to detect file paths which have objects on specific OSTs from <inputfile> generated by find_file_ost command.

convert_fid2path is available to only Administrators on CLIENT node.

Sample Output

```
/feffs/dir1/a  
/feffs/dir2/b
```

[OPTIONS]

-o <outfile>

print results to <outputfile>.

-m <mount_point>

specify the mount point of the FEFS file system.

[Error message at command abend]

If an error is detected during the search, such as an invalid argument, a message is output to the standard error and the search continues.

If an error occurs in the middle of a command, it prints an error message to standard error and returns a return value of 1.

Sample Output

```
/opt/FJSVfefs/sbin/convert_fid2path: Invalid argument
```

[RETURN VALUES]

The following status is returned.

- 0: Normal end
- 1: Abnormal end

A.2.15 force_intr

[NAME]

force_intr - control the target device or show the current status.

[SYNOPSIS]

```
/usr/sbin/force_intr [-v] -c -m {activate|deactivate|status} [-a] <target>...  
/usr/sbin/force_intr [-v] -s -m {activate|deactivate} <target>...  
/usr/sbin/force_intr [-v] -s -m status [-a] <target>...  
/usr/sbin/force_intr -h
```

[DESCRIPTION]

force_intr controls the specified target device. The target device is an FEFS device identifier in the following format:

<fsname>-[MDT | OST] <Hexadecimal 4-digit index>

If the target device is deactivated, the FEFS access being processed is terminated abnormally. In addition, the new FEFS access returns abnormally.

If the -c option is specified, disable/enable file access from the FEFS client to the specified target.

When the -s option is specified, file access to the target is disabled/enabled between FEFS servers. If disabled, file access between servers being transferred is interrupted and an error is returned to the FEFS client.

This command will fail if the FEFS service is not running.

force_intr is available to only Administrators.



After you activate with this command, you must run the lfs df command on all client nodes to force a reconnection between the client and server.

[OPTIONS]

-h

show help.

-v

show verbose messages.

-c

specify on the Client-Side.

-s

specify on the Server-Side. When a target device is deactivated, the target device is not used for new file creation.

-m {activate|deactivate|status}

activate command causes the target device to be validated.

deactivate command causes the target device to be invalidated.

status command shows the current status. "UP" is displayed when activated, and "IN" is displayed when deactivated.

Example

The following shows a display example when the status is checked on each node after executing a file system disconnection operation.

- Display example (Client node)

```
[Storage I/O node]
# /usr/sbin/force_intr -c -m status fefs-OST0000 fefs-OST0001
IN fefs-OST0000
IN fefs-OST0001
[Login node]
# /usr/sbin/force_intr -c -m status fefs-OST0000 fefs-OST0001
IN fefs-OST0000
IN fefs-OST0001
```

- Display example (Server node)

```
[MDS node]
# /usr/sbin/force_intr -s -m status fefs-OST0000 fefs-OST0001
IN fefs-OST0000-osc-MDT0001
IN fefs-OST0000-osc-MDT0002
```

```
IN fefs-OST0001-osc-MDT0001
IN fefs-OST0001-osc-MDT0002
```

-a

invalidate, validate or list status for all targets on the node.

[ARGUMENTS]

<target>

Target device name to be processed

[RETURN VALUES]

The following status is returned.

0: Normal end

1: Abnormal end

A.2.16 evict_client

[NAME]

evict_client - disconnects the specified client from the FEFS filesystem.

[SYNOPSIS]

```
/usr/sbin/evict_client [-h] <ipaddress ...>
```

[DESCRIPTION]

If one FEFS client panics while accessing a file, the file access may be waited for several minutes if the same file is accessed from another FEFS client. By specifying the FEFS client that panicked with this command, you can isolate it from the file system and quickly resolve the file access hang.

If you specify a working FEFS client, file access will return to an error. The client disconnected with evict_client can be failed back by accessing the file again.

evict_client is available to only Administrators on MDS or OSS node.

[OPTIONS]

-h

show help.

[ARGUMENTS]

ipaddress

Specifies the IP address assigned to the client. Separate multiple entries with a space.

Refer to the FEFS design sheet for the specified IP address. If NODETYPE is CCM or LN, specify the Primary Network IP ADDRESS value. For CN-SIO, specify the value of Tofu's IP ADDRESS.

[RETURN VALUES]

The following status is returned.

0: Normal end

1: Abnormal end

A.2.17 fefs_yaml2csv

[NAME]

fefs_yaml2csv - This command converts node information definition files for installer.

[SYNOPSIS]

```
/sbin/fefs_yaml2csv <infile> <outfile>
```

[DESCRIPTION]

This command outputs the specified <infile> to <outfile> in CSV format. You can import the output CSV file to the FEFSDesignSheet.xlsx.

For <infile>, specify the node information definition files or the node information definition files for FX server.

For details on the node information definition files and the node information definition files for FX server, see the *Job Operation Software Setup Guide*.

Only the users who have administrator privileges can use this command.

[RETURN VALUES]

This command returns the following status.

0: Normal end

1: Abnormal end

A.2.18 fefs_deactivate

[NAME]

fefs_deactivate - control the filesystem access or show the current filesystem status.

[SYNOPSIS]

```
/usr/sbin/fefs_deactivate -m {activate|deactivate|status} <path>
/usr/sbin/fefs_deactivate -m status -a
```

[DESCRIPTION]

fefs_deactivate is used to control requests for a filesystem.

<path> is a target filesystem mount point to activate or deactivate or show status.

If a target filesystem is deactivated, requests in progress for the filesystem will be interrupted.

fefs_deactivate is available to only Administrators.

[OPTIONS]

-m {activate|deactivate|status}

activate command causes the target filesystem to be validated.

deactivate command causes the target filesystem to be invalidated.

status command shows the current filesystem status.

Example

The following shows a display example when the status is checked on each node after executing a file system disconnection operation.

- Display example

```
[Storage I/O node]
# fefs_deactivate -m status /fefs
FS    ST MNT
FEFS IN /.fefs
LLIO IN /fefs
[Compute node]
# fefs_deactivate -m status /fefs
FS    ST MNT
LLIO IN /fefs
[Login node]
# fefs_deactivate -m status /fefs
FS    ST MNT
FEFS IN /fefs
```

Description of the header

FS: Display file system.

ST: Display the disconnecting/incorporating status."IN" is displayed when disconnected, and "UP" is displayed when not disconnected.

MNT: Display Intended mount point

-a

list status for all target filesystem

[RETURN VALUES]

The end status is returned as follows:

0: Normal end

1: Abnormal end

Appendix B Messages

B.1 Messages Output to System Log

The following shows the format of the messages output to system log by FEFS:

```
[ERR.] FEFS 0001 an error..  
(1) (2) (3) (4)
```

1. Message type

This shows an output level of messages. According to the message type, the message ID shown below is assigned.

- [ERR.]: ERROR message (0001 to 5999)
- [WARN]: WARNING message (6000 to 6999)
- [NOTE]: NOTICE message (7000 to 7999)

2. FEFS prefix

This is an identifier indicating that the output message is FEFS-related.

The contents are as follows:

- FEFS: Message output from FEFS main unit
- LNet: Message output from the FEFS communication layer

3. Message ID

This is an identification ID of messages. As is described in 1 above, the range of values is determined depending on the message type.

4. Message contents

These are the contents of the message.

[ERROR MESSAGES] (0001 to 5999)

[ERR.] FEFS 0001 *Obdname*: operation *Op* to node *Netdev* failed: rc = *Err*

Meaning

An error occurred during an operation between the FEFS client and server.

Obdname: OBD name

Op: Operation name

Netdev: Network device name

Err: Error code

Corrective action

Value of <i>Err</i>	Meaning	Corrective action
-2	The target file does not exist.	Confirm whether the file is present.
-11	A retry request is detected.	No action is necessary.
-12	A memory shortage has occurred.	Check the memory usage status.
-16	<i>Netdev</i> has failed to reconnect to the MDT/ OST.	No action is necessary.
-19	The target for operation does not exist.	No action is necessary.
-107	The connection information of the node on which this message is output, not in <i>Netdev</i> .	The node on which this message is output may have been evicted from the file system. Confirm whether the job

Value of <i>Err</i>	Meaning	Corrective action
		executed on the node on which this message is output is in IO error. The connection is restored automatically when the eviction from the file system occurs.
-110	A timeout has occurred for a disconnection notification from the MGS to a client.	No action is necessary.
-116	If <i>Op</i> is "mds_close", there is no information about the file closed on the MDS.	This message, if is output by itself, no action is necessary. If it is output with other messages, refer to how to deal with those messages.

Otherwise, contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[ERR.] FEFS 0003 Unknown option '*Opt*', won't mount

Meaning

The mount option is not correct.

Opt: Mount option

Corrective action

Specify the mount option correctly.

[ERR.] FEFS 0004 Illegal option value '*Opt=Val*', won't mount.

Meaning

The mount option is not correct.

Opt: Mount option

Val: Value of mount option

Corrective action

Specify the mount option correctly.

[ERR.] FEFS 0005 Simultaneous specification with noqos and qos,qos_cache options can't be performed, won't mount.

Meaning

The mount option is not correct. You cannot specify the noqos option at the same time as the qos, qos_cache option.

Corrective action

Specify either noqos option or qos, qos_cache option.

[ERR.] LNet 1000 Please specify EITHER 'networks' or 'ip2nets' but not both at once

Meaning

The communication layer definition file contains an error.

Corrective action

Confirm that the FEFS design sheet is correct.

[ERR.] LNet 1001 Error *Err* starting up LNI Net

Meaning

Communication layer initialization failed.

Err: Error code

Net: Network interface

Corrective action

Confirm that the network interface displayed in Net is enabled.

[ERR.] LNet 1002 Error parsing '*Def*'

Meaning

The communication layer definition file contains an error.

Def: Invalid definition

Corrective action

Confirm that the FEFS design sheet is correct.

[ERR.] LNet 1003 Duplicate network specified: *Type*

Meaning

The communication layer definition file contains an error.

Type: Duplicate network type

Corrective action

Confirm that the FEFS design sheet is correct.

[ERR.] LNet 1004 Can't parse networks: string too long

Meaning

The communication layer definition file contains an error.

Corrective action

Confirm that the FEFS design sheet is correct.

[ERR.] LNet 1005 Too many interfaces for net *Name*

Meaning

The Inet definition file contains an error. The specified number of network interfaces in the networks option exceeds the limit on the number that can be specified (16).

Name: Network name

Corrective action

Confirm that the FEFS design sheet is correct.

[ERR.] LNet 1006 Error *Err* enumerating local IP interfaces for ip2nets to match

Meaning

Communication layer initialization failed.

Err: Error code

Corrective action

If the error code is -12, the error is caused by a memory shortage. Check the memory usage status.

If the error code is other than -12, confirm that the FEFS design sheet is correct.

[ERR.] LNet 1007 No local IP interfaces for ip2nets to match

Meaning

The communication layer definition file contains an error.

Corrective action

Confirm that the FEFS design sheet is correct.

[ERR.] LNet 1008 Error *Err* parsing ip2nets

Meaning

The communication layer definition file contains an error.

Err: Error code

Corrective action

Confirm that the FEFS design sheet is correct.

[ERR.] LNet 1009 ip2nets does not match any local IP interfaces

Meaning

The communication layer definition file contains an error.

Corrective action

Confirm that the FEFS design sheet is correct.

[ERR.] FEFS 2000 libcfs_debug_init: *Err*

Meaning

Failed to initialize libcfs debug function when loading libcfs module

Err: Error code

Corrective action

Loading libcfs module failed. When *Err* is -12, a memory shortage has occurred on the computer. Take actions, such as reducing memory usage, and then restart the FEFS service. If the issue still persists, collect the system log file that contains this message, and contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[ERR.] FEFS 2001 init_fefslog: *Err*

Meaning

Failed to initialize FEFSLOG function when loading libcfs module

Err: Error code

Corrective action

Loading libcfs module failed. When *Err* is -22, some libcfs module parameters may be set invalid value. If so, reexamine and reset the parameters and restart FEFS service. When *Err* is -12, a memory shortage has occurred on the computer. Take actions, such as reducing memory usage, and then restart the FEFS service. If the issue still persists, collect the system log file that contains this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[ERR.] FEFS 2003 misc_register: error *Err*

Meaning

Failed to register miscellaneous devices when libcfs loading

Err: Error code

Corrective action

Loading libcfs module failed in miscellaneous devices registration. Collect the system log file that contains this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[ERR.] FEFS 2004 insert_proc: error *Err*

Meaning

Failed to create a procfs entry

Err: Error code

Corrective action

Loading libcfs module failed in creating a procfs entry. Collect the system log file that contains this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[ERR.] FEFS 2005 misc_deregister error *Err*

Meaning

Failed to deregister miscellaneous devices when removing libcfs module

Err: Error code

Corrective action

Collect the system log file that contains this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[ERR.] FEFS 2006 libcfs_debug_cleanup: *Err*

Meaning

The process of terminating the debug function failed.

Err: Error code

Corrective action

Collect the system log file that contains this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[ERR.] FEFS 2007 initialize workitem: error *Err*

Meaning

Failed to initialize FEFS service internally

Err: Error code

Corrective action

Loading libcfs module failed. Collect the system log file that contains this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[ERR.] FEFS 2008 Startup workitem scheduler: error: *Err*

Meaning

Failed to initialize FEFS service internally

Err: Error code

Corrective action

Loading libcfs module failed. Collect the system log file that contains this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[ERR.] FEFS 2009 cfs_crypto_register: error *Err*

Meaning

Failed to initialize FEFS service internally

Err: Error code

Corrective action

Loading libcfs module failed. Collect the log message that contains this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[ERR.] FEFS 2010 Cannot create proc entry: *Err*

Meaning

Failed to create "/proc/fefslog" entry on procfs when loading libcfs module

Err: Error code

Corrective action

Loading libcfs module failed. When *Err* is -12, a memory shortage has occurred on the computer. Take actions such as reducing memory usage and restart the FEFS service.

[ERR.] FEFS 2022 open: /proc/fefslog: *Msg: Err*

Meaning

Failed to open(2) /proc/fefslog entry in FEFSLOG daemon

Msg : The message generated by strerror(3) according to errno.

Err: Error code

Corrective action

FEFSLOG daemon might try to start its service without loading libcfs module. Load libcfs module or restart the FEFS service. If the same issue still persists, collect the system log file that contains this message, and contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[ERR.] FEFS 2023 open *File*: *Msg: Err*

Meaning

Failed to open(2) the FEFSLOG file in FEFSLOG daemon

File : The path to an FEFSLOG file

Msg : The message generated by strerror(3) according to errno.

Err: Error code

Corrective action

FEFSLOG daemon failed to open(2) an FEFSLOG file. *Err*, that is errno, must be the same written on the online manual article on open(2) system call. Refer it and follow the given guidance to fix the issue (For example, the following causes are thought. The daemon has no access permission for the file when *Err* is -13. The daemon has no retrieval permission for a part of the path of File. The *File* does not exist and the daemon has not write permission for the parent directory.). If the issue still persists, collect the system log file that contains this message, and contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[ERR.] FEFS 2024 ioctl: *Cmd: Msg: Err*

Meaning

Failed to ioctl(2) in FEFSLOG daemon.

Cmd: The command given to ioctl(2)

Msg : The message generated by strerror(3) according to errno.

Err: Error code

Corrective action

FEFSLOG daemon failed to ioctl(2) with *Cmd* command. Restart the FEFS service. If the issue still persists, collect the system log file that contains this message, and contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[ERR.] FEFS 2025 Portals memory leaked: *Bytes* bytes

Meaning

Memory leak has happened in libcfs module.

Bytes : The amount of leaked memory (byte)

Corrective action

There is no effect on running the FEFS service but libcfs module has a memory leak issue. Collect the system log file that contains this message, and contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[ERR.] FEFS 2200 QoS cannot allocate memory.

Meaning

Acquisition of the memory required for the QoS function failed.

Corrective action

Check the memory usage status.

[ERR.] FEFS 2201 QoS System error. func=*Func* route=*Route* code=*Err*

Meaning

A system error occurred in the QoS function.

Func: Function name

Route: Error route

Err: Error code

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[ERR.] FEFS 2202 QoS cannot be enabled.

Meaning

The QoS function cannot be enabled.

Corrective action

If you want to use QoS function, set the Lustre NRS-policy to fifo.

Execute the following command.

Example of execution on the MDS.

```
# lctl set_param mds.MDS.mdt.nrs_policies="fifo reg"
```

Example of execution on the OSS

```
# lctl set_param ost.OSS.ost_io.nrs_policies="fifo reg"
```

[ERR.] FEFS 2203 NRS cannot be enabled.

Meaning

The NRS (Network Request Scheduler) cannot be enabled.

Corrective action

The NRS is not supported.

[ERR.] FEFS 2400 fefsinfod-daemon is terminated abnormally. (Error code: *Err*)

Meaning

The statistics acquisition daemon terminated abnormally.

Err: Error code

Corrective action

Collect the system log file that contains this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[ERR.] FEFS 2410 Couldn't mount because of no required features.

Meaning

FEFS (MDS) mount failed.

Corrective action

Confirm that the MDT volume in FEFS design sheet is correct.

[ERR.] FEFS 2420 Lustre server(*Svr*) is unsupported version.

Meaning

The connect Lustre server is unsupported version.

Svr: Server

Corrective action

Confirm that the connect Lustre version is correct.

[WARNING MESSAGES] (6000 to 6999)

[WARN] FEFS 6420 gethostname: Cannot get hostname: *Msg*: *Err*

Meaning

Failed to gethostname(2) in FEFSLOG daemon

Msg : The message generated by strerror(3) according to errno.

Err: Error code

Corrective action

FEFSLOG daemon failed to gethostname(2) in its start process. FEFSLOG daemon can continue running but all the hostnames contained in FEFSLOG messages come to "UNKNOWN" during the error. When you find it inconvenient, refer the online manual article on gethostname(2), follow the given guidelines and restart FEFS service to fix the issue. If the issue still persists, collect the system log file that contains this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[WARN] FEFS 6421 close *File*: *Msg*: *Err*

Meaning

Failed to close(2) an FEFSLOG file in FEFSLOG daemon

File : The path to an FEFSLOG file

Msg : The message generated by strerror(3) according to errno.

Err: Error code

Corrective action

FEFSLOG daemon failed to close(2) an FEFSLOG file. Although FEFSLOG daemon can continue running there may be some missing log messages. Err, that is errno, must be the same written on the online manual article on close(2) system call. Refer it and follow the

given guidance to fix the issue. If the issue still persists, collect the system log file that contains this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[WARN] FEFS 6422 read /proc/efslog: Msg: Err

Meaning

Failed to read(2) /proc/efslog entry in FEFSLOG daemon

Msg : The message generated by strerror(3) according to errno.

Err : Error code

Corrective action

FEFSLOG daemon failed to read(2) /proc/efslog entry. Although FEFSLOG daemon can continue running there may be some missing log messages. Err, that is errno, must be the same written on the online manual article on read(2) system call. Refer it and follow the given guidance to fix the issue. If the issue still persists, collect the system log file that contains this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[WARN] FEFS 6423 write File: Msg: Err

Meaning

Failed to write(2) an FEFSLOG file in FEFSLOG daemon

File : The path to an FEFSLOG file

Msg : The message generated by strerror(3) according to errno.

Err : Error code

Corrective action

FEFSLOG daemon failed to write(2) the FILE file. Although FEFSLOG daemon can continue running there may be some missing log messages. Err, that is errno, must be the same written on the online manual article on write(2) system call. Refer it and follow the given guidance to fix the issue.

[WARN] FEFS 6424 close /proc/efslog: Msg: Err

Meaning

Failed to rotate log files in FEFSLOG

Msg : The message generated by strerror(3) according to errno.

Err : Error code

Corrective action

FEFSLOG daemon failed to close(2) when rotating log files. Although FEFSLOG daemon can continue running there may be some missing log messages. Err, that is errno, must be the same written on the online manual article on close(2) system call. Refer it and follow the given guidance to fix the issue. If the issue still persists, collect the system log file that contains this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[WARN] FEFS 6425 lost Bytes [Bytes] messages

Meaning

Abandon log messages when libcfs module removed

Bytes : The number of bytes of abandoned log messages

Corrective action

There is no effect on FEFS service, so that no need to care about the issue. But when you see the messages with considerable frequency it may be an FEFSLOG bug. In the case, collect the system log file that contains this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[WARN] FEFS 6426 strftime fails.

Meaning

Failed to strftime(3) in FEFSLOG daemon

Corrective action

FEFSLOG daemon failed to strftime(3) when translating log messages thought there is no effect on the FEFS service. When you see the issue with considerable frequency and find it inconvenient, collect the system log file that contains this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[WARN] FEFS 6427 snprintf: *Msg*: *Err*

Meaning

Failed to sprintf(3) in FEFSLOG daemon

Msg: The message generated by strerror(3) according to errno.

Err: Error code

Corrective action

FEFSLOG daemon failed to sprintf(3) when translating log messages thought there is no effect on the FEFS service. When you see the issue with considerable frequency and find it inconvenient, collect the system log file that includes this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[WARN] FEFS 6428 localtime: Cannot get localtime

Meaning

Failed to localtime(3) in FEFSLOG daemon

Corrective action

FEFSLOG daemon failed to localtime(3) when translating log messages thought there is no effect on the FEFS service. When you see the issue with considerable frequency and find it inconvenient, collect the system log file that includes this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

[WARN] FEFS 6600 QoS unknown request from ip=*Addr*

Meaning

An FEFS client issued a request not defined in the QoS definition file.

Addr: IP address of the requester

Corrective action

Review the definitions in the QoS definition file, and specify the appropriate IP address.

[NOTICE MESSAGES] (7000 to 7999)

[NOTE] FEFS 7001 Mounted *Fsname*-client

Meaning

The FEFS client mount process is complete.

Fsname: File system name

Corrective action

No action is necessary.

[NOTE] FEFS 7010 server umount {MGS | *Fsname*-(MDT | OST)}*Num* complete

Meaning

The FEFS server (MGS, MDS or OSS) unmount process is complete.

Fsname: File system name

Num: MDT index or OST index

Corrective action

No action is necessary.

[NOTE] FEFS 7011 Unmounted *Fsname*-client

Meaning

The FEFS client unmount process is complete.

Fsname : File system name

Corrective action

No action is necessary.

[NOTE] FEFS 7400 fefslog daemon is starting rotate fsize: *Fsize*, gen: *Gen*

Meaning

FEFSLOG daemon is starting.

Fsize : Size of file in which rotation of file occurs

Gen : Number of generations of retained logs

Corrective action

This is a normal message displayed when the FEFSLOG daemon starts. When the FEFS service starts, this message is output always.
No action is necessary.

[NOTE] FEFS 7401 Shutting down fefslogd ...

Meaning

The fefslog daemon has stopped.

Corrective action

This is a normal message displayed when the FEFSLOG daemon stops. When the FEFS service stops, this message is output always.
No action is necessary.

[NOTE] FEFS 7501 {MGS | *Fsname*-{MDT | OST}*Num*}: Will be in recovery for at least *Time*: or until *Numcli* client[s] reconnect[s]

Meaning

FEFS service (MGT, MDT or OST) failover has started. Failover will take at least *Time*, or *Numcli* FEFS clients are still being reconnected.

Fsname: File system name

Num: MDT index or OST index

Time: Time required for failover (mm:ss)

mm: minutes

ss: seconds

Numcli: Number of FEFS clients

Corrective action

No action is necessary.

[NOTE] FEFS 7502 {MGS | *Fsname*-{MDT | OST}*Num*}: haven't heard from client *Id* (at *Nid*) in *Time* seconds. I think it's dead, and I am evicting it. *Info*

Meaning

The client with NID *Nid* connected to the MGS, MDS or OSS has not responded for *Time* seconds. The client will be disconnected from FEFS.

Fsname: File system name

Nnrr: MDT index or OST index

Id: Unique ID

Nid: Client NID

Time: Period during which the client has not responded (seconds)

Info: Other information concerning the client

Corrective action

No action is necessary.

LDISKFS-fs warning (device *device*): ldiskfs_mb_check_ondisk_bitmap:*NNNN*: on-disk bitmap for group *group* corrupted: *blocknum* blocks free in bitmap, *gd* - in *gd*

Meaning

A file system inconsistency has been detected.

Description of parameters

device: device name

NNNN: line number

group: block group number

blocknum: block group number

gd: group descriptor number

Corrective action

The file system may be corrupted. Repair the file system inconsistency. See "[D.4.5 Executing fsck](#)".

B.2 Messages Output by Commands

If a problem occurs during the execution of a command, one of the following messages is output to the standard error output.

B.2.1 fefs_sync

ERROR: fefs_sync: /etc/opt/FJSVfefs/config not found.

Meaning

/etc/opt/FJSVfefs/config directory does not exist.

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

ERROR: fefs_sync: internal command failed. (cmd=XX error=XX cluster=XX)

Meaning

The internal command ended abnormally.

cmd=XX: *XX* will be filled with command name.

error=XX: *XX* will be filled with error code.

cluster=XX: *XX* will be filled cluster name.

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

ERROR: fefs_sync: cannot be used by this user.**Meaning**

The user does not have privileges to execute the fefs_sync command.

Corrective action

Review user's privilege to execute the fefs_sync command and the nodes to execute, and then execute it again.

ERROR: fefs_sync: setup failed. check *Filename* file.**Meaning**

The configuration file creation failed.

Description of parameters

Filename : Error message file

Corrective action

To identify the cause of the failure, check the contents of the file *Filename*.

If you can identify the cause of the failure, after taking corrective action for the cause, execute the fefs_sync command again.

If you cannot identify the cause of the failure, contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

ERROR: fefs_sync: mkfs failed. check *Filename* file.**Meaning**

Formatting of the volume failed.

Description of parameters

Filename : Error message file

Corrective action

To identify the cause of the failure, check the contents of the file *Filename*.

If you can identify the cause of the failure, after taking corrective action for the cause, execute the fefs_sync command again.

If you cannot identify the cause of the failure, contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

ERROR: fefs_sync: mount failed. check *Filename* file.**Meaning**

Mounting FEFS failed.

Description of parameters

Filename : Error message file

Corrective action

To identify the cause of the failure, check the contents of the file *Filename*.

If you can identify the cause of the failure, after taking corrective action for the cause, execute the fefs_sync command again.

If you cannot identify the cause of the failure, contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

ERROR: fefs_sync: umount failed. check *Filename* file.**Meaning**

Unmounting FEFS failed.

Description of parameters

Filename : Error message file

Corrective action

To identify the cause of the failure, check the contents of the file *Filename*.

If you can identify the cause of the failure, after taking corrective action for the cause, execute the fefs_sync command again.

If you cannot identify the cause of the failure, contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

ERROR: fefs_sync: stop failed. check *Filename* file.

Meaning

Stopping FEFS service failed.

Description of parameters

Filename : Error message file

Corrective action

To identify the cause of the failure, check the contents of the file *Filename*.

If you can identify the cause of the failure, after taking corrective action for the cause, execute the fefs_sync command again.

If you cannot identify the cause of the failure, contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

WARN: fefs_sync: Some of the nodes could not accessed.

Meaning

Some nodes could not execute.

Corrective action

A message above is followed by following inquiry message. Input 'y' if you continue, input 'n' if you cancel.

Do you want to continue? (y/n):

If you input 'y', process is continued.

If you input 'n'. process is stopped. Identify the node that could not execute.

Node ID which can not be run, you can check in the file that is output in the following message.

create downnodeid file: filename

After addressed to the cause, recover system on the basis of "[3.11 Recovering a Node That Has Encountered an Error](#)".

INFO: fefs_sync: formatting. (remain=N)

Meaning

Formatting of the volume is in progress.

After elapse of 30 minutes of formatting of the volume, this message is displayed at regular intervals.

It is not an unusual behavior.

*N*means the number of the remaining volumes.

Corrective action

No action is necessary.

Connect FEFS. [*nodetype*] (*cluster*)

Meaning

The connection with server volume is checked when mounting FEFS in the client.

After this message, the following message is output.

..... OK/AWAIT: <connection confirmed volume number> / <connection waiting volume number>

This message is displayed at regular intervals, but there is no problem. When the system confirms connection at all volumes, it displays 0 at the connection confirmed volume number. If it is not completed after elapse of 10 minutes, the following inquiry message is output. Input 'y' if you continue, input 'n' if you cancel.

Continue to wait? [y/n]:

If you input 'y', the processing continues.

If you input 'n', "Aborted." message is output and the processing stops.

Corrective action

No action is necessary.

B.2.2 fefsconfig

fefsconfig: Error: configuration file was not found.

Meaning

A configuration definition file for the FEFS setup tool was not found.

Corrective action

Distribute the configuration definition file for the FEFS setup tool to the appropriate directory.

fefsconfig: Error: failed to setup by bad configuration file.

Meaning

A configuration definition file for the FEFS setup tool contains an error.

Corrective action

Review the contents of the configuration definition file for the FEFS setup tool.

fefsconfig: Error: failed to make <Filename>.

Meaning

Creation of the FEFS configuration file <Filename> failed.

Description of parameters

<Filename> : FEFS configuration file name

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

fefsconfig: Error: this node will not mount the fefs. (bad design sheet)

Meaning

The file system definition is not found in the FEFS configuration definition file.

Corrective action

Confirm that the FEFS design sheet is correct.

fefsconfig: Error: this node must have network information. (bad design sheet)

Meaning

The definition of the network outside the FEFS configuration definition file contains an error.

Corrective action

Confirm that the FEFS design sheet is correct.

efsconfig: Error: this node defined both design sheet. (bad design sheet)

Meaning

Execution node is defined in both FEFSDesignSheet.xlsx and FEFSDesignSheet_External.xlsx.

Corrective action

Define it in either of the FEFS design sheets.

efsconfig: Error: no such network interface. (bad network interface: <Interface name>)

Meaning

The network interface name of the FEFS configuration definition file contains an error.

Description of parameters

<Interface name> : Interface name

Corrective action

Confirm that the network interface definition for the FEFS design sheet is correct.

efsconfig: Error: hostname no such network interface. (not tofu ip address)

Meaning

Tofu IP address of the network interface in the FEFS configuration definition files is not specified.

Description of parameters

hostname : Host name

Corrective action

Make sure that there is no error in the definition of the network interface in FEFS design sheet.

efsconfig: Error: no such network interface. (bad ip address: <IPaddress>)

Meaning

The network interface IP address of the FEFS configuration definition file contains an error.

Description of parameters

<IPaddress> : IP address

Corrective action

Confirm that the network interface definition for the FEFS design sheet is correct.

efsconfig: Error: hostname no such network interface. (bad tofu coordinate: tofu coordinate)

Meaning

The tofu coordinate of the FEFS configuration definition file contains an error.

Description of parameters

hostname : Host name

tofu coordinate : Tofu coordinate

Corrective action

Confirm that the network interface definition for the FEFS design sheet is correct.

efsconfig: Error: no such network interface. (bad design sheet)**Meaning**

The network interface definition of the FEFS configuration definition file contains an error.

Corrective action

Confirm that the network interface definition for the FEFS design sheet is correct.

efsconfig: Error: no such device. (bad MGT volume: <Devicename>)**Meaning**

The MGT volume defined in the FEFS configuration definition file is not found in the node.

Description of parameters

<Devicename> : Device name

Corrective action

Specify the correct device of "MGT VOLUME" in the FEFS design sheet.

efsconfig: Error: no such device. (bad MDT volume: <Devicename>)**Meaning**

The MDT volume defined in the FEFS configuration definition file is not found in the node.

Description of parameters

<Devicename> : Device name

Corrective action

Specify the correct device of "MDT VOLUME" in the FEFS design sheet.

efsconfig: Error: no such device. (bad OST volume: <Devicename>)**Meaning**

The OST volume defined in the FEFS configuration definition file is not found in the node.

Description of parameters

<Devicename> : Device name

Corrective action

Specify the correct device of "OST VOLUME" in the FEFS design sheet.

efsconfig: Error: no such device. (bad JOURNAL volume: <Devicename>)**Meaning**

The JOURNAL volume defined in the FEFS configuration definition file is not found in the node.

Description of parameters

<Devicename> : Device name

Corrective action

Specify the correct device of "JOURNAL VOLUME" in the FEFS design sheet.

efsconfig: Error: no such device. (bad SSD volume: <Devicename>)

Meaning

The SSD volume defined in the FEFS configuration definition files does not exist in the node.

Description of parameters

<Devicename> : Device name

Corrective action

Specify the correct device of "SSD VOLUME" in the FEFS design sheet.

fefsconfig: Error: IP address was not found (InfiniBand).

Meaning

The IP address assigned to the InfiniBand device was not found.

Corrective action

Check whether an IP address is assigned to the InfiniBand device, and then execute fefsconfig again.

fefsconfig: Error: multiple IP addresses were found (InfiniBand).

Meaning

Multiple IP addresses assigned to the InfiniBand device were found.

Corrective action

Check whether there is only one InfiniBand device assigned an IP address, and then execute fefsconfig again.

fefsconfig: Error: failed to make configuration file (*filename*).

Meaning

The configuration definition file for the FEFS setup tool could not be created.

Description of parameters

filename : The configuration definition file for the FEFS setup tool

Corrective action

Review the access privileges for the output directory and files under it, and then execute fefsconfig again.

fefsconfig: Error: exist configuration file.

Meaning

The configuration definition file for the FEFS setup tool already exists.

Corrective action

Before newly building FEFS, make sure to delete the previous configuration definition file for the FEFS setup tool.

fefsconfig: Error: fefs-mdt00 volume was not found.

Meaning

The volume whose name is fefs-mdt00 was not be found.

Corrective action

Make the volume whose name is fefs-mdt00, and then execute fefsconfig again.

fefsconfig: Error: new MDT volume was not found.

Meaning

The new MDT volume was not found.

Corrective action

Correct the RAID volume name of a MDT volume that you want to add, and then execute fefsconfig again.

fefsconfig: Error: new OST volume was not found.

Meaning

The new OST volume was not found.

Corrective action

Correct the RAID volume name of an OST volume that you want to add, and then execute fefsconfig again.

fefsconfig: Error: the number of mdt is large. (number of mdt: <number of mdt>)

Meaning

The number of MDTs defined in the FEFS configuration definition file is larger than the number of implemented MDTs.

Description of parameters

<number of mdt> : number of MDTs

Corrective action

Check the value of "NUMBER OF MDT" in the LLIO sheet of the FEFS design sheet.

B.2.3 fefs_mkfs

FATAL:Unknown option 'Opt'

Meaning

The specified command line option is invalid.

Description of parameters

Opt : Invalid command line option

Corrective action

Review the command line option.

FATAL: Failed to read previous Lustre data from Dev (Err)

Meaning

FEFS data cannot be read from the device.

Description of parameters

Dev : Device name

Err : Error code

Corrective action

The disk or file system may be damaged. Check it with the fsck.fefs_mds command.

FATAL: Failed to write local files

Meaning

Writing of FEFS setting information to the MDT failed.

Corrective action

Confirm that the MDT has no problems.

ERROR: fefsconfig was not completely finished.**Meaning**

Initialization of the volume failed, because of failure of creation of an FEFS configuration file.

Corrective action

Review the FEFS design sheet, and then create an FEFS configuration file again.

ERROR: fefs_setup was not completely finished.**Meaning**

Initialization of the volume failed, because of failure of FEFS setup.

Corrective action

Collect the system log file, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

ERROR: not exist /etc/opt/FJSVfefs/fefs_tab.**Meaning**

The FEFS configuration file does not exist.

Corrective action

Create an FEFS configuration file again from the configuration definition files.

ERROR: not mkfs target.**Meaning**

A volume associated with the specified parameter does not exist.

Corrective action

Confirm that the specified parameter is correct.

ERROR: Device *Dev* was previously formatted. Use --reformat to reformat it.**Meaning**

An attempt was made to format a formatted volume again.

Description of parameters

Dev: Device name

Corrective action

Check whether the device can be reformatted again.

For details on performing format, see "[4.19.2 Procedure for canceling the protection of data in a file system](#)."

ERROR: modprobe ldiskfs error. ErrorCode=Err**Meaning**

ldiskfs load failed.

Description of parameters

Err: Error code

Corrective action

ldiskfs load failed. Collect the system log file that includes this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

B.2.4 fefs_mount

warning: can't allocate memory

Meaning

Memory cannot be reserved.

Corrective action

Check the free space in memory.

warning: failed to open pipe: *Pipe*

Meaning

The pipe open operation failed.

Description of parameters

Pipe : Pipe path

Corrective action

Check the free space in memory.

warning: failed write data to pipe: *Pipe*

Meaning

Writing of data to the pipe failed.

Description of parameters

Pipe : Pipe path

Corrective action

Check the free space in memory.

warning: failed to close pipe: *Pipe*

Meaning

The pipe close operation failed.

Description of parameters

Pipe : Pipe path

Corrective action

Check the free space in memory.

can't allocate memory for options

Meaning

Memory for the command line options cannot be reserved.

Corrective action

Check the free space in memory.

Command buffer overflow

Meaning

An overflow occurred in the command buffer.

Corrective action

Check the free space in memory.

ERROR: fefsconfig was not completely finished.

Meaning

Mounting failed, because of failure of creation of an FEFS configuration file.

Corrective action

Review the FEFS design sheet, and then create an FEFS configuration file again.

ERROR: fefs_setup was not completely finished.

Meaning

Mounting failed, because of failure of FEFS setup.

Corrective action

Collect the system log file, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

ERROR: fefs_mkfs was not completely finished.

Meaning

Mounting failed, because of failure of FEFS setup.

Corrective action

Collect the system log file, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

ERROR: not exist /etc/opt/FJSVfefsfefs_tab.

Meaning

The FEFS configuration file does not exist.

Corrective action

Create an FEFS configuration file again from the configuration definition files

ERROR: fefs_mount failed. sts=Status.

Meaning

Mounting FEFS failed.

Description of parameters

Status: Internal process code

Corrective action

Collect the system log file that includes this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

ERROR: already in progress.

Meaning

Mounting or unmounting is in progress.

Corrective action

After waiting a moment, execute the command again.

ERROR: mount now in progress.

Meaning

An internal error occurred during FEFS mounting.

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk

modprobe module error. ErrorCode=Err

Meaning

Loading module failed.

Description of parameters

module: Module name

Err: Error code

Corrective action

Collect the system log file that contains this message, and then contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

B.2.5 gefssnap

[ERR.] FEFS 2750 gefssnap -d *Outputdir*: No such directory.

Meaning

The specified directory not found.

Description of parameters

Outputdir: -d The directory path specified in the -d option.

Corrective action

Review the specified -d option.

[ERR.] FEFS 2752 gefssnap Exist temporary directory(*Tmpdir*).

Meaning

The temporary space already exists.

Description of parameters

Tmpdir: Directory path of temporary space.

Corrective action

Confirm the temporary space.

[ERR.] FEFS 2753 gefssnap Cannot create temporary directory(*Tmpdir*).

Meaning

Temporary space cannot be created.

Description of parameters

Tmpdir: Directory path of temporary space.

Corrective action

Make sure conditions are such that the temporary space can be created.

[ERR.] FEFS 2754 gefssnap Cannot create output file.

Meaning

Output file cannot be created.

Corrective action

Confirm the output directory status.

B.2.6 Ifs command

Common

error: Com: No such file or directory - Filedir

Meaning

The specified file or directory does not exist.

Description of parameters

Com : Command name

Filedir : Specified file or directory

Corrective action

Review the file name or directory name.

error: Com: invalid path - Path

Meaning

The specified path is invalid.

Description of parameters

Com : Command name

Path : Specified path name

Corrective action

Review the path name.

error: Com: stat failed - Path: Err

Meaning

The stat operation failed.

Description of parameters

Com : Command name

Path : Specified path name

Err : Error code

Corrective action

Review the path name.

Com: invalid option -- 'Opt'

Meaning

The specified command line option is invalid.

Description of parameters

Com : Command name

Opt : Invalid command line option

Corrective action

Review the command line option.

Com: unrecognized option '*Opt*'

Meaning

The specified command line option is invalid.

Description of parameters

Com : Command name

Opt : Invalid command line option

Corrective action

Review the command line option.

Com: option requires an argument -- '*Opt*'

Meaning

For the command line option, no argument is specified.

Description of parameters

Com : Command name

Opt : Command line option requiring an argument

Corrective action

Review the command line option.

Com: option '--*Opt*' requires an argument

Meaning

For the command line option, no argument is specified.

Description of parameters

Com : Command name

Opt : Command line option requiring an argument

Corrective action

Review the command line option.

Com: option '--*Opt*' is ambiguous

Meaning

Command line option is ambiguous.

Description of parameters

Com : Command name

Opt : Ambiguous command line option

Corrective action

Review the command line option.

Ifs project

***dir* - project identifier is not set (inode=*ID1*, tree=*ID2*)**

Meaning

It is not the specified project ID.

Description of parameters

dir : Specified directory

ID1 : Project ID of the directory to be checked

ID2 : Specified project ID

Corrective action

Specify the correct project ID.

error: project inheritance flag is not set

Meaning

The inheritance flag is not set.

Corrective action

Set the inheritance flag if necessary.

QUOTA related common

error: use either -u, -g or -p

Meaning

You cannot specify the -u|g|p option at the same time.

Corrective action

Review the command line option.

error: missing quota argument(s)

Meaning

The specification of argument is incorrect.

Corrective action

Review the arguments.

error: missing quota info argument(s)

Meaning

The specification of argument is incorrect.

Corrective action

Review the arguments.

error: Com: too long path - Path

Meaning

The path name is too long.

Description of parameters

Com : Command name

Path : Path name

Corrective action

Review the path name.

error: Com: Not a directory - Path

Meaning

The specified path is not a directory.

Description of parameters

Com : Command name

Path : Path

Corrective action

Specify a correct path.

error: Com: Not on FEFS - Path

Meaning

The specified path is not on FEFS.

Description of parameters

Com : Command name

Path : Path

Corrective action

Specify a correct path.

error: can't find id for name Name

Meaning

No id corresponding to the specified name was found.

Description of parameters

Name : Specified name

Corrective action

Specify the correct name.

Permission denied.

Meaning

The user does not have sufficient privileges.

Corrective action

Perform the operation with the appropriate privileges.

Unexpected quotactl error: Err

Meaning

An unexpected error occurred.

Description of parameters

Err : Error code

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

error: -u, -g and -p can't be used more than once

Meaning

You cannot specify the -u|g|-p option at the same time.

Corrective action

Review the command line option.

error: bad block-grace: Value

Meaning

The specified block grace value is incorrect.

Description of parameters

Value : Grace value

Corrective action

Specify the correct grace value.

error: bad inode-grace: Value

Meaning

The specified inode grace value is incorrect.

Description of parameters

Value : Grace value

Corrective action

Specify the correct grace value.

error: neither -u, -g, nor -p was specified

Meaning

Neither -u, -g nor -p was specified.

Corrective action

Review the command line option.

error: unexpected parameters encountered

Meaning

An unexpected parameter is encountered.

Corrective action

Review the command line option.

setquota failed: Err

Meaning

The setquota operation failed.

Description of parameters

Err : Error code

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

error: bad limit value *Value***Meaning**

The specified limit value is incorrect.

Description of parameters

Value : Limit value

Corrective action

Specify the correct limit value.

warning: block softlimit is smaller than the minimal qunit size.**Meaning**

The specified block softlimit value is smaller than the minimal unit size (1025).

Corrective action

Specifying a value of the minimal unit size (1025) or more is recommended.

warning: block hardlimit is smaller than the minimal qunit size.**Meaning**

The specified block hardlimit value is smaller than the minimal unit size (1025).

Corrective action

Specifying a value of the minimal unit size (1025) or more is recommended.

warning: inode softlimit is smaller than the minimal qunit size.**Meaning**

The specified inode softlimit value is smaller than the minimal unit size (1025).

Corrective action

Specifying a value of the minimal unit size (1025) or more is recommended.

warning: inode hardlimit is smaller than the minimal qunit size.**Meaning**

The specified inode hardlimit value is smaller than the minimal unit size (1025).

Corrective action

Specifying a value of the minimal unit size (1025) or more is recommended.

error: at least one limit must be specified**Meaning**

No limit value is specified.

Corrective action

Specify at least one limit value.

error: setquota failed while retrieving current quota settings (*Err*)**Meaning**

An error occurred during the search for the current QUOTA setting.

Description of parameters

Err: Error code

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

lfs df

error: invalid path '*Path*': *Err*

Meaning

The specification of a mount point path name is invalid.

Description of parameters

Path : Path name

Err: Error code

Corrective action

Review the specification of the path name.

lfs find

err: find: filename|dirname must either precede options or follow options

Meaning

A file name and directory name can have either a preceding option or following option.

Corrective action

Review the command line option.

error: find: no filename|pathname

Meaning

No file name or path name was found.

Corrective action

Specify a file name or path name.

error: find failed for *Filedir*.

Meaning

The specification of a directory name or file name is invalid.

Description of parameters

Filedir: Directory name or file name

Corrective action

Review the specification of the directory name or file name.

error: can't get lov name.: *Msg (Err)*

Meaning

The specified path is not on FEFS.

Description of parameters

Msg : Error message

Err : Error code

Corrective action

Specify a path on FEFS.

lfs setstripe

error: setstripe: missing filename|dirname

Meaning

The specification of a file name or directory name is missing.

Corrective action

Specify a file name or directory name.

error: setstripe: bad size 'Size'

Meaning

The stripe size specification is invalid.

Description of parameters

Size : Invalid size value

Corrective action

Review the stripe size.

error: setstripe: bad stripe offset 'Offset'

Meaning

The specification of the OST where file writing starts is invalid.

Description of parameters

Offset : Invalid offset value

Corrective action

Review the specification of the OST where file writing starts.

error: setstripe: bad stripe count 'Num'

Meaning

The specification of the stripe count is invalid.

Description of parameters

Num : Invalid stripe count

Corrective action

Review the specification of the stripe count.

warning: stripe size 4G or larger is not currently supported and would wrap: Invalid argument (22)

Meaning

The stripe size that exceeds 4194240KiB(4GiB-64KiB) is invalid.

Corrective action

Specify a stripe size of 4194240KiB(4GiB-64KiB) or less.

error: bad stripe_size Size, must be an even multiple of 65536 bytes: Invalid argument (22)

Meaning

The specification of the stripe size is invalid. It must be an even multiple of 65536 bytes.

Description of parameters

Size : Invalid size value

Corrective action

Specify stripe size by an even multiple of 65536 bytes.

error: bad stripe_offset Offset: Invalid argument (22)

Meaning

The specification of the OST where file writing starts is invalid.

Description of parameters

Offset : Invalid offset value

Corrective action

Review the specification of the OST where file writing starts.

error: bad stripe_count Num: Invalid argument (22)

Meaning

The specification of the stripe count is invalid.

Description of parameters

Num : Invalid stripe count

Corrective action

Review the specification of the stripe count.

'Path' is not on a Lustre filesystem: Msg (Err)

Meaning

The specified mount point is not on FEFS.

Description of parameters

Path : Path name

Msg : Error message

Err : Error code

Corrective action

Specify a path on FEFS.

Pool 'Poolname' is not on filesystem 'Fservername'

Meaning

The specified OST_pool was not found.

The specified OST_pool does not exist on file system *Fservername*.

Description of parameters

Poolname : OST_pool name
Fsname : File system name

Corrective action

Review the OST_pool name.

pool '*Poolname*' does not exist

Meaning

The specified OST_pool was not found.

Description of parameters

Poolname : OST_pool name

Corrective action

Review the OST_pool name.

pool '*Poolname*' has no OSTs

Meaning

No OST is registered in the specified OST_pool.

Description of parameters

Poolname : OST_pool name

Corrective action

Register an OST to the pool, and then execute again

unable to open '*Path*': *Msg* (*Err*)

Meaning

Unable to open the specified path.

Description of parameters

Path : Path name
Msg : Error message
Err : Error code

Corrective action

Review the path name.

error on ioctl 0x4008669a for '*Path*' (*fd*): *Msg*

Meaning

Failed to execute the command.

Description of parameters

Path : Path name
fd : File descriptor
Msg : Error message

Corrective action

If the error message is "stripe already set," a file already exists on the specified path. Review the path name, and perform the operation again.

If the error message is "Invalid argument", review the setting parameters and try again.

If the problem still persists, collect the system log file that contains this message, and contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

error: setstripe: create stripe file '*Path*' failed

Meaning

Failed to set the stripe.

Description of parameters

Path : Path name

Corrective action

Refer to the corrective action of the message output at the same time.

lfs getstripe

error: getstripe: failed for *Filedir*.

Meaning

The specification of a directory name or file name is invalid.

Description of parameters

Filedir : Directory name or file name

Corrective action

Review the specification of the directory name or file name.

lfs getdirstripe

error: setup_oob_uuid: unknown obduuid: *uuid*

Meaning

The OBD name specified by *uuid* is invalid.

Corrective action

Review the OBD name specification.

error opening *dir*: No such file or directory (2)

Meaning

The directory specified by *dir* does not exist.

Corrective action

Review the directory name specification.

lfs pool_list

'*Path*' is not on a Lustre filesystem: *Msg (Err)*

Meaning

The specified mount point is invalid.

Description of parameters

Path : Path name

Msg : Error message

Err : Error code

Corrective action

Review the mount point of FEFS.

invalid path '*Path*': *Msg* (*Err*)

Meaning

The specified path is invalid.

Description of parameters

Path : Path name

Msg : Error message

Err : Error code

Corrective action

Review the path name.

Lustre filesystem '*Path*' not found: *Msg* (*Err*)

Meaning

The specified path is invalid.

Description of parameters

Path : Path name

Msg : Error message

Err : Error code

Corrective action

Review the path name.

Cannot open Poolpath: *Msg* (*Err*)

Meaning

Pool for specified path is invalid.

Description of parameters

Poolpath : Path name of the pool

Msg : Error message

Err : Error code

Corrective action

Review the path name.

lfs mkdir

error: mkdir: missing dirname

Meaning

No directory name is specified.

Corrective action

Specify a directory name.

error: mkdir: missing stripe offset and count.

Meaning

The index number or the stripe count of the MDT is not specified.

Corrective action

Specify the index number or the stripe count of the MDT.

error: mkdir: bad stripe offset '*index*'

Meaning

The index number of the MDT is not specified correctly.

Description of parameters

index : The index number of MDT

Corrective action

Specify the index number of the MDT correctly.

error: mkdir: bad stripe count '<count>'

Meaning

The specified stripe count is incorrect.

Corrective action

Specify the correct stripe count.

unable to open '*Path*': *Msg* (*Err*)

Meaning

The specified path could not be opened.

Description of parameters

Path : Path name

Msg : Error message

Err : Error code

Corrective action

If the error code is 12, the error is caused by a memory shortage. Thus, a memory shortage has occurred on the computer. Take actions, such as reducing memory usage, and then perform the operation again.

If the problem still persists, collect the system log file that contains this message, and contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

error on LL_IOC_LMV_SETSTRIPE '*Path*' (*fd*): *Msg*

Meaning

Stripe settings for the specified path could not be made.

Description of parameters

Path : Path name

fd : File descriptor

Msg : Error message

Corrective action

If the error message is "Cannot allocate memory," the error is caused by a memory shortage. Thus, a memory shortage has occurred on the computer. Take actions, such as reducing memory usage, and then perform the operation again.

If the "Operation not permitted" message is output, execute this command with root privileges.

If the "stripe already set" message is output, a file already exists on the specified path. Review the path name, and perform the operation again.

If the "File name too long" or "Inappropriate ioctl for device" message is output, e specified path, review the path name.

If the problem still persists, collect the system log file that contains this message, and contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

B.2.7 lctl command

Common

Com: invalid option -- 'Opt'

Meaning

The specified command line option is invalid.

Description of parameters

Com : Command name

Opt : Invalid command line option

Corrective action

Review the command line option.

lctl qos, and lctl sqos

QoS config-file not found. filepath=Path

Meaning

The QoS definition file specified by filepath was not found.

Description of parameters

Path : Path to the specified QoS definition file

Corrective action

Confirm the path to the QoS definition file.

QoS config-file error. code=Err line=Line

Meaning

The QoS definition file contains a syntax error.

Description of parameters

Err : Error code

Character string	Meaning
E_NO_MDS	The MDS section was not found.
E_NO_OSS	The OSS section was not found.
E_SEC_DOUBLE	The section name is duplicated.
E_SEC_INVALID	The section name is invalid.
E_SEC_END	The section end character "}" is missing.
E_ITEM_INVALID	The item name is invalid.
E_ITEM_DOUBLE	The item name is duplicated.
E_ITEM_NONE	The section does not contain a required specification item (qos or nodegrp).
E_VALUE_INVALID	The specified value for the item name is invalid.

Character string	Meaning
E_LINE_OVER	The number of characters in a line exceeds the maximum value (1024).
E_RATE_OVER	The total allocation rate specified for nodegrp exceeds 100%.
E_IP_INVALID	The specification format of the IP address in nodegrp contains a syntax error.
E_NODEGRP_INVALID	The definition of nodegrp[n] corresponding to usermax[n] or rootmax[n] is missing.
E_LIMIT_INVALID	There is an error in a specification concerning a limit value for cases where there are free server threads.

Line : Line containing the error

Corrective action

Correct the QoS definition file.

QoS cannot allocate memory.

Meaning

The memory required for QoS control cannot be reserved.

Corrective action

Check the free space in memory.

QoS System error. func=*Func* route=*Route* code=*Err*

Meaning

A system error occurred.

Description of parameters

Func : Function name

Route : Error route

Err : Error code

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

QoS status is already on.

Meaning

The lctl qos on command cannot be executed because QoS control is enabled.

Corrective action

Check the QoS status with the lctl qos stat command.

QoS status is already off.

Meaning

The lctl qos off command cannot be executed because QoS control is disabled.

Corrective action

Check the QoS status with the lctl qos stat command.

QoS filepath is not full path. filepath=*Path*

Meaning

The specified path to the QoS definition file is not an absolute path.

Description of parameters

Path : Path to the specified QoS definition file

Corrective action

Confirm the path to the QoS definition file.

QoS mds is not active.

Meaning

The MDS status is not the active state.

Corrective action

Check the MDS status.

QoS file copy error. filepath=Path

Meaning

The copy process of the QoS definition file failed.

Description of parameters

Path : Path to the QoS definition file

Corrective action

Confirm that writing to the /etc/opt/FJSVfefs directory is possible and that there is free space.

QoS command multiple exec error.

Meaning

The error is a duplicate start of the command.

Corrective action

Execute the command one by one.

QoS permission denied.

Meaning

The user does not have root privileges.

Corrective action

Execute this command with root privileges.

QoS invalid uid.

Meaning

The Specified User ID is incorrect.

Corrective action

Specify the correct user ID.

QoS status is off.

Meaning

QoS function is invalid.

Corrective action

Validate the QoS function.

QoS fefs-server is not active.

Meaning

MDS or OSS is not valid state.

Also, this command cannot be executed on the client node.

Corrective action

Check the MDS or the OSS status, and execute this command on the MDS or the OSS.

QoS oss is not active.

Meaning

The OSS status is not the active state.

Corrective action

Check the OSS status.

QoS cannot be enabled.

Meaning

The QoS function cannot be enabled.

Corrective action

If you want to use QoS function, set the Lustre NRS-policy to fifo.
execute the following command.

```
[MDS node]
# lctl set_param mds.MDS.mdt.nrs_policies="fifo reg"
[OSS node]
# lctl set_param ost.OSS.ost_io.nrs_policies="fifo reg"
```

QoS mds is not MDT0.

Meaning

MDS which executed the command isn't MDT0.

Corrective action

The lctl qos command must be executed on the MDS that mounts MDT0.

lctl cqos

CQoS invalid uid.

Meaning

The Specified User ID is incorrect.

Corrective action

Specify the correct user ID.

CQoS status is off.

Meaning

QoS function is invalid.

Corrective action

Validate the QoS function.

CQoS permission denied.

Meaning

The user does not have root privileges.

Corrective action

Execute this command with root privileges.

CQoS invalid mount-point.

Meaning

The specified mount point is incorrect.

Corrective action

Specify the correct mount point.

CQoS cannot allocate memory.

Meaning

The memory required cannot be reserved.

Corrective action

Check the free space in memory.

CQoS there is no stat data.

Meaning

Valid statistics does not exist.

Corrective action

None.

CQoS System error. func=*Func* route=*Route* code=*Err*

Meaning

A system error occurred.

Description of parameters

Func : function name

Route : error route

Err : error code

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

lctl pool command and related commands

Pool *fsname.poolname* not found

Meaning

The OST_pool named *fsname.poolname* was not found.

Corrective action

Specify the correct OST_pool.

Pool *fsname.poolname* already exists

Meaning

The OST_pool named *fsname.poolname* already exists.

Corrective action

Specify a new OST_pool.

Pool *fsname.poolname* not empty, please remove all members

Meaning

The OST_pool named *fsname.poolname* is not empty. Delete all the members.

Corrective action

Delete all the members, and then delete the OST_pool.

OST Name is already in pool *fsname.poolname*

Meaning

The specified OST name already exists in the OST_pool *fsname.poolname*.

Description of parameters

Name : OST name

Corrective action

Specify a new OST name.

OST Name is not part of the '*Fsname*' fs.

Meaning

The specified OST name is not part of the file system *Fsname*.

Description of parameters

Name : OST name

Fsname : File system name

Corrective action

Specify the correct OST name.

OST Name not found in pool *fsname.poolname*

Meaning

The specified OST name does not exist in the OST_pool *fsname.poolname*.

Description of parameters

Name : OST name

Corrective action

Specify the correct OST name.

Pool *Poolname* not found

Meaning

The specified OST_pool was not found.

Description of parameters

Poolname : OST_pool name

Corrective action

Review the OST_pool name.

argument Arg must be <fsname>.<poolname>

Meaning

The specification of argument is incorrect.

Description of parameters

Arg : The name of argument

Corrective action

Review the arguments.

pool_new: File name too long

Meaning

Pool name is too long.

Corrective action

Review the pool name.

pool_new: File exists

Meaning

The specified pool name already exists.

Corrective action

Check the pool name.

pool_destroy: Directory not empty

Meaning

The ODT_pool is not empty and cannot be deleted.

Corrective action

Remove all members of ODT_pool.

No device found for name MGS: *Msg*

Meaning

No device found for MGS.

Description of parameters

Msg : Error message

Corrective action

If the error message is "Invalid argument," execute this command on MGS.

If the error message is "Permission denied," execute this command with root privileges.

Name does not start with fname Fname

Meaning

The specified OST name is incorrect.

Description of parameters

Name : OST name
Fname : File system name

Corrective action

Review the OST name

Name does not start by Fname-OST nor OST

Meaning

The specified OST name is incorrect.

Description of parameters

Name : OST name
Fname : File system name

Corrective action

Review the OST name

ost's index in Index is not an hexa number

Meaning

The specified index of the OST is incorrect.

Description of parameters

Index : Index of the OST

Corrective action

Review the OST index.

ostname Name does not end with _UUID

Meaning

The specified OST name is incorrect.

Description of parameters

Name : OST name

Corrective action

Review the OST name

'Path' is not on a Lustre filesystem: Msg (Err)

Meaning

The specified path is not on FEFS.

Description of parameters

Path : Path name
Msg : Error message
Err : Error code

Corrective action

Specify a path on FEFS.

Lustre filesystem '*Path*' not found: *Msg* (*Err*)

Meaning

The specified path is invalid.

Description of parameters

Path : Path name

Msg : Error message

Err : Error code

Corrective action

Review the path name.

Cannot open Poolpath: *Msg* (*Err*)

Meaning

The specified path or pool name is invalid.

Description of parameters

Poolpath : Path name of the pool

Msg : Error message

Err : Error code

Corrective action

Review the path or pool name.

lctl ping

Can't parse process id "*Name*"

Meaning

The specification of NID is incorrect.

Description of parameters

Name : NID

Corrective action

Specify a valid NID.

Can't parse nid "*IP*"

Meaning

The specification of an IP address is incorrect.

Description of parameters

IP : IP address

Corrective action

Specify a valid IP address.

failed to ping *Name*: *Msg*

Meaning

Failed to ping.

Description of parameters

Name : NID
Msg : Error message

Corrective action

If the error message is " Input/output error," Review the parameter.

If the error message is " Permission denied," execute this command with root privileges.

lctl set_param

error: set_param: Parameter: Msg

Meaning

Failed to set the parameter.

Description of parameters

Parameter : Parameter
Msg : Error message

Corrective action

Review the parameter.

error: set_param: setting Parameter=Value: Msg

Meaning

Failed to set the parameter.

Description of parameters

Parameter : Parameter
Value : Setting value
Msg : Error message

Corrective action

Review the parameter and the setting value.

error: set_param: Msg opening Parameter

Meaning

Failed to set the parameter.

Description of parameters

Msg : Error message
Parameter : Parameter

Corrective action

Review the parameter.

lctl get_param

error: get_param: Parameter: Msg

Meaning

Failed to get the parameter.

Description of parameters

Parameter : Parameter

Msg : Error message

Corrective action

Review the parameter.

error: get_param: opening(' Parameter ') failed: Msg

Meaning

Failed to get the parameter.

Description of parameters

Parameter : Parameter

Msg : Error message

Corrective action

Review the parameter.

error: get_param: read(' Parameter ') failed: Msg

Meaning

Failed to get the parameter.

Description of parameters

Parameter : Parameter

Msg : Error message

Corrective action

Review the parameter.

Ictl Ifsck start and Ictl Ifsck stop

device name is too long. Valid length should be less than Maxsize

Meaning

Device name is too long.

Description of parameters

Maxsize: Length of device name that can be specified.

Corrective action

Confirm the device name.

invalid switch: -c 'str'. valid switches are: empty ('on'), or 'off' without space. For example: '-c', '-con', '-coff'

Meaning

Option argument is incorrect.

Description of parameters

Str: Invalid character string

Corrective action

Confirm the character string specified in the option.

Invalid option, '-h' for help.

Meaning

Invalid option specified.

Corrective action

Specify the correct option.

Must specify device to start LFSCK.

Meaning

The specified device contains an error.

Corrective action

Confirm that the specified device is correct.

Fail to pack ioctl data: rc = Err.

Meaning

Data required for lfsck cannot be created.

Description of parameters

Err: Error code

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

Fail to start LFSCK Msg

Meaning

lfsck cannot be started.

Description of parameters

Msg: Error message

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

Fail to stop LFSCK Msg

Meaning

lfsck cannot be stopped.

Description of parameters

Msg: Error message

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

B.2.8 fsck.Idiskfs

fsck.Idiskfs: Only one of the options -p/-a, -n or -y may be specified.

Meaning

The -p|n|y option cannot be specified multiple times at once.

Corrective action

Review the command line option.

Invalid non-numeric argument to -b ("superblock")

Meaning

The argument is non-numeric.

Description of parameters

superblock : Super block

Corrective action

Specify a correct numeric value.

Device is {mounted | in use}.

Meaning

The device is being mounted or in use.

Description of parameters

Device : Device

Corrective action

Stop the FEFS service, and then perform the operation.

fsck.Idiskfs: *Msg* while trying to open device

Meaning

FEFS device check and repair failed.

Description of parameters

Msg : Error message

device : Device

Corrective action

If the "Out of memory" message is output, a memory shortage has occurred on the computer. Take actions, such as reducing memory usage, and then perform the operation again.

If the "Operation not permitted" or "Permission denied" message is output, execute this command with root privileges.

If the "Is a directory" message is output, a directory was specified. Specify the correct device, and then perform the operation again.

If the "Bad magic number in super-block" message is output, the specified super block is incorrect. Specify the correct super block, and then perform the operation again.

If the "No such file or directory" message is output, a non-existent device was specified. Specify the correct device, and then perform the operation again.

If the problem still persists, collect the system log file that contains this message, and contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

fsck.Idiskfs: invalid option -- '*Opt*'

Meaning

The specified command line option is invalid.

Description of parameters

Opt : Invalid command line option

Corrective action

Review the command line option.

fsck.ldiskfs: unrecognized option '*Opt*'

Meaning

The specified command line option is invalid.

Description of parameters

Opt : Invalid command line option

Corrective action

Review the command line option.

fsck.ldiskfs: option requires an argument -- '*Opt*'

Meaning

The argument is not specified in command line option.

Description of parameters

Opt : Command line option requiring an argument

Corrective action

Review the command line option.

B.2.9 fefsbackup [PG]

Error : 0002 : Parameters not specified in config file "<config>".

Meaning

No parameters specified in configuration file.

Description of parameters

config : Config file name

Corrective action

Contact the system administrator to check the config file.

Error : 0005 : Cannot open config file "<config>".

Meaning

Failed to open the config file.

Description of parameters

config : Config file name

Corrective action

Check the config file.

Error : 0007 : Parameter <param> specification is wrong in config file "<config>".

Meaning

There is an error in the specification of <param> in the config file <config>.

Description of parameters

config : Config file name

param : Parameter name

Corrective action

Contact the system administrator to check the config file and WORK_DIR.

Error : 0008 : Internal error has occurred(<file>:<line>:<op> failed : <err>.)

Meaning

An error occurred during file access to the archive source file system.

Description of parameters

file : File name

line : Line number

op : System call name

err : Error message

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

Error : 0009 : Cannot connect to <host> : No route to host

Meaning

Could not connect to the destination node.

Description of parameters

host : Destination node name

Corrective action

Check the status of the destination node. Contact the system administrator.

Error : 0010 : Cannot connect to <host> : Connection refused.

Meaning

Connection refused from destination node.

Description of parameters

host : Destination node name

Corrective action

Failed to connect to node displayed in *host*. Check the status of the destination node. Contact the system administrator.

Error : 0011 : Cannot connect to <host> : Host key verification failed.

Meaning

The ssh key of the destination node has been changed.

Description of parameters

host : Destination node name

Corrective action

Contact the system administrator.

Error : 0013 : rsync unknown error.

Meaning

The rsync command terminated with an error.

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

Error : 0014 : tar unknown error.

Meaning

The tar command terminated with an error.

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

Error : 0015 : missing package <package>.

Meaning

Required package is missing.

Description of parameters

package : Package name

Corrective action

Contact the system administrator to install the package.

Error : 0017 : backupdir <backupdir> does not exist.

Meaning

The directory *backupdir* specified for BACKUP_ROOT in fefsbackup_rsync.conf cannot be found.

Description of parameters

backupdir : Directory name

Corrective action

Contact the system administrator and check BACKUP_ROOT in fefsbackup_rsync.conf.

Error : 0018 : Cannot connect to <host> : Name or service not known

Meaning

Could not connect to the destination node *host*.

Description of parameters

host : Destination node name

Corrective action

Check if the destination node can resolve the name. Contact the system administrator.

Error : 0019 : Cannot get request status information <request ID>.

Meaning

Failed while getting request status.

Description of parameters

request ID : Request ID name

Corrective action

If retrying fails, delete <*request ID*> and try again.

Error : 1005 : password authentication failed.

Meaning

Pass authentication failed.

Corrective action

Check your password.

Error : 1006 : cannot access <*pathlist*>: <*errmsg*>

Meaning

Cannot access pathlist.

Description of parameters

pathlist : Pathlist name

errmsg : Error contents

Corrective action

Confirm the error content displayed and remove the cause of preventing access.

Error : 1007 : pathlist <*pathlist*> is not a regular file

Meaning

pathlist is not a regular file.

Description of parameters

pathlist : Pathlist name

Corrective action

Confirm pathlist.

Error : 1008 : cannot specified -f option and target file at once.

Meaning

You cannot specify the path in the "-f" option and argument at the same time.

Corrective action

Confirm the specified option.

Error : 1009 : pathlist <*pathlist*> is an empty file.

Meaning

pathlist specified in the "-f" option is an empty file.

Description of parameters

pathlist : Pathlist name

Corrective action

Confirm pathlist.

Error : 1011 : target file is not found.

Meaning

The specified target does not contain any files.

Corrective action

Check the copy target or file.

Error : 1012 : request ID reached limit per one day. please use "-L" option.

Meaning

The number of request IDs reached the upper limit that can be automatically acquired a day.

Corrective action

Specify the request ID with the -L option.

Error : 1013 : label <request ID> request ID already used.

Meaning

The request ID name specified in the -L option already exists.

Description of parameters

request ID : Request ID name

Corrective action

Specify another request ID name.

Error : 1014 : label <request ID> is invalid.

Meaning

The character string specified in the "-L" option contains the characters that cannot be used.

Description of parameters

request ID : Request ID name

Corrective action

Confirm the characters that can be used in the "-L" option.

Error : 1015 : label <request ID> is too long.

Meaning

The character string specified in the "-L" option is too long.

Description of parameters

request ID : Request ID name

Corrective action

Confirm the number of characters that can be used in the "-L" option.

Error : 1019 : -d option must be specified, or invalid option is specified before -d option.

Meaning

The -d option has not been specified or invalid option has been specified before -d option.

Corrective action

You must specify the destination directory with the -d option. If you specify the -d option, check the options before the -d option.

Error : 1022 : request ID <request ID> is invalid.

Meaning

Invalid request ID was specified.

Description of parameters

request ID : Request ID name

Corrective action

Check the request ID with the list subcommand or the status subcommand.

Error : 1023 : request ID <request ID> status is RUNNING.

Meaning

The request ID specified in the "-R" option is running.

Description of parameters

request ID : Request ID name

Corrective action

Specify the request ID of which status is STOPPED in the status subcommand.

Error : 1024 : delete target must be specified.

Meaning

Delete target is not specified.

Corrective action

Specify the request ID of the archive information to be deleted.

Error : 1025 : There is a new line code in the file name <filename>.

Meaning

The target file <filename> contains a line feed.

Description of parameters

filename : File name

Corrective action

Change the file name to one that does not include a line feed.

Error : 1026 : This command is available to only Administrators.

Meaning

The execution user is invalid.

Corrective action

Execute as the root user.

Error : 1027 : The permission of the work directory is invalid.

Meaning

The temporary area permissions are invalid.

Corrective action

Make the permissions accessible.

Error : 1028 : The specified work directory does not exist.

Meaning

Temporary area does not exist.

Corrective action

Specify a correct temporary area.

Error : 1029 : subcommand is not specified.

Meaning

No subcommand specified.

Corrective action

Specify a subcommand.

Error : 1030 : Available subcommands are "list", "copy", "status" and "delete"

Meaning

Invalid subcommand is specified.

Corrective action

Specify any one of the subcommands: "list" "copy" "status" or "delete".

Error : 1031 : Exclusive options -L, -u and -R are specified.

Meaning

The -L, -u, and -R options cannot be specified at the same time.

Corrective action

Check the command specifications.

Error : 1032 : Invalid option.

Meaning

Invalid option specification.

Corrective action

Check the option specifications.

Error : 1033 : File was specified as a work directory.

Meaning

A file was specified in the temporary area.

Corrective action

Specify a directory for the temporary area.

Error : 1035 : destdir <path> does not exist.

Meaning

The destination directory does not exist.

Description of parameters

path : path name

Corrective action

Specify the correct directory. Also, the path name must be an absolute path.

Error : 1036 : <*path*> is not a directory.

Meaning

No directory was specified.

Description of parameters

path : Path name

Corrective action

Specify a directory.

B.2.10 File identification tool common

Permission denied

Meaning

The user does not have sufficient privileges.

Corrective action

Execute this command with root privileges.

Can't allocate memory

Meaning

The memory required for FEFS running cannot be reserved.

Corrective action

Secure enough memory and try again.

Can't open file (*File*)

Meaning

Failed to open file.

Description of parameters

File : Objects that failed to acquire

Corrective action

Secure enough memory and try again.

Invalid argument (*Arg*)

Meaning

Argument is incorrect.

Description of parameters

Arg : Value of the specified argument

Corrective action

Specify a correct value.

File already exists (*File*)

Meaning

The file already exists.

Description of parameters

File : File name

Corrective action

Specify a file name that does not exist.

Error while trying to resolve filename

Meaning

Internal error has occurred

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

Invalid option

Meaning

Invalid option specification.

Corrective action

Specify a correct option.

Too many arguments

Meaning

Too many arguments.

Corrective action

Only 1024 inode numbers can be specified in the argument. Check the argument.

Read inode failed (*inode*), err=*errno*

Meaning

Failed to read inode.

Description of parameters

inode: inode number

errno: Error number

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

Invalid inode number (*inode*)

Meaning

An invalid inode number was specified.

Description of parameters

inode: inode number

Corrective action

Specify the correct inode number and try again.

Device Msg while opening filesystem

Meaning

Failed to open file system.

Description of parameters

Device: Device name

Msg: Error message

Corrective action

Specify the correct device and try again.

Closing filesystem failed: *Msg*

Meaning

Failed to close file system.

Description of parameters

Msg: Error message

Corrective action

No action required.

Invalid execution environment

Meaning

Invalid execution environment.

Corrective action

Check the execution environment.

Invalid mount point *Mnt*

Meaning

The specified mount point is invalid.

Description of parameters

Mnt: Mount point

Corrective action

Specify the correct mount point and try again.

No such file *File*

Meaning

The input file does not exist.

Description of parameters

File: Input file

Corrective action

Specify the correct file and try again.

Can't convert path (*fid*)

Meaning

Could not convert fid to path.

Description of parameters

fid: FID

Corrective action

The path corresponding to fid may have been deleted. Try again.

Can't merge files

Meaning

Could not output file.

Corrective action

Check your disk space and try again.

<Input>: No such file or directory while opening filesystem

Meaning

The input file does not exist.

Description of parameters

Input: Input file

Corrective action

Specify the correct file and try again.

<inode>: Use no block

Meaning

Data block is not used. Data block is not used.

Description of parameters

inode: inode number

Corrective action

No action required.

Error while trying to resolve filename

Meaning

The specified MDT device is invalid.

Corrective action

Specify the correct MDT device.

B.2.11 find_file_ost

Read file failed (*file*), err=*errno*

Meaning

Failed to read inode.

Description of parameters

file: File information

errno: Error number

Corrective action

Contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

B.2.12 convert_fid2path

missing operand

Meaning

Missing operand.

Corrective action

Specify the operand.

Invalid inputfile (*file*)

Meaning

Bad input file.

Description of parameters

file: Input file name

Corrective action

Specify the correct input file.

B.2.13 force_intr

target doesn't match target name format

Meaning

The specified *target* does not match the target device naming convention.

Description of parameters

target: The specified target device name

Corrective action

Specify the correct target device name

Unknown command: *cmd*

Meaning

Unknown command specified.

Description of parameters

cmd: Command specified as argument to -m option (String)

Corrective action

Ensure that you specify deactivate, activate, or status as the command.

Specify either -c or -s**Meaning**

The specified option is invalid.

Corrective action

Ensure that either -c or -s is specified as an option.

Cannot specify -s and -a at the same time without status command**Meaning**

The specified option is invalid.

Corrective action

If you specify the -s and -a options together, ensure that the status subcommand is specified.

B.2.14 evict_client

no such nid correspond to <IP>**Meaning**

No node exists for the specified IP address.

Description of parameters

<IP> : IP address of the node to evict

Corrective action

Specify the correct IP address.

missing operand**Meaning**

Missing operand.

Corrective action

Specify the operand.

evict_client is available to only Administrators on MDS or OSS node.**Meaning**

Cannot run on this node.

Corrective action

Run on MDS or OSS.

B.2.15 fefs_yaml2csv

fefs_yaml2csv: '*Filename*': File not found.**Meaning**

The file specified for input not found.

Description of parameters

Filename : Input file name (node information definition files or node information definition files for FX server).

Corrective action

Review the file specified in the command.

fefs_yaml2csv: '*Filename*': File is invalid.

Meaning

The file specified for input not found.

Description of parameters

Filename : Input file name (node information definition files or node information definition files for FX server).

Corrective action

Review the file specified in the command.

fefs_yaml2csv: overwrite '*Filename*' ?

Meaning

The file specified for output already exists.

Description of parameters

Filename : Output file name

Corrective action

Enter y to overwrite the output file, or n to abort the command.

If you enter y, processing continues.

If you enter n, processing stops.

fefs_yaml2csv: '*Filename*': File is invalid format. (No essential key(s) *info*)

Meaning

The format of the file specified for the input is incorrect.

Description of parameters

Filename : Input file name (node information definition files or node information definition files for FX server).

info : The message varies depending on the file being processed.

"in [node]" ... node information definition file

"in [node(ft)]" ... node information definition files for FX server

Corrective action

Error converting node information definition file or node information definition files for FX server to CSV.

Verify that the file specified for the input is in the correct format.

B.2.16 fefs_deactivate

Unknown command: *cmd*

Meaning

Unknown command specified.

Description of parameters

cmd : Command specified as argument to -m option

Corrective action

Ensure that you specify activate, deactivate, or status as the command.

Invalid mount point *Mnt*

Meaning

The specified mount point is incorrect.

Description of parameters

Mnt : Mount point

Corrective action

Specify the correct mount point.

FEFS service isn't started.

Meaning

FEFS service isn't started.

Corrective action

Start FEFS service, and then execute fefs_deactivate again.

Deactivate processing failed for system *Msg*.

Meaning

Failed to disconnect the file system(FEFS/LLIO).

Description of parameters

system : FEFS or LLIO

Msg : error messages

Corrective action

Check the Specify mount point and try again.

If the issue still persists, collect the system log file that contains this message, and contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

Activate processing failed for system *Msg*.

Meaning

Failed to incorporate the file system(FEFS/LLIO).

Description of parameters

system : FEFS or LLIO

Msg : error messages

Corrective action

Check the Specify mount point and try again.

If the issue still persists, collect the system log file that contains this message, and contact a Fujitsu systems engineer (SE) or Fujitsu Support Desk.

Appendix C Settings Required after FEFS Build

Be sure to make the following settings after FEFS build.

C.1 Configuring the FEFS Script

This section describes the procedure for configuring the FEFS script. Unless otherwise noted, execute this procedure on the active system management node.

For the configuration procedure for an environment having multiple system management nodes, see "[C.2 Configuring the FEFS Script in an Environment Having Multiple System Management Nodes](#)".



The following setting must be performed when all nodes in the cluster have started and FEFS has been mounted on them.

1. Acquire the LNet NIDs of the FEFS clients.

Execute the following:

If there are multiple compute clusters, each serving as an FEFS client, execute the command on each of them.

```
# pmexec -c <compute-cluster-name> --nodetype <type-of-node-serving-as-FEFS-client> --stdout  
"lctl list_nids" > /etc/opt/FJSVfef/s/lnetid_list_<compute-cluster-name>
```

* At the end of the file name of the output file, write the name of the compute cluster, as shown above.

2. Acquire the IP addresses of the management network of the FEFS clients

Execute the following:

If the multiuse node of the PG node is an FEFS client, also acquire it.

If there are multiple compute clusters, each serving as an FEFS client, execute the commands on each of them.

```
# pashowclst -c <compute-cluster-name> -v -l --nodetype CCM --data | grep PG | awk -F ',' '{print  
$4,$7}' > tmp.txt  
# pashowclst -c <compute-cluster-name> -v -l --nodetype LN,CN --data | grep PG | awk -F ','  
'{print $4,$6}' >> tmp.txt  
# sort tmp.txt | uniq > /etc/opt/FJSVfef/s/mngnet_list_<compute-cluster-name>  
# rm -f tmp.txt
```

* If there are no PG nodes that are FEFS clients, create an empty file named /etc/opt/FJSVfef/s/mngnet_list_<compute-cluster-name>.

3. Acquire an MDT and OST list of FEFS servers

Execute the following:

If there are multiple storage clusters, each serving as an FEFS server, execute the commands on each of them.

```
# pmexec -c <storage-cluster-name> --nodetype MDS --stdout "ls /proc/fs/lustre/md[st]/" | egrep  
"^\$MDT.{4}\$" > /etc/opt/FJSVfef/s/reconnect_srv_gmnds_mdt_list_<storage-cluster-name>  
# pmexec -c <storage-cluster-name> --nodetype OSS --stdout "ls /proc/fs/lustre/obdfilter/" |  
egrep "^\$OST.{4}\$" > /etc/opt/FJSVfef/s/reconnect_srv_goss_ost_list_<storage-cluster-name>
```

* At the end of the file name of each output file, write the name of the storage cluster, as shown above.

4. Deploy the acquired files.

Deploy the following files acquired in steps 1-3 in the /etc/opt/FJSVfef/s directory on the standby system management node:

- /etc/opt/FJSVfef/s/lnetid_list_<compute-cluster-name>
- /etc/opt/FJSVfef/s/mngnet_list_<compute-cluster-name>
- /etc/opt/FJSVfef/s/reconnect_srv_gmnds_mdt_list_<storage-cluster-name>
- /etc/opt/FJSVfef/s/reconnect_srv_goss_ost_list_<storage-cluster-name>

5. Back up the system monitoring plug-in configuration file.

Back up the system monitoring plug-in configuration file /etc/opt/FJSVtcs/pamoplugin.conf on the active system management node.

6. Check the settings in the system monitoring plug-in configuration file.

Output the current settings of the system monitoring plug-in processing function to the /etc/opt/FJSVtcs/pamoplugin.conf file.

```
# pamopluginadm --show > /etc/opt/FJSVtcs/pamoplugin.conf
```

7. Edit the system monitoring plug-in configuration file

Add the following settings to the /etc/opt/FJSVtcs/pamoplugin.conf file on the active system management node.

- Add the following to the compute cluster definitions. If there are multiple compute clusters, each serving as an FEFS client, add the setting for each of them.

```
PluginCmd {  
    NodeType = "<type-of-node-serving-as-FEFS-client>"  
    ServiceName = "OS"  
    Status = "x"  
    Cmd = "bash /opt/FJSVfefefs/sbin/plugin_evict_client.sh -s <storage-cluster-name>"  
}  
PluginCmd {  
    NodeType = "<type-of-node-serving-as-FEFS-client>"  
    ServiceName = "OS"  
    Status = "-"  
    Cmd = "bash /opt/FJSVfefefs/sbin/plugin_evict_client.sh -s <storage-cluster-name>"  
}  
PluginCmd {  
    NodeType = "<type-of-node-serving-as-FEFS-client>"  
    ServiceName = "FEFS"  
    Status = "x"  
    Cmd = "bash /opt/FJSVfefefs/sbin/plugin_evict_client.sh -s <storage-cluster-name>"  
}
```

* If there are multiple types of nodes that serve as FEFS clients, specify those node types by delimiting them with a comma (,), as shown below. The following example assumes that nodes of the CCM, LN, and CN node types are serving as FEFS clients:

```
NodeType = "CCM,LN,CN"
```

```
PluginCmd {  
    NodeType = "GIO"  
    ServiceName = "OS"  
    Status = "x"  
    Cmd = "bash /opt/FJSVfefefs/sbin/plugin_lnet_router_dwn.sh -s <storage-cluster-name>"  
}
```

* The -s option specifies the storage cluster name to which the FEFS server, where the FEFS client under that compute cluster is mounted, is supposed to belong.

- Add the following to the storage cluster definitions.

```
PluginCmd {  
    NodeType = "MDS"  
    ServiceName = "OS"  
    Status = "x"  
    Cmd = "bash /opt/FJSVfefefs/sbin/plugin_reconnect_srv.sh -c <compute-cluster-name> --  
nodetype MDS"  
}  
PluginCmd {  
    NodeType = "OSS"  
    ServiceName = "OS"  
    Status = "x"  
    Cmd = "bash /opt/FJSVfefefs/sbin/plugin_reconnect_srv.sh -c <compute-cluster-name> --
```

```
    nodetype OSS"
}
```

* If there are multiple compute clusters that serve as FEFS clients in the global file system configured with an associated storage cluster, write the -c option for each compute cluster. The following example has FEFS clients of the global file system under the compute cluster calc1 and the compute cluster calc2.

```
PluginCmd {
    NodeType = "MDS"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfeffs/sbin/plugin_reconnect_srv.sh -c calc1 --nodetype MDS"
}
PluginCmd {
    NodeType = "OSS"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfeffs/sbin/plugin_reconnect_srv.sh -c calc1 --nodetype OSS"
}
PluginCmd {
    NodeType = "MDS"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfeffs/sbin/plugin_reconnect_srv.sh -c calc2 --nodetype MDS"
}
PluginCmd {
    NodeType = "OSS"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfeffs/sbin/plugin_reconnect_srv.sh -c calc2 --nodetype OSS"
}
```

8. Register the system monitoring plug-in configuration file.

Register the contents of the /etc/opt/FJSVtcs/pamopugin.conf file in the Job Operation Software.

```
# pamopuginadm --set
```

9. Check the settings in the system monitoring plug-in configuration file.

Check the settings to make sure that the /etc/opt/FJSVtcs/pamopugin.conf file registered in step 8 is registered correctly.

```
# pamopuginadm --show
```

10. Check crond of the MDS.

Confirm that crond of the MDS is on.

```
# pmexe -c <storage-cluster-name> --nodetype MDS --stdout "systemctl is-enabled crond.service"
```

If crond of the MDS is not on, execute the following command on the MDS to set crond to on.

```
[MDS node]
# systemctl enable crond.service
```

11. Check the status of crond of the MDS.

Confirm that crond of the MDS is running.

```
# pmexe -c <storage-cluster-name> --nodetype MDS --stdout "systemctl status crond.service"
```

If crond of the MDS is not running, start crond by executing the following command on the MDS.

```
[MDS node]
# systemctl start crond.service
```

12. Set crontab of the MDS.

Configure the FEFS script for cron of the MDS.

```
# pmexec -c <storage-cluster-name> --nodetype MDS --stdout "/opt/FJSVfef/sbin/activate_device_cron.sh"
```

13. Check the crontab setting.

Confirm that the script is correctly configured for cron of the MDS.

```
# pmexec -c <storage-cluster-name> --nodetype MDS --stdout "crontab -l"
```

Confirm the following setting:

```
*/1 * * * * bash /opt/FJSVfef/sbin/activate_device.sh
```

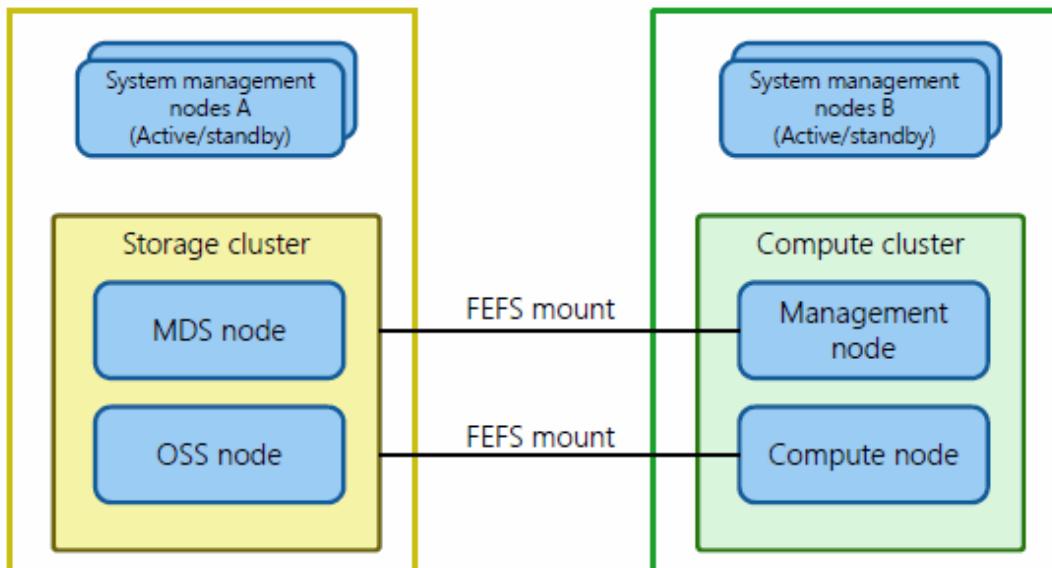
C.2 Configuring the FEFS Script in an Environment Having Multiple System Management Nodes

For an environment having multiple system management nodes, configure the FEFS script by using the procedure described here. Do not use this procedure for an environment that does not have multiple system management nodes.

In the following figure and description, the system management nodes (active/standby) that have a storage cluster are referred to as system management node A. Also, the system management nodes (active/standby) that have a compute cluster are referred to as system management node B.

An environment that has multiple system management nodes is an environment that has multiple system management nodes in a redundant configuration. The FEFS configuration has FEFS servers (MDS and OSS) belonging to the storage cluster under system management node A. FEFS clients (such as CCM, LN, and CN) belong to the compute cluster under system management node B. FEFS is mounted as shown.

Figure C.1 Environment that has multiple system management nodes



The following configuration procedure is based on the above figure.



- The following setting must be performed when all nodes in the cluster have started and FEFS has been mounted on them.
- If the storage cluster under system management node A has the same cluster name as the compute cluster under system management node B, the script configured in this procedure does not work. Do not set the same cluster name for the compute cluster and storage cluster.

1. Acquire the LNet NIDs of the FEFS clients.

Acquire the LNet NIDs of the FEFS clients by executing the following command on the active system management node B.
If there are multiple compute clusters, each serving as an FEFS client, execute the command on each of them.

```
[System management node B]
# pmexec -c <compute-cluster-name> --nodetype <type-of-node-serving-as-FEFS-client> \
--stdout "lctl list_nids" > /etc/opt/FJSVfefes/lnid_list_<compute-cluster-name>
```

* At the end of the file name of the output file, write the name of the compute cluster, as shown above.

2. Deploy the files.

Deploy the /etc/opt/FJSVfefes/lnid_list_<compute-cluster-name> file acquired in step 1 in also the /etc/opt/FJSVfefes/ directory on the standby node of system management node B and on the active/standby node of system management node A.

3. Acquire the IP addresses of the management network of the FEFS clients

Acquire the IP addresses of the management network of the FEFS clients by executing the following commands on the active system management node B.

If the multiuse node of the PG node is an FEFS client, also acquire it.

If there are multiple compute clusters, each serving as an FEFS client, execute the commands on each of them.

```
[System management node B]
# pashowclst -c <compute-cluster-name> -v -l --nodetype CCM --data | grep PG | awk -F ',' '{print
$4,$7}' > tmp.txt
# pashowclst -c <compute-cluster-name> -v -l --nodetype LN,CN --data | grep PG | awk -F ',' '{print
$4,$6}' >> tmp.txt
# sort tmp.txt | uniq > /etc/opt/FJSVfefes/mngnet_list_<compute-cluster-name>
# rm -f tmp.txt
```

* If there are no PG nodes that are FEFS clients, create an empty file named /etc/opt/FJSVfefes/mngnet_list_<compute-cluster-name>.

4. Deploy the files.

Deploy the /etc/opt/FJSVfefes/mngnet_list_<compute-cluster-name> file acquired in step 3 in also the /etc/opt/FJSVfefes/ directory on the standby node of system management node B and on the active/standby node of system management node A.

5. Acquire an MDT and OST list of FEFS servers.

Acquire an MDT and OST list of FEFS servers by executing the following command on system management node A.

```
[System management node A]
# pmexec -c <storage-cluster-name> --nodetype MDS --stdout "ls /proc/fs/lustre/md[st]/" | \
egrep "^\-MDT.{4}$" > /etc/opt/FJSVfefes/reconnect_srv_gmnds_mdt_list_<storage-cluster-name>
# pmexec -c <storage-cluster-name> --nodetype OSS --stdout "ls /proc/fs/lustre/obdfilter/" | \
egrep "^\-OST.{4}$" > /etc/opt/FJSVfefes/reconnect_srv_goss_ost_list_<storage-cluster-name>
```

* At the end of the file name of each output file, write the name of the compute or storage cluster, as shown above.

6. Deploy the files.

Deploy the following files acquired in step 5 in also the /etc/opt/FJSVfefes/ directory on the standby node of system management node A and on the active and standby nodes of system management node B:

- /etc/opt/FJSVfefes/reconnect_srv_gmnds_mdt_list_<storage-cluster-name>
- /etc/opt/FJSVfefes/reconnect_srv_goss_ost_list_<storage-cluster-name>

7. Back up the system monitoring plug-in configuration file.

Back up the system monitoring plug-in configuration file /etc/opt/FJSVtcs/pamopugin.conf on the active node of system management node A and system management node B.

8. Check the settings in the system monitoring plug-in configuration file.

Output the current settings of the system monitoring plug-in function to the /etc/opt/FJSVtcs/pamopugin.conf file.

```
[System management node A and System management node B]
# pamopuginadm --show > /etc/opt/FJSVtcs/pamopugin.conf
```

9. Edit the system monitoring plug-in configuration file.

a. Add the following settings to the /etc/opt/FJSVtcs/pamoplugin.conf file on the active node of system management node B.

1. Add the following to the compute cluster definitions. If there are multiple compute clusters, each serving as an FEFS client, execute the command on each of them.

```
PluginCmd {
    NodeType = "<type-of-node-serving-as-FEFS-client>"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfeffs/sbin/plugin_evict_client_multiple_execute.sh --ip
<representative-IP-address-of-system management node-A> -s <storage-cluster-name>"
}
PluginCmd {
    NodeType = "<type-of-node-serving-as-FEFS-client>"
    ServiceName = "OS"
    Status = "-"
    Cmd = "bash /opt/FJSVfeffs/sbin/plugin_evict_client_multiple_execute.sh --ip
<representative-IP-address-of-system management node-A> -s <storage-cluster-name>"
}
PluginCmd {
    NodeType = "<type-of-node-serving-as-FEFS-client>"
    ServiceName = "FEFS"
    Status = "x"
    Cmd = "bash /opt/FJSVfeffs/sbin/plugin_evict_client_multiple_execute.sh --ip
<representative-IP-address-of-system management node-A> -s <storage-cluster-name>"
}
```

* If there are multiple types of nodes that serve as FEFS clients, specify those node types by delimiting them with a comma (,), as shown below. The following example assumes that nodes of the CCM, LN, and CN node types are serving as FEFS clients:

```
NodeType = "CCM,LN,CN"
```

* *<representative-IP-address-of-system management node-A>* specified in the --ip option specifies the representative IP address of system management node A.
* *<storage-cluster-name>* specified in the -s option specifies the name of the storage cluster under system management node A.

2. Add the following to the compute cluster definitions.

```
PluginCmd {
    NodeType = "< type-of-node-serving-as-FEFS-client >"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfeffs/sbin/plugin_lnet_router_dwn_rmtcmd_execute.sh --ip
<representative-IP-address-of-system-administration-node-A> -s <storage-cluster-name>"}
```

* *<representative-IP-address-of-system management node-A>* specified in the --ip option specifies the representative IP address of system management node A.
* *<storage-cluster-name>* specified in the -s option specifies the name of the storage cluster under system management node A

- b. Add the following to the storage cluster definitions in the /etc/opt/FJSVpnavi/pamoexit.conf file on the active node of system management node A.

```
PluginCmd {
    NodeType = "MDS"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfeffs/sbin/012-0614-0913_TCS_SMM_rmtcmd_execute.sh --ip
<representative-IP address-of-system management node-B> -c <compute-cluster-name> --
```

```

    nodetype MDS"
}
PluginCmd {
    NodeType = "OSS"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefes/sbin/012-0614-0913_TCS_SMM_rmtcmd_execute.sh --ip
<representative-IP address-of-system management node-B> -c <compute-cluster-name> --
nodetype OSS"
}

```

* <representative-IP-address-of-system management node-B> specified in the --ip option specifies the representative IP address of system management node B.

* <compute-cluster-name> specified in the -c option specifies the name of the compute cluster under system management node B. If there are multiple compute clusters under system management node B, write the -c option for each compute cluster.

The following example has the compute clusters calc1 and calc2 under system management node B.

```

PluginCmd {
    NodeType = "MDS"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefes/sbin/plugin_reconnect_srv_rmtcmd_execute.sh --ip
xxx.xxx.xxx -c calc1 --nodetype MDS"
}
PluginCmd {
    NodeType = "OSS"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefes/sbin/plugin_reconnect_srv_rmtcmd_execute.sh --ip
xxx.xxx.xxx -c calc1 --nodetype OSS"
}
PluginCmd {
    NodeType = "MDS"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefes/sbin/plugin_reconnect_srv_rmtcmd_execute.sh --ip xxx.xxx.xxx -c calc2 --nodetype MDS"
}
PluginCmd {
    NodeType = "OSS"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefes/sbin/plugin_reconnect_srv_rmtcmd_execute.sh --ip
xxx.xxx.xxx -c calc2 --nodetype OSS"
}

```

10. Register the system monitoring plug-in configuration file.

Register the contents of the /etc/opt/FJSVtcs/pamopugin.conf file in the Job Operation Software on the active node of system management node A and system management node B.

```
[System management node A and System management node B]
# pamopuginadm --set
```

11. Check the settings in the system monitoring plug-in configuration file.

Check the settings to make sure that the /etc/opt/FJSVtcs/pamopugin.conf file registered in step 10 on the active node of system management node A and system management node B is registered correctly.

```
[System management node A and System management node B]
# pamopuginadm --show
```

C.3 Settings Required for Using ETERNUS

The following settings are required for an environment using ETERNUS. If ETERNUS is not used, these settings are not required.

C.3.1 Enabling ETERNUS NRDY measures on the MDS

1. Preparation

Referring to the *ETERNUS multipath driver individual surveillance setting manual for the FEFS environment*, which is included with the product, change the monitoring interval setting of the path status before enabling ETERNUS NRDY measures on the MDS.

2. Enable with a setting on the MDS.

The two methods of enabling ETERNUS NRDY measures on the MDS are as follows. Use either method for the setting.

- Setting for individual node types

Execute the following command on the system management node to create the /etc/opt/FJSVfefs/failover_sleep_file configuration file for the target node.

The following example makes the setting for the MDS in the clst cluster.

```
# pmexe -c clst --stdout --nodetype MDS "touch /etc/opt/FJSVfefs/failover_sleep_file"
```

- Setting for individual nodes

Log in to the target node, and create the /etc/opt/FJSVfefs/failover_sleep_file configuration file.

```
# touch /etc/opt/FJSVfefs/failover_sleep_file
```



Information

When ETERNUS NRDY measures are enabled on the MDS, the mount processing at failover is delayed by 7 minutes and 30 seconds.

C.3.2 Procedure for configuring the OSS autoboot script

Configure the plugin_fefs_autoboot_pairnode.sh script to automatically start the OSS when an event such as an ETERNUS CM (controller) restart on both the master CM and slave CM causes both nodes of the redundantly configured OSS to stop.

Use the following configuration procedure.

1. Preparation

For configuration instructions, refer to the documentation provided if you have installed Dump assist tool of the Fujitsu Linux Support Package.

If not, see the "Kernel Crash Dump Guide" published by Red Hat, Inc.



After a dump is collected, if you check the node status by using the -v option of the pashowclst command on the system management node, the command displays the following.

STATUS	REASON	PWR_STATUS
Stopped	-	off

2. Change parameters in the script.

Change parameters in the fefs_autoboot_pairnode.sh script to suit operation. The script has the following parameters:

- Dump collection completion wait setting
- Node startup suppression setting
- Node status check setting after node startup

After the script is applied, the script immediately reflects any changes made to its contents. For details on how to apply the script, see step "3. Apply the script."

The methods of setting each parameter are as follows.

[Dump collection completion wait setting]

Set the length of time to wait for power-off after the completion of dump collection.

The following settings are an example.

```
DUMP_CHK_TIME=40      # 40min          Time during which the completion of dump collection  
will be awaited  
DUMP_CHK_INTERVAL=5    # 5min           Interval at which the completion of dump collection  
will be checked  
DUMP_CHK_FORCE_RESET="Yes" # Yes(default)|No Action in the event of timeout
```

For DUMP_CHK_TIME, set the wait time for the completion of dump collection. The default value is 40 minutes. For DUMP_CHK_INTERVAL, set the interval for checking the completion of dump collection. The default value is 5 minutes. Specify these parameter values in units of minutes. For DUMP_CHK_FORCE_RESET, set the operation performed if power-off is not detected within the wait time (timeout time). The default operation is forced startup by a reset (setting value: Yes). If "No" is the setting value, processing ends without doing anything.

[Node startup suppression setting]

Set suppression so that the script retains and checks start records to prevent recurrences of node panic and restarts due to hardware failures, etc. after startup.

```
BOOT_RESERVE=24      # 24hour  Time range over which the record of the most recent starts will be  
checked  
BOOTED_COUNT=1       # 1 time    Number of start times before node startup is suppressed
```

The default setting is once every 24 hours.

For BOOT_RESERVE, set the range of time over which records of the most recent starts are checked. Set this value in units of hours. The default value is 24 hours (24hour). For BOOTED_COUNT, set how many starts occur before node startup is suppressed. Set a value within the range that is set in BOOT_RESERVE. Specify this value as a number of times. If the set value is "0", node startup is always suppressed.

[Node status check setting after node startup]

Set the length of time to wait for the node to enter the Running state after node startup.

```
BOOTSTATUS_CHK_TIME=40      # 40 minutes Status check time  
BOOTSTATUS_CHK_INTERVAL=5   # 5 minutes  Status check interval
```

For BOOTSTATUS_CHK_TIME, set the period of time to check the node status. The default value is 40 minutes. For BOOTSTATUS_CHK_INTERVAL, set the interval for the node status check. The default value is 5 minutes. Specify these parameter values in units of minutes.

If the node does not transition to the Running state within the set check period, the start is considered an abnormal start, and the following message is output to the system management node system log.

```
fefs_autoboot_pairnode.sh: [ERR.] STATUS Running was not able to be detected.  
(clstname:clstname NID:0xFFFFFFFF PAIR_NID:0xFFFFFFFF)
```

3. Apply the script.

a. Deploy the script.

Deploy the plugin_fefs_autoboot_pairnode.sh script in the /opt/FJSVfef/ directory on the active and standby system management nodes.

b. Back up the system monitoring plug-in configuration file.

Back up the system monitoring plug-in configuration file /etc/opt/FJSVtcs/pamoplugin.conf on the active system management node.

c. Check the settings in the system monitoring plug-in configuration file.

Output the current settings of the system monitoring plug-in function to the /etc/opt/FJSVtcs/pamoplugin.conf file.

```
# pamopluginadm --show > /etc/opt/FJSVtcs/pamoplugin.conf
```

- d. Edit system monitoring plug-in configuration file.

Add the following storage cluster definition to the /etc/opt/FJSVtcs/pamoplugin.conf file on the active system management node.

```
PluginCmd {
    ServiceName = "OS"
    NodeType = "OSS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefefs/plugin_fefs_autoboot_pairnode.sh"
}
```

- e. Register system monitoring plug-in configuration file.

Register the contents of the /etc/opt/FJSVtcs/pamoplugin.conf file in the Job Operation Software.

```
[Active system management node]
# pamopluginadm --set
```

- f. Check the settings in the system monitoring plug-in configuration file

Check if the /etc/opt/FJSVtcs/pamoplugin.conf file registered in step e is registered correctly.

```
[Active system management node]
# pamopluginadm --show
```



Information

To delete the script, execute the following procedure.

- a. Back up the abnormal exit configuration file.

The step is the same as step b for applying the script.

- b. Check the contents of settings in the abnormal exit configuration file.

The step is the same as step c for applying the script.

- c. Edit the abnormal exit processing configuration file.

Delete the following storage cluster definitions from the /etc/opt/FJSVpnavi/pamoexit.conf file on the active system management node.

```
PluginCmd {
    ServiceName = "OS"
    NodeType = "OSS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefefs/plugin_fefs_autoboot_pairnode.sh"
}
```

- d. Register the abnormal exit processing configuration file.

The step is the same as step e for applying the script.

- e. Check the contents of settings in the abnormal exit processing configuration file.

The step is the same as step f for applying the script.

```
[Active system management node]
# pamopluginadm --show
```

- f. Delete the script.

Delete the plugin_fefs_autoboot_pairnode.sh script from the /opt/FJSVfefefs/ directory on the active and standby system management nodes.

Corrective action for a node startup failure after both nodes of the redundantly configured OSS stopped

If the path to the redundantly configured OSS is blocked because of, for example, an ETERNUS CM restart, both nodes of the OSS are powered off after dump collection is completed.

When both nodes of the OSS have stopped, the OSS autoboot script powers on the stopped nodes, but they may be unable to start

normally such as because of an OSS startup failure on one of the nodes. Then, operation would not continue. In such cases, take the following action.

[Detection method]

1. Check the system log messages.

After power-on, if one of the nodes does not start normally and thus does not enter the Running state, the following message is output to the system management node system log (/var/log/messages).

```
plugin_fefs_autoboot_pairenode.sh: [ERR.] STATUS Running was not able to be detected.  
(clstname:clstname NID:0xxxxxxxxx PAIR_NID:0xxxxxxxxx)
```

clstname: Target cluster name

NID: Node ID of the node that did not start normally

PAIR_NID: Node ID of the paired node in a redundant configuration with the node that did not start normally



This procedure cannot be applied to cases where the message was output from both nodes, because neither node started normally. In this situation, check the power status and other conditions of both nodes of the OSS, eliminate the cause of the failure, and restart them.

2. Check the node status.

Check the STATUS and PWR_STATUS column values of the node that is the subject of the output message. Also, confirm that the STATUS column shows "Running" for the node paired with the target node.

You can check the target cluster name and the node IDs of the target node and paired node in the message that was output in step 1.

Example 1: After node startup, the FEFS service does not enter the "o" state.

```
[System management node]  
# pashowclst -c storage -n 0x00000009,0x0000000A  
[ CLST: storage ]  
[ NODE: 0x00000009 ]  
NODE      NODETYPE    STATUS     REASON      PWR_STATUS    ARCH_STATUS    SRV_STATUS  
0x00000009 OSS        Init       -           on            -             FEFSSR(b)  
[ NODE: 0x0000000A ]  
NODE      NODETYPE    STATUS     REASON      PWR_STATUS    ARCH_STATUS    SRV_STATUS  
0x0000000A OSS        Running   -           on            -             FEFSSR(o)
```

Example 2: The node has been powered on but remains stopped and does not start.

```
[System management node]  
# pashowclst -c storage -n 0x00000009,0x0000000A  
[ CLST: storage ]  
[ NODE: 0x00000009 ]  
NODE      NODETYPE    STATUS     REASON      PWR_STATUS    ARCH_STATUS    SRV_STATUS  
0x00000009 OSS        Stopped   -           off           -             -  
[ NODE: 0x0000000A ]  
NODE      NODETYPE    STATUS     REASON      PWR_STATUS    ARCH_STATUS    SRV_STATUS  
0x0000000A OSS        Running   -           on            -             FEFSSR(o)
```

[Corrective action]

If either of steps 1 and 2 in the detection method found an applicable condition, stop the node that did not start normally, and continue with one-sided operation where the FEFS service runs on only the paired node.

1. Stop the target node.

If "Running" in the STATUS column and "off" in the PWR_STATUS column are not shown for the node that did not start

normally, issue init to stop this node. If the PWR_STATUS column shows "off", this step is not necessary because the node has been powered off.

```
[System management node]
# papwrctl -c <storage-cluster-name> -n <nid-of-target-node> init
```

2. Put the FEFS service into one-sided operation.

Execute the following command to go from the target node to one-sided operation where the FEFS service runs on only the paired node.

```
[System management node]
# pmexe -c <storage-cluster-name> -n <nid-of-paired-node> /bin/fefs_failover --active -f
```

Except in this step, do not use this command.

3. Recover the FEFS service from one-sided operation.

After maintenance on the target node, recover the FEFS service from one-sided operation to return to the original operating state.

```
[System management node]
# paclstmgr -c <storage-cluster-name> --failback <nid-of-target-node>
```

C.4 Configuring periodic deletion of FEFS logs

This section describes the procedure for configuring periodic deletion of old log files under /var/opt/FJSVfef/s/dumplog on the FEFS server.

Unless otherwise noted, execute this procedure on the active system management node.

1. Check crond of the FEFS server.

Confirm that the crond of the FEFS server is enabled.

```
# pmexe -c <storage-cluster-name> --nodetype <type-of-node-serving-as-FEFS-server> \
--stdout "systemctl is-enabled crond.service"
```

If the crond of the FEFS server is not enabled, execute the following command to set the crond to enabled.

```
# pmexe -c <storage-cluster-name> --nodetype <type-of-node-serving-as-FEFS-server> \
--stdout "systemctl enable crond.service"
```

2. Check the status of crond of the FEFS server.

Confirm that the status of the crond of the FEFS server is running.

```
# pmexe -c <storage-cluster-name> --nodetype <type-of-node-serving-as-FEFS-server> \
--stdout "systemctl status crond.service"
```

If the status of the crond of the FEFS server is not running, execute the following command to start the crond.

```
# pmexe -c <storage-cluster-name> --nodetype <type-of-node-serving-as-FEFS-server> \
--stdout "systemctl start crond.service"
```

3. Set crontab of the FEFS server.

Configure the crond of the FEFS server for periodically deleting old log files that were created more than two months ago.

```
# pmexe -c <storage-cluster-name> --nodetype <type-of-node-serving-as-FEFS-server> \
--stdout "echo \"0 0 * * * /bin/find /var/opt/FJSVfef/s/dumplog/ -type f -mtime +60 -delete\" \
>> /var/spool/cron/root"
```

4. Check the crontab setting.

Confirm that the crond of the FEFS server is correctly configured.

```
# pmexe -c <storage-cluster-name> --nodetype <type-of-node-serving-as-FEFS-server> \
--stdout "crontab -l"
```

Confirm the following setting:

```
0 0 * * * /bin/find /var/opt/FJSVfefs/dumplog/ -type f -mtime +60 -delete
```

Appendix D Procedures to Recover the File System

This appendix describes recovery procedures for cases where a file system failure has occurred.

D.1 Introduction

This appendix assumes the following failures:

- Defective block
An unusable LBA (Logical Block Addressing) address on a device has been detected.
- Disk failure
A file server is not accessible due to a physical hard disk failure.
- File system crash
Due to a non-hardware cause, a file server cannot be mounted or file I/O is not being done normally. EIO has been returned.
- Both servers stopped
Both the active and standby server nodes (MGS, MDS, or OSS) have crashed.

D.2 Impact

MGS

If both MGS systems have crashed and the MGT disk failed or the file system crashed, the MGS cannot be remounted. After a server restart, operations such as restarting clients are performed. FEFS will fail to start.

MDS

The file system cannot continue operating in the following circumstances:

- Crash of both systems for the MDS mounted at MDT0 (index 0 on an FEFS design sheet)
- Failure of the disk allocated to MDT0

Also, the use of the multi MDS function disables access to the stored files and directories on the related disks in the following circumstances:

- Crash of both systems for the MDS with any mounted MDT not at MDT0
- Disk failure not at MDT0

OSS

If both OSS systems have crashed and the OST disk failed or the file system crashed, file I/O to a file stored on this OST will hang or be recovering from an error.

D.3 Failure Recovery Flow

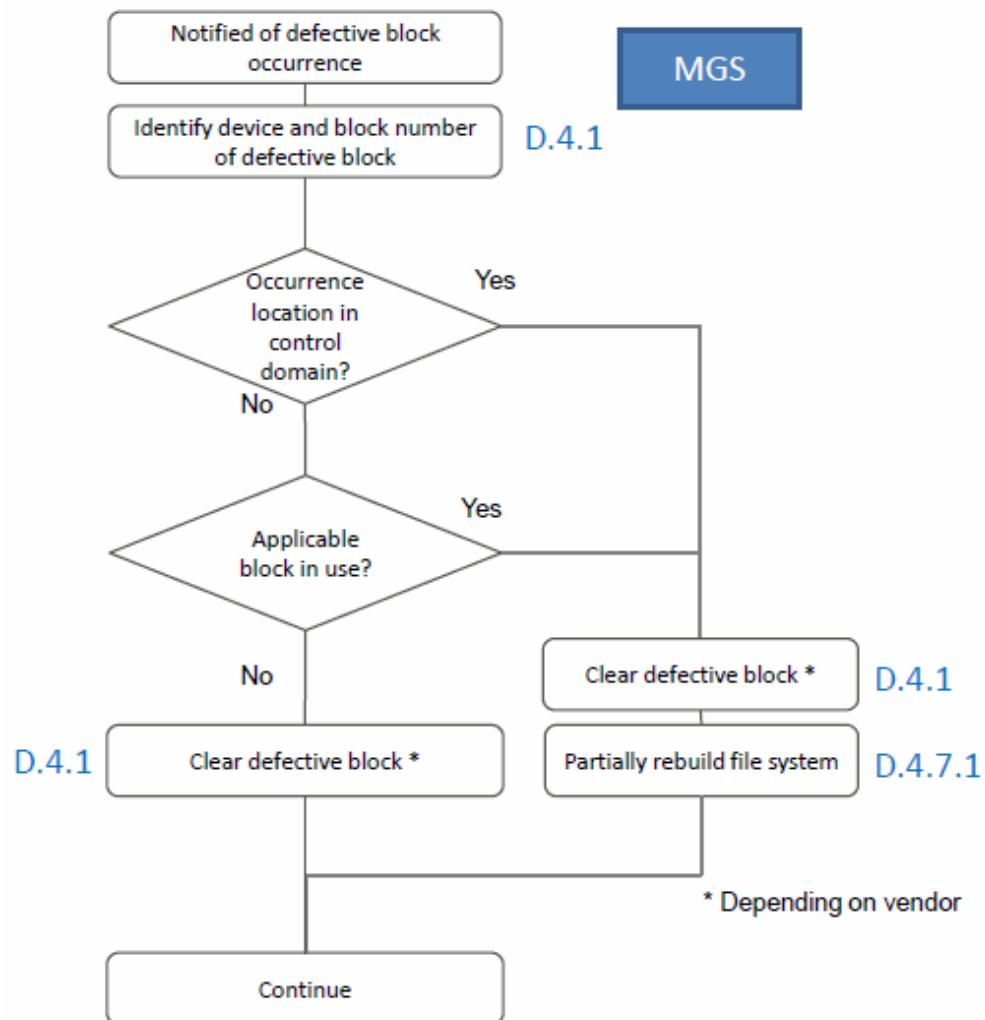
This section shows the flow when a failure has occurred on the MGS, MDS, or OSS.

The following sections show the operations in each flow. The numbers shown in blue in the figures are section numbers in "[D.4 Response Procedures](#)" describing the operations in detail.

D.3.1 Recovery flow when a defective block is detected

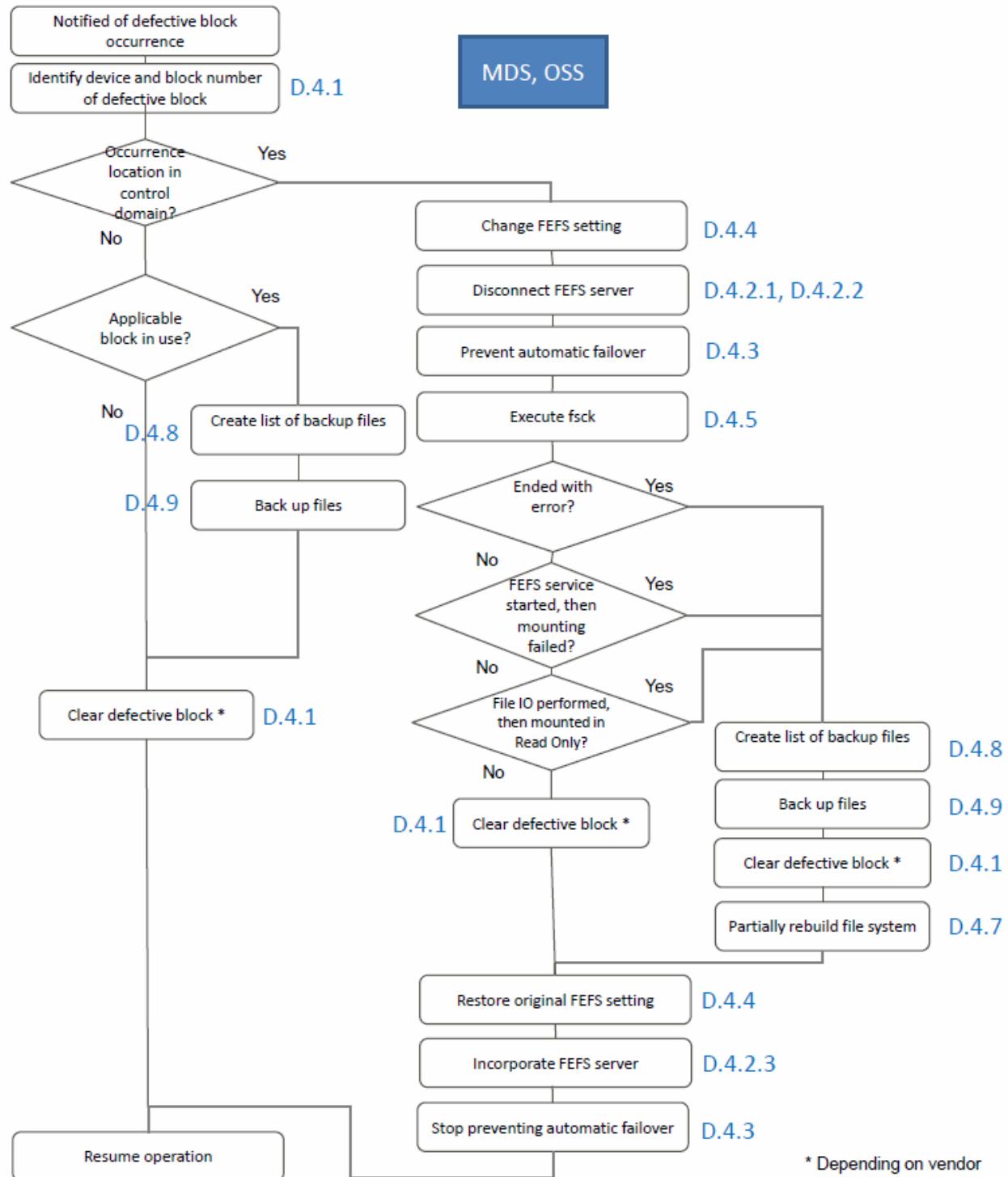
The following figure shows the MGT recovery flow when a defective block is detected.

Figure D.1 MGT recovery flow when a defective block is detected



The following figure shows the MDT and OST recovery flow when a defective block is detected.

Figure D.2 MDT/OST recovery flow when a defective block is detected



D.3.2 Recovery flow after a disk failure or file system crash

The following figures show the trouble handling flow after a disk failure or file system crash.

Figure D.3 OST recovery flow after a disk failure or file system crash

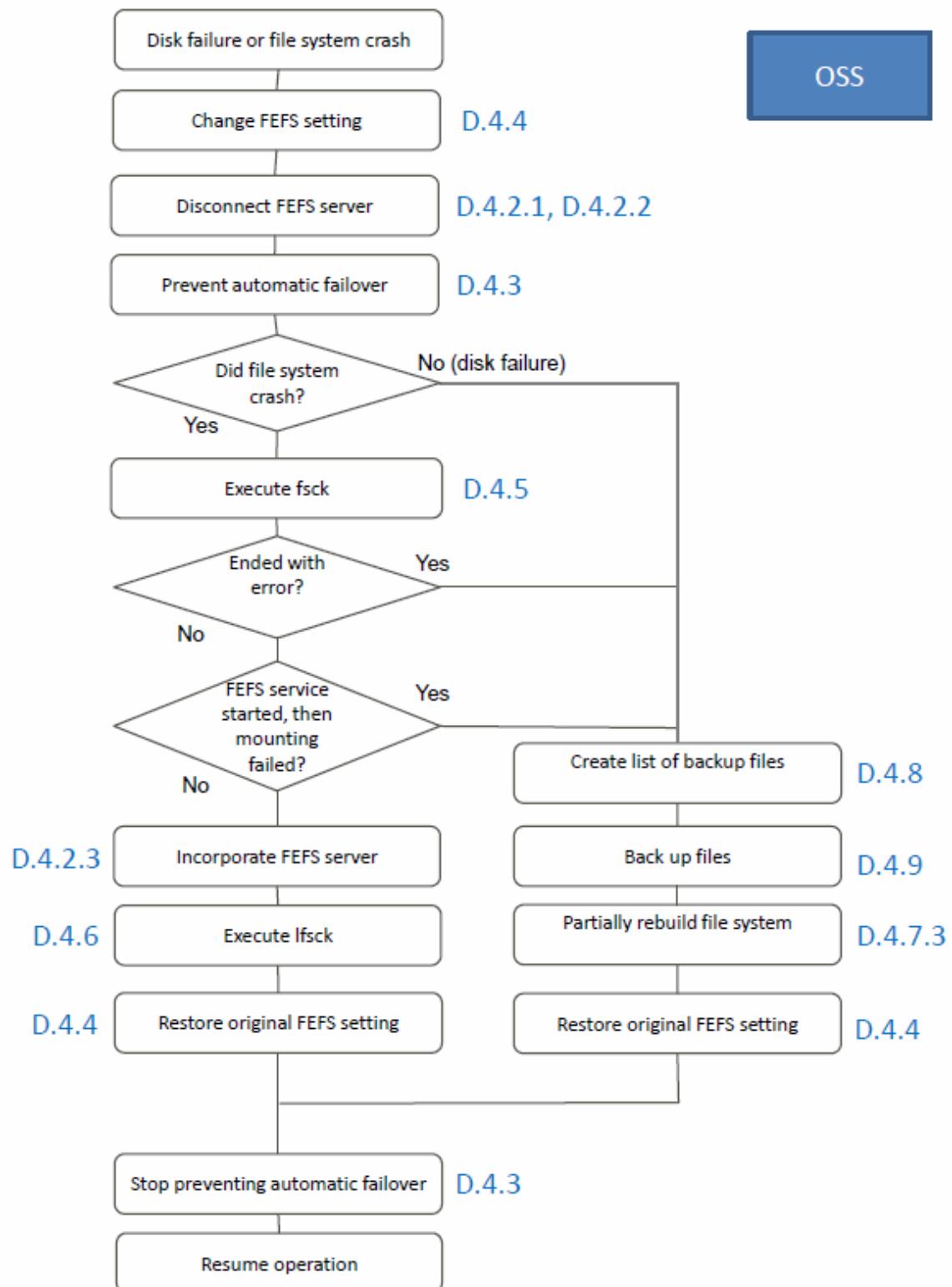


Figure D.4 MDT recovery flow after a disk failure or file system crash (including cases with an MGT operating as an MDT too)

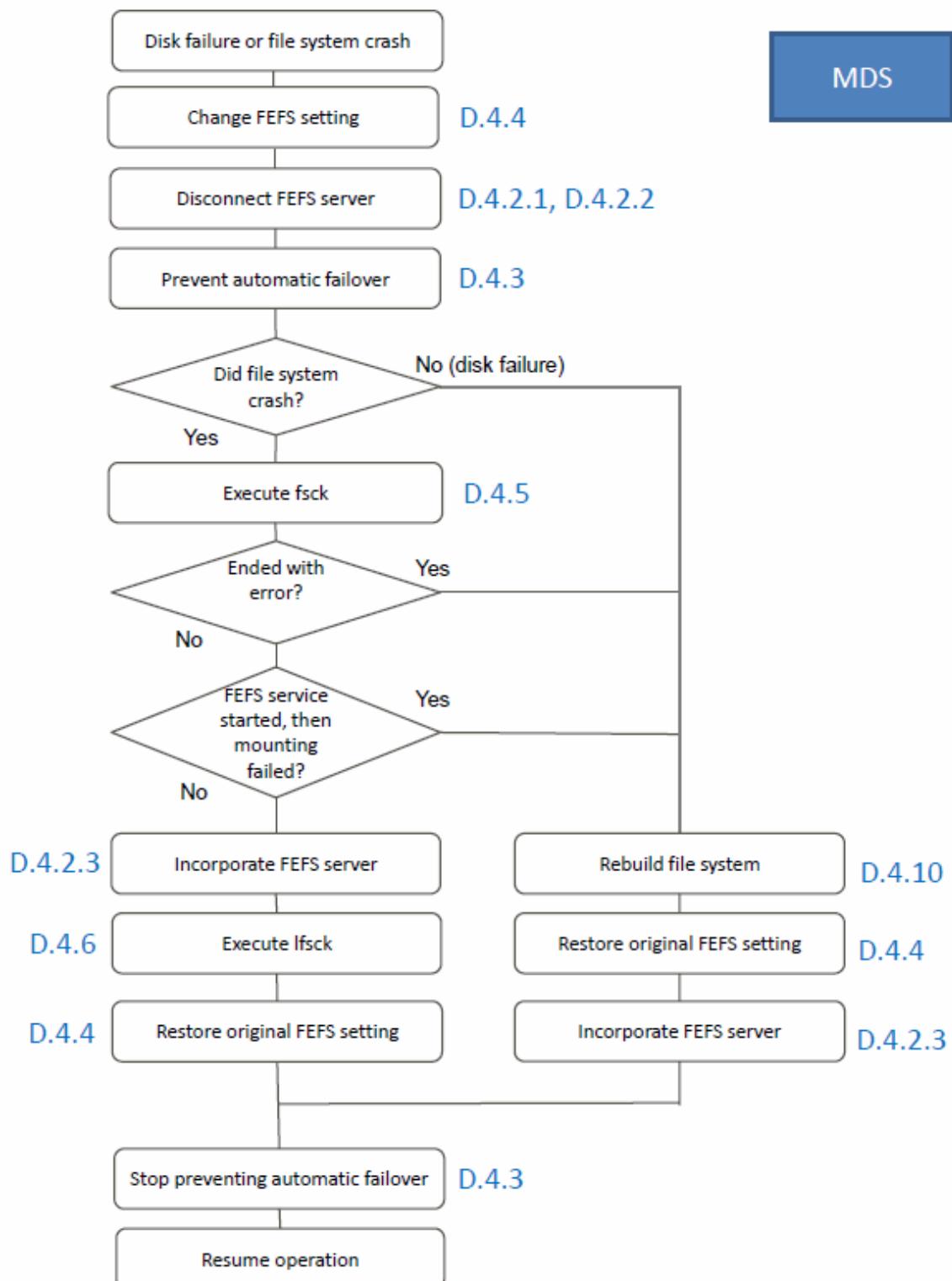
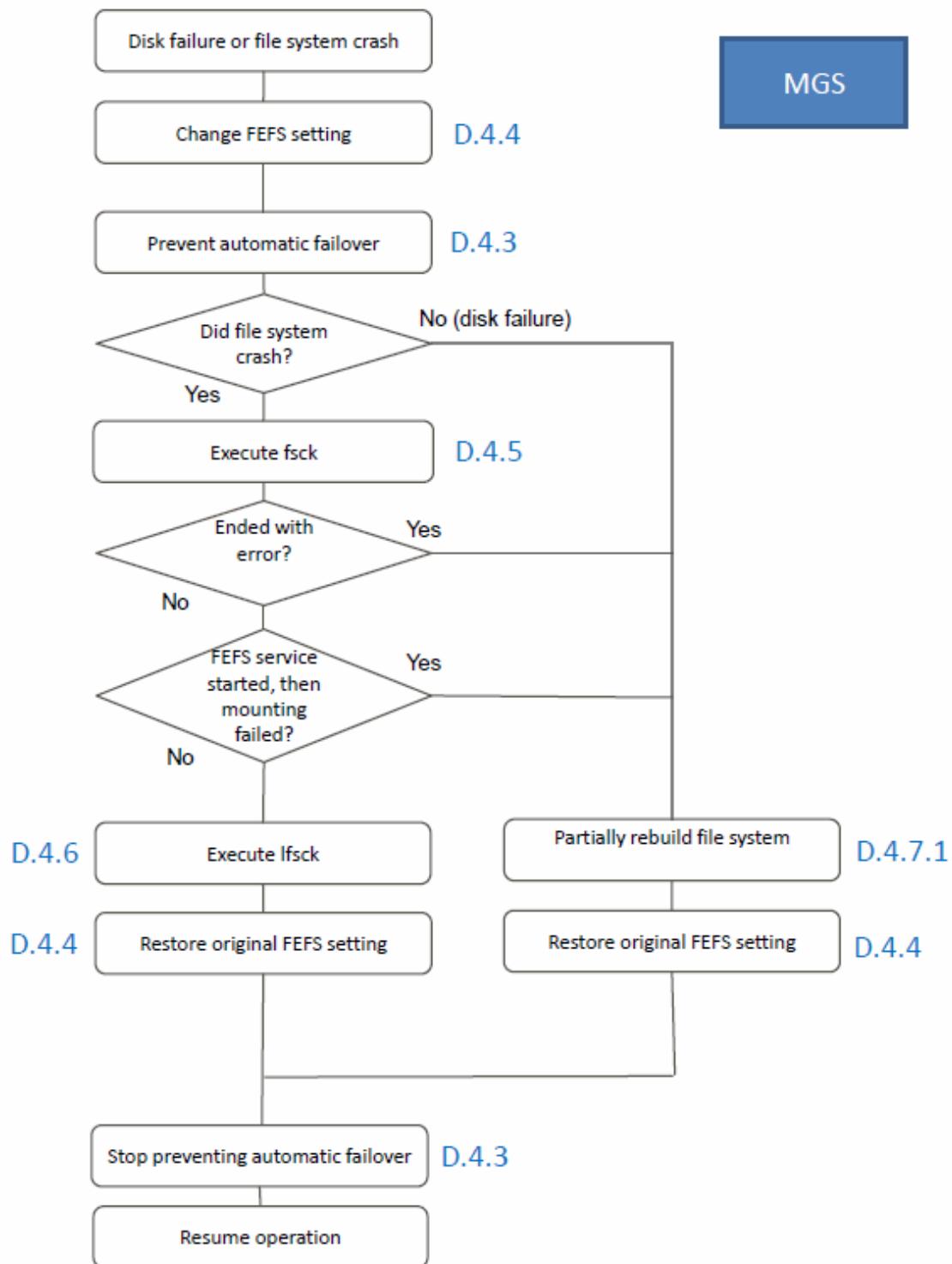


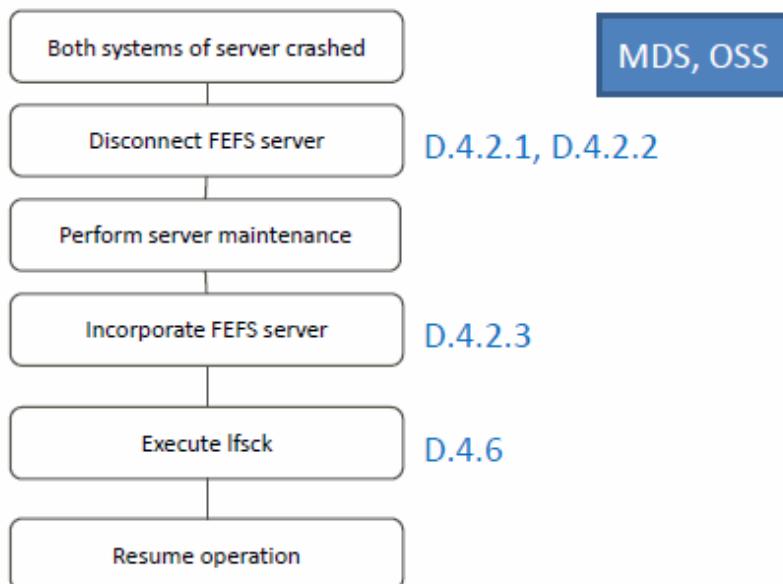
Figure D.5 MGT recovery flow after a disk failure or file system crash



D.3.3 Trouble handling flow when both systems stop

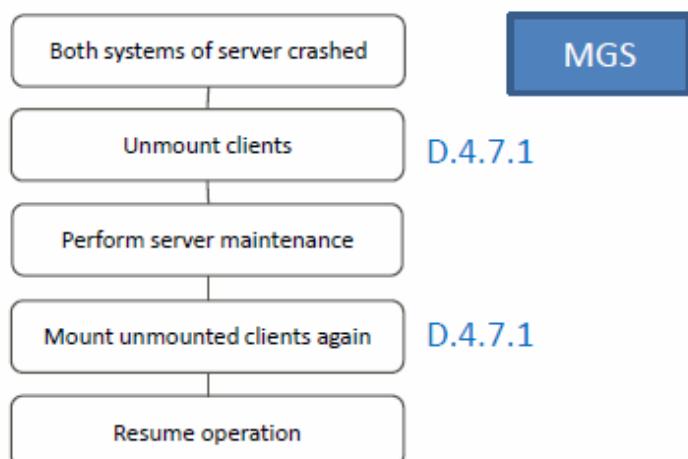
The following figure shows the trouble handling flow when both MDS and OSS systems stop.

Figure D.6 Trouble handling flow when both MDS/OSS systems stop (including cases with an MGS operating as an MDS too)



The following figure shows the trouble handling flow when both MGS systems stop.

Figure D.7 Trouble handling flow when both MGS systems stop



D.4 Response Procedures

D.4.1 Checking the status for a block number where a defective block occurred

If a defective block occurs, you need to identify the device and block number of this occurrence on the disk device.

The method of identification varies depending on the disk device, so check with the disk device vendor.

On the server using the device where the defective block occurred, use the debugfs.ldiskfs command in the following procedure to check the status.

If the icheck command outputs a value to Inode number, file data corruption has occurred.

If the testb command outputs "marked in use," destruction of the control domain has occurred.

```
# sync
# /opt/FJSVfefspogs/sbin/debugfs.ldiskfs <device> -R "icheck <block num>"
```

<device>: Device where a defective block occurred

<block num>: Block number where a defective block occurred

Execute the above command. If it displays "block not found," also execute the command below.

When displayed, "block not found" indicates damage to part of the file system.

```
# /opt/FJSVfefspogs/sbin/debugfs.ldiskfs <device> -R "testb <block num>"
```

Execute the above command. If "marked in use" appears as a result, it indicates that the specified block number is in use. Here, a block number in use indicates damage to the control domain of the file system.

If "not in use" appears as a result, it indicates that the specified block number is unused.

The following table outlines the information above.

Table D.1 Command execution results and file system status

Subcommand	Result	Status
icheck	Inode number displayed	File data corrupted
	"block not found" displayed	Damage to part of file system -> testb command
testb	"marked in use" displayed	Applicable block number in use, control domain damaged
	"not in use" displayed	Applicable block number unused

Examples of execution

1. File data corrupted (Checking block number 1544)

```
# sync
# /opt/FJSVfefspogs/sbin/debugfs.ldiskfs /dev/sdb -R "icheck 1544"
Block Inode number
1544 7
```

2. Control domain damaged (Checking block number 1545)

```
# sync
# /opt/FJSVfefspogs/sbin/debugfs.ldiskfs /dev/sdb -R "icheck 1545"
Block Inode number
1545 <block not found>
# /opt/FJSVfefspogs/sbin/debugfs.ldiskfs /dev/sdb -R "testb 1545"
Block 1545 marked in use
```

If that block is not used, the command displays the following.

```
# /opt/FJSVfefspogs/sbin/debugfs.ldiskfs /dev/sda7 -R "testb 10000"
debugfs 2.5.1 (01-Dec-2018)
Block 10000 not in use
```



The method of recovery varies depending on the place damaged by the defective block. Proceed with recovery according to the flow.

The following responses differ according to whether the target LBA address of the place where the defective block occurred is a data domain or the control domain, and whether the target LBA address is used or unused.

- Target LBA address is a data domain and unused

Clear the flag of the defective block. The method of clearing the flag varies depending on the disk device, so check with the disk device vendor.

- Target LBA address is a data domain and used
There are affected files. Identify the files.
 - Target LBA address is the control domain and unused
Clear the flag of the defective block. The method of clearing the flag varies depending on the disk device, so check with the disk device vendor.
 - Target LBA address is the control domain and used
File system operation is affected. Check the MGT, MDT, or OST recovery flow.
-

D.4.2 Disconnecting/incorporating an FEFS server

If both systems have crashed or a disk failure has occurred in an MDS or OSS failover pair, metadata access and file I/O operations hang until recovery. To resolve the hang, disconnect the FEFS server by performing the following procedures.

D.4.2.1 Confirming the target

Confirm the target to be disconnected.

File system crashed or disk failed

If a defective block caused damage to the control domain or a disk failure, check the device number of the target disk, the file system name, and the node type on the FEFS design sheet. The format is shown below.

```
<filesystem>-[MDT|OST] <index>
```

<filesystem>: File system name

<index>: Device number (4-digit hexadecimal number, zero padded)

- Example: Device number is 0001, file system name is fefs01, and node type is MDS

```
fefs01-MDT0001
```

Both systems of the server crashed

Use the following procedure to obtain the name of the target.

```
[Client node]
# lctl device_list -t | grep <IP addr> | awk '{print $4}' | cut -f 1,2 --delim="-"
```

<IP addr>: Specify respectively the IP address of the I/O network for the crashed server and the IP address of the failover pair.

- Example of execution

```
[Client node]
# lctl device_list -t | grep 172.31.211.60 | awk '{print $4}' | cut -f 1,2 --delim="-"
fef01-OST0000
# lctl device_list -t | grep 172.31.211.61 | awk '{print $4}' | cut -f 1,2 --delim="-"
fef01-OST0001
```

D.4.2.2 Disconnecting

1. Disconnecting

Use the following procedure to disconnect the intended target from the file system.

```
[System management node]
# pmexec -c <cluster> --stdout --nodetype SIO "/usr/sbin/force_intr -c -m deactivate <target>..."
# pmexec -c <cluster> --stdout --nodetype CCM,LN "/usr/sbin/force_intr -c -m deactivate
<target>..."
# pmexec -c <storage-cluster> --stdout --nodetype MDS,OSS "touch /var/opt/FJSVfefs/
stop_activate_device"
# pmexec -c <storage-cluster> --stdout --nodetype MDS "/usr/sbin/force_intr -s -m deactivate
<target>..."
```

<cluster>: Cluster name
<storage-cluster>: Storage cluster name
<target>: Intended target

- Example of execution

```
[System management node]
# pmexe -c compute --stdout --nodetype SIO "/usr/sbin/force_intr -c -m deactivate fefs-
OST0000 fefs-OST0001"
# pmexe -c compute --stdout --nodetype CCM,LN "/usr/sbin/force_intr -c -m deactivate fefs-
OST0000 fefs-OST0001"
# pmexe -c storage --stdout --nodetype MDS,OSS "touch /var/opt/FJSVfef/sstop_activate_device"
# pmexe -c storage --stdout --nodetype MDS "/usr/sbin/force_intr -s -m deactivate fefs-
OST0000 fefs-OST0001"
```



Note

- If the index number of the deactivated MDT (MDT0000) is 0, mounting is disabled. Consequently, the FEFS status will not be FEFS(o).
- If the index number of the deactivated MDT is other than 0, the df command cannot check the mount status. Check it with the lfs df command.
- If your system contains multiuse nodes, add NODETYPE defined in the FEFS design sheet as the argument to the --nodetype option. For details on NODETYPE, see "NODETYPE of multiuse nodes" in "[3.1.3.1 Filling in a Node sheet](#)."

2. Checking the status

After disconnecting the target, use the following procedure to check the status.

```
[System management node]
# pmexe -c <cluster> --stdout --nodetype SIO "/usr/sbin/force_intr -c -m status <target>..."
# pmexe -c <cluster> --stdout --nodetype CCM,LN "/usr/sbin/force_intr -c -m status <target>..."
# pmexe -c <storage-cluster> --stdout --nodetype MDS "/usr/sbin/force_intr -s -m status
<target>..."
```

<cluster>: Cluster name
<storage-cluster>: Storage cluster name
<target>: Intended target

Confirm that the status is "IN" for all targets.

- Example of execution

```
[System management node]
# pmexe -c compute --stdout --nodetype SIO "/usr/sbin/force_intr -c -m status fefs-OST0000
fefsf-OST0001"
# pmexe -c compute --stdout --nodetype CCM,LN "/usr/sbin/force_intr -c -m status fefs-OST0000
fefsf-OST0001"
# pmexe -c storage --stdout --nodetype MDS "/usr/sbin/force_intr -s -m status fefs-OST0000
fefsf-OST0001"
```

For details on output, see "[A.2.15 force_intr](#)".

D.4.2.3 Incorporating

1. Checking the status before incorporating

Before incorporating the target, use the following procedure to confirm that it is ready to be incorporated.

Incorporating an MDT

```
[System management node]
# pmexe -c <storage-cluster> --stdout --nodetype MDS "lctl get_param
mdt.<target>.recovery_status 2> /dev/null | grep status:"
```

<storage-cluster>: Storage cluster name

<target>: Intended target

Wait until status shows "COMPLETE" or "INACTIVE."

- Example of execution

```
[System management node]
# pmexe -c system4-storage --stdout --nodetype MDS "lctl get_param mdt.fefs-
MDT0000.recovery_status 2> /dev/null | grep status:"
[cmdline]
pmexe -c system4-storage --stdout --nodetype MDS lctl get_param mdt.fefs-
MDT0000.recovery_status 2> /dev/null | grep status:
[cluster]
system4-storage
<<<< ResultInformation >>>>
[0x00000005] status: COMPLETE
[0x00000006]
```

Incorporating an OST

```
[System management node]
# pmexe -c <storage-cluster> --stdout --nodetype OSS "lctl get_param
obdfilter.<target>.recovery_status 2> /dev/null | grep status:"
```

<storage-cluster>: Storage cluster name

<target>: Intended target

Wait until status shows "COMPLETE" or "INACTIVE."

- Example of execution

```
[System management node]
# pmexe -c system4-storage --stdout --nodetype OSS "lctl get_param obdfilter.fefs-
OST0000.recovery_status 2> /dev/null | grep status:"
[cmdline]
pmexe -c system4-storage --stdout --nodetype OSS lctl get_param obdfilter.fefs-
OST0000.recovery_status 2> /dev/null | grep status:
[cluster]
system4-storage
<<<< ResultInformation >>>>
[0x00000007] status: COMPLETE
[0x00000008]
```

2. Incorporating

Use the following procedure to incorporate the intended target into the file system.

```
[System management node]
# pmexe -c <storage-cluster> --stdout --nodetype MDS "/usr/sbin/force_intr -s -m activate
<target>..."
# pmexe -c <storage-cluster> --stdout --nodetype MDS,OSS "rm -f /var/opt/FJSVfefes/
stop_activate_device"
# pmexe -c <cluster> --stdout --nodetype SIO "/usr/sbin/force_intr -c -m activate <target>..."
# pmexe -c <cluster> --stdout --nodetype CCM,LN "/usr/sbin/force_intr -c -m activate <target>..."
```

<cluster>: Cluster name

<storage-cluster>: Storage cluster name

<target>: Intended target

- Example of execution

```
[System management node]
# pmexe -c storage --stdout --nodetype MDS "/usr/sbin/force_intr -s -m activate fefs-OST0000
fefes-OST0001"
# pmexe -c storage --stdout --nodetype MDS,OSS "rm -f /var/opt/FJSVfefes/stop_activate_device"
```

```
# pmexe -c compute --stdout --nodetype SIO "/usr/sbin/force_intr -c -m activate fefs-OST0000  
feffs-OST0001"  
# pmexe -c compute --stdout --nodetype CCM,LN "/usr/sbin/force_intr -c -m activate fefs-  
OST0000 fefs-OST0001"
```



If your system contains multiuse nodes, add NODETYPE defined in the FEFS design sheet as the argument to the --nodetype option. For details on NODETYPE, see "NODETYPE of multiuse nodes" in "[3.1.3.1 Filling in a Node sheet](#)."

3. Checking the status after incorporating

After incorporating the target, use the following procedure to check the status.

```
# pmexe -c <cluster> --stdout --nodetype SIO "/usr/sbin/force_intr -c -m status <target>..."  
# pmexe -c <cluster> --stdout --nodetype CCM,LN "/usr/sbin/force_intr -c -m status <target>..."  
# pmexe -c <storage-cluster> --stdout --nodetype MDS "/usr/sbin/force_intr -s -m status  
<target>..."
```

<cluster>: Cluster name

<storage-cluster>: Storage cluster name

<target>: Intended target

Confirm that the status is "UP" for all targets.

- Example of execution

```
# pmexe -c compute --stdout --nodetype SIO "/usr/sbin/force_intr -c -m status fefs-OST0000  
feffs-OST0001"  
# pmexe -c compute --stdout --nodetype CCM,LN "/usr/sbin/force_intr -c -m status fefs-OST0000  
feffs-OST0001"  
# pmexe -c storage --stdout --nodetype MDS "/usr/sbin/force_intr -s -m status fefs-OST0000  
feffs-OST0001"
```

For details on output, see "[A.2.15 force_intr](#)".

After completing the incorporation, execute the lfs df command on all client nodes.

```
[System management node]  
# pmexe -c <cluster> --stdout --nodetype SIO "/usr/bin/lfs df > /dev/null"  
# pmexe -c <cluster> --stdout --nodetype LN,CCM "lfs df > /dev/null"
```

<cluster>: Cluster name

D.4.3 Preventing automatic failover

See the *Job Operation Software Administrator's Guide for System Management*.

D.4.4 Changing the FEFS setting

The FEFS server enters the panic state upon detecting an MGT, MDT, or OST disk error.

As a workaround for the panic, change the mount setting to Read Only as shown below on all servers mounting the device where the disk error occurred, and restart the FEFS service.



- If the panic occurs immediately after startup, start in single-user mode.
- File access operations may temporarily hang when the service restarts.

Changing the setting

```
[FEFS server that detected a disk failure and the server that is paired for its failover]
# grep -l 'remount_ro' /etc/opt/FJSVfeffs/modprobe.conf | xargs sed -i.back -e 's/remount_ro=1/
remount_ro=0/g'
```

The execution of the above command changes the setting and saves the original file as /etc/opt/FJSVfeffs/modprobe.conf.back.
Restart the FEFS service to enable the setting.

Restoring the setting

```
[FEFS server that detected a disk failure and the server that is paired for its failover]
# mv /etc/opt/FJSVfeffs/modprobe.conf.back /etc/opt/FJSVfeffs/modprobe.conf
```

When restoring the setting, you also need to restart the service.

If a disk error is detected in the process of file I/O, for example, the setting for that disk changes to Read Only on the server. The lfs df command on a client will display "R" appended to the end of the respective entry for the disk. Also, a message indicating remounting in the Read Only state is output to /var/log/messages on the server.

```
[Client node]
# lfs df
UUID           1K-blocks   Used   Available Use% Mounted on
fefsf-MDT0000_UUID    5424664    40700   4897304  1% /fefsf[MDT:0]
fefsf-MDT0001_UUID    5424664    40324   4897680  1% /fefsf[MDT:1]
fefsf-OST0000_UUID   144009476   62256  136598996  0% /fefsf[OST:0]
fefsf-OST0001_UUID   144009476   62256  136598996  0% /fefsf[OST:1] R
filesystem_summary:   288018952   124512  273197992  0% /fefsf

[FEFS server that detected disk failure]
#less /var/log/messages
Jul 20 13:15:32 ossl kernel: LDISKFS-fs (sdb): Remounting filesystem read-only
```

With read-only mounting, any writing to the applicable OST causes an error.

For details on the impact on file access, see "[D.5 Impact on Access](#)."

D.4.5 Executing fsck

See the procedures from "[4.2 Stopping FEFS Servers and Clients](#)" to "[3\) Mounting OSTs](#)" in "[4.8.5 Repair of FEFS](#)," and perform repairs with the fsck.ldiskfs command.

If there is an error, recovery is not possible. Perform the operations in "[D.4.10 Rebuilding the file system](#)."

D.4.6 Executing lfsck

For the procedure for repairs with the lctl lfsck_start command, see "[4.8.5 Repair of FEFS](#)."

If there is an error, recovery is not possible. Perform the operations in "[D.4.10 Rebuilding the file system](#)."



Confirm that the forced I/O interrupt is canceled.

For the procedure, see "[3. Checking the status after incorporating](#)" in "[D.4.2.3 Incorporating](#)."

D.4.7 Partially rebuilding the file system

To perform a partial rebuild, prepare the FEFS design sheet used for the build.

D.4.7.1 Rebuilding an MGT

If an MGT fails, a client, MDT, or OST has not been successfully mounted.

Proceed with recovery by performing the following procedures.

Stopping the file system

```
[System management node]
# fefs_sync --stop --storage=<cluster> --compute=<cluster>
```

--storage: Specify a storage cluster name.
--compute: Specify a compute cluster name and a multiuse cluster name.



If multiple FEFS units are mounted, separately unmount the file system managing the MGT to be rebuilt.

Deleting FEFS client mount settings

In an environment with multiple file systems, if an FEFS client restart or other such operation starts a service while an MGT rebuild is in progress, mounting of that FEFS will fail. In this event, the FEFS status does not transition to FEFS(o).

To prevent this from happening, delete client information from the FEFS design sheet and temporarily delete it from the file system.

To continue operation using a different FEFS during the MGT rebuild, perform the following step.

To not mount the file system of the MGT being rebuilt, delete the nodes shown at "FX CLIENT" and "PG CLIENT" from the GFS design sheet defining the applicable MGS/MGT on the FEFS design sheet.

Distribute the configuration file to each client according to the build procedure. (See "[3.1.4 Creating configuration definition files for the FEFS setup tool](#)" and subsequent sections in "[3.1 Installation Flow](#)."

MGT rebuild

1. From the FEFS design sheet, confirm the applicable MGT device name and MGS.
2. Add "--replace" to MKFS OPTIONS in the MGS section of the GFS sheet defining this MGS/MGT on the FEFS design sheet, and re-create the configuration file.
3. When running with the Job Operation Software, place the MGS in maintenance mode.
See the *Job Operation Software Administrator's Guide for Maintenance*.
4. Save the file under /etc/opt/FJSVfefs/config on this MGS.
5. The re-created configuration file overwrites the file at /etc/opt/FJSVfefs/config on this MGS.
6. Execute the following commands on the MGS. If disk replacement is required, complete the replacement before performing this operation.

```
# fefsconfig --setup
# fefs_mkfs <volume>
```

<volume>: Volume name of the target MGT



If the MGT device name has changed, modify the FEFS design sheet, and distribute the settings again.

7. Clear the MDT or OST management files.
Execute the following commands on the MDT or OST of the rebuilt file system.

```
[MDS node, OSS node]
# export PATH="$PATH:/opt/FJSVfefsprogs/sbin"
# tunefs.lustre --writeconf <device>
```

<device>: Specify the targeted device.

8. Start the FEFS service.

Use the following procedure to start it.

```
[System management node]
# fefs_sync --start --storage=<cluster>[,...] --compute=<cluster>[,...]
```

--storage: Specify a storage cluster name.

--compute: Specify a compute cluster name and a multiuse cluster name.

9. When running with the Job Operation Software, cancel maintenance mode.

See the *Job Operation Software Administrator's Guide for Maintenance*.



Note

- If FEFS client mount settings have been deleted, undo the FEFS design sheet changes and redistribute the settings before starting the service.
- Since this procedure clears the OST pool settings, reconfigure the settings.

D.4.7.2 Partially rebuilding an MDT

An MDT cannot have a partial MKFS. A file system rebuild is required. See "D.4.10 Rebuilding the file system."

D.4.7.3 Partially rebuilding an OST

Advance preparation

Execute the following command for the OST to be partially rebuilt.

```
[System management node]
# pmexe -c <storage-cluster> --nodetype MDS --stdout "lctl set_param osp.<fsname>-OST<ostindex>-osc-MDT*.max_create_count=0"
```

<storage-cluster>: Storage cluster name

<fsname>: File system name

<ostindex>: OST index number (4-digit hexadecimal number, zero padded)

- Example of execution

```
[System management node]
# pmexe -c storage --nodetype MDS --stdout "lctl set_param osp.fefs01-OST0001-osc-MDT*.max_create_count=0"
```



If the storage cluster contains multiple file systems, narrow down the scope to the MDS that will build the OST to be partially rebuilt.

Partial rebuild

1. Confirm disconnection from operation.

See "2. Checking the status" in "D.4.2.2 Disconnecting."

2. From the FEFS design sheet, confirm the applicable OST device name and OSS.

3. Add "--replace" to MKFS OPTIONS in the OSS section of the GFS sheet defining this OSS/OST on the FEFS design sheet, and re-create the configuration file.

4. When running with the Job Operation Software, place the OSS in maintenance mode.

See the *Job Operation Software Administrator's Guide for Maintenance*.

5. Unmount this OST. Identify the mount point from the device name.

```
# umount -f <mount_point>
```

<mount_point>: File system mount point

6. Save the file under /etc/opt/FJSVfefs/config on this OSS.
7. The re-created configuration file overwrites the file at /etc/opt/FJSVfefs/config/ on this OSS.
8. Execute the following commands on this OSS. If disk replacement is required, complete the replacement before performing this operation.

```
# fefsconfig --setup  
# fefs_mkfs <volume>
```

<volume>: Volume name of the target OST

9. Execute the following command to restore the saved configuration file at /etc/opt/FJSVfefs/config/.

```
# fefsconfig --setup
```

10. Mount this OST.

```
# fefs_mount <mount_point>
```

<mount_point>: File system mount point

11. When running with the Job Operation Software, cancel maintenance mode.
See the *Job Operation Software Administrator's Guide for Maintenance*.
12. Incorporate all OSTs of the applicable OSS.
See "[D.4.2.3 Incorporating](#)."
13. Execute lfsck.
See "[4.8.5 Repair of FEFS](#)."



Note

- The execution of lfsck sets the QUOTA value again.
- The data saved on the faulty OST is not itself restored.

14. Restore backup files.
See "[D.4.9 Backing up files](#)."



Note

If the partial rebuild of the OST changed the OST device name, re-create and redistribute the FEFS design sheet.

Postprocessing

Execute the following command for the partially rebuilt OST.

```
[System management node]  
# pmexe -c <storage-cluster> --nodetype MDS --stdout "lctl set_param osp.<fsname>-OST<ostindex>-  
osc-MDT*.max_create_count=20000"
```

<storage-cluster>: Storage cluster name

<fsname>: File system name

<ostindex>: OST index number (4-digit hexadecimal number, zero padded)

D.4.8 Creating a list of backup files

Create a list of backup files in the following procedures.

MDT

The entire file system must be backed up. A list of backup files does not need to be created.

Perform the operations in "[D.4.9 Backing up files](#)."

OST

After a defective block occurs

From the number of the faulty block, identify the names of affected files on FEFS by performing the following procedure.

1. Perform the procedure in "[D.4.1 Checking the status for a block number where a defective block occurred](#)" to identify a inode number from the block number. If that block number is unused, no further work is necessary.
2. Specify an inode number, and execute the following command.

```
[OSS node]
# /opt/FJSVfefsporgs/sbin/fefs_ost2fid <Ost device> <inode number>
```

3. Execute the lfs fid2path command with the output FID.

The displayed path to a file using the faulty block is a relative path from the mount point.

```
[Client node]
# lfs fid2path <mount_point> <fid>
```

- Example of execution (for block number 1544)

```
[OSS node]
# sync
# /opt/FJSVfefsporgs/sbin/debugfs.ldiskfs /dev/sdb -R "icheck 1544"
Block   Inode number
1544    233
# /opt/FJSVfefsporgs/sbin/fefs_ost2fid /dev/sdb 233
233: [0x200000402:0x1:0x0]

[Client node]
# /usr/bin/lfs fid2path /fefs 0x200000402:0x1:0x0
file
```

The displayed path is a relative path from the mount point. Save this file. If an error such as EIO occurs while the file is being saved, the data cannot be read so it cannot be saved.

After an OST failure occurs

Identify the files on FEFS that use the faulty OST.

Use the following procedure to identify the files.

```
[MDS node]
# /opt/FJSVfefsporgs/sbin/find_file_ost -o <outfile> -d <mdt device> <ost index>
```

```
[Client node]
# /opt/FJSVfefsporgs/sbin/convert_fid2path -o <outfile> -m <mount_point> <infile ...>
```

- Example of execution

Repeat as many times as the number of MDTs.

/tmp/ost0000.out and /tmp/ost0000.err store lists of backup files.

```
[MDS node]
# /opt/FJSVfefsporgs/sbin/find_file_ost -o /tmp/mdt0000_ost0000 -d /dev/sdb 0
# /opt/FJSVfefsporgs/sbin/find_file_ost -o /tmp/mdt0001_ost0000 -d /dev/sdc 0
```

```

# scp /tmp/mdt0000_ost0000.out client:/tmp/mdt0000_ost0000.out
# scp /tmp/mdt0000_ost0000.err client:/tmp/mdt0000_ost0000.err
# scp /tmp/mdt0001_ost0000.out client:/tmp/mdt0001_ost0000.out
# scp /tmp/mdt0001_ost0000.err client:/tmp/mdt0001_ost0000.err

[Client node]
# /opt/FJSVfeffs/sbin/convert_fid2path -o /tmp/ost0000.out -m /feffs \
/tmp/mdt0000_ost0000.out /tmp/mdt0001_ost0000.out
# /opt/FJSVfeffs/sbin/convert_fid2path -o /tmp/ost0000.err -m /feffs \
/tmp/mdt0000_ost0000.err /tmp/mdt0001_ost0000.err
# wait

```

Delete temporary files.

```

[Client node]
# rm -f /tmp/mdt0000_ost0000.out /tmp/mdt0001_ost0000.out /tmp/mdt0000_ost0000.err /tmp/
mdt0001_ost0000.err

```

```

[MDS node]
# rm -f /tmp/mdt0000_ost0000.out /tmp/mdt0000_ost0000.err /tmp/mdt0001_ost0000.out /tmp/
mdt0001_ost0000.err

```

D.4.9 Backing up files

If the OST can be mounted, follow this procedure.

Execute the following using the list of affected files created in "D.4.8 Creating a list of backup files".

Incorporating OST on a login node

```

[System management node]
# pmexe -c <cluster> --stdout --nodetype LN "/usr/sbin/force_intr -c -m activate <target>..."
# pmexe -c <cluster> --stdout --nodetype LN "/usr/bin/lfs df > /dev/null"

```

<cluster>: Cluster name

<target>: Target

Backup

```

[Login node]
# /opt/FJSVfeffs/bin/fefsbackup copy -L ost0000_backup -n backup_node -d /backup_root \
-f /tmp/ost0000.out --ignore_err 2> /tmp/copy.err
# /opt/FJSVfeffs/bin/fefsbackup copy -L ost0000_backup -n backup_node -d /backup_root \
-f /tmp/ost0000.out --ignore_err 2>&1 | tee /tmp/copy.err
password: ***** *Enter ssh password
Copying request ost0000_backup is executing...
Total 2 files (2048 bytes) were copied from /tmp/ost0000.out file.
Copying request ost0000_backup is completed successfully.

# /opt/FJSVfeffs/bin/fefsbackup copy -L ost0000_backup2 -n backup_node -d /backup_root \
-f /tmp/ost0000.err --ignore_err 2> /tmp/copy2.err
password: ***** *Enter ssh password
Copying request ost0000_backup2 is executing...
Total 2 files (2048 bytes) were copied from /tmp/ost0000.err file.
Copying request ost0000_backup2 is completed successfully.

```

Check the following for files that could not be backed up. These files cannot be saved because there was an error during file access. It cannot be recovered.

```

# cat /tmp/copy.err
# cat /tmp/copy2.err

```

Disconnecting OST on a login node

```
[System management node]
# pmexec -c <cluster> --stdout --nodetype LN "/usr/sbin/force_intr -c -m deactivate <target>..."
```

<cluster>: Cluster name

<target>: Target

Restore

After recovery, return the saved data on the backup destination node as follows.:

```
[Client node]
# cd /backup_root
# /opt/FJSVfeffs/bin/fefsbackup copy -L ost0000_restore -n restore_node -d /feffs ./
```



If you have not saved the files, delete the files affected by the failed OST. An execution example when the created file list is mdt0000_ost0000.out is shown below.

```
[Client node]
# cat mdt0000_ost0000.out | xargs -n 1 unlink
```

D.4.10 Rebuilding the file system

After saving data, rebuild the file system.

Full backup

Before rebuilding, execute a full backup of the file system. The procedure is shown below.

Go to the mount point, then back up.

- Execution example

```
[Client node]
# cd /feffs
# /opt/FJSVfeffs/bin/fefsbackup copy -L fefs_backup -n backup_node -d /backup_root \
--ignore_err ./ 2> /tmp/copy.err
password: ***** *Enter ssh password
Copying request ost0000_backup is executing...
Total 2 files (2048 bytes) were copied from ./ directory.
Copying request fefs_backup is completed successfully.
```

Check the following for files that could not be backed up. These files cannot be saved because there was an error during file access. It cannot be recovered.

```
# cat /tmp/copy.err
```



- Check the capacity beforehand to examine the viability of backup.
- To restore a full backup result, do the following on the backup destination node.

```
[Client node]
# cd /backup_root
# /opt/FJSVfeffs/bin/fefsbackup copy -L fefs_restore -n restore_node -d /feffs ./
```

Rebuilding the file system

If disk replacement is required, replace the disk before proceeding with a rebuild. See "3.1.6 Building FEFS" to rebuild the file system.



Note

Rebuild the file system only after canceling the protection of data in the file system. For details, see "[4.19.2 Procedure for canceling the protection of data in a file system](#)."

D.5 Impact on Access

The following table shows the impact on file system access when a disk fails or both systems of a server crash.

Operation	OST			MDT		
	Read Only	Both systems failed Disk failure	Forced I/O interrupt	Read Only	Both systems failed Disk failure	Forced I/O interrupt
Open empty file	OK	Hang *1	Error *2	OK	Hang *1	Error *2
Create file	OK	Hang *1	OK	Error *1	Hang *1	Error *2
Delete file	OK	Hang *1	Error *2	Error *1	Hang *1	Error *2
read	OK	Hang *1	Error *2	OK	Hang *1	Error *2
write	Error *1	Hang *1	Error *2	OK	Hang *1	Error *2
Create/delete directory	OK	Hang *1	OK	Error *1	Hang *1	Error *2
stat (file)	OK	Hang *1	Error *2	OK	Hang *1	Error *2
stat (directory)	OK	OK	OK	OK	Hang *1	Error *2
statfs (df)	OK	OK	OK	OK	Hang	Error
lfs project	Error *1	Hang *1	Error *2	Error *1	Hang *1	Error *2
lfs setquota	OK	Hang	OK	OK	Hang	Error *2
lfs setstripe file	OK	Hang *1	OK	Error *1	Hang *1	Error *2
lfs setstripe directory	OK	OK	OK	Error *1	Hang *1	Error *2
lfs getstripe file/directory	OK	Hang *1	OK	OK	Hang	Error *2
lfs fid2path	OK	OK	OK	OK	Hang *1	Error *2
lfs find (search files on OST)	OK	OK	OK	OK	Hang	Error

*1 Aside from access to the two failed systems or the failed disk, files are normally accessible.

*2 In access to devices where I/O has not been forcibly interrupted, files are normally accessible.

Appendix E Procedures to Continue the Job Operation when File System Failure

This appendix describes procedures to continue the job operation for cases the file system failure has occurred.

Perform this procedure if you want to continue the job operation for cases the file system failure. If you do not want to continue job operation, see "[D.4.2 Disconnecting/incorporating an FEFS server](#)".

It describes the following procedures.

- Procedures to disconnecting/incorporating a file system in during operation
- Procedures to delete hanged jobs due to the file system failure
- Procedures to start node in during a file system failure

E.1 Disconnecting/Incorporating a File System

E.1.1 Disconnecting

By disconnecting the file system failure, the job operation can be continued on other file systems.

1. Disconnecting

Use the following procedure to disconnect the intended file system.

```
[System management node]
# pmexe -c <Cluster name> --stdout --nodetype CCM,LN "/usr/sbin/fefs_deactivate -m deactivate
<Intended mount point>"
```

- Example of execution

```
[System management node]
# pmexe -c compute --stdout --nodetype CCM,LN "/usr/sbin/fefs_deactivate -m deactivate /fefs"
```

2. Checking the status

After disconnecting the target, use the following procedure to check the status.

```
[System management node]
# pmexe -c <Cluster name> --stdout --nodetype CCM,LN "/usr/sbin/fefs_deactivate -m status
<Intended mount point>"
```

Confirm that the status is "IN" for intended file system.

- Example of execution

```
[System management node]
# pmexe -c compute --stdout --nodetype CCM,LN "/usr/sbin/fefs_deactivate -m status /fefs"
```

For details on output, see "[A.2.18 fefs_deactivate](#)".

E.1.2 Incorporating

The Job Operation remains continued, recovered the file system can be incorporated. About recovered the file system, see "[Appendix D Procedures to Recover the File System](#)".

1. Incorporating

Use the following procedure to incorporate the intended file system.

```
[System management node]
# pmexe -c <Cluster name> --stdout --nodetype CCM,LN "/usr/sbin/fefs_deactivate -m activate
<Intended mount point>"
```

- Example of execution

```
[System management node]
# pmexe -c compute --stdout --nodetype CCM,LN "/usr/sbin/fefs_deactivate -m activate /feefs"
```

2. Checking the status

After incorporating the target, use the following procedure to check the status.

```
[System management node]
# pmexe -c <Cluster name> --stdout --nodetype CCM,LN "/usr/sbin/fefs_deactivate -m status
<Intended mount point>"
```

Confirm that the status is "UP" for intended file system.

- Example of execution

```
[System management node]
# pmexe -c compute --stdout --nodetype CCM,LN "/usr/sbin/fefs_deactivate -m status /feefs"
```

For details on output, see "[A.2.18 fefs_deactivate](#)".

E.2 Procedures to Delete Hanged Jobs due to the File System Failure

The hanged jobs due to the file system failure, You can resolve the hang by the following procedure.

1. Disconnecting

"[E.1.1 Disconnecting](#)" is completed, resolve the hang by performing the following Disconnecting the node to which the hang job is assigned.

```
[System management node]
# pmexe -c <Cluster name> --stdout --bootgrp <Intended boot group range> "/usr/sbin/
feefs_deactivate -m deactivate <Intended mount point>"
```

- Example of execution

```
[System management node]
# pmexe -c compute --stdout --bootgrp 0x0101,0x0102 "/usr/sbin/fefs_deactivate -m deactivate /
feefs"
```

2. Checking the status

After disconnecting the target, use the following procedure to check the status.

```
[System management node]
# pmexe -c <Cluster name> --stdout --bootgrp <Intended boot group range> "/usr/sbin/
feefs_deactivate -m status <Intended mount point>"
```

Confirm that the status is "IN" for intended file system.

- Example of execution

```
[System management node]
# pmexe -c compute --stdout --bootgrp 0x0101,0x0102 "/usr/sbin/fefs_deactivate -m status /
feefs"
```

For details on output, see "[A.2.18 fefs_deactivate](#)".



- Disconnecting of the minimal unit size is BoB.
- When you perform this procedure, the job that uses the disconnected file system and the hanged job cause an error. The job ends with an error at PJM CODE 28 or 180.

E.3 Procedures to Start Node in during a File System Failure

If the file system fails while the client node is stopped, the failed file system must be disconnected in order for the FEFS service to start normally when the client node starts.

Node types of the client nodes are the following.

- Compute Node (CN)
- Compute Node and Storage I/O node (CN/SIO)
- Compute Node and Global I/O node (CN/GIO)
- Compute Node and Boot I/O node (CN/BIO)
- Login Node (LN)
- Compute Cluster Management node (CCM)

1. Starting the relevant node

Start the stopped node.

```
[System management node]
# papwrctl -c <Cluster name> -n <Node ID of target node> on
```

2. Disconnecting

The state of FEFS service remains FEFS(s). Use the following procedure to disconnect the intended file system.

```
[System management node]
# pmexe -c <Cluster name> --stdout -n <Node ID of target node> "/usr/sbin/fefs_deactivate -m deactivate <Intended mount point>"
```

- Example of execution

```
[System management node]
# pmexe -c compute --stdout -n 0x01010010 "/usr/sbin/fefs_deactivate -m deactivate /fefs"
```

3. Checking the status

After disconnecting the target, use the following procedure to check the status.

```
[System management node]
# pmexe -c <Cluster name> --stdout -n <Node ID of target node> "/usr/sbin/fefs_deactivate -m status <Intended mount point>"
```

Confirm that the status is "IN" for intended file system.

- Example of execution

```
[System management node]
# pmexe -c compute --stdout -n 0x01010010 "/usr/sbin/fefs_deactivate -m status /fefs"
```

For details on output, see "[A.2.18 fefs_deactivate](#)".

4. Checking the FEFS status

Confirm that the FEFS service has normally started using the pashowclst command.

```
[System management node]  
# pashowclst -c <Cluster name> -n <Node ID of target node >
```

If the FEFS status has changed to FEFSSR(o) and FEFS(o), the FEFS service has normally started.



- For nodes for which the disconnection procedure has been performed, the FEFS service will start normally when the node is restarted, even if this procedure is not performed.

Appendix F Materials Required for Troubleshooting

If a problem occurs, collect the following materials.

In the case of FX Server

Type of material to collect	Target node	Collected file/collection command
System log	All nodes	/var/log/messages*
PANIC DUMP	Node where dump was collected	/var/crash/OSdump-*
System material	All nodes	investigation data of OS collected by the pasnap command
FEFS material	All nodes	<outputdir>/fefssnap_<timestamp>.tgz created by the following executed command: # /usr/sbin/fefssnap -d <outputdir> timestamp: command execution time (yyyymmddHHMMSS) investigation data of FEFS collected by the pasnap command

In the case of PRIMERGY Server

Type of material to collect	Target node	Collected file/collection command
System log	All nodes	/var/log/messages*
PANIC DUMP	Node where dump was collected	/var/crash/127.0.0.1-XXX (for diskdump) * XXX is the dump collection date (year, month, day)
Kernel namelist	Node where dump was collected	/usr/lib/debug/lib/modules/version/vmlinux * version is the kernel version.
Kernel mapfile	Node where dump was collected	/boot/System.map-version * version is the kernel version.
System material	All nodes	investigation data of OS collected by the pasnap command
FEFS material	All nodes	<outputdir>/fefssnap_<timestamp>.tgz created by the following executed command: # /usr/sbin/fefssnap -d <outputdir> timestamp: command execution time (yyyymmddHHMMSS) investigation data of FEFS collected by the pasnap command



For details on collecting investigation data by the pasnap command, see the following manual.

Job Operation Software Administrator's Guide for System Management



- If LLIO is installed, the fefssnap command also collects data for LLIO.
- You may need an FEFS internal log, fefs.log, for troubleshooting. fefs.log may not exist, depending on the situation. To find out whether the fefs.log log has anything to output or whether log output has stopped, execute the ps command on the relevant node. You can collect a log when the following command is working.

```
lctl fefslog start /var/opt/FJSVfefsf/fs.log
```

Glossary

In addition to the terms below, see the *Job Operation Software Glossary* manual for terms related to the entire Technical Computing Suite.

ACL (Access Control List)

A list of individual users, their access privileges, and the file resources accessible to them

BoB (Bunch of Blade) [FX]

Control unit used in FX sever. A BoB is configured with 16 nodes.

by-id name

The device name generated from the unique identification information (e.g., serial number) assigned to a hard disk. The by-id name does not change unless the disk is replaced. This means that the same device name can always be used with the disk device.

compute node

A node on which jobs are executed. It is the smallest operation unit for computation.

external journal

A journal area, for the journaling function, created on an external device volume

failover

A function that, in the event of a server failure, enables an alternate server to take over processing and data from the faulty server

fallback

This function has the original server inherit processing and data back from the alternate server when resuming operation after the server failure was resolved.

FEFS (Fujitsu Exabyte File System)

A parallel distributed file system developed by Fujitsu

FEFS service monitoring daemon

A daemon process that monitors the FEFS service and acquires and reports the service status

FID

A management number for the inode inside FEFS

file block

A block or contiguous portion of file data

file data

The actual data as the file contents

HA configuration

A system configuration comprising a redundant server that reduces incidence of situation where the services can not be provided.

I/O node

A node connected to a disk and network device. For compute nodes, IO nodes provide file system functions and functions for networking outside the compute nodes.

inode

A data structure that stores file size, time stamp, UID, GID, and file block storage location information

internal journal

An internal journal area, for the journaling function, created within the MDT

LLIO (Lightweight Layered IO-Accelerator)

LLIO is a technology that makes high performance possible. LLIO establishes a storage layer using high-speed flash memory between FEFS and compute nodes, and uses the layer as the FEFS cache area and a temporary area for jobs.

LNet

A function for accessing file systems via multiple, different, commonly used networks such as Ethernet and InfiniBand

login node

A node on which users log in, and create or issue jobs.

MDS (Metadata Server)

A metadata server on which metadata is stored and managed

MDT (Metadata Target)

A logical volume on a disk device connected to the MDS. The MDT stores metadata.

metadata

Information other than file data. Such information includes the file size, timestamp, and owner.

MGS (Management Server)

A management server that manages the MDS, MDT, OSS, and OST configurations

MGT (Management Target)

A required disk unit for a MDS/MDT redundant configuration, which is an implementation of the Active/Active method

Multi-MDS

Configuration to which one file system is constructed with two MDS servers or more

multiuse node

A node used by the administrator for any purpose in Job Operation Software.

NID (Network Identifier)

The identifier of a network used by FEFS

object file

A file that stores the actual file data.

OSS (Object Storage Server)

A data server on which file data is stored and managed

OST (Object Storage Target)

A logical volume on a disk device connected to the OSS. The OST stores the actual file data.

OST_pool

A function for bundling multiple specified OSTs into a group and assigning files and directories to OSTs in the group

RDMA (Remote Direct Memory Access) communication

A communication method not involving the CPU to directly transfer data from the main memory of a node to the main memory of a remote node

Rolling update

Partial maintenance while continuing the job operation in the clusters of some compute nodes without stopping the system or the entire cluster when applying a package.

routing function

A function for routing between different networks

sparse file

A file that contains areas in the middle on which no data is written. The areas without data in a sparse file are not included in the disk usage counted against QUOTA because they do not consume any blocks on the OST.

Tofu interconnect D [FX]

Name of the Tofu interconnect in FX server. In this manual, for convenience, we simply refer to the Tofu interconnect.