FUJITSU

Fujitsu Software
Technical Computing Suite V4.0L20

FEFS ユーザーズガイド

J2UL-2554-01Z0(07)
2025年3月

まえがき

本書の目的

本書は、富士通が開発した共有ファイルシステムである FEFS の説明をしています。

本書は、システム管理者が FEFS の導入および運用管理ができることを目的としています。

本書の読者

本書は、FEFS の導入および運用管理を行うシステム管理者を対象としています。

本書を読むためには、Linux、ストレージ一般 および ETERNUS に関する知識が必要です。

本書の構成

本書は、以下に示す構成になっています。

第1章 概要

FEFS の概要および構成の説明をしています。

第2章 機能

FEFS の機能の説明をしています。

第3章 FEFS の導入と保守

FEFS の導入方法や保守について説明をしています。

第4章 運用方法

FEFS の運用方法の説明をしています。

付録A リファレンス

FEFS のシステムコール、コマンドのリファレンスマニュアルです。

付録B メッセージ

FEFS が出力するメッセージの説明をしています。

付録C FEFS の構築後に必要な設定

FEFS構築後に必要な設定の説明をしています。

付録D ファイルシステムの復旧手順

ファイルシステムの復旧手順について説明をしています。

付録E ファイルシステム故障発生時のジョブ運用継続手順

ファイルシステム故障時のジョブ運用継続手順について説明をしています。

付録F トラブル対処時に必要な資料

トラブル対処時に必要な資料の説明をしています。

用語集

FEFS における主な用語を説明しています。

本書の表記について

略称について

本書では、以下の略称を使用しています。

正式名称	略称
Windows(R) 8.1 Windows(R) 8.1 Pro Windows(R) 8.1 Enterprise	Windows 8.1、または Windows
Windows(R) 10 Home Windows(R) 10 Pro Windows(R) 10 Enterprise	Windows 10、または 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

単位の表現

本書では、単位を表現する際の接頭語は以下のとおりです。コマンドの表示や入力時の指定において注意してください。

接頭語	値	接頭語	値
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}

機種名の表現

本書では富士通製CPU A64FXを搭載した計算機を「FX サーバ」、FUJITSU server PRIMERGYを「PRIMERGY サーバ」(または単に「PRIMERGY」)と呼びます。
本書で説明する機能の一部には、対象機種によって仕様に差があります。このような機能の説明では、以下のように対象機種を略称で表記します。

[FX]: FX サーバを対象にした機能です。
[PG]: PRIMERGYサーバを対象にした機能です。

コマンド入力例におけるプロンプト

コマンド操作を行うために必要な管理者権限によって、プロンプトを区別しています。

- # は管理者権限 (スーパーユーザ) で実行することを意味します。
- $ は管理者権限以外で実行することを意味します。

マニュアル内のアイコンについて

本書では、以下のアイコンを使用しています。

注意

特に注意が必要な事項を説明しています。必ずお読みください。

参照

詳細な情報が書かれている参照先を示しています。

参考

FEFSに関連した参考記事を説明しています。

輸出管理規制について

本ドキュメントを輸出または第三者へ提供する場合は、お客様が居住する国および米国輸出管理関連法規等の規制をご確認のうえ、必要な手続きをおとりください。

商標

Lustreは米国 Seagate Technology LLC の登録商標です。
Linux®は米国及びその他の国におけるLinus Torvaldsの登録商標です。
Red Hat は米国およびその他の国において登録されたRed Hat, Inc. の商標です。
Microsoft、Windows、および Excelは、米国 Microsoft Corporation の、米国およびその他の国における登録商標または商標です。
そのほか、本書に記載されている会社名および製品名は、それぞれ各社の商標または登録商標です。

出版年月および版数

版数	マニュアルコード
2025年3月 第1.7版	J2UL-2554-01Z0(07)
2023年9月 第1.6版	J2UL-2554-01Z0(06)
2022年9月 第1.5版	J2UL-2554-01Z0(05)
2022年3月 第1.4版	J2UL-2554-01Z0(04)
2021年11月 第1.3版	J2UL-2554-01Z0(03)
2021年8月 第1.2版	J2UL-2554-01Z0(02)
2020年6月 第1.1版	J2UL-2554-01Z0(01)
2020年2月 初版	J2UL-2554-01Z0(00)

著作権表示

変更履歴

変更内容	変更箇所	版数
FEFS ログの定期削除の設定手順を追加しました。	C.4	第1.7版
FEFS クライアントの管理用ネットワークの IP アドレスを取得する手順を追加しました。	C.1 C.2	第1.6版
そのほか、誤記を修正しました。	-	
プロジェクト QUOTA 機能の注意事項を修正しました。	2.3.2	第1.5版
fefs_sync コマンドの --setup オプションの説明を修正しました。	A.2.1	第1.4版
サービス状態が変化した場合も evict スクリプトが動作するよう、計算クラスタの定義を追加しました。	C.1 C.2	第1.3版
ファイルシステム故障時のジョブ運用継続手順についての説明を追加しました。 また、fefs_deactivate コマンドのリファレンス、メッセージを追加しました。	付録E A.2.18 B.2.16	

変更内容	変更箇所	版数
ファイルシステムを追加する際の手順を修正しました。	4.16	第1.2版
データ管理ツールでのファイル転送時の注意事項を追加しました。	2.8	第1.1版
FEFSデザインシートがサポートする Windows のバージョンを以下に変更しました。 Windows 8.1、10	3.1.3 3.12.1	
手順1. LLIO SETTING セクションの指定方法について、説明を追加しました。 また、手順2のbとdに説明を追加しました。	3.1.3.2	
OS の更新パッケージ適用時の注意事項を「ジョブ運用ソフトウェア管理者向けガイド 保守編」の "第3章 " の "パッケージの削除" に移動したため、削除しました。	3.13	
データ管理ツールを使用する場合に実行するクライアントノードにインストールする必要があるパッケージを追加しました。	4.20.3	
fefsbackup.conf および fefsbackup_rsync.conf ファイルを設定する必要があるノードを明記しました。	4.20.4	
fefsbackup_rsync.conf の項目 MULTI_PUT_MAX の設定値に最大値を追加しました。	4.20.4.1	
データ管理ツールを実行するために必要な事前準備であるポート設定に関する説明を修正しました。 また、事前準備として、パスワードに関する説明を追加しました。	4.20.6	
copy サブコマンドのファイルシステム間のファイル転送手順について、以下を変更しました。 - 手順1 の説明を変更。 - 手順2 の表示メッセージを変更。 - 手順3 の実行例を修正。 - 手順1から4の実行例のオプション -d に指定するディレクトリのパス名を変更。 また、以下の注意事項を追加または変更しました。 - コピー時のファイル属性に関する注意事項を改善。 - ハードリンクファイルに関する注意事項を追加。 - シンボリックリンクファイルに関する注意事項を改善。 - オペランドの path または -f オプションの pathlist で指定するパス名に関する注意事項を追加。 - 転送処理がエラーになった場合のファイルパスに関する注意事項を追加。 - コマンドの同時実行に関する注意事項を追加。	4.20.8	
リクエストID 指定時の出力における表示パラメーター "Backed-up" の表示形式を修正しました。	4.20.9	
リクエストID 指定時の出力における表示パラメーター "Started" の表示形式を修正しました。	4.20.11	
fefsbackup コマンドのリファレンスについて、以下を追加または変更しました。 - コマンド実行権限に関する説明を追加。 - サブコマンド fefsbackup copy の機能説明を変更。 - オプション -L 指定時の注意事項を修正。 - オプション -u, --update *request_id* の説明を追加。 - オプション -v, --verbose の説明を修正。 - オプション --ignore_err の説明を改善。	A.2.11	
ファイルシステムの不整合を検出した際に出力されるシステムログメッセージを追加しました。	B.1	
fefsbackup コマンドのメッセージ 0002、0010、1035、1036 を追加しました。 メッセージ 1019 を変更しました。 メッセージ 0019、1022、1027 の対処を修正しました。	B.2.9	
"手順1 組込み前の状態確認" の「MDT を組込む場合」と「OST を組込む場合」のにおける組込み前の状態(state)について、説明を修正しました。 また、組込み完了後に全クライアントノードで lfs df コマンドを実行する手順を修正しました。	D.4.2.3	
ディスク異常を検知する対象ディスクに MGT を追加しました。 また、FEFSサーバのパニック回避するための設定変更または設定戻しを実施する対象ノードを変更しました。	D.4.4	
OST がマウントできている場合のファイルのバックアップ手順であることを明記しました。また、ログインノードで OST を組込みおよび切離し手順を追加し、かつ、バックアップ手順を修正しました。	D.4.9	

変更内容	変更箇所	版数
ファイルシステムのフルバックアップ手順の実行例を修正しました。	D.4.10	
そのほか、用語の統一、表記の揺れ、および誤記を修正しました。	-	

目　次

第1章 概要 1
1.1 FEFS の特長 1
1.2 FEFS のシステム構成 1
1.2.1 ハードウェア構成 1
1.2.2 サーバ構成 2
1.2.3 ネットワーク構成 6
1.3 FEFS のソフトウェア構成 6
1.4 上限値と下限値 7
1.5 注意事項 8

第2章 機能 10
2.1 ストライプ機能 10
2.1.1 ラウンドロビンとストライプ 10
2.1.2 ストライプ機能の効果 10
2.1.3 OST_pool 機能 11
2.2 マルチ MDS 機能 12
2.2.1 リモートディレクトリ 13
2.2.2 ストライプディレクトリ 13
2.3 QUOTA 機能 14
2.3.1 ユーザーまたはグループに対する QUOTA 機能 15
2.3.2 プロジェクト QUOTA 機能 16
2.4 QoS 機能 17
2.4.1 クライアント間の I/O 優先制御機能 18
2.4.2 ユーザー間フェアシェア機能 18
2.5 ACL (Access Control List) 機能 19
2.6 ジャーナリング機能 19
2.7 RAS機能/FEFS の状態確認機能 19
2.7.1 フェイルオーバ構成 19
2.7.2 FEFS サービス監視 21
2.7.3 FEFS の状態確認 21
2.7.4 サーバのフェイルオーバ (MGS、MDS、および OSS) 22
2.7.5 LNet マルチレール機能 22
2.7.6 LNet ルータ 23
2.8 データ管理ツール (fefsbackup コマンド) [PG] 23
2.9 FEFS 統計情報可視化機能 (fefssv.ph スクリプト) 24
2.10 FEFS以外のファイルシステムとの連携機能 25
2.10.1 外部システムへのNFSによる公開 26
2.10.2 Lustre 接続 [PG] 27

第3章 FEFS の導入と保守 28
3.1 導入の流れ 28
3.1.1 FEFS 構成の設計 29
3.1.2 FEFS パッケージの適用 33
3.1.3 FEFS デザインシートの作成 34
3.1.3.1 NODE シートの入力 34
3.1.3.2 LLIO シートの入力 37
3.1.3.3 GFS シートの入力 38
3.1.3.4 入力データのチェック 41
3.1.4 FEFSセットアップツール用構成定義ファイルの作成 42
3.1.5 FEFSセットアップツール用構成定義ファイルの配置 42
3.1.6 FEFSの構築 42
3.1.7 ファイルシステムのパーミッション変更 43
3.1.8 構築後に必要な設定 43
3.1.9 計算ノードの追加設定 43
3.1.10 ノード単位の構築方法 44
3.2 QoS 機能を有効にする設定 45

3.2.1 QoS 機能の有効化.....45
3.2.2 QoS 定義ファイルの設定.....47
3.3 ファイルロックを有効にする設定.....51
3.4 ACL 機能を有効にする設定.....51
3.5 user 拡張属性を有効にする設定.....51
3.6 フェイルオーバ機能を利用する場合の設定.....52
3.7 保守時の操作.....52
3.8 ローリングアップデート.....53
3.9 FEFS 統計情報可視化機能の設定.....54
3.10 NFS で公開する場合の設定.....55
3.11 構築に失敗したノードの構築方法.....55
3.12 外部ネットワークにおけるFEFSの構築方法.....55
3.12.1 外部ネットワーク用 FEFS デザインシートの作成.....56
3.12.1.1 NODE シートの入力.....57
3.12.1.2 NET シートの入力.....57
3.12.1.3 GFS シートの入力.....58
3.12.1.4 入力データのチェック.....59
3.12.2 FEFSセットアップツール用構成定義ファイルの作成.....59
3.12.3 FEFSセットアップツール用構成定義ファイルの配置.....59
3.12.4 外部ネットワークにおけるFEFSの構築.....59
3.12.4.1 FEFSサーバの設定およびルータの構築.....59
3.12.4.2 FEFSクライアントの設定.....60
3.13 注意事項.....61

第4章 運用方法.....63
4.1 FEFS サーバとクライアントの起動.....63
4.2 FEFS サーバとクライアントの停止.....63
4.3 ストライプ機能の設定.....63
4.3.1 ストライプの設定方法.....63
4.3.2 ストライプ設定の確認方法.....65
4.3.3 OST_pool の設定方法.....65
4.4 マルチ MDS の使い方.....68
4.4.1 リモートディレクトリの作成.....68
4.4.2 ストライプディレクトリ の作成.....69
4.5 QUOTA 機能の設定.....70
4.5.1 ユーザー・グループに対する QUOTA 設定.....70
4.5.2 プロジェクトに対する QUOTA 設定.....72
4.6 QoS機能の設定.....75
4.6.1 FEFSクライアントのQoS状態確認.....75
4.6.2 FEFSクライアントのQoS状態変更.....75
4.6.3 MDSのQoS状態確認.....76
4.6.4 MDS の QoS 定義の変更.....77
4.7 QoS機能のチューニング方法.....77
4.7.1 クライアントノード (メタ操作) の分析とチューニング.....78
4.7.2 クライアントノード (データ操作) の分析とチューニング.....79
4.7.3 MDS の分析とチューニング.....82
4.7.4 OSSの分析とチューニング.....84
4.8 ファイルシステム不整合の修復.....88
4.8.1 FEFS のサービス停止.....89
4.8.2 MGS 上での修復.....89
4.8.3 MDS 上での修復.....89
4.8.4 OSS上での修復.....89
4.8.5 FEFS の修復.....89
4.9 ACL の設定方法.....90
4.10 user 拡張属性の設定方法.....91
4.11 FEFS の状態確認.....92
4.12 フェイルオーバ.....93

4.13 MDS の追加 94
4.14 OSS の追加 95
4.15 クライアントの追加 96
4.16 ファイルシステムの追加 97
4.17 ファイルシステムの削除 98
4.18 ラック、BoB の追加 99
4.19 構築済みファイルシステムのデータの保護 100
4.19.1 ファイルシステムのデータを保護する手順 100
4.19.2 ファイルシステムのデータの保護を解除する手順 100
4.20 データ管理ツール (fefsbackup コマンド) の使い方 [PG] 101
4.20.1 サブコマンド概要 101
4.20.2 転送情報管理について 101
4.20.3 依存パッケージ 101
4.20.4 設定ファイル 102
4.20.4.1 設定ファイル詳細 102
4.20.5 管理情報の保管先の設計 103
4.20.6 事前準備 104
4.20.7 データ管理ツールの設定 104
4.20.8 copy サブコマンド 105
4.20.9 list サブコマンド 107
4.20.10 delete サブコマンド 108
4.20.11 status サブコマンド 108
4.21 JobStats機能 109
4.22 FEFS統計情報可視化機能 (fefssv.ph スクリプト) の利用方法 109
4.22.1 情報採取の方法 109
4.22.2 情報出力の方法 110
4.22.3 オプションと出力情報 113
4.23 Lustre 接続 [PG] 115
4.23.1 Lustre サーバと Lustre クライアントでの設定 115
4.23.2 Lustre クライアントから FEFS サーバのマウント 115
4.23.3 FEFS クライアントから Lustre サーバのマウント 115

付録A リファレンス 116
A.1 システムコール 116
A.2 コマンド 118
A.2.1 fefs_sync コマンド 118
A.2.2 fefsconfig コマンド 120
A.2.3 fefs_mkfs コマンド 121
A.2.4 fefs_mount コマンド 122
A.2.5 fefssnap コマンド 123
A.2.6 lfsコマンド 123
A.2.7 lctlコマンド 130
A.2.8 fsck.ldiskfs コマンド 145
A.2.9 tunefs.lustre コマンド 145
A.2.10 debugfs.ldiskfs コマンド 146
A.2.11 fefsbackup コマンド [PG] 147
A.2.12 fefs_ost2fid コマンド 150
A.2.13 find_file_ost コマンド 151
A.2.14 convert_fid2path コマンド 151
A.2.15 force_intr コマンド 152
A.2.16 evict_client コマンド 153
A.2.17 fefs_yaml2csv コマンド 154
A.2.18 fefs_deactivate コマンド 154

付録B メッセージ 156
B.1 システムログに出力されるメッセージ 156
B.2 コマンドの出力するメッセージ 167
B.2.1 fefs_sync コマンド 167

B.2.2 fefsconfig コマンド ... 170
B.2.3 fefs_mkfs コマンド ... 174
B.2.4 fefs_mount コマンド ... 176
B.2.5 fefssnap コマンド ... 178
B.2.6 lfs コマンド ... 179
B.2.7 lctl コマンド ... 192
B.2.8 fsck.ldiskfs コマンド ... 204
B.2.9 fefsbackup コマンド [PG] ... 205
B.2.10 ファイル特定ツール共通 ... 212
B.2.11 find_file_ost コマンド ... 216
B.2.12 convert_fid2path コマンド ... 216
B.2.13 force_intr コマンド ... 216
B.2.14 evict_client コマンド ... 217
B.2.15 fefs_yaml2csv コマンド ... 218
B.2.16 fefs_deactivate コマンド ... 219

付録C FEFS の構築後に必要な設定 ... 221
C.1 FEFS スクリプトの設定 ... 221
C.2 複数システム管理ノード環境での FEFSスクリプトの設定 ... 224
C.3 ETERNUS を利用する場合に必要な設定 ... 227
C.3.1 MDS で ETERNUS の NRDY 対策の有効化 ... 227
C.3.2 OSS の自動起動スクリプト設定手順 ... 228
C.4 FEFS ログの定期削除の設定 ... 232

付録D ファイルシステムの復旧手順 ... 233
D.1 はじめに ... 233
D.2 影響 ... 233
D.3 障害復旧フロー ... 233
D.3.1 不良ブロック検出時の復旧フロー ... 233
D.3.2 ディスク故障時またはファイルシステム破壊時の復旧フロー ... 235
D.3.3 両系停止時の障害対応フロー ... 238
D.4 対応手順 ... 239
D.4.1 不良ブロックが発生したブロック番号の状態確認 ... 239
D.4.2 FEFS サーバの切離し/組込み ... 241
D.4.3 自動フェイルオーバ抑止 ... 244
D.4.4 FEFS 設定変更 ... 244
D.4.5 fsckの実施 ... 245
D.4.6 lfsckの実施 ... 245
D.4.7 ファイルシステムの部分再構築 ... 245
D.4.8 バックアップファイル一覧の作成 ... 248
D.4.9 ファイルのバックアップ ... 250
D.4.10 ファイルシステムの再構築 ... 251
D.5 アクセス影響 ... 251

付録E ファイルシステム故障発生時のジョブ運用継続手順 ... 253
E.1 ファイルシステムの切離し/組込み手順 ... 253
E.1.1 切離し ... 253
E.1.2 組込み ... 253
E.2 ファイルシステム故障の影響でハングアップしたジョブの刈り取り手順 ... 254
E.3 ファイルシステム故障中のノード起動手順 ... 255

付録F トラブル対処時に必要な資料 ... 256

用語集 ... 258

第1章 概要

ここでは、FEFS の特長、構成および仕様を説明します。

1.1 FEFS の特長

FEFS は、オープンソースのファイルシステムである Lustre の技術に基づいた大規模および高性能な並列分散ファイルシステムです。以下の特長を持っています。

- 大規模
  10万ノードのクライアント、8EiBのファイルシステムサイズをサポートしています。
- 高性能
  ストライプ、ラウンドロビンの手法を使用して、ファイルデータをストレージに分散格納することによって、I/O性能を向上させています。
- 使いやすさ
  クライアント間の I/O 優先制御/ユーザー間フェアシェア(QoS機能)などを通じて、大量の I/O を行う他ユーザーの影響を抑止しています。
- 高信頼
  MGS (Management Server)、MDS (Meta Data Server)、およびOSS (Object Storage Server) のフェイルオーバ機能を持っています。
- 拡張性
  メタデータ領域/データ格納領域の動的な拡張が可能です。

1.2 FEFS のシステム構成

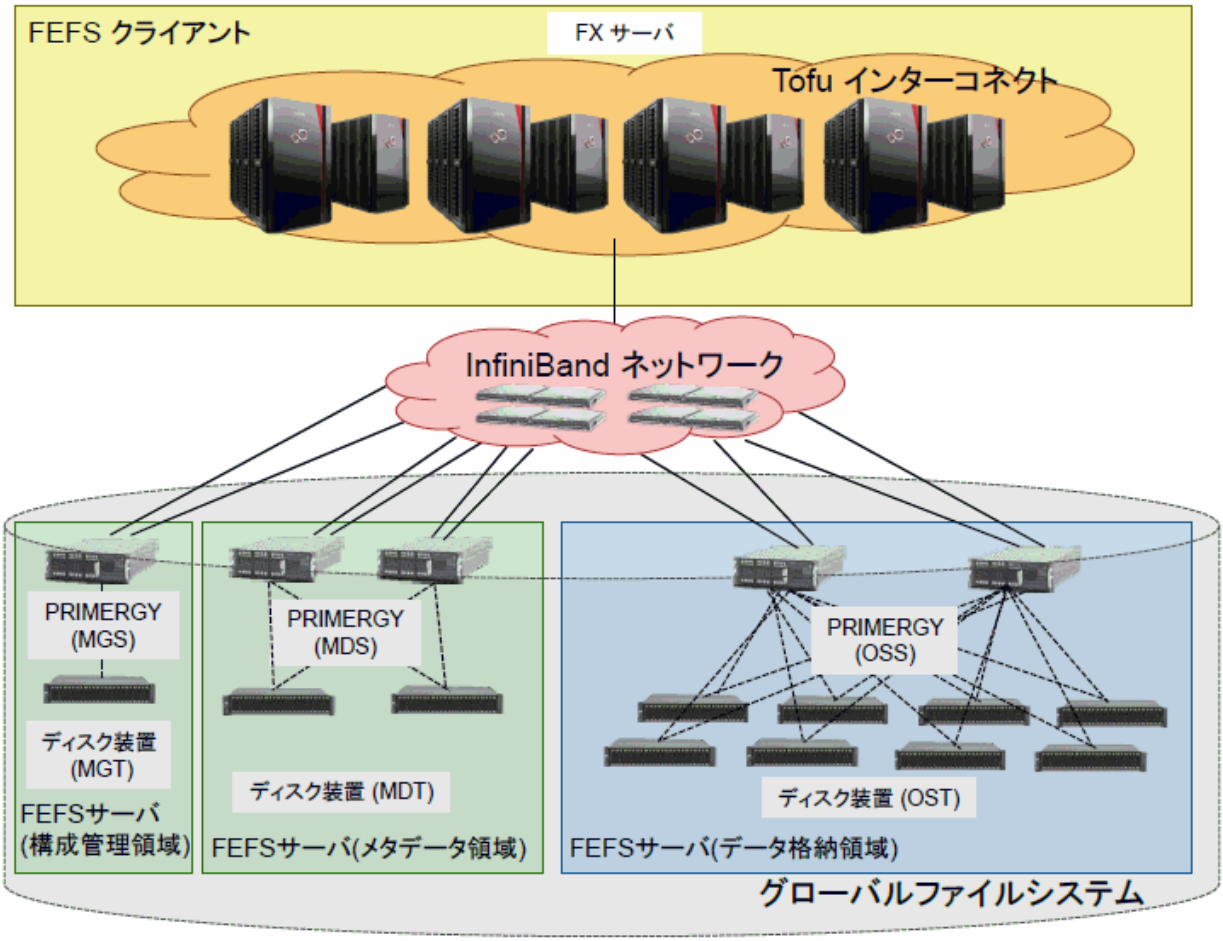
以下に、FEFS のシステム構成を示します。FEFS のクライアント構成、サーバ構成およびネットワーク構成を説明します。

1.2.1 ハードウェア構成

Tofuインターコネクトを導入したハードウェア構成を"図1.1 ハードウェア構成"に示します。

本構成では、計算ノードは FX サーバとなります。また、計算ノード群を接続するネットワークはTofuインターコネクトで構成されます。サーバ・クライアント間のネットワークはInfiniBandで構成されます。

図1.1 ハードウェア構成



FEFS クライアントとして利用できるノード種別は次のとおりです。

- 計算ノード (CN)
- 計算ノード兼ストレージ I/O ノード (CN/SIO)
- 計算ノード兼グローバル I/O ノード (CN/GIO)
- 計算ノード兼ブート I/O ノード (CN/BIO)
- ログインノード (LN)
- 計算クラスタ管理ノード (CCM)
- 多目的ノード

ノード種別についての詳細は、「ジョブ運用ソフトウェア 概説書」を参照してください。

1.2.2 サーバ構成

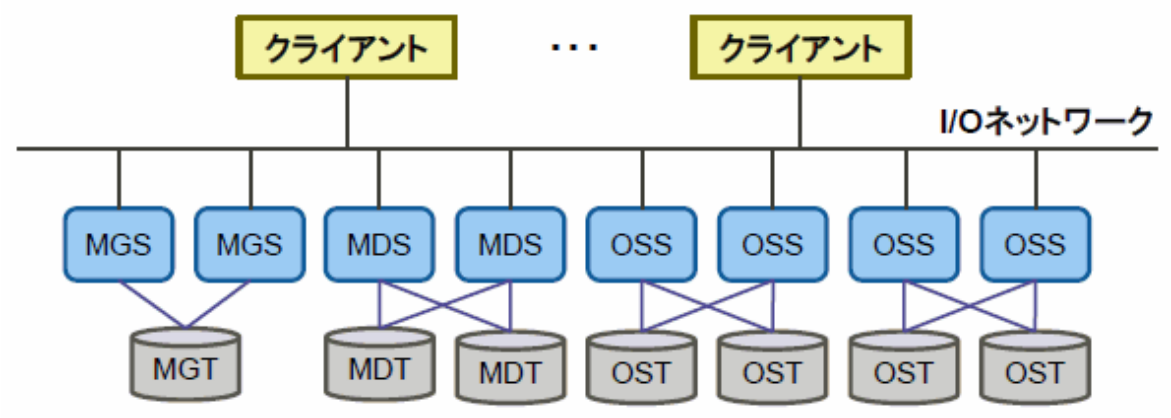
FEFS は、以下のサーバ/ディスク装置で構成されています。

- 管理サーバ (Management Server、以降 MGS と表記)
  MDS、MDT、OSS、および OST の構成を管理するサーバです。
- 管理ボリューム (Management Target、以降 MGT と表記)
  ファイルシステムの構成情報を格納するディスク装置です。
- メタデータサーバ (Metadata Server、以降 MDS と表記)
  メタデータ (inode、ストライプ情報、ディレクトリエントリ)を管理するサーバです。

- メタデータ論理ボリューム (Metadata Target、以降 MDT と表記)
  メタデータを格納するディスク装置です。
- オブジェクトストレージサーバ (Object Storage Server、以降 OSS と表記)
  ファイルデータ領域を管理するサーバです。
- オブジェクトストレージ論理ボリューム (Object Storage Target、以降 OST と表記)
  ファイルデータの実体を格納するディスク装置です。

以下は FEFS のサーバ構成の例です。

図1.2 サーバ構成 (MGS と MDS が異なるマシンで構築されている場合)



MGS/MGT構成

MGS は Active/Standby 方式の冗長構成をサポートしています。MGS は、MDS と同じマシン上で動作させることが可能ですが、独立して管理できるように、MGS と MDS をそれぞれ別のマシン上に構築することを推奨します。MDS と MGS を別サーバにする場合、または MDS の冗長構成を Active/Active 方式とする場合には独立したMGTが必要です。

MGT のボリューム構成

MGT は信頼性が求められるため、RAID1 (ミラーリング) 構成を推奨します。

MGT に必要なボリュームサイズは、以下の計算式で求められます。

```
200MiB × ファイルシステム数
```

MDS/MDT 構成

FEFS は MDS によりクライアントに対して統一したファイルツリー空間を提示し、アプリケーションからのファイルオープン要求を受けて集中管理することで、統一したファイルビューと排他制御を実現します。

FEFS は、MDS/MDT のペアによりファイルシステムのメタデータを管理します。MDS/MDT のペアを増やすことで、メタデータへのアクセスの負荷を分散できます。詳細は、"2.2 マルチ MDS 機能" を参照してください。

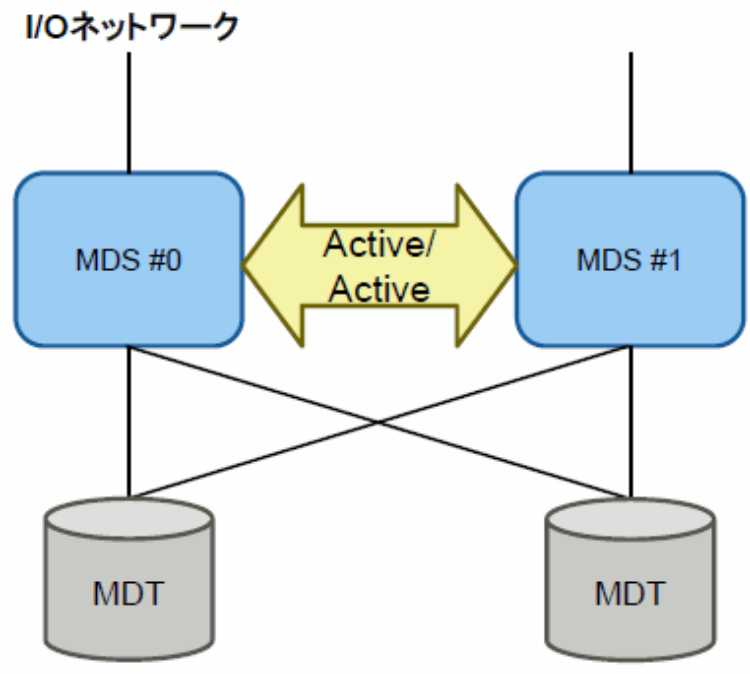
MDS は Active/Standby 方式とActive/Active 方式の両方の冗長構成をサポートしています。これにより、一方のノードが故障した場合は他方のノードに切り替えて、運用を継続できます。

なお、Active/Active 方式による冗長構成にする場合は、独立した MGT が必要となります。

冗長構成の詳細やノードの切り替えについては、"2.7 RAS機能/FEFS の状態確認機能" を参照してください。

MDS と MDT の構成例を以下に示します。

図1.3 MDS/MDT 構成



MDT のボリューム構成

MDT は、高速性と信頼性が求められるため、RAID10 (ミラーリング + ストライピング) 構成を推奨します。

MDT のボリュームサイズは、100GiB 以上を推奨します。

MDT に必要なボリュームサイズは、ファイルシステムに格納するファイル数 (inode 数) に応じて、以下の計算式で求めることができます。

```
(inode数 ＋ ACLまたはストライプを設定するファイル数) × 4KiB
 ＋ ACL領域 (※1)
 ＋ ストライプ領域 (※2)
 ＋ user拡張属性領域 (※3)

※1 ACL(Access Control List)を設定する場合、以下の計算式で求められるボリュームサイズが必要になります。
      ACLを設定するファイル数 × 1ファイルあたりの ACL エントリ数 × 8Byte

※2 ファイルにストライプを設定する場合、以下の計算式で求められるボリュームサイズが必要になります。
      ストライプを設定するファイル数 × 1ファイル あたりのストライプ数 × 24Byte

※3 user拡張属性を設定する場合、以下の計算式で求められるボリュームサイズが必要になります。
      user拡張属性を設定するファイル数 × 1ファイルあたりの user拡張属性サイズ(Byte)
```

注意

MDT 1個のボリュームサイズは、最大で 8TiB - 2KiB です。
上記の計算式で MDT に必要なボリュームサイズが、8TiB - 2KiB を超える場合は、以下の構成を検討してください。

- 複数ファイルシステム構成
- マルチ MDS 構成

システムの安定運用のため、容量不足が発生しないように、運用中は容量監視の対処をしてください。

参照

ACL 機能、ストライプ 機能、および user 拡張属性の詳細は、後述の機能説明や運用方法を参照してください。

OSS/OST 構成

FEFS は、OSS/OST によりファイルシステムの実データを管理します。

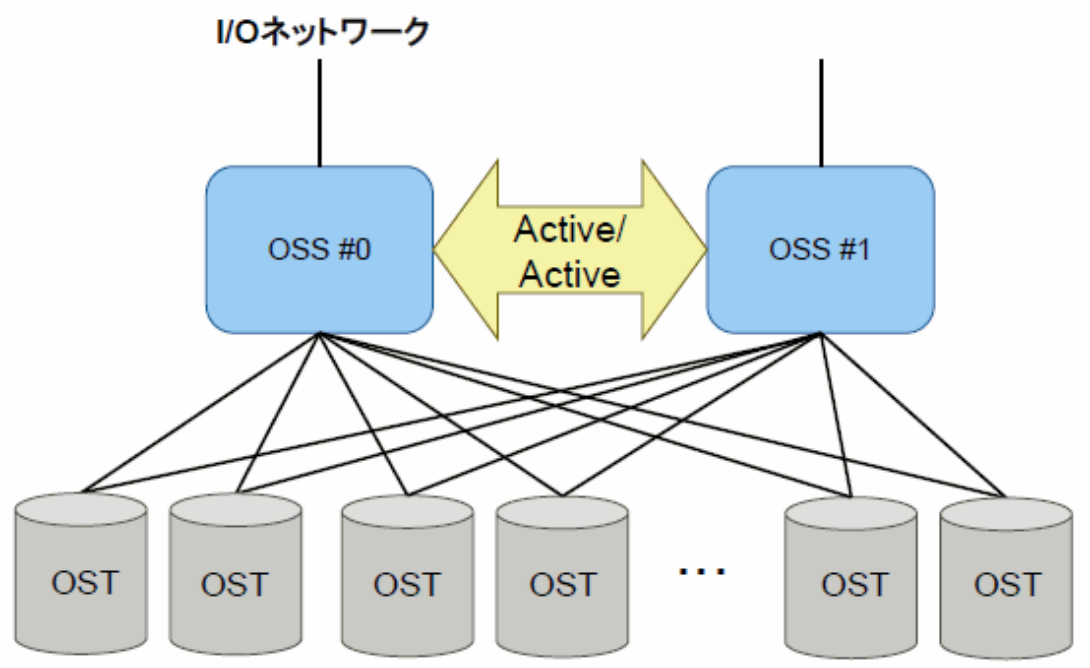
OSS は、クライアントが write したファイルデータをオブジェクトストレージ論理ボリューム (OST) に格納し、クライアントが read したファイルデータを OST から読み出してクライアントに転送します。OSS/OST のペアを増やすことでファイルシステムの容量とスループット性能を増やすことができます。

OSS は Active/Active 方式の冗長構成をサポートしています。これにより、ノードが故障した場合は、一方のノードに切り替えて、運用を継続できます。

冗長構成の詳細やノードの切り替えについては、"2.7 RAS機能/FEFS の状態確認機能" を参照してください。

OSS と OST の構成例を以下に示します。

図1.4 OSS/OST の構成



OST のボリューム構成

OST は容量と信頼性を両立するため RAID6 (ダブルパリティ) 構成を推奨します。

OST のボリュームサイズは、10GiB 以上を推奨します。

OST のボリュームサイズにより、利用可能な inode 数は以下のようになります。

```
OSTサイズ (※1) ÷ inodeレシオ (※2) = 利用可能なinode数
※1 OST 1個のボリュームサイズです。この値は、システム全体での合計値ではありません。
※2 inode レシオとは、OST のフォーマット時に、inode 1個あたりに割り当てるディスクサイズです。inode レシオは、OST のサイズによって異なります。OST サイズに応じた値は、以下のとおりです。

   OSTサイズ        inodeレシオ
 10GiB未満           16KiB
 10GiBから1TiB未満   68KiB
 1TiBから4TiB未満   256KiB
 4TiBから16TiB未満  512KiB
 16TiB以上            1MiB

 計算例1: OST のサイズが 300GiB の場合に利用可能な inode数
          300GiB ÷ 68KiB = 約4.6M個

 計算例2: OST のサイズが 2TiBの場合に利用可能な inode数
          2TiB ÷ 256KiB = 約8.3M個
```

計算例3: OST を 10個で構築するファイルシステムで、各 OST サイズが 2TiB の場合に利用可能な inode 数
(2TiB ÷ 256KiB) × 10 = 約83M個

OST のサイズに関係なく、inode レシオを指定したい場合は、FEFSデザインシート「GFS シート」の MKFS OPTION の OST OPTION に以下を指定してください。

inode レシオ	GFSシートの MKFS OPTION の OST OPTION
16KiB の場合	--mkfsoptions="-i 16384"
68KiB の場合	--mkfsoptions="-i 69905"
256KiB の場合	--mkfsoptions="-i 262144"
512KiB の場合	--mkfsoptions="-i 524288"
1MiB の場合	--mkfsoptions="-i 1048576"

注意

inode レシオには上記以外の値は指定しないでください。

1.2.3 ネットワーク構成

FEFS は、専用のネットワークドライバを使用して、以下をサポートします。

- Tofu インターコネクト[FX]
  - Tofu インターコネクトについての詳細は、「ジョブ運用ソフトウェア 概説書」を参照してください。
- InfiniBand (OFED)
  - RDMA (Remote Direct Memory Access) 通信を使用して、大規模データの高速な転送性能を実現します。
  - 複数の HCA カードを搭載した場合は、以下を実現します。
    - ラウンドロビンによる高バンド幅の転送
    - InfiniBand 故障における、有効な経路の選択およびファイルI/O の継続

FEFSでは、上記にある複数のネットワークを中継する機能を LNet (Lustre Networking) と呼ばれる通信層で実現しています。

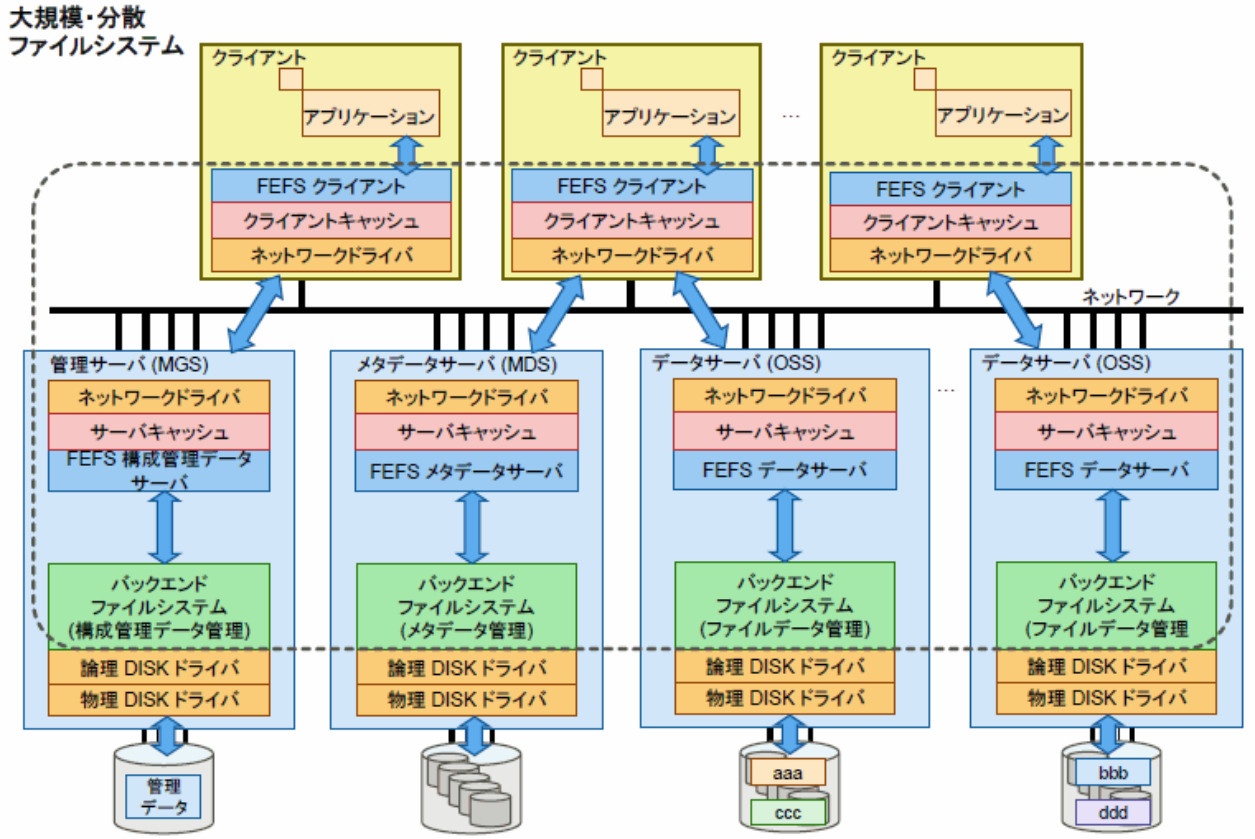
また、以下のように通信機能を使い分けています。

- 高速なデータ転送のため、カーネル・ワンサイド通信である RDMA の機能を使用します。
- そのほかの通信は、システムパケット (Send/Receive) の機能を使用します。

1.3 FEFS のソフトウェア構成

以下に、FEFS のソフトウェア構成を示します。

図1.5 ソフトウェア構成



1.4 上限値と下限値

FEFS の上限値および下限値を以下に示します。

表1.1 FEFS の上限値および下限値

項目	上限値または下限値
最大ファイルシステムサイズ	8EiB
最大ファイルサイズ	62.5PiB
最大ファイル数	16Ti個
単一ディレクトリ内の最大サブディレクトリおよびファイル数	10Mi個
最大メタボリューム (MDT) 数	4096個
最大メタボリューム (MDT) サイズ	8TiB - 2KiB
最小メタボリューム (MDT) サイズ	32MiB
最大ボリューム (OST) 数	8150個
最大ボリューム (OST) サイズ	2PiB
最大クライアント数	1Mi台
最大ブロックサイズ	4KiB
最大ストライプ数	4000個
最小ストライプサイズ	64KiB
最大ストライプサイズ	4194240KiB (4GiB-64KiB)
最大ファイル名の最大長	255

項目	上限値または下限値
パス名の最大長	4096 (終端の null 文字を含む)
QUOTA 最大ファイル数	8Ei個
QUOTA 最小ファイル数	1024 × MDT 数 (個)
QUOTA 最大ディスク容量	8EiB
QUOTA 最小ディスク容量	1024 × OST 数 (KiB)
ACL の最大エントリ数	32
ファイルシステム名の最大長	8
OST_pool名の最大長	15

1.5 注意事項

ここではFEFS を利用する際に留意すべき点について説明します。

使用用途

FEFS は、アプリケーションの作業領域および /home 配下の領域に使用できます。FEFS は、オペレーティングシステム自身が使用する / (ルートディレクトリ)、/var および /usr 配下の領域などには使用できません。

マウントポイント

同一の FEFS を同一 FEFS クライアントにおいて複数のマウントポイントへマウントするような使い方はしないでください。

同一ノード上での複数機能のサポート

同一ノード上で以下の兼用構成が可能です。

- MGS 兼 MDS 兼 OSS
- MGS 兼 MDS

サーバとクライアントを同一ノード上で動作させることは非サポートです。

注意

MDS と OSS を同一ノード上で動作させた場合、以下の構成は非サポートです。

- 複数ファイルシステム構成
- サーバの冗長構成

時刻の同期

ノード間で時刻がずれていると、運用に支障を来す場合があります。NTP などを使用してノード間の時刻を合わせてください。

ユーザー認証

MDS とクライアントでユーザー名/グループ名に対する uid/gid が異なると、ファイルシステムが正しく動作しません。FEFS を利用するユーザー/グループのuid/gid が、MDS とクライアントで同じになるように設定してください。

ファイルアクセス時に EIDRM エラー (Identifier removed: 識別子は削除されました) が発生する場合は、ユーザー認証が正常に行われていない可能性があります。LDAP によるユーザー認証を行っている場合は、LDAP サーバとの通信が正常に行われているかを確認してください。

多目的ノードで FEFS を利用する場合の注意事項

Technical Computing Suite では任意の用途に使える多目的ノードと多目的クラスタがあります。
ストレージクラスタの多目的ノードは、FEFSクライアント機能は利用できません。

多目的ノードと多目的クラスタについては、「ジョブ運用ソフトウェア 概説書」を参照してください。

FEFS 領域へのアクセスについて

FEFS 領域に対して ls -al のような stat 呼出しを伴うアクセスを行うと、処理が終了するまでの時間が数十秒かかる場合があります。

第2章 機能

ここでは、FEFS の機能を説明します。

2.1 ストライプ機能

2.1.1 ラウンドロビンとストライプ

FEFS では、ファイルのデータを OSS を介して OST に分散して格納します。

分散方式として、2つの方式を選択できます。

- ラウンドロビン方式

  ファイル単位で OST をラウンドロビンで選択して格納します。

  最大ファイルサイズは、物理的な OST の容量となります。

  デフォルトの分散方式です。

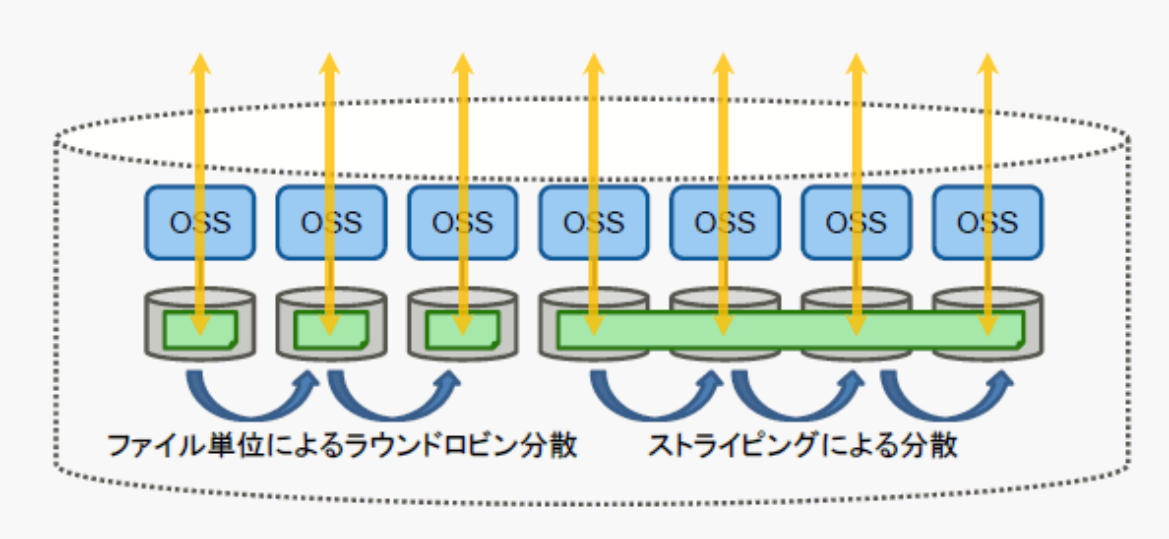
- ストライプ方式

  1つのファイルのデータを OST に分散させて格納します。

  ファイルサイズは、物理的な 1つの OST の容量を上回ることができます。

  この方式を使用するには、設定が必要です。設定方法については、"4.3 ストライプ機能の設定"を参照してください。

図2.1 ラウンドロビンとストライプ



2.1.2 ストライプ機能の効果

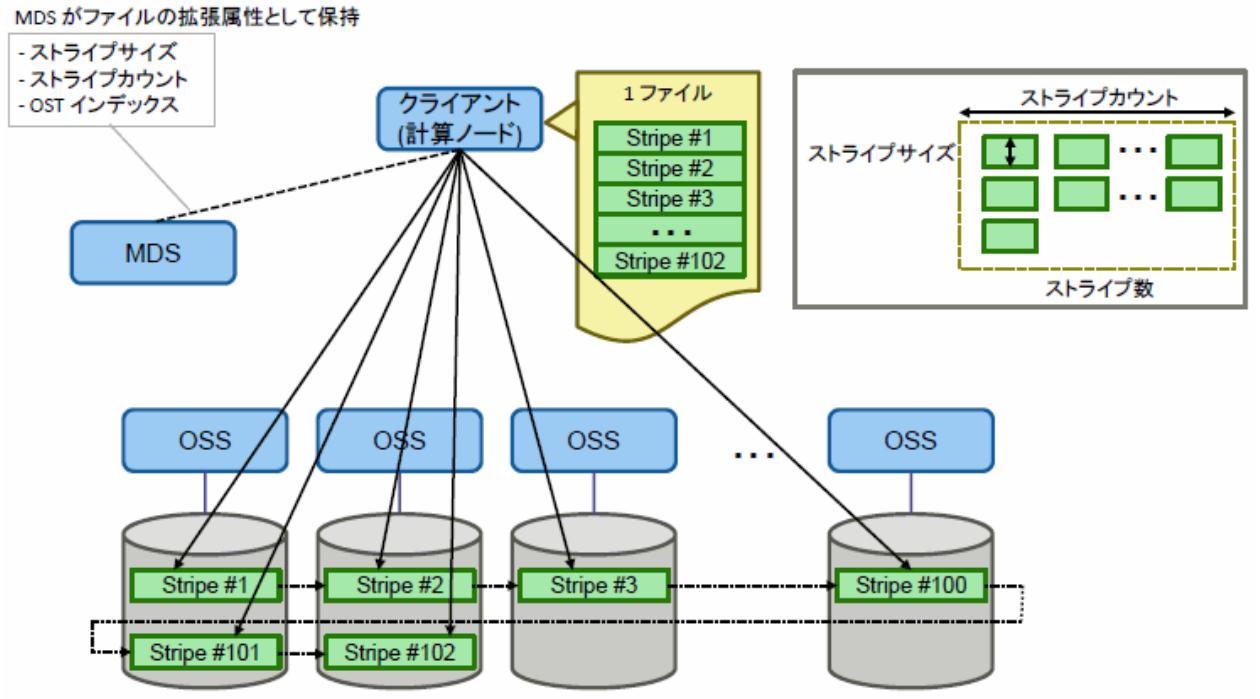
ストライプ機能は、1つのファイルのデータを先頭から指定したサイズ単位に複数の OST に分散して格納し、ファイルアクセスを分散させる機能です。

以下の効果があります。

- 物理的な 1つの OST の容量を超えるサイズのファイルが作成できます。

・ 1ファイルのデータを複数の OST に分散して格納することで、ファイルアクセスの帯域幅が向上します。

図2.2 ストライプ機能



なお、ストライプ機能を使用するには、設定が必要です。

分散させるデータのサイズ (ストライプサイズ) や分散させる OST の範囲 (ストライプカウント) などは、ストライプ機能を設定する際に指定できます。設定方法の詳細は、"4.3 ストライプ機能の設定" を参照してください。

2.1.3 OST_pool 機能

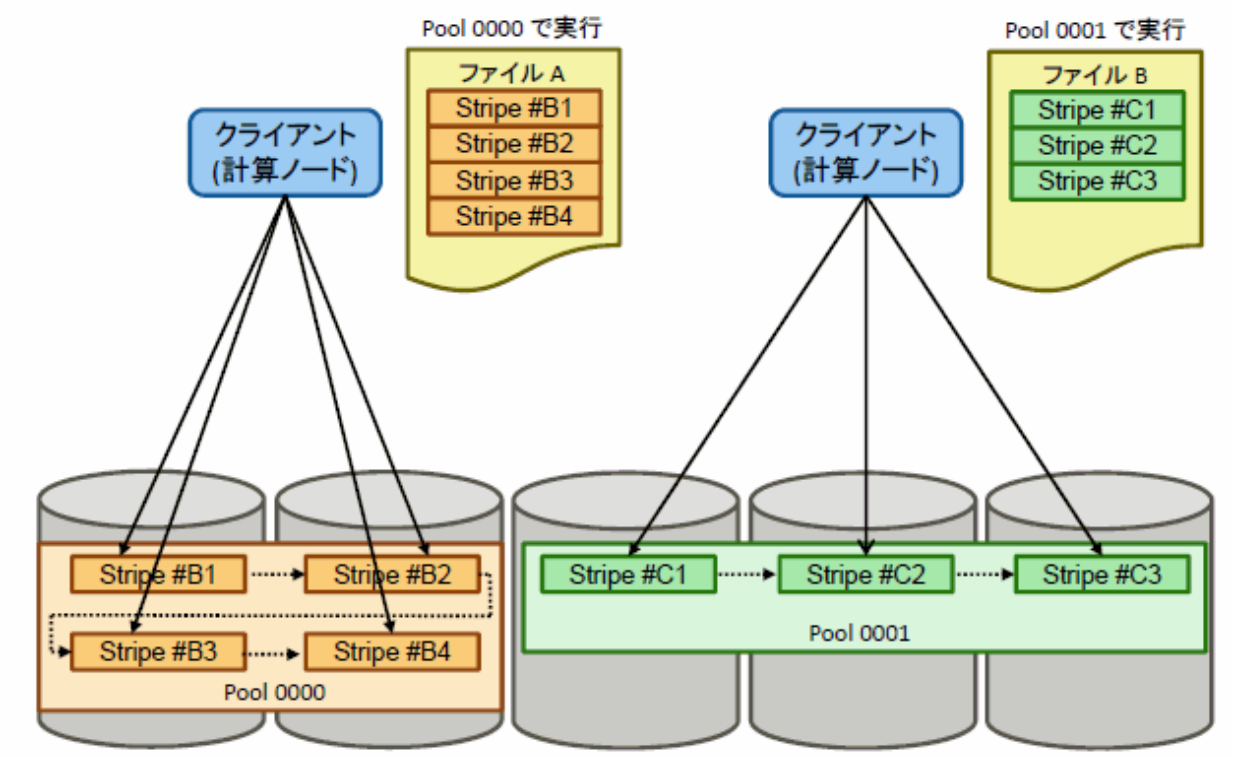
FEFS は、OST をグループ化して管理する OST_pool 機能を提供しています。

OST_pool 機能は、複数の OST をまとめて 1つのグループとして定義する機能です。

ストライプ機能でファイルを分散配置させる OST の範囲を指定できます。

以下は、OST_pool 機能の概念図です。

図2.3 OST_pool 機能

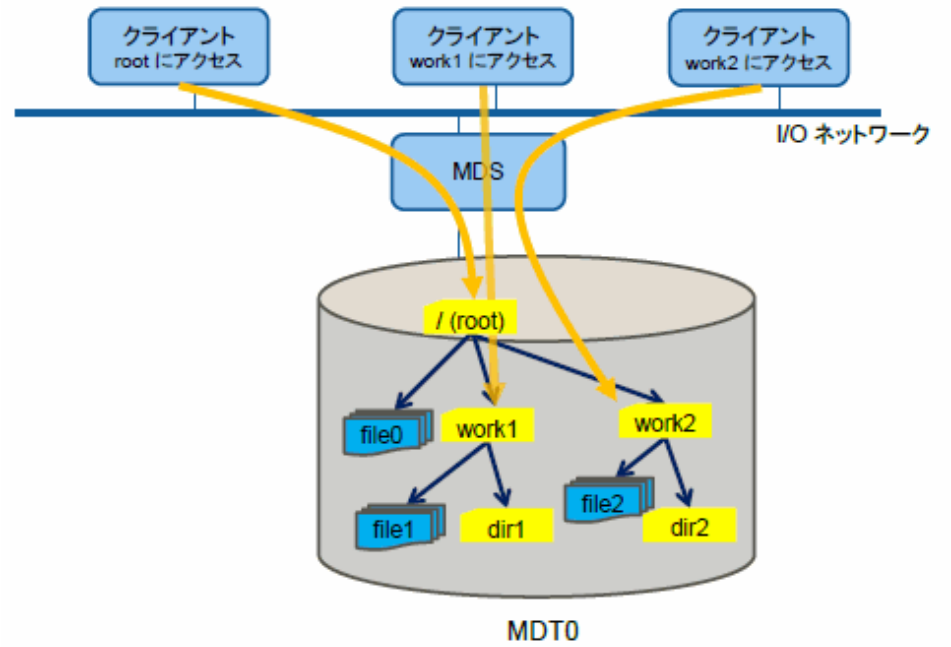


2.2 マルチ MDS 機能

FEFSでは、1つのファイルシステムを複数の MDS で構成するマルチ MDS 構成が可能です。

従来のシングルMDS構成の例を以下に示します。

図2.4 シングル MDS 構成の例



シングルMDS構成では、ファイルシステム全体のメタデータを 1台の MDS/MDT ペアで管理しているため、ファイルシステムの規模が大きくなるにつれて、メタデータアクセスが性能上のボトルネックとなります。マルチ MDS 構成では、メタデータを複数の MDS/MDT ペアで

管理することによって、アクセス負荷を分散でき、システム全体のパフォーマンスが向上します。また、MDT を増やすことによって、管理できるファイル数の上限を拡張することができます。

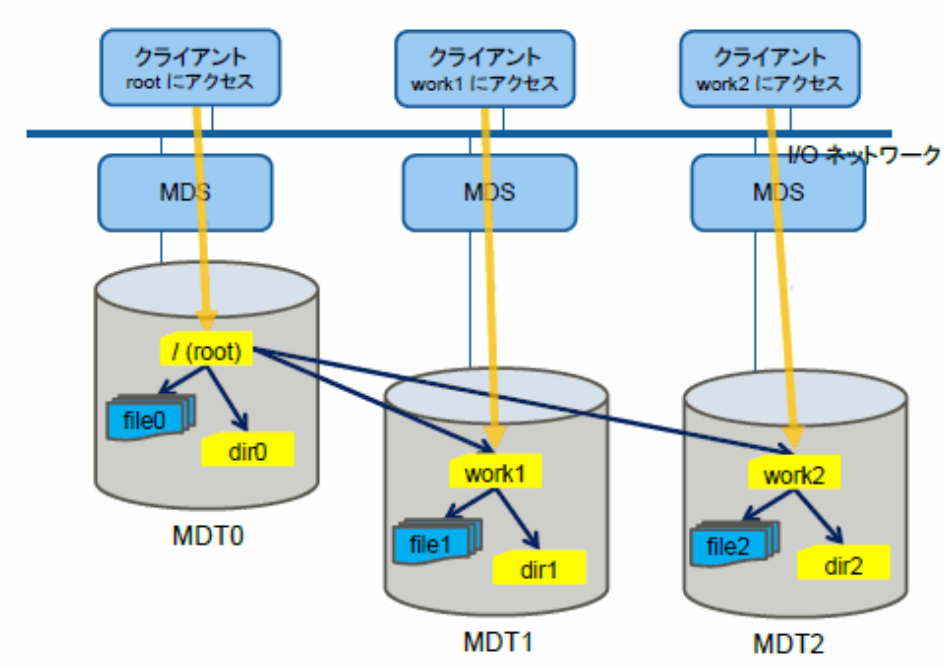
マルチ MDS構成を利用する方式には、リモートディレクトリとストライプディレクトリがあります。どちらかの方式で、メタデータを分散させる設定を明に行わない限り、すべてのメタデータは MDT0 だけに格納されます。

2.2.1 リモートディレクトリ

リモートディレクトリは、あるディレクトリのメタデータを、親ディレクトリのメタデータと異なる特定の MDT に格納する方式です。リモートディレクトリ配下にファイルを作成すると、リモートディレクトリと同じ MDT にそのメタデータを格納します。

リモートディレクトリの例を以下に示します。

図2.5 リモートディレクトリ



ルートディレクトリ (/) は、MDT0 をマウントする MDS が管理します。ここに、任意のMDT*n*をマウントする MDS を追加して、任意のサブディレクトリツリーを管理するよう設定できます。上の図のMDT1、MDT2 をマウントするMDS は、それぞれ /work1、/work2 を頂点とする部分木を管理します。

リモートディレクトリは、メタデータアクセス性能を必要とするユーザーやプロジェクトを、特定のMDTに割り付け、他のユーザーのメタデータアクセスからの影響を防ぐことができます。

注意

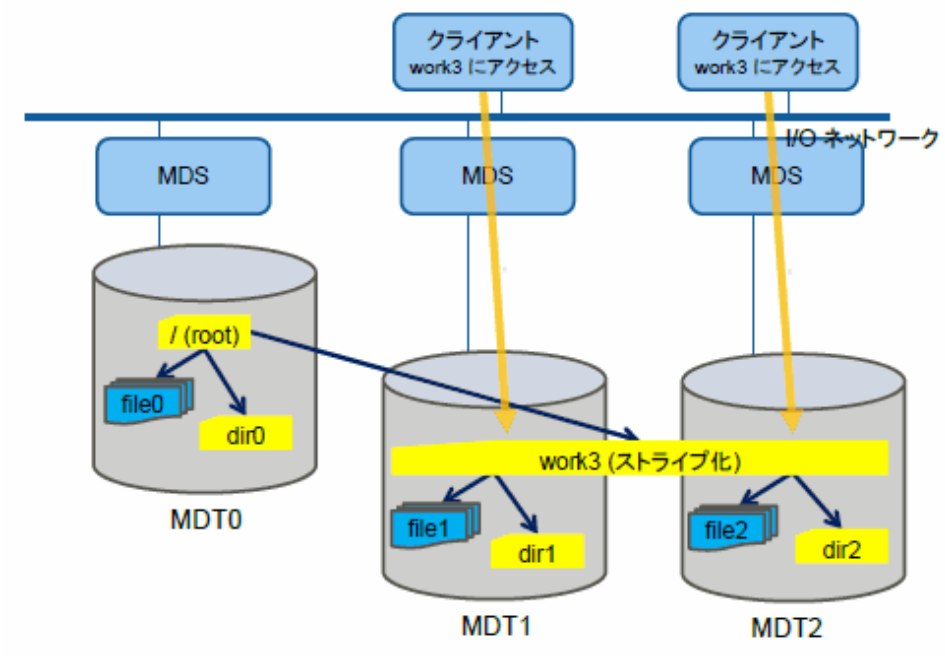
任意のMDTに割当てたディレクトリ配下のファイルを、別のMDTに割当てたディレクトリ配下に移動した場合に、移動元のMDTと移動先のMDTでinodeを消費します。例えば、上記の図でMDT1の/work1 配下のfile1をMDT2の/work2 配下に移動した場合に、MDT1 とMDT2 でそれぞれinodeを消費します。

2.2.2 ストライプディレクトリ

ストライプディレクトリは、あるディレクトリのメタデータを、指定した範囲の MDT に格納する方式です。ストライプディレクトリ配下にファイルを作成すると、指定した範囲内の 1つの MDT にそのメタデータを格納します。

ストライプディレクトリの例を以下に示します。

図2.6 ストライプディレクトリ



上の例では、/work3 の配下に存在するディレクトリ dir1とdir2、ファイル file1 と file2 のメタデータを、それぞれ MDT1とMDT2 に分散して格納しています。

ストライプさせる MDT の範囲(ストライプカウント)や、先頭の MDT 番号(ストライプインデックス)は、lfs mkdir コマンドでディレクトリを作成する際に指定します。FEFS は、ストライプディレクトリ配下のメタデータの MDT への割り当てを、ラウンドロビンで行います。

注意

- ストライプディレクトリ配下のファイルを移動した場合、別の MDT に割り当てられる場合があります。このとき、移動元の MDT と移動先の MDT の inode を消費します。

2.3 QUOTA 機能

QUOTA は、使用できるディスク容量とファイル数を制限するための機能です。QUOTA は、個々のユーザー、グループ、またはプロジェクトに対して、システム管理者が設定できます。

注意

- キャッシュとハードリミットについて
  lfs quota コマンドで表示されるディスク使用量は、OST の割り当てブロック数に基づいて計算されますが、ハードリミットを超えた値が表示されることがあります。
  FEFSでは高速化のため、書込みデータをメモリ上にキャッシュし、一定時間経過後に OST に書き込む処理を行っています。メモリ上にキャッシュした時点では OST 上の領域は消費しません。このため、QUOTA の制限値を超えた書込みが行われることがあります。これにより、lfs quota コマンドでディスクの使用状況を確認する際に、ハードリミットを超えた値が表示される場合がありますが、FEFS の動作としては問題ありません。
- ディスク使用量について
  lfs quota コマンドで表示されるディスク使用量は、lfs quota -v で表示される各 MDT と OST のディスク使用量の合計値です。QUOTA の制限値の対象となるのは OST のディスク使用量だけで、MDT のデータ量は制限値の対象となりません。このため、lfs quota コマンドでディスク使用状況を確認する際に、MDT のデータ量分の制限値を超えた値が表示される場合がありますが、FEFS の動作としては問題ありません。

- ソフトリミット超過時に表示される "*" について
  lfs quota コマンドで表示される情報で、ソフトリミットを超えた場合は数値の後に "*" が付加されます。

  lfs quota コマンドの出力例

```
# 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
```

  上記の状態から、ファイルを削除して inode 数がソフトリミット未満になった場合でも、"*" が消えないことがあります。
  "*" が表示された状態で、猶予期間 (grace) を超過した場合でも、inode 数 がソフトリミット未満であれば、ファイルの作成はできるため、FEFS の動作としては問題ありません。

- スパースファイルの扱いについて
  スパースファイル (Sparse File) とは、ファイルの途中に書込みがされていない領域が存在するファイルです。
  スパースファイルで書き込まれていない領域は、OST 上のブロックを消費しないため、QUOTA のディスク使用量としてカウントされません。

- QUOTA 制限値の誤差について
  ディスク使用量の QUOTA 制限は、設定した制限値に対してずれることがあります。その最大は以下となります。

  CN から書き込む場合 [FX]

```
4MiB × OST数 × クライアント数
```

  OST数: ディスク使用量の QUOTA 制限を設定するファイルシステムにおける OST の数
  クライアント数: ディスク使用量の QUOTA 制限を設定するファイルシステムをマウントするクライアントの数

  LN から書き込む場合 [PG]

```
512MiB × OST数 × クライアント数
```

  OST数: ディスク使用量の QUOTA 制限を設定するファイルシステムにおける OST の数
  クライアント数: ディスク使用量の QUOTA 制限を設定するファイルシステムをマウントするクライアントの数

  inode数のQUOTA制限値は、設定した制限値に対してずれる場合があります。その最大は以下となります。

```
1024 × MDT数
```

  MDT数: ディスク使用量の QUOTA 制限を設定するファイルシステムにおける MDT の数

- ストライプディレクトリ作成によるQUOTAの inode数の増加について
  ストライプディレクトリを作成すると、QUOTAの inode数が（ストライプカウント+1）増えます。

2.3.1 ユーザーまたはグループに対する QUOTA 機能

FEFS では、ext3 などのファイルシステムと同様に、以下の QUOTA 機能が利用できます。

- ファイル数およびディスク使用量の設定
- ソフトリミットおよびハードリミットの設定

また、FEFS 固有の QUOTA 機能には、以下の機能があります。

- QUOTA 管理用の専用コマンド ("lfs setquota" コマンドなど。管理用コマンドの詳細は、"A.2 コマンド" を参照してください)
- 複数の MDT で構成されたファイルシステムの場合に、MDT ごとのディスク使用量を表示する機能 ("lfs quota" コマンドの -v オプション)
- 複数の OST で構成されたファイルシステムの場合に、OST ごとのディスク使用量を表示する機能 ("lfs quota" コマンドの -v オプション)

2.3.2 プロジェクト QUOTA 機能

FEFSでは、ユーザーまたはグループに対してだけではなく、プロジェクトに対してもQUOTAを設定できます。「プロジェクト」は、ファイルやディレクトリなどを指す任意の inode の集合に、それぞれ同一の「プロジェクトID」を付与したものを言います。ユーザー名やグループ名と同様に、プロジェクトIDを単位として、QUOTA の設定・管理ができます。

プロジェクトIDを付与する対象は、通常は単一のディレクトリですが、複数のディレクトリに対して同一のプロジェクトIDを付与し、それらをまとめてQUOTA管理もできます。また、階層ディレクトリ構造において、各サブディレクトリに別のプロジェクトIDを付与もできます。従って、あるディレクトリに直接 QUOTA を設定するより柔軟な管理を行えると言えます。

以下は、あるディレクトリをプロジェクトに関連づけた場合の QUOTA 機能の運用イメージと適用シーンです。

図2.7 QUOTA の運用イメージ

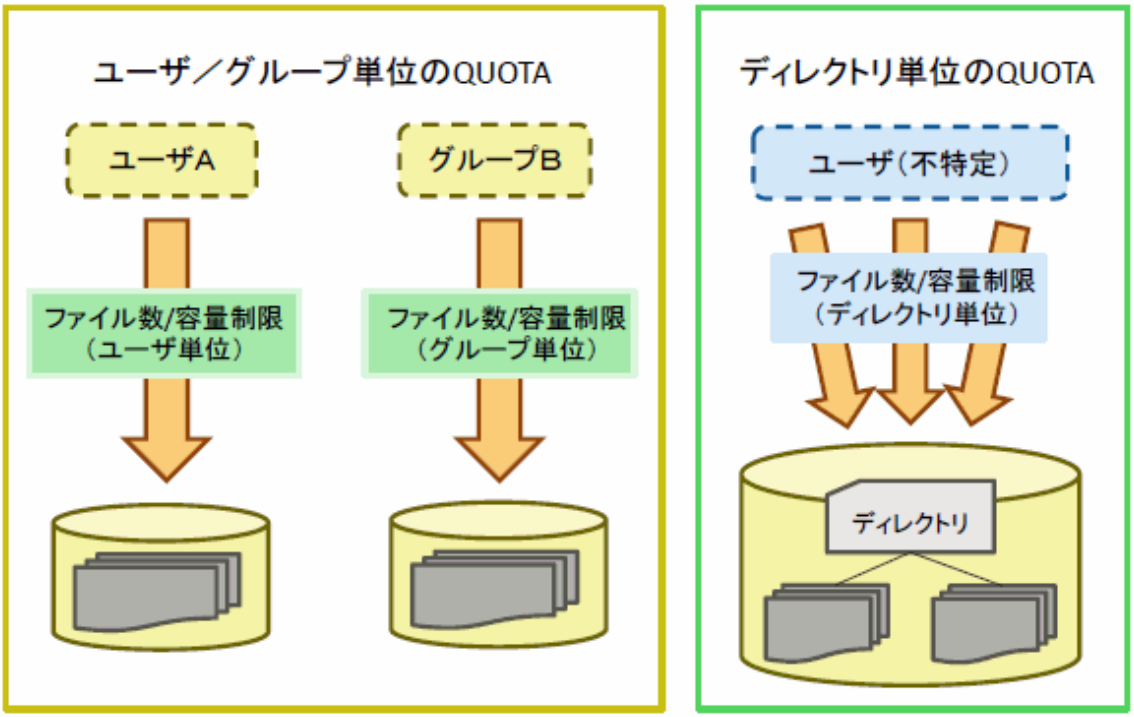
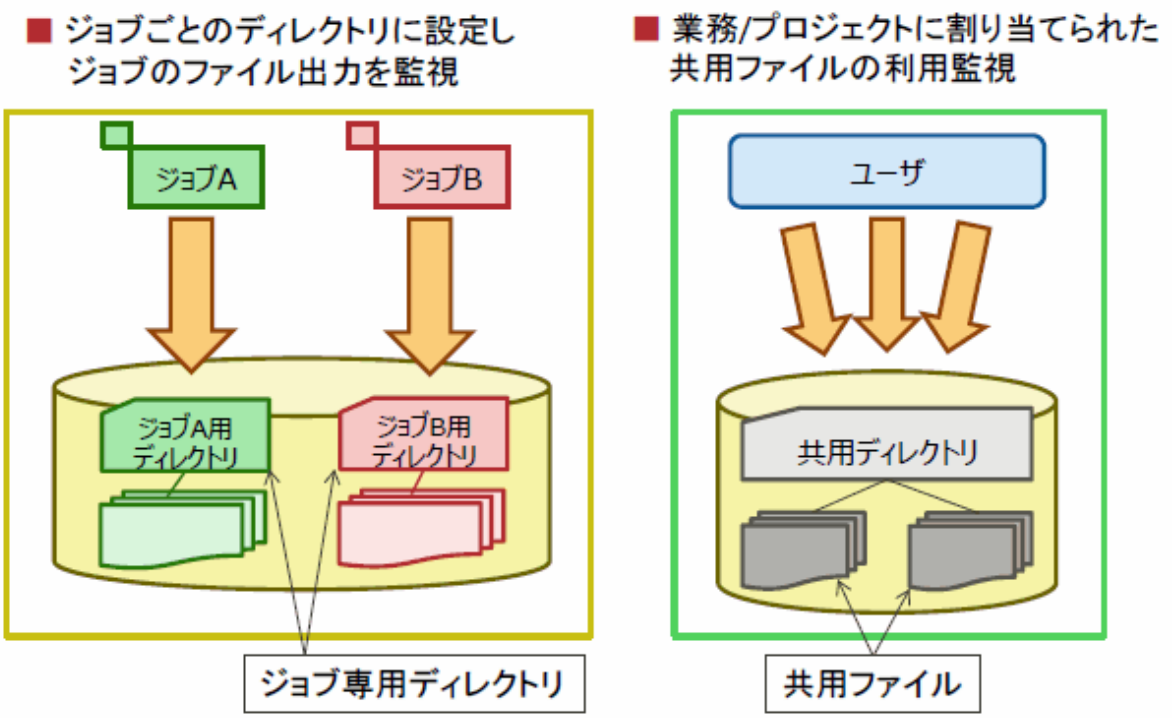


図2.8 プロジェクトQUOTA の適用シーン



注意

- rename、link システムコールについて
  異なるプロジェクト QUOTA が設定されている、または片方だけにプロジェクト QUOTA が設定されているディレクトリ間で、ファイルを rename システムコールで移動した場合に、移動先のプロジェクト QUOTA の QUOTA 制限値を超える場合でも制限値超過になりません。
  ディレクトリを rename システムコールで移動した場合、-1 が返ります (エラー番号 EXDEV) 。
  異なるプロジェクト QUOTA が設定されている、または片方だけにプロジェクト QUOTA が設定されているディレクトリ間で、ディレクトリやファイルを link システムコールでリンクファイルとして作成した場合、-1 が返ります (エラー番号 EXDEV) 。
- シンボリックリンク
  シンボリックリンクのディスク容量は、リンク先エントリーのサイズではなく inode が使用するサイズになります。また、ディレクトリのシンボリックリンクが存在した場合、そのリンク先ディレクトリ内のエントリーはディスク使用量および inode 数のカウントには含まれません。
- プロジェクトID として指定できる値は 1 ～ 4294967295 です。
- 各プロジェクトにユニークな値が割り振られるように、プロジェクトID を適切に管理してください。

2.4 QoS 機能

多数のユーザーが利用する大規模システムでは、特定のユーザーが大量にファイル I/O を行った場合でも、ほかのユーザーに悪影響を与えないことが求められます。また、計算ノードのジョブからファイルアクセスが行われている場合でも、ログインノード上のユーザーのレスポンスに影響しないことが求められます。

FEFS では、QoS (Quality of Service) 機能によりこれらの課題をクリアしています。QoS 機能には、以下の特長があります。

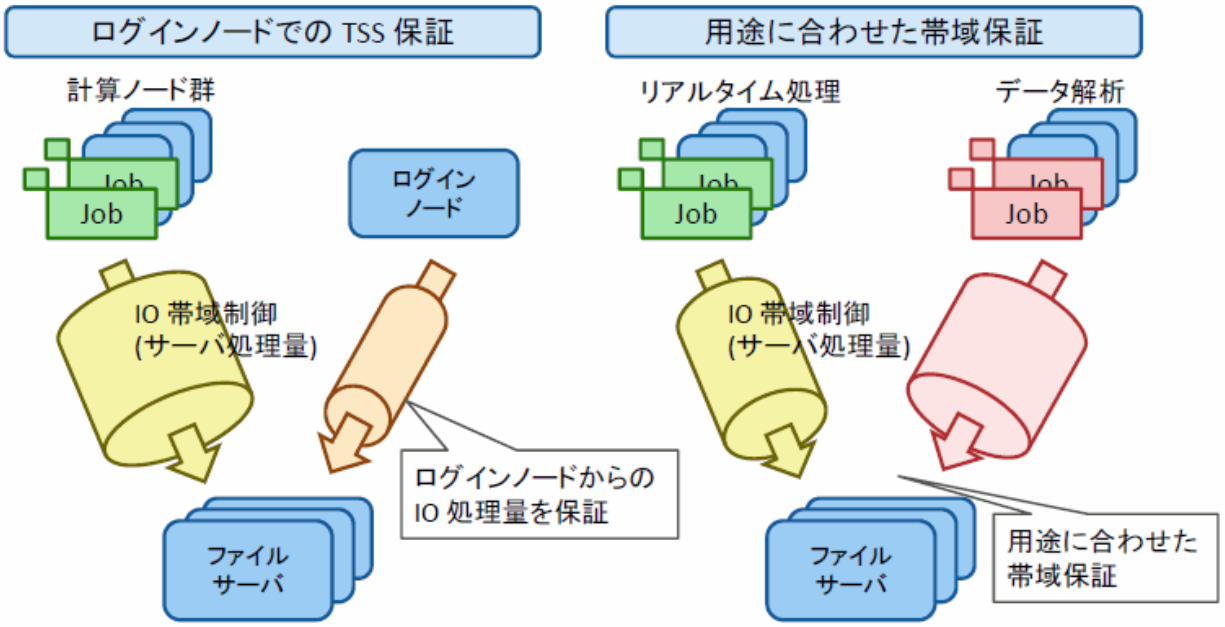
- クライアント間の I/O 優先制御機能
- ユーザー間フェアシェア機能

QoS 機能は、デフォルトでは無効になっています。QoS 機能を有効にする手順は、"3.2 QoS 機能を有効にする設定" を参照してください。

2.4.1 クライアント間の I/O 優先制御機能

クライアント間の I/O 優先制御機能では、サーバ側でクライアント群の I/O 処理量を制限できます。これにより、ジョブの I/O によるログインノードでの TSS レスポンス低下を防げます。また、クラスタの運用に合わせたクライアント群での帯域保証もできます。

図2.9 クライアント間の I/O 優先制御機能



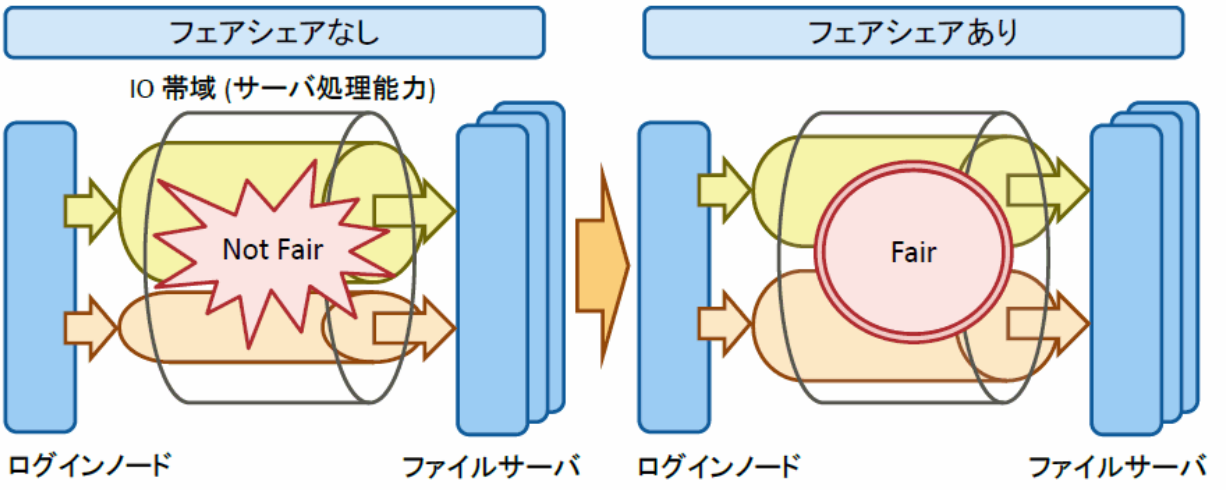
2.4.2 ユーザー間フェアシェア機能

ユーザー間フェアシェア機能では、クライアントおよびサーバでそれぞれ発行および処理する I/O リクエスト数を制御し、特定ユーザーによる I/O 資源の占有を防止します。

クライアント側では、1ユーザーが同時に発行できる I/O リクエストの上限を制限します。これにより、1ユーザーから大量に I/O リクエストが発行されることがなくなるため、I/O 帯域およびサーバ資源を占有することを抑止します。

また、計算ノードのように、複数クライアントから同じユーザーのアプリケーションが同時に I/O 要求を出すと、ファイルサーバ資源が占有される可能性があります。このため、ファイルサーバ側で 1ユーザーが利用できるサーバ処理能力を制御し、1ユーザーからの I/O 要求でサーバ資源が占有されることを防止します。

図2.10 ユーザー間フェアシェア機能



2.5 ACL (Access Control List) 機能

ACL は、従来の UNIX 形式のファイルアクセス制御 (ファイル所有者、グループ、およびそのほかのユーザー) に加えて、より柔軟なファイルアクセス制御を可能にする機能です。FEFSでは、個々のユーザーまたはグループに対するアクセス権限の設定ができます。

FEFS では、ext3 などのファイルシステムと同様に、以下の ACL 機能が利用できます。

- setfacl コマンドによる ACL の設定
- getfacl コマンドによる ACL 情報の表示

ACL 機能は、デフォルトでは無効になっています。ACL 機能を有効にする手順は "3.4 ACL 機能を有効にする設定" を参照してください。

2.6 ジャーナリング機能

FEFS には、電源の切断やシステムダウン時にファイルシステムの構成情報に矛盾が生じるのを避けるための、ジャーナリング機能があります。ジャーナリング機能は MDT および OST のそれぞれにあります。

また、ジャーナリング機能はジャーナルファイルへの記録処理のため定期的にディスクアクセスが必要になりますが、ジャーナル領域を外部デバイスにすることで、このオーバーヘッドを軽減できます。これを、外部ジャーナル機能と呼びます。

外部ジャーナル機能は、デフォルトで無効になっています。外部ジャーナル機能の設定方法は "3.1.3 FEFS デザインシートの作成" を参照してください。

注意

ジャーナルについては内部ジャーナルの使用を推奨します。
外部ジャーナルを使用する場合は、設定する外部ジャーナルサイズ分の空きメモリが必要になります。
外部ジャーナルの設定サイズによっては、MDS の性能や OSS 側のキャッシュ性能が低下する可能性があるため、設計の際は担当保守員 (SE) 、または当社 Support Desk に連絡してください。

2.7 RAS機能/FEFS の状態確認機能

フェイルオーバ機能による RAS 機能と FEFS の状態確認機能を提供します。

FEFS では、ハードウェアを二重化 (冗長構成) し、ソフトウェア制御によってサーバおよび I/O 通信の経路を切り替えることで、単点故障時にもファイルシステムとしてのサービスを持続できます。

フェイルオーバ機能により、FEFS のサーバに異常が発生した際も I/O は継続されるため、システムの運用継続を実現します。

2.7.1 フェイルオーバ構成

フェイルオーバは構成により、以下のように分類されます。

Active/Active 構成

フェイルオーバのペアとなるノードが、通常時は両ノードとも運用系となる構成です。
"図2.11 Active/Active のフェイルオーバ構成" に構成と動作概念を示します。
両ノードで FEFS サービスが起動しており、フェイルオーバすると正常なノード上にサービスが片寄せされた状態になります。

Active/Standby 構成

フェイルオーバのペアとなるノードが、通常時は運用系と待機系とに完全に分割される構成です。
"図2.12 Active/Standbyのフェイルオーバ構成" に構成と動作概念を示します。
通常時は、運用系ノードが FEFS サービスを提供し、待機系ノードは使用されていません。フェイルオーバすると、待機系ノードがFEFS サービスを提供します。

図2.11 Active/Active のフェイルオーバ構成

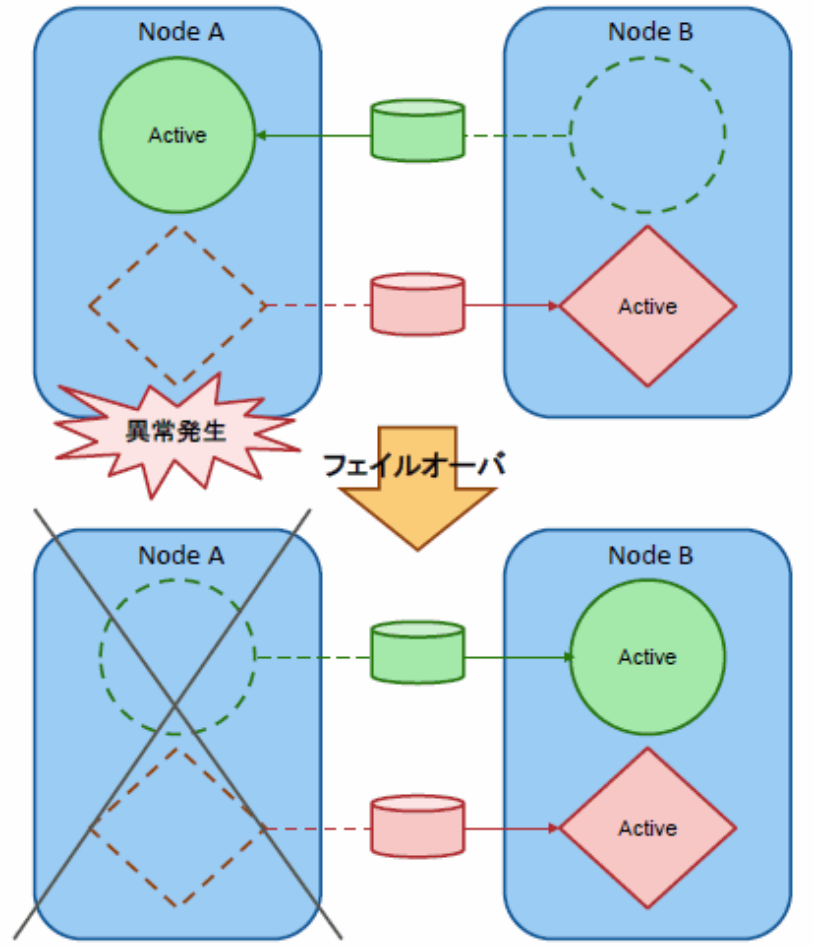
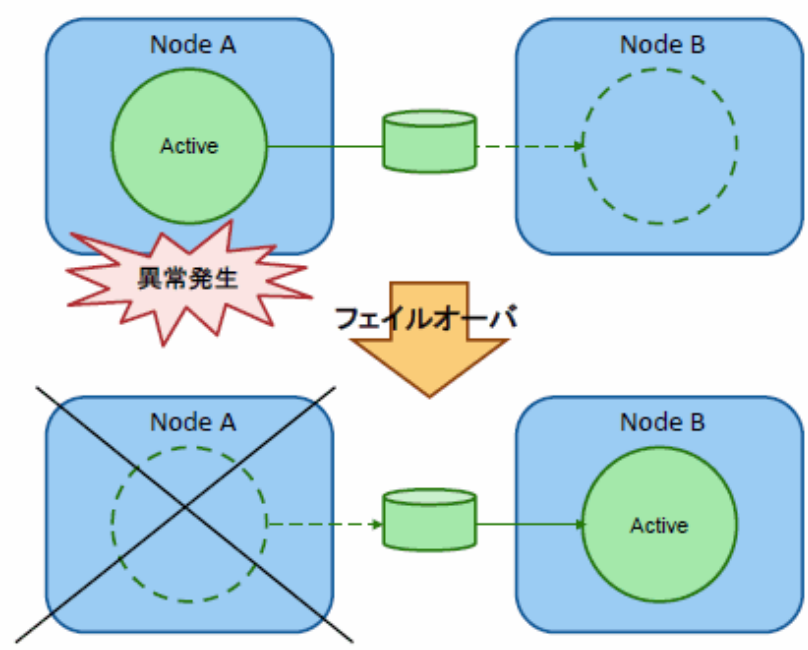


図2.12 Active/Standbyのフェイルオーバ構成



注意

- サービスが片寄せされた状態でも運用は行えますが、性能に影響があります。
- MDS を冗長構成にする場合、MDS と MDT (MGT) 間の接続はマルチパス構成にする必要があります。また、OSS を冗長構成にする場合、OSS と OST 間の接続はマルチパス構成にする必要があります。
- フェイルオーバ中は、ノードの切り替えが終了するまで一時的にジョブの I/O 処理が停止しますが、I/O 処理の再開後、ファイルアクセスは継続されます。

2.7.2 FEFS サービス監視

FEFS サービス監視は、ジョブ運用ソフトウェアとの連携時に利用可能な、FEFS 関連のサービス状態を監視・通知するための機能です。FEFS サービス監視デーモンが各ノードに常駐し、ジョブ運用ソフトウェアからの通知により FEFS 関連サービスの状態取得・通知を行います。

表2.1 FEFS サービス監視デーモンの監視対象(ハードウェア)

ノード種別	FC	InfiniBand	NVMe接続SSD※
MGS、MDS、OSS	○	○	－
CCM、LN、GIO[FX]、多目的ノード	－	○	－
SIO[FX]	－	－	○
CN	－	－	－

※: LLIOを利用する際に想定しているデバイスです。LLIOについて詳細は「LLIO ユーザーズガイド」を参照してください。

○:監視する
－:監視しない

表2.2 FEFS サービス監視デーモンの監視対象(ソフトウェア)

ノード種別	マウント状態		LLIO サービス	
	ストレージ	グローバル FS	サーバ	クライアント
MGS、MDS、OSS	○	－	－	－
CCM、LN	－	○	－	－
GIO[FX]	－	※	－	○
SIO[FX]	－	○	○	－
CN	－	※	－	○
多目的ノード	－	○	－	－

※: LLIO サービスが存在する環境では「－」、存在しない環境では「○」となります。
○:監視する
－:監視しない

FEFS のサービス監視機能は、サーバ機能と中継機能の監視を行う FEFSSR サービスと、クライアント機能の監視を行うFEFS サービスの2つに分かれています。詳細は、"4.11 FEFS の状態確認" および「ジョブ運用ソフトウェア 管理者向けガイド システム管理編」を参照してください。

2.7.3 FEFS の状態確認

pashowclst コマンドを使用して、各ノードの FEFS サービスの状態を確認できます。

参照

pashowclst コマンドによるサービスの状態確認方法は、以下のマニュアルを参照してください。

2.7.4 サーバのフェイルオーバ (MGS、MDS、および OSS)

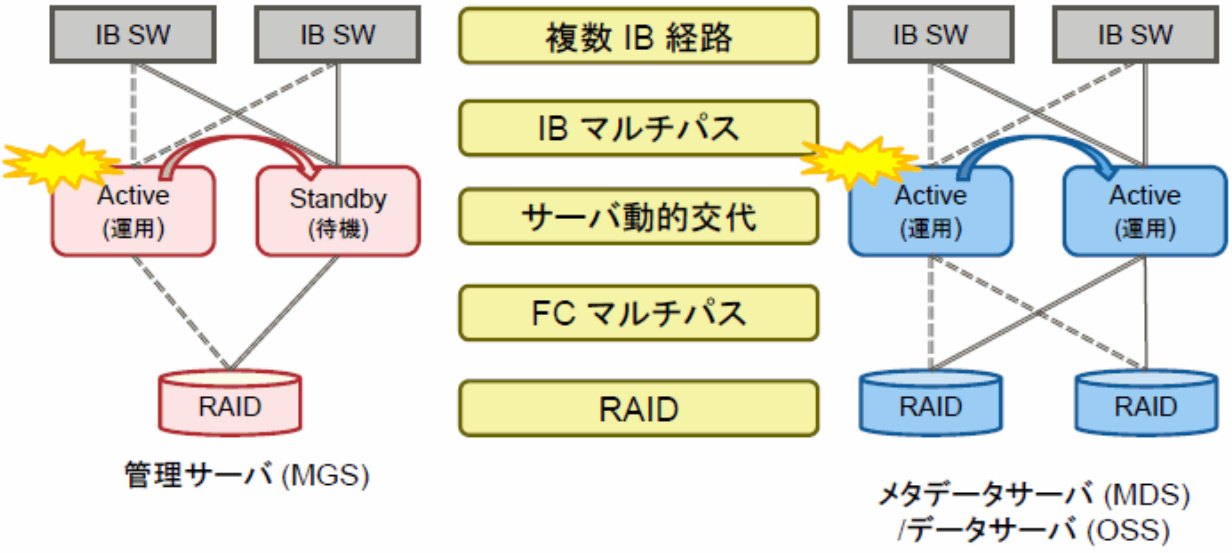
フェイルオーバ機能を利用するためには、冗長化された MGS、MDS、およびOSS が互いにフェイルオーバペアとなり、同じ MGT、MDT、および OST をマウント可能な状態にする必要があります。

ペアが共有している MGT、MDT、および OSTは、同時にはマウントされず、冗長化された MGS、MDS、および OSS の一方でマウントされます。

異常発生時に、MGT、MDT および OST をフェイルオーバペアのもう一方にマウントすることで I/O を継続できます。

FEFS 関連サービスに異常が発生した場合、各ノードに常駐している FEFS サービス監視デーモンがジョブ運用ソフトウェアへ異常を通知し、自動的にフェイルオーバが発生します。

図2.13 サーバのフェイルオーバ (MGS、MDS、および OSS)

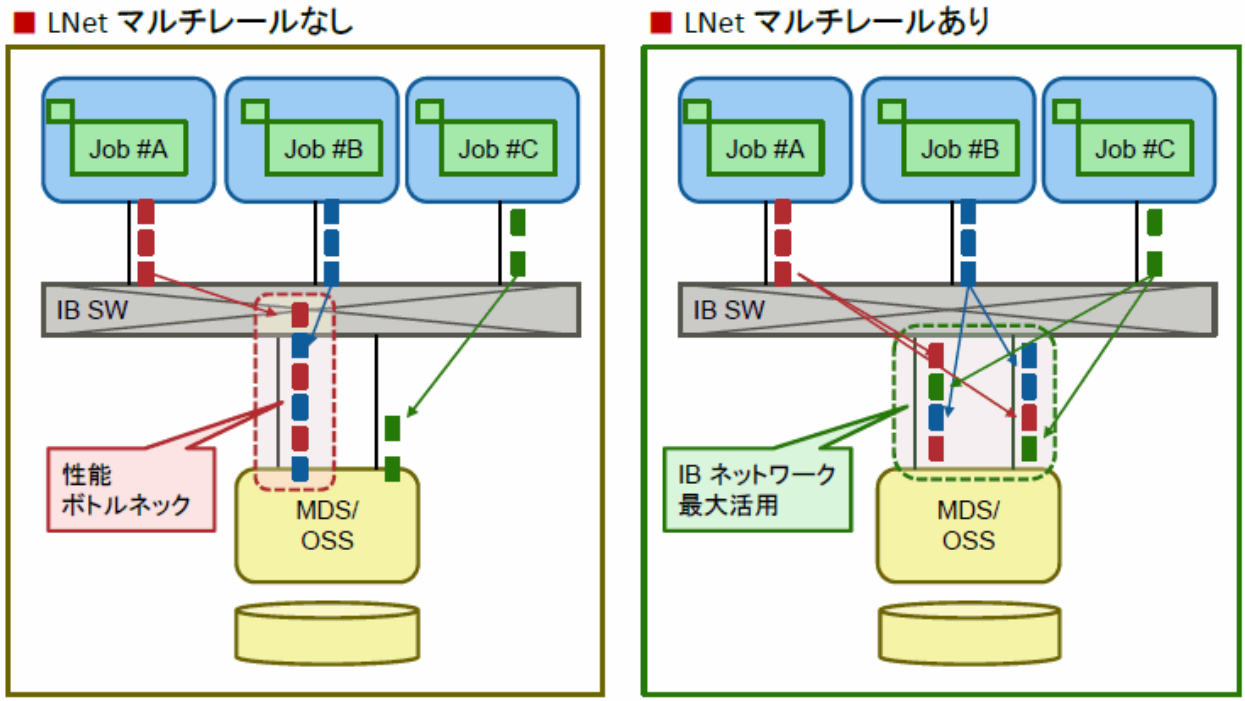


2.7.5 LNet マルチレール機能

FEFS では、サーバに搭載された複数の HCA に対して、高バンド幅と高い耐障害性を実現するために、複数の経路で同時に通信する機能をサポートします。本機能を LNet マルチレール機能と呼びます。

本機能を使用することで、InfiniBand 通信経路故障時に有効な経路が自動的に選択され、I/Oは継続されます。

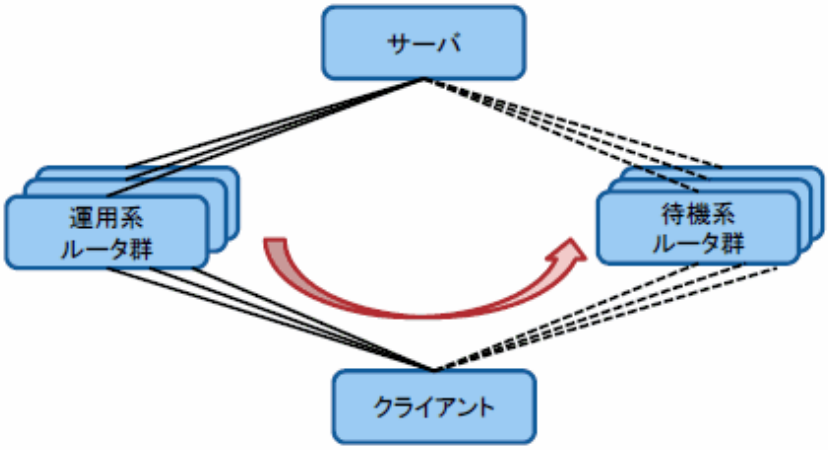
図2.14 LNet マルチレール機能



2.7.6 LNet ルータ

LNet ルータは、運用系 LNet ルータ群と待機系 LNet ルータ群を持ち、運用系 LNet ルータ群の異常時には待機系 LNet ルータ群に通信経路を切り替えて運用を継続します。

図2.15 LNet ルータにおける通信経路の切り替え



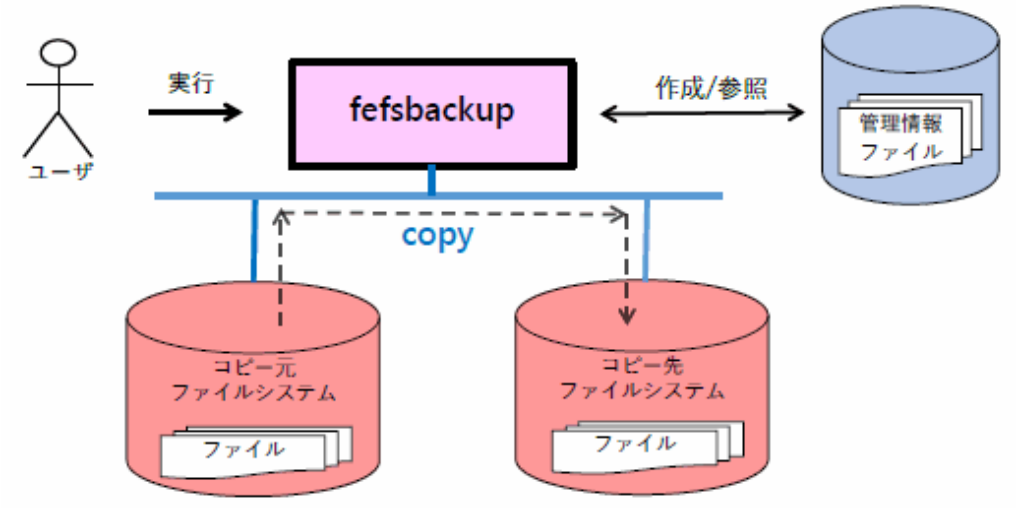
2.8 データ管理ツール (fefsbackup コマンド) [PG]

FEFSは、ファイルシステム間のデータ移行や、ディスク故障時のデータ退避を高速に行うために、データ管理ツールを提供します。データ管理ツールは、以下のような工夫をすることによって、高速なデータ転送を行います。

- rsync を並列実行する。
- サイズの小さなファイルが多数ある場合は、tar アーカイブにまとめて転送する。
- 2回目以降は差分ファイルのみを転送する。

fefsbackup コマンドの動作イメージを以下に示します。

図2.16 fefsbackupの動作イメージ



サブコマンドの詳細は、"4.20 データ管理ツール (fefsbackup コマンド) の使い方 [PG]" を参照してください。

注意

データ管理ツールは、京システムや Technical Computing Suite V2.0/V3.0 システムからのデータ移行をサポートしません。

データ管理ツールを利用するためには、サーバとクライアントに rsync がインストールされていることが必要です。

データ管理ツールがサポートするファイルシステムは、FEFS だけです。

データ管理ツールでファイル転送の実施中は、ファイルやディレクトリの更新、削除、追加を行わないでください。

2.9 FEFS 統計情報可視化機能 (fefssv.ph スクリプト)

FEFS統計情報可視化機能は、fefssv.ph スクリプトと呼ぶツールを利用することで、各サーバで FEFS 領域に対し実行されたFEFSリクエストの回数や I/O 量などを収集し、出力できます。

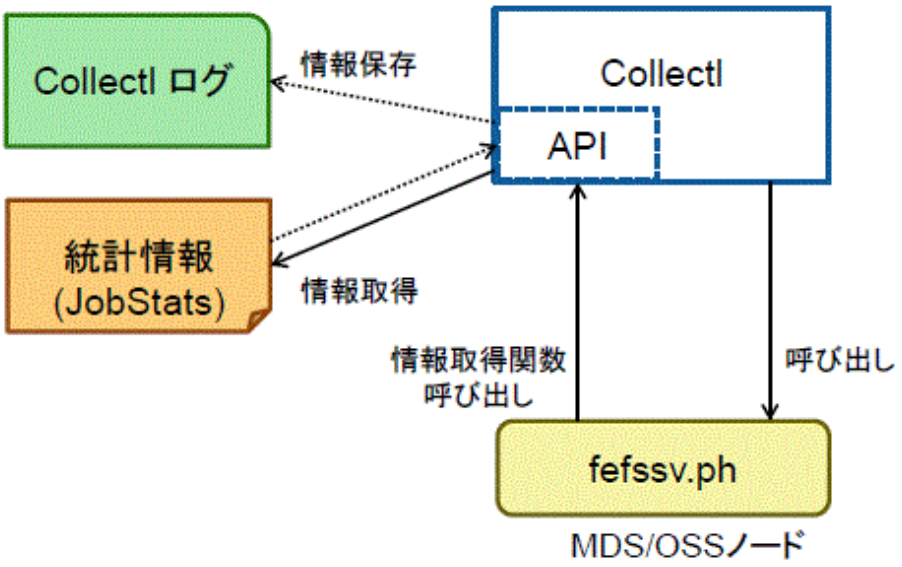
このツールの目的は、サーバ遅延などのトラブルが発生した際、各サーバ上の各種統計情報を把握・集計することによって、高負荷ジョブの特定をすることです。

fefssv.ph スクリプトは、監視目的で広く使われているパッケージである collectl から呼び出して使います。

情報の採取には、JobStats 機能を使います。JobStats についての詳細は "4.21 JobStats機能" を参照してください。

本機能の動作イメージを以下に示します。

図2.17 fefssv.ph の動作イメージ

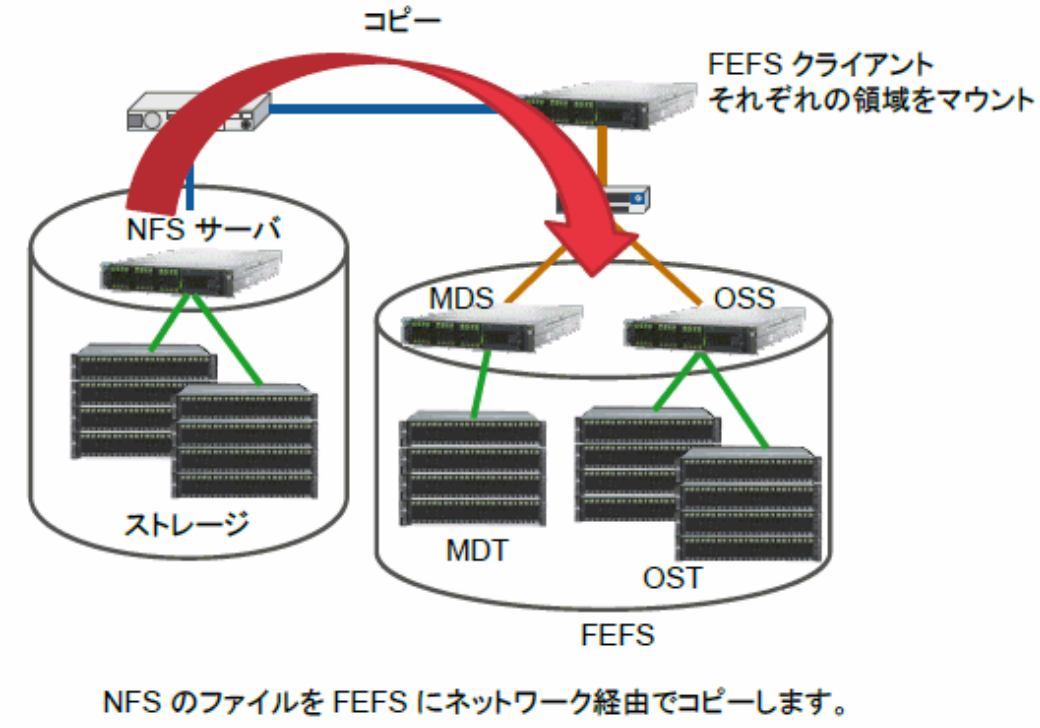


2.10 FEFS以外のファイルシステムとの連携機能

FEFS は、NFS および Lustre と連携できます。

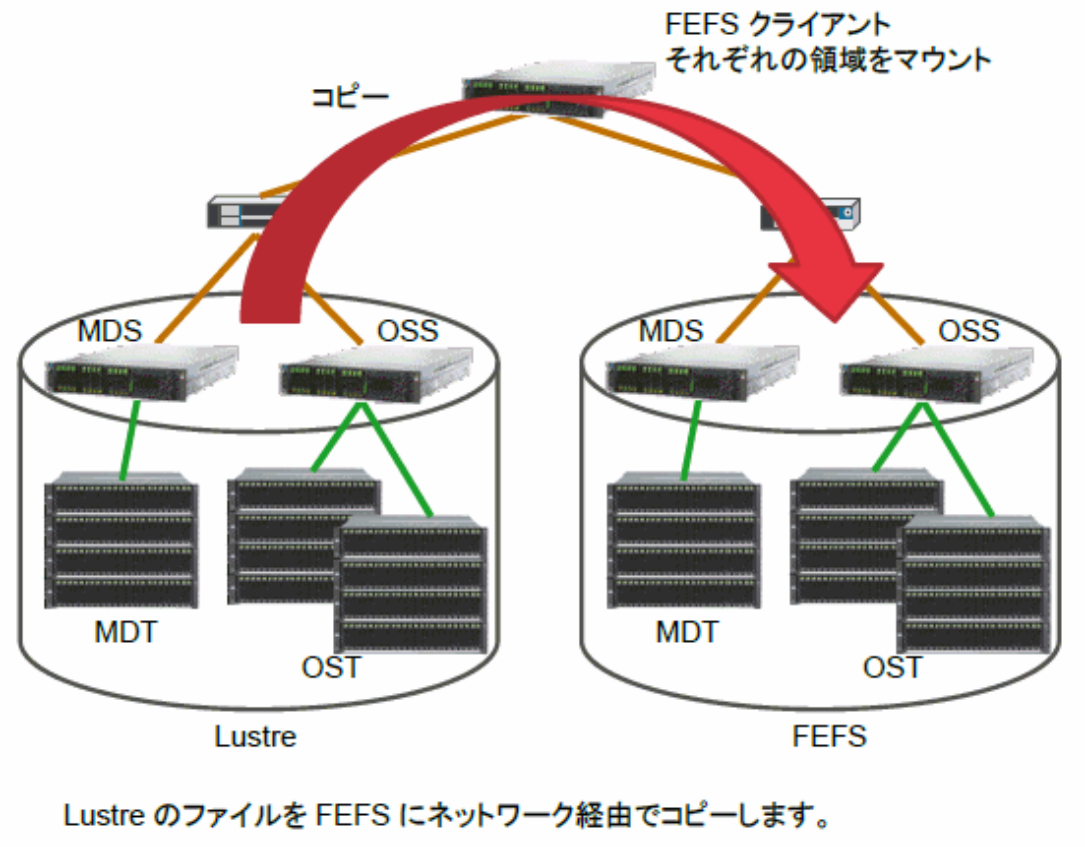
以下は、NFS から FEFS へのデータ移行と、Lustre から FEFS へのデータ移行の際の例を示しています。

図2.18 NFS から FEFS へのデータ移行



NFS のファイルを FEFS にネットワーク経由でコピーします。

図2.19 Lustre から FEFS へのデータ移行

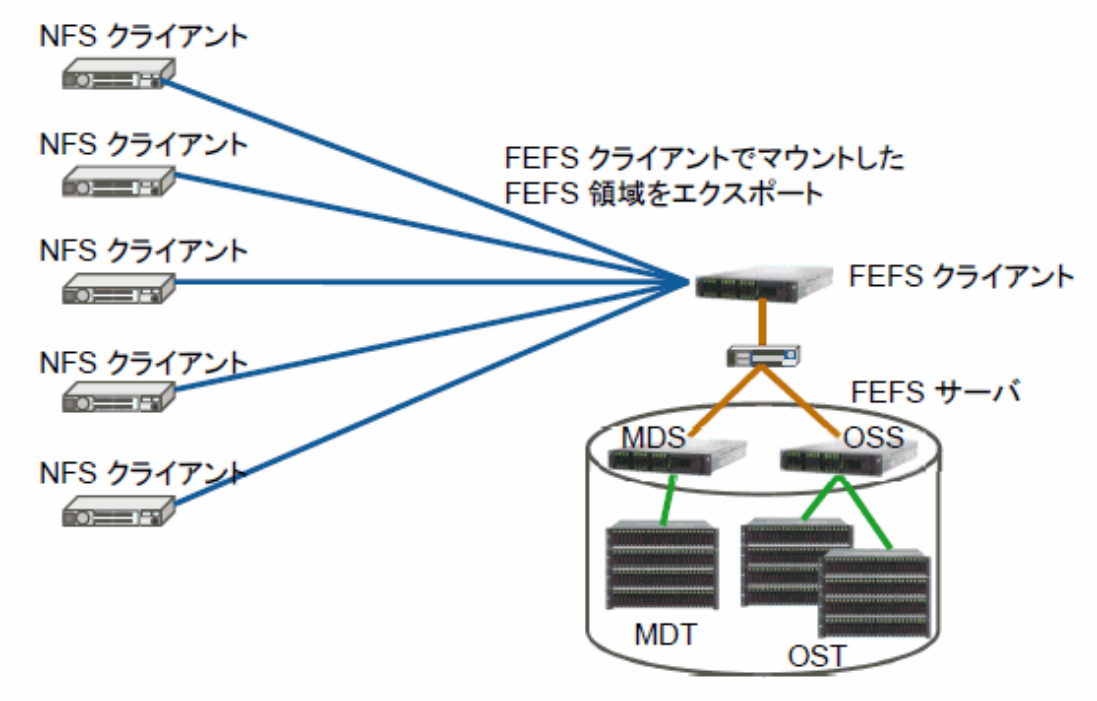


2.10.1 外部システムへのNFSによる公開

クライアントノードでマウントした FEFS を、NFS を使用して外部システムに公開できます。1つのFEFSサーバが提供するFEFSファイルシステムをFEFSクライアントからNFSで公開する場合、ファイルデータの一貫性・整合性を保つために、NFSサーバになれるのは1つのFEFSクライアントだけです。

FEFS を NFS で公開する場合、使用する NFS のバージョンは NFS バージョン 3 および バージョン 4 をサポートしています。

図2.20 NFSでの公開イメージ



参考

NFS クライアントノードからのファイルデータの一貫性・整合性に関しては、NFS の仕様に従います。

2.10.2 Lustre 接続 [PG]

FEFS クライアントから Lustre サーバをマウントする、または Lustre クライアントから FEFS サーバをマウントしてサーバ上のファイルにアクセスする機能を提供します。

サポートする Lustre の版数は 2.10.8です。

注意

- 同一ノード上に FEFS と Lustre を同時にインストールできません。
- FXサーバのFEFSクライアントから、Lustreサーバはマウントできません。
- Lustre 上では以下の機能は使用できません。
  - ― QoS 機能
  - ― LNet マルチレール機能

第3章 FEFS の導入と保守

ここでは、FEFS の導入および保守の方法を説明します。システムに必要なソフトのインストールは、システム管理のインストール機能がまとめて行います。以下の説明は、FEFS を構成するrpmパッケージ ("3.1.2 FEFS パッケージの適用" 参照) が適切なノードに配置されているものとして行います。

3.1 導入の流れ

FEFS の導入は、以下の順に行います。

1. FEFS 構成の設計
2. FEFS パッケージの適用
3. FEFS デザインシートの作成
4. FEFS セットアップツール用構成定義ファイルの作成
5. FEFS セットアップツール用構成定義ファイルの配置
6. FEFS の構築
7. ファイルシステムのパーミッション変更
8. 構築後に必要な設定

注意

「FEFS の構築」以降の操作を実施する場合は、計算ノード以外のFEFSがインストールされているすべてのノードが起動されている必要があります。

参照

導入時に指定するノード種別は、以下のとおりです。
各ノードの詳細は、「ジョブ運用ソフトウェア 概説書」を参照してください。

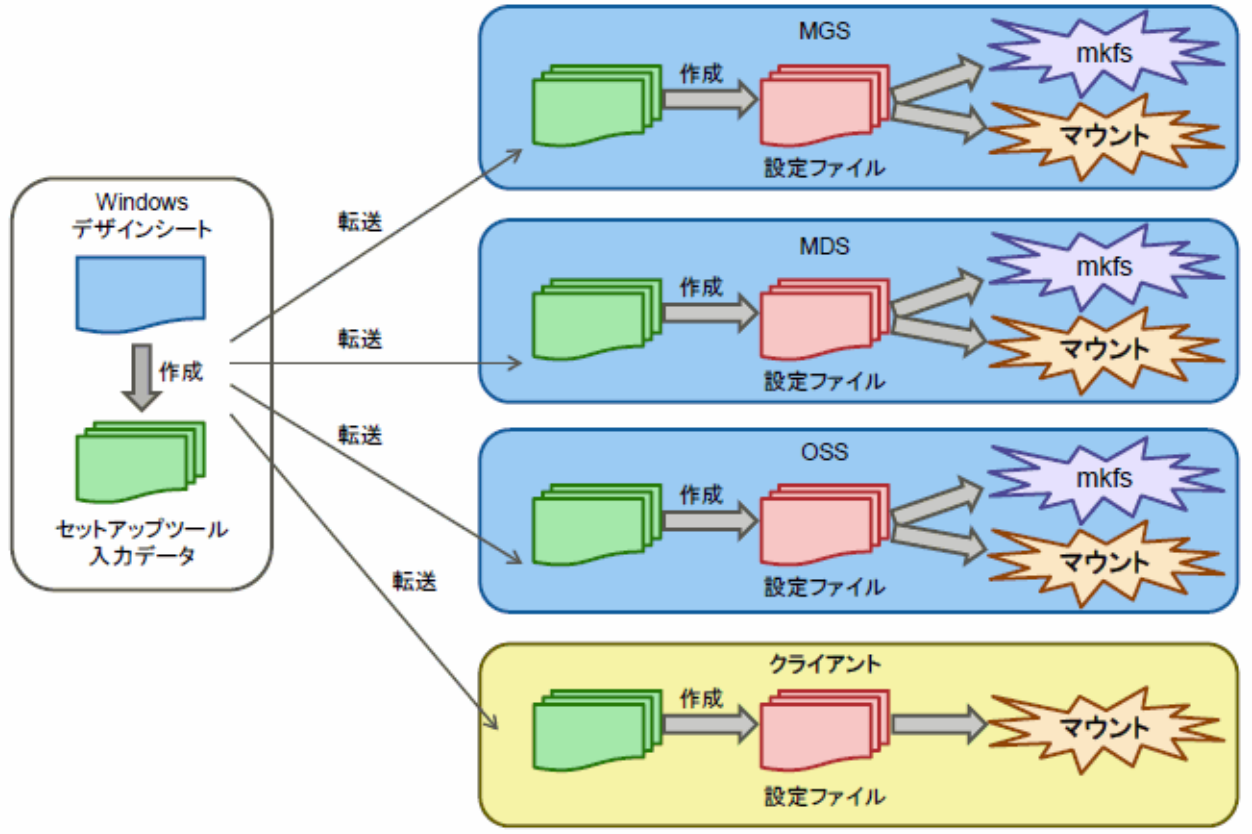
表3.1 ノード種別

種別	意味
MGS	管理サーバ
MDS	メタデータサーバ
OSS	オブジェクトストレージサーバ
SMM	システム管理ノード
CCM	計算クラスタ管理ノード
LN	ログインノード
CN	計算ノード[FX]
CN/BIO	BIOとCNの兼用ノード (FEFS デザインシートでは "CN-BIO") [FX]
CN/GIO	GIOとCNの兼用ノード (FEFS デザインシートでは "CN-GIO") [FX]
CN/SIO	SIOとCNの兼用ノード (FEFS デザインシートでは "CN-SIO") [FX]

※多目的ノードの扱いについては、"3.1.3.1 NODE シートの入力" の「注意」を参照してください。

FEFS 導入における FEFS セットアップツール用構成定義ファイルと FEFS 設定ファイルのイメージを以下に示します。

図3.1 FEFS導入の流れ



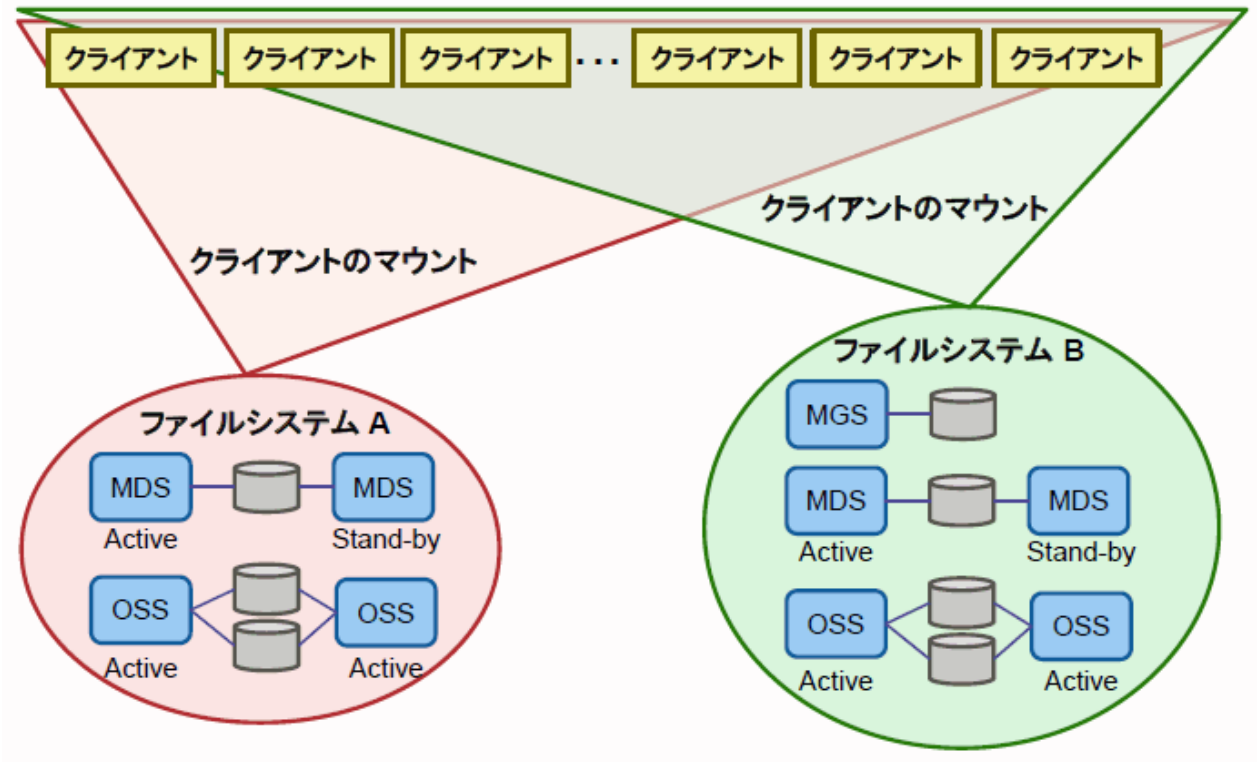
3.1.1 FEFS 構成の設計

FEFS の構成について、以下の観点で具体的に決定します。

- 機能・構成の選定
  - MDS および OSSでの外部ジャーナル機能の使用可否
  - QoS 機能の使用可否
  - ACL 機能の使用可否
  - QUOTA 機能の使用可否
- 構成の決定
  - ノード構成

    "図3.2 ノード構成決定" に示すように、MGS、MDS、OSS、およびクライアントの構成を決定します。
    - ファイルシステムの数
    - MGS、MDS および OSS の冗長化の有無
    - クライアントの台数
    - MGS、MDSの兼用

図3.2 ノード構成決定

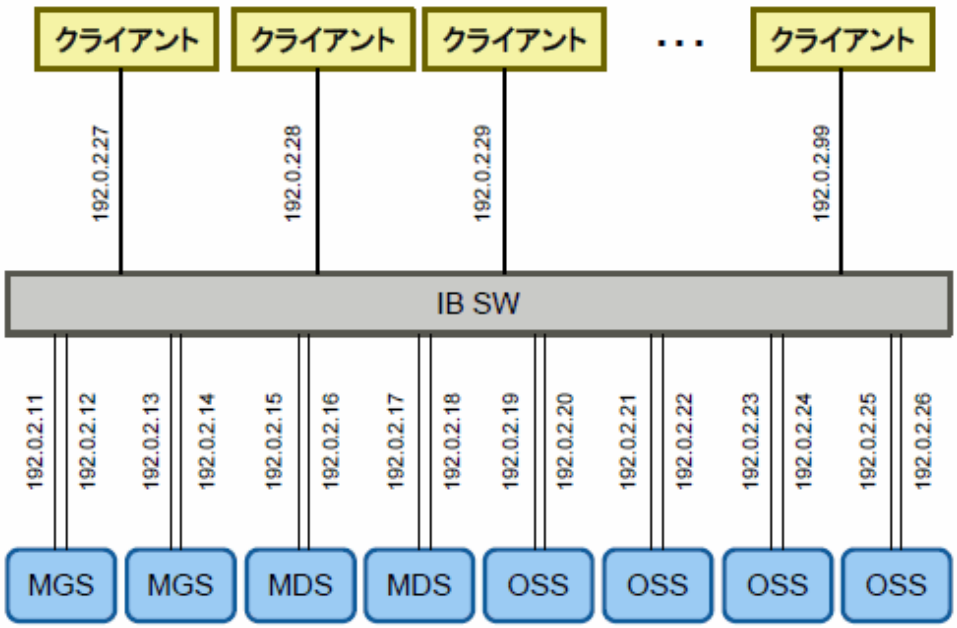


ー ネットワーク構成

"図3.3 ネットワーク構成決定" に示すように、利用する InfiniBand の構成と IPアドレスを決定します。

- MGS における InfiniBand の本数
- MDS における InfiniBand の本数
- OSS における InfiniBand の本数
- PG クライアントにおける InfiniBandの本数
- GIO兼CN における InfiniBandの本数
- IP アドレス

図3.3 ネットワーク構成決定



ー ボリューム構成

使用するボリュームの構成を決定します。

- MDT 関連ボリューム

  MDT ボリューム

  MGTボリューム (MDS とMGS を別サーバにする場合、または MDS の冗長構成を Active/Active 方式にする場合)

- OST 関連ボリューム

  OST ボリューム

参照

LLIO を導入する場合は、LLIO の構成も決定してください。詳細は、「LLIO ユーザーズガイド」の「3.1.1 LLIO 構成の設計」を参照してください。

参考

ボリューム情報は、以下の方法で確認できます。

- ボリューム情報確認

  ー ETERNUS マルチパスドライバ
  ボリュームは by-id 名で指定します。
  ボリューム名は、ノード上の以下のシンボリックリンクで確認できます。

```
/dev/disk/by-id/scsi-3600000e00d0000000002151900010000
```

  ー Device Mapper マルチパス
  ボリュームは /dev/mapper/<*ボリューム名*> で指定します。
  Device Mapper マルチパス使用時のボリューム名は、フレンドリ名が無効の場合は WWID(World Wide Identifier)、フレンドリ名が

有効の場合は "mpathN"(mpatha, mpathbなど) を指定してください。なお、これら以外のボリューム名には対応していません。
ボリューム名は、multipath -ll コマンドで確認できます。

- フレンドリ名が無効の場合

```
# multipath -ll
3600000e00d0000000001151a00000000 dm-0 FUJITSU,ETERNUS_DXL
size=818G features='1 queue_if_no_path' hwhandler='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
```

- フレンドリ名が有効の場合

```
# multipath -ll
mpatha (3600000e00d0000000001151a00000000) dm-0 FUJITSU,ETERNUS_DXL
size=398G features='1 queue_if_no_path' hwhandler='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
```

・ ボリューム名からストレージ機器の確認
ボリューム名からストレージ機器を判別する方法は、ストレージ機器の製品マニュアルを参照し確認してください。

例)ETERNUS でデバイス識別番号を確認する方法

ー ノードでデバイス識別番号と LUN (Logical Unit Number) を確認します。

- ETERNUS マルチパスの場合

```
/dev/disk/by-id/scsi-3600000e00d0000000021519 0001 0000
        [021519:デバイス識別番号]
        [0001:LUN]
```

- Device Mapper マルチパスの場合

```
# multipath -ll
  3600000e00d0000000021519 0001 0000 dm-1 FUJITSU,ETERNUS_DXL
      [021519:デバイス識別番号]
      [0001:LUN]
```

ー ETERNUS のデバイス識別番号を確認します。
ETERNUS に CLI (コマンドライン インターフェース) でログインし、エンクロージャステータスから取得します。

```
CLI> show enclosure-status
Enclosure View
Name []
Model Upgrade Status [Not Possible]
Model Name [ET08E21B]
Serial Number [XXXXXXXXXX]
Device Identification Number [021519]      ← デバイス識別番号
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 FEFS パッケージの適用

FEFSを構成するパッケージは以下のとおりです。

FEFS サーバパッケージ

1. FJSVfefsprogs-*.x86_64.rpm
2. FJSVfefs-modules-*.x86_64.rpm
3. FJSVfefs-osd-ldiskfs-modules-*.x86_64.rpm
4. FJSVfefs-osd-ldiskfs-mount-*.x86_64.rpm
5. FJSVfefs-*.x86_64.rpm

※ *には版数とリリース名が入ります。

パッケージと適用ノードの関係を以下に示します。

表3.2 FEFS サーバパッケージと適用ノード

パッケージ名	ノード種別		
	MGS	MDS	OSS
FJSVfefsprogs	○	○	○
FJSVfefs-modules	○	○	○
FJSVfefs-osd-ldiskfs-modules	○	○	○
FJSVfefs-osd-ldiskfs-mount	○	○	○
FJSVfefs	○	○	○

FEFS クライアントパッケージ[PG]

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

※ *には版数とリリース名が入ります。

パッケージと適用ノードの関係を以下に示します。

表3.3 FEFS クライアントパッケージと適用ノード (PGクライアント)

パッケージ名	ノード種別				
	SMM	CCM	LN	CCS※1	多目的※2
FJSVfefs-client-modules	○	○	○	○	○
FJSVfefs-client	○	○	○	○	○

※1: CCSとCNの兼用ノードとして利用する場合だけ、パッケージを適用してください。
※2: FEFSクライアントとして使う場合だけ、パッケージを適用してください。

FEFS クライアントパッケージ[FX]

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

※ *には版数とリリース名が入ります。

パッケージと適用ノードの関係を以下に示します。

表3.4 FEFS クライアントパッケージと適用ノード (FXクライアント)

パッケージ名	ノード種別
	CN
FJSVfefs-client-modules	○
FJSVfefs-client	○

参照

LLIO を導入する場合はLLIOの構成も決定してください。詳細は「LLIO ユーザーズガイド」の「3.1.2 LLIO パッケージの適用」を参照してください。

3.1.3 FEFS デザインシートの作成

FEFS デザインシートを、Windows 端末で作成してください。

FEFS デザインシートのひな形は、製品に同梱されています。FEFS デザインシートのファイル名は、"FEFSDesignSheet.xlsm" です。

FEFS デザインシート作成作業を始めるときは、最初に Excel のマクロを有効にしてください。

なお、セルの色が赤の入力項目は設定が必須の項目です。必ず値を入力してください。

注意

FEFS デザインシートは、以下の環境での入力をサポートしています。

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

これ以外の環境については、担当保守員(SE)または当社Support Deskに相談してください。

作成するシートは、以下の 3種類です。

- NODE(1)シート、NODE(2)シート
  ネットワーク情報など各ファイルシステムに依存しない構成情報を定義します。
  NODE(2) シートには、InfiniBand を 3本以上利用するノードの構成情報を定義します。
- LLIOシート
  LLIOを利用するための構成およびデバイスを定義します。未入力の場合はLLIOを利用しません。
- GFS シート
  グローバルファイルシステムの構成を定義します。ファイルシステムの数だけ作成してください。

以下、各シートの入力内容について説明します。

3.1.3.1 NODE シートの入力

各ファイルシステムで共通な情報を設定します。

1. NODE セクション
   FEFS が関連する全ノードのネットワーク情報を列挙してください。ジョブ運用ソフトウェアのインストール機能を利用している場合は、ジョブ運用ソフトウェアの定義ファイルの情報から NODE(1) シート、NODE(2) シートに入力するノード情報をインポートしてください。

   インポート方法については、後述の "ジョブ運用ソフトウェアのノード情報インポート" を参照してください。

以下は、NODE セクションの記入例です。

図3.4 NODE セクションの記入例

HOSTNAME	NODETYPE	CLSTNAME	Primary Network			Secondary Network			Tofu Coord						Tofu	SYSTEM ID
			N/I	IP ADDRESS	PREFIX	N/I	IP ADDRESS	PREFIX	X	Y	Z	A	B	C	IP ADDRESS	
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
   ホスト名を記述します。

b. NODETYPE
   ノード種別を記述します。プルダウンから選択してください。多目的ノードについては、下記の注意事項を参照してください。

c. CLSTNAME
   クラスタ名を定義します。

d. Primary Network、Secondary Network
   InfiniBand の情報を定義します。ネットワークインターフェース、IPアドレスおよびプレフィックスを入力してください。

e. Tofu Coord
   Tofu座標 (X,Y,Z,A,B,C) を定義します。

f. Tofu IP ADDRESS
   TofuインターコネクトのIPアドレスを定義します。

g. SYSTEM ID
   同一のTofu座標を持つノードが存在する場合に使用します。詳細は、下記の注意事項を参照してください。

注意

1HCA のノードは InfiniBand の情報を Primary Network に入力してください。

Tofuインターコネクトを持つノードは座標 (X,Y,Z,A,B,C) を入力してください。

FEFS で利用しないネットワークは入力する必要はありません。

NODE(1) シート、NODE(2) シートは InfiniBand の情報の数で使い分けてください。InfiniBand の情報が 3本以上ある場合だけ NODE(2) シートを使用し、それ以外はすべて NODE(1) シートを使用してください。NODE(2) シートは、Excel のマクロで追加します。

Excelマクロの「FEFS Design」>「Insert extra-IB node sheet」

多目的ノードの NODETYPE について

Technical Computing Suite には、多目的ノードというノード種別があります(詳細は、ジョブ運用ソフトウェアの「概説書」および「導入ガイド」を参照)。このノードは、NODETYPE の選択リストにありません。
3文字以上 5文字以下の半角英大文字および数字で設定してください。
NODETYPE は、ジョブ運用ソフトウェアのデザインシートで定義した名前と同じである必要があります。
命名ルールの詳細は、「ジョブ運用ソフトウェア 導入ガイド」を参照してください。
ただしジョブ運用ソフトウェアのノード情報をインポートする場合は、入力不要です。

ジョブ運用ソフトウェアのノード情報インポート

ジョブ運用ソフトウェアのノード情報を、NODE(1) シート、NODE(2) シートにインポートします。

1. インポート情報の生成
   「ノード情報定義ファイル」「FXサーバ用ノード情報定義ファイル」をFEFSデザインシートへ取り込むため、ファイル形式を変換します。
   ノード情報定義ファイルおよびFXサーバ用ノード情報定義ファイルが配置されたノード上で、以下のコマンドを実行してください。

```
# /sbin/fefs_yaml2csv <ノード情報定義ファイル> <出力ファイル>
# /sbin/fefs_yaml2csv <FXサーバ用ノード情報定義ファイル> <出力ファイル>
```

   ノード情報定義ファイルおよびFXサーバ用ノード情報定義ファイルは複数ファイルに分割されている場合があります。その場合は1ファイルごとにファイルの数だけ本コマンドを実行してください。
   「ノード情報定義ファイル」「FXサーバ用ノード情報定義ファイル」の詳細は「ジョブ運用ソフトウェア 管理者向けガイド システム管理編」を参照してください。

2. インポートするファイルの選択
   生成したインポート情報をFEFSデザインシートへ取り込むには Excelマクロを使用します。

   Excel マクロの「FEFS Design」>「Import installer data」

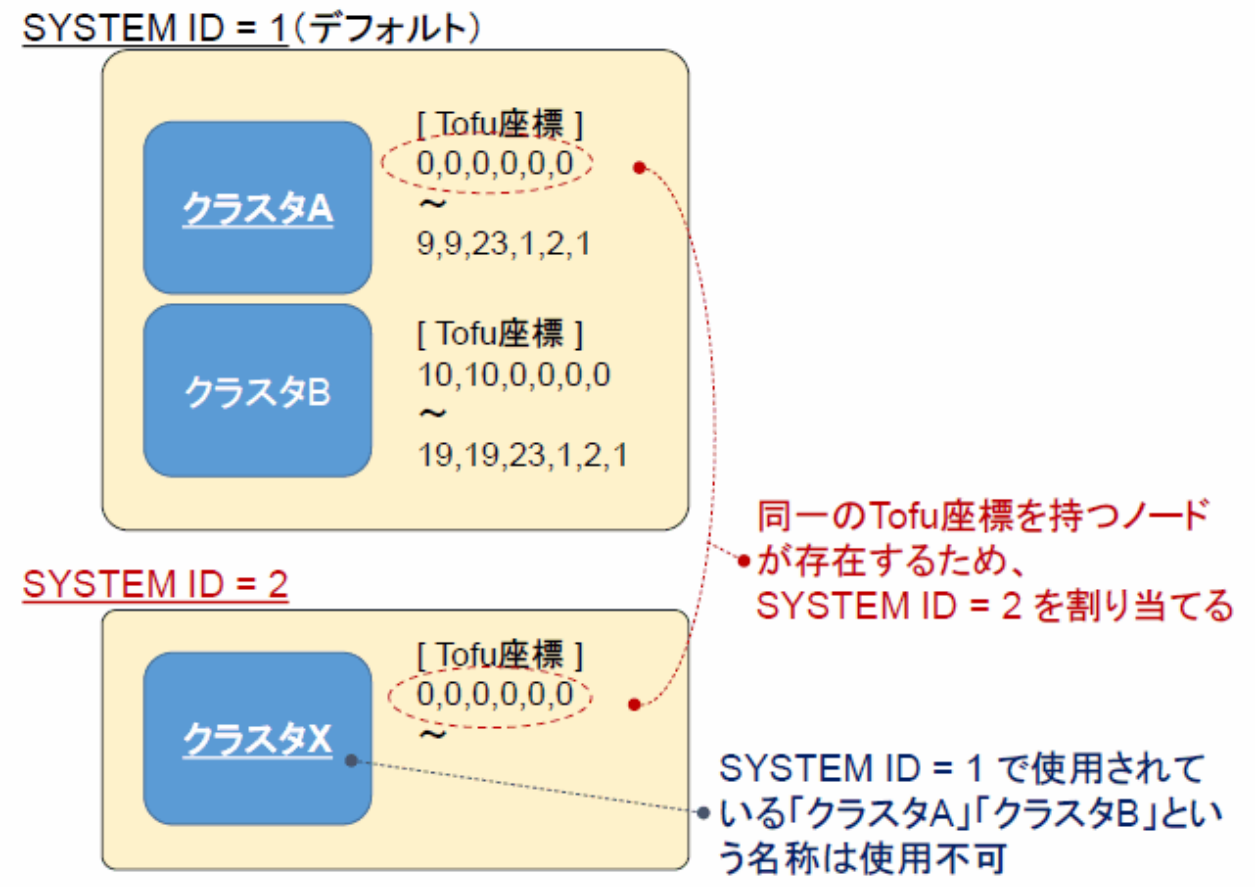
   ダイアログの出力に従って、1.で出力したファイルをインポートしてください。インポートが完了した場合は、"Import Complete" と表示されます。

   ※多目的ノードなど、FEFSを使用しないノードが含まれている場合は、当該ノードの定義をFEFSデザインシートから削除してください。

SYSTEM ID について

NODE(1) シートの SYSTEM ID 列は、通常は入力する必要はありません。同一のTofu座標を持つノードが存在する場合に使用してください。また、異なる SYSTEM ID 間で同一のクラスタ名を使用することはできません。

図3.5 SYSTEM ID の使用



3.1.3.2 LLIO シートの入力

LLIO を導入する場合は、LLIO のマウント情報およびストレージI/Oノード、第1階層ストレージデバイスの設定を行います。

1. LLIO SETTING セクション

図3.6 LLIO SETTING セクションの記入例

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

LLIO を利用する場合は、すべての項目に USE を指定してください。
LLIO を利用しない場合は初期値のままUNUSEとし、以降の FEFS デザインシートでの LLIO の設定は不要です。

2. LLIO セクション
以下の "LLIO セクションの記入例" に沿って説明します。

図3.7 LLIO セクションの記入例

■ 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]
   ジョブ内共有テンポラリ用のマウントポイントを指定してください。

   b. MOUNT OPTION [Shared temporary]
   ジョブ内共有テンポラリ用のマウントオプションを指定してください。
   ジョブ内共有テンポラリ用のマウントオプションを指定します。通常は変更する必要はありません。

   c. MOUNT POINT [Local temporary]
   ノード内テンポラリ用のマウントポイントを指定してください。

   d. MOUNT OPTION [Local temporary]
   ノード内テンポラリ用のマウントオプションを指定します。通常は変更する必要はありません。

   e. MOUNT OPTION [Global]
   ジョブ内第2階層ストレージのキャッシュのマウントオプションを指定します。通常は変更する必要はありません。

マウントポイントなしにマウントオプションを定義することはできません。定義が適切に行われている場合は、入力欄の右に "Configured" が入力されます。

3. SIO セクション
以下の "SIO セクションの記入例" に沿って説明します。

図3.8 SIO セクションの記入例

■ 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
   ストレージI/Oノードのホスト名を記述します。

   b. SSD VOLUME
   当該ストレージI/Oノードに接続された第1階層ストレージデバイスのパスを定義します。

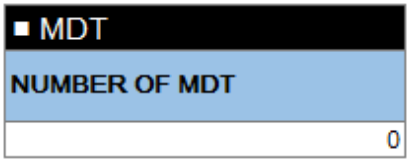
c. SSD SIZE (Mib)
第1階層ストレージが利用するデバイスサイズを Mib 単位で指定します。

d. MKFS OPTION
第1階層ストレージのmkfsオプションを指定します。通常は変更する必要はありません。

e. MOUNT OPTION
第1階層ストレージデバイスのマウントオプションを指定します。通常は変更する必要はありません。

4. MDT セクション
以下の "MDT セクションの記入例" に沿って説明します。

図3.9 MDTセクションの記入例



a. NUMBER OF MDT
共有テンポラリ向けに専用の MDT を用意している場合、専用の MDT 数を指定してください。
共有テンポラリ向けに専用の MDT を用意していない場合は、0 を指定してください。(デフォルト0)

参照

本項に出てくる LLIO 関連用語について、詳細は「LLIO ユーザーズガイド」を参照してください。

3.1.3.3 GFS シートの入力

GFS シートでは、グローバルファイルシステムの設定を行います。1シートにつき 1ファイルシステムの設定を行います。複数のファイルシステムを設定する場合は、ファイルシステムの数だけシートを追加して設定してください。

1. GFS シートの追加
GFS シートは、Excel のマクロで追加します。

Excelマクロの「FEFS Design」>「Insert global filesystem sheet」

2. FILESYSTEM セクション
以下の FILESYSTEM セクションの記入例に沿って説明します。

図3.10 FILESYSTEM セクションの記入例

■ FILESYSTEM	
FSNAME	fefs01
MOUNT POINT [FEFS]	/fefs01

a. FSNAME
ファイルシステム名を設定してください。

注意

システムでユニークである必要があります。また、ファイルシステム名は半角英数字 8文字以内である必要があります。

b. MOUNT POINT
クライアントのマウントポイントを指定してください。絶対パスで指定してください。

3. MGS セクション
   以下の "MGS セクションの記入例" に沿って説明します。
   本入力欄のMGS HOSTNAME (Active)、MGT VOLUME は必須の入力項目です。

図3.11 MGS セクションの記入例

■ MGS

MGS HOSTNAME (Active)	MGS HOSTNAME (Standby)	-	MGT VOLUME	-	MKFS OPTION	MOUNT OPTION
mgs		-	/dev/disk/by-id/scsi-3600000e00d1100000011129ae00000000	-	--reformat --verbose	defaults,retry=6

   a. MGS HOSTNAME (Active)
      MGT をマウントする運用系 MGS のノード名を指定してください。

      注意

      MGS を MDS と兼用する場合は、MDS セクション index0 の MDS HOSTNAME (Active) に同じノード名を指定してください。

   b. MGS HOSTNAME (Standby)
      MGSが HA 構成の場合は、対応する待機系 MGS のノード名を指定してください。

      注意

      MGS を MDS と兼用する場合は、MDS セクション index0 の MDS HOSTNAME (Standby) に同じノード名を指定してください。

   c. MGT VOLUME
      MGTのボリュームを指定してください。

      注意

      MGT を MDT と兼用する場合は、MDS セクション MDT VOLUME の index0 に MGT と同じボリューム名を指定してください。

   d. MKFS OPTION
      通常は変更する必要はありません。

   e. MOUNT OPTION
      通常は変更する必要はありません。

4. MDS セクション
   以下の "MDS セクションの記入例" に沿って説明します。
   本入力欄index0のMDS HOSTNAME(Active)、MDT VOLUME は必須の入力項目です。

図3.12 MDS セクションの記入例

■ 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-3600000e00d1100000011286800001000			
mds3	mds4	2	/dev/disk/by-id/scsi-3600000e00d1100000011286800002000			
mds4	mds3	3	/dev/disk/by-id/scsi-3600000e00d1100000011286800003000			

   a. MDS HOSTNAME(Active)
      通常運用時にMDTをマウントする運用系 MDSのノード名を指定してください。

      注意

      MGS をMDS と兼用する場合は、MDS HOSTNAME(Active)の index0 に MGS HOSTNAME(Active) と同じノード名を指定してください。

   b. MDS HOSTNAME(Standby)
      MDS が HA構成の場合は、対応する待機系 MDS のノード名を指定してください。

注意

MGS を MDS と兼用する場合は、MDS HOSTNAME(Standby) の index0 に MGS HOSTNAME(Standby) と同じノード名を指定してください。

c. MDT VOLUME
MDTのボリュームを指定してください。

注意

MGT を MDT と兼用する場合は、MDT VOLUME の index0 に MGT と同じボリューム名を指定してください。

d. JOURNAL VOLUME
外部ジャーナルを利用する場合は、ジャーナル用ボリュームを用意し、指定してください。

e. MKFS OPTION
通常は変更する必要はありません。

f. MOUNT OPTION
通常は変更する必要はありません。
以下の場合にオプションを追加してください。

- QoS 機能を有効にする場合
  QoS 定義ファイルのパスを追加します。QoS 機能を有効にする設定についての詳細は、"3.2.1 QoS 機能の有効化"を参照してください。

- ACL 機能を有効にする場合
  acl オプションを追加します。ACL 機能を有効にする設定についての詳細は、"3.4 ACL 機能を有効にする設定"を参照してください。

5. OSS セクション
以下の "OSS セクションの記入例" に沿って説明します。

図3.13 OSS セクションの記入例

■ OSS

OSS HOSTNAME (Active)	OSS HOSTNAME (Standby)	OST INDEX	OST VOLUME	JOURNAL VOLUME	MKFS OPTION	MOUNT OPTION
oss1	oss2	0	/dev/disk/by-id/scsi-3600000e00d110000001129ae00020000		--reformat --verbose	defaults,retry=6
oss2	oss1	1	/dev/disk/by-id/scsi-3600000e00d110000001129ae00021000			
oss3	oss4	2	/dev/disk/by-id/scsi-3600000e00d110000001129ae00022000			
oss4	oss3	3	/dev/disk/by-id/scsi-3600000e00d110000001129ae00023000			

a. OSS HOSTNAME(Active)
通常運用時に OST をマウントする運用系 OSS のノード名を指定してください。

b. OSS HOSTNAME(Standby)
OSSがHA構成の場合は、対応する待機系 OSS のノード名を指定してください。

c. OST VOLUME
OST のボリュームを指定してください。

d. JOURNAL VOLUME
外部ジャーナルを利用する場合は、ジャーナル用ボリュームを用意し、指定してください。

6. CLIENT セクション
以下の "CLIENT セクションの記入例" に沿って説明します。

図3.14 CLIENT セクションの記入例

■ CLIENT

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

各ノードのマウントオプションを指定します。通常は変更する必要はありません。
以下の場合にオプションを追加してください。

- QoS 機能を有効にする場合
  qos オプションを追加します。QoS 機能を有効にする設定についての詳細は、"3.2 QoS 機能を有効にする設定" を参照してください。
- user 拡張属性を有効にする場合
  user_xattr オプションを追加します。user 拡張属性を有効にする設定についての詳細は、"3.5 user 拡張属性を有効にする設定" を参照してください。

兼用ノード、多目的ノードのクライアントマウントオプション設定ルール
オプションを設定する場合は、以下のルールで設定してください。

- CCS(クライアントと兼用する場合) : CLIENT OPTION (PG Client) に設定してください。
- 多目的ノード : CLIENT OPTION (LN) に設定してください。

7. FX CLIENT セクション
   以下の "FX CLIENTセクションの記入例" に沿って説明します。

図3.15 FX CLIENT セクションの記入例

■ FX CLIENT

SYSTEM ID	Tofu Coord	
	X	Y
	0	0
	0	1

   a. SYSTEM ID
      FXシステムが複数存在する場合は、システム番号を入力してください。1システムの場合は入力不要です。

   b. TOFU Coord
      FXにおいてクライアントをマウントする本体装置のTOFU座標(X,Y)の一覧を入力してください。

8. PG CLIENT セクション
   PRIMERGY においてクライアントをマウントするノード名の一覧を指定してください。

図3.16 PG CLIENT セクションの記入例

■ PG CLIENT

HOSTNAME
client1
client2
client3
login1
login2

3.1.3.4 入力データのチェック

以下の Excel のマクロで入力データをチェックできます。入力データに不備がないか確認してください。

Excel マクロの「FEFS Design」>「Check」

不備がある場合は該当箇所が通知されます。該当箇所を修正して、再度確認してください。問題なければ、"OK: Check completed." と表示されます。

3.1.4 FEFSセットアップツール用構成定義ファイルの作成

以下の Excel のマクロでセットアップ用の入力データを作成できます。

Excelマクロの「FEFS Design」>「Create config files」

ダイアログに従って出力先フォルダを指定してください。指定したフォルダ配下に FEFS セットアップツール用構成定義ファイルが作成されます。

注意

LLIO を利用している場合は、この操作によって、LLIO の構成定義ファイルも作成されます。

参考

FEFS セットアップツール用構成定義ファイルから FEFS デザインシートを作成することができます(FEFSセットアップツール用構成定義ファイルのインポート機能)。

製品に同梱された FEFS デザインシートのひな形を使用し、以下の Excel マクロを実行してください。

Excelマクロの「FEFS Design」>「Import config files」

ダイアログの表示に従って、FEFS セットアップツール用構成定義ファイルを出力したフォルダ(「Create config files」に指定したフォルダ)を指定してください。FEFS デザインシートに FEFS セットアップツール用構成定義ファイルがインポートされます。

インポート時は以下の点に注意してください。

- インポートは白紙の FEFS デザインシートへ行います。すでに設定が入力されているFEFS デザインシートへ情報を追加することはできません。
- 複数のファイルシステムの設定が存在する場合、FEFSデザインシートへインポートする際の順番は任意です。

3.1.5 FEFSセットアップツール用構成定義ファイルの配置

運用系および待機系システム管理ノード上の以下のディレクトリに、FEFS セットアップツール用構成定義ファイルを配置してください。

/etc/opt/FJSVfefs/config 配下

3.1.6 FEFSの構築

FEFS の構築には、fefs_sync コマンドを使います。fefs_sync コマンドの詳細は、"A.2.1 fefs_sync コマンド" を参照してください。

以降の操作を実施するためには、計算ノード以外の FEFS がインストールされているすべてのノードが起動されている必要があります。

また、以降のコマンド実施については、ジョブ運用ソフトウェアがインストールされている必要があります。

以下の操作を運用系システム管理ノード上で実施してください。

参考

fefs_sync コマンドでエラーとなったノードへの復旧手順は、"3.11 構築に失敗したノードの構築方法" を参照してください。
fefs_sync コマンドで多目的クラスタを指定する場合は、--compute オプションに指定してください。

1. FEFS 設定ファイルの作成

   以下を実行してください。

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

   FEFS を構築するすべてのクラスタを指定してください。

--storage : ストレージクラスタ名を指定してください。
--compute : 計算クラスタ名および多目的クラスタ名を指定してください。

2. MGT、MDT、OST の初期化フォーマット

MGS、MDS、OSS のボリュームフォーマットを行います。

フォーマットする必要がある場合は、以下を実行してください。

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

FEFS を構築するすべてのクラスタを指定してください。

--storage : ストレージクラスタ名を指定してください。

3. FEFSサービスの起動

FEFSのサービスを起動し、MGS、MDS、OSS、CN、CCM、LN のマウントを行います。

以下を実行してください。

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

FEFS を構築するすべてのクラスタを指定してください。

--storage : ストレージクラスタ名を指定してください。
--compute : 計算クラスタ名および多目的クラスタ名を指定してください。

注意

LLIO を使用する構成の場合、CN でのマウントはサービスを起動したタイミングでは行われません。詳細については「LLIO ユーザーズガイド」を参照してください。

4. FEFS状態の確認

MGS、MDS、OSS、CN、CCM、LN において、FEFS のサービスが正常に起動されたことを pashowclst コマンドで確認してください。

以下を実行してください。

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

FEFS の状態が FEFSSR(o)および FEFS(o) に遷移していれば、FEFS のサービスは正常に起動されています。

3.1.7 ファイルシステムのパーミッション変更

1クライアント上でマウントポイントのパーミッションを設定してください。

ファイルシステムがマウントされている状態で行ってください(初期値は755です)。

3.1.8 構築後に必要な設定

FEFS のトラブルを事前に抑止するために、スクリプトを設定する必要があります。

"付録C FEFS の構築後に必要な設定" の手順を実行してください。

3.1.9 計算ノードの追加設定

ハードウェア故障などの理由により、ファイルシステム構築時に起動できなかった計算ノードがあった場合、あとで構築できます。運用系システム管理ノードで以下の手順を実施してください。

1. 該当ノードの運用切離し

対象となる計算ノードの設定を、非運用状態にします。

以下を実行してください。

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

<*cluster*>：ノードが属するクラスタを指定してください。
<*nodeid*>：対象のノードID を指定してください。

2. 該当ノードのノード起動

対象となる計算ノードを起動します。

以下を実行してください。

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

<*cluster*>：ノードが属するクラスタを指定してください。
<*nodeid*>：対象のノードID を指定してください。

3. FEFS 設定ファイルの作成

FEFS 設定ファイルを作成します。

以下を実行してください。

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

FEFS を構築するすべてのクラスタを指定してください。

--storage：ストレージクラスタ名を指定してください。
--compute：計算クラスタ名および多目的クラスタ名を指定してください。

4. 該当ノードのノード再起動

対象となる計算ノードを再起動します。

以下を実行してください。

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

<*cluster*>: ノードが属するクラスタを指定してください。
<*nodeid*>: 対象のノードIDを指定してください。

5. 該当ノードの運用組込み

対象となる計算ノードの設定を、運用状態にします。

以下を実行してください。

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

<*cluster*>：ノードが属するクラスタを指定してください。
<*nodeid*>：対象のノードID を指定してください。

3.1.10 ノード単位の構築方法

1ノードずつ、または一部のノードごとに構築を行いたい場合は、以下の手順で行ってください。

1. FEFS デザインシートの作成

"3.1.3 FEFS デザインシートの作成" を参照して作業を進めてください。

2. FEFS セットアップツール用構成定義ファイルの作成

Excel のマクロでセットアップ用の入力データを作成できます。

Excelマクロの「FEFS Design」>「Create config files」

3. FEFS セットアップツール用構成定義ファイルの配置

運用系および待機系システム管理ノード上の以下のディレクトリに、FEFS セットアップツール用構成定義ファイルを配置してください。

/etc/opt/FJSVfefs/config 配下

4. FEFS設定ファイルの作成

   システム管理ノードで、以下を実行してください。

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

   --storage : ストレージクラスタ名を指定してください。
   --compute : 計算クラスタ名を指定してください。
   --nodelist : 対象ノードのノードIDを列挙したファイルを指定してください。

5. ボリュームフォーマット

   構築ノードが、MGS、MDS、OSS の場合は、本手順を実施します。

   ボリュームのフォーマットを行う必要がある場合は、システム管理ノードで以下を実行してください。

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

   --storage : ストレージクラスタ名を指定してください。構築ノードが MGS、MDS、OSSの場合に指定します。
   --nodelist : ノードIDを列挙したファイルを指定してください。

6. FEFS の起動

   a. FEFS の起動
      システム管理ノードで以下を実行してください。

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

      --storage : ストレージクラスタ名を指定してください。
      --compute : 計算クラスタ名を指定してください。
      --nodelist : ノードIDを列挙したファイルを指定してください。

   b. FEFS 状態の確認
      ノードの FEFS サービスが正常に起動されたことを pashowclst コマンドで確認してください。

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

      ノードが属するクラスタとノードID を指定してください。

      <*cluster*> :ノードが属するクラスタを指定してください。
      <*nodeid*> : 対象のノードID を指定してください。

      FEFS の状態が FEFSSR(o)およびFEFS(o) に遷移していれば、FEFS のサービスは正常に起動されています。

参照

ノードの指定は、--nodeid オプションを指定することも可能です。詳細は "A.2.1 fefs_sync コマンド" を参照してください。

3.2 QoS 機能を有効にする設定

ここでは、QoS 機能を有効にする設定について説明します。

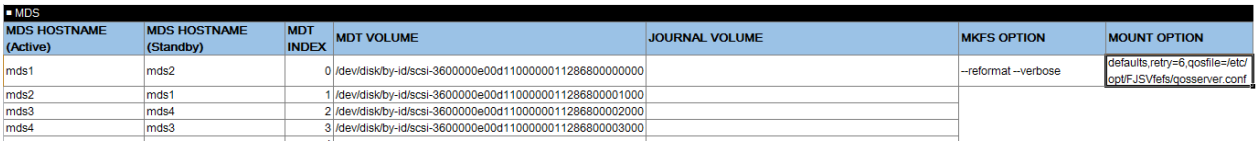
3.2.1 QoS 機能の有効化

QoS機能は、FEFS デザインシート作成時に、MDS および FEFSクライアントにQoS 機能を設定することで有効になります。

MDS の設定方法

GFS シートの MDS セクションの MOUNT OPTION の欄に、QoS 定義ファイルのパスを記述します。これにより、QoS機能が有効になります。QoS 定義ファイルの作成方法については、"3.2.2 QoS 定義ファイルの設定" を参照してください。

図3.17 QoS機能を有効にする場合(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-3600000e00d11000000112868000000000		--reformat --verbose	defaults,retry=6,qosfile=/etc/opt/FJSVfefs/qosserver.conf
mds2	mds1	1	/dev/disk/by-id/scsi-3600000e00d11000000112868000001000			
mds3	mds4	2	/dev/disk/by-id/scsi-3600000e00d11000000112868000002000			
mds4	mds3	3	/dev/disk/by-id/scsi-3600000e00d11000000112868000003000			

FEFS デザインシート作成後の手順については、"3.1.4 FEFSセットアップツール用構成定義ファイルの作成" 以下を参照してください。

参考

以下の手順で、MDS のマウント後に QoS 機能を有効にできますが、FEFS デザインシート作成時に有効にする手順を推奨します。

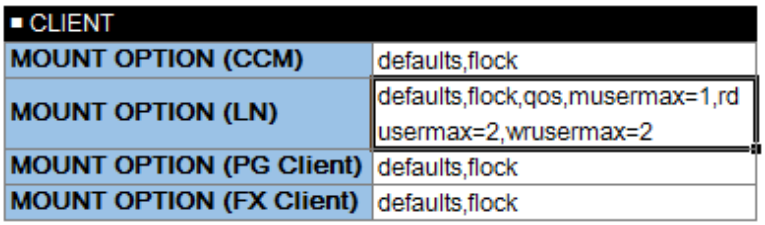
```
[MDSノード]
# lctl qos on /etc/opt/FJSVfefs/qosserver.conf
```

lctl qos onコマンドの詳細は、"A.2.7 lctlコマンド" のサブコマンド "lctl qos" を参照してください。

FEFSクライアントの設定方法

FEFS デザインシート作成時に、GFSシートの CLIENT セクションで設定します。MOUNT OPTION の欄に、qos または qos_cache オプションを指定することで、クライアントの QoS 機能が有効になります。クライアントの QoS 機能は、多数のユーザーで共用するクライアント (ログインノード) だけに設定することを推奨します。

図3.18 QoS機能を有効にする場合(FEFSクライアント)



■ CLIENT

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

FEFS デザインシート作成後の手順については、"3.1.4 FEFSセットアップツール用構成定義ファイルの作成" 以下を参照してください。

FEFS クライアントでは、以下の qos オプションを設定できます。

表3.5 FEFS クライアントで設定できる qos オプション一覧

オプション	説明
qos	クライアントノード上の QoS 機能 (リクエスト制御) を有効にします。 qos_cacheオプションと同時に指定できます。
qos_cache	クライアントノード上の QoS 機能 (キャッシュ制御) を有効にします。 qosオプションと同時に指定できます。
noqos	クライアントノード上の QoS 機能を無効にします (デフォルト)。
musermax=<数値>	MDS に対して、一般ユーザーが 1ユーザーあたりに同時発行可能なリクエスト数を指定します。 数値の指定可能範囲は 1 から 8 です。省略時は 1 です。 qos オプションと同時に指定してください。
rdusermax=<数値>	OSS に対して、一般ユーザーが 1ユーザーあたりに同時発行可能な read リクエスト数を指定します。 数値の指定可能範囲は 1 から 16 です。省略時は 2 です。 qos オプションと同時に指定してください。
wrusermax=<数値>	OSS に対して、一般ユーザーが 1ユーザーあたりに同時発行可能な write リクエスト数を指定します。 数値の指定可能範囲は 1 から 16 です。省略時は 2 です。

オプション	説明
	qos オプションと同時に指定してください。
mrootmax=<数値>	MDS に対して、root ユーザーが同時発行可能なリクエスト数を指定します。 数値の指定可能範囲は 1 から 8 です。省略時は 1 です。 qos オプションと同時に指定してください。
rdrootmax=<数値>	OSS に対して、root ユーザーが同時発行可能な read リクエスト数を指定します。 数値の指定可能範囲は 1 から 16 です。省略時は 2 です。 qos オプションと同時に指定してください。
wrrootmax=<数値>	OSS に対して、root ユーザーが同時発行可能な write リクエスト数を指定します。 数値の指定可能範囲は 1 から 16 です。省略時は 2 です。 qos オプションと同時に指定してください。
mclientmax=<数値>	MDS に対して、クライアントノード内で同時発行可能なリクエスト数を指定します。 数値の指定可能範囲は 1 から 16 です。省略時は 4 です。 qos オプションと同時に指定してください。
rdclientmax=<数値>	OSS に対して、クライアントノード内で同時発行可能な read リクエスト数を指定します。 数値の指定可能範囲は 1 から 32 です。省略時は 8 です。 qos オプションと同時に指定してください。
wrclientmax=<数値>	OSS に対して、クライアントノード内で同時発行可能な write リクエスト数を指定します。 数値の指定可能範囲は 1 から 32 です。省略時は 8 です。 qos オプションと同時に指定してください。
dpusermax=<数値>	クライアントノード上のクライアントキャッシュについて、一般ユーザーが 1ユーザーあたりに使用可能な割合 (%) を指定します。数値の指定可能範囲は 1 から 100 です。省略値は 10 です。 qos_cache オプションと同時に指定してください。
dprootmax=<数値>	クライアントノード上のクライアントキャッシュについて、rootユーザーが使用可能な割合 (%) を指定します。数値の指定可能範囲は 1 から 100 です。省略値は 10 です。 qos_cache オプションと同時に指定してください。

3.2.2 QoS 定義ファイルの設定

QoS 機能を利用するには、QoS定義ファイル /etc/opt/FJSVfefs/qosserver.conf を作成する必要があります。この作業は、root 権限を持つ管理者が MDS 上で行ってください。マルチ MDS 環境の場合は、MDT0 上で行ってください。

参考

QoS定義ファイルのサンプルとして、/etc/opt/FJSVfefs/qosserver.conf.sample があります。

1. QoS 定義ファイルの作成
   QoS 定義ファイルを作成します。
   以下は、QoS定義ファイルの書式です。

```
MDS{
  項目名=設定値
  項目名=設定値
    (中略)
}
OSS{
  項目名=設定値
```

```
  項目名=設定値
    （中略）
}
```

注意

- MDS、OSSセクションは記述が必須です。
- 1行の文字数は、空白・改行文字も含めて 1024文字以下にしてください。
- シャープ (#) から始まる行はコメント行です。行の途中からのコメントは記述できません。

以下は、QoS 定義ファイルの設定例です。

例1 : ユーザー間のフェアシェア

```
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
}
```

例2 : ログインノードの優先制御

```
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
}
```

以下は設定項目の詳細です。

表3.6 MDSセクションに記述可能な項目

指定項目	説明	省略時の動作
qos={on | off }	on : MDS で QoS 制御を行います。 off : MDS で QoS 制御を行いません。	必須パラメーターのため、省略できません。
nodegrp[1-10]=<*数値1*>%(<*数値2*>%) [*ipアドレス群*] (注)	ノード群に割り当てるサーバスレッド数の最大値 (割合)を指定します。<*数値1*> で指定した割合までサーバスレッドを割り当てます。<*数値1*> の指定範囲は 1 から 100 です。 nodegrp1 から nodegrp10 までの <*数値1*> の合計は、100 以下にする必要があります。 サーバスレッドに空きがある場合は、<*数値2*> で指定した割合までサーバスレッドを割り当てます。<*数値2*> の指定範囲は <*数値1*> 以上かつ 100 以下です。<*数値2*> を省略した場合は、<*数値1*> と同じ値を指定したものとします。こ	必須パラメーターのため、省略できません。1つ以上の nodegrp の指定が必要です。

指定項目	説明	省略時の動作
	の場合は、サーバスレッドに空きがあっても <*数値1*> で指定した割合までしかスレッドを割り当てません。 [*ipアドレス群*]にはFEFSクライアントの I/O 用インターコネクトの IPアドレスを指定します。	
usermax[1-10]=<*数値1*>%(<*数値2*>%)	nodegrp で定義したノード群内で、一般ユーザー 1ユーザーあたりに割り当て可能なサーバスレッド数の最大値 (割合) を指定します。<*数値1*> の指定範囲は 1 から 100 です。 サーバスレッドに空きがある場合は、<*数値2*> で指定した割合までサーバスレッドを割り当てます。<*数値2*> の指定範囲は、<*数値1*> 以上かつ 100 以下です。<*数値2*>を省略した場合は、<*数値1*> と同じ値を指定したものとします。この場合は、サーバスレッドに空きがあっても <*数値1*> で指定した割合までしかスレッドを割り当てません。	100% を指定したものとします。
rootmax[1-10]=<*数値1*>%(<数値2>%)	nodegrp で定義したノード群内で、root ユーザーに割り当て可能なサーバスレッド数の最大値 (割合) を指定します。<*数値1*> の指定範囲は 1 から 100 です。 サーバスレッドに空きがある場合は、<*数値2*> で指定した割合までサーバスレッドを割り当てます。<*数値2*> の指定範囲は <*数値1*> 以上かつ 100 以下です。<*数値2*> を省略した場合は、<*数値1*> と同じ値を指定したものとします。この場合は、サーバスレッドに空きがあっても <*数値1*> で指定した割合までしかスレッドを割り当てません。	100% を指定したものとします。

(注) nodegrp の *ipアドレス群* は、以下の書式で記述してください。

nodegrp の ipアドレス群の書式

```
<ipアドレス群> ::= <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"
```

以下は、nodegrp の定義例です。

例1: IPアドレスを1つずつ指定する場合

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

例2: IPアドレスを範囲指定する場合

```
nodegrp1 = 30% 192.0.2.[10-15]  -> 192.0.2.[10,11,12,13,14,15] と同じ意味になります。
nodegrp2 = 70% 198.51.100.*     -> 198.51.100.[0-255] と同じ意味になります。
```

例3: IPアドレスを一定間隔で範囲指定する場合

```
nodegrp1 = 30% 192.0.2.[10-20/3]        -> 192.0.2.[10,13,16,19] と同じ意味になります。
nodegrp2 = 70% 198.51.100.[50-100/10]   -> 198.51.100.[50,60,70,80,90,100] と同じ意味になります。
```

注意

nodegrp1 から nodegrp10 で重複した IPアドレスを定義した場合、nodegrp 番号の小さい方が優先されます。

例えば、以下の定義では、192.0.2.35 からの要求は、nodegrp1 として QoS 制御されます。

誤った定義例

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

192.0.2.35 からの要求を正しく QoS 制御するためには、以下のように記述します。

正しい定義例

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

以下の定義では、FEFS のすべてのクライアントからの要求は、nodegrp1 として QoS 制御されます。

誤った定義例

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

192.0.2.* からの要求を正しくQoS 制御するためには、以下のように記述します。

正しい定義例

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

表3.7 OSSセクションに記述可能な項目

指定項目	説明	省略時の動作
qos={on | off | same_mds }	on : OSS で QoS 制御を行います。 off : OSS で QoS 制御を行いません。 same_mds : MDS セクションの定義を使用します (OSS セクションの定義は無効となります)。ただし、OSS セクションでしか指定できないパラメーター (load_limit_usec) については、OSS セクションの定義を使用します。	必須パラメーターのため、省略できません。
nodegrp[1-10]=<*数値1*>%(<*数値2*>%) *[ipアドレス群]* (注)	MDS セクションと同様のため、MDS セクションの説明を参照してください。	必須パラメーターのため、省略できません。1つ以上の nodegrp の指定が必要です。
usermax[1-10]=<*数値1*>%(<*数値2*>%)	MDS セクションと同様のため、MDS セクションの説明を参照してください。	100% を指定したものとします。
rootmax[1-10]=<*数値1*>%(<*数値2*>%)	MDS セクションと同様のため、MDS セクションの説明を参照してください。	100% を指定したものとします。
load_limit_usec=<*数値*>	OST への 1回あたりのディスクアクセス時間の上限値を指定します。単位はマイクロ秒です。 ディスクアクセス時間が、このパラメーターで指定された値を超える場合は、usermax または rootmax で指定したサーバスレッド数の割合よりも、小さな値で QoS 制御を行います。 数値の指定可能範囲は 0 から 100000000 です。 0 を指定した場合は、I/O アクセス時間によるスレッド数制御は行われません。	0(無効) を指定したものとします。

(注) nodegrp のipアドレス群 は、前述の "nodegrp の ipアドレス群の書式" を参照して記述してください。

2. QoS定義ファイルの確認

作成したQoS 定義ファイルの構文に誤りがないか確認します。QoS 定義ファイルの確認は、MDS 上で、lctl qos check コマンドで行います。

例1 : 構文が正しい場合

```
[MDSノード]
# lctl qos check /etc/opt/FJSVfefs/qosserver.conf
QoS command was completed.
```

例2：構文が誤っている場合

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

lctl qos checkコマンドの詳細は、"A.2.7 lctlコマンド" のサブコマンド "lctl qos" を参照してください。

3. QoS 機能の有効化

   QoS 定義ファイルの構文が正しいことを確認したあと、"3.2.1 QoS 機能の有効化" の "MDS の設定方法" のとおりに QoS 機能を有効にしてください。

3.3 ファイルロックを有効にする設定

FEFS には、fcntl システムコールまたは flock システムコールによる勧告ロック機能があります。

ファイルロックを使用する場合は、GFSシートの CLIENT セクションの MOUNT OPTION の欄に、flock オプションを記述します。

図3.19 ファイルロックを有効にする場合

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

FEFS デザインシート作成後の手順については、"3.1.4 FEFSセットアップツール用構成定義ファイルの作成" 以下を参照してください。

3.4 ACL 機能を有効にする設定

FEFS デザインシート作成時にすべてのファイルシステムに適用したい場合は、GFS シートの MDS セクションの MOUNT OPTION の欄に、acl オプションを記述します。

図3.20 ACL機能を有効にする場合

■ 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-3600000e00d11000000112868000000000		--reformat --verbose	defaults,retry=6,acl
mds2	mds1	1	/dev/disk/by-id/scsi-3600000e00d11000000112868000001000			
mds3	mds4	2	/dev/disk/by-id/scsi-3600000e00d11000000112868000002000			
mds4	mds3	3	/dev/disk/by-id/scsi-3600000e00d11000000112868000003000			

FEFS デザインシート作成後の手順については、"3.1.4 FEFSセットアップツール用構成定義ファイルの作成" 以下を参照してください。

3.5 user 拡張属性を有効にする設定

user 拡張属性とは、システムコール setxattr およびシステムコール getxattr で設定・参照する情報で、名前空間を表す接頭辞が "user." の拡張属性です。user 拡張属性はデフォルトでは使用できません。user 拡張属性を使用する場合は、FEFS デザインシート作成時に、GFS シートの CLIENT セクションの MOUNT OPTION の欄に user_xattr オプションを指定します。

図3.21 user拡張属性を有効にする場合

■ 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

FEFS デザインシート作成後の手順については、"3.1.4 FEFSセットアップツール用構成定義ファイルの作成" 以下を参照してください。

3.6 フェイルオーバ機能を利用する場合の設定

MGS、MDS、およびOSS においてフェイルオーバ機能を利用する場合は、ノードがパニックしたときに自動的に再起動しないように設定を行う必要があります。

設定方法は、富士通Linuxサポートパッケージのダンプ支援ツールを導入している場合は、そのドキュメントを参照してください。

導入していない場合は、Red Hat 社が公開している「カーネルクラッシュダンプガイド」を参照してください。

3.7 保守時の操作

FEFSが構築されている環境において、モジュールのパラメーターを変更する場合は、一度 FEFS を停止し、モジュールをアンロードする必要があります。

FEFS の停止は、以下の順序で行ってください。

1. 事前準備
   「運用からの切離し」と「ソフトウェアメンテナンスモードへの移行」を実施します。詳細は、「ジョブ運用ソフトウェア 管理者向けガイド 保守編」の「ソフトウェア保守の事前準備」を参照してください。
2. クライアントの停止
   すべてのクライアント上で、以下を実行してください。

   ```
   # systemctl stop FJSVfefs
   ```
3. OSS の停止
   すべての OSS 上で以下を実行してください。

   ```
   # systemctl stop FJSVfefs
   ```
4. MDSの停止
   すべての MDS 上で以下を実行してください。

   ```
   # systemctl stop FJSVfefs
   ```
5. MGSの停止
   すべての MGS 上で以下を実行してください。

   ```
   # systemctl stop FJSVfefs
   ```

   モジュールをアンロード後、FEFSのパラメーターを変更し、FEFS を再起動してください。この際、すでにファイルシステムのフォーマットが完了している場合は、再度フォーマットを行う必要はありません。
6. 運用組込み
   事前準備で運用から切り離していた保守対象を運用に組み込みます。詳細は、「ジョブ運用ソフトウェア 管理者向けガイド 保守編」の「ソフトウェア保守後の運用組み込み」を参照してください。

注意

- MGS兼MDS構成の場合は、MGT をマウントするノードを最後に停止してください。

- ファイルシステムに不整合が生じた際、修復を行うときにもFEFSの停止が必要です。ファイルシステム修復についての詳細は、"4.8 ファイルシステム不整合の修復" を参照してください。
- FEFS の停止は、pasnap を実行中であれば完了後に行ってください。pasnap を実行中に FEFS の停止を行うと、ノードがパニックすることがあります。

3.8 ローリングアップデート

システム全体を停止せずにファイルシステムのパッケージを更新できます。
パッケージ適用は、以下の手順で行ってください。
FEFSのパッケージの種類については "3.1.2 FEFS パッケージの適用" を参照してください。
LLIOのパッケージの種類については「LLIO ユーザーズガイド」を参照してください。

注意

パッケージには、ローリングアップデートの可否情報が記載されています。適用済のパッケージに対しては、rpm -qi コマンドで、適用予定のパッケージに対しては、rpm -qpi コマンドで事前に作業の可否・条件などを確認しておいてください。詳細は「ジョブ運用ソフトウェア 管理者向けガイド 保守編」の「ローリングアップデートによるパッケージ適用」を参照してください。

事前準備

「運用からの切離し」と「ソフトウェアメンテナンスモードへの移行」を実施します。詳細は、「ジョブ運用ソフトウェア 管理者向けガイド 保守編」の「ソフトウェア保守の事前準備」を参照してください。

1. FEFS クライアントパッケージの適用

   操作は SIOグループ単位で行います。
   SIOグループの詳細は「ジョブ運用ソフトウェア 概説書」を参照してください。

   a. 対象範囲における FEFSサービスの停止
      システム管理ノードで以下を実行してください。

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

      --compute : 計算クラスタ名を指定してください。
      --nodeid : <*nodeid*> を含む SIO グループに対してコマンドが実行されます。

   b. 保守・適用作業
      a. で停止した範囲において パッケージ適用を行ってください。

   c. 対象範囲における FEFSサービスの起動
      システム管理ノードで以下を実行してください。

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

      --compute : 計算クラスタ名を指定してください。
      --nodeid : <*nodeid*> を含む SIO グループに対してコマンドが実行されます。

   ### 注意

   適用対象パッケージに LLIO パッケージが含まれる場合は、LLIO パッケージと FEFS クライアントパッケージとを上記の手順でまとめて適用できます。LLIO パッケージのみを適用する場合は、「LLIOユーザーズガイド」を参照してください。

2. FEFS サーバパッケージの適用

   複数ファイルシステムを構築している場合、保守をファイルシステム単位に行えます。

a. クライアントにおけるファイルシステムのアンマウント
保守の対象となるファイルシステムをクライアントからアンマウントします。
システム管理ノードで以下を実行してください。

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

--compute : 計算クラスタ名を指定してください。
--fsname : 保守対象となるファイルシステムのファイルシステム名を指定してください。

b. 保守対象のサーバ停止
保守対象のサーバで FEFSサービスを停止します。
システム管理ノードで以下を実行してください。

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

--storage : ストレージクラスタ名を指定してください。
--nodelist : ノードIDが列挙されたファイルを指定します。保守対象のサーバのノードIDを列挙して指定してください。

c. 保守・適用作業
b. で停止した範囲において パッケージ適用を行ってください。

d. 保守対象のサーバの起動
保守対象のサーバで FEFSサービスを起動します。
システム管理ノードで以下を実行してください。

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

--storage : ストレージクラスタ名を指定してください。
--nodelist : ノードIDが列挙されたファイルを指定します。保守対象のサーバのノードIDを列挙して指定してください。

e. クライアントにおけるファイルシステムのマウント
アンマウントしたファイルシステムを再マウントします。
システム管理ノードで以下を実行してください。

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

--compute : 計算クラスタ名を指定してください。
--fsname : 保守対象となるファイルシステムのファイルシステム名を指定してください。

運用組込み

事前準備で運用から切り離していた保守対象を運用に組み込みます。詳細は、「ジョブ運用ソフトウェア 管理者向けガイド 保守編」の「ソフトウェア保守後の運用組み込み」を参照してください。

3.9 FEFS 統計情報可視化機能の設定

"2.9 FEFS 統計情報可視化機能 (fefssv.ph スクリプト)" で述べたように、fefssv.ph は collectl から呼び出して使うスクリプトです。fefssv.ph スクリプトは FEFS のパッケージに同梱されているため、特に設定を行う必要はありません。collectl のオプションで fefssv.ph スクリプトを指定すれば利用できます。

注意

collectl のパッケージは別途入手して MDS および OSS にインストールする必要があります。
以下のパッケージがインストールされている場合、collectl のログファイルは圧縮して書き出されます。

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

なお、collectl は、バージョン 4.3.0 だけをサポートしています。

3.10 NFS で公開する場合の設定

NFS で公開する NFS サーバのディレクトリ /etc/exports で FEFS を公開する設定をする際は、以下のオプション指定を含めてください。

表3.8 FEFS を公開する設定時に必要なオプション

オプション名	備考
fsid=num	num には 0 以外の 32bit の整数値を公開ポイントごとに異なる値で指定してください。設定する値は 1 以上の小さな値を推奨します。
no_root_squash	no_root_squash を使用してexport した場合、NFS クライアントから root 権限でファイルシステム内の資源にアクセスできます。このため NFS export の設定でマウントできるクライアントの範囲に制限をかけて、意図しないノードからマウントできないようにしてください。

また、ログインユーザーのアカウント情報(ユーザー名、グループ名、uid/gid など)は、NFS サーバと NFS クライアントで同じになるように設定してください。

3.11 構築に失敗したノードの構築方法

FEFS構築中に故障していた、または、エラーが発生したノードを個別に構築できます。構築対象のノードのノードIDをファイルに列挙し、fefs_syncコマンドの --nodelist オプションに指定してください。

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

<*operation*>: --start, --stop などの操作オプションを指定します。詳細は "A.2.1 fefs_sync コマンド" を参照してください。
<*cluster*>: 故障ノードが属するクラスタを指定します。
<*nodeidlist*>: ノードIDを列挙したファイルを指定します。

FEFS 構築中に故障ノードがあった場合は、故障していたノードのノードIDが以下のファイルに列挙されます。故障からの復旧は、以下のファイルを退避し、退避したファイルを --nodelist オプションに指定してください。

fefs_syncコマンド名	ファイル名
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*> : 故障ノードが属するクラスタ

FEFS 構築中にエラーが発生したノードがあった場合は、エラーが発生していたノードのノードID が以下のファイルに列挙されます。エラーからの復旧は、以下のファイルを退避し、退避したファイルを fefs_sync --nodelist に指定してください。

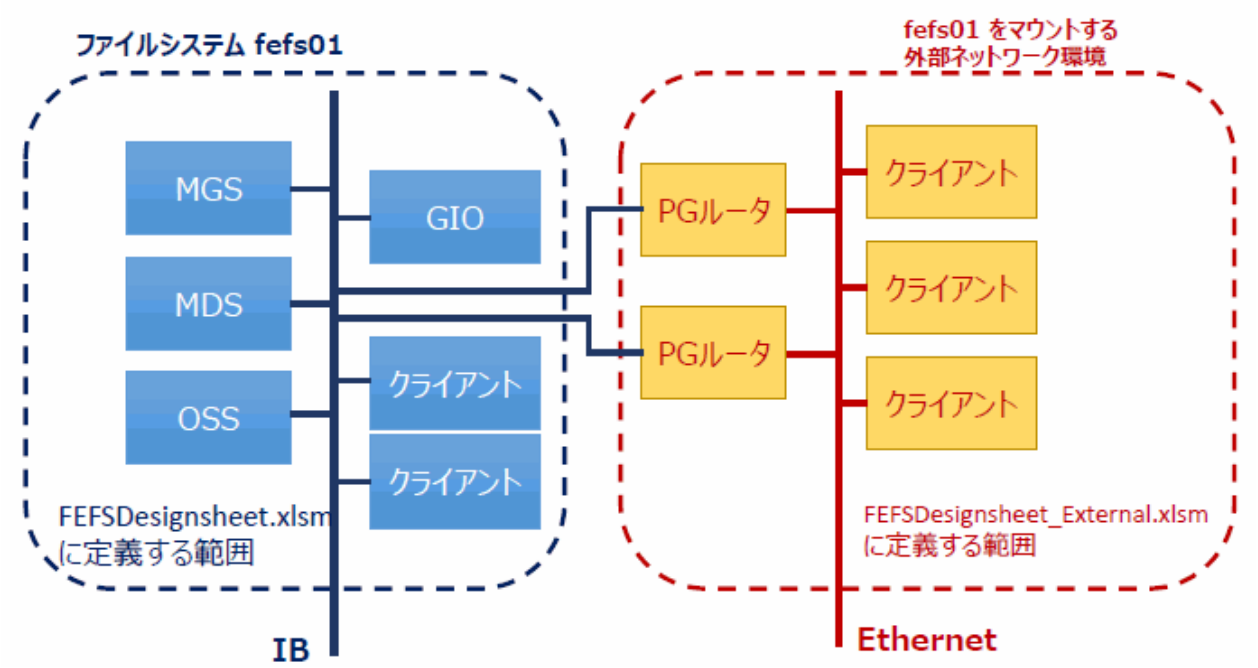
fefs_syncコマンド名	ファイル名
fefs_sync --setup	/var/opt/FJSVfefs/errornodeid_<*cluster*>_setup
fefs_sync --mkfs	/var/opt/FJSVfefs/errornodeid_<*cluster*>_mkfs
fefs_sync --mount	/var/opt/FJSVfefs/errornodeid_<*cluster*>_mount

<*cluster*> : エラーが発生したノードが属するクラスタ

3.12 外部ネットワークにおけるFEFSの構築方法

Technical Computing Suite で管理されていないノードからFEFSを利用できます。Technical Computing Suite で管理されていないネットワークを外部ネットワークと呼びます。FEFSと外部ネットワークとの通信にはEthernetを使用します。外部ネットワークを含むシステム構成のイメージを以下に示します。

図3.22 構成イメージ



以下では、外部ネットワークのクライアントからFEFSを利用するのに必要な設定を示します。外部ネットワークのクライアントと外部ネットワーク接続用ルータ (PGルータ) の設定を行います。

なお、本機能を利用するにあたり、外部ネットワーク接続用ルータとして多目的ノードを用意する必要があります。

3.12.1 外部ネットワーク用 FEFS デザインシートの作成

外部ネットワーク環境構築用のFEFS デザインシート(以降「外部 FEFSデザインシート」と表記)を、Windows 端末で作成してください。

外部FEFSデザインシートのひな形は製品に同梱されています。外部FEFSデザインシートのファイル名は、"FEFSDesignSheet_External.xlsm" です。

外部FEFSデザインシートの作成作業を始めるときは、最初に Excel のマクロを有効にしてください。

注意

FEFS デザインシートは、以下の環境での入力をサポートしています。

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

これ以外の環境については、担当保守員(SE)または当社Support Deskに相談してください。

作成するシートは、以下の3種類です。

- NODEシート
  ネットワーク情報などファイルシステムに依存しない構成情報を定義します。
- NETシート
  ルータとクライアントの構成を定義します。
- GFS シート
  グローバルファイルシステムの構成を定義します。ファイルシステムの数だけ作成してください。

以下、各シートの入力内容について説明します。

3.12.1.1 NODE シートの入力

各ファイルシステムで共通な情報を設定します。

1. 外部ネットワークにおいて構築を行う対象のルータおよびクライアントのネットワーク情報を列挙してください。FEFSデザインシート("FEFSDesignSheet.xlsm")に定義されたノードは定義する必要はありません。
2. 以下は、NODEシートの記入例です。

図3.23 NODE シートの記入例

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
ホスト名を定義します。

b. NODETYPE
ノード種別を定義します。
ルータ用のノードには、多目的ノードを指定してください。
多目的ノードについての詳細は、"3.1.3.1 NODE シートの入力" を参照してください。
Technical Computing Suite製品の範囲外のノードは入力不要です。

c. CLSTNAME
クラスタ名を定義します。Technical Computing Suite製品の範囲外のノードは入力不要です。

d. IB(1)、IB(2)、Ethernet
使用する InfiniBand または Ethernet の、ネットワークインターフェースおよびIPアドレスを定義します。

注意

ルータにおいて、LNet マルチレール機能は使用できません。

1HCAのノードは InfiniBand の情報を IB(1) に入力してください。

ルータ用のノードは、FEFSと接続するネットワークと、クライアントと接続するネットワークの情報の両方を定義してください。FEFS で利用しないネットワークは入力する必要はありません。

Technical Computing Suite製品の範囲内のノードは、ジョブ運用ソフトウェアのデザインシートで定義した名前と同じである必要があります。命名ルールの詳細は「ジョブ運用ソフトウェア 導入ガイド」を参照してください。

3.12.1.2 NET シートの入力

ルータとクライアントの構成を定義します。

以下は NET シートの記入例です。

図3.24 NET シートの記入例

■ ROUTER

HOSTNAME	LNET GROUP-ID	N/I SERVER	N/I CLIENT
router01	3	ib0	ib2
router02	3	ib0	ib2
router03	8	ib0	ens1f0
router04	8	ib0	ens1f0

■ CLIENT

HOSTNAME	LNET GROUP-ID
client001	3
client002	3
client003	8
client004	8

ROUTERセクションには、NODEシートで定義したノードのうち、ルータとして使用するノードを定義します。

a. HOSTNAME
ルータ用ノードのホスト名を定義します。

b. LNET GROUP-ID
ネットワークの範囲を定義するための識別子です。ネットワークの範囲ごとに異なる識別子を定義してください。また、ROUTERセクションとCLIENTセクションに定義した識別子は紐づいている必要があります。
入力できる数値は 1～65535 です。数値は通番・連番である必要はありません。

c. N/I(SERVER,CLIENT)
ルータノードにおける、サーバ側のネットワークインターフェースと、クライアント側のネットワークインターフェースを定義します。NODEシートに定義したネットワークインターフェースを指定します。

CLIENTセクションには、NODEシートで定義したノードのうち、クライアントとして使用するノードを定義します。

a. HOSTNAME
クライアントのホスト名を定義します。

b. LNET GROUP-ID
ネットワークの範囲を定義するための識別子です。ROUTERセクションに定義したノードに接続されるクライアントは、ルータと同一の識別子を入力してください。

3.12.1.3 GFS シートの入力

GFS シートでは、グローバルファイルシステムの設定を行います。1シートにつき 1ファイルシステムの設定を行います。複数のファイルシステムを設定する場合は、ファイルシステムの数だけシートを追加して設定してください。

1. GFS シートの追加
GFS シートは、Excel のマクロで追加します。

Excelマクロの「FEFS Design」>「Insert global filesystem sheet」

2. FILESYSTEM セクション
以下の FILESYSTEM セクションの記入例に沿って説明します。

図3.25 FILESYSTEM セクションの記入例

a. FSNAME
ファイルシステム名を設定してください。
FEFSデザインシート("FEFSDesignSheet.xlsm")で定義されたファイルシステム名と同一である必要があります。

b. MOUNT POINT
クライアントのマウントポイントを指定してください。絶対パスで指定してください。

3. CLIENT セクション
クライアントのマウントオプションを設定します。通常は変更する必要はありません。
以下の場合にオプションを追加してください。

－ QoS 機能を有効にする場合
qos オプションを追加します。QoS 機能を有効にする設定についての詳細は、"3.2 QoS 機能を有効にする設定" を参照してください。

ー user 拡張属性を有効にする場合
user_xattr オプションを追加します。user 拡張属性を有効にする設定についての詳細は、"3.5 user 拡張属性を有効にする設定" を参照してください。

4. PG CLIENT セクション
当該ファイルシステムをマウントするクライアントのホスト名を定義してください。

3.12.1.4 入力データのチェック

以下の Excel のマクロで入力データをチェックできます。入力データに不備がないか確認してください。

Excel マクロの「FEFS Design」>「Check」

不備がある場合は、該当箇所が通知されます。該当箇所を修正して、再度確認してください。問題がなければ、"OK: Check completed." と表示されます。

3.12.2 FEFSセットアップツール用構成定義ファイルの作成

以下の Excel のマクロでセットアップ用の入力データを作成できます。

Excelマクロの「FEFS Design」>「Create config files」

ダイアログが表示されますので、出力先フォルダを指定してください。指定したフォルダ配下に FEFS セットアップツール用構成定義ファイルが作成されます。

3.12.3 FEFSセットアップツール用構成定義ファイルの配置

運用系および待機系システム管理ノードの以下のディレクトリに、FEFS セットアップツール用構成定義ファイルを配置してください。

/etc/opt/FJSVfefs/config 配下

3.12.4 外部ネットワークにおけるFEFSの構築

Technical Computing Suite で管理された範囲内のノードと範囲外のノードの2段階で構築を行います。

1. FEFSサーバの設定およびルータの構築 (Technical Computing Suite 範囲内)
2. クライアントの設定 (Technical Computing Suite 範囲外)

外部ネットワーク上のクライアントは fefs_sync コマンドによる構築が行えないため、各ノード上で個々に構築を行います。

3.12.4.1 FEFSサーバの設定およびルータの構築

FEFSサーバの設定およびルータの構築にはfefs_syncコマンドを使います。fefs_syncコマンドの詳細は、"A.2.1 fefs_syncコマンド" を参照してください。

以降のコマンド実施については、ジョブ運用ソフトウェアがインストールされている必要があります。

以下の操作を運用系システム管理ノード上で実施してください。

注意

FEFS の構築(FEFS デザインシート "FEFSDesignSheet.xlsm" を使用した構築)が済んでいない場合は、"3.1 導入の流れ" を参考にして事前に構築を行ってください。

参考

fefs_sync コマンドでエラーとなったノードへの復旧手順は、"3.11 構築に失敗したノードの構築方法" を参照してください。
fefs_sync コマンドで多目的クラスタを指定する場合は、--compute オプションに指定してください。

1. FEFS 設定ファイルの作成

   "FEFSDesignSheet_External.xlsm" で設定した内容をFEFSへ反映します。
   システム管理ノード上で、以下を実行してください。

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

   FEFS を構築するすべてのクラスタを指定してください。

   --storage : ストレージクラスタ名を指定してください。
   --compute : 計算クラスタ名を指定してください。

2. FEFSサービスの起動

   ルータノードのFEFSサービスを起動します。
   以下を実行してください。

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

   ルータノードのクラスタとノードIDを指定してください。

   --compute : 計算クラスタ名を指定してください。
   --nodelist : ルータノードのノードIDが列挙されたファイルを指定してください。

3. FEFS状態の確認

   ルータノードにおいて、FEFS のサービスが正常に起動されたことを pashowclst コマンドで確認してください。

   以下を実行してください。

```
# pashowclst -v --nodetype <ルータノードのノード種別>
```

   FEFS の状態が FEFS(o) に遷移していれば、FEFS のサービスは正常に起動されています。

3.12.4.2 FEFSクライアントの設定

外部ネットワークにおけるFEFSクライアントの設定は、各ノードで個別に行います。以降のコマンド操作をすべてのクライアント上で行ってください。

1. FEFS設定ファイル作成

   a. FEFS構成定義ファイルの配置

      システム管理ノード上の FEFS セットアップツール用構成定義ファイルを、すべてのクライアントノードの以下のディレクトリに配布してください。

      /etc/opt/FJSVfefs/config配下

   b. FEFS設定ファイル作成

      クライアントノード上で、以下を実行してください。

```
# fefsconfig --setup
```

2. FEFSの起動

   a. FEFSサービスの起動

      すべてのクライアントノード上で以下を実行してください。

```
# systemctl start FJSVfefs
```

   b. FEFS状態の確認

      すべてのクライアントノード上で、FEFSがマウントされたことを確認してください。

```
# mount
```

3.13 注意事項

ARP キャッシュの枯渇について

多数のノードで構成される PRIMERGY ノードのクラスタでは、1つのノードから多数のノードに対する通信が発生します。これにより、通信元ノードの ARP キャッシュ数がデフォルトのままでは不足する可能性があります。ARP キャッシュが枯渇するとアドレス解決に失敗し、FEFS アクセスができなくなります。こうした事態を避けるために、以下のようにカーネルパラメーターのチューニングを検討してください。

- 設定対象ノード
  MGS ノード、MDS ノードおよび OSSノード。

- 設定するカーネルパラメーター
  ARP キャッシュ数に関する以下のカーネルパラメーターを設定します。

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

- パラメーターに設定する値
  必要となる ARP キャッシュ数は、以下の計算式で求められます。この値よりも大きい値を設定してください。

```
通信先ノードのI/F(IPアドレス)数 ×通信先ノード数
```

上記で算出した値が RHEL におけるカーネルパラメーター net.ipv4.neigh.default.gc_thresh3 のデフォルト値 (1024) を超える場合は、チューニングを行ってください。ただし、メモリ不足にならないよう調整する必要があります。

以下はパラメーターの算出例です。

```
通信先ノードに Ethernet が2つ実装されており、そのノードが 1000台の場合、通信元ノードで必要な ARP キャッシュ数は以下のように求められます。
2 (Ethernet) ×1000 (通信先ノード数) = 2000
```

カーネルパラメーターの変更は /etc/sysctl.conf ファイルを編集してください。詳細は RHEL の仕様を確認してください。

環境変数のLD_LIBRARY_PATHについて

プログラム実行時に必要な共有ライブラリをカレントディレクトリに配置しない場合は、環境変数のLD_LIBRARY_PATH にカレントディレクトリを含めないようにしてください。
FEFS 上にあるカレントディレクトリを検索パスに含めると、ファイルへの不必要なアクセスが増え、FEFS へのアクセスが高負荷となることがあります。
以下にカレントディレクトリが検索対象になる設定例を示します。

```
(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
```

なお、(a)、(d)、および (e) のように、明にカレントディレクトリ (".") を指定していなくても、カレントディレクトリが検索対象になるケースがあることに注意してください。詳細は、LD_LIBRARY_PATH の仕様を確認してください。

環境変数 LD_LIBRARY_PATH にカレントディレクトリの設定がない場合とある場合で、カレントディレクトリに対するアクセス回数の違いがあります。以下に例を示します。

例1: LD_LIBRARY_PATH にカレントディレクトリの設定がない場合

```
$ export LD_LIBRARY_PATH=
$ strace -e open dd if=/dev/zero of=./test.dat bs=4k count=1 |& grep ENOENT | wc -l
5
-> ENOENTで終了したopenシステムコールは5個
```

例2: LD_LIBRARY_PATH にカレントディレクトリの設定がある場合

```
$ export LD_LIBRARY_PATH=.
$ strace -e open dd if=/dev/zero of=./test.dat bs=4k count=1 |& grep ENOENT | wc -l
17
-> ENOENTで終了したopenシステムコールは17個
```

上記の例で、カレントディレクトリが FEFS 上の場合、FEFS に対して 12 (17 － 5) 個の ENOENT でエラーとなる open システムコールが実行されます。上記は 1ノードでの実行例ですが、計算ノード 100台でジョブを実行するようなケースでは、12 個 × 100台 = 1200 個の ENOENT でエラーとなる open システムコールが実行されます。

第4章 運用方法

ここでは、FEFSの運用方法を説明します。

4.1 FEFS サーバとクライアントの起動

FEFS サーバとクライアントは、以下のノード種別の順で起動してください。

ノード起動時に自動的にマウントされます。

1. MGS ノード、MDS ノード
2. OSS ノード
3. クライアントノード

注意

FEFS サービスの稼働中にランレベルの変更はできません。ランレベルの変更を行う場合は、FEFS サービスを停止してからランレベルを変更し、そのあと FEFS サービスを起動してください。

参照

ノードの起動に関する操作の詳細は、以下のマニュアルを参照してください。

「ジョブ運用ソフトウェア 管理者向けガイド システム管理編」

4.2 FEFS サーバとクライアントの停止

FEFS サーバとクライアントは、以下のノード種別の順で停止してください。

1. クライアントノード
2. OSS ノード
3. MGS ノード、MDS ノード

注意

FEFS の停止は、pasnap または fjsnap を実行中であれば完了後に行ってください。pasnap または fjsnapを実行中に FEFS の停止を行うと、ノードがパニックすることがあります。

参照

ノードの停止に関する操作の詳細は、以下のマニュアルを参照してください。

「ジョブ運用ソフトウェア 管理者向けガイド システム管理編」

4.3 ストライプ機能の設定

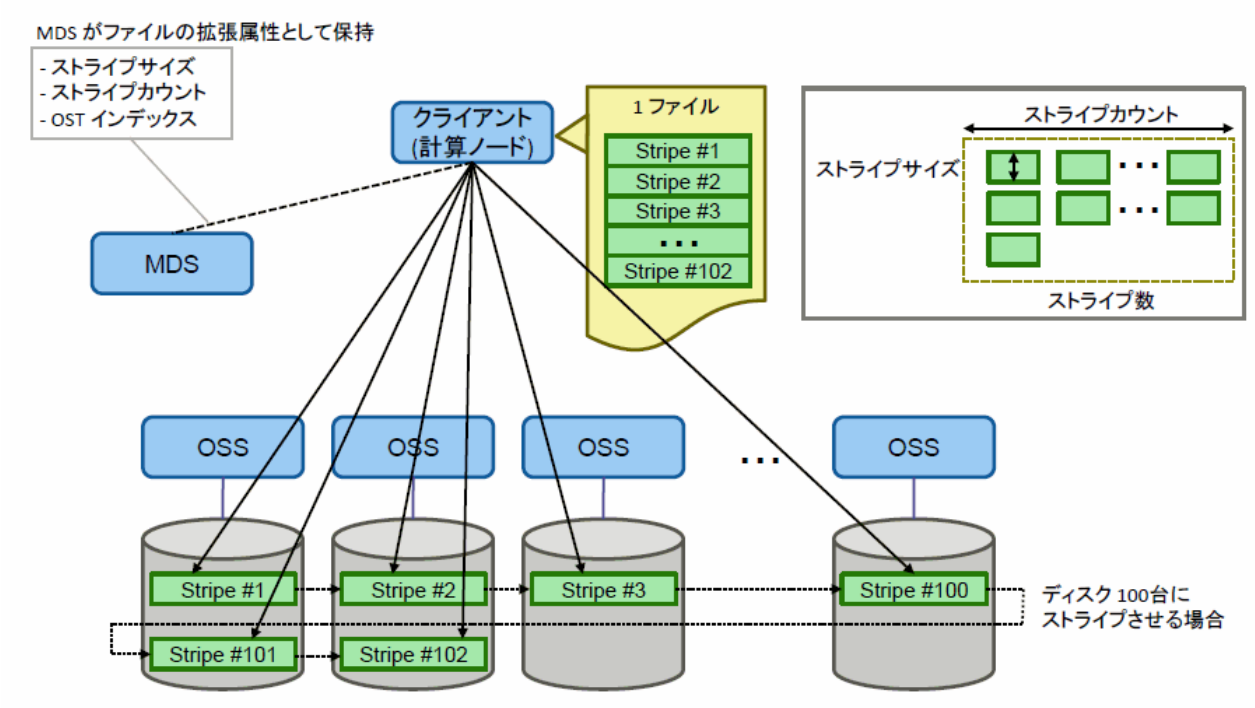
ここでは、ストライプ機能の設定について説明します。

4.3.1 ストライプの設定方法

ストライプの設定は、lfs setstripe コマンドで行います。本コマンドは、クライアントノードで実行します。

以下は、ディスク 100台にストライプさせる例です。

図4.1 ストライプの例



lfs setstripe コマンドでは、データを分散させるサイズ、データを分散させる範囲、データを分散させる数について指定できます。これらを指定したものをストライプパターンと呼びます。

- ストライプサイズ
  データを分散させるサイズを設定できます。
- ストライプカウント
  分散させる OST の範囲を設定できます。分散させる OST の範囲は、OST インデックスの ID および OST_pool 機能によって複数の OST をグループ化した OST のグループ (OST_pool) で指定します。なお、OST_pool は、あらかじめ設定しておく必要があります。
  OST_pool機能の設定については、"4.3.3 OST_pool の設定方法" を参照してください。
  "図4.1 ストライプの例" では、ディスク 100台にストライプさせており、ストライプカウントは 100 になります。

lfs setstripe コマンドで、ストライプパターンを持った新規ファイルを作成します。また、既存のディレクトリに対してもストライプパターンを設定できます。

以下は、ディレクトリ /fefs01 に対してストライプサイズ 2GiB、ストライプカウント 100 のストライプパターンを設定し、OST_pool "pool0001" に対してストライプさせる例です。

```
[クライアントノード]
# lfs setstripe -S 2g -c 100 -p pool0001 /fefs01
```

注意

- 指定したOSTの領域が枯渇している場合、別のOSTに設定される場合があります。
- ストライプサイズは、65536バイトの倍数で指定してください。

参考

lfs df コマンドで、OST インデックス出力情報を参照できます。

```
[クライアントノード]
# lfs df
```

```
UUID                 1K-blocks        Used   Available Use% Mounted on
gfefs0-MDT0000_UUID  189194456      130552   179298280   0% /fefs01[MDT:0]
gfefs0-OST0000_UUID  202702656       24792   192532828   0% /fefs01[OST:0]
gfefs0-OST0001_UUID  202702656       25304   192532312   0% /fefs01[OST:1]
gfefs0-OST0002_UUID  202702656       26328   192531288   0% /fefs01[OST:2]
gfefs0-OST0003_UUID  202702656       25816   192531804   0% /fefs01[OST:3]

filesystem summary:  810810624      102240   770128232   0% /fefs01
```

出力結果の各行で、右端の数字(下線部分)がOSTのインデックスのIDとなります。

参照

lfs setstripe コマンドの詳細は、"A.2.6 lfsコマンド" のサブコマンド "lfs setstripe" を参照してください。

4.3.2 ストライプ設定の確認方法

ストライプ設定の確認は、lfs getstripe コマンドで行います。指定したファイルやディレクトリのストライプパターンの情報を表示します。本コマンドは、クライアントノードで実行します。

以下は、ディレクトリ/fefs01のストライプパターンを表示する例です。

```
[クライアントノード]
# lfs getstripe /fefs01
/fefs01
stripe_count:   100 stripe_size:    2147483648 stripe_offset:  -1
```

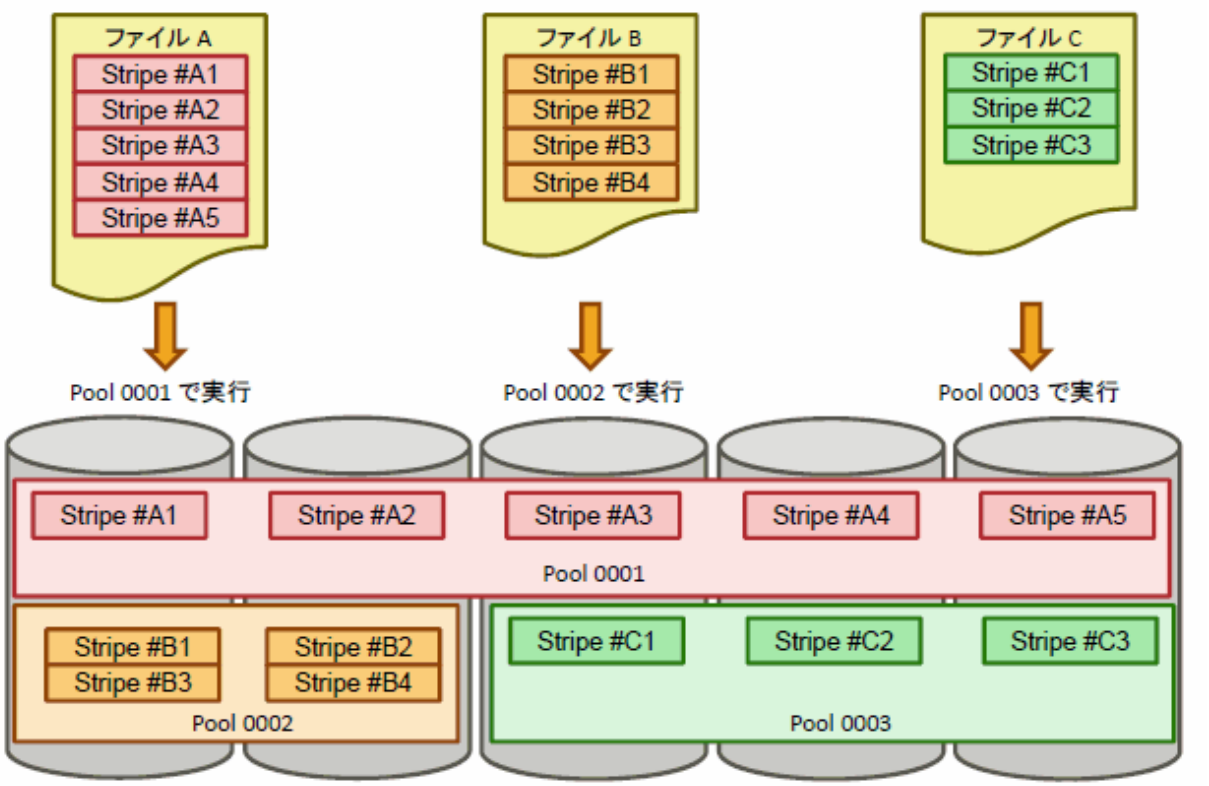
参照

lfs getstripe コマンドの詳細は、"A.2.6 lfsコマンド" のサブコマンド "lfs getstripe" を参照してください。

4.3.3 OST_pool の設定方法

OST_pool 機能は、あらかじめ OST をグループ化するものです。このグループ化したものを OST_Pool と呼びます。これは、ストライプ機能における分散範囲の指定に利用します。ここでは、OST_pool の設定方法について、説明します。

図4.2 OST_poolの例



1. MGS へクライアントをマウント (MGS と MDS が別ノードの場合だけ)
   MGS と MDS が別ノードの場合、OST_pool 機能を利用するためには、MGS へクライアントがマウントされている必要があります。
   MGS 上で、MGSへクライアントをマウントしてください。

   以下は、MGS の IB のIPアドレスが 192.0.2.81 の場合の例です。

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

   <*fsname*> : ファイルシステム名

2. 新規 OST_poolの作成
   lctl pool_new コマンドで、新しい OST_pool を作成します。本コマンドは MDS 上で実行します。ただし、MGSとMDS が別ノードの場合は、MGS 上で実行します。

   以下は、ファイルシステム fefs01 にpool0001というOST_poolを作成する例です。

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

   参照

   lctl pool_new コマンドの詳細は、"A.2.7 lctlコマンド" のサブコマンド "lctl pool_new" を参照してください。

3. OST_poolへのOSTの登録
   lctl pool_addコマンドで、OSTをOST_Poolに登録(エントリ)します。本コマンドは、MDS 上で実行します。ただし、MGSとMDS が別ノードの場合は、MGS 上で実行します。

   以下は、OST_pool "pool0001" に OST を登録する例です。

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

上記では、OST_pool "pool0001" に OST インデックスの ID が 0 の OST を登録しています。
本コマンドで指定する OST 名は、以下のように OST インデックスに対応しています。

OST*xxxx* : OST index *xxxx*

OST インデックスが 0 の場合、OST 名は OST0000 になります。pool0001 に所属させるすべての OST に対して、同じ操作を行ってください

参照

lctl pool_add コマンドの詳細は、"A.2.7 lctlコマンド" のサブコマンド "lctl pool_add" を参照してください。

4. OST_poolの情報表示
lctl pool_listコマンドで、OST_pool のリストと、OST_pool に登録されているOSTの一覧を表示します。本コマンドは、MDS 上で実行します。ただし、MGS と MDS が別ノードの場合は、MGS 上で実行します。OST_pool が登録されているファイルシステム名を指定すると、OST_Pool のリスト (OST_pool名) を表示できます。

以下は、ファイルシステム fefs01 に登録されている OST_pool の一覧を表示しています。

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

lctl pool_list コマンドに OST_pool の名 (<*fsname*>.<*Pool名*>) を指定すると、OST_pool に登録されている OST の一覧を表示できます。

以下は、OST_pool "pool0001" に登録されている OST の一覧を表示します。

```
[MGS ノード]
# lctl pool_list fefs01.pool0001
```

lfs pool_list コマンドでも、同じ情報を表示できます。本コマンドは、クライアントノードで実行します。

以下は、マウントポイント /fefs01 に対して登録されている OST_pool の一覧を表示しています。

```
[クライアントノード]
$ lfs pool_list /fefs01
```

OST_pool の pool0001 に登録されている OST の一覧を表示します。

```
[クライアントノード]
$ lfs pool_list fefs01.pool0001
```

参照

lctl pool_list コマンドの詳細は、"A.2.7 lctlコマンド" のサブコマンド "lctl pool_list" を参照してください。また、lfs pool_list コマンドの詳細は、"A.2.6 lfsコマンド" のサブコマンド "lfs pool_list" を参照してください。

5. OST_poolからのOSTの削除
lctl pool_remove コマンドで、OST_pool に登録されているOSTのエントリを削除します。本コマンドはMDS 上で実行します。ただし、MGS と MDS が別ノードの場合は、MGS 上で実行します。

以下は、OST_pool "pool0001" に登録されている "OST0000" を削除する例です。

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

参照

lctl pool_remove コマンドの詳細は、"A.2.7 lctlコマンド" のサブコマンド "lctl pool_remove" を参照してください。

6. OST_pool の削除
   lctl pool_destroy コマンドで OST_pool を削除します。本コマンドは MDS 上で実行します。ただし、MGS と MDS が別ノードの場合は、MGS 上で実行します。

   以下は、OST_pool "pool0001" を削除する例です。

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

参照

lctl pool_destroy コマンドの詳細は、"A.2.7 lctlコマンド" のサブコマンド "lctl pool_destroy"を参照してください。

7. クライアントのアンマウント (MGS と MDS が別ノードの場合だけ)

   MGS 上で、MGS へマウントしたクライアントをアンマウントします。

```
[MGS ノード]
# umount /mnt/fefs
```

注意

上記に示した OST_pool を操作するコマンドを実行する場合は、対象となる FEFS サーバ群にマウントしている FEFS クライアントをすべてアンマウントしてから行ってください。

4.4 マルチ MDS の使い方

4.4.1 リモートディレクトリの作成

マルチ MDS の環境で、ディレクトリを作成するMDTを指定する例を以下に示します。

MDT1、MDT2 はそれぞれインデックス番号1、インデックス番号2 であるものとします。

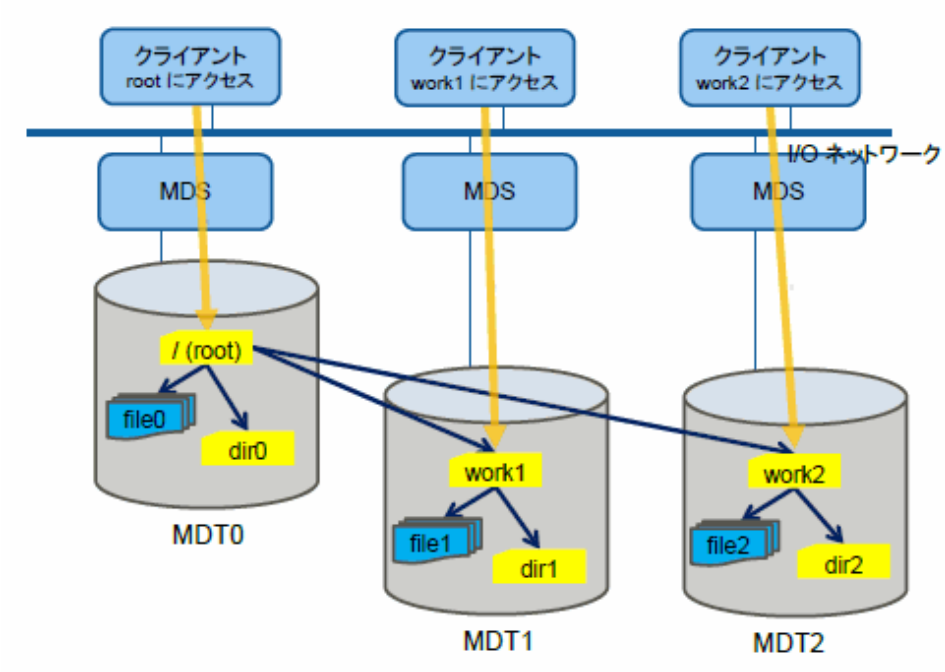
```
[クライアントノード]
# lfs mkdir -i 1 /mnt/fefs/work1
# lfs mkdir -i 2 /mnt/fefs/work2
```

これにより、MDT1に /work1 が、MDT2に /work2 がそれぞれ作られます。

注意

MDT数よりも大きな数を -i の指定値とした場合、MDT 数が指定されたものとして扱います。

図4.3 指定した MDT へのディレクトリ作成



4.4.2 ストライプディレクトリ の作成

マルチ MDS の環境で、ストライプディレクトリを作成してメタデータを複数の MDT に分散配置する例を以下に示します。

対象となる MDT の数は4つで、インデックス番号順に MDT1、MDT2、MDT3、MDT4となっているものとします。

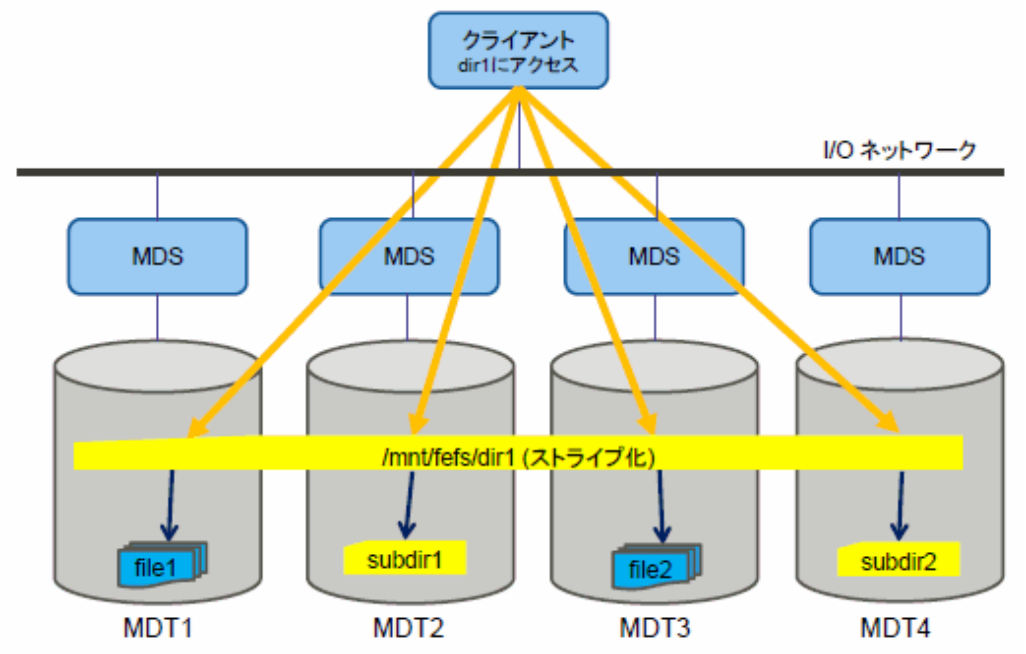
```
[クライアントノード]
# lfs mkdir -c 4 -i 1 /mnt/fefs/dir1
```

これにより、MDT1 を先頭とする 4つの MDT にストライプ化されたディレクトリが作成されます。
この後、ディレクトリ配下に作られたファイルは、分散管理されます。

注意

MDT数よりも大きな数を-c または -i の指定値とした場合、MDT 数が指定されたものとして扱います。

図4.4 ストライプディレクトリの作成



参照

コマンドの詳細は、"A.2.6 lfsコマンド" のサブコマンド "lfs mkdir" を参照してください。

ディレクトリのストライピング情報を確認するには lfs getdirstripe コマンドを使用します。以下に使用例を示します。

```
[クライアントノード]
$ 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]
```

上記は、ストライプカウントが 4、ストライプされた MDS の開始番号が 1 であることを意味します。

参照

コマンドの詳細は、"A.2.6 lfsコマンド" のサブコマンド "lfs getdirstripe" を参照してください

4.5 QUOTA 機能の設定

ここでは、QUOTA 機能の設定について説明します。

4.5.1 ユーザー・グループに対する QUOTA 設定

ファイルまたはディレクトリに対するQUOTA機能の設定方法を説明します。この設定は、特に断りがない限りroot権限を持つユーザーが行います。

1. 事前準備

   特に必要ありません。

2. 制限値の設定方法

QUOTA 機能を有効にすると、各ユーザー、グループに対する制限値の設定が可能となります。
以下は、ユーザーやグループごとの制限値の設定例です。

表4.1 ユーザーごとの制限値の設定例

ユーザー名	ディスク容量	ファイル数
user1	32GiB	制限なし
user2	制限なし	10000
user3	64GiB	5000

ユーザーのソフトリミットの猶予期間 (ブロック数): 1000秒
ユーザーのソフトリミットの猶予期間 (inode 数): 3日

表4.2 グループごとの制限値の設定例

グループ名	ディスク容量	ファイル数
group1	512GiB	制限なし
group2	制限なし	20000
group3	1TiB	30000

グループのソフトリミットの猶予期間 (ブロック数): 3日
グループのソフトリミットの猶予期間 (inode 数): 1000秒

注意

ソフトリミットの猶予期間は、個々のユーザーまたはグループに対してではなく、すべてのユーザーまたはグループに対して共通に設定されます。

これらの設定は、lfs setquotaコマンドで設定します。本コマンドは、クライアントノード上で実行します。ディスク容量またはファイル数を制限なしにする場合は、値に "0" を指定してください。

以下の例では、FEFSは /mnt/fefs にマウントしています。

```
[クライアントノード]
# lfs setquota -u user1 -B 33554432 /mnt/fefs                # user1の設定
# lfs setquota -u user2 -I 10000 /mnt/fefs                   # user2の設定
# lfs setquota -u user3 -B 67108864 -I 5000 /mnt/fefs        # user3の設定
# lfs setquota -g group1 -B 536870912 /mnt/fefs              # group1の設定
# lfs setquota -g group2 -I 20000 /mnt/fefs                  # group2の設定
# lfs setquota -g group3 -B 1073741824 -I 30000 /mnt/fefs    # group3の設定
# lfs setquota -t -u -b 1000 -i 3d /mnt/fefs                 # ユーザーソフトリミットの猶予期間の設定
# lfs setquota -t -g -b 3d -i 1000 /mnt/fefs                 # グループソフトリミットの猶予期間の設定
```

(注) ディスク容量の制限は、KiB単位で指定してください。

設定済の制限値を変更する場合も、手順は同じです。

参照

lfs setquota コマンドの詳細は、"A.2.6 lfsコマンド" のサブコマンド "lfs setquota" を参照してください。

注意

- QUOTA 設定の上限に達した状態で、上限値を上回る値で再設定しても、最初の上限値を超えてファイルを作成できないことがあります。その場合は、いったん 0を設定してから再設定すると、設定値が有効になります。

ー MDS に 1MDT、または OSS に 1OSTのノードがある場合、その MDT または OSTの QUOTA 制限値はデフォルトで有効になりません。有効にするためには、サーバマウント確認後に MGS ノードで以下のコマンドを実行します。

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

<*fsname*>: ファイルシステム名

3. 制限値の確認方法

制限値は、lfs quota コマンドで確認できます。本コマンドは、クライアントノード上で実行します。

以下は、ユーザー名: user1 の制限値を確認する例です。この場合は、lfs quota コマンドの -u オプションでユーザー名を指定します。

```
[クライアントノード]
# 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       -
```

以下は、自分のユーザー (ユーザー名:user1、所属するグループグループ名: group1) の制限値を確認する例です。この場合は、以下のように lfs quota コマンドを実行します。オプションの指定は不要です。

```
[クライアントノード]
$ lfs quota /mnt/fefs                    # 自分がuser1で、所属するグループがgroup1の場合
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       -
```

以下は、ソフトリミットの猶予期間を確認する例です。この場合は、以下のように lfs quota コマンドに -t オプションを指定します。

```
[クライアントノード]
$ lfs quota -t -u /mnt/fefs       # ユーザーソフトリミットの猶予期間を確認
Block grace time: 16m40s; Inode grace time: 3d
$ lfs quota -t -g /mnt/fefs       # グループソフトリミットの猶予期間を確認
Block grace time: 3d; Inode grace time: 16m40s
```

参照

lfs quotaコマンドの詳細は、"A.2.6 lfsコマンド" のサブコマンド "lfs quota" を参照してください。

4. QUOTA 機能を無効にする手順

QUOTA 機能を無効に変更するには、以下の手順で行ってください。

1. FEFSクライアント上で、全ユーザーおよびグループに設定している制限値をメモします。
2. FEFSクライアント上で、全ユーザーおよびグループに対して、"lfs setquota" でディスク容量およびファイル数の制限値に "0" を設定してください。
   再度、QUOTA機能を有効にする場合は、手順1.でメモした値に "lfs setquota" で設定してください。

4.5.2 プロジェクトに対する QUOTA 設定

プロジェクト QUOTA 機能の設定方法を説明します。この設定は、特に断りがない限り、root 権限を持つユーザーが行います。

1. 事前準備

lfs project コマンドを使って、プロジェクトID と継承フラグを設定します。ディレクトリに継承フラグが設定されていると、配下に作成されるディレクトリは、親ディレクトリのプロジェクトIDと継承フラグ自身を継承します。

以下の例では、ディレクトリ/mnt/fefs/dir1 にプロジェクトID 1000を設定し、その配下はすべてプロジェクトID 1000と継承フラグが設定されます。

```
[クライアントノード]
# lfs project -srp 1000 /mnt/fefs/dir1      # /mnt/fefs/dir1配下にすべてプロジェクトID 1000と継承フラグを設定する
# lfs project -d /mnt/fefs/dir1             # /mnt/fefs/dir1自身のプロジェクト設定を表示
    1000 P /mnt/fefs/dir1
# lfs project -r /mnt/fefs/dir1             # /mnt/fefs/dir1配下のプロジェクト設定を再帰的に表示
    1000 P /mnt/fefs/dir1/subdir11
    1000 P /mnt/fefs/dir1/file11
    1000 P /mnt/fefs/dir1/subdir11/subdir21
    1000 P /mnt/fefs/dir1/subdir11/file21
```

注意

- lfs project コマンドに -s オプションを付けないと、以後そのディレクトリ配下にファイル・ディレクトリを作成する際、それらのプロジェクトIDは設定されず、親ディレクトリの プロジェクトQUOTAカウントに計上されません。
- すでにプロジェクトIDが設定されているファイル、ディレクトリに対して再度 lfs project コマンドを実行すると、プロジェクトIDの値は上書きされます。

参照

lfs project コマンドの詳細は、"A.2.6 lfsコマンド" のサブコマンド "lfs project " を参照してください。

2. 制限値の設定方法

プロジェクトQUOTA 機能を有効にして、プロジェクトとディレクトリを関連づけると、ディレクトリごとにディスク容量やファイル数などの制限値を設定できます。

ディレクトリごとの制限値の設定例を示します。

表4.3 プロジェクトごとの制限値の設定例

プロジェクトID	ディレクトリ	ディスク容量	ファイル数
1000	/mnt/fefs/dir1	32GiB	制限なし
2000	/mnt/fefs/dir2	制限なし	10000
3000	/mnt/fefs/dir3	64GiB	5000

ソフトリミットの猶予期間 (ブロック数): 7200秒
ソフトリミットの猶予期間 (inode 数): 1日

注意

プロジェクトQUOTAのソフトリミットの猶予期間は、個々のプロジェクトごとでなく、ファイルシステムごとに設定します。

これらの設定は、lfs setquota コマンドで設定します。本コマンドは、クライアントノード上で実行します。

以下は、FEFS を /mnt/fefs にマウントしている例です。

```
[クライアントノード]
# mkdir /mnt/fefs/dir1                              # ディレクトリ /mnt/fefs/dir1 を作成
# lfs project -srp 1000 /mnt/fefs/dir1              # ディレクトリ /mnt/fefs/dir1 に プロジェクトID 1000 を設定
# lfs setquota -p 1000 -B 33554432 /mnt/fefs        # プロジェクトID 1000 に QUOTAを設定
# mkdir /mnt/fefs/dir2                              # ディレクトリ /mnt/fefs/dir2 を作成
# lfs project -srp 2000 /mnt/fefs/dir2              # ディレクトリ /mnt/fefs/dir2 に プロジェクトID 2000 を設定
# lfs setquota -p 2000 -I 10000 /mnt/fefs           # プロジェクトID 2000 に QUOTAを設定
# mkdir /mnt/fefs/dir3                              # ディレクトリ /mnt/fefs/dir3 を作成
```

```
# lfs project -srp 3000 /mnt/fefs/dir3                  # ディレクトリ /mnt/fefs/dir3 に プロジェクトID 3000 を設定
# lfs setquota -p 3000 -B 67108864 -I 5000 /mnt/fefs    # プロジェクトID 3000 に QUOTAを設定
# lfs setquota -t -p -b 7200 -i 1d /mnt/fefs            # プロジェクトQUOTAソフトリミットの猶予期間の設定
```

(注) ディスク容量の制限はKiB単位で指定してください。

設定済の制限値を変更する場合も、手順は同じです。

注意

- QUOTA 設定の上限に達した状態で、上限値を上回る値で再設定しても、最初の上限値を超えてファイルを作成できないことがあります。その場合は、いったん 0を設定してから再設定すると、設定値が有効になります。
- MDS に 1MDT、または OSS に 1OSTのノードがある場合、その MDT または OSTの QUOTA 制限値はデフォルトで有効になりません。有効にするためには、サーバマウント確認後に MGS ノードで以下のコマンドを実行します。

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

<*fsname*>: ファイルシステム名

lctl conf_param コマンドの詳細は、"A.2.7 lctlコマンド" のサブコマンド "lctl conf_param" を参照してください。

本コマンドは1つのファイルシステムにつき1度だけ実行すればよく、設定した値を削除する必要はありません。

参照

lfs setquota コマンドの詳細は、"A.2.6 lfsコマンド" のサブコマンド"lfs setquota" を参照してください。

参考

既存ディレクトリにプロジェクトQUOTAを設定する際、階層が深い場合や配下に多数のファイルを持っている場合は、あらかじめ空のディレクトリを作ってプロジェクトQUOTAと継承フラグを設定し、そこへファイルを移動させる方が、対象のディレクトリに lfs project -srp コマンドで直接プロジェクトQUOTAを設定するよりも、システムにかかる負荷は軽くなります。

3. 制限値の確認方法

プロジェクト QUOTA 機能の制限値は、lfs quota コマンドで確認できます。本コマンドは、クライアントノード上で実行します。

以下は、プロジェクトID 1000 の制限値を確認する例です。この場合は、lfs quota コマンドの -p オプションで確認対象のプロジェクトIDを指定します。

```
[クライアントノード]
$ 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       -
```

以下は、ソフトリミットの猶予期間を確認する例です。この場合は、lfs quota コマンドに -t オプションを指定します。

```
[クライアントノード]
$ lfs quota -t -p /mnt/fefs       # プロジェクト QUOTA ソフトリミットの猶予期間を確認
Block grace time: 2h; Inode grace time: 1d
```

参照

lfs quotaコマンドの詳細は、"A.2.6 lfsコマンド" のサブコマンド "lfs quota" を参照してください。

4. ファイルからプロジェクトIDを削除する方法

   プロジェクト QUOTA 機能を無効にするには、lfs project -C コマンドを使って、プロジェクトID をクリアします。以下の例では、/mnt/fefs/dir1 配下のプロジェクトID はすべてクリアされます。

```
[クライアントノード]
# lfs project -C -r /mnt/fefs/dir1
```

4.6 QoS機能の設定

4.6.1 FEFSクライアントのQoS状態確認

FEFSデザインシートに指定したQoS機能のオプションを確認する方法は、以下となります。

例1: QoS 機能の確認

FEFS デザインシートに qos を指定した場合は "1"、noqos を指定した場合は "0" が表示されます。

```
[クライアントノード]
# cat /proc/fs/lustre/mdc/<fsname>*/qos
1
```

<*fsname*>: FEFS デザインシートで指定したファイルシステム名

FEFS デザインシートに qos_cache を指定した場合は "1"、noqos を指定した場合は "0" が表示されます。

```
[クライアントノード]
# cat /proc/fs/lustre/mdc/<fsname>*/qos_cache
1
```

<*fsname*>: FEFS デザインシートで指定したファイルシステム名

例2: mclientmaxオプションの数値を確認する

mclientmax オプション (MDS に対して、クライアントノード内で同時発行可能なリクエスト数) が 4 であることが確認できます。

```
[クライアントノード]
# cat /proc/fs/lustre/mdc/<fsname>*/mclientmax
4
```

4.6.2 FEFSクライアントのQoS状態変更

FEFS デザインシートに指定した QoS 機能のオプションを変更する場合は、以下の手順で行ってください。

1. FEFS デザインシートの作成
   "3.2.1 QoS 機能の有効化" の "FEFSクライアントの設定方法" を参考に、FEFS デザインシートを更新してください。
2. FEFS セットアップツール用構成定義ファイルの作成
   "3.1.4 FEFSセットアップツール用構成定義ファイルの作成" の手順を実施してください。
3. FEFS セットアップツール用構成定義ファイルの配置
   "3.1.5 FEFSセットアップツール用構成定義ファイルの配置" の手順を実施してください。
4. FEFS 設定ファイルの作成
   以下を実行してください。

```
[システム管理ノード]
# fefs_sync --setup  --storage=<cluster>[,…] --compute=<cluster>[,…]
```

   FEFS を構築するすべてのクラスタを指定してください。

   --storage : ストレージクラスタ名を指定してください。
   --compute : 計算クラスタ名および多目的クラスタ名を指定してください。

5. システム停止
   以下を実行してください。

```
[システム管理ノード]
# pastop -c <cluster,cluster,…>
```

   -c オプションには、停止するクラスタ名をすべて指定してください。

注意

- システムの停止は、計算クラスタ → ストレージクラスタの順に行ってください。
- 多目的ノードで FEFS を起動している場合は、pastop コマンドを実行する前に、多目的ノードで FEFS を停止してください。

参照

pastopコマンドの詳細については、以下のマニュアルを参照してください。

「ジョブ運用ソフトウェア 管理者向けガイド システム管理編」

6. システム起動
   以下を実行してください。

```
[システム管理ノード]
# pastart -c <cluster,cluster,…>
```

   -c オプションには、起動するクラスタ名をすべて指定してください。

注意

- システムの起動は、ストレージクラスタ → 計算クラスタの順に行ってください。
- 多目的ノードで FEFS を使用する場合は、pastartコマンドを実行したあとに、多目的ノードで FEFS を起動してください。

参照

pastartコマンドの詳細については、以下のマニュアルを参照してください。

「ジョブ運用ソフトウェア 管理者向けガイド システム管理編」

4.6.3 MDSのQoS状態確認

MDS の QoS 定義ファイルで指定した QoS 機能の状態は、lctl qos stat コマンドで確認します。

QoS 機能が有効 (Enable) の場合、2行目に QoS 定義ファイルのパス名、3行目以降に QoS 定義ファイルの内容を表示されます。

```
[MDSノード]
# 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.
```

参照

lctl qos statコマンドの詳細は、"A.2.7 lctlコマンド" のサブコマンド "lctl qos" を参照してください。

4.6.4 MDS の QoS 定義の変更

QoS 定義ファイルの内容の変更は、lctl qos コマンドで行います。この操作は、MDS ノード上で、root 権限を持つユーザーが行います。

以下に手順を説明します。

1. QoS 定義ファイルの編集
   QoS 定義ファイルを編集します。

   ```
   [MDSノード]
   # vi /etc/opt/FJSVfefs/qosserver.conf
   ```

2. QoS 定義ファイルの構文チェック
   lctl qos check コマンドで、手順1. で編集した QoS 定義ファイルの構文チェックを行います。

   ```
   [MDSノード]
   # lctl qos check /etc/opt/FJSVfefs/qosserver.conf
   QoS command was completed.
   ```

3. QoS 機能の無効化
   lctl qos off コマンドで、QoS機能を無効にします。

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

4. QoS機能の有効化
   lctl qos on コマンドで、QoS 機能を有効にします。

   ```
   [MDSノード]
   # lctl qos on /etc/opt/FJSVfefs/qosserver.conf
   QoS command was completed.
   ```

参照

lctl qos check、lctl qos on、および lctl qos off コマンドの詳細は、"A.2.7 lctlコマンド" のサブコマンド "lctl qos" を参照してください。

4.7 QoS機能のチューニング方法

ここでは、QoS機能を適切に設定するために必要な作業について説明します。この作業は、root 権限を持つユーザーが行ってください。

ファイルアクセスが通常時と比べて遅くなる現象が発生した場合、遅くなっている原因がサーバ側にあるか、クライアント側にあるかの切り分けを行います。

以下に切り分け例を記述します。

例

```
現象          ：ログインノードでのコマンドレスポンスが遅い。
切り分け方法  ：別ノード（高負荷でないノード）から、同じコマンド(lsなど)を実行し、レスポンスを調べる。
```

別ノードでは正常な場合	―> クライアント側の問題の可能性が高い。
別ノードでも遅い場合	―> サーバ側の問題の可能性が高い。

クライアント側が原因の場合は、以下を参照してください。

- ファイルのメタ操作 (ls,touchコマンドなど) が遅い場合: "4.7.1 クライアントノード (メタ操作) の分析とチューニング"
- ファイルのデータ操作 (read/writeシステムコールなど) が遅い場合: "4.7.2 クライアントノード (データ操作) の分析とチューニング"

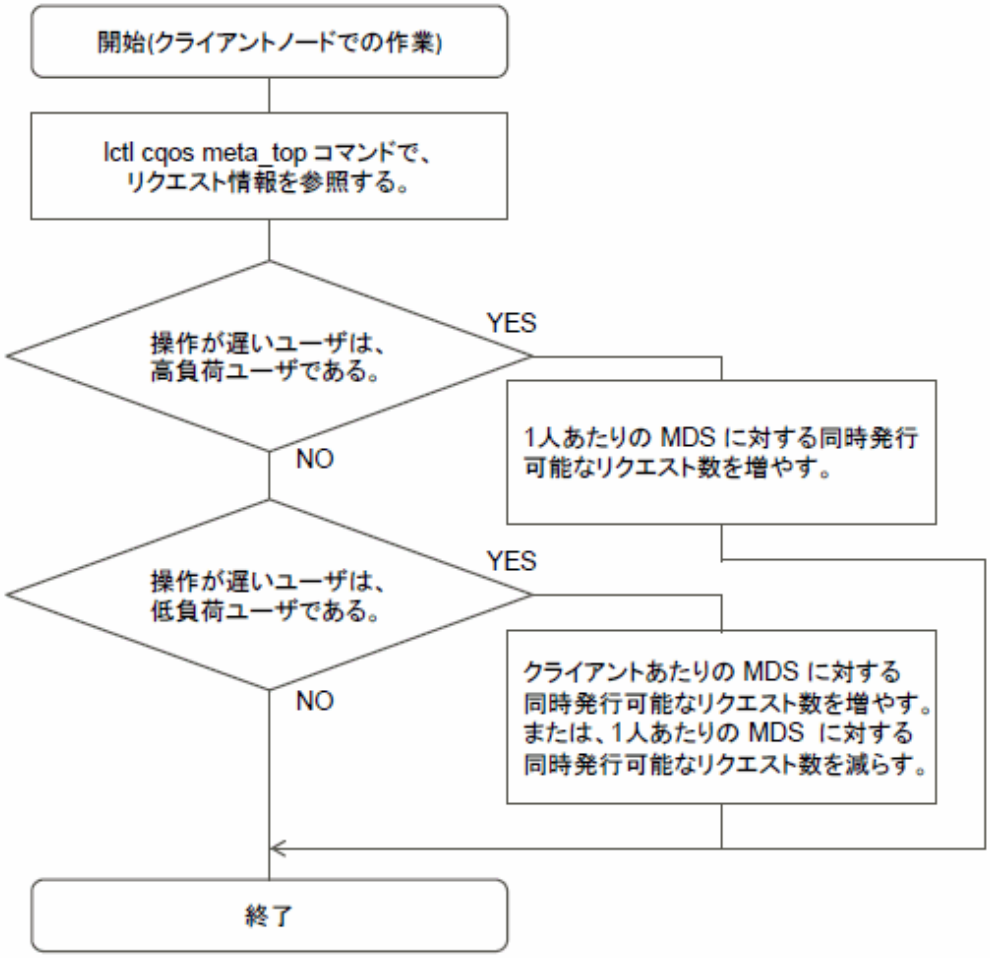
サーバ側が原因の場合は、以下を参照してください。

- ファイルのメタ操作 (ls,touchコマンドなど) が遅い場合: "4.7.3 MDS の分析とチューニング"
- ファイルのデータ操作 (read/writeシステムコールなど) が遅い場合: "4.7.4 OSSの分析とチューニング"

4.7.1 クライアントノード (メタ操作) の分析とチューニング

ファイルのメタ操作が遅い場合、以下の流れで調査を行ってください。

図4.5 クライアントノード (メタ操作) の分析とチューニング



具体的な調査例については、以下を参照してください。

1. ユーザーあたりの同時発行可能数の確認

   クライアントノードでの実行例

```
# lctl cqos meta_top /mnt/fefs/
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.
```

※各出力項目の説明は、"A.2.7 lctlコマンド"のサブコマンド "lctl cqos" を参照してください。

uid=1053 のユーザーは、total_wait_cnt の own が 9325のため、MDS に対して 1ユーザーが同時発行可能なリクエスト数 (上記例では musermax=1) を超えるリクエスト要求を行っています。uid=1053 以外のユーザーは、own が 0のため、上限値の 1を超えるリクエスト要求は行っていません。

ファイルのメタ操作が遅いユーザーID が、uid=1053 である場合は、musermax の値を増やすことで、MDS に対して 1ユーザーが同時発行可能なリクエスト数が増えるので、レスポンスが改善する場合があります。

musermax の値を変更する方法については、"4.6.2 FEFSクライアントのQoS状態変更" を参照してください。

2. クライアントあたりの同時発行可能数の確認

   クライアントノードでの実行例

```
# lctl cqos meta_top /mnt/fefs/
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.
```

uid=1071 のユーザーは、total_wait_cnt の own が 0のため、MDS に対して 1ユーザーが同時発行可能なリクエスト数 (上記例では musermax=3) を超えるリクエスト要求は行っていませんが、other が 2613のため、MDS に対してクライアントから同時発行可能なリクエスト数 (上記例ではmclientmax=4) の制限のために待たされたことがわかります。uid=1070と uid=1053 は own が 1以上のため、この 2ユーザーの同時発行数が多いために、uid=1071 のリクエストが待たされた可能性が高いと考えられます。

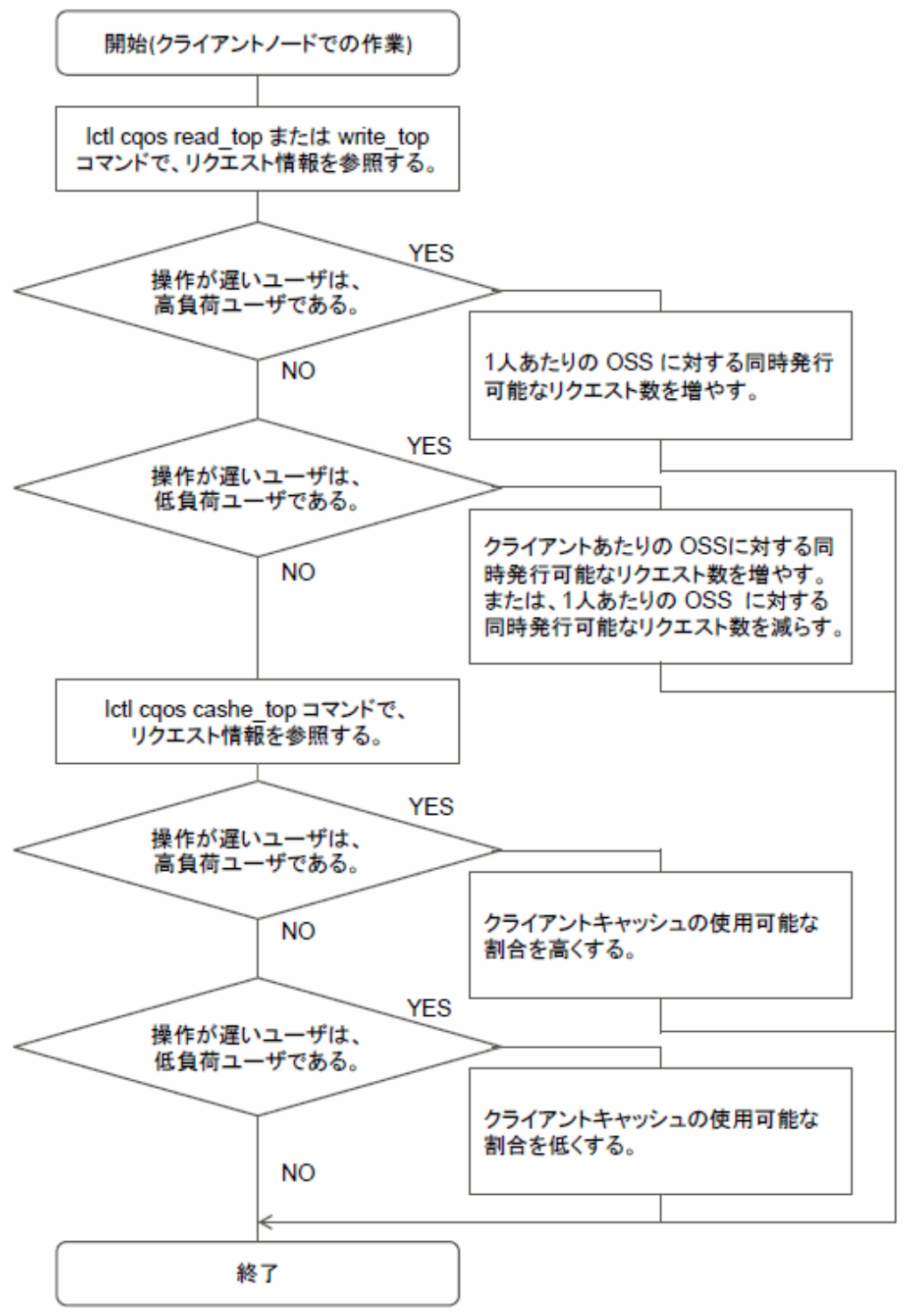
ファイルのメタ操作が遅いユーザーID が、uid=1071 である場合は、mclientmax を増やすことで、MDS に対して 1クライアントが同時発行可能なリクエスト数が増えるので、uid=1071 のレスポンスが改善する場合があります。または、musermax を減らすことで、MDS に対して 1ユーザーが同時発行可能なリクエスト数が減るので、uid=1070 と uid=1053 の同時発行数が抑えられ、uid=1071 のレスポンスが改善する場合があります。

mclientmax、または musermax の値を変更する方法については、"4.6.2 FEFSクライアントのQoS状態変更" を参照してください。

4.7.2 クライアントノード (データ操作) の分析とチューニング

ファイルのデータ操作が遅い場合、以下の流れで調査を行ってください。

図4.6 クライアントノード (データ操作) の分析とチューニング



具体的な調査例については、以下を参照してください。

1. ユーザーあたりの同時発行可能数の確認

   クライアントノードでの実行例

```
# 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.
```

※上記は write_top の出力例です。read の同時発行可能数を確認する場合は、lctl cqos サブコマンドの引数に read_top を指定してください。

※各出力項目の説明は、"A.2.7 lctlコマンド" のサブコマンド "lctl cqos" を参照してください。

uid=1053 のユーザーは、total_wait_cnt の own が 5035のため、OSS に対して 1ユーザーが同時発行可能な write リクエスト数 (上記例ではwrusermax=2) を超えるリクエスト要求を行っています。uid=1053 以外のユーザーは、own が 0 のため、上限値の 2を超えるリクエスト要求は行っていません。

ファイルのデータ操作が遅いユーザーID が、uid=1053 である場合は、wrusermax の値を増やすことで、OSS に対して 1ユーザーが同時発行可能な write リクエスト数が増えるので、レスポンスが改善する場合があります。

wrusermax の値を変更する方法については、"4.6.2 FEFSクライアントのQoS状態変更" を参照してください。

2. クライアントあたりの同時発行可能数の確認

   クライアントノードでの実行例

```
# lctl cqos write_top /mnt/fefs/
wrclientmax=8 wrrootmax=6 wrusermax=6
<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           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.
```

uid=1071 のユーザーは、total_wait_cnt の own が 0 のため、OSS に対して 1ユーザーが同時発行可能な write 数 (上記例ではwrusermax=6) を超えるリクエスト要求は行っていませんが、other が 499のため、OSS に対してクライアントから同時発行可能な write 数 (上記例ではwrclientmax=8) の制限のために待たされたことがわかります。uid=1070 と uid=1053 は own が 1以上であり、この 2ユーザーの同時発行数が多いために、uid=1071 のリクエストが待たされた可能性が高いと考えられます。

ファイルのデータ操作が遅いユーザーID が、uid=1071 である場合は、wrclientmax を増やすことで、OSS に対して 1クライアントが同時発行可能な write 数が増えるので、uid=1071 のレスポンスが改善する場合があります。または、wrusermax を減らすことで、OSS に対して 1ユーザーが同時発行可能な write 数が減るので、uid=1070 と uid=1053 の同時発行数が抑えられ、uid=1071 のレスポンスが改善する場合があります。

wrclientmax、または wrusermax の値を変更する方法については、"4.6.2 FEFSクライアントのQoS状態変更" を参照してください。

3. ユーザーあたりのキャッシュ使用量の確認

   クライアントノードでの実行例

```
# lctl cqos cache_top /mnt/fefs/
dprootmax=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    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.
```

uid=1053 のユーザーは、total_wait_cnt の own が 7149のため、1ユーザーが使用可能なクライアントキャッシュの割合 (上記例ではdpusermax=10) を超える書込み要求を行っています。uid=1053 以外のユーザーは、own が 0のため、上限値の 10%を超える書込み要求は行っていません。

ファイルのデータ操作が遅いユーザーID が、uid=1053 である場合は、dpusermax の値を増やすことで、1ユーザーが使用可能なクライアントキャッシュの割合が増えるので、レスポンスが改善する場合があります。

dpusermax の値を変更する方法については、"4.6.2 FEFSクライアントのQoS状態変更" を参照してください。

4. クライアントあたりのキャッシュ使用量の確認

   クライアントノードでの実行例

```
# lctl cqos cache_top /mnt/fefs/
dprootmax=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.
```

   uid=1071 のユーザーは、total_wait_cnt の own が 0のため、1ユーザーが使用可能なクライアントキャッシュの割合 (上記例では dpusermax=70) を超える要求は行っていませんが、other が 247 のため、他ユーザーによるクライアントキャッシュの使用により待たされたことがわかります。uid=1070 と uid=1053 は own が 1以上であり、この 2ユーザーのクライアントキャッシュ使用量が多いために、uid=1071 のリクエストが待たされた可能性が高いと考えられます。

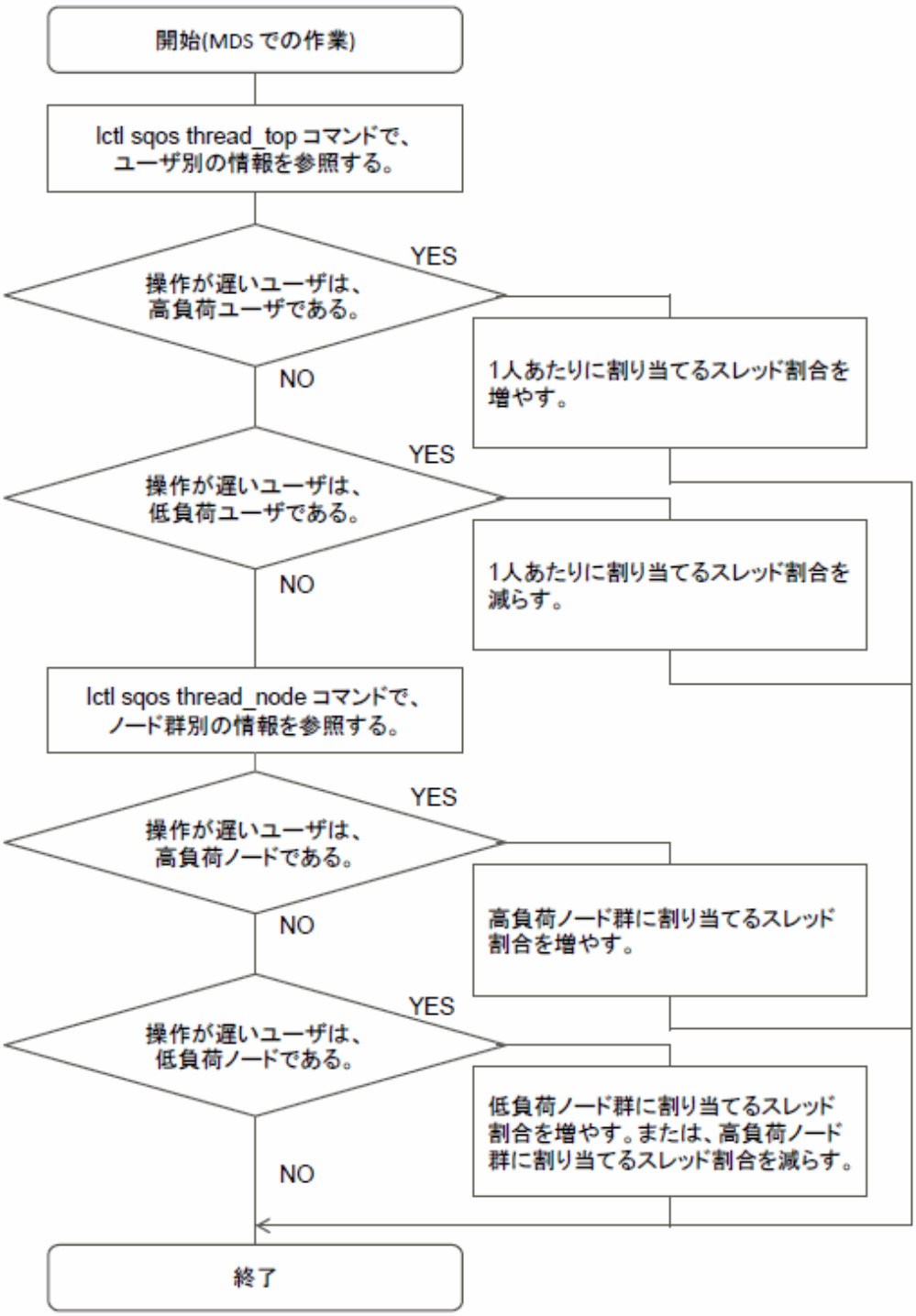
   ファイルのデータ操作が遅いユーザーID が、uid=1071 である場合は、dpusermax を減らすことで、1ユーザーが使用可能なクライアントキャッシュの割合が減るので、uid=1070 と uid=1053 のクライアントキャッシュ使用量が抑えられ、uid=1071 のレスポンスが改善する場合があります。

   dpusermax の値を変更する方法については、"4.6.2 FEFSクライアントのQoS状態変更" を参照してください。

4.7.3 MDS の分析とチューニング

ファイルのメタ操作が遅い場合、以下の流れで調査を行ってください。

図4.7 MDS の分析とチューニング



具体的な調査例については、以下を参照してください。

1. ユーザー単位の確認

   特定のユーザーで、ファイルのメタ操作が遅い場合、MDS で、以下のコマンドを実行してください。

   MDS での実行例

```
# 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.
```

※各出力項目の説明は、"A.2.7 lctlコマンド" のサブコマンド "lctl sqos" を参照してください。

uid=1053 のユーザーは、thread の max が lim と同じ 12となっているため、スレッド制限値の 12個まで使い切ったことがあります。thread の cur は 0 なので、現時点ではスレッドを使用していません。次に、wait_req の max が 133となっているため、スレッド割り当て待ちとなったリクエストキューの長さが最大で 133になったことがあります。wait_req の cur は 0なので、現時点ではリクエストキューは空です。

この出力例からは、uid=1053 が高負荷ユーザーで QoS によるスレッド実行数の抑制が行われ、uid=1053 以外のユーザーは、QoS によるスレッド実行数の抑制は行われていないことがわかります。

ファイルのメタ操作が遅いユーザーID が、高負荷ユーザー (uid=1053) である場合は、1ユーザーあたりに割り当てるスレッド割合を増やすことで、レスポンスが改善する場合があります。ファイルのメタ操作が遅いユーザーID が、低負荷ユーザー (uid=1053以外) である場合は、1ユーザーあたりに割り当てるスレッド割合を減らすことで、高負荷ユーザーの処理が抑えられ、低負荷ユーザーのレスポンスが改善する場合があります。

スレッド割合を変更するためには、QoS定義ファイルの usermax を修正する必要があります。QoS定義ファイルの修正方法については、"4.6.4 MDS の QoS 定義の変更" を参照してください。

2. ノード群単位の確認

特定のノード群で、ファイルのメタ操作が遅い場合、MDS で、以下のコマンドを実行してください。

MDSでの実行例

```
# 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             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
QoS command was completed.
```

※各出力項目の説明は、"A.2.7 lctlコマンド" のサブコマンド "lctl sqos" を参照してください。

nodegrp=1 のノード群は、thread の max が lim と同じ 4となっているため、スレッド制限値の 4個まで使い切ったことがあります。nodegrp=2 のノード群は、thread の max が lim より低い値であるため、スレッド制限値まで使用していません。

この出力例からは、nodegrp=1 は高負荷ノード群で、QoS によるスレッド実行数の抑制が行われ、nodegrp=2 は QoS によるスレッド実行数の抑制は行われていないことがわかります。

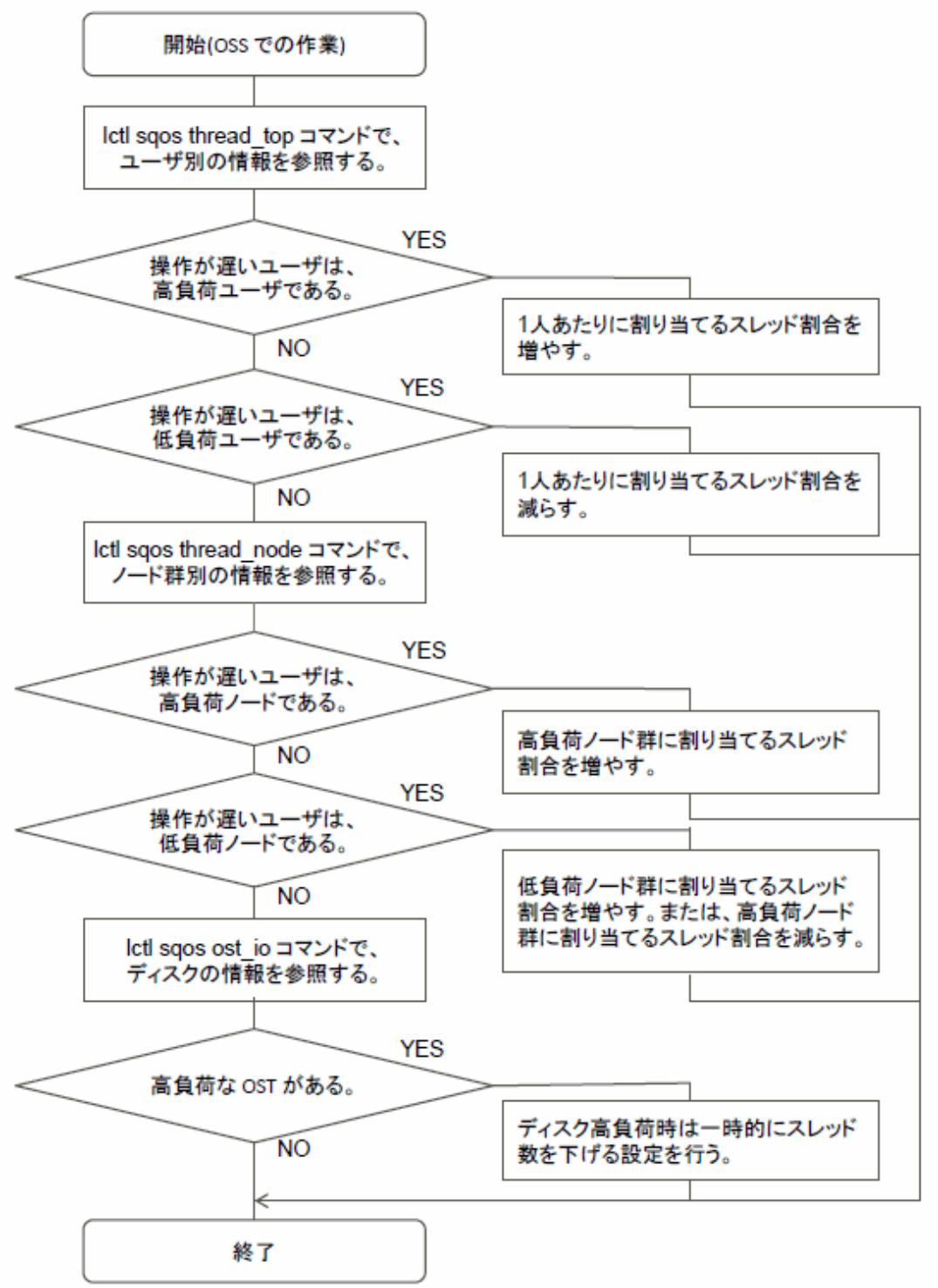
ファイルのメタ操作が遅いノード群が、高負荷ノード群 (nodegrp=1) である場合は、高負荷ノード群に割り当てるスレッド割合を増やすことで、レスポンスが改善する場合があります。ファイルのメタ操作が遅いノード群が、低負荷ノード群 (nodegrp=2) である場合は、低負荷ノード群に割り当てるスレッド割合を増やすか、または高負荷ノード群に割り当てるスレッド割合を減らすことで、レスポンスが改善する場合があります。

スレッド割合を変更するためには、QoS定義ファイルの nodegrp を修正する必要があります。QoS定義ファイルの修正方法については、"4.6.4 MDS の QoS 定義の変更" を参照してください。

4.7.4 OSSの分析とチューニング

ファイルのデータ操作が遅い場合、以下の流れで調査を行ってください。

図4.8 OSS の分析とチューニング



注意

MDS と OSS が同一ノードの場合、lctl sqosコマンドにossオプションを指定してください。lctl sqosコマンドの詳細については、"A.2.7 lctlコマンド" のサブコマンド "lctl sqos" を参照してください。

具体的な調査例については、以下を参照してください。

1. ユーザー単位での確認

   特定のユーザーで、ファイルのデータ操作が遅い場合、OSS で、以下のコマンドを実行してください。

   OSS での実行例

```
# 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.
```

   ※各出力項目の説明は、"A.2.7 lctlコマンド" のサブコマンド "lctl sqos" を参照してください。

   uid=1053 のユーザーは、thread の max が lim と同じ 10となっているため、スレッド制限値の 10個まで使い切ったことがあります。thread の cur は 0なので、現時点ではスレッドを使用していません。次に、wait_req の max が 71となっているため、スレッド割り当て待ちとなったリクエストキューの長さが最大で 71になったことがあります。wait_req の cur は 0なので、現時点ではリクエストキューは空です。

   この出力例からは、uid=1053 が高負荷ユーザーで QoS によるスレッド実行数の抑制が行われ、uid=1053 以外のユーザーは、QoS によるスレッド実行数の抑制は行われていないことがわかります。

   ファイルのデータ操作が遅いユーザーID が、高負荷ユーザー (uid=1053) である場合は、1ユーザーあたりに割り当てるスレッド割合を増やすことで、レスポンスが改善する場合があります。ファイルのデータ操作が遅いユーザーID が、低負荷ユーザー (uid=1053 以外) である場合は、1ユーザーあたりに割り当てるスレッド割合を減らすことで、高負荷ユーザーの処理が抑えられ、低負荷ユーザーのレスポンスが改善する場合があります。

   スレッド割合を変更するためには、QoS定義ファイルの usermax を修正する必要があります。QoS定義ファイルの修正方法については、"4.6.4 MDS の QoS 定義の変更" を参照してください。

2. ノード群単位での確認

   特定のノード群で、ファイルのデータ操作が遅い場合、OSS で、以下のコマンドを実行してください。

   OSS での実行例

```
# 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.
```

   ※各出力項目の説明は、"A.2.7 lctlコマンド" のサブコマンド "lctl sqos" を参照してください。

   nodegrp=1 のノード群は、thread の max が lim と同じ 51となっているため、スレッド制限値の 51個まで使い切ったことがあります。nodegrp=2 のノード群は、thread の max が lim より低い値であるため、スレッド制限値まで使用していません。

   この出力例からは、nodegrp=1 は高負荷ノード群で、QoS によるスレッド実行数の抑制が行われ、nodegrp=2 は QoS によるスレッド実行数の抑制は行われていないことがわかります。

   ファイルのデータ操作が遅いノード群が、高負荷ノード群 (nodegrp=1) である場合は、高負荷ノード群に割り当てるスレッド割合を増やすことで、レスポンスが改善する場合があります。ファイルのデータ操作が遅いノード群が、低負荷ノード群 (nodegrp=2) である場合は、低負荷ノード群に割り当てるスレッド割合を増やすか、または高負荷ノード群に割り当てるスレッド割合を減らすことで、レスポンスが改善する場合があります。

   スレッド割合を変更するためには、QoS定義ファイルの nodegrp を修正する必要があります。QoS定義ファイルの修正方法については、"4.6.4 MDS の QoS 定義の変更" を参照してください。

3. ディスク負荷の確認

   ファイルのデータ操作が遅い場合、OSS で、以下のコマンドを実行してください。

OSSでの実行例

```
# 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.
```

※各出力項目の説明は、"A.2.7 lctlコマンド" のサブコマンド "lctl sqos" を参照してください。

上記の出力例では、ディスクアクセスを行ったOSTが2つあり、fefs-OST0000は、ディスクへのアクセス要求が9228回あり、1回あたりのアクセス時間の平均は416984マイクロ秒です。fefs-OST0001は、ディスクへのアクセス要求が1025回あり、1回あたりのアクセス時間の平均は389789マイクロ秒です。

この出力例では、ディスクへの1回あたりのアクセス時間が0.3-0.4秒かかっているので、スレッド同士によるディスク競合が発生していると考えられます。このような状態では、ディスク待ちによる処理遅延が発生しやすいので、ディスクアクセス時間の上限値をQoS定義ファイルのload_limit_usecを指定して、同時にディスクアクセスするスレッド数を抑えてください。QoS定義ファイルの修正方法については、"4.6.4 MDS の QoS 定義の変更" を参照してください。

以下にload_limit_usecの指定有無による違いを説明します。

OSSでの実行例1 (load_limit_usec 指定なし)

```
# 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.
```

上記の出力例では、uid=1070は1スレッドしか使用していない低負荷ユーザーのため、スレッド割り当て待ち時間(wait_time)は短いですが、exec_timeのavgが203323マイクロ秒であり、ディスク待ちによる処理遅延が発生している可能性があります。uid=1070のディスク待ち時間を短縮するためには、2通りの方法があります。1つは、高負荷ユーザーのuid=1053が81スレッドを使用しているので、これを抑えるためQoS定義ファイルのusermaxで1ユーザーあたりのスレッド割合を下げることです。もう1つの方法は、QoS定義ファイルのload_limit_usecを指定するという方法で、ディスク高負荷時はusermaxで指定した割合よりも小さな値にスレッド数を抑える方法です。

以下は、QoS定義ファイルにload_limit_usec=30000を指定した場合の例です。

OSS での実行例2 (load_limit_usec=30000 指定あり)

```
# 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
   fefs-OST0000            18955        917113   86815
QoS command was completed.
```

上記の出力例では、io_time の avg が、load_limit_usec を指定することで、532015マイクロ秒から86815マイクロ秒に下がっています。これに伴って、低負荷ユーザーの uid=1070 の exec_time の avg は、203323マイクロ秒から28032マイクロ秒に下がり、uid=1070 のディスク待ち時間は短縮されています。

高負荷ユーザーの uid=1053 は、load_limit_usec を指定することで、ディスク高負荷時にスレッド数の上限値が抑えられるため、スレッドの待ち時間である wait_time の avg が 625マイクロ秒から 476457マイクロ秒に増加しています。

4.8 ファイルシステム不整合の修復

FEFSはジャーナリング機能を持っているため、通常はファイルシステムに不整合を生じることはありません。

ただし、ハードウェア故障などによってファイルシステムに不整合が生じる可能性があります。

このような場合、MGT、MDT、および OSTのマウント時にファイルシステム不整合が検出されます。

ファイルシステムの不整合が検出された場合、ファイルシステムの修復を行う必要があります。

ファイルシステムの修復は、fsck.ldiskfs コマンドで行います。

fsck.ldiskfs コマンドによる修復は、マウント時にファイルシステム不整合が検出されたボリュームに対してだけ実行することで修復できます。

- fsck.ldiskfs コマンドによる MGT の修復
- fsck.ldiskfs コマンドによる MDT の修復
- fsck.ldiskfs コマンドによる OST の修復
- lctl lfsck_start コマンドによる FEFS の修復

参照

fsck 関連コマンドの詳細は、"A.2.8 fsck.ldiskfs コマンド"、"lctl lfsck_start"、"lctl lfsck_stop" を参照してください。

以下に、MGT、MDT、および OST のファイルシステムの不整合を修復する例を示します。なお、これらの例は、以下の環境を想定しています。

表4.4 ファイルシステム修復作業のための環境例

ノードタイプ	ホスト名	利用概要	ボリューム名
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
			/dev/disk/by-id/scsi-36003005700a5388012be24d63761405f-part9

参考

ホスト名、ボリューム名の確認は、FEFS デザインシートの GFS シートで確認してください。

4.8.1 FEFS のサービス停止

ファイルシステムの修復を行う前には、FEFS サービスを停止する必要があります。FEFS サービスの停止方法は、"3.7 保守時の操作" を参照してください。

4.8.2 MGS 上での修復

MGS 上での修復手順です。

```
[mgs1ノード]
# /opt/FJSVfefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700a5a69012f2037006a14054-part6
```

4.8.3 MDS 上での修復

MDS 上での修復手順です。

```
[mds1ノード]
# /opt/FJSVfefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700adb66012c35b9708eb7a63-part6
[mds2ノード]
# /opt/FJSVfefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700adf5e012c35a3107cadfa1-part8
```

4.8.4 OSS上での修復

OSS 上での修復手順です。

```
[oss1ノード]
# /opt/FJSVfefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700a3b0f012fd9c6830808043-part6
# /opt/FJSVfefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700a3b0f012fd9c6830808043-part7
# /opt/FJSVfefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700a3b0f012fd9c6830808043-part8
# /opt/FJSVfefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700a3b0f012fd9c6830808043-part9
[oss2ノード]
# /opt/FJSVfefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700a5388012be24d63761405f-part6
# /opt/FJSVfefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700a5388012be24d63761405f-part7
# /opt/FJSVfefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700a5388012be24d63761405f-part8
# /opt/FJSVfefsprogs/sbin/fsck.ldiskfs -f -y /dev/disk/by-id/scsi-36003005700a5388012be24d63761405f-part9
```

4.8.5 FEFS の修復

MDT-OSTの矛盾をチェックし、FEFSの修復を行います。FEFSを修復するには、"lctl lfsck_start" コマンドを使用します。手順を以下に示します。

1) MGT のマウント

mgs1 上で以下を実行してください。

```
# systemctl start FJSVfefs
```

2) MDT のマウント

mds1 および mds2 上で以下を実行してください。

```
# systemctl start FJSVfefs
```

3) OST のマウント

oss1 および oss2 上で以下を実行してください。

```
# systemctl start FJSVfefs
```

4) クライアントのFEFSサービスの確認

すべてのクライアントでFEFSサービスが停止していることを確認します。

停止していないクライアントが存在する場合は、以下のコマンドを実行しサービスを停止します。

```
[クライアントノード]
# systemctl stop FJSVfefs
```

5) FEFSの修復

MDS 上でFEFS の修復を行います。

すべての MDS、OSS 上で MDT 、OST がマウントされていることを確認し、MDT0 をマウントする MDS 上で以下のコマンドを実行します。

ファイルシステムの修復が開始されると lctl lfsck_start コマンドは復帰しますが、非同期で修復処理が動作します。

```
[MDSノード]
# lctl lfsck_start -M <fsname>-MDT0000 -A
Started LFSCK on the device fefs-MDT0000: scrub layout namespace
```

※<*fsname*>は、FEFS のファイルシステム名を指定してください。

6) 修復の確認

ファイルシステムの修復の完了を確認する場合は、各 MDS、OSS ノード上で以下のコマンドを実行します。

status で "completed" が表示されていれば完了しています。

```
[MDSノード]
# lctl get_param osd-ldiskfs.<fsname>-MDT*.oi_scrub | grep -e status: -e =
osd-ldiskfs.<fsname>-MDT0000.oi_scrub=
status: completed                       # ファイルシステムの修復が完了した場合、"completed" が表示されます。
# lctl get_param mdd.<fsname>-MDT*.lfsck_namespace | grep -e status: -e =
mdd.fefs-MDT0000.lfsck_namespace=
status: completed                       # ファイルシステムの修復が完了した場合、"completed" が表示されます。
# lctl get_param mdd.<fsname>-MDT*.lfsck_layout | grep -e status: -e =
mdd.fefs-MDT0000.lfsck_layout=
status: completed                       # ファイルシステムの修復が完了した場合、"completed" が表示されます。
```

```
[OSSノード]
# lctl get_param osd-ldiskfs.<fsname>-OST*.oi_scrub | grep -e status: -e =
osd-ldiskfs.fefs-OST0000.oi_scrub=
status: completed                       # ファイルシステムの修復が完了した場合、"completed" が表示されます。
# lctl get_param obdfilter.<fsname>-OST*.lfsck_layout | grep -e status: -e =
obdfilter.fefs-OST0000.lfsck_layout=
status: completed                       # ファイルシステムの修復が完了した場合、"completed" が表示されます。
```

※<*fsname*> は、FEFS のファイルシステム名を指定してください。

7) クライアントのマウント

全 クライアント上で以下を実行してください。

```
# systemctl start FJSVfefs
```

4.9 ACL の設定方法

ACL の設定および ACL の情報の取得方法について説明します。これらの操作は、クライアントノードで行います。

- ACL の設定
  ACL の設定は、setfacl コマンドに -m オプションを指定して実行します。
  以下は、testfile ファイルに対してユーザー user1 に読込み、書込み、および実行権限を許可する例です。

```
[クライアントノード]
# setfacl -m user:user1:rwx testfile
```

  ACL の設定を削除する場合、setfacl コマンドに -x オプションを指定して実行します。
  以下は、上記で設定したACL を削除する例です。

```
[クライアントノード]
# setfacl -x user:user1: testfile
```

- ACL の情報取得
  設定されているACL情報は、以下のように getfacl コマンドを使用して取得します。

```
[クライアントノード]
# getfacl testfile
```

参照

setfacl コマンド、getfacl コマンドの詳細は、setfacl、getfacl コマンドのリファレンスマニュアルを参照してください。

4.10 user 拡張属性の設定方法

user 拡張属性の設定および user 拡張属性の情報の取得方法について setfattr/getfattr コマンドを利用する例にもとづいて説明します。これらの操作は、クライアントノードで行います。

- user 拡張属性の設定
  user 拡張属性の設定は、setfattr コマンドに -n オプションを指定して実行します。
  拡張属性の値も設定したい場合は、-v オプションを一緒に指定します。
  以下は、ファイル testfile に対して user 拡張属性 test に値 value0 を設定する例です。

```
[クライアントノード]
# setfattr -n user.test -v value0 testfile
```

- user 拡張属性の情報取得
  設定されている user 拡張属性情報は、以下のように getfattr コマンドを使用して取得します。

```
[クライアントノード]
# getfattr testfile              # testfileに設定されているuser拡張属性を表示する。
# file: testfile
user.test
# getfattr -n user.test testfile  # user拡張属性と値を合わせて表示する。
# file: testfile
user.test="value0"
```

- user 拡張属性の削除

  user 拡張属性の設定を削除する場合、setfattr コマンドに -x オプションを指定して実行します。
  以下は、上記で設定したuser 拡張属性を削除する例です。

```
[クライアントノード]
# setfattr -x user.test testfile
```

注意

setfattr コマンド、getfattr コマンドを利用するためには、実行するクライアントノードにattr パッケージを別途インストールする必要があります。

setfattr コマンド、getfattr コマンドの詳細は、setfattr、 getfattr コマンドのリファレンスマニュアルを参照してください。

4.11 FEFS の状態確認

ここでは、FEFS の状態確認方法について説明します。

pashowclst コマンドで、各ノードの FEFS のサービス状態を確認できます。

状態監視サービスには、FEFSSR サービスとFEFS サービスの2種類があります。

それぞれの状態監視対象ノードと状態監視項目を以下に示します。

表4.5 状態監視対象ノードと状態監視項目

監視サービス名	状態監視項目	状態監視対象ノード
FEFSSR	FEFS のサーバ機能	MGS ノード、MDS ノード、OSS ノード
	グローバルI/O ノードの中継機能	グローバルI/O ノード
	LLIO のサーバ機能	ストレージI/O ノード
FEFS	FEFS のクライアント機能	計算ノード、計算クラスタ管理ノード、ログインノード、多目的ノード

例) pashowclst コマンドの -n オプションで対象ノードのサービス状態を確認します。MDSの例を以下に示します。

```
[システム管理ノード]
# 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)
```

状態監視対象ノードで出力されるサービスの状態の意味を以下に示します。

表4.6 MGS、MDS、OSS ノードにおける FEFSSR 監視サービスの状態と意味

サービス状態	意味
o	監視項目がすべて正常
x	監視項目のどれかで異常
!	マルチパスドライバ、またはネットワークで縮退が発生
s	サービス初期化中
w	フェイルバックが可能
*	フェイルオーバが行われ、片寄せしている
f	フェイルオーバ処理を行っている

表4.7 グローバル I/O ノードにおける FEFSSR 監視サービスの状態と意味

サービス状態	意味
o	監視項目がすべて正常
x	FEFS サービスの停止、または異常
!	ネットワークで縮退または異常が発生
s	サービス初期化中
b	グローバルI/O ノードの中継機能が未設定状態

表4.8 ストレージ I/O ノードにおける FEFSSR 監視サービスの状態と意味

サービス状態	意味
o	監視項目がすべて正常
x	監視項目のどれかで異常
!	ネットワークで縮退または異常が発生
s	サービス初期化中
b	LLIOサーバが未設定状態

表4.9 計算ノード、計算クラスタ管理ノード、ログインノード、および多目的ノードにおける FEFS 監視サービスの状態と意味

サービス状態	意味
o	監視項目がすべて正常
x	監視項目のどれかで異常
!	ネットワークで縮退または異常が発生
s	サービス初期化中
a	SIO ノード、または GIO ノードの FEFSSR 監視サービスで異常があり使用できない
b	FEFS または LLIO クライアントが未設定状態

参照

SRV_STATUS の詳細については、「ジョブ運用ソフトウェア 管理者向けガイド システム管理編」の「システムの稼働状態の詳細表示」を参照してください。

4.12 フェイルオーバ

フェイルオーバはジョブ運用ソフトウェアと連携するという方法で実現します。

FEFS関連サービスに異常が発生した場合、各ノードに常駐しているFEFSサービス監視デーモンがジョブ運用ソフトウェアへ異常を通知し、自動的にフェイルオーバが発生します。

その際、異常発生ノードは停止状態となります。

異常が発生したノードはpashowclstコマンドの -dオプションで確認できます。

```
[システム管理ノード]
# pashowclst -c clstname -v -d
[ CLST: clstname ]
NODE        NODETYPE    STATUS      REASON      PWR_STATUS      ARCH_STATUS     SRV_STATUS
0xFFFF0003  MGS         SoftError   SrvDown     on              -               FEFSSR(x)
0xFFFF0004  MDS         SoftError   SrvDown     on              -               FEFSSR(x)
0xFFFF0005  OSS         SoftError   SrvDown     on              -               FEFSSR(x)
```

clstname: クラスタ名

異常発生ノードは、保守を実施後に片寄せされているノードからフェイルバックできます。

参照

保守時には手動で切り替えもできます。
フェイルオーバとフェイルバックについての詳細は、以下のマニュアルを参照してください。

「ジョブ運用ソフトウェア 管理者向けガイド 保守編」
「ジョブ運用ソフトウェア 管理者向けガイド システム管理編」

4.13 MDS の追加

運用中に MDS を新たに組み込む場合は、MDT をマウントする前に FEFS 情報を更新する必要があります。

以下は、MDS を追加する手順です。

参照

構築済みの MDS のデータを保護したい場合は、"4.19 構築済みファイルシステムのデータの保護" を参照してください。

1. FEFSデザインシートの更新
   初期導入時に作成した、または構築済みファイルシステムのデータの保護を実施したFEFSデザインシートに、MDS の情報を追加してください。
   設定方法は、"3.1.3 FEFS デザインシートの作成" を参照して作業を進めてください。

2. FEFS セットアップツール用構成定義ファイルの作成
   Excel のマクロでセットアップ用の入力データを作成できます。

   Excel マクロの「FEFS Design」>「Create config files」

   ダイアログが表示されますので、出力先フォルダを指定してください。指定したフォルダ配下にFEFSセットアップツール用構成定義ファイルが作成されます。

3. FEFS セットアップツール用構成定義ファイルの配置
   システム管理ノード上の /etc/opt/FJSVfefs/config 配下のディレクトリに、FEFS セットアップツール用構成定義ファイルを配置してください。

4. FEFS 設定ファイルの作成
   以下を実行してください。

   ```
   [システム管理ノード]
   # fefs_sync --setup  --storage=<cluster>[,…] --compute=<cluster>[,…]
   ```

   FEFS を構築するすべてのクラスタを指定してください。

   --storage : ストレージクラスタ名を指定してください。
   --compute : 計算クラスタ名および多目的クラスタ名を指定してください。

5. 追加MDSのボリュームフォーマット
   以下を実行してください。

   ```
   [システム管理ノード]
   # fefs_sync --mkfs --storage=<cluster> {--nodeid=<nodeid>|--nodelist=<nodeidlist>}
   ```

   追加する MDS のクラスタとノードID を指定してください。

   --storage : ストレージクラスタ名を指定してください。
   --nodeid : 追加するMDSのノードIDを指定してください。
   --nodelist : 追加するMDSのノードIDを列挙したファイルを指定してください。

6. 追加MDSの起動
   以下を実行してください。

   ```
   [システム管理ノード]
   # fefs_sync --start --storage=<cluster> {--nodeid=<nodeid>|--nodelist=<nodeidlist>}
   ```

   追加する MDS のクラスタとノードID を指定してください。

   --storage : ストレージクラスタ名を指定してください。
   --nodeid : 追加するMDSのノードID を指定してください。
   --nodelist : 追加するMDSのノードIDを列挙したファイルを指定してください。

7. FEFS の状態確認
   追加した MDS の FEFSSR のサービスが正常に起動されたことを pashowclst コマンドで確認してください。

```
[システム管理ノード]
# pashowclst -v --nodetype MDS
```

   FEFSSR の状態が FEFSSR(o) に遷移していれば、FEFSSR のサービスは正常に起動されています。

注意

MDS の動的削除は、未サポートです。

4.14 OSS の追加

運用中に OSS を新たに組み込む場合は、OST をマウントする前に FEFS 情報を更新する必要があります。

以下は、OSS を追加する手順です。

参照

構築済みの OSS のデータを保護したい場合は、"4.19 構築済みファイルシステムのデータの保護" を参照してください。

1. FEFS デザインシートの更新
   初期導入時に作成した、または構築済みファイルシステムのデータの保護を実施した FEFS デザインシートに、OSS の情報を追加してください。
   設定方法は、"3.1.3 FEFS デザインシートの作成" を参照して作業を進めてください。

2. FEFS セットアップツール用構成定義ファイルの作成
   Excel のマクロでセットアップ用の入力データを作成できます。

   Excel マクロの「FEFS Design」>「Create config files」

   ダイアログが表示されますので、出力先フォルダを指定してください。指定したフォルダ配下に FEFS セットアップツール用構成定義ファイルが作成されます。

3. FEFS セットアップツール用構成定義ファイルの配置
   システム管理ノード上の /etc/opt/FJSVfefs/config 配下のディレクトリに、FEFS セットアップツール用構成定義ファイルを配置してください。

4. FEFS 設定ファイルの作成

   以下を実行してください。

```
[システム管理ノード]
# fefs_sync --setup --storage=<cluster>[,…] --compute=<cluster>[,…]
```

   FEFS を構築するすべてのクラスタを指定してください。

   --storage : ストレージクラスタ名を指定してください。
   --compute : 計算クラスタ名および多目的クラスタ名を指定してください。

5. 追加 OSS のボリュームフォーマット
   以下を実行してください。

```
[システム管理ノード]
# fefs_sync --mkfs --storage=<cluster> {--nodeid=<nodeid>|--nodelist=<nodeidlist>}
```

   追加する OSS のクラスタとノードID を指定してください。

   --storage : ストレージクラスタ名を指定してください。
   --nodeid : 追加する OSS のノードID を指定してください。
   --nodelist : 追加する OSS のノードIDを列挙したファイルを指定してください。

6. 追加 OSS の起動
   以下を実行してください。

```
[システム管理ノード]
# fefs_sync --start --storage=<cluster>  {--nodeid=<nodeid> |--nodelist=<nodeidlist>}
```

   追加する OSS のクラスタとノードID を指定してください。

   --storage : ストレージクラスタ名を指定してください。
   --nodeid : 追加する OSS のノードID を指定してください。
   --nodelist : 追加するOSSのノードIDを列挙したファイルを指定してください。

7. FEFS の状態確認
   追加した OSS の FEFSSR のサービスが正常に起動されたことを pashowclst コマンドで確認してください。

```
[システム管理ノード]
# pashowclst -v --nodetype OSS
```

   FEFSSR の状態が FEFSSR(o) に遷移していれば、FEFSSR のサービスは正常に起動されています。

4.15 クライアントの追加

PG クライアントをあとから追加したい場合、以下の手順で構築できます。

1. FEFSデザインシートの更新

   初期導入時に作成した FEFS デザインシートに、PG クライアントの情報を追加してください。

   設定方法は、"3.1.3 FEFS デザインシートの作成" を参照して作業を進めてください。

2. FEFS セットアップツール用構成定義ファイルの作成

   以下の Excel のマクロでセットアップ用の入力データを作成できます。

   Excelマクロの「FEFS Design」>「Create config files」

3. FEFS セットアップツール用構成定義ファイルの配置

   運用系および待機系システム管理ノード上の以下のディレクトリに、FEFS セットアップツール用構成定義ファイルを配置してください。

   /etc/opt/FJSVfefs/config配下

4. FEFS 設定ファイルの作成

   システム管理ノードで、以下を実行してください。

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

   FEFS を構築するすべてのクラスタを指定してください。

   --storage : ストレージクラスタ名を指定してください。
   --compute : 計算クラスタ名を指定してください。

5. サービスの起動

   システム管理ノードで、以下を実行してください。

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

   --compute : 計算クラスタ名を指定してください。
   --nodelist : 追加したクライアントノードのノードID を列挙したファイルを指定してください。

参照

ノードの指定は、--nodelistの代わりに--nodeid オプションを指定することも可能です。詳細は "A.2.1 fefs_sync コマンド" を参照してください。

注意

多目的ノードにFEFSを追加した場合は、FEFSのセットアップ後に、以下の手順でジョブ運用ソフトウェアのサービス再起動が必要です。

```
[システム管理ノード]
# paclstmgr -c <cluster> --service  restart -n <nodeid>
```

<*cluster*>：クラスタ名を指定してください。
<*nodeid*>：多目的ノードのノードIDを指定してください。

paclstmgrコマンドの詳細は「ジョブ運用ソフトウェア 導入ガイド」を参照してください。

4.16 ファイルシステムの追加

FEFSのファイルシステムの追加は、以下の手順で行ってください。

参照

構築済みのデータを保護したい場合は、"4.19 構築済みファイルシステムのデータの保護" を参照してください。

1. FEFS デザインシートの更新
   初期導入時に作成した、または構築済みファイルシステムのデータの保護を実施したFEFSデザインシートにGFSシートを追加し、追加するファイルシステムの情報を入力してください。また、追加するMDSまたはOSSの情報をNODEシートに入力してください。
   設定方法は、"3.1.3 FEFS デザインシートの作成" を参照して作業を進めてください。

2. FEFS セットアップツール用構成定義ファイルの作成
   Excelのマクロでセットアップ用の入力データを作成できます。

   Excelマクロの「FEFS Design」>「Create config files」

   ダイアログが表示されますので、出力先フォルダを指定してください。指定したフォルダ配下にFEFSセットアップツール用構成定義ファイルが作成されます。

3. FEFS セットアップツール用構成定義ファイルの配置
   システム管理ノード上の /etc/opt/FJSVfefs/config ディレクトリ配下に、FEFSセットアップツール用構成定義ファイルを配置してください。

4. FEFS 設定ファイルの作成
   以下を実行してください。

   ```
   [システム管理ノード]
   # fefs_sync  --setup --storage=<cluster>[,…] --compute=<cluster>[,…]
   ```

   FEFSを構築するすべてのクラスタを指定してください。

   --storage：ストレージクラスタ名を指定してください。
   --compute：計算クラスタ名および多目的クラスタ名を指定してください。

5. 追加 MDT、OST のボリュームフォーマット

   以下を実行してください。

   ```
   [システム管理ノード]
   # fefs_sync --mkfs --storage=<cluster> --nodelist=<nodeidlist> --fsname=<fsname>
   ```

   追加するファイルシステムのクラスタとファイルシステム名を指定してください。

   --storage：ストレージクラスタ名を指定してください。
   --nodelist：追加するMDS、OSSのノードIDが列挙されたファイルを指定してください。
   --fsname：追加するファイルシステム名を指定してください。

注意

すでにフォーマットされて利用されているMDT、OST を追加する場合は、フォーマットする必要はありません。

6. FEFSサービス停止
   以下を実行してください。

```
[システム管理ノード]
# fefs_sync --stop --storage=<cluster> --compute=<cluster>
```

   --storage :ストレージクラスタ名を指定してください。
   --compute :計算クラスタ名を指定してください。

7. FEFSサービス起動
   以下を実行してください。

```
[システム管理ノード]
# fefs_sync --start --storage=<cluster> --compute=<cluster>
```

   --storage :ストレージクラスタ名を指定してください。
   --compute :計算クラスタ名を指定してください。

8. ファイルシステムのパーミッション変更
   任意の 1クライアント上でマウントポイントのパーミッションを設定してください。
   ファイルシステムがマウントされている状態で行ってください。
   (初期値は 755です。)

4.17 ファイルシステムの削除

FEFS のファイルシステムの削除は、以下の手順で行ってください。

注意

ファイルシステムを削除する前に必要なデータのバックアップをしてください。

1. FEFS デザインシートの更新
   FEFS デザインシートから削除するファイルシステムの GFS シートを削除してください。

2. FEFS セットアップツール用構成定義ファイルの作成
   Excel のマクロでセットアップ用の入力データを作成できます。

   Excel マクロの「FEFS Design」>「Create config files」

   ダイアログが表示されますので、出力先フォルダを指定してください。指定したフォルダ配下にFEFSセットアップツール用構成定義ファイルが作成されます。

3. FEFS セットアップツール用構成定義ファイルの配置
   システム管理ノード上の /etc/opt/FJSVfefs/config ディレクトリ配下に、FEFS セットアップツール用構成定義ファイルを配置してください。

4. FEFS 設定ファイルの作成
   以下を実行してください。

```
[システム管理ノード]
# fefs_sync --setup  --storage=<cluster>[,…] --compute=<cluster>[,…]
```

   FEFS を構築するすべてのクラスタを指定してください。

   --storage :ストレージクラスタ名を指定してください。
   --compute :計算クラスタ名および多目的クラスタ名を指定してください。

5. FEFSサービス停止
   以下を実行してください。

```
[システム管理ノード]
# fefs_sync --stop --storage=<cluster> --compute=<cluster>
```

   --storage :ストレージクラスタ名を指定してください。
   --compute :計算クラスタ名を指定してください。

6. FEFSサービス起動
   以下を実行してください。

```
[システム管理ノード]
# fefs_sync --start --storage=<cluster> --compute=<cluster>
```

   --storage :ストレージクラスタ名を指定してください。
   --compute :計算クラスタ名を指定してください。

4.18 ラック、BoB の追加

ラックまたはBoBの追加は、以下の手順で行ってください。

1. FEFS デザインシートの更新
   初期導入時に作成した FEFS デザインシートに ノードの情報を追加してください。
   設定方法は、"3.1.3 FEFS デザインシートの作成" を参照して作業を進めてください。

2. FEFS セットアップツール用構成定義ファイルの作成
   Excel のマクロでセットアップ用の入力データを作成できます。

   Excel マクロの「FEFS Design」>「Create config files」

   ダイアログが表示されますので、出力先フォルダを指定してください。指定したフォルダ配下に FEFS セットアップツール用構成定義ファイルが作成されます。

3. FEFS セットアップツール用構成定義ファイルの配置

   システム管理ノード上の /etc/opt/FJSVfefs/config ディレクトリ 配下に、FEFS セットアップツール用構成定義ファイルを配置してください。

4. FEFS 設定ファイルの作成
   以下を実行してください。

```
[システム管理ノード]
# fefs_sync --setup --storage=<cluster> --compute=<cluster>
```

   FEFS を構築するすべてのクラスタを指定してください。

   --storage : ストレージクラスタ名を指定してください。
   --compute : 計算クラスタ名を指定してください。

5. FEFSサービス起動
   以下を実行してください。

```
[システム管理ノード]
# fefs_sync --start --compute=<cluster> --nodelist=<nodeidlist>
```

   追加するラックまたはBoBのクラスタとノードのリストを指定してください。

   --compute : 計算クラスタ名を指定してください。
   --nodelist : 追加するノードのノードIDが列挙されたファイルを指定してください。

6. FEFSサービス起動
   追加したラックまたはBoBに搭載されたノードにおいて、FEFSのサービスが正常に起動されたことを確認します。
   以下を実行してください。

```
[システム管理ノード]
# pashowclst -v
```

   FEFSの状態が FEFSSR(o)および FEFS(o) に遷移したことを確認してください。

4.19 構築済みファイルシステムのデータの保護

構築済みファイルシステムが存在する環境に、ファイルシステム、MDS または OSS を追加する場合は、追加するボリュームに対してフォーマットが必要となるため、誤って構築済みファイルシステムを壊してしまう危険があります。これに備えて、構築済みのファイルシステムのデータを保護しておきます。

4.19.1 ファイルシステムのデータを保護する手順

以下は、構築済みのファイルシステムのデータを保護する手順です。

1. FEFS デザインシートの更新
   データを保護したいファイルシステムに対応する GFS シートの MKFS OPTION にある --reformat オプションを削除してください。

2. FEFS セットアップツール用構成定義ファイルの作成
   Excel のマクロでセットアップ用の入力データを作成できます。

   Excel マクロの「FEFS Design」>「Create config files」

   表示されるダイアログに、出力先フォルダを指定してください。

   指定したフォルダ配下に FEFS セットアップツール用構成定義ファイルが作成されます。

3. FEFS セットアップツール用構成定義ファイルの配置
   システム管理ノード上の /etc/opt/FJSVfefs/config ディレクトリ 配下に、FEFS セットアップツール用構成定義ファイルを配置してください。

4. FEFS 設定ファイルの作成

   以下を実行してください。

```
[システム管理ノード]
# fefs_sync --setup  --storage=<cluster>[,…] --compute=<cluster>[,…]
```

   FEFS を構築するすべてのクラスタを指定してください。

   --storage : ストレージクラスタ名を指定してください。
   --compute : 計算クラスタ名および多目的クラスタ名を指定してください。

注意

追加する MDS に接続されたディスク装置が、以前 LustreまたはFEFSでフォーマットされたことのある場合は、--reformat オプションを削除するとエラーになります。

--reformat オプションなしでセットアップを進めるためには、追加する MDS に、Lustre または FEFS でフォーマットされていないディスク装置を接続する必要があります。

4.19.2 ファイルシステムのデータの保護を解除する手順

以下は、構築済みファイルシステムのデータの保護を解除する手順です。

1. FEFSデザインシートの更新
   保護を解除したいファイルシステムに対応する GFS シートの MKFS OPTION に --reformat オプションを設定してください。

2. FEFS セットアップツール用構成定義ファイルの作成
   以下の Excel のマクロでセットアップ用の入力データを作成できます。

   Excelマクロの「FEFS Design」>「Create config files」

   ダイアログが表示されますので、出力先フォルダを指定してください。

   指定したフォルダ配下にFEFSセットアップツール用構成定義ファイルが作成されます。

3. FEFSセットアップツール用構成定義ファイルの配置
   システム管理ノード上の /etc/opt/FJSVfefs/config ディレクトリ配下に、FEFS セットアップツール用構成定義ファイルを配置してください。

4. FEFS設定ファイルの作成

   以下を実行してください。

```
[システム管理ノード]
# fefs_sync --setup  --storage=<cluster>[,…] --compute=<cluster>[,…]
```

   FEFS を構築するすべてのクラスタを指定してください。

   --storage : ストレージクラスタ名を指定してください。
   --compute : 計算クラスタ名および多目的クラスタ名を指定してください。

4.20 データ管理ツール (fefsbackup コマンド) の使い方 [PG]

本節ではデータ管理ツールの利用方法について述べます。

4.20.1 サブコマンド概要

fefsbackup コマンドのサブコマンドの機能概要を以下に示します。これらのサブコマンドは、特に断らない限り、FEFS クライアントで実行します。また、これらのサブコマンドは、管理者権限を持つユーザーだけが利用できます。

サブコマンドの詳細は "A.2.11 fefsbackup コマンド [PG]" を参照してください。

表4.10 サブコマンド機能概要

サブコマンド	機能
copy	コピー元ファイルシステム上のファイルをコピー先ファイルシステム上に転送
list	コピー情報の一覧を表示
delete	コピー情報の削除
status	現在実行中、およびエラー終了したコピー処理を表示

4.20.2 転送情報管理について

データ管理ツールは、copy 時にユニークなリクエストIDを管理情報として作成し、コピーしたファイルの情報を管理します。管理情報は設定ファイルで指定した保管先のノード上に保管します。

リクエストIDは、「日付_実行ノードのホスト名_連番」で自動生成します (例:2018/01/01 に hostAで 1回目の実行をした場合は 20180101_hostA_01)。ただし、-L オプションを指定した場合は、指定した文字列をリクエストIDとして使用します。

連番は同日中に実行した回数で、2桁です。1日に100回以上データ管理ツールを実行する場合は -L オプションでリクエストIDを指定して実行してください。

4.20.3 依存パッケージ

データ管理ツールを使用する場合は、実行するクライアントノードに以下のパッケージをインストールしてください。

```
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 設定ファイル

データ管理ツールで使用する設定ファイルの一覧を、以下に示します。

設定ファイルはインストール時には存在しないため、コピー元ノードに作成する必要があります。詳細は "4.20.6 事前準備" を参照してください。

表4.11 設定ファイル一覧

名称	概要
/etc/opt/FJSVfefs/fefsbackup/config/fefsbackup.conf	環境に依存しない共通処理に関連する設定の定義ファイル
/etc/opt/FJSVfefs/fefsbackup/config/fefsbackup_rsync.conf	ファイルシステム環境向けの処理に関する設定の定義ファイル

4.20.4.1 設定ファイル詳細

各定義ファイルの概要を説明します。

各設定項目は、以下のような形で指定します。

```
〈項目名〉 = 〈設定値〉
```

"=" の前後は半角スペース1文字が必要

使用しない項目名がある場合は項目名、設定値を削除してください。

fefsbackup.conf

環境に依存しない共通処理に関連する設定の定義ファイルです。設定項目について以下に示します。

表4.12 fefsbackup.conf 設定項目の概要

項目名	概要
WORK_DIR	作業用ディレクトリのルートディレクトリ(絶対パス)
WATCHDOG_INTERVAL	生存監視ファイルの更新間隔(秒) copyサブコマンドが実行中か否かを判断する為に、監視を行う周期。statusサブコマンドで情報を出力する為に使用する。
USE_TAR	tarアーカイブ転送を使用するか否かのフラグ
TAR_FILE_SIZE_MAX	tarファイルに格納するファイルサイズの最大値(バイト) ※
TAR_TOTAL_SIZE_MAX	1つのtarファイルに格納するファイルの合計サイズ(バイト) ※
TAR_FILE_NUM_MAX	1つのtarファイルに格納するファイル数 ※

※TAR_FILE_SIZE_MAX はtarアーカイブ化にするか否かのサイズの閾値。TAR_FILE_SIZE_MAX以下のファイルをtarアーカイブ化して転送します。

1つの tarアーカイブに含まれるファイルは、ファイルサイズの合計値が TAR_TOTAL_SIZE_MAX 以下またはファイル数が TAR_FILE_NUM_MAX 以下で決まります。どちらかの値を超えた場合、別の tar ファイルを作成し転送します。

表4.13 fefsbackup.conf の設定値

項目名	設定値	設定要否	デフォルト値
WORK_DIR	文字列(最大511文字)	必須	なし
WATCHDOG_INTERVAL	1～86,400の整数	任意	60
USE_TAR	0(無効) 1(有効)	任意	1
TAR_FILE_SIZE_MAX	0～2^63の整数	任意	1048576
TAR_TOTAL_SIZE_MAX	0～2^63の整数	任意	25000000000
TAR_FILE_NUM_MAX	1～2^63の整数	任意	1000000

fefsbackup_rsync.conf

ファイルシステム向けの処理に関連する設定の定義ファイルです。設定項目を以下に示します。

表4.14 fefsbackup_rsync.conf設定項目の概要

項目名	概要
ARCHIVE_HOST	管理情報の保管先ノードのIPアドレスまたはホスト名。Technical Computing Suite で管理されているノードを指定します。
BACKUP_ROOT	管理情報の保管先ノードのディレクトリパス。この配下に管理情報を保管します。
RSYNC_COMMAND_LOCAL	転送元ノードでのrsyncのパス。※
RSYNC_COMMAND_REMOTE	転送先ノードでのrsyncのパス。※
MULTI_PUT_MAX	ファイル転送の最大並列数。
SSH_PORT	sshで使用するポート番号。

※ 通常はデフォルトのパスにある rsync コマンドを実行しますが、特別に使用する場合(例えばバージョンが異なる rsync を使用したい場合など)、指定します。

表4.15 fefsbackup_rsync.conf の設定値

項目名	設定値	設定要否	デフォルト値
ARCHIVE_HOST	IPアドレスまたはホスト名	必須	なし
BACKUP_ROOT	文字列(最大1023)	必須	なし
RSYNC_COMMAND_LOCAL	文字列(最大1023)	任意	なし
RSYNC_COMMAND_REMOTE	文字列(最大1023)	任意	なし
MULTI_PUT_MAX	1～2^31の整数 (最大値:実行ノードのCPU数x2)	任意	1
SSH_PORT	0～65536の整数	任意	22

4.20.5 管理情報の保管先の設計

データ管理ツールを使用してファイルのコピーを行う場合、管理情報の保管先ノードに転送したファイルの情報を保管します。

保管先ノードで1つの管理情報で使用するディスク容量は以下です。

```
4096+((35+平均ファイル名長)*ファイル数)+(平均ディレクトリパス名長)
```

また一時作業領域として、1回のコピーで使用するディスク容量は最大で以下です。

```
(2 MB * MULTI_PUT_MAX)+((35+平均ファイル名長)*ファイル数+(35+ディレクトリパス名長)*ディレクトリ数)
```

複数のノード上でデータ管理ツールを実行する場合、設定ファイルのWORK_DIRにはFEFS上に設定してください。これにより、別のノードから、statusサブコマンドによる実行状態の確認や、異常時の再実行が可能となります。

4.20.6 事前準備

本項では、FEFSデータ管理ツールを実行するにあたり、事前に準備しておく必要がある項目を以下に示します。

- ポート設定
  転送元ノード、転送先ノード、および管理情報の保管先ノード上でsshで使用するポートを開けてください。ポート番号をデフォルト値(22)から変更する場合は設定ファイル(SSH_PORT)の変更をしてください。
- 管理用ディレクトリの作成
  コピーの管理情報の保管先ノード(設定ファイルの"ARCHIVE_HOST")に保管先のディレクトリ(設定ファイルの "BACKUP_ROOT" に指定したディレクトリ)をあらかじめ、作成しておきます。
- 作業用ディレクトリの作成
  データ管理ツールを実行するノード上で作業用ディレクトリ(設定ファイルの "WORK_DIR" に指定したディレクトリ)をあらかじめ、作成しておきます。
- 管理情報の保管先ノードのホストキー登録
  管理情報の保管先ノードに事前に接続してsshコマンドのknown_hostsに保管先ホストのホストキーを登録してください。また保管先ホストのホストキーが変更された場合も再登録が必要です。
- 管理者権限を持つユーザーのパスワード
  管理情報の保管先ノードと転送先ノードの管理者権限を持つユーザーのパスワードは同一にしてください。

4.20.7 データ管理ツールの設定

1. データ管理ツール設定ファイル作成

   データ管理ツールを実行するノード上で設定ファイルを作成してください。

   データ管理ツールの設定をするために設定ファイル "fefsbackup.conf"、"fefsbackup_rsync.conf" を作成します。

   設定ファイル "fefsbackup.conf"、"fefsbackup_rsync.conf" は、以下に配置してください。

   ```
   /etc/opt/FJSVfefs/fefsbackup/config/fefsbackup.conf
   /etc/opt/FJSVfefs/fefsbackup/config/fefsbackup_rsync.conf
   ```

   参照

   FEFS導入時点では "fefsbackup.conf"、"fefsbackup_rsync.conf" は存在していません。以下に示すサンプルファイルをコピーし、上記に示したパスに配置してください。

   "fefsbackup.conf" のサンプルファイル

   ```
   /etc/opt/FJSVfefs/fefsbackup/config/sample/fefsbackup.conf.sample
   ```

   "fefsbackup_rsync.conf" のサンプルファイル

   ```
   /etc/opt/FJSVfefs/fefsbackup/config/sample/fefsbackup_rsync.conf.sample
   ```

2. 設定内容の確認

   設定した内容を確認します。

   以下に設定例を表示します。

   設定ファイル"fefsbackup.conf" の表示例

   ```
   # cat /etc/opt/FJSVfefs/fefsbackup/config/fefsbackup.conf
   WORK_DIR  = /fefs/fefsbackup_workdir
   WATCHDOG_INTERVAL = 60
   ```

```
USE_TAR = 1
TAR_FILE_SIZE_MAX  = 1048576
```

設定ファイル "fefsbackup_rsync.conf" の表示例

```
# 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 サブコマンド

copy サブコマンドはファイルシステム間のファイルの転送を行います。以下に実行例を示します。

1. パスを指定して、同一ノード間のファイル転送を行います。

```
[クライアントノード][
# /opt/FJSVfefs/bin/fefsbackup copy -a -x -d /fefs/backupdir /home/projectA
password: ******** ※sshのパスワードの入力
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. -f オプションで指定した pathlist に記載されたファイル、ディレクトリを転送します。併せて転送情報を管理するリクエストID を指定します。

```
[クライアントノード]
# cat pathlist
/home/projectA/shell/test1.sh
/home/projectA/shell/test2.sh

# /opt/FJSVfefs/bin/fefsbackup copy -L projectA_Backup -f pathlist -n rhostA -d /fefs/backupdir
password: ******** ※sshのパスワードの入力
Copying request projectA_Backup is executing...
Total 2 files (2048 bytes) were copied from pathlist file.
Copying request projectA_Backup is completed successfully.
```

3. ユーザーが指定したリクエストIDで管理されるコピー済のデータに対し、指定したパス名配下の差分転送を行います。差分転送した情報は指定したリクエストIDで管理します。

```
[クライアントノード]
# /opt/FJSVfefs/bin/fefsbackup copy -v -u 20180604_lhostA_01 -n rhostA -d /fefs/backupdir /home/projectA
password: ******** ※sshのパスワードの入力
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. ユーザーが指定したリクエストIDで管理されるコピー済のデータに対し、指定した pathlist に記載されたファイル、ディレクトリについて差分転送を行います。

```
[クライアントノード]
# cat pathlist
/home/projectA/shell/test1.sh
/home/projectA/shell/test2.sh

# /opt/FJSVfefs/bin/fefsbackup copy -u 20180101_lhostA -f pathlist -n rhostA -d /fefs/backupdir
password: ******** ※sshのパスワードの入力
Copying request 20180101_lhostA is executing...
```

```
Total 2 files (2048 bytes) were copied from pathlist file..
Copying request 20180101_lhostA is completed successfully.
```

5. 転送処理がエラーなどで中断した場合、再実行可能であれば、リクエストIDを指定して再開します。

```
[クライアントノード]
# /opt/FJSVfefs/bin/fefsbackup copy -v -R 20180604_lhostA_01
password: ******** ※sshのパスワードの入力
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.
```

注意

- 再実行について
  ファイルシステムのアクセスエラーなどが発生した場合にデータ管理ツールは "Internal error has occurred" のメッセージとともにエラー終了します。この際、以下のメッセージが出力された場合、復旧後に -R オプションを使用してエラーとなったリクエストIDを指定し再度コピーを実行することでエラー箇所から処理を再開できます。

  ```
  Resume the command after checking the internal error cause.
  ```

  この際、-v オプション以外の指定内容については、再実行時の指定内容にかかわらず、初回実行時と同じ内容で実行します。

- コピー時のファイル属性(アクセス権、所有権、最終更新日時)について
  ファイルの属性(アクセス権、所有権、最終更新日時)は複写元ファイルの属性をコピーします。
  また、以下はコピーしません。

  ー プロジェクトID

  ー プロジェクトID継承フラグ

  ー ストライプディレクトリに関する情報

  ー ストライプに関する情報

- ハードリンクファイルについて
  ハードリンクのリンク情報のコピーは保証していません。
  設定ファイル fefsbackup.conf の設定値 TAR_FILE_SIZE_MAX 以下のファイルサイズのコピーでハードリンクが維持されることがあります。

- シンボリックリンクファイルについて
  シンボリックリンクの場合、参照されているファイルではなく、シンボリックリンクのファイルそのものをコピーします。

- 特殊ファイル、デバイスファイルについて
  名前付きソケットやFIFO、デバイスファイルなどの特殊ファイルもコピー対象となります。

- ファイル名について
  コピー対象に改行を含むファイル名またはディレクトリ名が存在するとエラーになります。

- オペランドの path または -f オプションの pathlist で指定するパス名について
  パス名の先頭に半角記号がある場合は、絶対パスで指定してください。または、先頭に半角記号を含まない親ディレクトリを指定してください。

- 転送処理がエラーになった場合のファイルパスについて
  ファイルのコピー失敗時のエラーメッセージのファイルのパスは、fefsbackup の copy サブコマンドの書式の path に絶対パスで指定しても、先頭の「/」が取り除かれたパスが出力されることがあります。
  先頭の「/」がパス名から取り除かれるのは、ファイルのコピーに失敗したファイルのサイズが設定ファイル fefsbackup.conf の設定値 TAR_FILE_SIZE_MAX に指定したサイズ以下の場合です。

・ コマンドの同時実行について
fefsbackup copy を -u オプションで同じリクエストIDを指定して同時に実行しないでください。

4.20.9 list サブコマンド

list サブコマンドは、コピーのリクエストID情報を出力します。以下に実行例を示します。

1. 引数なしで実行すると、リクエストID一覧を出力します。

```
[クライアントノード]
# /opt/FJSVfefs/bin/fefsbackup list
password: ******** ※sshのパスワードの入力
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    fefsbackup_test_data
20180628_hostA_01  copy     1K     1G     2018/06/28 13:21:51 shostA    dhostA    (fefsbackup_test_data.list)
```

引数なしで実行した場合の表示パラメーターの内容は以下の通りです。

表4.16 表示パラメーターの内容

パラメーター名	内容
Request ID	実施対象のリクエストID
Command	実行されたサブコマンド (コピー)
Files	コピーしたファイル数
Bytes	コピーしたバイト数
Backed-up	コピー指示をした時間 表示形式は YYYY/MM/DD hh:mm:ss
Src_Host	実行ノード名
Dest_Host	コピー先ノード名 (IPアドレス)
Target	コピー対象の path 括弧書きは各サブコマンド実行時に "-f" で指定した pathlist を表示

2. リクエストIDを指定すると、そのリクエストIDで管理されているファイルの一覧を出力します。

```
[クライアントノード]
# /opt/FJSVfefs/bin/fefsbackup list 20180709_hostA_01
password: ******** ※sshのパスワードの入力
dirA/fileA
dirB/fileA
dirB/fileB
```

3. -l オプションを付けると、詳細情報を出力します。

```
[クライアントノード]
# /opt/FJSVfefs/bin/fefsbackup list -l 20180709_hostA_01
password: ******** ※sshのパスワードの入力
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
```

リクエストID、-l オプション指定時の表示パラメーターの内容は以下の通りです。

表4.17 表示パラメーターの内容

パラメーター名	内容
アクセス権	1列目に該当ファイルのアクセス権を表示する。
オーナー	2列目に該当ファイルのオーナーを表示する。
グループ	3列目に該当ファイルのグループを表示する。

パラメーター名	内容
サイズ	4列目に該当ファイルのサイズをバイトで表示する。
更新日時	5列目に該当ファイルの更新日時を表示する。 表示形式はYYYY/MM/DD hh:mm
ファイル名	6列目に該当ファイル名を表示する。

4.20.10 delete サブコマンド

deleteサブコマンドは、指定されたコピーのリクエストIDを削除します。以下に実行例を示します。deleteサブコマンドでは管理情報を削除するのみでコピーしたファイルは削除されません。

1. 指定されたコピーのリクエストID を削除します。

```
[クライアントノード]
# /opt/FJSVfefs/bin/fefsbackup delete 20180604_lhostA_01
password: ******** ※sshのパスワードの入力
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 サブコマンド

status サブコマンドは、実行中、および、エラー終了したコピーのリクエストIDの一覧を出力します。リクエストIDが指定された場合は、そのリクエストIDの実行情報を出力します。以下に実行例を示します。

1. 引数なしで実行すると、そのユーザーの実行情報一覧を出力します。

```
[クライアントノード]
# /opt/FJSVfefs/bin/fefsbackup status
Request_id                          Command  Status
20180514_hostA_01                   copy     STOPPED
20180706_hostA_02                   copy     RUNNING
```

引数なしで実行した場合の表示パラメーターの内容は以下の通りです。

表4.18 表示パラメーターの内容

パラメーター名	内容
Request_id	リクエストID
Command	実行中またはエラー終了したサブコマンド
Status	実行状況。出力する状況は以下の通り。 ・RUNNING: 実行中の場合 ・STOPPED: システムトラブルなど内部処理異常で終了した場合

2. リクエストID を指定すると、そのリクエストIDの実行情報を出力します。

```
[クライアントノード]
# /opt/FJSVfefs/bin/fefsbackup status 20180706_hostA_02
RequestID  : 20180706_hostA_02
Command    : copy
Status     : RUNNING
Files      : 234
Bytes      : 45379
Started    : 2018/07/06 16:17:05
UserID     : 0
```

リクエストID 指定時の出力における表示パラメーターの内容は以下の通りです。

表4.19 表示パラメーターの内容

パラメーター名	内容
RequestID	実施対象のリクエストID
Command	実行中のサブコマンド
Status	実行状況。出力する状況は以下の通り。 ・RUNNING: 実行中の場合 ・STOPPED: システムトラブルなど内部処理異常で終了した場合
Files	転送完了済のファイル数
Bytes	転送完了済のバイト数
Started	実行を開始した時間。表示形式は YYYY/MM/DD hh:mm:ss
UserID	実行したユーザーID

4.21 JobStats機能

JobStatsは、ジョブプロセスごとに統計情報を記録する機能です。JobStats機能ではFEFSにアクセスするジョブプロセスの環境変数からジョブIDを取得し、そのジョブIDごとの統計情報を保存しています。保存された統計情報はメモリ上に蓄積されるため、JobStats機能により定期的にクリアされます。クリア間隔は以下のコマンドを実行することで確認できます。単位は秒です。

```
[MDSノード]
# cat /proc/fs/lustre/mdt/<fsname>-MDT<xxxx>/job_cleanup_interval
```

<fsname>：ファイルシステム名。
<xxxx>：MDTのインデックス番号。

```
[OSSノード]
# cat /proc/fs/lustre/obdfilter/<fsname>-OST<xxxx>/job_cleanup_interval
```

<fsname>：ファイルシステム名。
<xxxx>：OST のインデックス番号。

注意

- 複数計算クラスタで、それぞれの計算クラスタで独立したジョブIDが採番される環境においては、それらの独立したジョブIDが重複した場合に区別されず、同一のジョブIDとして統計情報が表示されます。
- JobStats 機能が使用するメモリ量は、10000本のジョブを実行した場合で MDS が約 12MB × MDT 数、OSS が 5MB × OST 数となります。
- 計算ノードでジョブ以外のプロセスがFEFSをアクセスした場合は、プロセス名.uidをジョブIDとして統計情報を収集します。

4.22 FEFS統計情報可視化機能 (fefssv.ph スクリプト) の利用方法

4.22.1 情報採取の方法

collectlの設定ファイル /etc/collectl.conf のcollectlデーモン起動オプション設定行(DaemonCommands)に以下を記述します。

```
[MDS ノードまたは OSS ノード]
DaemonCommands = -f <output dir> [<collectl option>] --import /opt/FJSVfefs/bin/fefssv.ph
```

<output dir>: 出力ディレクトリ名
<collectl option>: collectl コマンドのオプション

以下に設定例を示します。

```
[MDS ノードまたは OSS ノード]
DaemonCommands = -f /var/log/collectl -r00:00,7 -m -F60 -s+YZ -i10:60:300 --import /opt/FJSVfefs/bin/fefssv.ph
```

collectlサービスを自動起動させるために、以下を実行します。

```
[MDS ノードまたは OSS ノード]
# systemctl enable collectl
# systemctl start collectl
```

運用中に設定を変更する場合は、以下のようにcollectlサービスを再起動します。

```
[MDS ノードまたは OSS ノード]
# systemctl reload collectl
```

collectl コマンドのオプションについては、collectl(1) の man マニュアルを参照してください。

4.22.2 情報出力の方法

collectl のログファイルを入力にして、--import オプションで fefssv.ph スクリプトと表示したい内容を指定し、画面に表示をさせます。

```
[MDS ノードまたは OSS ノード]
$ collectl -p <data file> -s-all [<collectl option>] --import /opt/FJSVfefs/bin/fefssv.ph,{mdt|ost}
$ collectl -p <data file> -s-all [<collectl option>] --import /opt/FJSVfefs/bin/fefssv.ph,d,{mdt|ost}[,fs=<FS名>]
$ collectl -p <data file> -s-all [<collectl option>] --import /opt/FJSVfefs/bin/fefssv.ph,d,{mdt=<MDT名>|ost=<OST名>}
$ collectl -p <data file> -s-all [<collectl option>] ¥
--import /opt/FJSVfefs/bin/fefssv.ph,v,{mdt|ost}[,fs=<FS名>],[jobid=<ジョブid>]
$ collectl -p <data file> -s-all [<collectl option>] ¥
--import /opt/FJSVfefs/bin/fefssv.ph,v,{mdt=<MDT名>|ost=<OST名>}, [jobid=<ジョブid>]
```

<*data file*>: 出力ファイル名
<*collectl option*>: collectl コマンドのオプション
<*FS名*>: ファイルシステム名
<*MDT名*>: MDT名
<*OST名*>: OST名
<ジョブ*id*>: ジョブID
そのほか、fefssv.ph スクリプトのオプションについては、"4.22.3 オプションと出力情報" を参照してください。

以下に実行例を示します。

●MDSでの実行例

```
$ collectl -p /var/log/collectl/rx200-001-20160325-000000.raw.gz -s-all -oDm --from 11:03:20-11:06:20 --import (※)
/opt/FJSVfefs/bin/fefssv.ph,v,mdt=fefs-MDT0001

# Lustre Jobstats
#                              MDT_NAME      JOBID     open    close    mknod     link   unlink    mkdir    rmdir (※)
rename  getattr  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     1551        0 (※)
0       32        0        0        0        1        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
20160325 11:03:30.001     fefs-MDT0001         18        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
20160325 11:03:40.001     fefs-MDT0001         18        0        0        0        0        0    21837    40000 (※)
0    40004        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
20160325 11:03:50.001     fefs-MDT0001         18    40000    40000        0        0        0        0        0 (※)
0     5592        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
20160325 11:04:00.001     fefs-MDT0001        18        0        0        0        0    40000        5        4 (※)
0    34424        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
20160325 11:04:10.001     fefs-MDT0001        18        0        0        0        0        0    40000    34852 (※)
0    40012        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
20160325 11:04:20.001     fefs-MDT0001        18    34137    34134        0        0        0        0     5148 (※)
0        4        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
20160325 11:04:30.001     fefs-MDT0001        18     5863     5866        0        0    40000        5        4 (※)
0    40012        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
20160325 11:04:40.001     fefs-MDT0001        18        0        0        0        0        0    40000        4 (※)
0    40012        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
20160325 11:04:50.001     fefs-MDT0001        18    12468    12464        0        0        0        0    39996 (※)
0        4        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
20160325 11:05:00.001     fefs-MDT0001        18    27532    27536        0        0     6443        0        0 (※)
0    40004        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
20160325 11:05:10.001     fefs-MDT0001        18        0        0        0        0    33557    40005        4 (※)
0     9739        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
20160325 11:05:20.001     fefs-MDT0001        18        0        0        0        0        0        0    40000 (※)
0    30281        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
20160325 11:05:30.001     fefs-MDT0001        18    40000    40000        0        0        0        0        0 (※)
0    24632        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
20160325 11:05:40.001     fefs-MDT0001        18        0        0        0        0    40000     3950        4 (※)
0    15392        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
20160325 11:05:50.001     fefs-MDT0001        18        0        0        0        0        0    29325        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
20160325 11:06:00.001     fefs-MDT0001        18        0        0        0        0        0     6730    40000 (※)
0    40004        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
20160325 11:06:10.001     fefs-MDT0001        18    40000    40000        0        0        0        0        0 (※)
0    19208        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
20160325 11:06:20.001     fefs-MDT0001        18        0        0        0        0    40000        0        4 (※)
0    20804        0        0        0        0        0                0                0
```

備考) 紙面の都合で、上記表示例は (※) の箇所で改行しています。実際には、1行として表示されます。

●OSSでの実行例

```
$ collectl -p /var/log/collectl/rx200-003-20160325-000000.raw.gz -s-all -oDm --from 11:18:10-11:19:50 --import (※)
/opt/FJSVfefs/bin/fefssv.ph,v,ost=fefs-OST0001

# Lustre Jobstats
#                            OST_NAME      JOBID     read  read_bytes[B]    write write_bytes[B]  getattr  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
20160325 11:18:10.001     fefs-OST0001         18        0             0        0              0        0        0 (※)
0         0         0         0         0         0         0         0
20160325 11:18:10.001     fefs-OST0001         19        0             0      747      782237696        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
20160325 11:18:20.001     fefs-OST0001         18        0             0        0              0        0        0 (※)
0         0         0         0         0         0         0         0
20160325 11:18:20.001     fefs-OST0001         19        0             0      921      965738496        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
20160325 11:18:30.001     fefs-OST0001         18        0             0        0              0        0        0 (※)
0         0         0         0         0         0         0         0
20160325 11:18:30.001     fefs-OST0001         19        0             0      887      930086912        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
20160325 11:18:40.001     fefs-OST0001         18        0             0        0              0        0        0 (※)
0         0         0         0         0         0         0         0
20160325 11:18:40.001     fefs-OST0001         19        0             0      923      967835648        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
20160325 11:18:50.001     fefs-OST0001         18        0             0        0              0        0        0 (※)
0         0         0         0         0         0         0         0
20160325 11:18:50.001     fefs-OST0001         19        0             0      911      955252736        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
20160325 11:19:00.001     fefs-OST0001         18        0             0        0              0        0        0 (※)
0         0         0         0         0         0         0         0
20160325 11:19:00.001     fefs-OST0001         19        0             0      907      951058432        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
20160325 11:19:10.001     fefs-OST0001         18        0             0        0              0        0        0 (※)
0         0         0         0         0         0         0         0
20160325 11:19:10.001     fefs-OST0001         19        0             0      939      984612864        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
20160325 11:19:20.001     fefs-OST0001         18        0             0        0              0        0        0 (※)
0         0         0         0         0         0         0         0
20160325 11:19:20.001     fefs-OST0001         19        0             0      940      985661440        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
20160325 11:19:30.001     fefs-OST0001         18        0             0        0              0        0        0 (※)
0         0         0         0         0         0         0         0
20160325 11:19:30.001     fefs-OST0001         19        0             0      885      927989760        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
20160325 11:19:40.001     fefs-OST0001          18          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
20160325 11:19:50.001     fefs-OST0001          17          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
20160325 11:19:50.001     fefs-OST0001          19        209      219152384         0              0         0         0 (※)
0         0         0         0         0         0         0         0
```

備考) 紙面の都合で、上記表示例は (※) の箇所で改行しています。実際には、1行として表示されます。

参照

collectl のオプションについては、collectl(1) のリファレンスマニュアルを参照してください。

4.22.3 オプションと出力情報

本機能に指定可能な fefssv.ph のオプションは以下のとおりです。

表4.20 fefssv.ph オプション一覧

オプション	説明
v	出力形式の切り替えを行います。本オプションが指定された場合、詳細出力として各 MDT(OST) に対するジョブIDごとの統計情報を出力します。"d" オプションとは排他です。
d	出力形式の切り替えを行います。本オプションが指定された場合、ボリューム単位の統計情報を出力します。"v" オプションとは排他です。
mdt[=<MDT名>/<MDT名>/...]	出力情報の指定を行います。本オプションが指定された場合はMDTの統計情報を出力します。MDT 名が指定された場合は指定された MDT の統計情報だけを出力します。"/" 区切りで複数指定可能です。 本オプション、または "ost" の指定が必須です (ボリューム指定は任意)。 "ost" オプションとは排他です。 ボリューム指定をする場合、"v"、"d" オプションのどちらかの指定が必須です。
ost[=<OST名>/<OST名>/...]	出力情報の指定を行います。本オプションが指定された場合はOST の統計情報を出力します。OST 名が指定された場合は指定された OST の統計情報だけを出力します。"/" 区切りで複数指定可能です。 本オプション、または"mdt"の指定が必須です (ボリューム指定は任意)。 "mdt" オプションとは排他です。 ボリューム指定をする場合、"v"、"d" オプションのどちらかの指定が必須です。
fs=<FS名>[/<FS名>/...]	出力情報の指定を行います。本オプションが指定された場合、指定されたファイルシステムの統計情報だけを出力します。"/" 区切りで複数指定可能です。 "mdt"、"ost" オプションでボリューム指定がされていた場合、および "v"、"d" オプションのどちらも指定されていない場合、エラーになります。
jobid=<ジョブID>[/<ジョブID>/...]	出力情報を絞り込みます。本オプションが指定された場合、指定されたジョブの統計情報だけを出力します。"/"区切りで複数指定可能です。 "v" オプションの指定が必須です。

出力情報

本機能で出力する情報は以下のとおりです。各項目の集計単位についてはその時点での出力パターンによって異なります。

表4.21 MDSの出力情報

項目	内容
mdt_name	MDT名
jobid	ジョブID
open	単位時間内のopenの実行回数
close	単位時間内のcloseの実行回数
mknod	単位時間内のmknodの実行回数
link	単位時間内のlinkの実行回数
unlink	単位時間内のunlinkの実行回数
mkdir	単位時間内のmkdirの実行回数
rmdir	単位時間内のrmdirの実行回数
rename	単位時間内のrenameの実行回数
getattr	単位時間内のgetattrの実行回数
setattr	単位時間内のsetattrの実行回数
getxattr	単位時間内のgetxattrの実行回数
setxattr	単位時間内のsetxattrの実行回数
statfs	単位時間内のstatfsの実行回数
sync	単位時間内のsyncの実行回数
samedir_rename	単位時間内のsamedir_renameの実行回数
crossdir_rename	単位時間内のcrossdir_renameの実行回数

表4.22 OSSの出力情報

項目	内容
ost_name	OST名
jobid	ジョブID
read	単位時間内のreadの実行回数
read_bytes	単位時間内のread量(byte)
write	単位時間内のwriteの実行回数
write_bytes	単位時間内のwrite量(byte)
getattr	単位時間内のgetattrの実行回数
setattr	単位時間内のsetattrの実行回数
punch	単位時間内のpunchの実行回数
sync	単位時間内のsyncの実行回数
destroy	単位時間内のdestroyの実行回数
create	単位時間内のcreateの実行回数
statfs	単位時間内のstatfsの実行回数
get_info	単位時間内のget_infoの実行回数
set_info	単位時間内のset_infoの実行回数
quotactl	単位時間内のquotactlの実行回数

4.23 Lustre 接続 [PG]

FEFS と Lustre を接続するためには、以下の設定が必要です。

4.23.1 Lustre サーバと Lustre クライアントでの設定

FEFS と Lustre で通信を行うために、RPC のリクエスト受付ポート番号および RDMA のフラグメント数をそろえる必要があります。

Lustre サーバと Lustre クライアントの /etc/modprobe.d/lustre.conf ファイルに、以下の記述を追加してください。

```
options lnet accept_port=30988
options ko2iblnd service=30987
options ko2iblnd map_on_demand=16
```

4.23.2 Lustre クライアントから FEFS サーバのマウント

Lustre クライアントから FEFS サーバをマウントしてください。以下は、FEFS サーバの IB のアドレスが192.0.2.100 の場合の例です。

```
# mount -t lustre 192.0.2.100@o2ib:/<fsname1> /mountpoint
```

<fsname1>: マウントする FEFS サーバのファイルシステム名

4.23.3 FEFS クライアントから Lustre サーバのマウント

FEFS クライアントから Lustre サーバをマウントしてください。以下は、Lustre サーバの IB のアドレスが192.0.2.101 の場合の例です。

```
# mount -t lustre 192.0.2.101@o2ib:/<fsname2> /mountpoint
```

<fsname2>: マウントするLustreサーバのファイルシステム名

注意

FXサーバからのマウントはできません。

付録A　リファレンス

A.1　システムコール

FEFS で対応しているシステムコールの種類を以下の "表A.1 対応状況" に示します。

○：対応　△：勧告ロックだけ対応（強制ロックは非対応）　－：VFSレベルで対応　×：非対応　□：動作保証なし

表A.1 対応状況

システムコール	PG	FX
_llseek	○	○
access	○	○
bdflush	－	－
chdir	○	○
chmod	○	○
chown	○	○
chown32	○	○
chroot	○	○
close	○	○
creat	○	○
dup	－	－
dup2	－	－
execve	○	○
fchdir	○	○
fchmod	○	○
fchown	○	○
fchown32	○	○
fcntl	△	△
fcntl64	△	△
fdatasync	○	○
fgetxattr	○	○
flistxattr	○	○
flock	○	○
fremovexattr	○	○
fsetxattr	○	○
fstat	○	○
fstat64	○	○
fstatfs	○	○
fstatfs64	○	○
fsync	○	○
ftruncate	○	○
ftruncate64	○	○

システムコール	PG	FX
getdents	○	○
getdents64	○	○
getxattr	○	○
lchown	○	○
lchown32	○	○
lgetxattr	○	○
link	○	○
listxattr	○	○
llistxattr	○	○
lremovexattr	○	○
lseek	○	○
lsetxattr	○	○
lstat	○	○
lstat64	○	○
mkdir	○	○
mknod	○	○
mmap	○	○
mount	○	○
munmap	―	―
open	○	○
pipe	―	―
pivot_root	×	×
pread64	○	○
pwrite64	○	○
read	○	○
readdir	○	○
readlink	○	○
readv	○	○
removexattr	○	○
rename	○	○
rmdir	○	○
setrlimit	―	―
setxattr	○	○
stat	○	○
stat64	○	○
statfs	○	○
statfs64	○	○
swapon	□	□
swapoff	□	□

システムコール	PG	FX
symlink	○	○
sync	―	―
sysfs	○	×
truncate	○	○
truncate64	○	○
umount	○	○
umount2	○	○
unlink	○	○
utime	○	○
utimes	○	○
write	○	○
writev	○	○
FIEMAP ioctl	×	×
FIBMAP ioctl	×	×

A.2 コマンド

A.2.1 fefs_sync コマンド

【名前】

fefs_sync - FEFS 構築・更新を行うコマンド

【書式】

```
/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}
```

【説明】

fefs_sync は、FEFS を構築、更新するコマンドです。

FEFSを構築するサーバおよびクライアントに対して、FEFS構成定義ファイルの配布や、ディスクボリュームのフォーマット、マウントなどの操作を行います。システム管理ノード上から、指定したクラスタに対し、一括して操作できます。

システム管理ノードで実行してください。

操作オプションと共通オプションをそれぞれ指定してください。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

【操作オプション】

--setup [--directory=<*directory*>]

FEFSを構築するサーバおよびクライアントでFEFS設定ファイルを作成します。配布ファイルのディレクトリを変更したい場合は、--directoryオプションを指定してください。デフォルトのディレクトリは、/etc/opt/FJSVfefs/configです。

--mkfs

FEFSサーバ上でディスクボリュームのフォーマットを行います。

--mount

FEFSサーバ上ではディスクボリュームの、FEFSクライアント上ではFEFSクライアントのマウントを行います。

--umount

FEFSサーバ上ではディスクボリュームの、FEFSクライアント上ではFEFSクライアントのアンマウントを行います。

--start

FEFSサービスを起動します。

--stop

FEFSサービスを停止します。

--router --hostname=<*host*> --type=<*ope*>

ルーティング状態の変更を行います。

*host*にはルータのノード名を指定してください。

*ope*には以下の値を指定してください。

disable: 無効化する
enable: 有効化する
info: 状態表示する

--type=infoはノード保守時に異常が発生した際の調査情報の表示を目的としたオプションです。
担当保守員(SE)、または当社Support Deskより依頼があった際に実行してください。

【共通オプション】

--storage=<*cluster*>

ストレージクラスタ名を指定します。

--compute=<*cluster*>

計算クラスタ名および多目的クラスタ名を指定します。

--nodeid=<*nodeid*> [--siogrp|--giogrp]

ノードIDを指定します。--siogrpまたは--giogrp指定時は、指定したノードIDを含むSIOグループまたはGIOグループに対してコマンドを実行します。

--nodelist=<*nodelist*>

ノードIDが列挙されたファイルを指定します。ファイルには1行に1つのノードIDを記述してください。

--fsname=<*fsname*>

ファイルシステム名を指定します。

--nodetype=<*nodetype*>

ノードタイプを指定します。

--excludetype=<*nodetype*>

ノードタイプを指定します。本オプションを指定した場合、指定したノード種別にはコマンドが実行されません。

--nodegrp=<*nodegid*>

ノードグループIDを指定します。

--bootgrp=<*bootgid*>

ブートグループIDを指定します。

--model=<*model*>

機種を指定します。<*model*> には、以下の略称、またはユーザーが定義した任意機種を指定できます。

略称	機種
PG	PRIMERGY サーバ
FT	FX サーバ

--llio

操作の対象をLLIO に絞る場合に指定します。本オプションを指定した場合、FEFSへの操作は行われません。

【戻り値】

以下のステータスが返されます。

0: 正常終了

1: 異常終了

A.2.2 fefsconfig コマンド

【名前】

fefsconfig - FEFS 設定ファイルを生成するコマンド

【書式】

```
/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
```

【説明】

fefsconfig コマンドはFEFS 設定ファイルを生成するコマンドです。

FEFS セットアップツール用構成定義ファイルを元に、FEFS を構築するのに必要な、FEFS 設定ファイルを自動生成します。

本コマンドは、FEFS を構成するすべてのノード上で実行する必要があります。

また、FEFS セットアップツール用構成定義ファイルの作成もできます。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

【オプション】

--setup

FEFS 設定ファイルを作成します。

--make [MGS=<*mgs address*>]
[mount_option_mdt=<*mount option*>]
[mount_option_ost=<*mount option*>]
[mount_option_client=<*mount option*>]
[mount_point=<*mount point*>]

FEFS セットアップツール用構成定義ファイルを作成します。

指定可能なパラメーターについては以下のとおりです。

MGS=<*mgs address*>

接続するMGSのIBのIPアドレスを指定します。
MDS、OSS、クライアントで実行する場合はMGSのIPアドレス必須です。ただし、MGS と兼用しているノードでは不要です。

mount_option_mdt=<*mount option*>

MDT のマウントオプションに "*mount option*" で指定されたオプションを追加します。

デフォルト値は"defaults,retry=6"です。

mount_option_ost=<*mount option*>

OST のマウントオプションに "*mount option*" で指定されたオプションを追加します。

デフォルト値は "defaults,retry=6" です。

mount_option_client=<*mount option*>

クライアントのマウントオプションに "*mount option*" で指定されたオプションを追加します。

デフォルト値は "defaults,flock" です。

mount_point=<*mount point*>

FEFS のマウントポイントを指定します。

未指定の場合"/mnt/fefs"にマウントされます。

--mdtadd

FEFS セットアップツール用構成定義ファイルに追加する MDT の情報を追加します。

--ostadd

FEFS セットアップツール用構成定義ファイルに追加する OST の情報を追加します。

--cleanup

FEFS 設定ファイルを削除します。

本コマンドは、FEFS が不要になったノード上で実行する必要があります。

【戻り値】

以下のステータスが返されます。

0: 正常終了

1: 異常終了

A.2.3 fefs_mkfs コマンド

【名前】

fefs_mkfs - FEFS で利用するボリュームのフォーマットを行います。

【書式】

/sbin/fefs_mkfs {-a | -f *fsname* | *volume* }

【説明】

fefs_mkfs コマンドは、FEFS で利用するボリュームのフォーマットを行います。

本コマンドは、MGS、MDS、OSS 上で管理者権限を持つユーザーだけが利用できます。

【オプション】

-a

本コマンドを実行したノードで、FEFS が利用するすべてのフォーマットを行います。

ほかのオプションとの併用はできません。

-f *fsname*

本コマンドを実行したノードで、指定したファイルシステム名 (fsname) が利用するボリュームのフォーマットを行います。

ほかのオプションとの併用はできません。

volume

本コマンドを実行したノードで、指定したボリュームのフォーマットを行います。

ほかのオプションとの併用はできません。

【戻り値】

0: フォーマット成功

0以外: フォーマット失敗

A.2.4 fefs_mount コマンド

【名前】

fefs_mount - FEFS で利用するマウントポイントのマウントを行います。

【書式】

/sbin/fefs_mount { -a | -f *fsname* | *mountpoint* | *volume* }

【説明】

fefs_mount は、FEFS で利用するマウントポイントのマウントを行います。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

【オプション】

-a

本コマンドを実行したノードで、FEFS が利用するすべてのマウントポイントをマウントします。

ほかのオプションとの併用はできません。

-f *fsname*

本コマンドを実行したノードで、指定したファイルシステム名 (fsname) が利用するマウントポイントをマウントします。

ほかのオプションとの併用はできません。

mountpoint

本コマンドを実行したノードで、指定したマウントポイントをマウントします。

ほかのオプションとの併用はできません。

volume

本コマンドを実行したノードで、指定したボリュームをマウントします。

ほかのオプションとの併用はできません。

【戻り値】

以下のステータスが返されます。

0: 正常終了

0以外: 異常終了

A.2.5 fefssnap コマンド

【名前】

fefssnap - FEFS の調査に必要な資料の採取を行います。

【書式】

```
/usr/sbin/fefssnap -d <outputdir>
/usr/sbin/fefssnap --help
```

【説明】

fefssnap は FEFS の調査に必要な資料の採取を行います。

採取されたデータは tar + gzip の形式で圧縮され、以下の名前で指定したディレクトリに格納されます。

fefssnap_<タイムスタンプ>.tgz

※ タイムスタンプ : yyyymmddHHMMSS

本コマンドは、管理者権限を持つユーザーだけが利用できます。

【オプション】

-d <*outputdir*>

採取した資料を格納するディレクトリを指定します。

資料採取時、本オプションは必須オプションです。

--help

usage を表示して終了します。

【戻り値】

以下のステータスを返します。

0: 正常終了

1: 異常終了

A.2.6 lfsコマンド

lfs は FEFS のユーティリティコマンドで、以下のサブコマンドがあります。

lfs df

【名前】

lfs df - ディスク容量の使用状況を表示するコマンド

【書式】

```
/usr/bin/lfs df [-ih] [mount_point]
```

【説明】

そのノードに現在マウントされている各FEFSファイルシステム、または、パスが指定されている場合はパスを含むファイルシステムの使用情報をデフォルトで表示します。各 MDT と OST の現在の使用状況と合計を別々に表示し、各ファイルシステムの df (1) 出力に一致するファイルシステムごとの要約も表示します。

【オプション】

-i

inode の使用状況を表示します。

-h

読みやすい形式で表示します。

mount_point

ファイルシステムのマウントポイントを指定します。

lfs find

【名前】

lfs find - ファイル、ディレクトリを検索するコマンド

【書式】

```
/usr/bin/lfs find <dirname | filename>
                  [[!] {--name|-n} <pattern>]
                  [--obd <uuid>]
                  [--print0]
```

【説明】

与えられたパラメーターと一致するファイル、ディレクトリを検索します。

<*dirname*> で検索するディレクトリ、<*filename*> で検索するファイルを指定します。

【オプション】

[!] --name, -n <*pattern*>

名前が一致するファイル、ディレクトリを検索します。

!をつけた場合名前が一致しないファイル、ディレクトリを検索します。

--obd <*uuid*>

<*uuid*> で UUID を指定された OST 上のファイルを検索します。

--print0

ファイル名をフルパスで出力します。ファイル間は'¥0'(NULL文字)で区切ります。

lfs project

【名前】

lfs project - 指定したファイルまたはディレクトリのプロジェクト属性を変更または表示するコマンド

【書式】

```
/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...>
```

【説明】

ファイルまたはディレクトリに対し、プロジェクトIDとそのIDを継承するかを表す継承フラグを操作します。
本コマンドは、管理者権限を持つユーザーだけが利用できます。

lfs project [-d|-r] <*file*|*directory*...>

ファイルまたはディレクトリのプロジェクトID と継承フラグを表示します。

-d

ディレクトリ自身のプロジェクトIDと継承フラグを表示します。

-r

すべてのサブディレクトリのプロジェクトID と継承フラグを再帰的に表示します。

実行例

```
# 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
```

出力先頭の数字はプロジェクトIDを表します。Pは継承フラグが設定されている状態を表します。- は継承フラグが設定されていない状態を表します。

lfs project -s [-p *ID*] [-r] <*file*|*directory*...>

ファイルまたはディレクトリに対し、プロジェクトID と継承フラグを設定します。以後、指定した配下に作成される新しいファイルとサブディレクトリは、プロジェクトID と継承フラグを親から継承します。

注意

-rオプションを指定しない場合は、指定されたディレクトリとその直下のディレクトリまたはファイルにプロジェクトIDと継承フラグを設定します。
例えば、ディレクトリ /dir1 に対して -s オプション付き -r オプションなしで lfs project コマンドを実行すると、/dir1 直下にあるファイル、およびディレクトリ /dir1/dir2 には /dir1 と同じプロジェクトIDと継承フラグが設定されます。しかし、/dir1/dir2 配下のファイル、ディレクトリ /dir1/dir2/dir3、およびそれより深い階層にあるファイル・ディレクトリは影響を受けません。

-p *ID*

指定したファイルまたはディレクトリに対し、*ID* というプロジェクトIDを設定します。
プロジェクトID として指定できる値は 1 ～ 4294967295 です。0 を指定すると、プロジェクトID は無効となります。

注意

本コマンドは、上記 0 ～ 4294967295の領域を循環的に解釈します。従って、4294967296を指定して本コマンドを実行すると0 が設定され、プロジェクトIDは有効となりません。4294967297 を指定するとプロジェクトID は 1 となり、以下、4294967298 は 2、4294967299 は 3となります。また、-1を指定すると 4294967295が設定されます。

-r

サブディレクトリに対し、継承フラグを再帰的に設定します。
-p を指定した場合は、指定したプロジェクトIDをすべてのサブディレクトリに設定します。

lfs project -c [-d|-r [-p *ID*]] <*file*|*directory*...>

ファイルやディレクトリのプロジェクトIDと継承フラグをチェックし、異常値を出力します。

-p *ID*

指定したプロジェクトIDと異なるかをチェックします。-p の指定がない場合は、指定したディレクトリのプロジェクトIDに対してチェックします。指定したプロジェクト ID と異なる場合の出力例を以下に示します。

```
dir1/dir2/file1 - project identifier is not set (inode=6000, tree=6001)
```

-d

ディレクトリのプロジェクトIDと継承フラグをチェックします。継承フラグが設定されていない場合の出力例を以下に示します。

```
dir1/dir2/file1 - project inheritance flag is not set
```

-r

サブディレクトリに対し、プロジェクトIDと継承フラグを再帰的にチェックします。

lfs project -C [-r|-k] <*file*|*directory*...>

ファイルまたはディレクトリに設定されているプロジェクトID と継承フラグをクリアします。指定されたディレクトリとその直下のディレクトリまたはファイルのプロジェクトID と継承フラグをクリアします。それより深い階層にあるディレクトリまたはファイルのプロジェクトID と継承フラグはクリアされません。

-r

指定したディレクトリおよびサブディレクトリに対し、プロジェクトID と継承フラグを再帰的にクリアします。

-k

継承フラグだけをクリアします。

lfs quota

【名前】

lfs quota - ディスクの使用状況と使用限度、または QUOTA のソフトリミットの猶予期間を表示するコマンド

【書式】

```
/usr/bin/lfs quota [-v] [-u uname| -g gname| -p pid] <mount_point>
/usr/bin/lfs quota -t {-u|-g|-p} <mount_point>
```

【説明】

ディスクの使用状況と使用限度を表示します。

オプションに "-u" を指定した場合はユーザー名 *uname*、オプションに "-g" を指定した場合はグループ名 *gname*、オプションに "-p" を指定した場合は表示対象のプロジェクトID *pid*を指定します。オプションに "-u"、"-g"、"-p" のどれも指定しない場合は、lfs コマンドを実行したユーザー、および、所属するグループのディスクの使用状況と使用限度を表示します。

"-v" を付けて実行すると、MDT および OST 単位のディスクの使用状況と使用限度を表示します。

"-t" を付けて実行すると、QUOTA のソフトリミットの猶予期間を表示します。

<mount_point> はファイルシステムのマウントポイントを指定します。

【オプション】

-u

ユーザーの QUOTA の使用状況、ソフトリミットの猶予期間を表示します。

-g

グループの QUOTA の使用状況、ソフトリミットの猶予期間を表示します。

-p

プロジェクトの QUOTA の使用状況、ソフトリミットの猶予期間を表示します。

lfs setquota

【名前】

lfs setquota - QUOTAの設定を行うコマンド

【書式】

```
/usr/bin/lfs setquota {-u|-g|-p|} <name>
                      [--block-softlimit <block-soft/limit>]
                      [--block-hardlimit <block-hard/limit>]
                      [--inode-softlimit <inode-soft/limit>]
                      [--inode-hardlimit <inode-hard/limit>]
                      <mount_point>
/usr/bin/lfs setquota {-u|-g|-p|} <name>
                      [-b <block-soft/limit>]
```

```
                        [-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>
```

【説明】

QUOTA の設定を行います。

<*name*> はオプションに "-u" を指定した場合はユーザー名、オプションに "-g" を指定した場合はグループ名、オプションに "-p" を指定した場合は操作対象のプロジェクトIDを指定します。

<*block-softlimit*> は使用ブロック数のソフトリミットを設定します。KiB 単位で指定します。

<*block-hardlimit*> は使用ブロック数のハードリミットを設定します。KiB 単位で指定します。

<*inode-softlimit*> は inode 数のソフトリミットを設定します。

<*inode-hardlimit*> は inode 数のハードリミットを設定します。

"-t" を付けて実行すると、QUOTA のソフトリミットの猶予期間の設定を行います。

<*block-grace*> は、使用ブロック数のソフトリミットの猶予期間を設定します。猶予期間は、秒単位で指定するか、XXwXXdXXhXXmXXs のフォーマットで指定します。1w4d とすれば「1週間と4日」の意味になります。なお、猶予期間に 18446744073709551614秒を超える値は指定できません。また、猶予期間のデフォルト値は 7日です。

<*inode-grace*> は、inode数のソフトリミットの猶予期間を設定します。猶予期間は、秒単位で指定するか、XXwXXdXXhXXmXXs のフォーマットで指定します。1w4dとすれば「1週間と4日」の意味になります。なお、猶予期間に 18446744073709551614秒を超える値は指定できません。また、猶予期間のデフォルト値は 7日です。

<*block-grace*> と <*inode-grace*> の片方だけ指定した場合、指定しなかった側の猶予期間は変更されません。

<*mount_point*> はファイルシステムのマウントポイントを指定します。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

【オプション】

-u

ユーザーの QUOTA の設定を行います。

-g

グループの QUOTA の設定を行います。

-p

プロジェクトの QUOTA の設定を行います。

lfs setstripe

【名前】

lfs setstripe - ストライプパターンを設定するコマンド

【書式】

```
/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>
```

【説明】

lfs setstripe コマンドはストライプパターンを持った新規ファイルを作成、または既存のディレクトリのストライプパターンを設定します。<*dirname*> を指定することでディレクトリのストライプパターンを指定します。<*filename*> を指定することでファイルのストライプパターンを指定します。

【オプション】

--stripe-size,-S

ストライプサイズを設定します。

-S #k, -S #m, -S #g とすることで、サイズを KiB、MiB、GiB 単位で設定できます。

--stripe-count,-c

ストライプカウントを設定します。-1と設定した場合、すべての OST に書込みが行われます。

--stripe-index,-i

ファイル書込みを開始する OST を指定します。-1とした場合、ファイル書込みを開始する OST はランダムに選ばれます。

--pool,-p

OST_pool のストライプパターンを指定する場合に使用します。

lfs getstripe

【名前】

lfs getstripe - ストライプパターンの情報を表示するコマンド

【書式】

/usr/bin/lfs getstripe [--mdt-index|-M] <*dirname*|*filename*> ...

【説明】

指定したファイルやディレクトリのストライプパターンの情報を表示します。<*dirname*>を指定することでディレクトリのストライプパターンを表示します。<*filename*> を指定することでファイルのストライプパターンを表示します。

【オプション】

--mdt-index,-M

指定したファイルやディレクトリの MDT インデックス番号を表示します。

表示する内容は、指定したファイルやディレクトリを管理している MDT のインデックス番号です。

```
[クライアント]
$ lfs getstripe -M something
1
```

lfs getdirstripe

【名前】

lfs getdirstripe - ディレクトリのストライプパターンを一覧表示するコマンド

【書式】

/usr/bin/lfs getdirstripe [--mdt-count|-c] [--mdt-index|-i] [--recursive|-r] [{--obd|-O} <*uuid*>] <*dir*>...

【説明】

<*dir*> で指定したディレクトリのストライプパターン情報を獲得します。<*dir*> は複数指定可能です。

【操作オプション】

-c, --mdt-count

ディレクトリのストライプカウントだけを表示します。

-i, --mdt-index

ディレクトリのストライプインデックスだけを表示します。

-r, --recursive

指定されたディレクト内のすべてのサブディレクトリを再帰的にたどり、ストライピング情報を一覧表示します。

-O, --obd <*uuid*>

表示するストライプ情報を、<*uuid*> で UUID を指定された MDT 上にあるディレクトリのものに限定します。

lfs mkdir

【名前】

lfs mkdir - MDT 上にディレクトリを作成するコマンド

【書式】

/usr/bin/lfs mkdir [{-c | --count} <*stripe_count*>] [{-i | --index} <*mdt_idx*>] <*dir*>...

【説明】

MDT 上に <*dir*> で指定したディレクトリを作成します。

このコマンドはクライアントノードで実行してください。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

作成したディレクトリを削除する場合は、通常のディレクトリ削除と同様に rmdir コマンド (/bin/rmdir) で削除可能です。

【オプション】

-i,--index <*mdt_index*>

<*mdt_index*> を開始番号とする MDT に、ディレクトリを作成します。-c オプションが指定されていなければ、本オプションは必須です。

-c,--count <*stripe_count*>

ストライプカウントが <*stripe_count*> であるようなストライプディレクトリを作成します。-i オプションが指定されていなければ、本オプションは必須です。

lfs pool_list

【名前】

lfs pool_list - OST_pool のリスト、および OST_pool に登録された OST を表示するコマンド

【書式】

/usr/bin/lfs pool_list <*fsname*>[.<*poolname*>] | <*mount_point*>

【説明】

<*fsname*> により定義された OST_pool のリストを表示します。

<*fsname*>[.<*poolname*>] により定義された pool に含まれる OST リストを表示します。

<*mount_point*> はファイルシステムのマウントポイントを指定します。

OST_pool とは指定した複数の OST を束ねて1つのグループとし、ファイルやディレクトリをグループ内の OST に割り当てる機能です。

lfs fid2path

【名前】

lfs fid2path - FID に対応するファイルパス名を出力するコマンド

【書式】

/usr/bin/lfs fid2path [--link <*linkno*>] <*fsname*|*rootpath*> <*fid*>...

【説明】

指定された<*fid*> に対応する、<*rootpath*> でマウントされるか、または <*fsname*> という名前のファイルシステムのパス名を表示します。ファイルに複数のハードリンクがある場合は、--linkが指定するリンク番号だけに出力を制限しない限り、そのファイルのすべてのパス名を出力します(順不同で0から始まります)。複数の <*fid*> が指定されていても、各ファイルに必要なパス名が1つだけの場合は、--link 0 を使用してください。

このコマンドは root 権限で実行する必要があります。

lfs help

【名前】

lfs help - lfs ヘルプを表示します。

【書式】

```
/usr/bin/lfs help [command]
```

【説明】

lfs ヘルプを表示します。

【オプション】

[command]

ヘルプを表示するコマンドを指定します。

lfs --list-commands

【名前】

lfs --list-commands - サブコマンドの一覧を表示します。

【書式】

```
/usr/bin/lfs --list-commands
```

【説明】

サブコマンドの一覧を表示します。

A.2.7 lctlコマンド

lctl は FEFS の低レベル構成制御コマンドで、以下のサブコマンドがあります。

lctl device_list

【名前】

lctl device_list -ファイルシステム構成情報を表示します。

【書式】

```
/usr/sbin/lctl device_list [-t]
```

【説明】

ファイルシステムの構成情報を表示します。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

【オプション】

-t

クライアントノードで実行した場合、NID を付加して出力します。

lctl list_nids

【名前】

lctl list_nids - 有効な NID を表示します。

【書式】

/usr/sbin/lctl list_nids

【説明】

当該ノードから通信可能なサーバの NID をすべて表示します。表示例を以下に示します。

```
[MDSノード、OSSノード、クライアント]
# 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
```

本コマンドは、管理者権限を持つユーザーだけが利用できます。

lctl qos

【名前】

lctl qos - QoS 機能に関する設定を行います。

【書式】

/usr/sbin/lctl qos < on [*filepath*] | off | stat | check [*filepath*] >

【説明】

QoS 機能の設定を行います。このコマンドは MDS 上で実行してください。また、マルチ MDS 環境では、MDT0 上で実行してください。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

【引数】

on [*filepath*]

指定された QoS 定義ファイルで、QoS 制御を開始します。

filepath は絶対パスで指定します。*filepath* のデフォルトは /etc/opt/FJSVfefs/qosserver.conf です。

off

QoS 制御を終了します。

stat

1行目に QoS の状態を表示します。

"QoS is Enable." QoS 機能は有効です。

"QoS is Disable." QoS 機能は無効です。

QoS 機能が有効の場合、2行目には QoS 定義ファイルのパス名を表示し、3行目以降には QoS 定義ファイルの内容を表示します。

check [*filepath*]

指定された QoS 定義ファイルの構文チェックをします。

filepath は絶対パスで指定します。filepath のデフォルトは /etc/opt/FJSVfefs/qosserver.confです。

lctl sqos

【名前】

lctl sqos - サーバ側の QoS 機能に関する情報表示を行います。

【書式】

/usr/sbin/lctl sqos [oss] < thread_top | thread_all | thread_user <*user-id*> | thread_node | ost_io | clear >

【説明】

サーバ側のQoS機能の情報表示を行います。このコマンドは、MDSの情報を表示する時は、MDS上で実行してください。また、マルチMDS環境では、それぞれのMDS上で実行してください。OSSの情報を表示する時は、OSS上で実行してください。

表示される情報はQoS機能を有効にしてからの総計ですが、最後にファイルアクセスを行った日から30日を超えているユーザーについての情報は表示されません。また、ファイルアクセスを行ったユーザー数が1000人を超える場合は、1000人までの情報が表示対象となります。

このコマンドはサーバ側のQoS機能が無効の時は使用できません。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

【オプション】

oss

MDSとOSSが同一マシンの構成で、OSSの情報を表示する場合は、本オプションを指定してください。

【引数】

thread_top

クライアントからのリクエスト数が多いユーザーの情報を、多い順に10人まで表示します。

thread_all

全ユーザーの情報を表示します。表示順序は順不同です。

thread_user <*user-id*>

指定したユーザーの情報を表示します。

thread_node

QoS定義ファイルで定義したノード群ごとの情報を表示します。

ost_io

OST単位のアクセス状況を表示します。

このオプションはOSS上でコマンド実行する時だけ有効です。

clear

サーバ側のQoS機能の統計情報をクリアします。

統計情報はサーバ側でQoS機能を無効 (lctl qos off コマンド) にすることでもクリアされます。

thread_top の表示例

MDS での実行例

```
# 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   1079        1446260     0   2  12       0       2        638      18      48516      77  2013/08/14 09:34:49
 2   1078         766438     0   2  12       0       2        361      20      38156     102  2013/08/14 09:34:49
 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
 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.
```

OSS での実行例

```
# 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.
```

※各項目の説明は以下です。

項目名	説明
nodegrp=	QoS 定義ファイルの nodegrp[1-10] で定義したノード群に対応する番号です。 この番号が "undef" の場合は、QoS 定義ファイルで定義していないノードからのアクセスがあります。QoS 定義ファイルの nodegrp[1-10] で定義した IP アドレスの見直しを行ってください。
No.	exec_cnt の大きい順に10人まで表示します。
uid	ユーザー ID です。
exec_cnt	クライアントからのリクエストをサーバスレッドが実行した回数です。
thread	実行中のサーバスレッドの数です。 cur : 上記の現在値です。 max : 上記の最大値です。 lim : 実行可能なサーバスレッド数の上限値です。この値は、QoS 定義ファイルで指定したサーバスレッド数の最大値(割合)によって決定します。
wait_req	サーバスレッドの実行待ちとなっているリクエストの数です。 cur : 上記の現在値です。 max : 上記の最大値です。
wait_time(usec)	クライアントからのリクエストがサーバに到着してから、サーバスレッドによって実行が開始されるまでの待ち時間です。単位はマイクロ秒です。 max : 上記の最大値です。 avg : 上記の平均値です。
exec_time(usec)	クライアントからのリクエストがサーバスレッドによって実行された時の実行時間です。単位はマイクロ秒です。

項目名	説明
	max : 上記の最大値です。 avg : 上記の平均値です。
last_update	サーバスレッドを実行した最終日時です。

thread_all の表示例

MDS での実行例

```
# 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.
```

※各項目の説明は、前述の lctl sqos thread_top と同じです。

OSSでの実行例

```
# 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                1     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                2     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.
```

※各項目の説明は、前述の lctl sqos thread_top と同じです。

thread_user の表示例

MDS での実行例

```
# 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
QoS command was completed.
```

※各項目の説明は、前述の lctl sqos thread_top と同じです。

OSS での実行例

```
# 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
   2                   0     0   0 128       0       0          0        0          0        0
QoS command was completed.
```

※各項目の説明は、前述の lctl sqos thread_top と同じです。

thread_node の表示例

MDS での実行例

```
# 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.
```

※各項目の説明は、前述の lctl sqos thread_top と同じです。

OSS での実行例

```
# 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.
```

※各項目の説明は、前述の lctl sqos thread_top と同じです。

ost_io の表示例

OSS での実行例

```
# 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.
```

※各項目の説明は以下です。

項目名	説明
ost_name	ディスクアクセスを行った OST の名前です。
io_cnt	ディスクアクセスの実行回数です。
io_time(usec)	ディスクアクセス 1回あたりにかかった時間です。単位はマイクロ秒です。 max : 上記の最大値です。 avg : 上記の平均値です。

lctl cqos

【名前】

lctl cqos - クライアント側の QoS 機能に関する情報表示を行います。

【書式】

/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*>

【説明】

クライアント側の QoS 機能の情報表示を行います。このコマンドは、クライアントノード上で実行してください。

表示される情報は QoS 機能を有効にしてからの総計ですが、最後にファイルアクセスを行った日から 30日を超えているユーザーについての情報は表示されません。また、ファイルアクセスを行ったユーザー数が 1000人を超える場合は、1000人までの情報が表示対象となります。

このコマンドはクライアント側の QoS 機能が無効の時は使用できません。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

【引数】

meta_top

メタ操作の実行数が多いユーザーの情報を、多い順に 10人まで表示します。

read_top

readの実行数が多いユーザーの情報を、多い順に 10人まで表示します。

write_top

writeの実行数が多いユーザーの情報を、多い順に 10人まで表示します。

meta_all

メタ操作を行った全ユーザーの情報を表示します。表示順序は順不同です。

read_all

readを行った全ユーザーの情報を表示します。表示順序は順不同です。

write_all

writeを行った全ユーザーの情報を表示します。表示順序は順不同です。

meta_user <*user-id*>

指定したユーザーについて、メタ操作の情報を表示します。

read_user <*user-id*>

指定したユーザーについて、read の情報を表示します。

write_user <*user-id*>

指定したユーザーについて、write の情報を表示します。

cache_top

書込み用キャッシュの使用数が多いユーザーの情報を、多い順に 10人まで表示します。

cache_all

書込み用キャッシュを使用した全ユーザーの情報を表示します。表示順序は順不同です。

cache_user <*user-id*>

指定したユーザーについて、書込み用キャッシュの情報を表示します。

clear

クライアント側の QoS 機能 (リクエスト制御、キャッシュ制御) の統計情報をクリアします。

mount-point

FEFS のマウントポイントを指定します。

meta_top の出力例

```
# lctl cqos meta_top /mnt/fefs/
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    1079         1196513       95267         655      53656      174       5536      118  2013/09/04 19:33:19
 2    1078          494573       50474         717      43589      192      31710      311  2013/09/04 19:33:19
 3    1071          146154        2459         525      41067      338      50073      298  2013/09/04 19:33:19
 4    1072          112171          12         503       9534     1069      33345      308  2013/09/04 19:33:19
 5    1073           98563           9         500       1016      248      33034      321  2013/09/04 19:33:19
 6    1074           38232          19         356       1645      383      30111      210  2013/09/04 19:33:19
 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 19:33:16
10    1075             116           0           0          0        0          0        0  2013/09/04 19:32:35
CQoS command was completed.
```

※各項目の説明は以下です。

項目名	説明
mclientmax=4 mrootmax=1 musermax=1	FEFS デザインシートで指定した qos オプションの値です。各オプションの意味については、「FEFSクライアントの設定方法」を参照してください。
No.	exec_cnt の大きい順に 10人まで表示します。
uid	ユーザーID です。
exec_cnt	リクエストの実行回数です。
total_wait_cnt	リクエストを実行するまでに、QoS 機能で実行を待ち合わせた回数です。 own : 1ユーザーあたりに同時発行可能なリクエスト数の上限に到達したために、リクエストの実行を待ち合わせた回数です。 other : クライアントノード内で同時発行可能なリクエスト数の上限に到達したために、リクエストの実行を待ち合わせた回数です。
own_wtime(usec)	1ユーザーあたりに同時発行可能なリクエスト数の上限に到達した時に、リクエストの実行を待ち合わせた時間です。単位はマイクロ秒です。 max : 上記の最大値です。 avg : 上記の平均値です。
other_wtime(usec)	クライアントノード内で同時発行可能なリクエスト数の上限に到達した時に、リクエストの実行を待ち合わせた時間です。単位はマイクロ秒です。 max : 上記の最大値です。 avg : 上記の平均値です。
last_update	リクエストを実行した最終日時です。

read_top の出力例

```
# 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.
```

※各項目の説明は、前述の lctl cqos meta_top と同じです。

write_top の出力例

```
# 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   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.
```

※各項目の説明は、前述の lctl cqos meta_top と同じです。

meta_all の出力例

```
# lctl cqos meta_all /mnt/fefs/
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
     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.
```

※各項目の説明は、前述の lctl cqos meta_top と同じです。

read_all の出力例

```
# 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.
```

※各項目の説明は、前述の lctl cqos meta_top と同じです。

write_all の出力例

```
# 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.
```

※各項目の説明は、前述の lctl cqos meta_top と同じです。

meta_user の出力例

```
# 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.
```

※各項目の説明は、前述の lctl cqos meta_top と同じです。

read_user の出力例

```
# 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.
```

※各項目の説明は、前述の lctl cqos meta_top と同じです。

write_user の出力例

```
# 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.
```

※各項目の説明は、前述の lctl cqos meta_top と同じです。

cache_top の出力例

```
# lctl cqos cache_top /mnt/fefs/
dprootmax=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.
```

※各項目の説明は以下です。

項目名	説明
dprootmax=10 dpusermax=10	FEFS デザインシートで指定した qos オプションの値です。各オプションの意味については、「FEFSクライアントの設定方法」を参照してください。
No.	write_page_cnt の大きい順に 10人まで表示します。
uid	ユーザーID です。
write_page_cnt	クライアントキャッシュに書き込んだページ数です。
total_wait_cnt	クライアントキャッシュに書き込むまでに、QoS 機能で実行を待ち合わせた回数です。 own : 1ユーザーあたりに使用可能な割合の上限に到達したために、書込みを待ち合わせた回数です。 other : クライアントノード内で使用可能なクライアントキャッシュの上限に到達したために、書込みを待ち合わせた回数です。
own_wtime(usec)	1ユーザーあたりに使用可能な割合の上限に到達した時に、書込みを待ち合わせた時間です。単位はマイクロ秒です。 max : 上記の最大値です。 avg : 上記の平均値です。
other_wtime(usec)	クライアントノード内で使用可能なクライアントキャッシュの上限に到達した時に、書込みを待ち合わせた時間です。単位はマイクロ秒です。 max : 上記の最大値です。 avg : 上記の平均値です。
last_update	クライアントキャッシュに書き込んだ最終日時です。

cache_all の出力例

```
# lctl cqos cache_all /mnt/fefs/
dprootmax=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.
```

※各項目の説明は、前述の lctl cqos cache_top と同じです。

cache_user の出力例

```
# lctl cqos cache_user 1076 /mnt/fefs/
uid(name) = 1076(fefs_guest06)
dprootmax=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.
```

※各項目の説明は、前述の lctl cqos cache_top と同じです。

lctl pool_list

【名前】

lctl pool_list - OST_pool のリスト、および OST_pool に登録された OST を表示するコマンド

【書式】

/usr/sbin/lctl pool_list <*fsname*>[.<*poolname*>] | <*mount_point*>

【説明】

<*fsname*> により定義された OST_pool のリストを表示します。

<*fsname*>[.<*poolname*>] により定義された pool に含まれる OST リストを表示します。

<*mount_point*> はファイルシステムのマウントポイントを指定します。

OST_pool とは指定した複数のOSTを束ねて1つのグループとし、ファイルやディレクトリをグループ内の OST に割り当てる機能です。

lctl pool_new

【名前】

lctl pool_new - OST_pool を作成するコマンド

【書式】

/usr/sbin/lctl pool_new <*fsname*>.<*poolname*>

【説明】

<*fsname*>.<*poolname*> で定義された新しい OST_pool を作成します。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

lctl pool_destroy

【名前】

lctl pool_destroy - OST_pool を削除するコマンド

【書式】

/usr/sbin/lctl pool_destroy <*fsname*>.<*poolname*>

【説明】

<*fsname*>.<*poolname*> により定義された OST_pool を削除します。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

lctl pool_add

【名前】

lctl pool_add - OST_pool に OST を追加するコマンド

【書式】

/usr/sbin/lctl pool_add <*fsname*>.<*poolname*> <*ostname indexed list*>

【説明】

<*fsname*>.<*poolname*>で定義されたOST_poolに <*ostname indexed list*> で定義された OST を追加します。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

lctl pool_remove

【名前】

lctl pool_remove - OST_poolからOSTを削除するコマンド

【書式】

/usr/sbin/lctl pool_remove <*fsname*>.<*poolname*> <*ostname indexed list*>

【説明】

<*fsname*>.<*poolname*> より定義された OST_pool から <*ostname indexed list*>により定義された OST を削除します。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

lctl ping

【名前】

lctl ping - LNet 通信の疎通を確認します。

【書式】

/usr/sbin/lctl ping <*nid*>

【説明】

指定された <*nid*> の LNet 通信の疎通を確認します。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

lctl set_param

【名前】

lctl set_param - FEFSに関するパラメーター情報を設定する。

【書式】

/usr/sbin/lctl set_param <*parameter*>=<*value ...*>

【説明】

<*parameter*>に<*value*>を設定します。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

lctl get_param

【名前】

lctl get_param - FEFSに関するパラメーター情報を表示する。

【書式】

/usr/sbin/lctl get_param <*parameter ...*>

【説明】

<*parameter*>に設定された値を表示します。

lctl conf_param

【名前】

lctl conf_param - デバイスにパラメーターを設定します。

【書式】

/usr/sbin/lctl conf_param [-d] <*device*|*fsname*>.<*parameter*>=<*value*>

【説明】

MGS経由で任意のデバイスに対し恒久的なパラメーターを設定します。

<*device*|*fsname*> で指定されたデバイスまたはファイルシステムの <*parameter*> に <*value*> を設定します。

本コマンドはMGSノードで root 権限で実行する必要があります。

【オプション】

-d <device | fsname>.<parameter>

パラメーター設定を削除します（次の再起動時にはデフォルト値を使用します）。<*value*> に null を設定してもパラメーター設定を削除できます。

lctl lfsck_start

【名前】

lctl lfsck_start - 指定されたMDTデバイスで、lfsck を開始します。

【書式】

/usr/sbin/lctl lfsck_start -M <*fsname*>-MDT0000 -A
/usr/sbin/lctl lfsck_start -h

【説明】

MDT の矛盾をチェックし、ファイルシステムを修復します。

本コマンドは、MDT0 をマウントする MDS ノード上で実行してください。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

<*fsname*>は、FEFSのファイルシステム名を指定してください。

【オプション】

-M <*fsname*>-MDT0000

コマンドの対象となる MDT デバイスを指定します。

-A

システム内の使用可能なすべてのデバイスで、本コマンドを起動します。

-h

ヘルプ情報を表示します。

lctl lfsck_stop

【名前】

lctl lfsck_stop - 実行中のlfsckを停止します。

【書式】

```
/usr/sbin/lctl lfsck_stop -M <fsname>-MDT0000 -A
/usr/sbin/lctl lfsck_stop -h
```

【説明】

実行中のlfsckを停止します。

本コマンドは、MDT0 をマウントする MDSノード上で実行してください。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

<*fsname*>は、FEFSのファイルシステム名を指定してください。

【オプション】

-M <*fsname*>-MDT0000

コマンドの対象となる MDT デバイスを指定します。

-A

システム内の使用可能なすべてのデバイスで、本コマンドを終了します。

-h

ヘルプ情報を表示します。

lctl help

【名前】

lctl help - lctl ヘルプを表示します。

【書式】

```
/usr/sbin/lctl help [command]
```

【説明】

lctl ヘルプを表示します。

【オプション】

[command]

ヘルプを表示するコマンドを指定します。

lctl --list-commands

【名前】

lctl --list-commands - サブコマンドの一覧を表示します。

【書式】

```
/usr/sbin/lctl --list-commands
```

【説明】

サブコマンドの一覧を表示します。

A.2.8 fsck.ldiskfs コマンド

【名前】

fsck.ldiskfs - FEFS デバイスのチェック・修復を行います。

【書式】

/opt/FJSVfefsprogs/sbin/fsck.ldiskfs [-pnyfv] [-b *superblock*] [-B *blocksize*] [-j *external_journal*] *device*

【説明】

MGS/MDS/OSS ノード上で実行するコマンドで、MGT/MDT/OST に対しチェック・修復を行います。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

【オプション】

-p

質問なしでファイルシステムの自動的な修復を行います。

-n

ファイルシステムに何も変更を加えません。

-y

修復時に出る質問にすべて "yes" と答えたものとみなします。

-f

ファイルシステムがcleanであっても、強制的にチェックを実行します。

-v

詳細な表示を行います。

-b *superblock*

通常のスーパーブロックの代わりに *superblock* をスーパーブロックとして使用します。

-B *blocksize*

指定したブロックサイズでスーパーブロックを探すように強制します。

-j *external_journal*

外部ジャーナルがあるパス名を指定します。

【引数】

device

チェックするデバイスを指定します。

【終了コード】

fsck.ldiskfs が返す終了コードは以下の条件の合計です。

0: エラーなし
1: ファイルシステムエラーが修正された
2: ファイルシステムエラーが修正されたが、システム再起動が必要
4: 未修正のファイルシステムエラーあり
8: 操作エラー
16: 使用法または構文エラー
32: ユーザー要求により fsck.ldiskfs がキャンセルされた
128: 共有ライブラリエラー

A.2.9 tunefs.lustre コマンド

【名前】

tunefs.lustre - ディスクのFEFS設定情報を変更します。

【書式】

/usr/sbin/tunefs.lustre [option] *device*

【説明】

ターゲットディスクの設定情報を変更するために使用します。

本コマンドはディスクを再フォーマットしたりターゲット情報を消去したりすることはありませんが、設定情報を変更するとファイルシステムが使用できなくなる可能性があります。

ここで行った変更は、ターゲットが次にマウントされたときにだけファイルシステムに影響します。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

【オプション】

--writeconf

ファイルシステムの設定ログを消去し、それらを再生成します。本オプションは非常に危険です。

ファイルシステムの設定ログはOST/MDTがマウントされたタイミングで再生成されます。

操作順序は以下のとおりです。

1. 当該ファイルシステムの全クライアントをアンマウントします。
2. 当該ファイルシステムのMDTとOSTをすべてアンマウントします。
3. すべてのサーバ上で "tunefs.lustre --writeconf *device*" を実行します。
4. MDT と OST をマウントします。
5. クライアントをマウントします。

【引数】

device

対象となるデバイスを指定します。

A.2.10 debugfs.ldiskfs コマンド

【名前】

debugfs.ldiskfs - ファイルシステム・デバッガ。

【書式】

/opt/FJSVfefsprogs/sbin/debugfs.ldiskfs [-R *request*] [*device*]

【説明】

対話型ファイルシステム・デバッガです。

device はデバイスまたはファイルです。

本コマンドは、管理者権限を持つユーザーが、MGS/MDS/OSSで実行できます。

【オプション】

-R *request*

コマンド *request* を実行し、終了します。以下のコマンドをサポートします。

icheck

【書式】

icheck *block* ...

【機能】

コマンドラインで指定されたブロック (複数指定可) を使用しているiノードのリストを表示します。

testb

【書式】

testb *block* [*count*]

【機能】

ブロック番号 *block* が、ブロックビットマップで割り当て済みとして使用されているか確認します。オプションの引数 *count* を指定すると、ブロック番号 *block* から始まる *count* 分のブロックについて確認します。

quit

【機能】

debugfs.ldiskfs コマンドを終了します。

A.2.11 fefsbackup コマンド [PG]

【名前】

fefsbackup - データ管理ツール。

【書式】

/opt/FJSVfefs/bin/fefsbackup [--version][-h] [**サブコマンド**]

【説明】

データ管理ツールの本体で、以下に示すサブコマンドを機能として持ちます。
本コマンドは、管理者権限を持つユーザーだけが利用できます。

【オプション】

--version

データ管理ツールのバージョンを表示します。

-h, --help

ヘルプメッセージを表示します。

【サブコマンド】

以下のサブコマンドは、管理者権限を持つユーザーだけが利用できます。

fefsbackup copy

【書式】

/opt/FJSVfefs/bin/fefsbackup copy [-h]
/opt/FJSVfefs/bin/fefsbackup copy [-L *request_id*] [-v] [option] -d *destdir* [-n *node*] {*path* [*path*...] | -f *pathlist*}
/opt/FJSVfefs/bin/fefsbackup copy [-v] -u *request_id* [option] -d *destdir* [-n *node*] {*path* [*path*...] | -f *pathlist*}
/opt/FJSVfefs/bin/fefsbackup copy [-v] -R *request_id*

【機能】

FEFSファイルシステム間でファイルのコピーを高速に行います。
path に指定されたパスの先頭からのディレクトリ構成を -d オプションで指定されたディレクトリ配下にコピーします。

ファイルの属性 (アクセス権、所有権、最終更新日時) は複写元ファイルの属性をコピーします。
また、以下はコピーしません。

- ACLおよび拡張属性(コピーする場合はオプションで指定してください)
- ハードリンクのリンク情報(別ファイルとしてコピーされることがあります)
- プロジェクトID
- プロジェクトID 継承フラグ
- ストライプディレクトリに関する情報
- ストライプに関する情報

コピー先ディレクトリ配下にコピー元ディレクトリ配下のファイルと同じ名前のファイルが存在する場合、上書きを行います。
また、パス中に ".." ディレクトリを含む場合は、".." 以降に指定されたパス名配下をコピーします。

ファイルのコピーに失敗した場合、失敗したファイルのパス名およびエラーの情報を標準エラーに出力し、エラー終了します。

【オプション】

-h, --help

ヘルプメッセージを出力します。

-f, --file *pathlist*

指定された *pathlist* に記載されたファイル、ディレクトリパスをコピー対象として扱います。なお、このオプションが指定された場合、個別にパスが指定されていた場合はエラーとなります。*pathlist* には 1行にファイルパスを 1つ記述してください。

-L, --label *request_id*

コピー管理用のリクエストIDを指定します。リクエストIDはユニークなIDを指定してください。

注意

-L オプションを使用して指定可能なリクエストIDの条件は以下です。
使用可能文字: 半角英数字(a-z, A-Z, 0-9)、半角記号(.-_)
文字数: 1文字以上32文字以下
フォーマット: 先頭が半角記号(.-_)で始まらないこと

-R, --resume *request_id*

エラーなどで中断したコピー処理が再実行可能であれば、エラー箇所から処理を再開します。

-u, --update *request_id*

差分コピーの対象となるコピー済みリクエストIDを指定します。指定されたリクエストIDで管理されているファイルの最終更新日時、最終変更日時、またはサイズがコピー元ファイルシステムと異なる場合に、当該ファイルをコピー先ファイルシステムへ転送します。

-v, --verbose

転送完了済みファイル、ディレクトリを表示します。

-a, --acl

ACL の保存を行います。指定しない場合は保存されません。

-x, --xattr

拡張属性の保存を行います。指定しない場合は保存されません。

-n, --node *node*

コピー先ノードを指定します。指定しない場合は、ノード内転送を行います。コピー先ホスト名、またはIPアドレスを指定します。

-d, --destdir *destdir*

コピー先ディレクトリを指定します。本オプションは必須で、指定されない場合、エラーとなります。
コピー先ディレクトリは絶対パスを指定してください。

--ignore_err

ファイルのコピーに失敗した場合に、失敗したファイルを除いてコピーを継続します。デフォルトではコピーに失敗した場合はエラー終了します。

【終了コード】

0: 正常終了
1: 内部エラーが発生してコピーに失敗した場合
2: 使用法に誤りがある場合

fefsbackup list

【書式】

```
/opt/FJSVfefs/bin/fefsbackup list [-h]
/opt/FJSVfefs/bin/fefsbackup list [-l] request_id [path ...]
/opt/FJSVfefs/bin/fefsbackup list [-l] [-f pathlist] request_id
```

【機能】

コピーのリクエストID情報を出力します。リクエストIDの指定がない場合は、リクエストIDの一覧を出力します。コピーのリクエストIDを指定した場合は、そのリクエストIDで管理しているファイルの一覧を出力します。

引数で *path* または、-f オプションの *pathlist* を指定した場合は、指定されたファイルだけを出力します。

ディレクトリ指定された場合はそのディレクトリ配下のファイルの一覧を出力します。

【オプション】

-h, --help

ヘルプメッセージを出力します。

-l, --list

指定された場合、"ls -l" 相当の詳細情報を出力します。

-f, --file *pathlist*

指定された *pathlist* に記載されたファイル、ディレクトリパスを含む情報だけを出力します。なお、このオプションが指定された場合、個別にパスが指定されていた場合はエラーとなります。*pathlist* には 1行にファイルパスを 1つ記述してください。

【終了コード】

0: 正常終了
1: 内部エラーが発生して表示に失敗した場合
2: 使用法に誤りがある場合

fefsbackup delete

【書式】

```
/opt/FJSVfefs/bin/fefsbackup delete -h
/opt/FJSVfefs/bin/fefsbackup delete [-F] request_id
```

【機能】

指定されたコピーのリクエストIDを削除します。リクエストIDの管理情報の削除だけを行います。コピー済みファイルの削除は行いません。

【オプション】

-h, --help

ヘルプメッセージを出力します。

-F, --force

ユーザーに対する問い合わせに対し、すべて "yes" が入力されたとして扱います。

【終了コード】

0: 正常終了
1: 内部エラーが発生して削除に失敗した場合
2: 使用法に誤りがある場合

fefsbackup status

【書式】

```
/opt/FJSVfefs/bin/fefsbackup status -h
/opt/FJSVfefs/bin/fefsbackup status
/opt/FJSVfefs/bin/fefsbackup status request_id
```

【機能】

実行中、および、エラー終了したリクエストIDの一覧を出力します。

リクエストIDが指定された場合は、そのリクエストID の詳細な情報を出力します。

【オプション】

-h, --help

ヘルプメッセージを出力します。

【終了コード】

0: 正常終了
1: 内部エラーが発生して表示に失敗した場合
2: 使用法に誤りがある場合

A.2.12 fefs_ost2fid コマンド

【名前】

fefs_ost2fid - 指定したinode番号を使用しているファイルのFIDを表示します。

【書式】

/opt/FJSVfefsprogs/sbin/fefs_ost2fid <*OST device*> <*ino* ...>

【説明】

OST上のinode番号からFEFSファイルシステム上のFIDを標準出力に出力します。

本コマンドは、OSS上で管理者権限を持つユーザーだけが利用できます。

出力例

```
# /opt/FJSVfefsprogs/sbin/fefs_ost2fid /dev/sdb 13024 14254
13034: [0xf:0x2367de45:0x4000]
14254: [0x22:0x2367de46:0x4000]
```

【引数】

<OST device>

OSTデバイスのパス

<ino>

inode番号。複数指定する場合は " "(スペース) で区切ります。inode番号は最大1024個まで指定可能です。

【コマンド正常終了時】

FIDを標準出力に出力し、戻り値0を返す。

【コマンドの異常終了時エラーメッセージ】

FIDを取得できないinode番号が存在する場合、そのinode番号のFID欄には”No such file or directory”と出力し、コマンドは正常終了します。

なお引数不正など、検索の途中でエラーを検出した場合は標準エラーにメッセージを出力し、検索を続けます。

inode番号13648のFIDの読み込みに失敗した例を以下に示します。

出力例

```
# /opt/FJSVfefsprogs/sbin/fefs_ost2fid /dev/sdb 13648
Invalid EA entry in inode (13648)     #出力先は標準エラー
```

コマンド全体がエラーになった場合はエラーメッセージを標準エラーに出力し、戻り値1を返します。

【戻り値】

0: 正常終了
1: エラー終了

A.2.13 find_file_ost コマンド

【名前】

find_file_ost - 指定した OSTを使用しているファイルのパスを出力します。

【書式】

/opt/FJSVfefsprogs/sbin/find_file_ost -o <*filename*> -d <*device*> <*OST_index* ...>

【説明】

MDTデバイスから指定したOSTのファイルパスの一覧を内部形式でファイルに出力します。

本コマンドは、MDS 上で管理者権限を持つユーザーだけが利用できます。

【オプション】

-o <*filename*>

結果を <*filename*>.out に書き出します。失敗したファイルは <*filename*>.err に書き出します。

-d <*device*>

<*device*> にMDTのパスを指定してください。

<*OST_index*>

対象の OST を特定するのに、「OST インデックス」という 0で始まる番号を使います。FEFS デザインシート「GFSシート」の OSS セクションを参照し、故障したデバイス(OST VOLUME) に対応する "OST INDEX" の値を確認の上指定します。複数指定する場合は " " (スペース) で区切ります。OST インデックスは最大 1024個まで指定可能です。

【コマンドの異常終了時エラーメッセージ】

引数不正など、検索の途中でエラーを検出した場合は標準エラーにメッセージを出力し、検索を続けます。

コマンドの途中でエラーになった場合は、エラーメッセージを標準エラーに出力し、戻り値1を返します。

出力例

```
/opt/FJSVfefsprogs/sbin/find_file_ost: Permission denied
```

【コマンドの異常終了の対処】

検索に失敗したファイルの親ディレクトリ一覧が "<*outfile*>.err" に出力されるので、このファイルに対して、convert_fid2path コマンドを実行し、出力されたパス配下のファイルを復旧の対象とします。

【戻り値】

0: 正常終了
1: エラー終了

A.2.14 convert_fid2path コマンド

【名前】

convert_fid2path - find_file_ostが出力したファイルをファイルシステム上のファイルパスに変換します。

【書式】

/opt/FJSVfefs/sbin/convert_fid2path -o <*outfile*> -m <*mount_point*> <*infile* ...>

【説明】

find_file_ostコマンドが出力した <*infile*> から指定のOSTを使用しているファイルパスを /(root) からの絶対パスで出力します。

本コマンドは、クライアント上で管理者権限を持つユーザーだけが利用できます。

出力例

```
/fefs/dir1/a
/fefs/dir2/b
```

【オプション】

-o <*outfile*>

結果を <*outfile*> に書き出します。

-m <*mount_point*>

<*mount_point*> に FEFS をマウントしたパスを指定します。

【コマンドの異常終了時エラーメッセージ】

引数不正など、処理の途中でエラーを検出した場合は標準エラーにメッセージを出力し、処理を続けます。

コマンドの途中でエラーになった場合はエラーメッセージを標準エラーに出力し、戻り値1を返します。

出力例

```
/opt/FJSVfefs/sbin/convert_fid2path: Invalid argument
```

【戻り値】

0: 正常終了
1: エラー終了

A.2.15 force_intr コマンド

【名前】

force_intr - ターゲット・デバイスへのファイルアクセスを制御する、または、状態表示するコマンド。

【書式】

```
/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
```

【説明】

指定されたターゲット・デバイスを制御します。ターゲット・デバイスは以下の書式で表される FEFS デバイスの識別子です。

<fsname>-[MDT | OST] <16進4桁のインデックス>

ターゲット・デバイスを deactivate すると、処理中の FEFS アクセスは異常終了します。また、新たな FEFS アクセスは異常復帰します。-c オプションを指定した場合、FEFS クライアントから指定したターゲットへのファイルアクセスを無効化/有効化します。

-s オプションを指定した場合、FEFSサーバ間でターゲットへのファイルアクセスを無効化/有効化します。無効化された場合は、転送中のサーバ間のファイルアクセスを中断し FEFS クライアントにエラーを返します。

本コマンドは FEFS サービスが起動していない場合は失敗します。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

本コマンドで、activate を行った後、クライアントーサーバ間の再接続を促すために、全クライアントノードで lfs df コマンドを実行する必要があります。

【オプション】

-h

ヘルプメッセージを表示します。

-v

詳細なメッセージを表示します。

-c

FEFS クライアント上で実行する場合に指定します。

-s

FEFS サーバ上で実行する場合に指定します。deactivate されたターゲット・デバイスは新規ファイル作成時に使用されなくなります。

-m {activate|deactivate|status}

activate コマンドはターゲット・デバイスを有効化します。
deactivate コマンドはターゲット・デバイスを無効化します。
status コマンドは現在の状態を表示します。activate状態の場合は "UP"、deactivate状態の場合は "IN" と表示されます。その後に処理対象のデバイスが表示されます。

例

ファイルシステムの切離し操作を行った後、それぞれのノードで状態確認を行った場合の表示例を以下に示します。

- 表示例 (クライアントノード)

```
[ストレージI/Oノード]
# /usr/sbin/force_intr -c -m status fefs-OST0000 fefs-OST0001
IN fefs-OST0000
IN fefs-OST0001
[ログインノード]
# /usr/sbin/force_intr -c -m status fefs-OST0000 fefs-OST0001
IN fefs-OST0000
IN fefs-OST0001
```

- 表示例 (サーバノード)

```
[MDSノード]
# /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

接続しているすべてのターゲット・デバイスに対して実行します。

【引数】

<target>

処理対象となるターゲット・デバイス名

【戻り値】

以下のステータスが返されます。

0: 正常
1: エラー終了

A.2.16 evict_client コマンド

【名前】

evict_client - サーバ側でクライアント資源の回収を行う

【書式】

/usr/sbin/evict_client [-h] <*ipaddress* ...>

【説明】

あるFEFSクライアントがファイルアクセス中にパニックすると、別のFEFSクライアントから同じファイルをアクセスした場合にファイルアクセスが数分間待たされることがあります。本コマンドでパニックしたFEFS クライアントを指定することでファイルシステムから切り離し、ファイルアクセスハングを早期に解消できます。

動作しているFEFSクライアントを指定した場合はファイルアクセスが異常復帰します。evict_clientで切り離したクライアントは再度ファイルアクセスすることで切り戻すことができます。

本コマンドは、MDS または OSS 上で管理者権限をもつユーザーだけが利用できます。

【オプション】

-h

ヘルプメッセージを表示します。

【引数】

ipaddress

クライアントに割り当てられたIPアドレスを指定します。複数指定する場合はスペースで区切ります。

指定するIPアドレスはFEFS デザインシートを参照します。NODETYPE が CCM、LN の場合は、Primary Network の IP ADDRESS の値を指定します。CN-SIO の場合は、Tofu の IP ADDRESS の値を指定します。

【戻り値】

以下のステータスが返されます。

0: 正常
1: エラー終了

A.2.17 fefs_yaml2csv コマンド

【名前】

fefs_yaml2csv - インストーラ用ノード情報定義ファイルの変換コマンド

【書式】

/sbin/fefs_yaml2csv <*infile*> <*outfile*>

【説明】

指定された <*infile*> を <*outfile*> へ CSV形式で出力します。出力されたCSVファイルは、FEFSDesignSheet.xlsm へインポートできます。

<*infile*> には ノード情報定義ファイル または FXサーバ用ノード情報定義ファイルFXサーバ用ノード情報定義ファイル を指定してください。
ノード情報定義ファイル、FXサーバ用ノード情報定義ファイルについては、「ジョブ運用ソフトウェア 導入ガイド」を参照してください。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

【戻り値】

以下のステータスが返されます。

0: 正常終了
1: 異常終了

A.2.18 fefs_deactivate コマンド

【名前】

fefs_deactivate - ファイルシステムへのアクセスを制御する、または、状態表示するコマンド。

【書式】

/usr/sbin/fefs_deactivate -m {activate|deactivate|status} <*path*>
/usr/sbin/fefs_deactivate -m status -a

【説明】

fefs_deactivate コマンドはファイルシステムへのリクエストを制御します。

<*path*> には有効化、無効化または、状態表示するファイルシステムのマウントポイントを指定してください。
ファイルシステムを無効化すると該当のファイルシステムへのアクセスを中断します。

本コマンドは、管理者権限を持つユーザーだけが利用できます。

【オプション】

-m {activate|deactivate|status}

activate コマンドは指定したファイルシステムを有効化します。
deactivate コマンドは指定したファイルシステムを無効化します。
status コマンドは現在のファイルシステムの状態を表示します。

例

ファイルシステムの切離し操作を行った後、それぞれのノードで状態確認を行った場合の表示例を以下に示します。

- 表示例

```
[ストレージI/Oノード]
# fefs_deactivate -m status /fefs
FS   ST MNT
FEFS IN /.fefs
LLIO IN /fefs
[計算ノード]
# fefs_deactivate -m status /fefs
FS   ST MNT
LLIO IN /fefs
[ログインノード]
# fefs_deactivate -m status /fefs
FS   ST MNT
FEFS IN /fefs
```

ヘッダについての説明

FS : ファイルシステムを表示します。
ST : 切離し/組込み状態を表示します。切離されている場合は IN、切離されていない場合は UP を表示します。
MNT : 対象のマウントポイントを表示します。

-a

すべてのファイルシステムについての状態を表示します。

【戻り値】

以下のステータスが返されます。

0: 正常
1: エラー終了

付録B　メッセージ

B.1　システムログに出力されるメッセージ

以下は、FEFS がシステムログに出力するメッセージの形式です。

```
[ERR.] FEFS 0001 an error..
(1)    (2)  (3)  (4)
```

1. メッセージ種別

   メッセージの出力レベルを表します。種別によって、以下に示すメッセージID が付与されます。

   - [ERR.]: ERRORメッセージ (0001から5999)
   - [WARN]: WARNINGメッセージ(6000から6999)
   - [NOTE]: NOTICE メッセージ(7000から7999)

2. FEFS プレフィックス

   メッセージが FEFS 関連の出力であることを示す識別子です。

   内容は以下のとおりです。

   - FEFS: FEFS 本体が出力するメッセージです。
   - LNet: FEFS 通信レイヤが出力するメッセージです。

3. メッセージID

   メッセージの識別ID です。1. で述べたように値の範囲はメッセージの種別ごとに決まっています。

4. メッセージ内容

   メッセージの内容です。

[ERROR メッセージ] (0001から5999)

[ERR.] FEFS 0001 *Obdname*: operation *Op* to node *Netdev* failed: rc = *Err*

意味

FEFSクライアント、サーバ間のオペレーションでエラーが発生しました。

Obdname : OBD名
Op : オペレーション名
Netdev : ネットワークデバイス名
Err : エラーコード

対処

Err の値	意味	対処
-2	対象ファイルが存在しません。	ファイルの有無を確認してください。
-11	リトライ要求です。	対処不要です。
-12	メモリ不足です。	メモリの使用状況を確認してください。
-16	*Netdev* の MDT/OST への再接続に失敗しています。	対処不要です。
-19	操作対象が存在しません。	対処不要です。
-107	本メッセージが出力されたノードの接続情報が、*Netdev* にありません。	本メッセージが出力されたノードがファイルシステムから切り離された可能性があるため、本メッセージが出力されたノードで動作していたジョブなどが I/O エラーになっていないか確認してください。

Err の値	意味	対処
		なお、ファイルシステムから切り離されている場合は自動復旧されます。
-110	MGS からクライアントへの切離し通知がタイムアウトしたことを示しています。	対処不要です。
-116	*Op* が "mds_close" の場合は、MDS 上で close するファイルの情報がありません。	本メッセージは単独で出力されている場合は、対処不要です。 本メッセージがほかのメッセージと共に出力されている場合は、ほかのメッセージの対処方法を参照してください。

上記以外の場合は、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[ERR.] FEFS 0003 Unknown option '*Opt*', won't mount.

意味

マウントオプションに誤りがあります。

Opt : マウントオプション

対処

マウントオプションを正しく指定してください。

[ERR.] FEFS 0004 Illegal option value '*Opt=Val*', won't mount.

意味

マウントオプションに誤りがあります。

Opt : マウントオプション
Val : マウントオプションの値

対処

マウントオプションを正しく指定してください。

[ERR.] FEFS 0005 Simultaneous specification with noqos and qos,qos_cache options can't be performed, won't mount.

意味

マウントオプションに誤りがあります。noqos と qos,qos_cache オプションの同時指定はできません。

対処

noqos と qos,qos_cache オプションのどちらか一方を指定してください。

[ERR.] LNet 1000 Please specify EITHER 'networks' or 'ip2nets' but not both at once

意味

通信層の定義ファイルに誤りがあります。

対処

FEFS デザインシートに誤りがないか確認してください。

[ERR.] LNet 1001 Error *Err* starting up LNI *Net*

意味

通信層の初期化に失敗しました。

Err : エラーコード
Net : ネットワークインターフェース

対処

Netに表示されたネットワークインターフェースが有効になっているかを確認してください。

[ERR.] LNet 1002 Error parsing '*Def*'

意味

通信層の定義ファイルに誤りがあります。

Def: 誤りのある定義

対処

FEFS デザインシートに誤りがないか確認してください。

[ERR.] LNet 1003 Duplicate network specified: *Type*

意味

通信層の定義ファイルに誤りがあります。

Type: 重複しているネットワークタイプ

対処

FEFS デザインシートに誤りがないか確認してください。

[ERR.] LNet 1004 Can't parse networks: string too long

意味

通信層の定義ファイルに誤りがあります。

対処

FEFS デザインシートに誤りがないか確認してください。

[ERR.] LNet 1005 Too many interfaces for net *Name*

意味

通信層の定義ファイルに誤りがあります。指定可能なネットワークインターフェース数の最大値 (16) を超えています。

Name: ネットワーク名

対処

FEFS デザインシートに誤りがないか確認してください。

[ERR.] LNet 1006 Error *Err* enumerating local IP interfaces for ip2nets to match

意味

通信層の初期化処理に失敗しました。

Err: エラーコード

対処

エラーコードが -12 の場合は、メモリ不足が原因のため、メモリの使用状況を確認してください。

エラーコードが -12 以外の場合は、FEFS デザインシートに誤りがないか確認してください。

[ERR.] LNet 1007 No local IP interfaces for ip2nets to match

意味

通信層の定義ファイルに誤りがあります。

対処

FEFS デザインシートに誤りがないか確認してください。

[ERR.] LNet 1008 Error *Err* parsing ip2nets

意味

通信層の定義ファイルに誤りがあります。

Err: エラーコード

対処

FEFS デザインシートに誤りがないか確認してください。

[ERR.] LNet 1009 ip2nets does not match any local IP interfaces

意味

通信層の定義ファイルに誤りがあります。

対処

FEFS デザインシートに誤りがないか確認してください。

[ERR.] FEFS 2000 libcfs_debug_init: *Err*

意味

libcfsモジュールロード時にLIBCFSデバッグ機能の初期化処理に失敗しました。

Err: エラー番号

対処

libcfs モジュールのロードに失敗しています。*Err*が -12 の場合は計算機でメモリが不足しています。メモリ使用量を減らすなどの処置を行ってから FEFS サービスを再起動してください。それでも問題が解決しない場合は、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[ERR.] FEFS 2001 init_fefslog: *Err*

意味

libcfs モジュールロード時に FEFSLOG 機能の初期化処理に失敗しました。

Err: エラー番号

対処

libcfs モジュールのロードに失敗しています。*Err*が -22 の場合はモジュール・パラメーターに無効な値が設定されています。モジュール・パラメーターを再検討・再設定してから FEFS サービスを起動してください。*Err*が -12 の場合は計算機でメモリが不足しています。メモリ使用量を減らすなどの処置を行ってから FEFS サービスを再起動してください。それでも問題が解決しない場合は、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[ERR.] FEFS 2003 misc_register: error *Err*

意味

FEFS内部の登録処理に失敗しました。

Err: エラーコード

対処

FEFS内部モジュールのロードに失敗しています。当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[ERR.] FEFS 2004 insert_proc: error *Err*

意味

procファイルシステム上のFEFS内部のエントリの登録処理に失敗しました。

Err: エラーコード

対処

FEFS 内部モジュールのロードに失敗しています。当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[ERR.] FEFS 2005 misc_deregister error *Err*

意味

FEFS内部の登録削除処理に失敗しました。

Err: エラーコード

対処

当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[ERR.] FEFS 2006 libcfs_debug_cleanup: *Err*

意味

デバッグ機能の終了処理に失敗しました。

Err: エラーコード

対処

当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[ERR.] FEFS 2007 initialize workitem: error *Err*

意味

FEFSの初期化処理に失敗しました。

Err: エラーコード

対処

FEFS 内部モジュールのロードに失敗しています。当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[ERR.] FEFS 2008 Startup workitem scheduler: error: *Err*

意味

FEFSの初期化処理に失敗しました。

Err: エラーコード

対処

FEFS 内部モジュールのロードに失敗しています。当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[ERR.] FEFS 2009 cfs_crypto_regster: error *Err*

意味

FEFSの初期化処理に失敗しました。

Err: エラーコード

対処

FEFS内部モジュールのロードに失敗しています。当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[ERR.] FEFS 2010 Cannot create proc entry: *Err*

意味

モジュールロード時に/proc/fefslog エントリの作成・登録に失敗しました。

Err: エラー番号

対処

libcfs モジュールのロードに失敗しています。*Err*が -12 の場合は計算機でメモリが不足しています。メモリ使用量を減らすなどの処置を行ってから FEFS サービスを再起動してください。

[ERR.] FEFS 2022 open: /proc/fefslog: *Msg*: *Err*

意味

FEFSLOG デーモンで /proc/fefslog に対する open(2) が異常復帰しました。

Msg: strerror(3) で生成される、エラー番号に準じたメッセージ
Err: エラー番号

対処

libcfs モジュールが正常にロードされていない状態で、FEFSLOG デーモンの起動を試みた可能性があります。libcfs モジュールをロードしてから同様の操作を行うか、FEFS サービスを再起動してください。上記を実施後にも同様のメッセージが出力される場合、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[ERR.] FEFS 2023 open *File*: *Msg*: *Err*

意味

FEFSLOG デーモンで FEFS 専用ログファイルに対する open(2) が異常復帰しました。

File: FEFS 専用ログのファイルパス
Msg: strerror(3) で生成される、エラー番号に準じたメッセージ
Err: エラー番号

対処

FEFSLOG デーモンで *File* に対する open(2) が異常復帰しています。*Err* に該当するエラー番号は open(2) のオンラインマニュアルに記載されているエラー番号と同じものになるので、そちらを参照の上、適切な処置を行ってください(例えば *Err* が -13 の場合、ファイルに対するアクセスパーミッションがないか、*File* のディレクトリ部分のどれかのディレクトリ検索許可がなかった、またはファイルが存在せず、親ディレクトリへの書込み許可がなかったなどの原因が考えられます)。それでも問題が解決しない場合は、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[ERR.] FEFS 2024 ioctl: *Cmd*: *Msg*: *Err*

意味

FEFSLOG デーモンで ioctl(2) が異常復帰しました。

Cmd: ioctl(2) に対して実行されたコマンド
Msg: strerror(3) で生成される、エラー番号に準じたメッセージ
Err: エラー番号

対処

FEFSLOG デーモンで *Cmd* コマンドの ioctl(2) が異常復帰しています。FEFS サービスを再起動してください。それでも問題が解決しない場合は、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[ERR.] FEFS 2025 Portals memory leaked: *Bytes* bytes

意味

libcfs モジュールでメモリリークが発生しました。

Bytes : リークしたメモリ量 (byte 単位)

対処

運用影響はありませんが、libcfs モジュールでメモリリークが発生していました。メモリ操作関連の障害が疑われますので、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[ERR.] FEFS 2200 QoS cannot allocate memory.

意味

QoS 機能に必要なメモリの獲得に失敗しました。

対処

メモリの使用状況を確認してください。

[ERR.] FEFS 2201 QoS System error. func=*Func* route=*Route* code=*Err*

意味

QoS機能でシステムエラーが発生しました。

Func : 関数名
Route : エラールート
Err : エラーコード

対処

担当保守員 (SE)、または当社 Support Desk に連絡してください。

[ERR.] FEFS 2202 QoS cannot be enabled.

意味

QoS 制御を有効にできません。

対処

QoS 機能を使用する時は、Lustre の NRS (Network Request Scheduler) の policy を fifo に設定してください。

具体的には以下のコマンドを実行してください。

MDS 上で実行

```
# lctl set_param mds.MDS.mdt.nrs_policies="fifo reg"
```

OSS 上で実行

```
# lctl set_param ost.OSS.ost_io.nrs_policies="fifo reg"
```

[ERR.] FEFS 2203 NRS cannot be enabled.

意味

Lustre の NRS (Network Request Scheduler) を有効にできません。

対処

NRS は非サポート機能のため、使用しないでください。

[ERR.] FEFS 2400 fefsinfod-daemon is terminated abnormally. (Error code: *Err*)

意味

統計情報取得デーモンが異常終了しました。

Err: エラーコード

対処

当メッセージを含むシステムログファイルとFEFS の資料("付録Fトラブル対処時に必要な資料"参照)を採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[ERR.] FEFS 2410 Couldn't mount because of no required features.

意味

FEFS サーバ (MDS) のマウント処理に失敗しました。

対処

FEFS デザインシートの "MDT VOLUME" に正しいデバイスを指定してください。

[ERR.] FEFS 2420 Lustre server(*Svr*) is unsupported version.

意味

接続先の Lustre サーバのバージョンは、FEFS ではサポート範囲外のバージョンです。

Svr: 接続先サーバ

対処

接続先の Lustre サーバのバージョンを確認してください。

[WARNING メッセージ] (6000から6999)

[WARN] FEFS 6420 gethostname: Cannot get hostname: *Msg*: *Err*

意味

FEFSLOG デーモンで gethostname(2) が異常復帰しました。

Msg: strerror(3) で生成される、エラー番号に準じたメッセージ
Err: エラー番号

対処

FEFSLOG デーモンの起動処理中に gethostname(2) が異常復帰しています。FEFSLOG デーモンの処理は継続されますが、FEFS 専用ログファイルに書き込まれるログメッセージのホスト名は "UNKNOWN" と表記されてしまいます。これに問題がある場合は gethostname(2) のオンラインマニュアルで *Err* に対する対処方法を調べてから、その対処を行い、そのあとに FEFSサービスを再起動してください。それでも問題が解決しない場合は、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[WARN] FEFS 6421 close *File*: *Msg*: *Err*

意味

FEFSLOG デーモンで FEFS 専用ログファイルに対する close(2) が異常復帰しました。

File: FEFS 専用ログのファイルパス
Msg: strerror(3) で生成される、エラー番号に準じたメッセージ
Err: エラー番号

対処

FEFSLOG デーモンで *File* に対する close(2) が異常復帰しています。FEFSLOG デーモンの処理は継続可能ですが、FEFS 専用ログに対して出力されるべきであった一部のログメッセージが失われるなどの現象が発生している可能性があります。*Err* に該当するエラー番号は close(2) のオンラインマニュアルに記載されているエラー番号と同じものになるので、そちらを参照の上、適切な処置を行ってください。それでも問題が解決しない場合は、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[WARN] FEFS 6422 read /proc/fefslog: *Msg*: *Err*

意味

FEFSLOG デーモンで /proc/fefslog に対する read(2) が異常復帰しました。

Msg : strerror(3)で生成される、エラー番号に準じたメッセージ
Err : エラー番号

対処

FEFSLOG デーモンで *File* に対する read(2) が異常復帰しています。FEFSLOG デーモンの処理は継続可能ですが、FEFS 専用ログに対して出力されるべきであった一部のログメッセージが失われるなどの現象が発生している可能性があります。*Err* に該当するエラー番号は read(2) のオンラインマニュアルに記載されているエラー番号と同じものになるので、そちらを参照の上、適切な処置を行ってください。それでも問題が解決しない場合は、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[WARN] FEFS 6423 write *File*: *Msg*: *Err*

意味

FEFSLOG デーモンで FEFS 専用ログファイルに対する write(2) が異常復帰しました。

File : FEFS 専用ログのファイルパス
Msg : strerror(3) で生成される、エラー番号に準じたメッセージ
Err : エラー番号

対処

FEFSLOG デーモンで *File* に対する write(2) が異常復帰しています。FEFSLOG デーモンの処理は継続可能ですが、FEFS 専用ログに対して出力されるべきであった一部のログメッセージが失われるなどの現象が発生している可能性があります。*Err* に該当するエラー番号は write(2) のオンラインマニュアルに記載されているエラー番号と同じものになるので、そちらを参照の上、適切な処置を行ってください。

[WARN] FEFS 6424 close /proc/fefslog: *Msg*: *Err*

意味

FEFS のログローテーション処理中に問題が発生しました。

Msg : strerror(3) で生成される、エラー番号に準じたメッセージ
Err : エラー番号

対処

FEFSLOG デーモンで /proc/fefslog に対する close(2) が異常復帰しています。FEFSLOG デーモンの処理は継続可能ですが、FEFS 専用ログに対して出力されるべきであった一部のログメッセージが失われるなどの現象が発生している可能性があります。*Err* に該当するエラー番号は close(2) のオンラインマニュアルに記載されているエラー番号と同じものになるので、そちらを参照の上、適切な処置を行ってください。それでも問題が解決しない場合は、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[WARN] FEFS 6425 lost *Bytes* [Bytes] messages

意味

libcfs モジュールのアンロードに伴い、*Bytes*[Bytes] のメッセージが破棄されました。

Bytes : 破棄されたメッセージのバイト数

対処

サービスや運用自体には影響はないので対処は必要ありません。ただしサービスの停止の度に表示されるなど、頻繁にこのメッセージが表示される場合は FEFSLOG 機能の障害の可能性がありますので、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[WARN] FEFS 6426 strftime fails.

意味

strftime(3) が異常復帰したため、ログメッセージの時間変換に失敗しました。

対処

運用に影響はありません。稀に発生する分には特に対処の必要はありませんが、頻繁に発生するようでしたら、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[WARN] FEFS 6427 snprintf: *Msg*: *Err*

意味

snprintf(3) が異常復帰したため、ログメッセージの変換に失敗しました。

Msg : strerror(3) で生成される、エラー番号に準じたメッセージ
Err : エラー番号

対処

運用に影響はありません。稀に発生する分には特に対処の必要はありませんが、頻繁に発生するようでしたら、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[WARN] FEFS 6428 localtime: Cannot get localtime

意味

localtime(3) が異常復帰したため、ログメッセージの時間変換に失敗しました。

対処

運用に影響はありません。稀に発生する分には特に対処の必要はありませんが、頻繁に発生するようでしたら、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

[WARN] FEFS 6600 QoS unknown request from ip=*Addr*

意味

QoS 定義ファイルで定義されていない FEFS クライアントからの要求がありました。

Addr : 要求元の IPアドレス

対処

QoS定義ファイルの定義内容を見直し、適切なIPアドレスを指定してください。

[NOTICE メッセージ] (7000から7999)

[NOTE] FEFS 7001 Mounted *Fsname*-client

意味

FEFS クライアントのマウント処理が完了しました。

Fsname : ファイルシステム名

対処

対処不要です。

[NOTE] FEFS 7010 server umount {MGS | *Fsname*-{MDT | OST}*Num*} complete

意味

FEFSサーバ (MGS、MDS、またはOSS) のアンマウント処理が完了しました。

Fsname : ファイルシステム名
Num : MDTインデックスまたはOSTインデックス

対処

対処不要です。

[NOTE] FEFS 7011 Unmounted *Fsname*-client

意味

FEFS クライアントのアンマウント処理が完了しました。

Fsname : ファイルシステム名

対処

対処不要です。

[NOTE] FEFS 7400 fefslog daemon is starting .... rotate fsize: *Fsize*, gen: *Gen*

意味

FEFSLOG デーモンを起動しています。

Fsize : ファイルローテート契機になるファイルサイズ
Gen : 保持されるログの世代

対処

FEFSLOG デーモンが起動する際に表示される正常系のメッセージです。FEFS サービスの開始時に必ず表示されるもので、対処の必要はありません。

[NOTE] FEFS 7401 shutting down fefslogd ...

意味

FEFSLOG デーモンを終了します。

対処

FEFSLOG デーモンが終了する際に表示される正常系のメッセージです。FEFS サービスの停止時に必ず表示されるもので、対処は必要ありません。

[NOTE] FEFS 7501 {MGS | *Fsname*-{MDT | OST}*Num*}: Will be in recovery for at least *Time*: or until *Numcli* client[s] reconnect[s]

意味

FEFS サービス(MGT、MDT または OST) のフェイルオーバが開始されました。フェイルオーバには少なくとも *Time* 時間かかるか、まだ *Numcli* の FEFS クライアントが再接続しています。

Fsname : ファイルシステム名
Num : MDTインデックスまたはOSTインデックス
Time : フェイルオーバにかかる予定時間(*mm*:*ss*)

mm : 分
ss : 秒

Numcli : FEFS クライアント数

対処

対処不要です。

[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*

意味

MGS、MDSまたはOSS に接続していた *Nid* の NID をもつクライアントからの応答が *Time* 秒間ありません。FEFSからの切離し処理を行います。

Fsname : ファイルシステム名
Num : MDT インデックスまたは OST インデックス
Id : ユニークな ID
Nid : クライアントの NID
Time : クライアントから応答がなかった時間(秒)
Info : クライアントに関するそのほかの情報

対処

対処不要です。

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

意味

ファイルシステムの不整合を検出しています。

パラメーターの説明

device : デバイス名
nnnn : 行番号
group : ブロックグループ番号
blocknum :ブロック番号
gd :グループディスクリプタ番号

対処

ファイルシステムが壊れている可能性があります。ファイルシステムの不整合の修復を実施してください。"D.4.5 fsckの実施" を参照してください。

B.2 コマンドの出力するメッセージ

コマンド実行時に異常が発生した場合、以下のメッセージを標準エラー出力に出力します。

B.2.1 fefs_sync コマンド

ERROR: fefs_sync: /etc/opt/FJSVfefs/config not found.

意味

/etc/opt/FJSVfefs/config ディレクトリが存在しません。

対処

担当保守員 (SE)、または当社 Support Desk に連絡してください。

ERROR: fefs_sync: internal command failed. (cmd=*XX* error=*XX* cluster=*XX*)

意味

内部コマンドが異常終了しました。

cmd=*XX* : *XX*には内部コマンドを示すコード番号が入ります。
error=*XX* : *XX*にはエラーコードが入ります。
cluster=*XX* : *XX*にはクラスタ名が入ります。

対処

担当保守員 (SE)、または当社 Support Desk に連絡してください。

ERROR: fefs_sync: cannot be used by this user.

意味

fefs_sync コマンドを実行する権限がありません。

対処

fefs_sync コマンドを実行するユーザーの権限、実行するノードを見直し、再度実行してください。

ERROR: fefs_sync: setup failed. check *Filename* file.

意味

設定ファイル作成が失敗しました。

パラメーターの説明

Filename : 異常メッセージファイル

対処

Filename ファイルの内容を確認し、失敗した原因を特定してください。

原因を特定できた場合は、原因に対し対処したのち、fefs_sync コマンドを再度実行してください。

原因を特定できない場合は、担当保守員 (SE)、または当社 Support Desk に連絡してください。

ERROR: fefs_sync: mkfs failed. check *Filename* file.

意味

ボリュームのフォーマットが失敗しました。

パラメーターの説明

Filename : 異常メッセージファイル

対処

Filename ファイルの内容を確認し、失敗した原因を特定してください。

原因を特定できた場合は、原因に対し対処したのち、fefs_sync コマンドを再度実行してください。

原因を特定できない場合は、担当保守員 (SE)、または当社 Support Desk に連絡してください。

ERROR: fefs_sync: mount failed. check *Filename* file.

意味

マウントが失敗しました。

パラメーターの説明

Filename : 異常メッセージファイル

対処

Filename ファイルの内容を確認し、失敗した原因を特定してください。

原因を特定できた場合は、原因に対し対処したのち、fefs_sync コマンドを再度実行してください。

原因を特定できない場合は、担当保守員 (SE)、または当社 Support Desk に連絡してください。

ERROR: fefs_sync: umount failed. check *Filename* file.

意味

アンマウントが失敗しました。

パラメーターの説明

Filename : 異常メッセージファイル

対処

Filename ファイルの内容を確認し、失敗した原因を特定してください。

原因を特定できた場合は、原因に対し対処したのち、fefs_sync コマンドを再度実行してください。

原因を特定できない場合は、担当保守員 (SE)、または当社 Support Desk に連絡してください。

ERROR: fefs_sync: stop failed. check *Filename* file.

意味

FEFS サービス停止が失敗しました。

パラメーターの説明

Filename : 異常メッセージファイル

対処

Filename ファイルの内容を確認し、失敗した原因を特定してください。

原因を特定できた場合は、原因に対し対処したのち、fefs_sync コマンドを再度実行してください。

原因を特定できない場合は、担当保守員 (SE)、または当社 Support Desk に連絡してください。

WARN: fefs_sync: Some of the nodes could not accessed.

意味

一部のノードで実行できませんでした。

対処

上記メッセージに続いて、以下のとおり問い合わせのメッセージが出力されますので、継続する場合は y を、中止する場合は n を入力してください。

```
Do you want to continue? (y/n):
```

y を入力した場合は処理が継続されます。
n を入力した場合は処理が中止します。

実行できなかったノードID は、以下のメッセージで出力されるファイルで確認できます。

```
create downnodeid file: filename
```

原因に対し対処したのち、"3.11 構築に失敗したノードの構築方法" に基づいて復旧してください。

INFO: fefs_sync: formatting. (remain=*N*)

意味

ボリュームのフォーマットを実行中です。

ボリュームのフォーマットが 30分以上経過した場合、その後一定間隔で表示されますが、動作としては問題ありません。

ブロックサイズなどを小さく設定した場合などには発生することがあります。

*N*は残りのボリューム数です。

対処

必要ありません。

Connect FEFS. [*nodetype*](*cluster*)

意味

クライアントにおけるFEFSマウント時、サーバボリュームとの接続確認を行います。

本メッセージに続いて以下のメッセージが出力されます。

..... OK/AWAIT : <接続確認済みボリューム数> / <接続待ちボリューム数>

一定間隔で表示されますが動作としては問題ありません。すべてのボリュームで接続したことが確認できると、接続確認済みボリューム数に0が表示されます。10分経過しても完了しない場合は、以下の問い合わせが出力されますので、継続する場合はyを、中止する場合はnを入力してください。

```
Continue to wait? [y/n]:
```

yを入力した場合は処理が継続されます。
nを入力した場合は "Aborted." とメッセージが出力され、処理が中止します。

対処

必要ありません。

B.2.2 fefsconfig コマンド

fefsconfig: Error: configuration file was not found.

意味

FEFS セットアップツール用構成定義ファイルが見つかりません。

対処

FEFS セットアップツール用構成定義ファイルを適切なディレクトリに配布してください。

fefsconfig: Error: failed to setup by bad configuration file.

意味

FEFS セットアップツール用構成定義ファイルの内容に誤りがあります。

対処

FEFS セットアップツール用構成定義ファイルの内容を見直してください。

fefsconfig: Error: failed to make <*Filename*>.

意味

FEFS 設定ファイル <*Filename*> の作成に失敗しました。

パラメーターの説明

<*Filename*> : FEFS設定ファイル名

対処

担当保守員 (SE)、または当社 Support Deskに連絡してください。

fefsconfig: Error: this node will not mount the fefs. (bad design sheet)

意味

FEFS 構成定義ファイルに、ファイルシステム定義が存在しません。

対処

FEFS デザインシートに誤りがないか確認してください。

fefsconfig: Error: this node must have network information. (bad design sheet)

意味

FEFS 構成定義ファイルの外部のネットワークの定義に誤りがあります。

対処

FEFS デザインシートに誤りがないか確認してください。

---

fefsconfig: Error: this node defined both design sheet. (bad design sheet)

意味

実行ノードは、FEFSDesignSheet.xlsm と FEFSDesignSheet_External.xlsm 両方に定義されています。

対処

どちらかの FEFS デザインシートに定義してください。

---

fefsconfig: Error: no such network interface. (bad network interface: *<Interfacename>*)

意味

FEFS 構成定義ファイルのネットワークインターフェース名に誤りがあります。

パラメーターの説明

<Interfacename> : インターフェース名

対処

FEFS デザインシートのネットワークインターフェースの定義に誤りがないか確認してください。

---

fefsconfig: Error: *hostname* no such network interface. (not tofu ip address)

意味

FEFS 構成定義ファイルのネットワークインターフェースの tofu IP アドレスが設定されていません。

パラメーターの説明

hostname : ホスト名

対処

FEFS デザインシートのネットワークインターフェースの定義に誤りがないか確認してください。

---

fefsconfig: Error: no such network interface. (bad ip address: *<IPaddress>*)

意味

FEFS 構成定義ファイルのネットワークインターフェースの IP アドレスに誤りがあります。

パラメーターの説明

<IPaddress> : IPアドレス

対処

FEFS デザインシートのネットワークインターフェースの定義に誤りがないか確認してください。

---

fefsconfig: Error: *hostname* no such network interface. (bad tofu coordinate: *tofu coordinate*)

意味

FEFS 構成定義ファイルのネットワークインターフェースの tofu 座標に誤りがあります。

パラメーターの説明

hostname : ホスト名
tofu coordinate : tofu 座標

対処

FEFS デザインシートのネットワークインターフェースの定義に誤りがないか確認してください。

fefsconfig: Error: no such network interface. (bad design sheet)

意味

FEFS 構成定義ファイルのネットワークインターフェースの定義に誤りがあります。

対処

FEFS デザインシートのネットワークインターフェースの定義に誤りがないか確認してください。

fefsconfig: Error: no such device. (bad MGT volume: <*Devicename*>)

意味

FEFS 構成定義ファイルに定義された MGT ボリュームが ノードに存在しません。

パラメーターの説明

<*Devicename*> : デバイス名

対処

FEFS デザインシートの "MGT VOLUME" に正しいデバイスを指定してください。

fefsconfig: Error: no such device. (bad MDT volume: <*Devicename*>)

意味

FEFS 構成定義ファイルに定義された MDT ボリュームが ノードに存在しません。

パラメーターの説明

<*Devicename*> : デバイス名

対処

FEFS デザインシートの "MDT VOLUME" に正しいデバイスを指定してください。

fefsconfig: Error: no such device. (bad OST volume: <*Devicename*>)

意味

FEFS 構成定義ファイルに定義された OST ボリュームが ノードに存在しません。

パラメーターの説明

<*Devicename*> : デバイス名

対処

FEFS デザインシートの "OST VOLUME" に正しいデバイスを指定してください。

fefsconfig: Error: no such device. (bad JOURNAL volume: <*Devicename*>)

意味

FEFS 構成定義ファイルに定義された JOURNAL ボリュームが ノードに存在しません。

パラメーターの説明

<*Devicename*> : デバイス名

対処

FEFS デザインシートに "JOURNAL VOLUME" の正しいデバイスを指定してください。

fefsconfig: Error: no such device. (bad SSD volume: <*Devicename*>)

意味

FEFS 構成定義ファイルに定義された SSD ボリュームが ノードに存在しません。

パラメーターの説明

<*Devicename*> : デバイス名

対処

FEFS デザインシートに "SSD VOLUME" の正しいデバイスを指定してください。

---

fefsconfig: Error: IP address was not found (InfiniBand).

意味

InfiniBand に割り振られたIPアドレスが見つかりません。

対処

InfiniBand に IPアドレスが割り振られているかどうかを確認し、コマンドを再実行してください。

---

fefsconfig: Error: multiple IP addresses were found (InfiniBand).

意味

InfiniBand に割り振られた IPアドレスが複数見つかりました。

対処

IPアドレスが割り振られている InfiniBand が 1つだけか確認し、コマンドを再実行してください。

---

fefsconfig: Error: failed to make configuration file (*filename*).

意味

FEFS セットアップツール用構成定義ファイルが作成できませんでした。

パラメーターの説明

filename : FEFS セットアップツール用構成定義ファイル名

対処

出力先ディレクトリ、およびその配下のファイルのアクセス権限を見直し、コマンドを再実行してください。

---

fefsconfig: Error: exist configuration file.

意味

FEFS セットアップツール用構成定義ファイルがすでに存在します。

対処

新たに FEFS を構築する際に出力された場合は、以前の FEFS セットアップツール用構成定義ファイルを削除してください。

---

fefsconfig: Error: fefs-mdt00 volume was not found.

意味

MDT インデックス0 用のボリュームが見つかりませんでした。

対処

"fefs-mdt00" という名称で MDT用の RAID ボリュームを作成して、コマンドを再実行してください。

---

fefsconfig: Error: new MDT volume was not found.

意味

新しい MDT ボリュームが見つかりませんでした。

対処

新たに追加したい MDT ボリュームの RAID ボリューム名を正しく修正し、コマンドを再実行してください。

fefsconfig: Error: new OST volume was not found.

意味

新しい OST ボリュームが見つかりませんでした。

対処

新たに追加したい OST ボリュームの RAID ボリューム名を正しく修正し、コマンドを再実行してください。

fefsconfig: Error: the number of mdt is large. (number of mdt: *<number of mdt>*)

意味

FEFS 構成定義ファイルに定義された MDT 数が実装 MDT 数より大きくなっています。

パラメーターの説明

<number of mdt> : MDT 数

対処

FEFS デザインシートの LLIO シートの "NUMBER OF MDT" の値を見直してください。

B.2.3 fefs_mkfs コマンド

FATAL:Unknown option '*Opt*'

意味

無効なコマンドオプションが指定されています。

パラメーターの説明

Opt : 無効なコマンドオプション

対処

コマンドオプションを見直してください。

FATAL: Failed to read previous Lustre data from *Dev* (*Err*)

意味

デバイスから FEFS のデータを読み込むことができません。

パラメーターの説明

Dev : デバイス名
Err : エラーコード

対処

ディスクまたはファイルシステムが壊れている可能性があります。fsck.ldiskfs コマンドで確認してください。

FATAL: failed to write local files

意味

FEFSの設定情報のボリュームに対する書込みに失敗しました。

対処

ボリュームに問題がないか確認してください。

ERROR: fefsconfig was not completely finished.

意味

FEFS 設定ファイルの作成に失敗したため、ボリュームの初期化に失敗しました。

対処

FEFS デザインシートを見直し、構成定義ファイルを作成しなおしてください。

ERROR: fefs_setup was not completely finished.

意味

FEFS 構築に失敗したため、ボリュームの初期化に失敗しました。

対処

システムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

ERROR: not exist /etc/opt/FJSVfefs/fefs_tab.

意味

FEFS 設定ファイルが存在していません。

対処

FEFS 構成定義ファイルから FEFS 設定ファイルを作成しなおしてください。

ERROR: not mkfs target.

意味

指定したパラメーターに関連するボリュームがありません。

対処

正しいパラメーターが指定されているか確認してください。

ERROR: Device *Dev* was previously formatted. Use --reformat to reformat it.

意味

フォーマット済みのボリュームに対して、再度フォーマットが実行されました。

パラメーターの説明

Dev: デバイス名

対処

再度フォーマットを実施してもよいデバイスか確認してください。

フォーマットする場合は、"4.19.2 ファイルシステムのデータの保護を解除する手順" を参照してください。

ERROR: modprobe ldiskfs error. ErrorCode=*Err*

意味

ldiskfs モジュールのロードに失敗しました。

パラメーターの説明

Err: エラーコード

対処

ldiskfs モジュールのロードに失敗しています。当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

B.2.4 fefs_mount コマンド

warning: can't allocate memory

意味

FEFS が動作する際のメモリ確保ができません。

対処

メモリの空き状況を確認してください。

warning: failed to open pipe: *Pipe*

意味

パイプの open に失敗しました。

パラメーターの説明

Pipe : パイプのパス

対処

メモリの空き状況を確認してください。

warning: failed write data to pipe: *Pipe*

意味

パイプへのデータ書込みに失敗しました。

パラメーターの説明

Pipe : パイプのパス

対処

メモリの空き状況を確認してください。

warning: failed to close pipe: *Pipe*

意味

パイプの close に失敗しました。

パラメーターの説明

Pipe : パイプのパス

対処

メモリの空き状況を確認してください。

can't allocate memory for options

意味

コマンドオプション用のメモリを確保できません。

対処

メモリの空き状況を確認してください。

Command buffer overflow

意味

コマンドバッファがオーバーフローしました。

対処

メモリの空き状況を確認してください。

ERROR: fefsconfig was not completely finished.

意味

FEFS 設定ファイルの作成に失敗したため、マウントに失敗しました。

対処

FEFS デザインシートを見直し、構成定義ファイルを作成しなおしてください。

ERROR: fefs_setup was not completely finished.

意味

FEFS 構築に失敗したため、マウントに失敗しました。

対処

システムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

ERROR: fefs_mkfs was not completely finished.

意味

FEFS 構築に失敗したため、マウントに失敗しました。

対処

システムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

ERROR: not exist /etc/opt/FJSVfefs/fefs_tab.

意味

FEFS 設定ファイルが存在していません。

対処

FEFS 構成定義ファイルから FEFS 設定ファイルを作成しなおしてください。

ERROR: fefs_mount failed. sts=*Status*.

意味

FEFS のマウントに失敗しました。

パラメーターの説明

Status : 内部処理コード

対処

当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

ERROR: already in progress.

意味

FEFS のマウントまたはアンマウントの処理中です。

対処

しばらくしてから再度実行してください。

ERROR: mount now in progress.

意味

FEFS のマウントの処理中に内部異常が発生しました。

対処

担当保守員 (SE)、または当社 Support Deskに連絡してください。

modprobe *module* error. ErrorCode=*Err*

意味

モジュールのロードに失敗しました。

パラメーターの説明

module : モジュール名

Err : エラーコード

対処

当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

B.2.5 fefssnap コマンド

[ERR.] FEFS 2750 fefssnap -d *Outputdir*: No such directory.

意味

指定されたディレクトリが見つかりません。

パラメーターの説明

Outputdir : -d オプションで指定したディレクトリパス

対処

-d オプションの指定を見直してください。

[ERR.] FEFS 2752 fefssnap Exist temporary directory(*Tmpdir*).

意味

作業領域がすでに存在します。

パラメーターの説明

Tmpdir : 作業領域のディレクトリパス

対処

作業領域を確認してください。

[ERR.] FEFS 2753 fefssnap Cannot create temporary directory(*Tmpdir*).

意味

作業領域が作成できませんでした。

パラメーターの説明

Tmpdir : 作業領域のディレクトリパス

対処

作業領域が作成できる状態になっているか確認してください。

[ERR.] FEFS 2754 fefssnap Cannot create output file.

意味

出力ファイルが作成できませんでした。

対処

出力先ディレクトリの状態を確認してください。

B.2.6 lfs コマンド

共通

error: *Com*: No such file or directory - *Filedir*

意味

指定されたファイルまたはディレクトリは存在しません。

パラメーターの説明

Com : コマンド名
Filedir : 指定されたファイルまたはディレクトリ

対処

ファイル名またはディレクトリ名を見直してください。

error: *Com*: invalid path - *Path*

意味

指定されたパスは無効です。

パラメーターの説明

Com : コマンド名
Path : 指定されたパス名

対処

パス名を見直してください。

error: *Com*: stat failed - *Path*: *Err*

意味

stat に失敗しました。

パラメーターの説明

Com : コマンド名
Path : 指定されたパス名
Err : エラーコード

対処

パス名を見直してください。

Com: invalid option -- '*Opt*'

意味

無効なコマンドオプションが指定されました。

パラメーターの説明

Com : コマンド名
Opt : 無効なコマンドオプション

対処

コマンドオプションを見直してください。

Com: unrecognized option '*Opt*'

意味

無効なコマンドオプションが指定されました。

パラメーターの説明

Com : コマンド名
Opt : 無効なコマンドオプション

対処

コマンドオプションを見直してください。

Com: option requires an argument -- '*Opt*'

意味

コマンドオプションに引数が指定されていません。

パラメーターの説明

Com : コマンド名
Opt : 引数が必要なコマンドオプション

対処

コマンドオプションを見直してください。

Com: option '--*Opt*' requires an argument

意味

コマンドオプションに引数が指定されていません。

パラメーターの説明

Com : コマンド名
Opt : 引数が必要なコマンドオプション

対処

コマンドオプションを見直してください。

Com: option '--*Opt*' is ambiguous

意味

コマンドオプションが正しくが指定されていません。

パラメーターの説明

Com : コマンド名
Opt : 不正なコマンドオプション

対処

コマンドオプションを見直してください。

lfs project

dir - project identifier is not set (inode=*ID1*, tree=*ID2*)

意味

指定したプロジェクトID と異なります。

パラメーターの説明

dir : 指定したディレクトリ
ID1 : チェック対象ディレクトリのプロジェクトID
ID2 : 指定したプロジェクトID

対処

正しいプロジェクト ID を指定してください。

error: project inheritance flag is not set

意味

継承フラグが設定されていません。

対処

必要であれば、継承フラグを設定してください。

QUOTA関連共通

error: use either -u, -g or -p

意味

-u、-g、および -p オプションを同時に指定できません。

対処

コマンドオプションを見直してください。

error: missing quota argument(s)

意味

引数の指定が間違っています。

対処

引数を見直してください。

error: missing quota info argument(s)

意味

引数の指定が間違っています。

対処

引数を見直してください。

error: *Com*: too long path - *Path*

意味

パス名が長すぎます。

パラメーターの説明

Com : コマンド名
Path : パス名

対処

パスを確認してください。

---

error: *Com*: Not a directory - *Path*

意味

パスとして指定されているのはディレクトリではありません。

パラメーターの説明

Com : コマンド名
Path : パス

対処

正しいパスを指定してください。

---

error: *Com*: Not on FEFS - *Path*

意味

指定されたパスは FEFS ではありません。

パラメーターの説明

Com : コマンド名
Path : パス

対処

正しいパスを指定してください。

---

error: can't find id for name *Name*

意味

指定された名前に対応する ID がありません。

パラメーターの説明

Name : 指定された名前

対処

正しい名前を指定してください。

Permission denied.

意味

権限がありません。

対処

適切な権限で実行してください。

Unexpected quotactl error: *Err*

意味

予期せぬerrorが発生しました。

パラメーターの説明

Err : エラーコード

対処

担当保守員 (SE)、または当社 Support Desk に連絡してください。

---

error: -u, -g and -p can't be used more than once

意味

-u、-g、および -p オプションを同時に複数指定できません。

対処

コマンドオプションを見直してください。

---

error: bad block-grace: *Value*

意味

指定されたブロック猶予値が不正です。

パラメーターの説明

Value : 猶予値

対処

正しい猶予値を指定してください。

---

error: bad inode-grace: *Value*

意味

指定されたinode 猶予値が不正です。

パラメーターの説明

Value : 猶予値

対処

正しい猶予値を指定してください。

---

error: neither -u, -g, nor -p was specified

意味

-u、-g、および -p オプションのどれも指定されませんでした。

対処

コマンドオプションを見直してください。

---

error: unexpected parameters encountered

意味

予期せぬパラメーターに遭遇しました。

対処

コマンドオプションを見直してください。

---

setquota failed: *Err*

意味

setquota に失敗しました。

パラメーターの説明

Err : エラーコード

対処

担当保守員 (SE)、または当社 Support Desk に連絡してください。

---

error: bad limit value *Value*

意味

指定されたリミット値が不正です。

パラメーターの説明

Value : リミット値

対処

正しいリミット値を指定してください。

---

warning: block softlimit is smaller than the minimal qunit size.

意味

最小単位 1025 より小さい値の使用ブロック数のソフトリミット値が指定されました。

対処

最小単位 1025 以上の値を指定することを推奨します。

---

warning: block hardlimit is smaller than the minimal qunit size.

意味

最小単位 1025 より小さい値の使用ブロック数のハードリミット値が指定されました。

対処

最小単位 1025 以上の値を指定することを推奨します。

---

warning: inode softlimit is smaller than the minimal qunit size.

意味

最小単位 1025 より小さい値のinode数のソフトリミット値が指定されました。

対処

最小単位 1025 以上の値を指定することを推奨します。

---

warning: inode hardlimit is smaller than the minimal qunit size.

意味

最小単位(1025)より小さい値のinode数のハードリミット値が指定されました。

対処

最小単位(1025)以上の値を指定することを推奨します。

---

error: at least one limit must be specified

意味

リミット値が 1つも指定されていません。

対処

最低 1つはリミット値を指定してください。

---

error: setquota failed while retrieving current quota settings (*Err*)

意味

現在の QUOTA 設定を検索中にエラーが発生しました。

パラメーターの説明

Err : エラーコード

対処

担当保守員 (SE)、または当社 Support Desk に連絡してください。

lfs df

---

error: invalid path '*Path*': *Err*

意味

マウントポイントのパス名の指定が不正です。

パラメーターの説明

Path : パス名
Err : エラーコード

対処

パス名の指定を見直してください。

lfs find

---

err: find: filename|dirname must either precede options or follow options

意味

ファイル名やディレクトリ名は、前置オプションまたは後置オプションのどちらか片方しか取ることができません。

対処

コマンドオプションを見直してください。

---

error: find: no filename|pathname

意味

ファイル名またはパス名がありません。

対処

ファイル名またはパス名を指定してください。

---

error: find failed for *Filedir*.

意味

ディレクトリ名またはファイル名の指定が不正です。

パラメーターの説明

Filedir : ディレクトリ名またはファイル名

対処

ディレクトリ名またはファイル名の指定を見直してください。

error: can't get lov name.: *Msg* (*Err*)

意味

指定されたパスは FEFS ではありません。

パラメーターの説明

Msg : エラーメッセージ
Err : エラーコード

対処

FEFS のパスを指定してください。

lfs setstripe

error: setstripe: missing filename|dirname

意味

ファイル名またはディレクトリ名の指定がありません。

対処

ファイル名またはディレクトリ名を指定してください。

error: setstripe: bad stripe size '*Size*'

意味

ストライプサイズの指定が不正です。

パラメーターの説明

Size : 無効なサイズ値

対処

ストライプサイズを見直してください。

error: setstripe: bad stripe offset '*Offset*'

意味

ファイル書込みを開始するOSTの指定が不正です。

パラメーターの説明

Offset : 無効なオフセット値

対処

ファイル書込みを開始する OST の指定を見直してください。

error: setstripe: bad stripe count '*Num*'

意味

ストライプカウントの指定が不正です。

パラメーターの説明

Num : 無効なストライプカウント

対処

ストライプカウントの指定を見直してください。

warning: stripe size 4G or larger is not currently supported and would wrap: Invalid argument (22)

意味

4194240KiB (4GiB-64KiB) を超えるストライプサイズ指定は無効です。

対処

ストライプサイズは、4194240KiB (4GiB-64KiB) 以下で指定してください。

error: bad stripe_size *Size*, must be an even multiple of 65536 bytes: Invalid argument (22)

意味

ストライプサイズの指定に誤りがあります。65536バイトの倍数である必要があります。

パラメーターの説明

Size : 無効なサイズ値

対処

ストライプサイズの指定を 65536バイトの倍数で指定してください。

error: bad stripe offset *Offset*: Invalid argument (22)

意味

ファイル書込みを開始するOSTの指定が不正です。

パラメーターの説明

Offset : 無効なオフセット値

対処

ファイル書込みを開始するOSTの指定を見直してください。

error: bad stripe count *Num*: Invalid argument (22)

意味

ストライプカウントの指定が不正です。

パラメーターの説明

Num : 無効なストライプカウント

対処

ストライプカウントの指定を見直してください。

'*Path*' is not on a Lustre filesystem: *Msg* (*Err*)

意味

指定されたパスは FEFS ではありません。

パラメーターの説明

Path : パス名
Msg : エラーメッセージ
Err : エラーコード

対処

FEFS のパスを指定してください。

Pool '*Poolname*' is not on filesystem '*Fsname*'

意味

指定された OST_pool が見つかりません。

指定された OST_pool はファイルシステム *Fsname* に存在しません。

パラメーターの説明

Poolname : OST_pool名
Fsname : ファイルシステム名

対処

OST_pool の名前を見直してください。

---

pool '*Poolname*' does not exist

意味

指定された OST_pool が見つかりません。

パラメーターの説明

Poolname : OST_pool名

対処

OST_pool の名前を見直してください。

---

pool '*Poolname*' has no OSTs

意味

指定された OST_pool に OST が登録されていません。

パラメーターの説明

Poolname : OST_pool名

対処

OST を登録してから再実行してください。

---

unable to open '*Path*': *Msg* (*Err*)

意味

指定したパスをオープンできませんでした。

パラメーターの説明

Path : パス名
Msg : エラーメッセージ
Err : エラーコード

対処

パスを見直してください。

---

error on ioctl 0x4008669a for '*Path*' (*fd*): *Msg*

意味

コマンドの実行に失敗しました。

パラメーターの説明

Path : パス名
fd : ファイルディスクリプタ
Msg : エラーメッセージ

対処

エラーメッセージが "stripe already set" の場合は、指定したパスにすでにファイルが存在しています。パス名を見直して再実行してください。

エラーメッセージが "Invalid argument" の場合は設定パラメーターを見直して、再度実行してください。

それでも問題が解決しない場合は、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

---

error: setstripe: create stripe file '*Path*' failed

意味

ストライプの設定に失敗しました。

パラメーターの説明

Path : パス名

対処

このメッセージと同時に出力されているほかのメッセージを参考に対処してください。

lfs getstripe

---

error: getstripe: failed for *Filedir*.

意味

ディレクトリ名またはファイル名の指定が不正です。

パラメーターの説明

Filedir : ディレクトリ名またはファイル名

対処

ディレクトリ名またはファイル名の指定を見直してください。

lfs getdirstripe

---

error: setup_obd_uuid: unknown obduuid: *uuid*

意味

uuid で指定されたOBD 名が不正です。

対処

OBD 名の指定を見直してください。

---

error opening *dir*: No such file or directory (2)

意味

dir で指定されたディレクトリが存在しません。

対処

ディレクトリ名の指定を見直してください。

lfs pool_list

'*Path*' is not on a Lustre filesystem: *Msg* (*Err*)

意味

指定されたマウントポイントは無効です。

パラメーターの説明

Path : パス名
Msg : エラーメッセージ
Err : エラーコード

対処

FEFS のマウントポイントを見直してください。

---

invalid path '*Path*': *Msg* (*Err*)

意味

指定されたパスは無効です。

パラメーターの説明

Path : パス名
Msg : エラーメッセージ
Err : エラーコード

対処

パスを見直してください。

Lustre filesystem '*Path*' not found: *Msg* (*Err*)

意味

指定されたパスは無効です。

パラメーターの説明

Path : パス名
Msg : エラーメッセージ
Err : エラーコード

対処

パスを見直してください。

Cannot open *Poolpath*: *Msg* (*Err*)

意味

指定されたパスに対するプールは無効です。

パラメーターの説明

Poolpath : プールのパス名
Msg : エラーメッセージ
Err : エラーコード

対処

パスを見直してください。

lfs mkdir

---

error: mkdir: missing dirname

意味

ディレクトリ名の指定がありません。

対処

ディレクトリ名を指定してください。

---

error: mkdir: missing stripe offset and count.

意味

MDT のインデックス番号、または、ストライプカウントの指定がありません。

対処

MDT のインデックス番号、または、ストライプカウントを指定してください。

---

error: mkdir: bad stripe offset '*index*'

意味

MDT のインデックス番号が正しく指定されていません。

パラメーターの説明

index : MDT のインデックス番号

対処

MDT のインデックス番号を正しく指定してください。

---

error: mkdir: bad stripe count '<*count*>'

意味

ストライプカウントの指定が不正です。

対処

ストライプカウントを正しく指定してください。

---

unable to open '*Path*': *Msg* (*Err*)

意味

指定したパスをオープンできませんでした。

パラメーターの説明

Path : パス名
Msg : エラーメッセージ
Err : エラーコード

対処

指定したパスが正しいか確認してください。

エラーコードが 12 の場合は、計算機でメモリが不足しています。メモリ使用量を減らすなどの処理を行ってから再実行してください。

それでも問題が解決しない場合は、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

---

error on LL_IOC_LMV_SETSTRIPE '*Path*' (*fd*): *Msg*

意味

指定したパスに対するストライプ設定ができませんでした。

パラメーターの説明

Path : パス名
fd : ファイルディスクリプタ
Msg : エラーメッセージ

対処

エラーメッセージが "Cannot allocate memory" の場合は、計算機でメモリが不足しています。メモリ使用量を減らすなどの処理を行ってから再実行してください。

"Operation not permitted" のメッセージが出力された場合は、本コマンドを root 権限で実行してください。

"stripe already set" のメッセージが出力された場合は、指定したパスにすでにファイルが存在しています。パス名を見直して再実行してください。

" File name too long" , " Inappropriate ioctl for device" のメッセージが出力された場合は、パスを見直してください。

それでも問題が解決しない場合は、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

B.2.7 lctl コマンド

共通

Com: invalid option -- '*Opt*'

意味

無効なコマンドオプションが指定されました。

パラメーターの説明

Com : コマンド名
Opt : 無効なコマンドオプション

対処

コマンドオプションを見直してください。

lctl qosとlctl sqos

QoS config-file not found. filepath=*Path*

意味

filepathで指定された QoS 定義ファイルが見つかりません。

パラメーターの説明

Path : 指定された QoS 定義ファイルのパス

対処

QoS 定義ファイルのパスを確認してください。

QoS config-file error. code=*Err* line=*Line*

意味

QoS 定義ファイルに構文エラーがあります。

パラメーターの説明

Err : エラーコード

文字列	意味
E_NO_MDS	MDS セクションがありません。
E_NO_OSS	OSS セクションがありません。
E_SEC_DOUBLE	セクション名が重複しています。
E_SEC_INVALID	セクション名が不当です。
E_SEC_END	セクションの終了文字 ("}") がありません。

文字列	意味
E_ITEM_INVALID	項目名が不当です。
E_ITEM_DOUBLE	項目名が重複しています。
E_ITEM_NONE	セクション内に、指定必須項目 (qos,nodegrp) がありません。
E_VALUE_INVALID	項目名に対する設定値が不当です。
E_LINE_OVER	1行の最大文字数 1024 を超えています。
E_RATE_OVER	nodegrp で指定した割り当て率の合計が 100% を超えています。
E_IP_INVALID	nodegrp で指定した IPアドレスの指定形式に誤りがあります。
E_NODEGRP_INVALID	usermax[n] または rootmax[n] に対応する nodegrp[n] の定義がありません。
E_LIMIT_INVALID	サーバスレッドに空きがある場合の制限値に関する指定に誤りがあります。

Line : エラーを検出した行

対処

QoS定義ファイルを修正してください。

QoS cannot allocate memory.

意味

QoS 制御に必要なメモリの確保ができません。

対処

メモリの空き状況を確認してください。

QoS System error. func=*Func* route=*Route* code=*Err*

意味

システムエラーが発生しました。

パラメーターの説明

Func : 関数名
Route : エラールート
Err : エラーコード

対処

担当保守員 (SE)、または当社 Support Desk に連絡してください。

QoS status is already on.

意味

QoS 制御は有効のため、lctl qos on コマンドは実行できません。

対処

lctl qos stat コマンドで QoS の状態を確認してください。

QoS status is already off.

意味

QoS 制御は無効のため、lctl qos off コマンドは実行できません。

対処

lctl qos stat コマンドで QoS の状態を確認してください。

QoS filepath is not full path. filepath=*Path*

意味

QoS 定義ファイルのパスが絶対パスで指定されていません。

パラメーターの説明

Path : 指定されたQoS定義ファイルのパス

対処

QoS 定義ファイルのパスを確認してください。

QoS mds is not active.

意味

MDS が有効な状態ではありません。

対処

MDS の状態を確認してください。

QoS file copy error. filepath=*Path*

意味

QoS 定義ファイルのコピー処理に失敗しました。

パラメーターの説明

Path : QoS 定義ファイルのパス

対処

/etc/opt/FJSVfefs ディレクトリに書込みが可能、かつ空き容量があることを確認してください。

QoS command multiple exec error.

意味

コマンドの二重起動です。

対処

コマンドは1つずつ実行してください。

QoS permission denied.

意味

root 権限がありません。

対処

本コマンドは root 権限で実行してください。

QoS invalid uid.

意味

指定されたユーザーID が不正です。

対処

正しいユーザーID を指定してください。

QoS status is off.

意味

QoS機能が無効です。

対処

QoS機能を有効にしてください。

QoS fefs-server is not active.

意味

MDS または OSS が有効な状態ではありません。

また、クライアントノード上で本コマンドは実行できません。

対処

MDS または OSS の状態を確認し、MDS または OSS 上で本コマンドを実行してください。

QoS oss is not active.

意味

OSS が有効な状態ではありません。

対処

OSS の状態を確認してください。

QoS cannot be enabled.

意味

QoS 制御を有効にできません。

対処

QoS 機能を使用する時は、Lustre の NRS (Network Request Scheduler) の policy を fifo に設定してください。

具体的には以下のコマンドを実行してください。

```
[MDSノード]
# lctl set_param mds.MDS.mdt.nrs_policies="fifo reg"
[OSSノード]
# lctl set_param ost.OSS.ost_io.nrs_policies="fifo reg"
```

QoS mds is not MDT0.

意味

コマンドを実行した MDS が MDT0 ではありません。

対処

lctl qosコマンドは、MDT0をマウントする MDS で実行してください。

lctl cqos

CQoS invalid uid.

意味

指定されたユーザーID が不正です。

対処

正しいユーザーID を指定してください。

CQoS status is off.

意味

QoS 機能が無効です。

対処

QoS 機能を有効にしてください。

CQoS permission denied.

意味

root 権限がありません。

対処

本コマンドは root 権限で実行してください。

CQoS invalid mount-point.

意味

指定されたマウントポイントが不正です。

対処

正しいマウントポイントを指定してください。

CQoS cannot allocate memory.

意味

必要なメモリの確保ができません。

対処

メモリの空き状況を確認してください。

CQoS there is no stat data.

意味

有効な統計情報が存在しません。

対処

なし。

CQoS System error. func=*Func* route=*Route* code=*Err*

意味

システムエラーが発生しました。

パラメーターの説明

Func : 関数名
Route : エラールート
Err : エラーコード

対処

担当保守員 (SE)、または当社 Support Deskに連絡してください。

lctl pool系コマンド

Pool *fsname.poolname* not found

意味

fsname.poolname という名前の OST_pool が見つかりません。

対処

正しい OST_pool を指定してください。

Pool *fsname.poolname* already exists

意味

fsname.poolname という名前の OST_pool はすでに存在します。

対処

新しい OST_pool を指定してください。

Pool *fsname.poolname* not empty, please remove all members

意味

fsname.poolname という名前の OST_pool は空ではありません。すべてのメンバーを削除してください。

対処

メンバーをすべて削除してから OST_poolを削除してください。

OST *Name* is already in pool *fsname.poolname*

意味

指定された OST 名はすでに OST_pool *fsname.poolname* に存在します。

パラメーターの説明

Name : OST 名

対処

新しい OST 名を指定してください。

OST *Name* is not part of the '*Fsname*' fs.

意味

指定された OST 名はファイルシステム *Fsname* の一部ではありません。

パラメーターの説明

Name : OST 名
Fsname : ファイルシステム名

対処

正しい OST 名を指定してください。

OST *Name* not found in pool *fsname.poolname*

意味

指定された OST 名は OST_pool *fsname.poolname* にありません。

パラメーターの説明

Name : OST 名

対処

正しい OST 名を指定してください。

Pool *Poolname* not found

意味

指定された OST_pool が見つかりません。

パラメーターの説明

Poolname : OST_pool 名

対処

OST_pool の名前を見直してください。

argument *Arg* must be *<fsname>.<poolname>*

意味

引数の指定が間違っています。

パラメーターの説明

Arg : 引数名

対処

引数を見直してください。

pool_new: File name too long

意味

プール名が長すぎます。

対処

プール名を見直してください。

pool_new: File exists

意味

指定されたプール名はすでに存在します。

対処

プール名を見直してください。

pool_destroy: Directory not empty

意味

ODT_pool が空ではないので削除できません。

対処

ODT_pool のすべてのメンバーを削除してください。

No device found for name MGS: *Msg*

意味

MGS 用のデバイスを見つけることができませんでした。

パラメーターの説明

Msg : エラーメッセージ

対処

"Invalid argument" のエラーメッセージが出力された場合は、本コマンドを MGS 上で実行してください。

"Permission denied" のエラーメッセージが出力された場合は、本コマンドを root 権限で実行してください。

Name does not start with fsname *Fsname*

意味

指定された OST 名が不正です。

パラメーターの説明

Name : OST 名
Fsname : ファイルシステム名

対処

OST 名を見直してください。

Name does not start by *Fsname*-OST nor OST

意味

指定されたOST名が不正です。

パラメーターの説明

Name : OST 名
Fsname : ファイルシステム名

対処

OST 名を見直してください。

---

ost's index in *Index* is not an hexa number

意味

指定されたOSTのインデックスが不正です。

パラメーターの説明

Index : OST インデックス

対処

OST インデックスを見直してください。

---

ostname *Name* does not end with _UUID

意味

指定された OST 名が不正です。

パラメーターの説明

Name : OST 名

対処

OST 名を見直してください。

'*Path*' is not on a Lustre filesystem: *Msg* (*Err*)

意味

指定されたパスは FEFS ではありません。

パラメーターの説明

Path : パス名
Msg : エラーメッセージ
Err : エラーコード

対処

FEFS のパスを指定してください。

Lustre filesystem '*Path*' not found: *Msg* (*Err*)

意味

指定されたパスは無効です。

パラメーターの説明

Path : パス名
Msg : エラーメッセージ
Err : エラーコード

対処

パスを見直してください。

Cannot open *Poolpath*: *Msg* (*Err*)

意味

指定されたパスまたはプール名は無効です。

パラメーターの説明

Poolpath : プールのパス名
Msg : エラーメッセージ
Err : エラーコード

対処

パスまたはプール名を見直してください。

lctl ping

Can't parse process id "*Name*"

意味

NID の指定が間違っています。

パラメーターの説明

Name : NID

対処

有効な NID を指定してください。

Can't parse nid "*IP*"

意味

IPアドレスの指定が間違っています。

パラメーターの説明

IP : IP アドレス

対処

正しい IPアドレスを指定してください。

---

failed to ping *Name*: *Msg*

意味

pingの実行に失敗しました。

パラメーターの説明

Name : NID
Msg : エラーメッセージ

対処

"Input/output error" のメッセージが出力された場合は、パラメーターを見直してください。
"Permission denied" のメッセージが出力された場合は、本コマンドを root 権限で実行してください。

lctl set_param

---

error: set_param: *Parameter*: *Msg*

意味

パラメーターの設定に失敗しました。

パラメーターの説明

Parameter : パラメーター
Msg : エラーメッセージ

対処

パラメーターを見直してください。

---

error: set_param: setting *Parameter=Value*: *Msg*

意味

パラメーターの設定に失敗しました。

パラメーターの説明

Parameter : パラメーター
Value : 設定値
Msg : エラーメッセージ

対処

パラメーターと設定値を見直してください。

---

error: set_param: *Msg* opening *Parameter*

意味

パラメーターの設定に失敗しました。

パラメーターの説明

Msg : エラーメッセージ
Parameter : パラメーター

対処

パラメーターを見直してください。

lctl get_param

error: get_param: *Parameter*: *Msg*

意味

パラメーターの取得に失敗しました。

パラメーターの説明

Parameter : パラメーター
Msg : エラーメッセージ

対処

パラメーターを見直してください。

error: get_param: opening(' *Parameter* ') failed: *Msg*

意味

パラメーターの取得に失敗しました。

パラメーターの説明

Parameter : パラメーター
Msg : エラーメッセージ

対処

パラメーターを見直してください。

error: get_param: read(' *Parameter* ') failed: *Msg*

意味

パラメーターの取得に失敗しました。

パラメーターの説明

Parameter : パラメーター
Msg : エラーメッセージ

対処

パラメーターを見直してください。

lctl lfsck start と lctl lfsck stop

device name is too long. Valid length should be less than *Maxsize*

意味

デバイス名が長すぎます。

パラメーターの説明

Maxsize: 指定可能なデバイス名長

対処

デバイス名を確認してください。

invalid switch: -c ‘*str*’. valid switches are:
empty ('on'), or 'off' without space. For example:
'-c', '-con', '-coff'

意味

オプションの引数が不正です。

パラメーターの説明

Str: 無効な文字列

対処

オプションに指定した文字列を確認してください。

Invalid option, '-h' for help.

意味

無効なオプションが指定されました。

対処

正しいオプションを指定してください。

Must specify device to start LFSCK.

意味

デバイスの指定に誤りがあります。

対処

デバイスの指定に誤りがないか確認してください。

Fail to pack ioctl data: rc = *Err*.

意味

lfsckに必要なデータを作成できませんでした。

パラメーターの説明

Err: エラーコード

対処

担当保守員(SE)、または当社Support Desk に連絡してください。

Fail to start LFSCK *Msg*

意味

lfsckが開始できませんでした。

パラメーターの説明

Msg: エラーメッセージ

対処

担当保守員(SE)、または当社Support Desk に連絡してください。

Fail to stop LFSCK *Msg*

意味

lfsckが停止できませんでした。

パラメーターの説明

Msg: エラーメッセージ

対処

担当保守員(SE)、または当社Support Desk に連絡してください。

B.2.8 fsck.ldiskfs コマンド

fsck.ldiskfs: Only one of the options -p/-a, -n or -y may be specified.

意味

-p、-n、および -y オプションを同時に複数指定できません。

対処

コマンドオプションを見直してください。

Invalid non-numeric argument to -b ("*superblock*")

意味

引数が非数値です。

パラメーターの説明

superblock : スーパーブロック

対処

正しい数値を指定してください。

Device is {mounted | in use}.

意味

デバイスがマウント中か使用中です。

パラメーターの説明

Device : デバイス

対処

FEFS サービスを停止してから実行してください。

fsck.ldiskfs: *Msg* while trying to open *device*

意味

FEFS デバイスのチェック・修復に失敗しました。

パラメーターの説明

Msg : エラーメッセージ
device : デバイス

対処

"Out of memory" のメッセージが出力された場合は、計算機でメモリが不足しています。メモリ使用量を減らすなどの処理を行ってから再実行してください。

"Operation not permitted" または "Permission denied" のメッセージが出力された場合は、本コマンドを root権限で実行してください。

"Is a directory" のメッセージが出力された場合は、ディレクトリが指定されました。適切なデバイスを指定して再実行してください。

"Bad magic number in super-block" のメッセージが出力された場合は、正しくないスーパーブロックが指定されました。適切なスーパーブロックを指定して再実行してください。

"No such file or directory" のメッセージが出力された場合は、存在しないデバイスが指定されました。適切なデバイスを指定して再実行してください。

それでも問題が解決しない場合は、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

fsck.ldiskfs: invalid option -- '*Opt*'

意味

無効なコマンドオプションが指定されました。

パラメーターの説明

Opt : 無効なコマンドオプション

対処

コマンドオプションを見直してください。

fsck.ldiskfs: unrecognized option '*Opt*'

意味

無効なコマンドオプションが指定されました。

パラメーターの説明

Opt : 無効なコマンドオプション

対処

コマンドオプションを見直してください。

fsck.ldiskfs: option requires an argument -- '*Opt*'

意味

コマンドオプションに引数が指定されていません。

パラメーターの説明

Opt : 引数が必要なコマンドオプション

対処

コマンドオプションを見直してください。

B.2.9 fefsbackup コマンド [PG]

Error : 0002 : Parameters not specified in config file "<*config*>".

意味

設定ファイルにパラメーターが設定されていませんでした。

パラメーターの説明

config : 設定ファイル名

対処

システム管理者に連絡して、設定ファイルを確認してください。

Error : 0005 : Cannot open config file "<*config*>".

意味

設定ファイルのファイルオープンに失敗しました。

パラメーターの説明

config : 設定ファイル名

対処

システム管理者に連絡して、設定ファイルを確認してください。

Error : 0007 : Parameter <*param*> specification is wrong in config file "<*config*>".

意味

設定ファイル <*config*> の <*param*> の指定に誤りがあります。

パラメーターの説明

config : 設定ファイル名
param : パラメーター名

対処

システム管理者に連絡して、設定ファイルおよび WORK_DIR を確認してください。

Error : 0008 : Internal error has occurred(<*file*>:<*line*>:<*op*> failed : <*err*>.)

意味

アーカイブ元ファイルシステムへのファイルアクセスでエラーが発生しました。

パラメーターの説明

file : ファイル名
line : 行数
op : システムコール名
err : エラーメッセージ

対処

担当保守員 (SE)、または当社 Support Desk に連絡してください。

Error : 0009 : Cannot connect to <*host*> : No route to host

意味

接続先ノードに接続ができませんでした。

パラメーターの説明

host : 接続先ノード名

対処

host で表示されるノードに接続できません。接続先ノードの状態を確認してください。システム管理者に連絡してください。

Error : 0010 : Cannot connect to <*host*> : Connection refused.

意味

接続先ノードから接続が拒否されました。

パラメーターの説明

host : 接続先ノード名

対処

host で表示されるノードに接続できません。接続先ノードの状態を確認してください。システム管理者に連絡してください。

Error : 0011 : Cannot connect to <*host*> : Host key verification failed.

意味

接続先ノードの ssh キーが変更されています。

パラメーターの説明

host : 接続先ノード名

対処

host で表示されるノードに接続できません。接続先ノードのsshキーが変更されています。システム管理者に連絡してください。

Error : 0013 : rsync unknown error.

意味

rsyncコマンドがエラー終了しました。

対処

担当保守員(SE)、または当社Support Desk に連絡してください。

Error : 0014 : tar unknown error.

意味

tar コマンドがエラー終了しました。

対処

担当保守員(SE)、または当社Support Desk に連絡してください。

Error : 0015 : missing package <*package*>.

意味

必要なパッケージが不足しています。

パラメーターの説明

package : パッケージ名

対処

担当保守員(SE)、または当社Support Desk に連絡してください。

Error : 0017 : backupdir <*backupdir*> does not exist.

意味

fefsbackup_rsync.confのBACKUP_ROOTに指定されたディレクトリ *backupdir* が見つかりません。

パラメーターの説明

backupdir : ディレクトリ名

対処

システム管理者に連絡し、fefsbackup_rsync.confのBACKUP_ROOTを確認してください。

Error : 0018 : Cannot connect to <*host*> : Name or service not known

意味

接続先ノード<*host*>に接続ができませんでした。

パラメーターの説明

host :接続先ノード

対処

接続先ノードが名前解決できるかを確認してください。システム管理者に連絡してください。

Error : 0019 : Cannot get request status information <*request ID*>.

意味

リクエスト状態の取得中に失敗しました。

パラメーターの説明

request ID : リクエストID名

対処

再実行しエラーになる場合は、<*request ID*> を delete して再実行してください。

Error : 1005 : password authentication failed.

意味

パス認証に失敗しました。

対処

パスワードを確認してください。

Error : 1006 : cannot access <*pathlist*>: <*errmsg*>

意味

pathlist にアクセスできません。

パラメーターの説明

pathlist : pathlist名
errmsg : エラー内容

対処

表示されるエラー内容を確認してアクセスできない原因を取り除いてください。

Error : 1007 : pathlist <*pathlist*> is not a regular file

意味

pathlistが通常ファイルではありません。

パラメーターの説明

pathlist : pathlist名

対処

pathlistを確認してください。

Error : 1008 : cannot specified -f option and target file at once.

意味

-f オプションと引数でのパス指定は同時にはできません。

対処

オプションの指定を確認してください。

Error : 1009 : pathlist <*pathlist*> is an empty file.

意味

-f オプションで指定した <*pathlist*> が空ファイルです。

パラメーターの説明

pathlist : pathlist名

対処

pathlist を確認してください。

Error : 1011 : target file is not found.

意味

指定した対象にファイルが含まれていません。

対処

コピー対象またはファイルを確認してください。

Error : 1012 : request ID reached limit per one day. please use "-L" option.

意味

1日に自動取得できるリクエストIDの上限に達しました。

対処

-L オプションでリクエストIDを指定してください。

Error : 1013 : label <*request ID*> request ID already used.

意味

-L オプションで指定したリクエストID名はすでに存在します。

パラメーターの説明

request ID : リクエストID名

対処

別のリクエストID名を指定してください。

Error : 1014 : label <*request ID*> is invalid.

意味

-L オプションで指定された文字列に使用不可能な文字が含まれています。

パラメーターの説明

request ID : リクエストID名

対処

-L オプションで使用可能な文字について確認してください。

Error : 1015 : label <*request ID*> is too long.

意味

-L オプションで指定された文字列が長すぎます。

パラメーターの説明

request ID : リクエストID名

対処

-L オプションで使用可能な文字数について確認してください。

Error : 1019 : -d option must be specified, or invalid option is specified before -d option.

意味

-d オプションが指定されていない、または -d オプションの前に不正なオプションが指定されています。

対処

-d オプションでコピー先ディレクトリを必ず指定してください。-d オプションを指定している場合は、-d オプションより前のオプションに誤りがないか確認してください。

Error : 1022 : request ID <*request ID*> is invalid.

意味

無効なリクエストIDが指定されました。

パラメーターの説明

request ID : リクエストID名

対処

list サブコマンドおよび status サブコマンドでリクエストID を確認してください。

Error : 1023 : request ID <*request ID*> status is RUNNING.

意味

-Rオプションで指定したリクエストIDが実行中です。

パラメーターの説明

request ID : リクエストID名

対処

status サブコマンドで status が STOPPED のリクエストIDを指定してください。

Error : 1024 : delete target must be specified.

意味

delete対象が指定されていません。

対処

delete対象とするアーカイブ情報のリクエストIDを指定してください。

Error : 1025 : There is a new line code in the file name <*filename*>.

意味

対象のファイル <*filename*> に改行が含まれています。

パラメーターの説明

filename : ファイル名

対処

改行を含まないファイル名に変更してください。

Error : 1026 : This command is available to only Administrators.

意味

実行ユーザーが不正です。

対処

root ユーザーで実行してください。

Error : 1027 : The permission of the work directory is invalid.

意味

一時領域のパーミッションが不正です。

対処

アクセス可能なパーミッションにしてください。

Error : 1028 : The specified work directory does not exist.

意味

一時領域が存在しません。

対処

正しい一時領域を指定してください。

Error : 1029 : subcommand is not specified.

意味

サブコマンドが指定されていません。

対処

サブコマンドを指定してください。

Error : 1030 : Available subcommands are "list", "copy","status" and "delete"

意味

無効なサブコマンドが指定されました。

対処

"list"、"copy"、"status"、"delete"のいずれかサブコマンドを指定してください。

Error : 1031 : Exclusive options -L, -u and -R are specified.

意味

-L,-u,-Rオプションは同時に指定できません。

対処

コマンド指定を見直してください。

Error : 1032 : Invalid option.

意味

オプション指定が不正です。

対処

オプション指定を見直してください。

Error : 1033 : File was specified as a work directory.

意味

一時領域にファイルが指定されました。

対処

一時領域にディレクトリを指定してください。

Error : 1035 : destdir <*path*> does not exist.

意味

宛先に指定したディレクトリがありませんでした。

パラメーターの説明

path : パス名

対処

正しいディレクトリを指定してください。また、パス名には絶対パスを指定してください。

Error : 1036 : <*path*> is not a directory.

意味

ディレクトリが指定されませんでした。

パラメーターの説明

path : パス名

対処

ディレクトリを指定してください。

B.2.10 ファイル特定ツール共通

Permission denied

意味

権限がありません。

対処

rootユーザーで実行してください。

Can't allocate memory

意味

FEFSが動作する際のメモリ確保ができません。

対処

空きメモリを十分確保して再度実行してください。

Can't open file (*File*)

意味

ファイルのオープンに失敗しました。

パラメーターの説明

File:確保に失敗したオブジェクト

対処

空きメモリを十分確保して再度実行してください。

Invalid argument (*Arg*)

意味

引数が不正です。

パラメーターの説明

Arg:指定した引数の値

対処

正しい値を指定してください。

File already exists (*File*)

意味

ファイルがすでに存在しています。

パラメーターの説明

File:ファイル名

対処

存在しないファイル名を指定してください。

Error while trying to resolve filename

意味

内部エラーが発生しました。

対処

担当保守員 (SE)、または当社 Support Desk に連絡してください。

Invalid option

意味

オプション指定が不正です。

対処

正しいオプションを指定してください。

Too many arguments

意味

引数が多すぎます。

対処

引数には1024個のinode番号しか指定できません。引数を確認してください。

Read inode failed (*inode*), err=*errno*

意味

inodeの読み込みに失敗しました。

パラメーターの説明

inode: inode番号
errno: エラー番号

対処

担当保守員 (SE)、または当社 Support Desk に連絡してください。

Invalid inode number (inode)

意味

不正なinode番号が指定されました。

パラメーターの説明

inode: inode番号

対処

正しいinode番号を指定して、再実行してください。

Device Msg while opening filesystem

意味

ファイルシステムのオープンに失敗しました。

パラメーターの説明

Device: デバイス名
Msg: エラーメッセージ

対処

正しいデバイスを指定して、再実行してください。

Closing filesystem failed: *Msg*

意味

ファイルシステムのクローズに失敗しました。

パラメーターの説明

Msg: エラーメッセージ

対処

対処不要です。

Invalid execution environment

意味

実行環境が不正です。

対処

実行環境を見直してください。

Invalid mount point *Mnt*

意味

指定したマウントポイントが不正です。

パラメーターの説明

Mnt: マウントポイント

対処

正しいマウントポイントを指定して、再実行してください。

No such file *File*

意味

入力ファイルが存在しません。

パラメーターの説明

File: 入力ファイル名

対処

正しいファイルを指定して、再実行してください。

Can't convert path (*fid*)

意味

fidをパスに変換できませんでした。

パラメーターの説明

fid: FID

対処

fidに対応するパスが削除された可能性があります。再実行してください。

Can't merge files

意味

ファイルを出力できませんでした。

対処

ディスク容量に空きがあることを確認して、再実行してください。

<*Input*>: No such file or directory while opening filesystem

意味

入力ファイルが存在しません。

パラメーターの説明

Input: 入力ファイル名

対処

正しいファイルを指定して、再実行してください。

<*inode*>: Use no block

意味

データブロックを使用していません。

パラメーターの説明

inode: inode番号

対処

対処不要です。

Error while trying to resolve filename

意味

指定したMDTデバイスが不正です。

対処

正しいMDTデバイスを指定してください。

B.2.11 find_file_ost コマンド

Read file failed (*file*), err=*errno*

意味

inode の読み込みに失敗しました。

パラメーターの説明

file: ファイル情報
errno: エラー番号

対処

担当保守員 (SE)、または当社 Support Disk に連絡してください。

B.2.12 convert_fid2path コマンド

missing operand

意味

オペランドが不足しています。

対処

オペランドを指定してください.

Invalid inputfile (*file*)

意味

入力ファイルが間違っています。

パラメーターの説明

file: 入力ファイル名

対処

正しい入力ファイルを指定してください。

B.2.13 force_intr コマンド

target doesn't match target name format

意味

指定した *target* がターゲット・デバイスの命名規則に一致しません。

パラメーターの説明

target: 指定したターゲット・デバイス名

対処

正しいターゲット・デバイス名を指定してください。

Unknown command: *cmd*

意味

指定されたコマンドが不明です。

パラメーターの説明

cmd: -m オプションの引数として指定したコマンド（文字列）

対処

コマンドとしてdeactivate、activate、status のうちのどれかが指定されていることを確認してください。

Specify either -c or -s

意味

指定されたオプションが不正です。

対処

オプションとして-cか-sのうちのどちらかが指定されていることを確認してください。

Cannot specify -s and -a at the same time without status command

意味

指定されたオプションが不正です。

対処

-s オプションと -a オプションを同時に指定した場合は、status サブコマンドが指定されていることを確認してください。

B.2.14 evict_client コマンド

no such nid correspond to <IP>

意味

指定された IPアドレスに対応するノードは存在しません。

パラメーターの説明

<IP>: evictを行うノードのIPアドレス

対処

正しいIPアドレスを指定してください。

missing operand

意味

オペランドが不足しています。

対処

オペランドを指定してください。

evict_client is available to only Administrators on MDS or OSS node.

意味

このノードでは実行できません。

対処

MDS または OSS で実行してください。

B.2.15 fefs_yaml2csv コマンド

fefs_yaml2csv: '*Filename*': File not found.

意味

入力に指定されたファイルが見つかりません。

パラメーターの説明

Filename : 入力ファイル名（ノード情報定義ファイルまたは FX サーバ用ノード情報定義ファイル）

対処

コマンドに指定したファイルを見直してください。

fefs_yaml2csv: '*Filename*': File is invalid.

意味

入力に指定されたファイルが不正です。

パラメーターの説明

Filename : 入力ファイル名（ノード情報定義ファイルまたは FX サーバ用ノード情報定義ファイル）

対処

コマンドに指定したファイルを見直してください。

fefs_yaml2csv: overwrite '*Filename*' ?

意味

出力に指定されたファイルがすでに存在します。

パラメーターの説明

Filename : 出力ファイル名

対処

出力ファイルを上書きしてよい場合は y を、コマンドを中止する場合は n を入力してください。

y を入力した場合は処理が継続されます。
n を入力した場合は処理が中止します。

fefs_yaml2csv: '*Filename*': File is invalid format. (No essential key(s) *info*)

意味

入力に指定されたファイルのフォーマットが不正です。

パラメーターの説明

Filename : 入力ファイル名（ノード情報定義ファイルまたは FX サーバ用ノード情報定義ファイル）
info : 処理対象のファイルによってメッセージが異なります。
"in [node]" ... ノード情報定義ファイル
"in [node(ft)]" ... FX サーバ用ノード情報定義ファイル

対処

ノード情報定義ファイル、または FX サーバ用ノード情報定義ファイルを CSV へ変換する際のエラーです。
入力に指定したファイルが適切な形式になっているか確認してください。

B.2.16 fefs_deactivate コマンド

Unknown command: *cmd*

意味

指定されたコマンドが不明です。

パラメーターの説明

cmd : -m オプションの引数として指定したコマンド (文字列)

対処

コマンドとして activate、deactivate、status のうちのどれかが指定されていることを確認してください。

Invalid mount point *Mnt*

意味

指定したマウントポイントが不正です。

パラメーターの説明

Mnt : マウントポイント

対処

正しいマウントポイントを指定して、再実行してください。

FEFS service isn't started.

意味

FEFS サービスが起動していません。

対処

FEFS サービスを起動してから、コマンドを再実行してください。

Deactivate processing failed for *system Msg*.

意味

ファイルシステム (FEFS/LLIO) の切離し処理に失敗しました。

パラメーターの説明

system : FEFS or LLIO
Msg : エラーメッセージ

対処

指定したマウントポイントを見直して再実行してください。
それでも問題が解決しない場合は、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

Activate processing failed for *system Msg*.

意味

ファイルシステム (FEFS/LLIO) の組込み処理に失敗しました。

パラメーターの説明

system : FEFS or LLIO
Msg : エラーメッセージ

対処

指定したマウントポイントを見直して再実行してください。
それでも問題が解決しない場合は、当メッセージを含むシステムログファイルを採取し、担当保守員 (SE)、または当社 Support Desk に連絡してください。

付録C　FEFS の構築後に必要な設定

FEFS を構築したあと、必ず、以下の設定を行ってください。

C.1　FEFS スクリプトの設定

FEFS スクリプトの設定手順を示します。本手順は、特に断りがない限り、運用系システム管理ノード上で行ってください。

複数のシステム管理ノードが存在する環境での設定手順については、"C.2 複数システム管理ノード環境での FEFSスクリプトの設定"を参照してください。

注意

以下の設定は、クラスタ内のすべてのノードが起動している、かつ、FEFS マウントが完了している状態で行ってください。

1. FEFS クライアントの LNet の NID 取得
   以下を実行してください。
   FEFSクライアントとなっている計算クラスタが複数存在する場合は、FEFSクライアントとなっている計算クラスタ単位に実行してください。

```
# pmexe -c 〈計算クラスタ名〉 --nodetype 〈FEFSクライアントとなっているノードタイプ〉 ¥
--stdout "lctl list_nids" > /etc/opt/FJSVfefs/lnetid_list_〈計算クラスタ名〉
```

   ※ 出力するファイルのファイル名には、上記のように、ファイル名の最後に対象の計算クラスタ名を入れてください。

2. FEFS クライアントの管理用ネットワークの IP アドレス取得
   以下を実行してください。
   PG ノードの多目的ノードが FEFS クライアントとなっている場合は、あわせて取得してください。
   FEFS クライアントとなっている計算クラスタが複数存在する場合は、FEFS クライアントとなっている計算クラスタ単位に実行してください。

```
# pashowclst -c 〈計算クラスタ名〉 -v -l --nodetype CCM --data | grep PG | awk -F ',' '{print $4,$7}' > tmp.txt
# pashowclst -c 〈計算クラスタ名〉 -v -l --nodetype LN,CN --data | grep PG | awk -F ',' '{print $4,$6}' >> tmp.txt
# sort tmp.txt | uniq > /etc/opt/FJSVfefs/mngnet_list_〈計算クラスタ名〉
# rm -f tmp.txt
```

   ※FEFS クライアントとなっている PG ノードがない場合は、/etc/opt/FJSVfefs/mngnet_list_<*計算クラスタ名*> の空ファイルを作成してください。

3. FEFS サーバの MDT および OST のリスト取得
   以下を実行してください。
   FEFS サーバとなっているストレージクラスタが複数存在する場合は、FEFS サーバとなっているストレージクラスタ単位に実行してください。

```
# pmexe -c 〈ストレージクラスタ名〉 --nodetype MDS --stdout "ls /proc/fs/lustre/md[st]/" | ¥
egrep "¥-MDT.{4}$" > /etc/opt/FJSVfefs/reconnect_srv_gmds_mdt_list_〈ストレージクラスタ名〉
# pmexe -c 〈ストレージクラスタ名〉 --nodetype OSS --stdout "ls /proc/fs/lustre/obdfilter/" | ¥
egrep "¥-OST.{4}$" > /etc/opt/FJSVfefs/reconnect_srv_goss_ost_list_〈ストレージクラスタ名〉
```

   ※ 出力するファイルのファイル名には、上記のように、ファイル名の最後に対象のストレージクラスタ名を入れてください。

4. 取得したファイルの配置
   手順1～3 で取得した以下のファイルを、待機系システム管理ノードの /etc/opt/FJSVfefs/ ディレクトリにも配置してください。

   ー /etc/opt/FJSVfefs/lnetid_list_<*計算クラスタ名*>

   ー /etc/opt/FJSVfefs/mngnet_list_<*計算クラスタ名*>

   ー /etc/opt/FJSVfefs/reconnect_srv_gmds_mdt_list_<*ストレージクラスタ名*>

   ー /etc/opt/FJSVfefs/reconnect_srv_goss_ost_list_<*ストレージクラスタ名*>

5. システム監視プラグイン設定ファイルのバックアップ
   運用系システム管理ノードのシステム監視プラグイン設定ファイル /etc/opt/FJSVtcs/pamoplugin.conf をバックアップしてください。

6. システム監視プラグイン設定ファイルの設定内容の確認

   システム監視プラグイン処理機能の現在の設定内容を /etc/opt/FJSVtcs/pamoplugin.conf ファイルに出力します。

```
# pamopluginadm --show > /etc/opt/FJSVtcs/pamoplugin.conf
```

7. システム監視プラグイン設定ファイルの編集
   運用系システム管理ノード上の /etc/opt/FJSVtcs/pamoplugin.conf ファイルに以下の設定を追加します。

   a. 計算クラスタの定義に、以下を追加します。FEFS クライアントとなっている計算クラスタが複数存在する場合は、FEFS クライアントとなっている計算クラスタ単位に記述してください。

```
PluginCmd {
    NodeType = "<FEFS クライアントとなっているノードタイプ>"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefs/sbin/plugin_evict_client.sh -s <ストレージクラスタ名>"
}
PluginCmd {
    NodeType = "<FEFS クライアントとなっているノードタイプ>"
    ServiceName = "OS"
    Status = "-"
    Cmd = "bash /opt/FJSVfefs/sbin/plugin_evict_client.sh -s <ストレージクラスタ名>"
}
PluginCmd {
    NodeType = "<FEFS クライアントとなっているノードタイプ>"
    ServiceName = "FEFS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefs/sbin/plugin_evict_client.sh -s <ストレージクラスタ名>"
}
```

   ※ FEFS クライアントとなっているノードタイプが複数ある場合は、以下のように、コンマ (,) で区切って指定します。例えば、CCM、LN、および CN が FEFS クライアントとなっている場合は、以下のように指定します。

```
NodeType = "CCM,LN,CN"
```

```
PluginCmd {
    NodeType = "GIO"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefs/sbin/plugin_lnet_router_dwn.sh -s <ストレージクラスタ名>"
}
```

   ※ -s オプションは、当該計算クラスタ配下の FEFS クライアントがマウントしている FEFS サーバが属しているとなっているストレージクラスタ名を指定します。

   b. ストレージクラスタの定義に、以下を追加します。

```
PluginCmd {
    NodeType = "MDS"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefs/sbin/plugin_reconnect_srv.sh -c <計算クラスタ名> --nodetype MDS"
}
PluginCmd {
    NodeType = "OSS"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefs/sbin/plugin_reconnect_srv.sh -c <計算クラスタ名> --nodetype OSS"
}
```

※ -c オプションについては、当該ストレージクラスタで構成されるファイルシステムに対する FEFS クライアントが、複数の計算クラスタに存在する場合は、計算クラスタ単位に記述します。以下は、計算クラスタ calc1 配下と計算クラスタ calc2 配下に、FEFS クライアントが存在する場合の例です。

```
PluginCmd {
   NodeType = "MDS"
   ServiceName = "OS"
   Status = "x"
   Cmd = "bash /opt/FJSVfefs/sbin/plugin_reconnect_srv.sh -c calc1 --nodetype MDS"
}
PluginCmd {
   NodeType = "OSS"
   ServiceName = "OS"
   Status = "x"
   Cmd = "bash /opt/FJSVfefs/sbin/plugin_reconnect_srv.sh -c calc1 --nodetype OSS"
}
PluginCmd {
   NodeType = "MDS"
   ServiceName = "OS"
   Status = "x"
   Cmd = "bash /opt/FJSVfefs/sbin/plugin_reconnect_srv.sh -c calc2 --nodetype MDS"
}
PluginCmd {
   NodeType = "OSS"
   ServiceName = "OS"
   Status = "x"
   Cmd = "bash /opt/FJSVfefs/sbin/plugin_reconnect_srv.sh -c calc2 --nodetype OSS"
}
```

8. システム監視プラグイン設定ファイルの登録
   /etc/opt/FJSVtcs/pamoplugin.conf ファイルの記述内容をジョブ運用ソフトウェアに登録してください。

```
# pamopluginadm --set
```

9. システム監視プラグイン設定ファイルの設定内容の確認
   手順8 で登録した /etc/opt/FJSVtcs/pamoplugin.conf ファイルが正しく登録されているか、設定内容を確認してください。

```
# pamopluginadm --show
```

10. MDS の crond の確認
    MDS のcrond が on になっていることを確認してください。

```
# pmexe -c <ストレージクラスタ名> --nodetype MDS --stdout "systemctl is-enabled crond.service"
```

    MDS のcrond が enabled になっていない場合は、MDS 上で以下を実行して、crond を enable にしてください。

```
[MDS ノード]
# systemctl enable crond.service
```

11. MDS の crond の状態確認
    MDS の crond の状態が起動中になっていることを確認してください。

```
# pmexe -c <ストレージクラスタ名> --nodetype MDS --stdout "systemctl status crond.service"
```

    MDS の crond の状態が起動中となっていない場合は、MDS 上で以下を実行して、crond を起動してください。

```
[MDS ノード]
# systemctl start crond.service
```

12. MDSのcrontabの設定
    FEFS のスクリプトを MDS の cron に設定してください。

```
# pmexe -c <ストレージクラスタ名> --nodetype MDS --stdout "/opt/FJSVfefs/sbin/activate_device_cron.sh"
```

13. crontabの設定の確認
    MDS の cron にスクリプトが正しく設定されているか確認してください。

```
# pmexe -c 〈ストレージクラスタ名〉 --nodetype MDS --stdout "crontab -l"
```

    以下が設定されているか確認します。

    */1 * * * * bash /opt/FJSVfefs/sbin/activate_device.sh

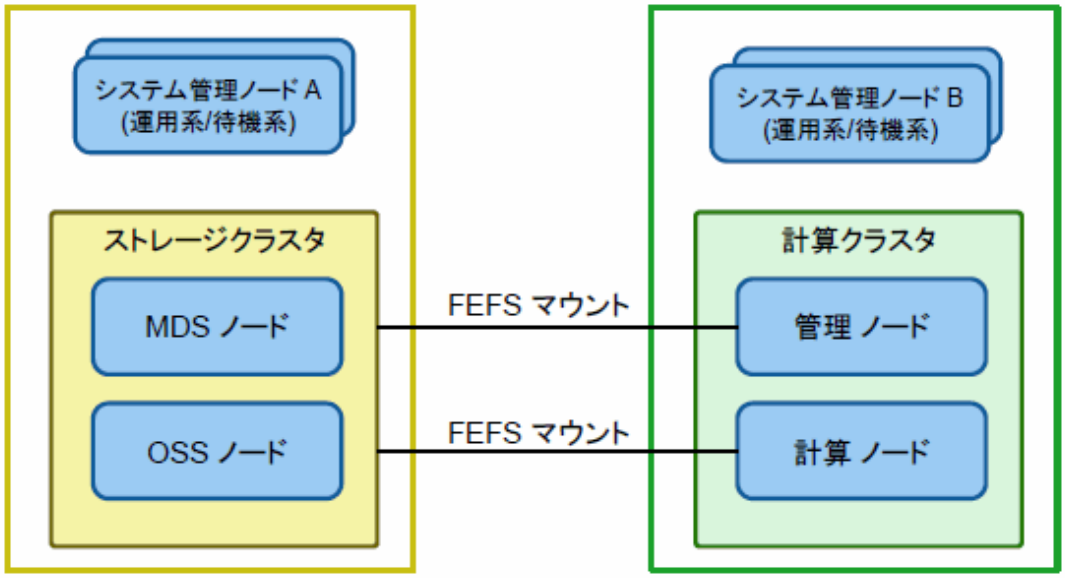
C.2 複数システム管理ノード環境での FEFSスクリプトの設定

複数のシステム管理ノードが存在する環境では、ここで説明する手順で FEFS スクリプトの設定を行ってください。複数のシステム管理ノードが存在する環境ではない場合は、本手順は行わないでください。

以下の図で、ストレージクラスタを持つシステム管理ノードA (運用系/待機系) を以降、システム管理ノードA と表記します。また、計算クラスタを持つシステム管理ノードB (運用系/待機系) を以降、システム管理ノードB と表記します。

複数のシステム管理ノードが存在する環境とは、冗長構成のシステム管理ノードが複数存在する環境、かつ、システム管理ノードA 配下のストレージクラスタに属する FEFS サーバ (MDS および OSS) で構成される FEFS を、システム管理ノードB 配下の計算クラスタに属する FEFS クライアント (CCM、LN、または CNなど) が、FEFSマウントしていることを意味します。

図C.1 複数のシステム管理ノードが存在する環境



上記の図を基に、以下に設定手順を示します。

注意

- 以下の設定は、クラスタ内のすべてのノードが起動している、かつ、FEFS マウントが完了している状態で行ってください。
- システム管理ノード A およびシステム管理ノード B 配下に、同じクラスタ名の計算クラスタ または ストレージクラスタが存在する場合は、本手順で設定されるスクリプトは動作しません。同じクラスタ名の計算クラスタまたはストレージクラスタは設定しないでください。

1. FEFS クライアント の LNetの NID 取得
   システム管理ノード B の運用系ノードで以下を実行して、FEFS クライアントの LNet の NID を取得してください。
   なお、FEFS クライアントとなっている計算クラスタが複数存在する場合は、FEFS クライアントとなっている計算クラスタ単位に実行してください。

```
[システム管理ノード B]
# pmexe -c 〈計算クラスタ名〉 --nodetype 〈FEFSクライアントとなっているノードタイプ〉 ¥
--stdout "lctl list_nids" > /etc/opt/FJSVfefs/lnetid_list_〈計算クラスタ名〉
```

※ 出力するファイルのファイル名には、ファイル名の最後に計算クラスタ名を入れてください。

2. ファイルの配置
手順1 で取得した /etc/opt/FJSVfefs/lnetid_list_<*計算クラスタ名*>ファイルを、システム管理ノード B の待機系ノードとシステム管理ノードA の運用系および待機系ノードの、/etc/opt/FJSVfefs/ ディレクトリにも配置してください。

3. FEFS クライアントの管理用ネットワークの IP アドレス取得
システム管理ノード B の運用系ノードで以下を実行して、FEFS クライアントの管理用ネットワークの IP アドレスを取得してください。
PG ノードの多目的ノードが FEFS クライアントとなっている場合は、あわせて取得してください。
FEFS クライアントとなっている計算クラスタが複数存在する場合は、FEFS クライアントとなっている計算クラスタ単位に実行してください。

```
[システム管理ノード B]
# pashowclst -c <計算クラスタ名> -v -I --nodetype CCM --data | grep PG | awk -F ',' '{print $4,$7}' > tmp.txt
# pashowclst -c <計算クラスタ名> -v -I --nodetype LN,CN --data | grep PG | awk -F ',' '{print $4,$6}' >> tmp.txt
# sort tmp.txt | uniq > /etc/opt/FJSVfefs/mngnet_list_<計算クラスタ名>
# rm -f tmp.txt
```

※ FEFS クライアントとなっている PG ノードがない場合は、/etc/opt/FJSVfefs/mngnet_list_<*計算クラスタ名*> の空ファイルを作成してください。

4. ファイルの配置
手順3 で取得した /etc/opt/FJSVfefs/mngnet_list_<*計算クラスタ名*>ファイルを、システム管理ノード B の待機系ノードとシステム管理ノード A の運用系および待機系ノードの /etc/opt/FJSVfefs/ ディレクトリにも配置してください。

5. FEFS サーバの MDT および OST のリスト取得
システム管理ノード A で以下を実行して、FEFS サーバの MDT および OST のリストを取得してください。

```
[システム管理ノード A]
# pmexe -c <ストレージクラスタ名> --nodetype MDS --stdout "ls /proc/fs/lustre/md[st]/" | ¥
egrep "¥-MDT.{4}$" > /etc/opt/FJSVfefs/reconnect_srv_gmds_mdt_list_<ストレージクラスタ名>
# pmexe -c <ストレージクラスタ名> --nodetype OSS --stdout "ls /proc/fs/lustre/obdfilter/" | ¥
egrep "¥-OST.{4}$" > /etc/opt/FJSVfefs/reconnect_srv_goss_ost_list_<ストレージクラスタ名>
```

※ 出力するファイルのファイル名には、ファイル名の最後に計算クラスタ名またはストレージクラスタ名を入れてください。

6. ファイルの配置
手順5 で取得した以下のファイルを、システム管理ノード A の 待機系ノードとシステム管理ノード B の運用系および待機系システム管理ノードの /etc/opt/FJSVfefs/ ディレクトリにも配置してください。

- /etc/opt/FJSVfefs/reconnect_srv_gmds_mdt_list_<*ストレージクラスタ名*>
- /etc/opt/FJSVfefs/reconnect_srv_goss_ost_list_<*ストレージクラスタ名*>

7. システム監視プラグイン設定ファイルのバックアップ
システム管理ノード A と システム管理ノード B の運用系ノードのシステム監視プラグイン設定ファイル /etc/opt/FJSVtcs/pamoplugin.conf をバックアップしてください。

8. システム監視プラグイン設定ファイルの設定内容の確認

システム監視プラグイン機能の現在の設定内容を /etc/opt/FJSVtcs/pamoplugin.conf ファイルに出力します。

```
[システム管理ノード A およびシステム管理ノード B]
# pamopluginadm --show > /etc/opt/FJSVtcs/pamoplugin.conf
```

9. システム監視プラグイン設定ファイルの編集

a. システム管理ノード B の運用系ノード上の /etc/opt/FJSVtcs/pamoplugin.conf ファイルに以下の設定を追加します。

1. 計算クラスタの定義に、以下を追加します。FEFS クライアントとなっている計算クラスタが複数存在する場合は、FEFS クライアントとなっている計算クラスタ単位に記述してください。

```
PluginCmd {
    NodeType = "<FEFSクライアントとなっているノードタイプ>"
    ServiceName = "OS"
    Status = "x"
```

```
    Cmd = "bash /opt/FJSVfefs/sbin/plugin_evict_client_multiple_execute.sh --ip <システム管理ノードAの代表IPアドレス> -s <ストレージクラスタ名>"
}
PluginCmd {
    NodeType = "<FEFS クライアントとなっているノードタイプ>"
    ServiceName = "OS"
    Status = "-"
    Cmd = "bash /opt/FJSVfefs/sbin/plugin_evict_client_multiple_execute.sh --ip <システム管理ノードAの代表IPアドレス> -s <ストレージクラスタ名>"
}
PluginCmd {
    NodeType = "<FEFS クライアントとなっているノードタイプ>"
    ServiceName = "FEFS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefs/sbin/plugin_evict_client_multiple_execute.sh --ip <システム管理ノードAの代表IPアドレス> -s <ストレージクラスタ名>"
}
```

※FEFS クライアントとなっているノードタイプが複数ある場合は、以下のように、コンマ(,)で区切って指定します。例えば、CCM、LN、および CN がFEFS クライアントとなっている場合は、以下のように指定します。

```
NodeType = "CCM,LN,CN"
```

※--ip オプションで指定する *<システム管理ノードA の代表IPアドレス>*は、システム管理ノード A の代表IPアドレスを指定します。

※-s オプションで指定する *<ストレージクラスタ名>*は、システム管理ノード A 配下のストレージクラスタ名を指定します。

2. 計算クラスタの定義に、以下を追記します。

```
PluginCmd {
    NodeType = "<FEFSクライアントとなっているノードタイプ>"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefs/sbin/plugin_lnet_router_dwn_rmtcmd_execute.sh --ip <システム管理ノードAの代表IPアドレス> -s <ストレージクラスタ名>"
}
```

※--ip オプションで指定する<システム管理ノードA の代表IPアドレス> は、システム管理ノード A の代表 IPアドレスを指定します。

※-s オプションで指定する<ストレージクラスタ名> は、システム管理ノード A 配下のストレージクラスタ名を指定します。

b. システム管理ノード A の運用系ノード上の /etc/opt/FJSVtcs/pamoplugin.conf ファイルのストレージクラスタの定義に、以下を追加します。

```
PluginCmd {
    NodeType = "MDS"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefs/sbin/plugin_reconnect_srv_rmtcmd_execute.sh --ip <システム管理ノードB の代表IPアドレス> -c <計算クラスタ名> --nodetype MDS"
}
PluginCmd {
    NodeType = "OSS"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefs/sbin/plugin_reconnect_srv_rmtcmd_execute.sh --ip <システム管理ノードB の代表IPアドレス> -c <計算クラスタ名> --nodetype OSS"
}
```

※--ip オプションで指定する *<システム管理ノードB の代表IPアドレス>*は、システム管理ノード B の代表 IPアドレスを指定します。

※-c オプションで指定する *<計算クラスタ名>*は、システム管理ノード B 配下の計算クラスタ名を指定します。システム管理

ノード B 配下の計算クラスタが複数存在する場合は、計算クラスタ単位に記述します。
以下は、システム管理ノード B 配下に、計算クラスタ calc1 と、計算クラスタ calc2 が存在する場合の例です。

```
PluginCmd {
    NodeType = "MDS"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefs/sbin/plugin_reconnect_srv_rmtcmd_execute.sh --ip <システム管理ノードB の代表IPアドレス> -c calc1 --nodetype MDS"
}
PluginCmd {
    NodeType = "OSS"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefs/sbin/plugin_reconnect_srv_rmtcmd_execute.sh --ip <システム管理ノードB の代表IPアドレス> -c calc1 --nodetype OSS"
}
PluginCmd {
    NodeType = "MDS"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefs/sbin/plugin_reconnect_srv_rmtcmd_execute.sh --ip <システム管理ノードB の代表IPアドレス> -c calc2 --nodetype MDS"
}
PluginCmd {
    NodeType = "OSS"
    ServiceName = "OS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefs/sbin/plugin_reconnect_srv_rmtcmd_execute.sh --ip <システム管理ノードB の代表IPアドレス> -c calc2 --nodetype OSS"
}
```

10. システム監視プラグイン設定ファイルの登録
    システム管理ノード A とシステム管理ノード B の運用系ノード上で、/etc/opt/FJSVtcs/pamoplugin.conf ファイルの記述内容をジョブ運用ソフトウェアに登録してください。

```
[システム管理ノード A およびシステム管理ノード B]
# pamopluginadm --set
```

11. システム監視プラグイン設定ファイルの設定内容の確認
    システム管理ノード A とシステム管理ノード B の運用系ノード上で、手順10 で登録した /etc/opt/FJSVtcs/pamoplugin.conf ファイルが正しく登録されているか、設定内容を確認してください。

```
[システム管理ノード A およびシステム管理ノード B]
# pamopluginadm --show
```

C.3 ETERNUS を利用する場合に必要な設定

ETERNUS を利用する環境においては、以下の設定が必要です。ETERNUS を利用しない場合は、本設定は不要です。

C.3.1 MDS で ETERNUS の NRDY 対策の有効化

1. 事前準備
   MDS で ETERNUS の NRDY 対策を有効する設定を行う前に、製品に同梱されている「FEFS 環境における ETERNUS マルチパスドライバ個別監視設定手順書」に従って、パスの状態監視間隔の設定を変更しておいてください。

2. MDS の設定の有効化
   MDS で ETERNUS の NRDY 対策を有効にするには、以下の 2通りの方法があります。どちらかの方法で設定してください。

   ー ノード種別単位で設定する方法
   システム管理ノード上で以下を実行して、対象ノードに設定ファイル /etc/opt/FJSVfefs/failover_sleep_file を作成してください。
   以下は、クラスタ clst の MDS に設定する例です。

```
# pmexe -c clst --stdout --nodetype MDS "touch /etc/opt/FJSVfefs/failover_sleep_file"
```

   ー ノード単位で設定する方法
   対象ノードにログインして、設定ファイル /etc/opt/FJSVfefs/failover_sleep_file を作成してください。

```
# touch /etc/opt/FJSVfefs/failover_sleep_file
```

参考

MDS で ETERNUS の NRDY 対策を有効にした場合は、フェイルオーバ時のマウント処理が、7分30秒待ち合わされます。

C.3.2 OSS の自動起動スクリプト設定手順

ETERNUS の CM (コントローラー) が マスタCM、スレーブCM ともに再起動されるなどの影響によって、冗長構成の OSS が両ノードともに停止した場合に自動的に OSS を起動させるスクリプト plugin_fefs_autoboot_pairnode.sh を設定してください。

以下に、設定手順を示します。

1. 事前準備
   本スクリプトの設定を行う前に、OSS のダンプ採取後の動作について設定しておく必要があります。
   すべての OSS のダンプ採取後の動作は、ノードの電源が OFF になるように設定してください。
   設定方法は、富士通Linuxサポートパッケージのダンプ支援ツールを導入している場合は、そのドキュメントを参照してください。
   導入していない場合は、Red Hat 社が公開している「カーネルクラッシュダンプガイド」を参照してください。

参考

ダンプ採取が行われたあとに、システム管理ノード上で pashowclst コマンドの -v オプションを使用してノードの状態を確認した場合、以下のように表示されます。

```
STATUS      REASON          PWR_STATUS
Stopped     -               off
```

2. スクリプト内のパラメーターの変更
   スクリプト plugin_fefs_autoboot_pairnode.sh 内のパラメーターを運用にあわせて変更してください。スクリプトは、/opt/FJSVfefs/sbin にあるものをコピーして使ってください。パラメーターには以下があります。

   ー ダンプ採取完了待ち合わせ設定

   ー ノードの起動抑止設定

   ー ノードの起動後のノード状態確認設定

   スクリプト適用後は、スクリプトを変更するとその内容は即時反映されます。スクリプトの適用方法は、手順 "3. スクリプトの適用" を参照してください。

   以下は各パラメーターの設定方法です。

   [ダンプ採取完了待ち合わせ設定]

   ダンプ採取が完了後、電源が off になるのを待ち合わせる時間を設定します。
   以下は、設定例です。

```
DUMP_CHK_TIME=40           # 40min             ダンプ採取完了待ち合わせ時間
DUMP_CHK_INTERVAL=5        # 5min              ダンプ採取完了チェック間隔
DUMP_CHK_FORCE_RESET="Yes" # Yes(default)|No   待ち合わせ時間内に電源 off が検出されないときの動作
```

DUMP_CHK_TIME には、ダンプ採取完了待ち合わせ時間を設定します。このデフォルト値は 40 分です。DUMP_CHK_INTERVALには、ダンプ採取完了チェック間隔を設定します。このデフォルト値は 5分です。これらは分単位で指定します。DUMP_CHK_FORCE_RESETには、待ち合わせ時間内に電源 off が検出されない場合 (タイムアウト時) の動作を設定します。デフォルトの動作は、リセットによる強制起動 (設定値は Yes) です。この設定値を "No" とした場合は、なにもせず処理を終了します。

[ノードの起動抑止設定]

スクリプトによる起動履歴を保持し照合することで、起動後にハードウェア故障などによるノードのパニックと再起動が繰り返されないように抑止する設定をします。

```
BOOT_RESERVE=24     # 24hour   直近の動作履歴をさかのぼって参照する範囲
BOOTED_COUNT=1      # 1 回     何回起動された場合にノードの起動を抑止するか
```

デフォルトの設定は、24時間に1回です。
BOOT_RESERVE には、直近の起動履歴をさかのぼって参照する範囲を設定します。この値は、時間単位で設定します。デフォルト値は、24時間 (24hour) です。BOOTED_COUNT には、BOOT_RESERVE で設定した範囲内で何回起動された場合にノードの起動を抑止するのかを設定します。この値は、回数を指定します。この設定値を "0" とした場合は、常に起動が抑止されます。

[ノードの起動後のノード状態確認設定]

ノードの起動後に、ノードの状態が Running になるのを待ち合わせる設定をします。

```
BOOTSTATUS_CHK_TIME=40     # 40min  ノード状態の確認時間
BOOTSTATUS_CHK_INTERVAL=5 # 5min   ノード状態のチェック間隔
```

BOOTSTATUS_CHK_TIME には、ノード状態を確認する時間を設定します。このデフォルト値は 40 分です。BOOTSTATUS_CHK_INTERVAL には、ノード状態をチェックする間隔を設定します。このデフォルト値は 5分です。これらの設定値は、分単位で指定します。

設定した確認時間内にノードの状態が Running に遷移しない場合は起動異常とみなし、システム管理ノードのシステムログに、以下のメッセージを出力します。

```
plugin_fefs_autoboot_pairnode.sh: [ERR.] STATUS Running was not able to be detected. (clstname:clstname NID:
0xXXXXXXXX PAIR_NID:0xXXXXXXXX)」
```

3. スクリプトの適用

a. スクリプトの配置
運用系と待機系のシステム管理ノード上に、スクリプトplugin_fefs_autoboot_pairnode.shを、ディレクトリ/opt/FJSVfefs/に配置してください。

b. システム監視プラグイン設定ファイルのバックアップ
運用系システム管理ノード上の、システム監視プラグイン設定ファイル /etc/opt/FJSVtcs/pamoplugin.conf をバックアップしてください。

c. システム監視プラグイン設定ファイルの設定内容の確認
システム監視プラグイン機能の現在の設定内容を /etc/opt/FJSVtcs/pamoplugin.conf ファイルに出力します。

```
# pamopluginadm --show > /etc/opt/FJSVtcs/pamoplugin.conf
```

d. システム監視プラグイン設定ファイルの編集
運用系システム管理ノード上の /etc/opt/FJSVtcs/pamoplugin.conf ファイルに、以下のストレージクラスタの定義を追加してください。

```
PluginCmd {
    ServiceName = "OS"
    NodeType = "OSS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefs/plugin_fefs_autoboot_pairnode.sh"
}
```

e. システム監視プラグイン設定ファイルの登録
/etc/opt/FJSVtcs/pamoplugin.conf ファイルの記述内容をジョブ運用ソフトウェアに登録してください。

```
[運用系システム管理ノード]
# pamopluginadm --set
```

f. システム監視プラグイン設定ファイルの設定内容の確認
手順 e で登録した /etc/opt/FJSVtcs/pamoplugin.conf ファイルが正しく登録されているか、設定内容を確認してください。

```
[運用系システム管理ノード]
# pamopluginadm --show
```

参考

スクリプトを削除する場合は、以下の手順で行ってください。

a. システム監視プラグイン設定ファイルのバックアップ
スクリプトの適用手順 b と同じです。

b. システム監視プラグイン設定ファイルの設定内容の確認
スクリプトの適用手順 c と同じです。

c. システム監視プラグイン設定ファイルの編集
運用系システム管理ノード上の /etc/opt/FJSVtcs/pamoplugin.conf ファイルに、以下のストレージクラスタの定義を削除してください。

```
PluginCmd {
    ServiceName = "OS"
    NodeType = "OSS"
    Status = "x"
    Cmd = "bash /opt/FJSVfefs/plugin_fefs_autoboot_pairnode.sh"
}
```

d. システム監視プラグイン設定ファイルの登録
スクリプトの適用手順 e と同じです。

e. システム監視プラグイン設定ファイルの設定内容の確認
スクリプトの適用手順 f と同じです。

```
[運用系システム管理ノード]
# pamopluginadm --show
```

f. スクリプトの削除
運用系と待機系のシステム管理ノードの /opt/FJSVfefs/ ディレクトリから、スクリプト plugin_fefs_autoboot_pairnode.sh を削除してください。

冗長構成の OSS の両ノードが停止したあと、ノードの起動に失敗した場合の対処方法

ETERNUS の CM の再起動などにより冗長構成の OSS へのパスが閉塞状態になった場合、ダンプ採取が完了した後に、OSSの両ノードの電源が停止 (OFF) されます。
OSS が両ノードともに停止されると、OSS の自動起動スクリプトによって停止したノードに電源が投入されますが、一方の OSS の起動が失敗するなど正常に起動できなかった場合は多重故障になるため、運用が継続されません。このような場合は、以下の対処を行ってください。

[検出方法]

1. システムログメッセージの確認
電源投入後、一方のノードが正常に起動されずノード状態が Running にならなかった場合、システム管理ノードのシステムログ (/var/log/messages) に以下のメッセージが出力されます。

```
plugin_fefs_autoboot_pairnode.sh: [ERR.] STATUS Running was not able to be detected. (clstname:clstname NID:0xXXXXXXXX PAIR_NID:0xXXXXXXXX)
```

clstname: 対象クラスタ名
NID: 正常に起動されなかったノードのノードID
PAIR_NID: 正常に起動されなかったノードの冗長構成のペアとなるノードのノードID

注意

両ノードからメッセージが出力されていた場合は、両ノードともに正常に起動しないような状態であるため、本手順では対処できません。このような場合は、OSS の両ノードの電源状態などを確認して、故障要因を取り除いたうえで再起動してください。

2. ノード状態の確認
メッセージが出力された対象ノードの STATUS 欄および PWR_STATUS 欄を確認してください。また、対象ノードのペアとなるノードの STATUS 欄が Running になっていることも確認してください。
対象クラスタ名や対象ノードおよび対象ノードのペアとなるノードの ノードID は手順1 で出力されたメッセージで確認できます。

例1: ノードを起動したあと、FEFS サービスの状態が "o" にならなかった場合

```
[システム管理ノード]
# 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)
```

例2: ノードの電源を投入しても停止したまま起動しなかった場合

```
[システム管理ノード]
# 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)
```

[対処方法]

検出方法の手順1 と 2 ともに該当する場合は、正常に起動しなかったノードを停止して、ペアとなるノードにFEFS サービスを片寄せして運用を継続させます。

1. 対象ノードの停止
正常に起動しなかったノードの STATUS 欄が Running にならず、PWR_STATUS 欄が off ではない場合は init を発行して対象ノードを停止させてください。PWR_STATUS 欄が off になっている場合は、ノードの電源は投入されていないため、本手順は不要です。

```
[システム管理ノード]
# papwrctl -c 〈ストレージクラスタ名〉 -n 〈対象ノードのノードID〉 init
```

2. FEFS サービスの片寄せ
以下のコマンドを実行して、対象ノードからペアとなるノードに FEFS サービスを片寄せします。

```
[システム管理ノード]
# pmexe -c 〈ストレージクラスタ名〉 -n 〈対象ノードのペアとなるノードのノードID〉 /bin/fefs_failover --active -f
```

なお、本コマンドは、この手順以外では使用しないでください。

3. 片寄せされた FEFS サービスの復旧
   対象ノードの保守後、FEFS サービスを片寄せされた状態から復旧させて、元の運用状態に戻してください。

```
[システム管理ノード]
# paclstmgr -c <ストレージクラスタ名> --failback <対象ノードのノードID>
```

C.4 FEFS ログの定期削除の設定

FEFS サーバの /var/opt/FJSVfefs/dumplog 配下の古いログファイルを定期的に削除する設定手順を示します。

本手順は、特に断りがない限り、運用系システム管理ノード上で行ってください。

1. FEFS サーバの crond の確認
   FEFS サーバの crond が enabled になっていることを確認してください。

```
# pmexe -c <ストレージクラスタ名> --nodetype <FEFSサーバとなっているノードタイプ> ¥
--stdout "systemctl is-enabled crond.service"
```

   FEFS サーバの crond が enabled になっていない場合は、以下を実行して、crond を enabled にしてください。

```
# pmexe -c <ストレージクラスタ名> --nodetype <FEFSサーバとなっているノードタイプ> ¥
--stdout "systemctl enable crond.service"
```

2. FEFS サーバの crond の状態確認
   FEFS サーバの crond の状態が起動中になっていることを確認してください。

```
# pmexe -c <ストレージクラスタ名> --nodetype <FEFSサーバとなっているノードタイプ> ¥
--stdout "systemctl status crond.service"
```

   FEFS サーバの crond の状態が起動中になっていない場合は、以下を実行して、crond を起動してください。

```
# pmexe -c <ストレージクラスタ名> --nodetype <FEFSサーバとなっているノードタイプ> ¥
--stdout "systemctl start crond.service"
```

3. FEFS サーバの crontab の設定
   2 か月以上前に作成された古いログファイルを定期的に削除するように、FEFS サーバの cron を設定してください。

```
# pmexe -c <ストレージクラスタ名> --nodetype <FEFSサーバとなっているノードタイプ> ¥
--stdout "echo ¥"0 0 * * * /bin/find /var/opt/FJSVfefs/dumplog/ -type f -mtime +60 -delete¥" ¥
>> /var/spool/cron/root"
```

4. crontab の設定の確認
   FEFS サーバの cron が正しく設定されていることを確認してください。

```
# pmexe -c <ストレージクラスタ名> --nodetype <FEFSサーバとなっているノードタイプ> --stdout "crontab -l"
```

   以下が設定されていることを確認します。

```
0 0 * * * /bin/find /var/opt/FJSVfefs/dumplog/ -type f -mtime +60 -delete
```

付録D　ファイルシステムの復旧手順

本章では、ファイルシステムに障害が発生した場合の復旧手順を示します。

D.1　はじめに

本章では、以下の障害を想定しています。

- 不良ブロック
  デバイス上に使用できない LBA (Logical Block Addressing) アドレスを検出した状態を指します。
- ディスク故障
  ハードディスクの物理的な故障により、ファイルサーバにアクセスできなくなっている状態を指します。
- ファイルシステム破壊
  ハードウェア以外の要因で、ファイルサーバにマウントできないか、またはファイルI/Oを正常に行えなくなっている状態を指します。EIO が返されます。
- 両系サーバ停止
  運用系および待機系のサーバノード (MGS、MDS、またはOSS) のどれもがダウンしている状態を指します。

D.2　影響

MGS

MGSの両系ダウンや、MGTのディスク故障、ファイルシステム破壊があった場合、再マウントができなくなります。サーバの再起動後やクライアントの再起動などを実施した際にFEFSの起動が失敗します。

MDS

以下の場合、ファイルシステムの運用継続はできません。

- MDT0 (FEFS デザインシート上のインデックス 0番) にマウントしている MDS の両系ダウン
- MDT0に割り当てているディスクの故障

また、マルチMDS機能を使用していると、以下の場合に当該ディスクに保管されているファイル、ディレクトリへのアクセスができなくなります。

- MDT0 以外のMDT をマウントしているMDS の両系ダウン
- MDT0 以外のディスク故障

OSS

OSSの両系ダウンやOSTのディスク故障、ファイルシステム破壊があった場合、当該OST上に保管されているファイルへのファイル I/Oがハングまたはエラー復帰します。

D.3　障害復旧フロー

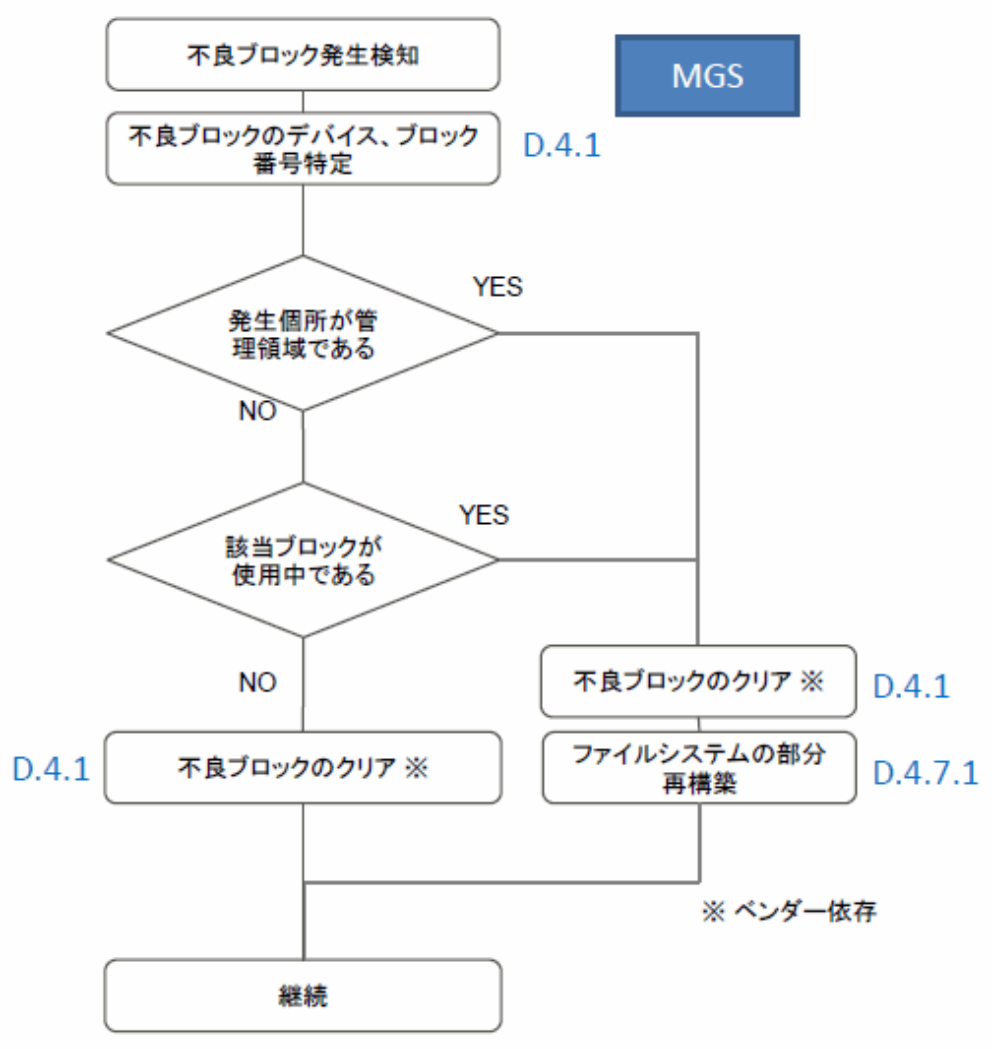
本節では MGS、MDS、OSS で障害が発生した場合のフローを示します。

次節で各フローのオペレーションを示します。"D.4 対応手順" で詳細を説明している節の番号を、図中に青字で示します。

D.3.1　不良ブロック検出時の復旧フロー

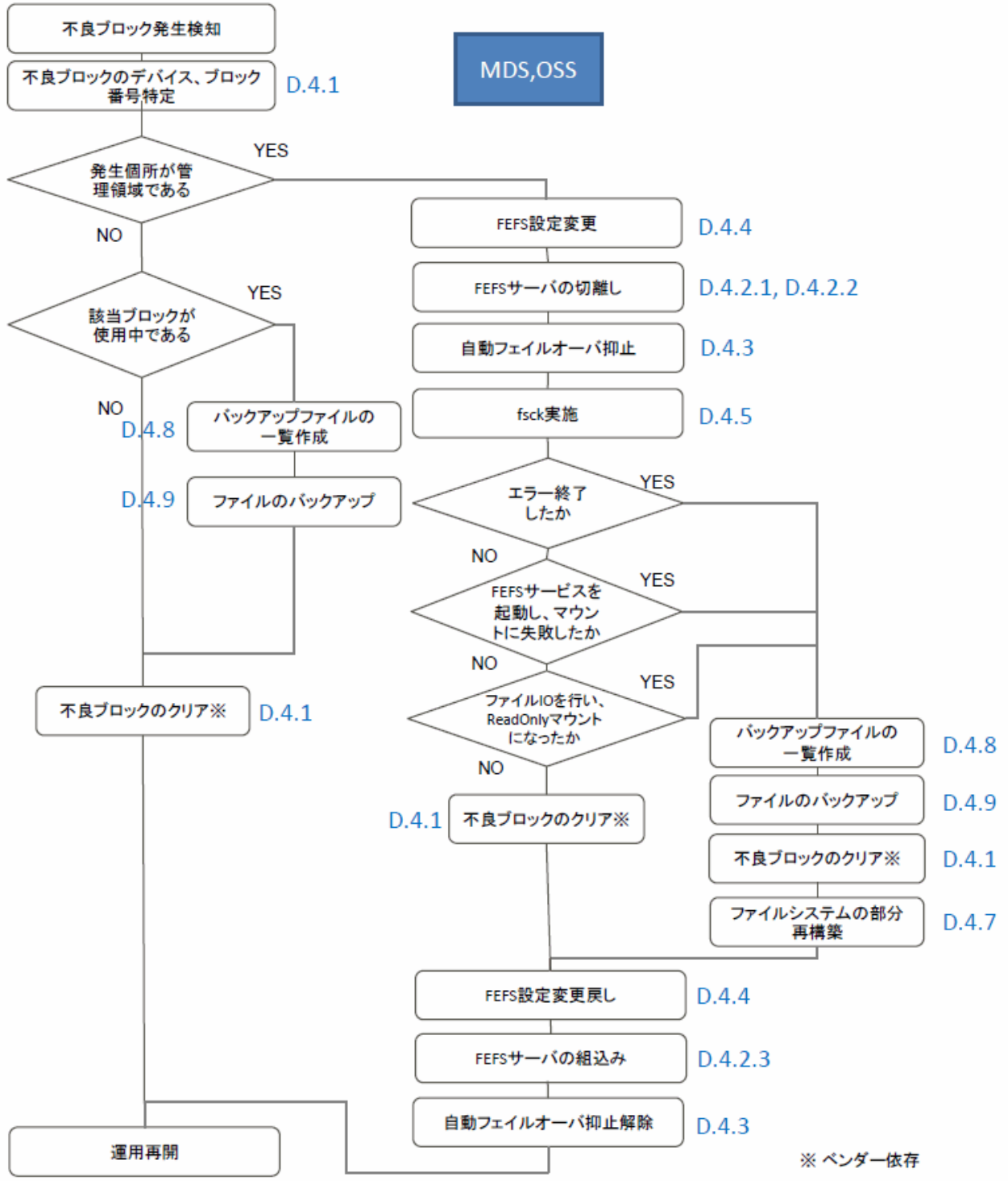
不良ブロック を検出したときのMGT復旧フローを以下に示します。

図D.1 不良ブロック検出時のMGT復旧フロー



不良ブロックを検出したときの MDT および OST の復旧フローを以下に示します。

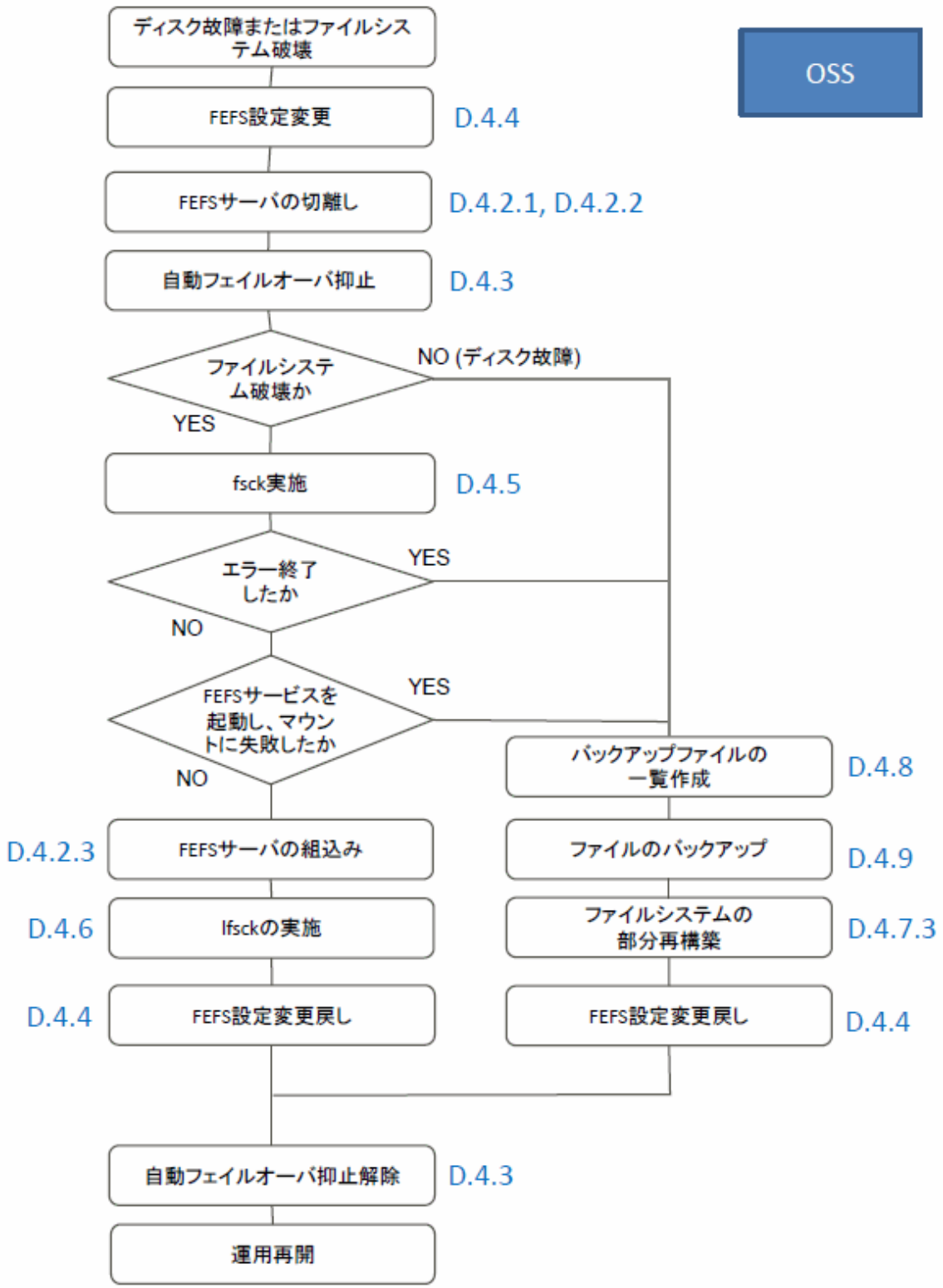
図D.2 不良ブロック検出時のMDT/OST復旧フロー



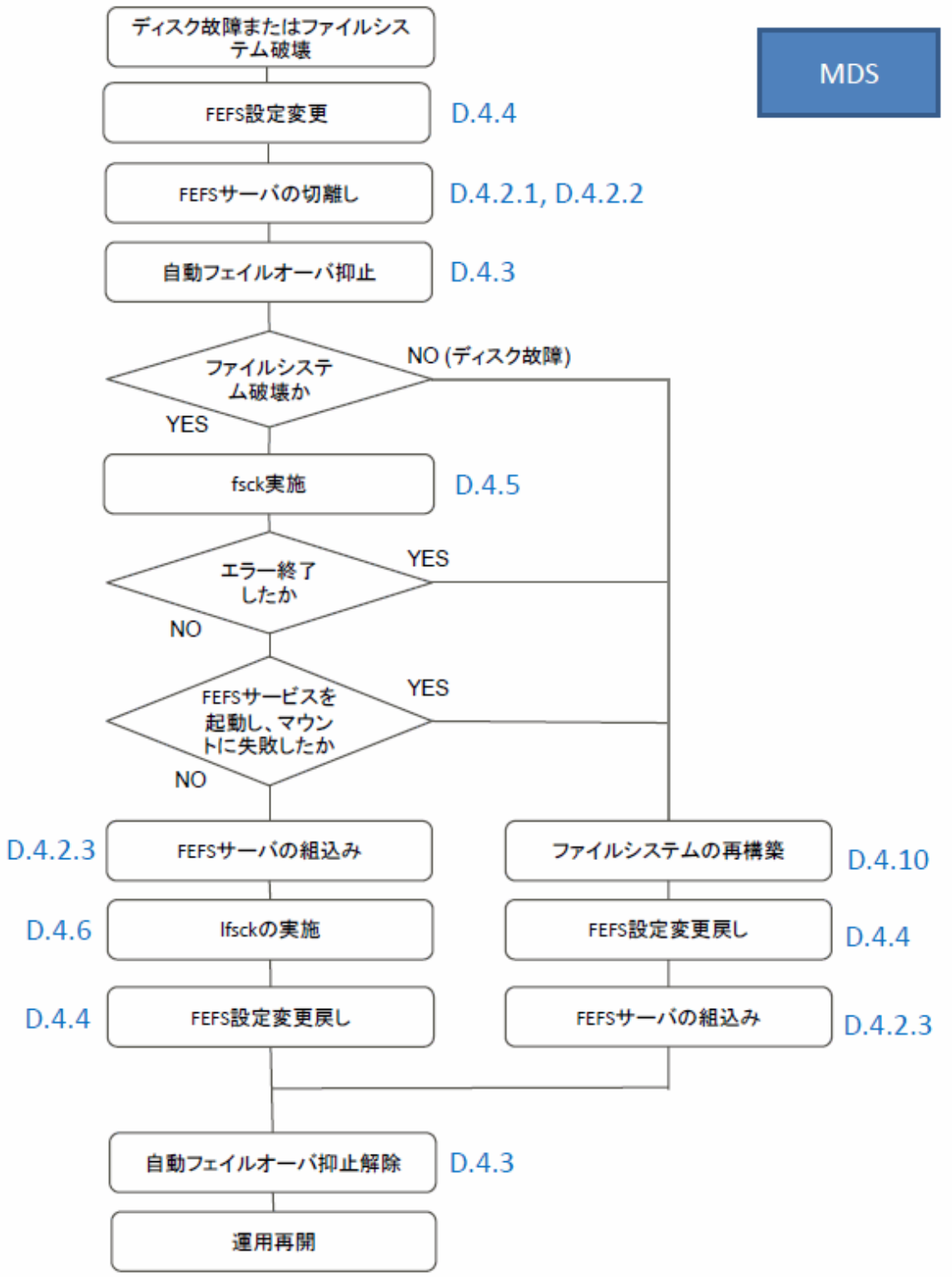
D.3.2 ディスク故障時またはファイルシステム破壊時の復旧フロー

ディスク故障時またはファイルシステム破壊時の障害対応フローを以下に示します。

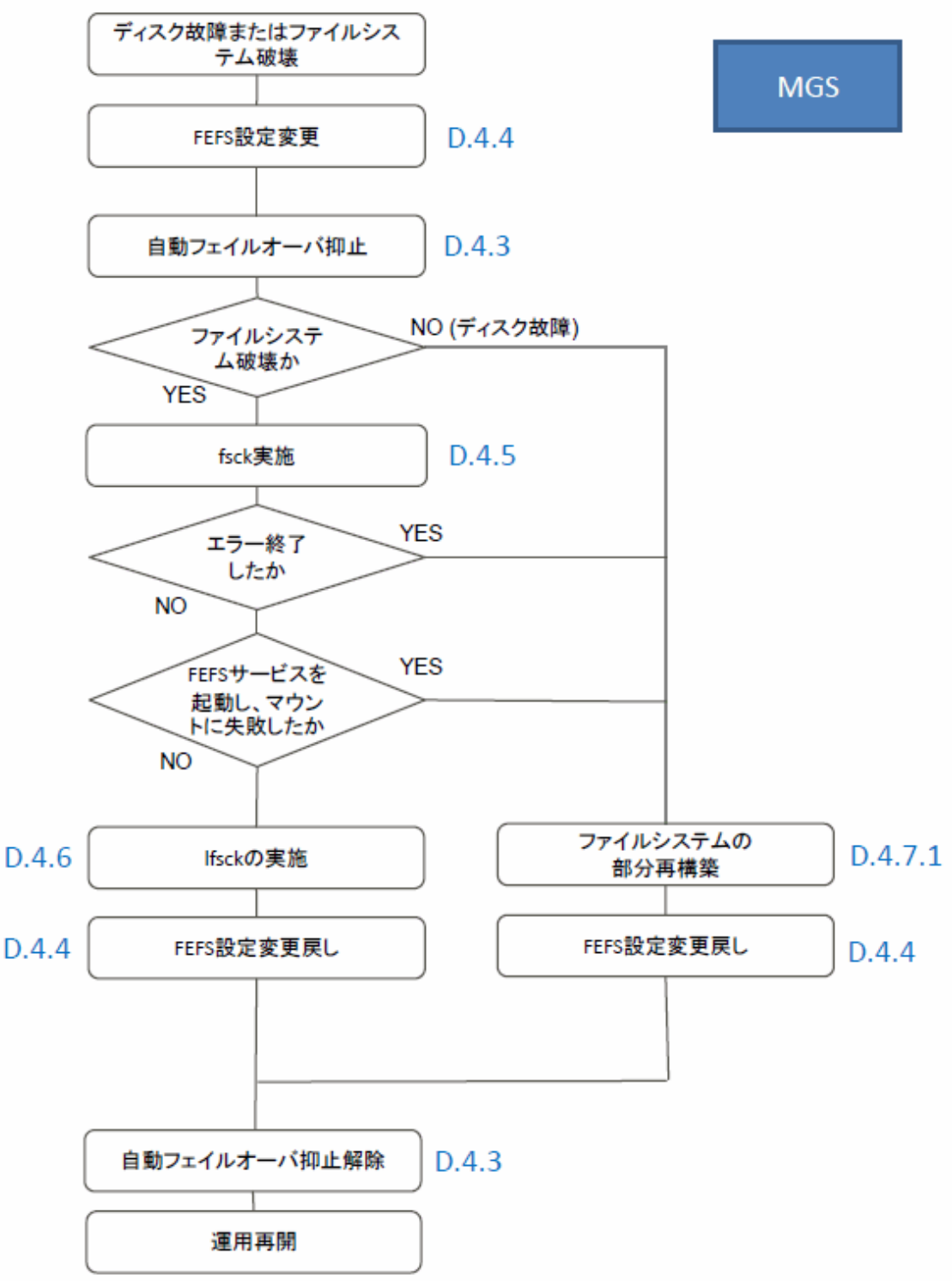
図D.3 ディスク故障時またはファイルシステム破壊時のOST復旧フロー



図D.4 ディスク故障時またはファイルシステム破壊時のMDT復旧フロー (MGT 兼 MDT の場合を含む)



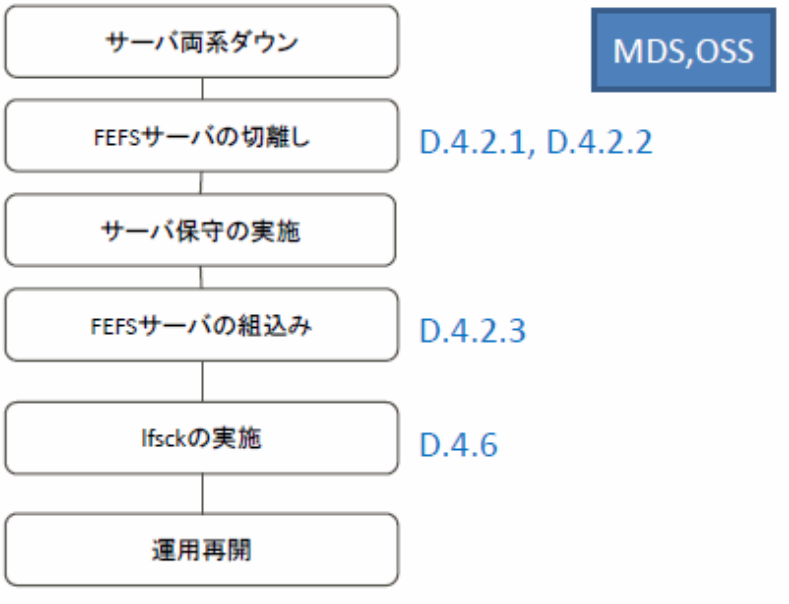
図D.5 ディスク故障時またはファイルシステム破壊時のMGT復旧フロー



D.3.3 両系停止時の障害対応フロー

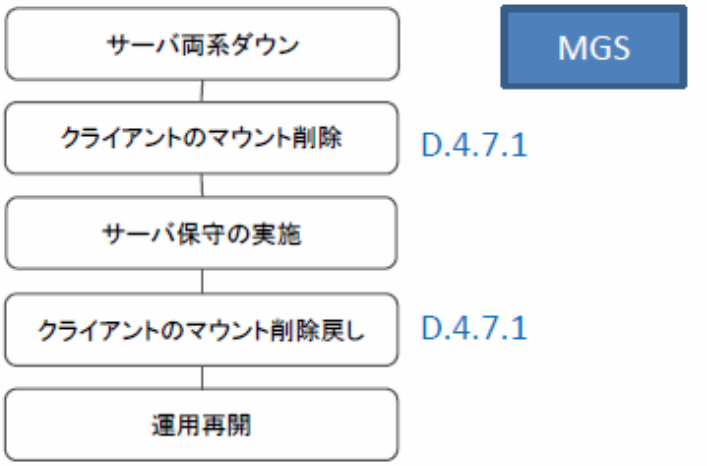
MDS, OSS 両系停止時の障害対応フローを以下に示します。

図D.6 MDS/OSS 両系停止時の障害対応フロー (MGS 兼 MDS の場合を含む)



MGS 両系停止時の障害対応フローを以下に示します。

図D.7 MGS 両系停止時の障害対応フロー



D.4 対応手順

D.4.1 不良ブロックが発生したブロック番号の状態確認

不良ブロックが発生した場合、ディスク装置上で発生したデバイスとブロック番号を特定する必要があります。

特定方法は各ディスク装置により異なるため、ディスク装置のベンダーに確認してください。

不良ブロックが発生したデバイスを使用しているサーバ上で以下の手順で debugfs.ldiskfs コマンドを使用して状態を確認してください。

icheck コマンドで Inode number に値が出力された場合はファイルデータ破損が発生しています。

testb コマンドで "marked in use" が出力された場合は管理領域破壊が発生しています。

```
# sync
# /opt/FJSVfefsprogs/sbin/debugfs.ldiskfs <device> -R "icheck <block num>"
```

<*device*>:不良ブロック発生デバイス
<*block num*>:不良ブロック発生ブロック番号

上記コマンドを実行し、"block not found" が表示された場合は、さらに以下のコマンドを実行してください。

"block not found" が表示された場合は、ファイルシステムの一部が破損した状態を示します。

```
# /opt/FJSVfefsprogs/sbin/debugfs.ldiskfs <device> -R "testb <block num>"
```

上記コマンドを実行し、"marked in use" と表示された場合は指定したブロック番号は使用中を示します。ここでブロック番号が使用中となった場合、ファイルシステムの管理領域が破損したことを示します。

"not in use" と表示された場合は指定したブロック番号が未使用を示します。

以上を表で示します。

表D.1 コマンドの実行結果とファイルシステムの状態

サブコマンド	結果	状態
icheck	"inode number" が表示	ファイルデータ破損
	"block not found" が表示	ファイルシステムの一部が破損 → testb コマンド
testb	"marked in use" が表示	該当ブロック番号使用中、管理領域破損
	"not in use" が表示	該当ブロック番号未使用

実行例

1. ファイルデータ破損の場合 (ブロック番号1544 をチェック)

```
# sync
# /opt/FJSVfefsprogs/sbin/debugfs.ldiskfs /dev/sdb -R "icheck 1544"
Block    Inode number
1544     7
```

2. 管理領域破損の場合 (ブロック番号1545 をチェック)

```
# sync
# /opt/FJSVfefsprogs/sbin/debugfs.ldiskfs /dev/sdb -R "icheck 1545"
Block    Inode number
1545     <block not found>
# /opt/FJSVfefsprogs/sbin/debugfs.ldiskfs /dev/sdb -R "testb 1545"
Block 1545 marked in use
```

当該ブロックが使用されていない場合は、以下のように表示されます。

```
# /opt/FJSVfefsprogs/sbin/debugfs.ldiskfs /dev/sda7 -R "testb 10000"
debugfs 2.5.1 (01-Dec-2018)
Block 10000 not in use
```

注意

不良ブロックによる破損箇所により、復旧方法が異なります。フローに従い復旧してください。

不良ブロック発生箇所の対象LBAアドレスがデータ領域か管理領域か、また対象LBAアドレスが使用か未使用かにより以下の対応が変わります。

- 対象LBAアドレスがデータ領域かつ未使用
  不良ブロックのフラグクリアを実施してください。フラグクリアの方法は各ディスク装置により異なるため、ディスク装置のベンダーに確認してください。
- 対象LBAアドレスがデータ領域かつ使用
  ファイルに影響があります。ファイルの特定を実施してください

・ 対象LBAアドレスが管理領域かつ未使用
不良ブロックのフラグクリアを実施してください。フラグクリアの方法は各ディスク装置により異なるため、ディスク装置のベンダーに確認してください。

・ 対象LBAアドレスが管理領域かつ使用
ファイルシステムの運用に影響があります。MGT、MDT、OSTの復旧フローを確認してください。

D.4.2 FEFS サーバの切離し/組込み

MDS、OSS のフェイルオーバペアで両系ダウンやディスク故障が発生した場合、復旧されるまでメタデータアクセスやファイルI/Oがハングします。ハングを解消する場合は以下の手順で FEFS サーバの切離しを行ってください。

D.4.2.1 対象の確認

切離しを行う対象を確認します。

ファイルシステム破壊、ディスク故障の場合

不良ブロックによる管理領域破損やディスク故障の場合は、FEFS デザインシートから、対象のディスクのデバイス番号、ファイルシステム名、ノード種別を確認します。フォーマットは以下のとおりです。

```
<filesystem>-[MDT|OST] <index>
```

<*filesystem*>: ファイルシステム名
<*index*>: デバイス番号 (ゼロ埋め 4桁 16進数)

ー 例: デバイス番号が 0001、ファイルシステム名が fefs01、ノード種別が MDS の場合

```
fefs01-MDT0001
```

サーバ両系ダウンの場合

以下の手順でターゲット名を取得します。

```
[クライアントノード]
# lctl device_list -t | grep <IP addr> | awk '{print $4}' | cut -f 1,2 --delim="-"
```

<*IP addr*>:ダウンしたサーバのI/Oネットワーク用のIPアドレス、フェイルオーバペアのIPアドレスをそれぞれ指定してください。

・ 実行例

```
[クライアントノード]
# lctl device_list -t | grep 172.31.211.60 | awk '{print $4}' | cut -f 1,2 --delim="-"
fefs01-OST0000
# lctl device_list -t | grep 172.31.211.61 | awk '{print $4}' | cut -f 1,2 --delim="-"
fefs01-OST0001
```

D.4.2.2 切離し

1. 切離し
以下の手順で対象のターゲットをファイルシステムから切り離します。

```
[システム管理ノード]
# pmexe -c <cluster> --stdout --nodetype SIO "/usr/sbin/force_intr -c -m deactivate <target>..."
# pmexe -c <cluster> --stdout --nodetype CCM,LN "/usr/sbin/force_intr -c -m deactivate <target>..."
# pmexe -c <storage-cluster> --stdout --nodetype MDS,OSS "touch /var/opt/FJSVfefs/stop_activate_device"
# pmexe -c <storage-cluster> --stdout --nodetype MDS "/usr/sbin/force_intr -s -m deactivate <target>..."
```

<*cluster*>:クラスタ名
<*storage-cluster*>:ストレージクラスタ名
<*target*>:対象のターゲット

ー 実行例

```
[システム管理ノード]
# 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/FJSVfefs/stop_activate_device"
# pmexe -c storage --stdout --nodetype MDS "/usr/sbin/force_intr -s -m deactivate fefs-OST0000 fefs-OST0001"
```

注意

ー インデックス番号が 0 の MDT(MDT0000) を deactivate した場合は、マウントができなくなります。そのため、FEFS の状態が FEFS(o) になりません。

ー インデックス番号が 0 以外のMDTを deactivate した場合、df コマンドでマウント状態を確認できません。lfs df コマンドで確認してください。

ー システムに多目的ノードが含まれる場合は、FEFS デザインシートで定義した NODETYPE を--nodetype オプションの引数に追加してください。NODETYPE について詳細は "3.1.3.1 NODE シートの入力" の「多目的ノードの NODETYPE について」を参照してください。

2. 状態確認
   切離し後に以下の手順で状態を確認します。

```
[システム管理ノード]
# 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*>:クラスタ名
<*storage-cluster*>:ストレージクラスタ名
<*target*>:対象のターゲット

すべてのターゲットに対して状態が "IN" であることを確認します。

ー 実行例

```
[システム管理ノード]
# pmexe -c compute --stdout --nodetype SIO "/usr/sbin/force_intr -c -m status fefs-OST0000 fefs-OST0001"
# pmexe -c compute --stdout --nodetype CCM,LN "/usr/sbin/force_intr -c -m status fefs-OST0000 fefs-OST0001"
# pmexe -c storage --stdout --nodetype MDS "/usr/sbin/force_intr -s -m status fefs-OST0000 fefs-OST0001"
```

出力について詳細は "A.2.15 force_intr コマンド" を参照してください。

D.4.2.3 組込み

1. 組込み前の状態確認
   組込み前に、組み込める状態になっているか以下の手順で確認します。

   MDT を組み込む場合

```
[システム管理ノード]
# pmexe -c <storage-cluster> --stdout --nodetype MDS "lctl get_param mdt.<target>.recovery_status 2> /dev/null |
grep status:"
```

<*storage-cluster*>:ストレージクラスタ名
<*target*>:対象のターゲット

status が COMPLETE または、INACTIVE になるまで待ちます。

ー 実行例

```
[システム管理ノード]
# 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]
```

OST を組み込む場合

```
[システム管理ノード]
# pmexe -c <storage-cluster> --stdout --nodetype OSS "lctl get_param obdfilter.<target>.recovery_status 2> /dev/
null | grep status:"
```

<*storage-cluster*>:ストレージクラスタ名
<*target*>:対象のターゲット

status が COMPLETE または、INACTIVE になるまで待ちます。

ー 実行例

```
[システム管理ノード]
# 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. 組込み
以下の手順で対象のターゲットをファイルシステムに組み込みます。

```
[システム管理ノード]
# 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/FJSVfefs/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*>:クラスタ名
<*storage-cluster*>:ストレージクラスタ名
<*target*>:対象のターゲット

ー 実行例

```
[システム管理ノード]
# pmexe -c storage --stdout --nodetype MDS "/usr/sbin/force_intr -s -m activate fefs-OST0000 fefs-OST0001"
# pmexe -c storage --stdout --nodetype MDS,OSS "rm -f /var/opt/FJSVfefs/stop_activate_device"
# pmexe -c compute --stdout --nodetype SIO "/usr/sbin/force_intr -c -m activate fefs-OST0000 fefs-OST0001"
# pmexe -c compute --stdout --nodetype CCM,LN "/usr/sbin/force_intr -c -m activate fefs-OST0000 fefs-OST0001"
```

注意

システムに多目的ノードが含まれる場合は、FEFS デザインシートで定義した NODETYPE を --nodetype オプションの引数に追加してください。NODETYPE について詳細は "3.1.3.1 NODE シートの入力" の「多目的ノードの NODETYPE について」を参照してください。

3. 組込み後の状態確認
   組込み後に以下の手順で状態を確認します。

```
# 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*>:クラスタ名
   <*storage-cluster*>:ストレージクラスタ名
   <*target*>:対象のターゲット

   すべてのターゲットに対して状態が "UP" であることを確認します。

   ー 実行例

```
# pmexe -c compute --stdout --nodetype SIO "/usr/sbin/force_intr -c -m status fefs-OST0000 fefs-OST0001"
# pmexe -c compute --stdout --nodetype CCM,LN "/usr/sbin/force_intr -c -m status fefs-OST0000 fefs-OST0001"
# pmexe -c storage --stdout --nodetype MDS "/usr/sbin/force_intr -s -m status fefs-OST0000 fefs-OST0001"
```

   出力について詳細は "A.2.15 force_intr コマンド" を参照してください。

   組込み完了後に全クライアントノードで lfs df コマンドを実施してください。

```
[システム管理ノード]
# pmexe -c <cluster> --stdout --nodetype SIO "/usr/bin/lfs df > /dev/null"
# pmexe -c <cluster> --stdout --nodetype LN,CCM "lfs df > /dev/null"
```

   <*cluster*>:クラスタ名

D.4.3 自動フェイルオーバ抑止

「ジョブ運用ソフトウェア 管理者向けガイド システム管理編」を参照してください。

D.4.4 FEFS 設定変更

MGT、MDT、OST のディスク異常を検出すると FEFSサーバはパニックします。

パニックを回避するにはディスク異常が発生したデバイスをマウントする全サーバでReadOnlyマウントとなるよう以下に示す方法で設定変更し、FEFSサービスを再起動してください。

注意

- 起動後すぐにパニックする場合は、シングルユーザモードで起動してください。
- サービス再起動時に一時的にファイルアクセスがハングする場合があります。

設定変更

```
[ディスク異常を検出したFEFSサーバとそのフェイルオーバのペアとなるサーバ]
# grep -l 'remount_ro' /etc/opt/FJSVfefs/modprobe.conf | xargs sed -i.back -e 's/remount_ro=1/ remount_ro=0/g'
```

上記を実施することで設定内容が変更され、/etc/opt/FJSVfefs/modprobe.conf.back に元のファイルが退避されます。
FEFSサービスの再起動により有効化されます。

設定戻し

```
[ディスク異常を検出したFEFSサーバとそのフェイルオーバのペアとなるサーバ]
# mv /etc/opt/FJSVfefs/modprobe.conf.back /etc/opt/FJSVfefs/modprobe.conf
```

設定を戻した場合もサービス再起動が必要です。

ファイルI/Oなどを行い、ディスク異常を検出した場合、サーバ上で当該ディスクはReadOnlyに変更となり、クライアントのlfs dfコマンドの各エントリの最後に "R" 付きで表示されます。また、サーバ上で /var/log/messages に ReadOnly 状態で再マウントされたことを示すメッセージが出力されます。

```
[クライアントノード]
# lfs df
UUID                   1K-blocks        Used   Available Use% Mounted on
fefs-MDT0000_UUID       5424664       40700     4897304   1% /fefs[MDT:0]
fefs-MDT0001_UUID       5424664       40324     4897680   1% /fefs[MDT:1]
fefs-OST0000_UUID     144009476       62256   136598996   0% /fefs[OST:0]
fefs-OST0001_UUID     144009476       62256   136598996   0% /fefs[OST:1] R
filesystem_summary:     288018952      124512   273197992   0% /fefs

[ディスク異常を検出したFEFSサーバ]
#less /var/log/messages
Jul 20 13:15:32 oss1 kernel: LDISKFS-fs (sdb): Remounting filesystem read-only
```

Read-Only マウントされた場合は当該 OST への書込みがエラーになります。

ファイルアクセスの影響の詳細については "D.5 アクセス影響" を参照してください。

D.4.5 fsckの実施

"4.8.1 FEFS のサービス停止" から "4.8.5 FEFS の修復" の "3) OST のマウント" までの手順を参照して、fsck.ldiskfs コマンドによる修復を実施してください。

エラーがあった場合、復旧不可です。"D.4.10 ファイルシステムの再構築" を実施してください。

D.4.6 lfsckの実施

lctl lfsck_start コマンドによる修復の手順は、"4.8.5 FEFS の修復" を参照してください。

エラーがあった場合、復旧不可です。"D.4.10 ファイルシステムの再構築"を実施してください。

注意

強制 I/O 中断が解除されていることを確認してください。

手順は、"D.4.2.3 組込み" の「3. 組込み後の状態確認」を参照してください。

D.4.7 ファイルシステムの部分再構築

部分再構築を実施する際は構築に使用した FEFS デザインシートを準備してください。

D.4.7.1 MGTの再構築

MGTが故障した場合、クライアント、MDT、OSTのマウントが正常に行われません。

以下に示す手順で復旧を行ってください。

ファイルシステムの停止

```
［システム管理ノード］
# fefs_sync --stop --storage=<cluster> --compute=<cluster>
```

--storage : ストレージクラスタ名を指定してください。
--compute : 計算クラスタ名および多目的クラスタ名を指定してください。

注意

複数のFEFSをマウントしている場合は、再構築を行うMGTの管理するファイルシステムを個別にアンマウントしてください。

FEFS クライアントのマウント設定の削除

複数ファイルシステムの環境において、MGTの再構築中にFEFSクライアントの再起動などでサービスの起動をした場合、当該FEFSのマウントに失敗します。この場合、FEFS の状態が FEFS(o) に状態遷移しません。

これを避けるため、クライアントの情報を FEFS デザインシートから削除して一時的にファイルシステムから削除します。

MGT 再構築中に別の FEFS を使用して運用を継続する場合は以下の手順を実施してください。

再構築するMGT のファイルシステムをマウントしないように、FEFS デザインシートの当該MGS/MGTを定義するGFS シートから "FX CLIENT" "PG CLIENT" に記載しているノードを削除してください。

構築手順("3.1 導入の流れ" の "3.1.4 FEFSセットアップツール用構成定義ファイルの作成" 以降を参照)に従い、各クライアントに設定ファイルを配布してください。

MGT の再構築

1. FEFS デザインシートより該当する MGT のデバイス名と MGS を確認します。
2. FEFS デザインシートの当該MGS/MGTを定義する「GFSシート」「MGSセクション」の MKFS OPTIONS に "--replace" を追加して設定ファイルを再作成します。
3. ジョブ運用ソフトウェアと連携している場合は、MGSをメンテナンスモードに移行します。
   「ジョブ運用ソフトウェア 管理者向けガイド 保守編」を参照してください。
4. 当該 MGS の /etc/opt/FJSVfefs/config 配下のファイルを退避します。
5. 再作成した設定ファイルを当該 MGS の /etc/opt/FJSVfefs/config/ に上書きします。
6. MGS 上で以下を実施します。ディスク交換が必要な場合は、本操作の前に交換を完了してください。

```
# fefsconfig --setup
# fefs_mkfs <volume>
```

   <*volume*> : 対象 MGT のボリューム名

注意

MGTのデバイス名が変更になる場合は FEFS デザインシートを修正し、設定を再度配布してください。

7. MDT、OST の管理ファイルのクリア
   再構築ファイルシステムの MDT、OST に対して以下のコマンドを実行してください。

```
[MDSノード、OSSノード]
# export PATH="$PATH:/opt/FJSVfefsprogs/sbin"
# tunefs.lustre --writeconf <device>
```

   <*device*> : 対象となるデバイスを指定します。

8. FEFSサービス起動

   以下の手順で起動を行います。

```
[システム管理ノード]
# fefs_sync --start --storage=<cluster>[,...] --compute=<cluster>[,...]
```

   --storage : ストレージクラスタ名を指定してください。
   --compute : 計算クラスタ名および多目的クラスタ名を指定してください。

9. ジョブ運用ソフトウェアと連携している場合は、メンテナンスモードを解除します。
   「ジョブ運用ソフトウェア 管理者向けガイド 保守編」を参照してください。

注意

- FEFS クライアントのマウント設定削除をしていた場合は、サービス起動前に FEFS デザインシートの変更を戻し、設定を再配布してください。

・ 本手順により OST poolの 設定はクリアされるため再設定を実施してください。

D.4.7.2 MDTの部分再構築

MDTの部分 MKFS はできません。ファイルシステムの再構築が必要です。"D.4.10 ファイルシステムの再構築" を参照してください。

D.4.7.3 OSTの部分再構築

事前準備

部分再構築を行う OST に対して以下を実施してください。

```
[システム管理ノード]
# pmexe -c <storage-cluster> --nodetype MDS --stdout "lctl set_param osp.<fsname>-OST<ostindex>-osc-
MDT*.max_create_count=0"
```

<*storage-cluster*>: ストレージクラスタ名
<*fsname*>: ファイルシステム名
<*ostindex*>: OSTのインデックス番号 (ゼロ埋め 4桁 16進数)

ー 実行例

```
[システム管理ノード]
# pmexe -c storage --nodetype MDS --stdout "lctl set_param osp.fefs01-OST0001-osc-MDT*.max_create_count=0"
```

注意

ストレージクラスタに複数のファイルシステムが含まれる場合は、部分再構築するOSTを構築するMDSに範囲を絞って実施してください。

部分再構築

1. 運用から切り離されていることを確認します。
   "D.4.2.2 切離し" の「2. 状態確認」を参照してください。
2. FEFS デザインシートより該当する OST のデバイス名と OSS を確認します。
3. FEFS デザインシートの当該 OSS/OST を定義する「GFSシート」「OSSセクション」のMKFS OPTIONS に "--replace" を追加して設定ファイルを再作成します。
4. ジョブ運用ソフトウェアと連携している場合は、OSS をメンテナンスモードに移行します。
   「ジョブ運用ソフトウェア 管理者向けガイド 保守編」を参照してください。
5. 当該 OST をアンマウントします。マウントポイントは、デバイス名から特定してください。

   ```
   # umount -f <mount_point>
   ```

   <*mount_point*> : ファイルシステムのマウントポイント
6. 当該 OSS の/etc/opt/FJSVfefs/config 配下のファイルを退避します。
7. 再作成した設定ファイルを当該OSSの/etc/opt/FJSVfefs/config/に上書きします。
8. 当該OSS 上で以下を実施します。ディスク交換が必要な場合は、本操作の前に交換を完了してください。

   ```
   # fefsconfig --setup
   # fefs_mkfs <volume>
   ```

   <*volume*> : 対象OSTのボリューム名
9. 退避した設定ファイル/etc/opt/FJSVfefs/config/を戻して、以下を実施します。

   ```
   # fefsconfig --setup
   ```

10. 当該 OST をマウントします。

```
# fefs_mount <mount_point>
```

<*mount_point*> : ファイルシステムのマウントポイント

11. ジョブ運用ソフトウェアと連携している場合は、メンテナンスモードを解除します。
「ジョブ運用ソフトウェア 管理者向けガイド 保守編」を参照してください。

12. 該当する OSS の全OSTを 組み込みます。
"D.4.2.3 組込み" を参照してください。

13. lfsck を実施します。
"4.8.5 FEFS の修復" を参照してください。

注意

- QUOTA の値は lfsckの実施によって再設定されます。
- 故障 OST に保存されていたデータ自体は復元されません。

14. バックアップファイルを復元します。
"D.4.9 ファイルのバックアップ" を参照してください。

注意

OSTの部分再構築でOSTのデバイス名が変わる場合は FEFS デザインシートを再作成し、配布しなおしてください。

事後処理

部分再構築を行った OST に対して以下を実施してください。

```
[システム管理ノード]
# pmexe -c <storage-cluster> --nodetype MDS --stdout "lctl set_param osp.<fsname>-OST<ostindex>-osc-
MDT*.max_create_count=20000"
```

<*storage-cluster*>: ストレージクラスタ名
<*fsname*>: ファイルシステム名
<*ostindex*>: OSTのインデックス番号 (ゼロ埋め 4桁 16進数)

D.4.8 バックアップファイル一覧の作成

以下の手順でバックアップファイル一覧を作成します。

MDTの場合

ファイルシステム全体のバックアップが必要です。バックアップファイル一覧の作成は不要です。

"D.4.9 ファイルのバックアップ" を実施してください。

OSTの場合

不良ブロックが発生した場合

異常があったブロック番号から影響するFEFS上のファイル名を以下の手順で特定します。

1. "D.4.1 不良ブロックが発生したブロック番号の状態確認" の手順を実行し、ブロック番号から inode 番号を特定します。当該ブロック番号が未使用であれば、以降の作業は不要です。

2. inode番号を指定して以下のコマンドを実行します。

```
[OSSノード]
# /opt/FJSVfefsprogs/sbin/fefs_ost2fid <Ost device> <inode number>
```

3. 出力されたFIDに対して、lfs fid2pathコマンドを実行します。
   異常があったブロックを使用しているファイルについて、マウントポイントからの相対パスが表示されます。

```
[クライアントノード]
# lfs fid2path <mount_point> <fid>
```

・ 実行例 (ブロック番号1544 の場合)

```
[OSSノード]
# sync
# /opt/FJSVfefsprogs/sbin/debugfs.ldiskfs /dev/sdb -R "icheck 1544"
Block   Inode number
1544    233
# /opt/FJSVfefsprogs/sbin/fefs_ost2fid /dev/sdb 233
233: [0x200000402:0x1:0x0]

[クライアントノード]
# /usr/bin/lfs fid2path /fefs 0x200000402:0x1:0x0
file
```

マウントポイントからの相対パスが表示されますので、このファイルを退避します。退避途中にEIOなどエラーが発生した場合はデータの読み込みができないため、退避はできません。

OST 故障が発生した場合

異常があったOSTを使用するFEFS上のファイルを特定します。

以下の手順でファイルを特定してください。

```
[MDSノード]
# /opt/FJSVfefsprogs/sbin/find_file_ost -o <outfile> -d <mdt device> <ost index>
```

```
[クライアントノード]
# /opt/FJSVfefs/sbin/convert_fid2path -o <outfile> -m <mount_point> <infile ...>
```

・ 実行例

MDTの数だけ繰り返します。
/tmp/ost0000.out、/tmp/ost0000.errにバックアップファイルの一覧が保管されます。

```
[MDSノード]
# /opt/FJSVfefsprogs/sbin/find_file_ost -o /tmp/mdt0000_ost0000 -d /dev/sdb 0
# /opt/FJSVfefsprogs/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

[クライアントノード]
# /opt/FJSVfefs/sbin/convert_fid2path -o /tmp/ost0000.out -m /fefs ¥
  /tmp/mdt0000_ost0000.out /tmp/mdt0001_ost0000.out
# /opt/FJSVfefs/sbin/convert_fid2path -o /tmp/ost0000.err -m /fefs ¥
  /tmp/mdt0000_ost0000.err /tmp/mdt0001_ost0000.err
# wait
```

一時ファイルを削除します。

```
[クライアントノード]
# rm -f /tmp/mdt0000_ost0000.out /tmp/mdt0001_ost0000.out /tmp/mdt0000_ost0000.err /tmp/mdt0001_ost0000.err
```

```
[MDSノード]
# rm -f /tmp/mdt0000_ost0000.out /tmp/mdt0000_ost0000.err /tmp/mdt0001_ost0000.out /tmp/mdt0001_ost0000.err
```

D.4.9 ファイルのバックアップ

OST がマウントできている場合は、本手順を実施してください。
ファイルをバックアップする場合は、"D.4.8 バックアップファイル一覧の作成" で作成された影響のあるファイルの一覧を使って以下を実行します。

ログインノードで OST の組込み

```
[システム管理ノード]
# 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*>:クラスタ名
<*target*>:対象のターゲット

バックアップ

```
[ログインノード]
# /opt/FJSVfefs/bin/fefsbackup copy -L ost0000_backup -n backup_node -d /backup_root ¥
-f /tmp/ost0000.out --ignore_err 2> /tmp/copy.err
password: ******** ※sshのパスワードの入力
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/FJSVfefs/bin/fefsbackup copy -L ost0000_backup2 -n backup_node -d /backup_root ¥
 -f /tmp/ost0000.err --ignore_err 2> /tmp/copy2.err
password: ******** ※sshのパスワードの入力
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.
```

バックアップできなかったファイルは以下を確認してください。これらのファイルはファイルアクセスでエラーがあったため退避不可能なファイルで、復旧することができません。

```
# cat /tmp/copy.err
# cat /tmp/copy2.err
```

ログインノードで OST の切離し

```
[システム管理ノード]
# pmexe -c <cluster> --stdout --nodetype LN "/usr/sbin/force_intr -c -m deactivate <target>..."
```

<*cluster*>:クラスタ名
<*target*>:対象のターゲット

リストア

復旧後、バックアップ先ノードで以下のようにして退避したデータを戻します。

```
[クライアントノード]
# cd /backup_root
# /opt/FJSVfefs/bin/fefsbackup copy -L ost0000_restore -n restore_node -d /fefs ./
```

注意

ファイルを退避していない場合は、故障したOSTにより影響を受けたファイルを削除します。作成済のファイル一覧が mdt0000_ost0000.out である場合の実行例を以下に示します。

```
[クライアントノード]
# cat mdt0000_ost0000.out | xargs -n 1  unlink
```

D.4.10 ファイルシステムの再構築

データの退避後にファイルシステムの再構築を実施してください。

フルバックアップ

再構築の前に、ファイルシステムをフルバックアップします。手順を以下に示します。
マウントポイントに移動してから、バックアップを行います。

・ 実行例

```
[クライアントノード]
# cd /fefs
# /opt/FJSVfefs/bin/fefsbackup copy -L fefs_backup -n backup_node -d /backup_root ¥
--ignore_err ./ 2> /tmp/copy.err
password: ******** ※sshのパスワードの入力
Copying request ost0000_backup is executing...
Total 2 files (2048 bytes) were copied from ./ directory.
Copying request fefs_backup is completed successfully.
```

バックアップできなかったファイルは以下を確認してください。これらのファイルはファイルアクセスでエラーがあったため退避不可能なファイルで、復旧することができません。

```
# cat /tmp/copy.err
```

注意

ー 実施前に容量を確認し、バックアップ可否を検討してください。

ー フルバックアップ結果をリストアするには、バックアップ先ノードで以下のようにします。

```
[クライアントノード]
# cd /backup_root
# /opt/FJSVfefs/bin/fefsbackup copy -L fefs_restore -n restore_node -d /fefs ./
```

ファイルシステムの再構築

ディスク交換が必要な場合は、交換を実施した後、"3.1.6 FEFSの構築" を参照してファイルシステムの再構築を実施してください。

注意

ファイルシステムの再構築は、ファイルシステムのデータ保護を解除後に実施してください。詳細は "4.19.2 ファイルシステムのデータの保護を解除する手順" を参照してください。

D.5 アクセス影響

ディスク故障時やサーバ両系ダウン時のファイルシステムのアクセス影響を以下に示します。

操作	OST			MDT		
	ReadOnly	両系故障 ディスク故障	強制IO中断	ReadOnly	両系故障 ディスク故障	強制IO中断
既存ファイルのオープン	OK	ハング※1	エラー※2	OK	ハング※1	エラー※2
ファイル作成	OK	ハング※1	OK	エラー※1	ハング※1	エラー※2
ファイル削除	OK	ハング※1	エラー※2	エラー※1	ハング※1	エラー※2
read	OK	ハング※1	エラー※2	OK	ハング※1	エラー※2
write	エラー※1	ハング※1	エラー※2	OK	ハング※1	エラー※2

操作	OST			MDT		
	ReadOnly	両系故障 ディスク故障	強制IO中断	ReadOnly	両系故障 ディスク故障	強制IO中断
ディレクトリ作成/削除	OK	ハング※1	OK	エラー※1	ハング※1	エラー※2
stat (ファイル)	OK	ハング※1	エラー※2	OK	ハング※1	エラー※2
stat (ディレクトリ)	OK	OK	OK	OK	ハング※1	エラー※2
statfs (df)	OK	OK	OK	OK	ハング	エラー
lfs project	エラー※1	ハング※1	エラー※2	エラー※1	ハング※1	エラー※2
lfs setquota	OK	ハング	OK	OK	ハング	エラー※2
lfs setstripe ファイル	OK	ハング※1	OK	エラー※1	ハング※1	エラー※2
lfs setstripe ディレクトリ	OK	OK	OK	エラー※1	ハング※1	エラー※2
lfs getstripe ファイル/ディレクトリ	OK	ハング※1	OK	OK	ハング	エラー※2
lfs fid2path	OK	OK	OK	OK	ハング※1	エラー※2
lfs find OST上のファイル検索	OK	OK	OK	OK	ハング	エラー

※1: 両系故障、ディスク故障以外へのファイルアクセスは正常にアクセス可能

※2: 強制I/O中断していないデバイスへのファイルアクセスは正常にアクセス可能

付録E　ファイルシステム故障発生時のジョブ運用継続手順

本章では、ファイルシステムに故障が発生した場合でも、ジョブ運用を継続できる手順を示します。

ファイルシステム故障時にジョブ運用を継続したい場合に本手順を実施してください。ジョブ運用を継続しない場合は、「D.4.2 FEFS サーバの切離し/組込み」を参照してください。

以下の手順を示します。

- 運用中のファイルシステム切離し・組込み手順
- ファイルシステム故障の影響でハングしたジョブの刈り取り手順
- ファイルシステム故障中のノード起動手順

E.1　ファイルシステムの切離し/組込み手順

E.1.1　切離し

故障したファイルシステムを切り離すことにより、それ以外のファイルシステムでジョブ運用を継続することができます。

1. 切離し
   以下の手順で対象のファイルシステムを切り離します。

   ```
   [システム管理ノード]
   # pmexe -c <クラスタ名> --stdout --nodetype CCM,LN "/usr/sbin/fefs_deactivate -m deactivate <対象のマウントポイント>"
   ```

   ー 実行例

   ```
   [システム管理ノード]
   # pmexe -c compute --stdout --nodetype CCM,LN "/usr/sbin/fefs_deactivate -m deactivate /fefs"
   ```

2. 状態確認
   切離し後に以下の手順で状態を確認します。

   ```
   [システム管理ノード]
   # pmexe -c <クラスタ名> --stdout --nodetype CCM,LN "/usr/sbin/fefs_deactivate -m status <対象のマウントポイント>"
   ```

   対象のファイルシステムに対して状態が "IN" であることを確認します。

   ー 実行例

   ```
   [システム管理ノード]
   # pmexe -c compute --stdout --nodetype CCM,LN "/usr/sbin/fefs_deactivate -m status /fefs"
   ```

   出力について詳細は "A.2.18 fefs_deactivate コマンド" を参照してください。

E.1.2　組込み

ジョブ運用を継続したまま、復旧したファイルシステムを組み込むことができます。ファイルシステムの復旧については、"付録D ファイルシステムの復旧手順" を参照してください。

1. 組込み
   以下の手順で対象のファイルシステムを組み込みます。

   ```
   [システム管理ノード]
   # pmexe -c <クラスタ名> --stdout --nodetype CCM,LN "/usr/sbin/fefs_deactivate -m activate <対象のマウントポイント>"
   ```

ー 実行例

```
[システム管理ノード]
# pmexe -c compute --stdout --nodetype CCM,LN "/usr/sbin/fefs_deactivate -m activate /fefs"
```

2. 状態確認
   組込み後に以下の手順で状態を確認します。

```
[システム管理ノード]
# pmexe -c <クラスタ名> --stdout --nodetype CCM,LN "/usr/sbin/fefs_deactivate -m status <対象のマウントポイント>"
```

対象のファイルシステムに対して状態が "UP" であることを確認します。

ー 実行例

```
[システム管理ノード]
# pmexe -c compute --stdout --nodetype CCM,LN "/usr/sbin/fefs_deactivate -m status /fefs"
```

出力について詳細は "A.2.18 fefs_deactivate コマンド" を参照してください。

E.2 ファイルシステム故障の影響でハングアップしたジョブの刈り取り手順

故障したファイルシステムの影響でハングアップしているジョブは、以下の手順によってハングを解消することができます。

1. 切離し
   「E.1.1 切離し」を実施後、ハングアップしているジョブが割り当たっているノードにおいて、以下の切離しを実施することで、ハングアップを解消します。

```
[システム管理ノード]
# pmexe -c <クラスタ名> --stdout --bootgrp <対象ブートグループ範囲> "/usr/sbin/fefs_deactivate -m deactivate <対象のマウントポイント>"
```

ー 実行例

```
[システム管理ノード]
# pmexe -c compute --stdout --bootgrp 0x0101,0x0102 "/usr/sbin/fefs_deactivate -m deactivate /fefs"
```

2. 状態確認
   切離し後に以下の手順で状態を確認します。

```
[システム管理ノード]
# pmexe -c <クラスタ名> --stdout --bootgrp <対象ブートグループ範囲> "/usr/sbin/fefs_deactivate -m status <対象のマウントポイント>"
```

対象のファイルシステムに対して状態が "IN" であることを確認します。

ー 実行例

```
[システム管理ノード]
# pmexe -c compute --stdout --bootgrp 0x0101,0x0102 "/usr/sbin/fefs_deactivate -m status /fefs"
```

出力について詳細は "A.2.18 fefs_deactivate コマンド" を参照してください。

注意

- 切離しの最小単位は BoB です。
- 本手順を行った際、切り離したファイルシステムを利用していたジョブ、ハングアップしていたジョブは、エラーになります。ジョブは PJM CODE 28 または 180 でエラー終了します。

E.3 ファイルシステム故障中のノード起動手順

クライアントノード停止中にファイルシステムが故障した場合、クライアントノード起動の際に FEFS サービスを正常に起動させるには、故障したファイルシステムを切り離す必要があります。

クライアントノードのノード種別は以下です。

- 計算ノード (CN)
- 計算ノード兼ストレージ I/O ノード (CN/SIO)
- 計算ノード兼グローバル I/O ノード (CN/GIO)
- 計算ノード兼ブート I/O ノード (CN/BIO)
- ログインノード (LN)
- 計算クラスタ管理ノード (CCM)

1. 該当ノードのノード起動
   停止していたノードを起動します。

   ```
   [システム管理ノード]
   # papwrctl -c <クラスタ名> -n <対象のノードID> on
   ```

2. 切離し
   FEFS サービスの状態は FEFS(s) のままとなります。以下の手順で対象のファイルシステムを切り離します。

   ```
   [システム管理ノード]
   # pmexe -c <クラスタ名> --stdout -n <対象のノードID> "/usr/sbin/fefs_deactivate -m deactivate <対象のマウントポイント>"
   ```

   ー 実行例

   ```
   [システム管理ノード]
   # pmexe -c compute --stdout -n 0x01010010 "/usr/sbin/fefs_deactivate -m deactivate /fefs"
   ```

3. 状態確認
   切離し後に以下の手順で状態を確認します。

   ```
   [システム管理ノード]
   # pmexe -c <クラスタ名> --stdout -n <対象のノードID> "/usr/sbin/fefs_deactivate -m status <対象のマウントポイント>"
   ```

   対象のファイルシステムに対して状態が "IN" であることを確認します。

   ー 実行例

   ```
   [システム管理ノード]
   # pmexe -c compute --stdout -n 0x01010010 "/usr/sbin/fefs_deactivate -m status /fefs"
   ```

   出力について詳細は "A.2.18 fefs_deactivate コマンド" を参照してください。

4. FEFS 状態の確認
   ノードの FEFS サービスが正常に起動されたことを pashowclst コマンドで確認してください。

   ```
   [システム管理ノード]
   # pashowclst -c <クラスタ名> -n <対象のノードID>
   ```

   FEFS の状態が FEFSSR(o) および FEFS(o) に遷移していれば、FEFS のサービスは正常に起動されています。

注意

- 切離し手順実施済みのノードは、ノード再起動時、本手順を実施せずとも FEFS サービスは正常に起動されます。

付録F　トラブル対処時に必要な資料

トラブルが発生したときは、以下の資料を採取してください。

FX サーバの場合

採取資料の種類	対象ノード	採取ファイル/採取コマンド
システムログ	全ノード	/var/log/messages*
PANIC DUMP	DUMP が採取されたノード	/var/crash/OSdump-*
システムの資料	全ノード	pasnapコマンドで採取された、OSの調査資料
FEFS の資料	全ノード	以下のコマンドを実行し、作成された <outputdir>/fefssnap_<タイムスタンプ>.tgz # /usr/sbin/fefssnap -d <outputdir> ※<タイムスタンプ>はコマンドの実行時間(yyyymmddHHMMSS)です。 pasnapコマンドで採取された、FEFSの調査資料

PRIMERGY サーバの場合

採取資料の種類	対象ノード	採取ファイル/採取コマンド
システムログ	全ノード	/var/log/messages*
PANIC DUMP	DUMP が採取されたノード	/var/crash/127.0.0.1-*XXX* (diskdump の場合) ※ *XXX* は採取された年月日です。
カーネルの namelist	DUMP が採取されたノード	/usr/lib/debug/lib/modules/*version*/vmlinux ※ *version* はカーネルのバージョンです。
カーネルの mapfile	DUMP が採取されたノード	/boot/System.map-*version* ※ *version* はカーネルのバージョンです。
システムの資料	全ノード	pasnap コマンドで採取された、OSの調査資料
FEFS の資料	全ノード	以下のコマンドを実行し、作成された *<outputdir>*/fefssnap_*<タイムスタンプ>*.tgz # /usr/sbin/fefssnap -d *<outputdir>* ※*<タイムスタンプ>*はコマンドの実行時間(yyyymmddHHMMSS)です。 pasnap コマンドで採取された、FEFS の調査資料

参照

pasnap コマンドの資料採取方法の詳細については、以下のマニュアルを参照してください。

「ジョブ運用ソフトウェア 管理者向けガイド システム管理編」

参考

- LLIO が組み込まれている場合、fefssnap コマンドはLLIO も資料採取の対象とします。

・ トラブルの調査のために、FEFSの内部ログである fefs.log が必要となることがあります。fefs.log は、状況によっては存在しない場合があります。fefs.log ログに、何も出力するものがないのか、ログ出力が停止しているのかを調査するには、該当ノード上で ps コマンドを実行してください。以下のコマンドが動作していればログ採取ができている状態です。

```
lctl fefslog start /var/opt/FJSVfefs/fefs.log
```

用語集

以下の用語に加えて、Technical Computing Suite 全体に関わる用語は、マニュアル「ジョブ運用ソフトウェア 用語集」を参照してください。

ACL (Access Control List)

個々の利用者が持つアクセス権限や、アクセス可能なファイル資源を列挙したリストです。

BoB (Bunch of Blade) [FX]

FXサーバの制御単位。16ノードで構成されます。

by-id 名

ハードディスクに設定されているユニークな識別情報 (シリアル番号など) から生成されるデバイス名です。ディスクを交換しない限り不変なので、常に同じデバイス名でデバイスを利用できます。

FEFS (Fujitsu Exabyte File System)

富士通が開発した並列分散ファイルシステムです。

FEFS サービス監視デーモン

FEFS サービスを監視し、サービスの状態取得・通知を行うデーモンプロセスです。

FID

FEFS 内部の inode に対する管理番号です。

HA 構成

サーバを冗長化することによって、サービス提供ができなくなる事態の発生頻度を低下させたシステム構成です。

I/Oノード

ディスク装置やネットワーク装置が接続され、計算ノードに対して、ファイルシステム機能やノード外へのネットワーク機能を持つノードです。

inode

ファイルサイズやタイムスタンプ、uid、gid、およびファイルブロックの格納場所を示す情報を保持するデータです。

LLIO (Lightweight Layered IO-Accelerator)

FEFSと計算ノードの間に高速なフラッシュメモリを用いたストレージ階層を設け、FEFSのキャッシュ領域やジョブの一時領域として使用することで高性能を実現する技術です。

LNet

Ethernet、InfiniBand など複数の異なるネットワークを共通に利用してファイルシステムにアクセスするための機能です。

MDS (Metadata Server)

メタデータを格納および管理するメタデータサーバです。

MDT (Metadata Target)

MDS に接続したディスク装置上の、メタデータを格納するための論理ボリュームです。

MGS (Management Server)

MDS、MDT、OSS、および OST の構成を管理する管理サーバです。

MGT (Management Target)

ファイルシステムの構成情報を格納するための論理ボリュームです。

NID (Network Identifier)

FEFS で使用するネットワークの識別子です。

OSS (Object Storage Server)

ファイルデータを格納および管理するデータサーバです。

OST (Object Storage Target)

OSS に接続したディスク装置上の、ファイルデータの実体を格納する論理ボリュームです。

OST_pool

指定した複数の OST を束ねて 1つのグループとし、ファイルやディレクトリをグループ内の OST に割り当てる機能です。

RDMA (Remote Direct Memory Access) 通信

あるノードの主記憶から、離れたノードの主記憶へ、CPU を介さず直接データ転送を行う通信方法です。

Tofu インターコネクト D [FX]

FX サーバにおけるTofuインターコネクトの呼称です。本書では便宜上、単にTofuインターコネクトと表現します。

オブジェクトファイル

ファイルデータの実体を格納するファイルです。

スパースファイル (Sparse File)

ファイルの途中に書込みがされていない領域が存在するファイルです。スパースファイルで書き込まれていない領域については、OST 上のブロックを消費しないため、QUOTA のディスク使用量としてカウントされません。

ファイルデータ

ファイルの内容であるデータの実体です。

ファイルブロック

ファイルデータを構成する各ブロックです。

フェイルオーバ

サーバに障害が発生した場合に、代替サーバが処理やデータを引き継ぐ機能です。

フェイルバック

元のサーバの障害が解消して稼働が再開されたときに、代替サーバから再び処理やデータを引き継ぐ機能です。

マルチ MDS

1つのファイルシステムを、2台以上の MDS サーバで構築することです。

メタデータ

ファイルサイズやタイムスタンプ、所有者などファイルデータ以外の情報です。

ローリングアップデート

パッケージ適用において、システムまたはクラスタ全体を停止せず、一部の計算ノードでクラスタ内のジョブ運用を継続しながら部分的な保守をすることです。

ログインノード

ユーザーがログインして、ジョブの作成や投入を行うためのノードです。

外部ジャーナル

ジャーナル機能のジャーナル領域を外部デバイスのボリュームに作成することです。

計算ノード

ジョブを実行するノード。計算を行う最小単位です。

多目的ノード

ジョブ運用ソフトウェアで、管理者が任意の用途で使うノードです。

中継機能

異なるネットワーク間を中継する機能です。

内部ジャーナル

ジャーナル機能のジャーナル領域を MDT 内部の領域に作成することです。