FUJITSU
shaping tomorrow with you

FUJITSU Software FEFS Lite（エフ・イー・エフ・エス　ライト）

高性能ファイルシステムをスモールスタートで

2013年9月16日

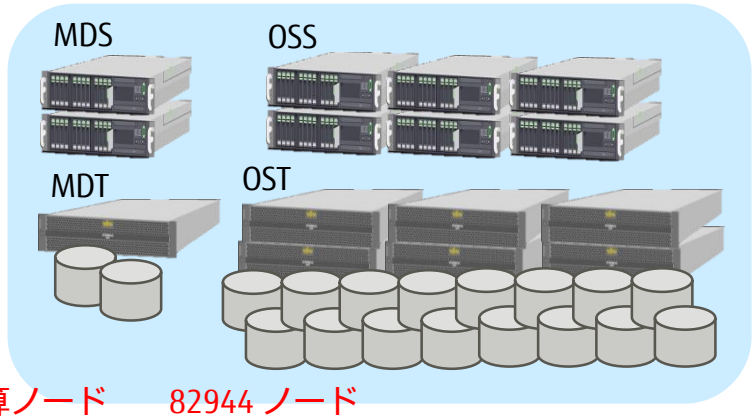
富士通株式会社
甲斐　俊彦

FEFSとは

- FEFSは理化学研究所様のスーパーコンピュータ「京」に導入
- 実測でI/O総スループット 1TB/s を達成

K computer

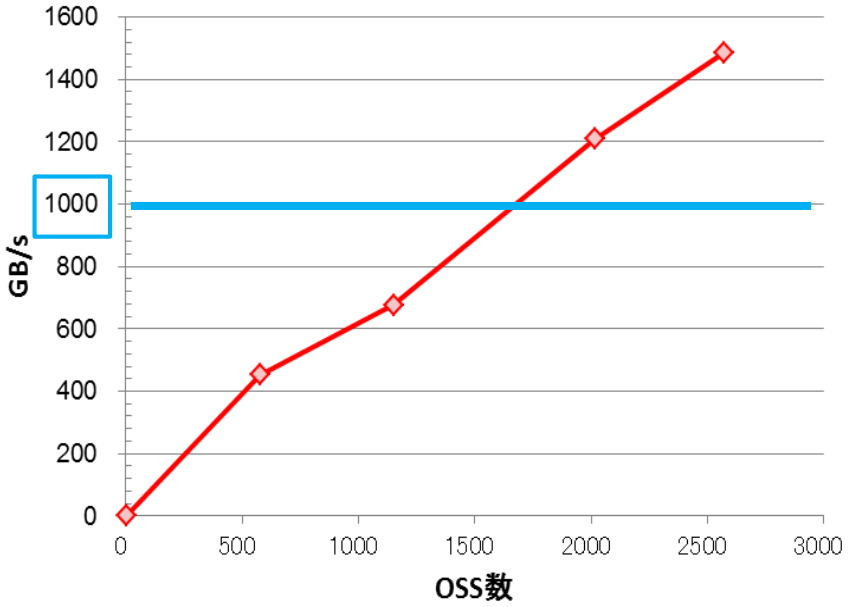
実行性能
2011年世界No.1
ｽﾊﾟｺﾝTOP500



計算ノード　82944 ノード
グローバルFS　30 PB以上
ローカルFS　11 PB以上

*理化学研究所様との共同調査

Read性能:約 1 TB/s超



測定方法：IOR、I/O長1MiB、ファイル/プロセス

※「京」は理化学研究所様と富士通が共同で開発したスーパーコンピュータです
※「京」は理化学研究所様の登録商標です

FEFSの特長

FUJITSU

■ 高性能

Lustre機能強化

- 世界最高クラスの1TB/sの総スループット性能(約1TB/sから)
- 1秒間に数万個のファイル作成が可能(一般Lustreの約3倍)

■ 高信頼性

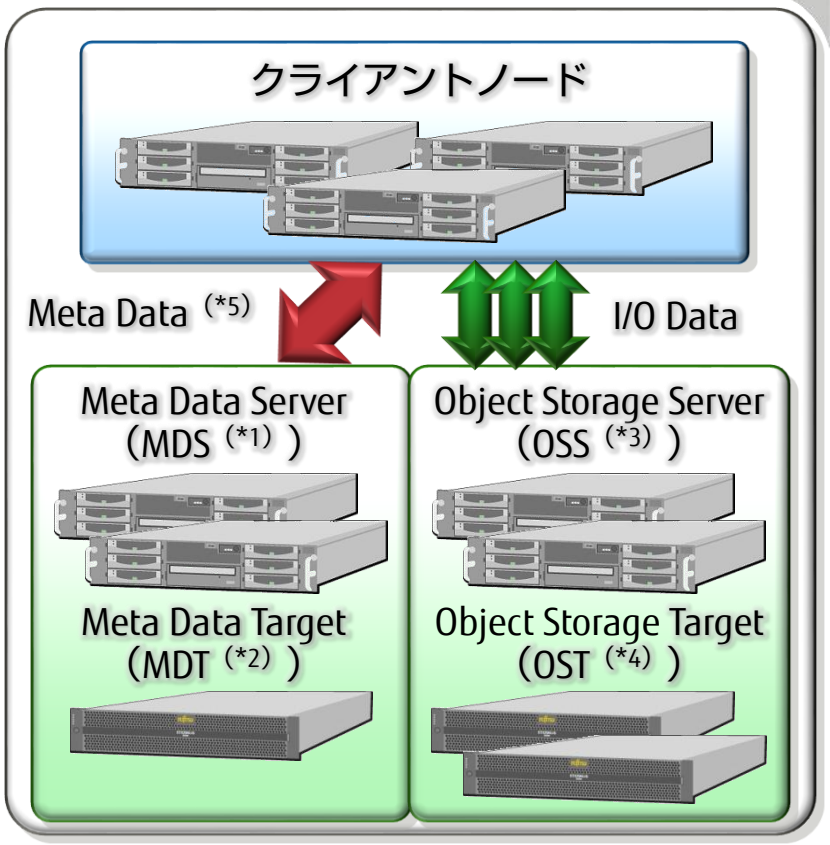
Lustre機能追加

- 物理的な冗長化(RAID構成、マルチパス、複数サーバ)が可能
- トラブル発生時切り替え(フェイルオーバー)が可能

■ 拡張性

Lustre機能強化

- 数TByteから最大8EByte（8,000,000TByte）規模まで拡張可能
- 数十台から最大100万台規模のクライアントノードからの利用が可能

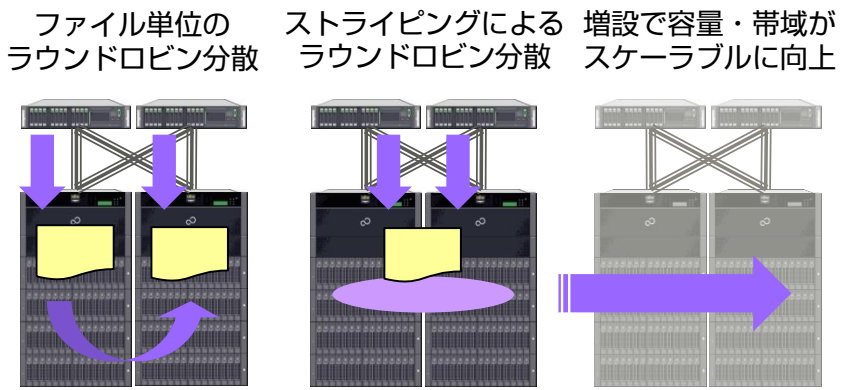


FEFS構成イメージ

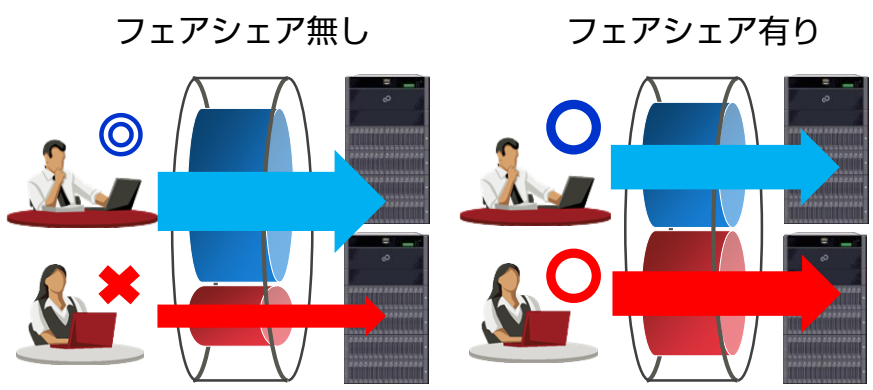
*1 MDS：Meta Data Server （メタデータを管理するサーバ）
*2 MDT：Meta Data Target （MDSに接続するストレージ）
*3 OSS：Object Storage Server（ファイルデータを制御するサーバ）
*4 OST：Object Storage Target（OSSに接続するストレージ）
*5 データについてのデータ （あるデータそのものではなく、そのデータに関連する情報）のこと。例えば、データの作成日時や作成者、データ形式、タイトル、注釈など。データを効率的に管理したり検索したりするために重要な情報

FEFSの機能

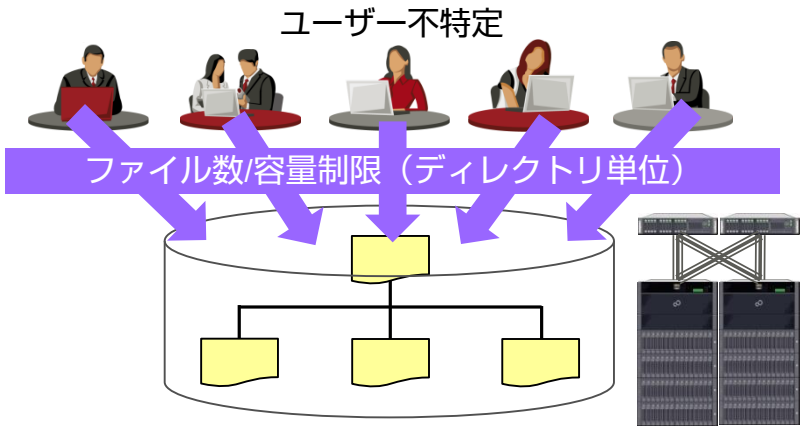
ラウンドロビン負荷分散(高バンド幅I/O)



フェアシェア/優先制御機能



ディレクトリ単位のクオータ指定

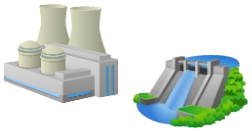


冗長機能・高可用性

HPCだけではなく様々な分野から関心

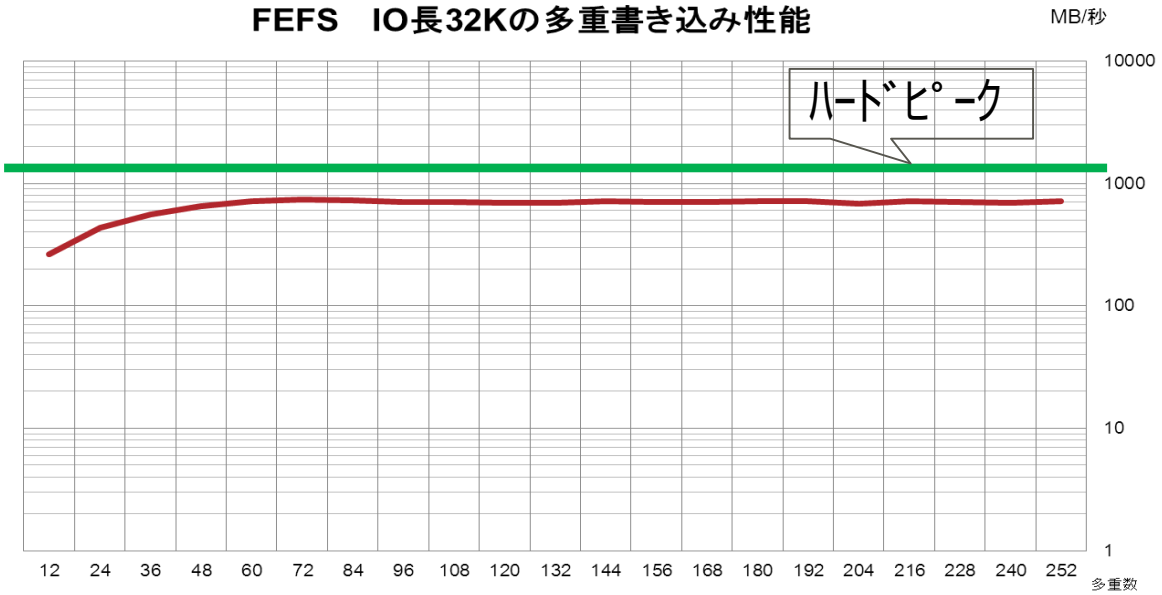
FUJITSU

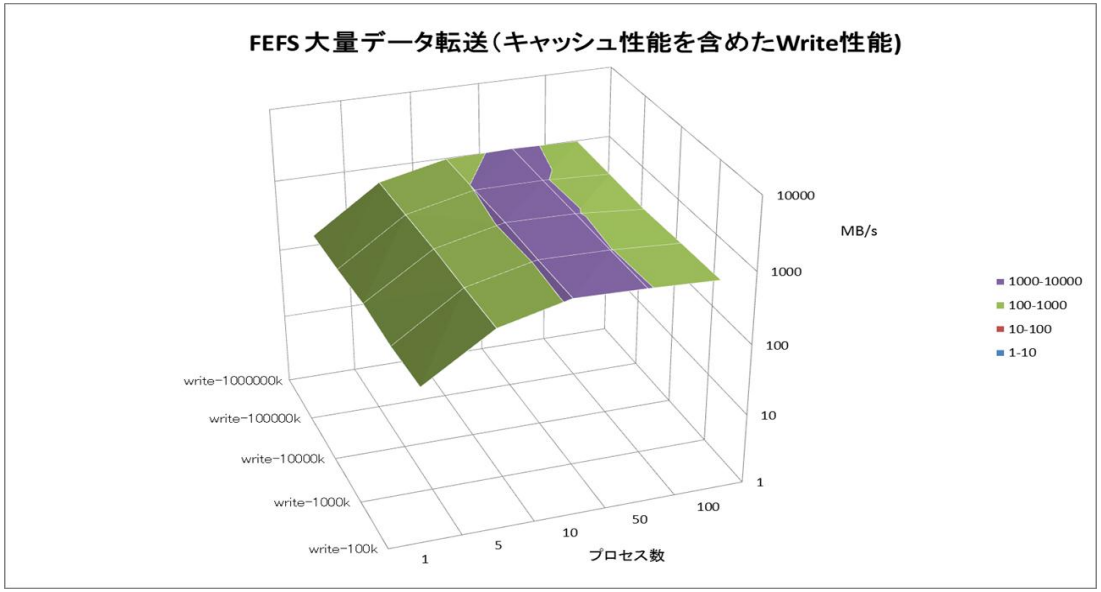
設計／製造	設計データの変換処理（半導体製造）
放送・通信	映像データ制作（メディア系企業）
インフラ・エネルギー	大量データバッチ処理（インフラ系企業）



- キーワードは・・・
  - **大容量のデータを高速に扱いたい**
  - **大量のクライアントからアクセスしたい**
  - **特別な仕組み、専用のインタフェース無くファイルを利用したい（今あるアプリをそのまま使いたい）**

普通のI/Oにも強いFEFS





I/O長32KBでも
ハード性能を引き出す

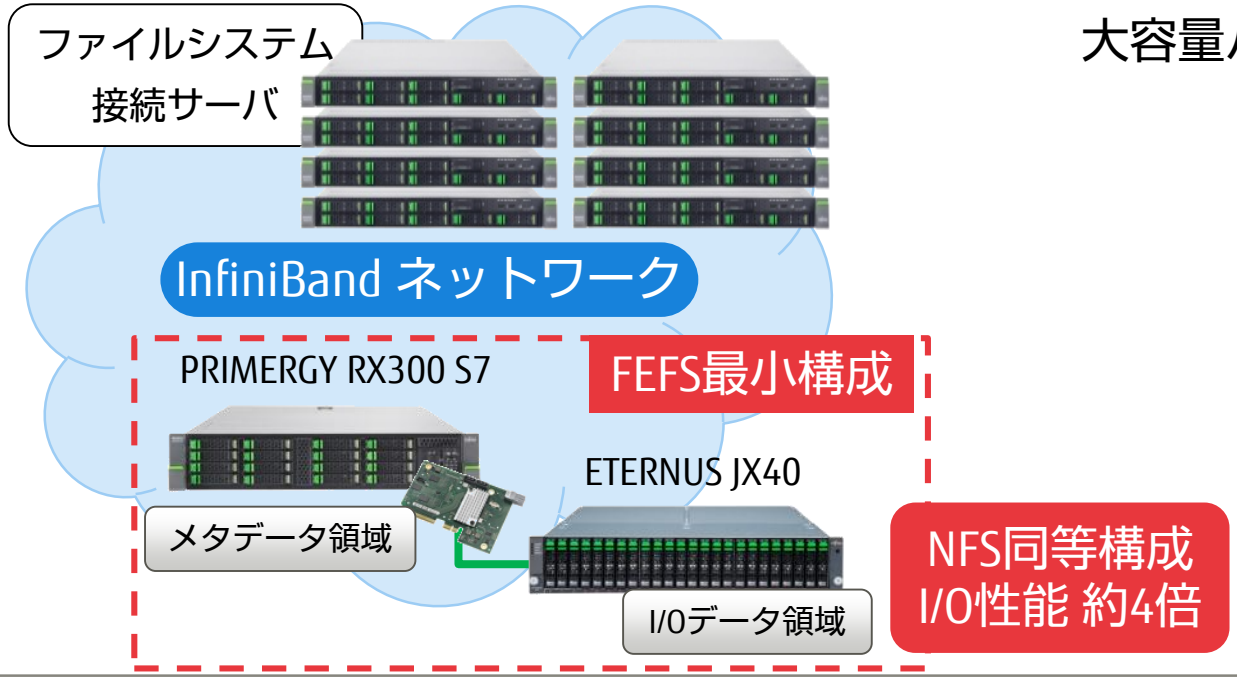
I/O長、多重度によらず
安定した性能

FEFSをより手軽に導入「FEFS Lite」

FUJITSU

- プログラムは従来の FEFS と同一
  ⇒　ラウンドロビン分散機能（高バンド幅I/O）、フェアシェア機能／優先制御機能、ディレクトリ単位のクォータ指定 等の機能を利用可能
- 16サーバまでのシステムに向けたエントリソリューション

  システム価格を従来の1/2



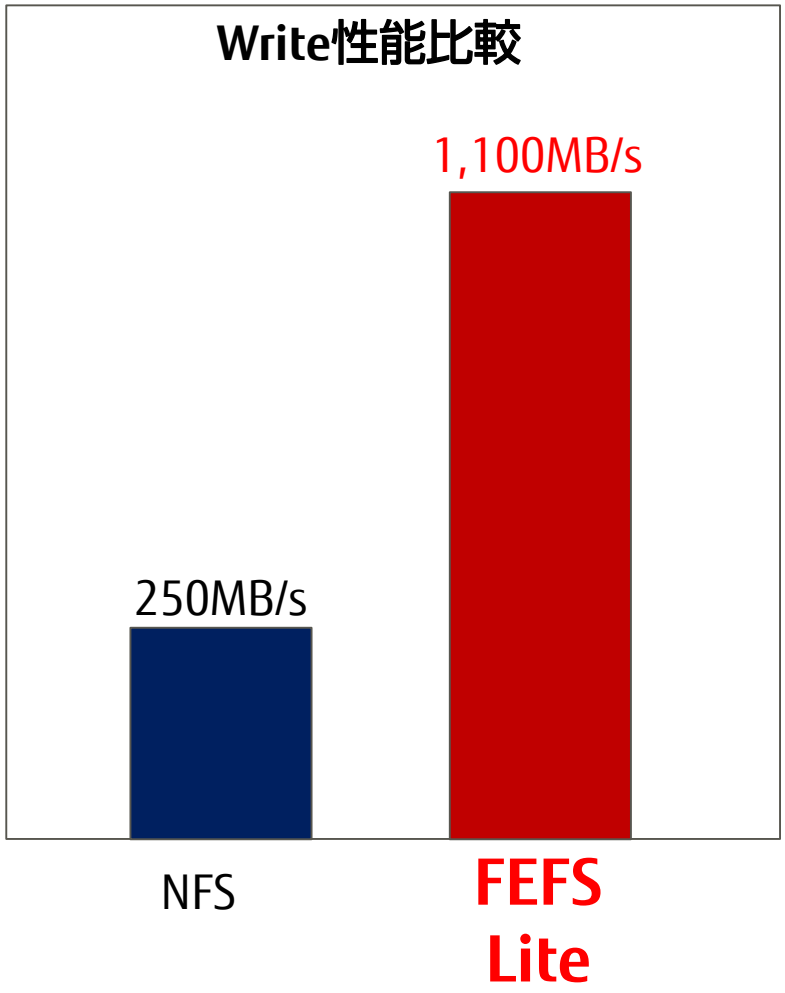
大容量ハードディスクキャビネット
2014年提供予定※

※2013年9月時点では未発表となるため、予告なく変更することがあります

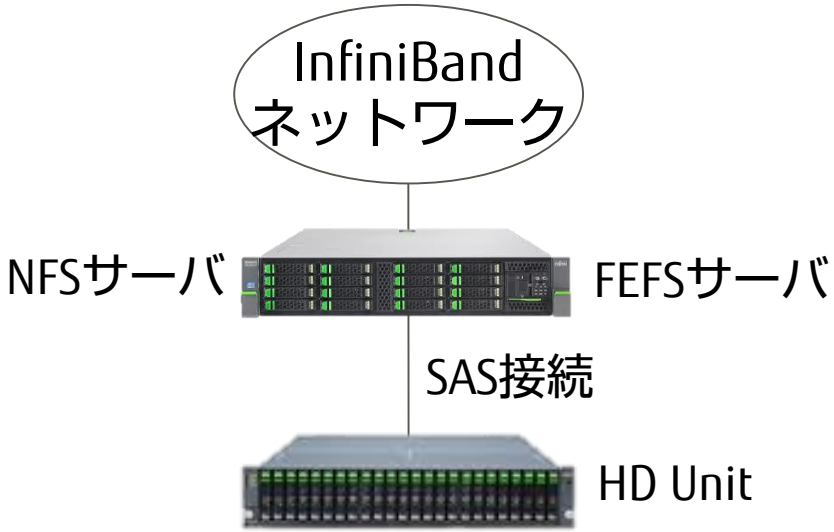
NFSとFEFS Liteの比較

FUJITSU

同一ハードウェア構成で4.4倍の性能



FEFSはInfiniBandの性能を使いきることが可能
1サーバから性能向上が可能



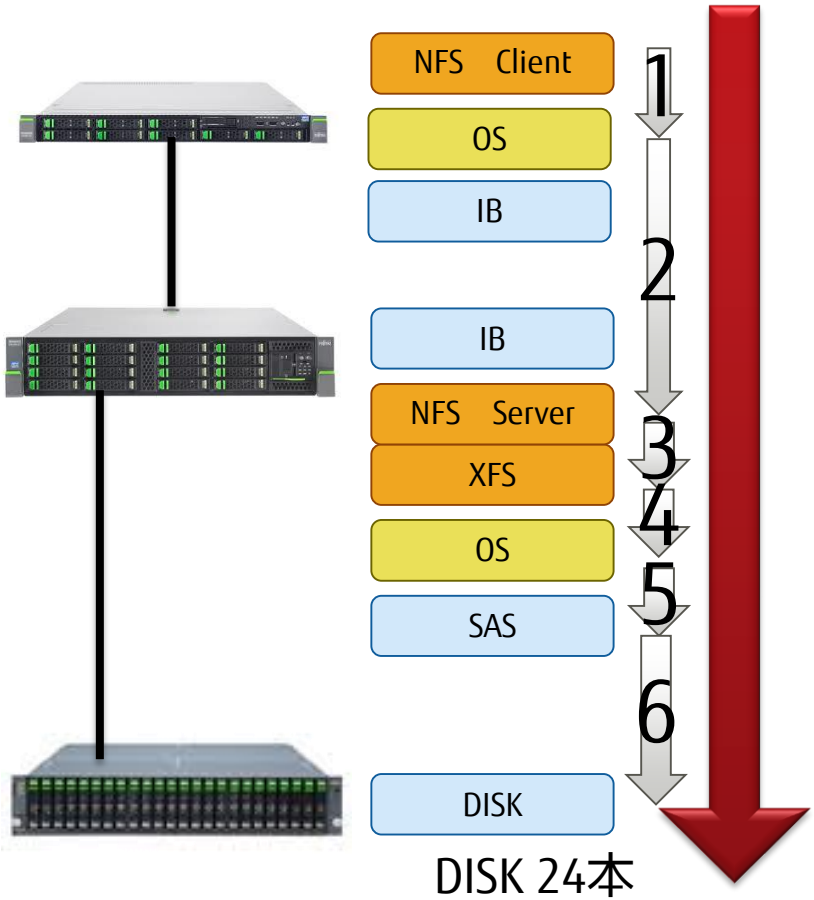
NFSとFEFS Liteの性能差解説

FUJITSU

NFS

Write 250MB/s

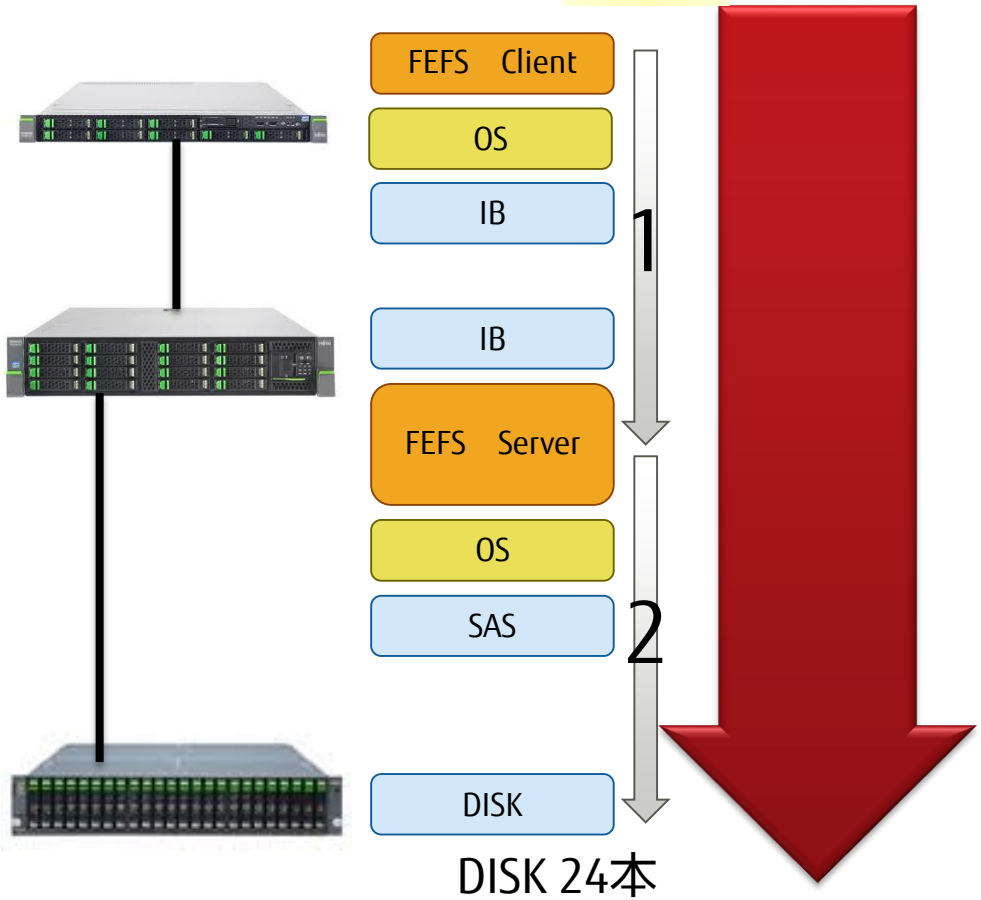
何度もコピー



FEFS Lite

Write 1100MB/s

直接転送

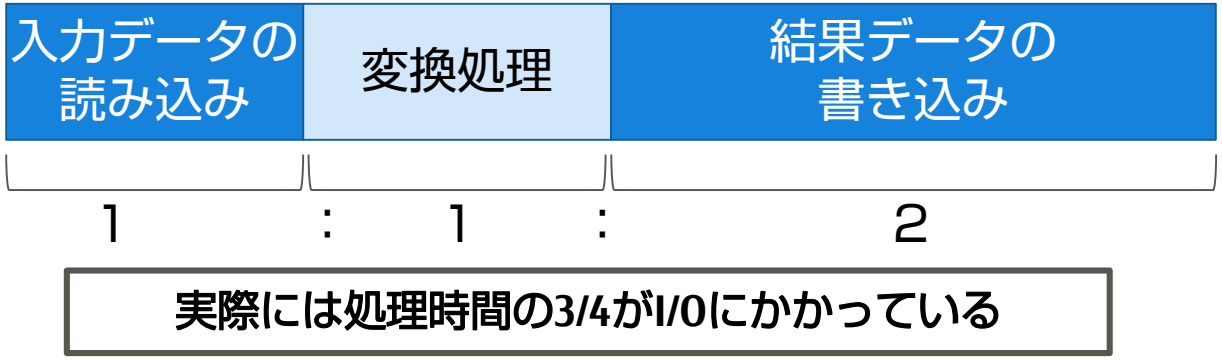


業務処理に占めるI/Oの「重さ」

FUJITSU

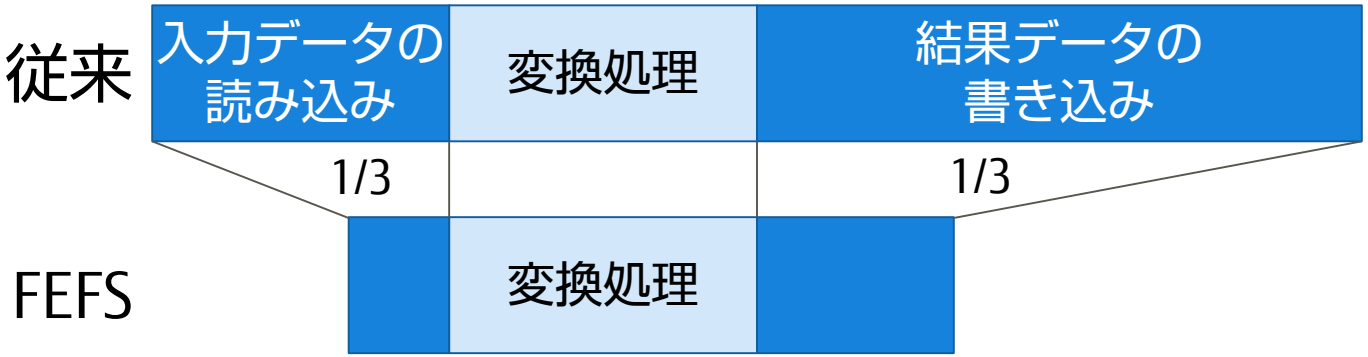
■ 設計データの変換処理の場合

100GB超のファイルを並列で読込み、変換処理後、1ファイルに結果を書込み



実際には処理時間の3/4がI/Oにかかっている

FEFSの利用で、I/O性能を3倍（600MB/s -> 2GB/s）に強化



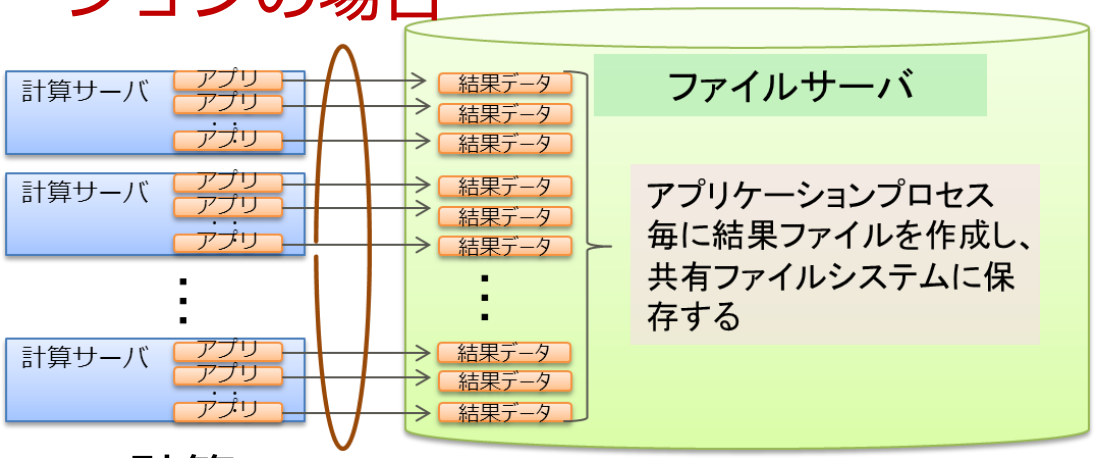
I/O性能が3倍になれば、システム全体の処理時間も1/2に短縮！

サーバ数台規模から発揮されるFEFSの力

FUJITSU

■ ある流体解析アプリケーションの場合・・・

計算プロセスがそれぞれ
結果ファイルを出力
⇒サーバ2台の処理でも
32多重でI/Oが発生



例）計算サーバ2台（32並列）での計算

	①計算処理	②結果ファイル出力	③合計
I/O先：NFS（RAID6（10+2））	360分.	257分.	617分.
I/O先：FEFS（RAID6（4+2）×4）	371分.	59分.	430分.

77%の削減

FEFS Lite

FUJITSU

■ システム構成例

PCクラスタ：クライアント（計算ノード,管理ノード, etc）

FEFS Lite（MDS兼OSS×1、OST×1）

RAID6: (10D+2P)×2

容量	6TB ～
性能	1.1GB/s ～
価格	\ 7,049,600 ～

商品名	個数	価格
FEFS Lite ライセンス価格(20ノードまで)	1	\ 2,300,000
PRIMERGY RX300 S7(MDS兼OSS)	1	\ 2,789,600
ETERNUS JX40(SAS 300GB x24)	1	\ 1,960,000

大容量モデル（MDS兼OSS×1、OST×1）※

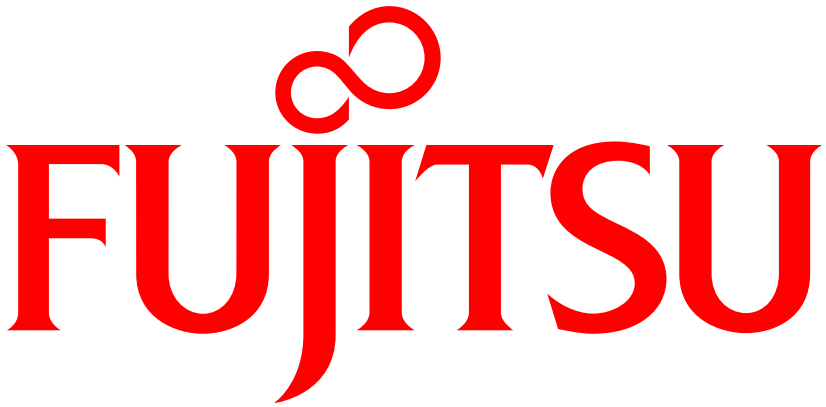
RAID6: (10D+2P)×5

容量	100TB ～
性能	2.0GB/s ～
価格	未発表

商品名	個数	価格
FEFS Lite ライセンス価格(20ノードまで)	1	\ 2,300,000
PRIMERGY RX300 S7後継(MDS兼OSS)	1	未発表
JX40後継(NL-SAS 2TB x 60)	1	未発表

※InfiniBandは、インターコネクトとFEFSパスを兼用します
※ラック、無停電電源装置は含まれておりません
※現調費、搬入費、保守費なども含んでおりません

※2013年9月時点では未発表となるため、
予告なく変更することがあります



shaping tomorrow with you