

2. ZFSの特長

ZFSの特長

◆優れたスケーラビリティ

- 世界初の128bitのファイルシステム
- 事実上無限大といえるファイルシステムの構築

◆簡素化された管理

- 管理コマンドの簡素化
- ストレージプールによるデバイスの隠蔽とストレージ管理の容易化

◆データの堅牢性

- RAID構成によるデータ冗長化
- Copy-On-Write (COW) トランザクションファイルシステムモデル
- チェックサムとデータの自己修復

優れたスケーラビリティ

■ ZFS… Zetta byte File System

↳ 10^{21} (1,000,000,000,000,000,000,000) byte (70bit)



128bit はさらに大きなサイズ!! (3.4×10^{38})

	UFS	ZFS
ファイルシステムサイズの上限	16TB	25京6000兆ZB (256,000,000,000,000,000,000,000TB)
ファイルシステム内に作成可能なファイル数	iノードの数分※	上限なし

※ファイルシステムのサイズにより異なる。16TBの場合、 15×10^6 個

➤ **実際のオペレーションにおいては無限と考えてよい**

参考: キロ (10^3)、メガ (10^6)、ギガ (10^9)、テラ (10^{12})、ペタ (10^{15})、エクサ (10^{18})、ゼタ (10^{21})、ヨタ (10^{24})

簡素化された管理(1)

ZFSの管理に必要なコマンドはわずか2つ

zfs コマンド

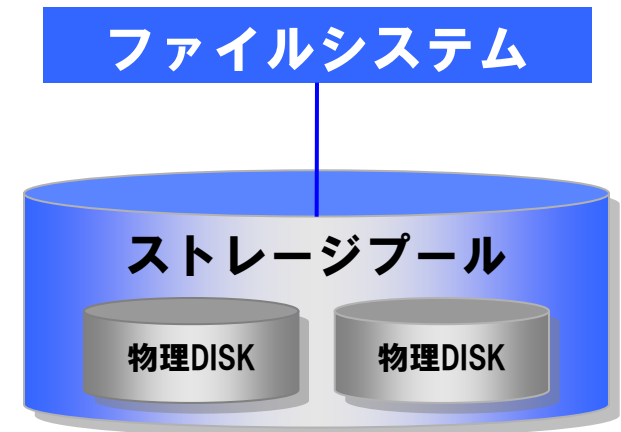
<主な操作>

- ・ストレージプール上にファイルシステム作成、マウント
- ・スナップショットやクローンの作成
- ・ファイルシステムのバックアップ、リストア

zpool コマンド

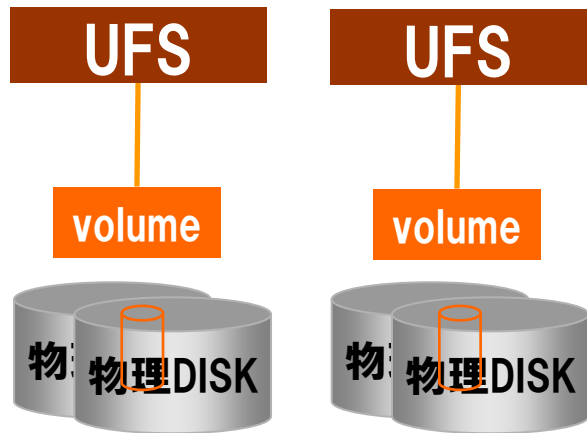
<主な操作>

- ・物理ディスクからストレージプールを作成
- ・ディスクの冗長構成(RAID)を設定
- ・データ整合性チェック(ディスクスクラブ)

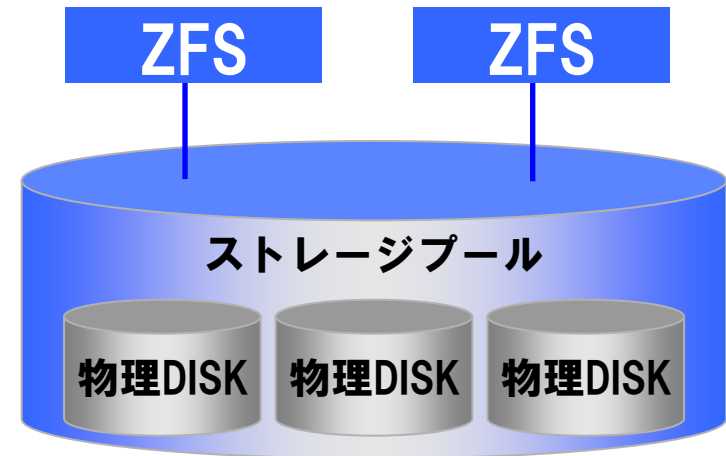


簡素化された管理(2)

UFSとZFSのファイルシステム構成の違い



- ファイルシステムごとにボリューム管理ソフト (GDS、SVMなど)による設定が必要。
- ファイルシステムのサイズ変更は、システムを停止しバックアップ/リストア。
- 設計時にパーティション構成やファイルシステムサイズを決める必要がある。
- 急なシステムダウンによりデータの不整合が発生することがある。



- ストレージプールによるディスク一元管理。ボリューム管理ソフトは不要。
- オンラインのままディスク増設可能。
- 設計時にパーティション構成やファイルシステムサイズを決める必要はない。
- 急なシステムダウンでもデータの不整合は発生しない。

<参考>UFSファイルシステムの作成

■ UFSファイルシステムは物理ディスクの領域と1対1に対応

4. ファイルシステムのマウント(mount)

- ・適切なマウントポイントに割り当てる

3. ファイルシステム作成(newfs、fsck)

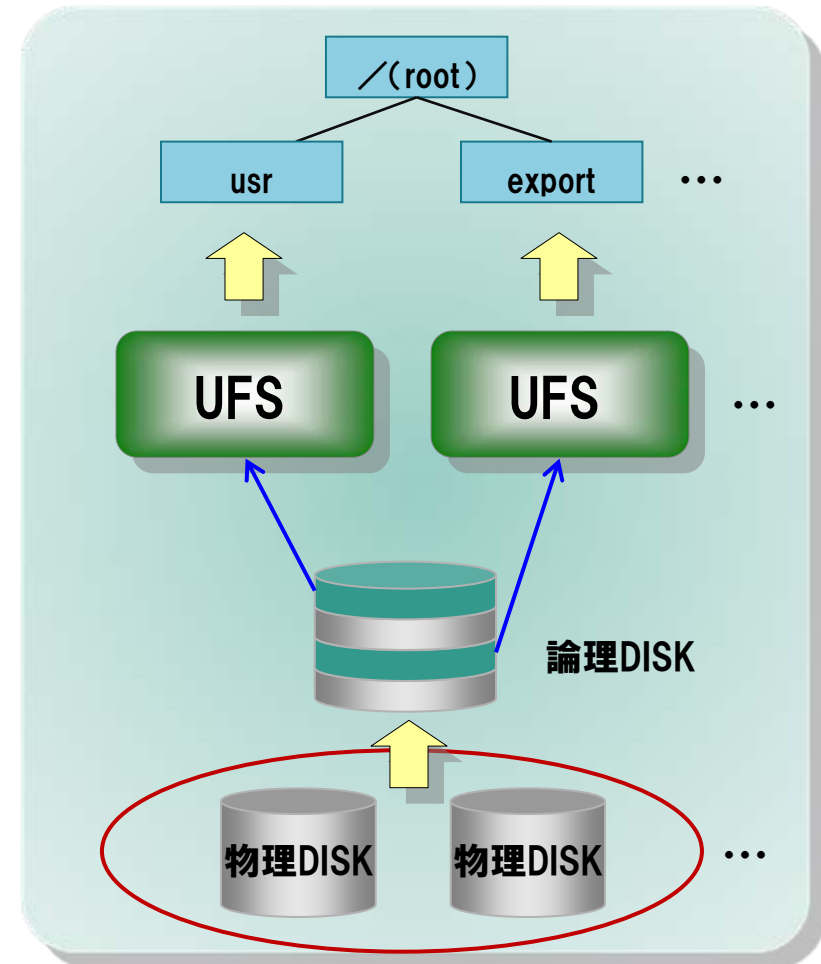
- ・パーティション単位にファイルシステムを作成し、整合性チェックを行う

2. パーティション情報作成(format)

- ・ディスク上の使用する区画(容量)を決める

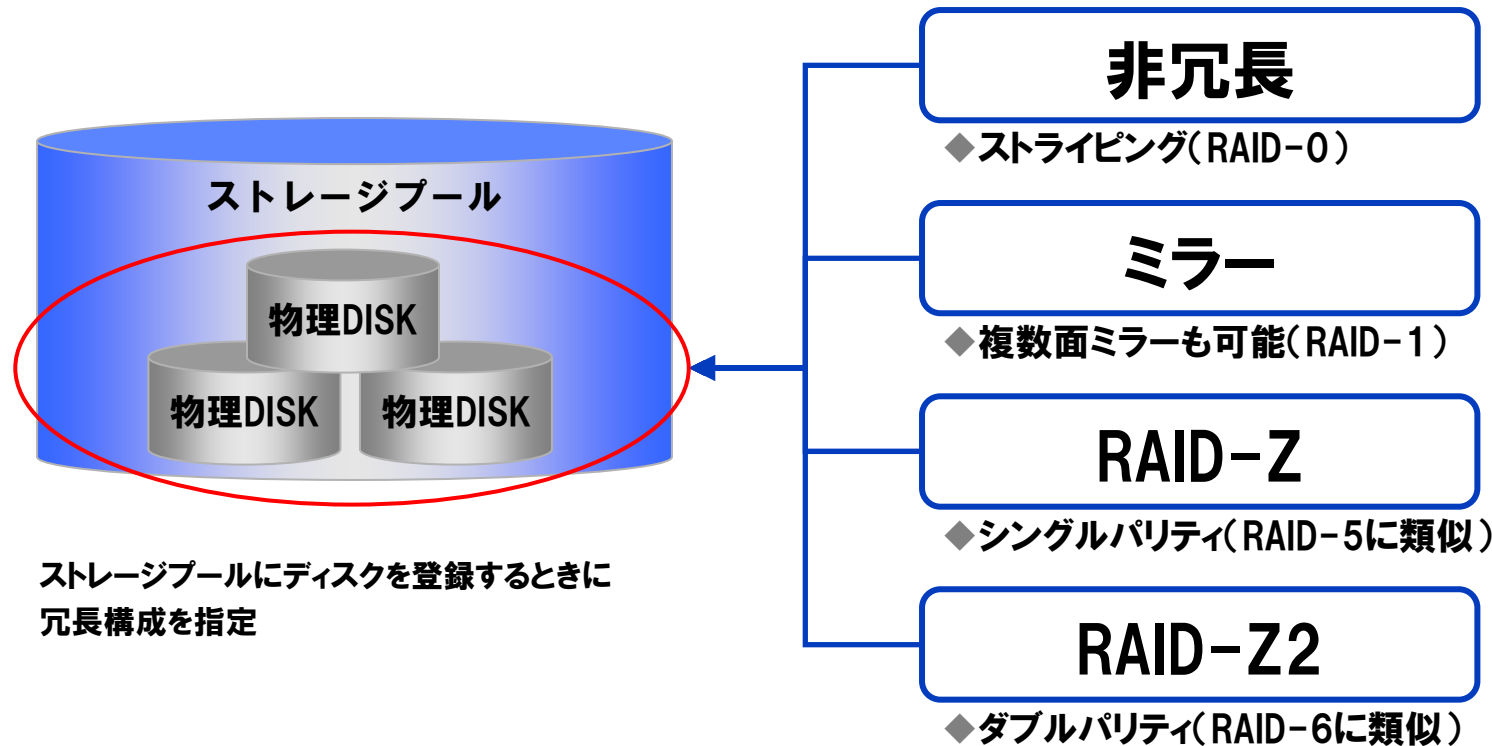
1. 物理ディスクの追加・認識

- ・ボリューム管理ソフトによる冗長構成や論理ディスク作成を行う



データの堅牢性(1)

ZFSでサポートするディスクの冗長構成



ストレージプールにディスクを登録するときに冗長構成を指定

データの読み書き時は、チェックサムによって不正データが検出されます。不正データが検出された場合、冗長構成(ミラー、RAID-Z、RAID-Z2)であれば複製データから自動的にデータを修復します(Self-Healing)

データの堅牢性(2)

堅牢性に優れたファイルシステムの仕組み

トランザクションファイルシステム

- ✓データの書き込みは「コピー・オン・ライト」の方式で管理
- ✓元のデータが上書きされることはない
- ✓一連の処理が完全に確定するか、完全に無視される

データの整合性を保証

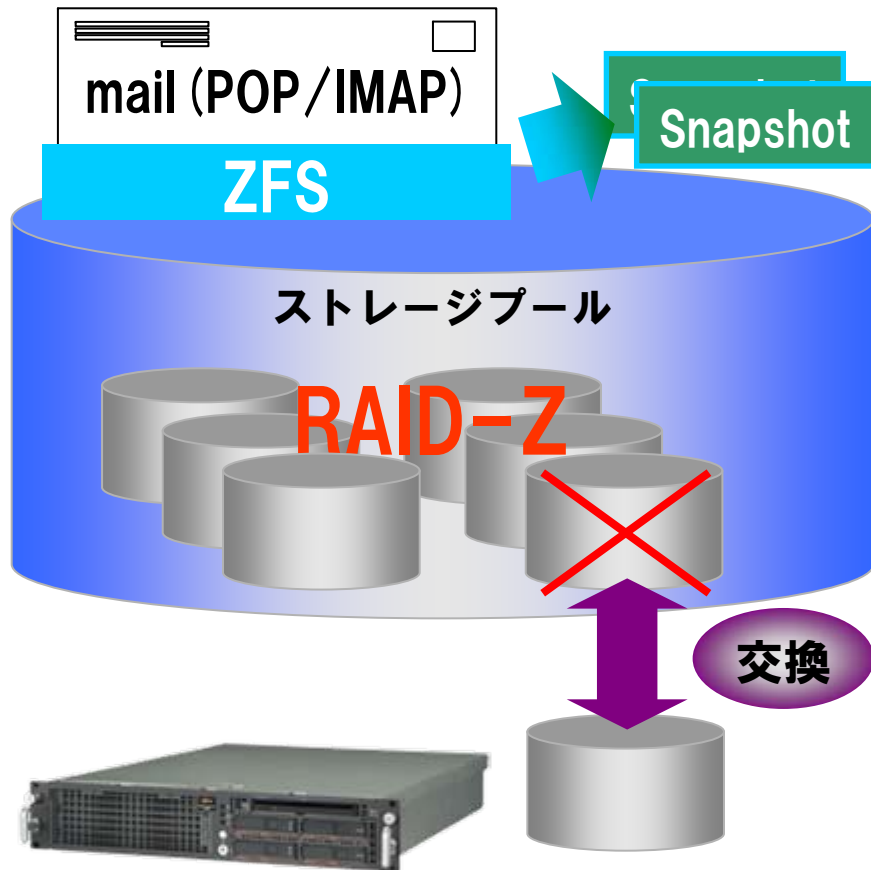
End-to-Endのチェックサム

- ✓データブロックのチェックサムは親ブロックに格納
- ✓異常検出時は冗長化したブロックからデータを回復
- ✓論理的な不整合(ソフトのバグ)でも検出・訂正

不正データの検出
データの自己修復

適用事例：弊社部門メールサーバ

OS標準機能のみで高信頼なメールサーバを構築



<特長>

- RAID-Zでストレージプールを作成し、信頼性とディスク容量を確保
- スナップショットによる世代管理
→作成時間を感じないほど短時間

<運用時>

- ディスク障害時は縮退して業務を継続
- オンラインでディスク交換を実施

商標について

- SPARC Enterpriseは、米国SPARC International, Inc.のライセンスを受けて使用している、同社の米国およびその他の国における商標または登録商標です。
- UNIXは、米国およびその他の国におけるオープン・グループの登録商標です。
- Sun、Sun Microsystems、Sunロゴ、SolarisおよびすべてのSolarisに関連する商標及びロゴは、米国およびその他の国における米国Sun Microsystems, Inc.の商標または登録商標であり、同社のライセンスを受けて使用しています。
- すべての SPARC 商標は、SPARC International, Inc. のライセンスを受けて使用している同社の米国およびその他の国における登録商標です。SPARC 商標が付いた製品は、Sun Microsystems, Inc. が開発したアーキテクチャーに基づくものです。
- SPARC64 は、米国 SPARC International, Inc. のライセンスを受けて使用している同社の登録商標です。
- ORACLE, SQL * Plus, SQL * Forms, SQL * Net, Pro * C, Pro * FORTRAN, Pro * COBOLは、ORACLE Corporationの登録商標もしくは商標です。
- その他各種製品名は、各社の製品名称、商標または登録商標です。

留意事項

- 本書の内容は、改善のため事前連絡なしに変更することがあります。
- 本書の内容は、細心の注意を払って制作致しましたが、本書中の誤字、情報の抜け、本書情報の使用に起因する運用結果に関しましては、責任を負いかねますので予めご了承願います。
- 本書に記載されたデータの使用に起因する第三者の特許権およびその他の権利の侵害については、当社はその責を負いません。
- 無断転載を禁じます。

