

White paper FUJITSU Storage ETERNUS AF series, ETERNUS DX series 重複排除／圧縮機能による ディスク容量効率化

FUJITSU Storage ETERNUS AF series, ETERNUS DX series は、重複排除／圧縮機能をサポートしています。本書では、重複排除／圧縮機能の概要と導入における検討ポイントについて解説するとともに、重複排除／圧縮機能を適用した容量削減の効果について説明します。



目次

はじめに	4
1. 重複排除／圧縮機能の概要	5
1.1. 重複排除と圧縮とは	5
1.2. ETERNUS AF/DX の重複排除／圧縮機能の特長	8
1.2.1. 重複排除／圧縮機能を適用した運用例	9
1.3. 重複排除／圧縮機能を導入前に検討すべき事項	11
1.4. 必要条件	12
1.5. 留意事項	13
2. 検証	14
2.1. 重複排除／圧縮機能を適用した場合のアクセス性能検証	14
2.1.1. 検証目的	14
2.1.2. 検証内容	14
2.1.3. 検証環境	14
2.1.4. 前提条件	15
2.1.5. 検証結果	15
2.1.6. 考察	15
2.2. データ用パーティションの容量削減率検証	16
2.2.1. 検証目的	16
2.2.2. 検証内容	16
2.2.3. 検証環境	17
2.2.4. 前提条件	18
2.2.5. 検証結果	18
2.2.6. 容量削減率の表示例	19
2.2.7. 考察	19
3. まとめ	20

図表目次 1

図 1 重複排除の動作イメージ	5
図 2 プリ・プロセス方式／ポスト・プロセス方式／インライン方式	6
図 3 圧縮の動作イメージ	7
図 4 重複排除／圧縮機能イメージ	8
図 5 重複排除／圧縮機能の構成管理	9
図 6 削減効果が高い業務データの物理容量削減	9
図 7 バックアップ環境における退避データの物理容量削減	10
図 8 バックアップボリュームの途中世代削除	10
図 9 故障した業務ボリュームのリストア	11
図 10 重複排除／圧縮機能を有効／無効に設定した際の性能検証	14
図 11 データ用パーティションのディレクトリ 1 つに格納	16
図 12 データ用パーティションのディレクトリ 2 つに同じファイルを格納	17
図 13 データパーティションの容量削減率 検証構成	17
図 14 容量削減率の表示例 (参考画像につき、検証時の構成や値とは関係がありません)	19

図表目次 2

表 1 固定長ブロック単位方式	5
表 2 可変長ブロック単位方式	5
表 3 主な各社製品の重複排除方式一覧	6
表 4 重複排除／圧縮機能の仕様一覧	9
表 5 Deduplication/Compression を有効に設定できる最大 TPP 数一覧	12
表 6 Deduplication/Compression システムボリュームの仕様一覧	13
表 7 アクセス性能検証の前提条件	15
表 8 標準的なシステム(Read: Write 比率 3:1)の測定結果	15
表 9 主に小さいデータをアクセスするシステム(Read: Write 比率 3:1)の測定結果	15
表 10 データ種別ファイル一覧	16
表 11 容量削減率検証の前提条件	18
表 12 パーティションのディレクトリ 1 つ(F:¥test)に格納した際の測定結果	18
表 13 パーティションのディレクトリ 2 つ(F:¥test と、F:¥test - Copy)に同じファイルを格納した際の測定結果	18

はじめに

近年多くの企業は、ビッグデータに代表されるようなビジネスデータの飛躍的な増加に対して大きな問題を抱えています。さらにストレージの初期投資や運用コストなどの費用負担も、同様に増加傾向をたどっています。システムに関する IT コストを削減するというニーズは、多くの企業にとって急務となっています。そこでシステムのリソースを最適化する手段として、重複排除と圧縮のテクノロジーが注目されています。先進的な重複排除と圧縮のテクノロジーを活用することで、加速しながら蓄積するデータを効率良くコンパクトに保管することができて、システムの TCO 削減にもつながります。

本書では、FUJITSU Storage ETERNUS AF series, ETERNUS DX series の重複排除／圧縮機能を適用した保管容量の効率化が、どの程度効果として見込めるのかを検証結果を交えて説明します。
なお、検証は 2016 年 9 月に実施した内容を記載しています。

■ 対象読者

本書では以下の方を対象にしています。

- ・FUJITSU Storage ETERNUS AF series, ETERNUS DX series で導入・運用コストの削減を検討している方
- ・FUJITSU Storage ETERNUS AF series, ETERNUS DX series の重複排除／圧縮機能を使って、どのように保管容量が効率化できるかを知りたい方

■ 対象機種

本書は、以下のストレージシステムを対象に記載しています。

- ・FUJITSU Storage ETERNUS AF series ※1

※1 重複排除／圧縮機能は、ETERNUS AF series のファームウェア V10L60 以降でサポートされます。

- ・FUJITSU Storage ETERNUS DX S3 series 以降 ※2 ※3 ※4 ※5

※2 ETERNUS DX60 S4/S3, DX100 S4/S3, DX8100 S4/S3, DX8700 S3 は対象外です。

※3 ETERNUS DX series で重複排除／圧縮機能を使用するには機能拡張メモリの搭載が必要です。

※4 重複排除／圧縮機能は、ETERNUS DX S3 series のファームウェア V10L60 以降でサポートされます。

※5 ETERNUS DX8900 S4 は圧縮機能のみサポートされます。

■ 略称

本書では以下の略称を用います。

- ・ETERNUS AF series.....FUJITSU Storage ETERNUS AF series
- ・ETERNUS DX seriesFUJITSU Storage ETERNUS DX series
- ・ETERNUS AF/DXETERNUS AF series と ETERNUS DX series
- ・TPPThin Provisioning Pool
- ・TPVThin Provisioning Volume
- ・重複排除／圧縮機能ETERNUS AF/DX の重複排除／圧縮機能
- ・Deduplication/Compression ボリューム.....Deduplication/Compression 設定を有効にした TPP 内で
Deduplication/Compression 設定を有効にした TPV
- ・Deduplication/Compression 無効ボリューム.....Deduplication/Compression 設定を有効にした TPP 内で
Deduplication/Compression 設定を無効にした TPV
- ・HDDハードディスクドライブ

■ 前提

本書は、2018 年 10 月時点の製品ラインナップ・製品情報で記載しています。

1. 重複排除／圧縮機能の概要

本章では、一般的な重複排除と圧縮の技術について解説しながら、重複排除／圧縮機能の概要を説明します。

1.1. 重複排除と圧縮とは

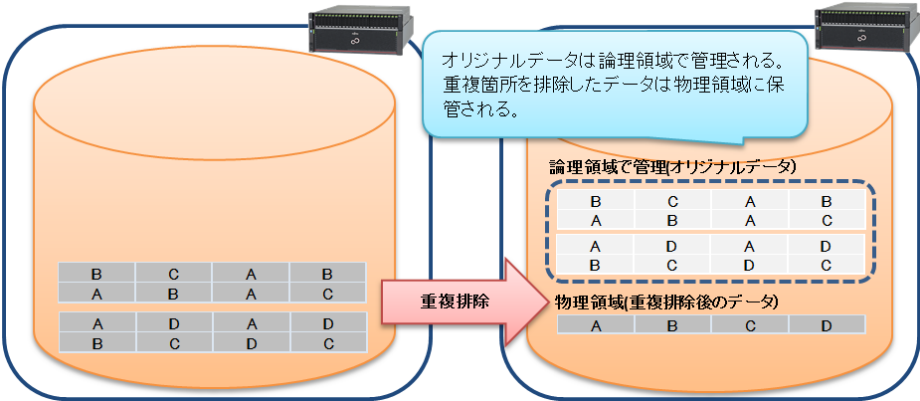
一般的な企業において増加するデータ量は、年次で 50%～60%といわれています。またデータ資産を保全する必要性やリスク回避の面からも、保管データの取り扱いが重要視されています。しかし膨大なデータをバックアップするには、IT コストの面で大きな負担となります。そこで、データを効率良く保管する手段として重複排除と圧縮の技術が注目されています。

重複排除と圧縮は、ともにデータ量を削減するという点で同じ用途で使用する機能ですが、原理や動作はそれぞれ異なります。まずは一般的な重複排除と圧縮の技術について説明します。

重複排除とは、同じ内容が複数存在するデータに対して自動的に重複箇所を解析して排除する技術です。重複排除機能を使用することで、バックアップ時のデータ転送量や格納容量は大幅に削減できます。

図 1 は重複排除の動作を示したものです。重複排除は対象データを小さなブロック単位に分割して比較しながら解析します。そして、変更のあったブロックは物理領域に保存され、重複箇所と認識されたブロックは物理領域には保存されず論理領域に元データと紐付けだけを行います。結果として、重複箇所が排除されたデータだけを物理領域に保存するので、データ容量を抑えて保管することができます。

図 1 重複排除の動作イメージ



重複排除で重要な要素となるのが「ブロック単位の大きさ」と、「プロセスを実行する場所とタイミング」です。この 2 点の組み合わせ方で、重複排除の効率は大きく変化します。

< 重複箇所を検出するブロック >

■ 固定長ブロック単位方式

書き込みされたデータを、固定されたブロック単位 (4KB, 8KB, 16KB, 128KB など) で分割して重複排除する方式です。ブロックのサイズを小さくすると詳細に比較することができて、重複排除の効果は格段に上がります。しかし、データ読み込みや復元をする際にハッシュ値を計算するため、システム負荷が高くなる傾向があります。

表 1 固定長ブロック単位方式

データ								統一したブロックサイズ (4KB) で検索
A	B	C	D	E	F	G	H	
4KB	4KB	4KB	4KB	4KB	4KB	4KB	4KB	

■ 可変長ブロック単位方式

書き込みされたデータを、数 KB～数百 KB の幅をもたせたサイズに分割して重複排除する方式です。固定長ブロック単位の仕組みに比べて多くの重複箇所を検出できます。しかし、可変長の各ブロックの終わりを検出する仕組みによっては、重複箇所を検出処理できない特殊なケースが頻出する場合があるため、重複排除の処理速度が遅くなる可能性があります。

表 2 可変長ブロック単位方式

データ				複数のブロックサイズ (4KB/8KB/16KB) で検索
A	B	C	D	
4KB	8KB	16KB	4KB	

＜重複排除を実行するプロセス＞

■ プリ・プロセス方式

サーバ側で重複箇所を検出して排除する方式です。
重複排除後のデータを転送するためデータ転送量は小さくなり、データ転送に必要な帯域を確保するための回線費用が削減できます。
ただし、サーバには大きな負荷がかかります。

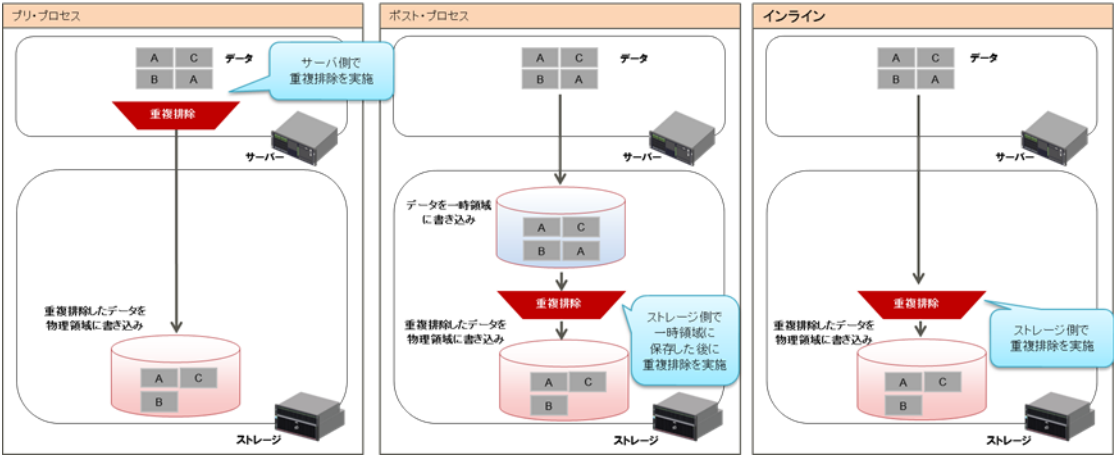
■ ポスト・プロセス方式

ストレージ側で重複箇所を検出して排除する方式です。
データを一時領域に保存した後に重複箇所を検出するため、I/O と検出処理は同期せずパフォーマンスへの影響は小さくなります。
ただし、重複排除前のデータを転送するためデータ転送量は大きく、I/O 負荷も高くなります。
また、一時保存領域が必要なため初期投資のディスクドライブの容量や数量が増え、重複排除処理の開始時間や終了時間、一時領域に必要なディスク容量を考慮した運用設計が必要となります。

■ インライン方式

サーバからデータ転送時にストレージ側で重複箇所を検出して排除する方式です。
ポスト・プロセス方式のような一時領域が不要なため、導入や運用の設計が容易になります。
ただし、重複排除前のデータを転送するためデータ転送量は大きく、I/O 負荷も高くなります。

図 2 プリ・プロセス方式／ポスト・プロセス方式／インライン方式



各社製品の重複排除方式は、以下のとおりです。

表 3 主な各社製品の重複排除方式一覧

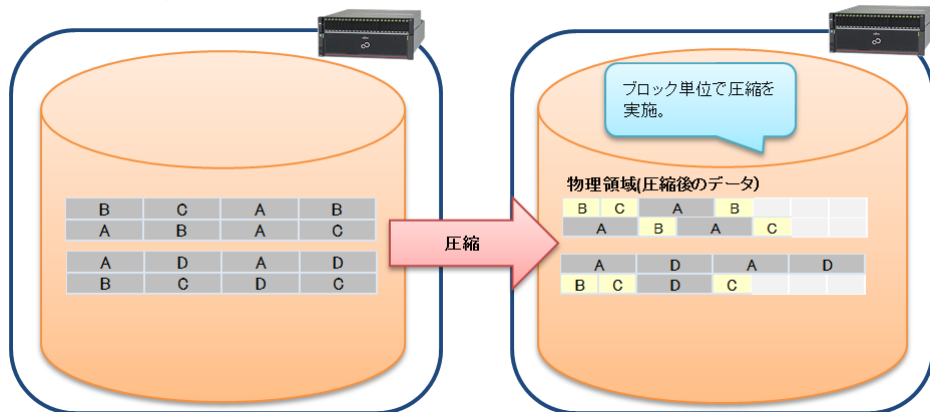
製品	重複排除を検出するブロック	重複排除を実行するプロセス
ETERNUS AF/DX	固定長ブロック単位方式(4KB)	インライン方式
NetApp FAS2500 シリーズ	固定長ブロック単位方式(4KB)	ポスト・プロセス方式
EMC VNX シリーズ	固定長ブロック単位方式(8KB)	ポスト・プロセス方式
EMC XtremIO	固定長ブロック単位方式(4KB)	インライン方式
Pure Storage FlashArray//m シリーズ	固定長ブロック単位方式(512byte)	インライン方式
HP 3PAR StoreServ 7450 Storage	固定長ブロック単位方式(16KB)	インライン方式

次に圧縮について説明します。

圧縮とは、データの実質的な性質を保ったまま、データ量を減らして別のデータに変換します。

圧縮の機能と聞いて想像するのは「ZIP」や「LHA」に代表されるファイル圧縮機能ですが、ストレージにおける圧縮の機能は全くの別物です。ストレージにおける圧縮の機能とは、小さなデータに分割してブロック単位で圧縮することです。圧縮の機能により保存データが小さくなると、ストレージの容量効率化が期待できます。

図 3 圧縮の動作イメージ



重複排除と圧縮の大きな違いは、重複排除はストレージ内の重複するデータブロックを削除する機能であることに対して、圧縮は異なるデータブロックをまとめて圧縮する機能であることです。また、重複排除と圧縮はそれぞれで容量削減の効果が高いデータが異なるため、両方の機能を効率良く使うことでより削減効果が高められます。

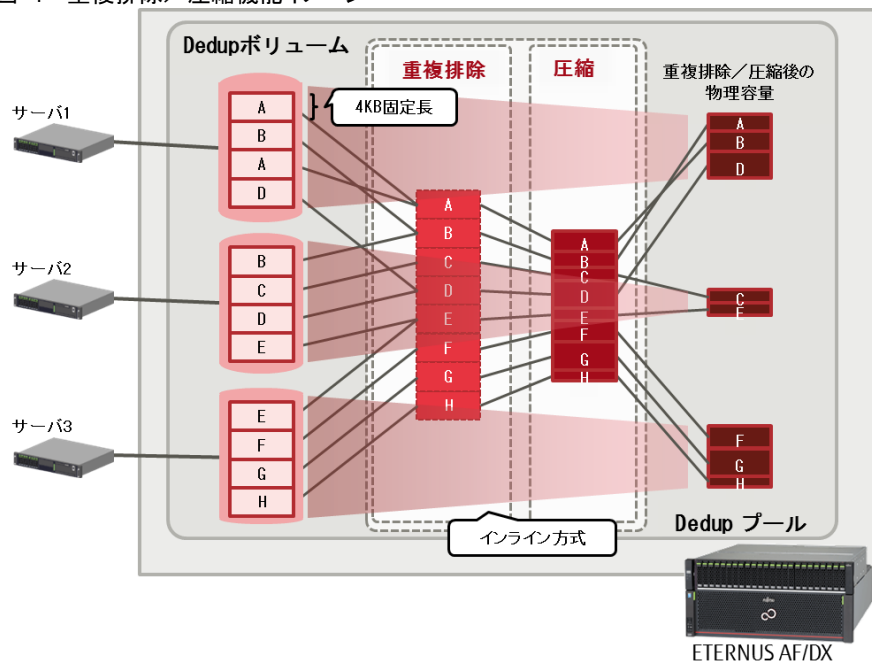
1.2. ETERNUS AF/DX の重複排除／圧縮機能の特長

ETERNUS AF/DX は重複排除の方式として、ブロックサイズの小さい 4KB の「固定長ブロック単位方式」と「インライン方式」を採用しています。インライン方式により一時領域を用意する必要がないため、導入や運用の設計が容易で初期投資のコストが抑えられます。

重複排除と圧縮の 2 つの機能を提供することで、ディスク容量のさらなる効率化を実現しています。

また、ETERNUS AF/DX は、重複排除機能のみ、または圧縮機能のみを単独で利用できます。格納するデータに対して効果的な機能を選択利用することで、より効率的にストレージシステムを運用できます。

図 4 重複排除／圧縮機能イメージ



重複排除／圧縮機能はストレージの容量削減を目的としているため、TPP をベースとした内部構造において物理容量の効率化を重視しています。

重複箇所の特定期間は、Deduplication/Compression 設定を有効にした TPP 内にあるすべての Deduplication/Compression ボリュームのデータです。重複箇所の特定期間には、「SHA-1」のハッシュアルゴリズムを利用します。

「SHA-1」は、世界的標準規格のハッシュ関数で、MD5 などと比べると走査時間はかかるものの判定精度は高くなります。

重複排除の方法は、まずデータをブロックごとに分割して、各ブロックのハッシュ値を計算します。この値が同じ場合は同一ブロックと判定し、既にそのブロックが物理領域に格納済みであれば、新たに物理領域には保存せず構成情報のみを管理します。

重複排除されたデータが呼び出された際は、構成情報に基づいて必要なブロックを集めて復元します。

■ 重複排除／圧縮機能の構成管理

重複排除／圧縮機能は、プール単位で重複排除／圧縮機能の構成情報を管理します。

そのため、ETERNUS Web GUI や ETERNUS CLI で表示される容量削減率も TPP 単位で表示します。

(以降、Deduplication/Compression 設定を有効にした TPP 単位の容量削減率を「容量削減率」とします)

容量削減率は、例えば 10TB のデータを圧縮して 3TB で書き込んだ場合「70%」と表示されます。

Deduplication/Compression 設定を有効化すると、サーバからは見えない内部ボリューム(以降、Deduplication/Compression システムボリューム)が自動的に作成されます。

Deduplication/Compression システムボリュームは重複排除／圧縮した後の物理情報や構成情報を管理する内部ボリュームです。

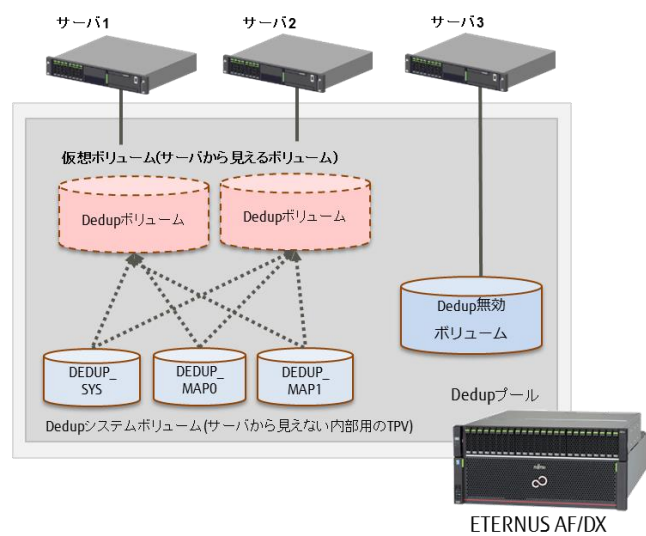
Deduplication/Compression システムボリュームには、DEDUP_SYS ボリュームと DEDUP_MAP ボリュームがあります。

- ・ DEDUP_SYS ボリューム
重複排除／圧縮機能を実施した実データおよびメタデータが格納されるボリュームです。
- ・ DEDUP_MAP ボリューム
Deduplication/Compression ボリュームのボリューム番号(または LBA)から、DEDUP_SYS ボリュームの物理位置情報を求めるためのマッピングテーブルが格納されるボリュームです。

その後、Deduplication/Compression ボリューム(サーバから見える仮想的なボリューム)を作成します。

Deduplication/Compression ボリュームは重複排除／圧縮する前の仮想ボリュームです。

図 5 重複排除／圧縮機能の構成管理



Deduplication/Compression ボリュームは、Deduplication/Compression 設定を有効にした TPP 内にボリュームを作成する際、Deduplication/Compression 設定を有効にすることで作成できます。
Deduplication/Compression ボリュームと Deduplication/Compression 無効ボリュームは混在できます。

表 4 重複排除／圧縮機能の仕様一覧

	ETERNUS AF/DX
ブロック長方式	固定長方式 (4KB)
プロセス方式	インライン方式
設定範囲	プール単位
重複チェック対象	Deduplication/Compression 設定を有効にした 1 つの TPP 内の、全 Deduplication/Compression ボリュームのデータ

1.2.1. 重複排除／圧縮機能を適用した運用例

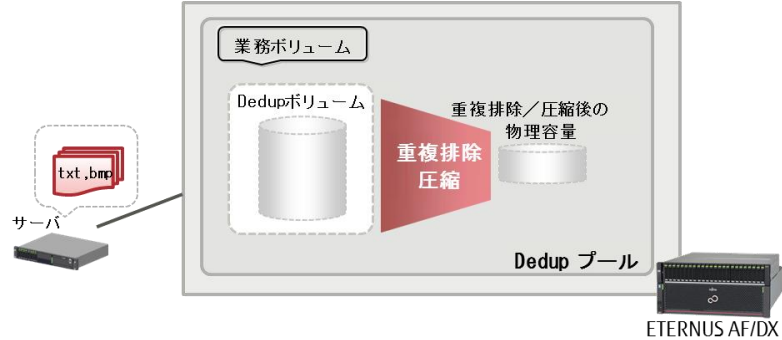
重複排除／圧縮機能を適用して削減効果が期待できる運用環境としては以下があります。

- ・ 文書データやテキストデータを扱うことが多い環境
- ・ OPC などを使ったフルバックアップの世代数が多い環境
- ・ フルバックアップの対象ボリュームに対してデータ更新が少ない環境

【運用例 1 効果が高い業務における保管容量の効率化】

文書データやテキストデータを扱うことが多い業務では、重複排除／圧縮機能の効果が高く、保管容量を大きく削減することが期待できます。
また、データ参照を主とする業務では、重複排除／圧縮機能のデータ読み込みに対するパフォーマンスへの影響は小さいため、大容量の参照データを効率的に保管することができます。

図 6 削減効果が高い業務データの物理容量削減



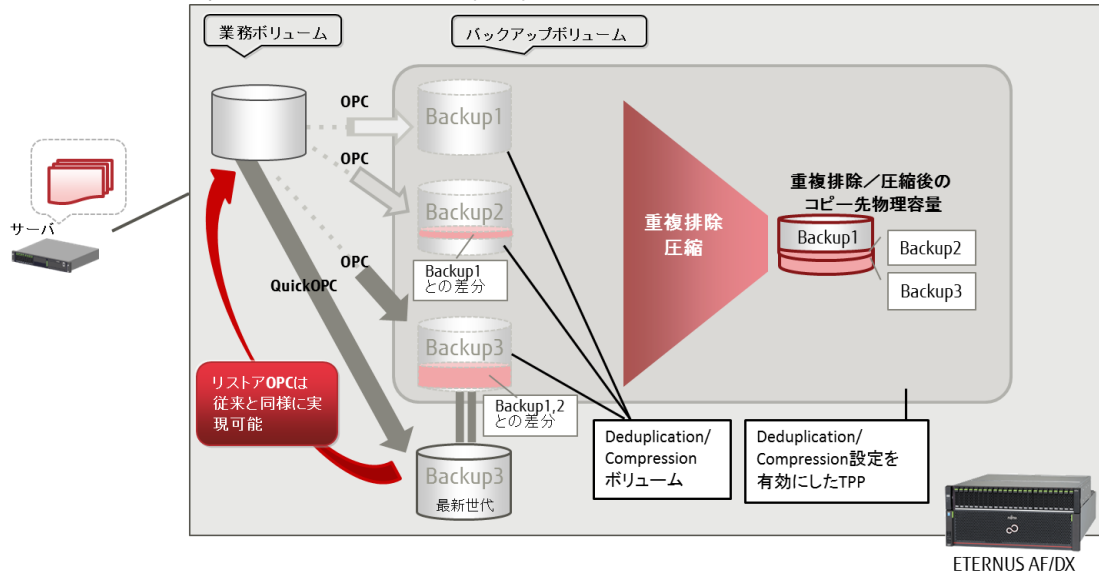
【運用例 2 OPC を使ったバックアップ環境における保管容量の効率化】

OPC を使ったバックアップ環境に重複排除／圧縮機能を適用することで、退避データの保管容量を大きく削減することが期待できます。

図 7 は、OPC によるバックアップを 3 世代保存することを前提としたものです。従来はコピー先容量を世代数分必要としていましたが、重複排除／圧縮機能を適用した場合は格段に少ない容量となり、従来の運用と比べて大きな削減効果があります。

最新バックアップ先を QuickOPC による差分コピーにすることで、リストア OPC の応答時間だけでなく、物理的なデータ復元時間も従来と同じようになり、迅速に復旧することができます。

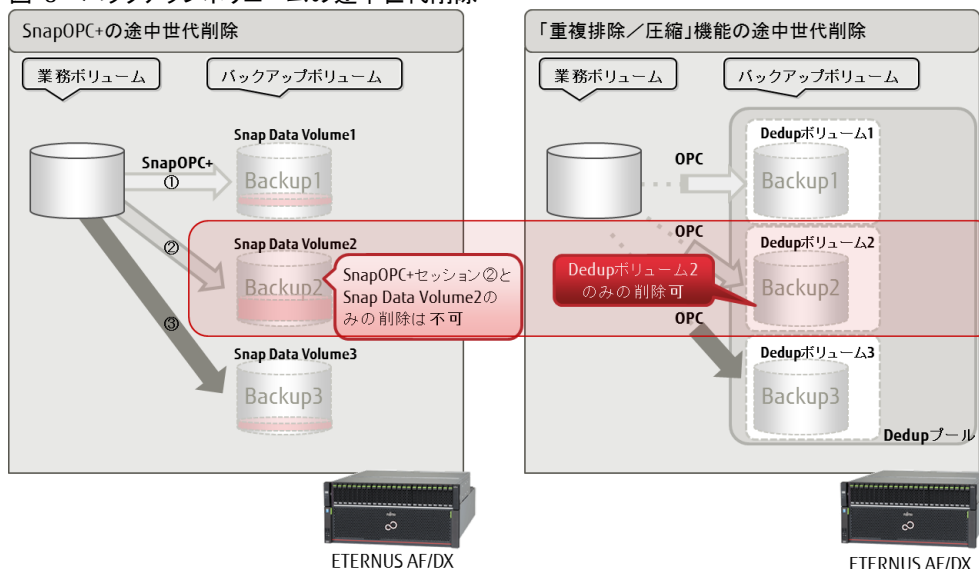
図 7 バックアップ環境における退避データの物理容量削減



また、重複排除／圧縮機能を使った OPC のバックアップは、SnapOPC+を使った世代バックアップと比べても、以下のメリットがあります。

- ・ SnapOPC+では、バックアップを時系列で差管理するため、容量枯渇や SnapOPC+世代数の制限、資産整理などの理由でユーザーが不要と判断した世代があっても、該世代より前の世代を全削除するか、すべての世代を残すことしかできません。しかし重複排除／圧縮機能を使った OPC では、すべての世代を常にフルバックアップで保持するため、不要となった世代は削除できます。これにより、ユーザーは容量枯渇や資産整理などの理由で不要と判断した世代を削除することが可能となり、より自由にバックアップボリュームを管理できるようになります。

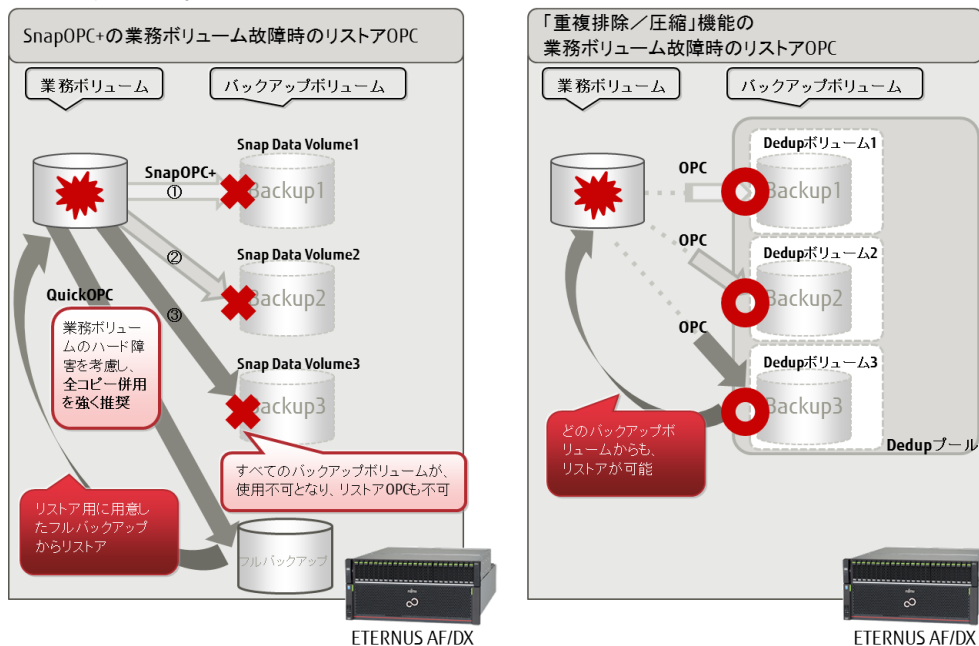
図 8 バックアップボリュームの途中世代削除



- ・ SnapOPC+では、業務ボリュームが故障するとバックアップボリュームも使用不可になってしまいます。そのため、リストア用に別途 QuickOPC を併用してフルバックアップを保持し、フルバックアップからリストア OPC で復旧する必要があります。

しかし重複排除／圧縮機能を使った OPC では、業務ボリュームが故障してもバックアップボリュームは使用可能です。そのため、どのバックアップボリューム先からもフルバックアップでリストアすることができます。これにより、ユーザーは業務ボリュームの故障を想定した運用設計が不要となります。

図 9 故障した業務ボリュームのリストア



1.3. 重複排除／圧縮機能を導入前に検討すべき事項

重複排除／圧縮機能は保管容量の削減に効果をもたらす場合がある一方で、パフォーマンスの低下を伴います。そのため重複排除／圧縮機能を導入するにあたり、この2点の影響バランスを考慮することが重要です。まずは該当システムにもたらす削減効果とパフォーマンス低下が、許容範囲であるかどうかを見極めたうえで適用してください。重複排除／圧縮機能に影響するものとして、以下があります。

■ データ

データの変更型（ファイル追加・データ更新・データ削除）、データの変更割合、ファイルのサイズ、ファイルの可能圧縮率などがあります。複数の仮想マシンシステムディスクイメージやメールデータは、一般的にある程度決まった構造や形式のため冗長化したデータが多く存在します。その結果、重複箇所を検知する割合が高くなるので、削減効果が期待できます。また、データの種別では、非圧縮フォーマットのデータ（テキスト、静止画像、音声）は圧縮率が高く、削減効果が期待できます。反対に、データの変更頻度が高い場合や、圧縮フォーマットのデータ（LZH、ZIP、CAB）を多く使用する環境には不向きです。

■ 性能

データへのアクセス頻度やI/O（シーケンシャルやランダム、サイズやパターン）などがあります。世代管理されたバックアップ環境などは、パフォーマンスへの影響が小さいため適しています。バッチ処理やオンライン業務環境などは、パフォーマンスへの影響が大きいため不向きです。

■ ハードウェア

CPU 使用率やキャッシュメモリサイズ、ディスクドライブの種類や速度などがあります。重複排除／圧縮機能はストレージに大きなパフォーマンス負荷を強いるため、以下の必要条件を参照して最適な構成設計の検討や対象機種の見積もりを行ってください。

1.4. 必要条件

重複排除／圧縮機能には、いくつかの必要条件があります。
以下の条件を守り、最適な構成を検討してください。

■ Deduplication/Compression モード

重複排除／圧縮機能を適用するには、ETERNUS AF/DX に対して重複排除／圧縮機能を設定する必要があります。
ETERNUS AF/DX の Deduplication/Compression モードを有効に設定してください。

■ 機能拡張メモリ

ETERNUS DX series は、重複排除／圧縮機能を利用するために機能拡張メモリを追加する必要があります。
機種ごと、また機種に搭載されているキャッシュメモリ容量により、必要な機能拡張メモリ容量が異なるため、最新の情報をご確認ください。

■ 対応ボリュームタイプ

TPV(SAN ボリュームのみ対象。NAS ボリュームは対象外)

■ Deduplication/Compression を有効に設定できる最大 TPP 数

Deduplication/Compression を有効に設定できる最大 TPP 数は、機種別ごとに上限があります。
表 5 を参照して Deduplication/Compression を有効に設定できる最大 TPP 数を確認してください。

表 5 Deduplication/Compression を有効に設定できる最大 TPP 数一覧

	ETERNUS AF250 S2/ AF250	ETERNUS AF650 S2/ AF650	ETERNUS DX200 S4/S3	ETERNUS DX500 S4/S3	ETERNUS DX600 S4/S3	ETERNUS DX8900 S4
Deduplication/Compression を有 効に設定できる最大 TPP 数	4	8	4	4	8	8

1.5. 留意事項

重複排除／圧縮機能には、いくつかの留意事項があります。

■ 既存ボリュームを Deduplication/Compression ボリュームに変更

RAID マイグレーションにより、既存ボリューム (Standard/TPV/FTV/WSV) を Deduplication/Compression ボリュームに変更することができます。また、その逆も可能です。既存ボリュームを Deduplication/Compression ボリュームに変更する場合、移行先には Deduplication/Compression ボリュームを指定してください。同一 TPP 内または別の TPP への変更が可能です。

■ バックアップ

リモートコピー (REC) やバックアップ／リストアにおいて、コピー元が Deduplication/Compression ボリュームの場合は、重複排除／圧縮前のデータが転送されます。

また、コピー先が Deduplication/Compression ボリュームの場合は、重複排除／圧縮機能を実施してデータを格納します。

■ Deduplication/Compression システムボリュームと、同一 TPP 内の Deduplication/Compression ボリュームの総論理容量

Deduplication/Compression システムボリュームには、以下の仕様があります。

また、Deduplication/Compression ボリュームの総論理容量は、以下のように同一 TPP 内の DEDUP_SYS ボリュームの論理容量で制限されます。

- ・TPP が 2RAID グループ以上の構成の場合、DEDUP_SYS ボリュームの論理容量の 10 倍が上限です。
- ・TPP が 1RAID グループの場合、DEDUP_SYS ボリュームの論理容量の 5 倍が上限です。

表 6 を参照して Deduplication/Compression システムボリュームの仕様を確認のうえ、Deduplication/Compression ボリュームの総論理容量の上限を確認してください。

表 6 Deduplication/Compression システムボリュームの仕様一覧

		DEDUP_SYS ボリューム	DEDUP_MAP ボリューム
ボリュームタイプ		TPV	TPV
ボリューム数	ETERNUS DX8900 S4	Deduplication/Compression 設定を有効にした TPP ごとに 1 個	不要
	ETERNUS DX8900 S4 以外		Deduplication/Compression 設定を有効にした TPP 内の RAID グループ数によって異なります。 ^{※6} ※6 TPP 内の RAID グループ数が 1 個の場合 1 個 TPP 内の RAID グループ数が 2 個以上の場合 2 個
ボリュームの論理容量 (1 個あたり)	ETERNUS DX8900 S4	32TB(初期状態)～最大 48PB	不要
	ETERNUS DX8900 S4 以外	8TB(初期状態) ～最大 128TB	5.38TB(固定)

2. 検証

本章では、重複排除／圧縮機能を適用した場合の容量削減効果について、2つの検証結果を示します。

2.1. 重複排除／圧縮機能を適用した場合のアクセス性能検証

本検証は、アクセス元を Windows OS でデータ用パーティション(一般的に Dドライブ以降)と想定しています。このパーティションに、重複排除／圧縮機能を適用した場合のアクセス性能を検証します。

2.1.1. 検証目的

重複排除／圧縮機能を適用した場合の性能データを測定することを目的とします。

2.1.2. 検証内容

パーティションへのアクセスはファイル参照が主体で、一部ファイル更新を行うモデルとして Read: Write 比率を 3:1 と想定します。パーティションに対して標準的なアクセスを行うシステムと、主に小さいデータをアクセスするシステムの2ケースについて測定します。

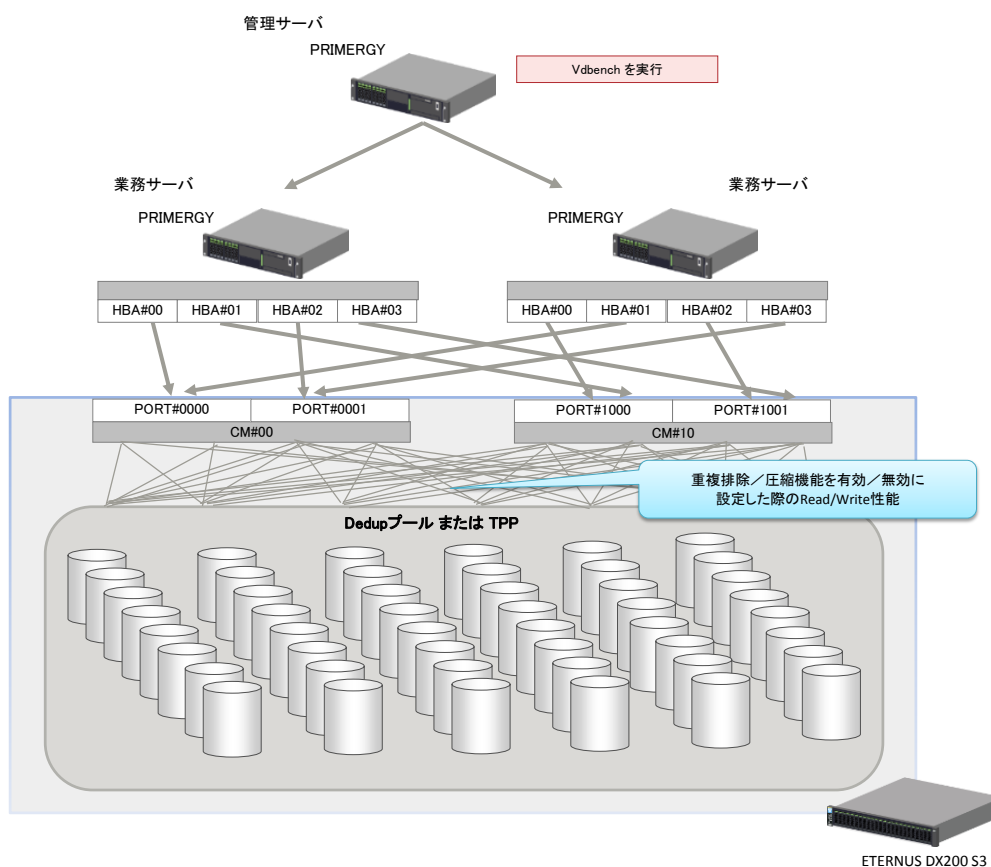
標準的なアクセスを行うシステムは、平均アクセスサイズを 8KB とし、主に小さいデータをアクセスするシステムの平均アクセスサイズは 4KB とします。

重複排除／圧縮機能は保管容量の削減をもたらす一方、I/O の状況によってはパフォーマンス低下のおそれがあるため、重複排除／圧縮機能を有効に設定した場合と、無効に設定した場合の検証結果を示します。

パーティションへのアクセスを行うツールとして、vdbench を使用します。

2.1.3. 検証環境

図 10 重複排除／圧縮機能を有効／無効に設定した際の性能検証



2.1.4. 前提条件

ストレージの設定は、以下のとおりです。

表 7 アクセス性能検証の前提条件

対象のハイブリッドストレージシステム装置	ETERNUS DX200 S3
Deduplication モード	有効
Pool	重複排除／圧縮機能を有効: Deduplication/Compression 設定を有効にした TPP に Deduplication/Compression ボリュームを作成 重複排除／圧縮機能を無効: Deduplication/Compression 設定を無効にした TPP に TPV を作成
RAID	HDD 5 個で構成する RAID5(4+1)を 4 つ作成
Logical Volume	16 個作成
Path	各 CM は 4 Path 構成で 2 つの業務サーバに接続(Total 8Path)

2.1.5. 検証結果

標準的なシステムの場合

表 8 標準的なシステム(Read: Write 比率 3:1)の測定結果

重複排除／圧縮機能	平均アクセスサイズ	スループット	IOPS	容量削減率
有効	8K	14.1MB/s	1803	60%
無効	8K	26.8MB/s	3425	0%

主に小さいデータをアクセスするシステムの場合

表 9 主に小さいデータをアクセスするシステム(Read: Write 比率 3:1)の測定結果

重複排除／圧縮機能	平均アクセスサイズ	スループット	IOPS	容量削減率
有効	4K	9.6MB/s	2457	60%
無効	4K	15.2MB/s	3890	0%

2.1.6. 考察

検証の結果から、重複排除／圧縮機能を適用しても、容量削減効果に対して十分な性能を示しています。容量の削減率が高いファイルを多く保管している環境で、容量効率化を重視するユーザーに有用な手段です。重複排除／圧縮機能の適用は、容量効率化と性能のバランスを考慮してご検討ください。

2.2. データ用パーティションの容量削減率検証

本検証は、Windows OS でデータ用パーティション(一般的に D ドライブ以降)を想定し、このパーティションに、複数のデータ種別ファイルを格納し、容量削減率を検証します。

2.2.1. 検証目的

重複排除／圧縮機能を適用して、複数のデータ種別ファイルを格納した際の容量削減率を測定します。

2.2.2. 検証内容

以下の複数のデータ種別ファイルを、データ用パーティションのディレクトリ 1 つに格納する(重複するファイルが存在しない)ケースと、ディレクトリ 2 つに同じファイルを格納する(ディレクトリのコピーにより、重複するファイルが存在する)ケースの 2 ケースについて容量削減率を測定します。

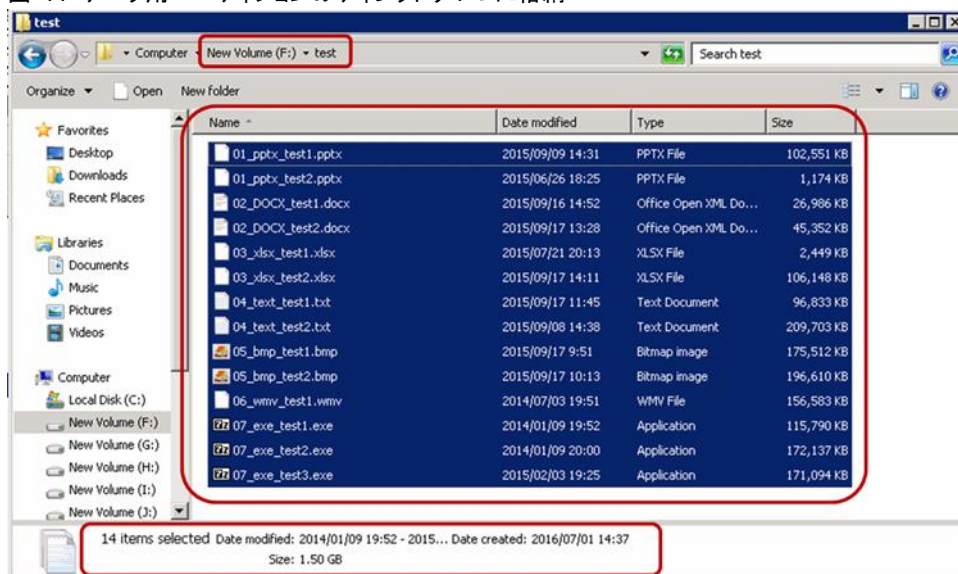
データ種別ファイル 7 種類、総容量は 1,539MB です。データ種別ファイルの詳細と、ファイルサイズは以下のとおりです。

表 10 データ種別ファイル一覧

	データ種別	ファイル	拡張子	ファイル数	ファイルサイズ合計
1	文書データ	パワーポイントファイル	pptx	2	101MB
2		ワードファイル	docx	2	70MB
3		エクセルファイル	xlsx	2	106MB
4		テキストファイル	txt	2	299MB
5	画像データ	ビットマップファイル	bmp	2	363MB
6	動画データ	ビデオファイル	wmv	1	152MB
7	プログラムデータ	実行ファイル	exe	3	448MB
		合計		14	1,539MB

【データ用パーティションのディレクトリ 1 つに格納する(重複するファイルが存在しない)ケース】

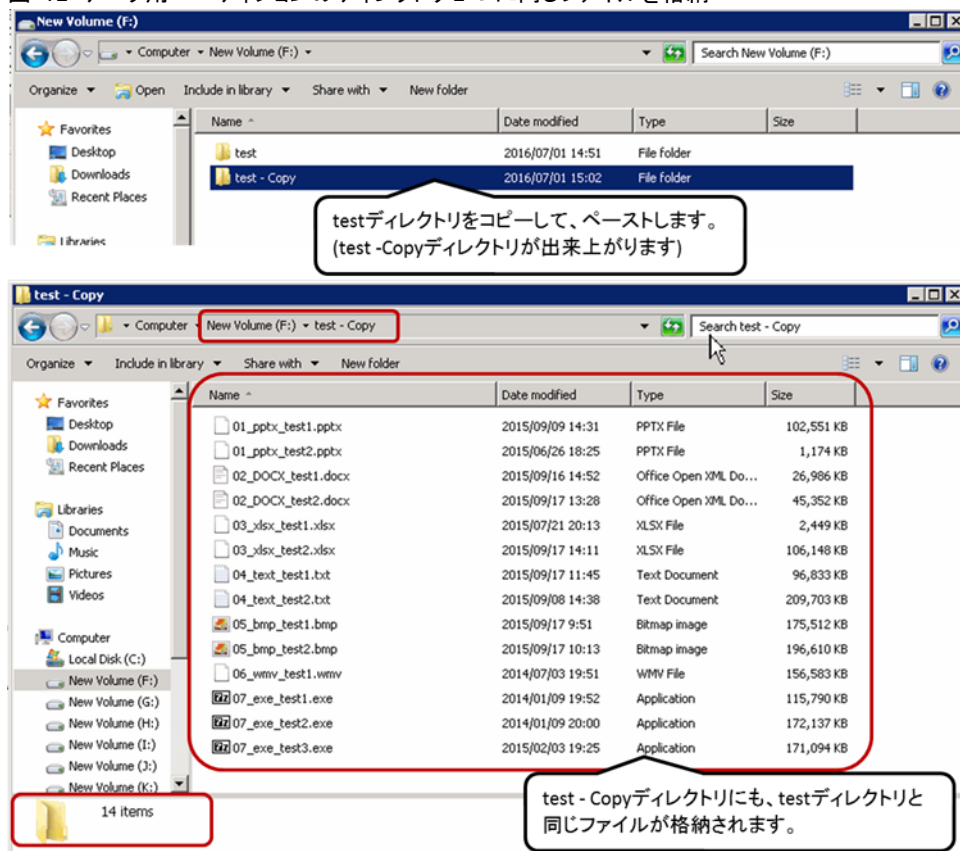
図 11 データ用パーティションのディレクトリ 1 つに格納



データ用パーティション(F:)の test ディレクトリにファイルを格納した状態です。
(合計 14 ファイル、合計ファイルサイズ 1,539MB(1.50GB))

【ディレクトリ 2 つに同じファイルを格納する(ディレクトリのコピーにより、重複するファイルが存在する)ケース】

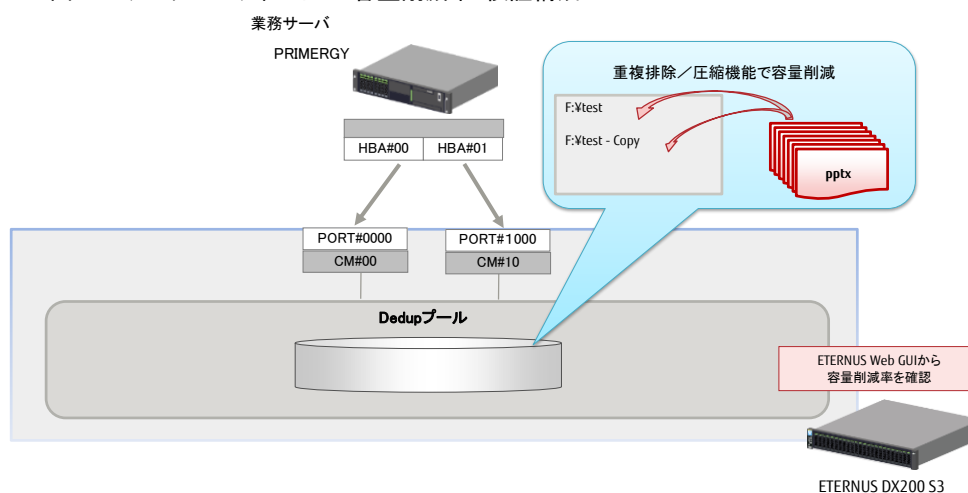
図 12 データ用パーティションのディレクトリ 2 つに同じファイルを格納



データ用パーティション(F:)の test ディレクトリと test - Copy ディレクトリに同じファイルを格納した状態です。

2.2.3. 検証環境

図 13 データパーティションの容量削減率 検証構成



2.2.4. 前提条件

ストレージの設定は、以下のとおりです。

表 11 容量削減率検証の前提条件

対象のハイブリッドストレージシステム装置	ETERNUS DX200 S3
Deduplication モード	有効
Pool	Deduplication/Compression を有効にした 512GB の TPP を 1 つ作成
RAID	HDD 4 個で構成する RAID1 (1+0) を 1 つ作成
Logical Volume	Deduplication/Compression を有効にした 256GB の Deduplication/Compression ボリュームを 1 つ作成 この Deduplication/Compression ボリュームをサーバに認識させ、クイックフォーマット実施。 そのボリュームに 262,014MB(255.87GB)の基本パーティションを作成し、 ドライブレター F:¥を割り当て
容量削減率	ETERNUS Web GUI の[シン・プロビジョニングプール]の 「Deduplication/Compression」が有効の場合に表示される値を参照

- ・ 容量削減率は、書き込まれた論理データの容量に対して、重複排除／圧縮機能により削減されたデータ容量の割合です。
- ・ 各測定は、データ領域を初期化するために Deduplication/Compression を有効にした TPP の作成から実施します。
Deduplication/Compression ボリュームをサーバに再認識させて測定します。

2.2.5. 検証結果

パーティションのディレクトリ 1 つ(F:¥test)に格納する(重複するファイルが存在しない)ケースの場合

表 12 パーティションのディレクトリ 1 つ(F:¥test)に格納した際の測定結果

Deduplication/Compression を有効にした TPP のサイズ	Deduplication/Compression ボリュームサイズ	パーティションサイ ズ	格納場所と 格納ファイルサイズ	容量削減 率	[格納ファイルサイズ × (1 - 容量削減率/100)] ^(*)
512GB	256GB	255.87GB	F:¥test に 1,539MB	32%	1,047MB

パーティションのディレクトリ 2 つ(F:¥test と、F:¥test - Copy)に同じファイルを格納する(重複するファイルが存在する)ケースの場合

表 13 パーティションのディレクトリ 2 つ(F:¥test と、F:¥test - Copy)に同じファイルを格納した際の測定結果
(F:¥test をディレクトリごとコピーし F:¥test - Copy を作成)

Deduplication/Compression を有効にした TPP のサイズ	Deduplication/Compression ボリュームサイズ	パーティションサイ ズ	格納場所と 格納ファイルサイズ	容量削減 率	[格納ファイルサイズ × (1 - 容量削減率/100)] ^(*)
512GB	256GB	255.87GB	F:¥test に 1,539MB、 F:¥test - Copy に 1,539MB 合計 3,078MB	65%	1,077MB

(*) [格納ファイルサイズ × (1 - 容量削減率/100)]で、容量削減後の概算値を求めることができます。

2.2.6. 容量削減率の表示例

ETERNUS Web GUI での容量削減率の表示例を示します。

「Deduplication/Compression」が Enable の場合、「Capacity Reduction Rate」に容量削減率が表示されます。

容量削減率は Deduplication/Compression 設定を有効にした TPP の瞬時値であり、サーバ側 OS のキャッシュがフラッシュされていないときなど、サーバ側で Write 処理が未完了データは除外されます。

図 14 容量削減率の表示例（参考画像につき、検証時の構成や値とは関係がありません）

No.	Name	Drive Type	RAID Level	Status	Used Status	Encryption	Deduplication/Compression	Capacity Reduction Rate
0		SSD		Available	Normal	-	Enable	0%
1		SSD		Available	Normal	-	Disable	-

2.2.7. 考察

検証の結果から、重複排除／圧縮機能は、同一ファイルが存在するファイルサーバや VDI など、文書ファイルや画像ファイルの管理に効果的です。

重複排除／圧縮機能を適用すると、利用者が操作や定義を行わずに容量削減が行われます。

さらに、重複排除／圧縮機能による保管容量の削減は、物理領域へデータ書き込み量と回数が減ります。これによりハードディスクやフラッシュなど、ハードウェア耐久性が全体的に向上し、信頼性の向上に結びつきます。

3. まとめ

重複排除／圧縮機能の効果が期待できる環境例を下記に示します。

- ・世代数を多く管理するバックアップ環境
- ・災害対策・法令順守などで大容量の退避データを管理している環境
- ・容量削減率が高いファイルを多く保管している環境

重複排除／圧縮機能は、保管容量を低減し、ストレージ容量の利用効率が向上するため、導入費用や初期投資の抑止、さらに保管コストや設置コストの削減を目指すうえで、重要なウエイトを占める機能です。

本書を参考にし、保管容量を低減するメリットを見極めて、適切な環境に適用してください。

重複排除／圧縮機能は、システムのコスト削減に大きく貢献します。

Contact

インターネット情報ページ
<http://www.fujitsu.com/jp/eternus/>

製品・サービスについてのお問い合わせは
富士通コンタクトライン 0120-933-200
受付時間 9:00～17:30
(土曜・日曜・祝日・当社指定の休業日を除く)

富士通株式会社
〒105-7123
東京都港区東新橋 1-5-2 汐留シティセンター

■商標登記について

EMC は、米国およびその他の国における EMC Corporation の登録商標です。
HP、HP のロゴ、HP に関連する商標及びロゴは、米国およびその他の諸国における Hewlett Packard Enterprise の登録商標です。
Intel、インテル、Xeon は、米国およびその他の国における Intel Corporation の商標です。
Microsoft、Windows、Windows Server 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。
NetApp は米国およびその他の国における NetApp, Inc. の登録商標または商標です。
Pure Storage は米国 Pure Storage, Inc. の米国における商標または登録商標です。
なお、本資料に記載されているシステム名、製品名等には、必ずしも商標表示(R)、(TM)を付記していません。
その他、一般製品名・社名は、各社の商標または登録商標です。

■免責事項について

富士通株式会社は、本書の内容に関して、いかなる保証もしません。また、本書の内容に関連したいかなる損害についてもその責任は負いません。