(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2021-99611
(P2021-99611A)
(43)公開日　令和3年7月1日(2021.7.1)

(51)Int.Cl.		F I			テーマコード（参考）
G06F 3/06	(2006.01)	G06F	3/06	301W	
		G06F	3/06	302A	
		G06F	3/06	302J	

審査請求 有　請求項の数 10　ＯＬ　（全 29 頁）

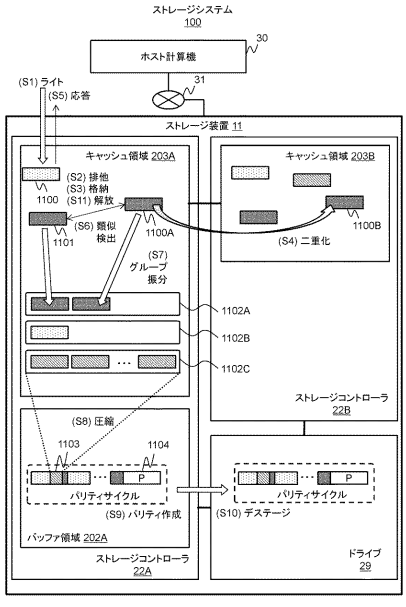
(21)出願番号　特願2019-230475 (P2019-230475)
(22)出願日　令和1年12月20日 (2019.12.20)

(71)出願人　000005108
株式会社日立製作所
東京都千代田区丸の内一丁目６番６号
(74)代理人　110000279
特許業務法人ウィルフォート国際特許事務所
(72)発明者　水島　永雅
東京都千代田区丸の内一丁目６番６号　株式会社日立製作所内
(72)発明者　吉原　朋宏
東京都千代田区丸の内一丁目６番６号　株式会社日立製作所内
(72)発明者　島田　健太郎
東京都千代田区丸の内一丁目６番６号　株式会社日立製作所内

(54)【発明の名称】ストレージシステムおよびストレージシステムのデータ圧縮方法

(57)【要約】
【課題】ランダムに書き込まれたデータの圧縮率を高め、アクセス性能を向上させる。
【解決手段】ストレージコントローラ２２Ａは、ドライブ２９にリードまたはライトされるデータを格納するキャッシュ領域２０３Ａを備え、キャッシュ領域２０３Ａに格納されドライブ２９に入力される複数のデータについて、そのデータ間の類似度に基づいてグループ化し、前記グループを選択し、選択したグループのデータをグループ単位で圧縮し、圧縮したデータをドライブ２９に格納させる。
【選択図】図１

図1



【特許請求の範囲】
【請求項１】
物理記憶領域を有するドライブと、前記ドライブに入出力するデータを処理するコントローラを備え、
前記コントローラは、
前記ドライブにリードまたはライトされるデータを格納するキャッシュ領域を備え、
前記コントローラは、
前記キャッシュ領域に格納され前記ドライブに入力される複数のデータについて、そのデータ間の類似度に基づいてグループ化し、
前記グループを選択し、
選択したグループのデータをグループ単位で圧縮し、
前記圧縮したデータを前記ドライブに格納させるストレージシステム。
【請求項２】
前記グループの選択は、前記グループ内のデータへのアクセス頻度および前記グループに含まれる複数のデータ間の類似度に基づいて行う請求項１に記載のストレージシステム。
【請求項３】
前記コントローラは、前記グループのアクセス頻度が第１設定値以下かつ前記グループに含まれるデータ間の類似度が第２設定値以上の場合、前記キャッシュ領域に格納されたデータのグループを選択して圧縮する請求項２に記載のストレージシステム。
【請求項４】
前記コントローラは、前記キャッシュ領域におけるダーティキャッシュ比率が所定値以上の場合、前記キャッシュ領域に格納されたデータのグループを選択して圧縮する請求項３に記載のストレージシステム。
【請求項５】
前記グループ化するデータは、別個のライト要求に基づいて前記キャッシュ領域に格納された複数のデータを含む請求項１に記載のストレージシステム。
【請求項６】
前記コントローラは、
前記ドライブに圧縮して格納された前記グループに含まれるデータ間の類似度を管理し、
前記ドライブに圧縮して格納されたデータを含むグループのうち、前記データ間の類似度が基準値以下のグループを選択し、
前記選択したグループに含まれるデータを前記ドライブから読み出して伸張し、前記伸張したデータを再グループ化及び再圧縮のために前記キャッシュ領域に格納する請求項１に記載のストレージシステム。
【請求項７】
前記コントローラは、
圧縮したデータの伸長は、先頭側から完了し、
前記圧縮時には、同一グループに含まれるデータについて、前記アクセス頻度の高いデータを前記アクセス頻度の低いデータに比べて前記グループの先頭により近い位置に配置する請求項２に記載のストレージシステム。
【請求項８】
前記コントローラは、前記データのハッシュ値に基づいて、前記類似度を評価し、
前記ハッシュ値の計算に用いる入力値は、前記データに含まれる同一長の複数の文字列であって、前記文字列は、前記データにおける出現頻度が所定数以上である請求項１に記載のストレージシステム。
【請求項９】
前記コントローラは、
ライト要求にかかるデータについて、重複排除判定を行い、

重複排除要の場合には、重複排除処理を行い、
重複排除不要の場合には、前記グループ化、圧縮及びドライブ格納を行う請求項１に記載のストレージシステム。
【請求項１０】
物理記憶領域を有するドライブと、前記ドライブに入出力するデータを処理するコントローラを備えるストレージシステムのデータ圧縮方法であって、
前記コントローラは、
前記ドライブにリードまたはライトされるデータを格納するキャッシュ領域を備え、
前記コントローラは、
前記キャッシュ領域に格納され前記ドライブに入力される複数のデータについて、そのデータ間の類似度に基づいてグループ化し、
前記グループを選択し、
選択したグループのデータをグループ単位で圧縮し、
前記圧縮したデータを前記ドライブに格納させるストレージシステムのデータ圧縮方法。

【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、ストレージシステムおよびストレージシステムのデータ圧縮方法に関する。
【背景技術】
【０００２】
ストレージシステムは、一般的に、１以上のストレージ装置を備える。１以上のストレージ装置の各々は、一般的に、記憶デバイスとして、例えば、ＨＤＤ（Ｈａｒｄ　Ｄｉｓｋ　Ｄｒｉｖｅ）又はＳＳＤ（Ｓｏｌｉｄ　Ｓｔａｔｅ　Ｄｒｉｖｅ）を備える。ストレージシステムは、ＳＡＮ（Ｓｔｏｒａｇｅ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）又はＬＡＮ（Ｌｏｃａｌ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）などのネットワーク経由で、１又は複数の上位装置（例えば、ホスト計算機３０計算機）からアクセスされる。一般的に、ストレージ装置は、ＲＡＩＤ（Ｒｅｄｕｎｄａｎｔ　Ａｒｒａｙ　ｏｆ　Ｉｎｄｅｐｅｎｄｅｎｔ　（ｏｒ　Ｉｎｅｘｐｅｎｓｉｖｅ）　Ｄｉｓｋｓ）技術に従う高信頼化方法を用いることで信頼性を向上している。
【０００３】
ストレージシステムには、物理的な記憶デバイスに保存するデータ量を削減してコストを低減するため、可逆圧縮機能を有するものがある。また、その圧縮率をできるだけ良くするために圧縮対象のデータに前処理を施すものもある。
【０００４】
特許文献１には、その一つの技術が開示されている。すなわち、特許文献１に開示された技術では、ストレージシステム外部から書き込まれたファイルおよびシーケンスデータをチャンクという単位に分割し、チャンク間の類似性を評価し、類似するチャンク順に並び替えて新たなシーケンスデータを作り、それをグループ単位で圧縮して記憶デバイスに保存する。また、元のファイルやシーケンスデータを復元するためのチャンク並び替えリストも記憶デバイスに保存する。
【先行技術文献】
【特許文献】
【０００５】
【特許文献１】米国特許第９３６７５５７号明細書
【発明の概要】
【発明が解決しようとする課題】
【０００６】

特許文献１に記載のストレージシステムは、一度に書き込まれたファイルおよびシーケンスデータを圧縮して格納する時に、類似性に基づく並び替えを行ってから圧縮することによって圧縮率を改善する。ストレージシステムへの書き込みデータは、そのようなシーケンシャルな性質のものだけでなく、ランダムで互いに無関係なものもある。

【０００７】

後者のような書き込みデータを扱う場合においても、類似性に基づいて圧縮率を改善するストレージシステムを実現することが課題である。また、それを実現する際に、ストレージシステムのアクセス性能を改善することも課題である。

【０００８】

本発明は、上記事情に鑑みなされたものであり、その目的は、ランダムに書き込まれたデータの圧縮率を高め、アクセス性能を向上させることが可能なストレージシステムおよびストレージシステムのデータ圧縮方法を提供することにある。

【課題を解決するための手段】

【０００９】

上記目的を達成するため、第１の観点に係るストレージシステムは、物理記憶領域を有するドライブと、前記ドライブに入出力するデータを処理するコントローラを備え、前記コントローラは、前記ドライブにリードまたはライトされるデータを格納するキャッシュ領域を備え、前記コントローラは、前記キャッシュ領域に格納され前記ドライブに入力される複数のデータについて、そのデータ間の類似度に基づいてグループ化し、前記グループを選択し、選択したグループのデータをグループ単位で圧縮し、前記圧縮したデータを前記ドライブに格納させる。

【発明の効果】

【００１０】

本発明によれば、ランダムに書き込まれたデータの圧縮率を高め、アクセス性能を向上させることができる。

【図面の簡単な説明】

【００１１】

【図１】図１は、実施形態に係るストレージシステムのデータ圧縮を伴うデータライト手順を示すブロック図である。

【図２】図２は、実施形態に係るストレージシステムのハードウェア構成例を示すブロック図である。

【図３】図３は、図２のボリューム管理テーブルの構成例を示す図である。

【図４】図４は、図２のプール構成管理テーブルの構成例を示す図である。

【図５】図５は、図２のＲＡＩＤ構成管理テーブルの構成例を示す図である。

【図６】図６は、図２のプール割当管理テーブルの構成例を示す図である。

【図７】図７は、図２のドライブ割当管理テーブルの構成例を示す図である。

【図８】図８は、図２のストレージ装置によって管理される論理記憶階層の構成例を示す図である。

【図９】図９は、図２のグループ管理テーブルの構成例を示す図である。

【図１０】図１０は、図２のメモリ割当管理テーブルの構成例を示す図である。

【図１１】図１１は、図２のＬＲＵリストテーブルとその遷移を示す図である。

【図１２】図１２は、図２のストレージ装置におけるキャッシュメモリのリソース配分比率の変化を示す図である。

【図１３】図１３は、図２のストレージ装置におけるスロットとグループの構成例を示す図である。

【図１４】図１４は、図２のストレージ装置が実行する類似ハッシュ算出処理を示す図である。

【図１５】図１５は、図２のストレージ装置が実行するデータリード処理を示すフローチャートである。

【図１６】図１６は、図２のストレージ装置が実行するデータライト処理を示すフローチ

ャートである。
【図１７】図１７は、図２のストレージ装置が実行する圧縮前データ構築処理を示すフローチャートである。
【図１８】図１８は、図２のストレージ装置が実行する圧縮・デステージ処理を示すフローチャートである。
【図１９】図１９は、図２のストレージ装置が実行するデータ削除リトライ処理を示すフローチャートである。
【図２０】図２０は、図２のストレージ装置におけるデータ伸張処理の過程を示す図である。
【発明を実施するための形態】
【００１２】
実施形態について、図面を参照して説明する。なお、以下に説明する実施形態は特許請求の範囲に係る発明を限定するものではなく、また、実施形態の中で説明されている諸要素およびその組み合わせの全てが発明の解決手段に必須であるとは限らない。
【００１３】
以下の説明では、「インターフェース部」は、ユーザインターフェース部と、通信インターフェース部とのうちの少なくとも１つを含んでよい。ユーザインターフェース部は、１以上のＩ／Ｏデバイス（例えば入力デバイス（例えばキーボード及びポインティングデバイス）と出力デバイス（例えば表示デバイス））と表示用計算機とのうちの少なくとも１つのＩ／Ｏデバイスを含んでよい。通信インターフェース部は、１以上の通信インターフェースデバイスを含んでよい。１以上の通信インターフェースデバイスは、１以上の同種の通信インターフェースデバイス（例えば１以上のＮＩＣ（Ｎｅｔｗｏｒｋ　Ｉｎｔｅｒｆａｃｅ　Ｃａｒｄ））であってもよいし、２以上の異種の通信インターフェースデバイス（例えばＮＩＣとＨＢＡ（Ｈｏｓｔ　Ｂｕｓ　Ａｄａｐｔｅｒ））であってもよい。
【００１４】
また、以下の説明では、「メモリ部」は、１以上のメモリを含む。少なくとも１つのメモリは、揮発性メモリであってもよいし、不揮発性メモリであってもよい。メモリ部は、主に、プロセッサ部による処理の際に使用される。
【００１５】
また、以下の説明では、「プロセッサ部」は、１以上のプロセッサを含む。少なくとも１つのプロセッサは、典型的には、ＣＰＵ（Ｃｅｎｔｒａｌ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ）である。プロセッサは、処理の一部又は全部を行うハードウェア回路（例えば、ＣＰＵの処理を補助するオフロードエンジン）を含んでもよい。
【００１６】
また、以下の説明では、「ｘｘｘテーブル」の表現にて情報を説明することがあるが、情報は、どのようなデータ構造で表現されていてもよい。すなわち、情報がデータ構造に依存しないことを示すために、「ｘｘｘテーブル」を「ｘｘｘ情報」と言うことができる。また、以下の説明において、各テーブルの構成は一例であり、１つのテーブルは、２以上のテーブルに分割されてもよいし、２以上のテーブルの全部又は一部が１つのテーブルであってもよい。
【００１７】
また、以下の説明では、同種の要素を区別しないで説明する場合には、参照符号のうちの共通符号を使用し、同種の要素を区別する場合は、参照符号（又は要素のＩＤ（例えば識別番号）を使用することがある。例えば、複数のストレージコントローラを区別しない場合には、「ストレージコントローラ２２」と記載し、各ストレージコントローラを区別する場合には、「ストレージコントローラ２２Ａ」、「ストレージコントローラ２２Ｂ」のように記載する。他の要素（例えばキャッシュ領域２０３、バッファ領域２０２、アドレス１１００等）も同様である。
【００１８】
また、以下の説明では、「ストレージシステム」は、１以上のストレージ装置を含む。

少なくとも１つのストレージ装置は、汎用的な物理計算機であってもよい。また、少なくとも１つのストレージ装置が、仮想的なストレージ装置であってもよいし、ＳＤｘ（Ｓｏｆｔｗａｒｅ－Ｄｅｆｉｎｅｄ ａｎｙｔｈｉｎｇ）を実行してもよい。ＳＤｘとしては、例えば、ＳＤＳ（Ｓｏｆｔｗａｒｅ Ｄｅｆｉｎｅｄ Ｓｔｏｒａｇｅ）（仮想的なストレージ装置の一例）又はＳＤＤＣ（Ｓｏｆｔｗａｒｅ－ｄｅｆｉｎｅｄ Ｄａｔａｃｅｎｔｅｒ）を採用することができる。

【００１９】

図１は、実施形態に係るストレージシステムのデータ圧縮を伴うデータライト手順を示すブロック図である。

図１において、ストレージシステム１００は、ホスト計算機３０及びストレージ装置１１を備える。ホスト計算機３０は、ネットワーク３１を介してストレージ装置１１に接続され、管理計算機（図示せず）によって管理される。

【００２０】

ストレージ装置１１は、１以上のボリューム（仮想的な記憶領域）を有する。ストレージ装置１１は、物理記憶領域を有するドライブ２９と、ドライブ２９を制御するストレージコントローラ２２Ａ、２２Ｂを備える。ストレージコントローラ２２Ａは、ドライブにおいてリードまたはライトされるデータをキャッシュするキャッシュ領域２０３Ａと、転送時のデータを一時的に保存するバッファ領域２０２を備える。ストレージコントローラ２２Ｂは、キャッシュ領域２０３Ａにキャッシュされたデータを二重化するキャッシュ領域２０３Ｂを備える。

【００２１】

ホスト計算機３０は、物理的な計算機でもよいし、物理的な計算機で実行される仮想的な計算機でもよい。ホスト計算機３０は、ストレージシステム１００において実行される仮想的な計算機でもよい。ホスト計算機３０からは、ストレージ装置１１のストレージコントローラ２２Ａ又はストレージコントローラ２２Ｂに対してデータの書き込みが行われる。

【００２２】

以下、このストレージシステム１００において、ホスト計算機３０からのデータライト命令に応じた処理手順について説明する。図１の例では、ホスト計算機３０からのライト命令をストレージコントローラ２２Ａが受領した場合について示す。

【００２３】

（Ｓ１）ストレージ装置１１は、ホスト計算機３０からネットワーク３１を介してライト命令を受信する。ライト命令は、データとデータの割当先アドレス１１００とを含んでいる。ストレージ装置１１がライト命令を処理する際のデータ単位は、例えば８ＫＢである。８ＫＢに満たない場合は、周辺アドレスの不足データを補うため、当該装置内の記憶済みデータを読み出して補完し、８ＫＢにしてから処理する。周辺アドレスに未書き込み部分がある場合は、その部分を不定値として管理し、暫定的にゼロデータで補完して処理する。８ＫＢより大きい場合は、１つ以上の８ＫＢデータと、８ＫＢに満たない残りデータとに分け、上記補完により全て８ＫＢデータにして個別に処理する。従って、以下の説明では、ストレージ装置１１が扱うデータサイズは８ＫＢであるものとする。ストレージ装置１１は、ライト命令受信後に、Ｓ２以降のライト処理を開始する。

【００２４】

（Ｓ２）ストレージ装置１１は、ライト命令に応答して、割当先アドレス１１００が示すスロットの排他を確保する。これによって、そのスロット内のデータが他のライト命令によって更新されることを防ぐ。「スロット」とは、ボリューム（ＶＯＬ）における領域単位の一種である。具体的には、本実施形態のスロットは、ドライブ２９への書き込みが行われたか否か及びバッファ領域２０２への転送が行われたか否か等の管理の単位となる。スロットのサイズは、例えば２５６ＫＢである。本実施形態では、この領域を「スロット」と呼ぶが、他の名称で呼ばれてもよい。

【００２５】

（Ｓ３）ストレージコントローラ２２Ａは、キャッシュ領域２０３Ａにおいて、データの割当先アドレス１１００に対応するアドレス１１００Ａにデータを格納する。

【００２６】

（Ｓ４）ストレージコントローラ２２Ａは、キャッシュ領域２０３Ａ内に格納されたデータをストレージコントローラ２２Ｂに転送する。ストレージコントローラ２２Ｂは、割当先アドレス１１００に対応するキャッシュ領域２０３Ｂ内のアドレス１１００Ｂに受領したデータを格納し、ストレージコントローラ２２Ａへ応答を返すことでストレージ装置１１内でのデータの二重化を完了する。

【００２７】

（Ｓ５）データの二重化を完了した後、ストレージ装置１１は、ホスト計算機３０に対してネットワーク３１を介してライト完了を応答する。この時点で、ホスト計算機３０は、ライトが完了したと認識する。

【００２８】

（Ｓ６）ストレージコントローラ２２Ａは、キャッシュ領域２０３Ａ内に格納されている複数のデータの中から、例えば、アドレス１１００Ａのデータの内容と類似関係にある１つ以上のデータを検出する。アドレス１１００Ａのデータの内容と類似関係にあるデータのアドレスを１１０１とする。

【００２９】

類似関係にあるデータを検出するために、８ＫＢのデータの内容から類似ハッシュと呼ばれる値（データ内容の特徴を表す小さいサイズ（例えば８Ｂ）の値）を算出する。例えば、類似ハッシュは、ＬＳＨ（Ｌｏｃａｌｔｙ　Ｓｅｎｓｉｔｉｖｅ　Ｈａｓｈ）である。２つのデータの内容が類似しているほど、互いの類似ハッシュの値は近くなる。類似ハッシュの値が近いことは、ハミング距離が短いことを意味する。具体的な類似ハッシュ算出方法は後述する。

【００３０】

（Ｓ７）ストレージコントローラ２２Ａは、キャッシュ領域２０３Ａにキャッシュされたデータ間の類似度に基づいて、キャッシュ領域２０３Ａにキャッシュされたデータをクルーピングしたグループを生成する。各グループにクルーピングされるデータは、ホスト計算機３０から別個のライト要求に基づいてライトされたデータを含んでいてもよい。このとき、ストレージコントローラ２２Ａは、グループごとに複数のデータを格納するため、１つ以上の領域をキャッシュ領域２０３Ａ内に確保する。グループごとに確保された領域のサイズは、例えば１２８ＫＢである。そして、ストレージコントローラ２２Ａは、例えば、ある１つのグループ１１０２Ａの領域に、アドレス１１００Ａのデータとアドレス１１０１のデータを転送する。転送するデータの数は２個以上であり、例えば１個のグループには、類似関係にある８ＫＢデータを１６個収納できる。別のグループ１１０２Ｂ、１１０２Ｃの領域には、（Ｓ６）の方法で検出した類似関係にあるデータをそれぞれ収納する。

【００３１】

また、後述するように、ストレージコントローラ２２Ａは、選択したデータをグルーピングする前に、重複排除処理を行って、その処理後の代表データのみをグルーピングしてもよい。重複排除処理は、内容が完全に等しいＮ（Ｎは、２以上の整数）個のデータを１個の代表データに割り当てる処理である。この重複排除処理により、実際の格納データ量をＮ分の１に削減することができる。

【００３２】

（Ｓ８）ストレージコントローラ２２Ａは、例えば、キャッシュ領域２０３Ａからドライブ２９へ書き出すグループ（例えば、グループ１１０２Ｃ）を選択し、選択したグループに含まれるデータをまとめて圧縮してバッファ領域２０２Ａ内のアドレス１１０３に追記的に格納する。

【００３３】

ここで、ストレージコントローラ２２Ａは、グループのアクセス頻度およびグループに

含まれるデータ間の類似度に基づいて、圧縮対象となるグループを選択する。例えば、ストレージコントローラ２２Ａは、グループのアクセス頻度が第１設定値以下かつグループに含まれるデータ間の類似度が第２設定値以上のグループを圧縮対象として選択する。これにより、アクセス頻度が低く、データ間の類似度が高いグループを圧縮対象として選択し、キャッシュ領域２０３Ａから落とすことができる。このため、ドライブ容量を削減することが可能となるとともに、アクセス頻度が高いデータをキャッシュ領域２０３Ａに残すことができ、キャッシュヒット率を向上させることが可能となることから、アクセス性能を向上させることができる。

【００３４】

さらに、ストレージコントローラ２２Ａは、キャッシュ領域２０３Ａにおけるダーティキャッシュ比率が上限値以上の場合、グループの圧縮を実行し、ダーティキャッシュ比率が上限値未満の場合、グループの圧縮を保留するようにしてもよい。これにより、キャッシュ領域２０３に新たにライトデータが受け入れられるＦｒｅｅ状態の領域を確保しつつ、より圧縮度を高めることが可能な類似したデータを可能な限りキャッシュ領域２０３に保持させることができ、データの圧縮率を高め、アクセス性能を向上させることが可能となる。

【００３５】

なお、この圧縮処理は、バッファ領域２０２Ａ内にＲＡＩＤパリティサイクル分以上の量が溜まるまでグループごとに実施される。このとき、ストレージコントローラ２２Ａは、選択したグループの圧縮後に暗号化処理を行って、その暗号化処理後の圧縮データをアドレス１１０３に格納してもよい。

【００３６】

（Ｓ９）ストレージコントローラ２２Ａは、バッファ領域２０２Ａ内に追記的に格納した１つ以上のグループの圧縮データがパリティサイクル分以上の量に達すると、それらの内容を保証するためのパリティを生成し、バッファ領域２０２Ａ内のアドレス１１０４へ格納する。

【００３７】

（Ｓ１０）ストレージコントローラ２２Ａは、バッファ領域２０２Ａ内の圧縮グループ（１つ以上のグループのパリティサイクル分以上の圧縮データ）及びそれに対応するパリティをドライブ２９へ送信し、ドライブ２９に書き出す（デステージ処理）。

【００３８】

（Ｓ１１）ストレージコントローラ２２Ａは、デステージ処理が完了すると、（Ｓ２）において確保したスロットの排他を解放する。

【００３９】

図２は、実施形態に係るストレージシステムのハードウェア構成例を示すブロック図である。

図２において、ストレージ装置１１は、１以上のストレージコントローラ２２と、１以上のストレージコントローラ２２に接続された複数のドライブ２９とを有する。

【００４０】

ストレージコントローラ２２は、ＦＥ＿Ｉ／Ｆ（フロントエンドインターフェースデバイス）２３、プロセッサ２４、メモリ２５、２１５、ＢＥ＿Ｉ／Ｆ（バックエンドインターフェースデバイス）２７、内部ネットワーク２６およびオフロードエンジン２１４を備える。ＦＥ＿Ｉ／Ｆ２３、プロセッサ２４、メモリ２５、ＢＥ＿Ｉ／Ｆ２７およびオフロードエンジン２１４は、内部ネットワーク２６を介して接続されている。

【００４１】

ＦＥ＿Ｉ／Ｆ２３は、ホスト計算機３０との通信を行う。ＢＥ＿Ｉ／Ｆ２７は、ドライブ２９との通信を行う。プロセッサ２４は、ストレージ装置１１全体を制御する。

【００４２】

メモリ２５は、プロセッサ２４で使用されるプログラム及びデータを格納する。メモリ２５は、プログラムを管理するプログラム領域２０１、データの転送時の一時的な保存領

域であるバッファ領域２０２、ホスト計算機３０からのライトデータ（ライト命令に応答して書き込まれるデータ）及びドライブ２９からのリードデータ（リード命令に応答して読み出されたデータ）を一時的に格納するキャッシュ領域２０３及び種々のテーブルを格納するテーブル管理領域２０６を有する。

【００４３】

テーブル管理領域２０６は、ボリュームに関する情報を保持するボリューム情報管理テーブル２０７、プールに関する情報を保持するプール構成管理テーブル２０８、ＲＡＩＤ構成に関する情報を保持するＲＡＩＤ構成管理テーブル２０９、プール割り当てに関する情報を保持するプール割当管理テーブル２１０、ドライブ割り当てに関する情報を保持するドライブ割当管理テーブル２１１、グループへのデータ割り当てに関する情報を保持するグループ管理テーブル２１２、メモリ割り当てに関する情報を保持するメモリ割当管理テーブル２１３及びスロットのアクセス頻度を管理するＬＲＵ（Ｌｅａｓｔ　Ｒｅｃｅｎｔｌｙ　Ｕｓｅｄ）リストテーブル２１７を格納する。

【００４４】

ドライブ２９は、不揮発性のデータ記憶媒体を有する装置であり、例えばＳＳＤ（Ｓｏｌｉｄ　Ｓｔａｔｅ　Ｄｒｉｖｅ）でもよいし、ＨＤＤ（Ｈａｒｄ　Ｄｉｓｋ　Ｄｒｉｖｅ）でもよい。複数のドライブ２９が、複数のＲＡＩＤグループ（パリティグループとも呼ばれる）を構成してよい。各ＲＡＩＤグループは、１以上のドライブ２９から構成される。

【００４５】

オフロードエンジン２１４は、プロセッサ２４の行う処理を補助するハードウェア回路であり、図１を用いて説明したデータライトにおいて実施する、類似検出、圧縮、重複排除、暗号化およびパリティ生成などの処理の一部または全部をプロセッサ２４よりも高速に実行する。同様に、データリード（図による説明は省略）において実施する、伸張、復号およびパリティによる回復（ドライブ障害時のみ）などの処理の一部または全部をプロセッサ２４よりも高速に実行する。

【００４６】

メモリ２１５は、オフロードエンジン２１４に直接接続された専用メモリである。メモリ２１５は、データの転送時の一時的な保存領域であるバッファ領域２１６を有する。メモリ２１５は、オフロードエンジン２１４で扱われるデータや、それらの処理に関する情報も保存する。メモリ２１５を設けることで、内部ネットワーク２６とプロセッサ２４を経由するメモリ２５とオフロードエンジン２１４間の転送量及び転送時間を削減し、性能向上を図ることができる。

【００４７】

ＦＥ＿Ｉ／Ｆ２３およびＢＥ＿Ｉ／Ｆ２７は、インターフェース部の一例である。メモリ２５、２１５は、メモリ部の一例である。プロセッサ２４とオフロードエンジン２１４は、プロセッサ部の一例である。

【００４８】

図３は、図２のボリューム管理テーブルの構成例を示す図である。

図３において、ボリューム管理テーブル２０７は、ボリューム毎にエントリを有する。各エントリは、ＶＯＬ＿ＩＤ４１、ＶＯＬ属性４２、ＶＯＬ容量４３及びプールＩＤ４４の情報を格納する。

【００４９】

ＶＯＬ＿ＩＤ４１は、ストレージ装置１１が管理している各ボリュームのＩＤである。ＶＯＬ属性４２は、各ボリュームの属性を示す。各ボリュームの属性は、例えば、シンプロビジョニングのボリュームか、通常割り当てのボリュームか、データ削減が有効（ＯＮ）か無効（ＯＦＦ）か、システムボリュームかなどである。システムボリュームは、類似関係にあるデータを含むグループをストレージ装置１１が管理するためのボリュームであり、ホスト計算機３０からは直接見えない領域である。ボリューム容量４３は、各ボリュームの容量を示す。プールＩＤ４４は、各ボリュームに関連付けられているプールのＩＤ

である。

【００５０】

プロセッサ２４は、デステージ処理において、ボリューム管理テーブル２０７のボリューム属性４２を参照することで、ホスト計算機３０からデータのライト／リード命令を受けた時に、どのような処理を必要とするボリュームか判定できる。例えば、ＶＯＬ＿ＩＤ４１が“１０”のボリュームは、ボリューム属性４２が“シンプロビジョニング、削減ＯＮ”であるため、必要に応じてスロットに動的にプールを割り当てながら、ホスト計算機３０のライト命令に対して図１で説明したデータライトを行う。

【００５１】

ＶＯＬ＿ＩＤ４１が“０”のボリュームは、ボリューム属性４２が“シンプロビジョニング、削減ＯＦＦ”であるため、必要に応じてスロットに動的にプールを割り当てながら、ホスト計算機３０のライト命令に対してデータ削減処理（重複排除及び圧縮）を除いたデータライトを行う。

【００５２】

ＶＯＬ＿ＩＤ４１が“３０”のボリュームは、ボリューム属性４２が“通常割当、削減ＯＦＦ”であるため、ボリューム作成時に割り当てたプール容量範囲で、ホスト計算機３０のライト命令に対してデータ削減処理（重複排除及び圧縮）を除いたデータライトを行う。

【００５３】

図４は、図２のプール構成管理テーブルの構成例を示す図である。

図４において、プールは、１以上のＲＡＩＤグループを基に構成された論理記憶領域である。プール構成管理テーブル２０８は、プール毎にエントリを有する。各エントリは、プールＩＤ５１、ＲＡＩＤグループＩＤ５２、プール容量５３及びプール使用容量５４の情報を格納する。

【００５４】

プールＩＤ５１は、ストレージ装置１１が管理している各プールのＩＤである。ＲＡＩＤグループＩＤ５２は、プールの基になっている１以上のＲＡＩＤグループの各々のＩＤである。プール容量５３は、プールの容量を示す。プール使用容量５４は、プールのプール容量のうち、ボリュームに割り当てられている領域の総量を示す。

【００５５】

図５は、図２のＲＡＩＤ構成管理テーブルの構成例を示す図である。

図５において、ＲＡＩＤ構成管理テーブル２０９は、ＲＡＩＤ構成を管理する対象のＲＡＩＤグループ毎にエントリを有する。各エントリは、ＲＡＩＤグループＩＤ６１、ＲＡＩＤレベル６２、ドライブＩＤ６３、ドライブ種別６４、容量６５及び使用容量６６の情報を格納する。

【００５６】

ＲＡＩＤグループＩＤ６１は、ストレージ装置１１が管理している各ＲＡＩＤグループのＩＤである。ＲＡＩＤレベル６２は、ＲＡＩＤグループに適用されるＲＡＩＤアルゴリズムの種別を示す。ドライブＩＤ６３は、ＲＡＩＤグループを構成する１以上のドライブの各々のＩＤである。ドライブ種別６４は、ＲＡＩＤグループを構成するドライブの種別（例えばＨＤＤかＳＳＤか）を示す。容量６５は、ＲＡＩＤグループの容量を示す。使用容量６６は、ＲＡＩＤグループの容量のうちの使用されている容量を示す。

【００５７】

図６は、図２のプール割当管理テーブルの構成例を示す図である。

図６において、プール割当管理テーブル２１０は、プールの割り当てを管理する対象のスロットのＶＯＬアドレス（ボリューム内のスロットを示すアドレス）毎にエントリを有する。各エントリは、ＶＯＬ＿ＩＤ７１、ＶＯＬアドレス７２、プールＩＤ７３、プールアドレス７４、圧縮前サイズ７５、圧縮後サイズ７６及び類似度７７の情報を格納する。

【００５８】

ＶＯＬ＿ＩＤ７１は、ストレージ装置１１が管理し、ＶＯＬアドレスによって識別され

るスロットが属するボリュームのＩＤである。ＶＯＬアドレス７２は、それらスロットのＶＯＬアドレスである。プールＩＤ７３は、各スロット（２５６ＫＢ）に含まれる２つのグループ（各１２８ＫＢ）のデータをそれぞれ圧縮したデータを格納するために割り当てられたデータ領域を含むプールの各ＩＤである。プールアドレス７４は、２つのグループのデータをそれぞれ圧縮したデータを格納するために割り当てられたデータ領域のアドレス（プールに属するアドレス）である。圧縮前サイズ７５は、各グループの圧縮前のデータサイズ（各１２８ＫＢ）を示す。圧縮後サイズ７６は、各グループの圧縮後のデータサイズを示す。類似度７７は、各グループを構成する１６個の８ＫＢのデータの相互の類似関係の強さを示す値である。類似度７７が０％は、内容に全く類似関係がないことを示し、類似度７７が１００％は、内容が完全に一致している（可能性が高い）ことを示す。一般に、類似度７７が高いグループほど圧縮率が高い（圧縮後サイズ７６が小さい）ことが期待できる。

【００５９】

ストレージ装置１１で用いられる可逆圧縮アルゴリズムは、スライド辞書（Ｓｌｉｄｉｎｇ　Ｄｉｃｔｉｏｎａｒｙ）圧縮をベースとしており、過去の文字列から一致する文字列を見つけて短い符号（発見位置までの距離と一致長）に置き換えてデータ量を削減する。例えば、３０バイト一致する文字列を見つけて、２バイトの短い符号に置き換えられれば２８バイト削減される。類似度が高いデータには共通する文字列が多く含まれており、圧縮するグループ内のデータの類似度が高いほど、この辞書圧縮が効果的に働くため、圧縮率が高くなる。

【００６０】

図７は、図２のドライブ割当管理テーブルの構成例を示す図である。

図７において、ドライブ割当管理テーブル２１１は、ドライブ割当を管理するプールアドレス毎にエントリを有する。各エントリは、プールＩＤ８１、プールアドレス８２、ＲＡＩＤグループＩＤ８３、ドライブＩＤ８４及びドライブアドレス８５の情報を格納する。

【００６１】

プールＩＤ８１は、プールアドレスが属するプールのＩＤである。プールアドレス８２は、ストレージ装置１１が管理しているプールアドレスである。ＲＡＩＤグループＩＤ８３は、プールアドレスが示すデータ領域の基になっているＲＡＩＤグループのＩＤである。ドライブＩＤ８４は、プールアドレスが示すデータ領域の基になっているドライブのＩＤである。ドライブアドレス８５は、プールアドレスに対応したドライブアドレスである。

【００６２】

図８は、図２のストレージ装置によって管理される論理記憶階層の構成例を示す図である。

図８において、ホストＶＯＬ１０００は、シンプロビジョニングが適用され、データ削減が有効なボリュームであり、ホスト計算機３０に提供される。システムＶＯＬ１００１は、データ削減時の圧縮単位であるグループを管理するためのボリュームであり、ホスト計算機３０からは直接見えない。

【００６３】

ホストＶＯＬ１０００内の各８ＫＢデータアドレスに、システムＶＯＬ１００１内のグループに含まれる各８ＫＢデータアドレスが対応付けられる。図８の例では、ＶＯＬ１０００内の内容「Ａ，Ｂ，Ｄ」を格納する３つのアドレスに、グループ１１０２Ｄ内の３つのアドレスが対応付けられ、ＶＯＬ１０００内の内容「Ｃ，Ｅ，Ｆ」を格納する３つのアドレスに、グループ１１０２Ｅ内の３つのアドレスが対応付けられている。内容「Ａ，Ｂ，Ｄ」と「Ｃ，Ｅ，Ｆ」を格納するアドレスの対応先がそれぞれ同じグループ１１０２Ｄ、１１０２Ｅに含まれているのは、図１の類似検出（Ｓ６）により、それぞれの内容に類似関係があると判断されたためである。

【００６４】

なお、重複排除処理が実施された場合は、ホストＶＯＬ１０００内の複数の８ＫＢデータアドレスに、システムＶＯＬ１００１の１つの８ＫＢデータアドレスが対応付けられることがある。図示していないが、異なるホストＶＯＬの複数の８ＫＢデータアドレスに、システムＶＯＬの１つの８ＫＢデータアドレスが対応付けられることもある。図８の例では、同一内容「Ａ」を格納する異なる２つの８ＫＢデータアドレス１１０５及び１１０６にグループ１１０２Ｄ内の１つの８ＫＢデータアドレス１１０７が対応付けられ、同一内容「Ｃ」を格納する異なる２つの８ＫＢデータアドレス１１０８及び１１０９にグループ１１０２Ｅ内の１つの８ＫＢデータアドレス１１１０が対応付けられている。

【００６５】

システムＶＯＬ１００１内のスロット（２５６ＫＢ）に含まれる２つグループ（各１２８ＫＢ）には、圧縮された当該グループデータの格納先として、プール１００２内のデータ領域がそれぞれ対応付けられる。図８の例では、グループ１１０２Ｄにデータ領域１１０３Ｄが対応付けられ、グループ１１０２Ｅにデータ領域１１０３Ｅが対応付けられている。なお、グループは圧縮されるため、各グループに対応するプール１００２内のデータ領域のサイズは、グループサイズ（１２８ＫＢ）以下である。

【００６６】

プール１００２の空間は、チャンクと呼ばれる単位で区分けされ、各チャンクはドライブアドレス空間１００３へ対応付けられる。ドライブアドレス空間１００３は、ＲＡＩＤグループ１００４を構成する複数のドライブ２９（例えば４台）によって提供される物理的なデータ格納空間である。パリティ生成処理で生成するパリティ１１０４のサイズは、チャンクサイズである。ドライブアドレス空間１００３は、複数のＲＡＩＤサイクルで構成される。ドライブ２９の故障によるデータ消失をＲＡＩＤ技術で回復するため、各ＲＡＩＤサイクルのパリティチャンクＰは、１つのドライブに収まるように対応付けられる。

【００６７】

ホストＶＯＬ１０００からシステムＶＯＬ１００１への割り当ては、図９のグループ管理テーブル２１２を基に管理される。そのテーブルの詳細は後述する。また、システムＶＯＬ１００１からプール１００２への割り当ては、図６のプール割当管理テーブル２１０を基に管理される。また、プール１００２からドライブアドレス空間１００３への割り当ては、図７のドライブ割当管理テーブル２１１を基に管理される。

【００６８】

図９は、図２のグループ管理テーブルの構成例を示す図である。

図９において、グループ管理テーブル２１２は、図８のホストＶＯＬ１０００の８ＫＢデータに対応するシステムＶＯＬ１００１のグループを管理するテーブルであり、管理対象の８ＫＢデータのアドレス毎にエントリを有する。各エントリは、ホストＶＯＬ＿ＩＤ９０１、ホストＶＯＬアドレス９０２、位置番号９０３、システムＶＯＬ＿ＩＤ９０４、システムＶＯＬアドレス９０５、グループ番号９０６及び位置番号９０７の情報を格納する。

【００６９】

ホストＶＯＬ＿ＩＤ９０１は、ストレージ装置１１が管理し、グループ対応の管理対象の８ＫＢデータが属するスロットが属するボリューム（つまり、ホストＶＯＬ１０００）のＩＤである。ホストＶＯＬアドレス９０２は、管理対象の８ＫＢデータが属するスロットのアドレスである。位置番号９０３は、管理対象の８ＫＢデータが属するスロット（２５６ＫＢ）の中で、当該８ＫＢデータが位置する場所を示す０～３１の範囲の番号であり、先頭を０、末尾を３１とする。

【００７０】

システムＶＯＬ＿ＩＤ９０４は、ストレージ装置１１が管理し、ホストＶＯＬ１０００の８ＫＢデータに対応する、システムＶＯＬ１００１のグループを含むスロットが属するボリュームのＩＤである。システムＶＯＬアドレス９０５は、管理対象の８ＫＢデータが対応する、システムＶＯＬ１００１のグループを含むスロットのアドレスである。グループ番号９０６は、管理対象の８ＫＢデータが対応する、システムＶＯＬ１００１のグルー

プの番号であり、０または１である。位置番号９０７は、管理対象の８ＫＢデータに対応する、システムＶＯＬ１００１のグループ（１２８ＫＢ）の中で、当該８ＫＢデータが対応する場所を示す０～１５の範囲の番号であり、先頭を０、末尾を１５とする。

【００７１】

なお、システムＶＯＬアドレス９０５、グループ番号９０６及び位置番号９０７において、いずれも“Ｎｏｎｅ”が設定されている第３行目のエントリの８ＫＢデータについては、図１の（Ｓ７）グループ振分で示したグルーピングが未完了であり、システムＶＯＬ１００１の対応先グループが未定の状態であることを示す。グルーピングが完了すると、これらの項目にアドレス及び番号が登録される。

【００７２】

また、第４行目と第７行目のエントリで管理される２つの８ＫＢデータは、システムＶＯＬ１００１の同じグループに対応付けられており、両データの内容が類似関係にあることを示す。第２行目と第６行目のエントリで管理される２つの８ＫＢデータも、システムＶＯＬ１００１の同じグループに対応付けられており、両データの内容が類似関係にあることを示す。第５行目と第８行目のエントリで管理される２つの８ＫＢデータは、システムＶＯＬ１００１の同じグループに対応付けられており、かつ位置番号も同一であるため、両データの内容が重複関係にあることを示す。

【００７３】

図１０は、図２のメモリ割当管理テーブルの構成例を示す図である。

図１０において、メモリ割当管理テーブル２１３は、キャッシュ領域２０３やバッファ領域２０２を利用するスロットについて、スロットアドレス毎にエントリを有する。各エントリは、ＶＯＬ＿ＩＤ９０８、ＶＯＬアドレス９０９、バッファ（ＢＦ）転送状態９１０、バッファ（ＢＦ）アドレス９１１、キャッシュ状態９１２及びキャッシュアドレス９１３の情報を格納する。

【００７４】

ＶＯＬ＿ＩＤ９０８は、ＶＯＬアドレスによって識別されるスロットが属するボリュームのＩＤである。ＶＯＬアドレス９０９は、ホストＶＯＬ１０００またはシステムＶＯＬ１００１のスロットアドレスである。

【００７５】

ＢＦ転送状態９１０は、システムＶＯＬ１００１の管理対象スロットに含まれるグループのデータを圧縮したデータを一時的に保持するため、バッファ領域２０２に転送されたか否かの状態であり、“未”はまだ転送されていないことを示し、“済”はすでに転送されたことを示す。ホストＶＯＬ１０００の管理対象スロットは、本項目を使わない。

【００７６】

ＢＦアドレス９１１は、圧縮されたシステムＶＯＬ１００１のグループデータが転送されたバッファ領域２０２内のアドレスを示す。ホストＶＯＬ１０００の管理対象スロットは、本項目を使わない。ＢＦ転送状態９１０が“未”の場合は、ＢＦアドレス９１１に“Ｎｏｎｅ”が設定され、管理対象スロットに含まれるグループデータが圧縮されていないことを意味する。ＢＦ転送状態９１０が“済”であり、ＢＦアドレス９１１が値を持つ場合、バッファ領域２０２に圧縮されたグループデータが保持されている状態を意味する。ＢＦ転送状態９１０が“済”であり、ＢＦアドレス９１１が“Ｎｏｎｅ”である場合、バッファ領域２０２にあった圧縮されたグループデータがすでにドライブ２９へ格納され、バッファ領域２０２の使用部分のアドレスは解放されたことを意味する。

【００７７】

キャッシュ状態９１２は、ホストＶＯＬ１０００またはシステムＶＯＬ１００１の管理対象スロットのデータについて、ドライブ２９への格納状態を示す。キャッシュ状態９１２が“Ｄｉｒｔｙ”のデータは、ドライブ２９への格納が未完了である状態を意味し、“Ｃｌｅａｎ”のデータはドライブ２９への格納が完了した状態を意味する。なお、ホストＶＯＬスロットのキャッシュ状態９１２では、３２個の８ＫＢデータ毎にキャッシュ状態を管理し、システムＶＯＬスロットのキャッシュ状態９１２では、２つの１２８ＫＢグル

ープ毎に状態を管理する。システムＶＯＬスロットのキャッシュ状態９１２が、２グループとも“Ｄｉｒｔｙ”から“Ｃｌｅａｎ”に変わる時、バッファ領域２０２の使用アドレスは解放され、ＢＦアドレス９１１に“Ｎｏｎｅ”が設定される。ただし“Ｃｌｅａｎ”になっても、キャッシュ領域２０３には、当該グループのデータは圧縮されていない状態で残る。当該データがキャッシュから落ちるまで、ホスト計算機３０からストレージ装置１１への当該データのリード命令はキャッシュヒットする。

【００７８】

キャッシュアドレス９１３は、ホストＶＯＬ１０００またはシステムＶＯＬ１００１の管理対象スロットのデータを格納するために割り当てられたキャッシュ領域２０３の部分のアドレスである。キャッシュ状態９１２が全て“Ｃｌｅａｎ”であるスロットは、キャッシュから落とすことができる。キャッシュから落とす場合、メモリ割当管理テーブル２１３からそのスロットのエントリを削除する。

【００７９】

図１１は、図２のＬＲＵリストテーブルとその遷移を示す図である。

図１１において、ＬＲＵリストテーブル２１７は、圧縮対象グループを選択するため、スロットのアクセス頻度を管理する。システムＶＯＬアドレス２１０１は、アクセス頻度を管理するスロットの属するシステムＶＯＬのＩＤを示す。システムＶＯＬアドレス２１０２は、アクセス頻度を管理するスロットのアドレスを示す。本リストの上位にあるスロットほどアクセス頻度が低いことを意味する。

【００８０】

あるデータがホスト計算機３０からアクセスされると、アクセスデータを含むスロット（例えばアドレス＝２６００）は本リストの最下位に移動し、そのスロットより下位だった全てのスロット（例えばアドレス＝３３００、２４００、４３００など）はそれぞれ1段上へ移動する。このような移動処理を行うと、アクセス頻度の低いデータを含むスロットほどリスト上位に集まる。圧縮対象グループは、例えば、本リストの最上位２１０３に位置するスロット（例えばアドレス＝４１００）に含まれる２つのグループから選択することができる。

【００８１】

図１２は、図２のストレージ装置におけるキャッシュメモリのリソース配分比率の変化を示す図である。

図１２において、図２のキャッシュ領域２０３は、ホストボリュームやシステムボリュームのスロットのデータを一時的に保持する。データの保持状態によって、キャッシュ領域２０３のメモリリソースは、以下の３つの状態に分けられる。

【００８２】

データが保持されており、そのデータがすでにドライブ２９に格納されている領域部分は、“Ｃｌｅａｎ”である。データが保持されており、そのデータがまだドライブ２９に格納されていない領域部分は、“Ｄｉｒｔｙ”である。キャッシュ領域２０３のうちデータが保持されていない領域部分は、“Ｆｒｅｅ”である。キャッシュ領域２０３は、それら３つの状態の配分比率１０１０で配分される。

【００８３】

ホスト計算機３０からデータのライト命令を受けて、キャッシュ領域２０３にデータが新たに保持される時、Ｆｒｅｅ状態のリソースが使われて、その領域部分がＤｉｒｔｙ状態に変化する（１０１１）。このとき、キャッシュ領域２０３のＦｒｅｅの比率が減り、Ｄｉｒｔｙの比率が増える。

【００８４】

キャッシュ領域２０３に保持したＤｉｒｔｙ状態のデータがドライブ２９に格納されると、そのデータの保持領域は、Ｄｉｒｔｙ状態からＣｌｅａｎ状態に変化する（１０１２）。このとき、キャッシュ領域２０３のＤｉｒｔｙの比率が減り、Ｃｌｅａｎの比率が増える。

【００８５】

キャッシュ領域２０３に新たにライトデータが受け入れられるＦｒｅｅ状態の領域を増やすためには、Ｃｌｅａｎ状態の保持領域を解放する必要がある。このとき、Ｃｌｅａｎ状態の保持領域のデータは、ドライブ２９に格納済みなので、Ｃｌｅａｎ状態の保持領域を解放してもストレージ装置１１としてデータは損失しない。これにより、その解放領域は、Ｃｌｅａｎ状態からＦｒｅｅ状態に変化する（１０１３）。このとき、キャッシュ領域２０３のＣｌｅａｎの比率が減り、Ｆｒｅｅの比率が増える。

【００８６】

キャッシュ領域２０３のメモリリソースは、以上のようなライフサイクルで管理される。図１２の１０１２で示すＤｉｒｔｙ状態からＣｌｅａｎ状態へ変化させるには、図１の（Ｓ８）圧縮～（Ｓ１０）デステージの処理を行う必要がある。ホスト計算機３０からデータのライト命令を受けてから、Ｆｒｅｅ状態のリソースを用意するため、（Ｓ８）圧縮～（Ｓ１０）デステージの処理によりＤｉｒｔｙ状態からＣｌｅａｎ状態へ変化させ、Ｃｌｅａｎ状態の領域の解放によってＦｒｅｅ状態へ変化させると、（Ｓ１）ライトから（Ｓ５）応答までに長い時間がかかる。

【００８７】

そこで、図２のストレージコントローラ２２は、Ｄｉｒｔｙ状態のリソースの比率がある閾値に達したことを契機に、キャッシュ領域２０３にあるグループデータに対して（Ｓ８）圧縮を開始して（Ｓ１０）デステージまでの処理を行い、Ｄｉｒｔｙ状態からＣｌｅａｎ状態へ事前に変化させる。これにより、ストレージコントローラ２２は、ライト命令を受けたらＣｌｅａｎ状態の領域の解放だけを行ってＦｒｅｅ状態のリソースを直ぐに用意し、（Ｓ１）ライトから（Ｓ５）応答までの時間を短縮することができる。

【００８８】

Ｄｉｒｔｙ状態のリソースの比率がある閾値に達した時に、（Ｓ８）圧縮の対象として選択するグループは、キャッシュ領域２０３において、その時点で最も長い時間アクセスされていないスロット（つまり、ＬＲＵスロット）に含まれるグループである。

【００８９】

グループの選択方法は、この方法に限定されない。アクセスされていない時間が長い順に複数のスロット（Ｍ（Ｍは２以上の整数）個）を選択し、それらスロットに含まれるグループ（２Ｍ個）の中で、図６のプール割当管理テーブルにおける類似度７７が最も高いグループを、（Ｓ８）圧縮の対象として選択してもよい。例えば、図１１のリスト上位範囲２１０４は、アクセスされていない時間が長い複数のスロット（Ｍ＝４）を含む。これらのスロットに含まれる８つのグループで最も類似度７７が高いグループを選択する。

【００９０】

この方法の利点を以下に説明する。類似度が低いグループデータは、圧縮率が低い（データの削減量が少ない）ことが予想される。そのため、そのグループデータは、圧縮せずにキャッシュ領域２０３に残す方がよい。もし、新たに受け入れたライトデータの中に、より類似したデータがあれば、キャッシュ領域２０３の上でグループを作り直して類似度７７を高めれば、圧縮率を改善できる可能性がある。一方、類似度が高いグループデータは、圧縮率が高い（データの削減量が多い）ことが期待できるため、キャッシュ領域２０３に残しても、新たに受け入れたライトデータの中に、より類似したデータがあることは期待できず、グループを作り直しても圧縮率が改善される可能性は低い。従って、最も類似度が高いグループを、（Ｓ８）圧縮の対象として選択するのが望ましい。

【００９１】

図１３は、図２のストレージ装置におけるスロットとグループの構成例を示す図である。

図１３において、ライト命令によって書かれる８ＫＢデータ１２０１は、ホスト計算機３０から見えるボリューム１２１０で管理される。ボリューム１２１０のアドレス空間は、複数のスロット１２００で構成され、ライト命令によって書かれる８ＫＢデータ１２０１の割当先アドレス１１００は、スロット１２００の１つが占めるアドレス空間に含まれる。

【００９２】
「スロットの排他を確保」とは、ホスト計算機３０からのリード命令及びライト命令で指定されたアドレスが示すスロットに対するリード及びライトを防ぐ操作であり、排他を確保したことをホスト計算機３０が認識するための情報が管理される。なお、この情報はビットマップ又は時間情報など識別できるものであれば種別は問わない。また、本実施形態において、「スロット」が、ボリューム（例えば、シンプロビジョニングに従うボリューム）における領域単位であるのに対し、「データ領域」は、スロットに割り当てられる領域（例えば、プール内の領域であるプール領域）である。例えば、ボリューム内の１個の２５６ＫＢスロットに割り当てられたプール領域には、８ＫＢデータを３２個格納できる。

【００９３】
類似関係にある１６個の８ＫＢデータを含むグループは、ホスト計算機３０から見えない別のボリューム１２２０で管理される。ボリューム１２２０のアドレス空間も、複数のスロット１２０２で構成される。ボリューム１２２０のスロット１２０２は、それぞれ２つのグループ１２０３で構成される。それら２つのグループ１２０３のアドレスは、それらを構成するスロット１２０２が占めるアドレス空間に含まれる。また、ある１つのグループ１２０３に１６個含まれる、類似関係にある８ＫＢデータ１２０４のアドレスは、そのグループ１２０３が占めるアドレス空間に含まれる。

【００９４】
図１４は、図２のストレージ装置が実行する類似ハッシュ算出処理を示す図である。なお、図１４では、簡単のため、類似ハッシュの対象データの数は２つ（データＡ、Ｂ）とし、各データのサイズは３７Ｂである場合を例にとる。データの数およびサイズが増えても、類似ハッシュの算出方法は同様である。

【００９５】
図１４において、各データＡ、Ｂについて、例えば、第Ｋ文字目から始まる連続３文字を３５個抽出する。ここで、Ｋは、１～３５である。以下、これらの３文字を単語と呼ぶ。２つのデータが共通の単語を多く含むかを判断することで類似性の強さを評価する。単語は、各データＡ、Ｂ内のどの位置に存在してもよいので、開始点（すなわち、Ｋ）を１バイトステップで変えながら単語を抽出する。

【００９６】
次に、各データＡ、Ｂにおいて、３５個の単語の出現頻度（以下、頻度と言う）を示す頻度順位表１４１、１４２を作成する。各頻度順位表１４１、１４２において、頻度の高いものから順に単語を並べる。ただし、頻度の少ない単語（例えば１回の単語）は、各頻度順位表１４１、１４２から除く。

【００９７】
図１４では、２つの頻度順位表１４１、１４２において共通する単語同士を線で結んでいる。共通する単語の数（積集合の要素数）を、両者の単語の和集合の要素数で割ることで、データＡ、Ｂ間の類似関係の強さＳがわかる。この例では、共通単語数が６、和集合の要素数が１１であるため、強さＳ＝６／１１である。

【００９８】
なお、頻度順位表１４１、１４２から頻度の少ない単語を除いた理由は、スライド辞書圧縮において頻度の少ない単語は一致文字列になる可能性が低く、頻度の少ない単語よりも多い単語の共通性を評価する方が、圧縮率を高くする上で効果的だからである。

【００９９】
なお、図１４の例では、単語のサイズを３バイトとしたが、それに限定しない。スライド辞書圧縮の多くが３～４バイト程度を一致文字列の最小検出長としているため、３～４バイト程度が望ましい。

【０１００】
全てのデータについて頻度順位表をその付属情報として管理し、互いの類似関係の強さＳの計算に用いると、ストレージ装置１１における管理情報の記憶量が多くなる。そこで

、全てのデータの頻度順位表を管理しなくても、その中のある２つのデータ間の類似関係の強さＳを近似的に評価できるｂ－ｂｉｔ　Ｍｉｎ－ｗｉｓｅ　ＨａｓｈというＬＳＨアルゴリズムを適用する。そのアルゴリズムを以下に示す。

【０１０１】

ｎ個のハッシュ関数を用意する。次に、データＡの頻度順位表１４１に出現する１０個の単語１４４をそれぞれｎ（ｎは２以上の整数）個のハッシュ関数に通してハッシュ値を計算する。この結果、データＡから１０×ｎ個のハッシュ値が生成される。

【０１０２】

そして、各ハッシュ関数ｋ（ｋは１～ｎ）からのハッシュ値１０個の中の最小値１４６を求め、ｎ個の最小値１４６を得る。それらｎ個の最小値１４６を並べてデータＡについての類似ハッシュ１４７を構成する。

【０１０３】

データＢにも同様の処理を行う。つまり、７×ｎ個のハッシュ値が生成され、各ハッシュ関数ｋからの７個のハッシュ値の中の最小値１４６を求め、ｎ個の最小値１４６を得る。それらｎ個の最小値１４６を並べてデータＢについての類似ハッシュ１４７を構成する。

【０１０４】

このようにして計算した各データＡ、Ｂについての２つの類似ハッシュ１４７のハミング距離をＨ、ハッシュ関数のビット長をｂとすると、類似関係の強さＳの近似値Ｊは、以下の式で与えられる。

【０１０５】

Ｊ＝｛（ｎ－Ｈ）／ｎ－（１／２）＾ｂ｝／｛１－（１／２）＾ｂ｝

ｎとｂが大きいほど、近似値Ｊの近似精度が向上する。

【０１０６】

以上の方法により、各データの類似ハッシュ１４７（各サイズはｎ×ｂビット）を管理しておけば、ある２つのデータ間の類似関係の強さＳを近似的に求めることができる。例えば、ｎ＝１６、ｂ＝８の場合、８ＫＢデータ当たり１６Ｂの付属情報（０．２％）を持つだけでよい。

【０１０７】

ある２つのデータの類似ハッシュ１４７のハミング距離Ｈ＝０であるとき、Ｊ＝１となる。つまり、両データの類似関係の強さＳは１００％である。ハミング距離Ｈ＝０（類似ハッシュ１４７が一致）のとき、両データは全く同じ内容である可能性がある。このため、重複排除処理のための重複データ検出に類似ハッシュ１４７を用いることも可能である。実際に重複しているか否かは、８ＫＢ全体を比較して判定する必要があるが、類似ハッシュ１４７が一致するか否かで、比較候補を絞り込むことができる。以上、類似ハッシュ算出方法の一例を説明したが、類似ハッシュ算出方法は、この方法に限定されない。

【０１０８】

図１５は、図２のストレージ装置が実行するデータリード処理を示すフローチャートである。

図１５において、リード処理は、図２のホスト計算機３０からネットワーク３１を介してストレージ装置１１がリード命令を受けた場合に開始される。リード命令では、例えば、ボリュームＩＤ、アドレス及びデータサイズが指定される。

【０１０９】

Ｓ１８０１において、プロセッサ２４は、指定アドレスから特定されるスロットの排他を確保する。なお、スロット排他確保時に他の処理がスロットの排他を確保している場合、プロセッサ２４は、一定の時間待ってから、Ｓ１８０１を行う。

【０１１０】

次に、Ｓ１８０２において、プロセッサ２４は、リード対象データがキャッシュ領域２０３に存在するか否かを判定し、Ｓ１８０２の判定結果が真の場合、Ｓ１８０７に進む。

【０１１１】

一方、Ｓ１８０２の判定結果が偽の場合、プロセッサ２４は、Ｓ１８０３において、ＲＡＩＤグループを構成するドライブから対象データ含む圧縮されたグループデータをリードする。この際、プロセッサ２４は、ホスト計算機３０が指定したボリュームＩＤとアドレスから、図６のプール割当管理テーブル２１０のプールＩＤ７３、プールアドレス７４及び圧縮後サイズ７６を特定し、図７のドライブ割当管理テーブル２１１からドライブＩＤ８４及びドライブアドレス８５を参照し、対象データを含む圧縮されたグループデータのドライブ上の格納場所及びサイズを特定する。

【０１１２】

次に、Ｓ１８０４において、プロセッサ２４は、ドライブからリードした圧縮されたグループデータをバッファ領域２０２にライトする。

【０１１３】

次に、Ｓ１８０５において、プロセッサ２４は、バッファ領域２０２上にライトした圧縮されたグループデータを伸張する。

【０１１４】

次に、Ｓ１８０６において、プロセッサ２４は、バッファ領域２０２上の伸張されたグループデータから指定サイズの対象データを抽出する。

【０１１５】

次に、Ｓ１８０７において、プロセッサ２４は、バッファ領域２０２上の対象データをホスト計算機３０に転送する。ホスト計算機３０は、このデータ転送が完了した時点でリード処理が終了したと認識する。

【０１１６】

次に、Ｓ１８０８において、プロセッサ２４は、Ｓ１８０１で確保していたスロット排他を解放する。

【０１１７】

なお、Ｓ１８０５の伸張処理は、プロセッサ２４の負荷低減のため、図２のオフロードエンジン２１４が高性能なハードウェアを用いて代わりに行ってもよい。その際に、オフロードエンジン２１４は、バッファ領域２０２の代わりにバッファ領域２１６を使うことにより、メモリ２５の転送帯域の消費を抑えることができる。

【０１１８】

図１６は、図２のストレージ装置が実行するデータライト処理を示すフローチャートである。なお、以下の説明では、例えば、図１のストレージコントローラ２２Ａに対応する図２のプロセッサ２４をプロセッサ２４Ａと記載するなど、図１の各ストレージコントローラ２２Ａ、２２Ｂに属するものをそれぞれ参照符号に付した「Ａ」及び「Ｂ」によって区別する。

【０１１９】

図１６において、データライト処理は、図２のホスト計算機３０からストレージ装置１１がライト命令を受信した場合に開始される。ライト命令では、例えば、ボリュームＩＤ、アドレス及びデータサイズが指定される。このデータライト処理は、図１の（Ｓ１）～（Ｓ５）で実行される。

【０１２０】

Ｓ１５０１において、プロセッサ２４Ａは、指定アドレスから特定されるスロットの排他を確保する。なお、スロット排他確保と同時に、プロセッサ２４Ａは、データのライト先とするキャッシュ領域２０３Ａの部分を割り当てる。

【０１２１】

次に、Ｓ１５０２において、プロセッサ２４Ａは、ホスト計算機３０に対してライト処理の準備ができたことを示す「Ｒｅａｄｙ」を応答する。プロセッサ２４Ａは、「Ｒｅａｄｙ」を受け取ったホスト計算機３０から、ライトデータを受信する。

【０１２２】

次に、Ｓ１５０３において、プロセッサ２４Ａは、ホスト計算機３０から受信したライトデータをキャッシュ領域２０３Ａに割り当てたスロット領域にライトする。

【０１２３】
次に、Ｓ１５０４において、プロセッサ２４Ａは、ストレージコントローラ２２Ａからストレージコントローラ２２Ｂに対してキャッシュ領域２０３Ａに格納したライトデータを転送し、キャッシュ領域２０３Ｂに格納することで二重化を行う。
【０１２４】
次に、Ｓ１５０５において、プロセッサ２４Ａは、図９のグループ管理テーブル２１２と図１０のメモリ割当管理テーブル２１３を更新する。ライトデータは、重複排除および類似グルーピング（両者を合わせて、圧縮前データ構築処理と呼ぶ）がまだ実施されていない。なお、類似グルーピングは、類似関係にあるデータのグルーピングである。そのため、プロセッサ２４Ａは、グループ管理テーブル２１２において、ライトデータのホストＶＯＬスロットに対応するシステムＶＯＬスロットのアドレス９０５、グループ番号９０６および位置情報９０７に“Ｎｏｎｅ”を設定する。また、プロセッサ２４Ａは、メモリ割当管理テーブル２１３で、ライトデータのホストＶＯＬスロットに対応するキャッシュ状態９１２において、ライトデータのキャッシュ状態を“Ｄｉｒｔｙ”に設定し、キャッシュアドレス９１３には、キャッシュ領域２０３Ａのライトデータ格納先アドレスを設定する。
【０１２５】
次に、Ｓ１５０６において、プロセッサ２４Ａは、ネットワーク３１を介してホスト計算機３０に対してライト処理が完了したとして完了応答を返却する。
【０１２６】
次に、Ｓ１５０７において、プロセッサ２４Ａは、Ｓ１５０１で確保していたスロットの排他を解放してライト処理を終了する。
【０１２７】
図１７は、図２のストレージ装置が実行する圧縮前データ構築処理を示すフローチャートである。
図１７において、圧縮前データ構築処理は、ライト処理が終了した後にプロセッサ２４が行う。圧縮前データ構築処理は、ライト処理終了直後に実行してもよいし、他に優先して行うべき処理があれば、その処理の後に実行してもよい。この圧縮前データ構築処理は、図１の（Ｓ６）～（Ｓ７）で実行される。
【０１２８】
Ｓ１６０１において、プロセッサ２４は、キャッシュ領域２０３に格納したホストＶＯＬ上のライトデータを選択する。
【０１２９】
次に、Ｓ１６０２において、プロセッサ２４は、選択したライトデータのアドレスから特定されるスロットの排他を確保する。
【０１３０】
次に、Ｓ１６０３において、プロセッサ２４は、例えば、図１４で説明した方法に従って、選択したデータの類似ハッシュを算出する。
【０１３１】
次に、Ｓ１６０４において、プロセッサ２４は、算出した類似ハッシュをハッシュテーブルに登録し、それまでに登録された１つ以上の類似ハッシュとの間でハミング距離Ｈを計算して、他のデータとの類似度を評価する。
【０１３２】
次に、Ｓ１６０５において、類似ハッシュが一致し、かつ内容も一致するデータがある場合は、プロセッサ２４は、そのデータとライトデータとの間で重複排除処理Ｓ１６０６を実行し、Ｓ１６１１に進む。
【０１３３】
重複排除処理Ｓ１６０６では、プロセッサ２４は、ライトデータと内容が一致するデータが含まれるシステムＶＯＬ上のスロットとグループを特定し、そのグループ内の一致するデータの位置番号を特定する。プロセッサ２４は、その特定結果に基づいて、Ｓ１６１

１で、図９のグループ管理テーブル２１２を更新する。すなわち、プロセッサ２４は、システムＶＯＬアドレス９０５、グループ番号９０６及び位置番号９０７の値が等しくなるように設定する。このとき、グループ管理テーブル２１２は、例えば、第５行目と第８行目のエントリの対象データは重複排除されていることを示す。

【０１３４】

一方、Ｓ１６０５において、類似ハッシュが一致するデータがない場合、または、類似ハッシュが一致し、かつ内容も一致するデータがない場合は、Ｓ１６０７に進む。

【０１３５】

次に、Ｓ１６０７において、プロセッサ２４は、ハミング距離Ｈの最小値があらかじめ規定した基準値以下であるかを判定する。判定結果が真であれば、プロセッサ２４は、類似関係にあるデータが存在すると判断し、Ｓ１６０８において、その最小値をもたらすデータが含まれるシステムＶＯＬ上のグループを選択し、Ｓ１６１０に進む。

【０１３６】

一方、Ｓ１６０７の判定結果が偽であれば、プロセッサ２４は、類似関係にあるデータは存在しないと判断し、Ｓ１６０９において、システムＶＯＬ上に新たにグループを作成する。この時、プロセッサ２４は、グループデータ格納先としてキャッシュ領域２０３の部分を割り当て、Ｓ１６１０に進む。

【０１３７】

次に、Ｓ１６１０において、プロセッサ２４は、システムＶＯＬ上の選択または作成されたグループに対応するキャッシュ領域部分に、ホストＶＯＬ上のライトデータを転送する。

【０１３８】

次に、Ｓ１６１０から遷移したＳ１６１１において、プロセッサ２４は、グループ管理テーブル２１２を更新する。すなわち、プロセッサ２４は、システムＶＯＬアドレス９０５に選択または作成したシステムＶＯＬ上のグループを含むスロットのアドレスを設定し、グループ番号９０６に当該グループの番号を設定し、位置番号９０７にライトデータが転送されたグループ内での位置を設定する。なお、ライトデータを転送したグループが、新しく作成したグループであり（つまり、Ｓ１６０９を経ており）、当該グループを含むスロットにおける最初のグループである場合は、さらに、プロセッサ２４は、メモリ割当管理テーブル２１３にそのスロットのエントリを追加する。すなわち、プロセッサ２４は、ＶＯＬ＿ＩＤ９０８にシステムＶＯＬのＩＤを設定し、ＶＯＬアドレス９０９に当該グループを含むスロットアドレスを設定する。また、ＢＦ転送状態９１０に“未”を、ＢＦアドレス９１１に“Ｎｏｎｅ”を設定する。また、キャッシュ状態９１２の当該グループのキャッシュ状態に“Ｄｉｒｔｙ”を設定する。また、キャッシュアドレス９１３にはキャッシュ領域２０３の割り当て部分のアドレスを設定する。

【０１３９】

次に、Ｓ１６１２において、プロセッサ２４は、ライトデータのアドレスから特定されるスロットの排他を解放して圧縮前データ構築処理を終了する。

【０１４０】

なお、Ｓ１６０３の類似ハッシュ算出、Ｓ１６０４のハッシュ登録・評価は、プロセッサ２４の負荷低減のため、図２のオフロードエンジン２１４が高性能なハードウェアを用いて代わりに行ってもよい。その際、オフロードエンジン２１４は、メモリ２１５をワークメモリとして使うことにより、メモリ２５の転送帯域の消費を抑えることができる。

【０１４１】

また、類似ハッシュを登録するためのハッシュテーブルは、キャッシュに乗っているデータの類似ハッシュを保持する。データがキャッシュから落とされるとき、そのデータの類似ハッシュをハッシュテーブルから削除する。

【０１４２】

また、Ｓ１６１０において、選択したグループのキャッシュ領域２０３にライトデータを格納する領域がない場合は、プロセッサ２４は、そのグループの所属データで最も他の

所属データとの類似性が低いデータを選択し、そのグループから除外してライトデータの格納先を確保する。すなわち、プロセッサ２４は、グループ管理テーブル２１２において、除外対象データのエントリのシステムＶＯＬアドレス９０５、グループ番号９０６及び位置番号９０７に“Ｎｏｎｅ”を設定する。これは、除外されたデータが再び重複排除および類似グルーピング（図１６参照）が施される準備ができたことを意味する。

【０１４３】

図１８は、図２のストレージ装置が実行する圧縮・デステージ処理を示すフローチャートである。

図１８において、圧縮・デステージ処理は、圧縮前データ構築処理が終了したグループデータについて、キャッシュ領域２０３の“Ｄｉｒｔｙ”比率が閾値以上となったことを契機として、非同期的に行われる。ただし、プロセッサ２４が低負荷な状態であれば、“Ｄｉｒｔｙ”比率が閾値未満であっても、デステージ処理を開始してもよい。また、“Ｄｉｒｔｙ”比率を問わず、タイマを用いて周期的にデステージ処理を開始してもよい。この圧縮・デステージ処理は、図１の（Ｓ８）～（Ｓ１０）で実行される。

【０１４４】

Ｓ１７０１において、プロセッサ２４は、キャッシュ領域２０３内で“Ｄｉｒｔｙ”状態のスロットに含まれるグループの中から、デステージ対象のグループを選択する。具体的には、プロセッサ２４は、上述の通り、ＬＲＵスロットに含まれるグループまたは最後のアクセスからの経過時間が長い複数スロットに含まれるグループのうち、最も類似度が高いグループを選択する。

【０１４５】

次に、Ｓ１７０２において、プロセッサ２４は、デステージ対象のグループが属するシステムＶＯＬスロットの排他を確保し、さらに、デステージするグループに転送されたデータが属するホストＶＯＬスロットの排他を確保する。

【０１４６】

次に、Ｓ１７０３において、プロセッサ２４は、デステージ対象のグループデータ（１２８ＫＢ）を読み出して、可逆圧縮処理を行う。

【０１４７】

次に、Ｓ１７０４において、プロセッサ２４は、圧縮されたグループデータ（１２８ＫＢ未満）の格納先としてバッファ領域２０２の部分を割り当て、その格納先へ圧縮グループデータをライトする。なお、バッファ領域２０２のアドレス割り当てでは、パリティサイクル分の圧縮グループが集まるまで割り当てを繰り返すことが明らかなため、あらかじめパリティサイクル分のアドレスを一括して割り当ててもよい。

【０１４８】

次に、Ｓ１７０５において、プロセッサ２４は、図１０のメモリ割当管理テーブル２１３のＢＦ転送状態９１０を“済”に更新する。また、ＢＦアドレス９１１に、割り当てたバッファ部分のアドレスを設定する。

【０１４９】

次に、Ｓ１７０６において、プロセッサ２４は、Ｓ１７０２で確保したスロットの排他を解放する。

【０１５０】

次に、Ｓ１７０７において、プロセッサ２４は、バッファ内の圧縮グループデータの蓄積量を計算する。圧縮グループデータの蓄積量がパリティサイクル分よりも小さい場合、プロセッサ２４は、Ｓ１７０１に戻ってデステージ対象のグループを追加で選択する。

【０１５１】

一方、パリティサイクル分の圧縮グループデータがバッファ領域２０２内に溜まった場合は、Ｓ１７０８に進む。なお、圧縮グループデータのサイズは可変であるため、バッファ領域２０２内の蓄積量が必ずしもパリティサイクル分に一致するとは限らない。パリティサイクル分に達する前にＳ１７０８へ処理を進めることもあり得る。

【０１５２】

次に、Ｓ１７０８において、プロセッサ２４は、バッファ領域２０２内に蓄積された圧縮グループデータ列からパリティを生成する。

【０１５３】

次に、Ｓ１７０９において、プロセッサ２４は、圧縮グループデータ列及び生成したパリティを、ＲＡＩＤグループを構成するドライブ２９に書き出す。プロセッサ２４は、ドライブ２９への書き出し後に、バッファ領域の使用していた部分を解放する。

【０１５４】

次に、Ｓ１７１０において、プロセッサ２４は、デステージ対象のグループを含むシステムＶＯＬスロットの排他を確保し、さらに、デステージ対象のグループに転送されたデータが属するホストＶＯＬスロットの排他を確保する。

【０１５５】

次に、Ｓ１７１１において、プロセッサ２４は、メモリ割当管理テーブル２１３で、デステージ対象のグループを含むシステムＶＯＬスロットのキャッシュ状態９１２において、当該グループのキャッシュ状態を“Ｃｌｅａｎ”に更新する。その結果、プロセッサ２４は、２個のグループのキャッシュ状態がともに“Ｃｌｅａｎ”となれば、ＢＦアドレス９１１を“Ｎｏｎｅ”に設定し、バッファ領域２０２の使用部分を解放する。このシステムＶＯＬスロットは“Ｃｌｅａｎ”状態となり、いつでもキャッシュから解放して“Ｆｒｅｅ”状態のキャッシュリソースを作ることができる。さらに、プロセッサ２４は、メモリ割当管理テーブル２１３で、デステージ対象のグループに転送された各データが属するホストＶＯＬスロットのキャッシュ状態９１２において、当該データのキャッシュ状態を“Ｃｌｅａｎ”に更新する。その結果、キャッシュ状態９１２が全て“Ｃｌｅａｎ”となったホストＶＯＬスロットは、いつでもキャッシュから解放して“Ｆｒｅｅ”状態のキャッシュリソースを作ることができる。

【０１５６】

次に、Ｓ１７１２において、プロセッサ２４は、デステージ対象のグループを含むホストＶＯＬスロットの排他を解放し、さらに、デステージ対象のグループに転送されたデータが属するホストＶＯＬスロットの排他を解放し、処理を終了する。

【０１５７】

以上の圧縮・デステージ処理において、キャッシュから解放されるグループは、アクセス頻度が低く、類似関係が強いデータが優先される。従って、ホスト計算機３０からのアクセスがキャッシュヒットする確率がより高くなるため、アクセス性能が向上する。また、ドライブ２９に格納されるデータの圧縮率がより高くなるため、データ削減効率が向上する。

【０１５８】

なお、Ｓ１７０３の可逆圧縮は、プロセッサ２４の負荷低減のため、図２のオフロードエンジン２１４が高性能なハードウェアを用いて代わりに行ってもよい。その際、オフロードエンジン２１４は、メモリ２１５のバッファ領域２１６をバッファ領域２０２の代わりとして使うことにより、メモリ２５の転送帯域の消費を抑えることができる。

【０１５９】

図１９は、図２のストレージ装置が実行するデータ削除リトライ処理を示すフローチャートである。

図１９において、データ削減リトライ処理は、図２のドライブ２９にデステージした圧縮グループを構成するデータを再度キャッシュ領域２０３に戻して、重複排除および類似グルーピングをリトライできる状態にする。デステージしたグループデータの類似度７７が悪かった場合、このデータ削減リトライ処理によって、前より類似関係の強いデータを含むグループが構築され得るため、ストレージ装置１１のデータ削減率を改善することができる。

【０１６０】

Ｓ１９０１において、プロセッサ２４は、図６のプール割当管理テーブル２１０で管理されているシステムＶＯＬスロットの１つを選択する。

【０１６１】
次に、Ｓ１９０２において、プロセッサ２４は、プール割当管理テーブル２１０で、そのシステムＶＯＬスロットに含まれるグループの類似度７７が、あらかじめ規定した基準値以下であるかを判定する。
【０１６２】
Ｓ１９０２の判定結果が偽であれば、プロセッサ２４は、そのシステムＶＯＬスロットのデータ削減は不要として処理を終了する。判定結果が真であれば、Ｓ１９０３に進む。
【０１６３】
次に、Ｓ１９０３において、プロセッサ２４は、そのシステムＶＯＬスロットの排他を確保する。さらに、図９のグループ管理テーブル２１２を参照し、当該システムＶＯＬスロット内の類似度が基準値以下のグループを構成するホストＶＯＬスロットのデータを特定し、そのホストＶＯＬスロットの排他も確保する。
【０１６４】
次に、Ｓ１９０４において、プロセッサ２４は、図１０のメモリ割当管理テーブル２１３を参照し、特定したデータがキャッシュ領域２０３に存在するか否かを判定する。
【０１６５】
Ｓ１９０４の判定結果が真の場合、プロセッサ２４は、メモリ割当管理テーブル２１３において、そのデータのキャッシュ状態９１２に“Ｄｉｒｔｙ”を設定する。また、グループ管理テーブル２１２において、当該ホストＶＯＬスロットのデータのエントリのシステムＶＯＬアドレス９０５、グループ番号９０６及び位置番号９０７に“Ｎｏｎｅ”を設定し、既存のシステムＶＯＬスロットのグループとの対応関係を切る。これは、当該データが重複排除および類似グルーピング（図１６参照）が施される準備ができたことを意味する。そして、Ｓ１９０９に進む。
【０１６６】
一方、Ｓ１９０４の判定結果が偽の場合、プロセッサ２４は、Ｓ１９０５において、ＲＡＩＤグループのドライブから、Ｓ１９０２で類似度が基準値以下と判定した圧縮グループデータをリードする。この際、プロセッサ２４は、図６のプール割当管理テーブル２１０のプールＩＤ７３、プールアドレス７４及び圧縮後サイズ７６を特定し、ドライブ割当管理テーブル２１１からドライブＩＤ８４及びドライブアドレス８５を参照し、その圧縮グループデータのドライブ上の格納場所及びサイズを特定する。
【０１６７】
次に、Ｓ１９０６において、プロセッサ２４は、その圧縮グループデータをバッファ領域２０２にライトする。
【０１６８】
次に、Ｓ１９０７において、プロセッサ２４は、バッファ領域２０２上の圧縮グループデータを伸張する。
【０１６９】
次に、Ｓ１９０８において、プロセッサ２４は、キャッシュ領域２０３内に当該データの格納用の領域を割り当て、バッファ領域２０２上の伸張されたグループから抽出したデータを、その割り当て領域にライトする。この際、プロセッサ２４は、メモリ割当管理テーブル２１３のＶＯＬ＿ＩＤ９０８に当該データの属するホストＶＯＬのＩＤを設定し、ＶＯＬアドレス９０９に当該データの属するホストＶＯＬスロットのアドレスを設定し、キャッシュ状態９１２に“Ｄｉｒｔｙ”を設定し、キャッシュアドレス９１３にその割り当て領域のアドレスを登録する。また、プロセッサ２４は、グループ管理テーブル２１２において、当該ホストＶＯＬスロットのデータのエントリを削除し、既存のシステムＶＯＬスロットのグループとの対応関係を切る。これは、当該データは重複排除や類似グルーピング（図１６参照）が施される準備ができたことを意味する。そして、Ｓ１９０９に進む。
【０１７０】
次に、Ｓ１９０９において、プロセッサ２４は、Ｓ１９０３で確保したシステムＶＯＬ

スロット及びホストＶＯＬスロットの排他も解放し、データ削減リトライ処理を終了する。

【０１７１】

なお、Ｓ１９０７の伸張処理は、プロセッサ２４の負荷低減のため、図２のオフロードエンジン２１４が高性能なハードウェアを用いて代わりに行ってもよい。その際、オフロードエンジン２１４は、メモリ２１５のバッファ領域２１６をバッファ領域２０２の代わりとして使うことにより、メモリ２５の転送帯域の消費を抑えることができる。

【０１７２】

図２０は、図２のストレージ装置におけるデータ伸張処理の過程を示す図である。

図２０において、プロセッサ２４は、類似関係にある複数の８ＫＢデータをグルーピングする処理において、データの並び順と伸張処理を最適化することで、データリード処理の性能を改善する。

【０１７３】

データリード処理では、ホスト計算機３０からリード命令を受けたデータがキャッシュ領域２０３にはなく、ドライブ２９に格納されていた場合、プロセッサ２４は、圧縮されたグループデータ１１０３を伸張して、１２８ＫＢサイズのグループ１１０２を復元し、その復元したデータからリード対象の８ＫＢデータを抽出してホスト計算機３０に応答する。

【０１７４】

このとき、プロセッサ２４は、圧縮されたグループデータの伸張処理において、先頭の８ＫＢデータから順に復元する。リード対象データが１１００Ｃの場合、プロセッサ２４は、復元までに範囲２００１の部分を伸張すればよい。リード対象データが１１００Ｄの場合、プロセッサ２４は、復元までに範囲２００２の部分を伸張すればよい。すなわち、リード対象データがグループデータの先頭に近いほど、プロセッサ２４は、短い時間でデータを復元できる。

【０１７５】

プロセッサ２４は、この特性を利用して、リード対象データが復元できた時点でグループデータの伸張処理を中断する。さらに、プロセッサ２４は、ホストＶＯＬのアドレス空間を複数の区間に分けて、所定時間範囲（例えば、過去１０時間）におけるそれぞれの区間のアクセス頻度を調べる。

【０１７６】

プロセッサ２４は、類似グルーピングの際に、ホスト計算機３０からのアクセス頻度が高い区間にあるデータほど、グループの先頭に近い（位置番号が小さい）位置に配置する。アクセス頻度が低い区間にあるデータほど、グループの末尾に近い（位置番号が大きい）位置に配置する。これにより、ストレージ装置１１は、ホスト計算機３０からのリード命令に対して、平均応答時間を短縮することができる。

【０１７７】

なお、本発明は上記した実施形態に限定されるものではなく、様々な変形例が含まれる。例えば、上記した実施形態は、本発明のより良い理解のために詳細に説明したのであり、必ずしも説明の全ての構成を備えるものに限定されものではない。

【０１７８】

また、上記の各構成、機能、処理部、処理手段等は、それらの一部又は全部を、例えば集積回路で設計する等によってハードウェアで実現してもよい。すなわち、類似ハッシュ算出、ハッシュ登録、ハッシュ評価及び可逆圧縮・伸張処理をオフロードエンジンの高速ハードウェアが代行して処理するという変形例以外の実施形態もある。

【０１７９】

また、上記の各構成、機能等は、プロセッサがそれぞれの機能を実現するプログラムを解釈し、実行することによってソフトウェアで実現してもよい。各機能を実現するプログラム、テーブル、ファイル等の情報は、不揮発性半導体メモリ、ＨＤＤ、ＳＳＤ等の記憶デバイス、または、ＩＣカード、ＳＤカード、ＤＶＤ等の計算機読み取り可能な非一時的

データ記憶媒体に格納することができる。

【０１８０】

また、制御線及び情報線は説明上必要と考えられるものを示しており、製品上必ずしも全ての制御線及び情報線を示しているとは限らない。実際には、ほとんど全ての構成が相互に接続されていると考えてもよい。

【０１８１】

なお、本発明は上記した実施形態に限定されるものではなく、様々な変形例が含まれる。例えば、上記した実施形態は本発明を分かりやすく説明するために詳細に説明したものであり、必ずしも説明した全ての構成を備えるものに限定されるものではない。また、ある実施形態の構成の一部を他の実施形態の構成に置き換えることが可能であり、また、ある実施形態の構成に他の実施形態の構成を加えることも可能である。また、各実施形態の構成の一部について、他の構成の追加・削除・置換をすることが可能である。また、上記の各構成、機能、処理部、処理手段等は、それらの一部又は全部を、例えば集積回路で設計する等によりハードウェアで実現してもよい。

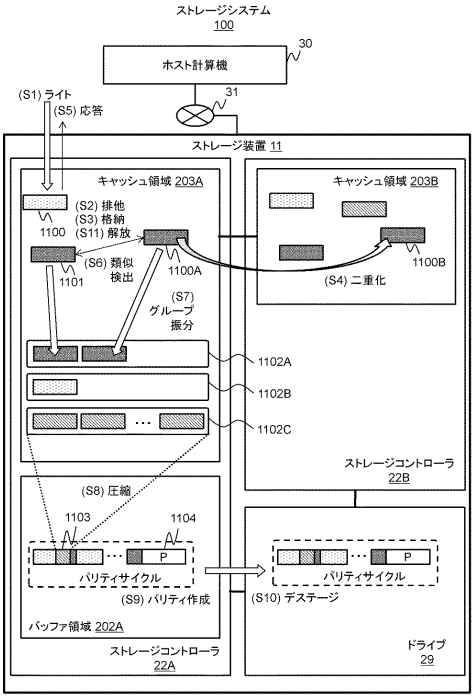
【符号の説明】

【０１８２】

１００　ストレージシステム、１１　ストレージ装置、２２、２２Ａ、２２Ｂ　ストレージコントローラ、２４　プロセッサ、２５、２１５　メモリ、２０２、２１６　バッファ領域、２０３、２０３Ａ、２０３Ｂ　キャッシュ領域、２１４　オフロードエンジン、２９　ドライブ、３０　ホスト計算機、３１　ネットワーク

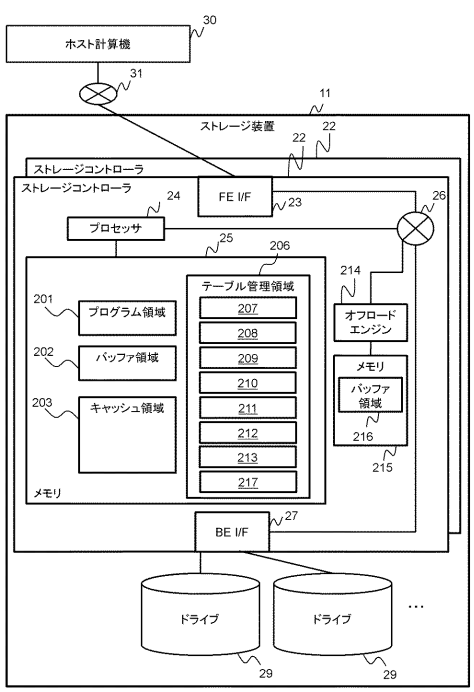
【図１】

図1



【図２】

図2



【図３】

図3

ボリューム管理テーブル
207

41 VOL ID	42 VOL属性	43 VOL容量	44 プール ID	45 ホスト 可視
0	シンプロビジョニング、削減OFF	100GB	0	Yes
10	シンプロビジョニング、削減ON	200GB	0	Yes
20	通常割当、削減OFF	500GB	1	Yes
30	システム（グループ管理用）	200GB	0	No
…	…	…	…	

【図４】

図4

プール構成管理テーブル
208

51 プールID	52 RAIDグループID	53 プール容量	54 プール使用容量
0	0 1	10TB	5TB
…	…	…	…

【図５】

図5

RAID構成管理テーブル
209

61 RAIDグループID	62 RAIDレベル	63 ドライブID	64 ドライブ種別	65 容量	66 使用容量
0	RAID5	0 1 2 3	SSD	5TB	3TB
…		…	…	…	…

【図６】

図6

プール割当管理テーブル
210

71 VOL ID	72 VOL アドレス	73 プール ID	74 プール アドレス	75 圧縮前 サイズ	76 圧縮後 サイズ	77 類似度
30	1000	0	05610	128KB	54KB	84%
		0	17480	128KB	29KB	97%
30	1300	0	07140	128KB	73KB	71%
		0	14990	128KB	61KB	80%
30	1500	0	06340	128KB	92KB	54%
		0	04720	128KB	77KB	66%
30	1800	0	11390	128KB	42KB	90%
		0	10570	128KB	39KB	92%
…	…	…	…	…	…	

【図７】

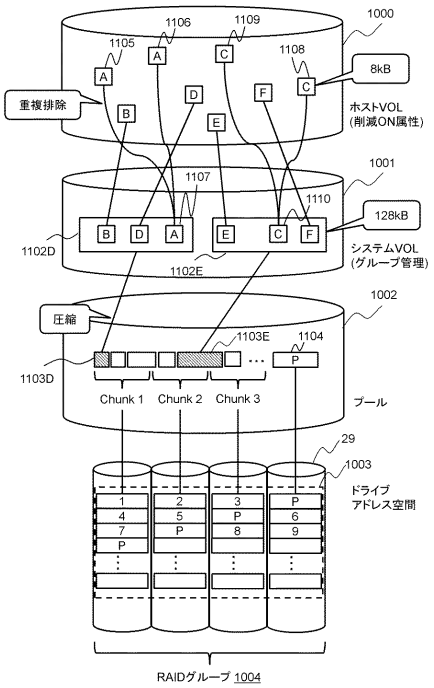
図7

ドライブ割当管理テーブル
211

81 プールID	82 プールアドレス	83 RAIDグループID	84 ドライブID	85 ドライブアドレス
0	05610	0	0	0360
0	17480	1	2	1530
0	07140	1	3	0940
0	14990	1	3	0730
0	06340	0	1	1830
0	04720	0	0	1220
0	11390	1	2	0790
…	…	…	…	…

【図８】

図8



【図９】

図9

グループ管理テーブル
212

901 ホスト VOL ID	902 ホスト VOL アドレス	903 位置 番号	904 システム VOL ID	905 システム VOL アドレス	906 グループ 番号	907 位置 番号
10	700	11	30	1300	1	7
10	500	8	30	1800	0	2
10	300	5	30	None	None	None
10	100	19	30	1500	1	10
10	200	10	30	1000	1	3
10	400	31	30	1800	0	5
10	600	17	30	1500	1	8
10	800	9	30	1000	1	3
…	…		…	…		…

【図１０】

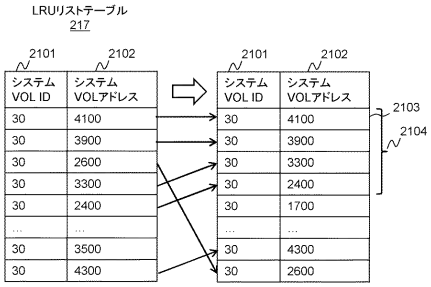
図10

メモリ割当管理テーブル
213

908 VOL ID	909 VOL アドレス	910 BF転送 状態	911 BF アドレス	912 キャッシュ 状態	913 キャッシュ アドレス
10	300			[Dirty, Dirty, … , Dirty]	2000
10	200			[Dirty, Dirty, … , Dirty]	2200
10	500			[Dirty, Dirty, … , Dirty]	2500
30	1300	未	None	[Dirty, Clean]	3000
30	1800	済	50	[Dirty, Dirty]	2100
30	1500	済	None	[Clean, Clean]	2300
30	1000	済	80	[Dirty, Clean]	3200
…	…	…	…	…	

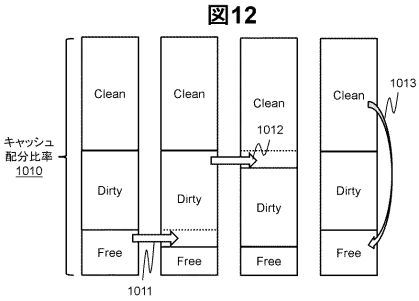
【図１１】

図11



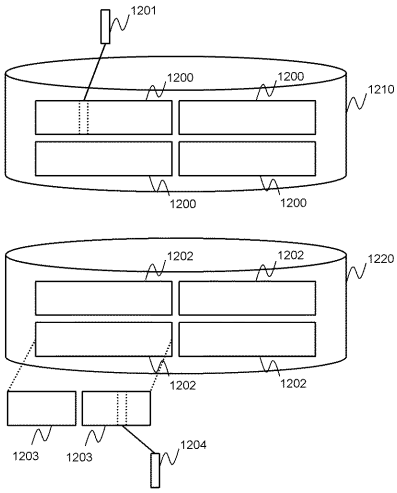
【図１２】

図12



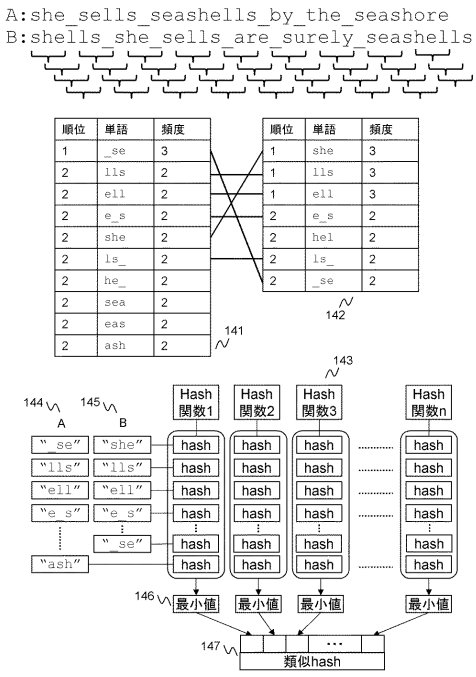
【図１３】

図13



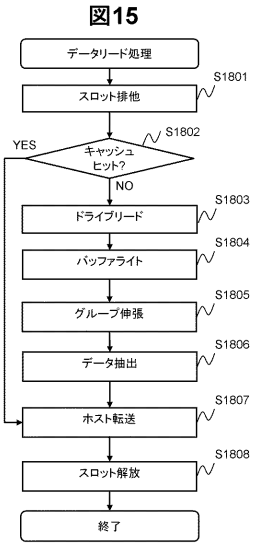
【図１４】

図14



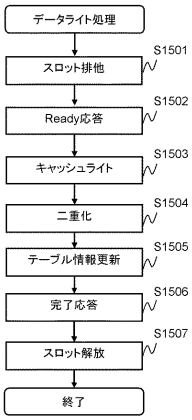
【図１５】

図15



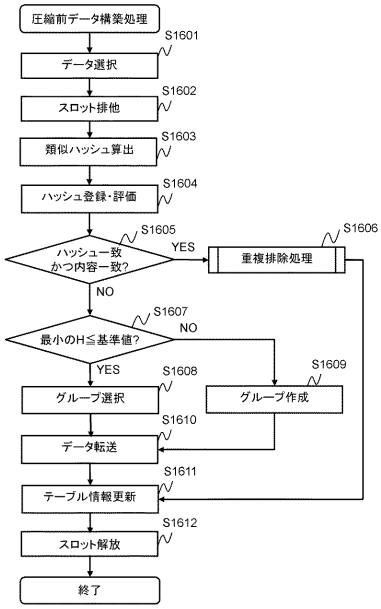
【図１６】

図16



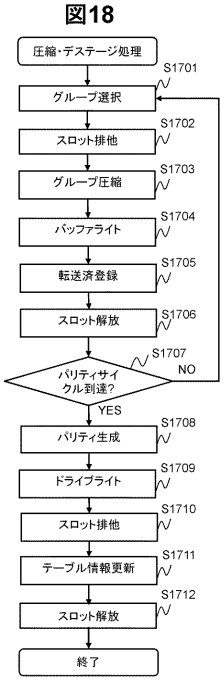
【図１７】

図17



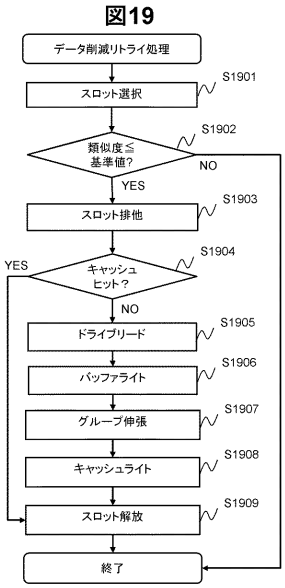
【図１８】

図18



【図１９】

図19



【図２０】

図20