

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号
特許第5112003号
(P5112003)

(45) 発行日 平成25年1月9日 (2013.1.9)

(24) 登録日 平成24年10月19日 (2012.10.19)

(51) Int.Cl.

G 0 6 F 3 / 0 6 (2006.01)

F I

G O 6 F 3 / 0 6 3 O 1 W

G O 6 F 3 / 0 6 5 4 O

請求項の数 7 (全 29 頁)

(21) 出願番号	特願2007-272900 (P2007-272900)	(73) 特許権者	000005108
(22) 出願日	平成19年10月19日 (2007.10.19)		株式会社日立製作所
(65) 公開番号	特開2009-104236 (P2009-104236A)		東京都千代田区丸の内一丁目6番6号
(43) 公開日	平成21年5月14日 (2009.5.14)	(74) 代理人	100093861
審査請求日	平成22年3月3日 (2010.3.3)		弁理士 大賀 真司
		(72) 発明者	武藤 淳一
			神奈川県小田原市中里322番2号 株式
			会社日立製作所 R A I D システム事業部内
		(72) 発明者	黒川 勇
			神奈川県小田原市中里322番2号 株式
			会社日立製作所 R A I D システム事業部内
		(72) 発明者	緒方 蘭
			神奈川県小田原市中里322番2号 株式
			会社日立製作所 R A I D システム事業部内

最終頁に続く

(54) 【発明の名称】 ストレージ装置及びこれを用いたデータ格納方法

(57) 【特許請求の範囲】

【請求項 1】

データを格納する記憶媒体を有するディスクデバイスと、前記ディスクデバイスを制御するディスクコントローラと、を備えるストレージ装置であって、

前記ディスクコントローラは、
ホスト装置に接続するためのチャネルアダプタと、
前記ディスクデバイスに接続されたディスクアダプタと、
前記チャネルアダプタと前記ディスクアダプタとの間でやり取りされるデータを一時的に格納するキャッシュメモリと、を備え、

前記ディスクアダプタは、
前記記憶媒体の固有記憶容量よりも大きな記憶容量が定義され、前記記憶媒体の記憶領域に対応付けられた実ボリューム及び前記実ボリューム以外の記憶領域が割り当てられた仮想ボリュームを含むデータボリュームと、前記仮想ボリュームに割り当てられた記憶領域に対するデータを格納するための記憶領域が割り当てられたプールボリュームとを形成するように、前記ディスクデバイスを制御し、

前記ディスクアダプタは、
所定の R A I D 構成による制御の下、前記ホスト装置から送られた書き込み要求に従う書き込みデータに基づいて、当該書き込みデータが属するパリティグループ内のパリティデータを生成し、

前記パリティグループ内の前記書き込みデータ及び前記生成されたパリティデータのそ

れぞれを圧縮して、圧縮データ及び圧縮パリティデータを生成し、

前記パリティグループ内の前記圧縮データ及び前記圧縮パリティデータのそれぞれが所定のサイズを超えるか否かを判断し、

前記圧縮データ及び前記圧縮パリティデータのうち前記所定のサイズを超えない分のデータを前記実ボリュームの記憶領域に格納し、前記圧縮データ及び前記圧縮パリティデータのうち前記所定のサイズを超える分のデータを前記仮想ボリュームに対応する前記プールボリュームの記憶領域に格納することを特徴とするストレージ装置。

【請求項 2】

請求項 1 記載のストレージ装置であって、

前記ディスクアダプタは、

前記書き込みデータを構成するデータブロック及び前記パリティグループ内の他のデータブロックに基づいて前記パリティデータのデータブロックを生成し、

前記パリティグループ内の前記データブロックのそれぞれに対して圧縮することにより得られる圧縮データ列が所定数を超える圧縮データブロックにより構成されているか否かに従って、前記所定のサイズを超えているか否かを判断し、

前記所定数を超えていない前記圧縮データブロックを前記実ボリュームの記憶領域に格納し、前記所定数を超えている前記圧縮データブロックを前記仮想ボリュームに対応する前記プールボリュームの記憶領域に格納することを特徴とするストレージ装置。

【請求項 3】

請求項 2 記載のストレージ装置であって、

前記ディスクアダプタは、

前記パリティグループ内の前記圧縮データブロックのそれぞれを、前記データボリュームを構成する物理的に異なる系列のディスクデバイスに分散させて格納することを特徴とするストレージ装置。

【請求項 4】

請求項 3 記載のストレージ装置であって、

前記ディスクアダプタは、

前記圧縮データ列を構成する前記圧縮データブロックは、前記データボリュームを構成する同一系列のディスクデバイスに形成される実ボリューム及び仮想ボリュームに格納することを特徴とするストレージ装置。

【請求項 5】

請求項 2 記載のストレージ装置であって、

前記パリティグループ内の前記データブロックのそれぞれを格納する、前記キャッシュメモリ上に形成された記憶領域のサイズは、前記実ボリュームの単位記憶領域のサイズよりも大きいことを特徴とするストレージ装置。

【請求項 6】

請求項 1 記載のストレージ装置であって、

前記ディスクアダプタは、

前記書き込み要求に従う書き込みデータに対する RAID 構成によるデータ格納方式を記憶し、当該記憶されたデータ格納方式に従って、前記書き込みデータを圧縮して、前記データボリュームに格納することを特徴とするストレージ装置。

【請求項 7】

データを格納する記憶媒体を有するディスクデバイスと、前記ディスクデバイスを制御するディスクコントローラと、を備えるストレージ装置におけるデータ格納方法であって、

前記データ格納方法は、

前記ディスクコントローラが、前記記憶媒体の固有記憶容量よりも大きな記憶容量が定義され、前記記憶媒体の記憶領域に対応付けられた実ボリューム及び前記実ボリューム以外の記憶領域が割り当てられた仮想ボリュームを含むデータボリュームと、前記仮想ボリュームに割り当てられた記憶領域に対するデータを格納するための記憶領域が割り当てら

10

20

30

40

50

れたプールボリュームとが形成されるように、前記ディスクデバイスを制御するステップと、

前記ディスクコントローラが、所定のＲＡＩＤ構成に従い、前記ホスト装置から送られた書き込み要求に従う書き込みデータに基づいて、当該書き込みデータが属するパリティグループ内のパリティデータを生成するステップと、

前記ディスクコントローラが、前記パリティグループ内の前記書き込みデータ及び前記生成されたパリティデータのそれぞれを圧縮して、圧縮データ及び圧縮パリティデータを生成するステップと、

前記ディスクコントローラが、前記パリティグループ内の前記圧縮データ及び前記圧縮パリティデータのそれぞれが所定のサイズを超えるか否かを判断するステップと、

前記ディスクコントローラが、前記圧縮データ及び前記圧縮パリティデータのうち前記所定のサイズを超えない分のデータを前記実ボリュームの記憶領域に格納するとともに、前記圧縮データ及び前記圧縮パリティデータのうち前記所定のサイズを超える分のデータを前記仮想ボリュームに対応する前記プールボリュームの記憶領域に格納するステップと、を含むことを特徴とするデータ格納方法。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、ストレージ装置及びこれを用いたデータ格納方法に関し、特に、ＲＡＩＤ制御の下、書き込みデータを圧縮して効率的に格納するストレージ技術に関する。

【背景技術】

【０００２】

ストレージ装置の限られた記憶資源を有効に活用するため、データを圧縮して、格納する技術が知られている。データを圧縮して格納する場合、圧縮後のデータのサイズは不定であり、圧縮前のデータの内容に大きく依存する。

【０００３】

下記特許文献１は、ディスクドライブにデータを圧縮して格納する際に、所定の格納エリアに格納しきれない圧縮データの一部をオーバーフロー用の格納エリアに格納する記憶装置を開示している。具体的には、特許文献１において、主処理装置からのデータをディスクドライブに書き込む場合に、当該データがデータ圧縮手段により圧縮され、書き込み先アドレスに基づいて、圧縮データがディスクドライブ内の固定サイズ格納エリアの該当ブロックに格納できるか否かを判定し、格納できないと判定される場合に、当該圧縮データの一部が該当ブロックに格納されるとともに、当該圧縮データの残りがオーバーフロー部格納エリアに格納される。

【０００４】

また、近年では、大容量かつ高信頼性等の観点から、ＲＡＩＤ（Redundant Arrays of Independent Disks）技術を採用したストレージ装置が主流を占めつつある。ＲＡＩＤには、ハードディスクドライブの構成に応じて、ＲＡＩＤレベル（ＲＡＩＤ０～６）が定義されている。例えば、ＲＡＩＤ１は、複数台のハードディスクドライブに、同時に同じ内容のデータを書き込む（ミラーリング）方式である。また、ＲＡＩＤ５は、ブロック単位のデータを複数のハードディスクドライブに誤り訂正符号データ（パリティデータ）とともに分散させて書き込む方式である。

【特許文献１】特開平５－１８９１５７号公報

【発明の開示】

【発明が解決しようとする課題】

【０００５】

ＲＡＩＤ構成は、実用的には、総合的なパフォーマンスに優れたＲＡＩＤ５の構成が主流である。ＲＡＩＤ５では、上述したように、耐障害性の観点から、パリティデータが存在する。従って、ＲＡＩＤ５の構成の下、書き込みデータを圧縮して格納する場合には、そのパリティデータも圧縮して格納することになる。パリティデータは、書き込みデータ

10

20

30

40

50

に比較してデータビットのランダム性が一般に高い傾向にあり、高い圧縮効率が期待できない。このため、圧縮効率の低下によるデータのオーバーフローが絶えず発生する可能性が高く、単に、従前のオーバーフロー用格納エリアを用意したのでは、すぐに使い切ってしまうおそれがある。また、予め十分なサイズのオーバーフロー用格納エリアを用意したのでは、ランニングコストが高くなり、経済的でない。

【 0 0 0 6 】

そこで、本発明は、所定の R A I D 構成の下で、書き込みデータに基づく圧縮データを効率的に格納することができるストレージ装置を提案することを目的としている。

【 0 0 0 7 】

具体的には、本発明は、圧縮データが所定の記憶領域に格納しきれない場合に、当該圧縮データのうち格納しきれない分（あふれ分）を、使用実績に応じた柔軟な拡張を可能にする記憶領域（拡張記憶領域）に格納するストレージ装置を提案する。

10

【 0 0 0 8 】

また、本発明は、ある R A I D 構成の下で、圧縮データの効率的な格納が期待できない場合に、効率的な格納が期待できる他の R A I D 構成により当該圧縮データを格納するストレージ装置を提案する。つまり、本発明は、パリティデータに対する圧縮効率に応じて、圧縮データを格納する最適な R A I D 構成を選択するストレージ装置を提案する。

【課題を解決するための手段】

【 0 0 0 9 】

上記課題を解決するために、本発明に係るストレージ装置は、以下の技術的特徴を含んで構成される。

20

【 0 0 1 0 】

すなわち、第 1 の観点に従う本発明は、データを格納する記憶媒体を有するディスクデバイスと、前記ディスクデバイスを制御するディスクコントローラと、を備えるストレージ装置である。前記ディスクコントローラは、ホスト装置に接続するためのチャネルアダプタと、前記ディスクデバイスに接続されたディスクアダプタと、前記チャネルアダプタと前記ディスクアダプタとの間でやり取りされるデータを一時的に格納するキャッシュメモリと、を備える。また、前記ディスクアダプタは、前記記憶媒体の固有記憶容量よりも大きな記憶容量が定義され、前記記憶媒体の記憶領域に対応付けられた実ボリューム及び前記実ボリューム以外の記憶領域が割り当てられた仮想ボリュームを含むデータボリュームと、前記仮想ボリュームに割り当てられた記憶領域に対するデータを格納するための記憶領域が割り当てられたプールボリュームとを形成するように、前記ディスクデバイスを制御する。

30

【 0 0 1 1 】

前記ディスクアダプタは、所定の R A I D 構成による制御の下、前記ホスト装置から送られた書き込み要求に従う書き込みデータに基づいて、当該書き込みデータが属するパリティグループ内のパリティデータを生成し、前記パリティグループ内の前記書き込みデータ及び前記生成されたパリティデータのそれぞれを圧縮して、圧縮データ及び圧縮パリティデータを生成する。また、前記ディスクアダプタは、前記パリティグループ内の前記圧縮データ及び前記圧縮パリティデータのそれぞれが所定のサイズを超えるか否かを判断し、前記圧縮データ及び前記圧縮パリティデータのうち前記所定のサイズを超えない分のデータを前記実ボリュームの記憶領域に格納し、前記圧縮データ及び前記圧縮パリティデータのうち前記所定のサイズを超える分のデータを前記仮想ボリュームに対応する前記プールボリュームの記憶領域に格納する。

40

【 0 0 1 2 】

また、本発明は、方法の発明として把握することができる。すなわち、第 2 の観点に従う本発明は、データを格納する記憶媒体を有するディスクデバイスと、前記ディスクデバイスを制御するディスクコントローラと、を備えるストレージ装置におけるデータ格納方法である。

【 0 0 1 3 】

50

前記データ格納方法は、前記ディスクコントローラが、前記記憶媒体の固有記憶容量よりも大きな記憶容量が定義され、前記記憶媒体の記憶領域に対応付けられた実ボリューム及び前記実ボリューム以外の記憶領域が割り当てられた仮想ボリュームを含むデータボリュームと、前記仮想ボリュームに割り当てられた記憶領域に対するデータを格納するための記憶領域が割り当てられたプールボリュームとが形成されるように、前記ディスクデバイスを制御するステップと、前記ディスクコントローラが、所定のRAID構成に従い、前記ホスト装置から送られた書き込み要求に従う書き込みデータに基づいて、当該書き込みデータが属するパリティグループ内のパリティデータを生成するステップと、前記ディスクコントローラが、前記パリティグループ内の前記書き込みデータ及び前記生成されたパリティデータのそれぞれを圧縮して、圧縮データ及び圧縮パリティデータを生成するステップと、前記ディスクコントローラが、前記パリティグループ内の前記圧縮データ及び前記圧縮パリティデータのそれぞれが所定のサイズを超えるか否かを判断するステップと、前記ディスクコントローラが、前記圧縮データ及び前記圧縮パリティデータのうち前記所定のサイズを超えない分のデータを前記実ボリュームの記憶領域に格納するとともに、前記圧縮データ及び前記圧縮パリティデータのうち前記所定のサイズを超える分のデータを前記仮想ボリュームに対応する前記プールボリュームの記憶領域に格納するステップと、を含む。

10

【発明の効果】**【0014】**

本発明によれば、所定のRAID構成の下で、書き込み圧縮データに基づく圧縮データを効率的に格納することができるようになる。

20

【0015】

また、本発明によれば、圧縮データが所定の記憶領域に格納しきれない場合に、拡張記憶領域に格納しているので、記憶領域の使用実績に応じた柔軟なディスクデバイスの構成を採用することができるようになる。

【0016】

さらに、本発明によれば、パリティを含むデータに対する圧縮効率に応じて最適なRAID構成を選択するので、圧縮データの効率的な格納ができるようになる。

【発明を実施するための最良の形態】**【0017】**

次に、本発明の実施の形態について、図面を参照しつつ説明する。

30

【第1の実施形態】**【0018】**

第1の実施形態では、RAID5の構成の下、書き込みデータを圧縮し、当該圧縮されたデータをデータボリューム上の実記憶領域に格納しきれない場合に、当該格納しきれない分を、ボリューム容量仮想化機能により提供される拡張記憶領域に格納するように構成されたストレージ装置が説明される。

【0019】

図1は、本発明の一実施形態に係るストレージシステムの構成を説明するための図である。同図に示すように、コンピュータシステム1は、ネットワーク2を介してホスト装置3に接続されたストレージ装置4を含んで構成されている。

40

【0020】

ネットワーク2は、例えばLAN、インターネット、又はSAN (Storage Area Network) であり、ネットワークスイッチやハブ等を含んで構成される。本実施形態では、ネットワーク2は、ファイバーチャネルプロトコルを用いたSAN (FC - SAN) で構成されているものとする。

【0021】

ホスト装置3は、例えば銀行の業務システムや航空機の座席予約業務システム等における中核をなすコンピュータである。ホスト装置3は、プロセッサと、メインメモリと、通信インターフェースと、ローカル入出力装置等のハードウェア資源を備え、また、デバイ

50

ストライバやオペレーティングシステム（OS）、アプリケーションプログラム等のソフトウェア資源を備えている（図示せず）。これによって、ホスト装置3は、プロセッサの制御の下、各種のプログラムを実行して、ハードウェア資源との協働作用により、所望の処理を実現する。例えば、ホスト装置3は、プロセッサの制御の下、業務アプリケーションプログラムを実行することにより、以下に詳述されるストレージ装置4にアクセスし、所望の業務システムを実現する。

【0022】

ストレージ装置4は、データストレージサービスをホスト装置3に提供する補助記憶装置である。ストレージ装置4は、データを格納するディスクデバイス5と、ディスクデバイス5の構成を管理して、それに対する書き込み又は読み出しといったI/O処理を制御するディスクコントローラ6と、を備える。ストレージ装置4はまた、ストレージ装置4全体を管理するための管理装置乃至はサービスプロセッサ（図示せず）を含むことができる。あるいは、ディスクコントローラ6が管理装置の機能を含むように、ディスクコントローラ6を構成してもよい。管理装置は、例えば、管理プログラムが実装された汎用コンピュータである。

【0023】

本実施形態のストレージ装置4は、後述するように、RAID制御の下、書き込みデータを圧縮し、その圧縮データをディスクデバイス5の所定の記憶領域に格納する。ストレージ装置4は、圧縮データが所定のブロックサイズを超える場合には、当該超えた分を拡張記憶領域（プールボリューム）に格納する。プールボリュームは、後述するように、ボリューム容量仮想化機能により提供される仮想ボリュームに対応付けられた、データの実体を格納する論理ボリュームである。

【0024】

ディスクデバイス5は、例えば複数のハードディスクドライブ（HDD）や半導体メモリ等の記憶媒体を含んで構成される物理デバイス（PDEV）である。複数のハードディスクの集合体をディスクアレイと呼ぶこともある。ディスクデバイス5には、ディスクコントローラ6の制御の下、ホスト装置3に提供するための1つ以上の論理デバイス（LDEV）が形成される。論理デバイスは、いくつかのハードディスクドライブを仮想的に1つのデバイスにまとめた仮想デバイスVDEVに対して形成されてもよい。

【0025】

論理デバイスは、ホスト装置3が認識しうる論理的な記憶装置である。論理デバイスには、論理ユニット（LU）が割り当てられてもよい。この場合、各論理デバイスは、後述するチャネルアダプタ61に設けられた各ポートに割り当てられ、これにより、ホスト装置3は、論理デバイスを論理ユニットとして認識する。各論理ユニットには、論理ユニット番号（LUN）が付与される。また、論理ユニットは、I/Oアクセスの最小単位であるブロックに分割され、各ブロックには、論理ブロックアドレス（LBA）が割り当てられる。これにより、ホスト装置3は、LUNおよびLBAからなる論理アドレスをストレージ装置4に与えることにより、特定の論理ユニットにおける任意のブロックに記憶されたデータに対してアクセスを行うことができる。論理デバイス及び論理ユニットは、ストレージ装置4が適用されるシステム環境により区別されうるが、本発明では両者を同義に扱うことができる。

【0026】

論理デバイスには、属性に応じた論理ボリューム（LVOLまたはVOL）が定義される。ただし、慣用的に両者は同義に扱われることもある。本実施形態では、論理ボリュームとして、データボリューム及びプールボリュームが定義される。後述するように、データボリュームは、実ボリューム及び仮想ボリュームを含んでいる。

【0027】

ディスクコントローラ6は、チャネルアダプタ（CHA）61と、キャッシュメモリ（CM）62と、ディスクアダプタ（DKA）63と、共有メモリ（SM）64と、を備え、これらのコンポーネントは、内部スイッチ（SW）65を介して相互に接続されている

10

20

30

40

50

。これらのコンポーネントは、高速性、耐障害性の観点等の観点から、多重化されることが望ましい。

【 0 0 2 8 】

ディスクコントローラ 6 は、ディスクデバイス 5 を制御して、データストレージサービスを実現するストレージ装置 4 のメインシステムボードである。本実施形態のディスクコントローラ 6 は、ホスト装置 3 からの I / O アクセス要求に基づく通常の I / O 処理機能に加え、データ圧縮 / 伸張機能、R A I D 機能、及びボリューム容量仮想化機能を実現する構成を含んでいる。

【 0 0 2 9 】

チャネルアダプタ 6 1 は、ネットワーク 2 を介してストレージ装置 4 をホスト装置 3 に通信可能に接続するためのコンポーネントである。本例では、2つのチャネルアダプタ 6 1 が搭載されている。各チャネルアダプタ 6 1 は、マイクロプロセッサ 6 1 1 と、ローカルメモリ 6 1 2 と、通信インターフェース (I / F) 6 1 3 と、I / O プロセッサ 6 1 4 と、を備える。

【 0 0 3 0 】

マイクロプロセッサ 6 1 1 は、ローカルメモリ 6 1 2 に格納されたマイクロプログラムを実行することにより、チャネルアダプタ 6 1 の動作を統括的に制御するチップ回路である。通信 I / F 6 1 3 は、ネットワーク 2 から通信パケット乃至はフレームを取り込み、または、ネットワーク 2 に通信パケットを送出する。通信 I / F 6 1 3 は、ネットワーク 2 に接続するための複数のポートを備える (図示せず) 。 I / O プロセッサ 6 1 4 は、内部スイッチ 6 5 を介して接続されるディスクコントローラ 6 内の他のコンポーネント (例えば、キャッシュメモリ 6 2 や共有メモリ 6 4) との間でデータの授受を制御する。

【 0 0 3 1 】

具体的には、マイクロプロセッサ 6 1 1 は、通信 I / F 6 1 3 を介してパケットを受け取ると、所定のプロトコルに応じたプロトコル変換処理を行って、内部データを取り出す。続いて、マイクロプロセッサ 6 1 1 は、当該内部データを解釈して、これが書き込みコマンドであれば、当該書き込みコマンドを共有メモリ 6 4 に、その書き込みデータ (すなわちユーザデータ) をキャッシュメモリ 6 2 に書き込むように、I / O プロセッサ 6 1 4 を制御する。I / O プロセッサ 6 1 4 は、これを受けて、書き込みコマンドを共有メモリ 6 4 に書き込むとともに、その書き込みデータをキャッシュメモリ 6 2 に書き込む。

【 0 0 3 2 】

また、マイクロプロセッサ 6 1 1 は、読み出しコマンドに従って、I / O プロセッサ 6 1 4 を介して読み出しデータをキャッシュメモリ 6 2 から取り出して、所定のプロトコルに応じたプロトコル変換処理を行ってパケットを生成し、通信 I / F 6 1 3 を介してネットワーク 2 に送出的る。

【 0 0 3 3 】

キャッシュメモリ 6 2 は、ホスト装置 3 に対して高いシステムパフォーマンスを提供するため、ホスト装置 3 とディスクデバイス 5 (論理ボリューム) との間でやり取りされるユーザデータを一時的に記憶 (キャッシュ) する。つまり、キャッシュメモリ 6 2 は、チャネルアダプタ 6 1 とディスクアダプタ 6 3 との間のデータの受け渡しに利用されるメモリである。キャッシュメモリ 6 2 は、例えば、D R A M 等の揮発性メモリで構成される。あるいは、キャッシュメモリ 6 2 は、フラッシュメモリ等の不揮発性メモリで構成されるかも知れない。

【 0 0 3 4 】

キャッシュメモリ 6 2 はまた、後述するように、圧縮された書き込みデータ (圧縮データ) を一時的に記憶するメモリとしても利用される。データ圧縮 / 伸張処理は、例えば、ディスクアダプタ 6 3 によって行われる。

【 0 0 3 5 】

ディスクアダプタ 6 3 は、ディスクチャネルを介して接続されたディスクデバイス 5 に対する I / O アクセスを行うコンポーネントである。本例では、2つのディスクアダプタ

10

20

30

40

50

63が搭載されている。各ディスクアダプタ63は、マイクロプロセッサ631と、ローカルメモリ632と、ディスクインターフェース(I/F)633と、I/Oプロセッサ634と、を備える。本実施形態では、ディスクアダプタ63は、I/O制御機能、データ圧縮/伸張機能、RAID機能、及びボリューム容量仮想化機能を実装している。これらの機能は、例えば、図2に示すようなファームウェアとして実現される。

【0036】

マイクロプロセッサ631は、ローカルメモリ632に格納されたマイクロプログラムを実行することにより、ディスクアダプタ63の動作を統括的に制御するチップ回路である。本実施形態では、マイクロプロセッサ631は、RAID制御の下、キャッシュメモリ62に格納された書き込みデータに対して所定の圧縮アルゴリズムに従った圧縮処理を行って圧縮データを生成し、これをディスクデバイス5の所定の記憶領域に書き込む。ここでいう所定の記憶領域とは、実ボリューム上の記憶領域(実記憶領域)である。圧縮データのうち当該実記憶領域に格納しきれない分のデータは、ボリューム容量仮想化機能により管理される拡張記憶領域(プールボリューム)に格納される。圧縮/伸張アルゴリズムは、例えば、LZW方式等の既知のアルゴリズムを適用することができる。

10

【0037】

ディスクI/F633は、物理デバイスとしてのディスクデバイス5にI/Oアクセスを行うためのインターフェースである。I/Oプロセッサ634は、内部スイッチ65を介して接続される他のコンポーネント(例えばキャッシュメモリ62や共有メモリ64)との間のデータの授受を制御する。

20

【0038】

より具体的には、マイクロプロセッサ631は、定期的(例えば数十msおき)又は不定期に、I/Oプロセッサ634を介して共有メモリ64を参照する。マイクロプロセッサ631は、共有メモリ64上の未処理の書き込みコマンドを見つけると、キャッシュメモリ62から対応する書き込みデータを取り出して、RAID制御の下で、圧縮処理を行い、圧縮データを生成する。つまり、書き込みデータに基づくパリティデータが生成され、このパリティデータを含む書き込みデータに対する圧縮データが生成される。圧縮データは、キャッシュメモリ62上の圧縮データ用ブロックに、再度、格納される。そして、マイクロプロセッサ631は、生成した圧縮データを、RAID制御の下で、データボリュームを形成しているディスクデバイス5上の所定の記憶領域(ブロック)に格納(デステージング)する。例えばRAID5(3D+1P)の構成にあれば、圧縮データは、そのパリティデータとともに、物理的に異なる系列のディスクデバイス5に分散されて格納される。また、このとき、マイクロプロセッサ63は、生成された圧縮データをデータボリュームにおいて規定される実記憶領域に格納しきれないかを判断し、格納しきれないと判断する場合には、当該圧縮データのうち格納しきれない分のデータを、ボリューム容量仮想化機能により管理される拡張記憶領域に格納する。

30

【0039】

また、マイクロプロセッサ631は、共有メモリ64上の読み出しコマンドに従い、RAID制御の下、所定の論理ボリュームから圧縮データを読み出して、キャッシュメモリ62に、一旦、格納する。そして、マイクロプロセッサ631は、伸張処理を行い、得られた伸張データ(元のデータ)をキャッシュメモリ522に、再度、書き込む(ステージング)。圧縮データの一部が拡張記憶領域に格納されている場合には、マイクロプロセッサ631は、適宜、拡張記憶領域から当該部分的な圧縮データを読み出し、一つの圧縮データに統合する。

40

【0040】

共有メモリ64は、ストレージ装置4内の各コンポーネントが参照すべき各種の情報を記憶する。共有メモリ64は、例えばDRAM等の揮発性メモリで構成される。

【0041】

内部スイッチ65は、クロスバースイッチ等により構成されるスイッチングデバイスである。内部スイッチ428は、入力されるデータ信号の競合を調停し、データ信号のバス

50

を切り替えて、送出元のモジュールと送出先のモジュールとのパスを構築する。なお、内部スイッチ 4 2 8 は、パケット交換方式のスイッチングデバイスであってもよい。

【 0 0 4 2 】

図 3 は、本発明の一実施形態に係るストレージ装置 4 における共有メモリ 6 4 の内容を示す図である。同図に示すように、共有メモリ 6 3 は、例えば、システム構成情報、キャッシュ管理テーブル、データブロック管理テーブル、物理デバイス管理テーブル等の各種の情報を記憶する。

【 0 0 4 3 】

システム構成情報は、ディスクデバイス 5 の構成情報やストレージ装置 4 内の各コンポーネント上で動作するマイクロプログラムのバージョン情報といったシステムの構成に関する情報である。ディスクデバイス 5 の構成情報としては、例えば、RAID 構成情報、物理デバイス / 論理デバイス (ボリューム) の構成情報、ボリューム容量仮想化構成情報を含む。物理デバイス / 論理デバイス構成情報は、例えば、セグメント管理テーブル及びアドレス管理テーブルとして定義される。また、ボリューム容量仮想化構成情報は、例えば、プールボリューム管理テーブルを含む。

【 0 0 4 4 】

システム構成情報は、システム管理者が操作する管理装置によって、設定・管理される。例えば、システム管理者は、管理装置を操作して、RAID 5 (3 D + 1 P) の構成を設定することができる。また、システム管理者は、ハードディスクドライブの増設に併せて、管理装置を操作して、ボリューム容量仮想化構成情報を設定する。

【 0 0 4 5 】

キャッシュ管理テーブルは、キャッシュメモリ 6 2 にキャッシュされたデータのディレクトリ情報を管理するテーブルである。キャッシュ管理テーブルは、書き込みデータを構成するデータブロックのそれぞれがディスクデバイス 5 (論理ボリューム) 上のどのブロックに対応するかを管理する。

【 0 0 4 6 】

上述したように、本実施形態では、論理ボリューム (データボリューム) には、書き込みデータそのものではなく、その圧縮データが格納される。従って、データボリューム上に上に規定された記憶領域のブロックサイズは、典型的には、平均的な圧縮率を考慮して、キャッシュメモリ 6 2 上に規定されたブロックサイズよりも小さく規定されている。また、ディスクアダプタ 6 3 は、キャッシュメモリ 6 2 の所定のブロックから読み出して圧縮したデータを、再度、キャッシュメモリ 6 2 の別のブロック (圧縮データ用ブロック) に格納してから、デステージングしている。このため、キャッシュ管理テーブルは、キャッシュメモリ 6 2 上の RAW 状態の書き込みデータ及び圧縮データの各論理アドレスと、データボリューム上の論理アドレスとを対応付けている。

【 0 0 4 7 】

データブロック管理テーブルは、RAID 制御によるパリティグループごとのデータブロックを管理するためのテーブルである。パリティグループとは、本明細書では、同時並列的にアレディスクに書き込まれるデータブロック群を指している。例えば、RAID 5 (3 D + 1 P) であれば、3 つのデータブロックと 1 つのパリティブロックからなるデータブロック群が同一のパリティグループに属する。

【 0 0 4 8 】

ディスクデバイス管理テーブルは、RAID 制御の下でデータボリュームを形成するストライプ構成された (すなわち、物理的に異なる系列として構成された) 「ディスクデバイス」のアドレス空間の使用状況を管理するためのテーブルである。ディスクデバイス管理テーブルは、ストライプ構成数だけ用意される。例えば、RAID 5 (3 D + 1 P) であれば、4 つのディスクデバイス管理テーブルが用意される。なお、慣用的な RAID の説明に従えば、ストライプ構成されたディスクデバイスは、個々の物理デバイスに等しく考えることができる。しかしながら、本実施形態では、ストライプ構成された「ディスクデバイス」は、ボリューム容量仮想化機能により物理記憶領域を持たない仮想ボリューム

10

20

30

40

50

を含む概念で説明される。従って、本明細書では、RAID制御の下でストライプ構成された個々のディスクデバイスを「疑似ディスクデバイス」と呼ぶことにする。RAID5(3D+1P)であれば、あるデータボリュームは、4つの疑似ディスクデバイス上に形成される。

【0049】

図4は、本発明の一実施形態に係るストレージ装置4におけるセグメント管理テーブルの一例を示している。セグメント管理テーブルは、ストレージ装置4が提供する論理ボリューム上の記憶領域空間を管理するためのテーブルである。同図に示すように、セグメント管理テーブルは、ディスクIDと、セグメント番号と、LBA開始アドレスと、LBA終了アドレスと、セグメントの使用状況とを含む。

10

【0050】

ディスクIDは、ディスクデバイス5を構成する物理デバイスを一意に識別するための識別子である。セグメント番号は、セグメントを一意に識別するための番号である。セグメントは、ディスクコントローラ6が管理するディスクデバイス5上の記憶領域である。LBA開始アドレス及びLBA終了アドレスは、ディスクデバイス5上のセグメントの物理的な開始位置及び終了位置を示すアドレスである。LBA開始アドレス及びLBA終了アドレスより、セグメントの大きさが求められる。使用状況は、セグメントが使用されているか否かを示す。セグメントが使用中の場合は“1”、未使用の場合は“0”の値が設定される。

【0051】

20

図5は、本発明の一実施形態に係るストレージ装置4の共有メモリ64に格納されたアドレス管理テーブルの一例を示している。アドレス管理テーブルは、ストレージ装置4が提供する論理ボリューム上の記憶領域(すなわち、論理アドレス空間)とディスクデバイス5上の実際の記憶領域(すなわち、物理アドレス空間)とを対応付けるテーブルである。ディスクアダプタ63は、アドレス管理テーブルを参照することによって、データの格納先として指定された論理アドレスを、ディスクデバイス5上の物理アドレスに変換し、その物理アドレスに従ってアクセスを行う。また、アドレス管理テーブルは、論理ボリュームのうち、プールボリュームとして使用されるセグメントを定義する。

【0052】

すなわち、同図に示すように、アドレス管理テーブルは、論理ユニット番号(LUN)、セグメント番号、LBA開始アドレス、LBA終了アドレス、及びプール割り当て状況を含む。論理ユニット番号は、ストレージ装置4がホスト装置3に提供する論理ボリュームを一意に識別するための番号である。セグメント番号は、セグメント管理テーブルにおいて管理されている記憶領域を一意に識別するための番号である。論理ボリュームLUは、複数のセグメントから構成される。LBA開始アドレス及びLBA終了アドレスは、セグメントの論理的な開始アドレス及び終了アドレスを示す。プール割り当て状況は、そのセグメントがプールボリュームとして使用されるか否かを示す。セグメントがプールボリュームとして使用される場合には、“1”が書き込まれる。

30

【0053】

図6は、本発明の一実施形態に係るストレージ装置4の共有メモリ64に格納されたデータボリュームを説明するための図である。同図に示すように、データボリュームDVOL0は、0~9999の論理アドレス空間を有し、第1のディスクデバイス及び第2のディスクデバイスにおけるそれぞれのセグメントが割り当てられている。第1のディスクデバイス5は、図4に示されたように、ディスクID「0」が割り当てられた物理デバイスであり、第2のディスクデバイス5は、ディスクID「1」が割り当てられた物理デバイスである。また、図5に示されるように、セグメント番号「1」で示されるセグメントは、データボリュームDVOL0上の0~1299の論理アドレス空間を提供する。また、セグメント番号「3」で示されるセグメントは、1300~1999の論理アドレス空間を提供する。このようにして、データボリュームDVOL0は、あたかも1つのディスクデバイスのように扱われる。

40

50

【 0 0 5 4 】

データボリュームD V O L 0の論理アドレス空間2 0 0 0 ~ 9 9 9 9 9は、物理デバイスとしてのディスクデバイスが割り当てられていない。ディスクコントローラ6は、データボリュームD V O L 0のその空間に対しても、見かけ上、データを格納する。このため、ストレージ装置4は、プールボリュームP V O Lと呼ばれる論理ボリュームをさらに備え、データボリュームD V O L 0内の記憶領域であって、ディスクデバイス5上のセグメントが割り当てられていない記憶領域（拡張記憶領域）に対して格納されるべきデータを、このプールボリュームP V O Lに格納する。つまり、データボリュームD V O L 0の論理アドレス空間2 0 0 0 ~ 4 9 9 9 9に格納されるデータの実体は、プールボリュームP V O Lに格納されることになる。このため、データボリュームの拡張記憶領域に対するアクセスは、内部的には、プールボリューム管理テーブルに従って、プールボリュームP V O Lに対して行われることになる。拡張記憶領域に格納されるべきデータは、物理デバイスにより実際に提供される記憶領域（実記憶領域）に格納しきれない圧縮されたデータブロックである。

10

【 0 0 5 5 】

同図は、ボリューム容量仮想化機能を適用されたデータボリュームD V O Lを簡略的に説明している。本実施形態では、ディスクデバイス5は、R A I D構成に応じてストライプ構成される。従って、データボリュームD V O Lは、実際は、物理的に異なる系列の4つの疑似ディスクデバイスによって提供される。

20

【 0 0 5 6 】

図7は、本発明の一実施形態に係るストレージ装置4における記憶領域の動的割り当てを概念的に説明するための図である。

【 0 0 5 7 】

上述したように、本実施形態に係るストレージ装置4のディスクデバイス5には、データボリュームD V O L及びプールボリュームP V O Lが形成される。データボリュームD V O Lは、ホスト装置3に提供される論理ボリュームである。データボリュームD V O Lは、実際に実装されているディスクデバイス5を構成する物理デバイスの固有記憶容量に依存して、実ボリュームと仮想ボリュームとに分けられている。つまり、実ボリュームは、ディスクデバイス5を構成する物理デバイスにより実際に提供されるセグメント上に形成された物理的な記憶領域を含む。これに対して、仮想ボリュームは、実際に提供されるセグメント以外の仮想的な記憶領域（拡張記憶領域）に対応する。従って、仮想ボリュームには、データの実体が格納されない。プールボリュームP V O Lは、仮想ボリュームに格納されるべきデータの実体を、将来的にディスクデバイスが追加的に実装され物理的な記憶領域が確保されて新たなデータボリュームD V O Lが定義されるまでの間、一時的に格納する記憶領域を提供する論理ボリュームである。

30

【 0 0 5 8 】

ストレージ装置4は、データボリュームV O Lにおける実ボリューム上の記憶領域に対しては、ホスト装置2が認識する論理アドレス（L U NおよびL B A）と、ディスクデバイス5における記憶領域上の物理アドレスとを一对一に関連付けている。これによって、ホスト装置2は、所定の論理アドレスを指定することによって、ディスクデバイス5上の所望の記憶領域にアクセスすることができる。

40

【 0 0 5 9 】

一方、プールボリュームP V O L上の記憶領域については、ホスト装置3が認識する論理アドレスと、実際にデータにアクセスするためのディスクデバイス5上の物理アドレスとを直接的に関連付けていない。ディスクアダプタ63は、データボリュームD V O Lにおける仮想ボリューム上の記憶領域を指定したデータの書き込み命令に従い、当該記憶領域をプールボリュームP V O Lに動的に割り当てる。また、ディスクアダプタ63は、データボリュームD V O Lにおける仮想ボリューム上の記憶領域からの読み出し命令に従い、プールボリュームP V O L上の当該記憶領域からデータを読み出す。

【 0 0 6 0 】

50

このように、ストレージ装置 4 は、プールボリューム管理テーブルを保持し、仮想ボリュームとプールボリュームとの間の動的な割り当てを管理している。プールボリューム管理テーブルは、例えば、ディスクコントローラ 6 の共有メモリ 6 4 に保持される。ストレージ装置 4 は、圧縮データを実ボリュームに格納しきれない場合には、当該格納しきれない分を、プールボリューム P V O L に格納するとともに、仮想ボリューム上の論理アドレスとプールボリュームのアドレスを示すポインタとを関連付けて、プールボリューム管理テーブルに登録する。

【 0 0 6 1 】

図 8 は、本発明の一実施形態に係るストレージ装置 4 における共有メモリ 6 4 に格納されたデータブロック管理テーブルの一例を示す図である。同図に示すように、データブロック管理テーブルは、スロット I D とデータボリューム内のアドレスとを対応付けている。スロット I D は、パリティグループ内のデータブロックを識別するための識別子である。これにより、データブロック管理テーブルは、パリティグループ内のデータブロックが、どの物理デバイスのどのアドレスに格納されるかを管理する。

【 0 0 6 2 】

図 9 は、本発明の一実施形態に係るストレージ装置 4 における共有メモリ 6 4 に格納されたディスクデバイス管理テーブルの一例を示す図である。同図に示すように、ディスクデバイス管理テーブルは、各疑似ディスクデバイス内の個々のアドレスで示される記憶領域とそこに格納されているデータブロック（スロット識別子）とを対応付けている。また、ディスクデバイス管理テーブルは、拡張記憶領域における使用状況を管理している。本例は、疑似ディスクデバイス（0）に対応するディスクデバイス管理テーブルを示している。すなわち、同図のディスクデバイス管理テーブルは、アドレス“x x x x”で示される記憶領域には“D 1 - 1”が格納され、拡張記憶領域 1 には“D 1 - 2”が格納されていることを示している。また、拡張記憶領域 2 は使用可能であることを示している。さらに、拡張記憶領域 n は、まだ、実際の記憶領域が割り当てられていないことを示している。

【 0 0 6 3 】

図 1 0 及び図 1 1 は、本発明の一実施形態に係るストレージ装置 4 におけるデータ書き込みメカニズムを説明するための図である。具体的には、これらの図は、ある書き込みデータが圧縮された後、R A I D 5（3 D + 1 P）の下で、異なる系列の疑似ディスクデバイスに格納される様子を示している。

【 0 0 6 4 】

データボリュームに対する書き込みデータは、ストレージ装置 4 内部で、ブロック化されたデータ（データブロック）として扱われる。単位ブロックのサイズは、例えば 6 4 K B である。本実施形態では、書き込みデータは圧縮されて格納されるため、データボリューム上では、当該サイズよりも小さなサイズ（例えば、1 6 K B）の単位記憶領域が規定される。キャッシュメモリ 6 2 上の 6 4 K B の各ブロックは、データボリューム上の 1 6 K B の各ブロックに対応付けられている。従って、6 4 K B のデータが圧縮されたときに、その圧縮データが 1 6 K B を超えなければ、キャッシュメモリ 6 2 上の 6 4 K B ブロックに対応付けられたデータボリューム上の 1 6 K B ブロック（基本ブロック）に収容できるが、その圧縮データが 1 6 K B を超える場合には、当該 1 6 K B ブロック以外の予備ブロックが必要になる。つまり、6 4 K B × 1 データブロックの圧縮データが 1 6 K B × 3 データブロックで構成される場合、超過分の 2 データブロックは、予備的なブロックに格納されることになる。

【 0 0 6 5 】

今、あるデータボリュームに対するデータ D 1 の書き込み要求があったとする。説明を簡略化するために、書き込みデータ D 1 は、ブロックサイズが 6 4 K B である単一のデータブロックであるとする。R A I D 制御の下、書き込みデータ D 1 をデータボリュームに格納するためには、同一のパリティグループ内のデータブロック D 2 及び D 3 とともにパリティを算出して、これらを格納する必要がある（図 1 0（A））。

【 0 0 6 6 】

データブロック D 1 ~ D 3 は、ディスクアダプタ 6 3 の制御の下、圧縮され、図 1 0 (B) に示すような圧縮データブロックが得られたとする。圧縮データブロックの単位サイズは 1 6 K B である。すなわち、圧縮前のデータブロック D 1 は、圧縮データブロック D 1 - 1 及び D 1 - 2 となり、圧縮前のデータブロック D 2 は、圧縮データブロック D 2 - 1 及び D 2 - 2 となり、そして、圧縮前のデータブロック D 3 は、圧縮データブロック D 3 - 1 となったとする。また、これら圧縮前データブロック D 1 ~ D 3 のパリティデータ P は、圧縮データブロック P 1 - 1 ~ P - 4 に変換されたとする。つまり、圧縮データブロック D 1 - 2、D 2 - 2、及び P - 2 ~ P - 4 は、データボリューム上の基本ブロックに格納しきれないデータブロックである。

10

【 0 0 6 7 】

本実施形態では、キャッシュメモリ 6 2 上の 6 4 K B ブロックは、データボリューム上の 1 6 K B 基本ブロックに関連付けられている。基本ブロックは、データボリュームの実ボリュームにおける記憶領域（実記憶領域）である。また、データの内容によっては、その圧縮データを実記憶領域に格納しきれない場合があるため、当該格納しきれない分のデータブロックは、仮想ボリュームのブロック（拡張記憶領域）に格納する。

【 0 0 6 8 】

すなわち、図 1 1 に示すように、圧縮データブロック D 1 - 1、D 2 - 1、D 3 - 1、及び P - 1 は、R A I D 制御の下、データボリュームを構成する R A I D グループにおける疑似ディスクデバイス (0) ~ 疑似ディスクデバイス (3) 上の実ボリュームにおけるブロック（実記憶領域）にそれぞれ格納される。また、圧縮データブロック D 1 - 2、D 2 - 2、及び P - 2 ~ P - 4 は、当該 R A I D グループにおける対応する疑似ディスクデバイス (0) ~ 疑似ディスクデバイス (4) 上の仮想ボリュームにおけるブロック（拡張記憶領域）にそれぞれ格納される。ただし、上述したように、仮想ボリュームは、データの実体を格納する物理デバイスのセグメントが割り当てられていない。従って、仮想ボリュームに格納されるデータの実体は、プールボリュームに格納されることになる。

20

【 0 0 6 9 】

図 1 2 は、本発明の一実施形態に係るストレージ装置 4 におけるデータ書き込み処理を説明するためのフローチャートである。具体的には、同図は、データ書き込み処理時におけるディスクアダプタ 6 3 の動作を説明している。

30

【 0 0 7 0 】

すなわち、ディスクアダプタ 6 3 は、共有メモリ 6 4 を参照し、未処理の書き込みコマンドを見つけると、キャッシュメモリ 6 2 上のそれに対応する書き込みデータを特定する。ディスクアダプタ 6 3 は、当該書き込みデータに基づいて、R A I D 構成に従ったパリティデータを生成する (S T E P 1 2 0 1)。つまり、本例では、ディスクアダプタ 6 3 は、R A I D 5 (3 D + 1 P) の構成に従い、3 つのデータブロックごとに 1 つのパリティデータブロックを生成する。これら 3 つのデータブロックと 1 つのパリティデータブロックにより 1 つのパリティグループが形成される。従って、1 つのデータブロックの書き込みであっても、パリティグループ内の残りの 2 つのデータブロックとともに新しいパリティデータブロックを生成する必要がある。これは一般に「ライトペナルティ」と呼ばれている。このフェーズでは、各データブロックは、6 4 K B のブロックサイズである。

40

【 0 0 7 1 】

続いて、ディスクアダプタ 6 3 は、当該パリティグループにおける各データブロック（パリティデータブロックを含む。）に対して圧縮処理を行って、圧縮データ列をそれぞれ生成する (S T E P 1 2 0 2)。圧縮データ列は、1 以上のデータブロック（圧縮データブロック）から構成される。ディスクアダプタ 6 3 は、各圧縮されたデータブロックをキャッシュメモリ 6 2 における圧縮データ用ブロックに書き込み、キャッシュ管理テーブルを更新する。圧縮データ用ブロックは、例えば、1 6 K B のブロックサイズである。

【 0 0 7 2 】

ディスクアダプタ 6 3 は、次に、圧縮データブロックを系列の異なる疑似ディスクデバ

50

イスに分散させて格納するために以下の処理を行う。

【 0 0 7 3 】

具体的には、ディスクアダプタ 6 3 は、まず、パリティサイクル変数 I に 0 をセットする (S T E P 1 2 0 3)。次に、ディスクアダプタ 6 3 は、データブロック管理テーブルに当該圧縮データブロックのエントリを登録する (S T E P 1 2 0 4)。すなわち、ディスクアダプタ 6 3 は、疑似ディスクデバイス (I) に格納されるべき圧縮データ列を構成する先頭の圧縮データブロックに対して、疑似ディスクデバイス (I) における実記憶領域のアドレスを割り当てる。つまり、圧縮データ列が圧縮データブロック D 1 - 1 及び D 1 - 2 で構成されている場合、当該圧縮データブロック D 1 - 1 について、疑似ディスクデバイス (I) における実記憶領域のアドレスをまず割り当てる。続いて、ディスクアダプタ 6 3 は、当該アドレスに従って、当該実記憶領域に圧縮データブロックを格納する (S T E P 1 2 0 5)。

10

【 0 0 7 4 】

ディスクアダプタ 6 3 は、次に、当該圧縮データ列が超過分の圧縮データブロックを含んでいるか否かを判断する (S T E P 1 2 0 6)。すなわち、圧縮データ列が 2 つ以上の圧縮データブロックで構成されるか否かが判断される。ディスクアダプタ 6 3 は、超過分の圧縮データブロックがないと判断する場合 (S T E P 1 2 0 6 の N o)、パリティサイクル変数 I の値を 1 つインクリメントし (S T E P 1 2 0 9)、パリティサイクル変数 I が所定のパリティサイクル数 (本例では 4) 以下であるか否かを判断する (S T E P 1 2 1 0)。パリティサイクル変数 I が所定のパリティサイクル数以下である場合 (S T E P 1 2 1 0 の N o)、パリティグループ内に未処理の圧縮データブロックがあるため、 S T E P 1 2 0 4 の処理に戻る。

20

【 0 0 7 5 】

ディスクアダプタ 6 3 は、超過分の圧縮データブロックがあると判断する場合 (S T E P 1 2 0 6 の Y e s)、データブロック管理テーブルに当該超過分の圧縮データブロックのエントリを登録する (S T E P 1 2 0 7)。すなわち、ディスクアダプタ 6 3 は、当該超過分圧縮データブロックに対して、疑似ディスクデバイス (I) における拡張記憶領域のアドレスを割り当てる。

【 0 0 7 6 】

続いて、ディスクアダプタ 6 3 は、超過分の圧縮データブロックを仮想ボリュームに格納するための処理を行う (S T E P 1 2 0 8)。仮想ボリュームに対する格納処理の詳細については、後述する。

30

【 0 0 7 7 】

その後、ディスクアダプタ 6 3 は、パリティサイクル変数 I の値を 1 つインクリメントし (S T E P 1 2 0 9)、パリティサイクル変数 I が所定のパリティサイクル数 (本例では 4) 以下であるか否かを判断する (S T E P 1 2 1 0)。パリティサイクル変数 I が所定のパリティサイクル数以下である場合 (S T E P 1 2 1 0 の N o)、パリティグループ内に未処理の圧縮データブロックがあるため、 S T E P 1 2 0 4 の処理に戻る。

【 0 0 7 8 】

パリティサイクル変数 I が所定のパリティサイクル数に達した場合 (S T E P 1 2 1 0 の Y e s)、ディスクアダプタ 6 3 は、当該データ書き込み処理を終了する。

40

【 0 0 7 9 】

図 1 3 は、本発明の一実施形態に係るストレージ装置 4 における仮想ボリュームに対するデータ書き込み処理を説明するためのフローチャートである。具体的には、同図は、図 1 2 に示した S T E P 1 2 0 8 の詳細を説明するためのフローチャートである。

【 0 0 8 0 】

すなわち、同図に示すように、ディスクアダプタ 6 3 は、まず、仮想ボリュームにプールボリューム上の記憶領域がすでに割り当てられているか否かを判断する (S T E P 1 3 0 1)。ディスクアダプタ 6 3 は、仮想ボリュームにプールボリューム上の記憶領域が割り当てられていないと判断する場合 (S T E P 1 3 0 1 の N o)、プールボリューム管理

50

テーブルに、超過分の圧縮データブロックについてのエントリを追加する（STEP 1302）。上述の例でいえば、圧縮データブロックD1 - 2について、仮想ボリュームのアドレスとプールボリュームのアドレスとを対応付ける。

【0081】

続いて、ディスクアダプタ63は、ディスクデバイス管理テーブルの拡張記憶領域にデータブロックを追加する（STEP 1303）。そして、ディスクアダプタ63は、超過分の圧縮データブロックを仮想ボリュームに格納する（STEP 1304）。この場合、実際には、ディスクアダプタ63の制御の下、当該超過分の圧縮データブロックは、プールボリュームを提供するディスクデバイス5に格納されることになる。

【0082】

10

以上のように本実施形態によれば、書き込みデータを圧縮して格納するために、キャッシュメモリ62上のブロックをディスクデバイス5上のより小さなブロックに固定的に割り当てた構成を採用するストレージ装置において、圧縮データがディスクデバイス5上の当該ブロックに格納しきれない場合であっても、当該圧縮データのうち格納しきれない分のデータを拡張記憶領域に格納することができるようになる。当該拡張記憶領域は、これに関連付けられたプールボリュームとして管理されているため、記憶領域の使用実績に応じた柔軟なディスクデバイスの構成を採用することができるようになる。

【第2の実施形態】

【0083】

第2の実施形態では、ストレージ装置4は、RAID1の構成の下、書き込みデータの圧縮データをミラーリングして格納する際に、複製された圧縮データを仮想ボリュームに格納するように構成されたストレージ装置4が説明される。

20

【0084】

図14及び図15は、本発明の一実施形態に係るストレージ装置4におけるデータ書き込みメカニズムを説明するための図である。具体的には、これらの図は、ある書き込みデータD4 ~ D6が圧縮された後、RAID1（4D）の下で、疑似ディスクデバイスに格納される様子を説明している。

【0085】

あるデータボリュームに対して書き込みデータD4 ~ D6を格納する場合を考える。説明を簡略化するために、書き込みデータD4 ~ D6は、ブロックサイズが64KBである単一のデータブロックであるとする。データブロックD4 ~ D6は、ディスクアダプタ63の制御の下、圧縮され、圧縮データブロックD4 - 1 ~ D6 - 1が得られたとする。また、圧縮データブロックD4 - 1 ~ D6 - 1の複製データ（ミラーリングデータ）を圧縮データブロックD4 - 1' ~ D6 - 1'で示す。圧縮データブロックのサイズは16KBである。従って、ストレージ装置4は、64KB × 3データブロックに対して、16KB × 6データブロックを格納することになる（図14）。

30

【0086】

本実施形態では、図15に示すように、これら圧縮データブロックを、疑似ディスクデバイスごとの実記憶領域に分散して格納するとともに、ミラーリングされた圧縮データブロックを拡張記憶領域に分散して格納する。ただし、本例のように、データサイクルが3であり、疑似ディスクデバイスが4つである場合には、ミラーリングされた圧縮後データブロックの一つを第4の疑似ディスクデバイスの実記憶領域に格納するようにしてもよい。

40

【0087】

図16は、本発明の一実施形態に係るストレージ装置4における共有メモリ64に格納されたディスクデバイス管理テーブルの一例を示す図である。同図に示すように、本実施形態のディスクデバイス管理テーブルは、スロットIDごとに、第1のアドレス及び第2のアドレスを対応付けている。第2のアドレスは、ミラーリングされた圧縮後データブロックを格納するデータボリュームのアドレスである。

【0088】

50

図17は、本発明の一実施形態に係るストレージ装置4におけるデータ書き込み処理を説明するためのフローチャートである。具体的には、同図は、データ書き込み処理時におけるディスクアダプタ63の動作を説明している。なお、本例では、RAIDグループは、4Dであるとする。

【0089】

すなわち、同図に示すように、ディスクアダプタ63は、共有メモリ64を参照し、未処理の書き込みコマンドを見つけると、キャッシュメモリ62上のそれに対応する書き込みデータを特定する。ディスクアダプタ63は、RAID1に従って、当該書き込みデータを構成するデータブロックに対して圧縮処理を行って、圧縮データブロックを生成すし(STEP1701)、さらに、当該圧縮データブロックを二重化する(STEP1702)。圧縮データブロック及びその二重化された圧縮データブロックは、キャッシュメモリ62における圧縮データ用ブロックに格納される。本例では、3つの圧縮データブロックを1データサイクルとして、二重化された圧縮後データブロックがそれに続いて格納される。

10

【0090】

ディスクアダプタ63は、次に、圧縮データブロックを系列の異なる疑似ディスクデバイスに分散させて格納するために以下の処理を行う。すなわち、ディスクアダプタ63は、まず、1データサイクル分の圧縮ブロックデータを異なる疑似ディスクデバイス(0)~疑似ディスクデバイス(2)に分散させて格納する(STEP1703)。次に、ディスクアダプタ63は、データサイクル変数Iに0をセットする(STEP1704)。

20

【0091】

ディスクアダプタ63は、データサイクル変数Iが0であるか否かを判断する(STEP1705)。つまり、ディスクアダプタ63は、疑似ディスクデバイス(0)に格納された圧縮データブロック(例えばD4-1)に対応する二重化された圧縮データブロック(例えばD4-1')を選択する。

【0092】

ディスクアダプタ63は、データサイクル変数Iが0であると判断する場合(STEP1705のYes)、当該二重化された圧縮データブロックを疑似ディスクデバイス(3)の実記憶領域に格納し(STEP1706)、データブロック管理テーブルに、当該二重化された圧縮データブロックのエントリを登録する(STEP1707)。

30

【0093】

これに対して、データサイクル変数Iが0以外である場合、ディスクアダプタ63は、当該二重化された圧縮後データブロックを、疑似ディスクデバイス(I-1)の拡張記憶領域に格納し(STEP1708)、データブロック管理テーブルに、当該二重化された圧縮データブロックのエントリを登録する(STEP1709)。つまり、疑似ディスクデバイス(0)以外の疑似ディスクデバイスに格納された圧縮データブロックに対応する二重化された圧縮データブロックは、1つ分だけローテイトシフトされた疑似ディスクデバイスに格納されることになる。

【0094】

そして、ディスクアダプタ63は、データサイクル変数Iの値を1つインクリメントして(STEP1710)、データサイクル変数Iが1データサイクルに達したか否かを判断する(STEP1711)。データサイクル変数Iが1データサイクルに達していない場合には(STEP1711のNo)、ディスクアダプタ63は、STEP1705の処理に戻る。一方、データサイクル変数Iが1データサイクルに達した場合には(STEP1711のYes)、当該書き込み処理を終了する。

40

【0095】

以上のように本実施形態によれば、RAID1の構成の下、書き込みデータの圧縮データをミラーリングして格納する際に、複製された圧縮データを仮想ボリュームに格納しているので、効率的なデータ格納を実現することができるようになる。

[第3の実施形態]

50

【0096】

第3の実施形態は、上記第1の実施形態で示された技術的特徴と上記第2の実施形態で示された技術的特徴との組み合わせである。具体的には、本実施形態では、ストレージ装置4は、データ書き込み処理時に、パリティグループごとにパリティデータブロックを生成し、各データブロックに対して圧縮処理を行う。ストレージ装置4は、圧縮処理の結果に基づいて、パリティを用いたデータ格納方式とミラーリングによるデータ格納方式のどちらが効率的なデータ格納であるかを判断し、効率的であるとされたデータ格納方式に従って、圧縮データブロックをデータボリュームに格納する。この場合、上述したように、圧縮データブロックは、適宜、仮想ボリュームに格納されることになる。

【0097】

図18及び図19は、本発明の一実施形態に係るストレージ装置4におけるデータ書き込みメカニズムを説明するための図である。具体的には、これらの図は、ある書き込みデータが圧縮された後、RAID5(3D+1P)またはRAID1(4D)の下で、疑似ディスクデバイスに格納される様子を示している。

【0098】

今、あるデータボリュームに対するデータD1の書き込み要求があり、続いて、データD5の書き込み要求があったとする。説明を簡略化するために、書き込みデータは、ブロックサイズが64KBである単一のデータブロックであるとする。RAID5制御の下、書き込みデータ(データブロック)をデータボリュームに格納するためには、それぞれ同一のパリティグループ内のデータブロックとともにパリティを算出して、これらを格納する必要はある。つまり、データブロックD1については、データブロックD2及びD3とともにパリティデータブロックP1が算出され、また、データブロックD5については、データブロックD4及びD6とともにパリティデータブロックP2が算出される。これらは、キャッシュメモリ62上の64KBブロックに格納された圧縮前のデータブロックである。

【0099】

データブロックD1～D3及びP1は、ディスクアダプタ63により圧縮され、図18(A)に示すようなデータブロックが得られたとする。これらは、キャッシュメモリ62上の16KBブロックに格納された圧縮データブロックである。上述したように、圧縮データブロックD1-2、D2-2、及びP1-2～P1-4は、データボリュームの実ボリュームにおける実記憶領域に格納しきれないデータブロックである。

【0100】

一方、データブロックD4～D6及びP2は、ディスクアダプタ63により圧縮され、図18(B)に示すような圧縮データブロックが得られたとする。つまり、データブロックD4～D6は、データボリュームの汁ボリュームにおける実記憶領域の1つに格納できるまでに圧縮された圧縮データブロックD4-1～D6-1に変換されている。また、データブロックP2は、圧縮効率が低く(圧縮率が高く)、圧縮データブロックP2-1～P2-4に変換されている。なお、本明細書では、圧縮率は、圧縮後のデータサイズを圧縮前のデータサイズで割った値をいうものとする。従って、圧縮率が低いということは、圧縮後のデータサイズが小さいことを意味し、圧縮効率が低いことになる。

【0101】

従って、データブロックD4～D6及びP2からは、7つの圧縮データブロックが生成される。これに対して、データブロックD4-1～D6-1が二重化されてデータボリュームに格納される場合、6個の圧縮データブロックが格納されれば足りる。このため、本実施形態のストレージ装置4は、書き込みデータを圧縮後、パリティを含めた圧縮データブロックの数をチェックし、より少ない数の圧縮データブロックを格納すれば足りる方式を選択する。

【0102】

従って、図19に示すように、圧縮データブロックD1-1、D2-1、D3-1、及びP-1は、RAID5制御の下、データボリュームを構成するRAIDグループにおけ

10

20

30

40

50

る疑似ディスクデバイス(0)～疑似ディスクデバイス(3)の実ボリュームにそれぞれ格納される。また、圧縮データブロックD1-2、D2-2、及びP-2～P-4は、当該RAIDグループにおける対応する疑似ディスクデバイス(0)～疑似ディスクデバイス(4)の仮想ボリュームにそれぞれ格納される。

【0103】

また、圧縮後データブロックD4-1～D6-1は、RAID1制御の下、ミラーリングされ、拡張記憶領域を利用しながら、疑似ディスクデバイス(0)～疑似ディスクデバイス(3)に格納される。

【0104】

図20は、本発明の一実施形態に係るストレージ装置4における共有メモリ64に格納されたディスクデバイス管理テーブルの一例を示す図である。具体的には、同図は、上述の例に従って、パリティを用いたデータ格納方式により書き込みデータが格納されたときのディスクデバイス管理テーブルを示している。ミラーリングによるデータ格納方式により書き込みデータが格納されたときのディスクデバイス管理テーブルは、図16に示したものと同一である。

10

【0105】

図21～図23は、本発明の一実施形態に係るストレージ装置4におけるデータ書き込み処理を説明するためのフローチャートである。具体的には、同図は、データ書き込み処理時におけるディスクアダプタ63の動作を説明している。

【0106】

20

すなわち、ディスクアダプタ63は、共有メモリ64を参照し、未処理の書き込みコマンドを見つけると、キャッシュメモリ62上のそれに対応する書き込みデータを特定する。ディスクアダプタ63は、当該書き込みデータに基づいて、RAID構成に従ったパリティデータを生成する(STEP2101)。つまり、本例では、ディスクアダプタ63は、RAID5(3D+1P)の構成に従い、3つのデータブロックごとに1つのパリティデータブロックからなるパリティグループを形成する。このフェーズでは、各データブロックは、64KBのブロックサイズである。

【0107】

続いて、ディスクアダプタ63は、形成したパリティグループ内の各データブロックに対して圧縮処理を行って、圧縮データ列をそれぞれ生成する(STEP2102)。ディスクアダプタ63は、各圧縮データ列を構成する圧縮データブロックのそれぞれをキャッシュメモリ62における圧縮データ用ブロックに書き込み、キャッシュ管理テーブルを更新する。圧縮データ用ブロックは、例えば、16KBのブロックサイズである。

30

【0108】

ディスクアダプタ63は、次に、パリティグループ内の圧縮データブロック(圧縮パリティデータブロックを含まない。)数が圧縮パリティデータブロック数よりも少ないか否かを判断する(STEP2103)。つまり、ディスクアダプタ63は、パリティを用いたデータ格納方式とミラーリングによるデータ格納方式のどちらが書き込みデータを効率的に格納できるか否かを判断する。

【0109】

40

ディスクアダプタ63は、パリティグループ内の圧縮データブロック数の方が圧縮パリティデータブロック数よりも少ないと判断する場合には、図22に示す処理を行う。図22に示す処理は、パリティを用いたデータ格納処理であり、図12に示したSTEP1203以降の処理と同じであるため、説明を省略する。

【0110】

これに対して、ディスクアダプタ63は、パリティグループ内の圧縮後データブロック数の方が圧縮後パリティデータブロック数よりも多いと判断する場合には、図23に示す処理を行う。図23に示す処理は、ミラーリングによるデータ格納処理であり、図17に示したSTEP1702以降の処理と同じであるため、説明を省略する。

【0111】

50

以上のように、本実施形態によれば、書き込みデータを圧縮して格納するに際して、パリティを含む圧縮データのサイズとパリティを用いずに二重化した圧縮データのサイズとに基づいて最適な R A I D 構成を選択するので、圧縮データの効率的な格納ができるようになる。

[第 4 の実施形態]

【 0 1 1 2 】

第 4 の実施形態は、第 3 の実施形態の変形であり、データの書き込み要求に対する前回のデータ格納方式を記憶しておき、当該データに対する別の書き込み要求があった場合に、当該記憶されたデータ格納方式を利用することを特徴としている。

【 0 1 1 3 】

図 2 4 (A) 及び (B) は、本発明の一実施形態に係るストレージ装置 4 における共有メモリ 6 4 に格納されたディスクデバイス管理テーブルの一例を示す図である。同図に示すように、本実施形態のディスクデバイス管理テーブルは、上述したディスクデバイス管理テーブルのほぼ同じであるが、前回のデータ格納方式を記憶する学習情報を保持する点が異なっている。すなわち、同図 (A) は、R A I D 5 を学習情報として保持した状態を示し、同図 (B) は、学習情報がない状態を示している。

【 0 1 1 4 】

図 2 5 は、本発明の一実施形態に係るストレージ装置 4 におけるデータ書き込み処理を説明するためのフローチャートである。

【 0 1 1 5 】

同図に示すように、ディスクアダプタ 6 3 は、共有メモリ 6 4 を参照し、未処理の書き込みコマンドを見つけると、キャッシュメモリ 6 2 上のそれに対応する書き込みデータを特定する。ディスクアダプタ 6 3 は、当該書き込みデータに対するデータブロック管理テーブルを参照して、現在の学習情報を取得し (S T E P 2 5 0 1)、現在の学習情報が “ R A I D 1 ” であるか否かを判断する (S T E P 2 5 0 2)。

【 0 1 1 6 】

ディスクアダプタ 6 3 は、現在の学習情報が “ R A I D 1 ” でないと判断する場合 (S T E P 2 5 0 2 の N o)、R A I D 5 の制御の下で、圧縮データの格納処理を行う。

【 0 1 1 7 】

すなわち、ディスクアダプタ 6 3 は、すでに述べたように、R A I D 構成に従ったパリティデータブロックを生成し (S T E P 2 5 0 3)、続いて、パリティグループ内の各データブロックを圧縮する (S T E P 2 5 0 4)。

【 0 1 1 8 】

ディスクアダプタ 6 3 は、次に、パリティグループ内の圧縮データブロック (圧縮パリティデータブロックを含まない。) 数が圧縮パリティデータブロック数よりも少ないか否かを判断する (S T E P 2 5 0 5)。つまり、ディスクアダプタ 6 3 は、パリティを用いたデータ格納方式とミラーリングを用いたデータ格納方式のどちらが書き込みデータを効率的に格納できるか否かを判断する。

【 0 1 1 9 】

ディスクアダプタ 6 3 は、パリティグループ内の圧縮データブロック数の方が圧縮パリティデータブロック数よりも多いと判断する場合には、ミラーリングによるデータ格納処理を行う (S T E P 2 5 0 9)。ミラーリングによるデータ格納処理は、図 2 3 に示した処理と同じである。

【 0 1 2 0 】

これに対して、ディスクアダプタ 6 3 は、パリティグループ内の圧縮データブロック数の方が圧縮パリティデータブロック数よりも少ないと判断する場合には、パリティを用いたデータ格納処理を行う (S T E P 2 5 0 6)。パリティを用いたデータ格納処理は、図 2 2 に示した処理と同じである。

【 0 1 2 1 】

そして、ディスクアダプタ 6 3 は、次回のデータ書き込み処理のため、データブロック

10

20

30

40

50

管理テーブルの学習情報を“RAID5”に設定する(STEP2507)。

【0122】

STEP2502において、現在の学習情報が“RAID1”であると判断する場合(STEP2502のYes)、RAID1の制御の下で、圧縮データの格納処理を行う。

【0123】

すなわち、ディスクアダプタ63は、まず、データブロックを圧縮(STEP2508)、続いて、ミラーリングによるデータ格納処理を行う(STEP2509)。

【0124】

ミラーリングによるデータ格納処理を終えた後、ディスクアダプタ63は、学習情報を更新するか否かを判断すべく、RAID構成に従ったパリティデータブロックを生成し(STEP2510)、続いて、パリティグループ内の各データブロックを圧縮する(STEP2511)。ディスクアダプタ53は、パリティグループ内の圧縮データブロック(圧縮後パリティデータブロックを含まない。)数が圧縮パリティデータブロック数よりも少ないか否かを判断する(STEP2512)。

10

【0125】

ディスクアダプタ63は、パリティグループ内の圧縮データブロック数の方が圧縮パリティデータブロック数よりも少ないと判断する場合には(STEP2512のYes)、データブロック管理テーブルの学習情報を“RAID5”に設定する(STEP2507)。

【0126】

20

これに対して、ディスクアダプタ63は、パリティグループ内の圧縮後データブロック数の方が圧縮後パリティデータブロック数よりも多いと判断する場合には(STEP2512のNo)、データブロック管理テーブルの学習情報を“RAID1”に設定する(STEP2513)。

【0127】

以上により、本実施形態によれば、前回のデータ格納方式を記憶しているため、さらに効率的なRAID構成の選択を行うことができるようになる。

【0128】

上記実施形態は、本発明を説明するための例示であり、本発明は上記実施形態に限定されるものでない。本発明は、その要旨を逸脱しない範囲でさまざまに形態で実施することができる。

30

【0129】

例えば、上記実施形態では、ディスクアダプタの動作をフローチャートの中でシーケンスに構成したが、動作果に矛盾が生じない限り、処理の順序を入れ替え、又は、並列に処理するように構成してもよい。

【0130】

また、上記実施形態では、RAID5を採用したが、パリティを用いる他のRAID構成として知られるRAID3、4、又は6であってもよい。

【産業上の利用可能性】

【0131】

40

本発明は、RAID構成を採用するストレージ装置に広く適用することができる。

【図面の簡単な説明】

【0132】

【図1】本発明の一実施形態に係るストレージシステムの構成を説明するための図である。

【図2】本発明の一実施形態に係るストレージ装置におけるディスクコントローラのディスクアダプタを機能的に説明するための図である。

【図3】本発明の一実施形態に係るストレージ装置における共有メモリの内容を示す図である。

【図4】本発明の一実施形態に係るストレージ装置におけるセグメント管理テーブルの一

50

例を示す図である。

【図 5】本発明の一実施形態に係るストレージ装置の共有メモリに格納されたアドレス管理テーブルの一例を示す図である。

【図 6】本発明の一実施形態に係るストレージ装置の共有メモリに格納されたデータボリュームを説明するための図である。

【図 7】本発明の一実施形態に係るストレージ装置における記憶領域の動的割り当てを概念的に説明するための図である。

【図 8】本発明の一実施形態に係るストレージ装置における共有メモリに格納されたデータブロック管理テーブルの一例を示す図である。

【図 9】本発明の一実施形態に係るストレージ装置における共有メモリに格納されたディスクデバイス管理テーブルの一例を示す図である。

10

【図 10】本発明の一実施形態に係るストレージ装置におけるデータ書き込みメカニズムを説明するための図である。

【図 11】本発明の一実施形態に係るストレージ装置におけるデータ書き込みメカニズムを説明するための図である。

【図 12】本発明の一実施形態に係るストレージ装置におけるデータ書き込み処理を説明するためのフローチャートである。

【図 13】本発明の一実施形態に係るストレージ装置における仮想ボリュームに対するデータ書き込み処理を説明するためのフローチャートである。

【図 14】本発明の一実施形態に係るストレージ装置におけるデータ書き込みメカニズムを説明するための図である。

20

【図 15】本発明の一実施形態に係るストレージ装置におけるデータ書き込みメカニズムを説明するための図である。

【図 16】本発明の一実施形態に係るストレージ装置における共有メモリに格納されたディスクデバイス管理テーブルの一例を示す図である。

【図 17】本発明の一実施形態に係るストレージ装置におけるデータ書き込み処理を説明するためのフローチャートである。

【図 18】本発明の一実施形態に係るストレージ装置におけるデータ書き込みメカニズムを説明するための図である。

【図 19】本発明の一実施形態に係るストレージ装置におけるデータ書き込みメカニズムを説明するための図である。

30

【図 20】本発明の一実施形態に係るストレージ装置における共有メモリに格納されたディスクデバイス管理テーブルの一例を示す図である。

【図 21】本発明の一実施形態に係るストレージ装置におけるデータ書き込み処理を説明するためのフローチャートである。

【図 22】本発明の一実施形態に係るストレージ装置におけるデータ書き込み処理を説明するためのフローチャートである。

【図 23】本発明の一実施形態に係るストレージ装置におけるデータ書き込み処理を説明するためのフローチャートである。

【図 24】、本発明の一実施形態に係るストレージ装置における共有メモリに格納されたディスクデバイス管理テーブルの一例を示す図である。

40

【図 25】本発明の一実施形態に係るストレージ装置におけるデータ書き込み処理を説明するためのフローチャートである。

【符号の説明】

【0133】

1 ... コンピュータシステム

2 ... ネットワーク

3 ... ホスト装置

4 ... ストレージ装置

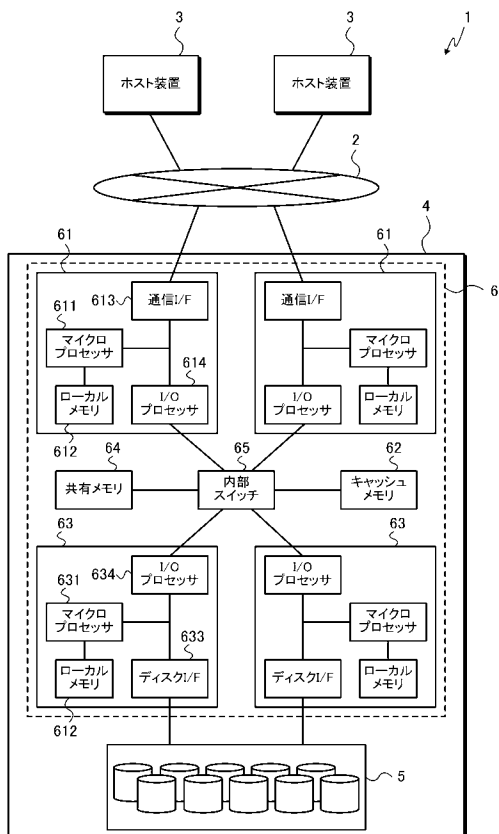
5 ... ディスクデバイス

50

- 6 ... ディスクコントローラ
- 6 1 ... チャンネルアダプタ
- 6 2 ... キャッシュメモリ
- 6 3 ... ディスクアダプタ
- 6 4 ... 共有メモリ
- 6 5 ... 内部スイッチ

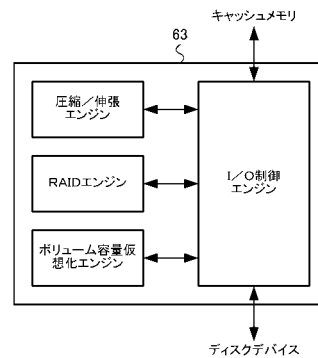
【図 1】

図1



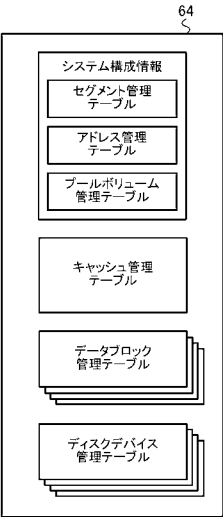
【図 2】

図2



【図 3】

図3



【図 4】

図4

セグメント管理テーブル

Disk ID	Segment #	LBA_start	LBA_end	In_use
0	0	0	499	1
0	1	500	1799	1
0	2	1800	2999	0
0	3	1800	2999	1
⋮	⋮	⋮	⋮	⋮
1	6	0	699	0
1	7	700	2999	1
⋮	⋮	⋮	⋮	⋮

【図 5】

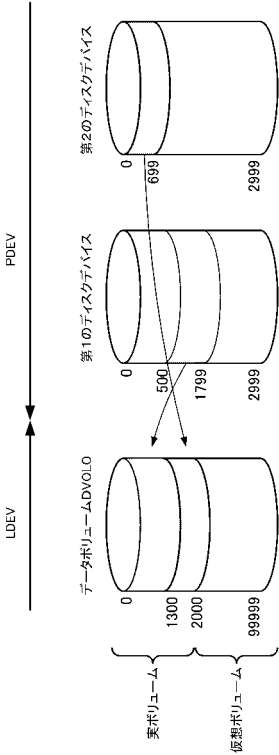
図5

アドレス管理テーブル

LUN	Segment #	LBA_start	LBA_end	Pool
0	1	0	1299	0
0	6	1300	1999	0
1	2	0	1199	0
2	8	0	299	0
⋮	⋮	⋮	⋮	⋮
9	8	19000	19999	1
⋮	⋮	⋮	⋮	⋮

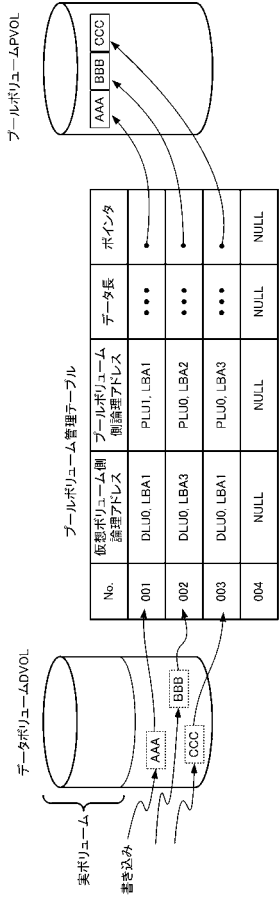
【図 6】

図6



【図 7】

図7



【図 8】

図8

データブロック管理テーブル

スロットID	アドレス
D1-1	DEVICE(0): xxxx
D2-1	DEVICE(1): xxxx
D3-1	DEVICE(2): xxxx
P1-3	DEVICE(3): xxxx
D1-2	DEVICE(0): xxxx
D2-2	DEVICE(1): xxxx
D3-2	UNUSED
P1-2	DEVICE(3): xxxx
D1-3	UNUSED
D2-3	UNUSED
D3-3	UNUSED
P1-3	DEVICE(3): xxxx
D1-4	UNUSED
D2-4	UNUSED
D3-4	UNUSED
P1-4	DEVICE(3): xxxx

【図 9】

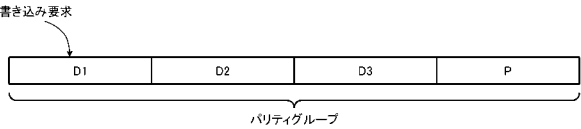
図9

ディスクデバイス管理テーブル

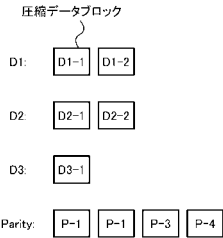
	アドレス	使用状況
実ボリューム	xxxx	D1-1
	:	:
	:	:
	:	:
	:	:
仮想ボリューム	拡張1	D2-1
	拡張2	Free
	拡張3	Free
	:	:
	拡張n	Null

【図 10】

図10



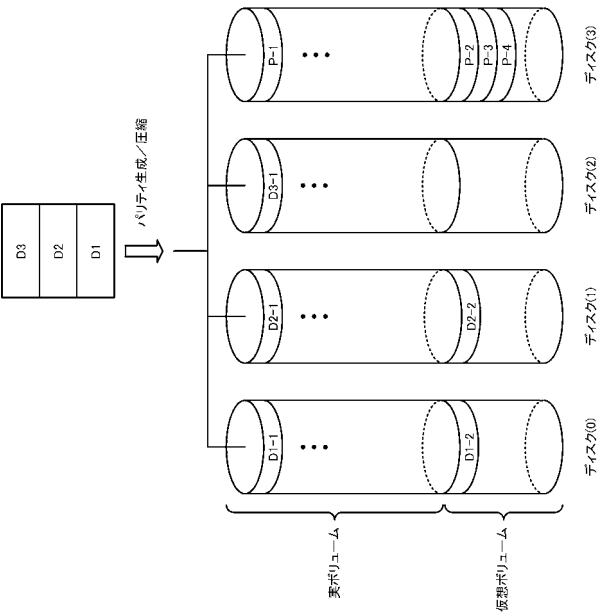
(A)



(B)

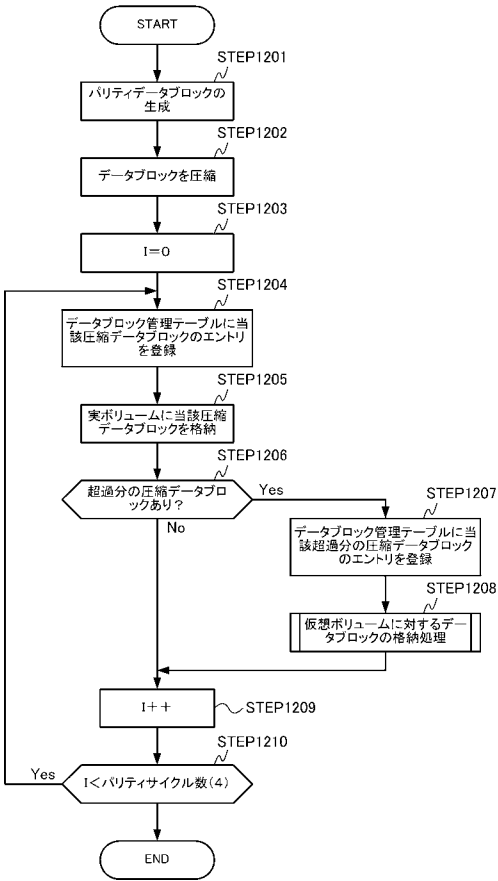
【図 11】

図11



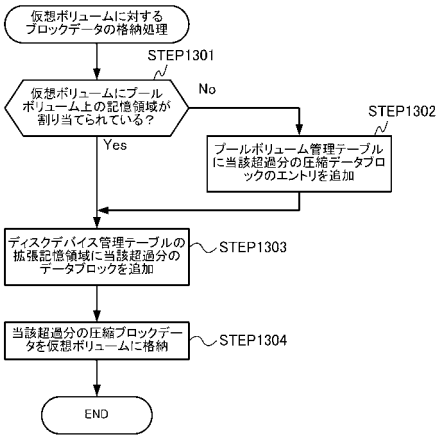
【図 1 2】

図12



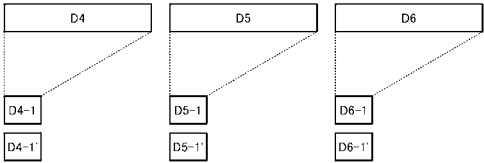
【図 1 3】

図13



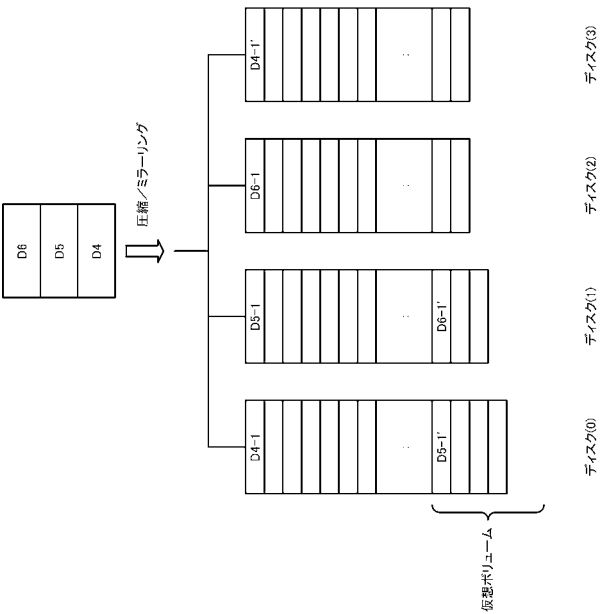
【図 1 4】

図14



【図 1 5】

図15



【図 1 6】

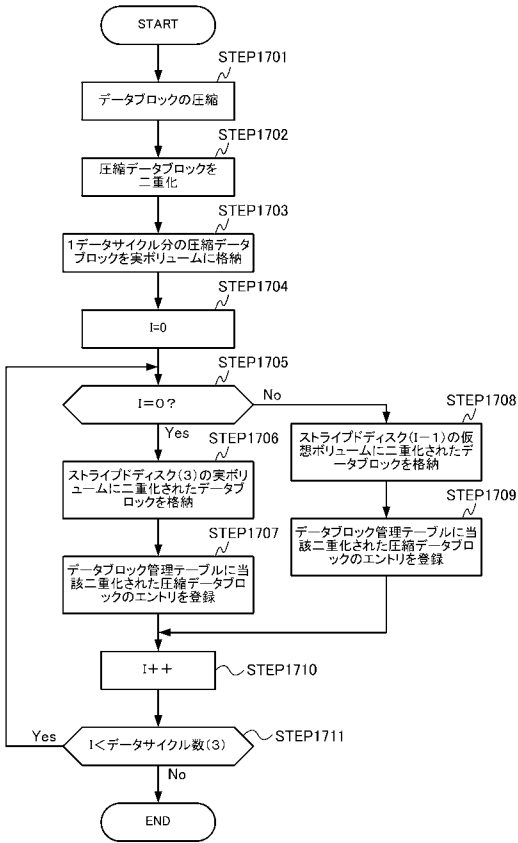
図16

データブロック管理テーブル

RAID LEVEL	RAID 1	
スロットID	アドレス1	アドレス2
D1-1	DEVICE(0): xxxxx	DEVICE(3): xxxxx
D2-1	DEVICE(1): xxxxx	DEVICE(0): xxxxx
D3-1	DEVICE(2): xxxxx	DEVICE(1): xxxxx
P1-3	UNUSED	UNUSED
D1-2	UNUSED	UNUSED
D2-2	UNUSED	UNUSED
D3-2	UNUSED	UNUSED
P1-2	UNUSED	UNUSED
D1-3	UNUSED	UNUSED
D2-3	UNUSED	UNUSED
D3-3	UNUSED	UNUSED
P1-3	UNUSED	UNUSED
D1-4	UNUSED	UNUSED
D2-4	UNUSED	UNUSED
D3-4	UNUSED	UNUSED
P1-4	UNUSED	UNUSED

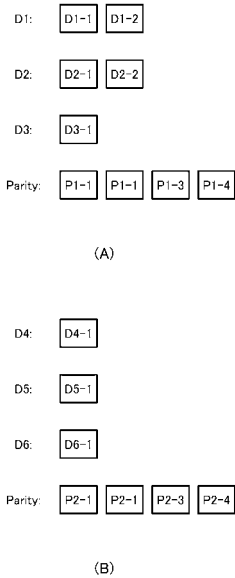
【図 17】

図17



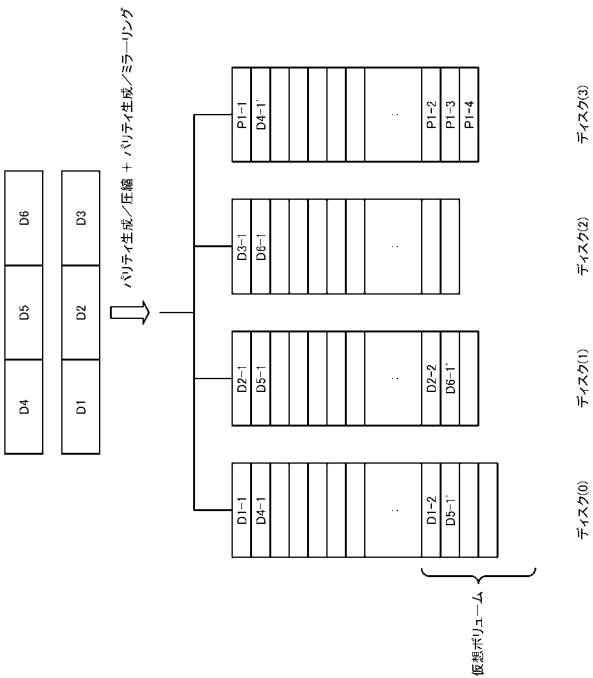
【図 18】

図18



【図 19】

図19



【図 20】

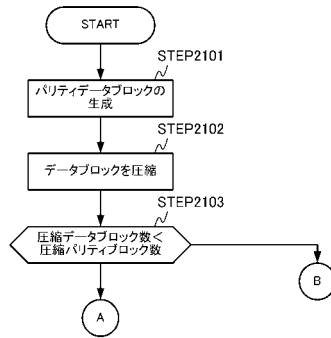
図20

データブロック管理テーブル

RAID LEVEL	RAID 5	
スロットID	アドレス	アドレス2
D1-1	DEVICE(0): xxxxx	UNUSED
D2-1	DEVICE(1): xxxxx	UNUSED
D3-1	DEVICE(2): xxxxx	UNUSED
P1-3	DEVICE(3): xxxxx	UNUSED
D1-2	DEVICE(0): xxxxx	UNUSED
D2-2	DEVICE(1): xxxxx	UNUSED
D3-2	UNUSED	UNUSED
P1-2	DEVICE(3): xxxxx	UNUSED
D1-3	UNUSED	UNUSED
D2-3	UNUSED	UNUSED
D3-3	UNUSED	UNUSED
P1-3	DEVICE(3): xxxxx	UNUSED
D1-4	UNUSED	UNUSED
D2-4	UNUSED	UNUSED
D3-4	UNUSED	UNUSED
P1-4	DEVICE(3): xxxxx	UNUSED

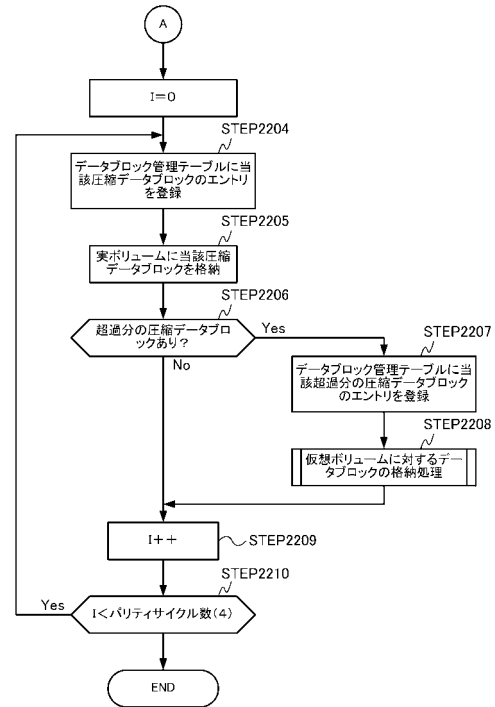
【図 2 1】

図21



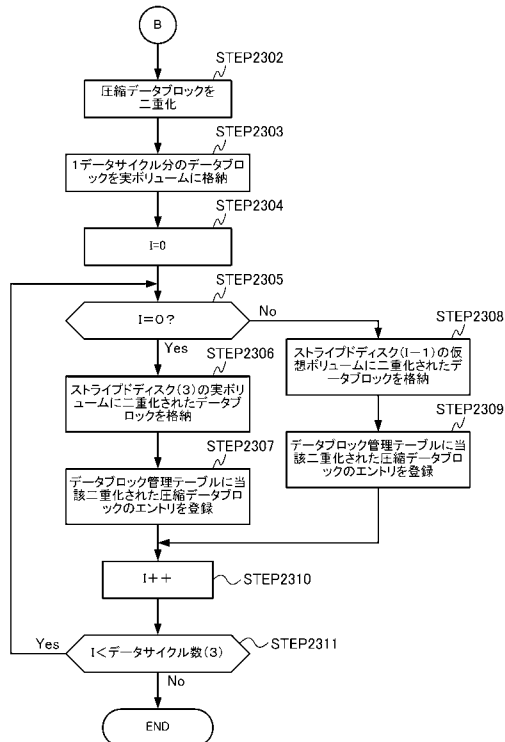
【図 2 2】

図22



【図 2 3】

図23



【図 2 4】

図24

データブロック管理テーブル

RAID 1		アドレス	
RAID LEVEL	装置情報	アドレス1	アドレス2
	スロットID	DEVICE(0): xxxx	DEVICE(3): xxxx
	D1-1	DEVICE(1): xxxx	DEVICE(0): xxxx
	D2-1	DEVICE(2): xxxx	DEVICE(1): xxxx
	D3-1	UNUSED	UNUSED
	P1-3	UNUSED	UNUSED
	D1-2	UNUSED	UNUSED
	D2-2	UNUSED	UNUSED
	D3-2	UNUSED	UNUSED
	P1-2	UNUSED	UNUSED
	D1-3	UNUSED	UNUSED
	D2-3	UNUSED	UNUSED
	D3-3	UNUSED	UNUSED
	P1-3	UNUSED	UNUSED
	D1-4	UNUSED	UNUSED
	D2-4	UNUSED	UNUSED
	D3-4	UNUSED	UNUSED
	P1-4	UNUSED	UNUSED

(B)

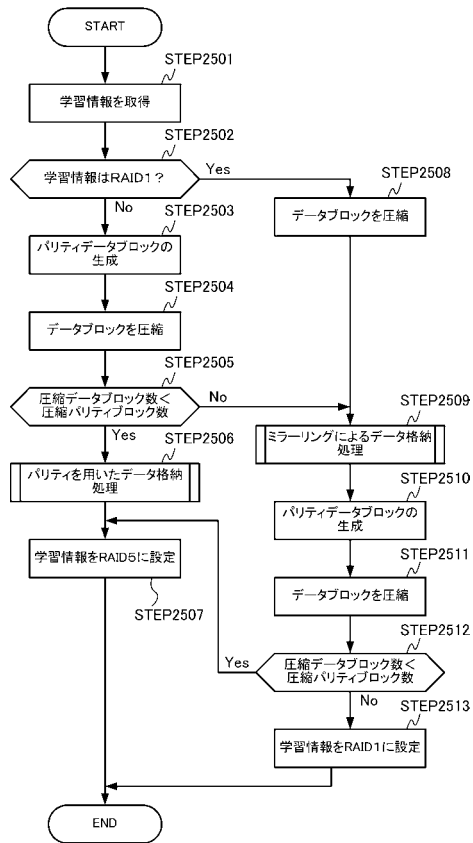
データブロック管理テーブル

RAID 5		アドレス	
RAID LEVEL	装置情報	アドレス1	アドレス2
	スロットID	DEVICE(0): xxxx	UNUSED
	D1-1	DEVICE(1): xxxx	UNUSED
	D2-1	DEVICE(2): xxxx	UNUSED
	D3-1	DEVICE(3): xxxx	UNUSED
	P1-3	DEVICE(0): xxxx	UNUSED
	D1-2	DEVICE(1): xxxx	UNUSED
	D2-2	UNUSED	UNUSED
	D3-2	UNUSED	UNUSED
	P1-2	DEVICE(3): xxxx	UNUSED
	D1-3	UNUSED	UNUSED
	D2-3	UNUSED	UNUSED
	D3-3	UNUSED	UNUSED
	P1-3	DEVICE(3): xxxx	UNUSED
	D1-4	UNUSED	UNUSED
	D2-4	UNUSED	UNUSED
	D3-4	UNUSED	UNUSED
	P1-4	DEVICE(3): xxxx	UNUSED

(A)

【図 25】

図25



フロントページの続き

(72)発明者 神道 一恵

神奈川県小田原市中里322番2号 株式会社日立製作所RAIDシステム事業部内

審査官 菅原 浩二

(56)参考文献 特開平05-189157(JP,A)

特開2004-013373(JP,A)

特開平07-261938(JP,A)

特開2007-199891(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F 3/06