

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4439960号
(P4439960)

(45) 発行日 平成22年3月24日(2010.3.24)

(24) 登録日 平成22年1月15日(2010.1.15)

(51) Int.Cl.

F I

G 0 6 F 12/00 (2006.01)

G 0 6 F 12/00 5 3 1 M

G 0 6 F 12/00 5 1 0 B

請求項の数 4 (全 14 頁)

(21) 出願番号 特願2004-83120 (P2004-83120)
 (22) 出願日 平成16年3月22日(2004.3.22)
 (65) 公開番号 特開2005-267569 (P2005-267569A)
 (43) 公開日 平成17年9月29日(2005.9.29)
 審査請求日 平成19年1月23日(2007.1.23)

(73) 特許権者 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 100122884
 弁理士 角田 芳末
 (72) 発明者 永田 幸司
 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

審査官 野田 佳邦

最終頁に続く

(54) 【発明の名称】 ストレージ装置

(57) 【特許請求の範囲】

【請求項 1】

上位装置と他のストレージ装置とに接続され、前記上位装置からデータを受ける上位インターフェースと、

前記上位インターフェースに接続され、前記上位装置との間で送受信されるデータを保存するメモリと、

前記メモリに接続され、前記上位装置との間で送受信されるデータを、前記メモリに読み書きするように制御する複数のディスクインターフェースと、

前記複数のディスクインターフェースに接続され、前記複数のディスクインターフェースの制御のもとに、前記上位装置から送られたデータが格納される複数のディスクドライブと、

前記複数のディスクドライブの記憶領域を用いて生成される第1の論理ボリュームへのデータの読み書きを制御し、前記第1の論理ボリュームにデータの書き込みがあった場合に当該書き込み前に格納されていたデータを差分データとして第2の論理ボリュームに書き込むように制御し、前記第1の論理ボリュームを複数の第1のサイズの第1の記憶領域に分割して管理し、当該複数の第1の記憶領域それぞれと前記第2の論理ボリュームに格納されている複数の差分データそれぞれとの対応関係を管理する第1のスナップショット管理テーブルを前記メモリの領域に設ける制御プロセッサと、

を有するストレージ装置であって、

前記制御プロセッサは、

10

20

前記第 1 のスナップショット管理テーブルを参照して、前記第 1 の論理ボリュームに格納されたデータと前記第 2 の論理ボリュームに格納された差分データとを用いて、前記第 1 の論理ボリュームと対応する過去のデータを有する仮想論理ボリュームを提供し、

前記仮想論理ボリュームと前記他のストレージ装置の有する第 3 の論理ボリュームとの間で、前記第 1 のスナップショット管理テーブルを参照して、前記第 1 の論理ボリュームに格納されたデータと前記第 2 の論理ボリュームに格納された差分データとを用いて、リモートコピーを実施し、

前記仮想論理ボリュームと前記第 3 の論理ボリュームがスプリット状態の時に、前記第 1 の論理ボリュームにデータ書き込みがあった場合には、

前記第 1 の論理ボリュームを複数の第 2 のサイズの第 2 の記憶領域に分割して管理し、当該複数の第 2 の記憶領域それぞれについてデータ書き込みの有無を管理するリモートコピー管理情報と、さらに、前記複数の第 2 の記憶領域それぞれと前記第 2 の論理ボリュームに格納されている複数の差分データそれぞれとの対応関係を管理する第 2 のスナップショット管理テーブルとを前記メモリに設け、

前記仮想論理ボリュームと前記第 3 の論理ボリュームがスプリット状態からペア状態になった時に、

前記第 2 のサイズの第 2 の記憶領域それぞれについて管理されている、前記リモートコピー管理情報および前記第 2 のスナップショット管理テーブルを参照して、前記第 1 の論理ボリュームに格納されたデータと前記第 2 の論理ボリュームに格納された差分データとを用いて、前記仮想論理ボリュームの有する前記第 1 の論理ボリュームと対応する過去のデータを前記第 3 の論理ボリュームに格納するように制御する

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において、

前記第 1 の論理ボリュームを複数の第 1 の記憶領域に分割する単位は 6 4 K B であり、

前記リモートコピー管理情報は、前記第 1 の論理ボリュームを複数の 1 M B の第 2 の記憶領域に分割して管理し、当該複数の第 2 の記憶領域それぞれについてデータ書き込みの有無を管理する第 1 のビットマップであり、

前記第 2 のスナップショット管理テーブルは、前記複数の第 2 の記憶領域それぞれと前記第 2 の論理ボリュームに格納されている複数の差分データそれぞれとの対応関係を管理する第 2 のビットマップであり、

前記制御プロセッサは、前記仮想論理ボリュームと前記第 3 の論理ボリュームがスプリット状態からペア状態になった時に、前記仮想論理ボリュームの有する前記第 1 の論理ボリュームと対応する過去のデータのうち、前記第 1 のビットマップと前記第 2 のビットマップの論理和によって示される第 2 の記憶領域に該当するデータを、前記第 3 の論理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 3】

請求項 1 記載のストレージ装置において、

前記制御プロセッサは、前記リモートコピーを指示された場合、前記第 1 のスナップショット管理テーブルを参照して、前記仮想論理ボリュームのリモートコピー対象のアドレスを決定し、当該アドレスに対応する実際のデータが前記第 1 の論理ボリューム又は前記第 2 の論理ボリュームのいずれに格納されているかを判定し、

前記第 1 の論理ボリュームに前記アドレスに対応する実際のデータが格納されている場合、前記複数の第 1 のサイズの第 1 の記憶領域に分割して管理された前記第 1 の論理ボリュームから該当するデータを転送し、

前記第 2 の論理ボリュームに前記アドレスに対応する実際のデータが格納されている場合、前記第 2 の論理ボリュームから該当する差分データを転送する

ことを特徴とするストレージ装置。

【請求項 4】

請求項 1 記載のストレージ装置において、
前記上位インターフェースと前記複数のディスクインターフェースとの間で送受信される前記データが一時的に記憶されるキャッシュメモリ、を更に備え、
前記仮想論理ボリュームと前記第 3 の論理ボリュームがスプリット状態からペア状態になった時に、前記仮想論理ボリュームの有する前記第 1 の論理ボリュームと対応する過去のデータとして、前記キャッシュメモリに記憶された該当データを前記第 3 の論理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。

【発明の詳細な説明】

【技術分野】

10

【0001】

本発明は、例えばホストコンピュータで作成され、ストレージ装置に書き込まれたデータのリモートコピーの作成に使用して好適なストレージ装置および情報管理システムに関する。詳しくは、いわゆるスナップショット機能を有するストレージ装置において、良好なリモートコピーの作成が行われるようにしたものである。

【背景技術】

【0002】

いわゆるスナップショット機能を有しているストレージ装置において、多世代のデータを複数のストレージ装置で保存し、障害が発生した場合には、これらの保存されたデータを用いて円滑にリストアを行う手段が提案されている（例えば、特許文献 1 参照。）。 20

【0003】

しかしながら、上述の特許文献 1 では、本発明が開示するスナップショット機能を有効に活用して、効率良くリモートコピーを作成する手段については、全く記載されていないものである。

【特許文献 1】特開 2003 - 242011 号公報

【発明の開示】

【発明が解決しようとする課題】

【0004】

従来のストレージ装置において、リモートコピーを作成する場合には、例えばリモートコピーの対象となる記憶領域（正ボリューム）と同じデータの記憶される副ボリュームを形成し、副ボリュームからリモートコピーを作成することが行われている。しかしながらこの方法では、副ボリュームを形成するために正ボリュームと同じ記憶容量が必要となり、極めて大容量の記憶領域が要求される。 30

【0005】

一方、いわゆるスナップショット機能を有しているストレージ装置においては、正ボリュームに対する読み書きが行われると、それ以前に記憶されていたデータをプール領域と呼ばれる記憶領域に保存し、さらにその保存データと元の正ボリュームとの関連をメモリ上に仮想ボリュームとして記憶することが行われている。これによれば、順次記憶される多世代のデータを少ない記憶容量で保存することができる。 40

【0006】

そこで、このようなスナップショット機能を有しているストレージ装置において、リモートコピーを作成する場合には、メモリ上の仮想ボリュームに基づいてリモートコピーの更新を行うことができる。すなわち、上述の仮想ボリュームには読み書きの行われたデータに関する記載があり、これを参照してリモートコピーの書き換えを行うことにより、リモートコピーの更新を行うことができる。 50

【0007】

ところが上述の仮想ボリュームは、例えばホストコンピュータのデータ処理の単位である 64 kB ごとにスナップショットの有無と、アドレスを保存しているものであり、現実にはどのデータにスナップショットが存在しているかを知ることはできない。したがって、このような仮想ボリュームからリモートコピーを作成する場合には、スナップショットの

有無を示すマップを生成する必要があるが、その生成には多くの時間が必要とされる。

【 0 0 0 8 】

すなわち、上述のスナップショット機能を有しているストレージ装置において、リモートコピーを作成する場合には、メモリ上の仮想ボリュームに基づいてリモートコピーの更新を行うことができるものの、それを行うにはスナップショットの有無を示すマップを生成する必要があり、その生成には時間が掛かるために、全体として効率の良いリモートコピーの作成を行うことができなかった。

【 0 0 0 9 】

本発明はこのような点に鑑みて成されたものであって、発明が解決しようとする問題点は、スナップショット機能を有しているストレージ装置において、リモートコピーを作成する場合に、従来の装置では、全体として効率の良いリモートコピーの作成を行うことができなかったというものである。

【課題を解決するための手段】

【 0 0 1 0 】

上記の課題を解決するため、本発明においては、スナップショット管理テーブルにより仮想論理ボリュームを生成し、仮想論理ボリュームを用いてリモートコピーの作成を行うことを第1の手段とするものである。また、仮想論理ボリュームを更新する際に、リモートコピーとの差分の生じた領域を示す第1のビットマップを作成し、さらにスナップショットの差分データを書き込む際に、差分データの領域を示す第2のビットマップを作成することを第2の手段とするものである。

【発明の効果】

【 0 0 1 1 】

これにより、従来の装置では、スナップショット機能を有しているストレージ装置において、リモートコピーを作成する場合に、全体として効率の良いリモートコピーの作成を行うことができなかったものを、本発明によれば、第1のビットマップおよび第2のビットマップを作成することで、これらを用いて、効率の良いリモートコピーの作成を行うことができ、したがって従来の問題点を解消することができるものである。

【発明を実施するための最良の形態】

【 0 0 1 2 】

以下、図面を参照して本発明を説明するに、図1は本発明のストレージ装置を適用したディスクアレイ装置、およびそれを用いた本発明の情報管理システムの一実施形態の構成を示すブロック図である。

【 0 0 1 3 】

図1において、この実施形態のディスクアレイ装置1は、ディスクアレイコントローラ10およびディスク20を含んで構成される。また、ディスクアレイ装置1は、S A N (Storage Area Network) 2を介して複数のホスト3に接続されると共に、L A N (Local Area Network) 4を介して管理用端末装置5に接続されている。

【 0 0 1 4 】

ディスクアレイコントローラ10は、制御プログラム103の動作によって、ディスク20に対するデータの入出力を制御する。また、ディスク20によってR A I D (Redundant Array of Independent Disks) が構成されており、記憶されるデータに冗長性を持たせている。このため、ディスクの一部に障害が生じても、記憶されたデータが消失しないようになっている。

【 0 0 1 5 】

さらに、ディスクアレイコントローラ10には、C P U 1 0 1、メモリ102、データ転送コントローラ104、フロントエンドインターフェース105、バックエンドインターフェース106、キャッシュメモリ107、及びL A N インターフェース108が設けられている。

【 0 0 1 6 】

メモリ102には制御プログラム103(図2参照)が記憶されており、C P U 1 0 1

10

20

30

40

50

が制御プログラム 103 を呼び出して実行することによって各種処理が行われる。データ転送コントローラ 104 は、CPU 101、フロントエンドインターフェース 105、バックエンドインターフェース 106、及びキャッシュメモリ 107 の間でデータを転送する。

【0017】

フロントエンドインターフェース 105 は、SAN 2 に対するインターフェースであって、例えば、ファイバチャネルプロトコルによって、ホスト 3 との間でデータや制御信号を送受信する。バックエンドインターフェース 106 は、ディスク 20 に対するインターフェースであって、例えば、ファイバチャネルプロトコルによって、ディスク 20 との間でデータや制御信号を送受信する。

10

【0018】

キャッシュメモリ 107 には、フロントエンドインターフェース 105 とバックエンドインターフェース 106 との間で送受信されるデータが一時的に記憶されるキャッシュが設けられている。すなわち、データ転送コントローラ 104 は、SAN 4 を介してディスクに読み書きされるデータをインターフェース 105、106 間で転送する。さらに、これらのディスクに読み書きされるデータをキャッシュメモリ 107 に転送する。

【0019】

LAN インターフェース 108 は、LAN 4 に対するインターフェースであって、例えば、TCP/IP プロトコルによって、管理用端末装置 5 との間でデータや制御信号を送受信することができる。SAN 2 は、例えばファイバチャネルプロトコルのような、データの転送に適するプロトコルで通信可能なネットワークである。

20

【0020】

ホスト 3 は、CPU、メモリ、記憶装置、インターフェース、入力装置及び表示装置が備わるコンピュータ装置であり、ディスクアレイ装置 1 から提供されるデータを利用して、データベースサービスやウェブサービス等を利用可能にする。LAN 4 は、ディスクアレイ装置 1 の管理用に用いられるもので、例えば、TCP/IP プロトコルによって、コンピュータ間でデータや制御情報を通信可能であり、例えばイーサネット（登録商標）が用いられる。

【0021】

管理用端末装置 5 は、CPU、メモリ、記憶装置、インターフェース、入力装置及び表示装置が備わるコンピュータ装置である。管理用端末装置 5 では管理プログラムが動作しており、該管理プログラムによってディスクアレイ装置 1 の動作状態を把握し、ディスクアレイ装置 1 の動作を制御する。

30

【0022】

なお、管理用端末装置 5 ではウェブブラウザ等のクライアントプログラムが動作しており、ディスクアレイ装置 1 から CGI (Common Gateway Interface) 等によって供給される管理プログラムによってディスクアレイ装置 1 の動作を制御してもよい。

【0023】

このようなディスクアレイ装置、およびそれを利用した情報管理システムにおいて、図 2 には、上述の制御プログラム 103 の一例を説明図で示す。

40

【0024】

図 2 において、ホスト 3 の通常 I/O 処理プログラム 301 から送られたデータ入出力要求は、ディスクアレイ装置 1 の制御プログラム 103 の R/W コマンド解析プログラム 111 によって解析され、スナップジョブプログラム 121 に送られる。スナップジョブプログラム 121 は、正ボリュームに対するデータ書き込み要求を受信すると、正ボリュームへのライト時に正ボリューム内の更新前のデータを、その格納場所であるプールボリュームに複写し、この複写後に正ボリュームの内容を更新する。

【0025】

さらに、スナップジョブプログラム 121 は、データが更新された正ボリューム内のブロックに対応する仮想ボリューム内のブロックを、スナップショット生成要求の受信時点

50

で正ボリューム内のブロックに対応する仮想ボリューム内のブロックを、スナップショット生成要求の受信時点で正ボリュームのデータ（すなわち更新前のデータ）が格納されているプールボリューム上のブロックと対応付けるようスナップショット管理テーブル（D D C B（Differential Data Control Block）204）を更新する。

【0026】

また、スナップリストアジョブプログラム122は、スナップショットの副から正ヘリストア処理をするものである。

【0027】

このようにして、ディスクアレイ装置1はスナップショットイメージを提供することが可能になる。そして、ホスト3が、通常I/O処理プログラム301によって、仮想ボリュームにアクセスすることによって、スナップショット生成要求発行時点での正ボリューム内の情報を利用することができる。

10

【0028】

また、通常I/O処理プログラム301から送られた制御コマンドは、その他コマンド解析プログラム112によって解析され、構成情報制御プログラム140に送られる。構成情報制御プログラム140のペア情報管理プログラム141は、スナップショット生成要求を受信すると、まずスナップショット管理テーブルに新しい仮想ボリュームの識別情報を登録する。この仮想ボリュームのブロックは、最初はスナップショット管理テーブルによって、一対一で対応付けられている。

【0029】

20

プールボリューム管理プログラム142は、後述するように、プール領域に登録されたボリュームの追加及び削除を管理する。プール管理プログラム150は、プールボリューム管理プログラム142に従ってプールそのものの管理を行う。また、WEBプログラム160は、ジョブをWEB上に展開するためのものである。

【0030】

さらにディスクアレイ装置1の制御プログラム103に設けられるRAIDマネージャプログラム111は、ホスト3のRAIDマネージャプログラム302と通信可能に接続されている。このRAIDマネージャプログラム121、302によって、スナップショットの生成やリモートコピーの生成、ペア状態の変更などを行うことができる。

【0031】

30

また、DAMPインターフェースプログラム122は、管理用端末装置5のDAMPプログラム501と通信可能に接続されており、このDAMPインターフェースプログラム122によって、管理用端末装置5のDAMPプログラム501との通信が行われて、ディスクアレイ装置1のRAIDの構成を管理することができる。

【0032】

さらに、このような装置を用いて、リモートコピーを行うには、図3のようなシステムが用いられる。図3において、ホストコンピュータ401に接続されるストレージ装置402には、正ボリューム411と複数のクイックシャドー412、413・・・41nが設けられている。これらの複数のクイックシャドー412、413・・・41nは、それぞれ例えば曜日ごとに分けて生成されるものである。

40

【0033】

そして、これらの複数のクイックシャドー412、413・・・41nに対して、それぞれ別のストレージ装置403、404・・・40nが設けられ、これらのストレージ装置403、404・・・40nにリモートコピーが行われる。

【0034】

さらに、図4は、本発明の実施の形態のスナップショットの管理方法の説明図である。この図4において、正ボリューム201は、通常の運用に供され、ホスト3からのデータ入出力の対象となる論理ユニット（P-VOL：Primary Volume）である。

【0035】

スナップショット差分ビットマップ202は、キャッシュメモリ107の管理領域に設

50

けられており、正ボリューム 201 のブロック（例えば、64 k バイト / ブロック）に対応するビットを有する。そして、正ボリューム 201 のブロックアドレスに対応する差分データがプールボリューム 206 に記録されている場合に、スナップショット差分ビットマップ 202 中の正ボリューム 201 に対応するビットが「1」となっている。このスナップショット差分ビットマップ 202 は、後述する正ボリュームアドレステーブル 203 及び DDCB 204 と共にスナップショット管理テーブルを構成する。

【0036】

したがって、正ボリューム 201 に対する書き込みが行われるときに、スナップショット差分ビットマップ 202 を参照することによって、更新前データをプールボリュームにコピーする必要があるかを判定することができる（すなわち、ビットが「1」であれば、スナップショット生成時のデータが既にプールボリューム 206 に書き込まれているので、正ボリューム 201 のデータをプールボリューム 206 にコピーする必要がない）。

10

【0037】

さらに、正ボリュームアドレステーブル 203 は、正ボリュームの 64 K バイトごとにその差分データの有無と、差分データの位置に従った情報を持つもので、キャッシュメモリ 107 の管理領域に設けられており、スナップショット差分ビットマップ 202 のビットと対応して DDCB 204 のアドレスが記録されている。

【0038】

DDCB 204 は、プールボリュームと一対で割り当てられており、キャッシュメモリ 107 の管理領域に設けられている。DDCB 204 は、プールボリューム 206 のブロック（例えば、64 k バイト / ブロック）毎に区切られて、その各々に管理テーブルが設けられている。この管理テーブルには、プールボリューム 206 のブロックに対応する位置に記録された差分データがどの世代のスナップショットのデータであるかが記録されたアドレスをたどることによって、複数世代の差分データを参照することができる。

20

【0039】

なお、DDCB 204 の使用されていない領域は、空きキューとしてリンクが設定されている。そして、この空きキューの量はキャッシュメモリ 107 に設けられた空きキューカウンタ 205 によって管理されている。

【0040】

プールボリューム 206 は、プール領域に登録されたボリュームによって構成される。このプールボリューム 206 によって、スナップショット作成時点における正ボリューム 201 のデータが論理的に複写されているように見せている。そして、プールボリューム 206 のデータがどの世代の差分データかは DDCB 204 の世代管理ビットマップによって知ることができる。

30

【0041】

よって、正ボリュームにデータを書き込むときは、まず、スナップショット差分ビットマップ 202 を参照して、更新前データをプールボリュームにコピーする必要があるかを判定する。そして、スナップショット差分ビットマップ 202 の対応ビットが「1」であれば、更新前データをプールボリュームにコピーする必要がないと判定し、正ボリュームにデータを書き込む。一方、スナップショット差分ビットマップ 202 の対応ビットが「0」であれば、更新前データをプールボリュームにコピーした後に、正ボリュームにデータを書き込む。

40

【0042】

そして、正ボリュームのブロックに対応する DDCB 204 のブロックに、新たに設定された差分データに対するリンクアドレスを設定する。そして、スナップショット差分ビットマップ 202 の対応ビットを「1」にし、プールボリューム 206 の当該更新前データを書き込んだアドレスに対応する DDCB 204 の世代管理ビットマップを設定する。そして、必要に応じて正ボリュームアドレステーブル 203 に、DDCB 204 のブロックのアドレスを設定する。さらに、DDCB 204 の空きキューを使用したので、空きキューカウンタ 205 を更新する。

50

【 0 0 4 3 】

また、仮想ボリューム (V - V O L : Virtual Volume) にアクセスするときには、正ボリュームアドレステーブル 2 0 3 を参照して、アクセス対象となる仮想ボリュームのブロックアドレス (正ボリュームのブロックアドレスに等しい) によって D D C B 2 0 4 のアドレスを特定して、 D D C B 2 0 4 の当該アドレスの世代管理ビットマップによって、アクセス対象の世代の差分データがあるかを特定する。

【 0 0 4 4 】

そして、所望の世代の差分データがあれば、当該 D D C B 2 0 4 のアドレスに対応するプールボリューム 2 0 6 のアドレスから差分データを読み出して、仮想ボリュームのイメージを提供する。一方、所望の世代の差分データでなければ、他の差分データに対するリンクアドレスを参照して、所望の世代の差分データを探す。そして、いずれの差分データも所望の世代のものでなければ、現在、正ボリュームに記録されているデータを仮想ボリュームのデータとして提供する。

10

【 0 0 4 5 】

さらに、上述の構成では、スナップショット機能により作成された仮想ボリュームの実際のデータは、スナップショットの正ボリュームと差分データを格納するプールボリュームに存在する。そのため、仮想ボリュームからのリモートコピーを実現する場合、リモートコピーの実行時に、正ボリュームとプールボリュームを選択してコピーを行う必要がある。また、ホストインターフェースと同じ領域をコピーする場合は、ホストインターフェースと同時に動作しないようにする必要がある。

20

【 0 0 4 6 】

そこでデータ転送の処理は、図 5 に示すような手順で実行される。すなわち図 5 において、カーネル側からリモートコピーの管理部へ、データ転送制御が指示されると、リモートコピーデータ転送ジョブが生成される。このリモートコピーデータ転送ジョブでは、最初に手順 (1) としてスナップショット用資源の確保が行われ、手順 (2) としてスナップショットのリードボリュームの判定 (正ボリュームか仮想ボリュームか) が行われる。

【 0 0 4 7 】

さらに手順 (3) として、ヒットかミスかの判定が行われ、ここでミスありのときは、手順 (4) でステージングジョブ生成の指示がカーネル側に行われる。これによりカーネル側の機能として、ステージングジョブが実行される。

30

【 0 0 4 8 】

また、手順 (3) で全てヒットのときは、手順 (5) でリモートコピー用資源が確保され、手順 (6) でリモートコピーコマンドがホスト制御によって発行される。これによって、副ボリューム側に対するリモートコピーの書き込みが行われる。

【 0 0 4 9 】

そして書き込みが終了すると、手順 (7) でリモートコピー用およびスナップショット用の資源が解放され、手順 (8) で結果報告が行われる。以後、スリープ状態とされ、次のジョブ開始時に手順 (1) から繰り返される。

【 0 0 5 0 】

上記の手順により、リモートコピーの差分データから次にコピーされる仮想ボリュームのアドレスが決定され、そのアドレスとスナップショット機能の差分データから、仮想ボリュームのデータが正ボリュームとプールボリュームのどちらにあるか判定する。そして、正ボリューム側にデータがある場合は、コピーの高速化のため大きなサイズ (ストライプ列単位) でコピーを行い、データの先読みも起動して、ディスクのインターフェース待ちにならずにコピーを行えるようにすることができる。

40

【 0 0 5 1 】

また、データがプールボリューム側にある場合は、スナップショット機能の差分の管理単位が 6 4 K B なので、6 4 K B 単位にディスクから読み出しを行い、バックアップコピー側へコピーを行う。なお、プールボリュームの場合は、通常連続してデータが配置されないため、先読みは行わない。また、スナップショットの正ボリュームからコピーの場合

50

に、正ボリュームへのホストインターフェースと同じ領域にアクセスする場合があるため、コピー範囲に対してロックを行い、ホストインターフェースとの排他を行う。

【 0 0 5 2 】

このようにして、スナップショット機能を有しているストレージ装置において、リモートコピーを作成する場合に、全体として効率の良いリモートコピーの作成を行うことができる。

【 0 0 5 3 】

また、図 6 には、書き込み処理のシーケンスを示す。図 6 において、最初に書き込みコマンドジョブが生成され、ここでは、対象ボリュームのペア状態のチェックが行われて、コピージョブが生成される。

10

【 0 0 5 4 】

そしてコピージョブでは、プール資源確保のために、正ボリュームの書き込みデータの可否が判定される。また、スナップショット用のセグメントが確保され、ステージングが行われる。なおこの際に、前に作成されたリモートコピーとの差分の生じた領域を示す第 1 のビットマップと、差分データの領域を示す第 2 のビットマップの何れか一方が作成される。

【 0 0 5 5 】

さらにステージングジョブでは、対象範囲のステージングが行われ、終了の報告が行われる。そして再びコピージョブで、スナップショット差分ビットマップ 2 0 2 のコピー済みのビットがオンにされて、コピージョブの終了が報告される。さらに再び書き込みコマンドジョブで、ステータスが送信され、スナップショット用の資源が解放される。

20

【 0 0 5 6 】

そしてこの装置において、スナップショット機能とリモートコピー機能とでそれぞれ差分管理を行う必要があるが、それぞれの差分管理を連携させる必要がある。また、スナップショット機能では、スプリット状態からペア状態への遷移を瞬時に終了させる必要があるが、スプリット状態でセットされたスナップショット機能の差分データをリモートコピーへの差分データへ瞬時に適用することは管理上難しい。

【 0 0 5 7 】

すなわち、上述の装置においては、リモートコピー機能では差分管理用データの容量削減のため、1 M B 単位で差分を管理しているが、スナップショット機能では差分の管理単位がホスト性能に直結するため 6 4 K B 単位で差分を管理している。これはスナップショット機能では、ホストインターフェースがたとえ 1 K B のコマンドでもホストの I / O ごとに差分管理の単位のコピーが発生するためである。

30

【 0 0 5 8 】

また、スナップショット機能では、正ボリュームと仮想ボリュームの関係は 1 : n の関係のため、差分管理が複雑になっている。そのため、上記の差分サイズを意識して差分ビットをセットする必要がある。さらに、スナップショット機能の差分データのリモートコピーの差分データへの適用時には、スプリット状態からペア状態への遷移時に瞬時に処理を終わらす必要があるが、上記のようにスナップショット側の差分データの形式が異なるため、リモートコピー側の差分データへそのまま適用できない。

40

【 0 0 5 9 】

そこで上述の装置では、リモートコピー用に別途新たな差分データ（カスケード差分：元々あるリモートコピー用差分データと同形式）を作成し、スナップショットの差分設定時に、同時にリモートコピーのカスケード差分もセットする。これにより、スプリット状態からペア状態への遷移時に、リモートコピーのカスケード差分をリモートコピーの差分データに加算することにより、瞬時に処理を終了させることができる。

【 0 0 6 0 】

すなわち図 7 において、図 7 の A に示すように、差分データの 6 4 K B ごとに作成されているスナップショット用のデータから、1 M B ごとのリモートコピー用の差分データの第 1 のビットマップに変換するのには、その変換処理時間が大きく必要とされる。そこで

50

、このようなデータの変換は、例えば仮想論理ボリュームの更新に合わせて行っておく。

【 0 0 6 1 】

また、スナップショット機能のスプリット時には、図 7 の B に示すように、差分データの 6 4 K B ごとに作成されているスナップショット用のデータの作成と同時に、新規差分データのカスケード差分データの第 2 のビットマップを作成する。このカスケード差分データの第 2 のビットマップは、上述の 1 M B ごとのリモートコピー用の差分データの第 1 のビットマップと同形式で作成される。

【 0 0 6 2 】

そして、図 7 の C に示すように、スナップショット機能のスプリット状態からペア状態への遷移時には、上述のカスケード差分データの第 2 のビットマップを、リモートコピー用の差分データの第 1 のビットマップに加算（論理和）して、この加算されたビットマップに基づいて、リモートコピーの作成を行う。

10

【 0 0 6 3 】

さらに、図 8 を用いて、上述の図 3 の一つのクイックシャドーから他のストレージ装置へのリモートコピーに関し、本発明の実際的な動作について説明する。図 8 において、正ボリューム 1 0 0 0 への書き込みが行われると、書き込み以前のデータがプールボリューム 2 0 0 0 に転記されると共に、その情報がスナップショットデータの記憶部 2 0 0 1 に記憶される。それと同時に、その差分の生じた領域のビットマップがカスケード差分の第 2 のビットマップ 2 0 0 2 として作成される。

【 0 0 6 4 】

20

このスプリット状態から、正ボリューム 1 0 0 0 と仮想ボリューム 3 0 0 0 がペア状態にされると、仮想ボリュームに 3 0 0 0 に、正ボリューム 1 0 0 0 とプールボリューム 2 0 0 0 の対応関係の情報が記憶されると共に、その差分の生じた領域のビットマップが差分データの第 1 のビットマップ 2 0 0 3 として加算される。なおこの加算の後、スナップショットデータの記憶部 2 0 0 1 とカスケード差分の第 2 のビットマップ 2 0 0 2 の情報は消去される。

【 0 0 6 5 】

そして、リモートコピーの作成時には、加算された第 1 のビットマップ 2 0 0 3 の情報に基づいて、リモートコピーボリューム 4 0 0 0 へのデータのリモートコピーが行われる。なおこのコピーが行われた後、第 1 のビットマップ 2 0 0 3 の情報は消去される。

30

【 0 0 6 6 】

したがってこの実施形態において、前に作成されたりモートコピーとの差分の生じた領域を示す第 1 のビットマップと、スナップショット機能の差分データの領域を示す第 2 のビットマップを作成することによって、スナップショット機能を有しているストレージ装置において、リモートコピーを作成する場合に、全体として効率の良いリモートコピーの作成を行うことができる。

【 0 0 6 7 】

これによって、従来の装置では、スナップショット機能を有しているストレージ装置において、リモートコピーを作成する場合に、全体として効率の良いリモートコピーの作成を行うことができなかったものを、本発明によれば、第 1 のビットマップおよび第 2 のビットマップを作成することで、これらを用いて、効率の良いリモートコピーの作成を行うことができ、したがって従来の問題点を解消することができるものである。

40

【 0 0 6 8 】

なお、上述の実施形態において、リモートコピーの更新を行う際には、第 2 のビットマップによって示される領域についてメモリに保存されたデータも用いて更新を行うようにする。これによって、いわゆるストレージ装置において、キャッシュメモリを設け、このキャッシュメモリに書き込まれた段階でホストコンピュータに書き込み終了を報告しているような装置においても、本発明を適用することができる。

【 0 0 6 9 】

また、上述の実施形態において、第 2 のビットマップは、リモートコピーを作成する体

50

制のときにのみ設けるものとする。これによって、例えば複数の仮想ボリュームを設けて、特定の期間ごとにそれらの仮想ボリュームを順番に用いているようなシステムにおいては、仮想ボリュームの更新を行わない期間は、差分データのビットマップを作らないことで、システム全体の記憶容量の削減を図ることができる。

【0070】

さらに本発明は、上述の説明した実施形態に限定されるものではなく、本発明の精神を逸脱することなく種々の変形が可能とされるものである。

【図面の簡単な説明】

【0071】

【図1】本発明の実施の形態のディスクアレイ装置のブロック図である。

10

【図2】本発明の実施の形態の制御プログラムの説明図である。

【図3】本発明の実施の形態のリモートコピーの説明図である。

【図4】本発明の実施の形態のプールボリュームの管理方法の説明図である。

【図5】本発明の実施の形態のデータ転送の処理の手順を示すシーケンス図である。

【図6】本発明の実施の形態の書き込み処理のシーケンス図である。

【図7】本発明の実施の形態のビットマップの加算処理の説明図である。

【図8】本発明の実施の形態の書き込み処理の説明図である。

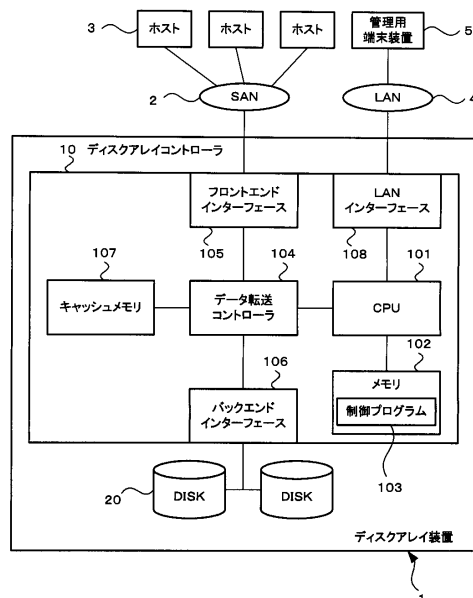
【符号の説明】

【0072】

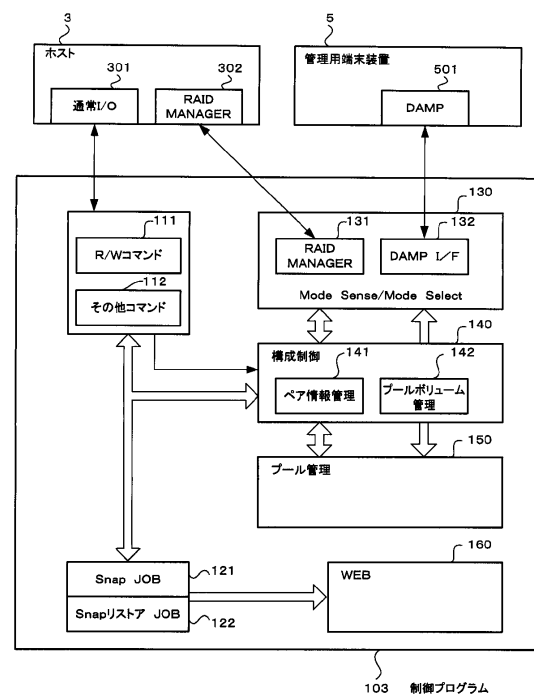
1 ... ディスクアレイ装置、2 ... SAN、3 ... ホスト、4 ... LAN、5 ... 管理用端末装置、10 ... ディスクアレイコントローラ、20 ... ディスク、101 ... CPU、102 ... メモリ、103 ... 制御プログラム、104 ... データ転送コントローラ、105 ... フロントエンドインターフェース、106 ... バックエンドインターフェース、107 ... キャッシュメモリ、108 ... LANインターフェース、109 ... RAIDコントローラ、110 ... RAID MANAGER、111 ... R/Wコマンド、112 ... その他コマンド、121 ... Snap JOB、122 ... Snapリスト JOB、130 ... Mode Sense/Mode Select、131 ... RAID MANAGER、132 ... DAMP L/F、140 ... 構成制御、141 ... ペア情報管理、142 ... プールボリューム管理、150 ... プール管理、160 ... WEB

20

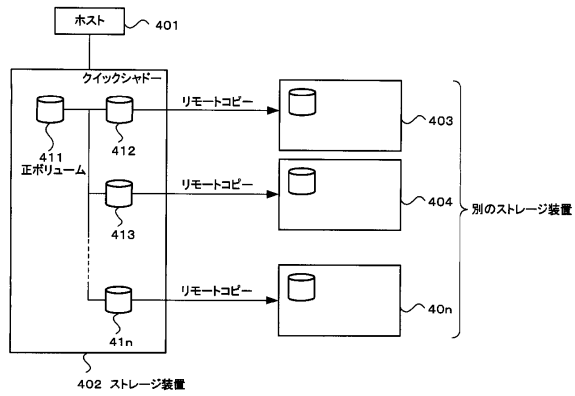
【図1】



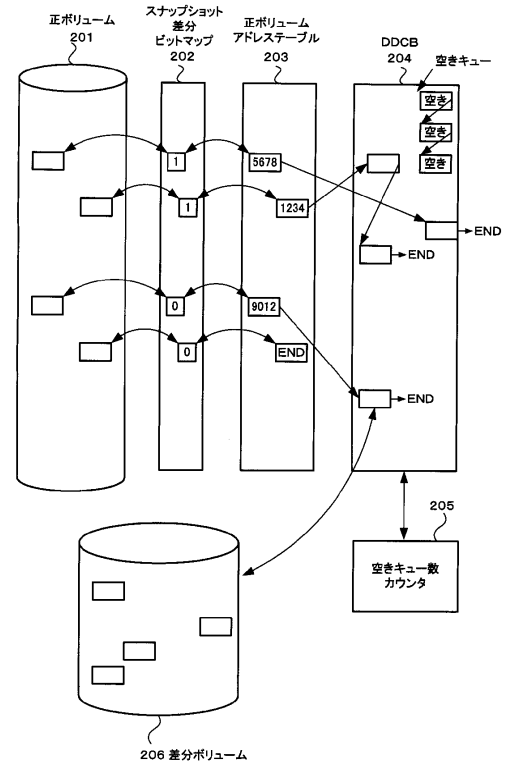
【図2】



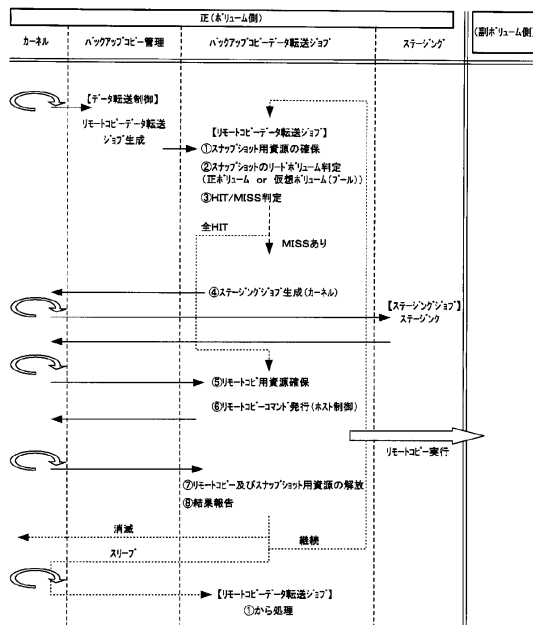
【図 3】



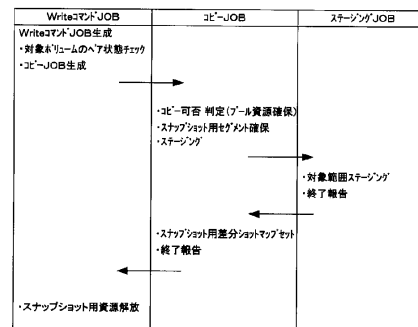
【図 4】



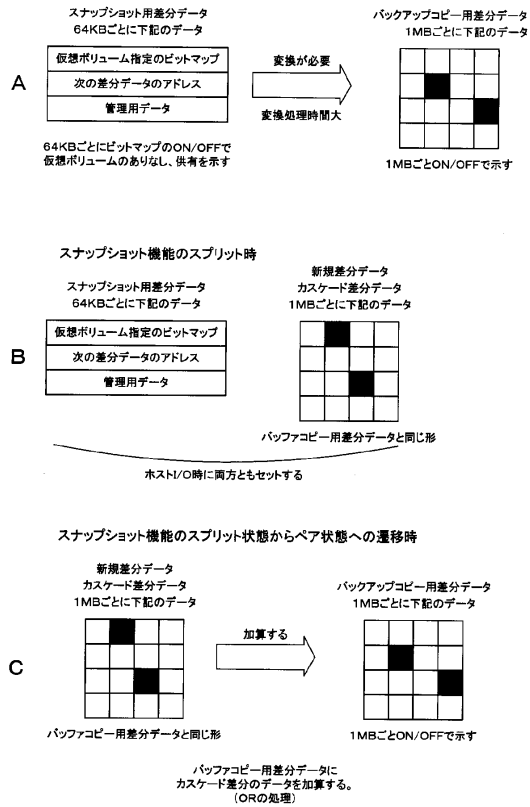
【図 5】



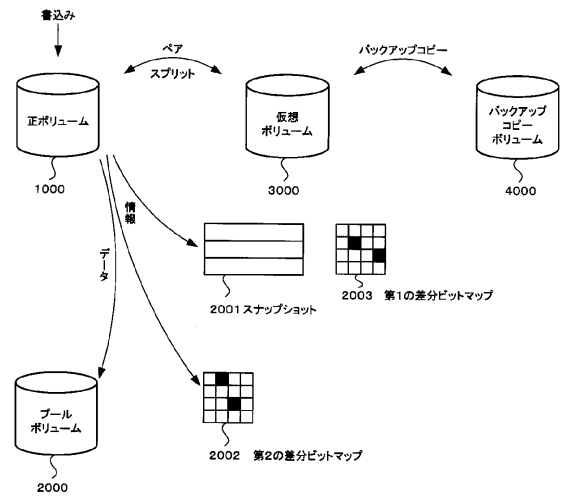
【図 6】



【図 7】



【図 8】



フロントページの続き

(56)参考文献 特開2002-278819(JP,A)
特開2003-242011(JP,A)
特開2001-051882(JP,A)

(58)調査した分野(Int.Cl., DB名)
G06F 12/00
G06F 3/06