

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4681249号
(P4681249)

(45) 発行日 平成23年5月11日(2011.5.11)

(24) 登録日 平成23年2月10日(2011.2.10)

(51) Int. Cl. F I
G 0 6 F 3 / 0 6 (2006.01)
 G 0 6 F 3 / 0 6 3 0 4 F
 G 0 6 F 3 / 0 6 3 0 1 J
 G 0 6 F 3 / 0 6 5 4 0

請求項の数 6 (全 19 頁)

(21) 出願番号 特願2004-116115 (P2004-116115)
 (22) 出願日 平成16年4月9日(2004.4.9)
 (65) 公開番号 特開2005-301628 (P2005-301628A)
 (43) 公開日 平成17年10月27日(2005.10.27)
 審査請求日 平成19年3月2日(2007.3.2)

(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 100075513
 弁理士 後藤 政喜
 (74) 代理人 100084537
 弁理士 松田 嘉夫
 (74) 代理人 100114236
 弁理士 藤井 正弘
 (72) 発明者 本間 裕季子
 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

最終頁に続く

(54) 【発明の名称】 ディスクアレイ装置

(57) 【特許請求の範囲】

【請求項1】

上位装置に接続され、前記上位装置からデータを受ける上位インタフェースと、
 データが格納される複数のディスク装置と、
 前記上位装置との間でやり取りされるデータの前記ディスク装置への読み書きを制御する複数のディスクインタフェースと、

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格納された過去のデータを世代毎の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2の論理ボリュームに格納されている前記世代毎の差分データの関係を管理する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ことによって前記差分データを管理する制御プロセッサと、を有するディスクアレイ装置であって、

前記制御プロセッサは、

前記第2の論理ボリュームに格納されたデータ量を管理し、

前記第2の論理ボリュームに格納されたデータ量が前記第2の論理ボリュームの容量の第1の割合を超えた場合、前記第2の論理ボリュームの容量を増加し、

前記第2の論理ボリュームに格納されたデータ量が、前記第1の割合より大きい第2の割合を超えた場合、前記第1の論理ボリュームに対するデータの書き込みを制限し、

前記第2の論理ボリュームに格納されたデータ量が、前記第2の割合より大きい第3の割合を超えた場合、前記第2の論理ボリュームに書き込まれた差分データのうち、特定の

世代の差分データを削除し、

前記第2の論理ボリュームに格納されたデータ量が、前記第3の割合より大きい第4の割合を超えた場合、前記第2の論理ボリュームに書き込まれた差分データのうち、前記特定の第1の論理ボリュームに対応するものを削除すること特徴とするディスクアレイ装置

【請求項2】

請求項1に記載のディスクアレイ装置であって、

前記制御プロセッサは、前記第2の論理ボリュームに格納されたデータ量が、前記第2の論理ボリュームの容量に達した場合、前記第2の論理ボリュームに書き込まれた全ての差分データを削除することを特徴とするディスクアレイ装置。

10

【請求項3】

請求項1に記載のディスクアレイ装置であって、

前記上位インタフェースに接続され、前記上位装置との間でやり取りされるデータ、及び前記上位装置との間でやり取りされるデータに関する制御情報を保存するメモリを有し

前記ディスクインタフェースは、前記上位装置との間でやり取りされ、前記メモリに格納されるデータを、前記ディスク装置に書き込むように制御することを特徴とするディスクアレイ装置。

【請求項4】

ディスクアレイ装置においてボリュームを管理する方法であって、

前記ディスクアレイ装置は、

上位装置に接続され、前記上位装置からデータを受ける上位インタフェースと、

データが格納される複数のディスク装置と、

前記上位装置との間でやり取りされるデータの前記ディスク装置への読み書きを制御する複数のディスクインタフェースと、

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格納された過去のデータを世代毎の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2の論理ボリュームに格納されている前記世代毎の差分データの間接関係を管理する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ことによって前記差分データを管理する制御プロセッサと、を有し、

20

前記方法は、

前記第2の論理ボリュームに格納されたデータ量を管理し、

前記第2の論理ボリュームに格納されたデータ量が前記第2の論理ボリュームの容量の第1の割合を超えた場合、前記第2の論理ボリュームの容量を増加し、

前記第2の論理ボリュームに格納されたデータ量が、前記第1の割合より大きい第2の割合を超えた場合、前記第1の論理ボリュームに対するデータの書き込みを制限し、

前記第2の論理ボリュームに格納されたデータ量が、前記第2の割合より大きい第3の割合を超えた場合、前記第2の論理ボリュームに書き込まれた差分データのうち、特定の世代の差分データを削除し、

30

40

前記第2の論理ボリュームに格納されたデータ量が、前記第3の割合より大きい第4の割合を超えた場合、前記第2の論理ボリュームに書き込まれた差分データのうち、前記特定の第1の論理ボリュームに対応するものを削除すること特徴とするボリューム管理方法

【請求項5】

請求項4に記載のボリューム管理方法であって、

前記第2の論理ボリュームに格納されたデータ量が、前記第2の論理ボリュームの容量に達した場合、前記第2の論理ボリュームに書き込まれた全ての差分データを削除することを特徴とするボリューム管理方法。

【請求項6】

50

請求項 4 に記載のボリューム管理方法であって、

前記ディスクアレイ装置は、前記上位インタフェースに接続され、前記上位装置との間でやり取りされるデータ、及び前記上位装置との間でやり取りされるデータに関する制御情報を保存するメモリを有し、

前記方法は、前記上位装置との間でやり取りされ、前記メモリに格納されるデータを、前記ディスク装置に書き込むように制御することを特徴とするボリューム管理方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ディスクアレイ装置に関し、特にスナップショットにおける差分LUの管理技術に関する。 10

【背景技術】

【0002】

コンピュータシステムにおいては、ストレージ装置に記憶されるデータをバックアップする場合など、元のデータを記憶した記憶ボリュームの複製を生成する場合がある、この場合、元のデータを記憶した記憶ボリュームと複製データを記憶した記憶ボリュームとは内容が一致し整合性が保たれる必要がある。そのためには複製の生成が完了するまでの間、元のデータを記憶した記憶ボリュームの更新を停止させる必要がある。しかし高可用性が求められるコンピュータシステムなどでは、元のデータを記憶した記憶ボリュームに対するアクセスを停止することができない場合がある。 20

【0003】

そこで、元のデータを記憶した記憶ボリュームと複製データを記憶した記憶ボリュームとの間である時点での整合性を保ち、ある時点の後に元のデータが更新されたとしてもある時点の元のデータを参照できるようにするスナップショットと呼ばれる技術が開発されている。

【0004】

スナップショットの技術によれば、整合性を保つべき時点の後に元のデータが更新される場合には、整合性を保つべき時点のデータを別の記憶ボリュームに記憶するようにする。つまり元のデータは、更新されなければそのままであり、更新される場合には整合性を保つべき時点のデータが別の記憶ボリュームに記憶される（例えば、特許文献1参照。） 30

【0005】

また、スナップショットを複数世代において利用可能にする方法が提案されている。

【0006】

これによれば、データ保持部1は、通常読み書きを行う。データ保持部2は、ある時点でのデータ保持部1のスナップショットイメージを保存する。データ保持部3は、データ保持部2にスナップショットイメージを保存した時点以降のデータ保持部1へのデータ書替えに伴う更新データと、その世代を示す情報と、その更新領域を示す情報とを含む履歴情報を保存する。データ読出し元選択部4は、読出されるべきスナップショットイメージの世代及び領域の指定に応じて、データ保持部3に保存される各履歴情報を参照し、読出されるべきスナップショットイメージの保存場所を知得し、その保存場所に応じてデータの読出し元をデータ保持部2及びデータ保持部3のいずれかに切替える（例えば、特許文献2参照。） 40

【特許文献1】特開2001-306407号公報

【特許文献2】特開2002-278819号公報

【発明の開示】

【発明が解決しようとする課題】

【0007】

しかし、スナップショット機能は、正LUの容量よりも差分LUの容量を減らせることが利点であるが、ホストからのデータの書き込みが多くなると差分データが多くなる。そ 50

して、プール領域（差分LU）の使用量が100%になるとスナップショットを維持できなくなる。このため、プール領域の使用量が閾値を超えたときに、管理者に警告を発する。この警告を受け取った管理者は、データを別ディスクへ退避して、プール領域の使用量を減らす、又はV-VOLを削除する等によって、運用で処理されていた。そのため、管理者の負担が大きく、プール領域の使用量が100%になることの回避に失敗するリスクを抱えていた。

【0008】

本発明は、プール領域の使用量が100%になることを回避して、全スナップショットイメージの破壊を防止することを目的とする。

【課題を解決するための手段】

【0009】

本発明は、上位装置に接続され、前記上位装置からデータを受ける上位インタフェースと、データが格納される複数のディスク装置と、前記上位装置との間でやり取りされるデータの前記ディスク装置への読み書きを制御する複数のディスクインタフェースと、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格納された過去のデータを世代毎の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2の論理ボリュームに格納されている前記世代毎の差分データの間隔を管理する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ことによって前記差分データを管理する制御プロセッサと、を有するディスクアレイ装置であって、前記制御プロセッサは、前記第2の論理ボリュームに格納されたデータ量を管理し、前記第2の論理ボリュームに格納されたデータ量が前記第2の論理ボリュームの容量の第1の割合を超えた場合、前記第2の論理ボリュームの容量を増加し、前記第2の論理ボリュームに格納されたデータ量が、前記第1の割合より大きい第2の割合を超えた場合、前記第1の論理ボリュームに対するデータの書き込みを制限し、前記第2の論理ボリュームに格納されたデータ量が、前記第2の割合より大きい第3の割合を超えた場合、前記第2の論理ボリュームに書き込まれた差分データのうち、特定の世代の差分データを削除し、前記第2の論理ボリュームに格納されたデータ量が、前記第3の割合より大きい第4の割合を超えた場合、前記第2の論理ボリュームに書き込まれた差分データのうち、前記特定の第1の論理ボリュームに対応するものを削除すること特徴とする。

【発明の効果】

【0011】

本発明によると、プール領域の使用量が100%になることを回避することができ、全スナップショットイメージの破壊を防止することができる。

【発明を実施するための最良の形態】

【0012】

以下、本発明の実施の形態を図面を参照して説明する。

【0013】

図1は、本発明の実施の形態のディスクアレイ装置の構成を示すブロック図である。

【0014】

本発明の実施の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及びディスク20を含んで構成される。また、ディスクアレイ装置1は、SAN2を介して複数のホスト3に接続され、LAN4を介して管理用端末装置5に接続されている。

【0015】

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RAID（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、記憶されたデータが消失しないようになっている。

【0016】

ディスクアレイコントローラ10には、CPU101、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LANインターフェース108が設けられている。

【0017】

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、CPU101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。

【0018】

データ転送コントローラ104は、CPU101、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。

10

【0019】

フロントエンドインターフェース105は、SAN2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。

【0020】

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。

【0021】

20

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。

【0022】

すなわち、データ転送コントローラ104は、SAN4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。

【0023】

LANインターフェース108は、LAN4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。

30

【0024】

SAN2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。

【0025】

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。

【0026】

LAN4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、以下同じ)が用いられる。

40

【0027】

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。なお、管理用端末装置5ではweb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。

50

【 0 0 2 8 】

図 2 は、本発明の実施の形態の制御プログラム 1 0 3 の説明図である。

【 0 0 2 9 】

ホスト 3 の通常 I / O 処理プログラム 3 0 1 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 1 0 3 の R / W コマンド解析プログラム 1 1 1 によって解析され、スナップジョブプログラム 1 2 1 に送られる。

【 0 0 3 0 】

スナップジョブプログラム 1 2 1 は、正 L U に対するデータ書き込み要求を受信すると、正 L U 内の更新前のデータを差分 L U に複写し、この複写後に正 L U の内容を更新する。更に、スナップジョブプログラム 1 2 1 は、データが更新された正 L U 内のブロックに対応する仮想 L U 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 L U のデータ（すなわち更新前のデータ）が格納されている差分 L U 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（差分情報管理ブロック 2 0 4 ）を更新する。

10

【 0 0 3 1 】

また、スナプリストアジョブプログラム 1 2 2 は、仮想 L U に対するアクセス要求を受信すると、スナップショット管理テーブルを参照して、仮想 L U のブロックと対応付けられている正 L U のブロック又は差分 L U のブロックにアクセスする。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 3 0 1 によって、仮想 L U 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 L U 内の情報を利用することができる。

20

【 0 0 3 2 】

また、通常 I / O 処理プログラム 3 0 1 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 1 1 2 によって解析され、構成情報制御プログラム 1 4 0 に送られる。

【 0 0 3 3 】

構成情報制御プログラム 1 4 0 のペア情報管理プログラム 1 4 4 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 L U の識別情報を登録し、ペア情報管理テーブルにスナップショットのペアを新規に登録し、差分ビットマップ 2 0 2 、正 L U アドレステーブル 2 0 3 、差分情報管理ブロック 2 0 4 を確保、初期化する。この仮想 L U 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、正 L U のブロックと一対一で対応付けられている。

30

【 0 0 3 4 】

プール L U 管理プログラム 1 4 2 は、後述するように、プール領域に登録された L U の追加及び削除を管理する。

【 0 0 3 5 】

プール管理プログラム 1 5 0 は、スナップショット差分管理テーブル（差分情報管理ブロック 2 0 4 ）の領域の確保、解放及び空きキューへの遷移等の管理を行う。

【 0 0 3 6 】

WEB プログラム 1 6 0 は、スナップショットで作成される各ペアの状態（ペア情報管理テーブル内の情報等からペアの障害有無）を WEB ブラウザに提供する。

40

【 0 0 3 7 】

ディスクアレイ装置 1 の制御プログラム 1 0 3 の R A I D マネージャプログラム 1 3 1 は、ホスト 3 の R A I D マネージャプログラム 3 0 2 と通信可能に接続されている。この R A I D マネージャプログラム 1 2 1 、 3 0 2 によって、スナップショットの生成やペア状態の変更などを行うことができる。

【 0 0 3 8 】

また、DAMP インターフェースプログラム 1 2 2 は、仮想 L U の削除処理等のディスクアレイの各種設定を行うユーザインターフェースを提供する。DAMP インターフェースプログラム 1 2 2 は、管理用端末装置 5 の DAMP プログラム 5 0 1 と通信可能に接続されている。この DAMP インターフェースプログラム 1 2 2 によって、管理用端末装置

50

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の構成の管理、プールへの自動追加、削除の設定が行われる。

【0039】

図3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。

【0040】

正LU201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。

【0041】

差分ビットマップ202は、キャッシュメモリ107の管理領域に設けられており、正LU201のブロック(例えば、64kバイト/ブロック)に対応するビットを有する。この差分ビットマップ202は、後述する正LUアドレステーブル203及び差分情報管理ブロック204と共にスナップショット管理テーブルを構成する。

10

【0042】

正LU201のブロックアドレスに対応する差分データが差分LU206に記録されている場合に、差分ビットマップ202中の正LU201に対応するビットが「1」となっている。従って、正LU201に対する書き込みが行われるときに、差分ビットマップ202を参照することによって、更新前データを差分LU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差分LU206に書き込まれているので、正LU201のデータを差分LU206にコピーする必要がない)。

20

【0043】

正LUアドレステーブル203は、キャッシュメモリ107の管理領域に設けられており、差分ビットマップ202のビットと対応して差分情報管理ブロック204のアドレスが記録されている。

【0044】

差分情報管理ブロック204は、差分LUと同じ容量を有し、キャッシュメモリ107の管理領域に設けられている。差分情報管理ブロック204は、差分LU206のブロック(例えば、64k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、差分LU206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ている。また、正LU201の当該ブロックに対応する他の差分データがあれば、その差分情報管理ブロック204上のアドレスへのリンク情報が記録されている。すなわち、差分情報管理ブロック204に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。

30

【0045】

なお、差分情報管理ブロック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。

【0046】

差分LU206は、プール領域に登録されたLUによって構成される。この差分LU206は、スナップショット作成時点における正LU201のデータが複写されている。そして、差分LU206のデータがどの世代の差分データかは差分情報管理ブロック204の世代管理ビットマップによって知ることができる。

40

【0047】

よって、正LUにデータを書き込むときは、まず、差分ビットマップ202を参照して、更新前データを差分LUにコピーする必要があるかを判定する。そして、差分ビットマップ202の対応ビットが「1」であれば、更新前データを差分LUにコピーする必要がないと判定し、正LUにデータを書き込む。一方、差分ビットマップ202の対応ビットが「0」であれば、更新前データを差分LUにコピーした後に、正LUにデータを書き込む。

50

【 0 0 4 8 】

そして、正LUのブロックに対応する差分情報管理ブロック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、必要に応じて正LUアドレステーブルに、差分情報管理ブロック204のブロックのアドレスを設定する。そして、差分ビットマップ202の対応ビットを「1」にし、差分LU206の当該更新前データを書き込んだアドレスに対応する差分情報管理ブロック204の世代管理ビットマップを設定する。さらに、差分情報管理ブロック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。

【 0 0 4 9 】

また、仮想LU(V-VOL:Virtual Volume)にアクセスするときには、正LUアドレステーブル203を参照して、アクセス対象となる仮想LUのブロックアドレス(正LUのブロックアドレスに等しい)によって差分情報管理ブロック204のアドレスを特定して、差分情報管理ブロック204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。そして、所望の世代の差分データがあれば、当該差分情報管理ブロック204のアドレスに対応する差分LU206のアドレスから差分データを読み出して、仮想LU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LUに記録されているデータを仮想LUのデータとして提供する。

【 0 0 5 0 】

図4は、本発明の実施の形態の空き容量チェック処理のフローチャートであり、正LUへのデータの書き込み処理等によって、正LUから差分LUに差分データをコピーするときに、プールLU管理プログラム132によって実行される。

【 0 0 5 1 】

まず、空きキューカウンタ205を参照して、プール(差分LU)の空き容量(既に使われている容量でもよい)を確認する(S101)。

【 0 0 5 2 】

次に、ステップS101で取得したプール領域の使用量と予め定めたLU追加閾値(70%)とを比較し、プール領域の使用量がLU追加閾値を超えているか否かを判定する(S102)。そして、プール領域の使用量がLU追加閾値(70%)を超えていれば、LU追加処理(図5)を実行する(S103)。

【 0 0 5 3 】

このLU追加閾値は、プール領域の使用量が100%に到達しないように余裕を持たせて70%程度とすることが望ましいが、ストレージ装置の稼働状態(例えば、データ入出力要求の量)によってLU追加閾値を変更することができる。

【 0 0 5 4 】

また、LUのプール領域への追加処理の実行中は通常時の値より高い閾値を設定することによって(例えば、通常時の70%に対して75%)、プール領域の使用量がLU追加閾値を超えたことによってLU追加処理が実行されているにもかかわらず、新たにLU追加処理の実行を防止することもできる。

【 0 0 5 5 】

次に、プール領域の使用量がI/O制限閾値(90%)を超えているか否かを判定する(S104)。その結果、プール領域の使用量がI/O制限閾値(90%)を超えていれば、ホスト3からのデータの書き込みを制限する(S105)。このI/O制限では、手動による仮想LUの削除処理を優先して、データ書き込み要求と差分データの削除処理とが1対1程度となるようにして、プール領域の使用量がこれ以上高くないようにする。

【 0 0 5 6 】

次に、プール領域の使用量が仮想LU解除閾値(95%)を超えているか否かを判定する(S106)。そして、プール領域の使用量が仮想LU解除閾値(95%)を超えてい

10

20

30

40

50

れば、仮想LU削除処理(図7)を実行する(S107)。この仮想LU削除処理によって運用の性能劣化を避けつつ、少量の空き容量を作ることができる。

【0057】

この仮想LU解除閾値は、プール領域の使用量が100%に到達する前に仮想LU削除処理を完了する必要があるため、若干の余裕を持たせて95%程度とすることが望ましいが、ストレージ装置の稼働状態(例えば、データ入出力要求の量)によって仮想LU解除閾値を変更することができる。

【0058】

また、1回目の判定では仮想LU解除閾値を95%とし、仮想LU削除処理中の2回目以後の判定では仮想LU解除閾値を高く設定する(例えば、96%とする)ことによって、プール領域の使用量が仮想LU解除閾値を超えたことによって仮想LU削除処理が実行されているにもかかわらず、新たに仮想LU削除処理が実行されないようにすることもできる。

10

【0059】

なお、プール領域の全容量が200GB以上の場合には、プール領域の使用量の割合ではなく、プール領域の残容量が10GB以下となったときに仮想LU削除処理を実行してもよい。

【0060】

次に、プール領域の使用量が正LU解除閾値(99%)を超えているか否かを判定する(S108)。この正LU解除閾値によって、プール領域の使用量が100%に到達する可能性が極めて高く、緊急かつ確実に空き容量を確保する必要があることを判定する。そして、プール領域の使用量が正LU解除閾値(99%)を超えていれば、仮想LU削除処理(図11)を実行する(S107)。

20

【0061】

なお、プール領域の全容量が200GB以上の場合には、プール領域の使用量の割合ではなく、プール領域の残容量が2GB以下となったときに仮想LU削除処理を実行してもよい。

【0062】

このように、複数段階で緊急度に応じた仮想LUの削除を実行するので、プール領域の使用量を確実に100%にさせないよういに行うことができる。

30

【0063】

次に、プール領域の使用量が全仮想LU解除閾値(100%)に到達したか否かを判定する(S110)。そして、プール領域の使用量が100%に達していれば、ペア状態管理テーブルに記録されている当該プール領域を使用している仮想LUの状態を全てエラー状態(PSE)にして、当該プール領域を使用している全ての仮想LUを削除する。

【0064】

このように、本発明の実施の形態の空き容量チェック処理では、管理者が介入することなく、指定条件に従って自動的に仮想LUを削除することができ、プール領域の使用量が100%となることを防止できる。また、プール領域の使用量に応じて段階的に仮想LUを削除するので、運用上の性能劣化を最小限に抑えることができる。

40

【0065】

図5は、本発明の実施の形態のLU追加処理のフローチャートであり、空き容量チェック処理(図4のステップS103)から呼び出されて(すなわち、プール領域の使用量がLU追加閾値(70%)を超えたときに)、プールLU管理プログラム132によって実行される。

【0066】

まず、プール自動追加対象管理テーブル143(図6)を参照して、プール自動追加対象にLUが登録されているか否かを判定する(S121)。そして、プール自動追加対象にLUが登録されていない場合は、追加すべきLUが存在しないので、LUを追加することなくこの処理を終了する。一方、プール自動追加対象にLUが登録されていれば、ステッ

50

プ S 1 2 2 に進む。

【 0 0 6 7 】

ステップ S 1 2 2 では、プール管理テーブル 1 4 1 (図 6) を参照して、プール領域に登録されている L U が最大数に達しているか否かを判定する。このプール領域の登録数の最大値はプール管理テーブル 1 4 1 の容量によって定められる。そして、プール管理テーブル 1 4 1 に登録されている L U の数が既に最大数であれば、プール管理テーブル 1 4 1 に新たに L U を追加することができないので、L U を追加することなくこの処理を終了する。一方、プール管理テーブル 1 4 1 に登録されている L U の数が最大数に満たなければ、ステップ S 1 2 3 に進む。

【 0 0 6 8 】

ステップ S 1 2 3 では、キャッシュメモリ 1 0 7 の管理領域に空きがあるか否かを判定する。これは、プール領域に L U を追加することに伴い差分情報管理ブロック 2 0 4 の容量を増加することが必要なためである。そして、キャッシュメモリ 1 0 7 の管理領域に空きがなければ、差分情報管理ブロック 2 0 4 の領域を増やすことができないので、L U を追加することなくこの処理を終了する。一方、キャッシュメモリ 1 0 7 の管理領域に空きがあれば、ステップ S 1 2 4 に進む。

【 0 0 6 9 】

S 1 2 1 ~ S 1 2 3 の処理によって L U を追加する条件が確認されたら、移動する L U をプール自動追加対象管理テーブル 1 4 3 から削除し、プール管理テーブル 1 4 1 に追加して、差分 L U として使用可能にする (S 1 2 4) 。

【 0 0 7 0 】

この移動する L U として、複数の L U をプール管理テーブル 1 4 1 に事前に指定しておくことができる。このようにすることで、必要な分だけ L U を追加することが可能となる。例えば、1 0 0 G B の L U を 1 0 個登録しておくこと、プール領域の容量が 3 0 0 G B 不足した場合に 1 0 0 G B の L U を 3 個だけプール領域に追加して、あとの 7 個の L U は、別の用途に使用可能となる。

【 0 0 7 1 】

その後、従来から存在した差分情報管理ブロック 2 0 4 の最後の空きキューに、新たに増加した差分情報管理ブロック 2 0 4 の領域の空きキューのアドレスを設定し、新たに増加した空きキューを従来からの空きキューにつなげる (S 1 2 5) 。

【 0 0 7 2 】

図 6 は、本発明の実施の形態の L U 追加処理の説明図である。

【 0 0 7 3 】

正 L U (L U 0) に対して、スナップショット生成時のデータのイメージを提供する仮想 L U (L U 2) が設けられている。この L U 2 は、正 L U のデータと、プール領域にある差分データとによって構成される仮想的な論理ユニットである。すなわち、L U 2 に対するアクセスがあると、そのデータが正 L U にあるのか、プール領域にある差分データなのかを判断して、差分 L U に対するアクセスを実現している。

【 0 0 7 4 】

プール領域に追加される L U は、プール自動追加対象管理テーブル 1 4 3 に登録されているもの (L U 6 、 L U 7) を、プール自動追加対象管理テーブル 1 4 3 から削除し、プール管理テーブル 1 4 1 に追加する。これによって、追加された L U が差分 L U として使用可能となる。

【 0 0 7 5 】

このように本発明の実施の形態の L U 追加処理では、プール領域の容量を増加するための L U を事前に複数用意しておき、プール領域の使用量がある閾値を越えた場合に、自動的に L U をプール領域に追加し、プール領域の容量を増加して、プール領域の使用量が 1 0 0 % になるのを防ぐ。また、プール領域の使用量が少なく、L U を追加する必要がない場合は、L U を他の用途に使用することができる。よって、不要なディスクをプール領域に事前に登録することなく、ディスクの容量を有効に活用することができる。

10

20

30

40

50

【 0 0 7 6 】

図7は、本発明の実施の形態の仮想LU削除処理（仮想LU単位）のフローチャートであり、空き容量チェック処理（図4のステップS107）から呼び出されて（すなわち、プール領域の使用量が仮想LU解除閾値（95%）を超えたときに）、プールLU管理プログラム132によって実行される。

【 0 0 7 7 】

まず、管理者が設定した情報に基づいて、削除する仮想LUを決定する（S131）。この削除するLUの設定には様々な方法がある。一つは、最も古い世代の仮想LUのデータを削除する方法である。また、管理者が指定した世代の仮想LUのデータを削除することもできる。なお、管理者が削除するLUを設定する画面は後述する（図9、図10）。 10

【 0 0 7 8 】

また、最も古い仮想LUを選択するためには、全ての正LUのペア情報管理テーブル（図8）を参照して、仮想LUのペア状態とスナップショット作成日時を検索して、ペア状態がスプリット状態（PSUS）となっている仮想LUの中で、最も古い仮想LUを選択する。PAIR状態では仮想LUの実体がないため、実体があるスプリット状態の仮想LUを削除対象としている。

【 0 0 7 9 】

そして、仮想LU単位の削除処理を登録し、この処理をバックグラウンドで実行する（S132）。この削除処理は、決定された仮想LUのデータを削除するものである。差分データは、複数世代に共通して用いられているものと、その削除対象の世代のみに用いられているものがあり、まず、差分データがこのいずれかであるかを判定する必要がある。差分情報管理ブロック204の世代管理微とマップを確認した結果、差分情報管理ブロック204が該当世代のみに使用されていれば、差分情報管理ブロック204のデータを更新して空きキューに変更する。一方、差分情報管理ブロック204を複数世代で共有されていれば、差分情報管理ブロック204の共有情報（世代管理ビットマップ）を更新する。 20

【 0 0 8 0 】

そして、ペア情報管理テーブル（図8）に記録されている当該削除対象となる仮想LUのペア状態をエラー状態（PSUE）に変更してペアを解除し、この仮想LUを使用不可にする（S133）。 30

【 0 0 8 1 】

以上説明した仮想LU削除処理はプール領域の使用量が仮想LU解除閾値（95%）を超えたときに実行されるものであるが、削除対象の仮想LUを削除しても、プール領域の使用量がI/O制限閾値（90%）以下にならない場合には、当該仮想LUの削除を行わないようにしてもよい。

【 0 0 8 2 】

このように仮想LU削除処理（仮想LU単位）は、削除対象の仮想LU（例えば、最も古い仮想LU等）のペア状態を解除し、スナップショットの維持に大きな影響を与えることなく、空き容量を作ることができる。また、管理者が設定した情報に基づいて削除する仮想LUを決定するので、古いデータであっても残すべき仮想LUを残して、LUの空き容量を増やすことができる。また、仮想LU削除処理をバックグラウンドで実行するので、運用の性能劣化を避けつつ、データの空き容量を作ることができる。 40

【 0 0 8 3 】

図8は、本発明の実施の形態のペア情報管理テーブルの説明図である。

【 0 0 8 4 】

ペア情報管理テーブルは、正LU毎に設けられており、正LUと各世代の仮想LUのペア状態と、その仮想LUの作成日時を記録して、正LUと仮想LUとの関係を管理している。例えば、0番の正LUの13番の仮想LUは2003年11月24日22時30分30秒に作成され、現在もペア状態（PAIR）が維持されている。すなわち、13番の仮想LUは最新の仮想LUであり、現在の差分データの書き込みが行われている。 50

【 0 0 8 5 】

また、0番の正LUの0番の仮想LUは2003年10月8日8時30分30秒にスナップショット作成要求が受け付けられることによって作成され、現在仮想LUが作成されたスプリット状態（PSUS）にある。このスプリット状態の仮想LUは削除対象となる。また、0番の正LUの9番の仮想LUは2003年9月20日10時10分10秒に作成されたが、既に削除された仮想LUであり、削除対象に登録されたエラー状態（PSE）となっている。

【 0 0 8 6 】

ペア情報管理テーブルによって、削除対象のLUがチェック、削除登録がされ、削除する仮想LUが決定される。なお、削除対象LUの決定後、実際に差分データの削除はプール管理プログラム150が行う。

10

【 0 0 8 7 】

図9は、本発明の実施の形態の仮想LU削除処理の設定画面の説明図である。

【 0 0 8 8 】

図9(a)は仮想LU削除情報設定のメイン画面であり、「削除可能」タブが選択された状態を示す。この画面では、削除優先順位について、仮想LUの作成日が「古い日付順から削除する。」のか、「指定した仮想LUを削除する。」のかを選択する。「指定した仮想LUを削除する。」が選択されると、「設定情報確認」ボタンがアクティブになる。そして「設定情報確認」ボタンを操作することによって、選択された正LUに関する削除優先順位設定サブ画面（図9(b)）に移行する。削除優先順位設定サブ画面では、削除チェック欄をチェックすることによって、仮想LU毎に削除の許可を設定することができる。また、削除チェック欄にチェックを入れることによって優先順位を選択することができる。なお、優先順位は「高」「中」「低」のランクではなく、個別に順序を設定することもできる。

20

【 0 0 8 9 】

図10は、本発明の実施の形態の仮想LU削除処理の設定画面の説明図である。

【 0 0 9 0 】

図10(a)は仮想LU削除情報設定のメイン画面であり、「削除禁止」タブが選択された状態を示す。この画面では、正LUを選択し、「設定情報確認」ボタンを操作することによって、選択された正LUに関する削除禁止設定サブ画面（図10(b)）に移行する。削除禁止設定サブ画面では、削除チェック欄をチェックすることによって、仮想LU毎に削除の禁止を設定することができる。

30

【 0 0 9 1 】

図11は、本発明の実施の形態の仮想LU削除処理（正LU単位）のフローチャートであり、空き容量チェック処理（図4のステップS109）から呼び出されて（すなわち、プール領域の使用量が正LU解除閾値（99%）を超えたときに）、プールLU管理プログラム132によって実行される。

【 0 0 9 2 】

まず、ペア情報管理テーブルを参照して仮想LUの削除中であるか否かを判定する。すなわち、ペア情報管理テーブルのペア状態欄が「PSUS」であれば仮想LUの削除中であると判定する（S141）。その結果、仮想LUの削除中であれば、仮想LUの削除を優先して速やかにプール領域の空き容量を確保するために、ステップS136に移行し、現在行われている削除処理を継続する。一方、仮想LUの削除中でなければ、新たに仮想LUの削除を実行する。

40

【 0 0 9 3 】

そこで、まず、仮想LUの削除に対応する正LUを決定する（S142）。この正LUは、管理者が予め定めたものを選択する他に、ホスト3からデータ書き込み要求を受けている正LUを選択することができる。現在データ書き込み要求を送っているホストは、プール領域の空き容量が減少している原因となっている可能性が高いので、データ書き込み要求に係る正LUを選択することによって、当該ホストからのアクセスを止めることができる

50

【 0 0 9 4 】

そして、全てのホスト 3 からデータ書き込みを停止する (S 1 4 3)。このデータ書き込みの停止によって、他の処理に優先して仮想 L U の削除処理を行うことができ、速やかに空き容量を確保することができる。

【 0 0 9 5 】

そして、選択された正 L U に関する仮想 L U の削除を実行する (S 1 4 4)。この削除処理は、選択された正 L U の正 L U アドレステーブル 2 0 3 を参照して、差分情報管理ブロック 2 0 4 に登録された内容を削除し、当該アドレスの差分情報管理ブロック 2 0 4 のデータを空きキューにすることによって、仮想 L U を削除し、プール領域の空き容量を確保する。この削除処理は、前述した仮想 L U 単位の削除処理 (図 7 の S 1 3 2) と比較して、差分情報管理ブロック 2 0 4 のデータ (世代管理ビットマップ) を参照して仮想 L U の世代を特定する必要がないので、迅速に仮想 L U を削除することができる。

【 0 0 9 6 】

そして、プール領域の空き容量が確保されたなら、ホスト 3 からの書き込み処理を再開する (S 1 4 5)。

【 0 0 9 7 】

このように仮想 L U 削除処理 (正 L U 単位) は、閾値を超えたときに削除対象のせい L U (例えば、データ入出力要求があった正 L U) に関する全ての仮想 L U のペア状態を P S U E にして、ペア状態を解除する。このペア解除処理は、バックグラウンドにて行うが、1 回の J O B で大量 (1 0 0 M B 程度) を削除できる。すなわち、選択された正 L U に関する全ての仮想 L U を削除するために全ホストからのデータ書き込みを一時的に停止して、正 L U 単位で全体 (又は、1 0 0 M B などの所定の容量ごと) にデータを削除して、一定のプール領域の空き容量を確保したうえで、ホストからのデータ書き込み要求を再開させる。よって、即座にかつ確実にプール領域の空き容量を増やすことができる。また、特定の正 L U に関して仮想 L U を削除するので、当該正 L U に関する特定のホストからのアクセスを止めることができる。

【 0 0 9 8 】

特許請求の範囲に記載した以外の本発明の観点の代表的なものとして、次のものがあげられる。

【 0 0 9 9 】

上位装置に接続され、前記上位装置からデータを受ける上位インタフェースと、
前記上位インタフェースに接続され、前記上位装置との間でやり取りされるデータ、及び前記上位装置との間でやり取りされるデータに関する制御情報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やり取りされるデータを、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フェースと、
前記複数のディスクインタフェースに接続され、前記複数のディスクインタ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格納された過去のデータを世代毎の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 2 の論理ボリュームに格納されている前記世代毎の差分データの関係を管理する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ことによって前記差分データを管理する制御プロセッサと、を有するディスクアレイ装置における制御方法であって、

前記制御プロセッサは、

前記第 2 の論理ボリュームに格納されたデータ量を管理し、

前記第 2 の論理ボリュームに格納されたデータ量が前記第 2 の論理ボリュームの容量の第 1 の割合を超えた場合、前記第 2 の論理ボリュームの容量を増加し、

前記第 2 の論理ボリュームに格納されたデータ量が前記第 2 の論理ボリュームの容量の

10

20

30

40

50

第2の割合を超えた場合、前記第1の論理ボリュームに対するデータの書き込みを制限し、

前記第2の論理ボリュームに格納されたデータ量が前記第2の論理ボリュームの容量の第3の割合を超えた場合、前記第2の論理ボリュームに書き込まれた差分データのうち、特定の世代の差分データを削除し、

前記第2の論理ボリュームに格納されたデータ量が前記第2の論理ボリュームの容量の第4の割合を超えた場合、前記第2の論理ボリュームに書き込まれた差分データのうち、前記第1の論理ボリュームに対応するものを削除し、

前記第2の論理ボリュームに格納されたデータ量が前記第2の論理ボリュームの容量の第5の割合を超えた場合、前記第2の論理ボリュームに書き込まれた全ての差分データを削除するディスクアレイ装置の制御方法。

10

【0100】

さらに、前記制御プロセッサは、最も古い世代の差分データを削除対象として決定するディスクアレイ装置の制御方法。

【0101】

さらに、前記制御プロセッサは、1回目の判定では前記第1の所定値を超えたか否かの判定をし、差分データ削除処理中の2回目以後の判定では前記第1の所定値より大きい第2の所定値を用いて、前記第2の論理ボリュームに格納されたデータ量が前記第2の所定値を超えたか否かの判定をするディスクアレイ装置の制御方法。

【0102】

20

さらに、前記制御プロセッサは、前記削除する差分データに対応する前記第1の論理ボリュームを決定する前に、前記差分データの削除処理中か否かを判定し、前記差分データの削除処理中であれば、現在実行中の削除処理を優先して実行するディスクアレイ装置の制御方法。

【図面の簡単な説明】

【0103】

【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。

【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差分LUの管理方法の説明図である。

【図4】本発明の実施の形態の空き容量チェック処理のフローチャートである。

30

【図5】本発明の実施の形態のLU追加処理のフローチャートである。

【図6】本発明の実施の形態のLU追加処理の説明図である。

【図7】本発明の実施の形態の仮想LU削除処理のフローチャートである。

【図8】本発明の実施の形態のペア情報管理テーブルの説明図である。

【図9】本発明の実施の形態の仮想LU削除処理の設定画面の説明図である。

【図10】本発明の実施の形態の仮想LU削除処理の設定画面の説明図である。

【図11】本発明の実施の形態の仮想LU削除処理のフローチャートである。

【符号の説明】

【0104】

1 ディスクアレイ装置

40

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

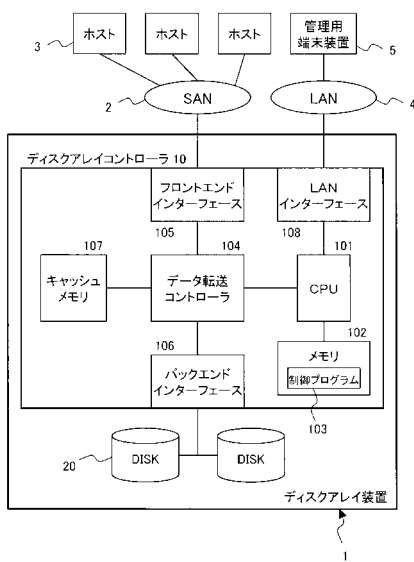
103 制御プログラム

104 データ転送コントローラ

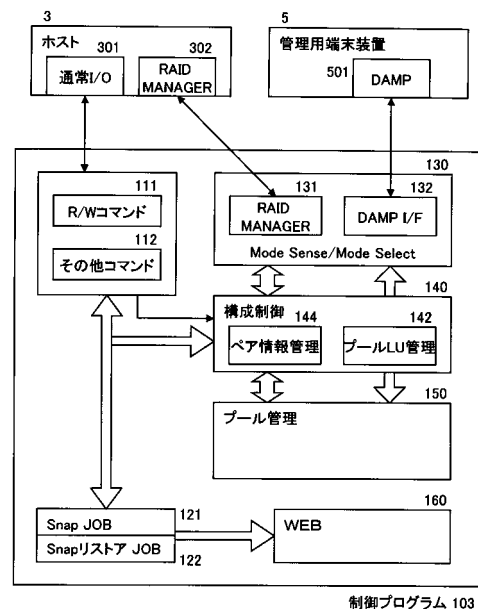
50

- 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 データバッファ
- 108 LANインターフェース

【図1】

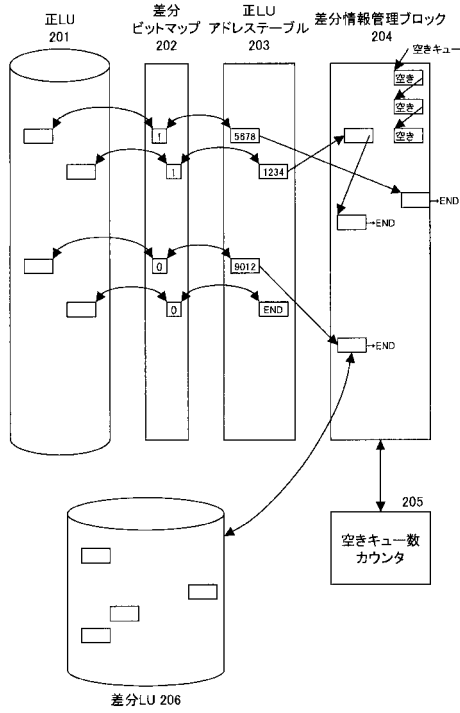


【図2】

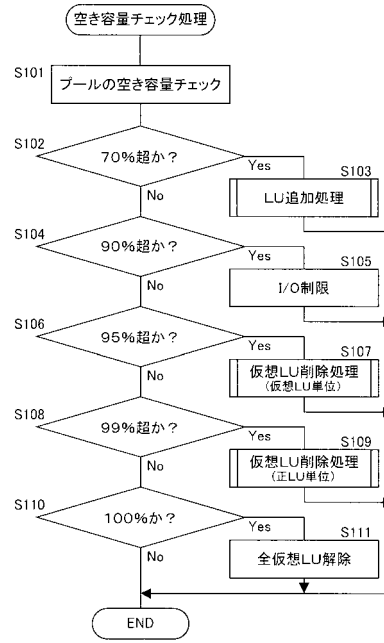


制御プログラム 103

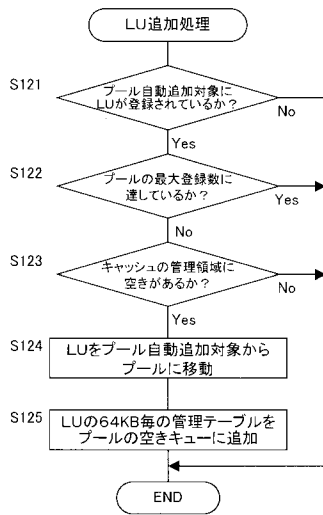
【図3】



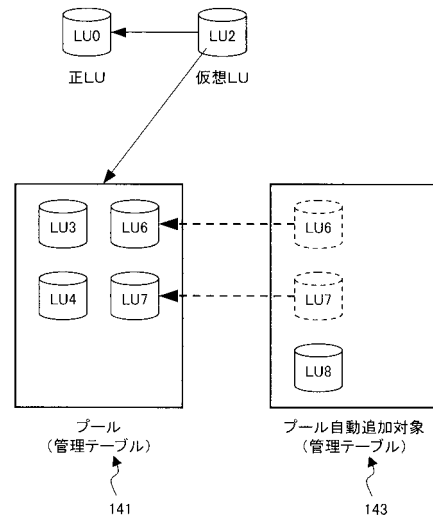
【図4】



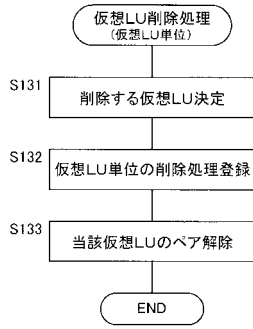
【図5】



【図6】



【 図 7 】

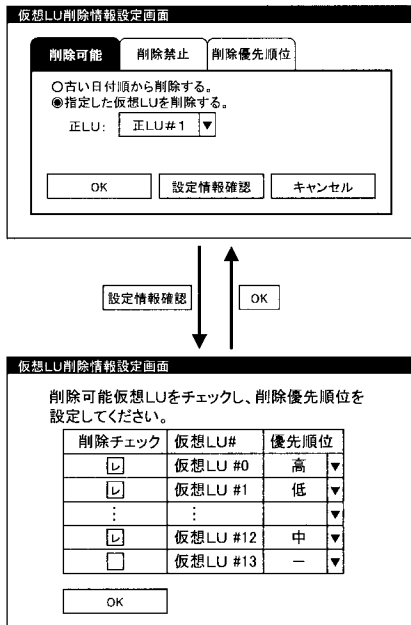


【 図 8 】

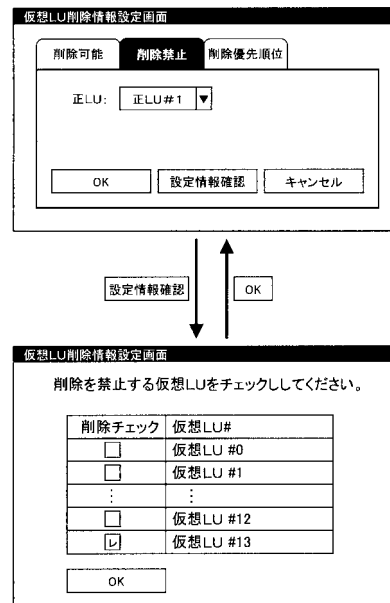
正LU	ペア状態	仮想LU#	snapshot作成日時
0	PSUS	0	20031008083030
0	PSUS	1	20031009084040
0	PSUS	2	20031010091010
0	PSUS	3	20031011084545
0	PSUS	4	20031012082020
0	PSUS	5	20031013091515
0	PSUS	6	20031014090505
0	PSUS	7	20030930141420
0	PSUS	8	20031007182030
0	PSUE	9	20030920101010
0	PSUS	10	20031121223030
0	PSUS	11	20031122223030
0	PSUS	12	20031123223030
0	PAIR	13	20031124223030

ペア情報管理テーブル

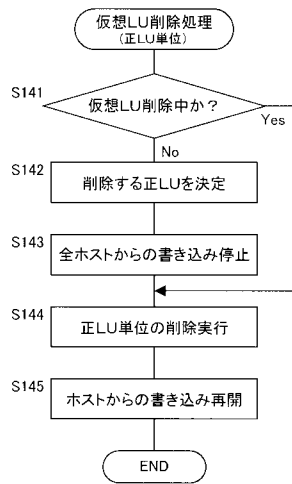
【 図 9 】



【 図 10 】



【図 11】



フロントページの続き

(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 RAIDシステム事業部内

審査官 木村 貴俊

(56)参考文献 特開2005-050024(JP,A)

特開2002-278819(JP,A)

米国特許出願公開第2004/0030951(US,A1)

米国特許出願公開第2004/0068636(US,A1)

(58)調査した分野(Int.Cl., DB名)

G06F 3/06 - 3/08

G06F 12/00

G06F 13/10 - 13/14