

(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2007-102579

(P2007-102579A)

(43) 公開日 平成19年4月19日(2007.4.19)

(51) Int. Cl. F I テーマコード (参考)
G06F 3/06 (2006.01) G06F 3/06 304F 5B065
 G06F 3/06 306Z

審査請求 未請求 請求項の数 19 O L (全 44 頁)

(21) 出願番号	特願2005-292842 (P2005-292842)	(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
(22) 出願日	平成17年10月5日 (2005.10.5)	(74) 代理人	110000062 特許業務法人第一国際特許事務所
		(72) 発明者	出口 彰 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所 内
		(72) 発明者	岩村 卓成 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所 内

最終頁に続く

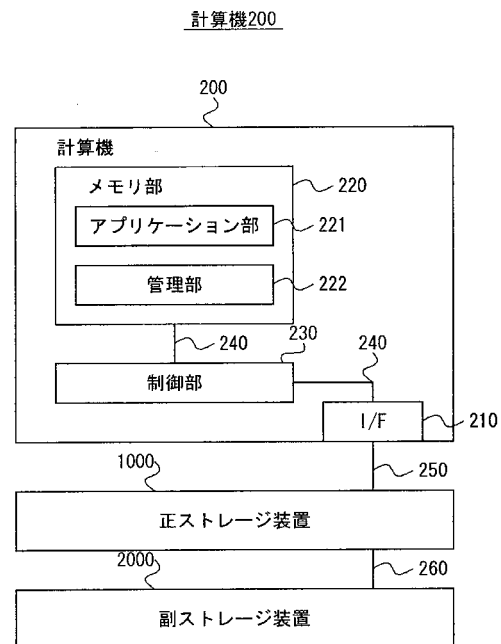
(54) 【発明の名称】 情報処理システム、情報処理システムの制御方法及びストレージ装置

(57) 【要約】

【課題】 リモートコピー機能がデータコピー状況（副ボリュームへのデータコピー停止中、差分コピー中など）にあわせて、リモートコピーペアの副ボリュームを構成するHDDの電源制御を行う。これによって、HDDの長寿命化、ストレージ装置の低電力化を図る。

【解決手段】 情報処理装置と、第一のボリュームを備え、情報処理装置と接続される第一の記憶装置と、第二のボリュームを備え、第一の記憶装置と接続される第二の記憶装置とを有し、前記第二の記憶装置は、第一の記憶装置に対するリモートコピー処理と連携して前記第二のボリュームに関する電源制御を行うことにより、リモートコピーと電源制御を効率よく実施する。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

情報処理装置と、
第一のボリュームを備え、前記情報処理装置と接続される第一の記憶装置と、
第二のボリュームを備え、前記第一の記憶装置と接続される第二の記憶装置とを有し、
前記第一のボリュームから前記第二のボリュームへのリモートコピーを実施する情報処理システムにおいて、

前記第一の記憶装置から前記第二の記憶装置へのリモートコピー処理と連携して前記第二のボリュームに関する電力制御を行うことを特徴とする情報処理システム。

【請求項 2】

請求項 1 に記載の情報処理システムにおいて、

前記情報処理装置から前記第一の記憶装置へのリモートコピー処理の指示に応じて、前記第一の記憶装置から前記第二の記憶装置へのリモートコピー処理と連携して前記第二のボリュームに関する電力制御を行うことを特徴とする情報処理システム。

【請求項 3】

請求項 1 に記載の情報処理システムにおいて、

前記第二の記憶装置に接続された第二の情報処理装置を有し、前記第二の情報処理装置から前記第二の記憶装置へのリモートコピー処理の指示に応じて、前記第一の記憶装置から前記第二の記憶装置へのリモートコピー処理と連携して前記第二のボリュームに関する電力制御を行うことを特徴とする情報処理システム。

【請求項 4】

請求項 1 に記載の情報処理システムにおいて、

前記第二の記憶装置は、更新情報記憶部を備え、前記第一のボリュームに対する更新情報を第二の記憶装置の前記更新情報記憶部へコピーするリモートコピーを実施し、

前記連携するリモートコピー処理がリモートコピー停止指示処理であるときに、前記更新情報記憶部にコピーされた更新情報の前記第二のボリュームへのリストア処理が完了した後に前記第二のボリュームに関する電力削減を指示することを特徴とする情報処理システム。

【請求項 5】

請求項 4 に記載の情報処理システムにおいて、

前記連携するリモートコピー処理がリモートコピー停止指示処理であるときに、前記第一の記憶装置は、前記第二の記憶装置にリモートコピー停止指示を通知することを特徴とする情報処理システム。

【請求項 6】

請求項 1 に記載の情報処理システムにおいて、

前記第一の記憶装置はさらに、第三のボリュームを備え、

前記第二の記憶装置はさらに、第四のボリュームを備え、

前記第三のボリュームから前記第四のボリュームへのリモートコピーが実施され、

前記連携するリモートコピー処理が前記第一のボリュームから前記第二のボリュームに対するリモートコピーと前記第三のボリュームから前記第四のボリュームに対するリモートコピーの形成コピー処理であるときに、前記第二のボリュームの電力削減を解除して第一のコピーを行い、前記第一のコピーが完了した後に前記第二のボリュームの電力削減を指示し、前記第四のボリュームの電力削減を解除して前記第三のボリュームから前記第四のボリュームへの第二のコピーを行うことを特徴とする情報処理システム。

【請求項 7】

請求項 6 に記載の情報処理システムにおいて、

前記第二のボリュームの電力削減を解除して行う第一のコピーが第一の差分コピーであり、前記第四のボリュームの電力削減を解除して行う第二のコピーが第二の差分コピーであることを特徴とする情報処理システム。

【請求項 8】

10

20

30

40

50

請求項 7 に記載の情報処理システムにおいて、前記第三のボリュームと前記第四のボリュームの電力削減の解除の指示は、前記第一の記憶装置と前記第二の記憶装置との間のネットワーク帯域を考慮して行われることを特徴とする情報処理システム。

【請求項 9】

請求項 1 に記載の情報処理システムにおいて、
前記第二の記憶装置は更新情報記憶部を備え、
前記第一のボリュームに対する更新情報を第二の記憶装置の前記更新情報記憶部へコピーするリモートコピーを実施し、
前記連携するリモートコピー処理が前記更新情報記憶部の更新情報を前記第二のボリュームへ反映するリストア処理であるときに、前記リストア処理を一次中止し、その間は前記第二のボリュームに対して電力削減を指示することを特徴とする情報処理システム。 10

【請求項 10】

情報処理装置と接続される第一の記憶装置の第一のボリュームから、前記
第一の記憶装置と接続される第二の記憶装置の第二のボリュームへのリモートコピーを実施する情報処理システムの制御方法において、
前記第一の記憶装置から前記第二の記憶装置へのリモートコピー処理と連携して前記第二のボリュームに関する電力制御を行うことを特徴とする情報処理システムにおける制御方法。

【請求項 11】

請求項 10 に記載の情報処理システムにおける制御方法において、 20
前記リモートコピーが、前記第一のボリュームに対する更新情報を前記第二の記憶装置の更新情報記憶部へコピーするリモートコピーであり、
前記連携するリモートコピー処理がリモートコピー停止指示処理であるときに、前記更新情報記憶部にコピーされた更新情報の前記第二のボリュームへのリストア処理が完了した後に前記第二のボリュームに関する電力削減を指示するステップを備えることを特徴とする情報処理システムにおける制御方法。

【請求項 12】

請求項 10 に記載の情報処理システムにおける制御方法において、
前記第一の記憶装置の第三のボリュームから前記第二の記憶装置の第四のボリュームへのリモートコピーが実施され、 30
前記連携するリモートコピー処理が前記第一のボリュームから前記第二のボリュームに対するリモートコピーと前記第三のボリュームから前記第四のボリュームに対するリモートコピーの形成コピー処理であるときに、前記第二のボリュームの電力削減を解除して第一の差分コピーを行うステップと、
前記第一の差分コピーが完了した後に前記第二のボリュームの電力削減を指示するステップと、
前記第四のボリュームの電力削減を解除して前記第三のボリュームから前記第四のボリュームへの第二の差分コピーを行うステップと、
を備えることを特徴とする情報処理システムにおける制御方法。

【請求項 13】

情報処理装置に接続され、第一のボリュームを備えた正記憶装置に接続される副記憶装置において、 40
前記副記憶装置は、第二のボリュームを備えており、
前記第一のボリュームから前記第二のボリュームへのリモートコピーが実施される際に、前記正記憶装置から前記副記憶装置へのリモートコピー処理と連携して前記第二のボリュームに関する電力制御を行うことを特徴とする副記憶装置。

【請求項 14】

請求項 13 に記載の副記憶装置において、
前記副記憶装置は、更新情報記憶部を備え、前記第一のボリュームに対する更新情報を前記副記憶装置の前記更新情報記憶部へコピーするリモートコピーを実施し、前記連携す 50

るリモートコピー処理がリモートコピー停止指示処理であるときに、前記更新情報記憶部にコピーされた更新情報の前記第二のボリュームへのリストア処理が完了した後に前記第二のボリュームに関する電力削減を指示することを特徴とする副記憶装置。

【請求項 15】

請求項 14 に記載の副記憶装置において、

前記連携するリモートコピー処理がリモートコピー停止指示処理であるときに、

前記正記憶装置から、前記副記憶装置へのリモートコピー停止指示の通知に応じて電力削減の指示を行うことを特徴とする副記憶装置。

【請求項 16】

請求項 13 に記載の副記憶装置において、

前記正記憶装置は、さらに、第三のボリュームを備え、

前記副記憶装置は、さらに、第四のボリュームを備え、

前記第三のボリュームから前記第四のボリュームへのリモートコピーが実施され、前記連携するリモートコピー処理が前記第一のボリュームから前記第二のボリュームに対するリモートコピーと前記第三のボリュームから前記第四のボリュームに対するリモートコピーの形成コピー処理であるときに、

前記第二のボリュームの電力削減を解除して第一のコピーを行い、前記第一のコピーが完了した後に前記第二のボリュームの電力削減を指示し、前記第四のボリュームの電力削減を解除して前記第三のボリュームから前記第四のボリュームへの第二のコピーを行うことを特徴とする副記憶装置。

10

20

【請求項 17】

請求項 16 に記載の副記憶装置において、

前記第二のボリュームの電力削減を解除して行う前記第一のコピーが第一の差分コピーであり、前記第四のボリュームの電力削減を解除して行う前記第二のコピーが第二の差分コピーであることを特徴とする副記憶装置。

【請求項 18】

請求項 17 に記載の副記憶装置において、前記第三のボリュームと前記第四のボリュームの電力削減の解除の指示は、前記正記憶装置と前記副記憶装置との間のネットワーク帯域を考慮して行われることを特徴とする副記憶装置。

【請求項 19】

請求項 13 に記載の副記憶装置において、

前記副記憶装置は更新情報記憶部を備え、

前記第一のボリュームに対する更新情報が前記副記憶装置の前記更新情報記憶部へコピーされるリモートコピーを実施し、

前記連携するリモートコピー処理が前記更新情報記憶部の更新情報を前記第二のボリュームへ反映するリストア処理であるときに、前記リストア処理を一次中止し、その間は前記第二のボリュームに対して電力削減を指示することを特徴とする副記憶装置。

30

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ストレージ装置を備えた情報処理システム及びストレージ装置を備えた情報処理システムの制御方法に関わり、特にディスク装置（以後、HDDと呼ぶ）の稼働期間の延長と、ストレージ装置の消費電力低減を実現するための技術に関する。

40

【背景技術】

【0002】

現在、企業情報システムのデータを長期保存するというニーズが拡大してきている。その背景として、法規制によって、金融、医療、製薬など分野においては、取引データ、医療データなどの長期保存が義務付けられてきているということがある。

【0003】

従来、長期データ保存メディアとしてテープがある。しかし、テープは、湿気やほこり

50

などの影響を受けやすいなどのことから、HDDなどに比較して管理が大変である。さらに、テープに格納されるデータをソフトウェアが利用可能な状態にするまでの時間もHDDに比較して長時間を要するというデメリットが知られている。

【0004】

特に最近では、HDDのビットコストの低下などから、データの長期保存用のストレージ装置にHDDを使用するケースも多く見られる。この場合、大きな課題となるのが、HDDの長寿命化とストレージ装置の低電力化である。

【0005】

特許文献1に記載されている技術では、ストレージ装置の低電力化を目的として、ホストコンピュータからのアクセスがなくなってから予め設定された時間経過後、HDDの電源の停止、または、節電モードへの移行を行う。

10

【0006】

次に、ストレージ装置のある記憶領域(コピー元記憶領域)に格納されているデータを、コピー元記憶領域とは異なる記憶領域(コピー先記憶領域)にコピーする技術がある。これは、コピー元記憶領域に書き込まれるデータを、コピー元記憶領域とコピー先記憶領域に2重書きする技術である(以後、リモートコピーと呼ぶ)。

特許文献2には、コピー元記憶領域に書き込まれるデータを、非同期にコピー先記憶領域に書き込む技術が開示されている。

【特許文献1】特開2000-293314号公報

【特許文献2】特開平11-85408号公報

20

【発明の開示】

【発明が解決しようとする課題】

【0007】

データの長期保存用のストレージ装置へのデータコピーには、リモートコピーなどが利用される可能性が高い。しかし、特許文献1と特許文献2に記載の技術では、リモートコピー実施中の電源制御の技術は開示されていない。本発明では、リモートコピー行いつつも効率良く電源制御を実施する方法を提供することを目的とする。

【課題を解決するための手段】

【0008】

本発明は、情報処理装置と、第一のボリュームを備え、前記情報処理装置と接続される第一の記憶装置と、第二のボリュームを備え、前記第一の記憶装置と接続される第二の記憶装置と、前記第一のボリュームから前記第二のボリュームへのリモートコピーを実施する情報処理システムにおいて、前記第一の記憶装置に対するリモートコピー処理と連携して前記第二のボリュームに関する電力制御を行うことを特徴とする。

30

【発明の効果】

【0009】

本発明では、リモートコピー部が、HDDの電源のON/OFFを制御することにより、リモートコピー部のコピー先記憶領域へのデータコピー状況(コピー先記憶領域へのデータコピー停止中、初期化コピー中など)にあわせてHDDの電源を制御することができる。よって、従来技術に比較して、HDDの長寿命化、ストレージ装置の低電力化を図ることができる。

40

【発明を実施するための最良の形態】

【0010】

以下、本発明の実施形態を説明する。以下に説明する実施形態は一例であって、本発明の実施形態はこれに限定されるものではない。

【実施例1】

【0011】

図1は、本発明の実施例1に係る情報処理システムのシステム構成の一例を示した図である。図1において、計算機200と正ストレージ装置1000は、ネットワーク250を介して接続されている。これによって、計算機200は、正ストレージ装置1000に

50

ライト/リードの要求を行うことができる。さらに、正ストレージ装置 1000 と副ストレージ装置 2000 はネットワーク 260 を介して接続されている。これによって、正ストレージ装置 1000 に格納されるデータを副ストレージ装置 2000 にコピーすることができる。ネットワーク 250、および、ネットワーク 260 は、ネットワーク障害に対応するために複数本であってもよい。

【0012】

計算機 200 は、メモリ部 220、CPU 230、インタフェース 210（以後、I/F と呼ぶ）を有する。CPU 230 とメモリ部 220、I/F 210 は装置内ネットワーク 240 によって接続されている。メモリ部 220 には、OS（Operating System）、ミドルウェア、アプリケーション部 221、管理部 222 等のソフトウェアが記録されており、CPU 230 がこれを実行する。なお、管理部 222 は、計算機 200 から正ストレージ装置 1000 へ各種設定を行うためのソフトウェアである。

10

【0013】

I/F 210 は、正ストレージ装置 1000 に対してデータのライトやリードといった I/O 処理（以下、単に I/O 要求と呼ぶ）を要求するためのインタフェースである。I/F 210 は、正ストレージ装置 1000 とネットワーク 250 を介して接続されている。なお、ネットワーク 250 としてはファイバーチャネルや ATM、イーサネット等が考えられるが、計算機 200 と正ストレージ装置 1000 の間で I/O 要求及びその応答を送受信できる媒体であればこれ以外であってもよい。

【0014】

図 4 は計算機 200 から正ストレージ装置 1000 に発行されるライト要求 300 の一例を示す図である。計算機 200 から発行されるライト要求 300 には、ライトデータ 310、ライト対象ボリュームの識別情報であるライト対象ボリューム ID 330、当該ライト対象ボリューム内におけるライト位置を示すライトアドレス 320、ライトデータ 310 のデータ長 340 が格納される。

20

【0015】

図 2 には、正ストレージ装置 1000 の構成の一例を示している。そして、図 3 には、副ストレージ装置 2000 の構成の一例を示している。正ストレージ装置 1000 と副ストレージ装置 2000 の構成要素/各部の多くは同一であるため、同一の構成要素/各部をまとめてストレージ装置 1000、2000 の構成要素/各部として説明する。本実施形態では、正ストレージ装置 1000 から副ストレージ装置 2000 へのリモートコピーを行う場合を考える。すなわち、正ストレージ装置 1000 はリモートコピーペアのコピー元ボリュームを有しており、副ストレージ装置 2000 はコピー先ボリュームを有している。

30

【0016】

ストレージ装置 1000、2000 は、大きく分けてメモリ部 1020、CPU 1030、キャッシュ 1040、1 つ以上の HDD（Hard Disk Drive）1060、ディスク制御モジュール 1100、I/F 1130、電源 1150 から構成されている。そして、CPU 1030 と I/F 1130、キャッシュ 1040、メモリ部 1020、ディスク制御モジュール 1100 は装置内ネットワーク 1050 で接続されている。さらに、HDD 1060 とディスク制御モジュール 1100 も装置内ネットワーク 1050 で接続されている。

40

【0017】

そして、電源 1150 とメモリ部 1020、キャッシュ 1040、HDD 1060 は、電源線 1160 で接続されている。電源 1150 は、メモリ部 1020、キャッシュ 1040、HDD 1060 へ電源線 1160 を使用して、電源を供給している。

【0018】

キャッシュ 1040 は、HDD 1060 に格納される使用頻度の高いデータなどを格納することにより、すべてのデータを HDD に格納する場合に比べて、計算機 200 からのリードやライトなどの処理を高速化するために利用する。I/F 1130 は、正ストレ

50

ジ装置 1000 と副ストレージ装置 2000 との間でデータの送受信や制御メッセージの送受信のためのインタフェースである。また、ストレージ装置 1000、2000 は複数個の I/F 1130 を有していてもよい。

【0019】

HDD 1060 には、計算機 200 によって実行されるソフトウェアが利用するデータを格納している。また、ディスク制御モジュール 1100 に全ての HDD 1060 が接続されている。1つ以上の HDD 1060 を、パリティグループ 1120 という単位でまとめて RAID (Redundant Arrays of Inexpensive Disks) のような方法で高信頼化してもよい。1つ以上の HDD 1060 からなるパリティグループ 1120 を1つ以上の論理的な領域に分けて、それぞれを論理的な記憶デバイスとして扱ってもよい。論理的な記憶デバイスをボリューム 1070 と呼ぶ。

10

【0020】

正ストレージ装置 1000 特有の構成要素、メモリ部 1020 内に有する各部および、制御情報に関して述べる。正ストレージ装置 1000 は、上記のストレージ装置 1000、2000 共通の構成要素に加えて、I/F 1010 を有する。I/F 1010 は、CPU 1030 とネットワーク 1050 によって接続されている。I/F 1010 は、計算機 200 から正ストレージ装置 1000 に対して発行される、ライトやリードといった I/O 要求を受け付けるためのインタフェースである。また、ストレージ装置 1000 は、複数個の I/F 1010 を有していてもよい。

【0021】

メモリ部 1020 は、大きく分けて制御部 1080 と制御情報部 1090 から構成される。正ストレージ装置 1000 の制御部 1080 は計算機 200 からライト要求 300 を受信して、これを処理するライトデータ受領部 1801 と、リモートコピーを実行するためのリモートコピー部 1082 を有する。なお、制御部 1080 が有する各部は、メモリ部 1020 に格納されているプログラムであり、正ストレージ装置 1000 の CPU 1030 によって実行される。しかし、これらプログラムの一部または全てをハードウェアによって実行してもよい。

20

【0022】

正ストレージ装置 1000 の制御情報部 1090 は、ペア情報テーブル 1091、ボリューム管理表 500、パリティグループ表 510、ボリューム構成情報 1092、差分管理表 1093、更新情報記憶部 1094、SEQ# 情報 1095 を有する。更新情報記憶部 1094 には後述の更新情報 1300 が記憶される。なお、制御情報部 1094 に含まれる各種情報は、後述するボリューム 1070 に格納してもよい。なお、正ストレージ装置 1000 と副ストレージ装置 2000 の更新情報記憶部 1094 は、キャッシュ 1040 の一部または全てを用いて実現することが考えられるが、これ以外にもパリティグループ 1120 やボリューム 1070 の一部または全ての領域を併用してもよい。

30

【0023】

副ストレージ装置 2000 特有の構成要素、メモリ部 1020 内に有する各部および、制御情報に関して述べる。副ストレージ装置 2000 は、上述したストレージ装置 1000、2000 共通の構成要素に加えて、電源制御モジュール 1140 を有する。電源制御モジュール 1140 は、制御部と HDD 1060 と装置内ネットワーク 1050 によって接続されている。電源制御モジュール 1140 は、後述する電源操作部 1083 からの指示により、HDD 1060 の電源の ON/OFF を切り替えるためのモジュールである。なお、明細書中では、あるパリティグループ 1120 に属する全ての HDD 1060 の電源を ON することを、単に「パリティグループ 1120 を起動する」と表現する。さらに、パリティグループ 1120 に属する全ての HDD 1060 の電源を OFF することを、単に「パリティグループ 1120 を停止する」と表現する。さらに、「HDD 1060 の電源の OFF」とは、HDD 1060 が消費電力を抑える動作状態に遷移することであり、例えば、HDD 1060 内部のプラッタの回転を停止したり、HDD 1060 内部のコ

40

50

ントローラーの消費電力を削減したり、実際にHDD1060の電源供給を停止することが考えられる。また、「HDD1060の電源のON」とはHDD1060が消費電力を抑えた状態から通常通りにライト/リード要求を処理できるような動作状態に復帰することである。

【0024】

メモリ部1020は、大きく分けて制御部1080と制御情報部1090から構成される。副ストレージ装置2000の制御部1080は、ライトデータ受領部1801、リモートコピー部1082を有しており、これらは、正ストレージ装置1000が有する各部と同一である。副ストレージ装置2000の制御部1080は、さらに電源操作部1083とコピー電源連携部1084を有している。

10

【0025】

副ストレージ装置2000の制御情報部1090は、ペア情報テーブル1091、ボリューム管理表500、パリティグループ表510、更新情報記憶部1094を有しており、これらは、正ストレージ装置1000のものと同じである。副ストレージ装置2000の制御情報部1090は、さらに目標電力量1097と更新情報量閾値1098を有している。

【0026】

目標電力量1097について説明する。本発明の実施形態では、リモートコピーペアの差分コピーを行う際に、副ストレージ装置2000で、一度に起動するパリティグループ1120を一定数に制限することで、低電力化を図る。本実施形態では、目標電力量1097の値を超えないように起動するパリティグループ1120の数を制御する。なお、目標電力量1097は、管理者が管理部222などを介して、副ストレージ装置2000に設定しておくものとする。

20

【0027】

更新情報量閾値1098について説明する。本発明の実施形態では、更新情報記憶部1094に正ストレージ装置1000から転送される更新情報1300を一定量滞留させておき、更新情報量閾値1098を超えた場合に、更新情報1300中のライトデータを副ストレージ装置2000のコピー先ボリュームへ書き込むように制御する。詳細は、本発明の実施形態の詳細を説明するときに説明する。

【0028】

なお、副ストレージ装置2000が有する電源制御モジュール1140、電源操作部1083、コピー電源連携部1084、目標電力量1097、および、更新情報量閾値1098は、正ストレージ装置1000も有してよい。さらに、副ストレージ装置2000も、差管理表1093、SEQ#情報1095を有してもよい。それらを有している場合には、副ストレージ装置2000から、正ストレージ装置1000へのリモートコピーを行うことも可能である。

30

【0029】

リモートコピーは、任意のボリューム間で形成することができる。そして、本発明の実施形態はリモートコピーのコピー先ボリュームを構成するHDD1060の電源を制御する。そのため、ここで、ボリュームとHDD1060の関係を説明する。まず、前述したように複数個のHDD1060を、パリティグループ1120という単位でまとめており、この関係を管理するのがパリティグループ表510である。

40

【0030】

次に、パリティグループ1120を1つ以上の論理的な領域であるボリューム1070に分けて利用しており、この関係を管理するのがボリューム管理表500である。図6には、パリティグループ表510の一例を示している。

【0031】

パリティグループ表510は、属性としてパリティグループID、HDD-ID、RAIDレベル、電源状態、使用電力量を持つ。パリティグループIDは、パリティグループ1120に対して割り当てられ、当該パリティグループ1120を一意に識別可能な情報

50

である。HDD-IDは、パリティグループIDによって示されるパリティグループ1120に属するHDD1060の識別子を示している。なお、HDD-IDの値は、ある1つのストレージ装置1000、2000において、HDD1060を一意に識別可能な識別情報である。

図6の例では、パリティグループ「1」にHDD-IDが「0」のHDD1060とHDD-IDが「1」のHDD1060が属していることを示している。RAIDレベルは、パリティグループ1120に属するHDD1060が構成しているRAIDのレベルを示している。電源状態は、パリティグループ1120に属するHDD1060の電源がONであるかOFFであるかを示している。使用電力量は、パリティグループ1120に属する全てのHDD1060の電源をONにした場合に必要となる電力量を示している。

10

【0032】

図5には、ボリューム管理表500の一例を示している。ボリューム管理表500は、属性としてボリュームID、容量、パリティグループID、MAIDフラグ、電源停止可能フラグを持つ。ボリュームIDは、ボリューム1070に対して割り当てられ、当該ボリュームを一意に識別可能な識別情報である。容量は、ボリュームIDによって示されるボリューム1070に格納可能なデータ量を示している。

パリティグループIDは、ボリュームIDによって示されるボリューム1070が属するパリティグループ1120のIDである。MAIDフラグは、ボリュームIDによって示されるボリューム1070が属するパリティグループ1120内のHDD1060に対する電源のON/OFFをストレージ装置1000、2000が制御可能であるか否かを示している。MAIDフラグの値が「true」の場合は、ストレージ装置1000、2000はHDD1060の電源のON/OFFを制御可能であることを示している。MAIDフラグの値が「false」の場合は、ストレージ装置1000、2000はHDD1060の電源のON/OFFを制御不能であることを示している。

20

【0033】

次に、1つのパリティグループ1120には、1つ以上のボリューム1070が属することができる。図5の例では、パリティグループ1にボリューム1とボリューム2が属している。このため、ボリューム1にとっては、パリティグループ1120を停止可能な状態にあり、ボリューム2にとっては、パリティグループ1120内のHDD1060へアクセスしているなどの理由から停止不可能な状態にあるという場合がある。

30

よって、本実施形態では、ボリューム1070がパリティグループ1120を利用するときには、電源停止可能フラグを「false」にすることで、パリティグループ1120を利用していることを宣言する。そして、電源停止可能フラグを「true」にすることで、パリティグループ1120を利用していないことを宣言する。

【0034】

そして、本実施形態では、同一パリティグループ1120に属する全てのボリュームの電源停止可能フラグを参照して、全てが「true」の場合にパリティグループ1120内の全てのHDD1060の電源をOFFするように制御する。

さらに、電源停止可能フラグの少なくとも1つ以上が「false」である場合には、パリティグループ1120内の全てのHDD1060の電源をOFFしないように制御する。

40

なお、MAIDフラグが「false」の場合には、ボリューム1070を構成するHDD1060の電源制御は不可能である。よってその場合の電源停止可能フラグは常に「false」とする。

【0035】

レコード501は、1つのボリューム1070に関するボリュームID、容量、パリティグループID、MAIDフラグ、電源停止可能フラグから成る情報である。本発明の実施形態では、以上の情報によってHDD1120からボリューム1070を設定し、計算機200へ提供する。

【0036】

50

では、リモートコピーの概要を述べる。図7には、リモートコピー部1082の構成の一例を示している。一般的には、リモートコピーには、同期リモートコピー、および、非同期リモートコピーがあるが、本実施例では非同期リモートコピーのみを例として説明する。よって、リモートコピー部1082は、非同期リモートコピー部120から成る。非同期リモートコピー120は、更新情報作成部121、データ転送部122、パージ部123、更新情報受領部124、リストア部125、ペア形成部126、サスペンド部127、リシンク部128、差分コピー部129を有する。なお、非同期リモートコピー部1082が有する各部の説明は後述する。ここでは、非同期リモートコピーの概要を述べる。

【0037】

リモートコピーは、任意のボリューム間で形成することが可能である。以後の説明では、リモートコピーのコピー元となるボリューム1070を正ボリューム、コピー先となるボリューム1070を副ボリュームと呼ぶことにする。コピー元ボリュームを有するストレージ装置を正ストレージ装置、コピー先ボリュームを有するストレージ装置を副ストレージ装置と呼ぶことにする。さらに、コピー元ボリュームとコピー先ボリュームの組をリモートコピーペアと呼ぶ。

【0038】

正ストレージ装置1000が、計算機200からライト要求300を受領したときに、ライト対象ボリュームを正ボリュームとするリモートコピーペアが存在する場合に、リモートコピー部1082を起動することでリモートコピーを実行する。ライト対象ボリュームを正ボリュームとするリモートコピーペアが存在するか否かの判定にはペア情報テーブル1091を使用する。

【0039】

図8には、ペア情報テーブル1091の一例を示している。ペア情報テーブル1091は、属性として正ボリュームID、副ボリュームID、ペア状態、相手ストレージ装置ID、リモートコピータイプを有している。正ボリュームIDは、リモートコピーペアの正ボリュームに割り当てられ、正ストレージ装置1000内で当該ボリュームを一意に識別可能な識別情報である。副ボリュームIDは、リモートコピーの副ボリュームに割り当てられ、副ストレージ装置2000内で当該ボリュームを一意に識別可能な識別情報である。

【0040】

ペア状態は、正ボリュームと副ボリュームに対して定義されているリモートコピーペアの状態を示している(Initial-Copying、Duplex、Duplex-pending、Suspendの状態があるが、これらは、後述する)。

【0041】

相手ストレージ装置IDは、正ボリュームを有する正ストレージ装置1000においては、副ストレージ装置のIDであり、副ボリュームを有する副ストレージ装置2000においては、正ストレージ装置のIDである。リモートコピータイプは、正ボリュームIDと副ボリュームIDで示される2つのボリューム間で定義されているリモートコピーのタイプを示しており、同期または非同期のいずれかの値を取り得る。

【0042】

レコード700は、1つのリモートコピーペアに関する正ボリュームID、副ボリュームID、ペア状態、相手ストレージ装置ID、リモートコピータイプからなる情報である。すなわち、ペア情報テーブル1091は、レコード700の集合である。

【0043】

では、以降には、リモートコピーの使用方法を述べる。まず、ユーザは計算機200の管理部222などから、リモートコピーペアの形成指示を正ストレージ装置1000へ発行する。リモートコピーペアの形成指示を受領した正ストレージ装置1000は、正ボリュームと副ボリュームに格納されているデータを同一にするために、正ボリュームに格納されている全てのデータを副ストレージ装置へコピーする(初期化コピー)。

10

20

30

40

50

【0044】

さらに、リモートコピーペアの形成指示を受領した正ストレージ装置は、ペア情報テーブル1091へ正ボリュームID、副ボリュームID、ペア状態、相手ストレージ装置ID、リモートコピータイプからなるレコード700を追加する。これにより、以降、正ストレージ装置1000が計算機200から新たに形成したリモートコピーペアの正ボリュームに対するライト要求300を受領した場合には、リモートコピー部1082が起動されるようになる。つまり、ライト要求300のライトデータ310を非同期に副ボリュームへ転送することで、正ボリュームと副ボリュームに格納されるデータを同一に保つことができる。なお、リモートコピーペアの形成指示を受領し、初期化コピーが完了するまでのペア状態のことをInitial-Copying状態と呼ぶ。

10

【0045】

初期化コピーが完了すると、リモートコピーペアのペア状態は、Duplex状態となる。正ストレージ装置1000が、計算機200からDuplex状態であるリモートコピーペアの正ボリュームに対するライト要求300を受領したときには、非同期リモートコピー部120が、ライト要求300を受領したタイミングとは非同期に副ストレージ装置2000の副ボリュームにもライトデータ310を書き込む。

【0046】

Duplex状態とは、正ストレージ装置1000から副ストレージ装置2000へ転送中、および、未転送データを除いて正ボリュームの内容と副ボリュームの内容が同一である状態のことである。ただし、例えば、ボリューム毎に一意的な識別子が、ボリュームに格納される場合等では、このような識別子が保存されているボリューム内の特定の領域について、正ボリュームと副ボリュームのデータの内容が同一でない場合があっても良いものとする。なお、このような同一性のことを以後の説明では、同期状態と呼ぶことにする。

20

【0047】

以下に、正ストレージ装置1000から副ストレージ装置2000へ、ライト要求300とは非同期にライトデータ310を書き込む方法を概説する。非同期リモートコピー部120が実行する正ボリュームから副ボリュームへのデータのコピー方法としては、(1)正ストレージ装置1000は、ライト要求300を受領する度に、ライト要求中のライトデータ310、ライトアドレス320などを含む更新情報1300を作成し、計算機200へライト完了を報告する。(2)正ストレージ装置1000は、計算機200へのライト完了の報告とは、独立したタイミングで、更新情報1300を副ストレージ装置2000に転送する。(3)副ストレージ装置2000は、正ストレージ装置1000から受領した更新情報1300に含まれているライトアドレス320に基づいて副ボリュームにライトデータ310を書き込む。このような方法が考えられる。

30

【0048】

さらに、本実施形態では、この発展形として、正ストレージ装置1000がライト要求300を受領した順序情報(後述のSEQ#情報1095のこと)を、更新情報1300に付与する。さらに、副ストレージ装置2000では、当該順序情報を利用して、正ストレージ装置1000がライト要求300を受領した順序に従って、ライトデータ310を副ボリュームへ書き込む。

40

【0049】

すなわち、本実施形態における非同期リモートコピーでは、正ストレージ装置1000には、副ストレージ装置2000に未転送の更新情報1300が更新情報記憶部1094に滞留している。さらに、副ストレージ装置2000においても、更新情報1300は、一旦副ストレージ装置2000の更新情報記憶部に1094格納され、その後、更新情報1300内のライトデータが副ボリュームに書き込まれる。

【0050】

また、非同期リモートコピー部120をより効率化するために、正ボリューム内の同一アドレスに対して、ある期間内にライト要求300が複数回発行された場合には、全てのライト要求300について更新情報1300を副ストレージ装置2000に送信するので

50

はなく、当該期間内に発行された最後のライト要求 300 に対する更新情報 1300 のみを副ストレージ装置 2000 に送信するとしても良い。

【0051】

また、ストレージ装置 1000、2000 がキャッシュ 1040 を有する場合には、ライトデータ 310 が格納されている正ストレージ装置 1000 のキャッシュ 1040 上の記憶領域を示すポインタを、更新情報 1300 に付与しておき、キャッシュ 1040 上のライトデータ 310 が別のライト要求 300 によって更新されるのを契機に、キャッシュ 1040 から更新情報記憶部 1094 へライトデータ 310 をコピーして更新情報 1300 にライトデータ 310 を付与しても良い。即ち、正ストレージ装置 1000 のキャッシュ 1040 上でライトデータ 310 の更新が起こる直前まで、更新情報 1300 作成のためのデータのコピーを遅延させてもよい。これにより、更新情報 1300 のライトデータと、キャッシュ上のライトデータを共有することができる。

10

【0052】

次に、計算機 200 からリモートコピーペアのサスペンド指示を正ストレージ装置へ発行することができる。これは、正ボリュームから副ボリュームへのデータ転送を一旦停止するための指示である。このように、正ボリュームから副ボリュームへのデータ転送を一旦停止し、正ボリュームと副ボリュームの同一性が保証されなくなったりリモートコピーペアのペア状態のことを S u s p e n d 状態と呼ぶ。

【0053】

正ストレージ装置が S u s p e n d 状態であるリモートコピーペアの正ボリュームに対するライト要求 300 を受領したときには、ライト要求 300 中のライトアドレス 320 を差分管理表 1093 に記録してもよい。すなわち、正ボリュームと副ボリュームで異なるデータが格納されているアドレスを記録する。この差分管理表 1093 を利用することで、正ボリュームと副ボリュームを再び同期状態にする場合には、正ボリュームと副ボリュームの間で格納されているデータが異なる領域についてのみコピーすることが可能となる（差分コピーと呼ぶ）。

20

【0054】

計算機 200 から、S u s p e n d 状態のリモートコピーペアの正ボリュームと副ボリュームを再び同期状態とするためのリシンク指示を発行することができる。これは、先に述べたように差分コピーによって実現することができる。リモートコピーペアのリシンク指示を受領し、差分コピーが完了するまでのペア状態のことを D u p l e x - P e n d i n g 状態と呼ぶ。また、差分コピーが完了し、正ボリュームと副ボリューム間に格納されているデータ内容が同期状態となった後に、ペア状態は D u p l e x 状態となる。

30

【0055】

なお、以上の説明では I n i t i a l - C o p y i n g 状態と D u p l e x - P e n d i n g 状態は別々な状態としたが、これらをまとめて1つの状態として表示、遷移させても良い。本実施形態中では、I n i t i a l - C o p y i n g 状態と D u p l e x - P e n d i n g 状態をまとめて、D u p l e x - P e n d i n g 状態として扱う。さらには、本実施形態では、I n i t i a l - C o p y i n g 状態に実行する初期化コピーと D u p l e x - P e n d i n g 状態に実行する差分コピーもまとめて、差分コピーとして扱う。すなわち、初期化コピーは正ボリュームの全てのアドレスが差分管理表 1093 に記録されているものとして扱う。

40

【0056】

では、次に、計算機 200 から S u s p e n d 状態であるリモートコピーペアの正ボリュームに対するライト要求 300 を受領したときに、当該ライト要求 300 のライトアドレス 320 を記録する差分管理表 1093 の説明を行う。図 9 には、差分管理表 1093 の一例を示している。

【0057】

まず、ボリュームアドレスは、ボリュームの全領域を一定サイズの小領域に分割して、小領域の先頭から順に付与する番号のことである。差分ビットは、S u s p e n d 状態の

50

ときなどに、ボリュームアドレスによって示される小領域が計算機 200 によって更新されたか否かを示している。差分ビットの値が「on」のときは更新されたことを示し、差分ビットの値が「off」のときは更新されていないことを示す。ボリュームアドレスと当該アドレスに対する差分ビットの組をレコード 800 とする。差分管理表 1093 は、レコード 800 の集合から成る。

なお、ある 1 つの差分管理表 1093 に関して、差分ビットが「on」であるレコード 800 が 1 つ以上ある場合に、差分が記録されていると表現する。さらに、ある 1 つの差分管理表 1093 に関して、全てのレコード 800 の差分ビットが「off」である場合に、差分が記録されていないと表現する。また、差分管理表 1093 のレコード 800 の差分ビットを「on」に変更することを、その差分管理表 1093 に差分を記録すると表現する。

【0058】

最後に、コピー済みボリュームアドレスは、差分コピーがコピー完了済み領域と未完了領域を区別するために利用する。図 9 に示すコピー済みボリュームアドレスは、ボリュームアドレス「0」と「1」に関しては差分コピーが完了しており、ボリュームアドレス「2」をコピー中であり、ボリュームアドレス「3」以降は差分コピーが未完了である状態を示している。なお、差分管理表 1093 を利用して差分コピーを実行する差分コピー部 129 の詳細は後述する。

【0059】

本発明の実施形態の概要を述べる。第 1 の本発明の実施形態（リモートコピーのサスペンド指示と連携した電源制御）の特徴を図 10 に示している。リモートコピーの概要説明で述べたように、リモートコピーのペア状態が Suspend 状態になった場合には、正ボリュームに対するライト要求 300 に含まれるライトデータ 310 は、副ボリュームに書き込まれない。さらに、副ストレージ装置 2000 がデータ長期保存用のストレージ装置である場合には、副ストレージ装置 2000 に計算機が接続されていない構成もあるため、Suspend 状態になった直後の副ボリュームを使用するケースは少ないと考えられる。図 10 に示す発明の特徴はこうした背景を利用して電源制御を効率よく行うものである。

【0060】

まず、正ストレージ装置 1000 が計算機 200 からリモートコピーペアに対するサスペンド要求を受領する。図 10 の例は、パリティグループ 1 に属している 3 つのリモートコピーペアのサスペンドを行う場合を示している。サスペンド要求を受領すると、正ストレージ装置 1000 のリモートコピー部 1082 は、正ストレージ装置 1000 から副ストレージ装置 2000 への更新情報 1300 の転送をやめる。この段階で、副ストレージ装置 2000 に新たな更新情報 1300 は到着しなくなる。次に、副ストレージ装置 2000 のリモートコピー部 1082 は、副ストレージ装置 2000 に滞留する更新情報 1300 のライトデータを全て副ボリュームに書き込んだ後、副ボリュームが属しているパリティグループ 1 内の HDD 1060 の電源を OFF する。

【0061】

図 10 下側には、このときのパリティグループ表 510 の内容を示しており、サスペンドしたリモートコピーペアの副ボリュームが属しているパリティグループ 1 の電源状態は OFF となる。また、パリティグループ 2 に属している 3 つのリモートコピーペアは、正ボリュームから副ボリュームへのデータコピーを継続しているため、パリティグループ 2 の電源状態は ON のままである。

【0062】

このように、リモートコピーペアをサスペンドする操作にあわせて HDD 1060 の電源を制御することで、以降、I/O が発生しないことを副ストレージ装置 2000 が認識した直後に副ボリュームを構成する HDD 1060 の電源を OFF することができ、従来技術に比較して、さらなる HDD 1060 の長寿命化、ストレージ装置 2000 の低電力化を実現できる。

10

20

30

40

50

【 0 0 6 3 】

次に、第2の本発明の実施形態（ストレージ装置間のネットワーク帯域を考慮した電源制御）の特徴を説明する。一般的には、正ストレージ装置1000と副ストレージ装置2000の間のネットワーク260の転送帯域幅は、正ストレージ装置1000または副ストレージ装置2000が有する全てのHDD1060の転送帯域幅の合計よりも小さな値となる。以下に説明する特徴では、こうした傾向を利用して、形成コピーまたは差分コピー中のONするHDD1060の数を削減する。

【 0 0 6 4 】

図11は正ストレージ装置1000と副ストレージ装置2000の間のネットワーク帯域が100Mbps（bit per second）であり、正ストレージ装置1000のパーティグループ1のリード性能、および、副ストレージ装置2000のパーティグループ1へのライト性能が共に100Mbpsである場合のシステム状態を示している。また、パーティグループ1120を構成するHDD1060の転送帯域幅が100Mbps以上であるため、1つのパーティグループ間のデータコピーでネットワーク260を充足することができるものとする。

第2の本発明の実施形態の特徴によれば、副ストレージ装置2000では、まず、パーティグループ1のみ起動して、パーティグループ1に副ボリュームが属しているリモートコピーペアの差分コピーを実施する。差分コピーが完了すると、パーティグループ1を停止し、差分コピーが未完了のリモートコピーペアの副ボリュームが属しているパーティグループ2を起動する。そして、パーティグループ2に副ボリュームが属しているリモートコピーペアの差分コピーを実施する。

【 0 0 6 5 】

図11下側には、パーティグループ1に副ボリュームが属しているリモートコピーの差分コピーを実施しているときの、パーティグループ表510の内容を示している。パーティグループ1の電源状態はONであるが、差分コピーを実施していないパーティグループ2の電源状態はOFFである。

【 0 0 6 6 】

最後に、図12を用いて第3の本発明の実施形態（非同期リモートコピーの処理特性を利用した電源制御）の特徴を説明する。非同期リモートコピーでは、正ストレージ装置1000は、計算機200からライト要求300を受領したときに、計算機200に完了メッセージを返し、その後、副ストレージ装置2000へライトデータ310を転送する。そのため、正ストレージ装置1000や副ストレージ装置2000の更新情報記憶部1094に、計算機200が正ストレージ装置1000へ発行したライトデータ310を含む更新情報1300を蓄積することができる。本特徴ではこうした特性を利用して電源制御を行う。

【 0 0 6 7 】

本特徴を実現するためには、まず、リモートコピーの処理を実行し、正ストレージ装置1000と副ストレージ装置2000の更新情報記憶部1094にライトデータ310を含む更新情報1300を、意図的に蓄積させる。すなわち、副ボリュームに対するライトデータ310の書き込みは、一定時間または、一定量のライトデータ310が蓄積するまで待つ。そして、ライトデータ310の書き込みを停止させている間は副ボリュームが含まれるパーティグループ1120の電源をOFFすることで、消費電力の削減を図る。本特徴は特に、更新情報記憶部1094の一部領域としてパーティグループ1120の全てまたは一部を利用可能である場合、特に有効な消費電力削減方法となる。

【 0 0 6 8 】

まず、以降の説明で利用する、SEQ#情報1095、更新情報1300に関して説明する。図13には、更新情報作成部121、差分コピー部129、サスペンド部127が、更新情報1300を作成するときに割り当てる数値（SEQ#）を管理しているSEQ#情報1095の一例を示している。SEQ#は、正ストレージ装置1000が計算機200から受領するライト要求300の順序と、副ストレージ装置2000で、ライトデー

10

20

30

40

50

タ 3 1 0 を副ボリュームに書き込む順序を同一にするために利用する。

【 0 0 6 9 】

図 1 4 には、非同期リモートコピーを行う際に、正ストレージ装置 1 0 0 0 から副ストレージ装置 2 0 0 0 へ転送する更新情報 1 3 0 0 の一例を示す。更新情報 1 3 0 0 は、S E Q #、ライトデータ、副ボリューム I D、ライトアドレス、データ長、副ストレージ装置 I D、サスペンドフラグ、コピー完了フラグを有する。

【 0 0 7 0 】

更新情報 1 3 0 0 は、計算機 2 0 0 から正ストレージ装置 1 0 0 0 に書き込まれるライトデータ 3 1 0 を副ストレージ装置へ転送するために使用する。

【 0 0 7 1 】

更新情報 1 3 0 0 に格納する値を以下に示す。ライトデータ、ライトアドレス、データ長には、正ストレージ装置 1 0 0 0 が計算機 2 0 0 から受領するライト要求 3 0 0 に格納されている情報を格納する。副ボリューム I D、副ストレージ装置 I D は更新情報作成部 1 2 1 がペア情報テーブル 1 0 9 1 から取得した値を格納する。

これによって、副ストレージ装置 2 0 0 0 にもライトデータを書き込むことができる。また、S E Q # は、上述した S E Q # 情報 1 0 9 5 から取得した値を格納する。サスペンドフラグ、コピー完了フラグは、サスペンド処理、リシンク処理で利用するためのものであり、後述する。

【 0 0 7 2 】

リモートコピー部 1 0 8 2 が実行する処理は、リモートコピーペアのペア状態によって異なるが、どの状態においても計算機からライト要求 3 0 0 を受領することができる。以降には、(1) リモートコピーペアのペア状態が Dup l e x である場合のライト要求 3 0 0 の処理、(2) リモートコピーペアをサスペンドする処理、および、ペア状態が S u s p e n d 状態である場合のライト要求 3 0 0 の処理、(3) リモートコピーペアをリシンクする処理、および、Dup l e x - P e n d i n g 状態である場合のライト要求の処理の順で説明していく。

(1) リモートコピーペアのペア状態が Dup l e x 状態である場合のライト要求の処理
ペア状態が Dup l e x 状態であるリモートコピーペアの正ボリュームに対して発行されるライト要求 3 0 0 のライトデータは、ライト要求 3 0 0 を受領するタイミングとは、非同期に副ストレージ装置 2 0 0 0 の副ボリュームにも書き込む。これは、正ストレージ装置 1 0 0 0 のライトデータ受領部 1 0 8 1、更新情報作成部 1 2 1、データ転送部 1 2 2、ページ部 1 2 3、副ストレージ装置 2 0 0 0 の更新情報受領部 1 2 4、リストア部 1 2 5 の連携によって実現される。

【 0 0 7 3 】

図 1 5 には、ライトデータ受領部 1 0 8 1 の処理の一例を示している。なお、本処理を行う前に、後ほど説明する差分コピーによって、正ボリュームのデータと副ボリュームのデータは一致しているものとする。ライトデータ受領部 1 0 8 1 は計算機 2 0 0 から、ライト要求 3 0 0 を受領する (ステップ 8 0 1)。このとき、図 4 に示すようにライト要求 3 0 0 としてライトデータ受領部 1 0 8 1 は、ライトデータ 3 1 0、ライト対象ボリューム I D 3 3 0、ライトアドレス 3 2 0、データ長 3 4 0 を受け取る。

【 0 0 7 4 】

次に、ライトデータ受領部 1 0 8 1 は、ライト対象ボリューム I D が示すボリューム 1 0 7 0 を正ボリュームとするリモートコピーペアに関するレコード 7 0 0 を取得する (ステップ 8 0 2)。具体的には、ライトデータ受領部 1 0 8 1 は、ライト対象ボリューム I D 3 3 0 をキーとしてペア情報テーブル 1 0 9 1 を検索し、正ボリューム I D にライト対象ボリューム I D と同じ I D が登録されているレコード 7 0 0 を探す。

【 0 0 7 5 】

続けて、ライトデータ受領部 1 0 8 1 は、ライト対象ボリューム I D 3 3 0 が示すボリューム 1 0 7 0 を正ボリュームとした非同期リモートコピーが定義されているかを調べる (ステップ 8 0 3)。具体的には、ステップ 8 0 2 で得たレコード 7 0 0 のリモートコピー

10

20

30

40

50

タイプが「非同期」であるかどうかを調べる。なお、ステップ802でライト対象ボリュームID330が示すボリューム1070を正ボリュームとしたリモートコピーペアに関するレコード700が存在しなかった場合には、ステップ803の結果は「no」となる。

【0076】

ステップ803の結果が「no」の場合、ライトデータ受領部1081はステップ804及び805をスキップして、ステップ806を実行する。

【0077】

ステップ803の結果が「yes」の場合、ライトデータ受領部1081は、ライト要求300の内容とステップ803で調べたレコード700をパラメータとして、更新情報作成部121を呼び出す(ステップ804)。そして、ライトデータ受領部1081は、更新情報作成部121からの応答を待つ(ステップ805)。

10

【0078】

ライトデータ受領部1081は、更新情報作成部121から応答を受けると、ライト要求300中のライトデータ310をライト対象ボリュームに書き込む(ステップ806)。なお、ボリューム1070への書き込みはストレージ装置内部のキャッシュ1040への書き込みで代替してもよい。

【0079】

最後に、ライトデータ受領部1081は、ライトの要求元である計算機200へライト要求完了を報告する(ステップ807)。すなわち、計算機200は、この時点でライトデータ310の書き込みが完了したと認識する。

20

【0080】

次に更新情報作成部121の処理について説明する。図16は更新情報作成部121の処理の一例を示す図である。更新情報作成部121は図15に示すライトデータ受領部1081のステップ804において、ライトデータ受領部1081によって起動される。

【0081】

更新情報作成部121は、計算機200から正ストレージ装置1000が受領したライト要求300中のライトデータ310にSEQ#を割り当て、更新情報1300を作成する部である。

【0082】

まず、更新情報作成部121は、パラメータとして渡されたレコード700のペア状態を調べる(ステップ1201)。

30

【0083】

ここでは、リモートコピーペアは、Duplex状態であるとしたので、更新情報作成部121はステップ1202をスキップしてステップ1203以降の処理を実行する。

【0084】

更新情報作成部121は、SEQ#情報1095から現在のSEQ#を得て、次の更新情報1300に付与するSEQ#の値を用意しておくために、SEQ#情報1095に記録されているSEQ#に1を加える(ステップ1203)。

【0085】

次に、更新情報作成部121は、更新情報1300を作成し、更新情報記憶部1094へ格納する(ステップ1204)。

40

【0086】

このとき、更新情報1300に格納する値に関して述べる。更新情報1300のSEQ#は、ステップ1203で得たSEQ#を格納する。ライトデータ、ライトアドレス、データ長は、ライト要求300のライトデータ310、ライトアドレス320、データ長340を格納する。そして、副ボリュームIDには、ライトデータ受領部1081から渡されるレコード700中の副ボリュームIDを格納し、副ストレージ装置IDには、ライトデータ受領部1081から渡されるレコード700中の相手ストレージ装置IDを格納する。最後に、サスペンドフラグ、コピー完了フラグは、共に「off」とする。最後に、

50

更新情報作成部 121 はライトデータ受領部 1081 へ完了を報告する (ステップ 1206)。

【0087】

ここまでの処理によって、正ストレージ装置 1000 は、計算機 200 から受領したライト要求 300 に対して、受領した順番にシーケンシャルな番号を付与した更新情報 1300 を作成し、更新情報記憶部 1094 に格納できる。よって、以後の処理において、ライト要求 300 を受領したタイミングとは非同期に、更新情報記憶部 1094 に格納されている更新情報 1300 を副ストレージ装置 2000 へ転送し、副ボリュームにライトデータを書き込むことによって、非同期リモートコピーを実現することができる。

【0088】

図 17 には、更新情報作成部 121 が、正ストレージ装置 1000 の更新情報記憶部 1094 に格納した更新情報 1300 を副ストレージ装置 2000 へ送信するデータ転送部 122 の処理の一例を示す。データ転送部 122 は、更新情報作成部 121 によって更新情報 1300 が正ストレージ装置 1000 の更新情報記憶部 1094 に格納されるまで待つ (ステップ 1301)。

なお、データ転送部 122 は更新情報 1300 が更新情報記憶部 1094 に格納される度に以降の処理を実行するのではなく、ある容量以上の更新情報 1300 が更新情報記憶部 1094 に格納された時点で以降の処理を実行する等、更新情報 1300 が更新情報記憶部 1094 に格納されるタイミングとは異なるタイミングで以降の処理を実行してもよい。

【0089】

次に、データ転送部 122 は、正ストレージ装置 1000 の更新情報記憶部 1094 から、1つまたは複数個の更新情報 1300 を読み出し (ステップ 1302)、読み出した更新情報 1300 を副ストレージ装置 2000 へ送信する (ステップ 1303)。そして、データ転送部 122 はステップ 1301 に戻り、上述の処理を繰り返す。

【0090】

図 18 には、正ストレージ装置 1000 から更新情報 1300 を受領する副ストレージ装置 2000 の更新情報受領部 124 の処理の一例を示している。更新情報受領部 124 は、正ストレージ装置 1000 のデータ転送部 122 から更新情報 1300 を受け取る (ステップ 1801)。そして、更新情報受領部 124 は、受け取った 1つまたは複数の更新情報 1300 を、副ストレージ装置 2000 の更新情報記憶部 1094 へ格納し、処理を終了する (ステップ 1802)。

【0091】

図 19 は、副ストレージ装置 2000 の更新情報記憶部 1094 に格納されている更新情報 1300 の内容に基づき、ライトデータ 310 を副ボリュームへ書き込むリストア部 125 の処理の一例を示す図である。なお、図中には記載されていないが、リストア部はリストア済み SEQ# と呼ばれる、過去にリストアが完了した更新情報 1300 の最新の SEQ# を保持している。

【0092】

リストア部 125 は、副ストレージ装置 2000 の更新情報記憶部 1094 から、1つまたは複数個の更新情報 1300 を読み出す (ステップ 2301)。次に、リストア部 125 は、読み出した更新情報 1300 を更新情報 1300 が含む SEQ# を用いてソートし、さらに、リストア済み SEQ# から連続している SEQ# のうち、最新の値を求め、それ以下の SEQ# を持つ更新情報 1300 を決定する (ステップ 2302)。そして、ステップ 2302 で決定した更新情報 1300 を、ソートした順番に、副ボリュームに書き込み、リストア済み SEQ# をリストアが完了した SEQ# に更新する (ステップ 2303)。

【0093】

このようにして、正ストレージ装置 1000 が計算機 200 から受領したライト要求 300 の順序を保証しながら、ライトデータを副ボリュームに書き込むことが実現される。

10

20

30

40

50

なお、高速化の一環として、ステップ 2303 の書き込み処理を並列に行っても良い。ただし、この場合は、ステップ 2302 で決定した SEQ # までの書き込みが全て終わるまで次の書き込みを禁止することと、必ず決められた SEQ # までには書き込みを終わるようにして、ライトの順序を守る。

【0094】

リストア部 125 は、リストア済み SEQ # を正ストレージ装置 1000 のページ部 123 へ通知する (ステップ 2304)。最後に、リストア部 125 は、正ストレージ装置 1000 のページ部 123 へ通知した SEQ # 以下の SEQ # を持つ更新情報 1300 を副ストレージ装置 2000 の更新情報記憶部 1094 から削除する (ステップ 2305)。そして、ステップ 2301 へ戻る。なお、更新情報 1300 の削除は実際にデータを削除しなくてもよく、制御情報の操作で仮想的に削除したものとできればよい。

10

【0095】

なお、リストア部 125 は、ステップ 2301 からステップ 2305 の処理を、常に繰り返し実行しているとしたが、更新情報記憶部 1094 に更新情報 1300 が格納されていない場合には、処理を終了してもよい。さらに、リストア部 125 はリモートコピー部 1052 の他の部から必要に応じて起動されることも可能である。

【0096】

図 20 は、副ストレージ装置 2000 のリストア部 125 から、SEQ # を受け取り、正ストレージ装置 1000 の更新情報記憶部 1094 から、不要となった更新情報 1300 を削除するページ部 123 の処理の一例を示す図である。ページ部 123 は、副ストレージ装置 2000 のリストア部 125 からのメッセージを受ける (ステップ 1901)。次に、ページ部 123 は、副ストレージ装置 2000 のリストア部 125 から受領したメッセージが含んでいる SEQ # 以下の SEQ # を持つ更新情報 1300 を正ストレージ装置 1000 の更新情報記憶部 1094 から削除する (ステップ 1902)。なお、更新情報 1300 の削除は実際にデータを削除しなくてもよく、制御情報の操作で仮想的に削除したものとできればよい。

20

【0097】

以上のように、計算機 200 が正ストレージ装置 1000 の正ボリュームに書き込むライトデータ 310 を、ライト要求 300 とは非同期に副ストレージ装置 2000 に転送し、副ボリュームに SEQ # 順に書き込むことによって、正ストレージ装置 1000 と副ストレージ装置 2000 の間で書き込み順序を保証した非同期リモートコピーを実現することができる。

30

【0098】

(2) リモートコピーペアをサスペンドする処理、および、ペア状態が S u s p e n d 状態である場合のライト要求の処理

リモートコピーペアをサスペンド指示とは、リモートコピーペアの正ボリュームから、副ボリュームへのデータコピーを一旦停止するための指示であり、S u s p e n d 状態とはコピーが停止した状態を示す。このサスペンド指示は、計算機 200 から正ストレージ装置 1000 へ、正ボリューム ID、副ボリューム ID、正ストレージ装置 ID、副ストレージ装置 ID を指定して発行する。

40

【0099】

図 21 には、計算機 200 からサスペンド指示を受領し、正ストレージ装置 1000 においてサスペンド処理を実行するサスペンド部 127 の処理の一例を示している。正ストレージ装置 1000 で、行う必要がある処理は以下である。

1) サスペンド対象のリモートコピーペアのペア状態を S u s p e n d 状態へと変更する。

2) 正ストレージ装置 1000 の更新情報記憶部 1094 に蓄積した更新情報 1300 を全て転送し、転送が完了したことを副ストレージ装置 2000 へ知らせる。

3) S u s p e n d 状態となったリモートコピーペアの正ボリュームに対するライト要求 300 の処理を変更する。具体的には、ライトアドレス 320 の記録を開始するように

50

変更する。

【0100】

以下に、処理ステップを説明する。サスペンド部127は、計算機200からサスペンド指示を受領する(ステップ2001)と、ペア情報テーブル1091へアクセスして、サスペンド対象であるリモートコピーペアのペア状態をSuspend状態へと変更する(ステップ2002)。

【0101】

この時点で、計算機200からサスペンド対象のリモートコピーペアの正ボリュームに対するライト要求300に対応する更新情報1300は、作成されなくなり、代わりにライト要求300のライトアドレス320が差分管理表1093に記録され始める。ライトアドレス320を差分管理表1093に記録する処理の詳細は後述する。 10

【0102】

次に、サスペンド部127は、SEQ#情報1095から現在のSEQ#を得て、次の更新情報1300に付与するSEQ#の値を用意しておくために、SEQ#情報1095に記録されているSEQ#に1を加える(ステップ2003)。

【0103】

最後に、サスペンド部127は、更新情報1300を作成して、更新情報記憶部1094に格納する(ステップ2004)。

【0104】

ステップ2304で作成した更新情報1300は、当該更新情報1300が、正ストレージ装置1000に滞留する最後の更新情報1300であることを、副ストレージ装置2000へ知らせるための特殊な更新情報1300である。この更新情報1300を、更新情報(サスペンド)1300と表現する。他の更新情報1300と区別するために、更新情報1300中のサスペンドフラグの値を「on」とする。 20

【0105】

更新情報1300に格納する値に関して述べる。SEQ#は、ステップ2003で得たSEQ#を格納し、ライトアドレス、ライトデータ、データ長には何も格納しない。副ボリュームID、および、副ストレージ装置IDは、サスペンド部127が計算機200からパラメータとして受け取る副ボリュームID、および、副ストレージ装置IDを格納する。コピー完了フラグは「off」とし、サスペンドフラグを「on」とする。 30

【0106】

次に、サスペンド部127が、ステップ2002でペア状態をSuspend状態としたリモートコピーペアの正ボリュームに対する計算機200からのライト要求300の処理に関して説明する。計算機200が正ボリュームに対して発行した、ライト要求300は、ライトデータ受領部1081によってDuplex状態の場合と同様に処理され、図15のステップ804で更新情報作成部121が呼び出される。

【0107】

図16に説明を移すと、更新情報作成部121の処理では、ステップ1201でリモートコピーペアのペア状態をチェックした結果、Suspend状態であると判定される。よって、更新情報作成部121は、ライト要求300に含まれるライトアドレス320と同一の値を持つレコード800を差分管理表1093から探し、当該レコード800の差分ビットを「on」へ変更する(ステップ1205)。そして、更新情報作成部121は、ライトデータ受領部1081へ完了メッセージを送り、処理を終了する(ステップ1206)。すなわち、更新情報1300を作成せず、代わりに、ライト要求300のライトアドレス320を差分管理表1093に記録する。 40

【0108】

リモートコピーペアの正ボリュームと副ボリュームを、再び同期状態にするときには、差分管理表1093を利用して、正ボリュームと副ボリュームの間で格納されているデータが異なる領域についてのみコピーすることが可能となる(差分コピー)。

【0109】

次に副ストレージ装置 2000 でのサスペンド処理を説明する。副ストレージ装置 2000 で行う必要がある処理は以下である。(1) サスペンド対象のリモートコピーペアのペア状態を Suspend 状態へと変更する

図 22 には、更新情報(サスペンド) 1300 を検出し、リモートコピーペアのペア状態を Suspend 状態へと変更するリストア部 125 の処理の一例を示している。リストア部 125 は、Duplex 状態と同じく、更新情報記憶部 1094 から更新情報 1300 を取り出して副ボリュームに書き込む処理を行っているが、書き込みを行う時にあわせて更新情報(サスペンド) 1300 を書き込み対象としたかどうかを確認する。

【0110】

もし、サスペンドフラグの値が「on」である更新情報(サスペンド) 1300 を見つけると(ステップ 1501)、ペア情報テーブル 1091 へアクセスして、サスペンド対象となっているリモートコピーペアのペア状態を Suspend 状態へ変更する(ステップ 1502)。そして、正ストレージ装置 1000 の更新情報記憶部 1094 から不要となった更新情報 1300 を削除するために、更新情報(サスペンド) 1300 の SEQ# を、正ストレージ装置 1000 のページ部 123 へ通知する(ステップ 1503)。最後に、副ストレージ装置 2000 の更新情報記憶部 1094 から、更新情報(サスペンド) 1300 が有する SEQ# 以下の SEQ# を有する更新情報 1300 を削除する(ステップ 1504)。

【0111】

以上のように、正ストレージ装置 1000 のサスペンド部 127 と副ストレージ装置 2000 のリストア部 125 の連携によってリモートコピーペアのサスペンドは実現される。なお、上記説明では、更新情報(サスペンド) 1300 をペア状態の遷移のための通知情報と、サスペンド以前の更新情報 1300 の転送が完了したことを示す情報の両方の意味を持っていた。しかし、ペア状態遷移の通知情報は別な方法で副ストレージ装置 2000 に通知してもよい。

【0112】

(3) リモートコピーペアをリシンクする処理、および、ペア状態が Duplex - Pending 状態である場合のライト要求の処理

リモートコピーペアのリシンク指示とは、Suspend 状態にあるリモートコピーペアの正ボリュームと副ボリュームに格納されているデータを、再び同期状態とするための指示である。このリシンク指示は、計算機 200 から正ストレージ装置 1000 へ、正ボリューム ID、副ボリューム ID、正ストレージ装置 ID、副ストレージ装置 ID を指定して発行する。

【0113】

このリシンク指示を受けた後は、正ストレージ装置 1000 と副ストレージ装置 2000 は、計算機からのライト順序とは無関係な順序で差分管理表 1093 に記録されたアドレスに格納されているデータを正ボリュームから副ボリュームへコピーするため、副ボリュームのライト順序は維持されない。そのため、差分コピーが実行中の本状態を Duplex - Pending 状態として計算機 200 に情報提供している。

【0114】

図 23 には、計算機 200 からリシンク指示を受領し、正ストレージ装置 1000 においてリシンク処理を実行するリシンク部 128 の処理の一例を示している。正ストレージ装置 1000 で、行う必要がある処理は以下である。

(1) リシンク指示を受領すると、リシンク対象のリモートコピーペアのペア状態を Duplex - Pending とし、リシンク動作が正しく開始したことを利用者に示す。

(2) 差分管理表 1093 に記録されている情報を元に、正ボリュームと副ボリュームの間で格納されているデータが異なる領域についてのみデータをコピーする。

(3) Duplex - Pending 状態となったリモートコピーペアの正ボリュームに対するライト要求の処理を変更する。具体的には、ライト要求 300 のライトアドレス 320 が、差分コピーが完了している領域の場合と、差分コピーが未完了の領域の場合とで

異なる処理を実行する。詳細は後述する。

(4) 差分コピーが完了すると、副ストレージ装置 2000 へ、差分コピー完了を通知し、リシンク対象のリモートコピーペアのペア状態を Duplex 状態へと変更する。

【0115】

以下に、処理ステップを説明する。リシンク部 128 は、計算機 200 からリシンク指示を受領する (ステップ 2101) と、ペア情報テーブル 1091 へアクセスして、リシンク対象であるリモートコピーペアのペア状態を Duplex - Pending 状態へと変更する (ステップ 2102)。

【0116】

次に、リシンク部 128 は、副ストレージ装置 2000 のペア状態テーブル 1091 のペア状態を変更するために、副ストレージ装置 2000 へリシンク指示を通知する (ステップ 2103)。最後に、リシンク部 128 は、差分コピーを実行する差分コピー部 129 を呼び出し、処理を終了する。 (ステップ 2104)

【0117】

では、次に、正ストレージ装置 1000 のリシンク部 128 のステップ 2104 から呼び出される差分コピー部 129 の処理を説明する。図 26 には、差分管理表 1093 を利用して差分コピーを実施する差分コピー部 129 の処理の一例を示している。

最初に、差分コピー部 129 は、差分管理表 1093 の先頭 (ボリュームアドレス「0」を持つレコード 800 の位置) を示すようにコピー済みボリュームアドレスの値を「0」とする (ステップ 1101)。次に、差分コピー部 129 は、コピー済みボリュームアドレスが示しているレコード 800 の差分ビットが、「on」であるか「off」であるかを調べる (ステップ 1102)。

【0118】

差分ビットの値が「on」の場合、レコード 800 のボリュームアドレスが示す領域に関して、正ボリュームと副ボリュームに格納されているデータが異なっている。差分コピー部 129 は、SEQ# 情報 1095 から現在の SEQ# を取得して、SEQ# 情報 1095 に格納されている SEQ# に 1 を加える (ステップ 1103)。

【0119】

そして、差分コピー部 129 は、レコード 800 のボリュームアドレスが示す領域に格納されているデータと、前ステップ 1103 で得た SEQ# を有する更新情報 1300 を作成し、更新情報記憶部に格納する (ステップ 1104)。

次に、差分コピー部 129 は、現在コピー済みボリュームアドレスが示しているレコード 800 の差分ビットの値を「on」から「off」へ変更する (ステップ 1105)。

差分ビットの値が「off」の場合には、レコード 800 のボリュームアドレスが示す領域に関して、正ボリュームと副ボリュームに格納されているデータは既に同一であるため、ステップ 1103、1104、1105 をスキップして、ステップ 1106 以降を実行する。

【0120】

差分コピー部 129 は、コピー済みボリュームアドレスの示すアドレスが差分管理表 1093 の末尾であるかをチェックする (ステップ 1106)。末尾でない場合には、差分コピー部 129 は、コピー済みボリュームアドレスに 1 を加え、ステップ 1102 へ戻る (ステップ 1107)。ステップ 1101 から 1107 の処理によって、差分管理表 1093 のレコード 800 を全スキャンし、差分ビットが「on」となっているアドレスに格納されているデータを有する更新情報 1300 を作成することができる。よって、当該更新情報 1300 のライトデータを副ボリュームに書き込むことで、正ボリュームと副ボリュームに格納されるデータを同期状態にすることができる。

【0121】

引き続き、ステップ 1108 以降の差分コピーの完了を副ストレージ装置 2000 に通知するための処理と、正ストレージ装置 1000 のペア情報テーブル 1091 を更新する処理について説明する。差分コピー部 129 は、正ストレージ装置 1000 のペア情報テ

ープル1091へアクセスし、差分コピーを行っているリモートコピーペアのペア状態を、Duplex状態へ変更する(ステップ1108)。次に、ステップ1103と同じく、差分コピー部129は、SEQ#情報1095から現在のSEQ#を取得して、次の更新情報1300に付与するSEQ#の値を用意しておくために、SEQ#情報1095に記録されているSEQ#に1を加える(ステップ1109)。

そして、差分コピー部129は、更新情報1300を作成し、作成した更新情報1300を正ストレージ装置1000の更新情報記憶部1094へ格納する(ステップ1110)。この更新情報1300は、差分コピーの完了を副ストレージ装置2000へ知らせるための特殊な更新情報1300である。この更新情報1300を、更新情報(コピー完了)1300と表現する。他の更新情報1300と区別するために、更新情報1300中のコピー完了フラグの値を「on」とする。 10

【0122】

更新情報1300に格納する値に関して述べる。更新情報1300内のSEQ#は、ステップ1109で得た値を格納し、ライトアドレス、ライトデータ、データ長には何も格納しない。サスペンドフラグは「off」とする。コピー完了フラグは「on」とする。

【0123】

次に、Duplex-Pending状態中のリモートコピーペアの正ボリュームに対する計算機200のライト要求300の処理に関して述べる。図15で示したとおり、計算機200が正ボリュームに対して発行したライト要求300は、ライトデータ受領部1081によってDuplex状態の場合と同様に処理され、ステップ804で更新情報作成部121が呼び出される。 20

【0124】

次に、図16に説明を移すと、更新情報作成部121の処理では、ステップ1201でリモートコピーペアのペア状態をチェックした結果、Duplex-Pending状態であると判定される。

【0125】

その場合、更新情報作成部121は、ライト要求300のライトアドレス320が示す領域が、差分コピーが既に完了しているか否かを判断する(ステップ1202)。ステップ1202の判定は、更新情報作成部121が、差分管理表1093のコピー済みボリュームアドレスとライト要求300のライトアドレス320を比較することにより実現できる。 30

【0126】

ライト要求300のライトアドレス320が示す領域が、差分コピーが未完了の領域である場合には、更新情報作成部121はSuspend状態の場合と同様、ステップ1205及びステップ1206の処理を実行する。差分コピー未完了の領域である場合には、差分管理表1093に差分を記録しておけば、その後、差分コピー部129によって正ボリュームから副ボリュームへコピーされるためである。

【0127】

一方、ライト要求300のライトアドレス320が示す領域が、差分コピーが既に完了している領域である場合には、更新情報作成部121はステップ1203以降の処理を実行する。すなわち、更新情報作成部121が、更新情報1300を作成し、副ストレージ装置2000へライトデータ310を転送する。 40

【0128】

次に副ストレージ装置2000での処理を説明する。副ストレージ装置2000で行う必要がある処理は以下である。

- (1) 正ストレージ装置1000から、リシンク指示の通知を受領して、リシンク対象のリモートコピーペアのペア状態をDuplex-Pendingへと変更する。
- (2) 正ストレージ装置1000から受領する更新情報1300のライトデータを副ボリュームに書き込む。
- (3) 差分コピー完了を検出して、リシンク対象のリモートコピーペアのペア状態をDu 50

plexへと変更する。

【0129】

図24には、正ストレージ装置1000のリシンク部128のステップ2103からリモートコピーペアのリシンク指示の通知を受領する副ストレージ装置2000のリシンク部128の処理の一例を示している。副ストレージ装置2000のリシンク部128は、正ストレージ装置1000からリシンク指示の通知を受領する(ステップ2201)。そして、副ストレージ装置2000のペア情報テーブル1091へアクセスして、リシンク対象であるリモートコピーペアのペア状態をDuplex-Pending状態へと変更する(ステップ2202)。

【0130】

これにより、副ストレージ装置2000におけるリモートコピーペアのペア状態もDuplex-Pendingとなり、差分コピーが正しく開始されていることをユーザに示すことができる。

【0131】

図25には、更新情報(コピー完了)1300を検出し、リモートコピーペアのペア状態をDuplex状態へと変更するリストア部125の処理の一例を示している。リストア部125は、コピー完了フラグの値が「on」である更新情報1300を見つけ(ステップ1601)、ペア情報テーブル1091へアクセスして、リシンク対象となっているリモートコピーペアのペア状態をDuplex状態へ変更する(ステップ1602)。以上のようにして、正ストレージ装置1000のリシンク部127、差分コピー部129、副ストレージ装置2000のリシンク部127の連携によってリモートコピーペアのリシンクは実現される。

【0132】

最後に、電源操作部1083の処理の内容を述べておき、その後、本発明の実施形態の処理内容を述べる。図27には、電源操作部1083の処理の一例を説明する。これは、パラメータとしてボリュームIDと要求内容(電源ON、または、電源OFF)を受領し、要求内容に応じてボリューム1070が属するパリティグループ1120を起動、停止する。

【0133】

電源操作部1083は、電源操作要求を受ける(ステップ1401)。次に、電源操作部1083は、ボリューム管理表500を参照することで、パラメータとして受け取ったボリュームIDが示すボリュームが属しているパリティグループIDを得る(ステップ1402)。そして、電源操作部1083は、パラメータとして渡される要求内容をチェックすることで、電源ONの要求であるか、電源OFFの要求であるかを判定する(ステップ1403)。

【0134】

電源ONの要求である場合には、電源操作部1083は、ボリューム管理表500へアクセスして、パラメータとして渡されるボリュームIDを持つレコード501を見つける。そして、当該レコード501の電源停止可能フラグを「false」へ変更する(ステップ1404)。これによって、ボリューム1070が、ステップ1402で決定したパリティグループ1120を利用中であることを宣言する。

【0135】

次に、電源操作部1083は、パリティグループ表510へアクセスし、ステップ1402で決定したパリティグループ1120の電源状態を「ON」へと変更し、電源制御モジュール1140へ当該パリティグループ1120内の全HDD1060の電源をONにするように指示する(ステップ1405)。なお、既に電源状態がONである場合は、ステップ1405はスキップする。最後に、電源操作の要求元へ電源操作完了を報告する(ステップ1410)。

【0136】

次に、電源OFFの場合を説明する。まず、電源操作部1083は、ボリューム管理表

10

20

30

40

50

500へアクセスして、パラメータとして渡されるボリュームIDを持つレコード501を見つける。そして、当該レコード501の電源停止可能フラグを「true」へ変更する(ステップ1406)。これによって、ボリューム1070にとっては、当該ボリューム1070が属するパリティグループ1120が停止されてもよいことを宣言する。

【0137】

次に、電源操作部1083は、再度ボリューム管理表500へアクセスし、ステップ1402で決定したパリティグループ1120に属する全てのボリュームを見つける(ステップ1407)。そして、電源操作部1083は、前ステップ1407で見つけた全てのボリュームの電源停止可能フラグを参照し、全てのボリュームの電源停止可能フラグが「true」であるかどうかをチェックする(ステップ1408)。

10

【0138】

全てのボリュームの電源停止可能フラグが「true」の場合には、パリティグループ1120に含まれる全てのボリューム1070は電源を停止してもよいことを意味しているため、電源操作部1083は、パリティグループ表510へアクセスし、ステップ1402で決定したパリティグループ1120の電源状態を「OFF」へと変更し、電源制御モジュール1140へ当該パリティグループ1120内のHDD1060の電源をOFFにするように指示する(ステップ1409)。

【0139】

次に、電源停止可能フラグに、1つ以上「false」がある場合には、パリティグループ1120内のHDD1060を利用しているボリューム1070が存在することを意味しているため、ステップ1409をスキップして、電源操作部1083は、電源操作の要求元へ電源操作完了を報告する(ステップ1410)。

20

【0140】

以後、これまでに説明してきたリモートコピーと電源制御を連携させる処理について説明する。リモートコピーのサスペンド指示と連携した電源制御に関して説明する。本連携制御では、副ストレージ装置2000は、計算機200から発行されるリモートコピーペアのサスペンド指示にあわせてリモートコピーペアの副ボリュームを構成するHDD1060を可能であるならば電源OFFとする。

【0141】

これは、正ストレージ装置1000のサスペンド部127と副ストレージ装置2000のサスペンド部127の連携によって実現される。まず、図28に正ストレージ装置1000のサスペンド部127の処理の一例を示している。図21で説明した電源制御と連携しないサスペンド部127との相違点は、ステップ2001の直後に、副ストレージ装置2000のサスペンド部127に、サスペンド指示を受領したことを通知するステップ2701を追加している点である。

30

【0142】

次に、図29に副ストレージ装置2000のサスペンド部127の処理の一例を示している。以下に、各ステップの詳細を説明する。サスペンド部127は、正ストレージ装置のサスペンド部127から、サスペンド指示の通知を受領する(ステップ1701)。

【0143】

次に、サスペンド部127は、サスペンド対象となっているリモートコピーペアの副ボリュームを構成するHDD1060の電源操作が可能であるか否かをチェックする(ステップ1702)。これは、当該副ボリュームに関するレコード501をボリューム管理表500から探し、当該レコード501のMAIDフラグを参照することでチェックする。MAIDフラグの値が「true」の場合には、ボリュームを構成するHDD1060の電源制御は可能であると判断する。

40

【0144】

ここで、サスペンド対象となっているリモートコピーペアの副ボリュームを構成するHDD1060の電源制御が不可能である場合には、処理を終了して、図19に示しているリストア部125に処理を任せる。

50

【0145】

サスペンド対象となっているリモートコピーペアの副ボリュームを構成するHDD1060の電源制御が可能である場合には、サスペンド部127は、正ストレージ装置、および、副ストレージ装置2000の更新情報記憶部1094に滞留する全ての更新情報1300のライトデータが副ボリュームに書き込まれるまで待つ(ステップ1703)。これは、副ストレージ装置2000のリストア部125がサスペンド対象のリモートコピーペアのペア状態をSuspendに変更すると、全ての更新情報1300が副ボリュームに書き込まれたと判断する。

【0146】

最後に、サスペンド部127は、サスペンド対象となっているリモートコピーペアの副ボリュームが属しているパリティグループ1120の停止を電源操作部1083へ要求する(ステップ1704)。

【0147】

以上のようにして、計算機200から正ストレージ装置1000がリモートコピーペアのサスペンド指示を受領したときには、サスペンド要求を受領する前に正ストレージ装置1000が計算機200から受領したライト要求300のライトデータ310を全て、副ストレージ装置2000の副ボリュームに書き込んだ直後、可能であるならば副ボリュームが属するパリティグループ1120内のHDD1060の電源をOFFするように制御することができる。よって、HDD1060の長寿命化や低電力化が期待できる。

【0148】

ストレージ装置間のネットワーク帯域を考慮した電源制御に関して説明する。本連携制御では、計算機からリモートコピーペアのリシンク(または、ペア形成)指示を受領し、差分コピーを実施するとき、副ストレージ装置2000で起動するパリティグループ数を一定数に制御する。これは、正ストレージ装置1000のリシンク部128と副ストレージ装置2000のリシンク部128の連携によって実現される。

【0149】

正ストレージ装置1000のリシンク部128は、計算機200からリモートコピーペアのリシンク指示を受領すると、以下の処理を実行する。なお、以下に示している項目は、従来技術のリシンク部128が実行する処理と異なる項目のみを示している。

(1) 副ストレージ装置2000へリシンク指示を受領したことを通知するとき、差管理表1093に差分が記録されていないリモートコピーペア、すなわち、実際には正ボリュームと副ボリュームが既に同期状態であるリモートコピーペアを探し、当該リモートコピーペアの情報も副ストレージ装置へ通知する。

(2) 副ストレージ装置から、差分コピー開始可能を通知されると、差管理表1093に記録されている情報を元に、正ボリュームと副ボリュームの間で格納されているデータが異なる領域についてのみデータをコピーする。

【0150】

図30に正ストレージ装置1000のリシンク部128の処理の一例を示している。以下に、各ステップの詳細を説明する。リシンク部128は、計算機200から1つ以上のリモートコピーペアのリシンク指示を受領する(ステップ2401)。このとき、リシンク対象となっているリモートコピーペアの情報をパラメータとして受領する。リモートコピーペアの情報とは、正ボリュームID、副ボリュームID、正ストレージ装置ID、副ストレージ装置IDである。

【0151】

次に、リシンク部128は、正ストレージ装置1000のペア情報テーブル1091へアクセスし、リシンク対象となっているリモートコピーペアのペア状態をDuplex-Pending状態へと変更する(ステップ2402)。

【0152】

そして、リシンク部128は、リシンク対象となっているリモートコピーペアのうち、差管理表1093に差分が記録されていないリモートコピーペアを見つける。当該リモ

10

20

30

40

50

ートコピーペアは、既に正ボリュームと副ボリュームに格納されているデータは同期状態となっているため、ペア情報テーブル1091へアクセスし、ペア状態をDuplex状態へと変更する(ステップ2403)。

【0153】

続けて、リシンク部128は、リシンク指示を受領したことを、副ストレージ装置2000へ通知する。このとき、パラメータとして、前ステップ2403で見つけた差分が記録されていないリモートコピーペアの情報と、計算機200から受領したリシンク対象のリモートコピーペアの情報を渡す。(ステップ2404)。

【0154】

リシンク部128の以降の処理は、副ストレージ装置2000から応答を受け(ステップ2405)、応答内容をチェックする(ステップ2406)。副ストレージ装置2000から受け取る応答内容が「差分コピー完了通知」ならば、処理を終了する。副ストレージ装置2000からの応答内容が「差分コピー開始可能通知」ならば、当該通知と共に通知されるリモートコピーペアの差分コピーを差分コピー部129へ要求する(ステップ2407)。そして、ステップ2405へ戻り、「差分コピー完了通知」を受領するまでステップ2405からステップ2407を繰り返す。

【0155】

では、次に副ストレージ装置2000のリシンク部128に関して述べる。副ストレージ装置2000のリシンク部128は、正ストレージ装置1000のリシンク部128からリモートコピーペアのリシンク指示の通知を受領すると、以下の処理を実行する。なお、以下に示している項目は、従来技術のリシンク部128が実行する処理と異なる項目のみを示している。

(1) 正ストレージ装置1000の差分管理表1093に差分が記録されていないリモートコピーペアのペア状態をDuplex状態へと変更する。

(2) リシンク対象のリモートコピーペアのうち、副ボリュームを構成するHDD1060の電源制御が不可能である副ボリュームを探し、差分コピー開始可能を正ストレージ装置1000へ通知する。

(3) 目標電力量1097で起動可能な一定数のパリティグループ1120を起動し、当該パリティグループ1120に副ボリュームが属しているリモートコピーペアの差分コピー開始可能を正ストレージ装置1000へ通知する。さらに、差分コピー開始可能を通知したリモートコピーペアの差分コピー完了を待ち。当該リモートコピーペアの副ボリュームが属するパリティグループ1120を停止する。

【0156】

図31に副ストレージ装置2000のリシンク部128の処理の一例を示している。以下に、各ステップの詳細を説明する。リシンク部128は、正ストレージ装置1000から、リモートコピーペアのリシンク指示の通知を受領する(ステップ2501)。このとき、1つ以上のリモートコピーペアの情報と、1つ以上の差分が記録されていないリモートコピーペアの情報をパラメータとして受領する。

【0157】

次に、リシンク部128は、副ストレージ装置2000のペア情報テーブル1901へアクセスし、リシンク対象のリモートコピーペアのペア状態をDuplex-Pending状態へ、差分が記録されていないリモートコピーペアのペア状態をDuplex状態へと変更する(ステップ2502)。この段階で、正ストレージ装置1000の差分管理表1093に差分が記録されていないリモートコピーペアのリシンク処理は終了する。

【0158】

次に、リシンク部128は、パラメータとして受領した、リシンク対象のリモートコピーペアのうち、副ボリュームを構成するHDD1060の電源制御が可能であるか否かをチェックし、副ボリュームを構成するHDD1060の電源制御が不可能であるリモートコピーペアをリストアップする(ステップ2503)。

【0159】

10

20

30

40

50

リモートコピーペアの副ボリュームを構成するHDD1060の電源制御が不可能である場合には、副ボリュームを構成するHDD1060は常に電源ONであるため、リシンク部128は、正ストレージ装置1000のリシンク部128へ、ステップ2503でリストアップしたリモートコピーペアの差分コピー開始可能を通知する(ステップ2504)。上記通知によって、副ボリュームを構成するHDD1060の電源制御が不可能であるリモートコピーペアの差分コピーが正ストレージ装置1000の差分コピー部129によって開始される。

【0160】

副ストレージ装置2000のリシンク部128は、以降のステップで、差分が記録されており、かつ、副ボリュームを構成するHDD1060の電源制御が可能であるリモートコピーペアに関する処理を実行する。まず、リシンク部128は、パラメータとして受け取ったリモートコピーペアのうちで、差分コピーが未だに開始していないリモートコピーの副ボリュームが属しており、かつ、停止状態であるパリティグループ1120をリストアップする(ステップ2505)。次に、リシンク部128は、目標電力量1097で起動可能なパリティグループ1120を起動する。

具体的には、ステップ2505で得たパリティグループ1120のリストの先頭から順番にパリティグループ1120の使用電力量を合計していく。パリティグループ1120の使用電力量は、パリティグループ表510から取得し、その時点までの使用電力量の合計が目標電力量1097以下であるならば、当該パリティグループ1120の起動を電源操作部1083へ要求する。そして、使用電力量の合計が目標電力量1097を超えると、パリティグループ1120のリストを先頭から順にスキャンしていく処理を終了する(ステップ2506)。なお、使用電力量の計算をするに当たって、電源制御が不可能なパリティグループの使用電力量を加算してもよい。

【0161】

ステップ2506で起動を要求したパリティグループ1120の起動が完了すると、リシンク部128は、正ストレージ装置1000からパラメータとして渡されるリモートコピーペアのうち、差分コピーが未だに開始していないリモートコピーペアの副ボリュームに関するレコード501をボリューム管理表500から取得する。そして、レコード501のパリティグループIDを参照することで、副ボリュームが起動したパリティグループ1120に属するか否かを調べる。副ボリュームが起動したパリティグループ1120に属する場合には、レコード501の電源停止可能フラグを「false」へ変更する(ステップ2507)。これによって、当該副ボリュームが起動したパリティグループ1120を利用していることを宣言する。

【0162】

そして、リシンク部128は、ステップ2507で見つけたリモートコピーペアの差分コピー開始可能を、正ストレージ装置1000のリシンク部128へ通知する(ステップ2508)。このとき、ステップ2507で見つけたリモートコピーペアの情報をパラメータとして渡す。上記の通知によって、副ボリュームを構成するHDD1060の電源をONにしたリモートコピーペアの差分コピーが正ストレージ装置1000の差分コピー部129によって開始される。

【0163】

次に、リシンク部128は、差分コピー開始可能を通知したリモートコピーペアの差分コピーが完了するまで待つ。(ステップ2509)。これは、副ストレージ装置2000のリストア部125がリシンク対象のリモートコピーペアのペア状態をDuplexに変更すると、差分コピーが完了したと判断する。リシンク部128は、差分コピーが完了したリモートコピーペアの副ボリュームが属するパリティグループ1120の停止を電源操作部1083へ要求し、応答を待つ(ステップ2510)。

【0164】

電源操作部1083からの応答を受けると、正ストレージ装置1000のリシンク部128からパラメータとして受けた全てのリモートコピーペアの差分コピーが開始している

10

20

30

40

50

か否かをチェックする（ステップ2511）。正ストレージ装置から受け取ったリシンク対象となっているリモートコピーペアで差分コピーが未だに開始していないリモートコピーペアが存在する場合には、ステップ2505へ戻って、以降のステップを繰り返す。この場合、ステップ2510で、差分コピーが完了したリモートコピーペアの副ボリュームが属するパリティグループ1120を停止するため、差分コピー未開始のリモートコピーペアの副ボリュームが属する別のパリティグループ1120を、目標電力量1097を超えずに起動することが可能である。また、ステップ2511で、全ての差分コピーが開始している場合には、正ストレージ装置1000のリシンク部128へ差分コピー完了を通知して処理を終了する（ステップ2512）。

【0165】

以上のように、副ストレージ装置2000で、一度に起動するパリティグループ1120数を一定数に制御し、起動したパリティグループ1120に副ボリュームが属するリモートコピーペアの差分コピーを行う。よって、長時間を要する差分コピーを行う際に、常に全てのパリティグループ1120を起動しておく必要はなくなる。さらに、リシンク対象のリモートコピーペアのうちで、差分が記録されているリモートコピーペアの副ボリュームが属するパリティグループ1120のみを起動するように制御することができる。

【0166】

以上のことから、副ボリュームを構成するHDD1060の長寿命化や、低電力化を実現することができる。なお、本処理はリモートコピーペアを新たに形成する場合にも適用することができる。

【0167】

正ストレージ装置1000が計算機200からリモートコピーペアのペア形成指示を受領した場合には、正ストレージ装置1000は正ボリュームに格納される全てのデータを副ボリュームにコピーする（初期化コピー）。リモートコピーペアの形成処理は、以下の点においてリモートコピーペアのリシンク処理と異なる。

【0168】

(1) ペア情報テーブル1091には、新たに形成するリモートコピーに関するレコード700を追加する。(2) 正ボリュームに格納される全てのデータを副ボリュームにコピーすることで、正ボリュームと副ボリュームを同期状態にする。(ただし、先に断ったように、これも差分コピーと呼ぶ)

第2の発明の実施形態をリモートコピーペアの形成に適用する場合の処理を、リシンク処理の場合との違いを中心に説明する。図32には、正ストレージ装置1000で動作するペア形成部126の処理の一例を示している。

【0169】

まず、正ストレージ装置1000のペア形成部126が計算機200からリモートコピーペアのペア形成指示を受領する。このとき、正ストレージ装置1000のペア形成部126は、計算機200からパラメータとして、正ボリュームID、正ストレージ装置ID、副ボリュームID、副ストレージ装置ID、リモートコピータイプをパラメータとして受ける（ステップ3001）。

【0170】

ペア形成部126は、パラメータとして受け取ったリモートコピーペアの情報1つに対して、1つのレコード700を作成し、正ストレージ装置1000のペア情報テーブル1091に追加する（ステップ3002）。正ボリュームID、副ボリュームID、レプリケーションタイプはパラメータとして渡される値である。相手ストレージ装置IDは、副ストレージ装置IDとし、ペア状態は、「Duplex-Pending」とする。

【0171】

そして、副ストレージ装置2000のペア形成部126へペア形成指示を通知する。このとき、計算機200から受領したパラメータの全てを副ストレージ装置2000のペア形成部126に渡す（ステップ3003）。リモートコピーペアのペア形成の処理では、正ボリュームに格納されている全てのデータを副ボリュームへコピーする。そのため、リ

10

20

30

40

50

モートコピーペアのリシンクで説明した、リモートコピーペアに差分が記録されているかをチェックするための処理は不要である。

【0172】

次に、ペア形成部126は、リシンク部128のステップ2405と同様に、副ストレージ装置2000からの応答を受け(ステップ3004)、応答内容をチェックする(ステップ3005)。応答内容としては、リシンク部128の場合と同様に、「差分コピー完了通知」または、「差分コピー開始可能通知」のいずれかである。差分コピー完了通知を受領した場合には、ペア形成部126は、リシンク部128と同じく処理を終了する。

【0173】

差分コピー開始可能通知を受領した場合には、ペア形成部126は、副ストレージ装置2000から通知されるリモートコピーペアの差分管理表1093の差分ビットを全て「on」へ変更する(ステップ3006)。そして、ペア形成部は、リシンク部128と同じく、差分コピー部129を呼び出すことで差分コピーを開始する。ステップ3006で、差分管理表1093の差分ビットを全て「on」へ変更することによって、差分コピー部129は、正ボリュームに格納される全てのデータを副ボリュームへコピーすることとなる。

【0174】

図33に移って、副ストレージ装置2000で動作するペア形成部126の処理の一例を説明する。副ストレージ装置2000のペア形成部126は、正ストレージ装置1000のペア形成部126から、リモートコピーペアのペア形成指示の通知を受領する。このとき、パラメータとして、正ボリュームID、正ストレージ装置ID、副ボリュームID、副ストレージ装置ID、リモートコピータイプをパラメータとして受け取る(ステップ3101)。

【0175】

ペア形成部126は、パラメータとして受け取ったリモートコピーペアの情報1つに対して、1つのレコード700を作成し、副ストレージ装置2000のペア情報テーブル1091に追加する。正ボリュームID、副ボリュームID、レプリケーションタイプは正ストレージ装置1000から受領した値である。相手ストレージ装置IDは、正ストレージ装置1000から受領した正ストレージ装置IDとし、ペア状態は、「Duplex-Pending」とする(ステップ3102)。

【0176】

次に、ペア形成部126の以降のステップに関して述べる。ステップ2503、ステップ2504は、副ボリュームの電源制御が不可能であるリモートコピーペアを決定し、正ストレージ装置1000へ差分コピー開始可能を通知するためのステップである。さらに、ステップ2504からステップ2511は、目標電力量1097で起動可能な一定数のパリティグループ1120を起動し、当該パリティグループ1120に副ボリュームが属するリモートコピーペアの差分コピー開始可能を正ストレージ装置へ通知するためのステップである。これらのステップは、副ストレージ装置2000のリシンク部128で説明したステップ2503からステップ2511と同じ処理である。

【0177】

以上のように、正ストレージ装置1000のペア形成部126と副ストレージ装置2000のペア形成部126、電源操作部1083が連携することで、計算機200からペア形成が要求されたリモートコピーペアの差分コピーを実施し、正ボリュームと副ボリュームの内容を同期状態にすることができる。さらに、副ストレージ装置2000で一度に起動するパリティグループ1120数を一定数に制御する点などに関してはリシンク部128と同じである。よって、副ボリュームを構成するHDD1060の長寿命化や、低電力化を実現することができる。

【0178】

なお、本発明の実施形態では、起動するパリティグループ1120の数を決定するに当たって、目標電力量1097に加えて、正ストレージ装置1000と副ストレージ装置200

10

20

30

40

50

00間のネットワーク260の転送帯域を比較対象にすることも考えられる。即ち、図31のステップ2506にて起動を決定したパリティグループ1120をさらに以下の方法によって選別する。

【0179】

(A)ステップ2506で選択したパリティグループ1120のそれぞれの転送帯域幅を求める。この値はパリティグループ1120に含まれるHDD1060の数などで求めても良いし、管理者が手で入力を行っても良い。また、正ストレージ装置1000や副ストレージ装置2000内部で測定を行っても良い。

【0180】

(B)(A)にて求めた転送帯域幅を参照しながら、ネットワーク260の転送帯域を満たし、かつ最小の電力使用量となるようにパリティグループ1120を選択する。この選択はステップ2505にてリストアップしたパリティグループ1120の転送帯域幅を先頭から加算していき、その加算値がネットワーク260の転送帯域幅を超えた直後まで加算対象としたパリティグループ1120を選択すればよい。

【0181】

非同期リモートコピーの処理特性を利用した電源制御では、リモートコピー部1082が、更新情報1300を、更新情報記憶部1094に一定時間、または、一定量滞留させている間、すなわち、副ボリュームにアクセスが発生しない間は、副ボリュームを構成するHDD1060は電源をOFFとする。そして、リモートコピー部1082がライトデータを副ボリュームに書き込む直前に、副ボリュームが属するパリティグループ1120を起動し、ライトデータの書き込みが完了した直後にパリティグループ1120を停止することで、HDD1060の長寿命化、低電力化する。

【0182】

本実施形態では、副ストレージ装置2000のコピー電源連携部1084が、副ストレージ装置2000の更新情報記憶部1094に更新情報1300を滞留させておき、更新情報記憶部1094の使用率が、予め設定された更新情報量閾値1098を越えるのを契機として、ライトデータを副ボリュームへ書き込む場合を例として説明する。

【0183】

図34には、コピー電源連携部1084の処理の一例を示している。以下に、処理の詳細を述べる。

コピー電源連携部1084は、副ストレージ装置2000の更新情報記憶部1094の使用率と更新情報量閾値1098を比較する。更新情報記憶部1094の使用率が更新情報量閾値1098を超えていれば、以降のステップに進む。更新情報記憶部1094の使用率が更新情報量閾値1098を超えていなければ、以降のステップには進まず、本ステップを繰り返す(ステップ2801)。なお、更新情報記憶部1094の使用率が更新情報量閾値1098を超えていない場合に、一定時間処理を停止したあと、ステップ2801を再度実行するなどの効率化を行ってもよい。

【0184】

更新情報記憶部1094の使用率が更新情報閾値1098を超えた場合には、コピー電源連携部1084は、副ストレージ装置2000の更新情報記憶部1084に格納されている更新情報1300をSEQ#を用いてソートし、さらに、SEQ#が連続している部分(更新情報集合に含まれる更新情報1300をSEQ#順に並べた際に、SEQ#に抜けが生じない部分)に含まれる更新情報1300を決定する(ステップ2802)。

【0185】

そして、コピー電源連携部1084は、前ステップ2802で決定した更新情報1300のライトデータを書き込むべき副ボリュームを決定し、当該副ボリュームが属するパリティグループ1120をリストアップする。具体的には、更新情報1300に含まれる副ボリュームIDと同一の値を持つレコード501をボリューム管理表500から探し、そのパリティグループIDをリストアップする(ステップ2803)。

コピー電源連携部1084は、前ステップ2803でリストアップしたパリティグループ

ブ 1 1 2 0 の起動を電源操作部 1 0 8 3 へ要求し、電源操作部 1 0 8 3 からの応答を待つ (ステップ 2 8 0 4)。

【0186】

次に、コピー電源連携部 1 0 8 4 は、リストア部 1 2 5 を起動し、リストア部 1 2 5 がステップ 2 8 0 2 で決定した全ての更新情報 1 3 0 0 のライトデータを、副ボリュームへ書き込むのを待つ (ステップ 2 8 0 5)。これは、リストア部 1 2 5 が処理を終了することを監視することや、更新情報記憶部 1 0 9 4 の使用率を監視することで実現できるし、リストア部 1 2 5 に通知させることでも実現することもできる。

次に、コピー電源連携部 1 0 8 4 は、ステップ 2 8 0 4 で起動したパリティグループ 1 1 2 0 の停止を電源操作部 1 0 8 3 へ要求し、電源操作部 1 0 8 3 からの応答を待つ (ステップ 2 8 0 6)。

【0187】

以上のようにして、リモートコピー部 1 0 8 2 が、副ストレージ装置 2 0 0 0 の更新情報記憶部 1 0 9 4 に、一定量の更新情報 1 3 0 0 を滞留させておき、更新情報記憶部 1 0 9 4 の使用率が更新情報量閾値 1 0 9 8 を超えたときに副ボリュームに更新情報 1 3 0 0 のライトデータを書き込むように制御する場合にも、HDD 1 0 6 0 の起動を、実際にライトデータを副ボリュームに書き込むときのみに限定することができる。よって、HDD の長寿命化、および、ストレージ装置の低電力化を実現することができる。

【0188】

なお、第 3 の発明の実施形態の説明では、副ストレージ装置 2 0 0 0 の更新情報記憶部 1 0 9 4 の使用率に基づいて、ライトデータを副ボリュームに書き込むとした。しかし、更新情報記憶部 1 0 9 4 の使用率ではなく、時間に基づいて、ライトデータを副ボリュームに書き込むことも考えられる。このように、ライトデータを副ボリュームに書き込むタイミングは、本説明で用いた例に限定されない。

【0189】

なお、図には示していないが、正ストレージ装置 1 0 0 0 は電源制御可能な副ストレージ装置 2 0 0 0 に加えて電源制御しない通常のストレージ装置が副ストレージ装置 2 0 0 0 となる可能性もあるため、上記に説明した処理を実行する前に、副ストレージ装置 2 0 0 0 が電源制御可能かどうかを問い合わせ、その結果として電源制御と連携した処理を実行するかどうかを決定してもよい。

【0190】

また、本発明の実施形態では、正ボリュームと副ボリュームは別々なストレージ装置に含まれているが、同一のストレージ装置に含まれ、リモートコピーの代わりとしてローカルコピーに適用してもよい。さらに、本発明の実施形態は正ストレージ装置 1 0 0 0 が電源制御可能な場合も適用可能である。例えば、リシンク操作を受け付けた場合、上記の説明では副ストレージ装置 2 0 0 0 の電源制御のみ行っているが、それと合わせて (リモート、または、ローカル) コピーペアの正ボリュームを構成する HDD 1 0 6 0 の電源 ON を行ってもよい。

【実施例 2】

【0191】

実施例 1 では、リモートコピーペアの差分コピーを実行する場合に、副ストレージ装置 2 0 0 0 では、リモートコピー部 1 0 8 2 が一度に起動するパリティグループ 1 1 2 0 数を一定数に制御する方法を述べた。

【0192】

本実施例 2 では、障害発生時など緊急の場合に、目標電力量 1 0 9 7 を無視して、差分コピーの速度を優先する方法を述べる。

一般に、計算機 2 0 0 や正ストレージ装置 1 0 0 0 に障害が発生したとき、副ストレージ装置 2 0 0 0 に格納されているデータを利用して業務継続を行う。よって、計算機 2 0 0 や、正ストレージ装置 1 0 0 0 の一部に障害が発生している場合に、実施例 1 のように、一度に一定数のパリティグループのみを起動し、差分コピーを行う方法や、一定時間リ

10

20

30

40

50

ストア部 125 を停止させる方法は時間がかかる。

なお、ここでの正ストレージ装置の一部に障害とは、正ストレージ装置 1000 が副ストレージ装置 2000 にデータをコピーすることができる状態とする。

【0193】

本実施例 2 では、図 35 に示すように、障害発生時に業務を引き継ぐための計算機 400 を配置して、ネットワーク 250 を使用して副ストレージ装置 2000 と接続する。計算機 400 の構成は計算機 200 と同一である。なお、副ストレージ装置 2000 には、計算機 400 から I/O を受け付けるための I/F 1010 を設置するものとする。

【0194】

そして、差分コピー実施中に、計算機 200 や、正ストレージ装置の一部に障害が発生している場合に、計算機 400 が副ストレージ装置 2000 に、目標電力量 1097 を無視して差分コピーを実施するように指示する。以後、「高速差分コピー指示」と表現する。

【0195】

図 36 に、高速差分コピーの要求を受領した副ストレージ装置 2000 が実行する処理の一例を示している。

副ストレージ装置 2000 は、計算機 400 から高速差分コピー指示を受領する（ステップ 2601）。そして、副ストレージ装置 2000 は、副ストレージ装置 2000 のペア情報テーブル 1091 を参照し、ペア状態が Duplex - Pending 状態であるリモートコピーペアを探す（ステップ 2602）。

【0196】

次に、副ストレージ装置 2000 は、ステップ 2602 で見つけたリモートコピーペアの副ボリュームが属するパリティグループ 1120 をリストアップする。具体的には、副ボリューム ID と同一の値を持つレコード 501 をボリューム管理表 500 から探し、そのパリティグループ ID をリストアップする（ステップ 2603）。

【0197】

副ストレージ装置 2000 は、リストアップしたパリティグループ 1120 の起動を、電源操作部 1083 へ要求し（ステップ 2604）、電源操作部 1083 からの応答を待つ（ステップ 2605）。

そして、副ストレージ装置 2000 は、ステップ 2602 で見つけたリモートコピーペアの副ボリュームに関するレコード 501 をボリューム管理表 500 から探し、当該レコード 501 の電源停止可能フラグを「false」へ変更する。最後に、副ストレージ装置 2000 は、正ストレージ装置 1000 へ、ステップ 2602 で見つけたリモートコピーペアの差分コピー可能を通知する。

なお、差分コピー可能の通知を受領した正ストレージ装置 1000 の動作は、リシンク部 128、および、ペア形成部 126 で述べた。

【0198】

以上のようにして、計算機 200 や正ストレージ装置 1000 の一部に障害が発生した場合に、差分コピーを急ぐことによって、万一正ストレージ装置 1000 から副ストレージ装置 2000 へのデータコピーが不可能となった場合に失われるデータを最小限に抑えることができる。

【図面の簡単な説明】

【0199】

【図 1】本発明の実施例 1 である情報処理システムの構成、および、計算機の構成の一例を示す図である。

【図 2】正ストレージ装置の構成の一例を示す図である。

【図 3】副ストレージ装置の構成の一例を示す図である。

【図 4】ライト要求の一例を示す図である。

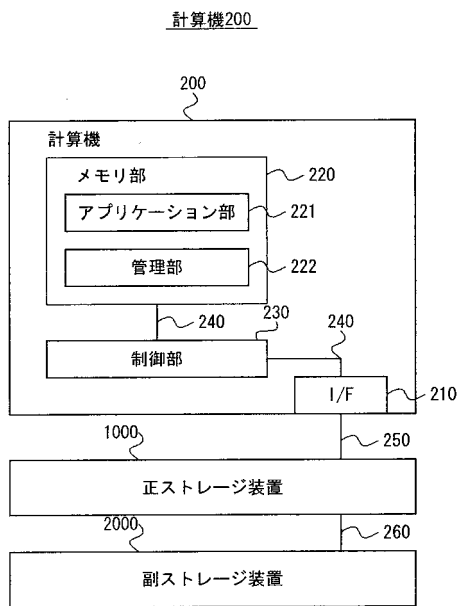
【図 5】ボリューム管理表の一例を示す図である。

【図 6】パリティグループ表の一例を示す図である。

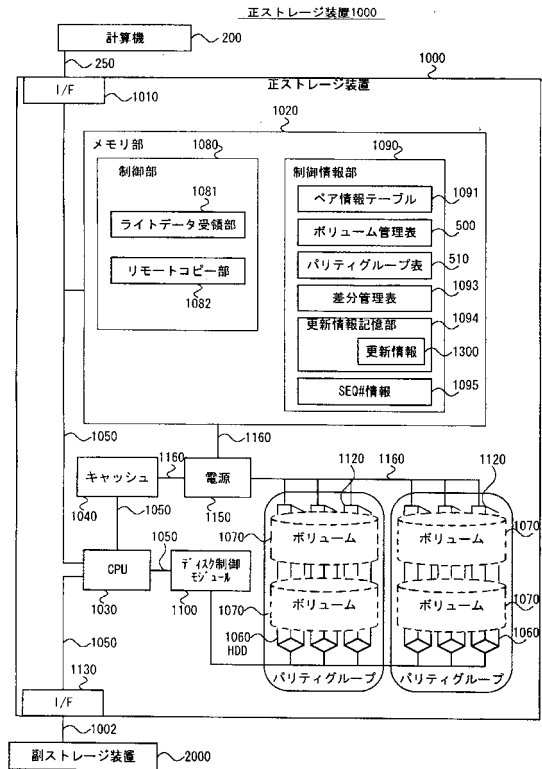
- 【図 7】リモートコピー部の構成の一例を示す図である。
- 【図 8】ペア情報テーブルの一例を示す図である。
- 【図 9】差分管理表の一例を示す図である。
- 【図 10】第 1 の本発明の実施形態の概要図である。
- 【図 11】第 2 の本発明の実施形態の概要図である。
- 【図 12】第 3 の本発明の実施形態の概要図である。
- 【図 13】SEQ # 情報の一例を示す図である。
- 【図 14】更新情報の一例を示す図である。
- 【図 15】ライトデータ受領部による処理の一例を示す図である。
- 【図 16】更新情報作成部による処理の一例を示す図である。 10
- 【図 17】データ転送部による処理の一例を示す図である。
- 【図 18】更新情報受領部による処理の一例を示す図である。
- 【図 19】リストア部による処理の一例を示す図である。
- 【図 20】パージ部による処理の一例を示す図である。
- 【図 21】正ストレージ装置のサスペンド部による処理の一例を示す図である。
- 【図 22】サスペンドを検出した場合のリストア部による処理の一例を示す図である。
- 【図 23】正ストレージ装置のリシンク部による処理の一例を示す図である。
- 【図 24】副ストレージ装置のリシンク部による処理の一例を示す図である。
- 【図 25】差分コピー完了を検出した場合のリストア部による処理の一例を示す図である 20
- 。
- 【図 26】差分コピー部による処理の一例を示す図である。
- 【図 27】電源操作部による処理の一例を示す図である。
- 【図 28】電源制御を考慮した正ストレージ装置のサスペンド部による処理の一例を示す図である。
- 【図 29】電源制御を考慮した副ストレージ装置のサスペンド部による処理の一例を示す図である。
- 【図 30】電源制御を考慮した正ストレージ装置のリシンク部による処理の一例を示す図である。
- 【図 31】電源制御を考慮した副ストレージ装置のリシンク部による処理の一例を示す図 30
- である。
- 【図 32】電源制御を考慮した正ストレージ装置のペア形成部による処理の一例を示す図である。
- 【図 33】電源制御を考慮した副ストレージ装置のペア形成部による処理の一例を示す図である。
- 【図 34】コピー電源制御連携部による処理の一例を示す図である。
- 【図 35】本発明の実施例 2 である情報処理システムの構成、および、計算機の構成の一例を示す図である。
- 【図 36】高速差分コピーの要求を受けた副ストレージ装置の処理の一例を示す図である 40
- 。
- 【符号の説明】 40
- 【0200】
- | | |
|------|--------------|
| 200 | 計算機 |
| 300 | ライト要求 |
| 400 | 計算機 |
| 500 | ボリューム管理表 |
| 510 | パリティグループ表 |
| 1000 | 正ストレージ装置 |
| 1060 | HDD |
| 1080 | 制御部 |
| 1081 | ライトデータ受領部 50 |

- 1 0 8 2 リモートコピー部
- 1 0 8 3 電源操作部
- 1 0 8 4 コピー電源連携部
- 1 0 9 0 制御情報部
- 1 0 9 1 ペア情報テーブル
- 1 0 9 3 差分管理表
- 1 0 9 4 更新情報記憶部
- 1 0 9 5 SEQ#情報
- 1 1 0 0 ディスク制御モジュール
- 1 1 2 0 パリティグループ
- 1 1 4 0 電源制御モジュール
- 1 1 5 0 電源
- 1 1 7 0 ボリューム
- 1 3 0 0 更新情報
- 2 0 0 0 副ストレージ装置

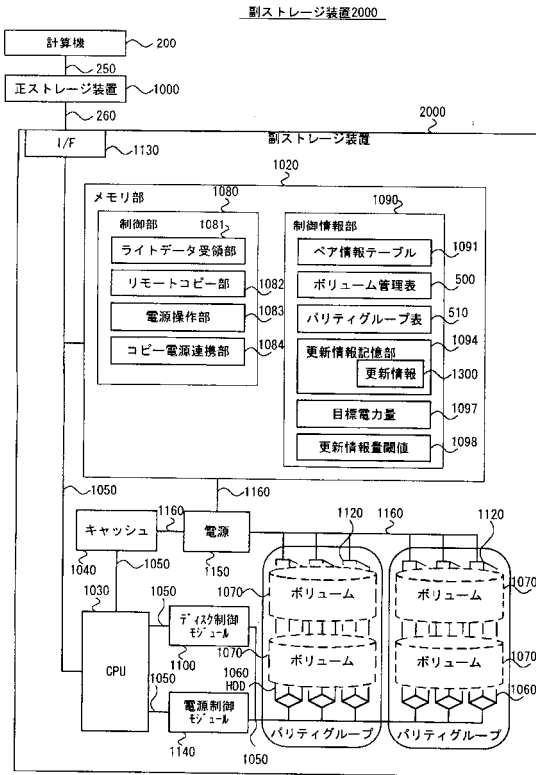
【図1】



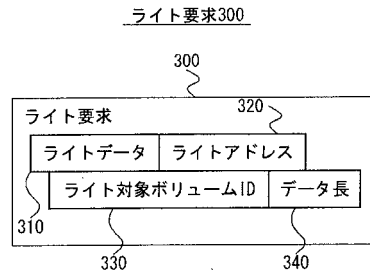
【図2】



【図3】



【図4】



【図5】

ボリューム管理表500

ボリュームID	容量(G Byte)	パリティグループID	MAIDフラグ	電源停止可能フラグ
ボリューム1	300	パリティグループ1	true	true
ボリューム2	300	パリティグループ1	true	false
ボリューム3	700	パリティグループ2	false	-
...

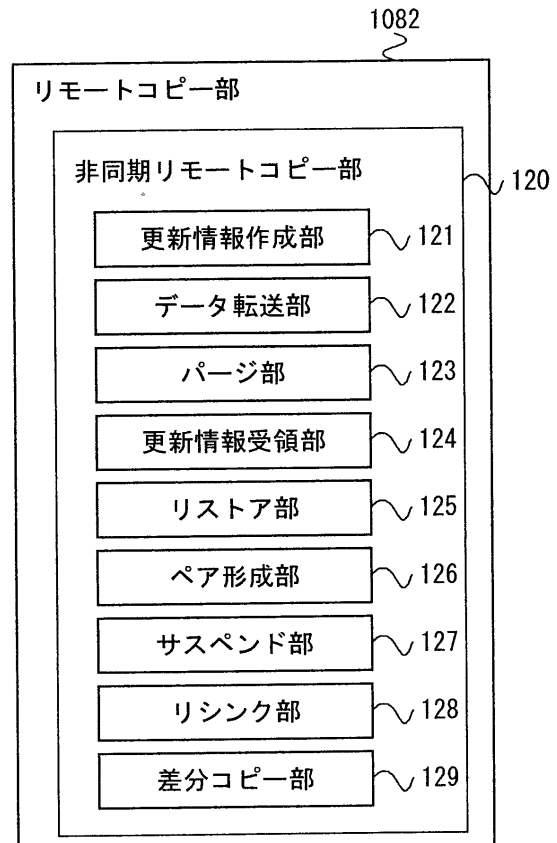
レコード501

【図6】

パリティグループ表510

パリティグループID	HDD-ID	RAIDレベル	電源状態	使用電力量
パリティグループ1	0	1	ON	30W
	1			
パリティグループ2	2	5	OFF	60W
	3			
	4			
パリティグループ3	6	1	-	30W
	7			
...

【図7】



【図8】

ペア情報テーブル1091

1091				
正ボリュームID	副ボリュームID	ペア状態	相手ストレージシステムID	リモートコピータイプ
ボリューム1	ボリューム2	Duplex	ストレージシステム1	非同期
ボリューム3	ボリューム4	Duplex	ストレージシステム2	非同期
...

コード700

【図9】

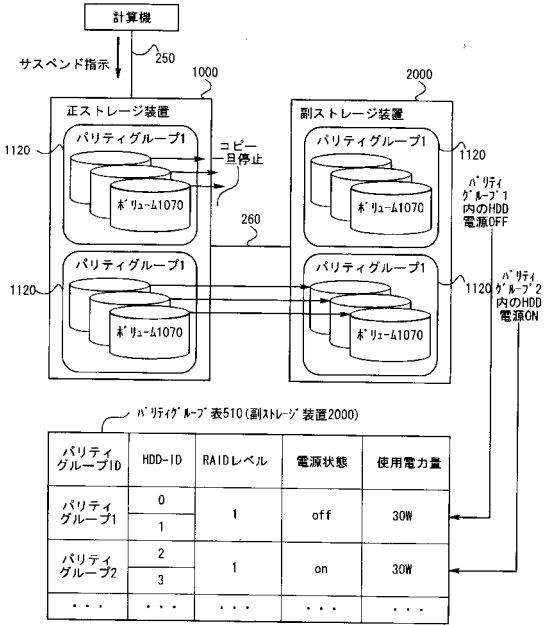
コピー済みボリュームアドレス

2

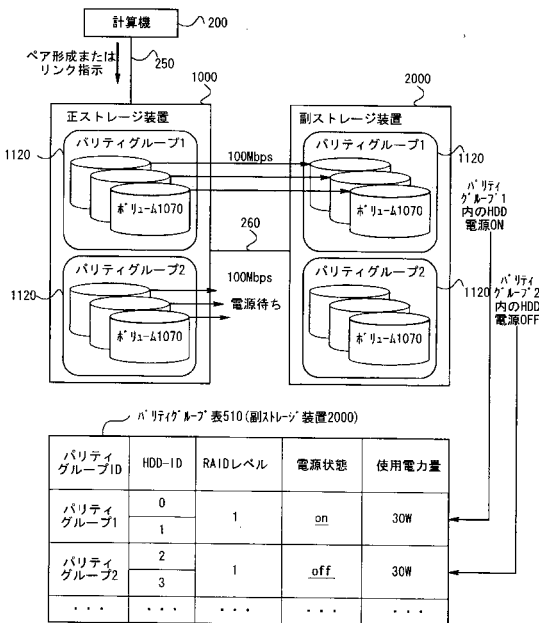
1093	
ボリュームアドレス	差分ビット
0	off
1	off
2	on
3	off
4	on
5	on
...	...

コード800

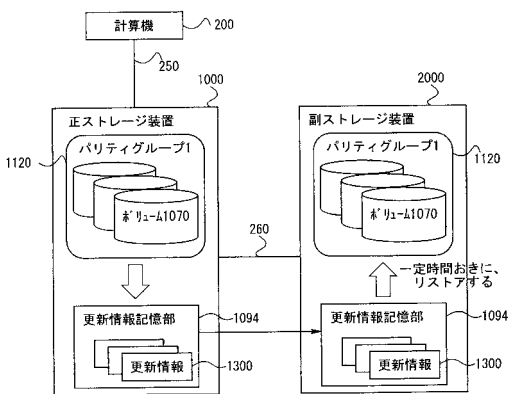
【図10】



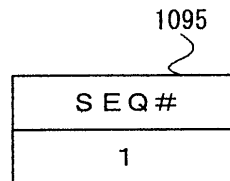
【図11】



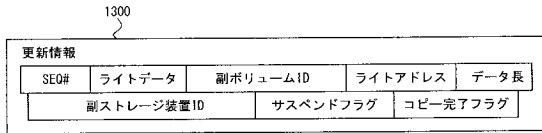
【図12】



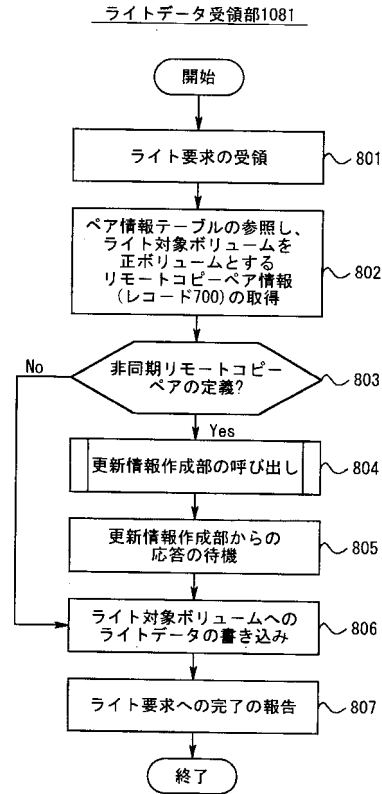
【図13】



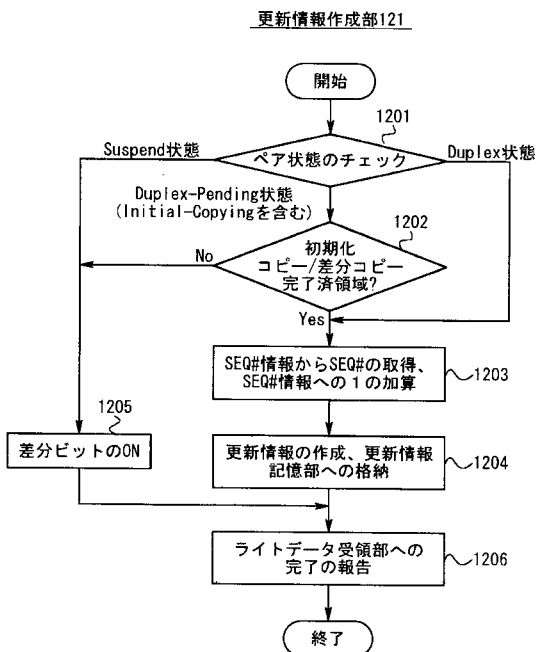
【図14】



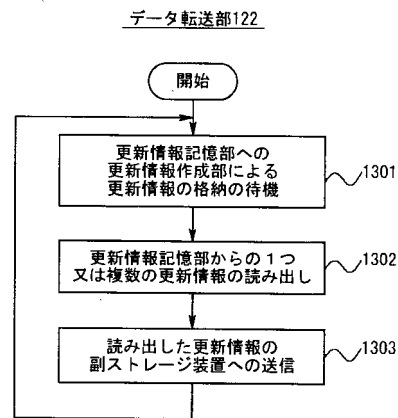
【図15】



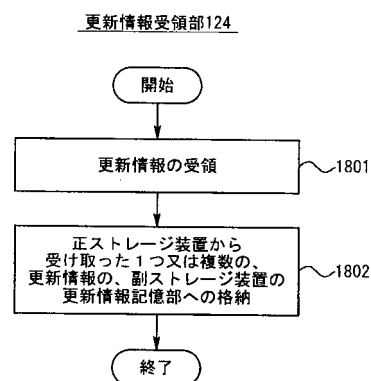
【図16】



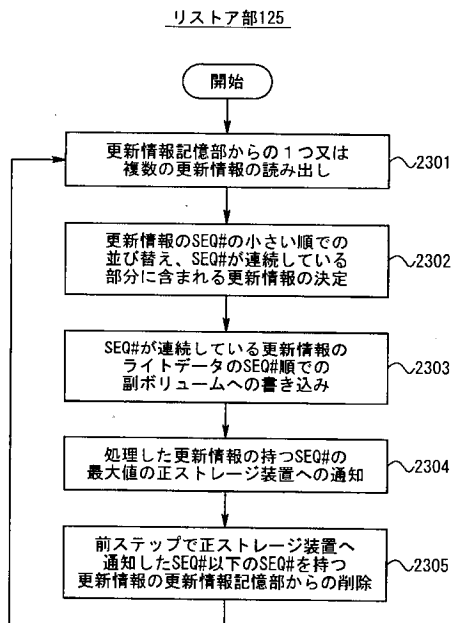
【図17】



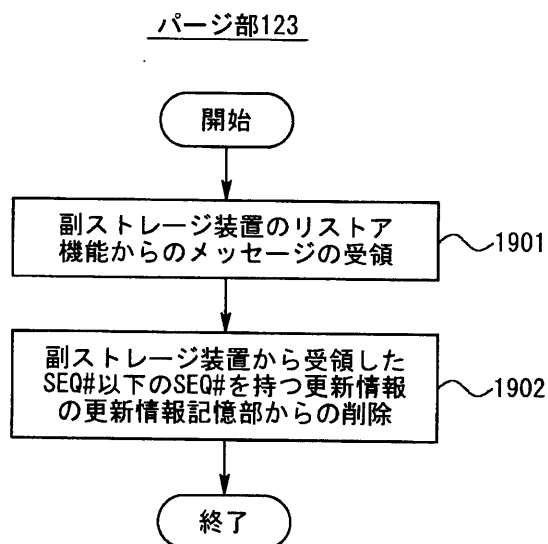
【図18】



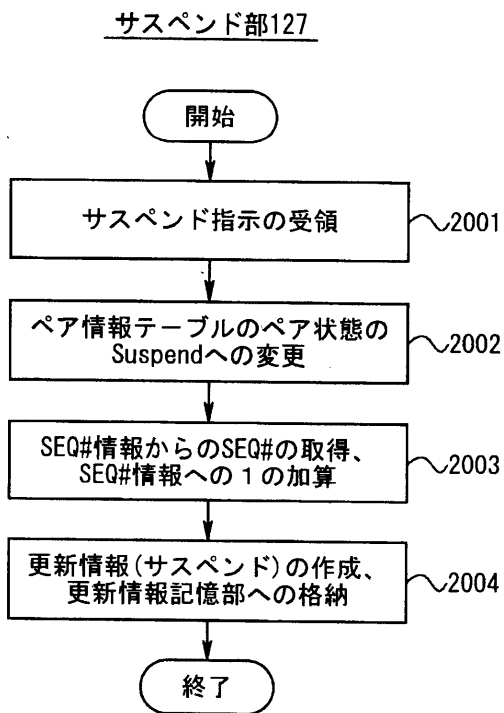
【図19】



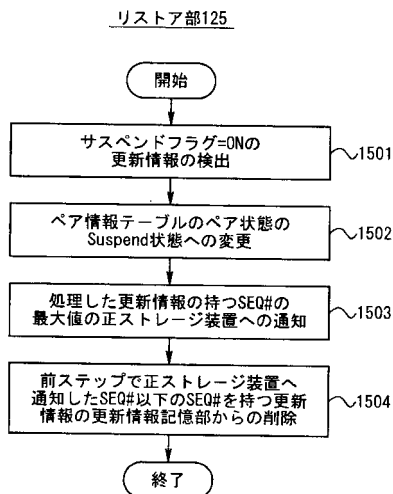
【図20】



【図21】

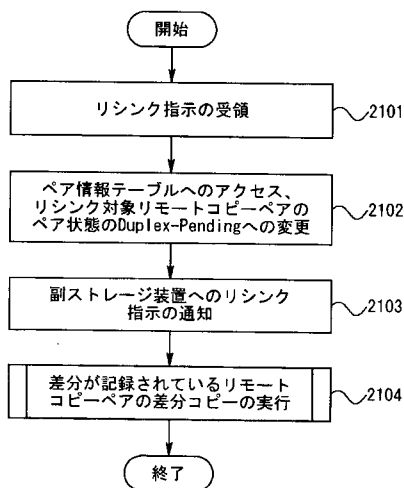


【図22】



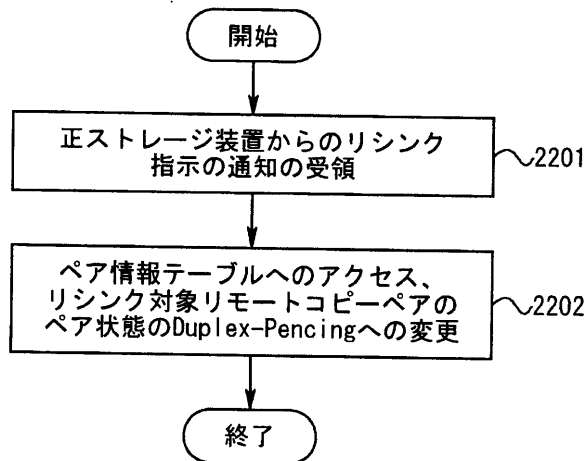
【図 2 3】

リシンク部(正ストレージ装置)128



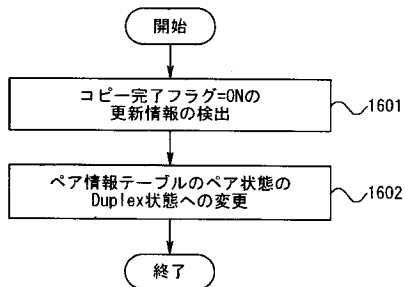
【図 2 4】

リシンク部(副ストレージ装置)128



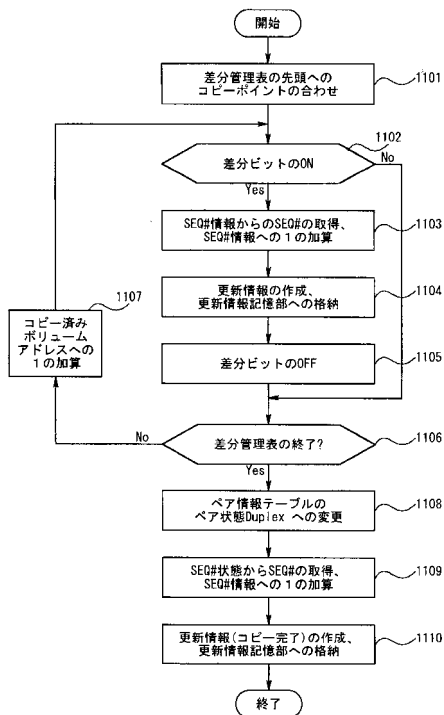
【図 2 5】

リストア部125

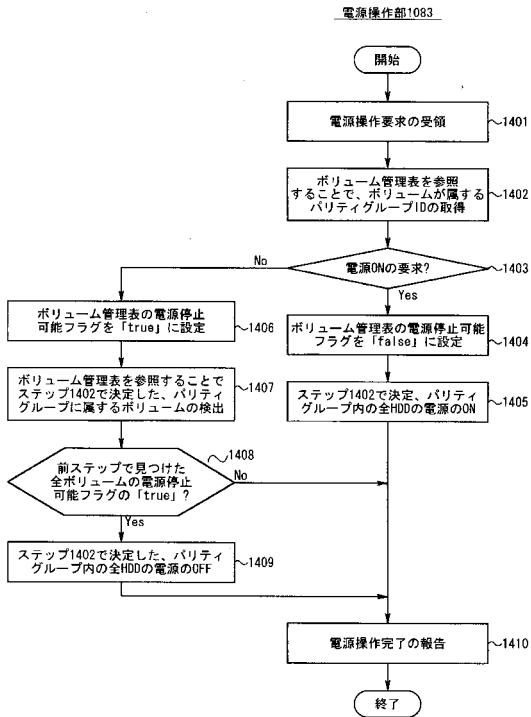


【図 2 6】

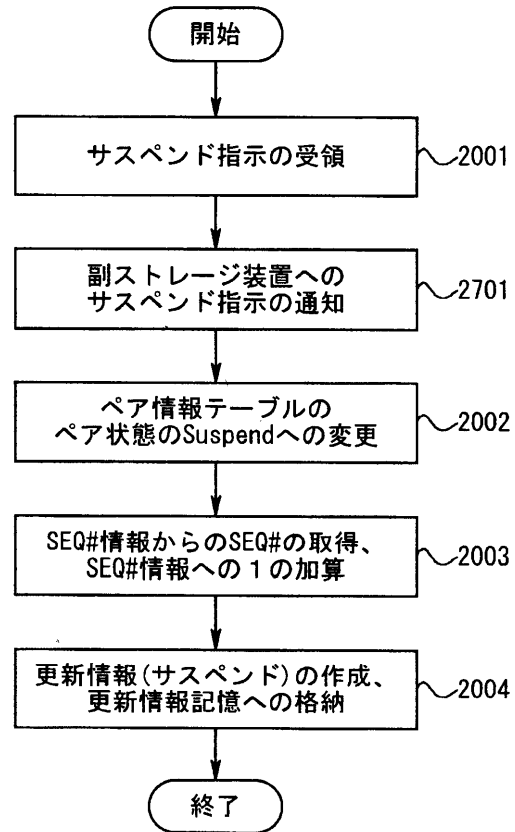
差分コピー部129



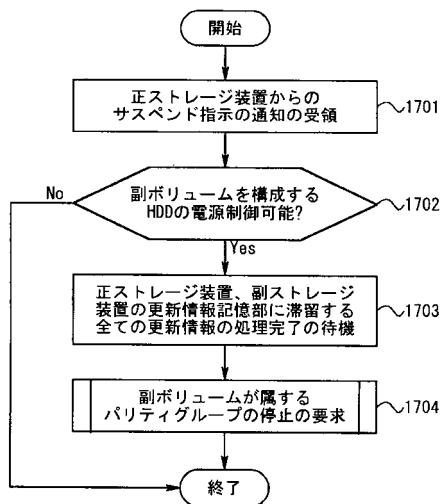
【図 27】



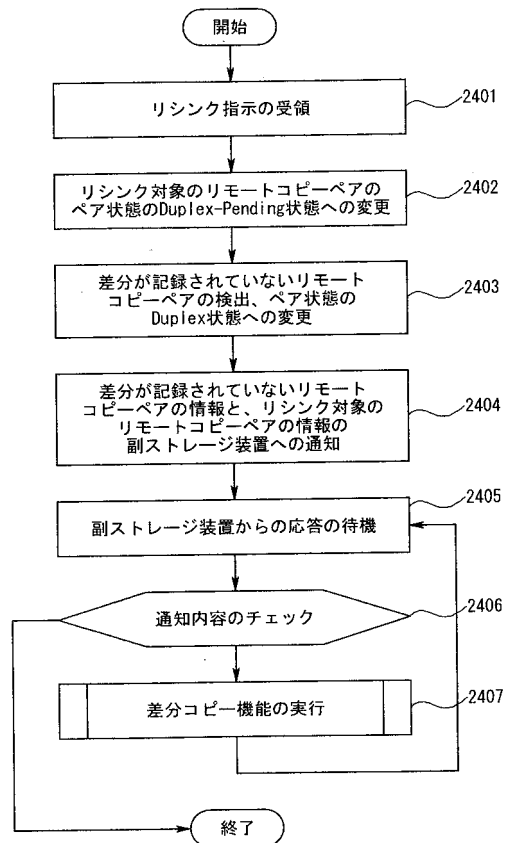
【図 28】



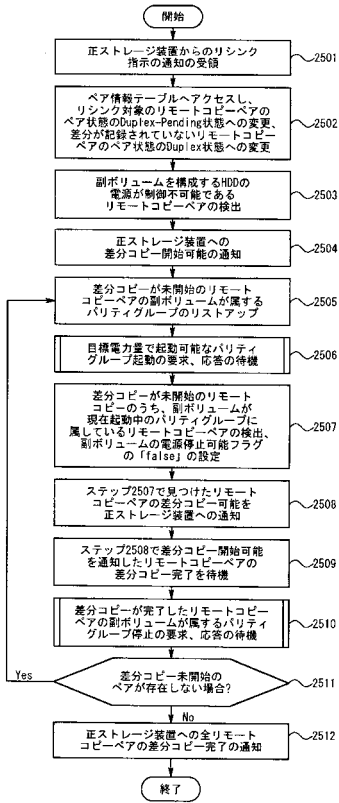
【図 29】



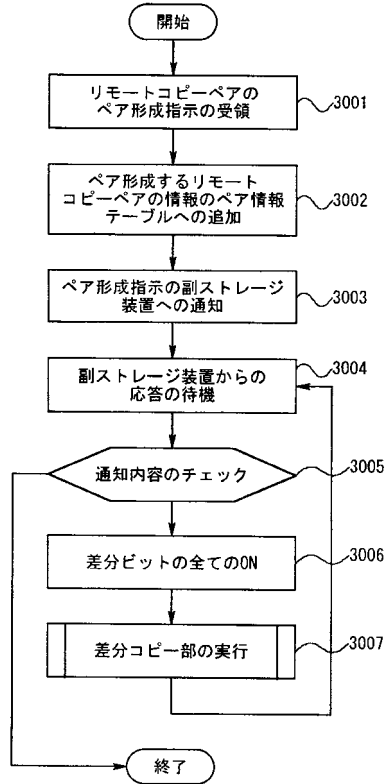
【図 30】



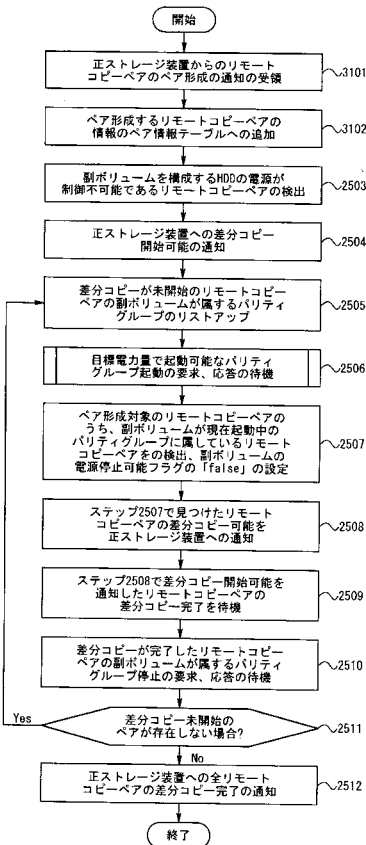
【図 3 1】



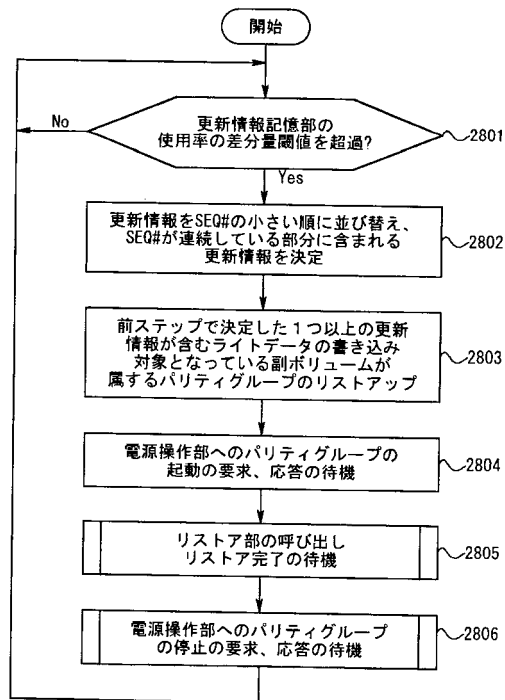
【図 3 2】



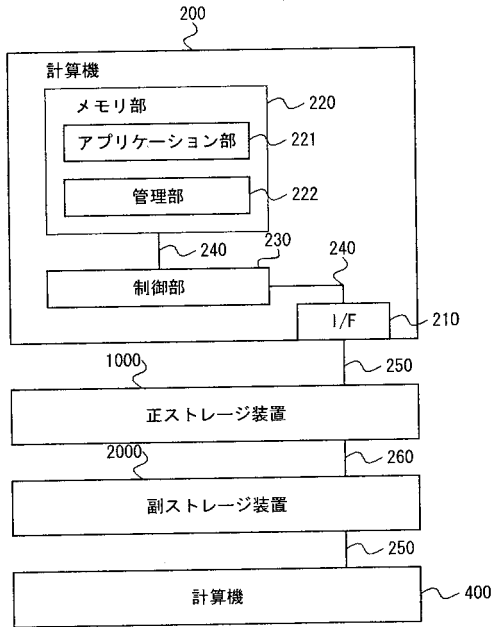
【図 3 3】



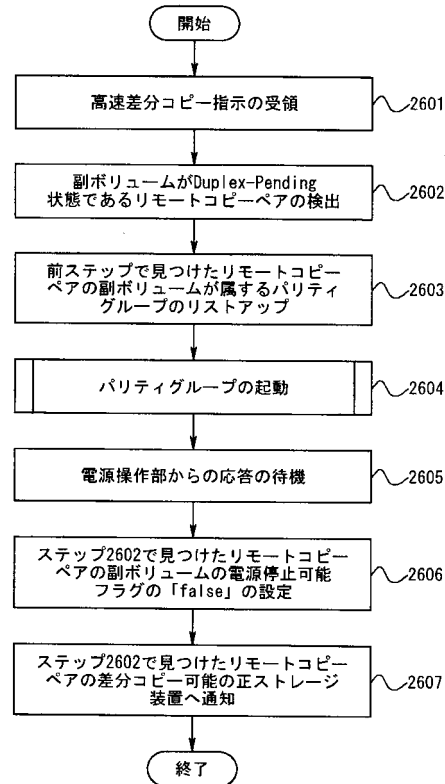
【図 3 4】



【図 3 5】



【図 3 6】



【手続補正書】

【提出日】平成18年2月23日(2006.2.23)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】0022

【補正方法】変更

【補正の内容】

【0022】

正ストレージ装置1000の制御情報部1090は、ペア情報テーブル1091、ボリューム管理表500、パリティグループ表510、ボリューム構成情報1092、差分管理表1093、更新情報記憶部1094、SEQ#情報1095を有する。更新情報記憶部1094には後述の更新情報1300が記憶される。

なお、制御情報部1090に含まれる各種情報は、後述するボリューム1070に格納してもよい。なお、正ストレージ装置1000と副ストレージ装置2000の更新情報記憶部1094は、キャッシュ1040の一部または全てを用いて実現することが考えられるが、これ以外にもパリティグループ1120やボリューム1070の一部または全ての領域を併用してもよい。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】0036

【補正方法】変更

【補正の内容】

【0036】

では、リモートコピーの概要を述べる。図7には、リモートコピー部1082の構成の一例を示している。一般的には、リモートコピーには、同期リモートコピー、および、非

同期リモートコピーがあるが、本実施例では非同期リモートコピーのみを例として説明する。よって、リモートコピー部 1082 は、非同期リモートコピー部 120 から成る。非同期リモートコピー部 120 は、更新情報作成部 121、データ転送部 122、ページ部 123、更新情報受領部 124、リストア部 125、ペア形成部 126、サスペンド部 127、リシンク部 128、差分コピー部 129 を有する。なお、非同期リモートコピー部 1082 が有する各部の説明は後述する。ここでは、非同期リモートコピーの概要を述べる。

【手続補正 3】

【補正対象書類名】明細書

【補正対象項目名】0095

【補正方法】変更

【補正の内容】

【0095】

なお、リストア部 125 は、ステップ 2301 からステップ 2305 の処理を、常に繰り返し実行しているとしたが、更新情報記憶部 1094 に更新情報 1300 が格納されていない場合には、処理を終了してもよい。さらに、リストア部 125 はリモートコピー部 1082 の他の部から必要に応じて起動されることも可能である。

【手続補正 4】

【補正対象書類名】明細書

【補正対象項目名】0184

【補正方法】変更

【補正の内容】

【0184】

更新情報記憶部 1094 の使用率が更新情報閾値 1098 を超えた場合には、コピー電源連携部 1084 は、副ストレージ装置 2000 の更新情報記憶部 1094 に格納されている更新情報 1300 を SEQ# を用いてソートし、さらに、SEQ# が連続している部分（更新情報集合に含まれる更新情報 1300 を SEQ# 順に並べた際に、SEQ# に抜けが生じない部分）に含まれる更新情報 1300 を決定する（ステップ 2802）。

フロントページの続き

(72)発明者 山本 康友

神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所システム開発研究所内

(72)発明者 丸山 哲也

神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所システム開発研究所内

Fターム(参考) 5B065 BA01 CA11 CE22 CE30 EA11 EA27 EA33 ZA14