

(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2007-141047

(P2007-141047A)

(43) 公開日 平成19年6月7日(2007.6.7)

(51) Int. Cl.	F I	テーマコード (参考)
<b>G06F 3/06 (2006.01)</b>	G06F 3/06	5B065
<b>G06F 12/00 (2006.01)</b>	G06F 12/00 531M	5B082

審査請求 未請求 請求項の数 20 O L (全 30 頁)

(21) 出願番号 特願2005-335655 (P2005-335655)  
 (22) 出願日 平成17年11月21日 (2005.11.21)

(71) 出願人 000005108  
 株式会社日立製作所  
 東京都千代田区丸の内一丁目6番6号  
 (74) 代理人 110000062  
 特許業務法人第一国際特許事務所  
 (72) 発明者 高岡 伸光  
 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 兼田 泰典  
 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  
 Fターム(参考) 5B065 BA01 CE01 ZA14  
 5B082 DE06 HA08

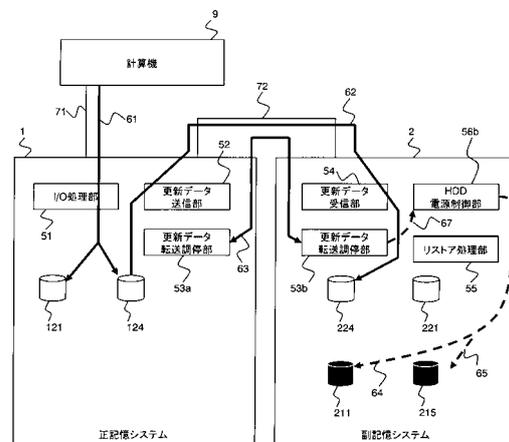
(54) 【発明の名称】 計算機システム及び計算機システムのデータ複製方法

(57) 【要約】

【課題】 リモートコピーを用いた計算機システムにおいて、必要な記憶装置に電源を投入するように制御して、副記憶システムの電力消費量をより低減し、記憶装置の寿命を延長することができる計算機システムを提供する

【解決手段】 上位装置が第一の記憶システムへデータを送信し、第一の記憶システムはデータを第一の論理ボリュームへ記録し、第一の記憶システムはデータを含む更新データを生成し、第二の記憶システムは、第一の記憶システムから更新データを受領する契機において、第二の論理ボリュームを構成する記憶装置の電源が切断されていれば第二の論理ボリュームを構成する記憶装置の電源を投入して前記更新データを前記第二の論理ボリュームに記録し、さらに、第二の記憶システムは第二の論理ボリュームに記録された更新データを読み出して第三の論理ボリュームへ記録することを特徴とする。

【選択図】 図1



## 【特許請求の範囲】

## 【請求項 1】

第一の論理ボリュームを有する第一の記憶システムと、  
前記第一の記憶システムに接続され、第二の論理ボリュームと第三の論理ボリュームを有する第二の記憶システムと、  
前記第一の記憶システムに接続された上位装置と、  
を有する計算機システムにおいて、  
前記上位装置は、前記第一の記憶システムへデータを送信する送信部を備え、  
前記第一の記憶システムは、前記データを前記第一の論理ボリュームへ記録し、前記データを含む更新データを生成する第一の処理部を備え、  
前記第二の記憶システムは、前記第一の記憶システムから前記更新データを受領する契機において、前記第二の論理ボリュームを構成する記憶装置の電源が切断されていれば、前記第二の論理ボリュームを構成する記憶装置の電源を投入する電源制御部と前記更新データを前記第二の論理ボリュームに記録する調停部とを備え、  
さらに前記第二の記憶システムは、前記第二の論理ボリュームに記録された更新データを読み出して前記第三の論理ボリュームへ記録する第二の処理部を備えていることを特徴とする計算機システム。

10

## 【請求項 2】

請求項 1 記載の計算機システムであって、  
前記更新データを読み出して前記第三の論理ボリュームへ記録する契機は、前記第二の論理ボリュームに記録される更新データの量により決定することを特徴とする計算機システム。

20

## 【請求項 3】

請求項 1 記載の計算機システムであって、  
前記第一の記憶システムは、前記更新データを記録する第四の論理ボリュームを有しており、  
前記更新データを受領する契機は、前記第四の論理ボリュームに記録される更新データの量により決定することを特徴とする計算機システム。

## 【請求項 4】

請求項 1 記載の計算機システムであって、  
前記第一の論理ボリュームと前記第三の論理ボリュームはペアを構成しており、前記更新データを受領する契機は、前記ペアの状態に基づき決定することを特徴とする計算機システム。

30

## 【請求項 5】

請求項 1 記載の計算機システムであって、  
前記更新データを受領する契機は、前記記憶装置の電源の状態に基づき決定することを特徴とする計算機システム。

## 【請求項 6】

請求項 1 記載の計算機システムであって、  
前記第二の記憶システムは、前記記憶装置の電源を切断する契機に至った場合に、前記電源制御部により、前記記憶装置の電源を切断することを特徴とする計算機システム。

40

## 【請求項 7】

請求項 1 記載の計算機システムであって、  
前記第二の記憶システムは、前記記憶装置の電源を切断する契機に至った場合に、前記電源制御部により、前記記憶装置の電源を切断し、前記電源を切断する契機は、前記記憶装置に対する最終アクセス回数または電源投入回数に基づき決定することを特徴とする計算機システム。

## 【請求項 8】

第一の論理ボリュームを有する第一の記憶システムと、

50

前記第一の記憶システムに接続され、第二の論理ボリュームを有する第二の記憶システムと、

前記第一の記憶システムに接続された上位装置と、を有する計算機システムにおいて、前記上位装置は、前記第一の記憶システムへデータを送信する送信部を備え、

前記第一の記憶システムは、前記データを前記第一の論理ボリュームへ記録する第一の処理部と、前記データを含む更新データを生成して送信する送信部を備え、

前記第二の記憶システムは、前記第一の記憶システムから前記更新データを受領する契機において、前記第二の論理ボリュームを構成する記憶装置の電源が切断されていれば、前記第二の論理ボリュームを構成する記憶装置の電源を投入する電源制御部を備え、

さらに前記第二の記憶システムは、前記送信部から送信された更新データを受信し、前記更新データに含まれている前記第一の論理ボリュームからのデータを前記第二の論理ボリュームへ記録する受信部を備えていることを特徴とする計算機システム。

10

【請求項 9】

請求項 8 記載の計算機システムであって、

前記更新データを受領する契機は、前記第一の論理ボリュームに記録されているデータのうち、変更されたデータの割合が一定の値を超えた時点に基づき決定することを特徴とする計算機システム。

【請求項 10】

請求項 8 記載の計算機システムであって、

前記第二の記憶システムに、更に、外部記憶システムが接続されて、

20

前記第二の記憶システムの前記第二の論理ボリュームは、前記外部記憶システムの論理ボリュームの仮想論理ボリュームを形成しており、

第二の記憶システムの電源制御部は、前記仮想論理ボリュームに対する電源制御の要求を受けた場合に、前記外部記憶システムの前記論理ボリュームの電源の制御を要求し、前記外部記憶システムの電源制御部は、前記論理ボリュームを構成している記憶装置の電源を制御することを特徴とする計算機システム。

【請求項 11】

第一の論理ボリュームを有する第一の記憶システムと、

前記第一の記憶システムに接続され、第二の論理ボリュームと第三の論理ボリュームを有する第二の記憶システムと、

30

前記第一の記憶システムに接続された上位装置と、

を有する計算機システムのデータ複製方法において、

前記上位装置は、前記第一の記憶システムへデータを送信し、

前記第一の記憶システムは、前記データを前記第一の論理ボリュームへ記録し、前記データを含む更新データを生成し、

前記第二の記憶システムは、前記第一の記憶システムから前記更新データを受領する契機において、前記第二の論理ボリュームを構成する記憶装置の電源が切断されていれば、前記第二の論理ボリュームを構成する記憶装置の電源を投入し、前記更新データを前記第二の論理ボリュームに記録し、

さらに前記第二の記憶システムは、前記第二の論理ボリュームに記録された更新データを読み出して前記第三の論理ボリュームへ記録することを特徴とする計算機システムのデータ複製方法。

40

【請求項 12】

請求項 11 記載の計算機システムのデータ複製方法であって、

前記更新データを読み出して前記第三の論理ボリュームへ記録する契機は、前記第二の論理ボリュームに記録される更新データの量により決定することを特徴とする計算機システムのデータ複製方法。

【請求項 13】

請求項 11 記載の計算機システムのデータ複製方法であって、

前記第一の記憶システムは、前記更新データを記録する第四の論理ボリュームを有して

50

おり、

前記更新データを受領する契機は、前記第四の論理ボリュームに記録される更新データの量により決定することを特徴とする計算機システムのデータ複製方法。

【請求項 14】

請求項 11 記載の計算機システムのデータ複製方法であって、

前記第一の論理ボリュームと前記第三の論理ボリュームはペアを構成しており、前記更新データを受領する契機は、前記ペアの状態に基づき決定することを特徴とする計算機システムのデータ複製方法。

【請求項 15】

請求項 11 記載の計算機システムのデータ複製方法であって、

前記更新データを受領する契機は、前記記憶装置の電源の状態に基づき決定することを特徴とする計算機システムのデータ複製方法。

10

【請求項 16】

請求項 11 記載の計算機システムのデータ複製方法であって、

前記第二の記憶システムは、前記記憶装置の電源を切断する契機に至った場合に、前記電源制御部により、前記記憶装置の電源を切断することを特徴とする計算機システムのデータ複製方法。

【請求項 17】

請求項 11 記載の計算機システムのデータ複製方法であって、

前記第二の記憶システムは、前記記憶装置の電源を切断する契機に至った場合に、前記電源制御部により、前記記憶装置の電源を切断し、前記電源を切断する契機は、前記記憶装置に対する最終アクセス回数または電源投入回数に基づき決定することを特徴とする計算機システムのデータ複製方法。

20

【請求項 18】

第一の論理ボリュームを有する第一の記憶システムと、

前記第一の記憶システムに接続され、第二の論理ボリュームを有する第二の記憶システムと、

前記第一の記憶システムに接続された上位装置と、を有する計算機システムのデータ複製方法において、

前記上位装置は、前記第一の記憶システムへデータを送信し、

前記第一の記憶システムは、前記データを前記第一の論理ボリュームへ記録し、前記データを含む更新データを生成して送信し、

30

前記第二の記憶システムは、前記第一の記憶システムから前記更新データを受領する契機において、前記第二の論理ボリュームを構成する記憶装置の電源が切断されていれば、前記第二の論理ボリュームを構成する記憶装置の電源を投入し、

さらに前記第二の記憶システムは、前記送信部から送信された更新データを受信し、前記更新データに含まれている前記第一の論理ボリュームからのデータを前記第二の論理ボリュームへ記録することを特徴とする計算機システムのデータ複製方法。

【請求項 19】

請求項 18 記載の計算機システムのデータ複製方法であって、

前記更新データを受領する契機は、前記第一の論理ボリュームに記録されているデータのうち、変更されたデータの割合が一定の値を超えた時点に基づき決定することを特徴とする計算機システムのデータ複製方法。

40

【請求項 20】

請求項 18 記載の計算機システムのデータ複製方法であって、

前記第二の記憶システムに、更に、外部記憶システムが接続されて、

前記第二の記憶システムの前記第二の論理ボリュームは、前記外部記憶システムの論理ボリュームの仮想論理ボリュームを形成しており、

第二の記憶システムの電源制御部は、前記仮想論理ボリュームに対する電源制御の要求を受けた場合に、前記外部記憶システムの前記論理ボリュームの電源の制御を要求し、前

50

記外部記憶システムの電源制御部は、前記論理ボリュームを構成している記憶装置の電源を制御することを特徴とする計算機システムのデータ複製方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、複数の記憶システムを有する計算機システムに関し、特に、一方の記憶システムに記録されているデータを他方の記憶システムへコピーする機能、すなわちリモートコピー機能を有する計算機システムにおいて、記憶装置の電源を制御して電力消費量の抑制が可能な計算機システム及びデータ複製方法に関する。

【背景技術】

【0002】

本発明の背景技術として、二つの技術を挙げる。一つは、リモートコピー技術であり、一つは、記憶装置電源制御技術である。

【0003】

まず、リモートコピー技術について述べる。

リモートコピー技術は、計算機等の上位装置が記憶システムに書き込むデータ（オリジナルデータ）を、別の記憶システムにもコピーする技術である。このような技術は、ディザスタリカバリ（大規模自然災害により破壊された情報システムの復元）や、データのバックアップのために必要とされる。

【0004】

特許文献1は、リモートコピー技術の一形態を開示している。同公報が開示するリモートコピー技術を以下に説明する。前提として、下記(1)(2)(3)の要素からなる計算機システムを仮定する。

(1) 計算機。

(2) (1)の計算機に接続され、(1)の計算機がデータを記録し、変更する記憶システム。前記データをここではオリジナルデータと呼ぶ。

(3) (2)の記憶システムに接続され、オリジナルデータのコピーを保持する記憶システム。

【0005】

上記の計算機システムにおいて、(1)の計算機は、(2)の記憶システムに対してオリジナルデータの変更を要求する。すると、(2)の記憶システムは、オリジナルデータの更新に関する情報（ジャーナル）を生成し、自身の記憶領域に一旦蓄積する。(3)の記憶システムは、(2)の記憶システムに対してジャーナルの送信を要求し、受信したジャーナルを用いてコピーを更新する。

(3)の記憶システムは、計算機からのアクセスとは非同期的に、ジャーナルが一旦蓄積された記憶領域からジャーナルを取得する（ジャーナルリード）。これにより、計算機によるオリジナルデータへのアクセスに、リモートコピーを原因とする性能低下が生じることを防いでいる。

【0006】

以降、リモートコピー技術による記憶システムのデータコピー機能を、単にリモートコピーと呼ぶ。コピー対象のオリジナルのデータを記録している記憶システムを正記憶システムと呼び、コピーされたデータを記録している記憶システムを副記憶システムと呼ぶ。

【0007】

上位装置からのデータの書き込みを、オリジナルデータとそのコピーに同時に反映するリモートコピーを同期リモートコピーと呼び、そうではない場合、すなわち、オリジナルデータへの書き込みとコピーへの書き込みを、同時に実行することを保証しないリモートコピーを、非同期リモートコピーと呼ぶ。

【0008】

次に、記憶装置電源制御技術について述べる。

記憶装置電源制御技術は、記憶システムが備える記憶装置の電源を、記憶装置のアクセ

10

20

30

40

50

ス頻度やアクセスの有無に応じて切断することで、記憶システムが消費する電力量の低減と、記憶装置が故障するまでの期間（以下、可動期間という）の延長を図る技術である。

ここで、記憶装置とは、電子データの記録と読み出しが可能な装置のことであり、本明細書ではハードディスクドライブに注目する。この技術には、記憶装置の電源が切断されているために発生するアクセスの遅延の低減や、記憶装置の電源が切断された状態を出来る限り維持する工夫が含まれる。

【0009】

特許文献2は、記憶装置電源制御技術の一形態を開示している。同公報が開示する記憶装置電源制御技術は、以下のようなものである。

記憶システムは、指定した記憶装置（\*）の電源の投入/切断を外部の計算機から要求できるインタフェースを備える。前記記憶システムに接続し、前記記憶システムにデータを記録する計算機は、前記記憶システムの記憶装置の内、当面はアクセスしないことが分っているものについて、前記記憶システムに対して、前記インタフェースを使って記憶装置の電源切断を要求する。（\*）補足：実際には、論理ボリューム（後述）を指定する。

【0010】

また、特許文献3が開示する技術は、以下のようなものである。

記憶システムは、記憶装置へのアクセスが無い限りは、原則的に記憶装置の電源を切断しておく。電源が切断された状態の記憶装置を多くするために、RAID（Redundant Arrays of Inexpensive Disks）レベル4を修正したデータ記録方式を用いる。アクセスの遅延を低減するために、前記データ記録方式におけるパリティディスク（エラー訂正情報を記録する記憶装置）およびメタデータディスク（記憶システム内部の構成情報と、データのキャッシュが記録された記憶装置）の電源を、常時投入しておく。

【特許文献1】特開2005-18506号公報

【特許文献2】特開2005-157710号公報

【特許文献3】米国特許出願公開第2004/0054939号明細書

【発明の開示】

【発明が解決しようとする課題】

【0011】

リモートコピー技術を使用して、主記憶システムに記録されたデータを副記憶システムへコピーする際、コピーされたデータを記録する記憶装置は、更新が必要ではない間は電源を切断できる。そこで、副記憶システムへの記憶装置電源制御技術の適用を考える。

【0012】

特許文献2に記載の技術では、記憶装置の電源切断と投入を計算機が指示する。このため、特許文献1に記載されているような非同期リモートコピー、すなわち、データのコピーを、計算機からのアクセスとは非同期的に記憶システム間で実行するリモートコピーには、特許文献2に記載の技術は適さない。

【0013】

特許文献3に記載の技術では、常時電源が投入されている記憶装置を必要とする。このような記憶装置は、記憶システムの電力消費量の増加と、記憶システム内の記憶装置の平均可動時間の減少を招く。

【0014】

本発明の目的は、上記を鑑み、リモートコピー技術を使用する計算機システムにおいて、必要な記憶装置に電源を投入するように制御して、副記憶システムの電力消費量をより低減し、記憶装置の寿命を延長することができる計算機システムを提供することである。

【課題を解決するための手段】

【0015】

上記の課題を解決するために、本発明の計算機システムでは、上位装置が、前記第一の記憶システムへデータを送信する送信部を備え、第一の記憶システムは、前記データを第一の論理ボリュームへ記録し、前記データを含む更新データを生成する第一の処理部を備え、第二の記憶システムは、第一の記憶システムから更新データを受領する契機において

、第二の論理ボリュームを構成する記憶装置の電源が切断されていれば、第二の論理ボリュームを構成する記憶装置の電源を投入する電源制御部と前記更新データを前記第二の論理ボリュームに記録する調停部とを備え、さらに、第二の記憶システムは、第二の論理ボリュームに記録された更新データを読み出して第三の論理ボリュームへ記録する第二の処理部を備えることを特徴とする。

【発明の効果】

【0016】

本発明により、リモートコピーを使用した計算機システムにおいて、副記憶システムの電力消費量をより低減し、記憶装置の寿命を延長することができる計算機システムを提供することができる。

10

【発明を実施するための最良の形態】

【0017】

以降、図面を参照して、本発明の実施形態を説明する。

【実施例1】

【0018】

図1は、本発明の第一の実施形態の論理的な構成を示している。

本発明の第一の実施形態は、正記憶システム1、副記憶システム2、上位装置である計算機9からなる計算機システムである。

本明細書において、上位装置とは、記憶システムに対してデータの記録を要求する装置一般を指す。記憶システムもまた上位装置である。

20

【0019】

正記憶システム1と計算機9の間には、通信路71が構成される。正記憶システム1と副記憶システム2の間には、通信路72が構成される。これらの通信路を構成するには、ファイバチャネルによるストレージエリアネットワーク、LAN、公衆回線等のネットワークを使用する。通信路71と通信路72は同一のネットワークを使用して構成されるものであってもよい。

【0020】

正記憶システム1と副記憶システム2は、記憶装置であるハードディスクドライブ（以降、HDDと表記）を複数個備えている。

正記憶システム1は、そのHDDから構成される複数の論理ボリュームを有し、とりわけ論理ボリューム121と124を有する。正記憶システム1の論理ボリューム121は、計算機9から一つのブロックデバイスとしてアクセスできる。すなわち、計算機9は、論理ボリューム121へのデータの書き込みと、論理ボリューム121からのデータの読み出しを実行することができる。正記憶システム1の論理ボリューム124は、論理ボリューム121の更新データ（後述）を記録するために、正記憶システム1が内部で使用する。論理ボリューム124には、計算機9からはアクセスできない。

30

【0021】

副記憶システム2は、複数の論理ボリュームを有し、とりわけ論理ボリューム221と224を有する。論理ボリューム221は、正記憶システム1の論理ボリューム121に記録されたデータのコピーを記録する論理ボリュームである。論理ボリューム224は、論理ボリューム121の更新データを記録するために、副記憶システム2が内部で使用する。

40

【0022】

上記の更新データとは、上位装置からの論理ボリュームへのデータの書き込みを、副論理ボリュームに記録されているデータのコピーに対しても適用するためのデータである。更新データは、正記憶システム1で生成され、副記憶システム2へ送信される。更新データには、書き込まれたデータの内容、ペア（後述）を識別するための情報、データ更新の順序に関する情報などを含む。

【0023】

この実施形態1では、正記憶システム1から副記憶システム2へ更新データを転送する

50

ためのプロトコルとして、副記憶システム 2 が正記憶システム 1 から更新データを「読み出す」形態のプロトコルを採用する。

すなわち、副記憶システム 2 が正記憶システム 1 に対して更新データの送信を要求することで、正記憶システム 1 は副記憶システム 2 に対して更新データを送信する。

#### 【0024】

副記憶システム 2 は、複数の HDD を備え、とりわけ HDD 211 と 215 を備える。副記憶システム 2 の論理ボリューム 224 は、HDD 211 (および、その他 3 つの HDD) の記憶領域から構成され、論理ボリューム 221 は、HDD 215 (および、その他 3 つの HDD) から構成される。正記憶システム 1 の論理ボリュームも同様に HDD の記憶領域より構成される。図 1 では HDD 211 と 215 以外の HDD の記載を省略している。

10

#### 【0025】

記憶システム内の各 HDD を識別するために、各 HDD には、記憶システム内で一意の識別番号 (ID) が割当てられている。

HDD 211 と 215 は、通常は電源が切断されており、それらの論理ボリュームへのデータの記録が必要となった場合に電源を投入する。

#### 【0026】

論理ボリューム 121 に記録されるデータは、正記憶システム 1 と副記憶システム 2 のリモートコピー機能により、副記憶システム 2 の論理ボリューム 221 にコピーされる。

以降、コピーされるオリジナルデータを記録している論理ボリュームを正論理ボリューム、オリジナルデータのコピーを記録している論理ボリュームを副論理ボリュームと呼ぶ。

20

正論理ボリューム 121 と副論理ボリューム 221 の関係をペアと呼ぶ。

#### 【0027】

計算機システムを管理する者 (以降、ユーザと言う) は、正論理ボリューム 121 と副論理ボリューム 221 のペアの確立を、正記憶システム 1 もしくは副記憶システム 2 に指示することが出来る。ペアの確立が指示されると、正記憶システム 1 と副記憶システム 2 は、指示されたペアの正論理ボリューム 121 と副論理ボリューム 221 を、ペアの情報として内部の共有メモリに記録する。

#### 【0028】

ペアの確立の指示は、正記憶システム 1 もしくは副記憶システム 2 を管理するための管理端末機器を通して実行できる。図 1 では管理端末機器の記載を省略しているが、このような仕組みが実施可能であることは、当業者には明かである。

30

#### 【0029】

正記憶システム 1 と副記憶システム 2 の間のリモートコピーにおいては、更新データを一時的に記録する論理ボリューム 124 および 224 を使用する。更新データを記録する論理ボリュームを更新データ論理ボリュームと呼ぶ。また、正記憶システム 1 の更新データ論理ボリュームを正更新データ論理ボリューム、副記憶システム 2 の更新データ論理ボリュームを副更新データ論理ボリュームと呼ぶ。

#### 【0030】

更新データ論理ボリュームを使用して、論理ボリューム 121 へ書き込まれたデータを論理ボリューム 221 へ適用する処理の流れは次のようになる。正記憶システム 1 は、計算機 9 から論理ボリューム 121 へのデータの記録を指示されると、論理ボリューム 121 へデータを記録するとともに、更新データを生成し更新データ論理ボリューム 124 に記録する。その後、正記憶システム 1 は、しかるタイミングで更新データを副記憶システム 2 に送信する。副記憶システム 2 は、受信した更新データを更新データ論理ボリューム 224 に記録し、しかるタイミングで、更新データに基づき論理ボリューム 221 へ記録されたデータを更新する。

40

#### 【0031】

上でしかるタイミングという言葉を使用したが、このタイミングを決定する処理につい

50

ては後に詳細に説明する。また、データの更新の処理が上記のようであるため、計算機 9 からの正論理ボリューム 1 2 1 へのデータの書きこみは、副論理ボリューム 2 2 1 へ遅延して適用され、副論理ボリューム 2 2 1 のデータは、正論理ボリューム 1 2 1 のある過去の時点でのデータと一致する。このようなデータの一致性を、本明細書では断りなく一致性という場合がある。

【 0 0 3 2 】

なお、上記のタイミングは、正論理ボリューム 1 2 1 へのデータ記録処理とは無関係に決定される。したがって、本実施形態 1 のリモートコピーは非同期リモートコピーの一形態である。

上記の処理を実施するにあたり、更新データ論理ボリューム 1 2 4 および 2 2 4 の関係付け、論理ボリューム 1 2 1 と更新データ論理ボリューム 1 2 4 の関係付け、論理ボリューム 2 2 1 と更新データ論理ボリューム 2 2 4 の関係付けを、正記憶システム 1 と副記憶システム 2 に指示する必要がある。これらの関係付けは、ペアの確立の場合と同様に、ユーザが管理端末機器を使用して指示する。

10

【 0 0 3 3 】

正論理ボリューム 1 2 1 と副論理ボリューム 2 2 1 のペアの確立が確立した直後は、正論理ボリューム 1 2 1 のデータは、副論理ボリューム 2 2 1 にはその一部もコピーされていない。このため、正論理ボリューム 1 2 1 のデータを副論理ボリューム 2 2 1 へ全くコピーする必要がある。この、ペア確立直後のデータのコピーを初期コピーと呼ぶ。初期コピー後、上に述べたデータの更新の処理へ移行することで、正論理ボリューム 1 2 1 のデータと副論理ボリューム 2 2 1 のデータの間の一致性が維持される。

20

【 0 0 3 4 】

正記憶システム 1 および副記憶システム 2 への初期コピーの開始の指示は、ペア確立の場合と同様に、ユーザが管理端末機器を使用して行う。ペア確立の指示が初期コピー開始の意味を含むものであってもよい。

【 0 0 3 5 】

以降、副記憶装置の HDD 2 1 1 および 2 1 5 の電源を制御しつつ、正論理ボリューム 1 2 1 と副論理ボリューム 2 2 1 の一致性を維持する処理の流れ（フロー F 1）を説明する。正論理ボリューム 1 2 1 と副論理ボリューム 2 2 1 の間の初期コピーは行われているものとする。また、副記憶システム 2 の HDD 2 1 1 と 2 1 5 の電源は切断されているものとする。

30

【 0 0 3 6 】

本実施形態 1 では、正更新データ論理ボリューム 1 2 4 に記録された更新データの量に基づき、更新データを送信するタイミングを決定する。

[ フロー F 1 ]

( F 1 - 1 ) 計算機 9 は、正記憶システム 1 に対して、正論理ボリューム 1 2 1 へのデータの書き込みを要求する。この要求は、I / O 処理部 5 1 によって処理される。

( F 1 - 2 ) I / O 処理部 5 1 は、計算機 9 からのデータの書き込みの要求に対して、正論理ボリューム 1 2 1 へ書き込むとともに、その要求から更新データを生成し、更新データを正更新データ論理ボリューム 1 2 4 へ書き込む。

40

【 0 0 3 7 】

ここまでのステップにおけるデータの流れを、矢印 6 1 によって示している。

( F 1 - 3 ) 副記憶システム 2 の更新データ転送調停部 5 3 b (以降、更新データ転送調停部 (副) と記載) は、更新データ転送調停部 (正) 5 3 a から正更新データ論理ボリューム 1 2 4 の使用率を定期的を取得する (矢印 6 3)。正更新データ論理ボリューム 1 2 4 の使用率が一定の閾値以上である場合は、正記憶システム 1 から、正更新ボリューム 1 2 4 に記録されている更新データを取得することを決定する。

( F 1 - 4 ) この時、副更新データ論理ボリューム 2 2 4 を構成している HDD 2 1 1 の電源は停止しているため、受信した更新データを記録することができない。このため、更新データ転送調停部 (副) 5 3 b は、まず HDD 電源制御部 5 6 b に対して、副更新デ

50

ータ論理ボリューム 224 を構成している HDD の電源を投入することを指示する（破線矢印 67）。

（F1-5） HDD 電源制御部 56b は、上記の指示を受け、HDD 211 の電源を投入する（破線矢印 64）。

（F1-6） HDD の電源が投入され、副更新データ論理ボリューム 224 へのデータの記録が可能となると、更新データ転送調停部（副）53b は、更新データ受信部 54 に対して、更新データの取得を指示する。

（F1-7） 更新データ受信部 54 は、上記指示を受け、正記憶システム 1 の更新データ送信部 52 に対して、正更新データ論理ボリューム 124 に記録された更新データの転送を指示する。更新データ転送部 52 は、上記の指示を受け、正更新データ論理ボリューム 124 に記録された更新データを転送する。更新データ受信部 54 は、受信した更新データを副更新データ論理ボリューム 224 に記録する。

（F1-8） 副更新データ論理ボリューム 224 の使用率が定められた閾値以上の場合は、更新データを副論理ボリューム 221 に適用する。

#### 【0038】

更新データを副論理ボリューム 221 に適用する場合、更新データ転送調停部 53b は、リストア処理部 55 に対して、更新データの適用を指示する。ここで、副論理ボリューム 221 の HDD の電源は切断されているため、更新データ転送調停部 53b は、まず HDD 電源制御部 56b に対して副論理ボリューム 221 の電源の投入を指示する。

（F1-9） HDD 電源制御部 56b は、上記指示を受けて、HDD 215 の電源を投入する（破線矢印 65）。

（F1-10） リストア処理部 55 は、更新データ転送調停部（副）53b からの指示を受け、副更新データ論理ボリューム 224 に記録された更新データを、副論理ボリューム 221 に適用する。

（F1-11） 副記憶システム 2 は、HDD の電源を、適切なタイミングで切断する。このタイミングは、HDD へのアクセスの有無や、HDD への電源投入回数に基づき決定される。

#### 【0039】

以上が、本実施形態 1 の概要である。以降、本実施形態 1 の詳細を、より詳細な図を参照しながら説明する。

図 2 は、正記憶システム 1 の構成を示す図である。

正記憶システム 1 は、ホストアダプタ 101 および 102、ディスクアダプタ 109、共有メモリ 103、キャッシュメモリ 110 を備えたコントローラ 100 と、HDD 111、112、113、114、115、116、117、118 とを有する。ホストアダプタ 101 および 102、ディスクアダプタ 109、共有メモリ 103、キャッシュメモリ 110 は、バス 119 により CPU 126 に接続されており、相互に通信可能となっている。ディスクアダプタ 109 は、HDD 接続線 120 によって HDD 上記 8 つの HDD に接続されており、これによりこれらの HDD の制御（データの読み出し、書き込み、電源切断、電源投入）が可能となっている。

#### 【0040】

ホストアダプタ 101 および 102 は、I/O 処理部 51、更新データ送信部 52、更新データ転送調停部 53、更新データ受信部 54、リストア処理部 55 を有する。これらを CPU 126 が実行できるプログラムとし、ホストアダプタ 101 および 102 内の各プログラムをコントローラ 100 の CPU 126 が実行するように構成し、ホストアダプタ 101 および 102 の上位装置からの要求（データの書き込み、読み出し）と更新データの転送に関する処理を実行することができる。

#### 【0041】

また、図 2 においては、ホストアダプタ 101 および 102 が行う上位装置からの要求（データの書き込み、読み出し）と、更新データの転送に関する処理をコントローラ 100 の CPU 126 が実行するものとして説明したが、ホストアダプタ 101 および 102

10

20

30

40

50

のそれぞれに、CPUおよびメモリを設置して、I/O処理部51、更新データ送信部52、更新データ転送調停部53、更新データ受信部54、リストア処理部55のプログラムを、ホストアダプタ101および102上のそれぞれのCPUが実行し、それぞれ、上位装置からの要求(データの書き込み、読み出し)と、更新データの転送に関する処理を行うようにすることができる。

【0042】

更には、ホストアダプタ101および102のI/O処理部51、更新データ送信部52、更新データ転送調停部53、更新データ受信部54、リストア処理部55を、それぞれICチップ等のハードウェアコンポーネントで構成し、これらを論理制御回路で接続して、ホストアダプタ101および102をそれぞれ、独立に作動可能なハードウェアコン

10

【0043】

ホストアダプタ101は、ネットワーク41に接続されている。図1の通信路71は、ネットワーク41を用いて構成されている。ホストアダプタ102は、ネットワーク42に接続されている。図1の通信路72は、ネットワーク42を用いて構成されている。ネットワーク41と42は、同一のネットワークであってもよい。

【0044】

ディスクアダプタ109は、HDDの電源の制御と、論理ボリュームの制御を実行する

ディスクアダプタ109のHDD電源制御部56および論理ボリューム制御部57をプログラムとして、CPU50とRAM58を用いて実行することができる。また、ディスクアダプタ109上にCPU50及びRAM58を設けて、HDD電源制御部56および論理ボリューム制御部57のプログラムを実行するように構成すること、あるいは、ディスクアダプタ109のHDD電源制御部56および論理ボリューム制御部57をハードウェアコンポーネントとすることもできる。

20

【0045】

共有メモリ103は、コントローラ100のCPU126、あるいは、ホストアダプタ101および102およびディスクアダプタ109からアクセスできるメモリである。共有メモリ103には、HDDグループ管理情報104、論理ボリューム情報105、更新データグループ管理情報106、ペア情報107、更新データグループ進捗情報108が記

30

【0046】

HDDグループ管理情報104は、正記憶システム1が有するHDDグループの情報を記録している。HDDグループとは、HDDの障害時に、HDDに記録されたデータを復元するための冗長データを含む、HDDの集合である。本実施形態では、HDDグループ内のHDDを連続した記憶領域みなした上で、その記憶領域を幾つかの領域に分割し、分割した領域を論理ボリュームとする。

論理ボリューム情報105は、正記憶システム1が有する論理ボリュームの構成に関する情報を記録している。

【0047】

更新データグループ管理情報106は、正記憶システム1が有する更新データ論理ボリュームと、更新データグループに関する情報を記録している。更新データグループとは、正記憶システム1と副記憶システム2の更新データ論理ボリューム124および224の関連性と、その関連性の上にペアを成す論理ボリューム121と221を含む、ペアを管理するためのグループである。

40

ペア情報107は、正記憶システム1に構成されているペアの情報を記録している。

更新データグループ進捗情報108は、更新データを記録する更新データ論理ボリューム124のアドレスと、更新データを副記憶システム2に送信するために更新データを読み出す更新データ論理ボリューム124のアドレスを記録している。

【0048】

50

キャッシュメモリ 110 は、ホストアダプタ 101、102 および ディスクアダプタ 109 の間でのデータの受渡しのために使用するメモリである。また、キャッシュメモリは、計算機 9 等の上位装置からの論理ボリュームへのアクセス要求を高速に（応答時間を短く）処理するためのキャッシュ領域としての用途もある。

#### 【0049】

図 3 は、副記憶システム 2 の構成を示す図である。副記憶システム 2 は、正記憶システム 1 とほぼ同様の構成であり、ホストアダプタ 201 および 202、ディスクアダプタ 209、共有メモリ 203、キャッシュメモリ 210 を備えたコントローラ 200 と、HDD 211、212、213、214、215、216、217、218 とを有する。ホストアダプタ 201 および 202、ディスクアダプタ 209、共有メモリ 203、キャッシュメモリ 210 は、バス 219 により CPU 226 に接続されており、相互に通信可能となっている。ディスクアダプタ 209 は、HDD 接続線 220 によって HDD 上記 8 つの HDD に接続されており、これによりこれらの HDD の制御（データの読み出し、書き込み、電源切断、電源投入）が可能となっている。

副記憶システム 2 の各構成要素は正記憶システムの各構成要素と同様に動作し、正記憶システム 1 と同様に、コントローラ 200 の CPU 226 による制御、ホストアダプタ 201 及びホストアダプタ 202 上に設置した CPU による制御、あるいは、ハードウェアコンポーネントとしてのホストアダプタ 201、ホストアダプタ 202 による制御等が可能のように構成することができる。

副記憶システム 2 は、正記憶システム 1 と異なる構成としては、コントローラ 200 内の共有メモリ 203 内に HDD 電源制御情報 225 が記録されている。また、ホストアダプタ 201 はネットワーク 42 を介して正記憶システム 1 に接続され、ホストアダプタ 202 は、ネットワーク 43 に接続されている。

#### 【0050】

図 4 は、正記憶システム 1 の論理ボリューム管理情報 105 の内容をテーブルの形式で示す図である。論理ボリューム管理情報 105 は、記憶システム 1 が有する論理ボリュームの構成に関する情報を記録している。

一つの論理ボリュームの構成に関する情報は、論理ボリューム ID、論理ボリュームタイプ、論理ボリューム状態、容量、物理アドレスの各情報からなる。

論理ボリューム ID は、記憶システムが有する論理ボリュームを識別するための識別子（ID）である。本実施形態では、論理ボリューム ID は、記憶システム内で重複しない一意の整数とする。理解を助けるために、本明細書では論理ボリューム ID と、図 1 の符号の指示番号を一致させている（図 1 に記載のない論理ボリュームもある）。

#### 【0051】

論理ボリュームタイプは、論理ボリュームの用途を示す数値である。論理ボリュームタイプには、ホスト（10）、正更新データ（20）、副更新データ（21）がある（括弧内は、対応する数値）。計算機 9 からアクセスされる論理ボリュームの場合、論理ボリュームタイプはホスト（10）である。論理ボリュームが正更新データ論理ボリュームの場合、論理ボリュームタイプは正更新データ（20）である。論理ボリュームが副更新データ論理ボリュームの場合、論理ボリュームタイプは副更新データ（21）である。

論理ボリュームがペアの正論理ボリューム 121 である場合は、論理ボリュームタイプの値に 1 を加算し、副論理ボリュームである場合は、論理ボリュームタイプの値に 2 を加算する。

理解を助けるために、本明細書の図においては、論理ボリュームタイプの数値を記載する代わりに、ホスト、ホスト（正）、ホスト（副）、更新データ（正）、更新データ（副）等と記載する。

#### 【0052】

論理ボリューム状態は、論理ボリュームの状態を示す数値である。論理ボリュームの状態には、正常（1）、異常（2）がある（括弧内は、対応する数値）。理解を助けるために、本明細書の図においては、論理ボリューム状態の数値を記載する代わりに、正常、異常

10

20

30

40

50

等と記載する。

論理ボリュームが正常な状態にあるとは、論理ボリュームに対してデータを正しく記録でき、またデータを正しく読み出せることを言う。論理ボリュームが異常な状態にあるとは、論理ボリュームが正常でない状態を言う。論理ボリュームの異常な状態は、HDD等の障害により発生しうる。

【0053】

容量は、論理ボリュームに記録できる最大のデータ量である。

物理アドレスは、論理ボリュームを構成しているHDDグループの識別番号(ID)、および、HDDグループにおける論理ボリュームの開始アドレスである。図4では「HDDグループのID：開始アドレス」の形式で表記している。HDDグループのIDは、HDDグループを構成する先頭のHDDの識別番号と同一とする。 10

【0054】

図4に示す例では、論理ボリュームは、計算機9からアクセスされる論理ボリュームであり、ペアの正論理ボリューム121であり、ボリューム状態が正常、容量が100ギガバイト(GB)、HDDグループ111の先頭(アドレス0)からデータが格納されることを示している。また、論理ボリューム情報105には、その他の論理ボリューム124、122、123の構成に関する情報が記録されている。

【0055】

図5は、副記憶システム2の論理ボリューム管理情報205の内容をテーブル形式で示す図である。論理ボリューム管理情報205のデータ形式は、正記憶システム1の論理ボリューム管理情報105と同じであるため、説明を省略する。 20

【0056】

図6は、正記憶システム1のHDDグループ管理情報104に記録されている情報をテーブル形式で示す図である。HDDグループ管理情報104は、正記憶システム1の各HDDグループの構成情報を記録している。

各HDDグループの構成情報は、HDDグループID、開始HDD、HDD数、RAIDレベルの情報からなる。

HDDグループIDは、記憶システム内のHDDグループそれぞれに一意に割り当てる識別番号(ID)である。

開始HDDは、HDDグループを構成するHDDの、識別番号が最も小さいものの、HDD IDである。理解を助けるため、HDDグループIDと開始HDDは同じ値を取るものとしている。 30

HDD数は、HDDグループに含まれるHDDの数である。本実施形態では、どのHDDグループも、開始HDDによって示されるHDD IDから、HDD数に示される数の、HDD IDが連続しているHDDから構成されるものとする。

RAIDレベルは、HDDグループに適用されている、RAID (Redundant Array of Inexpensive Disks) によるデータ冗長化方式のレベルを示す数字である。

図6では、HDDグループ管理情報104は、HDDグループIDが111であり、開始HDDのIDが111、HDD数が4、RAIDレベルが5であるHDDグループの情報を記録している。また、HDDグループ管理情報104は、HDDグループIDが115であるHDDグループの情報を記録している。 40

【0057】

ID111のHDDグループと、その論理ボリュームの構成は、図7のように表わすことができる。図7において、111、112、113、114の番号により指示される実線の矩形は、それぞれHDD111、112、113、114の記憶領域を表わしている。図中で、記憶領域のアドレスは、上から下へ付けられているとする。121と122の番号により指示される点線の矩形は、それぞれ論理ボリューム121と122が、HDD111、112、113、114の記憶領域から構成されていることを表わしている。論理ボリューム121は、HDDグループの先頭の記憶領域から100GB分の領域を占有し、論理ボリューム122は、HDDグループの(論理的な)アドレス0xC80000 50

0 から 1 0 0 G B 分の領域を占有している。

【 0 0 5 8 】

図 8 は、副記憶システム 2 の H D D グループ管理情報 2 0 4 に記録される情報を、テーブルの形式で示している。副記憶システム 2 の H D D グループ管理情報 2 0 4 は、正記憶システム 1 の H D D グループ管理情報 1 0 4 と形式は同じであるため、説明を省く。

【 0 0 5 9 】

図 9 は、正記憶システム 1 の更新データグループ管理情報 1 0 6 に記録する情報を、テーブル形式で示す図である。更新データグループ管理情報 1 0 6 は、正記憶システム 1 が制御している更新データグループの構成に関する情報を記録している。

一つの更新データグループの構成に関する情報は、更新データグループ I D、更新データグループ状態、ペア I D、更新データ論理ボリューム I D、更新番号の各情報から成る。

更新データグループ I D は、正記憶システム 1 と副記憶システム 2 の間で、更新データグループを一意に識別するための識別番号 ( I D ) である。

更新データグループ状態は、更新データグループの状態を示す数値である。更新データグループの状態には、正常 ( 1 )、異常 ( 2 ) がある ( 括弧内は、状態を表す数値 )。更新データグループが正常な状態とは、更新データグループを構成している更新データ論理ボリュームに対する更新データの書き込み、同更新データ論理ボリュームからのデータの読み出し、更新データの送信が正常に実行できる状態を言い、更新データグループが異常な状態とは、これらの動作を実行できないことを言う。更新データグループが異常な状態は、H D D の障害、ネットワークの障害等により発生しうる。

理解を助けるために、本明細書においては、更新データグループの状態を数値で記載する代わりに、正常、異常等と記載する。

【 0 0 6 0 】

ペア I D は、更新データグループに関連付けられているペアの識別番号 ( I D ) である。更新データグループにペアが関連付けられている場合、そのペアの更新データは、関連付けられた更新データグループを構成する更新データ論理ボリュームに記録される。

なお、ペア I D は、正記憶システム 1 と副記憶システム 2 の間で、ペアを一意に識別するために、共通の値を用いる。

更新データ論理ボリューム I D は、更新データグループを構成する更新データ論理ボリュームの I D である。更新データ論理ボリュームは、複数でもよい。

更新番号は、更新データ論理ボリュームに最後に記録された更新データを示す、更新データに割り振られた番号である。更新データには、連続した一意の番号を割り振る。

図 9 の例は、更新データグループ I D が 1 の更新データグループは、状態が正常、関連付けられたペアの I D は 1、更新データ論理ボリュームは更新データ論理ボリューム 1 2 4 であることを示している。

【 0 0 6 1 】

図 1 0 は、副記憶システム 2 の更新データグループ管理情報 2 0 6 に記録する情報を、テーブル形式で示す図である。更新データグループ管理情報 2 0 6 のデータ形式は、正記憶システム 1 の更新データグループ管理情報 1 0 6 と同じであるため、説明を省く。ただし、更新データグループ管理情報 2 0 6 の更新番号は、更新データ論理ボリュームから読みだすべき更新データの番号を記録する。

また、更新データグループ管理情報 2 0 6 には、正閾値と副閾値のデータを追加している。正閾値は、更新データグループに含まれる正更新データ論理ボリューム)の使用率 ( 単位 : % ) に対する閾値である。副閾値は、更新データグループに含まれる、副更新データ論理ボリュームの使用率に対する閾値である。これらの値は、副更新データ論理ボリューム、および副更新データを構成する H D D の電源を投入するか否かを決定するために使用する。これらの値を使用した処理については、図 1 8 および図 1 9 を用いて後に詳細に説明する。

【 0 0 6 2 】

10

20

30

40

50

図 1 1 は、正記憶システム 1 のペア情報 1 0 7 に記録するペアに関する情報を、テーブル形式で示した図である。各ペア情報は、正記憶システム 1 と副記憶システム 2 の間の各ペアの構成に関する情報を記録する。

ペア情報 1 0 7 は、各ペアに関してペア I D、ペア状態、正記憶システム 1 I D、正論理ボリューム I D、副記憶システム 2 I D、副論理ボリューム I D、更新データグループ I D の情報を記録する。

ペア I D は、正記憶システム 1 と副記憶システム 2 の間でペアを一意に識別するための識別番号 ( I D ) である。

ペア状態は、ペアの状態を示す数値である。ペアの状態には、正常 ( 1 )、異常 ( 2 )、停止 ( 3 ) がある ( 括弧内は、対応する数値 )。ペアの状態が正常であるとは、ペアの更新データが正記憶システム 1 と副記憶システム 2 の間で正常に送受信できる状態を言う。ペアの状態が異常であるとは、ネットワークや HDD 等の障害により、正記憶システム 1 と副記憶システム 2 の間で更新データを正常に送受信できない状態を言う。ペアの状態が停止であるとは、正論理ボリューム 1 2 1 へのデータの書き込みが、副論理ボリュームへ更新されない状態を言う。ただし、未送信の更新データに関しては、正記憶システム 1 と副記憶システム 2 の間で送受信が実行される可能性がある。

理解を助けるために、本明細書においては、ペア状態の数値を記載する代わりに、正常、異常、停止等と記載する。

正記憶システム 1 I D は、ペアの正論理ボリューム 1 2 1 を有する正記憶システム 1 の識別番号 ( I D ) である。説明を簡単にするために、本明細書では、図 1 における正記憶システム 1 の符号の指示番号 1 を、正記憶システム 1 の識別番号とする。同様に、副記憶システム 2 の識別番号を、2 とする。

正論理ボリューム I D は、ペアの正論理ボリュームの I D である。

副記憶システム 2 I D は、ペアの副論理ボリュームを有する副記憶システム 2 の識別番号 ( I D ) である。

副論理ボリューム I D は、ペアの副論理ボリュームの I D である。

更新データグループ I D は、ペアに関連付けられている更新データグループの識別番号である。

#### 【 0 0 6 3 】

図 1 1 の例では、ペア I D が 1 であるペアは、ペア状態が正常、ペアの正記憶システムが正記憶システム 1、ペアの正論理ボリュームが論理ボリューム 1 2 1、ペアの副記憶システムが 2、ペアの副論理ボリュームが論理ボリューム 2 2 1、ペアの更新データグループが、I D 1 で識別される更新データグループであることを示している。

副記憶システム 2 のペア情報 2 0 7 には、正記憶システム 1 との間で構成されているペア ( I D 1 ) の情報を、正記憶システム 1 のペア情報 1 0 7 と同じ形式で記録する。

論理ボリューム情報 1 0 5 および 2 0 5 の論理ボリュームタイプ、更新データグループ管理情報 1 0 6 および 2 0 6、ペア情報 1 0 7 および 2 0 7 は、ユーザが管理端末機器を用いてペアの構成を指示した時、その指示に基づき内容を変更する。

#### 【 0 0 6 4 】

図 1 2 は、正記憶システム 1 の更新データグループ進捗情報 1 0 8 に記録する情報を、テーブル形式で示している。更新データグループ進捗情報 1 0 8 は、各更新データグループにおける更新データの送受信の進捗を管理するための情報である ( 副記憶システム 2 にも更新データグループ進捗情報 2 0 8 があるが、これは更新データの副論理ボリュームへの適用の進捗を管理する情報として使用する )。

更新データグループは、更新データグループ I D、書き込みアドレス、読み出しアドレスの情報を各更新データグループに関して記録する。

更新データグループ I D は、更新データグループの識別番号である。

書き込みアドレスは、更新データを記録すべき更新データ論理ボリュームの識別番号と、そのアドレスからなる情報である。

10

20

30

40

50

読み込みアドレスは、更新データを最後に読み出した更新データ論理ボリュームの識別番号と、そのアドレスからなる情報である。

【0065】

図12では、書き込みアドレスと読み込みアドレスを、「更新データ論理ボリュームの識別番号：更新データ論理ボリュームのアドレス」の形式で記載している。

本実施形態において、更新データ論理ボリュームをいわゆるリングバッファのように使用する。上記の書き込みアドレスは、更新データを次に記録するアドレスであり、更新データを記録するごとに加算する。また読み込みアドレスは、更新データを次に読み出すアドレスであり、更新データをそのアドレスから読み出すごとに加算する。これらのアドレスが更新データグループに含まれる記憶領域の容量を超えた場合には、適切な先頭アドレスにリセットする。

更新データグループが複数の更新データ論理ボリュームを含む場合、更新データ論理ボリュームの識別番号が小さい順に、連続した記憶領域として扱う。

【0066】

図12の例では、更新データグループIDが1の更新データグループは、書き込みアドレスが更新データ論理ボリューム124の500、読み込みアドレスが更新データ論理ボリューム124の100となっている。

ここで、更新データグループの使用率について説明する。更新データグループの使用率とは、記憶システムにおいて、更新データグループに属する更新データ論理ボリュームに記録可能な更新データの最大数に対して、実際に記録されている更新データの数の比率のことを言う。この値は、記憶システム個別に算出する値であり、正更新データ論理ボリュームと副更新データ論理ボリュームの合算から算出する値ではない。

更新データグループに記録可能な更新データの最大値は、更新データグループに含まれる更新データ論理ボリュームの容量の合計と、更新データのデータ長（後で述べるように、更新データは固定長のデータである）により定まる。本実施形態では、更新データ論理ボリュームから読み出され、正常に処理された更新データは破棄されるものとする。更新データグループに記録されている更新データの数は、更新データグループ進捗情報の書き込みアドレスと読み込みアドレスから算出する。これらの値から、更新データグループの使用率を算出できる。

【0067】

図13は、副記憶システム2の更新データグループ進捗情報208に記録する情報を、テーブル形式で示す図である。更新データグループ進捗情報208は、正記憶システム1の更新データグループ進捗情報108とデータ形式が同じであるため、説明を省略する。

【0068】

図14は、副記憶システム2のHDD電源制御情報225に記録する情報を、テーブル形式で示す図である。HDD電源制御情報225は、副記憶システム2が有するHDDの個々に関して、HDD ID、最終アクセス時刻、電源投入回数の各情報を記録している。

HDD IDは、HDDを識別するためのIDである。

最終アクセス時刻は、HDDへのアクセス（データの書き込みまたはデータの読み出し）が発生した時刻である。この情報は、ディスクアダプタ209の論理ボリューム制御部57が、HDDへのアクセスを行う度に、その時刻を記録する。

電源投入回数は、電源が切断されたHDDへ電源を投入した回数であり、0以上の値をとる。また、この値は一定の時間間隔ごとに減少するようにしている。この理由は後述する。これらの処理はディスクアダプタ209のHDD電源制御部56が行う。

【0069】

図15は、更新データのデータ形式800を示す図である。更新データは、更新番号、ペアID、アドレス、データ長、ライトデータの各情報を含む。

更新番号は、更新データに割り振られた、一意の連続した番号である。更新データの更新番号は、更新データを生成する際に、直前に生成した更新データの更新番号の次の値と

10

20

30

40

50

なるように決定する。更新番号として取り得る値には最大値を定め、前記方法により更新番号を決定する際、最大値よりも大きい値となる場合は、その更新番号を0とする。前記最大値は、更新データ論理ボリュームに記録される更新データの更新番号が重複しない程度の大きさとする。

ペアIDは、更新データの適用の対象である副論理ボリュームが構成しているペアの識別番号である。

アドレスは、更新データに基づく更新を実施する論理ボリュームの、更新適用の対象となる先頭アドレスである。

データ長は、更新データに含まれるライトデータの有効なデータ長である。

ライトデータは、更新時に副論理ボリュームに書き込むデータの内容である。

10

更新番号、ペアID、アドレス、データ長からなる情報を、更新データヘッダと呼ぶ。

更新データは固定長のデータである。更新データヘッダとライトデータを含むデータ量が、更新データの固定長のサイズよりも大きい場合、ライトデータを適切に二つ以上に分割し、二つ以上の更新データを生成する。逆に、更新データヘッダとライトデータを含むデータ量が、更新データの固定長のサイズよりも小さい場合、ライトデータの余りの領域には0をセットする。

#### 【0070】

以降、本実施形態1における処理の流れを、フローチャートを参照しつつ説明する。

図16は、正記憶システム1のI/O処理部51における、計算機9からのデータ書き込み指示を受けた際の処理の流れを示すフローチャートである。

20

正記憶システム1のホストアダプタ101は、計算機9よりデータの書き込みの要求を受けると、I/O処理部51にてステップS101からの処理を開始する。計算機9からの要求には、情報として、データを書き込む対象の論理ボリュームを指定する情報、書き込むデータ(ライトデータ)とそのデータ長、データの書き込みを開始する論理ボリュームのアドレスが含まれる。

#### 【0071】

以降の説明では、計算機9は、正論理ボリューム121へのデータの書き込みを要求することを想定する。以降、特に断わらない限り、I/O処理部51と言う場合はホストアダプタ101のI/O処理部を指す。また、計算機9よりデータの書き込みを指示された論理ボリュームを対象論理ボリュームと呼ぶ。

30

ステップS101において、I/O処理部51は、計算機9からの要求内容を、キャッシュメモリ110に記録する。ステップS102において、I/O処理部51は、対象論理ボリュームがデータを書き込める論理ボリュームであるかどうかを判定する。データを書き込める論理ボリュームとは、論理ボリュームタイプがホスト、ホスト(正)のいずれかである論理ボリュームを言う。この判定には、論理ボリューム管理情報105に記録されている、論理ボリューム121の論理ボリュームタイプ情報を使用する(図4参照)。

#### 【0072】

図4では、論理ボリューム121の論理ボリュームタイプは、ホスト(正)である。この場合、処理はステップS104に進む。もし論理ボリューム121の論理ボリュームタイプがホスト(副)等である場合、I/O処理部51は計算機9に対して、要求が無効であることを通知し処理を終了する(ステップS103)。

40

ステップS104において、I/O処理部51は、対象論理ボリュームの論理ボリューム状態が正常であるかどうかを判定する。この判定には、論理ボリューム管理情報105に記録されている、論理ボリューム121の論理ボリューム状態情報を使用する。

#### 【0073】

図4では、論理ボリューム121の論理ボリューム状態は正常である。この場合、処理はステップS106に進む。もし論理ボリューム121の論理ボリュームタイプが異常である場合、I/O処理部51は、計算機9に対して対象論理ボリュームが異常であることを通知して処理を終了する(ステップS105)。

ステップS106において、I/O処理部51は、対象論理ボリュームがペアの正論理

50

ボリューム 1 であるかどうかを判定する。この判定には、論理ボリューム管理情報 1 0 5 に記録されている、論理ボリューム 1 2 1 の論理ボリュームタイプ情報を使用する。

【 0 0 7 4 】

図 4 では、論理ボリューム 1 2 1 の論理ボリュームタイプは、ホスト ( 正 ) であり、論理ボリューム 1 2 1 はペアの正論理ボリューム 1 である。この場合、処理はステップ S 1 0 7 に進む。もし論理ボリューム 1 2 1 の論理ボリュームタイプがホスト ( 正 ) でないならば、処理はステップ S 1 0 8 に進む。

ステップ S 1 0 7 において、I / O 処理部 5 1 は、計算機 9 からの要求から更新データを生成し、更新データ論理ボリュームに記録する。この処理の詳細については、図 1 7 を用いて後述する。ステップ S 1 0 8 において、I / O 処理部 5 1 は、対象論理ボリューム 10 10 に対して、計算機 9 から指示されたデータを記録する。このステップでは、ディスクアダプタ 1 0 9 の論理ボリューム制御部 5 7 を使用する。I / O 処理部 5 1 は、ステップ S 1 0 1 において計算機 9 からの要求内容を記録したキャッシュメモリ 1 1 0 上のアドレスと、論理ボリュームへのデータの書き込みの要求を、ディスクコントローラ 1 0 9 の論理ボリューム制御部 5 7 に通知する。この要求を受けて、ディスクコントローラ 1 0 9 の論理ボリューム制御部 5 7 は、論理ボリューム管理情報 1 0 6 から論理ボリュームの構成 ( H D D の構成 ) を得、キャッシュメモリ 1 1 0 に記録されている計算機 9 からの要求内容を参照し、論理ボリューム 1 2 1 ( を構成する H D D ) にライトデータを記録する。

その後、計算機 9 に対して正常に処理が終了したことを通知して処理を終了する ( ステップ S 1 0 9 ) 。

【 0 0 7 5 】

図 1 7 は、図 1 6 のステップ S 1 0 7 ( 更新データ記録処理 ) の詳細な処理の流れを示すフローチャートである。

更新データを記録する処理は、ステップ S 2 0 1 より開始する。なお、説明を簡単にするために、以降は論理ボリューム、更新データ論理ボリュームが異常状態である場合の処理の説明を省略する。

【 0 0 7 6 】

ステップ S 2 0 1 において、対象論理ボリュームが構成しているペアを特定する。この特定処理ではペア情報 1 0 7 を用いる ( 図 1 1 参照 ) 。ペア情報 1 0 7 を探索し、正論理ボリューム 1 が 1 2 1 であるペアを見つけ出す。

ステップ S 2 0 2 において、前ステップで特定したペアと関連付けられている更新データグループを特定する。更新データグループは、ペア情報 1 0 7 を参照することで特定できる。

ステップ S 2 0 3 ) において、前ステップで特定した更新データグループの書き込みアドレスと読み込みアドレスを、更新データグループ進捗情報 1 0 8 から取得する。

ステップ S 2 0 4 において、前ステップで取得した書き込みアドレスと読み込みアドレスから、更新データグループに含まれる更新データ論理ボリュームに、更新データ論理ボリュームに更新データを書き込めるだけの空き記憶領域があるかどうかを判定する。

【 0 0 7 7 】

空き記憶領域がない場合、ステップ S 2 0 5 に進む。ステップ S 2 0 5 は、更新データを書き込める記憶領域ができることを期待し、一定時間待機する。その後、再度ステップ S 2 0 4 の判定を行う。空き記憶領域がない場合の別の処理方法として、ペアのペア状態を " 停止 " に変更し、更新データの記録を停止する、または副記憶システム 2 に対して更新データを取得するように要求する方法を採用することもできる。

ステップ S 2 0 6 において、計算機 9 からのデータ書き込み要求に基づいて更新データを生成し、キャッシュメモリ 1 1 0 に記録する。

【 0 0 7 8 】

更新データを構成する諸情報を次のように決定する。更新番号を、1 0 6 に記録されている、上記で特定した更新データグループの更新番号とする。ペア ID を、上記で特定したペアの ID とする。アドレスを、計算機 9 に指示されたアドレスとする。データ長を、

10

20

30

40

50

計算機 9 に指示されたライトデータのデータ長とする。ライトデータを計算機 9 に指示されたライトデータとする。

上記の諸情報は、ステップ S 1 0 1 でキャッシュメモリに記録した、計算機 9 からの要求内容から得る。

ステップ S 2 0 7 において、I/O 処理部 5 1 は、ディスクアダプタ 1 0 9 の論理ボリューム制御部 5 7 に対し、キャッシュメモリ 1 1 0 に記録した更新データを、前ステップで取得した書き込みアドレス（更新データ論理ボリュームとそのアドレス）に記録することを指示する。ディスクアダプタ 1 0 9 の論理ボリューム制御部 5 7 は、前記指示を受け、更新データ論理ボリュームに更新データを書き込む。

#### 【 0 0 7 9 】

ステップ S 2 0 8 において、I/O 処理部 5 1 は、更新データグループ進捗情報 1 0 8 に記録されている更新データグループの書き込みアドレスを、次の書き込みアドレス（次回更新データを書き込む更新データ論理ボリュームおよびそのアドレス）へ変更する。また、更新データグループ管理情報 1 0 6 の更新番号を次の値にセットする。次の書き込みアドレスは、現在の書き込みアドレスに、更新データのデータ長を加算して算出する。

この加算の結果、次の書き込みアドレスが最大値を超える場合は、次の書き込みアドレスは、更新データグループの先頭の更新データ論理ボリュームと、その先頭アドレス（0）となる。書き込みアドレスの最大値は、更新データグループに含まれる更新データ論理ボリュームの容量により定まる。

以上の処理を実行した後、処理を終了する（ステップ S 2 0 9 ）。

対象論理ボリュームが複数のペアを構成している場合、各ペアに関して上記処理を繰り返す。

#### 【 0 0 8 0 】

図 1 8 は、副記憶システム 2 において、正記憶システム 1 に記録されている更新データを取得し、副更新データ論理ボリュームへ記録する処理（更新データ転送処理）の流れを示すフローチャートである。

この処理では、副更新データ論理ボリュームを構成している HDD の電源の状態に応じた処理を行う。この処理は、副記憶システム 2 の更新データ転送調停部 5 3（以降、副更新データ転送調停部）が、全更新データグループを対象に一定の時間間隔で定期的に行う。

#### 【 0 0 8 1 】

ステップ S 3 0 1 より、更新データ転送処理を開始する。

ステップ S 3 0 2 において、副記憶システム 2 に設定されている更新データグループの内、一つを選択する。選択した更新データグループを、以降対象更新データグループと呼ぶ。

ステップ S 3 0 3 において、対象更新データグループの更新データを、正記憶システム 1 から取得するか否かを決定する。本実施形態では、この判定方法の一つとして、正記憶システム 1 における対象更新データグループの使用率に基づく方法を採用する。以下、その処理の流れを説明する。

#### 【 0 0 8 2 】

副更新データ転送調停部は、（正記憶システム 1 の）対象更新データグループの使用率を、正記憶システム 1 のホストアダプタ 1 0 2 の更新データ転送調停部 5 3（以降、正更新データ転送調停部）より取得する。この使用率を、更新データグループ管理情報 2 0 6 に記録されている対象更新データグループの正閾値よりも大きいかどうかを判定する。もし大きければ、更新データを取得することとし、小さければ取得は行わない。

#### 【 0 0 8 3 】

正更新データ転送調停部は、更新データグループ進捗情報 1 0 8 に記録されている書き込みアドレスと読み込みアドレス、および更新データグループに含まれる正更新データ論理ボリュームの容量の合計から、使用率を算出する。

10

20

30

40

50

ステップ S 3 0 4 において、前ステップでの決定にしたがい、S 3 0 5 またはステップ S 3 0 7 に進む。

【 0 0 8 4 】

ステップ S 3 0 5 において、更新データを記録すべき更新データ論理ボリュームに関し、その更新データ論理ボリュームを構成している H D D の電源を投入する。このステップを以下に示す。

【 0 0 8 5 】

副更新データ転送調停部は、更新データグループ進捗情報 2 0 8 に記録されている、対象更新データグループの書き込みアドレスから、更新データを記録すべき更新データ論理ボリュームを特定する。そして、その更新データ論理ボリュームを構成している個々の H D D に対して、電源が切断されていれば、その電源を投入するように、ディスクアダプタ 2 0 9 の H D D 電源制御部 5 6 に指示する。論理ボリュームを構成する H D D は、論理ボリューム管理情報 2 0 5 より取得できる。各 H D D の電源の状態は、ディスクアダプタ 2 0 9 の H D D 電源制御部 5 6 が把握しており、副更新データ転送調停部は、H D D の電源状況を、ディスクアダプタ 2 0 9 に問い合わせることで認識することができる。

10

【 0 0 8 6 】

ディスクアダプタ 2 0 9 の H D D 電源制御部 5 6 は、上記指示を受け、指定された論理ボリュームを構成する H D D の電源を投入する。

ステップ S 3 0 6 において、正記憶システム 1 より更新データを取得し、更新データ論理ボリュームに書き込む。このステップを以下に示す。

20

【 0 0 8 7 】

副更新データ転送調停部は、正更新データ転送調停部に対して、対象更新データグループの更新データを送信するように要求する。この要求を受け、正更新データ調停部は、ホストアダプタ 1 0 2 の更新データ送信部 5 2 に対し、対象更新データグループの更新データの送信を指示する。以降、特に断わらない限り、更新データ送信部 5 2 は、ホストアダプタ 1 0 2 の更新データ送信部を指す。

【 0 0 8 8 】

更新データ送信部 5 2 は、対象更新データグループに含まれる正更新データ論理ボリュームに記録された更新データであって、最後に副記憶システム 2 に送信された更新データの次の更新データを、副記憶システム 2 へ送信する。このような更新データが記録された更新データ論理ボリュームとそのアドレスは、更新データグループ進捗情報 1 0 8 の読み込みアドレスより判断できる。

30

【 0 0 8 9 】

対象更新データグループに更新データが含まれていない場合、更新データ送信部 5 2 は、更新データが存在しないことを副記憶システム 2 に通知する。更新データの送信が終了した場合は、更新データグループ進捗情報 1 0 8 の読み込みアドレスの値を、送信した更新データが記録されていたアドレスに変更する。

【 0 0 9 0 】

上記のように正記憶システム 1 により送信され、副記憶システム 2 により受信された更新データは、副記憶システム 2 のホストアダプタ 2 0 1 の更新データ受信部 5 4 が、更新データ論理ボリュームに書き込む。

40

【 0 0 9 1 】

更新データ受信部 5 4 が更新データを受信し、適切な更新データ論理ボリュームに書き込む処理は、図 1 7 に示したステップ S 2 0 3 ~ S 2 0 9 とほぼ同じであるため説明を省略する。ただし、ステップ S 2 0 6 のように更新データの生成は行わず、更新データ論理ボリュームに記録する更新データを、正記憶システム 1 から送信された更新データとする。

【 0 0 9 2 】

ステップ S 3 0 6 の更新データ受信処理は、正更新データ論理ボリュームの使用率が、例えば 5 % 未満程度となるまで繰り返す。

50

ステップS 3 0 2 からステップS 3 0 6 までの処理を、副記憶システム 2 の全ての更新データグループに対して実施する（ステップS 3 0 7）。

以上のステップにより、更新データを取得する処理を完了する（ステップS 3 0 8）。

【0093】

上記のステップS 3 0 3 では、更新データを取得するか否かを、正記憶システム 1 における更新データグループの使用率に基づいて判断していたが、副記憶システム 2 において管理される情報や状態を判断のために使用することで、HDDの電源の制御をより効率的にすることができる。

【0094】

HDDの電源の投入と切断を過剰に行うと、スピナップ、スピンドアウンによるスピンドルモータへの負荷を高め、HDDの寿命の短縮を招くことになる。そこで、HDDの電源の投入/切断の回数を低減するために、副更新データ論理ボリュームのHDDの電源がすでに投入されているならば、正記憶システム 1 における更新データグループの使用率に関わらず更新データを取得し、副更新データ論理ボリュームに記録する、という判断を行なってもよい。

10

【0095】

副更新データ論理ボリューム 2 2 4 を構成するHDDの電源は、同じHDDグループに構成されている他の論理ボリューム（たとえば、論理ボリューム 2 2 2）へのアクセスを原因として、投入されることがある。上記の判断方法は、このような状況を効果的に活用する。

20

【0096】

図 1 9 は、更新データ論理ボリュームに記録されている更新データを、副論理ボリュームに記録されているデータに適用する処理（以降、リストア処理と言う）の処理の流れを示すフローチャートである。

リストア処理は、副記憶システム 2 のリストア処理部 5 5 が実行する。ホストアダプタ 2 0 1 と 2 0 2 の両方がリストア処理部 5 5 を持つが、どちらのリストア処理部 5 5 がリストア処理を実行してもよい。ここでは、ホストアダプタ 2 0 1 のリストア処理部 5 5 がリストア処理を実行するものとし、以降、特に断わらない限りは、リストア処理部 5 5 とは、ホストアダプタ 2 0 1 のリストア処理部 5 5 のことを指す。

【0097】

リストア処理は、前述の更新データ取得処理の終了直後に実行する。したがって、リストア処理実行時には、更新データグループに含まれる（副記憶システム 2 の）HDDの電源が投入されている。

30

ステップS 4 0 1 より、リストア処理を開始する。

ステップS 4 0 2 において、副記憶システム 2 に構成されている更新データグループを、リストア処理の対象として一つ選択する。ここで選択した更新データグループを、対象更新データグループと呼ぶ。

【0098】

ステップS 4 0 3 において、対象更新データグループのリストア処理を実行するか否かを判断する。この判断は、ここでは対象更新データグループの（副記憶システム 2 における）使用率、対象更新データグループを構成するHDDの電源の状態、対象更新データグループに関連づけられている副論理ボリュームの電源の状態を基にする。

40

【0099】

更新データグループの使用率は、更新データグループ進捗情報 2 0 8 に記録されている、対象更新データグループの書き込みアドレスと読み込みアドレス、および対象更新データグループに記録可能な更新データの最大数から算出する。このように算出した使用率が、更新データグループ管理情報 2 0 6 に記録されている対象更新データグループの副閾値を超えているならば、その更新データグループのリストア処理を行うものとする。

【0100】

ただし、対象更新データグループに関連づけられている副論理ボリュームのいずれかが

50

、そのHDDの電源が投入されているならば、その副論理ボリュームに関しては、対象更新データグループの使用率に関わらず、リストア処理を行うものとする。これにより、HDDに対する電源投入/切断の回数を低減する。

#### 【0101】

ステップS404において、前ステップで対象更新データグループに対するリストア処理を実行することを決定したならば、ステップS405へ進む。そうでなければ、ステップS411へ進む。

ステップS405において、対象更新データグループの更新データ論理ボリュームに記録された更新データを、キャッシュメモリに読み出す。この処理は、図18に示した更新データの取得処理のステップS306において、正記憶システム1が、副記憶システム2 10  
に対して更新データを送信するために、正更新データ論理ボリュームから更新データを読み出す処理とほぼ同様である。リストア処理部55は、対象更新データグループに含まれる副更新データ論理ボリュームに記録された更新データであって、直前に読み出した更新データの次の更新データを読み出す。

#### 【0102】

ステップS406において、上記のように読み出した更新データに対して、その更新データヘッダから、更新を実施すべき副論理ボリュームを特定し、副論理ボリュームへの更新データの適用(リストア処理)を実行するかどうかを判断する。もし副論理ボリュームを構成するHDDの電源が切断されているならば、リストア処理のためにHDDの電源を投入する必要があるが、前述のように、HDDの電源の投入/切断を過剰に繰り返すこと 20  
は好ましくない。そこで、ここでは、HDDの電源が投入されている副論理ボリュームに対してはリストア処理を実行し、HDDの電源が投入されていない副論理ボリュームに対しては、既定量(ユーザが設定できるものとする)以上の更新データが無く、かつ既定の目標復旧時点(RPO: Recovery Point Objective、ユーザが設定できるものとする)に達していない場合、リストア処理を延期することとする。

上記以外に、更新データグループ内の複数の副論理ボリューム間で、データの更新の順序をホストからの書き込みの順序と一致するようにし、データ間の時間的整合性を保つようにすることもできる。この場合、更新データの副論理ボリュームへの適用は、必ず更新データが更新データ論理ボリュームに記録された順序で行う。更新データグループにおける更新データ適用の過程で、処理中の更新データが対象としている副論理ボリュームのH 30  
D Dの電源が切断されている場合がある。このような場合、更新データ転送調停部53が、当該更新データグループにおける以降の更新データの適用を延期するか、もしくは電源が切断されている当該HDDの電源を投入して、更新データの適用を続行するかを判断する。上記の判断は、更新データ転送調停部53が、更新データの量に基づいて行う。

#### 【0103】

更新データヘッダにはペアIDが含まれている(図15参照)。これを用いて、ペア情報207から、ペアの副論理ボリュームのIDを取得できる。

ステップS407において、副論理ボリュームへのリストア処理を行うならばステップS408へ進み、そうでなければステップS409へ進む。

#### 【0104】

ステップS408において、副論理ボリュームに対して、更新データの内容に基づきデータの更新を実行する。この処理は以下の通りである。

リストア処理55は、リストア処理を実行する対象の副論理ボリュームを構成するHDDの電源が切断されているならば、そのHDDの電源を投入するように、ディスクアダプタ209のHDD電源制御部56に対して指示を行う。

ディスクアダプタ209のHDD電源制御部56は、指定された論理ボリュームのHDDを特定し、その電源を投入する。HDDを特定するために、論理ボリューム管理情報205を参照する。

#### 【0105】

リストア処理部55は、ディスクアダプタ209の論理ボリューム制御部57に対して 50

、ステップ S 4 0 5 にてキャッシュメモリ 2 1 0 に記録した更新データに基づき、副論理ボリュームのデータを更新するように指示する。この指示のパラメータとして、キャッシュメモリ 2 1 0 上の更新データを記録しているアドレスと、副論理ボリュームの ID を含める。ディスクアダプタ 2 0 9 の論理ボリューム制御部 5 7 は、キャッシュメモリ 2 1 0 に記録されている更新データが含むライトデータを、上記 ID により指定された副論理ボリュームの、更新データに記録されているアドレスから書き込む。

【 0 1 0 6 】

ステップ S 4 0 9 において、更新データの副論理ボリュームへの適用により、対象更新データグループの使用率が、既定値（たとえば 5 %、ユーザが設定できるものとする）以下になったかどうかを判定する。使用率が既定値以上である場合は、残りの更新データに関する処理を行うために S 4 0 5 へ戻る。

10

ステップ S 4 0 2 からステップ S 4 0 9 までの処理を、副記憶システム 2 の全ての更新データグループに対して実施する（ステップ S 4 1 0）。

【 0 1 0 7 】

図 2 0 は、HDD の電源を定期的に切断する処理（HDD 電源切断処理）の流れを示すフローチャートである。この処理は、副記憶システム 2 のディスクアダプタ 2 0 9 の HDD 電源制御部 5 6 が実行する。主記憶システム 1 にて同様の処理を行うこともできる。

HDD 電源切断処理は、一定の時間間隔で、副記憶システム 2 が有する全ての HDD を対象として実行する。

【 0 1 0 8 】

20

ステップ S 5 0 1 より、HDD 電源切断処理を開始する。

ステップ S 5 0 2 において、電源切断処理の対象とする HDD の ID を、HDD 電源制御情報 2 2 5 に記録されている HDD ID の中から選択する。ステップ S 5 0 2 からステップ S 5 0 7 までの処理は、HDD 電源制御情報 2 2 5 に記録されているすべての HDD ID を対象とするループ処理であり、HDD 電源制御情報 2 2 5 に記録されている順で、ループの先頭に戻る度に毎回異なる HDD ID を選択する。選択した HDD を対象 HDD と呼ぶ。

【 0 1 0 9 】

ステップ S 5 0 3 において、対象 HDD が最後にアクセスされた時刻から、一定時間（例えば 1 5 分、ユーザが設定できるものとする）が経過しているかどうかを判定する。この判定は、HDD 電源制御情報 2 2 5 に記録されている最終アクセス時刻と、現在時刻を比較することにより行う。一定時間が経過していればステップ S 5 0 4 に進み、そうでなければステップ S 5 0 6 に進む。

30

【 0 1 1 0 】

ステップ S 5 0 4 において、HDD 電源制御情報 2 2 5 に記録されている、対象 HDD の HDD 電源投入回数が、一定値（例えば 4、ユーザが設定できるものとする）以下かどうかを判定する。一定値以下であれば、ステップ S 5 0 5 に進み、そうでなければ、ステップ S 5 0 6 に進む。このステップの目的は、HDD に対する過剰な電源の投入 / 切断の繰り返しを防止することである。HDD 電源制御情報 2 2 5 の HDD 電源投入回数情報は、HDD の電源投入回数を制限するために記録している値であり、HDD の電源が投入されるごとに 1 加算され、一定の時間間隔（例えば 2 時間、ユーザが設定できるものとする）で 1 減算されるようにしている。

40

【 0 1 1 1 】

ステップ S 5 0 5 において、対象 HDD の電源を切断する。

ステップ S 5 0 6 において、HDD 電源制御情報 2 2 5 の HDD 電源投入回数情報に記録されている値を、1 減算する時刻に達しているか否かを、判定する。この時刻に達しているならばステップ S 5 0 7 に進み、そうでなければステップ S 5 0 8 に進む。この判定は、例えば、偶数の時間（1 4 : 0 0 など）の時刻を超たかどうかを毎回検査する方法により実施する。

【 0 1 1 2 】

50

ステップ S 5 0 7 において、対象 H D D の、H D D 電源制御情報 2 2 5 の H D D 電源投入回数に記録されている値を 1 減算する。

ステップ S 5 0 8 において、H D D 電源制御情報 2 2 5 に記録されている H D D について、ステップ S 5 0 3 から S 5 0 7 までの処理を行なったかどうかを判定する。全ての H D D について処理を行ったならば、H D D 電源切断処理を終了し (ステップ S 5 0 9 )、そうでなければ、ステップ S 5 0 2 に戻って、次の H D D を選択する。

【実施例 2】

【0 1 1 3】

次に、本発明の第二の実施形態を、図 2 1 を参照しながら説明する。

図 2 1 は、本発明の第二の実施形態の概要を表わすブロック図である。

第二の実施形態については、第一の実施形態との違いのみを示す。第二の実施形態と、第一の実施形態との違いは以下の通りである。

【0 1 1 4】

第二の実施形態では、更新データ論理ボリュームを使用しない非同期リモートコピーを使用する。第二の実施形態における I / O 処理部 5 1 では、計算機 9 からのデータ書き込み要求に基づく更新データを生成しない。その代わりに、正論理ボリューム 1 2 1 へデータを書き込む際に、書き込むデータと、正論理ボリューム 1 2 1 のどのアドレスにデータを記録したかの情報を、キャッシュメモリ 1 1 0 上のログ 1 2 5 に記録する。

【0 1 1 5】

更新データは、正記憶システム 1 から副記憶システム 2 へ送信するときに、正記憶システム 1 の更新データ送信部 5 2 により生成される。更新データ送信部 5 2 は、ログ 1 2 5 と正論理ボリューム 1 に記録されているデータより、更新データを生成する。

【0 1 1 6】

第一の実施形態では、更新データ論理ボリュームの使用率に基づき、更新データを送信するタイミングを決定していた。第二の実施形態では、正論理ボリューム 1 2 1 に記録されているデータの内、前回更新データを副記憶システム 2 に送信した時点から (計算機 9 からの要求により) 変更されたものの、全データ量に占める割合が一定の値を超えた時点、更新データを送信するタイミングとする。ログ 1 2 5 を参照することで、正論理ボリューム 1 2 1 のデータの更新された割合を判定できる。

【0 1 1 7】

上記、更新データ送信のタイミングに至ると、正記憶システム 1 の更新データ転送調停部 5 3 a は、副記憶システム 2 の更新データ調停部 5 3 に対して、更新データの送信を開始することを通知する。副記憶システム 2 の更新データ調停部 5 3 は、この通知を受けると、更新データの適用対象である副論理ボリューム 2 2 1 を構成する H D D の電源を (電源が切断されているならば) 投入する (破線矢印 6 7、6 4)。

ログ 1 2 5 と正論理ボリューム 1 に記録されているデータより生成された更新データは正記憶システム 1 の更新データ送信部 5 2 により送信され、副記憶システム 2 の更新データ受信部 5 4 により受信される。更新データに含まれている正論理ボリューム 1 からのデータは副論理ボリューム 2 2 1 に記録される。

【0 1 1 8】

あるいは、副記憶システム 2 は、すでに H D D の電源が投入されている副論理ボリュームへ優先的に更新データを適用するために、上記通知への返答において、更新データを優先的に送信することを所望するペアを、正記憶システム 1 へ通知してもよい。また、更新データを優先的に送信することを所望するペア、もしくは所望しないペアを、更新データの送信のタイミングに至る前に、副記憶システム 2 が、正記憶システム 1 へ通知してもよい。この通知は、副論理ボリュームを構成する H D D の電源が投入された、または切断されたことを契機とする。このような状況は、副論理ボリュームを構成している H D D グループ上に構成されている別の論理ボリュームに対して、計算機 9 等の上位装置からアクセスされる事等により発生する。

【実施例 3】

10

20

30

40

50

## 【0119】

次に、本発明の第三の実施形態を、図22を参照しながら説明する。

図22は、本発明の第三の実施形態の概要を表わすブロック図である。

第三の実施形態では、第二の実施形態における副記憶システム2（以降、記憶システム（2）と表記）を、仮想化機能を備える記憶システムとする。仮想化機能とは、外部記憶システムの論理ボリュームの一つまたは複数から、仮想的な論理ボリューム（仮想論理ボリューム）を形成する機能である。仮想論理ボリュームへのアクセスは、記憶システム（2）が、それを構成する外部の論理ボリュームへのアクセス要求に変換した上で、外部記憶システムへ要求することで、仮想ボリュームを実現する。記憶システム（2）の副論理ボリューム222は、記憶システム3（以降、記憶システム（3）と表記）の論理ボリューム321から形成される仮想ボリュームとなっている。記憶システム（2）と、外部記憶システム（3）の間には、通信路73が形成されている。

10

## 【0120】

第三の実施形態では、記憶システム（2）のHDD電源制御部56は、（仮想）論理ボリューム222に対する電源制御（電源の切断または投入）の要求を受けた場合、記憶システム（3）に対して、仮想論理ボリューム222を形成している論理ボリューム321の電源の制御を要求する（破線矢印67）。記憶システム（3）は、HDD電源制御部56cを備えており、HDD電源制御部56cは、前記電源制御の要求を受けて、論理ボリューム321を構成しているHDD311の電源を制御する（破線矢印66）。

## 【0121】

記憶システム（2）を、HDDを内部に持たない、仮想化機能のみを備える仮想化システムに置き換えて実施してもよい。また、第一の実施形態のように、更新データ論理ボリュームを使用したリモートコピーを使用し、更新データ論理ボリュームないしは副論理ボリュームないしはそれらの両方を仮想論理ボリュームとして、本発明を実施することもできる。

20

## 【0122】

第一の実施形態において、更新データの送信を、副記憶システム2が要求することで開始しているが、正記憶システム1が、副記憶システム2に更新データを送信することを通知することで、更新データの送信を開始してもよい。また、正記憶システム1が、副記憶システム2に、更新データの送信を通知することなく、副記憶システム2に更新データを送信するようにもできる。

30

## 【0123】

本明細書に述べた各実施形態では、更新データを送信するタイミングを、主としてデータ量に基づき決定している。上記タイミングの決定の判断を行うための情報として、ある時点でのデータ量を用いるほか、データ量の増加率を算出し、用いることができる。また、データ量およびデータ量の増加率と時刻の関係を統計的に算出してHDDの電源の投入が効率的となる時刻を計算し、HDDの電源の制御を効率的にすることもできる。また、HDDの電源を投入する時刻を、ユーザが記憶システムに事前に設定しておき、記憶システムは、この時刻に至ってHDDの電源を投入するようにしてもよい。

## 【0124】

そのほか、ユーザがペア状態を変更した時点、更新データを送信するタイミングとすることもできる。また、ユーザが、正論理ボリューム1への計算機9からのデータの書き込みが当面ないと判断し、正記憶システム1および副記憶システム2に対して更新データの転送を指示できる仕組みを設けることもできる。また、特定の時刻や時間間隔で、副記憶システム2が必ず更新データを受領するようにし、RPOが過去に成り過ぎることを防止することがより好ましい。

40

## 【0125】

本明細書に述べた各実施形態では、HDDの電源を切断するタイミングを、HDDへのアクセスの有無により判断しているが、図18や図19に示した処理が終了した時点、（副更新データ論理ボリュームや、副論理ボリュームへのデータの書き込みが終了した時点

50

)で、HDDの電源を切断するようにしてもよい。

【図面の簡単な説明】

【0126】

【図1】本発明の一実施形態の論理的な構成を示す図である。

【図2】正記憶システム1の構成を示す図である。

【図3】副記憶システム2の構成を示す図である。

【図4】正記憶システム1の論理ボリューム管理情報を示す図である。

【図5】副記憶システム2の論理ボリューム管理情報を示す図である。

【図6】正記憶システム1のHDDグループ管理情報を示す図である。

【図7】正記憶システム1のHDDグループの一つを表わす図である。

【図8】副記憶システム2のHDDグループ管理情報を示す図である。

【図9】正記憶システム1の更新データグループ管理情報を示す図である。

【図10】副記憶システム2の更新データグループ管理情報を示す図である。

【図11】正記憶システム1のペア情報を示す図である。

【図12】正記憶システム1の更新データグループ進捗情報を示す図である。

【図13】副記憶システム2の更新データグループ管理情報を示す図である。

【図14】副記憶システム2のHDD電源制御情報を示す図である。

【図15】更新データの形式を示す図である。

【図16】上位装置からの書き込み要求に対する処理の内容を示すフローチャートである。

。

【図17】更新データを正更新データ論理ボリュームへ記録する処理の内容を示すフローチャートである。

【図18】更新データを受信し、副更新データ論理ボリュームへ記録する処理の内容を示すフローチャートである。

【図19】更新データ論理ボリュームに記録されている更新データを副論理ボリュームに記録されているデータに適用するリストア処理の流れを示すフローチャートである。

【図20】HDDの電源を切断する処理の内容を示すフローチャートである。

【図21】本発明の第二の実施形態の論理的な構成を示す図である。

【図22】本発明の第三の実施形態の論理的な構成を示す図である。

【符号の説明】

【0127】

1 正記憶システム

2 副記憶システム

3 記憶システム

9 計算機

41, 42, 43 ネットワーク

51 I/O処理部

52 更新データ送信部

53, 53a, 53b 更新データ転送調停部

54 更新データ受信部

55 リストア処理部

56, 56b, 56c HDD電源制御部

57 論理ボリューム制御部

71, 72, 73 通信路

100, 200 コントローラ

101, 102, 201, 202 ホストアダプタ

103, 203 共有メモリ

104, 204 HDDグループ管理情報

105, 205 論理ボリューム管理情報

106, 206 更新データグループ管理情報

10

20

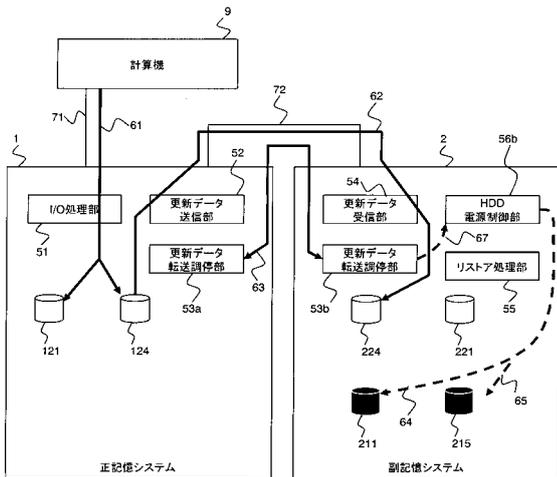
30

40

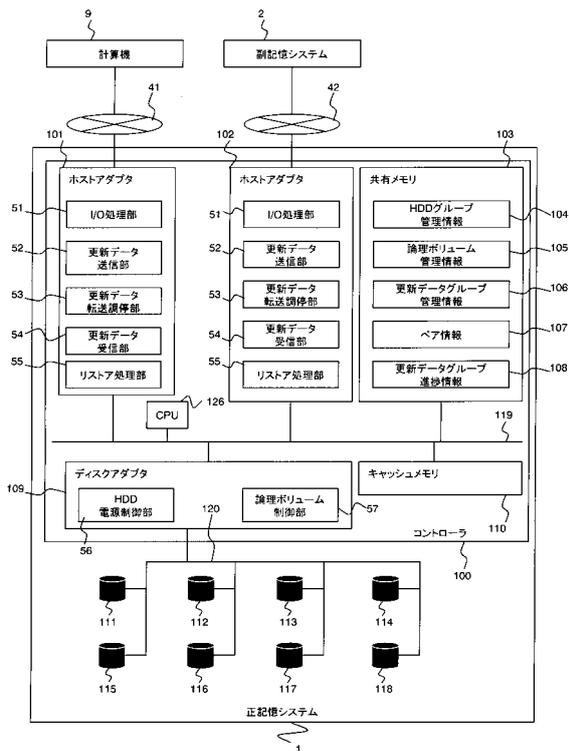
50

- 1 0 7 , 2 0 7      ペア情報
- 1 0 8 , 2 0 8      更新データグループ進捗情報
- 1 0 9 , 2 0 9      ディスクアダプタ
- 1 1 0 , 2 1 0      キャッシュメモリ
- 1 1 1 ~ 1 1 8 , 2 1 1 ~ 2 1 8      記憶装置
- 1 1 9 , 2 1 9      バス
- 1 2 0 , 2 2 0      HDD 接続線
- 1 2 1      正論理ボリューム
- 1 2 4      正更新データ論理ボリューム
- 1 2 5      ログ
- 1 2 6      CPU
- 2 2 1      副論理ボリューム
- 2 2 2      仮想論理ボリューム
- 2 2 4      副更新データ論理ボリューム
- 2 2 5      HDD 電源制御情報
- 2 2 6      CPU
- 3 1 1      外部記憶システム
- 3 2 1      論理ボリューム

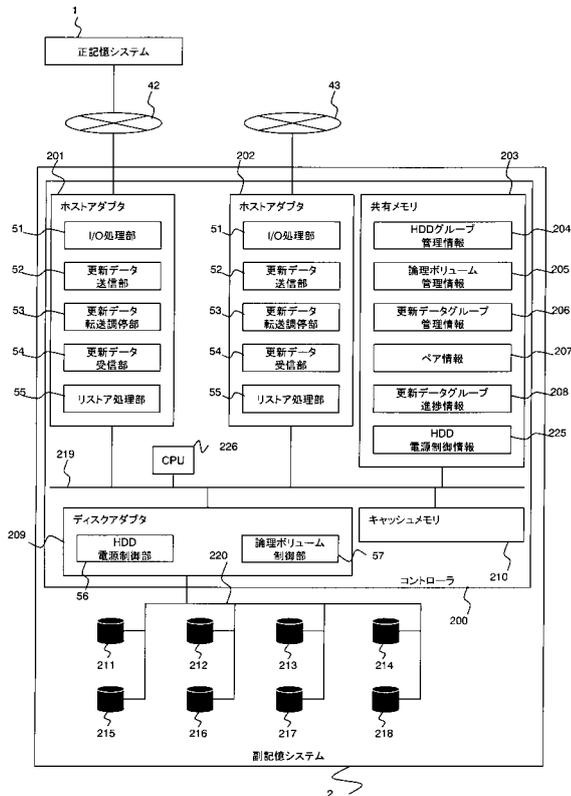
【 図 1 】



【 図 2 】



【図3】



【図4】

論理ボリュームID	論理ボリュームタイプ	論理ボリューム状態	容量	物理アドレス
121	ホスト(正)	正常	100GB	111:0
124	更新データ(正)	正常	100GB	115:0
122	ホスト	正常	100GB	111:C800000
123	ホスト	正常	100GB	115:C800000

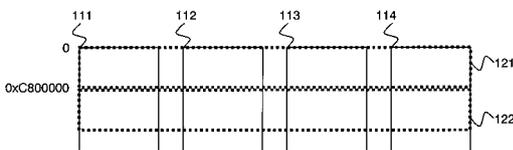
【図5】

論理ボリュームID	論理ボリュームタイプ	論理ボリューム状態	容量	物理アドレス
224	更新データ(副)	正常	100GB	211:0
221	ホスト(副)	正常	100GB	215:0
222	ホスト	正常	100GB	211:C800000
223	ホスト	正常	100GB	215:C800000

【図6】

HDDグループID	開始HDD	HDD数	RAIDレベル
111	111	4	5
115	115	4	5

【図7】



【図8】

HDDグループID	開始HDD	HDD数	RAIDレベル
211	211	4	5
215	215	4	5

【図12】

更新データグループID	書き込みアドレス	読み込みアドレス
1	124:500	124:400

【図9】

更新データグループID	更新データグループ状態	ペアID	更新データ論理ボリュームID	更新番号
1	正常	1	124	50

【図13】

更新データグループID	書き込みアドレス	読み込みアドレス
1	224:100	224:100

【図10】

更新データグループID	更新データグループ状態	ペアID	更新データ論理ボリュームID	更新番号	閾値1	閾値2
1	正常	1	224	40	75	75

【図14】

HDD ID	最終アクセス時刻	電源投入回数
211	7/21 0:05	2
215	7/20 23:36	4
230	5/4 0:00	0
231	5/6 0:00	0

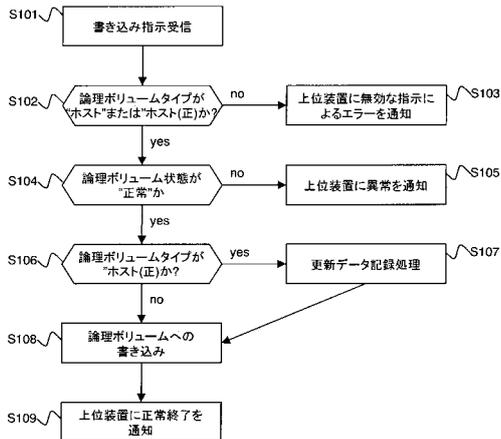
【図11】

ペアID	ペア状態	正統システムID	正統論理ボリュームID	副記憶システムID	副論理ボリュームID	更新データグループID
1	正常	1	121	2	221	1

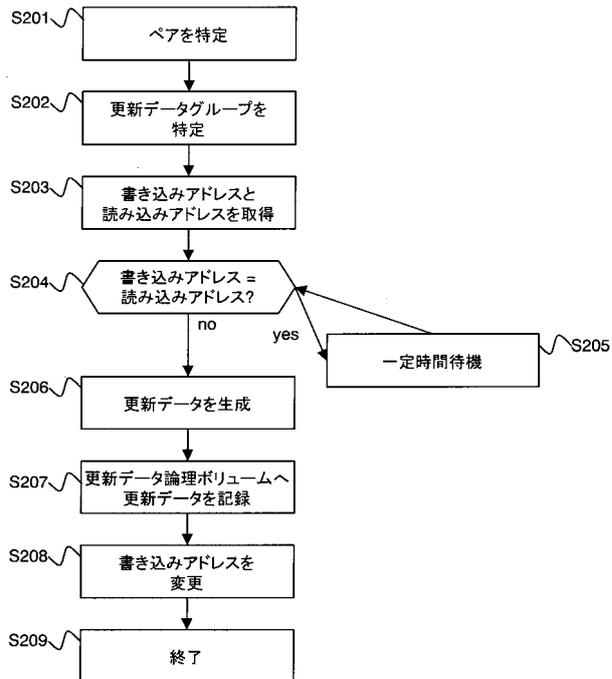
【図15】

更新番号	ペアID	アドレス	データ長	ライトデータ

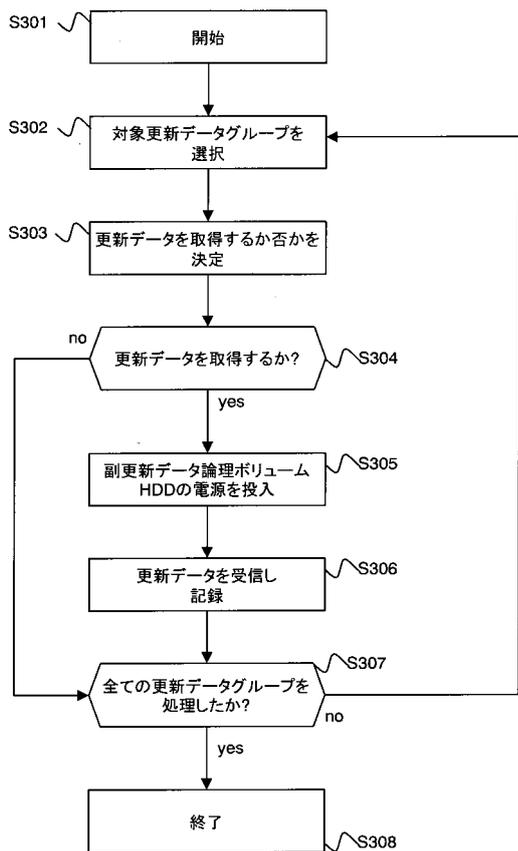
【 図 1 6 】



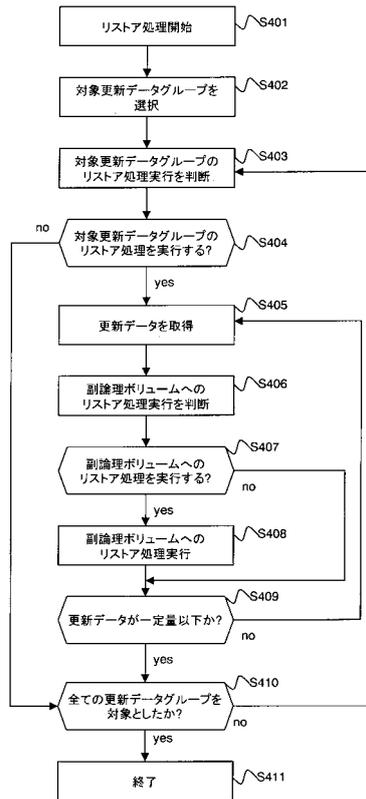
【 図 1 7 】



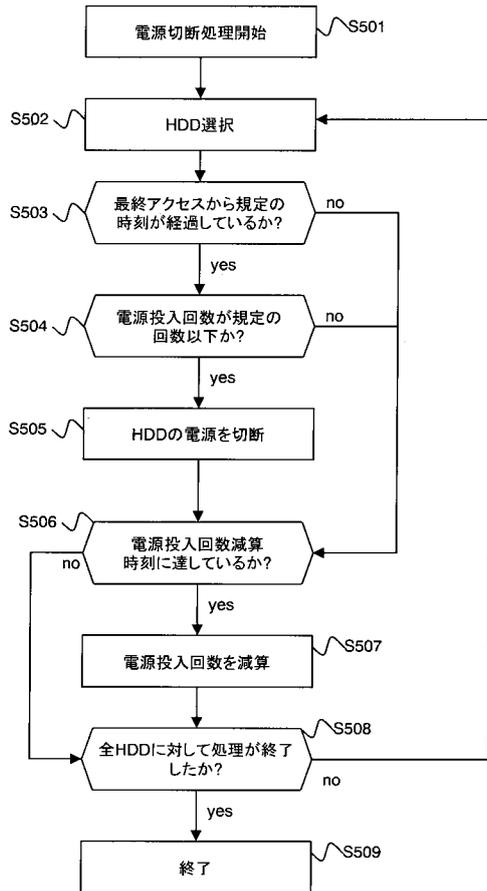
【 図 1 8 】



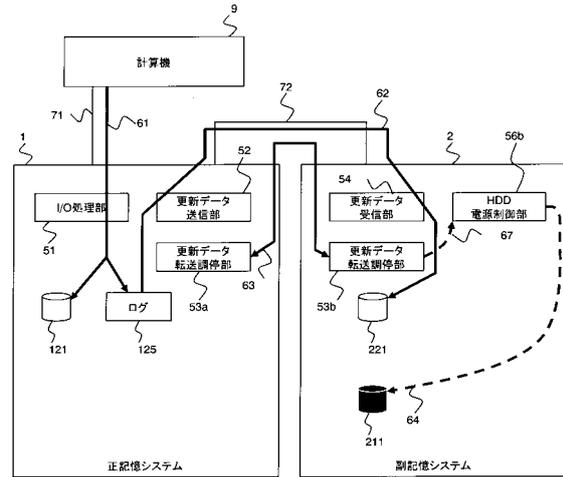
【 図 1 9 】



【図20】



【図21】



【図22】

