

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第5026102号
(P5026102)

(45) 発行日 平成24年9月12日 (2012. 9. 12)

(24) 登録日 平成24年6月29日 (2012. 6. 29)

(51) Int. Cl.

F I

G 0 6 F 3/06 (2006. 01)

G 0 6 F 12/08 (2006. 01)

G 0 6 F 12/16 (2006. 01)

G 0 6 F 3/06 3 O 4 Z

G 0 6 F 3/06 3 O 2 A

G 0 6 F 12/08 5 O 1 F

G 0 6 F 12/08 5 4 1 C

G 0 6 F 12/08 5 5 7

請求項の数 18 (全 28 頁) 最終頁に続く

(21) 出願番号 特願2007-27620 (P2007-27620)
 (22) 出願日 平成19年2月7日 (2007. 2. 7)
 (65) 公開番号 特開2008-192028 (P2008-192028A)
 (43) 公開日 平成20年8月21日 (2008. 8. 21)
 審査請求日 平成21年8月10日 (2009. 8. 10)

(73) 特許権者 000005108
 株式会社日立製作所
 東京都千代田区丸の内一丁目6番6号
 (74) 代理人 110000279
 特許業務法人ウィルフォート国際特許事務所
 (72) 発明者 飯田 純一
 神奈川県小田原市中里322番2号 株式
 会社日立製作所 R A I Dシステム事業部
 内
 (72) 発明者 姜 小明
 神奈川県小田原市中里322番2号 株式
 会社日立製作所 R A I Dシステム事業部
 内

最終頁に続く

(54) 【発明の名称】 ストレージ制御装置及びデータ管理方法

(57) 【特許請求の範囲】

【請求項 1】

外部装置からライトアクセス要求を受信し、ライトアクセス要求対象のデータを記憶装置へ書き込む制御を行うストレージ制御装置であって、

電源からの電力の供給を行う電力供給部と、

電力を供給可能に蓄えるバッテリーと、

電力の供給を受けてデータを記憶可能な揮発メモリであるキャッシュメモリと、

電力の供給を受けずともデータを記憶し続けることのできる不揮発メモリと、

前記電力の供給を受けて、ライトアクセス要求を受信する要求受信部と、

前記ライトアクセス要求の対象のデータをキャッシュデータとして前記キャッシュメモリに格納する第1データ格納部と、

前記キャッシュメモリの前記キャッシュデータのうちの前記記憶装置への反映がされていないダーティデータのデータ量が所定のしきい値を超えるか否かを判定する判定部と、

前記しきい値を超えると判定された場合に、前記キャッシュメモリの前記ダーティデータの少なくとも一部を前記記憶装置へ格納する第2データ格納部と、

前記電力供給部から供給される電力の電圧異常を検出する電源監視部と、

前記電源監視部により前記電圧異常が検出された場合に、前記バッテリーからの前記電力の供給を受けて、前記キャッシュメモリに格納されているダーティデータを前記不揮発メモリに退避させる退避格納部と、

前記電源監視部により前記電圧異常が検出された場合に、前記バッテリーからの電力を用

10

20

いて前記キャッシュメモリ及び前記退避格納部への電力の供給を維持する電力供給制御部とを有し、

前記不揮発メモリを保持するメモリボードが着脱可能であり、

更に、前記不揮発メモリに退避されたパーティデータを、前記キャッシュメモリに復旧させるデータ復旧部と、

装着されていた前記メモリボードの第1識別情報を保持する装着ボード情報保持部と、不揮発メモリを保持するメモリボードが新たに装着されたことを検出する装着検出部と、

前記新たに装着されたメモリボードの第2識別情報を取得するボード情報取得部と、

取得した前記第2識別情報と、前記装着ボード情報保持部が保持する前記第1識別情報とが一致するか否かを判定する装着判定部と、

一致しないと判定した場合に、前記データ復旧部によるパーティデータの復旧をさせずに、前記新たに装着されたメモリボードの不揮発メモリのデータの初期化を実行する初期化部とを有するストレージ制御装置。

【請求項2】

前記しきい値は、前記不揮発メモリの容量に基づいて決定されている請求項1に記載のストレージ制御装置。

【請求項3】

前記しきい値を前記不揮発メモリの容量に基づいて決定するしきい値決定部を有する請求項2に記載のストレージ制御装置。

【請求項4】

前記退避格納部は、前記パーティデータを暗号化して前記不揮発メモリに退避させる請求項1乃至請求項3のいずれか1項に記載のストレージ制御装置。

【請求項5】

前記退避格納部は、前記パーティデータのデータ量を変化させない暗号化を実行する請求項4に記載のストレージ制御装置。

【請求項6】

前記退避格納部は、前記パーティデータを圧縮して前記不揮発メモリに退避させる請求項1乃至請求項5のいずれか1項に記載のストレージ制御装置。

【請求項7】

前記パーティデータの前記キャッシュメモリにおける格納先情報を前記不揮発メモリに退避させる格納先情報格納部を更に有する請求項1乃至請求項6のいずれか1項に記載のストレージ制御装置。

【請求項8】

前記不揮発メモリを保持するメモリボードを着脱可能な複数のスロットを備え、

前記装着ボード情報保持部は、前記第1識別情報と、前記第1識別情報のメモリボードが装着されていたスロットの第1スロット識別情報とを対応付けて記憶し、

前記ボード情報取得部は、前記第2識別情報と、当該第2識別情報のメモリボードが装着されたスロットの第2スロット識別情報とを取得し、

前記装着判定部は、前記第1識別情報及び前記第1スロット識別情報と、前記第2識別情報及び前記第2スロット識別情報とが一致するか否かを判定し、

前記初期化部は、前記第1識別情報及び前記第1スロット識別情報と、前記第2識別情報及び前記第2スロット識別情報とが一致しない場合に、前記データ復旧部によるパーティデータの復旧をさせずに、前記新たに装着されたメモリボードの不揮発メモリのデータの初期化を実行する

請求項7に記載のストレージ制御装置。

【請求項9】

前記不揮発メモリは、複数の不揮発メモリデバイスを有し、

前記退避格納部は、前記複数の不揮発メモリデバイスによって構成されるRAIDグループに対して、前記パーティデータを格納する請求項1乃至請求項8のいずれか1項に記

10

20

30

40

50

載のストレージ制御装置。

【請求項 10】

前記退避格納部は、前記パーティデータを所定の大きさの複数のデータ単位に分割して、前記 RAID グループの複数の前記不揮発メモリデバイスに分散させて格納するとともに、所定数のデータ単位のデータに基づいて作成されるパリティを前記 RAID グループの前記不揮発メモリデバイスに格納する請求項 9 に記載のストレージ制御装置。

【請求項 11】

前記電力供給制御部は、前記パーティデータの前記不揮発メモリへの退避が終了した後に、前記バッテリーから前記揮発メモリへの電力供給を切断する

請求項 1 乃至請求項 10 のいずれか 1 項に記載のストレージ制御装置。

10

【請求項 12】

前記キャッシュメモリは、複数の揮発メモリデバイスにより構成されており、

前記電力供給制御部は、前記退避格納部により前記パーティデータの退避を終えた各前記揮発メモリデバイスから電力供給を逐次切断していく

請求項 1 乃至請求項 11 のいずれか 1 項に記載のストレージ制御装置。

【請求項 13】

前記電力供給制御部は、前記電源監視部により前記電圧異常が検出された場合に、前記要求受信部に電力を供給することなく、前記キャッシュメモリ及び前記退避格納部に電力を供給する

請求項 1 乃至請求項 12 のいずれか 1 項に記載のストレージ制御装置。

20

【請求項 14】

前記キャッシュメモリ、前記不揮発メモリ、前記退避格納部、及び前記電力供給制御部は、同一のボードに備えられ、

前記バッテリーは、前記ボードに対して電力を供給可能となっている

請求項 13 に記載のストレージ制御装置。

【請求項 15】

前記退避格納部と前記第 2 データ格納部とは、異なるデバイスによって構成されており、

前記電力供給制御部は、前記電源監視部により前記電圧異常が検出された場合に、前記第 2 データ格納部に電力を供給することなく、前記退避格納部に電力を供給する

請求項 1 乃至請求項 14 のいずれか 1 項に記載のストレージ制御装置。

30

【請求項 16】

前記キャッシュメモリ及び前記不揮発メモリの組を複数有し、

前記第 1 データ格納部は、前記ライトアクセス要求対象のデータを各組の前記キャッシュメモリのそれぞれに格納し、

前記退避格納部は、前記電源監視部により前記電圧異常が検出された場合に、前記バッテリーからの前記電力の供給を受けて、前記複数のキャッシュメモリのいずれか 1 つのキャッシュメモリからパーティデータを読み出して、前記複数の不揮発メモリに分散させて退避させる

請求項 1 乃至請求項 15 のいずれか 1 項に記載のストレージ制御装置。

40

【請求項 17】

外部装置からライトアクセス要求を受信し、ライトアクセス要求対象のデータを記憶装置へ書き込む制御を行うストレージ制御装置によるデータ管理方法であって、

前記ストレージ制御装置は、電源からの電力の供給を行う電力供給部と、電力を供給可能に蓄えるバッテリーと、電力の供給を受けてデータを記憶可能な揮発メモリであるキャッシュメモリと、電力の供給を受けずともデータを記憶し続けることのできる不揮発メモリと、前記電力供給部から供給される電力の電圧異常を検出する電源監視部とを備えており、

ライトアクセス要求を受信し、前記ライトアクセス要求の対象のデータをキャッシュデータとして前記キャッシュメモリに格納し、前記キャッシュメモリの前記キャッシュデー

50

タのうちの前記記憶装置への反映がされていないダーティデータのデータ量が所定のしきい値を超えるか否かを判定し、前記しきい値を超えると判定された場合に、前記キャッシュメモリの前記ダーティデータの少なくとも一部を前記記憶装置へ格納し、前記電源監視部により前記電圧異常が検出された場合に、前記バッテリーからの電力を用いて前記キャッシュメモリへの電力の供給を維持させ、前記バッテリーからの前記電力の供給を受けて、前記キャッシュメモリに格納されているダーティデータを前記不揮発メモリに退避させ、

前記不揮発メモリを保持するメモリボードが着脱可能であり、前記不揮発メモリに退避されたダーティデータを、前記キャッシュメモリに復旧させるデータ復旧部を備え、装着されていた前記メモリボードの第1識別情報を保持し、不揮発メモリを保持するメモリボードが新たに装着されたことを検出し、前記新たに装着されたメモリボードの第2識別情報を取得し、取得した前記第2識別情報と、保持している前記第1識別情報とが一致するか否かを判定し、一致しないと判定した場合に、前記データ復旧部によるダーティデータの復旧をせずに、前記新たに装着されたメモリボードの不揮発メモリのデータの初期化を実行するデータ管理方法。

10

【請求項18】

外部装置からライトアクセス要求を受信し、ライトアクセス要求対象のデータを記憶装置へ書き込む制御を行うストレージ制御装置であって、

電源からの電力の供給を行う電源回路と、

電力を供給可能に蓄えるバッテリーと、

電力の供給を受けてデータを記憶可能な揮発メモリであるキャッシュメモリと、

20

電力の供給を受けずともデータを記憶し続けることのできる不揮発メモリと、

前記外部装置からのライトアクセス要求を受信するインタフェースと、

前記インタフェースに接続されるとともに、前記キャッシュメモリとのデータの入出力可能な第1プロセッサと、

前記キャッシュメモリと前記不揮発メモリとの間でデータの入出力可能な第2プロセッサと、

前記電源回路からの電圧異常を検出する電源監視制御部と、
を備え、

前記第1プロセッサは、前記インタフェースが受信したライトアクセス要求を受け付け、前記ライトアクセス要求の対象のデータをキャッシュデータとして前記キャッシュメモリに格納し、前記キャッシュメモリの前記キャッシュデータのうちの前記記憶装置への反映がされていないダーティデータのデータ量が所定のしきい値を超えるか否かを判定し、前記しきい値を超えると判定した場合に、前記キャッシュメモリの前記ダーティデータの少なくとも一部を前記記憶装置へ格納し、

30

前記第2プロセッサは、前記電源監視制御部により前記電圧異常が検出された場合に、前記バッテリーからの前記電力の供給を受けて、前記キャッシュメモリに格納されているダーティデータを前記不揮発メモリに退避させ、

前記電源監視制御部は、前記電圧異常を検出した場合に、前記バッテリーからの電力を用いて前記キャッシュメモリ及び前記第2プロセッサへの電力の供給を維持し、

前記不揮発メモリを保持するメモリボードが着脱可能であり、

40

更に、前記不揮発メモリに退避されたダーティデータを、前記キャッシュメモリに復旧させるデータ復旧部を有し、

前記第2プロセッサは、装着されていた前記メモリボードの第1識別情報を保持し、不揮発メモリを保持するメモリボードが新たに装着されたことを検出し、前記新たに装着されたメモリボードの第2識別情報を取得し、取得した前記第2識別情報と、前記装着ボード情報保持部が保持する前記第1識別情報とが一致するか否かを判定し、一致しないと判定した場合に、前記データ復旧部によるダーティデータの復旧をさせずに、前記新たに装着されたメモリボードの不揮発メモリのデータの初期化を実行するストレージ制御装置。

50

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、例えば、電源障害が発生した際に、揮発メモリに格納されたデータを不揮発メモリに退避するストレージ制御装置及びデータ管理方法に関する。

【背景技術】

【0002】

ストレージ制御装置には、例えば、ハードディスクドライブのような記憶デバイスが複数接続されている。ストレージ制御装置は、ホスト計算機からライトコマンドを受信し、複数の記憶装置のうちの少なくとも一つの記憶装置に対してデータを書き込み、また、ホスト計算機からのリードコマンドを受信し、複数の記憶装置のうちの少なくとも一つの記憶装置からデータを読み出してホスト計算機に送信する。

10

【0003】

このようなストレージ制御装置においては、ライトコマンドに従って記憶装置に書き込まれるデータを一時的に記憶するため、あるいは、リードコマンドに従って記憶装置から読み出されたデータを一時的に記憶するためにキャッシュメモリが備えられている。

【0004】

このキャッシュメモリとしては、一般には、電力が供給されることによりデータを記憶しておくことのできる揮発メモリが使用される。

【0005】

20

揮発メモリをキャッシュメモリとして備えているストレージ制御装置においては、例えば、外部電源の障害が発生した等によってキャッシュメモリへの電力の供給が行われなくなると、キャッシュメモリに記憶されたデータが失われてしまう。

【0006】

そこで、このような外部電源の障害等に対応するために、ストレージ制御装置において電力を供給することのできるバッテリーを備え、外部電源の障害が発生した場合には、バッテリーからキャッシュメモリに対して電力を供給することにより、キャッシュメモリに記憶されたデータを保持することが行われている。

【0007】

しかしながら、障害が解消されるまでキャッシュメモリへの電力の供給を維持する必要があるため、バッテリーの容量を比較的大きくしておく必要がある。このため、ストレージ制御装置の製造原価が増大してしまう問題が発生する。

30

【0008】

これに対して、キャッシュメモリのデータを不揮発メモリに退避させることにより、バッテリーによってキャッシュメモリへの電力の供給を障害の解消まで維持しなくてもデータを保全することのできる技術が開示されている（例えば、特許文献1）。

【0009】

【特許文献1】特開2004-21811号公報

【発明の開示】

【発明が解決しようとする課題】

40

【0010】

例えば、キャッシュメモリに記憶されたデータを不揮発メモリに退避させることを想定した場合において、キャッシュメモリのすべてのデータを適切に退避させる場合には、キャッシュメモリの容量と同等の容量を有する不揮発メモリを用意しておく必要がある。この場合には、ストレージ制御装置の製造原価が増大してしまう。

【0011】

一方、製造原価を抑えるために、キャッシュメモリよりも容量の少ない不揮発メモリを用意してデータを退避させるようにした場合にあっては、キャッシュメモリのデータを不揮発メモリに適切に退避させることができず、必要なデータが消失してしまうという事態が発生する虞がある。

50

【 0 0 1 2 】

そこで、本発明は、上記課題に鑑みなされたものであり、その目的は、不揮発メモリの容量を抑えつつ、キャッシュメモリに格納されたデータを適切に退避させることのできる技術を提供することにある。

【課題を解決するための手段】

【 0 0 1 3 】

上記課題を解決するため、本発明の一観点に従うストレージ制御装置は、キャッシュメモリ中に格納されているデータには、記憶装置に反映されているデータ（クリーンデータ）と、記憶装置に反映されていないデータ（ダーティデータ）とがあることに着目してなされたものである。すなわち、本発明の一観点に従うストレージ制御装置は、キャッシュメモリに格納されているダーティデータのデータ量に基づいて、ダーティデータを記憶装置に格納するか否かを決定し、電圧異常の場合に、キャッシュメモリのダーティデータを不揮発メモリに退避する。

10

【 0 0 1 4 】

具体的には、本発明の一観点に従うストレージ制御装置は、外部装置からライトアクセス要求を受信し、ライトアクセス要求対象のデータを記憶装置へ書き込む制御を行うストレージ制御装置であって、電源からの電力の供給を行う電力供給部と、電力を供給可能に蓄えるバッテリーと、電力の供給を受けてデータを記憶可能な揮発メモリであるキャッシュメモリと、電力の供給を受けずともデータを記憶し続けることのできる不揮発メモリと、前記電力の供給を受けて、ライトアクセス要求を受信する要求受信部と、前記ライトアクセス要求の対象のデータをキャッシュデータとして前記キャッシュメモリに格納する第1データ格納部と、前記キャッシュメモリの前記キャッシュデータのうちの前記記憶装置への反映がされていないダーティデータのデータ量が所定のしきい値を超えるか否かを判定する判定部と、前記しきい値を超えると判定された場合に、前記キャッシュメモリの前記ダーティデータの少なくとも一部を前記記憶装置へ格納する第2データ格納部と、前記電力供給部から供給される電力の電圧異常を検出する電源監視部と、前記電源監視部により前記電圧異常が検出された場合に、前記バッテリーからの前記電力の供給を受けて、前記キャッシュメモリに格納されているダーティデータを前記不揮発メモリに退避させる退避格納部と、前記電源監視部により前記電圧異常が検出された場合に、前記バッテリーからの電力を用いて前記キャッシュメモリ及び前記退避格納部への電力の供給を維持する電力供給制御部とを有する。

20

30

【発明を実施するための最良の形態】

【 0 0 1 5 】

本発明の実施形態について、図面を参照して説明する。なお、以下に説明する実施形態は特許請求の範囲にかかる発明を限定するものではなく、また実施形態の中で説明されている特徴の組み合わせの全てが発明の解決手段に必須であるとは限らない。

< 第1実施形態 >

図1は、本発明の第1実施形態に係る計算機システムの構成図である。

【 0 0 1 6 】

計算機システムは、1台以上のホスト装置10と、一台以上のディスクサブシステム100とを備えている。ホスト装置10とディスクサブシステム100とは、ネットワーク20を介して接続されている。ネットワークとしては、S A N (Storage Area Network)、L A N (Local Area Network)、インターネット、専用回線、公衆回線等であってもよく、データ通信を行うことのできるネットワークであればよい。また、ネットワーク20におけるプロトコルとしては、ファイバチャネルプロトコルや、T C P / I P プロトコルであってもよく、ホスト装置10とディスクサブシステム100との間でデータのやりとりが可能なプロトコルであれば、任意のプロトコルであってもよい。なお、ネットワーク20に代えて、ケーブルによってホスト装置10とディスクサブシステム100とを直接接続するようにしてもよい。

40

【 0 0 1 7 】

50

ホスト装置 10 は、図示しない CPU (Central Processing Unit)、図示しないメモリ、キーボード等の入力装置、ディスプレイ等を備えている。ホスト装置 10 は、例えば、汎用のコンピュータ (パーソナルコンピュータ) によって構成することができる。ホスト装置 10 には、アプリケーションプログラム 11 が備えられている。また、ホスト装置 10 には、ネットワーク 20 との接続を可能にするポート (PORT) 12 が備えられている。

【0018】

ホスト装置 10 の CPU がアプリケーションプログラム 11 を実行することにより、ディスクサブシステム 100 に対して、データの書き込みアクセス (ライトアクセス) や、データの読み出しアクセス (リードアクセス) することができる。

10

【0019】

ディスクサブシステム 100 は、ストレージ制御装置の一例としてのディスク制御装置 200 と、記憶装置 300 と、複数の電源回路 400 と、複数のバッテリー 500 とを有している。

【0020】

記憶装置 300 は、複数の記憶デバイス 310 を含む。記憶デバイス 310 は、例えば、ハードディスクドライブ (HDD) である。ディスクサブシステム 100 においては、複数の記憶デバイス 310 の記憶空間を基に 1 又は複数の論理ボリュームを提供することができる。また、ディスクサブシステム 100 においては、複数の記憶デバイス 310 の内の 2 以上の記憶デバイス 310 で RAID (Redundant Array of Independent Disks) グループを構成して、RAID グループの記憶空間を論理ボリュームとして提供することもできる。

20

【0021】

電源回路 400 は、例えば、外部の商用電源から供給される電力をディスク制御装置 200 の各部に供給する。本実施形態では、図面右側の電源回路 400 は、ディスク制御装置 200 の破線で区切られた右側の各部に電力を供給し、図面左側の電源回路 400 は、ディスク制御装置 200 の破線で区切られた左側の各部に電力を供給するようになっている。

【0022】

バッテリー 500 は、電力を蓄積しており、ディスク制御装置 200 の所定の部位に供給できるようになっている。本実施形態では、図面右側のバッテリー 500 は、図面右側のメモリボード 250 上の各部に電力を供給できるようになっており、図面左側のバッテリー 500 は、図面左側のメモリボード 250 上の各部に電力を供給できるようになっている。

30

【0023】

ディスク制御装置 200 は、複数のチャネルアダプタ 210 と、複数の I/O プロセッサ 220 と、コントロールユニット 230 と、接続部 240 と、複数のメモリボード 250 と、複数のディスクアダプタ 270 とを有している。チャネルアダプタ 210 と、I/O プロセッサ 220 と、コントロールユニット 230 と、メモリボード 250 と、ディスクアダプタ 270 は、接続部 240 を介してそれぞれ接続されている。

【0024】

接続部 240 は、チャネルアダプタ 210 と、I/O プロセッサ 220 と、コントロールユニット 230 と、メモリボード 250 と、ディスクアダプタ 270 との間の相互通信を可能にする。接続部 240 は、例えば、スイッチング動作によってデータ伝送を行うクロスバスイッチであってもよい。

40

【0025】

チャネルアダプタ 210 は、ネットワーク 20 と接続するためのポート 211 を有している。チャネルアダプタ 210 は、ポート 211 を介して接続されているホスト装置 10 との間での通信を実行する。本実施形態では、ホスト装置 10 との間でのデータの読み出し (データリード) 及びデータの書き込み (データライト) における各種情報の送受信を実行する。

50

【 0 0 2 6 】

コントロールユニット 2 3 0 は、接続部 2 4 0 を介して、チャンネルアダプタ 2 1 0 と、I/O プロセッサ 2 2 0 と、コントロールユニット 2 3 0 と、メモリボード 2 5 0 と、ディスクアダプタ 2 7 0 とに対してアクセス可能であり、これら各部を管理者が保守管理するために用いられる。コントロールユニット 2 3 0 は、例えば、管理者が入力するキーボード、マウス等の入力装置、CPU、ROM、RAM、ハードディスクドライブ、情報を表示出力するためのディスプレイ等を備えていてもよい。本実施形態では、コントロールユニット 2 3 0 は、ディスク制御装置 2 0 0 のメモリボード用スロットに装着されたメモリボード 2 5 0 の固有識別子をメモリボード 2 5 0 から取得するとともに、装着されているスロットの識別番号（スロット番号）とを取得し、それらに対応付けて記憶する。

10

【 0 0 2 7 】

ディスクアダプタ 2 7 0 は、記憶装置 3 0 0 の各記憶デバイス 3 1 0 と接続するためのポート 2 7 1 を有している。ディスクアダプタ 2 7 0 は、記憶デバイス 3 1 0 との間でデータの授受を行う。

【 0 0 2 8 】

I/O プロセッサ 2 2 0 は、メモリボード 2 5 0 上の共有メモリ 2 5 4（図 2 参照）に読み出されたプログラムを実行することにより各種制御処理を実行する。I/O プロセッサ 2 2 0 は、チャンネルアダプタ 2 1 0 と、メモリボード 2 5 0 と、ディスクアダプタ 2 7 0 との間でのデータの受け渡しを制御する。例えば、チャンネルアダプタ 2 1 0 が受信したデータをメモリボード 2 5 0 上のキャッシュメモリ 2 5 3（図 2 参照）に格納させる制御を行う。また、I/O プロセッサ 2 2 0 は、キャッシュメモリ 2 5 3 に格納されたデータを、ディスクアダプタ 2 7 0 に渡したり、チャンネルアダプタ 2 1 0 に渡したりする制御を行う。また、I/O プロセッサ 2 2 0 は、ディスクアダプタ 2 7 0 が記憶デバイス 3 1 0 から取得したデータをキャッシュメモリ 2 5 3 に格納させる制御を行う。また、I/O プロセッサ 2 2 0 は、不揮発メモリ 2 5 5 に退避格納したデータをキャッシュメモリ 2 5 3 に復旧するための処理を行う。

20

【 0 0 2 9 】

メモリボード 2 5 0 は、ディスク制御装置 2 0 0 のメモリボード用スロットに対して着脱可能になっている。

【 0 0 3 0 】

30

図 2 は、本発明の第 1 実施形態に係るメモリボードの構成図である。

【 0 0 3 1 】

メモリボード 2 5 0 は、メモリコントローラ 2 5 1 と、プロセッサ（プロセサ）2 5 2 と、キャッシュメモリ 2 5 3 と、共有メモリ 2 5 4 と、不揮発メモリ 2 5 5 と、不揮発メモリ 2 5 6 と、電圧監視制御部 2 5 7 とを有する。

【 0 0 3 2 】

メモリコントローラ 2 5 1 は、接続部 2 4 0 に接続されるとともに、プロセッサ 2 5 2 と、キャッシュメモリ 2 5 3 と、共有メモリ 2 5 4 と、不揮発メモリ 2 5 5 と、不揮発メモリ 2 5 6 とに接続されている。

【 0 0 3 3 】

40

メモリコントローラ 2 5 1 は、I/O プロセッサ 2 2 0、コントロールユニット 2 3 0 又は、プロセッサ 2 5 2 の制御によって、接続部 2 4 0 から送信されるデータをメモリボード 2 5 0 内のメモリ（キャッシュメモリ 2 5 3、共有メモリ 2 5 4、不揮発メモリ 2 5 5、又は不揮発メモリ 2 5 6）に格納する処理や、メモリボード 2 5 0 内のメモリに格納されたデータを接続部 2 4 0 へ送信する処理や、メモリボード 2 5 0 内のメモリ間でのデータの受け渡しの処理を実行する。また、メモリコントローラ 2 5 1 は、不揮発メモリ 2 5 5 に退避格納するデータについて暗号化する処理を行う。本実施形態では、メモリコントローラ 2 5 1 は、データのデータ量が変わらない暗号化、例えば、シーザー暗号を用いて暗号化している。

【 0 0 3 4 】

50

電圧監視制御部 257 は、電源回路 400 からメモリボード 250 に供給される電力の電圧を監視し、電圧に異常がないか、例えば、所定の電圧以下でないかを判定し、異常があると判定した場合には、その旨をプロセッサ 252 に通知すると共に、バッテリー 500 からの電力がメモリボード 250 の所定の部位（例えば、プロセッサ 252、メモリコントローラ 251、キャッシュメモリ 253、共有メモリ 254、及び不揮発メモリ 255、256）へ供給されるように制御する。また、電圧監視制御部 257 は、後述するデータ退避処理においてバッテリー 500 からの電力の供給を切断する制御を行う。

【0035】

キャッシュメモリ 253 は、揮発メモリ、例えば、DRAM (Dynamic Random Access Memory) である。キャッシュメモリ 253 は、チャンネルアダプタ 210 により受信されたデータや、ディスクアダプタ 270 により記憶デバイス 310 から取得されたデータを一時的に記憶する。キャッシュメモリ 253 は、例えば、個別に入出力動作が可能な複数のキャッシュメモリデバイスによって構成されている。

10

【0036】

プロセッサ 252 は、共有メモリ 254 に読み出されたプログラムを実行することにより各種制御処理を実行する。例えば、プロセッサ 252 は、キャッシュメモリ 253 に格納されたキャッシュデータを不揮発メモリ 255 に退避格納する処理を実行する。

【0037】

共有メモリ 254 は、揮発メモリであり、各種情報を記憶する。記憶する情報としては、例えば、ホスト装置とのやりとりが行われるデータに関する構成情報 261（例えば、データが記憶デバイス 310 のどこに格納されるかを示す情報）及び制御情報 260（例えば、データがキャッシュメモリ 253 のどこに格納されるかを示す情報）とがある。

20

【0038】

不揮発メモリ 255 及び 256 は、例えば、フラッシュメモリ、MRAM (Magnetoresistive Random Access Memory)、PRAM (Phase change RAM) 等の電源を供給せずともデータを記憶しておくことのできるメモリである。

【0039】

不揮発メモリ 255 は、例えば、キャッシュメモリ 253 や共有メモリ 254 に格納されたデータを退避格納するために用いられる。本実施形態では、不揮発メモリ 255 には、ダーティデータを格納することとなるので、当該不揮発メモリ 255 には、キャッシュメモリ 255 に格納されたダーティデータを格納することのできるだけの容量を有しておく必要がある。このことは、データ退避を確実に行うためには、不揮発メモリ 255 の容量によってキャッシュメモリ 255 に格納させておくことのできるダーティデータの量が決まることを意味している。また、本実施形態では、メモリボード 255 毎に、キャッシュメモリ 253 のデータを当該メモリボード 250 内の不揮発メモリ 255 に退避格納させることができるので、各メモリボード 250 毎に確実にデータを退避格納しておくことができる。不揮発メモリ 256 は、不揮発メモリ 255 に退避格納されたデータを元の状態に復旧するためのアドレス管理テーブル 262 と、メモリボード 250 を一意に識別する固有識別子 263（例えば、メモリボード 250 の製造番号）を記憶する。

30

【0040】

次に、共有メモリ 254 に記憶された構成情報及び制御情報の一例について図面を参照して説明する。

40

【0041】

図 3 は、本発明の第 1 実施形態に係る制御情報及び構成情報の一例を示す図である。図 3 A は、本発明の第 1 実施形態に係る制御情報の一例を示し、図 3 B は、本発明の第 1 実施形態に係る構成情報の一例を示す。

【0042】

制御情報 260 は、図 3 A に示すように、論理アドレス 2601 と、キャッシュアドレス 2602 と、ステージング bit 2603 と、ダーティ bit 2604 とを対応付けたレコードを含む。

50

【 0 0 4 3 】

論理アドレス 2 6 0 1 には、データを特定するための論理的なアドレス（論理アドレス）が格納される。論理アドレスとしては、例えば、ホスト装置 1 0 から送信されるアクセス要求中の L U N（Logical Unit Number）と L B A（Logical Block Address）との組み合わせがある。本実施形態では、論理アドレスを単位として管理しているため、各レコードが管理するデータのデータ量は、論理アドレスに対応する所定のデータ量となっている。

【 0 0 4 4 】

キャッシュアドレス 2 6 0 2 には、対応するデータが格納されているキャッシュメモリ 2 5 3 のアドレスが格納される。

10

【 0 0 4 5 】

ステージング b i t 2 6 0 3 には、対応するデータが記憶デバイス 3 1 0 に格納されたデータと一致しているか否かを示すビットが格納される。例えば、ステージング b i t 2 6 0 3 には、対応するデータが記憶デバイス 3 1 0 に格納されたデータと一致している場合には、“ 1 ” が格納され、一致しない場合には、“ 0 ” が格納される。

【 0 0 4 6 】

ダーティ b i t 2 6 0 4 には、対応するデータが記憶デバイス 3 1 0 に反映されているデータ（クリーンデータ）か、又は反映されていないデータ（ダーティデータ）かを示すビットが格納される。例えば、ダーティ b i t 2 6 0 4 には、対応するデータがクリーンデータである場合には、“ 0 ” が格納され、ダーティデータである場合には、“ 1 ” が格納される。ダーティ b i t 2 6 0 4 に“ 0 ” が格納されているデータ、すなわちクリーンデータは、記憶デバイス 3 1 0 に存在する。このため、キャッシュメモリ 2 5 3 への電力供給が停止して、キャッシュメモリ 2 5 3 から消去されてしまっても記憶デバイス 3 1 0 から取り出すことができる。したがって、電力供給が停止しても、ディスクサブシステム 1 0 0 からデータが消失してしまうことはない。一方、ダーティ b i t 2 6 0 4 に“ 1 ” が格納されているデータ、すなわちダーティデータは、キャッシュメモリ 2 5 3 上に存在するが、記憶デバイス 3 1 0 に反映されていない。このため、キャッシュメモリ 2 5 3 への電力供給が停止してしまうと、データが消失してしまい、ディスクサブシステム 1 0 0 から完全に消失してしまうこととなる。そこで、ディスク制御装置 2 0 0 においては、電圧異常が発生した場合に、ダーティデータをキャッシュメモリ 2 5 3 から不揮発メモリ 2 5 5 に退避することとしている。このように、ダーティデータを不揮発メモリ 2 5 5 に退避させるので、データがディスクサブシステム 1 0 0 から完全に消失してしまうことを防止することができる。本実施形態では、退避する対象をダーティデータとし、クリーンデータを退避する対象としていないので、データの退避に必要な不揮発メモリの容量を低減することができるとともに、データ退避処理を迅速に行うことができる。

20

30

【 0 0 4 7 】

構成情報 2 6 1 は、図 3 B に示すように、論理アドレス 2 6 1 1 と、物理アドレス 2 6 1 2 とを対応付けたレコードを含む。

【 0 0 4 8 】

論理アドレス 2 6 1 1 には、データを特定するための論理アドレスが格納される。論理アドレスとしては、例えば、ホスト装置 1 0 から送信されるアクセスコマンド中の L U N（Logical Unit Number）と L B A（Logical Block Address）との組み合わせがある。物理アドレス 2 6 1 2 には、対応する論理アドレスのデータが格納されている記憶デバイス 3 1 0 及び当該記憶デバイス 3 1 0 における格納領域を示す物理的なアドレス（物理アドレス）が格納される。

40

【 0 0 4 9 】

次に、不揮発メモリ 2 5 6 に格納されたアドレス管理テーブルの一例について図面を参照して説明する。

【 0 0 5 0 】

図 4 は、本発明の第 1 実施形態に係るアドレス管理テーブルの一例を示す図である。図

50

4 Aは、本発明の第1実施形態に係る制御情報についてのアドレス管理テーブルの一例を示し、図4Bは、本発明の第1実施形態に係る構成情報についてのアドレス管理テーブルの一例を示し、図4Cは、本発明の第1実施形態に係るキャッシュデータについてのアドレス管理テーブルの一例を示す。

【0051】

アドレス管理テーブル262は、不揮発メモリ255に格納された制御情報のアドレスを管理するための制御情報のアドレス管理テーブル262Aと、不揮発メモリ255に格納された構成情報のアドレスを管理するためのアドレス管理テーブル262Bと、不揮発メモリ255に格納されたキャッシュデータのアドレスを管理するためのアドレス管理テーブル262Cとを含む。

10

【0052】

制御情報のアドレス管理テーブル262Aは、不揮発メモリアドレス2621と、共有メモリアドレス2622と、データ長2623とを対応付けたレコードを含む。

【0053】

不揮発メモリアドレス2621には、制御情報の格納に割り当て可能な不揮発メモリ255上のアドレス（不揮発メモリアドレス）が格納される。共有メモリアドレス2622には、対応する不揮発メモリアドレスからの記憶領域が割り当てられた制御情報が共有メモリ254上で格納されていたアドレス（共有メモリアドレス）が格納される。データ長2623には、対応する制御情報についての不揮発メモリ255上におけるデータ長が格納される。

20

【0054】

構成情報のアドレス管理テーブル262Bは、不揮発メモリアドレス2624と、共有メモリアドレス2625と、データ長2626とを対応付けたレコードを含む。

【0055】

不揮発メモリアドレス2624には、構成情報の格納に割り当て可能な不揮発メモリ255上のアドレス（不揮発メモリアドレス）が格納される。共有メモリアドレス2625には、対応する不揮発メモリアドレスからの記憶領域が割り当てられた構成情報が共有メモリ254上で格納されていたアドレス（共有メモリアドレス）が格納される。データ長2626には、対応する構成情報についての不揮発メモリ255上におけるデータ長が格納される。

30

【0056】

キャッシュデータのアドレス管理テーブル262Cは、不揮発メモリアドレス2627と、キャッシュメモリアドレス2628と、データ長2629とを対応付けたレコードを含む。

【0057】

不揮発メモリアドレス2627には、キャッシュデータの格納に割り当て可能な不揮発メモリ255上のアドレス（不揮発メモリアドレス）が格納される。キャッシュメモリアドレス2628には、対応する不揮発メモリアドレスからの記憶領域が割り当てられたキャッシュデータがキャッシュメモリ253上で格納されていたアドレス（キャッシュメモリアドレス）が格納される。データ長2629には、対応するキャッシュデータについての不揮発メモリ255上におけるデータ長が格納される。

40

【0058】

次に、本発明の第1実施形態に係るディスク制御装置の処理動作について説明する。

【0059】

図5Aは、本発明の第1実施形態に係るライトアクセス要求時処理のフローチャートである。

【0060】

ライトアクセス要求時処理は、ホスト装置10から送信されたライトアクセス要求をディスクサブシステム100のチャネルアダプタ210がポート211を介して受信し、当該ライトアクセス要求をI/Oプロセッサ220が取得した場合に、実行が開始される。

50

【 0 0 6 1 】

まず、I/Oプロセッサ220が、チャンネルアダプタ210からライトアクセス要求を取得すると(ステップS11)、I/Oプロセッサ220が、ライトアクセス要求の対象のデータ(ライトデータ)をチャンネルアダプタ210から取得し、当該ライトデータをキャッシュメモリ253に書き込む(ステップS12)。次いで、I/Oプロセッサ220は、共有メモリ254の構成情報261の当該ライトデータに対応するレコードを更新する(ステップS13)。すなわち、I/Oプロセッサ220は、構成情報261中のライトデータに対応するレコードのキャッシュアドレス2602に、ライトデータを格納したキャッシュメモリ253のキャッシュアドレスを格納し、更に、ダーティビット2604にダーティデータであることを示す“1”を格納する。

10

【 0 0 6 2 】

次いで、I/Oプロセッサ220は、キャッシュメモリ253に格納されているダーティデータのデータ量を検出し、予め設定されているしきい値(ライトキャッシュしきい値)を超えるか否かを判定する(ステップS14)。ここで、キャッシュメモリ253に格納されているダーティデータのデータ量は、共有メモリ254の構成情報261を参照し、構成情報261のダーティビット2604に“1”が格納されているアドレス数に基づいて検出することができる。また、ライトキャッシュしきい値は、キャッシュメモリ253のダーティデータのデータ量が当該しきい値以下であれば、当該ダーティデータを不揮発メモリ255へ確実に退避格納できることを示すしきい値である。このライトキャッシュしきい値は、例えば、管理者により入力された指示に応じてコントロールユニット230が設定するようにしてもよく、また、コントロールユニット230が不揮発メモリ255のデータ容量に基づいて、一のしきい値に設定するようにしてもよく、また、コントロールユニット230がストレージサブシステム100の動作状態及び不揮発メモリ255のデータ容量に基づいて、動的にしきい値を設定するようにしてもよい。ライトキャッシュしきい値としては、不揮発メモリ255の容量の、例えば、50～80パーセントの間のいずれかの容量に設定してもよい。

20

【 0 0 6 3 】

ステップS14での判定の結果、ダーティデータのデータ量がライトキャッシュしきい値を超えると判定した場合(ステップS14のYes)には、I/Oプロセッサ220は、少なくとも一部のダーティデータをデステージさせる(ステップS15)。すなわち、I/Oプロセッサ220は、キャッシュメモリ253のダーティデータの少なくとも一部を記憶デバイス310に格納させる。これによって、当該データは、キャッシュメモリ253の内容が記憶デバイス310に反映されることとなる。なお、デステージさせるダーティデータとしては、例えば、アクセス頻度が少ないダーティデータを対象としてもよく、また、前回のアクセスから最も時間が経過しているダーティデータを対象としてもよい。

30

【 0 0 6 4 】

次いで、I/Oプロセッサ220は、共有メモリ254の構成情報261のデステージしたデータに対応するレコードを更新する(ステップS16)。すなわち、I/Oプロセッサ220は、構成情報261中のデステージしたデータに対応するレコードのダーティビット2604にクリーンデータであることを示す“0”を格納し、ライトアクセス要求時処理を終了する。これによって、後述するデータ退避処理において、キャッシュメモリ253内のダーティデータを不揮発メモリ255に確実に退避格納できることとなる。

40

【 0 0 6 5 】

一方、ステップS14での判定の結果、ダーティデータのデータ量がライトキャッシュしきい値を超えないと判定した場合(ステップS14のNo)には、キャッシュメモリ253内のダーティデータが不揮発メモリ255に確実に退避格納できることを示しているので、I/Oプロセッサ220は、ライトアクセス要求時処理を終了する。なお、キャッシュメモリ253のダーティデータを記憶デバイス310に格納するデステージ処理(ステップS15及びステップS16と同様な処理)は、ライトアクセス要求時処理において

50

だけでなく、例えば、I/Oプロセッサ220の処理負荷が軽い場合等において、適宜I/Oプロセッサ220が実行するようにしてもよい。

【0066】

図5Bは、本発明の第1実施形態に係るリードアクセス要求時処理のフローチャートである。

【0067】

リードアクセス要求時処理は、ホスト装置10から送信されたリードアクセス要求をディスクサブシステム100のチャネルアダプタ210がポート211を介して受信し、当該リードアクセス要求をI/Oプロセッサ220が取得した場合に、実行が開始される。

【0068】

まず、I/Oプロセッサ220が、チャネルアダプタ210からリードアクセス要求を取得すると(ステップS21)、I/Oプロセッサ220が、リードアクセス要求の対象のデータ(リードデータ)がキャッシュメモリ253に格納されているか否かを判定する(ステップS22)。ここで、リードデータがキャッシュメモリ253に格納されているか否かは、例えば、共有メモリ254中の制御情報260に、リードアクセス要求内に含まれる論理アドレスに対応するレコードが格納されているか否かを確認することによって判定することができる。

【0069】

ステップS22での判定の結果、キャッシュメモリ253に格納されていると判定した場合(ステップS22のYes)には、I/Oプロセッサ220が、キャッシュメモリ253から対応するキャッシュデータを読み出して、チャネルアダプタ210により要求元のホスト装置10に読み出したデータを送信させ(ステップS23)、リードアクセス要求時処理を終了する。

【0070】

一方、ステップS22での判定の結果、キャッシュメモリ253に格納されていないと判定した場合(ステップS22のNo)には、I/Oプロセッサ220が、対応するデータをステージさせる(ステップS24)。すなわち、I/Oプロセッサ220が、対応するデータを格納している記憶デバイス310から当該データを読み出して、キャッシュメモリ253に格納する。次いで、I/Oプロセッサ220は、共有メモリ254の構成情報261中の読み出したデータに対応するレコードを更新する(ステップS25)。すなわち、I/Oプロセッサ220は、共有メモリ254の制御情報260中に、読み出したデータに対応するレコードを追加し、当該レコードのキャッシュアドレス2602に、読み出したデータを格納したキャッシュメモリ253のキャッシュアドレスを格納し、更に、ダーティビット2604にクリーンデータであることを示す“0”を格納する。そして、I/Oプロセッサ220が、当該キャッシュメモリ253に読み出したデータをチャネルアダプタ210により要求元のホスト装置10に送信させて、リードアクセス要求時処理を終了する。

【0071】

図6は、本発明の第1実施形態に係るディスクサブシステムにおけるステージ及びデステージを説明する図である。

【0072】

ステージとは、図6に示すように、記憶デバイス310に格納されたデータをキャッシュメモリ253に格納することをいい、デステージとは、図6に示すように、キャッシュメモリ253に格納されたキャッシュデータを記憶デバイス310に反映させることをいう。

【0073】

次に、本発明の第1実施形態に係るストレージサブシステム100におけるデータ退避処理を説明する。

【0074】

図7は、本発明の第1実施形態に係るデータ退避処理のフローチャートである。

【 0 0 7 5 】

データ退避処理は、電圧監視制御部 2 5 7 が、電源障害、例えば、電源回路 4 0 0 から供給される電圧が異常値を示していることを検出したことによって開始される（ステップ S 3 1）。電圧監視制御部 2 5 7 は、メモリボード 2 5 0 のプロセッサ 2 5 2 に、電圧異常が発生したことを通知するとともに、メモリボード 2 5 0 の各部に供給する電力を、電源回路 4 0 0 から供給される電力からバッテリー 5 0 0 から供給される電力に切り替える（ステップ S 3 2）。これによって、メモリボード 2 5 0 の各部は、バッテリー 5 0 0 から供給される電力によって動作を継続することができる。このため、キャッシュメモリ 2 5 3 及び共有メモリ 2 5 4 は、データの記憶を維持しておくことができる。なお、以下の処理は、バッテリー 5 0 0 の電力がメモリボード 2 5 0 のみに供給されていればよい。このため、バッテリー 5 0 0 に蓄えておくべき電力量を低減することができる。

10

【 0 0 7 6 】

プロセッサ 2 5 2 は、共有メモリ 2 5 4 の制御情報 2 6 0 を参照し（ステップ S 3 3）、キャッシュメモリ 2 5 3 中の 1 つのキャッシュメモリデバイス进行处理対象として未退避のダーティデータがあるか否かを判定する（ステップ S 3 4）。

【 0 0 7 7 】

ステップ S 3 4 の判定の結果、キャッシュメモリデバイスに未退避のダーティデータがあると判定した場合（ステップ S 3 4 の Y e s）には、プロセッサ 2 5 2 は、当該キャッシュメモリデバイスからダーティデータを読み出し（ステップ S 3 5）、不揮発メモリ 2 5 6 のアドレス管理テーブル 2 6 2 に従って、当該ダーティデータを格納する不揮発メモリ 2 5 5 のアドレス（不揮発メモリアドレス）を決定し、当該不揮発メモリアドレスに対応するレコードのキャッシュメモリアドレス 2 6 2 8 に、当該ダーティデータが格納されていたキャッシュメモリ 2 5 3 のアドレスを格納するとともに、データ長 2 6 2 9 に、当該ダーティデータのデータ長を格納する（ステップ S 3 6）。

20

【 0 0 7 8 】

次いで、プロセッサ 2 5 2 は、ダーティデータと共に、当該ダーティデータを格納する不揮発メモリアドレスをメモリコントローラ 2 5 1 に渡す。メモリコントローラ 2 5 1 は、プロセッサ 2 5 2 から渡されたダーティデータを暗号化して（ステップ S 3 7）、不揮発メモリ 2 5 5 の指定された不揮発メモリアドレスに格納する（ステップ S 3 8）。このように、ダーティデータは、暗号化されて不揮発メモリ 2 5 5 に格納されているので、不揮発メモリ 2 5 5 からデータが読取られても、当該データからもとのデータを容易に把握することができないため、情報が漏洩することを適切に防止することができる。

30

【 0 0 7 9 】

次いで、プロセッサ 2 5 2 は、同じキャッシュメモリデバイスを対象に、上記したステップ S 3 3 からの処理を繰り返す。このように処理を繰り返すことによって、同じキャッシュメモリデバイスに格納されているすべてのダーティデータを不揮発メモリ 2 5 5 に退避させることができる。

【 0 0 8 0 】

一方、ステップ S 3 4 の判定の結果、キャッシュメモリデバイスに未退避のダーティデータがないと判定した場合（ステップ S 3 4 の N o）には、対象のキャッシュメモリデバイスにダーティデータが存在しない、又は当該キャッシュメモリデバイスのすべてのダーティデータを退避させたことを示しているため、プロセッサ 2 5 2 は、電圧監視制御部 2 5 7 により、当該キャッシュメモリデバイスへの電力供給を切断させ（ステップ S 3 9）、ダーティデータを退避する処理の対象となる他のキャッシュメモリデバイスがあるか否かを判定する（ステップ S 4 0）。

40

【 0 0 8 1 】

ステップ S 4 0 の結果、ダーティデータを退避する処理の対象となる他のキャッシュメモリデバイスがある場合（ステップ S 4 0 の Y e s）には、他のキャッシュメモリデバイスに対して上記同様のステップ S 3 3 からの処理を実行する。

【 0 0 8 2 】

50

一方、ステップS 4 0の結果、ダーティデータを退避する処理の対象となる他のキャッシュメモリデバイスがない場合（ステップS 4 0のNo）には、キャッシュメモリ2 5 3のすべてのダーティデータの退避が完了したことを意味しているので、プロセッサ2 5 2は、共有メモリ2 5 4から構成情報2 6 1と、制御情報2 6 0のうちのダーティデータに関する制御情報を読み出し（ステップS 4 1）、不揮発メモリ2 5 6の制御情報のアドレス管理テーブル2 6 2 A及び構成情報のアドレス管理テーブル2 6 2 Bに従って、構成情報及び制御情報を格納する不揮発メモリ2 5 5のアドレス（不揮発メモリアドレス）を決定し、当該不揮発メモリアドレスに対応するレコードの共有メモリアドレス2 6 2 2、2 6 2 5に、当該構成情報又は制御情報が格納されていた共有メモリ2 5 4のアドレスを格納するとともに、データ長2 6 2 3、2 6 2 6に、当該構成情報又は制御情報のデータ長を格納する（ステップS 4 2）。

10

【0083】

次いで、プロセッサ2 5 2は、構成情報及び制御情報と共に、当該構成情報及び制御情報を格納する不揮発メモリアドレスをメモリコントローラ2 5 1に渡す。メモリコントローラ2 5 1は、プロセッサ2 5 2から渡された構成情報及び制御情報を暗号化して（ステップS 4 3）、不揮発メモリ2 5 5の指定された不揮発メモリアドレスに格納する（ステップS 4 4）。次いで、プロセッサ2 5 2は、電圧監視制御部2 5 7により、当該メモリボード2 5 0への電力供給を切断させる（ステップS 4 5）。

【0084】

本実施形態においては、上記データ退避処理に係るメモリコントローラ2 5 1、キャッシュメモリ2 5 3、共有メモリ2 5 4、不揮発メモリ2 5 5、2 5 6、及びプロセッサ2 5 2が同一のメモリボード上2 5 0にあるので、データ退避処理を迅速に行うことができる。

20

【0085】

図8は、本発明の第1実施形態に係るデータの退避を説明する図である。

【0086】

上記した図7に示すデータ退避処理が実行されると、キャッシュメモリ2 5 3に格納されたダーティデータであるデータd 2が不揮発メモリ2 5 5に退避される。また、共有メモリ2 5 4の構成情報2 6 1も不揮発メモリ2 5 5に退避される。また、共有メモリ2 5 4の制御情報2 6 0の中でデータd 2の制御情報も不揮発メモリ2 5 5に退避される。この際には、不揮発メモリ2 5 6には、不揮発メモリ2 5 5に退避させた、データd 2、構成情報、及び制御情報の元の格納先を示すアドレス管理テーブル2 6 2が格納される。

30

【0087】

図9は、本発明の第1実施形態に係るデータ復旧処理のフローチャートである。

【0088】

データ復旧処理は、ディスク制御装置2 0 0の電源が復帰した場合に開始され（ステップS 5 1）、まず、I/Oプロセッサ2 2 0が、不揮発メモリ2 5 6内のアドレス管理テーブル2 6 2を参照し（ステップS 5 2）、復旧すべきデータ等（キャッシュデータ、構成情報、制御情報）が格納されているか否かを判定する（ステップS 5 3）。なお、データ等が格納されているか否かは、アドレス管理テーブル2 6 2の共有メモリアドレス2 6 2 2、2 6 2 5、又はキャッシュメモリアドレス2 6 2 8にアドレスが格納されているか否かによって判定できる。

40

【0089】

この判定の結果、復旧すべきデータ等が格納されている場合（ステップS 5 3のYes）には、I/Oプロセッサ2 2 0は、メモリコントローラ2 5 1にアドレス管理テーブル2 6 2に従って不揮発メモリ2 5 5の対応するアドレスからデータ等をリードさせ、更に当該リードしたデータ等を復号化させ、該当するデータ等のアドレスを揮発メモリ（キャッシュメモリ2 5 3又は共有メモリ2 5 4）用のアドレスに変換する（ステップS 5 6）。すなわち、アドレス管理テーブル2 6 2から対応するデータ等の共有メモリアドレス2 6 2 2、2 6 2 5、又はキャッシュメモリアドレス2 6 2 8を取得する。

50

【 0 0 9 0 】

次いで、I/Oプロセッサ220は、メモリコントローラ251により、変換したアドレスに基づいてデータ等を共有メモリ254又はキャッシュメモリ253に書き込ませ（ステップS57）、復旧すべき他のデータ等があるか否かを判定し（ステップS58）、復旧すべき他のデータがある場合（ステップS58のYes）には、上記ステップS54からの処理を繰り返し実行することにより、データ退避前のダーティデータ及びダーティデータに関する構成情報及び制御情報を元の状態に復旧する。これにより、ダーティデータについてデータ退避前と同様に各種処理に利用することができるようになる。

【 0 0 9 1 】

一方、復旧すべきデータ等が格納されていない場合（ステップS53のNo）、又は復旧すべき全てのデータの復旧を終えた場合（ステップS58のNo）には、通常のI/O処理に移行する（ステップS59）。

【 0 0 9 2 】

本実施形態にかかるメモリボード250は、上述のようにディスク制御装置200に対して着脱可能であり、また、メモリボード250の不揮発メモリ255には、キャッシュデータが退避される。このため、キャッシュデータが退避されたメモリボード250がディスク制御装置200から取り外されて、他のディスク制御装置200に装着されるとキャッシュデータの内容が漏洩してしまう虞がある。そこで、本実施形態では、以下のデータ復旧判定処理を実行するようにして、メモリボード250に退避されているデータが漏洩しないようにしている。

【 0 0 9 3 】

図10は、本発明の第1実施形態に係るデータ復旧判定処理のフローチャートである。

【 0 0 9 4 】

I/Oプロセッサ220は、ディスク制御装置200のメモリボードスロットにメモリボード250が挿入されたことを検出すると（ステップS61）、装着されたメモリボード250の不揮発メモリ256からメモリボード250の固有識別子263を取得し、当該固有識別子263と、当該メモリボード250が装着されたスロットの番号とに基づいてデータ復旧の要否を判定する（ステップS62）。本実施形態では、I/Oプロセッサ220は、コントロールユニット230から以前に装着されていたメモリボード250の固有識別子及びスロット番号を取得し、新たに装着されたメモリボード250の固有識別子263及びスロット番号と一致するか否かにより復旧の要否を判定している。すなわち、固有識別子263及びスロット番号が一致する場合には、メモリボード250を一時的に取り除き、同一のメモリボード250を再び同じスロットに挿入したことを意味しているので、データの復旧を行うこととし、固有識別子263が違う場合には、メモリボードが当該ディスク制御装置200に以前に装着されていたものではないので、当該メモリボードのデータが漏洩することを防止するためにデータの復旧を行わず、また、固有識別子が同じでもスロット番号が違うと、データ退避時とは違う接続をしていることを意味しているので、データの復旧を行わないこととしている。

【 0 0 9 5 】

上記判定の結果、データ復旧が必要ないと判定した場合（ステップS63のNo）には、確実にデータの漏洩を防止するために、I/Oプロセッサ220は、メモリコントローラ251により、不揮発メモリ255のデータを初期化し、例えば、全ての記憶領域に“0”を書き込み（ステップS64）、通常I/Oの処理へ移行する（ステップS66）。

【 0 0 9 6 】

一方、データ復旧が必要であると判定した場合（ステップS63のYes）には、データ復旧処理（ステップS65：図9のステップS52以降の処理と同様）を実行し、通常I/Oの処理へ移行する（ステップS66）。

【 0 0 9 7 】

次に、上記第1実施形態に係る計算機システムの変形例について説明する。

【 0 0 9 8 】

図 11 は、本発明の変形例に係る計算機システムの構成図である。なお、第 1 実施形態と同様な機能部分には、同一番号を付し、説明を省略する。

【0099】

変形例に係るディスクサブシステム 101 のディスク制御装置 201 は、第 1 実施形態に係るディスク制御装置 200 におけるチャンネルアダプタ 210 に代えてチャンネルアダプタ 212 を備え、ディスクアダプタ 270 に代えてディスクアダプタ 272 を備え、メモリボード 250 の共有メモリ 254 をメモリボード 250 と異なる共有メモリボード 265 に備え、I/O プロセッサ 220 を取り除いた構成となっている。

【0100】

チャンネルアダプタ 212 は、チャンネルアダプタ 210 に対してプロセッサ 213 を更に備えている。ディスクアダプタ 272 は、ディスクアダプタ 270 にプロセッサ 273 を更に備えている。プロセッサ 213 及びプロセッサ 273 は、I/O プロセッサ 220 によって実現されていた処理を分散して実行する。

【0101】

このようなディスク制御装置 201 においても、上記した図 7、図 9、図 10 と同様な処理を実行することができ、同様な効果を得ることができる。この場合においては、図 7、図 9 においてプロセッサ 252 が実行していた処理は、例えば、プロセッサ 213、273 のいずれかが実行すればよく、また、図 10 における I/O プロセッサ 220 の処理は、例えば、プロセッサ 213、273 のいずれかが実行すればよい。

< 第 2 実施形態 >

図 12 は、本発明の第 2 実施形態に係る計算機システムの構成図である。なお、第 1 実施形態と同様な機能部分には、同一番号を付すこととする。

【0102】

ディスク制御装置 202 は、同一の構成を有する複数のクラスタ 203 を有している。各クラスタ 203 は、例えば、一つの制御ボードサブシステム制御ボードにより構成され、チャンネルアダプタ 210 と、I/O プロセッサ 280 と、サブシステムコントローラ 281 と、揮発メモリ 282 と、不揮発メモリ 283 と、ディスクアダプタ 270 と、電圧監視制御部 257 とを有している。

【0103】

電源回路 400 は、例えば、外部の商用電源から供給される電力をディスク制御装置 202 の各部に供給する。本実施形態では、電源回路 400 は、多重化されておらず、複数のクラスタ 203 の各部に電力を供給するようになっている。なお、電源回路 400 を複数備えるようにして、それぞれが各クラスタ 203 に電力を供給するようにしてもよい。

【0104】

バッテリー 500 は、電力を蓄積しており、ディスク制御装置 202 の所定の部位に供給できるようになっている。本実施形態では、本実施形態では、バッテリー 500 は、多重化されておらず、複数のクラスタ 203 の所定の部位に電力を供給するようになっている。なお、バッテリー 500 を複数備えるようにして、それぞれが各クラスタ 203 の所定の部位に電力を供給するようにしてもよい。

【0105】

I/O プロセッサ 280 は、揮発メモリ 282 に読み出されたプログラムを実行することによりクラスタ 203 の全体の動作を制御する。I/O プロセッサ 280 は、チャンネルアダプタ 210 と、揮発メモリ 282 と、不揮発メモリ 283 と、ディスクアダプタ 270 との間でのサブシステムコントローラ 281 によるデータの受け渡しを制御する。例えば、チャンネルアダプタ 210 が受信したデータを揮発メモリ 282 に格納させる制御を行う。また、I/O プロセッサ 280 は、揮発メモリ 282 に格納されたデータを、ディスクアダプタ 270 に渡したり、チャンネルアダプタ 210 に渡したりする制御を行う。また、I/O プロセッサ 280 は、ディスクアダプタ 270 が記憶デバイス 310 から取得したデータを揮発メモリ 282 に格納させる制御を行う。また、I/O プロセッサ 280 は、不揮発メモリ 283 に退避格納したデータを揮発メモリ 282 に復旧するための処理を

行う。

【 0 1 0 6 】

サブシステムコントローラ 2 8 1 は、チャンネルアダプタ 2 1 0、ディスクアダプタ 2 7 0、揮発メモリ 2 8 2、不揮発メモリ 2 8 3、プロセッサ 2 8 0、及び他のクラスタ 2 0 3 のサブシステムコントローラ 2 8 1 に接続されており、各部でやりとりするデータの中継を行う。この構成により、サブシステムコントローラ 2 8 1 は、I / O プロセッサ 2 8 0 の制御によって、チャンネルアダプタ 2 1 0 を介してホスト装置 1 0 から受信したライトデータを揮発メモリ 2 8 2 に格納させると共に、ライトデータを他のクラスタ 2 0 3 のサブシステムコントローラ 2 8 1 に送信して、ライトデータを他のクラスタ 2 0 3 側の揮発メモリ 2 8 2 に格納させることができる。また、本実施形態では、サブシステムコントローラ 2 8 1 は、I / O プロセッサ 2 8 0 の制御によって、データを他のクラスタ 2 0 3 のサブシステムコントローラ 2 8 1 に送信して、他のクラスタ 2 0 3 の不揮発メモリ 2 8 3 に格納させることや、他のクラスタ 2 0 3 の不揮発メモリ 2 8 3 からデータを読み出すことができる。

10

【 0 1 0 7 】

また、サブシステムコントローラ 2 8 1 は、不揮発メモリ 2 8 3 の複数の不揮発メモリデバイス 2 8 3 1 (図 1 3 参照) により構成される R A I D グループへのデータ格納処理を実行する。例えば、サブシステムコントローラ 2 8 1 は、R A I D グループに格納する際には、格納対象のデータを所定のデータ単位に分割すると共に、複数個 (例えば、3 つ) のデータ単位毎に、それらデータ単位毎のパリティをパリティ生成回路 2 8 1 1 により生成し、これら複数個のデータ単位と作成したパリティとを R A I D グループ内の異なる不揮発メモリデバイス 2 8 3 1 に格納する。本実施形態では、サブシステムコントローラ 2 8 1 は、データ単位と、パリティとを暗号化して不揮発メモリデバイス 2 8 3 1 に格納する。

20

【 0 1 0 8 】

次に、揮発メモリ 2 8 2 及び不揮発メモリ 2 8 3 について詳細に説明する。

【 0 1 0 9 】

図 1 3 は、本発明の第 2 実施形態に係るストレージ制御装置の一部を詳細に説明する図である。

【 0 1 1 0 】

揮発メモリ 2 8 2 は、第 1 実施形態の共有メモリ 2 5 4 及びキャッシュメモリ 2 5 3 と同様な各種情報を記憶する。記憶する情報としては、例えば、ホスト装置 1 0 とのやりとりが行われるデータに関する構成情報 2 6 1 (例えば、記憶デバイス 3 1 0 のどこに格納されるかを示す情報) 及び制御情報 2 6 0 (例えば、揮発メモリ 2 8 2 のどこに格納されるかを示す情報) とがある。また、揮発メモリ 2 8 2 は、チャンネルアダプタ 2 1 0 により受信されたデータや、ディスクアダプタ 2 7 0 により記憶デバイス 3 1 0 から取得されたデータを一時的に記憶する。揮発メモリ 2 8 2 は、例えば、個別に入出力動作が可能な複数の揮発メモリデバイスによって構成されている。

30

【 0 1 1 1 】

不揮発メモリ 2 8 3 は、例えば、フラッシュメモリ、M R A M (Magnetoresistive Random Access Memory)、P R A M (Phase change RAM) 等の電源を供給せずともデータを記憶しておくことのできるメモリである。不揮発メモリ 2 8 3 は、複数の不揮発メモリデバイス 2 8 3 1 によって構成されている。不揮発メモリ 2 8 3 は、例えば、揮発メモリ 2 8 2 に格納されたデータ (キャッシュデータ、構成情報、制御情報) を退避格納するために用いられる。本実施形態では、キャッシュデータ、構成情報、及び制御情報については、複数のクラスタ 2 0 3 の不揮発メモリ 2 8 3 の複数の不揮発メモリデバイス 2 8 3 1 によって構成された R A I D グループに対して格納するようにしている。R A I D グループに格納する際には、例えば、R A I D レベル 2 ~ 5 のいずれであってもよい。これらの R A I D レベルであれば、不揮発メモリ 2 8 3 に必要となる容量を抑えつつ、データの信頼性を向上することができる。また、不揮発メモリ 2 8 3 は、不揮発メモリ 2 8 3 に退避格

40

50

納されたデータを元の状態に復旧するためのアドレス管理テーブル 262 を記憶する。なお、アドレス管理テーブル 262 における不揮発メモリアドレス 2621、2624、2627 は、第 2 実施形態では、不揮発メモリ 283 の物理的なアドレスでなく、RAID グループの論理的な記憶領域でのアドレス（論理アドレス）となっている。サブシステムコントローラ 281 は、この論理アドレスから物理的なアドレス（すなわち、どの不揮発メモリデバイス 2831（他のクラスタ 203 の不揮発メモリデバイス 2831 も含む）のどのアドレスであるか）を特定することができるようになっている。また、アドレス管理テーブル 262 の共有メモリアドレス 2622、共有メモリアドレス 2625、及びキャッシュメモリアドレス 2628 には、揮発メモリ 282 におけるアドレスが格納される。

10

【0112】

図 14 は、本発明の第 2 実施形態に係るデータ退避処理のフローチャートである。

【0113】

データ退避処理は、電圧監視制御部 257 が、電源障害、例えば、電源回路 400 から供給される電圧が異常値を示していることを検出したことによって開始される（ステップ S71）。電圧監視制御部 257 は、I/O プロセッサ 280 に、電圧異常が発生したことを通知するとともに、クラスタ 203（サブシステム制御ボード）の各部に供給する電力を、電源回路 400 から供給される電力からバッテリー 500 から供給される電力に切り替える（ステップ S72）。これによって、サブシステム制御ボードの各部は、バッテリー 500 から供給される電力によって動作を継続することができる。このため、揮発メモリ 282 は、データの記憶を維持しておくことができる。

20

【0114】

I/O プロセッサ 280 は、揮発メモリ 282 の制御情報 260 を参照し（ステップ S73）、揮発メモリ 282 中の 1 つの揮発メモリデバイスを処理対象として未退避のダーティデータがあるか否かを判定する（ステップ S74）。

【0115】

ステップ S74 の判定の結果、揮発メモリデバイスに未退避のダーティデータがあると判定した場合（ステップ S74 の Yes）には、I/O プロセッサ 280 は、当該揮発メモリデバイスからダーティデータを読み出し（ステップ S75）、不揮発メモリ 283 のアドレス管理テーブル 262 に従って、当該ダーティデータを格納する不揮発メモリ 283 の論理アドレスを決定し、当該不揮発メモリ 283 の論理アドレスに対応するレコードのキャッシュメモリアドレス 2628 に、当該ダーティデータが格納されていた揮発メモリ 282 のアドレスを格納し、また、データ長 2629 に当該ダーティデータのデータ長を格納する（ステップ S76）。

30

【0116】

次いで、I/O プロセッサ 280 は、ダーティデータと共に、当該ダーティデータを格納する不揮発メモリ 283 の論理アドレスをサブシステムコントローラ 281 に渡す。サブシステムコントローラ 281 は、I/O プロセッサ 280 から渡されたダーティデータを所定の大きさ（データ量）のデータ単位に分割して、複数のデータ単位毎に、それらデータ単位に対応するパリティを生成し（ステップ S77）、複数のデータ単位及び作成したパリティを暗号化する（ステップ S78）。次いで、サブシステムコントローラ 281 は、指定された不揮発メモリ 283 の論理アドレスに基づいて、各データ単位及びパリティのそれぞれを格納する物理アドレスを特定し、対応する物理アドレスが示す不揮発メモリデバイス 2831 に各データ単位及びパリティを格納する（ステップ S79）。本実施形態では、他のクラスタ 203 の不揮発メモリデバイス 2813 にも格納される。これによって、データ及びそれらデータに対応するパリティが複数の不揮発メモリデバイス 2831 に亘って分散して格納されることになる。したがって、いずれか 1 つの不揮発メモリデバイス 2831 に障害が発生しても元のデータが復元できることとなる。また、ダーティデータは、暗号化されて不揮発メモリ 283 に格納されているので、不揮発メモリ 283 からデータが読取られても、当該データからもとのデータを容易に把握することができ

40

50

ないため、情報が漏洩することを適切に防止することができる。

【 0 1 1 7 】

次いで、I / O プロセッサ 2 8 0 は、同じ揮発メモリデバイスを対象に、上記したステップ S 7 3 からの処理を繰り返す。このように処理を繰り返すことによって、同じ揮発メモリデバイスに格納されているすべてのダーティデータを不揮発メモリ 2 8 3 に退避させることができる。

【 0 1 1 8 】

一方、ステップ S 7 4 の判定の結果、揮発メモリデバイスに未退避のダーティデータがないと判定した場合（ステップ S 7 4 の N o ）には、対象の揮発メモリデバイスにダーティデータが存在しない、又は当該揮発メモリデバイスのすべてのダーティデータを退避させたことを示しているの、I / O プロセッサ 2 8 0 は、電圧監視制御部 2 5 7 により、当該揮発メモリデバイスへの電力供給を切断させ（ステップ S 8 0 ）、ダーティデータを退避する処理の対象となる他の揮発メモリデバイスがあるか否かを判定する（ステップ S 8 1 ）。

10

【 0 1 1 9 】

ステップ S 8 1 の結果、ダーティデータを退避する処理の対象となる他の揮発メモリデバイスがある場合（ステップ S 8 1 の Y e s ）には、他の揮発メモリデバイスに対して上記同様のステップ S 7 3 からの処理を実行する。

【 0 1 2 0 】

一方、ステップ S 8 1 の結果、ダーティデータを退避する処理の対象となる他の揮発メモリデバイスがない場合（ステップ S 8 1 の N o ）には、揮発メモリ 2 8 2 のすべてのダーティデータの退避が完了したことを意味しているの、I / O プロセッサ 2 8 0 は、揮発メモリ 2 8 2 から構成情報 2 6 1 と、制御情報 2 6 0 のうちのダーティデータに関する制御情報を読み出し（ステップ S 8 1 ）、不揮発メモリ 2 8 3 のアドレス管理テーブル 2 6 2 に従って、当該構成情報 2 6 1 及び制御情報 2 6 0 を格納する不揮発メモリ 2 8 3 の論理アドレスを決定し、当該不揮発メモリの論理アドレスに対応するレコードの共有メモリアドレス 2 6 2 2 又は 2 6 2 5 に、当該構成情報 2 6 1 又は制御情報 2 6 0 が格納されていた揮発メモリ 2 8 2 のアドレスを格納し、また、データ長 2 6 2 3、2 6 2 6 に当該構成情報 2 6 1 又は制御情報 2 6 0 のデータ長を格納する（ステップ S 8 3 ）。

20

【 0 1 2 1 】

次いで、I / O プロセッサ 2 8 0 は、構成情報 2 6 1 及び制御情報 2 6 0 と共に、格納する不揮発メモリ 2 8 3 の論理アドレスをサブシステムコントローラ 2 8 1 に渡す。サブシステムコントローラ 2 8 1 は、I / O プロセッサ 2 8 0 から渡された構成情報 2 6 1 及び制御情報 2 6 0 を所定の大きさ（データ量）のデータ単位に分割して、所定数のデータ単位毎に、それらデータ単位に対応するパリティを生成し（ステップ S 8 4 ）、複数のデータ単位及び作成したパリティを暗号化する（ステップ S 8 5 ）。次いで、サブシステムコントローラ 2 8 1 は、指定された不揮発メモリ 2 8 3 の論理アドレスに基づいて、各データ単位及びパリティのそれぞれを格納する物理アドレスを特定し、対応する物理アドレスが示す不揮発メモリデバイス 2 8 3 1 に各データ単位及びパリティを格納する（ステップ S 8 6 ）。これによって、データ及びそれらデータに対応するパリティが複数の不揮発メモリデバイス 2 8 3 1 に亘って分散して格納されることになる。したがって、いずれか 1 つの不揮発メモリデバイス 2 8 3 1 に障害が発生しても元のデータが復元できることとなる。

30

40

【 0 1 2 2 】

次いで、I / O プロセッサ 2 8 0 は、電圧監視制御部 2 5 7 により、クラスタ 2 0 3 のすべての部位への電力供給を切断させる（ステップ S 8 7 ）。

【 0 1 2 3 】

図 1 5 は、本発明の第 2 実施形態に係るデータ復旧処理のフローチャートである。

【 0 1 2 4 】

データ復旧処理は、ディスク制御装置 2 0 2 の電源が復帰した場合に開始され（ステッ

50

プS91)、まず、I/Oプロセッサ280が、不揮発メモリ283内のアドレス管理テーブル262を参照し(ステップS92)、復旧すべきデータ等(キャッシュデータ、構成情報、及び制御情報)が格納されているか否かを判定する(ステップS93)。なお、データ等が格納されているか否かは、アドレス管理テーブル262の共有メモリアドレス2622、2625、又はキャッシュメモリアドレス2628にアドレスが格納されているか否かによって判定できる。

【0125】

この判定の結果、復旧すべきデータ等が格納されている場合(ステップS93のYes)には、I/Oプロセッサ280は、サブシステムコントローラ281にアドレス管理テーブル262から取得した不揮発メモリ283の論理アドレスを渡す。サブシステムコントローラ281は、当該論理アドレスに対応する物理アドレスを特定し、当該物理アドレスが示す不揮発メモリ283からデータ等をリードし、更に当該リードしたデータ等を復号化し(ステップS95)、各データ等のパリティチェックを行う(ステップS96)。これによって、所定数のデータ単位と、これに対応するパリティとが所定の関係を有する場合には、データが正しいことを意味しているのでそのまま次の処理を行う一方、所定の関係を有していない場合には、データを再生して次の処理を行う。

【0126】

次いで、サブシステムコントローラ281は、複数のデータ単位をもとの並びに並べることによりもとのデータにし、I/Oプロセッサ280に渡す。I/Oプロセッサ280は、アドレス管理テーブル262の共有メモリアドレス2622、共有メモリアドレス2625、又はキャッシュメモリアドレス2628からデータが退避時に格納されていた揮発メモリ282のアドレスを取得する(ステップS97)。次いで、I/Oプロセッサ280は、サブシステムコントローラ281により、取得した揮発メモリ282のアドレスに不揮発メモリ283から取得したデータを格納させる(ステップS98)。なお、本実施形態では、他のクラスタのサブシステムコントローラ281に対しても、同様に、他のクラスタ203の揮発メモリ282に対してデータを格納させる。これにより、複数のクラスタの不揮発メモリ282の状態を同じにすることができる。

【0127】

次いで、I/Oプロセッサ280は、復旧すべき他のデータ等があるか否かを判定し(ステップS99)、復旧すべき他のデータがある場合(ステップS99のYes)には、上記ステップS94からの処理を繰り返し実行することにより、データ退避前のダートデータ及びダートデータに関する構成情報及び制御情報を元の状態に復旧する。これにより、ダートデータについてデータ退避前と同様に各種処理に利用することができるようになる。

【0128】

一方、復旧すべきデータ等が格納されていない場合(ステップS93のNo)、又は復旧すべき全てのデータの復旧を終えた場合(ステップS99のNo)には、通常のI/O処理に移行する(ステップS100)。

【0129】

以上、本発明を一実施形態に基づいて説明したが、本発明は上述した実施形態に限られず、他の様々な態様に適用可能である。

【0130】

例えば、上記各実施形態において、記憶デバイス310として、ハードディスクドライブ(HDD)を用いた例を示していたが、本発明はこれに限られず、例えば、ハードディスクドライブの少なくとも一部、又は全部を、DVDドライブ、磁気テープドライブ、フラッシュメモリデバイス等のデータを記憶可能な他の記憶デバイスに置き換えるようにしてもよい。

【0131】

また、上記第1実施形態において、共有メモリ254を揮発メモリとして説明していたが、本発明はこれに限られず、例えば、不揮発メモリとしてもよい。共有メモリ254を

10

20

30

40

50

不揮発メモリとした場合にあっては、データ退避を行う際に制御情報 2 6 0、構成情報 2 6 1 の退避処理を行わずに済む。

【 0 1 3 2 】

また、上記第 1 実施形態では、キャッシュメモリ 2 5 3 と共有メモリ 2 5 4 とを物理的に分離した構成として説明したが、これに限られず、キャッシュメモリ 2 5 3 と共有メモリ 2 5 4 とを 1 つの集合体としてもよい。

【 0 1 3 3 】

また、上記第 1 実施形態では、不揮発メモリ 2 5 5 と不揮発メモリ 2 5 6 とを物理的に分離した構成として説明したが、これに限られず、不揮発メモリを 1 つの集合体として構成してもよい。

10

【 0 1 3 4 】

上記各実施形態では、データ量が変わらない暗号化を用いていたが、本発明はこれに限られず、例えば、データ量が変わる暗号化を行うようにしてもよい。なお、この場合には、アドレス管理テーブル 2 6 2 に格納するデータ長を暗号化後のデータ長にするようにする必要がある。

【 0 1 3 5 】

また、上記各実施形態において、キャッシュメモリ 2 5 3 のダーティデータをそのままのデータ長で不揮発メモリ 2 5 5 に格納していたが、本発明はこれに限られず、例えば、キャッシュメモリ 2 5 3 のダーティデータを圧縮した後に、不揮発メモリ 2 5 5 に格納するようにしてもよい。このようにすると、不揮発メモリ 2 5 5 における記憶効率を向上させることができると共に、不揮発メモリ 2 5 5 への書き込み処理に要する時間を短縮することができる。

20

【 0 1 3 6 】

また、上記第 2 実施形態において、各クラスタ 2 0 3 に同様に不揮発メモリ 2 8 3 を備えるようにし、複数のクラスタ 2 0 3 の不揮発メモリ 2 8 3 により構成された R A I D グループにダーティデータを退避するようにしていたが、本発明はこれに限られず、例えば、一方のクラスタ 2 0 3 にのみ不揮発メモリ 2 8 3 を備え、ダーティデータの退避に用いるようにしてもよい。

【図面の簡単な説明】

【 0 1 3 7 】

30

【図 1】本発明の第 1 実施形態に係る計算機システムの構成図である。

【図 2】本発明の第 1 実施形態に係るメモリボードの構成図である。

【図 3】図 3 A は、本発明の第 1 実施形態に係る制御情報及び構成情報の一例を示す図である。図 3 B は、本発明の第 1 実施形態に係る制御情報及び構成情報の一例を示す図である。

【図 4】図 4 A は、本発明の第 1 実施形態に係る制御情報についてのアドレス管理テーブルの一例を示す図である。図 4 B は、本発明の第 1 実施形態に係る構成情報についてのアドレス管理テーブルの一例を示す図である。図 4 C は、本発明の第 1 実施形態に係るキャッシュデータについてのアドレス管理テーブルの一例を示す図である。

【図 5】図 5 A は、本発明の第 1 実施形態に係るライトアクセス要求時処理のフローチャートである。図 5 B は、本発明の第 1 実施形態に係るリードアクセス要求時処理のフローチャートである。

40

【図 6】本発明の第 1 実施形態に係るディスクサブシステムにおけるステージ及びデステージを説明する図である。

【図 7】本発明の第 1 実施形態に係るデータ退避処理のフローチャートである。

【図 8】本発明の第 1 実施形態に係るデータの退避を説明する図である。

【図 9】本発明の第 1 実施形態に係るデータ復旧処理のフローチャートである。

【図 1 0】本発明の第 1 実施形態に係るデータ復旧判定処理のフローチャートである。

【図 1 1】本発明の変形例に係る計算機システムの構成図である。

【図 1 2】本発明の第 2 実施形態に係る計算機システムの構成図である。

50

【図 1 3】本発明の第 2 実施形態に係るストレージ制御装置の一部を詳細に説明する図である。

【図 1 4】本発明の第 2 実施形態に係るデータ退避処理のフローチャートである。

【図 1 5】本発明の第 2 実施形態に係るデータ復旧処理のフローチャートである。

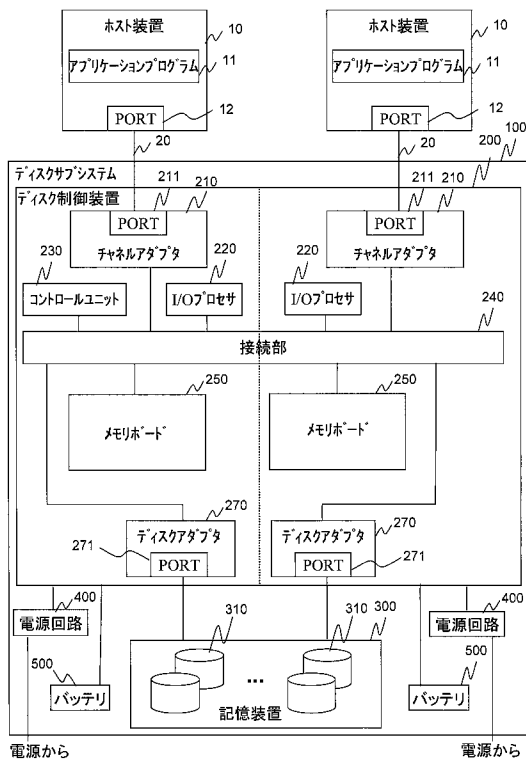
【符号の説明】

【 0 1 3 8 】

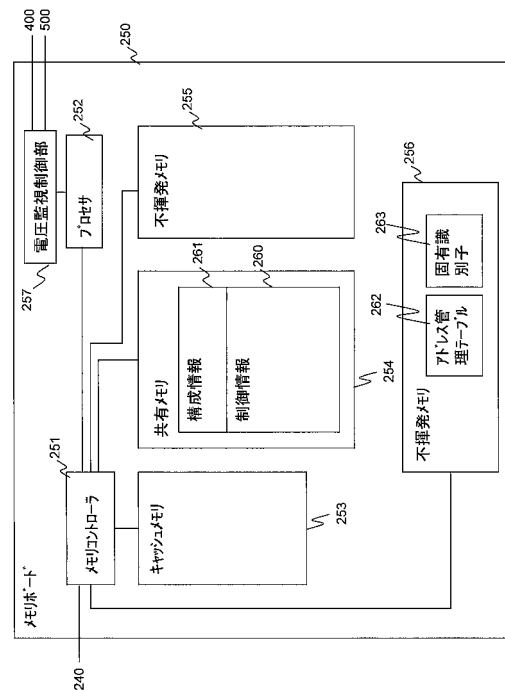
10 ホスト装置、20 ネットワーク、100 ディスクサブシステム、200 ディスク制御装置、210 チャネルアダプタ、220 I/Oプロセッサ、230 コントロールユニット、240 接続部、250 メモリボード、251 メモリコントローラ、252 プロセッサ、253 キャッシュメモリ、254 共有メモリ、255 不揮発メモリ、256 不揮発メモリ、257 電圧監視制御部、270 ディスクアダプタ、300 記憶装置、310 記憶デバイス、400 電源回路、500 バッテリ。

10

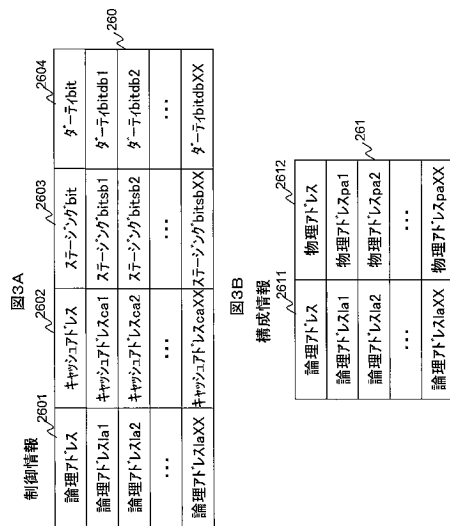
【図 1】



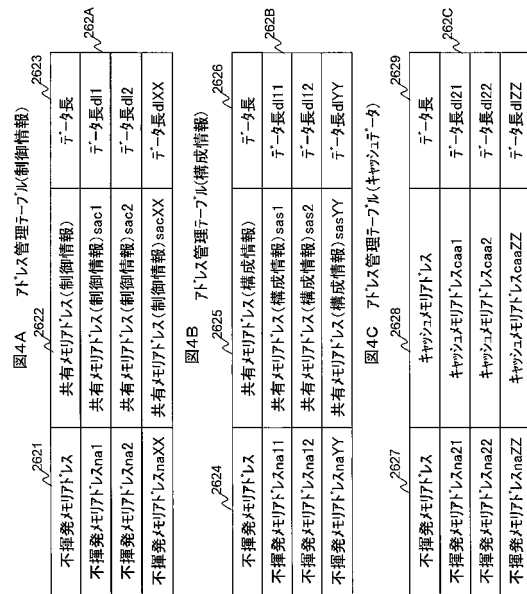
【図 2】



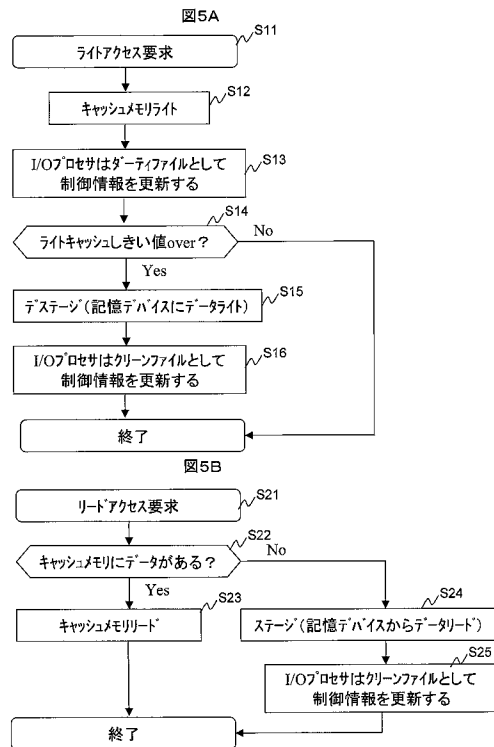
【図3】



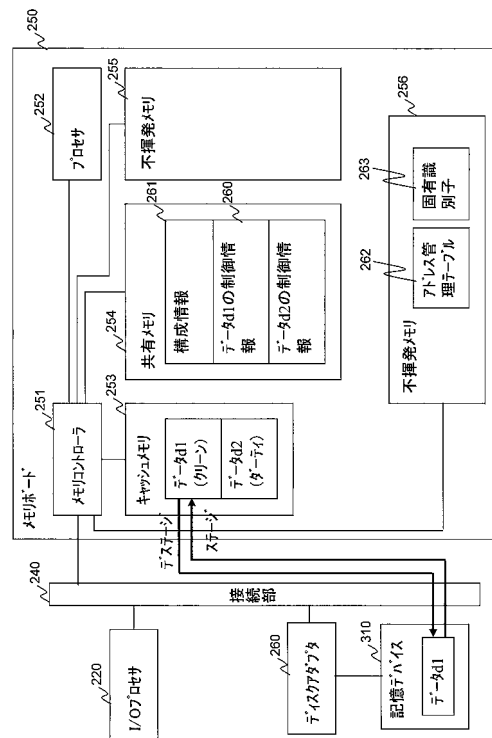
【図4】



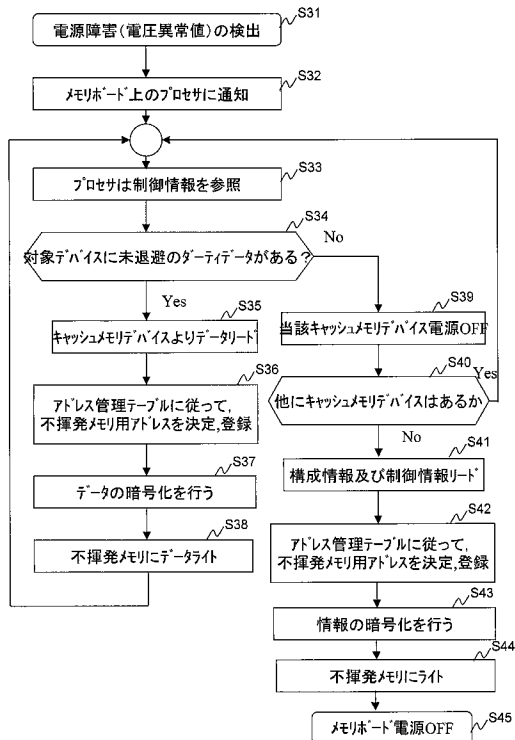
【図5】



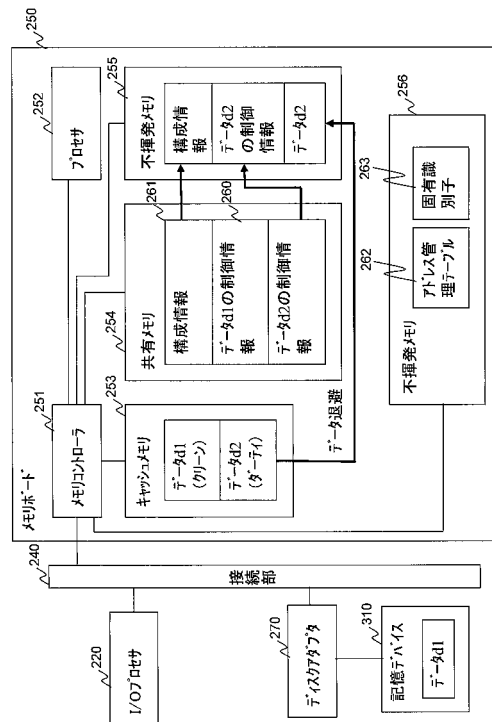
【図6】



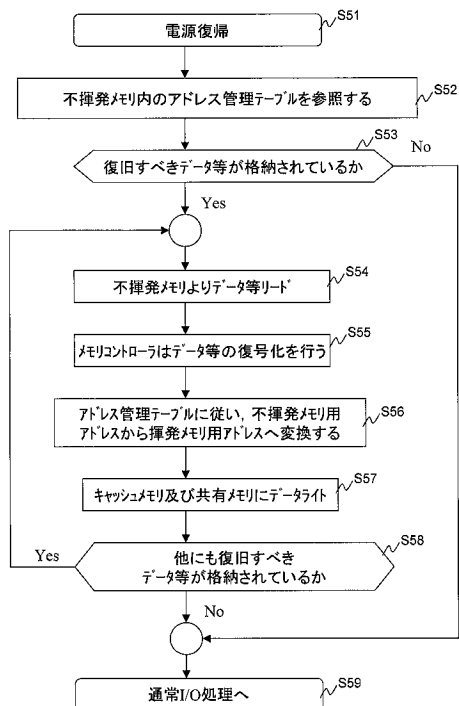
【 図 7 】



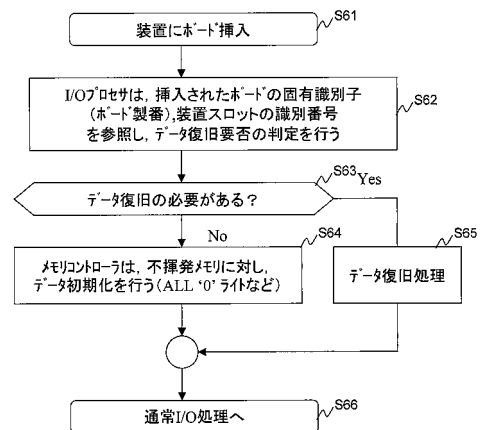
【 図 8 】



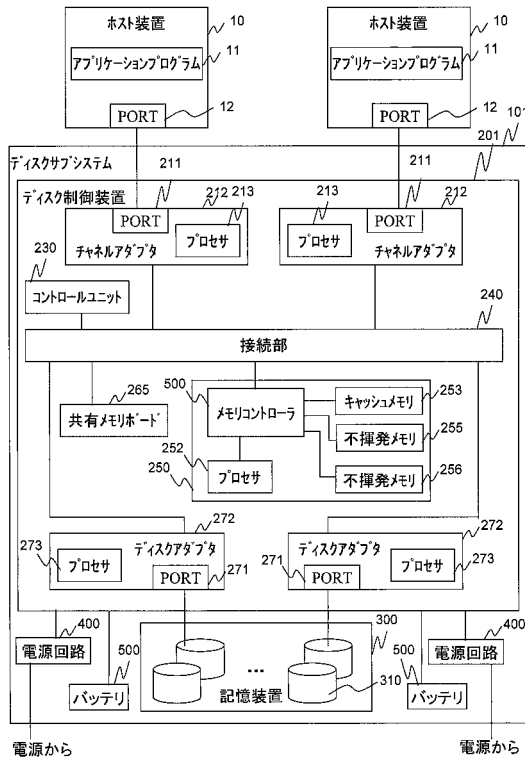
【 図 9 】



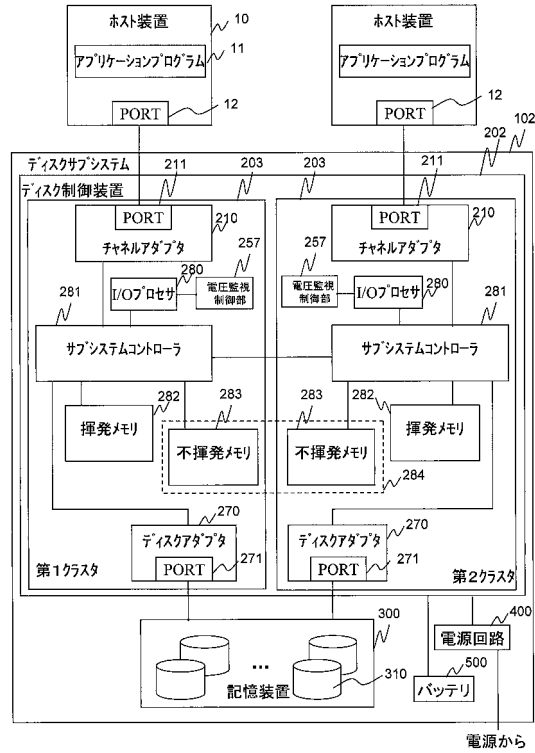
【 図 1 0 】



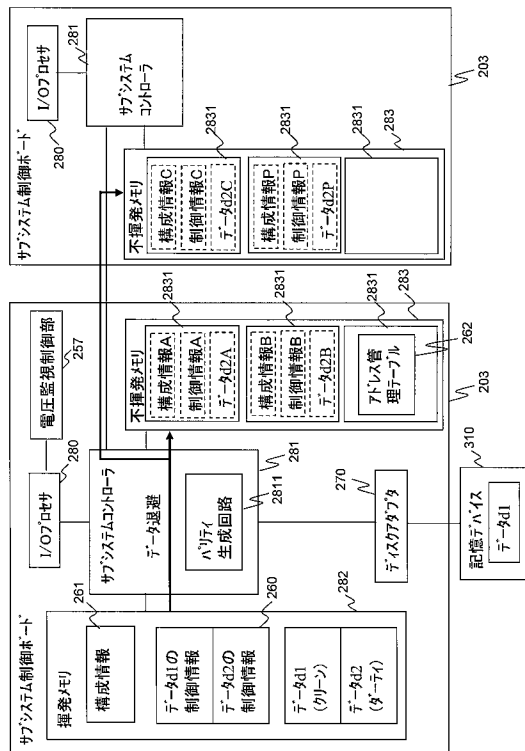
【 図 1 1 】



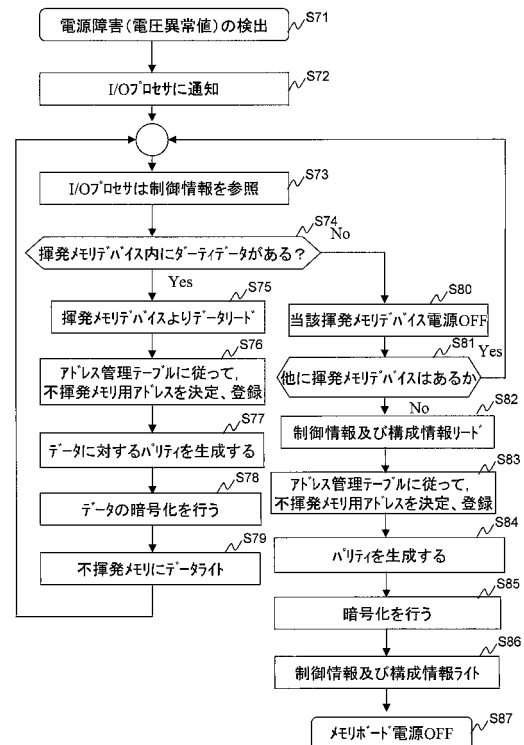
【 図 1 2 】



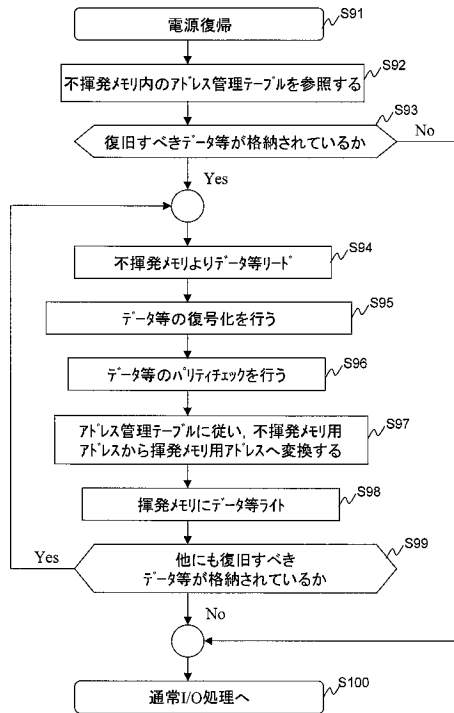
【 図 1 3 】



【 図 1 4 】



【図 15】



フロントページの続き

(51)Int.Cl. F I
G 0 6 F 12/08 5 0 7 H
G 0 6 F 12/08 5 4 3 B
G 0 6 F 12/16 3 4 0 Q

審査官 古河 雅輝

(56)参考文献 特開 2 0 0 6 - 3 1 3 4 0 7 (J P , A)
特開 2 0 0 5 - 0 4 4 0 1 0 (J P , A)
特開 2 0 0 4 - 2 8 7 4 7 6 (J P , A)
特開平 1 1 - 1 7 5 4 0 6 (J P , A)
特開平 0 6 - 3 0 9 2 3 4 (J P , A)
特開 2 0 0 0 - 2 0 7 1 3 7 (J P , A)
特開 2 0 0 5 - 3 3 2 4 7 1 (J P , A)
特開 2 0 0 4 - 0 2 1 8 1 1 (J P , A)
特開平 0 8 - 2 2 1 3 2 6 (J P , A)
特開 2 0 0 5 - 2 2 2 3 8 3 (J P , A)

(58)調査した分野(Int.Cl. , D B 名)
G 0 6 F 3 / 0 6 - 3 / 0 8
G 0 6 F 1 2 / 0 0
G 0 6 F 1 2 / 0 8 - 1 2 / 1 2
G 0 6 F 1 2 / 1 6
G 0 6 F 1 3 / 1 0 - 1 3 / 1 4