

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4919752号
(P4919752)

(45) 発行日 平成24年4月18日(2012.4.18)

(24) 登録日 平成24年2月10日(2012.2.10)

(51) Int. Cl.		F I			
G06F	1/32	(2006.01)	G06F	1/00	332B
G06F	3/06	(2006.01)	G06F	1/00	332Z
			G06F	3/06	
			G06F	3/06	540
			G06F	3/06	304R

請求項の数 11 (全 35 頁)

(21) 出願番号	特願2006-267135 (P2006-267135)	(73) 特許権者	000005108
(22) 出願日	平成18年9月29日 (2006.9.29)		株式会社日立製作所
(65) 公開番号	特開2008-90352 (P2008-90352A)		東京都千代田区丸の内一丁目6番6号
(43) 公開日	平成20年4月17日 (2008.4.17)	(74) 代理人	110000279
審査請求日	平成21年7月13日 (2009.7.13)		特許業務法人ウィルフォート国際特許事務所
		(72) 発明者	堀 雅則
			神奈川県小田原市中里322番2号 株式会社日立製作所 RAIDシステム事業部内
		(72) 発明者	佐藤 雅彦
			神奈川県小田原市中里322番2号 株式会社日立製作所 RAIDシステム事業部内

最終頁に続く

(54) 【発明の名称】 記憶制御装置

(57) 【特許請求の範囲】

【請求項1】

第1コントローラと、複数の第1記憶装置とを含む第1筐体と、
第1筐体と接続し、第2コントローラと、各々、所定時間アクセスされない場合にスピンドアウン状態となり、スピンドアウン状態にてアクセスを受領した場合にスピンドアアップ状態となる複数の第2記憶装置とを含む第2筐体と、
前記第2筐体と接続し、第3コントローラと、各々、前記所定時間アクセスされない場合にスピンドアウン状態となり、スピンドアウン状態にてアクセスを受領した場合にスピンドアアップ状態となる複数の第3記憶装置とを含む第3筐体と、を含み、
前記第1コントローラは、前記複数の第2記憶装置の全てがスピンドアウン状態の場合に、前記複数の第2記憶装置への電源供給を停止する制御をし、
前記第1コントローラは、前記複数の第2記憶装置への電源供給を停止した後に、前記第3筐体への電源供給を停止している場合に、前記第2筐体への電源供給を停止する制御をする、ことを特徴とする記憶制御装置。

10

【請求項2】

前記第1コントローラは、前記複数の第2記憶装置への電源供給を停止する場合に、前記第2コントローラへの電源供給を停止しない、ことを特徴とする請求項1に記載の記憶制御装置。

20

【請求項 3】

前記第 1 コントローラは、前記複数の第 3 記憶装置の全てがスピンダウン状態の場合に、前記複数の第 3 記憶装置への電源供給を停止する制御をし、

前記第 1 コントローラは、前記複数の第 3 記憶装置への電源供給を停止した後に、前記第 3 筐体にさらに他の筐体が接続されていない場合には、前記第 3 筐体への電源供給を停止する制御をする、ことを特徴とする請求項 1 に記載の記憶制御装置。

【請求項 4】

前記第 1 筐体は、前記第 2 筐体及び前記第 3 筐体の構成と電源供給状態とを管理するためのシステム構成情報を保持しており、前記第 1 コントローラは、前記システム構成情報に基づいて、前記複数の第 2 記憶装置及び前記複数の第 3 記憶装置への電源供給と前記第 2 筐体及び前記第 3 筐体への電源供給を制御する、ことを特徴とする請求項 1 に記載の記憶制御装置。

10

【請求項 5】

前記第 1 筐体と前記第 2 筐体及び前記第 3 筐体に接続され、外部電源と前記第 2 筐体及び前記第 3 筐体とを接続するスイッチを含む電源供給回路部を、さらに含み、

前記第 1 コントローラは、前記外部電源と前記第 2 筐体及び前記第 3 筐体とを接続する前記スイッチをオフさせることにより、前記第 2 筐体及び前記第 3 筐体への電源供給を停止する制御をする、ことを特徴とする請求項 1 に記載の記憶制御装置。

20

【請求項 6】

第 1 筐体が基本筐体であり、第 2 筐体及び第 3 筐体が増設筐体である、ことを特徴とする請求項 1 に記載の記憶制御装置。

【請求項 7】

前記第 1 筐体は、前記上位装置に接続され、前記上位装置による利用の頻度が予め設定された所定値以下のデータを、前記第 1 筐体から接続上の距離が遠い順番で、前記第 2 筐体内の前記複数の第 2 記憶装置又は前記第 3 筐体内の前記複数の第 3 記憶装置に記憶させる請求項 1 に記載の記憶制御装置。

30

【請求項 8】

前記複数の第 3 記憶装置は前記複数の第 2 記憶装置よりも低性能の記憶装置である、ことを特徴とする請求項 1 に記載の記憶制御装置。

【請求項 9】

前記第 1 筐体と前記第 2 筐体と、前記第 2 筐体と前記第 3 筐体とは、ダイジェーション接続されており、

前記第 1 筐体は、前記第 2 筐体及び前記第 3 筐体への電源供給を停止させた後で、前記第 3 筐体に信号の伝達が求められている場合には、前記複数の第 2 記憶装置への電源供給を停止させた状態にて前記第 2 コントローラに電源を供給する、ことを特徴とする請求項 1 に記載の記憶制御装置。

40

【請求項 10】

前記第 1 筐体は、電源供給がそれぞれ停止されている前記第 2 筐体及び前記第 3 筐体に電源を供給する場合には、前記第 2 筐体及び前記第 3 筐体への電源供給を時間差をつけて実行する請求項 1 に記載の記憶制御装置。

【請求項 11】

前記第 1 筐体は、電源供給が停止されている前記第 2 筐体に、定期的にまたは不定期に

50

電源を供給することにより、正常に動作するか否かを診断する請求項 1 に記載の記憶制御装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、記憶制御装置に関する。

【背景技術】

【0002】

記憶制御装置は、多数のハードディスクドライブをアレイ状に接続した記憶部を備えており、サーバ等のホストコンピュータ（以下「ホスト」）に、論理的な記憶領域（論理ボリューム）を提供する。信頼性等を高めるために、記憶制御装置では、RAID（Redundant Array of Independent Disks）に基づく冗長化された記憶領域をホストに提供する。

10

【0003】

ところで、企業等で管理すべきデータは日々増大する。従って、データ量の増大に応じて、記憶容量を増加させることのできる記憶制御装置が好ましい。そこで、制御機能を司る基本筐体に、記憶領域を提供する増設筐体を複数個接続させることにより、ユーザの希望に応じた記憶領域を提供可能な記憶制御装置が提案されている（特許文献 1）。

【0004】

この第 1 文献に記載の従来技術では、基本筐体への電源供給の開始または停止と、各増設筐体への電源供給の開始または停止とを連動させる。これにより、記憶制御装置を完全に停止させる場合には、基本筐体への電源供給を停止させるだけで、各増設筐体への電源供給を全て自動的に停止させることができる。

20

【0005】

第 2 文献に記載の他の従来技術では、RAIDグループを構成する各磁気ディスク装置を節電モードに移行させるようになっている（特許文献 2）。

【特許文献 1】特開 2001 - 339853 号公報

【特許文献 2】特開 2000 - 293314 号公報

【発明の開示】

【発明が解決しようとする課題】

【0006】

前記第 1 文献に記載の従来技術では、基本筐体への電源供給動作と各増設筐体の電源供給動作とを連動させるだけのものであり、記憶制御装置の電力消費量を低減させるという観点からは効果が乏しい。

30

【0007】

これに対し、前記第 2 文献に記載の従来技術では、一定期間アクセスされない磁気ディスク装置を節電モードに移行させるため、記憶制御装置の電力消費量を低減させることができる。

【0008】

しかし、第 2 文献に記載の従来技術では、RAIDグループを構成する各磁気ディスク装置の電力消費量を節減するに留まり、さらなる改善の余地がある。即ち、各磁気ディスク装置が節電モードに移行した場合でも、筐体内の制御回路部や冷却ファン、電源回路には電源が供給されており、稼働を続けている。従って、磁気ディスク装置以外の各部における電力消費量について改善の余地がある。

40

【0009】

本発明は、上記の問題点を鑑みてなされたもので、その目的は、筐体単位で電源の供給及び停止を行うことにより、より一層電力消費量を低減できるようにした記憶制御装置を提供することにある。本発明の他の目的は、各筐体がデイジーチェーン接続されている場合に、信号伝達の機能を維持しながら、筐体単位での省電力を実現することができるようにした記憶制御装置を提供することにある。本発明のさらなる目的は、後述する実施形態の記載から明らかになるであろう。

50

【課題を解決するための手段】

【0010】

上記課題を解決すべく、本発明の一つの観点に従う、複数の筐体を備えた記憶制御装置では、各筐体のうちいずれか1つの筐体は、上位装置との間のデータ通信及び他の筐体の動作を制御するための第1筐体となっており、各筐体のうち第1筐体以外の他の筐体は、それぞれ複数の記憶デバイスを備える第2筐体となっている。そして、第1筐体は、記憶制御装置の構成及び電源供給状態を管理するためのシステム構成情報を保持しており、このシステム構成情報に基づいて、第2筐体毎にそれぞれ個別に電源の供給または停止を制御する。

【0011】

第1筐体は、各記憶デバイスのうち、予め設定された所定時間以上のアクセスが無い記憶デバイスについては、電力消費を抑制させる節電モードに移行させてもよい。

【0012】

第1筐体は、各第2筐体のうち、筐体内の全ての記憶デバイスが節電モードに移行している所定の第2筐体が検出された場合には、当該所定の第2筐体内の全ての記憶デバイスへの電源供給を停止させてもよい。

【0013】

第1筐体は、各第2筐体のうち、筐体内の全ての記憶デバイスが節電モードに移行している所定の第2筐体が検出された場合には、当該所定の第2筐体への電源供給を停止させてもよい。

【0014】

第1筐体は、所定の第2筐体内の全ての記憶デバイスへの電源供給を停止させた後で、所定の第2筐体への電源供給を停止させてもよい。

【0015】

本発明の実施形態では、第1筐体と各第2筐体とは、デジチェーン接続されており、第1筐体は、所定の第2筐体内の全ての記憶デバイスへの電源供給を停止させた後で、所定の第2筐体よりも後段に位置する他の第2筐体への電源供給が停止されていると判定された場合、または、所定の第2筐体が接続上の末端に位置すると判定された場合のいずれかの場合に、第2筐体への電源供給を停止させる。

【0016】

第1筐体は、上位装置による利用の頻度が予め設定された所定値以下のデータを、第1筐体から接続上の距離が遠い順番で、各第2筐体内の記憶デバイスに記憶させることもできる。

【0017】

各第2筐体のうち、第1筐体との接続上の距離が遠い第2筐体の各記憶デバイスには相対的に低性能の記憶デバイスをそれぞれ用い、第1筐体との接続上の距離が近い第2筐体の各記憶デバイスには相対的に高性能の記憶デバイスをそれぞれ用いてもよい。

【0018】

本発明の実施形態では、第1筐体と各第2筐体とは、デジチェーン接続されており、第1筐体は、所定の第2筐体内の全ての記憶デバイスへの電源供給を停止させた後で、所定の第2筐体に信号の伝達が求められている場合には、所定の第2筐体が信号を伝達するために必要な機能以外の他の機能への電源供給を停止させる。

【0019】

第1筐体は、電源供給がそれぞれ停止されている複数の第2筐体に電源を供給する場合には、これら複数の第2筐体に時間差をつけて電源を供給させる。

【0020】

第1筐体は、電源供給が停止されている第2筐体に、定期的にまたは不定期に電源を供給することにより、正常に動作するか否かを診断してもよい。

【0021】

本発明の他の観点に従う、複数の筐体を備えた記憶制御装置は、第1筐体と、それぞれ

10

20

30

40

50

複数の記憶デバイスを有する第2筐体と、第1筐体及び各第2筐体にそれぞれ外部電源を供給するための電源供給回路部とを備える。第1筐体は、記憶制御装置の構成及び電源供給状態に関するシステム構成情報を管理するシステム構成情報管理部と、システム構成情報管理部により管理されるシステム構成情報に基づいて、各第2筐体への電源供給及び上位装置と各記憶デバイスとの間のデータ入出力を処理する第1制御部と、電源供給回路部を介して供給される外部電源を所定の電圧に変換することにより、第1制御部に所定の電源を供給する第1筐体内電源部と、を少なくとも備える。各第2筐体は、上位装置により利用されるデータを記憶する複数の記憶デバイスと、第1制御部からの指示に応じて、各記憶デバイスへのデータ入出力及び第2筐体電源部の動作を制御する第2制御部と、電源供給回路部を介して供給される外部電源を所定の電圧に変換することにより、各記憶デバイス及び第2制御部に電源をそれぞれ供給する第2筐体内電源部と、を少なくとも備える。電源供給回路部は、少なくとも各第2筐体内電源部と外部電源との間の電力供給経路の途中にそれぞれ設けられるスイッチ部を備え、該各スイッチ部は、第1制御部からの指示に応じて各電力供給経路をそれぞれ連通または遮断させる。第1制御部は、システム構成情報に基づいて、各第2筐体のうち当該第2筐体内に存在する各記憶デバイスへのアクセスが不要であると判定された所定の第2筐体については、この所定の第2筐体に接続されているスイッチ部を動作させることにより、所定の第2筐体の第2筐体内電源部と外部電源との間の電力供給経路を遮断させる。

10

【0022】

本発明の実施形態では、第1筐体と各第2筐体とはデジチェーン接続されており、第1制御部は、(1)各記憶デバイスのうち、予め設定された所定時間以上のアクセスが無い記憶デバイスについては、電力消費を抑制させる節電モードに移行させる第1モードと、(2)各第2筐体のうち、筐体内の全ての記憶デバイスが節電モードに移行している所定の第2筐体が発出された場合には、当該所定の第2筐体内の全ての記憶デバイスへの電源供給を停止させる第2モードと、(3)所定の第2筐体内の全ての記憶デバイスへの電源供給を停止させた後で、所定の第2筐体への電源供給を停止させる第3モードとを、それぞれ実行することができる。

20

【発明を実施するための最良の形態】**【0023】**

以下、図面に基づき、本発明の実施の形態を説明する。図1は、本実施形態に係る記憶制御装置の全体概要を示す説明図である。

30

【0024】

この記憶制御装置は、例えば、複数種類の筐体1, 2を重ねて配置することにより、構成されている。1つの種類の筐体は、基本筐体1であり、他の1つの種類の筐体は、増設筐体2である。基本筐体1は第1筐体に該当し、増設筐体2は第2筐体に該当する。基本筐体1と各増設筐体2とは、デジチェーン接続されている。即ち、基本筐体1と各増設筐体2との間の信号やデータの伝達は、隣接する筐体を順番に介して行われる。

【0025】

基本筐体1は、記憶制御装置全体の制御を司る筐体であり、上位装置としてのホスト4及び各増設筐体2と接続されている。基本筐体1は、ホスト4から発行されたコマンドに応じて、増設筐体2内のディスクドライブ2Bにアクセスし、データの読み書きを行う。

40

【0026】

基本筐体1は、例えば、I/O処理部1Aと、電源制御部1B及びシステム構成情報管理部1Cとを備えて構成される。なお、後述の実施例からも明らかなように、基本筐体1は、ディスクドライブを備えることもできる。

【0027】

I/O処理部1Aは、ホスト4から発行されたコマンドに基づいて、各増設筐体2内のディスクドライブ2Bにアクセスすることにより、データを書き込んだり、または、データを読み出したりする。

【0028】

50

電源制御部 1 B は、基本筐体 1 が有する制御機能の一つであり、各増設筐体 2 に関する電源の供給及び停止を制御する。電源制御部 1 B は、複数種類の省電力モード F 1 ~ F 4 をそれぞれのタイミングで実行させることができる。

【 0 0 2 9 】

ドライブ単体の省電力を実現させるモード F 1 では、例えば、一定時間アクセスされないディスクドライブ 2 B のモータへの通電を停止させたり、電子回路の一部を停止させることにより、ドライブ毎の省電力を実現する。

【 0 0 3 0 】

全ドライブへの通電を停止させるモード F 2 では、同一筐体に設けられている全てのディスクドライブ 2 B への通電を停止させることにより、筐体内のディスクドライブ全体について省電力を実現する。

10

【 0 0 3 1 】

制御機能以外の通電停止モード F 3 では、例えば、増設筐体 2 に設けられている制御基板 2 A の実現する制御機能以外の他の機能への通電を停止させることにより、さらなる省電力を実現する。このモード F 3 では、増設筐体 2 の制御機能のみが維持され、他の機能は停止する。制御機能としては、例えば、基本筐体 1 からの指示を受領し、他の増設筐体 2 に伝達する機能を挙げることができる。モード F 3 では、これらの信号伝達機能を制御基板 2 A が部分的に作動させることが可能な場合、制御基板 2 A の回路のうち信号伝達機能を実現するために必要な回路以外の他の回路への通電を停止させることができる。また、モード F 3 では、例えば、制御基板 2 A の備える諸機能のうち、冷却ファン 2 D の作動を制御する機能は、停止させることも可能である。しかし、例えば、冷却ファン 2 D による冷却機能の実現が、制御基板 2 A の正常な動作に好ましい場合、モード F 3 では、冷却ファン 2 D への通電を継続させることもできる。即ち、モード F 3 では、信号伝達機能及び該機能を支援する他の機能にのみ電源を供給し、それ以外の各機能への電源供給を停止させる。

20

【 0 0 3 2 】

筐体全体への通電を停止させるモード F 4 では、電源供給回路部 3 のスイッチ 3 A を開くことにより、外部電源 5 から増設筐体 2 への電源供給を停止させる。このモード F 4 では、増設筐体 2 の各機能は停止され、最も省電力効果が高くなる。

【 0 0 3 3 】

電源制御部 1 B は、自己診断モード F 5 を備えている。自己診断モード F 5 では、筐体全体への通電が停止された増設筐体 2 に対して、定期的にまたは不定期に通電することにより、その増設筐体 2 が正常に動作するか否かを診断する。

30

【 0 0 3 4 】

電源制御部 1 B は、データ再配置部 F 6 を備えている。データ再配置部 F 6 は、ホスト 4 からの利用状況に基づいて、各増設筐体 2 内の各ディスクドライブ 2 B に記憶されているデータの記憶先を制御する。即ち、データ再配置部 F 6 は、利用頻度の高いデータほど基本筐体 1 に近い増設筐体 2 内に位置し、利用頻度の高いデータほど基本筐体 1 に遠い増設筐体 2 に位置するように、データマイグレーションを実行する。ここで、基本筐体 1 に近い、または、遠いとは、デイジーチェーン接続された構成において、接続上の物理的な距離が近いか遠いかを意味する。

40

【 0 0 3 5 】

システム構成情報管理部 1 C は、記憶制御装置のシステム構成に関する情報を記憶し、管理している。システム構成に関する情報としては、例えば、ディスクドライブ 2 B の配置に関する情報、RAID グループや論理ボリュームの構成に関する情報、増設筐体 2 と電源供給回路部 3 との接続構成に関する情報等を挙げることができる。本実施形態では、記憶制御装置に関する全ての構成情報を基本筐体 1 が一元的に管理する。電源制御部 1 B は、システム構成情報を参照することにより、目的の増設筐体 2 への通電停止等を行う。

【 0 0 3 6 】

電源供給回路部 3 は、例えば、外部電源 5 からの交流電源を各筐体 1 , 2 にそれぞれ分

50

配するための回路である。ここでは、基本筐体 1 への電源分配については省略しているが、後述の実施例から明らかなように、電源供給回路部 3 を介して、外部電源 5 と基本筐体 1 とは接続されている。

【 0 0 3 7 】

電源供給回路部 3 は、各増設筐体 2 と外部電源 5 とをそれぞれ接続するスイッチ 3 A を備えている。従って、目的の増設筐体 2 への電源供給を停止させる場合は、その増設筐体 2 に接続されたスイッチ 3 A を開く。これにより、外部電源 5 から目的の増設筐体 2 への電力供給経路が遮断される。

【 0 0 3 8 】

電源供給回路部 3 は、各筐体 1, 2 と独立した回路基板として構成可能である。これに限らず、電源供給回路 3 を各筐体 1, 2 内に設けることもできる。即ち、例えば、スイッチ 3 A を各増設筐体 2 内に設ける構成でもよい。

【 0 0 3 9 】

各増設筐体 2 は、例えば、制御基板 2 A と、複数のディスクドライブ 2 B と、電源部 2 C と、冷却ファン 2 D とを備えて構成される。制御基板 2 A は、基本筐体 1 からの指示に基づいて、ディスクドライブ 2 B へのデータ入出力を行う。また、制御基板 2 A は、基本筐体 1 からの指示を隣接する他の増設筐体 2 の制御基板 2 A に伝達する。

【 0 0 4 0 】

ディスクドライブ 2 B は、記憶デバイスに該当する。ここでは、便宜上、ディスクドライブと呼ぶが、ディスク以外の媒体を用いる記憶デバイスでもよい。ディスクドライブ 2 B としては、例えば、ハードディスクドライブ、半導体メモリデバイス、フラッシュメモリデバイス、光ディスクドライブ、光磁気ディスクドライブ、磁気テープデバイス等を用いることができる。

【 0 0 4 1 】

電源部 2 C は、筐体内電源部に該当する。筐体内電源部 2 C は、例えば、外部電源 5 から電源供給回路部 3 を介して入力された交流電源を、所定の電圧を有する直流電源に変換し、制御基板 2 A, 各ディスクドライブ 2 B 及び冷却ファン 2 D にそれぞれ供給するものである。電源部 2 C は、複数種類の電圧を出力可能である。また、外部電源 5 は、交流電源に限らず、直流電源であってもよい。後述の実施例から明らかとなるように、電源部 2 C は、制御基板 2 A からの指示に応じて、各ディスクドライブ 2 B への通電を停止させることができるように構成することができる。

【 0 0 4 2 】

冷却ファン 2 D は、増設筐体 2 内に冷却風を導いて冷却させるものである。冷却ファン 2 D は、制御基板 2 A からの信号によって動作する。即ち、制御基板 2 A は、例えば、冷却ファン 2 D の動作開始や動作停止、回転数を制御することができる。なお、空冷機構として冷却ファン 2 D を挙げたが、増設筐体 2 は水冷機構を備える構成でもよい。水冷機構の場合は、例えば、冷却水循環ポンプや熱交換用ファンの動作を制御することにより、増設筐体 2 の電力消費量を節減する。

【 0 0 4 3 】

図 1 の右側に示すように、基本筐体 1 と接続上の距離が短い増設筐体 2 (# 1) には、ホスト 4 からの利用頻度が比較的高いデータが配置される。この増設筐体 2 (# 1) は、高利用頻度層を形成する。この高利用頻度層には、高信頼性ドライブが用いられる。高信頼性ドライブとは、例えば、FC (Fibre Channel) ディスク等のような高信頼性のディスクドライブである。高利用頻度層の増設筐体 2 (# 1) の各ディスクドライブ 2 B は、利用頻度が高いため、モード F 1 または F 2 が比較的多く使用されるであろう。

【 0 0 4 4 】

基本筐体 1 と接続上の距離が遠い増設筐体 2 (# 3, # 4) には、ホスト 4 からの利用頻度が比較的小さいデータが配置される。この増設筐体 2 (# 3, # 4) は、低利用頻度層を形成する。利用頻度の小さいデータとしては、例えば、アーカイブデータやバックアップデータ等を挙げることができる。この低利用頻度層には、低信頼性ドライブが用いら

10

20

30

40

50

れる。低信頼性ドライブとは、例えば、SATA (Serial AT Attachment) ディスク等のような低性能のディスクドライブである。低利用頻度層の増設筐体 2 (# 3 , # 4) の各ディスクドライブ 2 B は、利用頻度が少ないため、モード F 3 , F 4 が比較的多く使用されるであろう。

【 0 0 4 5 】

基本筐体 1 と接続上の距離が中程度に離れている増設筐体 2 (# 2) には、ホスト 4 からの利用頻度が中程度のデータが配置される。この増設筐体 2 (# 2) は、中利用頻度層を形成する。この中利用頻度層には、例えば、高信頼性ドライブまたは低信頼性ドライブのいずれか、または、中程度の信頼性を有するドライブが使用される。中利用頻度層の増設筐体 2 (# 2) は、ホスト 4 からの利用頻度が中程度であるため、モード F 2 , F 3 が比較的多く使用されるであろう。

10

【 0 0 4 6 】

本実施形態では、基本筐体 1 にデジチェーン接続される各増設筐体 2 を、低利用頻度層、中利用頻度層、高利用頻度層の 3 つの階層に分ける場合を例示した。これに限らず、低利用頻度層及び高利用頻度層の 2 つの階層に分ける構成でもよい。または、4 つ以上の階層を設ける構成でもよい。

【 0 0 4 7 】

図 1 に示すように、増設筐体 2 (# 1) では、各部に電源が供給されている。増設筐体 2 (# 2) では、所定時間以上アクセスされていないディスクドライブ 2 B について、モード F 1 が適用されている。所定時間以上アクセスされていないディスクドライブ 2 B では、例えば、スピンドルモータの回転を停止または低下させたり、一部の電子回路を停止させる。

20

【 0 0 4 8 】

増設筐体 2 (# 3) では、モード F 3 が適用されている。即ち、制御基板 2 A 以外の各ディスクドライブ 2 B 及び冷却ファン 2 D への通電がそれぞれ停止されている。増設筐体 2 (# 4) では、モード F 4 が適用されている。このモードでは、スイッチ 3 A (4) が開いており、外部電源 5 と増設筐体 2 (# 4) との間の電力供給経路が遮断されるため、増設筐体 2 (# 4) への電源供給が停止される。

【 0 0 4 9 】

各筐体 1 , 2 はデジチェーン接続されている。従って、各筐体 1 , 2 間での信号伝達を維持するために、モード F 4 は、基本筐体 1 と接続上の距離が遠い増設筐体 2 から順番に適用される。モード F 2 またはモード F 3 のいずれかが適用されている増設筐体 2 は、モード F 4 に無条件で直ちに移行するわけではなく、後段の増設筐体 2 が機能を停止している場合 (モード F 4 が適用されている場合) にのみ、モード F 4 に移行することができる。

30

【 0 0 5 0 】

本実施形態の記憶制御装置では、例えば、モード F 1 モード F 2 モード F 3 モード F 4 のように、順番に遷移する。しかし、常に必ずしもこのように遷移するわけではなく、ホスト 4 からの利用状況に応じて、例えば、モード F 3 からモード F 1 へ、モード F 2 からモード F 1 に遷移する場合もある。

40

【 0 0 5 1 】

基本筐体 1 との接続上の距離が遠い増設筐体 2 (# 3 , # 4) に利用頻度の低いデータを配置させることにより、これら接続上の距離の遠い増設筐体 2 (# 3 , # 4) にモード F 4 が適用される可能性を高めることができる。この結果、増設筐体 2 (# 3 , # 4) への電源供給を停止させて、記憶制御装置全体の電力消費量を従来技術のそれよりも低減させることができる。

【 0 0 5 2 】

また、モード F 4 の適用される増設筐体 2 (# 3 , # 4) では、低信頼性ドライブを使用するため、信頼性の低いディスクドライブ 2 B の寿命を長くすることができ、高信頼性ドライブとの寿命の差を少なくすることもできる。低信頼性ドライブの場合は、累積作動

50

時間によって寿命が設定される場合が多いため、作動時間を短くすることにより、寿命を延ばすことができる。これにより、高信頼性ドライブと低信頼性ドライブとの交換時期をほぼ一致させて、保守作業の作業効率を高めることができる。

【 0 0 5 3 】

本実施例では、電源供給が停止された増設筐体 2 に定期的または不定期に通電し、正常に動作するか否かを診断する。従って、長期間停止されている増設筐体 2 の信頼性を確認しながら、記憶制御装置全体の電力消費量を低減することができる。以下、本実施形態の記憶制御装置を詳細に説明する。

【 実施例 1 】

【 0 0 5 4 】

図 2 は、記憶制御装置の外観構成を示す斜視図である。記憶制御装置は、例えば、基本ラック 1 0 と増設ラック 1 1 とをケーブル等で接続することにより、構成される。基本ラック 1 0 は、基本筐体 1 0 0 及び複数の増設筐体 2 0 0 を備える。基本ラック 1 0 は、記憶制御装置の最小構成であり、基本ラック 1 0 のみで記憶制御装置の機能を実現することができる。増設ラック 1 1 は、複数の増設筐体 2 0 0 を備える。基本ラック 1 0 の記憶容量が足りなくなった場合に、増設ラック 1 1 を基本ラック 1 0 に接続することにより、記憶容量を増加させることができる。

【 0 0 5 5 】

図 3 は、基本筐体 1 0 0 を前面側から見た斜視図である。基本筐体 1 0 0 のケーシング 1 0 1 には、その前面側 1 0 1 F に、複数のバッテリー 1 4 0 及び複数のディスクドライブ 2 3 0 がそれぞれ設けられる。基本筐体 1 0 0 は、ディスクドライブ 2 3 0 を備える構成でもよいし、ディスクドライブ 2 3 0 を備えない構成でもよい。ディスクドライブ 2 3 0 を備えない場合、前面 1 0 1 F の所定箇所は、バッテリー等の他の構成部品が搭載されるか、又は、化粧板等で施蓋される。

【 0 0 5 6 】

図 4 は、基本筐体 1 0 0 を後面側から見た斜視図である。ケーシング 1 0 1 の後面側 1 0 1 R には、例えば、複数のコントローラ 1 1 0 と、複数の AC/DC 電源部 1 2 0 と、複数の冷却ファン 1 3 0 とがそれぞれ取り付けられる。冷却ファン 1 3 0 は、複数のファンが一体化されたファンアッセンブリとして構成されており、ケーシング 1 0 1 の左右両側にそれぞれ配置される。コントローラ 1 1 0 は、ケーシング 1 0 1 の中央に上下に並んで配置される。AC/DC 電源部 1 2 0 は、ケーシング 1 0 1 の下側に、左右に並んで配置される。なお、以下の増設筐体 2 0 0 の説明でも同様であるが、基本筐体 1 0 0 及び増設筐体 2 0 0 の構成は、図示するものに限定されない。

【 0 0 5 7 】

図 5 は、増設筐体 2 0 0 を前面側から見た斜視図である。増設筐体 2 0 0 のケーシング 2 0 1 には、その前面側 2 0 1 F に、複数のディスクドライブ 2 3 0 を取り付けすることができる。本実施例では、# 0 ~ # 1 4 の合計 1 5 台のディスクドライブ 2 3 0 が取り付けられる場合を例に挙げて説明する。1 つの筐体当たりのディスクドライブ搭載数は、1 5 に限定されない。ディスクドライブ 2 3 0 が取り付けられていない空間は、化粧板等で施蓋することにより、外部の塵埃等がケーシング 2 0 1 内に侵入するのを防止する。

【 0 0 5 8 】

図 6 は、増設筐体 2 0 0 を後面側から見た斜視図である。ケーシング 2 0 1 の後面側 2 0 1 R には、複数の制御基板 2 1 0 及び複数の AC/DC 電源部 2 2 0 がそれぞれ取り付けられている。なお、図 6 の構成においては、冷却ファンは AC/DC 電源部 2 2 0 内に搭載されているが、図 8 に示すように、増設筐体 2 0 0 には、冷却ファン 2 0 2 を設けることができる。即ち、冷却ファンと AC/DC 電源部とを別々の構成部品としてもよい。

【 0 0 5 9 】

図 7 は、記憶制御装置の電源供給構造を模式的に示す説明図である。各ラック 1 0 , 1 1 内には、各筐体 1 0 0 , 2 0 0 へ外部からの電源を分配するための電源分配回路部 3 0 0 が設けられている。以下の説明では、電源分配回路部を PDB (Power Distribution Box

10

20

30

40

50

)と呼ぶ。

【0060】

記憶制御装置は、複数の電源供給システムを備えている。1つのシステムは、PDB#0で示すシステムであり、他の1つのシステムはPDB#1で示すシステムである。即ち、記憶制御装置の電源供給構造は、冗長化されている。

【0061】

各PDB300は、例えば、4個等のような所定数の筐体に外部からの電源を分配することができる。従って、1つのラック内に多数の筐体に取り付けられる場合、複数の電源供給システムのそれぞれにおいて、複数のPDB300が必要となる。なお、これに代えて、ラックに装着可能な最大数の筐体に外部からの電源を分配可能な単一のPDBを、各電源供給系

10

【0062】

各PDB300は、所定数のスイッチ310を備えている。各スイッチ310は、それぞれ対応する筐体のAC/DC電源部に接続される。図中では、AC/DC電源部を「PS」と略記して示している。

【0063】

基本筐体100の有する各電源部120は、それぞれ別システムのPDB300のスイッチ310に接続されている。同様に、各増設筐体200の有する各電源部220も、それぞれ別システムのPDB300のスイッチ310にそれぞれ接続されている。従って、いずれか一方のシステムに障害が発生した場合でも、他方のシステムから基本筐体100及び各増設筐体200

20

【0064】

基本筐体100のコントローラ110は、図中に点線で示すように、信号線P1を介して、各PDB300の各スイッチ310にそれぞれ接続されており、各スイッチ310の開閉を個別に操作することができる。コントローラ110は、目的の増設筐体200への電源供給を停止させる場合、その増設筐体200に接続された各スイッチ310をそれぞれ開くことにより、外部電源との電源部220との間の電源供給経路を遮断させる。

【0065】

なお、各コントローラ110と各PDB300とは、有線接続でもよいし、無線接続でもよい。無線接続の場合は、PDB300内に、コントローラ110からの無線信号に応じて各スイッチ310を開閉させるための回路を設ければよい。

30

【0066】

図8は、記憶制御装置の内部構成を模式的に示す説明図である。基本筐体100の各コントローラ110は、通信ネットワークCN1を介して、ホスト13にそれぞれ接続されている。また、各コントローラ110は、図9に示すように、通信ネットワークCN2を介して管理端末14にもそれぞれ接続されている。

【0067】

ホスト13は、例えば、図外のクライアント端末からの要求に応じて、記憶制御装置にアクセスし、データの読み書きを行う。ホスト13としては、例えば、メインフレームコンピュータやサーバコンピュータ等を挙げることができる。通信ネットワークCN1としては、例えば、LAN(Local Area Network)、SAN(Storage Area Network)、インターネットあるいは専用回線等を挙げることができる。

40

【0068】

LANを用いる場合、ホストコンピュータ13と記憶制御装置とは、TCP/IP(Transmission Control Protocol/Internet Protocol)に従って通信を行う。SANを用いる場合、ホストコンピュータ13と記憶制御装置とは、ファイバチャネルプロトコルに従って通信を行う。また、ホストコンピュータ13がメインフレームコンピュータである場合、例えば、FICON(Fibre Connection:登録商標)、ESCON(Enterprise System Connection:登録商

50

標)、ACONARC(Advanced Connection Architecture:登録商標)、FIBARC(Fibre Connection Architecture:登録商標)等の通信プロトコルに従ってデータ転送が行われる。

【0069】

さらに、図9に示すように、記憶制御装置には、管理用の通信ネットワークCN2を介して、管理端末14を接続することもできる。管理端末14は、記憶制御装置の各種設定を行うためのコンピュータ端末である。管理端末14は、図外の管理サーバに接続することもできる。管理サーバは、複数の記憶制御装置を一括して管理可能なコンピュータ装置として構成することができる。

【0070】

記憶制御装置の制御系統も、電源供給系統と同様に冗長化されている。#0のコントローラ110と#1のコントローラ110とは、それぞれ配下にある全てのディスクドライブ230にアクセス可能である。即ち、いずれか一方の系統に障害が発生した場合でも、他方の系統からデータアクセスが可能となっている。記憶制御装置の全体動作は基本筐体100内のコントローラ110によって制御される。従って、各コントローラ110は、基本筐体100内の各ディスクドライブ230のみならず、増設筐体200内の各ディスクドライブ230にもそれぞれアクセス可能である。

10

【0071】

基本筐体100の各電源部120は、各コントローラ110及び基本筐体100内の各ディスクドライブ230に、それぞれ所定電圧の直流電源を供給する。電源部120は、直流5ボルトや直流12ボルト等のような複数種類の電圧を出力可能である。なお、図中では、説明の便宜上、下段側のディスクドライブ230にのみ電源が供給されるかのように示されているが、各電源部120は、筐体内の全てのディスクドライブ230に電源を供給する。

20

【0072】

図8中の下側には増設筐体200が示されている。紙面の都合上、1つの増設筐体200のみを示しているが、上述のように、基本筐体100は、複数の増設筐体200とデジチェーン接続することができる。

【0073】

増設筐体200の電源部220は、電源部120と同様に、外部から入力された電源を所定電圧の直流電源に変換して出力する。筐体内に存在する全てのディスクドライブ230、制御基板210及び冷却ファン202には、各電源部220から直流電源がそれぞれ供給される。なお、説明の便宜のために、各電源部220から直流電源がそれぞれ供給されると述べたが、例えば、通常時においては、いずれか一方の電源部220(例えば、#0の電源部)のみから各ディスクドライブ230や制御基板210等に直流電源が供給される構成でもよい。ディスクドライブ230や制御基板210等の負荷から各電源部220を見たときのインピーダンスに差をつけることにより、いずれか一方の電源部220から電源を供給させることができる。例えば、いずれかの電源部220からの出力に、他方の電源部220よりも多い段数のダイオードを設ければよい。これにより、障害発生時には、スイッチ切替等を行うことなく、他方の電源部220に自動的に切り替わる。

30

【0074】

ディスクドライブ230は、例えば、ATA(AT Attachment)ディスク、SCSI(Small Computer System Interface)ディスク、FC(Fibre Channel)ディスク等のようなハードディスクドライブとして構成される。これに限らず、例えば、半導体メモリドライブ(フラッシュメモリデバイスを含む)、光ディスクドライブ、光磁気ディスクドライブ等の他の記憶デバイスを用いてもよい。

40

【0075】

例えば、4個1組、8個1組等のような所定数のディスクドライブ230によって、RAIDグループ240が構成される。RAIDグループ240は、別々の筐体に存在する複数のディスクドライブ230から構成することもできるし、同一筐体に存在する1つまたは複数のディスクドライブ230から構成することもできる。

50

【 0 0 7 6 】

図 9 に示すように、RAIDグループ 2 4 0 には、1 つまたは複数の論理ボリューム 2 5 0 を設けることができる。論理ボリューム 2 5 0 は、ホスト 1 3 からアクセスされる対象であり、ホスト 1 3 により使用されるデータを記憶している。

【 0 0 7 7 】

図 9 は、1 つのコントローラ 1 1 0 に着目した場合のブロック図である。コントローラ 1 1 0 は、例えば、ホストインターフェース回路 1 1 1 と、ドライブインターフェース回路 1 1 2 と、データ転送回路 1 1 3 と、キャッシュメモリ 1 1 4 と、プログラムメモリ 1 1 5 と、プロセッサ（図中「MPU」と略記）1 1 6 と、ブリッジ 1 1 7 と、LAN インターフェース 1 1 8 とを備えて構成される。

10

【 0 0 7 8 】

ホストインターフェース回路 1 1 1 は、ホスト 1 3 との間の通信を制御する。ホスト 1 3 から発行された各種コマンドやデータは、ホストインターフェース回路 1 1 1 によって受信される。ディスクドライブ 2 3 0 から読み出されたデータやコマンドの処理完了を告げる通知は、ホストインターフェース回路 1 1 1 からホスト 1 3 に送信される。

【 0 0 7 9 】

ドライブインターフェース回路 1 1 2 は、各ディスクドライブ 2 3 0 との間の通信を制御する。ドライブインターフェース回路 1 1 2 は、論理ブロックアドレス（LBA）とディスクドライブ 2 3 0 の物理的なアドレスとの変換操作等を行う。

【 0 0 8 0 】

データ転送回路 1 1 3 は、コントローラ 1 1 0 内のデータ転送を制御するための回路である。データ転送回路 1 1 3 は、ホストインターフェース回路 1 1 1 とキャッシュメモリ 1 1 4 との間のデータ転送や、ドライブインターフェース回路 1 1 2 とキャッシュメモリ 1 1 4 との間のデータ転送を制御する。

20

【 0 0 8 1 】

プロセッサ 1 1 6 は、一つまたは複数のプロセッサコアを備えている。プロセッサ 1 1 6 は、プログラムメモリ 1 1 5 に記憶された電源管理プログラム等の種々のプログラムを読み込んで実行することにより、後述の電源制御等を実現する。

【 0 0 8 2 】

キャッシュメモリ 1 1 4 は、ホスト 1 3 から受信したデータやホスト 1 3 により読み出されたデータを記憶する。キャッシュメモリ 1 1 4 には、記憶制御装置のシステム構成に関する情報も記憶されている。キャッシュメモリ 1 1 4 の記憶内容は、バッテリー 1 4 0 によってバックアップされる。

30

【 0 0 8 3 】

ブリッジ 1 1 7 は、プロセッサ 1 1 6 とプログラムメモリ 1 1 5 とを接続する。また、プロセッサ 1 1 6 は、ブリッジ 1 1 7 を介してデータ転送回路 1 1 3 に接続される。

【 0 0 8 4 】

LAN インターフェース 1 1 8 は、管理端末 1 4 とコントローラ 1 1 0 とを接続するための回路である。ユーザは、管理端末 1 4 を介して、コントローラ 1 1 0 に各種の指示を与えたり、記憶制御装置の各種状態に関する情報を読み出して画面に表示させることができる。

40

【 0 0 8 5 】

図 1 0 は、増設筐体 2 0 0 のブロック図である。先に、AC/DC電源部 2 2 0 の構成を説明する。電源部 2 2 0 は、例えば、電圧変換回路 2 2 1 と、複数の出力ダイオード 2 2 2 A ~ 2 2 2 C と、出力制御スイッチ 2 2 3 とを備えている。

【 0 0 8 6 】

電圧変換回路 2 2 1 は、PDB 3 0 0 を介して入力された交流電源を複数種類の直流電圧 V 1 , V 2 に変換して出力する。例えば、V 1 は直流 5 ボルト程度、V 2 は直流 1 2 ボルト程度にそれぞれ設定される。出力ダイオード 2 2 2 A ~ 2 2 2 C は、逆流防止のために設けられている。スイッチ 2 2 3 は、直流電圧を出力させるか否かを制御するためのもの

50

で、制御基板 210 によって切り替えられる。

【0087】

各ディスクドライブ 230 には、V1, V2 がそれぞれ供給される。各制御基板 210 及び各冷却ファン 202 には、V2 がそれぞれ供給される。図 10 に示すように、2つの電源部 220 の出力は、それぞれ OR 接続されている。従って、いずれか一方の電源部 220 に障害が生じた場合でも、他方の電源部 220 から制御基板 210 やディスクドライブ 230 等に所定電圧の電源が供給される。各電源部 220 からの出力 V1, V2 が停止した場合、筐体内の全てのディスクドライブ 230 への電源供給が同時に停止する。

【0088】

各制御基板 210 は、信号線 P2 を介して、各電源部 220 内の各スイッチ 223 にそれぞれ接続されている。各制御基板 210 は、各スイッチ 223 をそれぞれ個別に開閉させることができる。また、各制御基板 210 は、信号線 P3 を介して、各冷却ファン 202 にもそれぞれ接続されている。各制御基板 210 は、各冷却ファン 202 の作動を開始または停止、あるいは回転数を調節することができる。

10

【0089】

図 11 は、1つの制御基板 210 に着目した構成説明図である。制御基板 210 は、例えば、マイクロプロセッサ部 211 と、筐体間通信インターフェース部 212 と、制御信号出力回路 213 とを備えて構成される。

【0090】

マイクロプロセッサ部 211 は、制御基板 210 の動作を制御するものであり、例えば、資源管理機能 211A、コマンド処理機能 211B、電源制御機能 211C、冷却ファン制御機能 211D をそれぞれ実現する。

20

【0091】

資源管理機能 211A は、例えば、増設筐体 200 内の各資源（ディスクドライブ 230, 電源部 220 等）の配置や状態等を管理する機能である。コマンド処理機能 211B は、コントローラ 110 から入力された指示を実行するための機能である。即ち、コマンド処理機能 211B は、コントローラ 110 からデータの書込みを要求された場合は、指定されたディスクドライブ 230 の指定されたアドレスにデータを書き込む。また、コマンド処理機能 211B は、コントローラ 110 からデータの読出しを要求された場合は、指定されたディスクドライブ 230 の指定されたアドレスから、指定された長さのデータを

30

【0092】

読み出してコントローラ 110 に転送する。電源制御機能 211C は、コントローラ 110 からの指示に基づいて、スイッチ 223 を開閉させることにより、電源部 220 の出力を制御する機能である。冷却ファン制御機能 211D は、コントローラ 110 からの指示に基づいて、冷却ファン 202 の作動を制御する機能である。

【0093】

筐体間通信インターフェース部 212 は、隣接する他の筐体との間の通信を行うものである。基本筐体 100 に対して各増設筐体 200 はダイジーチェーン接続されており、基本筐体 100 のコントローラ 110 から出力されたコマンドやデータは、隣接する増設筐体 200 を順番に介して伝達されていく。図 11 中に示す上位筐体とは、接続上の上位に位置する筐体である。即ち、基本筐体 100 と接続上の距離が近い方の筐体である。下位筐体とは、接続上の下位に位置する筐体であり、基本筐体との距離が長い方の筐体を意味する。

40

【0094】

図 12 は、論理ボリューム 250 を管理するためのテーブル T1 の構成を示す説明図である。このテーブル T1 は、図 13 に示す電源供給状態管理テーブル T2 と共に、システム構成情報を構成する。これら各テーブル T1, T2 は、例えば、キャッシュメモリ 114 内に記憶される。

【0095】

50

ボリューム管理テーブル T 1 は、例えば、ボリューム番号 (LU#) I 1 1 と、容量 I 1 2 と、RAIDグループ番号 I 1 3 と、RAIDレベル I 1 4 と、ドライブ番号リスト I 1 5 と、最新アクセス時刻 I 1 6 とのような項目を対応付けて管理する。なお、図に示す以外の項目を管理する構成でもよい。また、テーブル T 1 を複数のテーブルに分割して管理する構成でもよい。

【 0 0 9 6 】

論理ボリューム番号 I 1 1 は、記憶制御装置内に設けられている全ての論理ボリューム 2 5 0 を一意に特定するための識別情報である。容量 I 1 2 は、論理ボリューム 2 5 0 の記憶容量を示す情報である。RAIDグループ番号 I 1 3 は、論理ボリューム 2 5 0 が設けられている RAIDグループ 2 4 0 を一意に特定するための識別情報である。RAIDレベル I 1 4 とは、RAIDグループ 2 4 0 の RAIDレベルを示す情報である。RAIDレベルとしては、例えば、RAID 1 , RAID 5 , RAID 6 等を挙げるができる。ドライブ番号リスト I 1 5 とは、RAIDグループ 2 4 0 を構成しているディスクドライブ 2 3 0 を特定するための情報である。最新アクセス時刻 I 1 6 とは、ホスト 1 3 が論理ボリューム 2 5 0 に最後にアクセスした時刻を示す情報である。なお、紙面の都合上、図示していないが、例えば、所定の期間毎のアクセス頻度や、各論理ボリューム 2 5 0 の空き容量等をテーブル T 1 で管理することもできる。

【 0 0 9 7 】

図 1 3 は、各筐体 1 0 0 , 2 0 0 における電源の供給状態を管理するためのテーブル T 2 を示す説明図である。この電源供給状態管理テーブル T 2 は、例えば、筐体番号 1 2 1 と、筐体への電源供給状態 I 2 2 と、ドライブ番号 I 2 3 と、スピン状態 I 2 4 と、第 1 系統 PDB のスイッチ (図中「SW」と略記) 番号 I 2 5 と、そのスイッチ状態 I 2 6 と、第 2 系統 PDB のスイッチ番号 I 2 7 と、そのスイッチ状態 I 2 8 とをそれぞれ対応付けて管理している。

【 0 0 9 8 】

筐体番号 I 2 1 とは、記憶制御装置を構成する各筐体 1 0 0 , 2 0 0 を一意に特定するための識別情報である。筐体への電源供給状態 I 2 2 とは、その筐体に外部からの電源が供給されているか否かを示す情報である。ドライブ番号 I 2 3 とは、その筐体に設けられている各ディスクドライブ 2 3 0 を特定するための識別情報である。

【 0 0 9 9 】

スピン状態 I 2 4 とは、そのディスクドライブ 2 3 0 がスピニングアップしているか、スピンドアウンしているかを示す情報である。スピニングアップ状態の場合は「ON」で示され、スピンドアウン状態の場合は「OFF」で示される。

スピニングアップ状態とは、ディスクが所定速度で回転している状態であり、データの読み書きが可能である。スピンドアウン状態とは、ディスクが所定速度以下で回転または停止している状態であり、この状態ではデータの読み書きを行うことはできない。

【 0 1 0 0 】

第 1 系統 PDB のスイッチ番号 I 2 5 及び第 2 系統 PDB のスイッチ番号 I 2 7 とは、その筐体に接続されている第 1 電源供給系統 (# 0 の系統) 及び第 2 電源供給系統 (# 1 の系統) の PDB 内スイッチ 3 1 0 をそれぞれ特定するための情報である。スイッチ状態 I 2 6 , I 2 8 とは、I 2 5 , I 2 7 で特定されたスイッチ 3 1 0 の開閉状態をそれぞれ示す情報である。スイッチ 3 1 0 が閉じている場合は「ON」で示され、スイッチ 3 1 0 が開いている場合は「OFF」で示される。本実施例では、信頼性を高めるために、スイッチ 3 1 0 は、常時開いている a 接点として構成されている。

【 0 1 0 1 】

図 1 4 は、コントローラ 1 1 0 によって実行されるデータ再配置処理を示すフローチャートである。この処理は、例えば、ホスト 1 3 からのデータアクセスが少ない時間帯等を選んで実行させることができる。

【 0 1 0 2 】

コントローラ 1 1 0 は、データ再配置対象の筐体番号に、最前段の筐体の番号をセット

10

20

30

40

50

する（S1）。最前段の筐体とは、デジチェーン接続上の順番が最も高位の筐体を意味する。従って、本処理の開始時点では、基本筐体100の番号がセットされる。

【0103】

コントローラ110は、テーブルT1を用いることにより、S1で設定された対象筐体内に存在する各ボリューム250の利用状態をそれぞれ検出する（S2）。コントローラ110は、検出された利用頻度が所定値よりも低いボリューム250が存在するか否かを判定する（S3）。所定値は、予めユーザ等によって設定される。所定値は、記憶階層のレベル毎にそれぞれ用意される。図1の概要図で述べたように、記憶制御装置の有する記憶領域を、例えば、高利用頻度層、中利用頻度層、低利用頻度層等のように、階層化して管理することができる。これらの利用頻度に応じた各記憶階層毎に、それぞれ移動判定用の所定値（閾値）が予め用意されている。

10

【0104】

利用頻度が所定値よりも低いボリュームが検出された場合（S3:YES）、コントローラ110は、対象筐体の1つ後ろに位置する筐体内に、空いている容量が存在するか否かを判定する（S4）。コントローラ110は、接続上の下位に位置する筐体内に、S3で検出されたボリューム250を収容するだけの空き容量があるか否かを判定する。

【0105】

隣接する下位筐体に所定サイズの空きが存在する場合（S4:YES）、コントローラ110は、S3で検出されたボリューム250のデータを、S4で検出された記憶領域に移動させる（S5）。例えば、コントローラ110は、移動先の筐体（隣接する下位筐体）の各ディスクドライブ230が有する記憶領域に、所定サイズのボリュームを移行先ボリュームとして新たに形成する。そして、コントローラ110は、対象筐体内の利用頻度の低いボリューム250を移行元ボリュームとして、移行元ボリューム内のデータを移行先ボリュームにコピーさせる。コピー完了後に、移行元ボリュームのデータは消去される。

20

【0106】

コントローラ110は、記憶制御装置内の全ての筐体についてデータ再配置の必要性を判断するまで、S2～S5を繰り返す（S6:NO）。即ち、コントローラ110は、対象筐体の筐体番号を1つ後ろにずらし（S7）、必要があれば、移動判定用の所定値を変更する（S8）。

【0107】

例えば、基本筐体100についてのデータ再配置が判断された直後の場合は、S7では基本筐体100に隣接する増設筐体200（#1）がセットされる。S7でセットされた増設筐体200と直前に判定された対象筐体とが別々の記憶階層に位置する場合は、移動判定用の閾値も変更される。以下、前記同様にして、基本筐体100から下位の各増設筐体200に向けて、S2～S6がそれぞれ順番に繰り返されていく。全ての筐体についてデータ再配置を判定した場合（S6:YES）、本処理は終了する。

30

【0108】

図14に示すような処理を行うことにより、利用頻度の低いボリュームは、より基本筐体100からの接続上の距離の遠い増設筐体200に集められていく。なお、図14では、図示を省略しているが、逆向きのデータ再配置を行うこともできる。即ち、例えば、予め設定された所定値よりも利用頻度の高いボリューム250を、基本筐体100に近い増設筐体200に移動させることもできる。なお、データ再配置は、ユーザが管理端末14から指示を与えることによっても、実行させることができる。

40

【0109】

図15は、データ再配置が行われた場合の様子を模式的に示す説明図である。図15に示すように、基本筐体100から最も遠い増設筐体200（#3）には、利用頻度の最も少ないバックアップデータが保存される。基本筐体100には、最も利用頻度の高いデータが保存される。図15では、ホスト13による利用頻度の高いデータをアクセスデータと表現している。また、基本筐体100に隣接する他の増設筐体200（#2）等には、アーカイブデータが保存される。

50

【0110】

このように、本実施例では、記憶制御装置の物理的な配置と論理的な記憶階層とが一致している。物理的な配置とは、接続上の配置である。図2に示したように、各筐体は、重ねられたような状態でラックに収容されるため、接続上の距離と物理的な位置関係とが対応する。

【0111】

基本筐体100との接続上の距離が近い増設筐体200ほど、ラック内では基本筐体100に近い場所に格納される。これに対し、記憶制御装置の有する記憶領域を利用頻度に応じて階層化する記憶階層は、論理的な存在であり、本来は筐体の物理的な位置関係に依存しない。従って、従来であれば、記憶階層は、筐体100, 200の物理的な配置と無関係に設定されるため、ユーザは、筐体の物理的位置関係のみから記憶階層の構成を知ることができない。これに対し、本実施例では、基本筐体100に各増設筐体200をディジーチェーン接続し、基本筐体100に近い筐体(本実施例のように、基本筐体100がディスクドライブ230を備える場合は、基本筐体100そのものを含む。)から順番に、利用頻度の高いデータを記憶させ、利用頻度の低いデータほど接続上の距離の遠い増設筐体200に記憶させる。従って、本実施例では、記憶階層の構成を筐体の物理的配置によって可視化することができ、ユーザは、記憶制御装置の外観から容易にわかる物理的な配置によって、記憶階層の所在を知ることができ、使い勝手が向上する。また、このように接続上の距離が遠い筐体内に、利用頻度の低いデータを集めることにより、後述するように、筐体単位の電源供給停止を効果的に行うことができる。

また、例えば、ホスト13が、利用頻度の判明しているデータを、ディスク装置の適切な物理的位置(記憶階層)に予め配置しておくことにより、より一層、節電効果を高めることもできる。

【0112】

図16は、電源供給に関する接続構成の初期設定を確認するための処理を示すフローチャートである。どの筐体にどのPDB300のスイッチ310を接続するかは、ユーザによって決定することができる。ユーザは、自らが行った接続に関する構成を、管理端末14からコントローラ110のテーブルT2に登録することができる。しかし、ヒューマンエラーを完全に排除するのは難しい。そこで、図16に示す処理を実行することにより、ユーザの登録した電源供給の接続構成が正しいか否かを検査する。

【0113】

本処理は、第1電源供給システム及び第2電源供給システムのそれぞれについて、別々に実施される。コントローラ110は、判断対象のスイッチ310の番号として、最終段のスイッチ310の番号を設定する(S10)。接続上の末端に位置する筐体200に接続されている第1電源システムのスイッチ310の番号である。

【0114】

コントローラ110は、S10で設定されたスイッチ310に制御信号を出力し、このスイッチ310をオフさせる(S11)。コントローラ110は、S10で設定されたスイッチ310に接続されているはずの電源部220から停電を示す警報信号が出力されたか否かを確認する(S12)。各電源部120, 220は、外部からの電源供給が途絶えた場合に、警報信号を外部に出力できるように構成されている。

【0115】

S10で設定されたスイッチ310に接続されているはずの電源部220から警報信号が出力された場合(S12:YES)、コントローラ110は、ユーザにより登録された情報が正しいものと判断する(S13)。S10で設定されたスイッチ310に接続されているはずの電源部220から警報信号が出力されなかった場合(S12:NO)、コントローラ110は、ユーザにより登録された情報が誤っていると判断する(S14)。ユーザから登録された情報に誤りが検出された場合、PDB300によって電源供給を停止させないようにする(S15)。

【0116】

コントローラ 110 は、全てのスイッチ 310 について接続状態の判定を行ったか否かを判断する (S16)。未判定のスイッチ 310 が存在する場合 (S15: YES)、コントローラ 110 は、対象スイッチ番号を 1 つ前のスイッチ 310 の番号に変更し (S17)、S11 ~ S15 を繰り返す。全てのスイッチ 310 について検査を完了した場合 (S15: YES)、コントローラ 110 は本処理を終了する。なお、S15 でも示した通り、接続状態に誤りが検出された場合、少なくとも接続の誤っている箇所に関するスイッチ操作は行わない。電源供給に関する接続に誤りが検出された場合、その誤りが正されるまでの間、PDB300 による電源供給の停止を禁止することができる。

【0117】

図 17 は、1 つのモードによる電源供給停止処理を示すフローチャートである。この処理では、各ディスクドライブ 230 毎に省電力を行う。本処理は、図 1 中で述べたモード F1 の一例である。

10

【0118】

コントローラ 110 は、テーブル T1 を参照することにより、所定時間以上アクセスされていないディスクドライブ 230 が存在するか否かを判定する (S20)。ユーザは、例えば、1 時間等のように、任意の値を所定時間として設定できる。

【0119】

コントローラ 110 は、所定時間以上アクセスされていないディスクドライブ 230 を検出した場合 (S20: YES)、そのディスクドライブ 230 をスピンドアウン状態に変更させるための指示を、そのディスクドライブ 230 の存在する増設筐体 200 に向けて発行する (S21)。コントローラ 110 は、S20 で検出されたディスクドライブ 230 のスピンドアウン状態をスピンドアウンとしてテーブル T2 に記録する。コントローラ 110 から発行された指示 (コマンド) は、隣接する増設筐体 200 を順番に辿りながら、目的の増設筐体 200 まで伝達される。

20

【0120】

コントローラ 110 からの指示を受領した増設筐体 200 の制御基板 210 は、指定されたディスクドライブ 230 をスピンドアウンさせる (S23)。また、制御基板 210 は、それが可能であれば、冷却ファン 202 の回転数を低下させる (S24)。

【0121】

例えば、筐体内で他のディスクドライブ 230 が稼働している場合は、筐体内の温度を保つために、冷却ファン 202 の回転数は一定値に保持される。これに対し、S23 で、1 つまたは複数のディスクドライブ 230 をスピンドアウンさせたことにより、冷却能力を低下させることが可能な場合、制御基板 210 は、冷却ファン 202 の回転数を低下させることもできる。なお、冷却ファン 202 の回転数を低下させるか否かは、コントローラ 110 によって決定してもよいし、あるいは、増設筐体 200 の制御基板 210 によって決定してもよい。

30

【0122】

図 18 は、他のモードによる電源供給停止処理を示すフローチャートである。本処理では、筐体内の全ディスクドライブ 230 への通電停止 (図 1 中のモード F2) と、制御機能以外の通電停止 (図 1 中のモード F3) と、筐体全体への電源供給停止 (図 1 中のモード F4) とがそれぞれ示されている。

40

【0123】

コントローラ 110 は、電源供給を停止させるか否かの判断を行う対象筐体の番号に、接続上の最後尾に位置する増設筐体 200 の番号をセットする (S30)。コントローラ 110 は、テーブル T2 を用いることにより、対象筐体内の全ディスクドライブ 230 が、スピンドアウン状態になっているか否かを判定する (S31)。図 17 で述べた通り、各ディスクドライブ 230 へのアクセスが所定時間以上無い場合は、各ディスクドライブ 230 は個別にスピンドアウン状態に置かれる。スピンドアウン状態に置かれたディスクドライブ 230 についてデータアクセスが発生した場合、そのディスクドライブ 230 は、スピンドアウン状態に変更される。

50

【 0 1 2 4 】

対象筐体内の全ディスクドライブ 2 3 0 がスピンドウン状態であると判定された場合 (S31: YES)、コントローラ 1 1 0 は、対象筐体の制御基板 2 1 0 に向けて、全ディスクドライブ 2 3 0 への通電停止を指示する (S 3 2)。対象筐体の制御基板 2 1 0 は、電源部 2 2 0 内のスイッチ 2 2 3 を開くことにより、出力 V 1 , V 2 をそれぞれ停止させ、筐体内の全ディスクドライブ 2 3 0 への通電を停止させる (S 3 3)。なお、この場合、コントローラ 1 1 0 は、全ディスクドライブ 2 3 0 への通電停止のみならず、信号伝達機能以外の全ての機能の停止を要求することもできる。例えば、コントローラ 1 1 0 は、冷却ファン 2 0 2 の停止を要求することができる。

【 0 1 2 5 】

対象筐体内の各ディスクドライブ 2 3 0 のうちいずれかが 1 つ以上のディスクドライブ 2 3 0 がスピンドアアップ状態の場合 (S31:NO)、S 3 4 , S 3 5 をスキップして、S 3 6 に移る。

【 0 1 2 6 】

コントローラ 1 1 0 は、対象筐体内の全ディスクドライブ 2 3 0 への通電停止を指示した後で、対象筐体よりも後側に位置する各下位筐体が全て電源供給停止状態であるか否かを判定する (S 3 4)。対象筐体よりも下位の各筐体に各系統の PDB 3 0 0 から電源が供給されていない場合 (S34: YES)、コントローラ 1 1 0 は、対象筐体に接続された各系統の PDB 3 0 0 のスイッチ 3 1 0 をそれぞれオフさせ、外部からの電源供給を停止させる (S 3 5)。これにより、対象筐体は機能を停止する。この際、コントローラ 1 1 0 は、まず最初に、第 1 系統 (PDB#0) の PDB 3 0 0 のスイッチ 3 1 0 をオフにし、電源供給の停止を知らせる警報信号が第 1 系統側の電源部 2 2 0 から出力されるのを確認する。第 1 系統側での電源供給が正常に停止したことを確認してから、コントローラ 1 1 0 は、第 2 系統 (PDB#1) の PDB 3 0 0 のスイッチ 3 1 0 をオフにする。

【 0 1 2 7 】

コントローラ 1 1 0 は、記憶制御装置内の全筐体について判断したか否かを判定し (S 3 6)、未判定の筐体が存在する場合 (S36:NO)、対象筐体の番号を、対象筐体に隣接する 1 つ前の筐体の番号に変更する (S 3 7)。即ち、対象筐体に隣接する上位の筐体の番号がセットされる。そして、S 3 1 ~ S 3 6 が繰り返される。

【 0 1 2 8 】

このように、本実施例では、基本筐体 1 0 0 からの距離が遠い増設筐体 2 0 0 から順番に、筐体そのものへの通電を停止させるか否かを判定していく。データ再配置処理で述べたように、本実施例では、接続上の距離が遠い増設筐体 2 0 0 に利用頻度の低いデータを配置させる。従って、接続上の距離が遠い下位の増設筐体 2 0 0 ほど、データアクセスされる可能性が低下するため、筐体そのものへの電源供給を停止できる機会が増大する。従って、記憶制御装置全体としての電力消費量を従来よりも低減することができる。

【 0 1 2 9 】

図 1 9 は、ホスト 1 3 から発行されたライトコマンドを処理するためのフローチャートを示す。

【 0 1 3 0 】

コントローラ 1 1 0 は、ホスト 1 3 からライトコマンド及びライトデータを受領すると (S 4 0)、ライトデータをキャッシュメモリ 1 1 4 に記憶させる (S 4 1)。コントローラ 1 1 0 は、ライトデータをキャッシュメモリ 1 1 4 に記憶させた時点で、ホスト 1 3 にライトコマンドの処理が完了した旨を通知する (S 4 2)。コントローラ 1 1 0 は、ライトコマンドに明示された論理アドレスをディスクドライブ 2 3 0 の物理的なアドレスに変換し (S 4 3)、アクセス対象のディスクドライブ 2 3 0 を特定する (S 4 4)。以下の説明では、便宜上、アクセス対象のディスクドライブ 2 3 0 をアクセス先ドライブと表現する場合がある。

【 0 1 3 1 】

コントローラ 1 1 0 は、テーブル T 2 を用いることにより、S 4 4 で特定されたアクセ

10

20

30

40

50

ス先ドライブにアクセス可能か否かを判定する（S45）。アクセス先ドライブがスピニング状態であり、かつ、アクセス先ドライブを有する増設筐体200までの通信経路が正常な場合は、直ちにアクセス先ドライブにアクセス可能である。アクセス先ドライブにアクセス可能であると判定された場合（S45:YES）、コントローラ110は、アクセス先ドライブにライトデータを転送する（S46）。ライトデータは、基本筐体100から隣接する増設筐体200を順番に介して、アクセス先ドライブを有する増設筐体200に伝達される。制御基板210は、ライトデータを受領すると、指定されたディスクドライブ230（アクセス先ドライブである）にライトデータを書き込む。

【0132】

これに対し、アクセス先ドライブへのアクセスが不能であると判定された場合（S45:NO）、後述する電源供給開始処理が実行される（S100）。このように、本実施例では、ライトデータをアクセス先ドライブに転送する前に、アクセス先ドライブにアクセス可能であるか否かを判断し、アクセス不能な場合は電源供給開始処理を実行する。これに代えて、後述の実施例のように、アクセス先ドライブへのライトデータの書込みができなかった場合に、電源供給開始処理を実行させることもできる。

【0133】

図20は、ホスト13から発行されたリードコマンドを処理するためのフローチャートを示す。コントローラ110は、ホスト13からリードコマンドを受領すると（S50）、アクセス先ドライブを特定する（S51）。ここでのアクセス先ドライブとは、読出しを要求されたデータを記憶しているディスクドライブ230である。

【0134】

コントローラ110は、ホスト13から要求されたデータがキャッシュメモリ114に記憶されているか否かを判定する（S52）。以下の説明では、ホスト13により読出しが要求されているデータをリードデータと呼ぶ場合がある。

【0135】

リードデータがキャッシュメモリ114に記憶されている場合（S52:YES）、後述のS53、S54をスキップして、キャッシュメモリ114に記憶されているリードデータをホスト13に送信する（S55）。

【0136】

リードデータがキャッシュメモリ114に記憶されていない場合（S52:NO）、コントローラ110は、テーブルT2を用いることにより、アクセス先ドライブにアクセス可能か否かを判断する（S53）。アクセス先ドライブにアクセス可能であると判定された場合（S53:YES）、コントローラ110は、アクセス先ドライブからリードデータを読み出してキャッシュメモリ114に記憶させる（S54）。コントローラ110は、キャッシュメモリ114に記憶されたリードデータをホスト13に送信する（S55）。

【0137】

これに対し、アクセス先ドライブにアクセス不能であると判定された場合（S53:NO）、コントローラ110は、電源供給開始処理を実行する（S100）。

【0138】

図21は、電源供給開始処理（S100）の詳細を示すフローチャートである。コントローラ110は、対象筐体の番号として、アクセス先ドライブを有する増設筐体200の番号をセットする（S101）。

【0139】

コントローラ110は、テーブルT2を用いることにより、基本筐体100からS101でセットされた対象筐体までの伝達経路上に存在する各増設筐体200の電源供給状態をそれぞれ取得する（S102）。

【0140】

コントローラ110は、伝達経路上の各増設筐体200にPDB300から外部電源がそれぞれ供給されているか否かを判定する（S103）。経路上に位置する増設筐体200に外部電源が供給されていないと判定された場合（S103:NO）、コントローラ110は、

10

20

30

40

50

電源供給が停止されている各増設筐体 200 に接続されている PDB 300 内のスイッチ 310 を順番にオンさせていく (S104)。即ち、コントローラ 110 は、伝達経路上に位置する各増設筐体 200 に対して、その上位側から順番に外部電源を供給し、起動させていく。但し、S104 では、外部電源を経路途中の各増設筐体 200 に供給させるだけであり、これら経路途中の各増設筐体 200 内のディスクドライブ 230 をスピニングさせない。経路上の各制御基板 210 を起動させて、コマンドやデータの伝達機能のみを復活させれば足りるためである。

【0141】

経路上の各増設筐体 200 に PDB 300 から外部電源が供給されている場合 (S103: YES)、S104 はスキップされる。コントローラ 110 は、対象筐体に対して、対象筐体内の全ディスクドライブ 230 に電源部 220 から直流電源 V1, V2 を供給するように指示する (S105)。対象筐体の制御基板 210 は、S105 で発行された指示を受領すると、電源部 220 内のスイッチ 223 をオンさせることにより、各ディスクドライブ 230 に電源をそれぞれ供給させる (S106)。

10

【0142】

コントローラ 110 は、アクセス先ドライブのスピニングを対象筐体の制御基板 210 に指示する (S107)。この指示を受領した制御基板 210 は、アクセス先ドライブのみをスピニングさせる (S108)。

【0143】

図 22 は、自己診断処理を示すフローチャートである。アクセス頻度の低いデータを記憶している増設筐体 200 への通電を長時間停止した場合、通電停止中に増設筐体 200 に障害が発生しても検出することができず、必要な場合にデータを取り出せない可能性がある。そこで、本実施例では、定期的にまたは不定期に、休止中の増設筐体 200 に通電することにより、正常に動作するか否かを確認する。この自己診断処理は、予め設定された所定周期で定期的に行うこともできるし、ユーザが任意に指定した時刻で不定期に行うこともできる。

20

【0144】

コントローラ 110 は、テーブル T2 を用いて、所定時間以上、PDB 300 から外部電源が供給されていない増設筐体 200 が存在するか否かを判定する (S60)。コントローラ 110 は、所定時間以上通電されていない増設筐体 200 を発見した場合 (S60: YES)、その増設筐体 200 に接続されている PDB 300 内のスイッチ 310 をオンさせて (S61)、その増設筐体 200 に外部電源を供給する。

30

【0145】

コントローラ 110 は、診断用コマンドとして予め設定されているコマンド (スピニングコマンドやスピンドアウンコマンド等) を増設筐体 200 に送信し、その増設筐体 200 内の各ディスクドライブ 230 の動作を確認する (S62)。コントローラ 110 は、その増設筐体 200 内の全ディスクドライブ 230 が正常に動作するか否かを判定する (S63)。コントローラ 110 は、全ディスクドライブ 230 が正常に動作すると判定された場合 (S63: YES)、S60 に戻る。いずれか 1 つ以上のディスクドライブ 230 が正常に動作しない場合 (S63: NO)、コントローラ 110 は、その増設筐体 200 に異常が生じている旨の警報を出力する (S64)。この警報は、管理端末 14 を介して、ユーザに通知される。また、管理端末 14 が管理サーバに接続されている場合、この警報は、管理端末 14 から管理サーバに送信される。

40

【0146】

なお、診断対象の増設筐体 200 に診断用のコマンドを送信するために、基本筐体 100 と診断対象の増設筐体 200 との間の伝達経路が有効となるように、経路上の各増設筐体 200 の制御基板 210 が起動される。

【0147】

図 23 は、各増設筐体 200 における節電モードの変化を模式的に示す。図 23 (a) に示すように、最初は、各ディスクドライブ 230, 制御基板 210, 冷却ファン 202

50

のそれぞれに電源部 220 から直流電源が供給されている。図 23 (b) に示すように、所定時間アクセスされなかったディスクドライブ 230 (#1) は、スピンドウン状態に変更される。

【0148】

図 23 (c) に示すように、スピンドウン状態となるディスクドライブ 230 が増加し、図 23 (d) に示すように、全てのディスクドライブ 230 がスピンドウン状態に変化すると、図 23 (e) に示すように、PDB 300 から増設筐体 200 への外部電源の供給そのものが停止する。このように、本実施例では、ディスクドライブ 230 の利用状況に応じて、電力消費量を段階的に節減させていくことができ、最終的には、ほぼ電力消費量を零とすることができる。

10

【0149】

図 24 ~ 図 27 は、増設筐体 200 に対する外部電源の供給停止の順序を示す説明図である。図 24 に示すように、各増設筐体 200 内のいずれか 1 つ以上のディスクドライブ 230 についてそれぞれデータアクセスが発生している場合、各 PDB 300 のスイッチ 310 はそれぞれオン状態となっている。従って、各増設筐体 200 には、それぞれ外部電源が供給される。

【0150】

図 25 に示すように、最も利用頻度の低いデータを記憶する増設筐体 200 (#3) は、内部の各ディスクドライブ 230 へのデータアクセスが発生しにくいいため、比較的早い時期に、筐体への外部電源の供給が停止されるであろう。

20

【0151】

図 26 に示すように、次に利用頻度の低いデータを記憶する増設筐体 200 (#2) も、増設筐体 200 (#3) に続いて、データアクセスが途絶え、全ディスクドライブ 230 がスピンドウン状態にされる。この時点で、下位の増設筐体 200 (#3) は既に電源供給が停止されている。従って、増設筐体 200 (#2) への通電も停止させることができる。

【0152】

図 27 に示すように、増設筐体 200 (#1) 内の各ディスクドライブ 230 に対するデータアクセスが所定時間以上途絶えると、増設筐体 200 (#1) 内の各ディスクドライブ 230 はそれぞれスピンドウン状態にされる。この時点で、下位の増設筐体 200 (#2, #3) への通電が停止されているので、増設筐体 200 (#1) への外部電源の供給も停止される。

30

【0153】

このように構成される本実施例の効果では、ディスクドライブ 230 の利用状況に応じて、各増設筐体 200 への外部電源の供給そのものを停止させるため、従来よりも電力消費量を低減することができる。本実施例の構成から得られる他の効果は、既に述べた通りである。

【実施例 2】

【0154】

図 28 に基づいて、本発明の第 2 実施例を説明する。本実施例を含む以下の各実施例は、第 1 実施例の変形例に該当する。本実施例では、アクセス先ドライブへのアクセスができなかった場合にのみ、電源供給開始処理を実行する。

40

【0155】

図 28 は、本実施例による記憶制御装置のコントローラ 110 が実行するライト処理のフローチャートである。このフローチャート中、S70 ~ S74 は、図 19 に示す S40 ~ S44 にそれぞれ対応する。従って、重複した説明を割愛し、本実施例の特徴を中心に説明する。

【0156】

コントローラ 110 は、アクセス先ドライブが特定されると (S74)、アクセス先ドライブを備える増設筐体 200 にライトデータを転送する (S75)。ライトデータの書

50

込が正常に完了したことを確認した場合 (S76:YES)、コントローラ 110 は、本処理を終了する。

【0157】

アクセス先ドライブへのライトデータの書込みが正常に完了しなかった場合 (S76:NO)、コントローラ 110 は、テーブル T2 を用いることにより、アクセス先ドライブがスピニング状態であるか否かを判定する (S77)。

【0158】

アクセス先ドライブがスピンドアウン状態であると判定された場合 (S77:NO)、コントローラ 110 は、電源供給開始処理 (S100) を実行する。アクセス先ドライブがスピニング状態であるにもかかわらず、ライトデータを書き込めなかった場合 (S77:YES)、
10

【0159】

このように構成される本実施例も前記第 1 実施例と同様の効果を奏する。

【実施例 3】

【0160】

図 29 に基づいて、第 3 実施例を説明する。本実施例では、図 19 に示したライト処理と電源供給開始処理とを同一のフローチャート内で実現させる。

【0161】

コントローラ 110 は、ホスト 13 からライトコマンドを受領すると (S200)、ライトデータをキャッシュメモリ 114 に記憶させ (S201)、処理完了をホスト 13 に通知する。コントローラ 110 は、テーブル T2 を用いることにより、アクセス先ドライブの状態を確認し (S202)、アクセス先ドライブが動作中であるか否かを判定する (S203)。
20

【0162】

アクセス先ドライブが動作している場合 (S203:YES)、S211 に移る。アクセス先ドライブが動作していない場合 (S203:NO)、コントローラ 110 は、アクセス先ドライブがスピンドアウン状態であるか否かを判定する (S204)。アクセス先ドライブがスピンドアウン状態の場合 (S204:YES)、S210 に移る。

【0163】

アクセス先ドライブがスピンドアウン状態ではない場合 (S204:NO)、コントローラ 110 は、電源部 220 からアクセス先ドライブへの電源供給が停止されているか否かを判定する (S205)。電源部 220 から直流電源が各ディスクドライブ 230 に供給されている場合 (S205:YES)、S209 に移る。
30

【0164】

電源部 220 からの直流電源の出力が停止されているわけではない場合 (S205:NO)、コントローラ 110 は、アクセス先ドライブを有する対象筐体にそれぞれ接続されている各 PDB 300 内のスイッチ 310 をオンにする (S206)。コントローラ 110 は、対象筐体に隣接する上位の増設筐体 200 への電源供給が停止中であるか否かを判断する (S207)。
40

【0165】

隣接する上位の増設筐体 200 が電源供給停止中の場合 (S207:YES)、コントローラ 110 は、その上位の増設筐体 200 に接続されている各スイッチ 310 をそれぞれオンにして、外部電源を供給させる (S208)。S207 及び S208 を繰り返すことにより、基本筐体 100 とアクセス先ドライブを有する対象筐体との間に位置する各増設筐体 200 に順番に外部電源が供給されていき、伝達経路が形成される。

【0166】

隣接する上位の増設筐体 200 が電源供給の停止中ではない場合 (S207:NO)、コントローラ 110 は、対象筐体内の各ディスクドライブ 230 へ電源部 220 から直流電源を供給するように、制御基板 210 に指示する (S209)。
50

【 0 1 6 7 】

続いて、コントローラ 1 1 0 は、アクセス先ドライブをスピンドルアップさせるためのスピンドルアップコマンドを発行し (S 2 1 0)、アクセス先ドライブにライトデータを転送して記憶させる (S 2 1 1)。

【 0 1 6 8 】

このように構成される本実施例も前記第 1 実施例と同様の効果を奏する。

【実施例 4】

【 0 1 6 9 】

図 3 0 ~ 図 3 2 に基づいて、第 4 実施例を説明する。本実施例では、複数の記憶制御装置を備えるストレージシステムにおいて、1つの記憶制御装置 1 0 0 0 が他の記憶制御装置 1 1 0 0 , 1 2 0 0 の有する記憶領域を仮想化してホスト 1 3 に提供する。

10

【 0 1 7 0 】

図 3 0 に示すストレージシステムは、複数の記憶制御装置 1 0 0 0 , 1 1 0 0 , 1 2 0 0 を備えている。記憶制御装置 1 0 0 0 は、他の各記憶制御装置 1 1 0 0 , 1 2 0 0 に接続されている接続元の記憶制御装置であり、他の各記憶制御装置 1 1 0 0 , 1 2 0 0 を管理下においている。

【 0 1 7 1 】

接続元の記憶制御装置 1 0 0 0 は、接続先の記憶制御装置 1 1 0 0 , 1 2 0 0 がそれぞれ有する記憶領域を、記憶制御装置 1 0 0 0 の内部で仮想化し、それがあたかも記憶制御装置 1 0 0 0 の内部ボリュームであるかのように、ホスト 1 3 に提供する。従って、接続元の記憶制御装置 1 0 0 0 は、各記憶制御装置 1 0 0 0 , 1 1 0 0 , 1 2 0 0 のシステム構成に関する情報を一元的に管理している。

20

【 0 1 7 2 】

各記憶制御装置 1 0 0 0 , 1 1 0 0 , 1 2 0 0 は、それぞれ同一の基本構造を備えて構成することもできるし、それぞれ異なる構造を備えてもよい。本実施例では、記憶制御装置 1 0 0 0 , 1 1 0 0 , 1 2 0 0 は、前記第 1 実施例で述べたような構成をそれぞれ備えているものとする。

【 0 1 7 3 】

接続元の記憶制御装置 1 0 0 0 の基本筐体 1 0 0 と、接続先の各記憶制御装置 1 1 0 0 , 1 2 0 0 の各基本筐体 1 0 0 とは通信経路 P 4 を介して接続されている。この通信経路 P 4 は、例えば、SAN等の通信ネットワークとして構成することができる。また、接続元の記憶制御装置 1 0 0 0 の基本筐体 1 0 0 と、接続先の各記憶制御装置 1 1 0 0 , 1 2 0 0 の各 PDB 3 0 0 とは、信号線 P 5 を介して接続されている。

30

【 0 1 7 4 】

図 3 1 は、本実施例による電源供給停止処理を示すフローチャートである。この処理は、接続元記憶制御装置 1 0 0 0 の基本筐体 1 0 0 内のコントローラ 1 1 0 によって実行される。

【 0 1 7 5 】

接続元のコントローラ 1 1 0 は、処理対象の記憶制御装置として、装置番号の最も少ない記憶制御装置 1 1 0 0 を選択する (S 3 0 0)。コントローラ 1 1 0 は、対象記憶制御装置 1 1 0 0 の有する全てのボリューム 2 5 0 について、アクセス不要であるか否かを判定する (S 3 0 1)。

40

【 0 1 7 6 】

記憶制御装置 1 1 0 0 の有する全ボリューム 2 5 0 についてアクセス不要であると判断された場合 (S301:YES)、接続元のコントローラ 1 1 0 は、記憶制御装置 1 1 0 0 の基本筐体 1 0 0 に向けて、シャットダウンコマンドを発行する (S 3 0 2)。

【 0 1 7 7 】

シャットダウンコマンドを受領した記憶制御装置 1 1 0 0 のコントローラ 1 1 0 は、シャットダウン処理を実行する (S 3 0 3)。シャットダウン処理では、例えば、下位の増設筐体 2 0 0 から順番に電源の供給を停止させていき、最後に基本筐体 1 0 0 の動作も停

50

止させる。

【0178】

接続元のコントローラ110は、記憶制御装置1100に接続されている各PDB300内の各スイッチ310を下位側から順番にオフさせる(S304)。これにより、接続先の記憶制御装置1100の基本筐体100及び各増設筐体200への外部電源の供給が停止される。

【0179】

接続元のコントローラ110は、全ての接続先記憶制御装置について判断したか否かを判定する(S305)。未判定の記憶制御装置が存在する場合(S305:NO)、コントローラ110は、対象の記憶制御装置の装置番号を次の接続先記憶制御装置1200の装置番号に変更し(S306)、S301～S305を繰り返す。

10

【0180】

図32は、電源供給を開始する場合の処理を示すフローチャートである。接続元のコントローラ110は、対象の記憶制御装置に接続された各PDB300内の各スイッチ310を上位側から順番にオンさせる(S310)。

【0181】

接続先の記憶制御装置では、各系統のPDB300から外部電源が供給されると、まず最初に、基本筐体100を起動させ(S311)、次に、外部電源が供給された順番で、各増設筐体200を起動させる(S312)。

【0182】

このように構成される本実施例も前記第1実施例と同様の効果を奏する。これに加えて、本実施例では、複数の記憶制御装置1000, 1100, 1200を備えるストレージシステムにおいて、より一層、電力消費量を節減することができる。

20

【実施例5】

【0183】

図33は、第5実施例に係るストレージシステムを示す説明図である。本実施例では、接続先の各記憶制御装置1100, 1200の各PDB300に電源を供給する電源装置15の動作を、接続元記憶制御装置1000の基本筐体100が制御する。

【0184】

このように構成される本実施例も前記第4実施例と同様の効果を奏する。

30

【0185】

なお、本発明は、上述した実施の形態に限定されない。当業者であれば、本発明の範囲内で、種々の追加や変更等を行うことができる。例えば、各実施例を適宜組み合わせることができる。

【図面の簡単な説明】

【0186】

【図1】本発明の実施形態の概念を示す説明図である。

【図2】記憶制御装置の外観を示す斜視図である。

【図3】基本筐体を前側から見た斜視図である。

【図4】基本筐体を後側から見た斜視図である。

40

【図5】増設筐体を前側から見た斜視図である。

【図6】増設筐体を後側から見た斜視図である。

【図7】記憶制御装置の電源供給構造を示す説明図である。

【図8】記憶制御装置の構成を簡略化して示す説明図である。

【図9】コントローラの構成を示す説明図である。

【図10】増設筐体の構成を示す説明図である。

【図11】制御基板の構成を示す説明図である。

【図12】ボリュームを管理するためのテーブルを示す説明図である。

【図13】電源供給状態を管理するためのテーブルを示す説明図である。

【図14】データ再配置処理を示すフローチャートである。

50

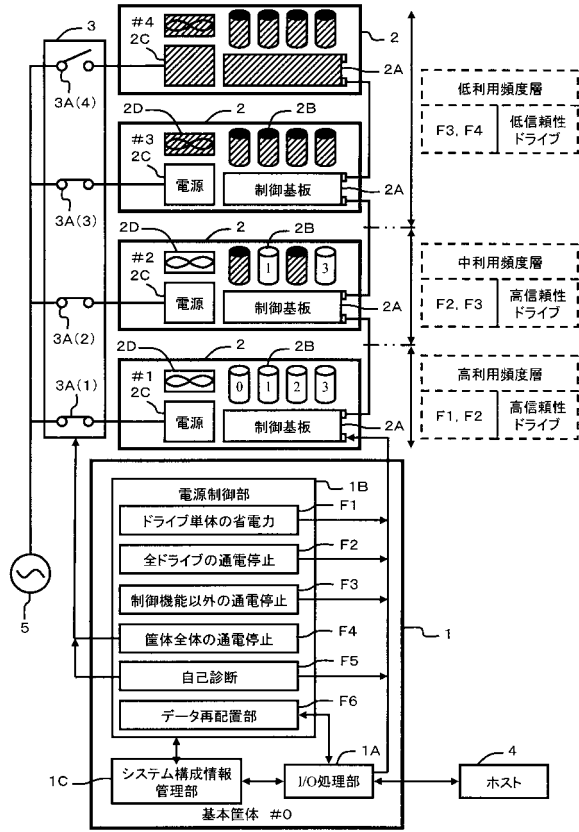
- 【図15】利用頻度に応じてデータが配置される様子を示す説明図である。
- 【図16】電源供給の接続を確認するための処理を示すフローチャートである。
- 【図17】各ドライブ毎に節電するための処理を示すフローチャートである。
- 【図18】増設筐体毎に電源供給を停止等するための処理を示すフローチャートである。
- 【図19】ライト処理を示すフローチャートである。
- 【図20】リード処理を示すフローチャートである。
- 【図21】電源供給を開始する場合の処理を示すフローチャートである。
- 【図22】通電停止中の増設筐体を診断するための処理を示すフローチャートである。
- 【図23】ディスクドライブの利用状況に応じて節電モードが段階的に変化していく様子
を示す説明図である。 10
- 【図24】PDBの各スイッチの状態が変化する様子を示す説明図である。
- 【図25】図24に続く説明図である。
- 【図26】図25に続く説明図である。
- 【図27】図26に続く説明図である。
- 【図28】第2実施例に係るライト処理を示すフローチャートである。
- 【図29】第3実施例に係るライト処理及び電源供給開始処理を示すフローチャートであ
る。
- 【図30】第4実施例に係る記憶制御装置が用いられるストレージシステムの全体構成を
示す説明図である。 20
- 【図31】記憶制御装置単位で電源供給を停止させる処理を示すフローチャートである。
- 【図32】記憶制御装置単位で電源供給を再開させる処理を示すフローチャートである。
- 【図33】第5実施例に係る記憶制御装置が用いられるストレージシステムの全体構成を
示す説明図である。

【符号の説明】

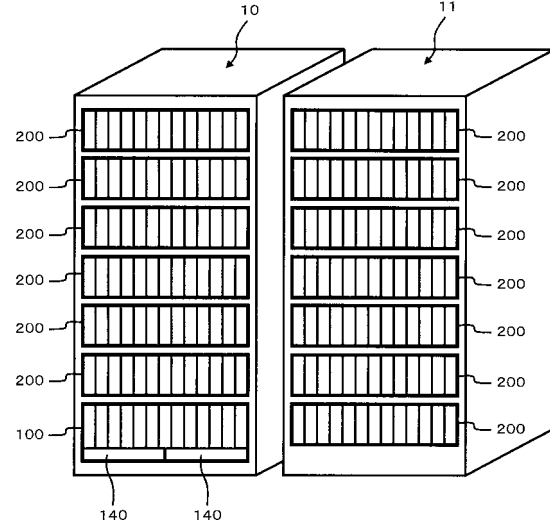
【0187】

1 ... 基本筐体、1 A ... I/O処理部、1 B ... 電源制御部、1 C ... システム構成情報管理部
、2 ... 増設筐体、2 A ... 制御基板、2 B ... ディスクドライブ、2 C ... 筐体内電源部、2 D
... 冷却ファン、3 ... 電源供給回路部、3 A ... スイッチ、4 ... ホスト、5 ... 外部電源、1 0
... 基本ラック、1 1 ... 増設ラック、1 3 ... ホストコンピュータ、1 4 ... 管理端末、1 5 ...
電源装置、1 0 0 ... 基本筐体、1 0 1 ... ケーシング、1 1 0 ... コントローラ、1 1 1 ... ホ
ストインターフェース回路、1 1 2 ... ドライブインターフェース回路、1 1 3 ... データ転
送回路、1 1 4 ... キャッシュメモリ、1 1 5 ... プログラムメモリ、1 1 6 ... プロセッサ、
1 1 7 ... ブリッジ、1 1 8 ... LANインターフェース、1 2 0 ... AC/DC電源部、1 3 0 ... 冷却
ファン、1 4 0 ... バッテリー、2 0 0 ... 増設筐体、2 0 1 ... ケーシング、2 0 2 ... 冷却ファ
ン、2 1 0 ... 制御基板、2 1 1 ... マイクロプロセッサ部、2 1 1 A ... 資源管理機能、2 1
1 B ... コマンド処理機能、2 1 1 C ... 電源制御機能、2 1 1 D ... 冷却ファン制御機能、2
1 2 ... 筐体間通信インターフェース部、2 1 3 ... 制御信号出力回路、2 2 0 ... AC/DC電源
部、2 2 1 ... 電圧変換回路、2 2 2 A ~ 2 2 2 C ... 出力ダイオード、2 2 3 ... 出力制御ス
イッチ、2 3 0 ... ディスクドライブ、2 4 0 ... RAIDグループ、2 5 0 ... 論理ボリューム、
3 0 0 ... 電源分配回路部 (PDB)、3 1 0 ... スイッチ、1 0 0 0 , 1 1 0 0 , 1 2 0 0 ...
記憶制御装置 40

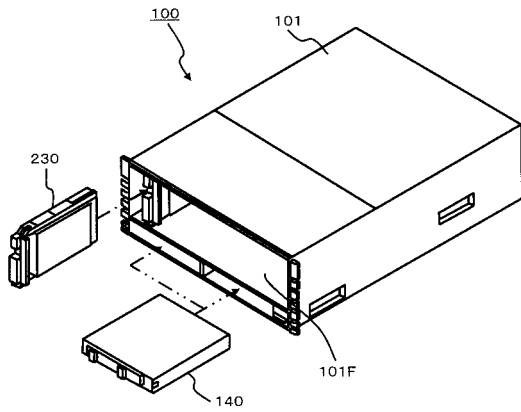
【図1】



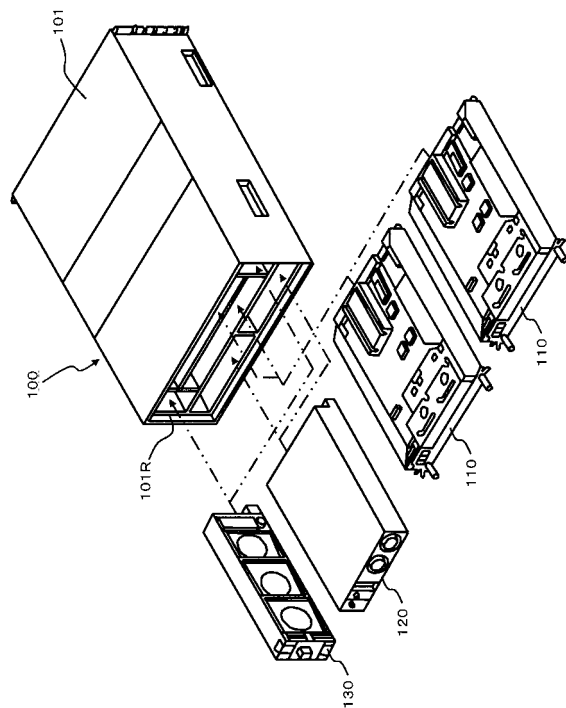
【図2】



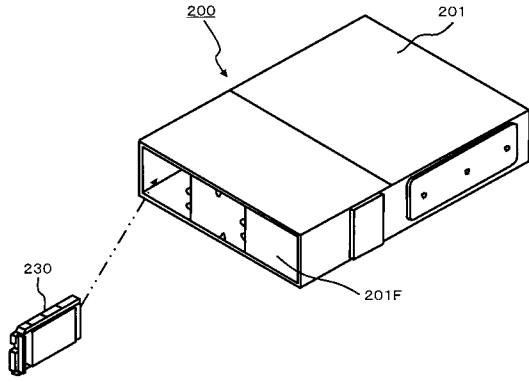
【図3】



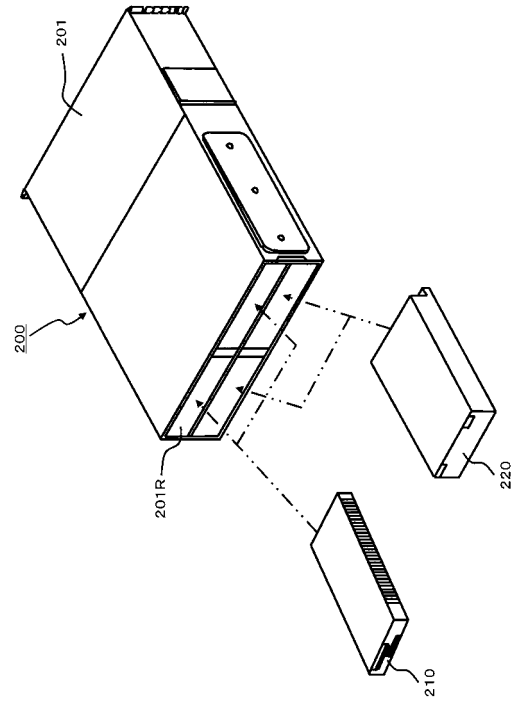
【図4】



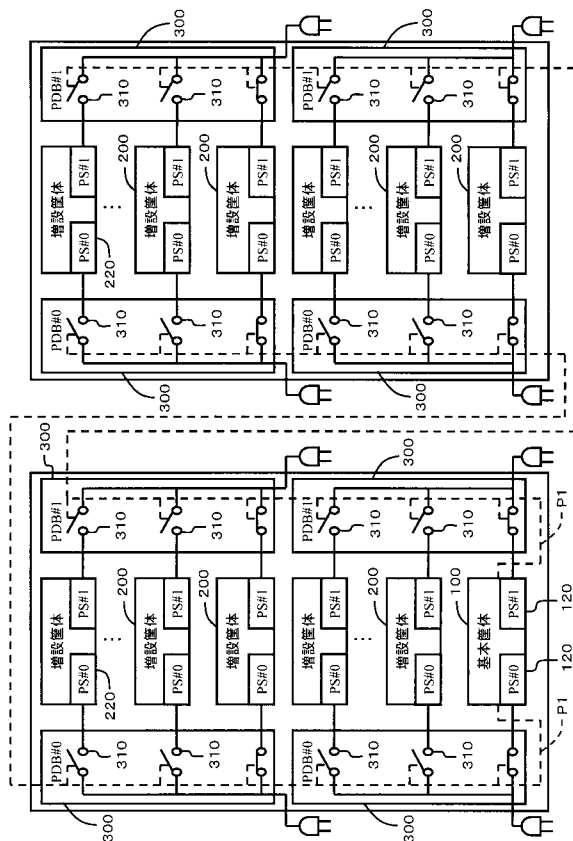
【図5】



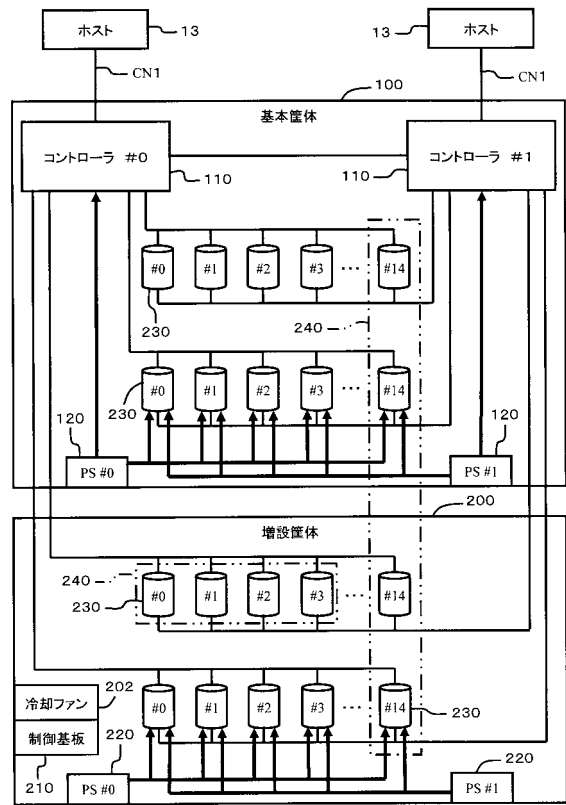
【図6】



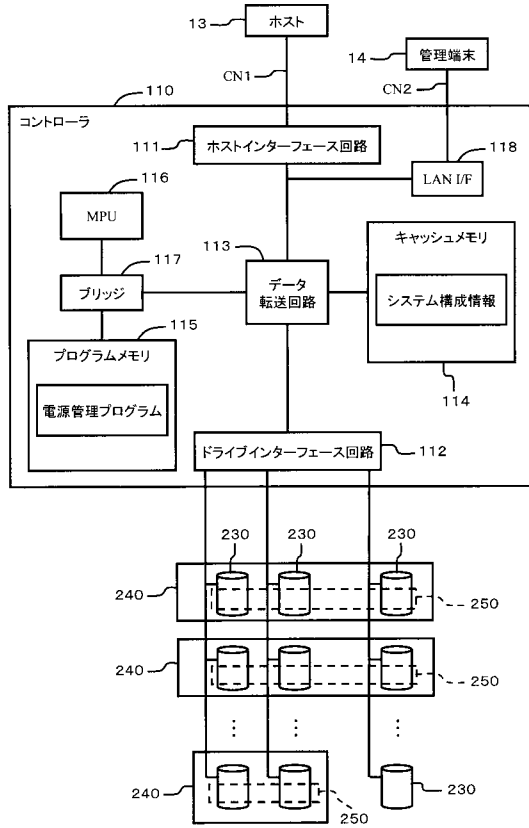
【図7】



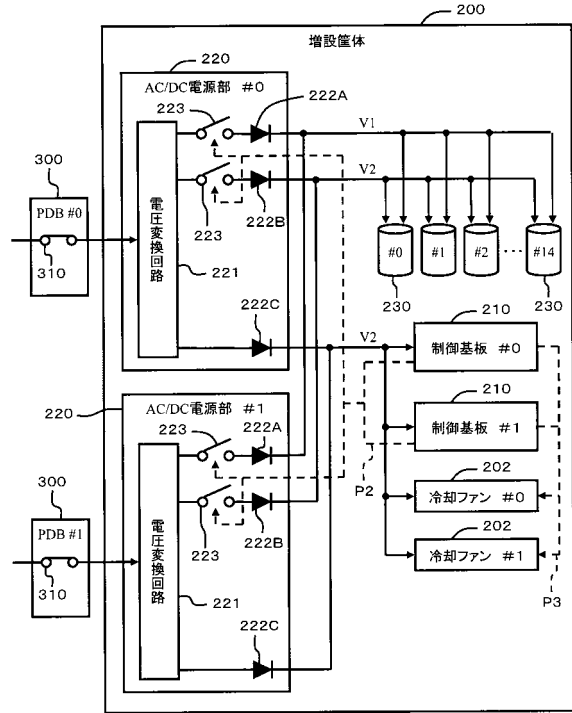
【図8】



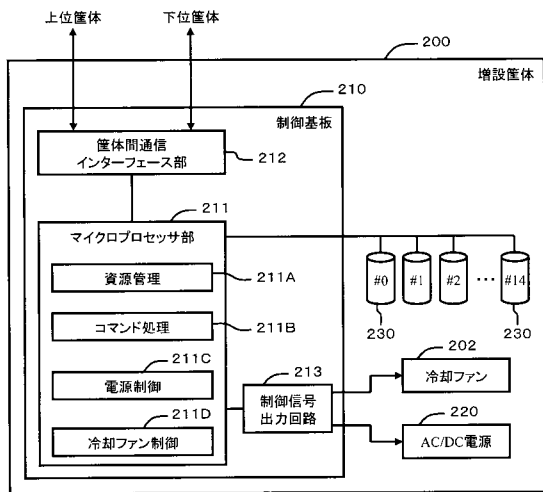
【図9】



【図10】



【図11】



【図12】

ボリューム管理テーブル (Volume Management Table)

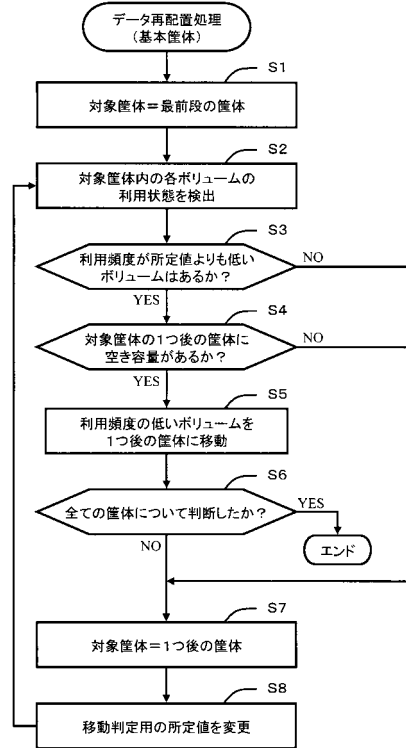
I11	I12	I13	I14	I15	I16
LU#	容量	RAIDグループ#	RAIDレベル	ドライブ#リスト	最新アクセス時刻
LU#0	200MB	RG01	RAID5	HDD01,02,03,04	2006/09/25 12:00:00
...

【図13】

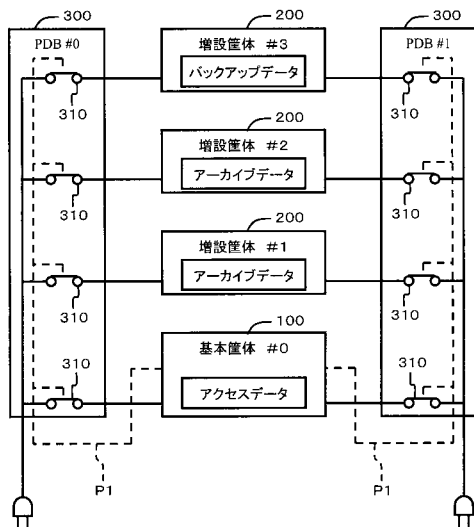
電源供給状態管理テーブル

筐体#	筐体への電源供給状態	ドライブ#	スピンドル状態	第1PDB SW#	SW状態	第2PDB SW#	SW状態
#0	ON	HDD00	ON	01-0	ON	11-0	ON
		HDD01	ON				
		HDD02	ON				
					
		HDD14	ON				
#1	ON	HDD00	ON	01-1	ON	11-1	ON
		HDD01	OFF				
		HDD02	ON				
					
		HDD14	ON				
#2	ON	HDD00	ON	01-2	ON	11-2	ON
		HDD01	OFF				
		HDD02	ON				
					
		HDD14	ON				
#3	ON	HDD00	ON	01-3	ON	11-4	ON
		HDD01	OFF				
		HDD02	ON				
					
		HDD14	ON				
...							

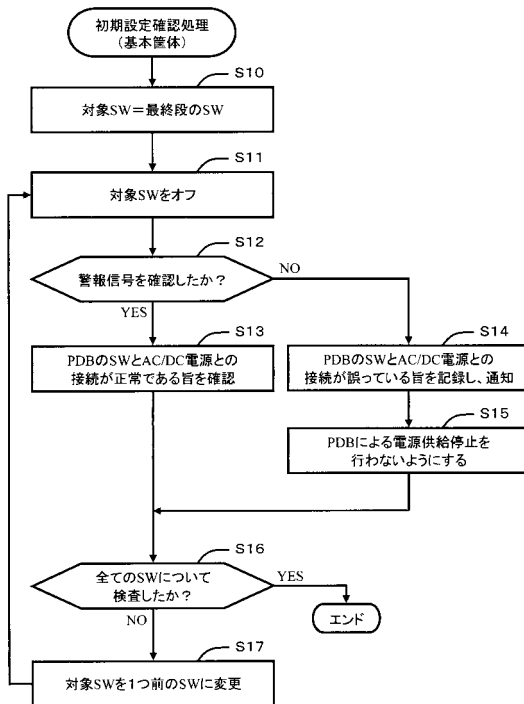
【図14】



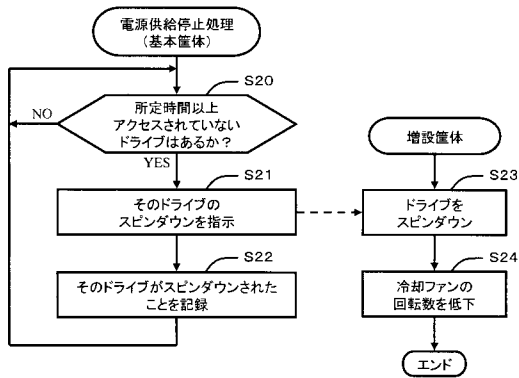
【図15】



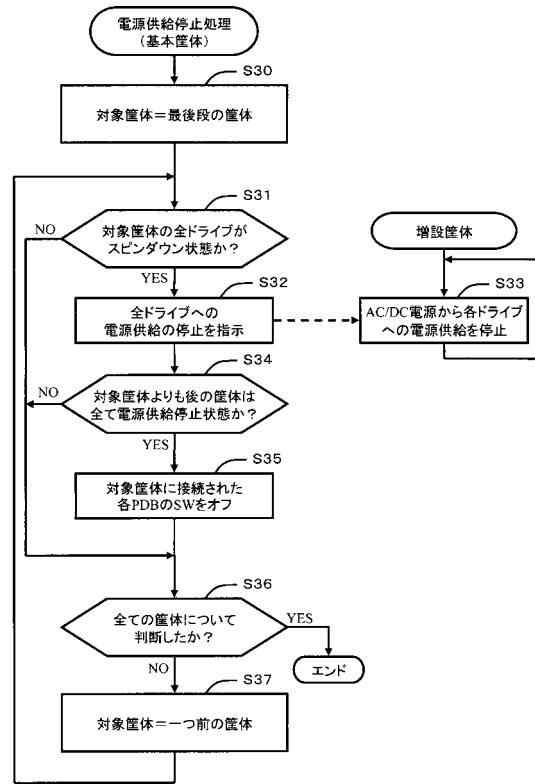
【図16】



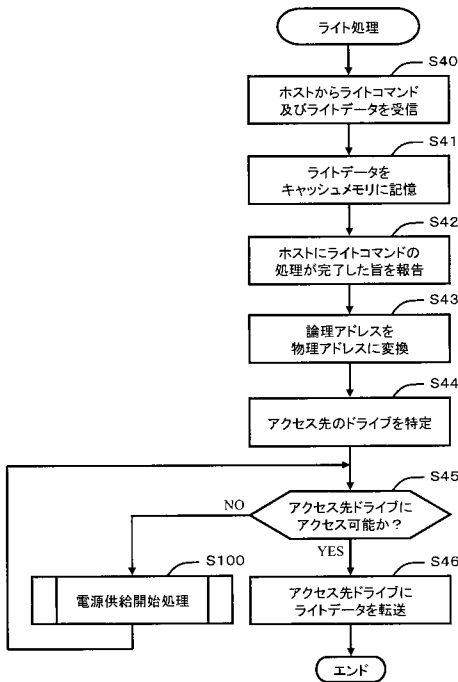
【図17】



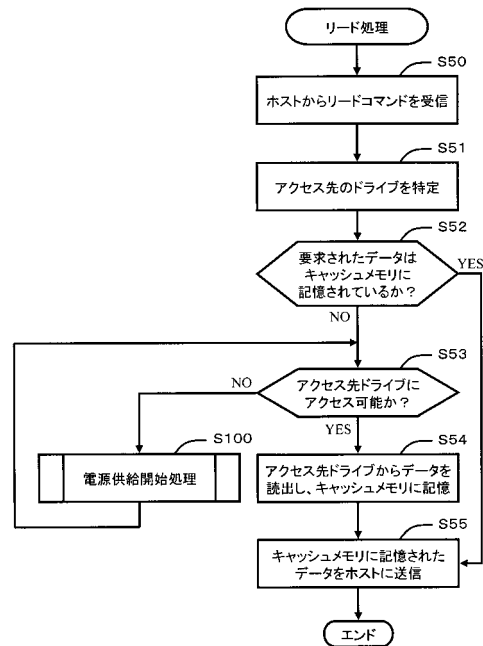
【図18】



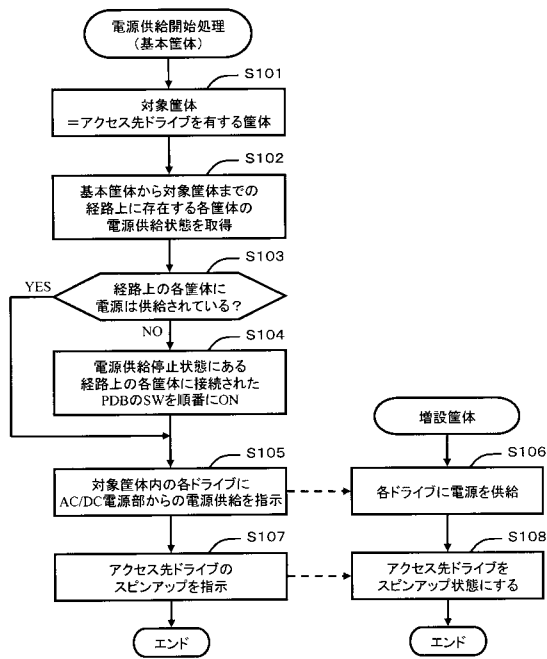
【図19】



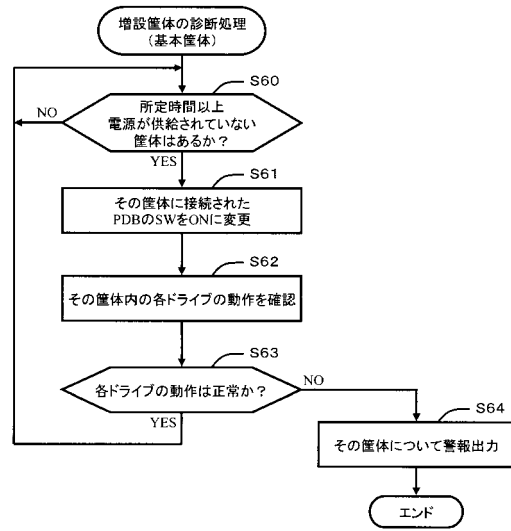
【図20】



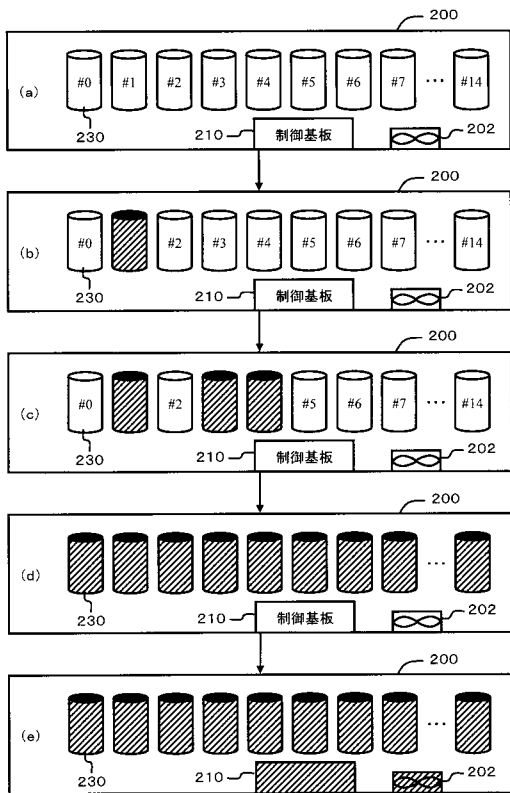
【図 2 1】



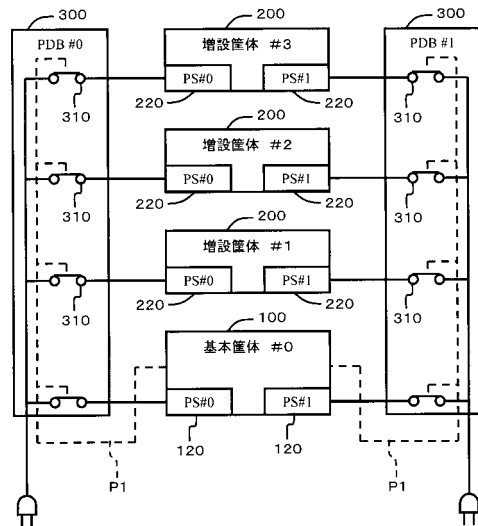
【図 2 2】



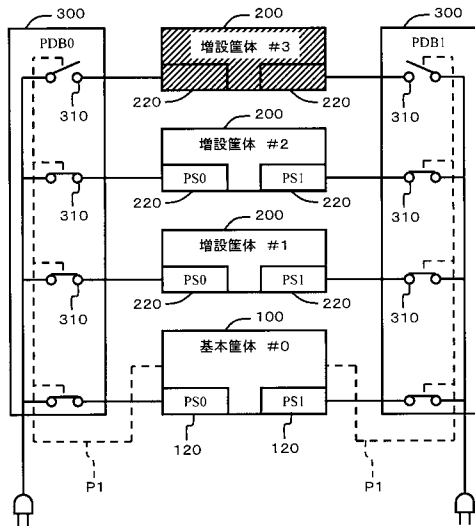
【図 2 3】



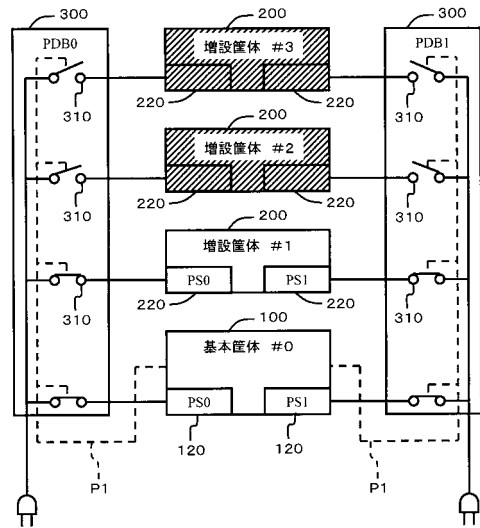
【図 2 4】



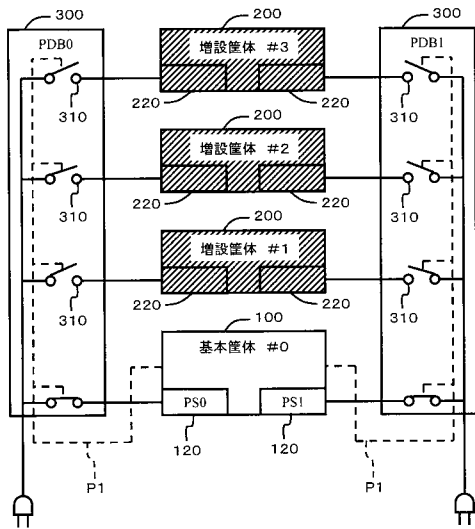
【図 25】



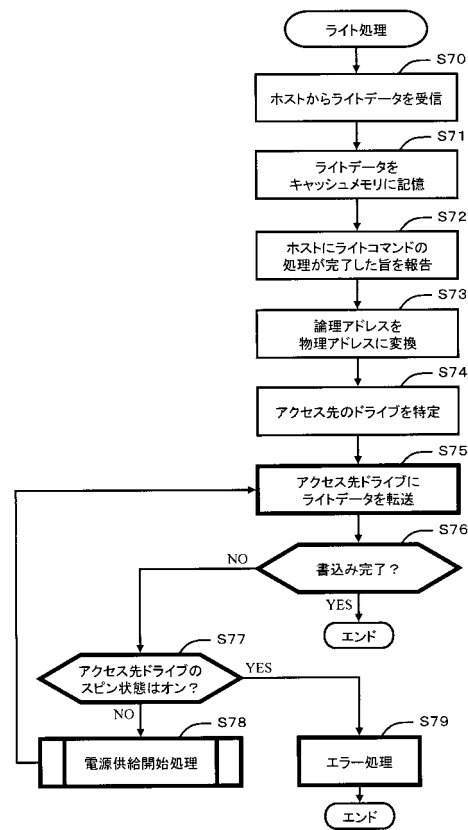
【図 26】



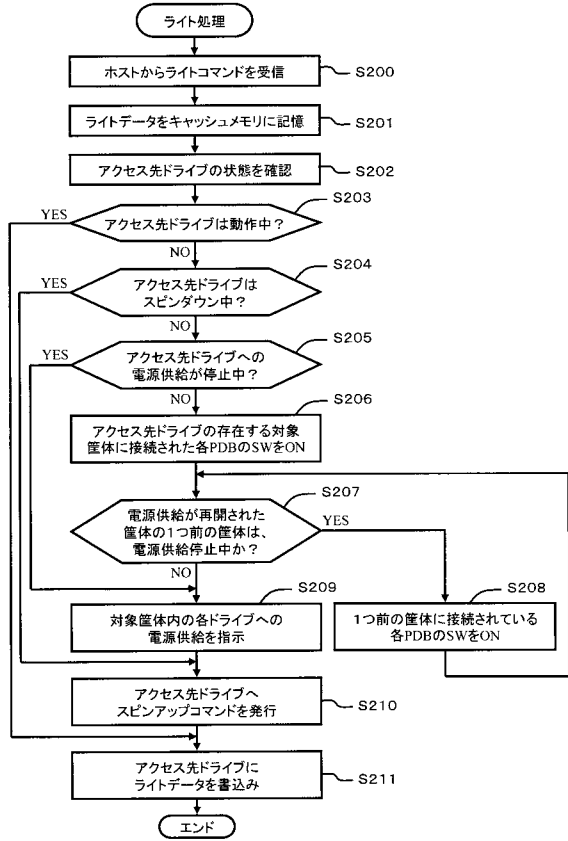
【図 27】



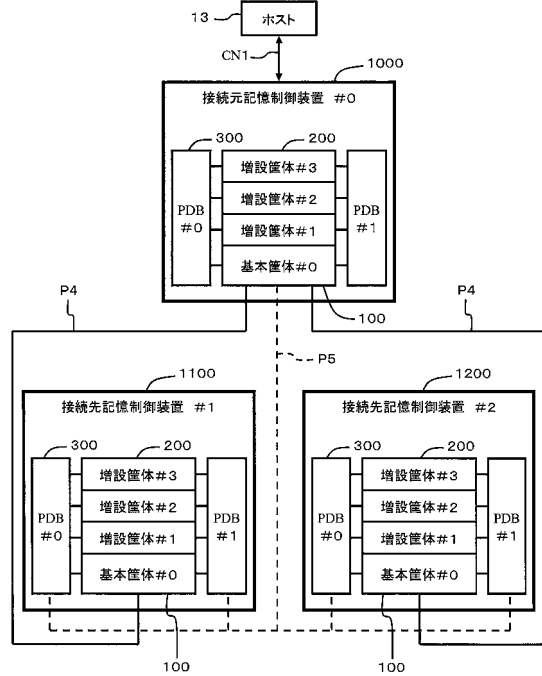
【図 28】



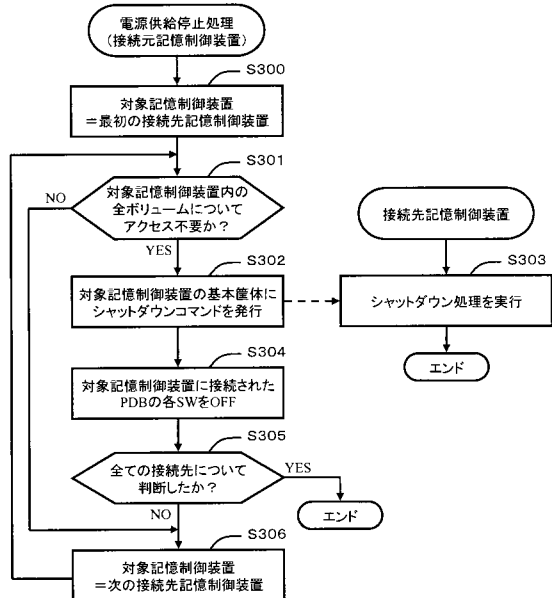
【図 29】



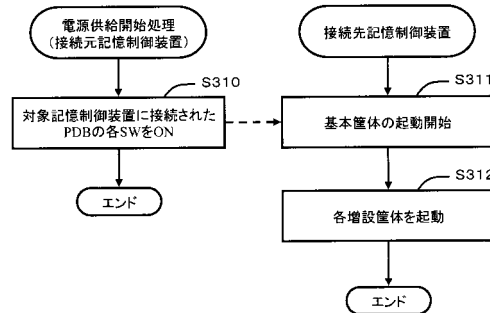
【図 30】



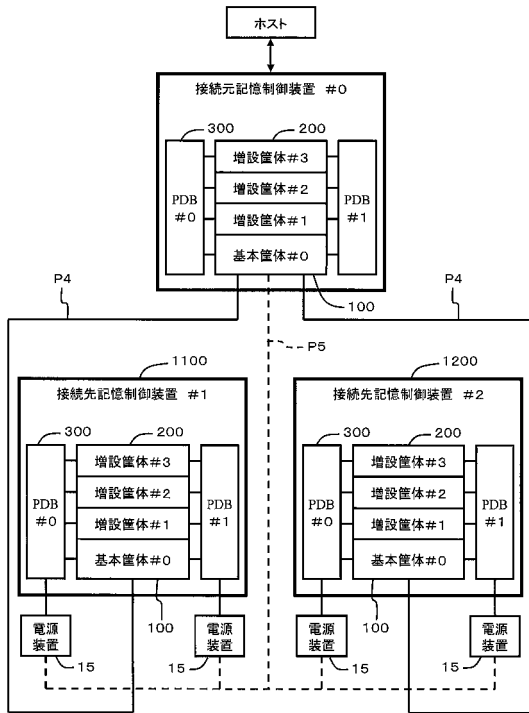
【図 31】



【図 32】



【図 33】



フロントページの続き

(72)発明者 山梨 明人

神奈川県足柄上郡中井町境781番地 日立コンピュータ機器株式会社内

審査官 三浦 みちる

(56)参考文献 特開平08-221219(JP,A)
特開2000-293314(JP,A)
特開2001-337789(JP,A)
特開2002-297320(JP,A)
特開2004-348876(JP,A)
国際公開第2006/098036(WO,A1)

(58)調査した分野(Int.Cl., DB名)

G06F 1/32

G06F 3/06