(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4958641号
(P4958641)
(45)発行日　平成24年6月20日(2012.6.20)　(24)登録日　平成24年3月30日(2012.3.30)

(51)Int.Cl.　F I
G06F 3/06 (2006.01)　G06F 3/06 301E

請求項の数 18 (全 28 頁)

(21)出願番号　特願2007-142426(P2007-142426)
(22)出願日　平成19年5月29日(2007.5.29)
(65)公開番号　特開2008-299434(P2008-299434A)
(43)公開日　平成20年12月11日(2008.12.11)
審査請求日　平成22年2月8日(2010.2.8)

(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　100093861
弁理士　大賀　眞司
(72)発明者　永田　幸司
神奈川県小田原市中里３２２番２号　株式会社日立製作所ＲＡＩＤシステム事業部内
(72)発明者　松井　佑光
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　小林　正明
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】記憶制御装置及びその制御方法

(57)【特許請求の範囲】
【請求項１】
記憶装置が提供する記憶領域上にデータを入出力するための論理的な領域でなる複数の論理ボリュームを作成し、作成した前記論理ボリュームをホスト計算機に提供すると共に、前記ホスト計算機からの要求に応じて、前記論理ボリュームに対するデータの入出力を制御する記憶制御装置において、
それぞれ自己が担当する前記論理ボリュームに対するデータの入出力を制御する複数の制御部と、
各前記制御部の負荷状況を監視する負荷監視部と、
前記負荷監視部の監視結果に基づいて、各前記制御部の負荷を均等化するように、前記論理ボリュームの担当を他の前記制御部に切り替える切替え部と
を備え、
前記切替え部は、
コピーペアの作成時に、当該コピーペアを形成する前記論理ボリュームの担当と、当該論理ボリュームと相互に関連する他の前記論理ボリュームの担当とが異なるときには、一方の前記コピーペアを形成する論理ボリューム又は前記他の論理ボリュームの担当を、他方の前記他の論理ボリューム又は前記コピーペアを形成する論理ボリュームを担当する前記制御部に切り替える
ことを特徴とする記憶制御装置。
【請求項２】

前記論理ボリュームごとの担当の前記制御部を規定した担当情報を記憶する記憶部を備え、
前記制御部は、
前記担当情報に基づいて自己が担当する前記論理ボリュームを認識し、
前記切替え部は、
前記担当情報における対応する前記論理ボリュームの担当を変更するようにして、当該論理ボリュームの担当を他の前記制御部に切り替える
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３】
前記論理ボリュームに対するデータの入出力を制御する複数のコントローラを備え、
前記制御部は、
各前記コントローラにそれぞれ設けられたマイクロプロセッサを構成するコアである
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４】
前記切替え部は、いずれかの前記論理ボリュームの担当を他の前記制御部に切り替えた後、所定時間内における前記論理ボリュームの担当の切り替えを抑止する
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５】
前記切替え部は、
担当切替え候補の論理ボリュームと相互に関連する他の前記論理ボリュームが存在するときには、担当切替え候補の論理ボリューム及び当該他の論理ボリュームの担当を一括して同じ前記他の制御部に切り替える
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６】
前記切替え部は、
前記コピーペアが内部コピーを行う前記論理ボリューム同士のペアであるときには、副側の前記論理ボリュームの担当を、正側の前記論理ボリュームを担当する前記制御部に切り替える
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】
前記切替え部は、
前記コピーペアが前記論理ボリュームと、当該論理ボリュームのスナップショットのデータを格納するプールボリュームとのペアであるときには、正側の前記論理ボリュームの担当を、前記プールボリュームを担当する前記制御部に切り替える
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８】
前記切替え部は、
前記コピーペアが、正側及び副側の各前記論理ボリュームがそれぞれ異なる記憶装置内に作成され、正側の前記論理ボリュームに対するデータの書き込みとは非同期に正側及び副側の各前記論理ボリューム間でデータをコピーする非同期リモートコピーのペアであるときには、正側の前記論理ボリュームの担当を、当該正側の論理ボリュームのスナップショットのデータを格納するプールボリュームの担当の前記制御部に切り替え、副側の前記論理ボリュームの担当を、当該副側の論理ボリュームのスナップショットのデータを格納するプールボリュームを担当する前記制御部に切り替える
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】
前記記憶装置と通信するための通信ポートを備え、
正側及び副側の各前記論理ボリュームがそれぞれ異なる記憶装置内に作成され、前記ホスト計算機から正側の前記論理ボリュームに対するデータの書き込みと同期して、当該データを正側及び副側の各前記論理ボリューム間でコピーするリモートコピーの正側の前記

論理ボリュームを担当する前記制御部と、前記リモートコピーの正側の論理ボリュームが設けられた前記記憶装置と通信するための前記通信ポートを担当する前記制御部とが異なる場合、リンクサービス系のコマンド単位の処理を実行させるためのコマンドについては、前記リモートコピーの正側の前記論理ボリュームを担当する前記制御部が発行し、前記リンクサービス系のポート単位の処理を実行させるためのコマンドについては、当該通信ポートを担当する前記制御部が発行する

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１０】

記憶装置が提供する記憶領域上にデータを入出力するための論理的な領域でなる複数の論理ボリュームを作成し、作成した前記論理ボリュームをホスト計算機に提供すると共に、前記ホスト計算機からの要求に応じて、前記論理ボリュームに対するデータの入出力を制御する記憶制御装置の制御方法において、

前記記憶制御装置は、それぞれ自己が担当する前記論理ボリュームに対するデータの入出力を制御する複数の制御部を備え、

各前記制御部の負荷状況を監視する第１のステップと、

監視結果に基づいて、各前記制御部の負荷を均等化するように、前記論理ボリュームの担当を他の前記制御部に切り替える第２のステップと

を備え、

コピーペアの作成時に、当該コピーペアを形成する前記論理ボリュームの担当と、当該論理ボリュームと相互に関連する他の前記論理ボリュームの担当とが異なるときには、一方の前記コピーペアを形成する論理ボリューム又は前記他の論理ボリュームの担当を、他方の前記他の論理ボリューム又は前記コピーペアを形成する論理ボリュームを担当する前記制御部に切り替える

ことを特徴とする記憶制御装置の制御方法。

【請求項１１】

前記記憶制御装置は、前記論理ボリュームごとの担当の前記制御部を規定した担当情報を記憶する記憶部を備え、

前記制御部は、前記担当情報に基づいて自己が担当する前記論理ボリュームを認識し、

前記第２のステップでは、

前記担当情報における対応する前記論理ボリュームの担当を変更するようにして、当該論理ボリュームの担当を他の前記制御部に切り替える

ことを特徴とする請求項１０に記載の制御方法。

【請求項１２】

前記記憶制御装置は、前記論理ボリュームに対するデータの入出力を制御する複数のコントローラを備え、

前記制御部は、各前記コントローラにそれぞれ設けられたマイクロプロセッサを構成するコアである

ことを特徴とする請求項１０に記載の制御方法。

【請求項１３】

いずれかの前記論理ボリュームの担当を他の前記制御部に切り替えた後、所定時間内における前記論理ボリュームの担当の切り替えを抑止する第３のステップを備える

ことを特徴とする請求項１０に記載の制御方法。

【請求項１４】

前記第２のステップでは、

担当切替え候補の前記論理ボリュームと相互に関連する他の前記論理ボリュームが存在するときには、担当切替え候補の論理ボリューム及び当該他の論理ボリュームの担当を一括して同じ前記他の制御部に切り替える

ことを特徴とする請求項１０に記載の制御方法。

【請求項１５】

前記コピーペアが内部コピーを行う前記論理ボリューム同士のペアであるときには、副

側の前記論理ボリュームの担当を、正側の前記論理ボリュームを担当する前記制御部に切り替える
ことを特徴とする請求項１０に記載の制御方法。
【請求項１６】
前記コピーペアが前記論理ボリュームと、当該論理ボリュームのスナップショットのデータを格納するプールボリュームとのペアであるときには、正側の前記論理ボリュームの担当を、前記プールボリュームを担当する前記制御部に切り替える
ことを特徴とする請求項１０に記載の制御方法。
【請求項１７】
前記コピーペアが、正側及び副側の各前記論理ボリュームがそれぞれ異なる記憶装置内に作成され、正側の前記論理ボリュームに対するデータの書き込みとは非同期に正側及び副側の各前記論理ボリューム間でデータをコピーする非同期リモートコピーのペアであるときには、正側の前記論理ボリュームの担当を、当該正側の論理ボリュームのスナップショットのデータを格納するプールボリュームの担当の前記制御部に切り替え、副側の前記論理ボリュームの担当を、当該副側の論理ボリュームのスナップショットのデータを格納するプールボリュームを担当する前記制御部に切り替える
ことを特徴とする請求項１０に記載の制御方法。
【請求項１８】
前記記憶制御装置は、前記記憶装置と通信するための通信ポートを備え、
正側及び副側の各前記論理ボリュームがそれぞれ異なる記憶装置内に作成され、前記ホスト計算機から正側の前記論理ボリュームに対するデータの書き込みと同期して、当該データを正側及び副側の各前記論理ボリューム間でコピーするリモートコピーの正側の前記論理ボリュームを担当する前記制御部と、前記リモートコピーの正側の論理ボリュームが設けられた前記記憶装置と通信するための前記通信ポートを担当する前記制御部とが異なる場合、リンクサービス系のコマンド単位の処理を実行させるためのコマンドについては、前記リモートコピーの正側の前記論理ボリュームを担当する前記制御部に発行させ、前記リンクサービス系のポート単位の処理を実行させるためのコマンドについては、当該通信ポートを担当する前記制御部に発行させる
ことを特徴とする請求項１０に記載の制御方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、記憶制御装置及びその制御方法に関し、特に、ホスト計算機からのデータの入出力を制御するマイクロプロセッサが複数のコア（計算機エンジン）からなる記憶制御装置に適用して好適なものである。
【背景技術】
【０００２】
記憶制御装置は、記憶装置が提供する記憶領域上にデータを入出力するための論理的な領域でなる複数の論理ボリュームを作成し、作成した論理ボリュームをホスト計算機に提供する機能を有する。このような記憶制御装置は、ホスト計算機からの要求に応じて、論理ボリュームに対するデータの入出力を制御するコントローラを備えて構成される。
【０００３】
従来より、この種の記憶制御装置に対して、ホスト計算機から与えられるコマンドに対する処理の信頼性及び応答性の向上が求められている。そこで、近年では、記憶制御装置に複数のコントローラを設け、これら複数のコントローラによって処理を分散して実行する技術が提案されている。
【０００４】
特許文献１には、一方のコントローラが受信したコマンドに基づく処理を、そのコマンドを受信したコントローラが備えるロプロセッサと、他方のコントローラが備えるプロセッサとの両方のプロセッサを利用して実行する技術が開示されている。また特許文献２に

は、デュアルコントローラ構成の記憶制御装置が記載されている。
【特許文献１】特開平９－１４６８４２号公報
【特許文献２】特開平１１－３１２０５８号公報
【発明の開示】
【発明が解決しようとする課題】
【０００５】
これら特許文献１及び２では、多重化したコントローラを備える記憶制御装置について開示されている。各コントローラには、そのコントローラがデータ入出力処理等を担当すべき論理ボリュームが対応付けられており、コントローラが担当外の論理ボリュームを対象とするコマンドをホスト計算機から受信した場合、そのコントローラはそのコマンドに応じた処理をかかる論理ボリュームを担当するコントローラに依頼する。
【０００６】
ところが、このように複数のコントローラを記憶制御装置に設けた場合、ホスト計算機からのデータ入出力要求が各コントローラに均等に発行されていても、そのデータ入出力処理自体は対象となる論理ボリュームを担当するコントローラによって行なわれるため、特定の論理ボリュームにアクセスが集中するような場合に、負荷がその論理ボリュームを担当するコントローラに偏ってしまう問題があった。そして、このように負荷が特定のコントローラに偏った場合、そのコントローラが担当する論理ボリュームに対するデータ入出力処理の処理速度が低下することとなる。
【０００７】
また上述のように複数のコントローラを記憶制御装置に設けた場合、論理ボリューム間でデータをコピーする際に、コピー元である正側の論理ボリュームと、コピー先である副側の論理ボリュームとを異なるコントローラが担当することもあり得る。しかしながら、このようにコピー元の論理ボリュームを担当するコントローラと、コピー先の論理ボリュームを担当するコントローラとが異なる場合、コピー処理の際にコントローラ間での通信が必要となり、その分コントローラの負荷が増加する問題があった。そしてこのようなコントローラの負荷の増加は、そのコントローラが担当する論理ボリュームに対するデータ入出力処理の処理速度の低下を引き起こす原因ともなり得る。
【０００８】
従って、上述のようなデータ入出力処理の処理速度の低下を防止するためには、コントローラ間で負荷の偏りが生じないように、またコピー元の論理ボリューム及びコピー先の論理ボリュームを同じコントローラが担当するように、論理ボリュームに対する担当のコントローラを設定する必要があり、この設定作業のために多くの労力及び時間を要する問題があった。
【０００９】
本発明は以上の点を考慮してなされたもので、コントローラ間の負荷の偏り等に起因するデータ入出力処理性能の劣化を有効かつ簡易に防止し得る記憶制御装置及びその制御方法を提供しようとするものである。
【課題を解決するための手段】
【００１０】
かかる課題を解決するため本発明においては、記憶装置が提供する記憶領域上にデータを入出力するための論理的な領域でなる複数の論理ボリュームを作成し、作成した前記論理ボリュームをホスト計算機に提供すると共に、前記ホスト計算機からの要求に応じて、前記論理ボリュームに対するデータの入出力を制御する記憶制御装置において、それぞれ自己が担当する前記論理ボリュームに対するデータの入出力を制御する複数の制御部と、各前記制御部の負荷状況を監視する負荷監視部と、前記負荷監視部の監視結果に基づいて、各前記制御部の負荷を均等化するように、前記論理ボリュームの担当を他の前記制御部に切り替える切替え部とを備えることを特徴とする。
【００１１】
また本発明においては、記憶装置が提供する記憶領域上にデータを入出力するための論

理的な領域でなる複数の論理ボリュームを作成し、作成した前記論理ボリュームをホスト計算機に提供すると共に、前記ホスト計算機からの要求に応じて、前記論理ボリュームに対するデータの入出力を制御する記憶制御装置の制御方法において、前記記憶制御装置は、それぞれ自己が担当する前記論理ボリュームに対するデータの入出力を制御する複数の制御部を備え、各前記制御部の負荷状況を監視する第１のステップと、監視結果に基づいて、各前記制御部の負荷を均等化するように、前記論理ボリュームの担当を他の前記制御部に切り替える第２のステップとを備えることを特徴とする。

【発明の効果】

【００１２】

本発明によれば、各制御部の負荷を均等化することができ、制御部間の負荷の偏り等に起因するデータ入出力処理性能の劣化を有効かつ簡易に防止し得る記憶制御装置及びその制御方法を実現できる。

【発明を実施するための最良の形態】

【００１３】

以下図面について、本発明の一実施の形態を詳述する。

【００１４】

（１）ストレージシステムの構成

図１は、本実施の形態によるストレージシステムを示す。このストレージシステム１は、ホスト計算機２Ａ，２Ｂが記憶制御装置３を介して複数の記憶装置４Ａ～４Ｄに接続されることにより構成されている。

【００１５】

ホスト計算機２Ａ，２Ｂは、例えばＣＰＵ（Central Processing Unit）やメモリなどの情報処理資源を備えたコンピュータであり、具体的にはパーソナルコンピュータ、ワークステーション又はメインフレームなどから構成される。ホスト計算機２Ａ，２Ｂは、記憶制御装置３にアクセスするための通信ポート（例えばＬＡＮカードやホストバスアダプタに設けられたポート）を備えており、この通信ポートを介してデータの入出力要求コマンドを記憶制御装置３に送信することができる。

【００１６】

記憶制御装置３は、それぞれ異なるホスト計算機２Ａ，２Ｂが接続された０系及び１系のコントローラ６Ａ，６Ｂと、これらコントローラ６Ａ，６Ｂ間を通信可能な状態に接続するコントローラ間接続経路５とから構成される。

【００１７】

コントローラ間接続経路５としては、例えば最大８レーンで１レーン当たりのデータ転送量が２〔Gbit/sec〕と高速なデータ通信を実現するＰＣＩ（Peripheral Component Interconnect）－Ｅｘｐｒｅｓｓ規格に準拠したバスを適用することができる。０系及び１系のコントローラ６Ａ，６Ｂ間における各種データやコマンドのやり取りは、すべてこのコントローラ間接続経路５を介して行なわれる。

【００１８】

各コントローラ６Ａ，６Ｂは、それぞれ自コントローラに接続されたホスト計算機２Ａ，２Ｂからの要求に応じて記憶装置４Ａ～４Ｄに対するデータの入出力を制御する機能を有するもので、ホスト通信制御部１０Ａ，１０Ｂ、データ転送制御部１１Ａ，１１Ｂ、キャッシュメモリ１２Ａ，１２Ｂ、ブリッジ１３Ａ，１３Ｂ、ローカルメモリ１４Ａ，１４Ｂ、マイクロプロセッサ１５Ａ，１５Ｂ、記憶装置通信制御部１６Ａ，１６Ｂ及び記憶装置側スイッチ１７Ａ，１７Ｂなどから構成される。

【００１９】

このうちホスト通信制御部１０Ａ，１０Ｂは、ホスト計算機２Ａ，２Ｂとの間の通信を制御するインタフェースであり、複数の通信ポート２０Ａ，２０Ｂ及びホスト通信プロトコルチップ２１Ａ，２１Ｂを備える。

【００２０】

通信ポート２０Ａ，２０Ｂは、そのコントローラ６Ａ，６Ｂをネットワークやホスト計

算機２Ａ，２Ｂに接続するためのものであり、例えばＩＰ（Internet Protocol）アドレスやＷＷＮ（World Wide Name）などの固有のネットワークアドレスがそれぞれ割り当てられる。

【００２１】

ホスト通信プロトコルチップ２１Ａ，２１Ｂは、ホスト計算機２Ａ，２Ｂとの間の通信時におけるプロトコル制御を行う。このためホスト通信プロトコルチップ２１Ａ，２１Ｂとしては、例えばホスト計算機２Ａ，２Ｂとの間の通信プロトコルがファイバーチャネル（ＦＣ：Fibre Channel）プロトコルである場合にはファイバーチャネル変換プロトコルチップ、かかる通信プロトコルがｉＳＣＳＩプロトコルである場合にはｉＳＣＳＩプロトコルチップというように、ホスト計算機２Ａ，２Ｂとの間の通信プロトコルに適合したものが採用される。

【００２２】

またホスト通信プロトコルチップ２１Ａ，２１Ｂには、複数のマイクロプロセッサとの間で通信を行い得るマルチＣＰＵ機能が搭載されている。これによりホスト通信プロトコルチップ２１Ａ，２１Ｂは、必要時に０系のコントローラ６Ａ内のマイクロプロセッサ１５Ａと、１系のコントローラ６Ｂ内のマイクロプロセッサ１５Ｂとの双方と通信を行うことができる。

【００２３】

データ転送制御部１１Ａ，１１Ｂは、０系及び１系のコントローラ６Ａ，６Ｂ間のデータ転送と、０系又は１系のコントローラ６Ａ，６Ｂ内の各要素間でのデータ転送とを制御する機能を備える。またデータ転送制御部１１Ａ，１１Ｂは、自系のマイクロプロセッサ１５Ａ，１５Ｂからの指示により、ホスト計算機２Ａ，２Ｂから与えられたライトデータを、指定されたキャッシュメモリ１２Ａ，１２Ｂに複写（多重書き）する機能を備える。具体的には、０系又は１系のマイクロプロセッサ１５Ａ，１５Ｂは、自系のキャッシュメモリ１２Ａ，１２Ｂにデータを格納した場合、このデータを他系のキャッシュメモリ１２Ｂ，１２Ａにも書き込む（二重書き）。

【００２４】

またデータ転送制御部１１Ａ，１１Ｂは、後述する自系及び他系のローカルメモリ１４Ａ，１４Ｂ内の共有領域４２Ａ，４２Ｂ（図４）に格納された情報が常に同じとなるように、一方の共有領域４２Ａ，４２Ｂ内の情報を更新したときには、他方の共有領域４２Ｂ，４２Ａ内の情報も同様に更新する。

【００２５】

ブリッジ１３Ａ，１３Ｂは、自系のマイクロプロセッサ１５Ａ，１５Ｂ及びローカルメモリ１４Ａ，１４Ｂをそれぞれ自系のデータ転送制御部１１Ａ，１１Ｂに接続する中継装置であり、ホスト通信制御部１０Ａ，１０Ｂ、データ転送制御部１１Ａ，１１Ｂ、記憶装置通信制御部１６Ａ，１６Ｂ及びキャッシュメモリ１２Ａ，１２Ｂ間を接続するバスを流れるデータのうちの対応するデータのみを抽出してマイクロプロセッサ１５Ａ，１５Ｂやローカルメモリ１４Ａ，１４Ｂに転送する。

【００２６】

マイクロプロセッサ１５Ａ，１５Ｂは、それぞれ自系のコントローラ６Ａ，６Ｂ全体の動作制御を司る機能を有する。これらマイクロプロセッサ１５Ａ，１５Ｂは、後述のようにローカルメモリ１４Ａ，１４Ｂに保持されたライトコマンドやリードコマンドに応じて、予め自己に対して排他的に割り当てられた論理ボリューム（以下、これを「担当論理ボリューム」と呼ぶ）に対するデータの入出力等の処理を行う。

【００２７】

各マイクロプロセッサ１５Ａ，１５Ｂは、それぞれ複数のコア１８Ａ（１８ＡＸ，１８ＡＹ），１８Ｂ（１８ＢＸ，１８ＢＹ）を備える。図１の例においては、マイクロプロセッサ１５ＡはＸコア１８ＡＸとＹコア１８ＡＹとを備えており、マイクロプロセッサ１５ＢＸはコア１８ＢＸとＹコア１８ＢＹとを備えている。各コア１８Ａ，１８Ｂは適切なソフトウェア環境において複数のソフトウェア・スレッドを並列に実行する。各コア１８Ａ

，１８Ｂには、それぞれ自マイクロプロセッサの担当論理ボリュームのうちの１又は２以上の論理ボリュームが割り当てられる。そして各コア１８Ａ，１８Ｂは、自己に割り当てられた論理ボリュームに対するデータの入出力処理の制御等を担当する。

【００２８】

各コア１８Ａ，１８Ｂに対する論理ボリュームの割り当ては、各コア１８Ａ，１８Ｂの負荷状況や、ホスト計算機２Ａ，２Ｂから与えられる、論理ボリュームとその論理ボリュームに対するコマンドに応じた処理を担当すべきコア（以下、これを担当コアと呼ぶ）１８Ａ，１８Ｂとを指定した担当指定コマンドの受信等によって動的に変更することが可能である。また、各コア１８Ａ，１８Ｂに対する論理ボリュームの割り当ては、記憶制御装置３及びホスト計算機２Ａ，２Ｂ間の接続経路や、記憶制御装置３及び記憶装置４Ａ～４Ｄ間の接続経路等における障害発生の有無によっても動的に変更することができる。

【００２９】

ローカルメモリ１４Ａ，１４Ｂは、各種制御プログラムを格納するために用いられるほか、ホスト計算機２Ａ，２Ｂから与えられるリードコマンドやライトコマンド等の各種コマンドを一時的に保持するために用いられる。マイクロプロセッサ１５Ａ，１５Ｂは、このローカルメモリ１４Ａ，１４Ｂに格納されたリードコマンドやライトコマンドを、当該ローカルメモリ１４Ａ，１４Ｂに格納された順番で処理する。マイクロプロセッサ１５Ａ，１５Ｂ自体を制御するマイクロプログラムは同じコントローラ６Ａ，６Ｂ内のローカルメモリ１４Ａ，１４Ｂに保存される。

【００３０】

このローカルメモリ１４Ａ,１４Ｂは、ホスト計算機２Ａ，２Ｂから送信されるコマンドを順番に処理するためのキュー構造を備える。このキュー構造について、図２を用いて詳しく説明する。図２は、コントローラ６Ａにおけるキュー構造の概要を示している。図２では、１系のコントローラ６Ｂと、この１系のコントローラ６Ｂに接続されるコントローラ間接続経路５と、記憶装置４Ａ～４Ｄとが省略されている。ローカルメモリ６Ａに対する以下の説明は、他系のローカルメモリ６Ｂにそのまま適用可能である。

【００３１】

かかるキュー構造は、ホスト通信制御部１０Ａのホスト通信プロトコルチップ２１Ａが受領したホスト計算機２Ａからのコマンドを格納するための受領キュー２００と、プロセッサ１５ＡのＸコア１８ＡＸ又はＹコア１８ＡＹが処理すべきコマンドを格納するためのコマンドキュー２０２とを備える。

【００３２】

受領キュー２００は、ホスト通信制御部１０Ａの各通信ポート２０Ａにそれぞれ対応させて設けられており、これら通信ポート２０Ａに対してホスト計算機２Ａが発行したコマンドがその通信ポート２０Ａと対応付けられた受領キュー２００に格納される。例えば図２において、「ａ」という受領キュー２００には、ホスト計算機２Ａが「Ａ」という通信ポート２０Ａに対して発行したコマンドが格納される。また「ｂ」という受領キュー２００には、ホスト計算機２Ａが「Ｂ」という通信ポート２０Ａに対して発行したコマンドが格納され、「ｃ」という受領キュー２００には、ホスト計算機２Ａが「Ｃ」という通信ポート２０Ａに対して発行したコマンドが格納される。さらに「ｄ」という受領キュー２００には、ホスト計算機２Ａが通信ポート２０Ａに対して発行したコマンドが格納される。このような受領キュー２００へのコマンドの格納は、ホスト通信プロトコルチップ２１Ａによって行なわれる。

【００３３】

各受領キュー２００に格納されたコマンドは、その受領キュー２００と対応付けられた通信ポート２０Ａの担当として設定されたコア１８Ａによって処理される。すなわち、受領キュー２００ごとにその受領キュー２００に格納されたコマンドを処理するコア１８Ａが予め設定されている。例えば図２では、「Ａ」及び「Ｂ」という受領キュー２００に格納されたコマンドはＸコア１８ＡＸによって処理され、「Ｃ」及び「Ｄ」という受領キュー２００に格納されたコマンドはＹコア１８ＡＹによって処理される。ＸコアＸ１８ＡＸ及

びＹコア１８ＡＹは、自己が担当する通信ポート２０Ａと対応付けられた受領キュー２００からコマンドを順に読み出して、そのコマンドがいずれの論理ボリュームＬＵに対するコマンドであるか、いずれのコア１８Ａが実行すべきコマンドであるかなどの分析を行い、そのコマンドを対応するコマンドキュー２０２に格納する。

【００３４】

コマンドキュー２０２は、１つの論理ボリュームＬＵに対して複数設けられる。そして、論理ボリュームＬＵに対するコマンドがその論理ボリュームＬＵに対応付けられたコマンドキュー２０２に格納される。例えば図２の例では、ＬＵＮが「０」である論理ボリュームＬＵに対して「Ａ」及び「Ｂ」というコマンドキュー２０２が設けられ、当該論理ボリュームＬＵに対するコマンドがこれら「Ａ」及び「Ｂ」というコマンドキュー２０２に格納される。また、ＬＵＮが「１」である論理ボリュームＬＵには「Ｃ」及び「Ｄ」というコマンドキュー２０２が設けられ、当該論理ボリュームＬＵに対するコマンドがこれら「Ｃ」及び「Ｄ」というコマンドキュー２０２に格納される。ＬＵＮが「０」である論理ボリュームＬＵに対するリード又はライトコマンドの処理はＸコア１８ＡＸが担当し、ＬＵＮが「１」である論理ボリュームＬＵに対するコマンドの処理はＹコア１８ＡＹが担当する。

【００３５】

各コマンドキュー２０２へのコマンドの移動がどのコア１８Ａによって行なわれるか、さらに各コマンドキュー２０２に格納されたコマンドの処理を担当するコア１８Ａは、ローカルメモリ１４Ａに格納された特性表２０４に従う。

【００３６】

このように、本実施の形態においては、論理ボリュームＬＵごとにコマンド処理を行うコア１８Ａが定められているため、Ｘコア１８ＡＸとＹコア１８ＡＹとが同一の論理ボリュームＬＵへの処理について競合することなく、Ｘコア１８ＡＸとＹコア１８ＡＹとがそれぞれ独立して異なる論理ボリュームＬＵへのＩＯアクセスを処理できる。かくするにつき、例えば、１つの論理ボリュームＬＵに対する処理が２つのコア１８Ａで競合し、処理の負荷が掛かっていない一方のコア１８Ａが他方のコア１８Ａの処理を待たなければならないといった状況を避けることができるため、コントローラ６Ａ，６Ｂ全体としてのホスト計算機２Ａ，２Ｂからのコマンドを高速・高効率で処理することができる。

【００３７】

図１に戻りさらに説明を続ける。キャッシュメモリ１２Ａ，１２Ｂは、主としてホスト計算機２Ａ，２Ｂ及び記憶装置４Ａ～４Ｄ間や、０系及び１系のコントローラ６Ａ，６Ｂ間で転送されるデータを一時的に記憶するために用いられる。

【００３８】

記憶装置通信制御部１６Ａ，１６Ｂは、各記憶装置４Ａ～４Ｄとの通信を制御するインタフェースであり、記憶装置通信プロトコルチップ２２Ａ，２２Ｂを備える。この記憶装置通信プロトコルチップ２２Ａ，２２Ｂとしては、例えば記憶装置４Ａ～４ＤとしてＦＣハードディスクドライブを適用する場合にはＦＣプロトコルチップ、ＳＡＳハードディスクドライブを適用する場合にはＳＡＳプロトコルチップを適用できる。また記憶装置４Ａ～４ＤとしてＳＡＴＡハードディスクドライブを適用する場合には、記憶装置通信プロトコルチップ２２Ａ，２２ＢとしてＦＣプロトコルチップ又はＳＡＳプロトコルチップ、ＳＡＴＡハードディスクドライブを適用する場合にはＳＡＴＡプロトコル変換チップを介して接続する構成とすることもできる。

【００３９】

記憶装置側スイッチ１７Ａ，１７Ｂは、通信相手となる記憶装置４Ａ～４Ｄを切り替えるためのスイッチであり、例えばＳＡＳ－ＥｘｐａｎｄｅｒやＦＣループスイッチを適用することができる。また記憶装置側スイッチ１７Ａ，１７Ｂに代えて、例えばＦＣループで記憶装置４Ａ～４Ｄと接続する構成とすることもできる。

【００４０】

記憶装置４Ａ～４Ｄは、例えば複数のハードディスクドライブ、具体的にはＦＣハード

ディスクドライブ、ＳＡＳハードディスクドライブ及び又はＳＡＴＡハードディスクドライブ等が搭載されたディスクアレイ装置として構成される。これら複数のハードディスクドライブが提供する記憶領域上には、データを読み書きするための論理的な記憶領域である論理ボリュームを複数設定することができる。

【００４１】

論理ボリュームについては、複数の記憶装置４Ａ～４Ｄ間でＲＡＩＤ技術を適用することで信頼性及び応答性を高めることができる。具体的には、「ＲＡＩＤ０」、「ＲＡＩＤ１」、「ＲＡＩＤ３」、「ＲＡＩＤ５」、「ＲＡＩＤ６」又は「ＲＡＩＤ０＋１」等の様々なＲＡＩＤレベルを論理ボリュームごとに設定することができる。

【００４２】

さらに記憶装置４Ａ～４Ｄに搭載する記憶デバイスとしては、ハードディスクドライブに代えて、フラッシュメモリ等の半導体メモリや、光ディスク装置を適用することができる。またフラッシュメモリとしては、安価で書込み速度が比較的低速であり、かつライト回数制限が低い第１のタイプのものと、高価でライトコマンド処理が当該第１のタイプよりも高速であり、かつライト回数制限が第１のタイプよりも高い第２のタイプとのいずれのタイプのフラッシュメモリを用いることもできる。さらにこれら第１及び又は第２のタイプのフラッシュメモリと、ハードディスクドライブとを記憶デバイスとして記憶装置４Ａ～４Ｄに混載する構成とすることもできる。

【００４３】

次に、ストレージシステム１の記憶制御装置３において行われるコマンド処理について説明する。このストレージシステム１の場合、記憶制御装置３を構成する０系及び１系の各コントローラ６Ａ，６Ｂのメモリ空間には、それぞれ自系のコントローラ６Ａ，６Ｂ内のメモリ領域に加えて、他系のコントローラ６Ｂ，６Ａ内のメモリ領域もマッピングされている。これにより本実施の形態においては、０系及び１系のコントローラ６Ａ，６Ｂが他系のコントローラ６Ｂ，６Ａ内のメモリ領域にも直接アクセスできるようになされている。

【００４４】

そして、０系及び１系のコントローラ６Ａ，６Ｂ内のマイクロプロセッサ１５Ａ，１５Ｂの各コア１８ＡＸ，１８ＢＸは、ホスト計算機２Ａ，２Ｂから他系のコントローラ６Ｂ，６Ａに割り当てられている論理ボリュームを対象とするライトコマンドやリードコマンドが与えられたときには、かかるライトコマンド又はリードコマンドを他系のローカルメモリ１４Ｂ，１４Ａに書き込むようにして当該他系のコントローラ６Ｂ，６Ａに転送する。

【００４５】

このようにこのストレージシステム１においては、ホスト計算機２Ａ，２Ｂから０系又は１系のコントローラ６Ａ，６Ｂに対して他系のコントローラ６Ｂ，６Ａが実行すべきライトコマンドやリードコマンドが与えられたときに、そのライトコマンドやリードコマンドを他系のローカルメモリ１４Ｂ，１４Ａに直接書き込むようにして中継するため、当該ライトコマンド又はリードコマンドの受け渡しのために０系及び１系のコントローラ６Ａ，６Ｂ同士が通信を行なう必要がなく、その分迅速なライトコマンド処理又はリードコマンド処理を行うことができる。

【００４６】

図３（Ａ）は、０系（CTL0）及び１系（CTL1）のマイクロプロセッサ１５Ａ，１５Ｂがそれぞれ認識するコントローラ６Ａ，６Ｂにおけるメモリ空間でのメモリマップを示す。また図３（Ｂ）は、０系及び１系のホスト通信プロトコルチップ１０Ａ，１０Ｂがそれぞれ認識するメモリ空間のメモリマップを示す。

【００４７】

図３（Ａ）において、０系のマイクロプロセッサ１５Ａが認識するメモリ空間は、「CTL0 Memory Map（CTL0プロセッサ側）」のフィールド３０Ｂに登録されたメモリ又はレジスタがそれぞれ提供する記憶領域（「ADR」フィールド３０Ａに格納されたアドレス範囲

の領域）からなるメモリ空間であり、１系のマイクロプロセッサ１５Ｂが認識するメモリ領域は、「CTL1 Memory Map（CTL1 プロセッサ側）」のフィールド３０Ｃに登録されたメモリ又はレジスタがそれぞれ提供する記憶領域（「ADR」フィールド３０Ａに格納されたアドレス範囲の領域）からなるメモリ空間である。

【００４８】

また図３（Ｂ）において、０系のホスト通信プロトコルチップ１０Ａが認識するメモリ空間は、「CTL0 Memory Map（CTL0 ホスト通信プロトコルチップ側）」のフィールド３１Ｂに登録されたメモリやレジスタがそれぞれ提供する記憶領域（「ADR」フィールド３１Ａに格納されたアドレス範囲の領域）からなるメモリ空間であり、１系のホスト通信プロトコルチップ１０Ｂが認識するメモリ領域は、「CTL1 Memory Map（CTL1 ホスト通信プロトコルチップ側）」のフィールド３１Ｃに登録されたメモリがそれぞれ提供する記憶領域（「ADR」フィールド３１Ａに格納されたアドレス範囲の領域）からなるメモリ空間である。

【００４９】

図３（Ａ）及び（Ｂ）のメモリマップのデータのうち、自系のローカルメモリ１４Ａ，１４Ｂのアドレス（「0x0_00000000-0x0_7FFFFFFF」）及び自系のブリッジ１３Ａ，１３Ｂのアドレス（「0x0_80000000-0x0_FFFFFFFF」）は、それぞれ自系のマイクロプロセッサ１５Ａ，１５Ｂに接続された図示しないフラッシュメモリに第１のアドレス情報として予め格納され、図３（Ａ）のこれ以外のアドレスは自系のデータ転送制御部１１Ａ，１１Ｂ内に設けられた図示しないレジスタに第２のアドレス情報として予め格納されている。

【００５０】

そして、０系及び１系の各マイクロプロセッサ１５Ａ，１５Ｂ及びホスト通信プロトコルチップ２１Ａ，２１Ｂは、それぞれこれらフラッシュメモリ及びレジスタに格納された第１及び第２のアドレス情報に基づいてメモリ空間を図３（Ａ）や図３（Ｂ）のようにそれぞれ認識でき、この認識結果に基づいて、自系内のローカルメモリ１４Ａ，１４Ｂ及びキャッシュメモリ１２Ａ，１２Ｂ等に加えて、他系のローカルメモリ１４Ａ，１４Ａやキャッシュメモリ１２Ｂ，１２Ａなどにもアクセスすることができる。

【００５１】

図４（Ａ）及び（Ｂ）は、それぞれ０系及び１系のコントローラ６Ａ，６Ｂにおけるローカルメモリ１４Ａ，１４Ｂのメモリ構成をそれぞれ示す。この図４（Ａ）及び（Ｂ）に示すように、コントローラ６Ａ，６Ｂは、かかる０系及び１系のローカルメモリ１４Ａ，１４Ｂの記憶領域を、自系コントローラ受領コマンド格納領域４０Ａ，４０Ｂ、他系コントローラ受領コマンド格納領域４１Ａ，４１Ｂ及び共有領域４２Ａ，４２Ｂの別個の３つの記憶領域に分けて管理している。自系コントローラ受領コマンド領域４０Ａ,４０Ｂは、さらに自Ｘコア用領域４０ＡＸ，４０ＢＸと自Ｙコア用領域４０ＡＹ，４０ＢＹとに分けて管理され、他系コントローラ受領コマンド格納領域４１Ａ,４１Ｂも同様に他Ｘコア用領域４１ＡＸ，４１ＢＸと他Ｙコア用領域４１ＡＹ，４１ＢＹとに分けて管理される。

【００５２】

このうち自系コントローラ受領コマンド格納領域４０Ａ，４０Ｂでは、自Ｘコア用領域４０ＡＸ，４０ＢＸ内に制御情報格納領域４０ＡＸＡ，４０ＢＸＡが設けられると共に、自Ｙコア用領域４０ＡＹ，４０ＢＹ内に制御情報格納領域４０ＡＹＡ，４０ＢＹＡが設けられており、自系のホスト通信制御部１０Ａ，１０Ｂが受信したライトコマンドやリードコマンド等の各種コマンドがこれら制御情報格納領域４０ＡＸＡ，４０ＢＸＡ，４０ＡＹＡ，４０ＢＹＡに格納される。

【００５３】

また他系コントローラ受領コマンド格納領域４１Ａ，４１Ｂでは、他Ｘコア用領域４１ＡＸ，４１ＢＸ内に制御情報格納領域４１ＡＸＡ，４１ＢＸＡが設けられると共に、他Ｙコア用領域４１ＡＹ，４１ＢＹ内に制御情報格納領域４１ＡＹＡ，４１ＢＹＡが設けられており、他系のコントローラ６Ｂ，６Ａのホスト通信制御部１０Ｂ，１０Ａが受信した一般的なコマンド及び制御情報がこれら制御情報格納領域４１ＡＸＡ，４１ＡＹＢ，４１Ｂ

ＸＡ，４１ＢＹＡに格納される。

【００５４】

さらに他系コントローラ受領コマンド格納領域４１Ａ，４１Ｂでは、他Ｘコア用領域４１ＡＸ，４１ＢＸ内に実行処理格納領域４１ＡＸＢ，４１ＢＸＢが設けられると共に、他Ｙコア用領域４１ＡＹ，４１ＢＹ内に実行処理格納領域４１ＡＹＢ，４１ＢＹＢが設けられており、他系コントローラ６Ｂ，６Ａの担当論理ボリュームに関する処理を自系コントローラ６Ａ，６Ｂが新たに担当する際に、他系コントローラ６Ｂ，６Ａ内で受信コマンドが解析された後の実行処理内容がこれらの実行処理格納領域４１ＡＸＢ，４１ＢＸＢ，４１ＡＹＢ，４１ＢＹＢに格納される。これは他系コントローラ６Ｂのマイクロプロセッサ１５Ａ，１５Ｂが障害などによって閉塞したような場合に行なわれる。

【００５５】

さらに共有領域３２Ａ，３２Ｂには、担当管理テーブル４３Ａ，４３Ｂ、内部コピーペア管理テーブル４４Ａ，４４Ｂ、スナップショットペア管理テーブル４５Ａ，４５Ｂ、非同期リモートコピーペア管理テーブル４６Ａ，４６Ｂ、リモートコピーペア管理テーブル４７Ａ，４７Ｂ及びプール管理テーブル４８Ａ，４８Ｂが格納される。これらテーブルの詳細については、後述する。

【００５６】

（２）負荷分散機能

（２－１）各種テーブルの構成

次に、かかる記憶制御装置３に搭載された負荷分散機能について説明する。本実施の形態の場合、記憶制御装置３には、例えばペアを構成する論理ボリューム同士のように、同じコア１８Ａ，１８Ｂが担当することが望ましい相互に関連する複数の論理ボリュームについては、コピーペアの作成時に、これら複数の論理ボリュームの担当を一括して１つのコア１８Ａ，１８Ｂに切り替える第１の負荷分散機能が搭載されている。

【００５７】

また本実施の形態の場合、記憶制御装置３には、各コントローラ６Ａ，６Ｂのマイクロプロセッサ１５Ａ，１５Ｂの各コア１８Ａ，１８Ｂの負荷状況を監視し、各コア１８Ａ，１８Ｂの負荷を均等化するように、負荷の大きいコア１８Ａ，１８Ｂに割り当てられた論理ボリュームの担当を、負荷の少ないコア１８Ａ，１８Ｂに切り替える第２の負荷分散機能が搭載されている。この際、記憶制御装置３は、同じコア１８Ａ，１８Ｂが担当することが望ましい相互に関連する複数の論理ボリュームについては、第１の負荷分散機能と同様に、これら複数の論理ボリュームの担当を一括して１つのコア１８Ａ，１８Ｂに切り替える。

【００５８】

このような第１及び第２の負荷分散機能を実現するための手段として、記憶制御装置３の各コントローラ６Ａ，６Ｂのローカルメモリ１４Ａ，１４Ｂにおける共有領域４２Ａ，４２Ｂには、それぞれ上述のように担当管理テーブル４３Ａ，４３Ｂ、内部コピーペア管理テーブル４４Ａ，４４Ｂ、スナップショットペア管理テーブル４５Ａ，４５Ｂ、非同期リモートコピーペア管理テーブル４６Ａ，４６Ｂ、リモートコピーペア管理テーブル４７Ａ，４７Ｂ及びプール管理テーブル４８Ａ，４８Ｂが格納される。

【００５９】

担当管理テーブル４３Ａ，４３Ｂは、記憶装置４Ａ～４Ｄ内に形成された各論理ボリュームの担当コントローラ６Ａ，６Ｂ及び担当コアを管理するためのテーブルであり、論理ボリュームごとにその生成時に作成される。各コア１８Ａ，１８Ｂは、これら各論理ボリュームの担当管理テーブル４３Ａ，４３Ｂに基づいて、自コアがその論理ボリュームの担当であるか否かを認識する。

【００６０】

この担当管理テーブル４３Ａ，４３Ｂは、図５に示すように、「担当コントローラ」欄４３ＣＡ及び「担当コア」欄４３ＣＢから構成される。このうち「担当コントローラ」欄４３ＣＡには、その論理ボリュームに対するリードコマンドやライトコマンドなどのコマ

ンドに応じた処理（コマンド処理）を担当するコントローラ（以下、これを担当コントローラと呼ぶ）６Ａ，６Ｂに付与された識別番号が格納される。また「担当コア」欄４３ＣＢには、そのコントローラ６Ａ，６Ｂ内のマイクロプロセッサ１５Ａ，１５Ｂが備える１８Ａ，１８Ｂのうち、その論理ボリュームの担当コア１８Ａ，１８Ｂに付与された識別番号が格納される。

【００６１】

内部コピーペア管理テーブル４４Ａ，４４Ｂは、内部コピーを行う論理ボリュームのペア（以下、これを内部コピーペアと呼ぶ）を管理するためのテーブルであり、論理ボリュームごとに作成される。なお、内部コピーとは、同じ記憶装置４Ａ～４Ｄ内の論理ボリューム間で行われるデータのコピーのことをいう。この内部コピーペア管理テーブル４４Ａ，４４Ｂは、図６に示すように、「ペア状態」欄４４ＣＡ、「ペア属性」欄４４ＣＢ、「相手ＬＵＮ」欄４４ＣＣ、「ミラー番号」欄ＣＤ、「コピースピード」欄４４ＣＥ、「グループ種別」欄４４ＣＦ、「グループＩＤ」欄４４ＣＧ及び「副ＬＵＮ」欄４４ＣＨから構成される。

【００６２】

「ペア状態」欄４４ＣＡには、その論理ボリュームが形成するコピーペアのペア状態を表すコードが格納される。このようなペア状態としては、その論理ボリュームについて内部コピーペアのペア定義がされていない状態（「SMPL」）と、初期コピー中又は再同期コピー中の状態（「COPY」）と、内部コピーペアの正側及び副側のデータ内容が一致した状態（「PAIR」）と、内部コピーペアの正側の内容が副側に反映されない状態（「PSUS」）となどがある。

【００６３】

また「ペア属性」欄４４ＣＢには、その論理ボリュームがコピーペアを形成している場合に、その論理ボリュームのペア属性（正側又は副側）を表すコードが格納される。さらに「相手ＬＵＮ」欄４４ＣＣには、その論理ボリュームが内部コピーペアの副側に設定されている場合に、正側の論理ボリュームの論理ボリューム番号が格納される。この「相手ＬＵＮ」欄４４ＣＣは、その論理ボリュームが内部コピーペアの正側であるときには参照できない。

【００６４】

さらに本実施の形態の場合、内部コピーペアについては、最大８個までの副側を設定できる。そして、その論理ボリュームの副側としての順位を表す番号が「ミラー番号」欄４４ＣＤに格納される。また「コピースピード」欄４４ＣＥには、内部コピーにおける初期コピーや再同期コピーの際のコピースピードとして設定されたスピードに応じたコードが格納される。

【００６５】

さらに本実施の形態の場合、複数の内部コピーペアを１つのグループとして設定できる。そして「グループ種別」欄４４ＣＦには、その設定が有効か無効かを表すコードが格納される。また、その論理ボリュームがいずれかのグループに属する場合には、そのグループに付与されたＩＤが「グループＩＤ」欄４４ＣＧに格納される。

【００６６】

なお、「ミラー番号」欄４４ＣＤ、「コーススピード」欄４４ＣＥ、「グループ種別」欄４４ＣＦ及び「グループＩＤ」欄４４ＣＧに格納された数値は、対応する論理ボリュームが内部コピーペアの副側に設定されているときにのみ有効値となる。

【００６７】

さらに「副ＬＵＮ」欄４４ＣＨには、その論理ボリュームが内部コピーペアの正側に設定されている場合に、副側の論理ボリュームの論理ボリューム番号が格納される。なお、上述のように本実施の形態においては、内部コピーペアの副側を最大８個まで設定することができるため、「副ＬＵＮ」欄４４ＣＨには、副側として設定された各論理ボリュームの論理ボリューム番号がそれぞれ格納されることとなる。

【００６８】

一方、スナップショットペア管理テーブル４５Ａ，４５Ｂは、論理ボリュームと、その論理ボリュームのスナップショットのデータ（更新前のデータ）を格納するプールボリュームとのペア（以下、これをスナップショットペアと呼ぶ）を管理するためのテーブルであり、論理ボリュームごとに作成される。このスナップショットペア管理テーブル４５Ａ，４５Ｂは、図７に示すように、「ペア状態」欄４５ＣＡ、「ペア属性」欄４５ＣＢ、「相手ＬＵＮ」欄４５ＣＣ、「副ＬＵ数」欄４５ＣＤ、「副ＬＵＮ」欄４５ＣＥ、「ミラー番号」欄４５ＣＦ、「グループＩＤ種別」欄４５ＣＧ、「グループＩＤ」欄４５ＣＨ及び「プールボリューム番号」欄４５Ｉから構成される。

【００６９】

「ペア状態」欄４５ＣＡには、その論理ボリュームのスナップショットペアに関するペア状態を表すコードが格納される。また「ペア属性」欄４５ＣＢには、その論理ボリュームが他の論理ボリュームとスナップショットペアに設定されている場合に、当該論理ボリュームのペア属性（正側又は副側）を表すコードが格納される。

【００７０】

さらに「相手ＬＵＮ」欄４５ＣＣには、その論理ボリュームがスナップショットペアの副側（つまりプールボリューム）に設定されている場合に、正側の論理ボリュームの論理ボリューム番号が格納される。この「相手ＬＵＮ」欄４５ＣＣは、その論理ボリュームがスナップショットペアの正側であるときには参照できない。

【００７１】

さらに「副ＬＵ数」欄４５ＣＤには、その論理ボリュームと共にスナップショットペアを形成している副側の論理ボリューム（つまりプールボリューム）の数が格納される。本実施の形態の場合、かかるスナップショットペアの副側として、最大３２個の論理ボリュームを設定することができる。また「副ＬＵＮ」欄４５ＣＥには、その論理ボリュームが形成するスナップショットペアの副側の論理ボリュームの論理ボリューム番号が格納される。この「副ＬＵＮ」欄４５ＣＥは、対応する論理ボリュームのペア属性が正側である場合にのみ参照することができる。

【００７２】

さらに「ミラー番号」欄４５ＣＦには、その論理ボリュームがスナップショットペアの副側である場合に、当該論理ボリュームの副側としての順位を表す番号が格納される。さらに本実施の形態の場合、複数のスナップショットペアを１つのグループとして設定できる。そして「グループＩＤ種別」欄４５ＣＧには、その設定が有効か無効かを表すコードが格納される。また、その論理ボリュームがいずれかのグループに属する場合には、そのグループに付与されたＩＤが「グループＩＤ」欄４５ＣＨに格納される。

【００７３】

さらに「ＰＯＯＬ番号」欄４５ＣＩには、その論理ボリュームがスナップショットペアの正側である場合に、副側の論理ボリューム（プールボリューム）の識別番号であるプールボリューム番号が格納される。この「ＰＯＯＬ番号」欄４５ＣＩは、対応する論理ボリュームがスナップショットペアの正側である場合にのみ参照することができる。

【００７４】

非同期リモートコピーペア管理テーブル４６Ａ，４６Ｂは、非同期リモートコピーを行う論理ボリュームのペア（以下、これを非同期リモートコピーペアと呼ぶ）を管理するためのテーブルであり、論理ボリュームごとに作成される。ここで非同期リモートコピーとは、正側及び副側の各論理ボリュームがそれぞれ異なる記憶装置４Ａ～４Ｄ内に作成され、正側の論理ボリュームに対するデータの書き込みとは非同期に正側及び副側の各論理ボリューム間でデータをコピーするコピー方法である。この場合、正側及び副側の記憶装置４Ａ～４Ｄでは、それぞれ正側の論理ボリュームや副側の論理ボリュームのスナップショットを定期的に作成する。

【００７５】

かかる非同期リモートコピーペア管理テーブル４６Ａ，４６Ｂは、「ペア状態」欄４６ＣＡ、「ペア属性」欄４６ＣＢ、「コンシステンシグループＩＤ」欄４６ＣＣ、「相手Ｌ

ＵＮ」欄４６ＣＤ、「コピースピード」欄４６ＣＥ及び「プール番号」欄４６ＣＦから構成される。

【００７６】

このうち「ペア状態」欄４６ＣＡには、対応する論理ボリュームの非同期リモートコピーに関するペア状態を表すコードが格納される。また「ペア属性」欄４６ＣＢには、その論理ボリュームが他の論理ボリュームと非同期リモートコピーペアに設定されている場合に、当該論理ボリュームのペア属性（正側又は副側）を表すコードが格納される。

【００７７】

さらに「コンシステンシグループＩＤ」欄４６ＣＣには、その論理ボリュームが形成するコンシステンシグループに付与された識別番号（コンシステンシグループ番号）が格納される。なお、コンシステンシグループとは、例えばデータベースの目次データが格納された論理ボリュームと、そのデータベースの検索データが格納された論理ボリュームと、ログデータが格納された論理ボリュームとなどのように、関連するデータがそれぞれ格納された複数の論理ボリュームをまとめたグループをいう。

【００７８】

「相手ＬＵＮ」欄４６ＣＤには、その論理ボリュームが非同期リモートコピーペアの副側に設定されている場合に、正側の論理ボリュームの論理ボリューム番号が格納され、「コピースピード」欄４６ＣＥには、非同期リモートコピーにおける初期コピーや再同期コピーの際のコピースピードとして設定されたスピードに応じたコードが格納される。さらに「プールボリューム番号」欄４６ＣＦには、その論理ボリュームのスナップショットのデータが保存されるプールボリュームのプールボリューム番号が格納される。

【００７９】

リモートコピーペア管理テーブル４７Ａ，４７Ｂは、リモートコピーを行う論理ボリュームのペア（以下、これをリモートコピーペアと呼ぶ）を管理するためのテーブルであり、論理ボリュームごとに作成される。ここでリモートコピーとは、正側の記憶装置４Ａ～４Ｄ内に設けられた正側の正側の論理ボリュームに書き込まれたデータを、当該論理ボリュームへのデータ書き込みと同期して、副側の記憶装置４Ａ～４Ｄ内に作成された副側の論理ボリュームにコピーするコピー方法である。このとき正側及び副側の記憶装置４Ａ～４Ｄでは、それぞれ正側又は副側の論理ボリュームのスナップショットを定期的に作成する。

【００８０】

なお、かかるリモートコピーペア管理テーブル４７Ａ，４７Ｂは、非同期リモートコピーペア管理テーブル４６Ａ，４６Ｂと同様の構成を有するものであるため、ここでの説明は省略する。

【００８１】

他方、プール管理テーブル４８Ａ，４８Ｂは、プールボリュームを形成する論理ボリュームを管理するためのテーブルであり、いずれかのプールボリュームに属する論理ボリュームごとに作成される。このプール管理テーブル４８Ａ，４８Ｂは、図９に示すように、「論理ボリューム番号」欄４８ＣＡ、「先頭プールフラグ」欄４８ＣＢ及び「所属プール番号」欄４８ＣＣから構成される。

【００８２】

そして「論理ボリューム番号」欄４８ＣＡには、対応するプールボリュームに所属する各論理ボリュームの論理ボリューム番号が格納され、「先頭プールフラグ」欄４８ＣＢには、その論理ボリュームがそのプールボリュームにおいける先頭部分のアドレスが割り振られた論理ボリュームであるか否かを表すコードが格納される。また「所属プールボリューム番号」欄４８ＣＣには、その論理ボリュームが所属するプールボリュームのプール番号が格納される。

【００８３】

（２－２）第１の負荷分散機能に関するコアの処理

図１０は、上述の第１の負荷分散機能に関する各コア１８Ａ，１８Ｂの具体的な処理内

容を示している。

【００８４】

各コア１８Ａ，１８Ｂは、ホスト計算機２Ａ，２Ｂから内部コピーペア、スナップショットペア、リモートコピーペア又は非同期リモートコピーペアを形成すべき旨のコマンドが与えられると、ローカルメモリ１４Ａ，１４Ｂに格納された対応する制御プログラムに基づいてこの図１０に示すペア形成処理を開始し、まず、正側となる論理ボリュームの担当管理テーブル４３Ａ，４３Ｂと、副側となる論理ボリュームの担当管理テーブル４３Ａ，４３Ｂとを参照して、これら正側の論理ボリューム及び副側の論理ボリュームの担当コントローラ６Ａ，６Ｂ及び担当コア１８Ａ，１８Ｂの識別番号を取得する（ＳＰ１）。

【００８５】

続いてコア１８Ａ，１８Ｂは、正側又は副側の論理ボリュームの担当を変更すべきか否かを判断する（ＳＰ２）。具体的に、コア１８Ａ，１８Ｂは、かかるペア作成コマンドが内部コピーペアを作成すべき旨のコマンドであった場合には、正側の論理ボリュームの担当コア１８Ａ，１８Ｂと、副側の論理ボリュームの担当コア１８Ａ，１８Ｂとが同じであるか否かを判断する。またコア１８Ａ，１８Ｂは、かかるペア作成コマンドがスナップショットペアを作成すべき旨のコアンドであった場合には、正側の論理ボリュームの担当コア１８Ａ，１８Ｂと、そのスナップショットのデータを格納する副側のプールボリュームの担当コア１８Ａ，１８Ｂとが同じであるか否かを判断する。さらにコア１８Ａ，１８Ｂは、かかるペア作成コマンドが非同期リモートコピーペア又はリモートコピーペアを作成すべき旨のコマンドであった場合には、自己が担当する正側又は副側の論理ボリュームの担当コア１８Ａ，１８Ｂと、その論理ボリュームのスナップショットのデータを格納するプールボリュームの担当コア１８Ａ，１８Ｂとが同じであるか否かを判断する。

【００８６】

そしてコア１８Ａ，１８Ｂは、この判断において肯定結果を得るとステップＳＰ４に進み、これに対して否定結果を得ると、対応する正側又は副側の論理ボリュームの担当コア１８Ａ，１８Ｂを副側又は正側の論理ボリュームの担当コア１８Ａ，１８Ｂに切り替える（ＳＰ３）。

【００８７】

具体的にコア１８Ａ，１８Ｂは、かかるペア作成コマンドが内部コピーペアを作成すべき旨のコマンドであった場合には、論理ボリューム切替えテーブル４３Ａ，４３Ｂ（図５）上の副側の論理ボリュームの担当コントローラ６Ａ，６Ｂ及び担当コア１８Ａ，１８Ｂを、正側の論理ボリュームの担当コントローラ６Ａ，６Ｂ及び担当コア１８Ａ，１８Ｂと同じ担当コントローラ６Ａ，６Ｂ及び担当コア１８Ａ，１８Ｂに変更する。またコア１８Ａ，１８Ｂは、かかるペア作成コマンドがスナップショットペアを作成すべき旨のコマンドであった場合には、論理ボリューム切替えテーブル４３Ａ，４３Ｂ上の正側の論理ボリュームの担当コントローラ６Ａ，６Ｂ及び担当コア１８Ａ，１８Ｂを、副側であるプールボリュームの論理ボリュームの担当コントローラ６Ａ，６Ｂ及び担当コア１８Ａ，１８Ｂと同じ担当コントローラ６Ａ，６Ｂ及び担当コア１８Ａ，１８Ｂに変更する。さらにコアは、かかるペア作成コマンドが非同期リモートコピーペア又はリモートコピーペアを作成すべき旨のコマンドであった場合には、自己が担当する正側又は副側の論理ボリュームの担当コントローラ６Ａ，６Ｂ及び担当コア１８Ａ，１８Ｂを、その論理ボリュームのスナップショットのデータを格納するプールボリュームの担当コントローラ６Ａ，６Ｂ及び担当コア１８Ａ，１８Ｂに変更する。

【００８８】

そしてコア１８Ａ，１８Ｂは、この後、対応する記憶装置４Ａ～４Ｄを制御して正側の論理ボリュームのデータを副側の論理ボリュームに初期コピーさせる（内部コピーペアの場合）などの所定のペア形成処理を実行し、この後このペア形成処理を終了する。

【００８９】

（２－３）第２の負荷分散機能に関するコアの処理

一方、図１１は、上述の第２の負荷分散機能に関する各コア１８Ａ，１８Ｂの具体的な

処理内容を示すフローチャートである。各コア１８Ａ，１８Ｂは、ローカルメモリ１４Ａ，１４Ｂに格納された対応する制御プログラムに基づき、周期的（以下、５秒毎とする）にこの図１１に示す第２の負荷分散処理を実行する。

【００９０】

実際上、コア１８Ａ，１８Ｂは、この第２の負荷分散処理を開始すると、まず、ストレージシステム１が自コアに割り当てられた論理ボリュームを他のコア１８Ａ，１８Ｂの担当に切り替えるための前提条件を満たしているか否かを判断する（ＳＰ１０）。

【００９１】

本実施の形態の場合、かかる前提条件として、性能モニタの対象項目に対する監視設定がすべてオン（つまりすべての対象項目について性能データを取得する設定）になっていることや、いずれかの論理ボリュームの担当コア１８Ａ，１８Ｂを切り替えてから１分を超えていることなどが規定されている。前者は、負荷分散処理を行う場合、かかる対象項目すべてを監視する必要があるからであり、後者は、頻繁に論理ボリュームの担当コア１８Ａ，１８Ｂの切り替えが発生すると、担当コア１８Ａ，１８Ｂの切替え処理時におけるオーバヘッドにより逆に性能が劣化するおそれがあるからである。なお、性能モニタは、マイクロプロセッサ１５Ａ，１５Ｂが常時行っているデータ入出力性能に関する監視処理であり、この性能モニタの対象項目としては、自マイクロプロセッサ１５Ａ，１５Ｂ内の各コア１８Ａ，１８Ｂがそれぞれ担当している各論理ボリュームごとの１秒当たりのデータ入出力回数（ＩＯＰＳ）や、自マイクロプロセッサ１５Ａ，１５Ｂ内の各コア１８Ａ，１８Ｂの１秒当たりの使用率（＝処理時間／モニタ時間）などがある。

【００９２】

コア１８Ａ，１８Ｂは、ステップＳＰ１０の判断において否定結果を得るとこの第２の負荷分散処理を終了し、これに対して肯定結果を得ると、かかる性能モニタにより得られた各対象項目の性能データの中から自コアの負荷状況を表すデータを取得し、このデータに基づいて自コアの負荷状況を数値化する（ＳＰ１１）。具体的に、コア１８Ａ，１８Ｂは、かかる性能モニタにより得られた自コアの１秒ごとの使用率を５秒ごとに平均化したもの（以下、これを平均使用率と呼ぶ）を、過去３分間分算出する。

【００９３】

その後コア１８Ａ，１８Ｂは、ステップＳＰ１１において算出した自コアの過去３分間の平均使用率が、予め定められた第１の切替え条件を満たしているか否かを判断する（ＳＰ１２）。本実施の形態の場合、かかる第１の切替え条件として、第１に、過去３分間における５秒毎の自コアの平均使用率（合計３６個）のうち、値が８０％以上であるものが８個以上あること、第２に、過去３分間における１分ごとのすべての期間において自コアの平均使用率が８０％を超える期間がなかったこと、第３に、過去３分間において５秒毎の自コアの平均使用率の平均が７０％以上であることなどが規定されている。

【００９４】

そしてコア１８Ａ，１８Ｂは、自コアの過去３分間における５秒後との平均使用率がかかる第１の切替え条件のいずれも満たさない場合には、この第２の負荷分散処理を終了する。

【００９５】

これに対してコア１８Ａ，１８Ｂは、自コアの過去３分間における５秒ごとの平均使用率がかかる第１の切替え条件のいずれか１つでも満たす場合には、他のコア１８Ａ，１８Ｂ（他のコントローラ６Ｂ，６Ａ内のマイクロプロセッサ１５Ｂ，１５Ａが備えるコア１８Ｂ，１８Ａを含む）の中から自コアが担当している論理ボリュームの担当切替え先とするコア１８Ａ，１８Ｂを決定する（ＳＰ１３～ＳＰ１５）。

【００９６】

具体的には、コア１８Ａ，１８Ｂは、まず、他のコア１８Ａ，１８Ｂ（自プロセッサ内の他のコア１８Ａ，１８Ｂも含む）のそれぞれについて、その過去３分間における５秒後との平均使用率をステップＳＰ１１と同様に算出する（ＳＰ１３）。

【００９７】

続いてコア１８Ａ，１８Ｂは、これら他のコア１８Ａ，１８Ｂの中に予め定められた第２の切替え条件を満たすコア１８Ａ，１８Ｂが存在するか否かを判断する（ＳＰ１４）。本実施の形態の場合、かかる第２の切替え条件として、第１に、そのコア１８Ａ，１８Ｂの過去３分間における５秒ごとの平均使用率のうち８０％となったものが１つもないこと、第２に、そのコア１８Ａ，１８Ｂの過去３分間における５秒ごとの平均使用率の平均が４０％以下であることが規定されている。

【００９８】

そしてコア１８Ａ，１８Ｂは、他のコア１８Ａ，１８Ｂの中にかかる第２の切替え条件をすべて満たすコア１８Ａ，１８Ｂが存在する場合には、そのコア１８Ａ，１８Ｂを、自コアの担当論理ボリュームの担当切替え先として決定する（ＳＰ１５）。

【００９９】

なおコア１８Ａ，１８Ｂは、ステップＳＰ１４の判断においてかかる第２の切替え条件をすべて満たす他のコア１８Ａ，１８Ｂが複数存在したときには、自コントローラ６Ａ，６Ｂ内の他のコア１８Ａ，１８Ｂを優先的に担当切替え先として決定する。これはコア１８Ａ，１８Ｂ間の担当切替え処理がコントローラ６Ａ，６Ｂ間での担当切替え処理に比べてオーバヘッドが少ないからである。またコア１８Ａ，１８Ｂは、かかる他のコア１８Ａ，１８Ｂが自コントローラ６Ａ，６Ｂ内に存在しないときには、他のコントローラ６Ｂ，６Ａ内のマイクロプロセッサ１５Ｂ，１５Ａが備える他のコア１８Ｂ，１８Ａのうち、過去３分間における５秒ごとの平均使用率が最も低いコア１８Ｂ，１８Ａを担当切替え先として決定する。これは他のコントローラ６Ｂ，６Ａの場合、かかる平均使用率が最も低いコア１８Ｂ，１８Ａを担当切替え先とした方が負荷分散を図れるからである。

【０１００】

続いてコア１８Ａ，１８Ｂは、自己が担当する論理ボリュームの中から、ステップＳＰ１５において決定した他のコア１８Ａ，１８Ｂに担当を切り替える論理ボリュームの候補を選択する（ＳＰ１６～ＳＰ１９）。

【０１０１】

この場合において、かかる候補を選択する方法としては、そのコア１８Ａ，１８Ｂが担当している各論理ボリュームについて、その論理ボリュームに対するそのコア１８Ａ，１８Ｂの使用率（＝（その論理ボリュームに関する処理時間）／モニタ時間）をそれぞれ算出し、この使用率（以下、これをコア使用率と呼ぶ）が最も大きい論理ボリュームを選択する方法が考えられる。しかしながら、現状では論理ボリュームごとのコア使用率を取得することは難しく、また論理ボリュームごとのコア使用率を取得しようとすると、数多くの場所に情報取得用の処理を追加する必要があり、性能的に問題が発生する。

【０１０２】

そこで、本実施の形態においては、かかる候補の論理ボリュームを選択する方法として、現状の性能モニタで採用している論理ボリュームごとの１秒当たりのデータ入出力数の比率から、そのコア１８Ａ，１８Ｂが担当している論理ボリュームごとのコア使用率の近似値を算出する（ＳＰ１６、ＳＰ１７）。

【０１０３】

具体的には、コア１８Ａ，１８Ｂは、自己が担当する各論理ボリュームについて、次式

【数１】

$$スコア＝Rm＋Rh×Th＋W×Tw \quad \cdots\cdots(1)$$

によりコア使用率をスコア化する（ＳＰ１６）。この（１）式において、その論理ボリュームに格納されたデータについてリードミスした（当該データが読出し対象のデータがキャッシュメモリ１２Ａ，１２Ｂに格納されていなかった）１秒当たりの回数の過去１分間での平均値、Ｒｈはリードヒットした（読出し対象のデータがキャッシュメモリ１２Ａ，１２Ｂに格納されていた）１秒当たりの回数の過去１分間での平均値、Ｗはその論理ボリュームにデータを書き込んだ１秒当たりの回数の過去１分間での平均値、Ｔｈはリードヒットのオーバヘッド係数、Ｔｗはライトのオーバヘッド係数をそれぞれ表す。

【０１０４】

なお、本実施の形態の場合、Ｔｈについては0.3が適用される。またＴｗについては、記憶装置４Ａ～４ＤのＲＡＩＤレベルやその論理ボリュームが形成するコピーペアのペア状態に応じた値が適用される。例えばＲＡＩＤレベルが「１」のときには2．5、「５」のときには５、「０」のときには１が適用される。ただし、その論理ボリュームが内部コピーペア、リモートコピーペア又は非同期リモートコピーペアの正側であり、かつその内部コピーペア、リモートコピーペア又は非同期リモートコピーペアのペア状態が正側及び副側のデータが一致している状態（「PAIR」）である場合には、上記値にそれぞれ２、1.5又は４を乗じた値がＴｗとして適用される。また、その論理ボリュームがスナップショットコピーペアの正側であり、かつそのスナップショットコピーペアのペア状態がスプリット状態（「PSUS」）である場合には、上記値に４を乗じた値がＴｗとして適用される。

【０１０５】

続いてコア１８Ａ，１８Ｂは、ステップＳＰ１６において算出した論理ボリュームごとのスコアに基づいて、これら論理ボリュームごとのスコアの比率と、自コア全体の使用率とからかかる論理ボリュームごとのコア使用率を算出する（ＳＰ１７）。

【０１０６】

そしてコア１８Ａ，１８Ｂは、この後、上述のようにして得られた論理ボリュームごとのコア使用率に基づいて、担当を切り替える論理ボリュームを選択する（ＳＰ１８）。具体的にコア１８Ａ，１８Ｂは、担当を切り替える論理ボリュームとして、担当を切り替えた後の担当切替え元のコア使用率及び担当切替え先のコア使用率を均一化するため、そのコア使用率が、担当切替え元のコア（自コア）の過去１分間のコア使用率と、担当切替え先のコアの過去１分間のコア使用率との差の半分に最も近いという要件（以下、これを対象論理ボリューム選択要件と呼ぶ）を満たす論理ボリュームを選択する。

【０１０７】

この際、コア１８Ａ，１８Ｂは、そのとき対象としている論理ボリュームの内部コピーペア管理テーブル４４Ａ，４４Ｂ（図６）、スナップショットペア管理テーブル４５Ａ，４５Ｂ（図７）、非同期リモートコピーペア管理テーブル４７Ａ，４６Ｂ（図８）及びリモートコピーペア管理テーブル４８Ａ，４８Ｂ（図４）をそれぞれ参照して、その論理ボリュームが内部コピーペア、スナップショットペア、非同期リモートコピーペア又はリモートコピーペアのいずれかを構成しているか否かと、その論理ボリュームがコンシステンシグループを構成しているか否かとを判定する。

【０１０８】

そしてコア１８Ａ，１８Ｂは、その論理ボリュームがいずれかのコピーペアを構成していたり、コンシステンシグループを構成しているときには、かかるコピーペアの相手側の論理ボリュームや、かかるコンシステンシグループを構成する他の論理ボリューム、さらに相手側の論理ボリュームがプールボリュームであるときにはそのプールボリュームにマッピングされた他の論理ボリュームなど、その論理ボリュームと相互に関連する他の論理ボリュームを１つの論理ボリューム群として、かかる対象論理ボリューム選択要件を満たす論理ボリューム群を選択する。

【０１０９】

ただし、この対象論理ボリューム選択要件を満たす論理ボリューム又は論理ボリューム群であっても、±３％以内のクロス論理ボリューム（リードコマンドやライトコマンドの９０％以上が、担当でないコントローラ６Ｂ，６Ａ側からきている論理ボリューム）が存在する場合には、担当でないコントローラ６Ｂ，６Ａを担当コントローラ６Ａ，６Ｂにすることを優先するため、このクロス論理ボリュームの中から最も近いものを選択する。

【０１１０】

その後コア１８Ａ，１８Ｂは、ステップＳＰ１８において選択した論理ボリューム又は論理ボリューム群全部の担当をステップＳＰ１５において選択したコア１８Ａ，１８Ｂに切り替えることができるか否かを判断する（ＳＰ１９）。具体的には、本実施の形態に場合、その論理ボリューム又は論理ボリューム群全部の担当を切り替えた後の切替え先のコア１８Ａ，１８Ｂの使用率が、切り替え前の切替え元のコア１８Ａ，１８Ｂの使用率の－

１０％以上になると予想される場合には担当を切り替えることができない。これは、例えばかかる切替え元のコア１８Ａ，１８Ｂの使用率が７０％、切替え先のコア１８Ａ，１８Ｂの使用率が４０％であるものとすると、切替え対象の論理ボリューム又は論理ボリューム群としてそのコア使用率（論理ボリューム群のときには当該論理ボリューム群全体でのコア使用率）が３０％の論理ボリューム又は論理ボリューム群を選択した場合、担当を切り替えた後も負荷バランスが不均衡であることに変わりがないからである。

【０１１１】

そしてコア１８Ａ，１８Ｂは、ステップＳＰ１９の判断において否定結果を得ると、５つ目の候補を選択し終えたか否かを判断する（ＳＰ２０）。そしてコア１８Ａ，１８Ｂは、この判断において否定結果を得ると、同様の処理を最大５回繰り返す（ＳＰ１８～ＳＰ２０－ＳＰ１９）。最大回数を設けた理由は、同様の処理をさらに続けると処理時間が長くなることから、これを回避するためのである。なお、この最大回数は５回以外であっても良い。そしてコア１８Ａ，１８Ｂは、ステップＳＰ１８～ステップＳＰ２０の処理を５回繰り返しても担当切替え条件を満たす論理ボリューム又は論理ボリューム群を選択できなかった場合には、担当切替え対象となる論理ボリューム又は論理ボリューム群が存在しないものとして所定のエラー処理を行った後（ＳＰ２１）、この第２の負荷分散処理を終了する。

【０１１２】

一方、コア１８Ａ，１８Ｂは、かかるステップＳＰ１８～ステップＳＰ２０の処理により切替え対象の論理ボリューム又は論理ボリューム群を選択できた場合には、その論理ボリューム又は論理ボリューム群の担当をステップＳＰ１５において決定したコア１８Ａ，１８Ｂに切り替えることができるか否かを判断する（ＳＰ２２）

【０１１３】

この場合において、例えばステップＳＰ２４について後述する担当切替え抑止時間内ではかかる担当の切り替えを行なえないため、このステップＳＰ２２において否定結果を得ることになる。かくしてこのときコア１８Ａ，１８Ｂは、かかる担当の切り替えを行なうことなく、この第２の負荷分散処理を終了する。

【０１１４】

これに対してコア１８Ａ，１８Ｂは、ステップＳＰ２２の判断において肯定結果を得ると、かかる担当の切り替えを行なう。具体的は、コア１８Ａ，１８Ｂは、他コントローラ６Ｂ，６Ａのローカルメモリ１４Ｂ，１４Ａの共有領域４２Ａ，４２Ｂ（図４）に格納されている、その論理ボリューム又はその論理ボリューム群を構成する各論理ボリュームの担当管理テーブル４３Ａ，４３Ｂ（図５）上の「担当コア」欄４３ＣＡに格納されたコードと、「担当コントローラ」欄４３ＣＢに格納されたコードとを、それぞれステップＳＰ１５において選択したコア１８Ａ，１８Ｂのコードと、当該コア１８Ａ，１８Ｂを備えるマイクロプロセッサ１５Ａ，１５Ｂが属するコントローラ６Ａ，６Ｂのコードとに書き換える。またコア１８Ａ，１８Ｂは、これと併せて、他コントローラ６Ｂ，６Ａのローカルメモリ１４Ｂ，１４Ａの共有領域４２Ｂ，４２Ａに格納されている担当管理テーブル４３Ｂ，４３Ａについても同様に更新する。

【０１１５】

そしてコア１８Ａ，１８Ｂは、この後記憶制御装置３内のすべてのコア１８Ａ，１８Ｂについて、上述のように論理ボリュームの担当を切り替える処理を一定時間（これを担当切替え抑止時間と呼ぶ）だけ抑止するための設定を行い（ＳＰ２４）、この後この第２の負荷分散処理を終了する。

【０１１６】

（３）コマンド発行分担機能

次に、本実施の形態による記憶制御装置の各コントローラに搭載されたもう１つの特徴的な機能であるコマンド発行分担機能について説明する。

【０１１７】

従来、複数のコントローラを備える記憶制御装置では、リモートコピーを行う場合に、

イニシエータ側（コマンド発行側）の記憶装置４Ａ～４Ｄを制御する処理と、ターゲット側（コマンド受領側）の記憶装置４Ａ～４Ｄを制御する処理との双方の処理を行う。この場合において、従来の記憶制御装置では、イニシエータ側の記録装置４Ａ～４Ｄを制御する処理のうち、ＳＣＳＩコマンドの発行は、正側の論理ボリュームの担当コントローラ（より具体的には担当コア）が行い、それ以外のリンクサービス系のコマンドの発行は、そのコマンドをターゲット側の記憶装置４Ａ～４Ｄに送信する際に使用する通信ポート２０Ａを担当しているコントローラ（より具体的には担当コア）が行なっている。

【０１１８】

しかしながら、本実施の形態のストレージシステム１では、正側の論理ボリュームの担当コントローラ６Ａ，６Ｂでないコントローラ６Ａ，６Ｂが担当している通信ポート２０Ａを介してＳＣＳＩコマンドを発行することが可能であり、実際上、負荷や障害等の影響によりそのようなケースが発生することもあり得る。

【０１１９】

このような状況のもと、リンクサービス系のコマンドのうち、ＳＣＳＩコマンドと関連して発行されるコマンドについてまでも、当該コマンドを送信する際に使用する通信ポート２０Ａを担当しているコントローラ６Ａ，６Ｂが行なうものとすると、正側の論理ボリュームの担当コントローラ６Ａ，６Ｂと、そのコマンドを発行する通信ポート２０Ａを担当するコントローラ６Ｂ，６Ａとが異なる場合に、これらコントローラ６Ａ，６Ｂ間でコマンドの受け渡しのための通信が発生し、その分の負荷がコントローラ６Ａ，６Ｂに発生する問題がある。

【０１２０】

そこで、本実施の形態の記憶制御装置には、リンクサービス系のコマンドのうち、ＳＣＳＩコマンドと関連して発行されるコマンドについては正側の論理ボリュームの担当コントローラ６Ａ，６Ｂが発行し、リンクサービス系のこれ以外のコマンド（特にポート単位の処理に関するコマンド）については、そのコマンドをターゲット側の記憶装置４Ａ～４Ｄに送信する際に使用する通信ポート２０Ａを担当しているコントローラ６Ｂ，６Ａが発行を担当するコマンド発行分担機能が搭載されている。

【０１２１】

具体的には、例えば発行したＳＣＳＩコマンドに関してタイムアウトが発生した場合などに、ターゲット側の記憶装置４Ａ～４ＤにそのＳＣＳＩコマンドに基づく処理を強制終了させる目的で使用されるＡＢＴＳ（Abort Sequence）コマンドについては、正側の論理ボリュームの担当コントローラ６Ａ，６Ｂが発行を担当する。一方、ＡＢＴＳコマンドを発行しても繰り返しタイムアウトになった場合などに、ターゲット側の記憶装置４Ａ～４Ｄに対応する通信ポート２０Ａをリセットさせる目的で使用されるＰＬＯＧＩ（Port Log In）コマンドについては、当該ＰＬＯＧＩコマンドをターゲット側に送信する際に使用する通信ポート２０Ａを担当するコントローラ６Ｂ，６Ａが発行を担当する。

【０１２２】

このような本実施の形態によるコマンド発行分担機能に関する具体的な記憶制御装置内での一連の処理の流れを図１２に示す。なお、以下においては、これから形成しようとするコピーペアの正側の論理ボリュームの担当コントローラは１系のコントローラ６Ｂであり、０系のコントローラ６Ａが担当している通信ポート２０Ａを介して記憶制御装置３がイニシエータ側の記憶装置４Ａ～４Ｄにログインしようとしているものとする。また以下においては、処理主体を０系又は１系のコントローラ６Ａ，６Ｂとして説明するが、より具体的にはその０系又は１系のコントローラ６Ａ，６Ｂ内のコア１８Ａ，１８Ｂが処理を行うことは言うまでもない。

【０１２３】

図１２（Ａ）は、０系のコントローラ６Ａがかかる通信ポート（「0A」）２０Ａを介してイニシエータ側の記憶装置４Ａ～４Ｄにログインするためのプログインコマンド（「PLOGI」）を発行し、この結果、その記憶装置４Ａ～４Ｄの所定のポートにログインした状態を示している。０系のコントローラ６Ａは、かかるログインを完了すると、ログイン

先の記憶装置４Ａ～４ＤのＩＤや当該記憶装置４Ａ～４Ｄにおけるログイン先のポートＩＤ及びフレームサイズなどのログイン情報を自コントローラ内のローカルメモリ１４Ａの共有領域４２Ａと、１系のコントローラ６Ｂ内のローカルメモリ１４Ｂの共有領域４２Ｂとにそれぞれ格納する。なお、図１２（Ａ）において、ログイン管理テーブル５０は、０系のコントローラ６Ａが「0A」という通信ポート２０Ａを介して１又は複数のイニシエータ側の記憶装置４Ａ～４Ｂにログインしている場合における、これらログインに関するログイン情報が格納されている自コントローラ内のローカルメモリ１４Ａの共有領域４２Ａ上のアドレスを管理するためのテーブルである。

【０１２４】

このとき１系のコントローラ６Ｂは、図１２（Ｂ）に示すように、自コントローラ内のローカルメモリ１４Ｂの共有領域４２Ｂを参照して、０系のコントローラ６Ａがターゲット側の記憶装置４Ａ～４Ｂにログインしたか否かをチェックしている（「ログインチェック」）。そして１系のコントローラ６Ｂは、共有領域４２Ｂに格納されたログイン情報に基づいてかかるログインを確認すると、記憶制御装置３内において使用しているターゲットＩＤをファイバーチャネル規格に応じたＩＤに変換する変換処理等の必要な処理（「ＩＤ変換等」）を行う。さらに１系のコントローラ６Ｂは、この後、これからＳＣＳＩコマンドを送信することを意味するＳＣＳＩコマンド（「VU（Vender Unique）コマンド」）を０系のコントローラ６Ａ内のローカルメモリの共有領域及び１系のコントローラ内のローカルメモリの共有領域に書き込む。この結果、このＳＣＳＩコマンドが０系のコントローラ６Ａによりかかる通信ポート２０Ｂを介してイニシエータ側の記憶装置４Ａ～４Ｄに送信される。

【０１２５】

一方、この後、かかるＳＣＳＩコマンドについてタイムアウトが発生した場合、図１２（Ｃ）に示すように、１系のコントローラ６Ｂは、かかるＳＣＳＩコマンドに対する処理をリセットし、これと併せてかかるＳＣＳＩコマンドに基づく処理を中断させるためのＡＢＴＳコマンドを０系のコントローラ６Ａのローカルメモリ１４Ａの共有領域４２Ａに書き込む。この結果、かかるＡＢＴＳコマンドが０系のコントローラ６Ａによりかかる通信ポート２０Ａを介してイニシエータ側の記憶装置４Ａ～４Ｄに送信される。またかかるＡＢＴＳコマンドによりイニシエータ側の記憶装置４Ａ～４ＤがかかるＳＣＳＩコマンドに対する処理を中断した場合、１系のコントローラ６Ｂは、ＲＲＱコマンドを０系のコントローラ６Ａのローカルメモリ１４Ａの共有領域４２Ａに書き込む。この結果、かかるＡＢＴＳコマンドが０系のコントローラ６Ａによりかかる通信ポート２０Ａを介してイニシエータ側の記憶装置４Ａ～４Ｄに送信される。

【０１２６】

これに対して、かかるＡＢＴＳコマンドやＲＲＱコマンドについてタイムアウトが発生した場合、図１２（Ｄ）に示すように、１系のコントローラ６Ｂは最終的な手段として、リセットとしての意味合いをもつＰＬＯＧＩコマンドをイニシエータ側の記憶装置４Ａ～４Ｄに発行するように０系のコントローラ６Ａに依頼する。かくしてこの依頼を受けた０系のコントローラ６Ａは、自己が管理する対応するかかる通信ポート２０Ａを介してイニシエータ側の記憶装置４Ａ～４ＤにＰＬＯＧＩコマンドを発行する。

【０１２７】

（４）本実施の形態の効果

以上のように本実施の形態によるストレージシステム１では、記憶制御装置３の各コントローラ６Ａ，６Ｂのマイクロプロセッサ１５Ａ，１５Ｂを構成する各コア１８Ａ，１８Ｂが、それぞれ自己の負荷状況を監視し、負荷が大きくなったときには自己が担当する論理ボリュームの担当を負荷の小さい他のコア１８Ａ，１８Ｂに切り替えるため、コア１８Ａ，１８Ｂ間における負荷の均等化を図ることができる。かくするにつき、コア１８Ａ，１８Ｂ間の負荷の偏り等に起因するデータ入出力処理性能の劣化を有効かつ簡易に防止することができる。

【０１２８】

また本実施の形態によるストレージシステム１では、コピーペアの形成時や上述のようにして負荷分散を図るときに、相互に関連する複数の論理ボリュームの担当を一括して１つのコア１８Ａ，１８Ｂに切り替えるため、コピー処理の際にコントローラ６Ａ，６Ｂ間での通信が発生せず、かかる通信に起因するコントローラ６Ａ，６Ｂの負荷の発生を未然に防止することができる。かくするにつき、このような負荷の発生によるデータ入出力性能の劣化を有効かつ簡易に防止することができる。

【０１２９】

さらに本実施の形態によるストレージシステム１では、リンクサービス系のコマンドのうち、ＳＣＳＩコマンドと関連して発行されるコマンドについては正側の論理ボリュームの担当コントローラ６Ａ，６Ｂが発行し、リンクサービス系のポート単位の処理に関するコマンドについては、そのコマンドをターゲット側の記憶装置４Ａ～４Ｄに送信する際に使用する通信ポート２０Ａを担当しているコントローラ６Ｂ，６Ａが発行を担当するため、これらコントローラ６Ａ，６Ｂ間でコマンドの受け渡しのための通信が発生するのを未然に回避することができる。かくするにつき、かかる通信に起因するコントローラ６Ａ，６Ｂの負荷の発生を未然に防止し、このような負荷の発生によるデータ入出力性能の劣化を有効かつ簡易に防止することができる。

【０１３０】

（５）他の実施の形態

なお上述の実施の形態においては、本発明を、マイクロプロセッサ１５Ａ，１５Ｂがそれぞれ２つのコア（１８ＡＸ，１８ＡＹ，１８ＢＸ，１８ＢＹ）により構成された記憶制御装置３に適用するようにした場合について述べたが、本発明はこれに限らず、マイクロプロセッサ１５Ａ，１５Ｂが１つのコアから構成された記憶制御装置や、マイクロプロセッサ１５Ａ，１５Ｂが３つ以上のコアから構成された記憶制御装置にも広く適用することができる。

【０１３１】

また上述の実施の形態においては、本発明を、コントローラ６Ａ，６Ｂが２つ設けられた記憶制御装置３に適用するようにした場合について述べたが、本発明はこれに限らず、マイクロプロセッサが複数のコアから構成されている場合には、コントローラが１つだけ設けられた記憶制御装置３にも本発明を広く適用することができる。

【０１３２】

さらに上述の実施の形態においては、自己が担当する論理ボリュームに対するデータの入出力を制御する制御部と、各制御部の負荷状況を監視する負荷監視部と、負荷監視部の監視結果に基づいて、各制御部の負荷を均等化するように、論理ボリュームの担当を他の制御部に切り替える切替え部とを同じ１つのコア１８Ａ，１８Ｂにより構成するようにした場合について述べたが、本発明はこれに限らず、各コア１８Ａ，１８Ｂの負荷状況をそれぞれ監視する負荷監視部としての機能を有するコアをかかるコア１８Ａ，１８Ｂとは別に設け、その負荷監視部の監視結果に基づいて各コア１８Ａ，１８Ｂが上述の切替え部としての処理を実行するようにしても良い。

【０１３３】

さらに上述の実施の形態においては、本発明を図１のように構成された記憶制御装置３に適用するようにした場合について述べたが、本発明はこれに限らず、記憶装置が提供する記憶領域上に作成した論理ボリュームをホスト計算機に提供すると共に、当該論理ボリュームに対するデータの入出力を制御する種々の構成の記憶制御装置に広く適用することができる。

【図面の簡単な説明】

【０１３４】

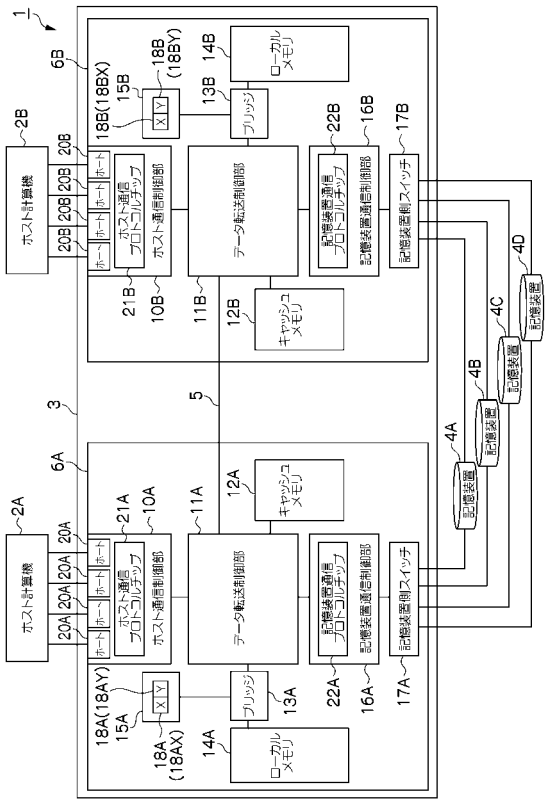
【図１】本実施の形態によるストレージシステムの全体構成を示すブロック図である。

【図２】記憶制御装置のローカルメモリにおけるキュー構造の説明に供する概念図である。

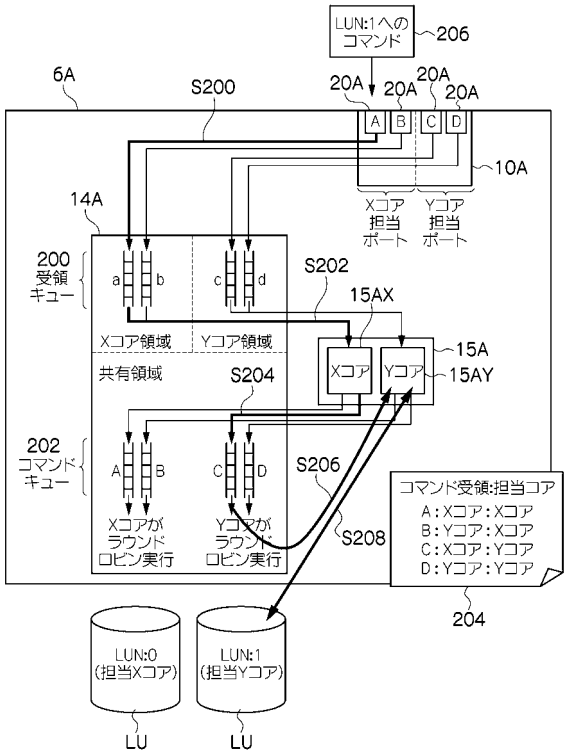
【図３】（Ａ）は、０系及び１系のマイクロプロセッサが認識するメモリ空間の一例を示

す図表であり、（Ｂ）は、０系及び１系のホスト通信プロトコルチップがそれぞれ認識するメモリ空間の一例を示す図表である。
【図４】（Ａ）は０系のローカルメモリのメモリ構造を示す概念図であり、（Ｂ）は１系のローカルメモリのメモリ構造を示す概念図である。
【図５】担当管理テーブルの説明に供する概念図である。
【図６】内部コピーペア管理テーブルの説明に供する概念図である。
【図７】スナップショットコピーペア管理テーブルの説明に供する概念図である。
【図８】非同期リモートコピーペア管理テーブルの説明に供する概念図である。
【図９】プール管理テーブルの説明に供する概念図である。
【図１０】第１の負荷分散処理の説明に供するフローチャートである。
【図１１】第２の負荷分散処理の説明に供するフローチャートである。
【図１２】コマンド発行分担機能の説明に供する概念図である。
【符号の説明】
【０１３５】
１……ストレージシステム、２Ａ，２Ｂ……ホスト計算機、３……記憶制御装置、４Ａ～４Ｄ……記憶装置、６Ａ，６Ｂ……コントローラ、１４Ａ，１４Ｂ……ローカルメモリ、１５Ａ，１５Ｂ……マイクロプロセッサ、１８Ａ，１８ＡＸ，１８ＡＹ，１８Ｂ，１８ＢＸ，１８ＢＹ……コア、２０Ａ，２０Ｂ……通信ポート、４０Ａ，４０Ｂ……自系コントローラ受領コマンド格納領域、４１Ａ，４１Ｂ……他系コントローラ受領コマンド格納領域、４２Ａ，４２Ｂ……共有領域、４３Ａ，４３Ｂ……担当管理テーブル、４４Ａ，４４Ｂ……内部コピーペア管理テーブル、４５Ａ，４５Ｂ……スナップショットペア管理テーブル、４６Ａ，４６Ｂ……非同期リモートコピーペア管理テーブル、４７Ａ，４７Ｂ……リモートコピーペア管理テーブル、４８Ａ，４８Ｂ……プール管理テーブル。

【図１】
図1

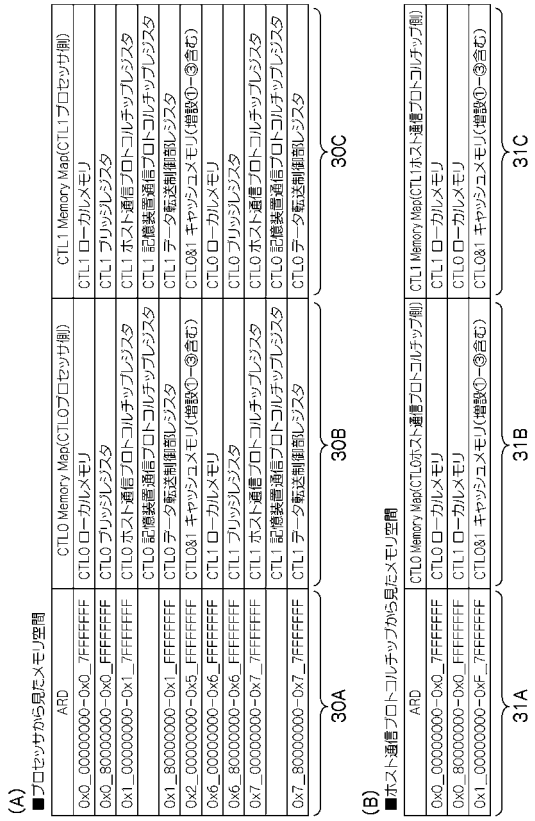


【図２】
図2



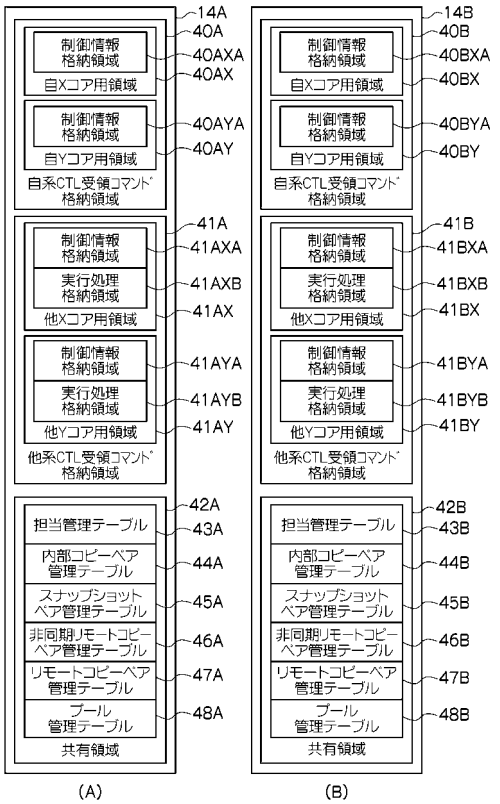
【図３】

図3



【図４】

図4



【図５】

図5

項目	詳細	
担当コントローラ	0x00：コントローラ0が担当 0x01：コントローラ1が担当 0xFF：LU未定義	43CA
担当コア	0x00：Xコアが担当 0x01：Yコアが担当 0xFF：LU未定義	43CB

43A(43B)

【図６】

図6

項目	詳細	
ペア状態	00：SMPL・・・ペア未定義 01：COPY・・・イニシャルコピー中／再同期中 02：PAIR・・・正副のデーター致 03：PSUS・・・スプリット状態 04：PSUE・・・障害発生／サスペンド 05：MIX・・・ペア状態混在（正のみありえる）	44CA
ペア属性	00：正VOL 01：副VOL	44CB
相手LUN	相手LUのLUN ※副のみ有効値	44CC
ミラー番号	0～7 ※副のみ有効値	44CD
コピースピード	81～85 ：遅い 86～8A ：普通 8B～8F ：早い ※副のみ有効値	44CE
グループ種別	00：グループID無効 01：グループID有効 ※副のみ有効値	44CF
グループID	0～255	44CG
副LUN	副LUのLUN 0～2047(S)／4095(M,H) 初期値(FFFF)H：未定義 ※正のみ有効値	44CH

44A(44B)

【図７】
図7

項目	詳細	
ペア状態	00：SMPL・・・ペア未定義 01：COPY・・・リストアデータコピー中 02H：PAIR・・・正副のデータ一致 03：PSUS・・・スプリット状態 04：PSUS・・・障害発生／サスペンド 05：MIX・・・ペア状態混在 06：SMPLS・・・SMPL(副VOLイメージ形成済み)	45CA
ペア属性	(00)H：正VOL (01)H：副VOL	45CB
相手LUN	相手LUのLUN ※副のみ有効値	45CC
副LUN数	1～32	45CD
副LUN	副LUのLUN 初期値(FFFF)H：未定義 ※正のみ有効値	45CE
ミラー番号	0xFFFF：無効	45CF
グループID	0x01：グループID有効 0x00：グループID無効	45CG
グループID	0～255	45CH
POOL#	使用するPOOLのPOOL#	45CI

45A(45B)

【図８】
図8

項目	詳細	
ペア状態	00：SMPL・・・ペア未定義 01：COPY・・・イニシャルコピー中／再同期中 02：PAIR・・・非同期コピー維持中 03：PSUS・・・サスペンド状態 04：PSUE・・・障害によるサスペンド状態	46CA
ペア属性	00：正VOL 01：副VOL	46CB
CTグループID	0～15	46CC
相手LUN	TCEの相手ペアLUN 0～2047(M)／4095(H)	46CD
コピースピード	81～85 ：遅い 86～8A ：普通 8B～8F ：早い	46CE
POOL#	使用するPOOLのPOOL#	46CF

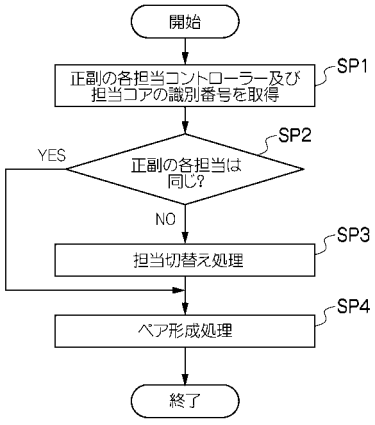
46A(46B)

【図９】
図9

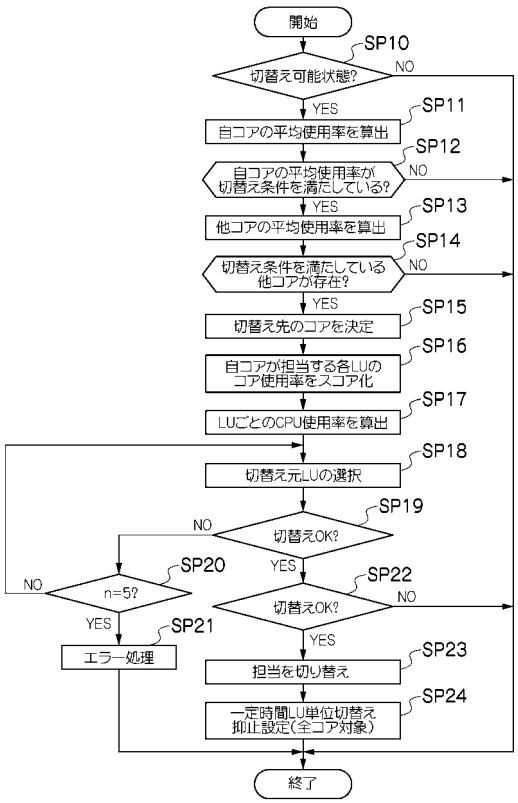
項目	詳細	
LU番号	プールに所属しているLU番号	48CA
先頭プールフラグ	0x01：先頭 0x00：先頭以外	48CB
所属POOL#	使用するPOOLのPOOL#	48CC

48A(48B)

【図１０】
図10

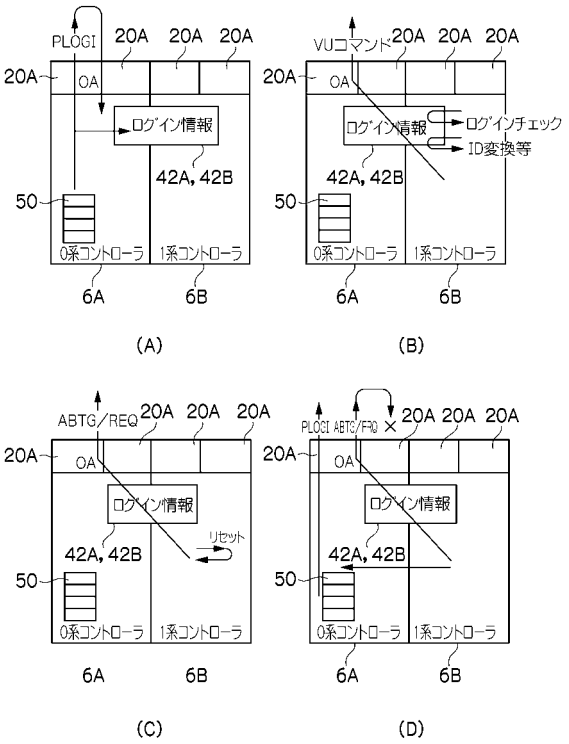


【図１１】
図11



【図１２】

図12


PLOGI
20A
20A
20A
20A
OA
ログイン情報
42A, 42B
50
0系コントローラ
1系コントローラ
6A
6B
(A)
VUコマンド
ログインチェック
ID変換等
(B)
ABTG/REQ
リセット
(C)
PLOGI ABTG/FRQ
×
(D)

フロントページの続き

審査官　木村　雅也

(56)参考文献　特開２００７－１１５０１９（ＪＰ，Ａ）
特開２００６－２７７４８７（ＪＰ，Ａ）
特開平０９－１４６８４２（ＪＰ，Ａ）
特開平１１－３１２０５８（ＪＰ，Ａ）
特開２００４－１４５８５５（ＪＰ，Ａ）
米国特許第０６６７５２６８（ＵＳ，Ｂ１）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　　３／０６