(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4121255号
(P4121255)
(45)発行日　平成20年7月23日(2008.7.23)　(24)登録日　平成20年5月9日(2008.5.9)

(51)Int.Cl.　F I
G06F 12/00 (2006.01)　G06F 12/00　501B

請求項の数 3　(全 22 頁)

(21)出願番号　特願2001-177180(P2001-177180)
(22)出願日　平成13年6月12日(2001.6.12)
(65)公開番号　特開2002-366398(P2002-366398A)
(43)公開日　平成14年12月20日(2002.12.20)
審査請求日　平成16年10月6日(2004.10.6)

(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　100093492
弁理士　鈴木　市郎
(74)代理人　100078134
弁理士　武　顕次郎
(72)発明者　平川　裕介
神奈川県川崎市麻生区王禅寺１０９９番地　株式会社　日立製作所　システム開発研究所内
(72)発明者　荒川　敬史
神奈川県川崎市麻生区王禅寺１０９９番地　株式会社　日立製作所　システム開発研究所内

最終頁に続く

(54)【発明の名称】クラスタ構成記憶システム

(57)【特許請求の範囲】
【請求項１】
複数の記憶システムノードを１つの記憶システムとして運用可能とするクラスタ構成記憶システムにおいて、
前記記憶システムノード内部及び前記記憶システムノード相互間のアクセス情報を採取する手段と、採取したアクセス情報を保守員に提示する手段と、保守員からのポートの設定指示を受け付ける手段と、前記指示に基づいて前記クラスタ構成記憶システム内のポート設定を変更する手段とを有し、
前記クラスタ構成記憶システム内のポート設定を変更する手段は、論理ポートと論理ボリュームとの間の仮想的なパスである論理パス毎の単位時間当りのデータ転送量またはデー転送時間を示すアクセス頻度を保持するアクセス情報を参照し、全論理パスの中でデータ転送量が規定値以上の論理パスの１つを選択し、選択した論理パスが記憶システムノード間のデータ転送を必要とするか否かをチェックし、記憶システムノード間のデータ転送を必要とした場合、
前記論理パスを経て前記ホストコンピュータがアクセスしている論理ボリュームを格納している記憶装置を有する第１の記憶システムノードに未使用ポートがあり、その未使用ポートが当該ホストコンピュータに接続可能である場合、あるいは、当該ホストコンピュータに接続された未使用ポートがある場合に、前記ホストコンピュータが使用するノードを、それまで使用していた第２の記憶システムノードのポートから前記第１の記憶システムノードの前記未使用ポートとするように、ポート設定を変更することを特徴とするクラ

スタ構成記憶システム。
【請求項２】
複数の記憶システムノードを１つの記憶システムとして運用可能とするクラスタ構成記憶システムにおいて、
前記記憶システムノード内部及び前記記憶システムノード相互間のアクセス情報を採取する手段と、採取したアクセス情報に基づいて、ポートの設定を決定する手段と、前記クラスタ構成記憶システム内のポート設定を変更する手段とを有し、
前記クラスタ構成記憶システム内のポート設定を変更する手段は、論理ポートと論理ボリュームとの間の仮想的なパスである論理パス毎の単位時間当りのデータ転送量またはデー転送時間を示すアクセス頻度を保持するアクセス情報を参照し、全論理パスの中でデータ転送量が規定値以上の論理パスの１つを選択し、選択した論理パスが記憶システムノード間のデータ転送を必要とするか否かをチェックし、記憶システムノード間のデータ転送を必要とした場合、
前記論理パスを経て前記ホストコンピュータがアクセスしている論理ボリュームを格納している記憶装置を有する第１の記憶システムノードに未使用ポートがあり、その未使用ポートが当該ホストコンピュータに接続可能である場合、あるいは、当該ホストコンピュータに接続された未使用ポートがある場合に、前記ホストコンピュータが使用するノードを、それまで使用していた第２の記憶システムノードのポートから前記第１の記憶システムノードの前記未使用ポートとするように、ポート設定を変更することを特徴とするクラスタ構成記憶システム。
【請求項３】
複数の記憶システムノードを１つの記憶システムとして運用可能とするクラスタ構成記憶システムにおいて、
前記記憶システムノード内部及び前記記憶システムノード相互間のアクセス情報を採取する手段と、前記クラスタ構成記憶システムを利用するホストコンピュータにクラスタ構成記憶システム内のポート情報及び前記アクセス情報を提供する手段と、前記ホストコンピュータからのポートの設定指示を受け付ける手段と、前記指示に基づいてクラスタ構成記憶システム内のポート設定を変更する手段とを有し、
前記クラスタ構成記憶システム内のポート設定を変更する手段は、論理ポートと論理ボリュームとの間の仮想的なパスである論理パス毎の単位時間当りのデータ転送量またはデー転送時間を示すアクセス頻度を保持するアクセス情報を参照し、全論理パスの中でデータ転送量が規定値以上の論理パスの１つを選択し、選択した論理パスが記憶システムノード間のデータ転送を必要とするか否かをチェックし、記憶システムノード間のデータ転送を必要とした場合、
前記論理パスを経て前記ホストコンピュータがアクセスしている論理ボリュームを格納している記憶装置を有する第１の記憶システムノードに未使用ポートがあり、その未使用ポートが当該ホストコンピュータに接続可能である場合、あるいは、当該ホストコンピュータに接続された未使用ポートがある場合に、前記ホストコンピュータが使用するノードを、それまで使用していた第２の記憶システムノードのポートから前記第１の記憶システムノードの前記未使用ポートとするように、ポート設定を変更することを特徴とするクラスタ構成記憶システム。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は、記憶システムに係り、特に、複数の記憶システムを１つの記憶システムとして運用可能とするクラスタ構成の記憶システムに関する。
【０００２】
【従来の技術】
記憶システムに関する従来技術として、例えば、特開平１１－１６７５２１号公報等に記載された技術が知られている。この従来技術は、記憶システムを使用する上位装置（ホス

トコンピュータ）に対するインタフェース（ホストアダプタ、ＣＨＡ）、記憶システム内の磁気ディスク装置等の記憶装置に対するインタフェース（ディスクアダプタ、ＤＫＡ）、キャッシュメモリ（ＣＡＣＨＥ）、管理メモリ（ＳＭ）の相互間をコモンパス方式で接続して構成したものである。

【０００３】

図１５は従来技術による記憶システムの構成例を示すブロック図であり、以下、図１５を参照して従来技術について説明する。図１５において、１００は記憶システム、１１０はホストアダプタ（ＣＨＡ）、１２０はディスクアダプタ（ＤＫＡ）、１３０はキャッシュメモリ（ＣＡＣＨＥ）、１４０は管理メモリ（ＳＭ）、１５０は記憶装置（ＨＤＤ）、１６０はコモンパス、１７０は接続線である。

【０００４】

従来技術による記憶システム１００は、図１５に示すように、ホストアダプタ１１０、ディスクアダプタ１２０、キャッシュメモリ１３０、管理メモリ１４０、記憶装置１５０、コモンパス１６０、接続線１７０より構成される。ホストアダプタ１１０、ディスクアダプタ１２０、キャッシュメモリ１３０、管理メモリ１４０は、コモンパス１６０により相互間が接続されている。コモンパス１６０は、コモンパス１６０の障害時のために２重化されている。ディスクアダプタ１２０と記憶装置１５０との間は、１つのディスクアダプタ１２０あるいは１つの接続線１７０の障害時にも記憶装置１５０を使用できるように、１つの記憶装置１５０に２つのディスクアダプタ１２０が異なる接続線１７０で接続されている。

【０００５】

ホストアダプタ１１０は、図示しないホストコンピュータとキャッシュメモリ１３０との間のデータ転送を制御する。ディスクアダプタ１２０は、キャッシュメモリ１３０と記憶装置１５０との間のデータ転送を制御する。キャッシュメモリ１３０は、ホストコンピュータから受信したデータ、あるいは、記憶装置１５０から読み取ったデータを一時的に蓄えるメモリである。管理メモリ１４０は、全てのホストアダプタ１１０とディスクアダプタ１２０とが共有するメモリである。また、図示していないが、記憶システム１００の設定、監視、保守等を行なうために保守端末（ＳＶＰ）が全てのホストアダプタ１１０とディスクアダプタ１２０とに専用線を用いて接続されている。

【０００６】

前述したような構成を持つ記憶システム１００のシステム構成を拡張する場合、ホストアダプタ１１０、ディスクアダプタ１２０、キャッシュメモリ１３０、管理メモリ１４０、記憶装置１５０等の構成要素が新たに追加される。例えば、ホストコンピュータとの接続数を増やす場合、ホストアダプタ１１０が新たにコモンパス１６０に接続される。また、記憶システム１００の記憶容量を増やす場合、記憶装置１５０を追加する、あるいは、ディスクアダプタ１２０を新たにコモンパス１６０に接続して記憶装置１５０を追加する。

【０００７】

【発明が解決しようとする課題】

前述した従来技術は、記憶システムのシステム拡張に当たり、ホストアダプタ、ディスクアダプタ、キャッシュメモリ、管理メモリ及び記憶装置等の記憶システムの構成要素を増設することにより対応しているため、記憶システムの拡張性が記憶システムの構成要素の最大搭載数に制限されている。この結果、前述の従来技術は、大規模な記憶システムの要求に応じるために、記憶システムの最大搭載数を大きくすると、小規模な記憶システムの要求時にコスト及び設置面積が大きくなってしまうという問題点を有している。

【０００８】

前述したような問題点を解決する方法として、クラスタ構成の記憶システムが考えられる。クラスタ構成記憶システムは、複数の前述したような記憶システムを接続した構成で、ホストコンピュータからは１つの記憶システムとして運用可能とした記憶システムである。以下、クラスタ構成の記憶システムを構成する記憶システムを記憶システムノードと呼ぶ。クラスタ構成の記憶システムは、小規模な記憶システムの要求時には、少数の記憶シ

ステムノードでクラスタ構成記憶システムを構成し、記憶システムの規模を拡大する場合、クラスタ構成記憶システムに記憶システムノードを増設していくことにより対応することができる。このように、クラスタ構成記憶システムは、小規模なシステムから大規模なシステムまで対応することができ、また、ホストコンピュータからは１つの記憶システムとして運用可能であるため、管理が容易になるという利点を有している。

【０００９】

しかし、クラスタ構成の記憶システムは、ホストコンピュータからのアクセス命令を受信する記憶システムノードとアクセス対象のデータを保持する記憶システムノードとが異なる場合、記憶システムノード相互間のデータ転送が必要となり、アクセス性能が低下するという問題点を有している。

【００１０】

本発明の目的は、前述したようなクラスタ構成の記憶システムの問題点を解決し、アクセス性能を向上させることができるクラスタ構成の記憶システムを提供することにある。

【００１１】

【課題を解決するための手段】

本発明によれば前記目的は、複数の記憶システムノードを１つの記憶システムとして運用可能とするクラスタ構成記憶システムにおいて、前記記憶システムノード内部及び前記記憶システムノード相互間のアクセス情報を採取する手段と、採取したアクセス情報を保守員に提示する手段と、保守員からのポートの設定指示を受け付ける手段と、前記指示に基づいて前記クラスタ構成記憶システム内のポート設定を変更する手段とを有し、前記クラスタ構成記憶システム内のポート設定を変更する手段は、論理ポートと論理ボリュームとの間の仮想的なパスである論理パス毎の単位時間当りのデータ転送量またはデー転送時間を示すアクセス頻度を保持するアクセス情報を参照し、全論理パスの中でデータ転送量が規定値以上の論理パスの１つを選択し、選択した論理パスが記憶システムノード間のデータ転送を必要とするか否かをチェックし、記憶システムノード間のデータ転送を必要とした場合、前記論理パスを経て前記ホストコンピュータがアクセスしている論理ボリュームを格納している記憶装置を有する第１の記憶システムノードに未使用ポートがあり、その未使用ポートが当該ホストコンピュータに接続可能である場合、あるいは、当該ホストコンピュータに接続された未使用ポートがある場合に、前記ホストコンピュータが使用するノードを、それまで使用していた第２の記憶システムノードのポートから前記第１の記憶システムノードの前記未使用ポートとするように、ポート設定を変更することにより達成される。

【００１２】

前述において、アクセス情報は、例えば、記憶システムノード内部及び記憶システムノード相互間の単位時間内のデータ転送量またはデータ転送に必要な時間等でであり、アクセス命令を受信した論理ポートとデータの管理単位である論理ボリュームとの組み合わせ単位に採取する。このようにアクセス情報を採取することにより、どの論理ポートからどの論理ボリュームへのアクセスが記憶システムノード間のデータ転送を多く必要としているかが判る。そして、保守員は、保守端末を用いて前記アクセス情報を参照することができ、記憶システムノード相互間のデータ転送の頻度が高い論理ポートと論理ボリュームとの組み合わせを認識でき、その論理ポートの設定変更、その論理ボリュームの再配置の検討が可能となる。さらに、前記クラスタ構成記憶システムは、保守員からの指示により、ポート設定変更、記憶システムノード相互間での論理ボリュームの再配置を行い、前記記憶システムノード相互間のデータ転送の頻度をより小さくすることができ、これにより、クラスタ構成記憶システムのアクセス性能の向上を図ることができる。

【００１３】

また、本発明は、前記クラスタ構成記憶システム、あるいは、前記クラスタ構成記憶システム内の保守端末に、前記アクセス情報に基づいてポート設定変更及び論理ボリュームの再配置を検討する機能を持たせることができ、これにより、保守員の負担を減らすようにすることができる。

【００１４】
また、本発明は、クラスタ構成記憶システムに、前記クラスタ構成記憶システムを使用するホストコンピュータに対して前記アクセス情報を提供する機能、前記ホストコンピュータからポート設定指示、論理ボリュームの再配置指示を受け付ける機能を持たせることにより、ホストコンピュータあるいはホストコンピュータの管理者が、前記アクセス情報及び運用状態に基づいて、ポート設定変更、論理ボリュームの再配置を検討し、クラスタ構成記憶システムにポート設定指示及び論理ボリュームの再配置指示を行うようにすることができる。これにより、保守員には判断できない、高度な条件下でのポート設定変更、論理ボリュームの再配置を行うことが可能となり、クラスタ構成記憶システムのアクセス性能を向上させることができる。

【００１５】
【発明の実施の形態】
以下、本発明による記憶システムの実施形態を図面により詳細に説明する。以下に説明する本発明の実施形態は、複数の記憶システムを有するクラスタ構成記憶システムがアクセス情報を採取し、クラスタ構成の記憶システム内の保守端末を通じて保守員に提示し、このアクセス情報に基づく保守員の再配置指示により、アクセス命令を受信する論理ポートの変更、あるいは、クラスタ構成の記憶システム内でデータの記憶装置への再配置を行うものである。なお、以下の説明において、クラスタ構成の記憶システムを構成する記憶システムを記憶システムノードと呼ぶこととする。また、保守端末は、クラスタ構成の記憶システム内に備えられていても、外部に備えられていてもよく、いずれの場合も、各記憶システムノードに接続可能であればよい。

【００１６】
図１は本発明の一実施形態によるクラスタ構成の記憶システムの構成を示すブロック図、図２は記憶システムノードの構成を示すブロック図、図３はポート情報の例を説明する図、図４は物理位置情報の例を説明する図、図５は未使用記憶容量情報の例を説明する図、図６はアクセス情報の例を説明する図である。図１～図６において、２００はクラスタ構成記憶システム、２１０－１～２１０－ｎは記憶システムノード２１０、２２０は接続線、２３０は論理ボリューム、２４０はポート、３１０はデータ転送コントローラ（ＤＴＣ）、４００はポート情報、５００は物理位置情報、６００は未使用記憶容量情報、７００はアクセス情報であり、他の符号は図１５の場合と同一である。

【００１７】
本発明の一実施形態によるクラスタ構成記憶システム２００は、接続線２２０により相互に接続されている複数の記憶システムノード２１０－１～２１０－ｎ（以下の説明では、これらを区別する必要のない場合、あるいは、全体を示す場合、単に２１０と記す）と図示しない保守端末とから構成される。保守端末は、専用線を用いて全ての記憶システムノード２１０と接続されている。クラスタ構成記憶システム２００は、ポート情報４００、物理位置情報５００、未使用記憶容量情報６００、アクセス情報７００を有する。これらの情報は、保守端末から参照することができる。クラスタ構成記憶システム２００の記憶領域は、分割して管理されおり、分割した記憶領域を論理ボリューム２３０と呼ぶ。各論理ボリューム２３０のフォーマット形式及び記憶容量は、保守端末を用いて指定することができる。論理ボリューム２３０は、記憶システムノード２１０内であれば、複数の記憶装置１５０に分割して保持することができる。論理ボリューム２３０のクラスタ構成記憶システム２００内の物理的な格納位置（物理アドレス）は、後述する物理位置情報５００に保存されている。

【００１８】
記憶システムノード２１０は、基本的に図１５により説明したものと同様な内部構造を有し、記憶システムノード２１０と図１５により説明した記憶システム１００との差異は、記憶システムノード２１０が他の記憶システムノード２１０との通信のためのデータ転送制御コントローラ３１０を備えている点である。そして、記憶システムノード２１０は、１つ以上のホストアダプタ１１０、１つ以上のディスクアダプタ１２０、１つ以上のキャ

ッシュメモリ１３０、１つ以上の管理メモリ１４０、１つ以上の記憶装置１５０、２つ以上のコモンパス１６０、１つ以上の接続線１７０、１つ以上のデータ転送制御コントローラ３１０を備えて構成される。ホストアダプタ１１０、ディスクアダプタ１２０、キャッシュメモリ１３０、管理メモリ１４０、データ転送制御コントローラ３１０はコモンパス１６０により相互間が接続されている。コモンパス１６０は、コモンパス１６０の障害時のために２重化されてある。ディスクアダプタ１２０と記憶装置１５０とは接続線１７０によって接続されている。また、図示していないが、クラスタ構成記憶システム２００の設定、監視、保守等を行なうために保守端末（ＳＶＰ）が全てのホストアダプタ１１０とディスクアダプタ１２０とに専用線を用いて接続されている。

【００１９】

ホストアダプタ１１０は、図示しないホストコンピュータとキャッシュメモリ１３０との間のデータ転送を制御する。ホストアダプタ１１０は、ホストコンピュータとの接続のための複数のポート２４０を持ち、さらに、１つのポート２４０は、１つ以上の論理的なポート（以下、論理ポートとよぶ）を持つ。ディスクアダプタ１２０は、キャッシュメモリ１３０と記憶装置１５０との間のデータ転送を制御する。キャッシュメモリ１３０は、ホストコンピュータから受信したデータあるいは記憶装置１５０から読み出したデータを一時的に保持するメモリである。管理メモリ１４０は、クラスタ構成記憶システム２００内の全てのホストアダプタ１１０とディスクアダプタ１２０とが共有するメモリである。ホストアダプタ１１０及びディスクアダプタ１２０は、データ転送制御コントローラ３１０及び接続線２２０を用いて、他の記憶システムノード２１０内のホストアダプタ１１０、ディスクアダプタ１２０との通信及び他の記憶システムノード２１０内のキャッシュメモリ１３０、管理メモリ１４０の使用が可能である。

【００２０】

ポート情報４００は、論理ポートを有する記憶システムノード番号及びポート番号と、その論理ポートを使用してアクセスする論理ボリューム番号とホストコンピュータ番号とを保存する。ポート情報４００は、ホストアダプタ１１０が使用可能なメモリ、例えば、管理メモリ１４０あるいはホストアダプタ１１０の内部メモリに保存される。図３に示すポート情報４００の例において、ホストコンピュータと接続していないポートの論理ポートの論理ボリューム番号及びホストコンピュータ番号には、その論理ポートが未使用であることを表すために、論理ボリューム番号及びホストコンピュータ番号で使用しない数値が設定される。使用していない論理ポートの論理ボリューム番号には当該論理ポートが未使用であることを表すために、論理ボリューム番号で使用しない数値が設定される。

【００２１】

図３に示す例の場合、その論理ポートが使用されていないことを示すため“０”が使用されている。図３に示すポート情報において、論理ポート３の情報は、論理ポート３が記憶システムノード１のポート２内の論理ポートであることを示し、ホストコンピュータ２が論理ボリューム１をアクセスするために使用していることを示している。論理ポート５の情報は、論理ポート５が記憶システムノード２のポート１内の論理ポートであることを示し、記憶システムノード２のポート１はどのホストコンピュータにも接続されていないことを示している。また、論理ポート８の情報は、論理ポート８が記憶システムノード２のポート２内の論理ポートであることを示し、その論理ポートがどの論理ボリュームにも使用されていないことを示している。

【００２２】

物理位置情報５００は、論理ボリューム２３０の物理アドレス情報、フォーマット形式、容量、状態情報、再配置先物理アドレス情報、再配置完了位置、その論理ボリュームをアクセスするために使用する論理ポート番号を保持する。物理位置情報５００は、ホストアダプタ１１０から参照可能なメモリ、例えば、管理メモリ１４０あるいはホストアダプタ１１０の内部メモリに保存される。図４に示す物理位置情報５００において、物理アドレス情報は、クラスタ構成記憶システム２００内の論理ボリューム２３０の物理的な格納位置を示す情報であり、例えば、記憶システムノード番号と記憶システムノード内部での物

理位置とからなる。図示例では、これらがカンマによって区切られて示されている。状態情報は、正常、再配置中等の論理ボリューム２３０の論理的な状態を表す。再配置先物理アドレス情報と再配置完了位置とは、状態情報が再配置中のときのみ有効である。再配置先物理アドレス情報は、後述する論理ボリュームの再配置決定処理により求められた再配置先の論理ボリュームの物理アドレスである。再配置完了位置は、データの再配置処理が終了した論理ボリューム内の位置である。

【００２３】

前述において、状態情報が正常の場合、論理ボリュームの物理アドレスは、物理アドレス情報が用いられ、状態情報が再配置中の場合、論理ボリュームの物理アドレスは、アクセス対象のデータの論理ボリューム内での位置（アクセス命令内の論理アドレス）により物理アドレス情報、あるいは、再配置先物理アドレス情報のどちらか一方が使用される。例えば、論理ボリュームの先頭からデータの再配置処理を行う場合、論理アドレスが再配置完了位置より前であれば、アクセス対象のデータが既に再配置されているため、論理ボリュームの物理アドレスは、再配置先物理アドレス情報が用いられる。一方、論理アドレスが再配置完了位置より後の場合、アクセス対象のデータが再配置されていないため、論理ボリュームの物理アドレスは、物理アドレス情報が用いられる。

【００２４】

１つの論理ボリュームは、複数のホストコンピュータから同時にアクセスすることが可能であるため、１つの論理ボリュームの項目に対し、論理ポート番号が複数存在する。図４に示す例において、論理ボリューム１の情報は、記憶システムノード１内の先頭から０の位置から格納されており、フォーマット形式がＯＰＥＮ３、容量が３ＧＢ、データの再配置処理を行っておらず、論理ポート２、３を通じてホストコンピュータからアクセスされることを示している。また、論理ボリューム２の情報は、記憶システムノード２内の先頭から０の位置から格納されており、フォーマット形式がＯＰＥＮ６、容量が６ＧＢ、データの再配置処理を行っており、再配置先が記憶システムノード１内の先頭から５００の位置から格納されている論理ボリュームで、データの再配置処理が論理ボリュームの先頭から３００の位置まで終了しており、論理ポート１を通じてホストコンピュータからアクセスされることを示している。

【００２５】

未使用記憶容量情報６００は、記憶システムノードの未使用の記憶容量を保持する。未使用記憶容量情報６００は、例えば、管理メモリ１４０に保存される。この情報は、記憶システムノード２１０に新たに論理ボリュームを作成できる否かを調べるために使用される。図５に示す未使用容量情報６００は、例えば、記憶システムノード１に２０ＧＢの未使用の記憶容量があることを示している。

【００２６】

アクセス情報７００は、論理パス毎にアクセス頻度を保存する。論理パスとは、論理ポートと論理ボリュームとの間の仮想的なパスであると定義する。以下、論理パスの論理ポートを有する記憶システムノードを論理パスのフロントノード、論理パスの論理ボリュームを格納する記憶システムノードを論理パスのエンドノードと呼ぶ。アクセス頻度は、例えば、論理パスの論理ポートを使用して、所定の時間を単位時間として、単位時間、例えば、６０秒または３０秒内に論理パスの論理ボリュームをアクセスしている時間、データ量等であってよい。アクセス頻度は、ホストコンピュータからのアクセス命令の実行時にホストアダプタ１１０あるいはディスクアダプタ１２０が更新する。アクセス情報７００は、ホストアダプタ１１０あるいはディスクアダプタ１２０が使用可能なメモリ、例えば、管理メモリ１４０、ホストアダプタ１１０の内部メモリ、ディスクアダプタ１２０の内部メモリに保存される。さらに、アクセス情報７００は、予め指定された時間毎あるいは保守員の指示によって保守端末に転送され、予め指定された期間あるいは保守員の指示で保守端末内に保存される。保守端末に保存されたアクセス情報７００は、保守端末から参照することができ、保守員は、アクセス情報７００に基づいて、後述する論理ポートの設定変更決定処理及び論理ボリュームの再配置決定処理を実行する。

【００２７】
図６に示すアクセス情報７００の例において、例えば、アクセス頻度をデータ転送量とした場合、図６の論理パス１のアクセス頻度は、論理ポート１を使用して単位時間内に論理ボリューム２にアクセスするデータ量が２０（単位は、任意であるが、例えば、ＭＢ等であってよい）であることを示す。ポート情報４００が図３、物理位置情報５００が図４に示すようなものである場合、論理ポート１は、記憶システムノード１にあり、論理ボリューム２は記憶システムノード２にあるため、記憶システムノード相互間に２０のデータ転送が生じていることが判る。図６に示す論理パス２のアクセス頻度は、論理ポート２を使用して論理ボリューム１にアクセスするデータ量が１５であることを示す。ポート情報４００が図３、物理位置情報５００が図４に示すようなものである場合、論理ポート２は、記憶システムノード１にあり、論理ボリューム１は記憶システムノード１にあるため、記憶システムノード内部に１５のデータ転送が生じていることが判る。

【００２８】
図７はクラスタ構成記憶システム２００の動作を説明する図であり、以下、これについて説明する。まず、リード／ライト処理時のクラスタ構成記憶システム２００の動作について説明する。

【００２９】
ホストアダプタ１１０は、ホストコンピュータからアクセス命令７１０を受信する。このホストコンピュータからのアクセス命令７１０は、リード（またはライト）の命令、リード（またはライト）対象のデータの論理ボリューム２３０内での位置（論理アドレス）、データ量等を含んでいる。ホストアダプタ１１０は、アクセス命令７１０を受信すると、まず、アクセス対象のデータのクラスタ構成記憶システム２００内での物理的な格納位置（物理アドレス）を求める（物理位置算出処理７２０）。アクセス対象の物理アドレスは、アクセス対象を含む論理ボリューム２３０の物理アドレスとアクセス命令７１０内の論理アドレスとにより一意に決定される。

【００３０】
物理位置算出処理７２０の後、ホストアダプタ１１０は、アクセス処理７３０を行う。一例として、アクセス命令７１０がリード命令の場合で説明する。ホストアダプタ１１０は、このホストアダプタ１１０を有する記憶システムノード内あるいは物理位置算出処理７２０で求めた物理アドレスに対応するディスクアダプタ１２０を有する記憶システムノード内のキャッシュメモリ１３０にアクセス命令７１０内のデータ量と等しいメモリ領域を確保する。ホストアダプタ１１０は、物理位置算出処理７２０で求めた物理アドレスに対応するディスクアダプタ１２０に対し、そのキャッシュメモリ１３０へのアクセス対象のデータの読み出しを命令する。ホストアダプタ１１０から命令を受けたディスクアダプタ１２０は、記憶装置１５０からアクセス対象のデータをキャッシュメモリ１３０に読み出し、ホストアダプタ１１０に転送完了を報告する（アクセス処理７４０）。ホストアダプタ１１０は、キャッシュメモリ１３０からホストコンピュータにデータを送信し、リード処理が完了する。その後、ホストアダプタ１１０あるいはディスクアダプタ１２０は、アクセス対象の論理パスに対応するアクセス情報７００のアクセス頻度を変更する。このアクセス頻度の更新は、常時行ってもよいし、保守員が保守端末を用いてアクセス情報７００の更新の可否を指示してもよい。

【００３１】
保守員は、定期的あるいは必要に応じて、論理ポートの設定変更決定処理７７０を行い、論理ポートの設定変更による記憶システムノード間のデータ転送量の削減の可能性を検討する。保守員は、この検討の結果、削減が可能と判断すれば、保守端末を通じて、クラスタ構成記憶システム２００に再配置指示７５０を行って、再配置処理７６０を実行させる。

【００３２】
また、保守員は、定期的あるいは必要に応じて、論理ボリュームの再配置決定処理７８０を行い、論理ボリュームの再配置による記憶システムノード間のデータ転送量の削減の可

能性を検討する。保守員は、この検討の結果、削減が可能と判断すれば、保守端末を通じて、クラスタ構成記憶システム２００に論理ボリュームの再配置指示７５０を行って、再配置処理７６０を実行させる。

【００３３】

図８は前述した物理位置算出処理７２０での処理動作を説明するフローチャートであり、次に、これについて説明する。

【００３４】

（１）ホストアダプタ１１０は、まず、ポート情報４００を用いて、アクセス命令７１０を受信した論理ポート番号から論理ボリューム番号を求める（ステップ９００）。

【００３５】

（２）次に、論理ボリューム番号と物理位置情報５００とから論理ボリューム２３０の物理アドレスを算出するため、まず、アクセス対象の論理ボリューム２３０の状態情報を参照して、その論理ボリュームが再配置中か否かを調べる（ステップ９１０）。

【００３６】

（３）ステップ９１０の調べで、その論理ボリュームが再配置中であった場合、アクセス命令７１０の論理アドレスとアクセス対象の論理ボリューム２３０の再配置完了位置とを比較し、アクセス命令７１０の論理アドレスが再配置完了位置より前、すなわち、アクセス位置が再配置済みか否かを判定する（ステップ９２０）。

【００３７】

（４）ステップ９２０の判定で、アクセス位置が再配置済みであった場合、論理ボリューム２３０の物理アドレスとして、再配置先物理アドレス情報を求め、これを用いることを決定する（ステップ９３０）。

【００３８】

（５）ステップ９１０の調べで、その論理ボリュームが再配置中でなかった場合、または、ステップ９２０の判定で、アクセス命令７１０の論理アドレスが再配置完了位置より後、すなわち、アクセス位置が再配置済みでなかった場合、論理ボリューム２３０の物理アドレスとして、現在の物理アドレス情報を求め、これを用いることを決定する（ステップ９４０）。

【００３９】

（６）次に、ステップ９３０の処理またはステップ９４０の処理で求めた論理ボリューム２３０の物理アドレスにアクセス命令７１０の論理アドレスを加えてアクセス対象の物理アドレスを求める（ステップ９５０）。

【００４０】

図９は図７における論理ポートの設定変更決定処理７７０での処理動作を説明するフローチャート、図１０は論理ポートの設定変更決定処理７７０の中で作成される論理パス集合Ｂ１～Ｂ３の例を説明する図、図１３は論理ポートの設定変更決定処理７７０、後述する論理ボリュームの再配置決定処理７８０の中で作成される再配置指示について説明する図であり、以下、アクセス情報７００のアクセス頻度をデータ転送量として、これらについて説明する。

【００４１】

（１）アクセス情報７００を参照し、全論理パスの中でデータ転送量が規定値以上の論理パスがあるか否かをチェックし、なければ変更不要として処理を終了する(ステップ１０１０、１０２０）。

【００４２】

（２）ステップ１０１０のチェックで、データ転送量が規定値以上の論理パスがが複数存在する場合、例えば、データ転送量が大きいものから順に選択して以後の処理を行う。いま、選択した論理パスを第１論理パスと呼び、この第１論理パスの論理ポートを論理ポートＡ、論理ボリュームを論理ボリュームＡとし、論理ポートＡを有するポートをポートＡ、論理ポートＡを有する記憶システムノードをフロントノードＡ、論理ボリュームＡを格納する記憶システムノードをエンドノードＡとする。さらに、論理ポートＡを用いて論理

ボリュームＡをアクセスするホストコンピュータをホストコンピュータＡとする。これらの情報は、ポート情報４００及び物理位置情報５００から得ることができる。そして、第１論理パスが記憶システムノード間のデータ転送を必要とするか否かをチェックし、不要であった場合、変更不要として処理を終了する(ステップ１０１５、１０２０）。

【００４３】

（３）ステップ１０１５のチェックで、第１論理パスが記憶システムノード間のデータ転送を必要とした場合、ポート情報４００を用いて、エンドノードＡに未使用のポートが有り、そのポートとホストコンピュータＡとを新たに接続することができるかいなかを判定し、その未使用ポートとホストコンピュータＡとを新たに接続することができる場合、図１３に示す再配置指示７５０－Ｃを作成する（ステップ１０２５、１０３５）。

【００４４】

再配置指示７５０－Ｃは、論理ポート１と論理ポート２との２つのパラメータを持ち、論理ポート１を用いて論理ボリュームＡをアクセスすることをやめ、論理ポート２を用いて論理ボリュームＡをアクセスすることをクラスタ構成記憶システム２００に指示する命令であり、再配置指示７５０－Ｃの論理ポート１に論理ポートＡを、論理ポート２にそのポートの任意の論理ポートを設定する。

【００４５】

（４）ステップ１０２５の判定で、その未使用ポートとホストコンピュータＡとを新たに接続することができなかった場合、ポート情報４００を用いて、エンドノードＡにホストコンピュータＡと接続されている未使用の論理ポートがあるか否かをチェックし、未使用の論理ポートがあった場合、図１３に示す再配置指示７５０－Ｃを作成する。ここでは、再配置指示７５０－Ｃの論理ポート１に論理ポートＡを、論理ポート２に当該論理ポートを設定する（ステップ１０３０、１０３５）。

【００４６】

（５）ステップ１０３０のチェックで、エンドノードＡにホストコンピュータＡと接続されている未使用の論理ポートが存在しなかった場合、全論理パスから論理パスのフロントノードがエンドノードＡと等しく、論理パスの論理ボリュームをアクセスするホストコンピュータがホストコンピュータＡと等しい論理パス集合Ｂを取得する。さらに、論理パス集合Ｂを３つの集合Ｂ１～Ｂ３に分割する。論理パス集合Ｂ１は、論理パス集合Ｂのうち、論理パスのエンドノードがフロントノードＡと等しい論理パスとする。論理パス集合Ｂ２は、論理パス集合Ｂのうち、論理パスのエンドノードがエンドノードＡと等しくない論理パスとする。論理パス集合Ｂ３は、論理パス集合Ｂのうち、論理パスのエンドノードがエンドノードＡと等しい論理パスとする（ステップ１０４０）。

【００４７】

論理パス集合Ｂ１～Ｂ３の例を図１０に示しており、この図１０において、第１論理パスのフロントノードは２、エンドノードは１、論理ボリュームは１、ホストコンピュータはＡとする。論理パス集合Ｂ１の例は、フロントノード１、エンドノード２の論理パスである。論理パス集合Ｂ２の例は、フロントノード１、エンドノード３の論理パスである。論理パス集合Ｂ３の例は、フロントノード１、エンドノード１の論理パスである。前述した論理パス集合Ｂ１～Ｂ３は、この順に、論理パスの設定変更を行ったときに、記憶システムノード相互間のデータ転送量の削減効果が大きい。このことは、後述するたの論理パス集合においても同様である。

【００４８】

（６）ステップ１０４０の処理の後、論理パス集合Ｂ１が存在するか否かをチェックし、論理パス集合Ｂ１が存在する場合、論理パス集合Ｂ１の中で、例えば、最もデータ転送量が多い論理パス（以下、この論理パスを第２論理パスと呼ぶ）を選択する。これは、第１論理パスと第２論理パスとの論理ポートを入れ替えることにより、第１論理パスと第２論理パスとに関する記憶システムノード相互間のデータ転送量を削減することができるからである。このような第２論理パスが存在する場合、図１３に示す再配置指示７５０－Ｄを作成する（ステップ１０４５、１０６５）。

【００４９】
再配置指示７５０－Ｄは、論理ポート１と論理ポート２との２つのパラメータを持ち、それまで論理ポート１を使用してアクセスしていた論理ボリュームを論理ポート２を用いてアクセスし、それまで論理ポート２を使用してアクセスしていた論理ボリュームを論理ポート１を用いてアクセスすることをクラスタ構成記憶システム２００に指示する命令である。再配置指示７５０－Ｄの論理ポート１と論理ポート２に第２論理パスの論理ポート番号と第１論理パスの論理ポート番号を設定する。
【００５０】
（７）ステップ１０４５のチェックで、論理パス集合Ｂ１が存在しなかった場合、論理パス集合Ｂ２が存在するか否かをチェックし、論理パス集合Ｂ２が存在した場合、論理パス集合Ｂ２の中で、任意の論理パス（以下、この論理パスを第３論理パスと呼ぶ）を選択する。これは、第１論理パスと第３論理パスとの論理ポートを入れ替えることにより、第１論理パスに関する記憶システムノード間のデータ転送量を削減できるからである。このような第３論理パスが存在する場合、図１３に示す再配置指示７５０－Ｄを作成する。ここでは、再配置指示７５０－Ｄの論理ポート１と論理ポート２に第３論理パスの論理ポート番号と第１論理パスの論理ポート番号を設定する（ステップ１０５０、１０６５）。
【００５１】
（８）ステップ１０５０のチェックで、論理パス集合Ｂ２が存在しなかった場合、論理パス集合Ｂ３が存在するか否かをチェックし、論理パス集合Ｂ３が存在した場合、論理パス集合Ｂ３の中で、例えば、最もデータ転送量が少ない論理パス（以下、この論理パスを第４論理パスとよぶ）を選択する。但し、第４論理パスが第１論理パスよりデータ転送量が大きい場合は選択しない。これは、第１論理パスと第４論理パスとの論理ポートを入れ替えることにより、第４論理パスに関する記憶システムノード間のデータ転送量は増えるが、第１論理パスに関する記憶システムノード間のデータ転送量を削減することができるからである。このような第４論理パスが存在する場合は、図１３に示す再配置指示７５０－Ｄを作成する。ここでは、再配置指示７５０－Ｄの論理ポート１と論理ポート２に第４論理パスの論理ポート番号と第１論理パスの論理ポート番号を設定する（ステップ１０５５、１０６５）。
【００５２】
（９）ステップ１０５５のチェックで、論理パス集合Ｂ３が存在しなかった場合、あるいは、論理パス集合Ｂ３が存在したとしても第４論理パスが存在しなかった場合、論理ポートの設定変更による記憶システムノード間のデータ転送量の削減が不可能と判断して処理を終了する（ステップ１０６０）。
【００５３】
図１１は図７における論理ボリュームの再配置決定処理７８０での処理動作を説明するフローチャート、図１２は論理ボリュームの再配置決定処理７８０の中で作成されるノード間パス集合Ａ１～Ａ３の例を説明する図であり、以下、アクセス情報７００のアクセス頻度をデータ転送量であるとして、これらについて説明する。
【００５４】
（１）まず、アクセス情報７００から全てのノード間パスのデータ量を計算する。ノード間パスとは、命令を受信する論理ポートを有する記憶システムノードと論理ボリュームとの間の仮想的なパスであると定義する。また、ノード間パスからの命令を受信する論理ポートを有する記憶システムノードをフロントノードと呼び、ノード間パスの論理ボリュームを格納する記憶システムノードをエンドノードと呼ぶ。ノード間パスのデータ量は、データの再配置により変化する記憶システムノード間のデータ転送量であり、ノード間パスのフロントノード内の任意の論理ポートを使用してノード間パスの論理ボリュームにアクセスするデータ転送量の総和から、ノード間パスのエンドノード内の任意の論理ポートを使用してノード間パスの論理ボリュームにアクセスするデータ転送量の総和を引くことにより算出することができる。例えば、図３に示すポート情報４００、図４に示す物理位置情報５００、図６に示すアクセス情報７００を用いた場合のフロントノード４、論理ボリ

ューム５との間のノード間パスのデータ転送量は、論理パス９のデータ転送量と論理パス１０とのデータ転送量の和から論理パス６のデータ転送量を引いたものである。ノード間パスのデータ量が正の場合、ノード間パスの論理ボリュームに関する記憶システムノード間のデータ転送量が記憶システム内部のデータ転送量より大きく、論理ボリュームの再配置を行うことにより、記憶システムノード間のデータ転送量を削減することができることを意味する。また、ノード間パスのデータ量が負の場合、ノード間パスの論理ボリュームに関する記憶システムノード間のデータ転送量が記憶システム内部のデータ転送量より小さく、論理ボリュームの再配置を行うことにより、記憶システムノード間のデータ転送量が増えることを意味する（ステップ１２０５）。

【００５５】

（２）次に、全てのノード間パスの中でデータ転送量が規定値以上のノード間パスがあるか否かを調べ、該当するノード間パスがなければ、再配置不要として処理を終了する(ステップ１２１０、１２２０）。

【００５６】

（３）ステップ１２１０のチェックで、該当するノード間パスが存在し、それが複数存在する場合、例えば、データ転送量が大きいものから順に選択して以後の処理を行う。いま、選択したノード間パスを第１ノード間パスと呼び、この第１ノード間パスのフロントノードをフロントノードＡ、論理ボリュームを論理ボリュームＡとし、論理ボリュームＡを格納する記憶システムノードをエンドノードＡとする。これらの情報は、ポート情報４００及び物理位置情報５００から得ることができる。そして、第１ノード間パスが記憶システムノード間のデータ転送を必要とするか否かを判定し、不要であった場合、再配置不要として処理を終了する（ステップ１２１５、１２２０）。

【００５７】

（４）ステップ１２１５の判定で、第１ノード間パスが記憶システムノード間のデータ転送を必要とした場合、未使用記憶容量情報６００を用いて、フロントノードＡの未使用記憶容量が論理ボリュームＡの容量以上あるか否かをチェックして、フロントノードＡの未使用記憶容量が論理ボリュームＡの容量以上であった場合、図１３に示す再配置指示７５０－Ａを作成する（ステップ１２２５、１２３５）。

【００５８】

再配置指示７５０－Ａは、論理ボリューム１と記憶システムノード２との２つのパラメータをもち、記憶システムノード２に論理ボリューム１と同一フォーマット形式及び同一容量の論理ボリューム２を作成し、論理ボリューム１のデータを論理ボリューム２にコピーし、コピー終了後、論理ボリューム１の物理アドレス情報と論理ボリューム２の物理アドレス情報とを入れ替え、論理ボリューム１の記憶領域を開放することをクラスタ構成記憶システム２００に指示する命令である。ここでは、再配置指示７５０－Ａの論理ボリューム１に論理ボリュームＡを、記憶システムノード２にフロントノードＡを設定する。

【００５９】

（５）ステップ１２２５のチェックで、フロントノードＡの未使用領域が論理ボリュームＡの容量未満であった場合、全ノード間パスからノード間パスの論理ボリュームを格納する記憶システムノード２１０がフロントノードＡと等しく、ノード間パスの論理ボリュームのフォーマット形式及び容量が論理ボリュームＡと等しいノード間パス集合Ａを取得する。さらに、ノード間パス集合Ａを３つの集合Ａ１～Ａ３に分割する。ノード間パス集合Ａ１は、ノード間パス集合Ａのうち、ノード間パスのフロントノードがエンドノードＡと等しいノード間パスとする。ノード間パス集合Ａ２は、ノード間パス集合Ａのうち、ノード間パスのフロントノードがフロントノードＡと等しくないノード間パスとする。ノード間パス集合Ａ３はノード間パス集合Ａのうち、ノード間パスのフロントノードがフロントノードＡと等しいノード間パスとする（ステップ１２４０）。

【００６０】

ノード間パス集合Ａ１～Ａ３の例を図１２に示しており、この図１２において、第１ノード間パスは、フロントノード２、論理ボリューム１であるとする。ノード間パス集合Ａ１

の例は、フロントノード１、論理ボリューム１０のノード間パスである。ノード間パス集合Ａ２の例は、フロントノード３、論理ボリューム１２のノード間パスである。ノード間パス集合Ａ３の例は、フロントノード２、論理ボリューム１１のノード間パスである。

【００６１】

（６）ステップ１２４０の処理の後、ノード間パス集合Ａ１が存在するか否かをチェックし、ノード間パス集合Ａ１が存在する場合、ノード間パス集合Ａ１の中で、データ転送量が最も大きいノード間パス（以下、このノード間パスを第２ノード間パスとよぶ）を選択する。これは、第１ノード間パスの論理ボリュームＡを第２ノード間パスのエンドノードに格納し、第２ノード間パスの論理ボリュームをエンドノードＡに格納することにより、第１ノード間パスと第２ノード間パスに関する記憶システムノード間のデータ転送量を削減することができるからである。そして、このような第２ノード間パスが存在する場合、図１３に示す再配置指示７５０－Ｂを作成する（ステップ１２４５、１２８０）。

【００６２】

再配置指示７５０－Ｂは、論理ボリューム１と論理ボリューム２との２つのパラメータを持ち、論理ボリューム１のデータと論理ボリューム２のデータとを入れ替え、入れ替え終了後、論理ボリューム１の物理アドレス情報と論理ボリューム２の物理アドレス情報とを入れ替えることをクラスタ構成記憶システム２００に指示する命令である。ここでは、再配置指示７５０－Ｂの論理ボリューム１と論理ボリューム２に論理ボリュームＡと第２ノード間パスの論理ボリュームを設定する。

【００６３】

（７）ステップ１２４５のチェックで、ノード間パス集合Ａ１が存在しなかった場合、ノード間パス集合Ａ２が存在するか否かをチェックし、ノード間パス集合Ａ２が存在した場合、ノード間パス集合Ａ２の中で、任意のノード間パス（以下、このノード間パスを第３ノード間パスと呼ぶ）を選択する。これは、第１ノード間パスの論理ボリュームＡを第３ノード間パスのエンドノードに格納し、第三ノード間パスの論理ボリュームをエンドノードＡに格納することにより、第１ノード間パスに関する記憶システムノード間のデータ転送量を削減することができるからである。そして、このような第３ノード間パスが存在する場合、再配置指示７５０－Ｂを作成する。ここでは、再配置指示７５０－Ｂの論理ボリューム１と論理ボリューム２に論理ボリュームＡと第３ノード間パスの論理ボリュームを設定する（ステップ１２５０、１２８０）。

【００６４】

（８）ステップ１２５０のチェックで、ノード間パス集合Ａ２が存在しなかった場合、ノード間パス集合Ａ３が存在するかチェックし、ノード間パス集合Ａ３が存在した場合、ノード間パス集合Ａ３の中で、第１ノード間パスのデータ転送量より小さく、データ転送量が最も小さいノード間パス（以下、このノード間パスを第４ノード間パスとよぶ）を選択する。これは、第１ノード間パスの論理ボリュームＡを第４ノード間パスのエンドノードに格納し、第４ノード間パスの論理ボリュームをエンドノードＡに格納することにより、第４ノード間パスに関する記憶システムノード間のデータ転送量は増えるが、第１ノード間パスに関する記憶システムノード間のデータ転送量を削減することができるからである。このような第４ノード間パスが存在する場合、再配置指示７５０－Ｂを作成する。ここでは、再配置指示７５０－Ｂの論理ボリューム１と論理ボリューム２に論理ボリュームＡと第４ノード間パスの論理ボリュームを設定する（ステップ１２５５、１２８０）。

【００６５】

（９）ステップ１２５５のチェックで、ノード間パス集合Ａ３が存在しなかった場合、あるいは、存在しても、第４ノード間パスが存在しなかった場合、論理ボリュームの再配置による記憶システムノード間のデータ転送量の削減が不可能であると判断して処理を終了する（ステップ１２６０）。

【００６６】

次に、クラスタ構成記憶システム２００が再配置指示７５０を受信して実行する再配置処理７６０について説明する。クラスタ構成記憶システム２００は、再配置指示７５０－Ａ

を受信すると、記憶システムノード２に論理ボリューム１と同一フォーマット形式及び容量の論理ボリューム２を作成し、未使用記憶容量情報６００の記憶システムノード２の未使用容量から論理ボリューム１の容量を減じる。そして、後述する再配置処理７６０－Ｂと同様の方法で論理ボリューム１のデータを論理ボリューム２にコピーする。データのコピー終了後、論理ボリューム１の物理アドレス情報を論理ボリューム２の物理アドレス情報に変更し、論理ボリューム１の領域を開放し、未使用記憶容量情報６００の論理ボリューム１を格納する記憶システムノード１の未使用容量に論理ボリューム１の容量を加える。以上の処理は、論理ボリューム１へのアクセスを停止せずに行うことができる。

【００６７】

図１４はクラスタ構成記憶システム２００が再配置指示７５０－Ｂの受信時に実行する論理ボリュームの再配置処理７６０－Ｂでの処理動作を説明するフローチャートであり、以下、これについて説明する。すでに説明したように、再配置指示７５０－Ｂは、論理ボリューム１と論理ボリューム２とからなる。ここでは、論理ボリューム１を格納する記憶システムノード２１０をノードＡ、論理ボリューム２を格納する記憶システムノード２１０をノードＢとして説明する。

【００６８】

（１）まず、ノードＡ及びノードＢは、再配置処理の１回の処理単位分のメモリをキャッシュメモリ１３０（キャッシュメモリＡ、キャッシュメモリＢとする）に確保する。キャッシュメモリＡ及びキャッシュメモリＢは、どの記憶システムノード２１０のキャッシュメモリ１３０であってもよい（ステップ１５００）。

【００６９】

（２）次に、論理ボリューム１と論理ボリューム２の状態情報を「再配置中」に設定する。そして、論理ボリューム１と論理ボリューム２との再配置完了位置を先頭位置に初期化し、論理ボリューム１の再配置先物理アドレス情報を論理ボリューム２の物理アドレス情報に設定し、論理ボリューム２の再配置先物理アドレス情報を論理ボリューム１の物理アドレス情報に設定する（ステップ１５１０、１５２０）。

【００７０】

（３）論理ボリューム１と論理ボリューム２との再配置完了位置を調べ、全領域の再配置が完了しているか否かをチェックし、全領域の再配置が完了してした場合、論理ボリューム１と論理ボリューム２との物理アドレス情報を交換し、状態情報を正常に戻し、ステップ１５００で確保したキャッシュメモリ１３０を開放して処理を終了する（ステップ１５３０、１５７０）。

【００７１】

（４）ステップ１５３０のチェックで、全領域の再配置が完了していなかった場合、ノードＡのディスクアダプタ１２０が、論理ボリューム１の再配置完了位置が示しているデータ位置から再配置処理の１回の処理単位分のデータを記憶装置１５０からキャッシュメモリＡへ読み込む。同様に、ノードＢのディスクアダプタ１２０が、論理ボリューム２の再配置完了位置が示しているデータ位置から再配置処理の１回の処理単位分のデータを記憶装置１５０からキャッシュメモリＢへ読み込む（ステップ１５４０）。

【００７２】

（５）次に、ノードＡのディスクアダプタ１２０が、ステップ１５４０の処理でキャッシュメモリＢに格納されたデータを読み取り、そのデータを論理ボリューム１の再配置完了位置が示しているデータ位置に書き込む。同様に、ノードＢのディスクアダプタ１２０が、ステップ１５４０の処理でキャッシュメモリＡに格納されたデータを読み取り、そのデータを論理ボリューム２の再配置完了位置が示しているデータ位置に書き込む（ステップ１５５０）。

【００７３】

（６）ステップ１５５０の処理の後、１回の処理単位分だけ論理ボリューム１と論理ボリューム２との再配置完了位置を進めて、ステップ１５３０からの処理に戻って処理を繰り返す（ステップ１５６０）。

【００７４】
前述で説明した再配置処理は、論理ボリューム１及び論理ボリューム２へのアクセスを停止することなく行うことができる。
【００７５】
前述では、クラスタ構成記憶システム２００が再配置指示７５０－Ｂの受信時に実行する再配置処理について説明したが、前述の例において、別の再配置指示を受信した場合、クラスタ構成記憶システム２００は、次のように動作する。
【００７６】
クラスタ構成記憶システム２００は、再配置指示７５０－Ｃを受信すると、ポート情報４００の論理ポート２の論理ボリューム番号に論理ポート１の論理ボリュームＡを設定し、論理ポート１の論理ボリューム番号を未使用に変更する。クラスタ構成記憶システム２００は、変更終了後、保守端末を通じて、保守員に変更終了を知らせる。以上の処理中、論理ボリュームＡへのアクセスを停止する必要がある。
【００７７】
クラスタ構成記憶システム２００は、再配置指示７５０－Ｄを受信すると、ポート情報４００の論理ポート１の論理ボリューム番号を論理ポート２の論理ボリュームＡに変更し、論理ポート２の論理ボリューム番号を論理ポート１の論理ボリュームＢに変更する。クラスタ構成記憶システム２００は、変更終了後、保守端末を通じて、保守員に変更終了を知らせる。以上の処理中、論理ボリュームＡ及び論理ボリュームＢへのアクセスを停止する必要がある。
【００７８】
保守員は、新たに論理ポートに論理ボリュームを設定する場合、保守端末を用いて全論理パスのアクセス情報７００を参照し、極力記憶システムノード相互間のデータ転送量が増加しないように、論理ポートの設定を行って論理ボリュームを作成する。
【００７９】
次に、ホストコンピュータＡが未使用の論理ポートを用いて既存の論理ボリュームＡを使用する場合について説明する。
【００８０】
まず、論理ボリュームＡを格納している記憶システムノードＡを調べる。次に、記憶システムノードＡに未使用のポートが存在するか否かを調べる。記憶システムノードＡに未使用のポートが存在し、ホストコンピュータＡとそのポートとを新たに接続することができる場合、記憶システムノード相互間のデータ転送が生じないため、ホストコンピュータＡとその未使用ポートとを接続し、そのポートの任意の論理ポートに論理ボリュームＡを設定する。条件を満たすポートが存在しない場合、記憶システムノードＡ内のホストコンピュータＡと接続されたポートに未使用の論理ポートが存在するか否かを調べる。条件を満たす論理ポートが存在する場合、記憶システムノード相互間のデータ転送が生じないため、その論理ポートに論理ボリュームＡを設定する。
【００８１】
前述で条件を満たす論理ポートが存在しない場合、必ず記憶システムノード間のデータ転送が生じるため、任意の未使用の論理ポートに論理ボリュームＡを設定する。この場合、保守員が前述した検討を行わずに、保守端末から論理ボリュームＡ及び使用してよいポートを指定してもよく、また、保守端末が前述の検討を行って、処理結果を保守端末上に表示させ、処理結果の論理ポートに論理ボリュームＡを設定してもよい。さらに、保守員の代わりに、保守端末が自動的に検討結果の論理ポートに論理ボリュームＡを設定してもよい。
【００８２】
次に、ホストコンピュータＡが新たに作成する論理ボリュームＡを使用する場合について説明する。
【００８３】
まず、論理ボリュームＡの容量以上の未使用記憶容量を有する記憶システムノード集合Ａ

を調べる。次に、記憶システムノード集合Ａ内に未使用のポートを有する記憶システムノードＢを調べる。そして、記憶システムノードＢが存在し、ホストコンピュータＡと記憶システムノードＢの有する未使用のポートＢとを新たに接続することができる場合、記憶システムノード相互間のデータ転送が生じないため、記憶システムノードＢに論理ボリュームＡを作成し、ホストコンピュータＡとポートＢとを接続し、ポートＢの任意の論理ポートに論理ボリュームＡを設定する。条件を満たす記憶システムノードが存在しない場合、記憶システムノード集合Ａ内に、ホストコンピュータＡと接続され、ポート内に未使用の論理ポートを有する記憶システムノードＣを調べる。記憶システムノードＣが存在する場合、記憶システムノード相互間のデータ転送が生じないため、記憶システムノードＣに論理ボリュームＡを作成し、記憶システムノードＣの未使用の論理ポートに論理ボリュームＡを設定する。記憶システムノードＣが存在しない場合、必ず記憶システムノード相互間のデータ転送が生じるため、記憶システムノード集合Ａの任意の記憶システムノードに論理ボリュームＡを作成し、任意の未使用の論理ポートに論理ボリュームＡを設定する。この場合、保守員が前記検討を行わずに、保守端末に作成する論理ボリュームのフォーマット形式及び容量と使用してよいポートを指定してもよく、また、保守員が保守端末に前記検討を行わせ、検討結果を保守端末上に表示させて、検討結果の記憶システムノードに論理ボリュームＡを作成し、検討結果の論理ポートに論理ボリュームＡを設定してもよい。さらに、保守員の代わりに、保守端末が自動的に、検討結果の記憶システムノードに論理ボリュームＡを作成し、検討結果の論理ポートに論理ボリュームＡを設定してもよい。

【００８４】

前述で説明した本発明の実施形態は、複数の記憶システムを１つの記憶システムとして運用可能とするクラスタ構成記憶システム２００において、論理パス毎にアクセス情報７００を採取し、保守端末を用いて保守員にアクセス情報７００を提示し、論理ポートの設定変更決定処理７７０及び論理ボリュームの再配置決定処理７８０に基づく保守員の判断により、論理ポートの設定変更処理及び論理ボリュームの再配置処理を行わせることができるので、クラスタ構成記憶システム２００を構成する記憶システムノード相互間の通信負荷を抑えることができ、クラスタ構成記憶システム２００のアクセス性能の向上を図ることができる。

【００８５】

前述した本発明の実施形態は、論理ポートの設定変更及び論理ボリュームの再配置処理の両方を行うものとして説明したが、本発明は、前述した本発明の実施形態に比較してその効果が小さくなるが、いずれか一方の処理を行うだけてあってもよい。そして、この場合、システムの構成を簡易化することができるので、システム全体のコストの低減を図ることができる。

【００８６】

次に、前述で説明した本発明の実施形態の変形例について説明する。この変形例は、保守員が論理ポートの設定変更決定処理７７０を行うのではなく、クラスタ構成記憶システム２００の保守端末、クラスタ構成記憶システム２００内の１つのホストアダプタ１１０あるいはディスクアダプタ１２０がアクセス情報７００を参照して論理ポートの設定変更決定処理７７０を行い、論理ポートの設定変更による記憶システムノード相互間のデータ転送量の削減が可能な場合に、処理結果のポートの設定変更をホストコンピュータの管理者あるいは保守員に提案するようにしたものである。ホストコンピュータの管理者あるいは保守員は、運用上問題であると判断したら、提案されたポート設定を実行する。

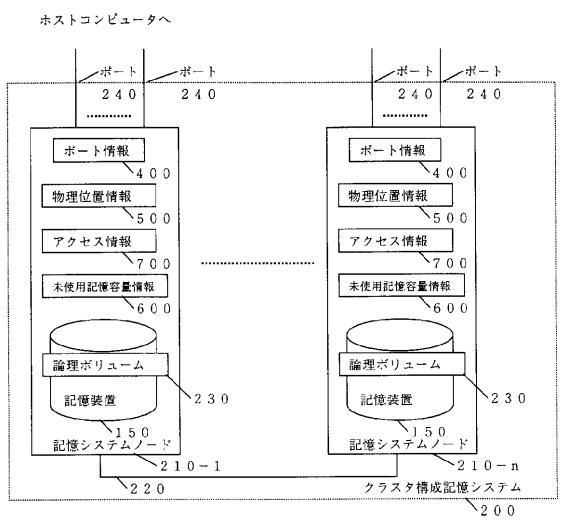
【００８７】

次に、前述で説明した本発明の実施形態の他の変形例について説明する。この他の変形例は、保守員が論理ボリュームの再配置決定処理７８０を行うのではなく、クラスタ構成記憶システム２００の保守端末、クラスタ構成記憶システム２００内の１つのホストアダプタ１１０あるいはディスクアダプタ１２０がアクセス情報７００を参照して論理ボリュームの再配置決定処理７８０を行い、論理ボリュームの再配置処理を自動的に行うようにしたものである。

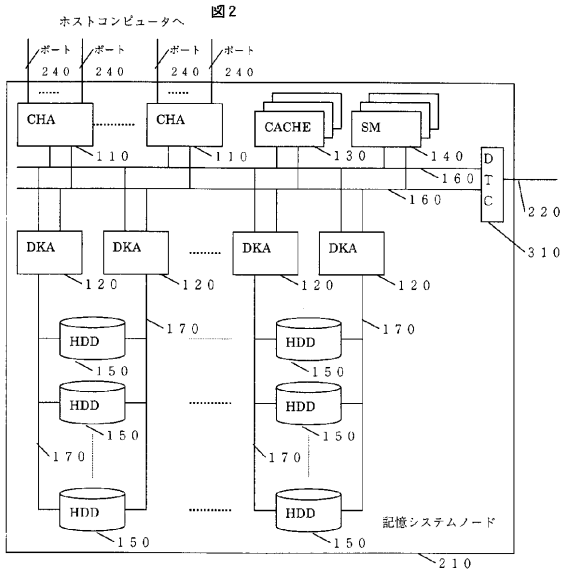
【００８８】
前述した本発明の実施形態の２つの変形例によれば、保守員の負担を低減することが可能となる。
【００８９】
次に、前述で説明した本発明の実施形態のさらに他の変形例について説明する。このさらに他の変形例は、クラスタ構成記憶システム２００がアクセス情報７００を採取し、クラスタ構成記憶システム２００を使用するホストコンピュータにポート情報４００、物理位置情報５００、未使用記憶容量情報６００、アクセス情報７００を提供し、ホストコンピュータからの再配置指示７５０により、再配置処理７６０を行うようにしたものである。そして、ホストコンピュータあるいはホストコンピュータの管理者は、前述で説明した実施形態の場合と同様に、論理ポートの設定変更決定処理７７０と論理ボリュームの再配置決定処理７８０と運用状況とによりポートの設定変更及び論理ボリュームの再配置を行うか否かを決定する。これにより、保守員には判断することが困難な高度な条件下でのポートの設定変更及びデータの再配置を行うことが可能となる。例えば、負荷が低いときに論理ボリュームの再配置を行う等の状況に応じた運用が可能となる。
【００９０】
【発明の効果】
以上説明したように本発明によれば、クラスタ構成記憶システムを構成する記憶システムノード相互間のデータ転送の頻度、データ転送量を削減することができ、クラスタ構成記憶システムのアクセス性能を向上させることができる。
【図面の簡単な説明】
【図１】本発明の一実施形態によるクラスタ構成の記憶システムの構成を示すブロック図である。
【図２】記憶システムノードの構成を示すブロック図である。
【図３】ポート情報の例を説明する図である。
【図４】物理位置情報の例を説明する図である。
【図５】未使用記憶容量情報の例を説明する図である。
【図６】アクセス情報の例を説明する図である。
【図７】クラスタ構成記憶システムの動作を説明する図である。
【図８】物理位置算出処理での処理動作を説明するフローチャートである。
【図９】図７における論理ポートの設定変更決定処理での処理動作を説明するフローチャートである。
【図１０】論理ポートの設定変更決定処理の中で作成される論理パス集合Ｂ１～Ｂ３の例を説明する図である。
【図１１】図７における論理ボリュームの再配置決定処理での処理動作を説明するフローチャートである。
【図１２】論理ボリュームの再配置決定処理の中で作成されるノード間パス集合Ａ１～Ａ３の例を説明する図である。
【図１３】論理ポートの設定変更決定処理、論理ボリュームの再配置決定処理の中で作成される再配置指示について説明する図である。
【図１４】クラスタ構成記憶システムが再配置指示の受信時に実行する論理ボリュームの再配置処理での処理動作を説明するフローチャートである。
【図１５】従来技術による記憶システムの構成例を示すブロック図である。
【符号の説明】
１００　記憶システム
１１０　ホストアダプタ（ＣＨＡ）
１２０　ディスクアダプタ（ＤＫＡ）
１３０　キャッシュメモリ（ＣＡＣＨＥ）
１４０　管理メモリ（ＳＭ）
１５０　記憶装置（ＨＤＤ）

１６０　コモンパス
１７０　接続線
２００　クラスタ構成記憶システム
２１０－１～２１０－ｎ　記憶システムノード
２２０　接続線
２３０　論理ボリューム
２４０　ポート
３１０　データ転送コントローラ（ＤＴＣ）
４００　ポート情報
５００　物理位置情報
６００　未使用記憶容量情報
７００　アクセス情報

【図１】



【図２】



【図３】

図3

論理ポート番号	記憶システムノード番号	ポート番号	論理ボリューム番号	ホストコンピュータ番号
1	1	1	2	1
2	1	1	1	1
3	1	2	1	2
4	1	2	3	2
5	2	1	0	0
6	2	1	0	0
7	2	2	4	2
8	2	2	0	2
9	3	1	5	1
10	3	1	6	1
11	3	2	6	3
12	3	2	0	3
13	4	1	5	1
14	4	1	0	1
15	4	2	5	4
16	4	2	0	4

400　ポート情報

【図４】

図4

論理ボリューム番号	物理アドレス情報	フォーマット形式	容量（GB）	状態情報	再配置先物理アドレス情報	再配置完了位置	論理ポート番号
1	1,0	OPEN3	3	正常	0,0	0	2,3
2	2,0	OPEN6	6	再配置中	1,500	300	1
3	1,500	OPEN6	6	再配置中	2,0	300	4
4	2,1000	OPEN3	3	正常	0,0	0	7
5	3,0	OPEN3	3	正常	0,0	0	9,13,15
6	3,500	OPEN3	3	正常	0,0	0	10,11

500　物理位置情報

【図５】

図5

記憶システムノード番号	未使用容量（GB）
1	20
2	40
3	10
4	0

600　未使用記憶容量情報

【図６】

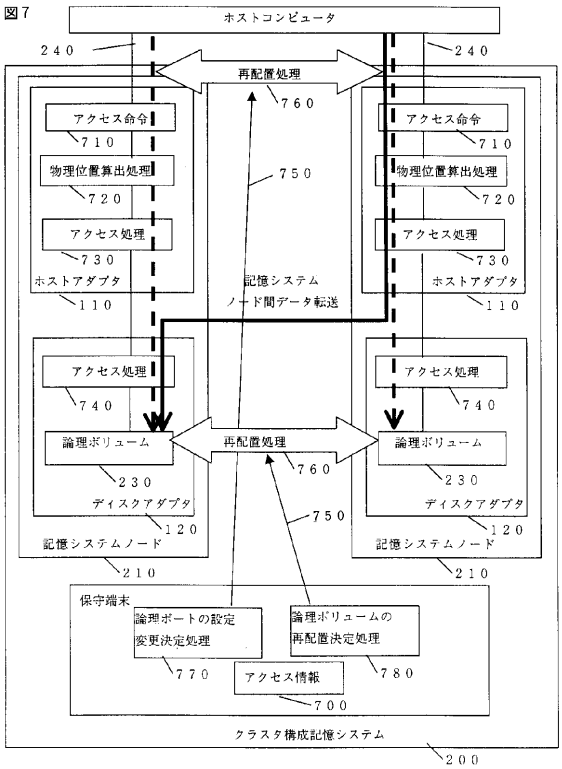
図6

論理パス番号	論理ポート番号	論理ボリューム番号	アクセス頻度
1	1	2	20
2	2	1	15
3	3	1	15
4	4	3	10
5	7	4	13
6	9	5	10
7	10	6	5
8	11	6	2
9	13	5	11
10	15	5	6

700　アクセス情報

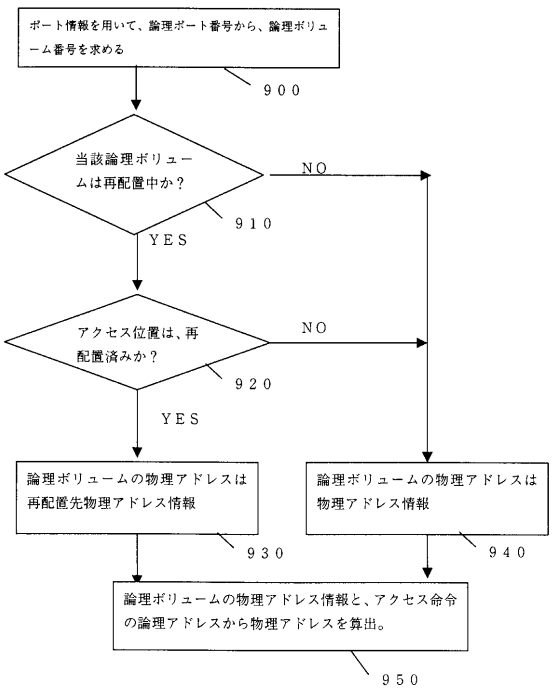
【図７】

図7



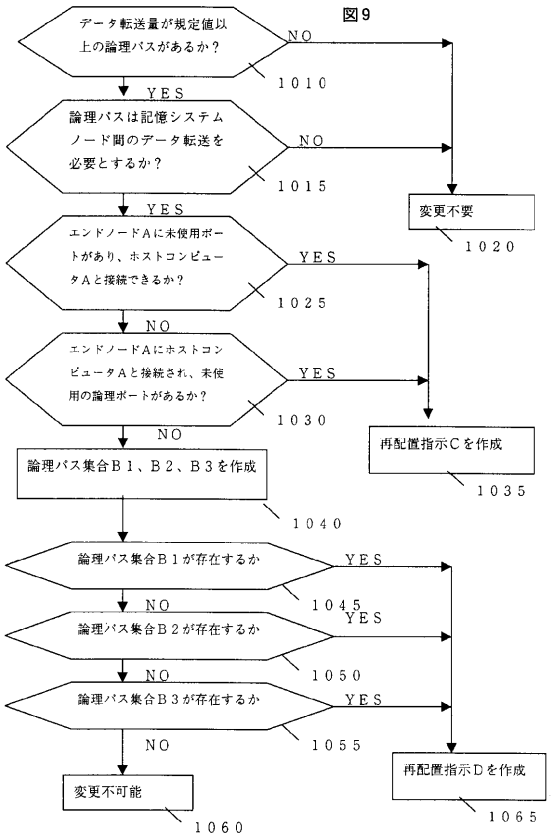
【図８】

図8



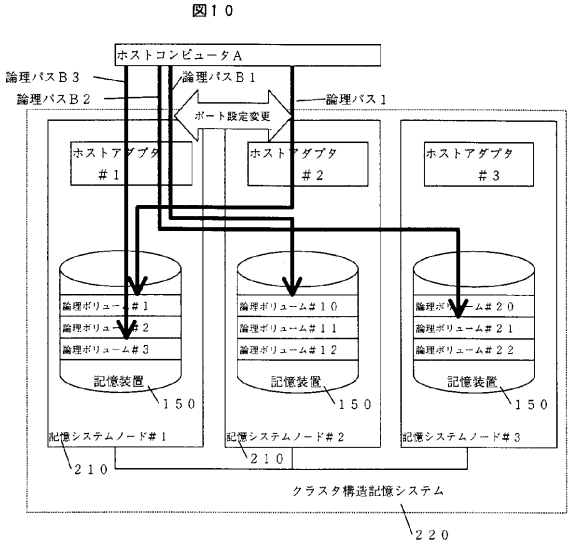
【図９】

図9



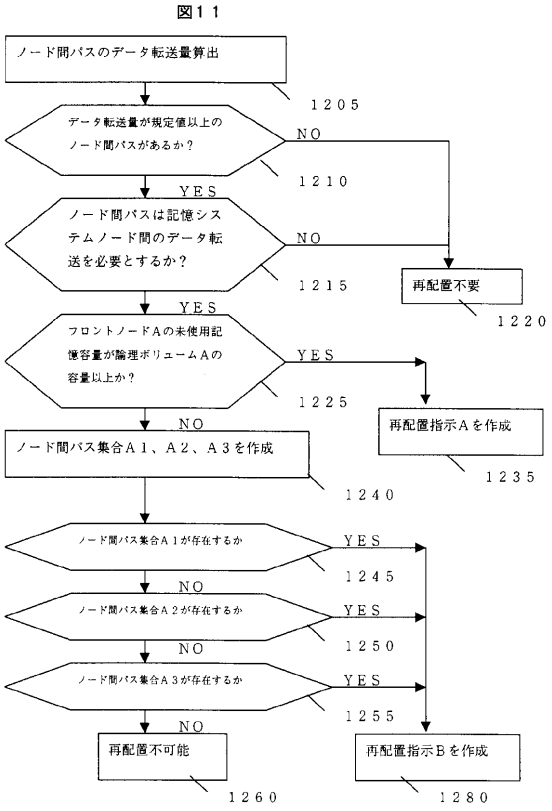
【図１０】

図10



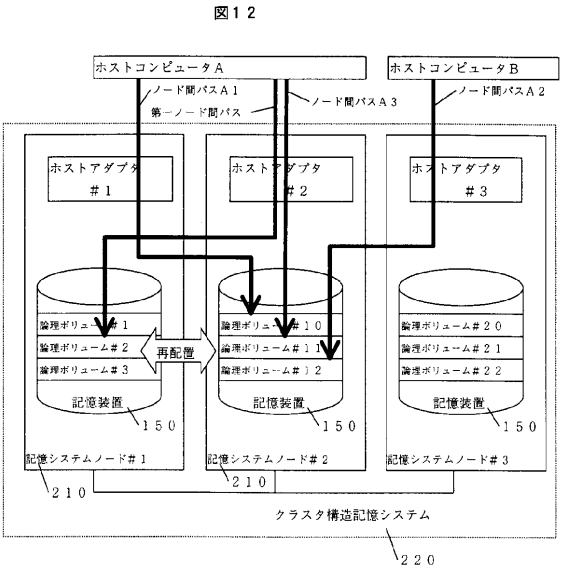
【図１１】

図11



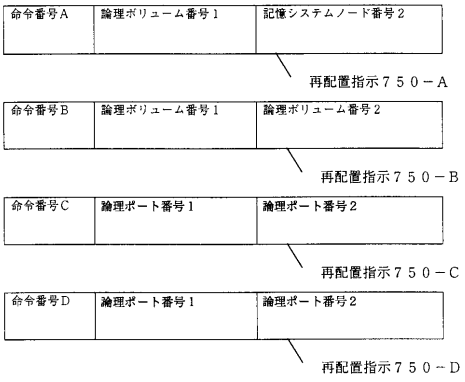
【図１２】

図12



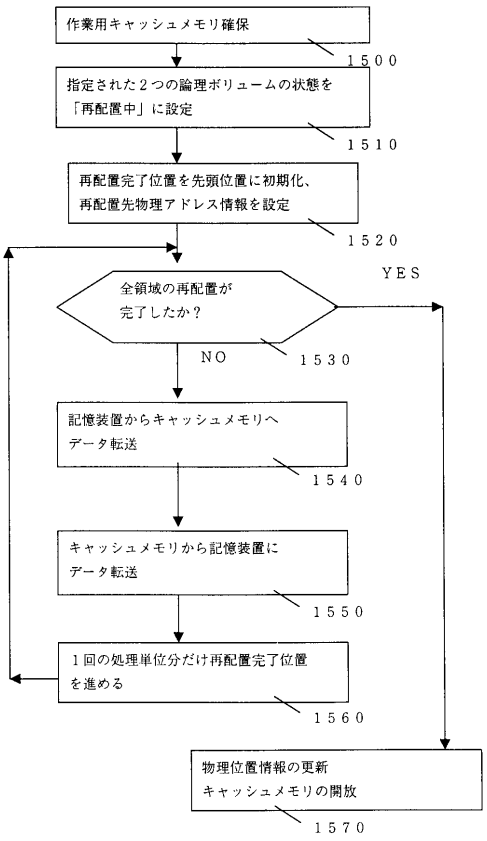
【図１３】

図１３



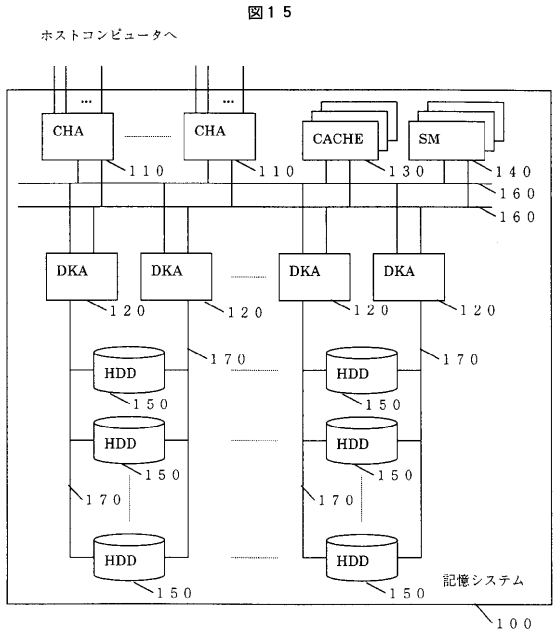
【図１４】

図１４



【図１５】

図１５



フロントページの続き

(72)発明者　大枝　高
神奈川県川崎市麻生区王禅寺１０９９番地　株式会社　日立製作所　システム開発研究所内
(72)発明者　荒井　弘治
神奈川県小田原市中里３２２番地２号　株式会社　日立製作所　ＲＡＩＤシステム事業部内

審査官　浜岸　広明

(56)参考文献　特開平１１－０９５９３４（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
G06F 12/00