(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第3641872号

(P3641872)

(45)発行日　平成17年4月27日(2005.4.27)　(24)登録日　平成17年2月4日(2005.2.4)

(51)Int.Cl.[7]	F I		
G06F 3/06	G06F	3/06	540
G06F 12/08	G06F	3/06	301X
	G06F	3/06	302E
	G06F	3/06	304N
	G06F	12/08	320

請求項の数 12　(全 16 頁)

(21)出願番号　特願平8-85370
(22)出願日　平成8年4月8日(1996.4.8)
(65)公開番号　特開平9-274544
(43)公開日　平成9年10月21日(1997.10.21)
審査請求日　平成12年3月10日(2000.3.10)
審判番号　不服2002-4389(P2002-4389/J1)
審判請求日　平成14年3月14日(2002.3.14)

(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　100095511
弁理士　有近　紳志郎
(72)発明者　山本　康友
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　山本　彰
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】記憶装置システム

(57)【特許請求の範囲】

【請求項１】

データ処理装置が直接アクセスを行う論理的記憶装置が配置された物理的記憶装置を含む複数の物理的記憶装置と、前記データ処理装置と前記複数の物理的記憶装置の間のデータ転送を制御する記憶制御装置とを有する記憶装置システムにおいて、

予め定めた指標に基づいて、前記論理的記憶装置を、前記複数の物理的記憶装置のうち、該論理的記憶装置が配置された物理的記憶装置とは異なる物理的記憶装置に再配置する際に、再配置先の物理的記憶装置に、前記論理的記憶装置に格納されたデータ全体を連続的に格納する論理的記憶装置再配置手段を有することを特徴とする記憶装置システム。

【請求項２】

データ処理装置が直接アクセスを行う論理的記憶装置が配置された物理的記憶装置を含む複数の物理的記憶装置と、前記データ処理装置と前記複数の物理的記憶装置の間のデータ転送を制御する記憶制御装置とを有する記憶装置システムにおいて、

前記データ転送の制御の運用中にデータ処理装置の論理的記憶装置へのアクセス情報を指標として採取するアクセス情報採取手段と、前記指標に基づいて前記論理的記憶装置を、前記複数の物理的記憶装置のうち、該論理的記憶装置が配置された物理的記憶装置とは異なる物理的記憶装置に再配置する際に、再配置先の物理的記憶装置に、前記論理的記憶装置に格納されたデータ全体を連続的に格納する論理的記憶装置再配置手段とを有することを特徴とする記憶装置システム。

【請求項３】

請求項２に記載の記憶装置システムにおいて、前記アクセス情報が、前記データ処理装置から前記論理的記憶装置へのアクセス頻度情報を含むことを特徴とする記憶装置システム。

【請求項４】
請求項２または請求項３に記載の記憶装置システムにおいて、前記アクセス情報が、前記データ処理装置から前記論理的記憶装置へのアクセスパターン情報を含むことを特徴とする記憶装置システム。

【請求項５】
請求項１に記載の記憶装置システムにおいて、前記指標が、前記論理的記憶装置に求められる信頼性であることを特徴とする記憶装置システム。

【請求項６】
請求項１から請求項５のいずれかに記載の記憶装置システムにおいて、前記指標を保守員に提示する指標提示手段と、保守員からの再配置指示を受け付ける再配置指示受付手段とを具備したことを特徴とする記憶装置システム。

【請求項７】
請求項１から請求項５のいずれかに記載の記憶装置システムにおいて、データ処理装置からの再配置指示を受け付ける再配置指示受付手段を具備したことを特徴とする記憶装置システム。

【請求項８】
複数の物理的記憶装置と、データ処理装置と前記複数の物理的記憶装置の間のデータ転送を制御する記憶制御装置とを有する記憶装置システムにおいて、
前記複数の物理的記憶装置のうちの第一の物理的記憶装置群にまたがって設定され、前記データ処理装置が直接アクセスを行う論理的記憶装置を有し、
前記記憶制御装置は、予め定めた指標に基づいて、前記第一の物理的記憶装置群とは異なる第二の物理的記憶装置群を選択すると共に、前記第二の物理的記憶装置群に、前記論理的記憶装置に格納されたデータ全体を連続的に格納する論理的記憶装置再配置手段を有することを特徴とする記憶装置システム。

【請求項９】
請求項１から請求項８のいずれかに記載の記憶装置システムにおいて、再配置中の論理的記憶装置にデータ処理装置からのアクセスがあったとき、再配置中の論理的記憶装置の再配置完了領域と再配置未完領域とを識別し、前記アクセス位置が前記再配置完了領域ならば再配置先の物理的記憶装置にアクセスさせ、前記アクセス位置が前記再配置未完領域ならば再配置前の物理的記憶装置にアクセスさせるアクセス位置切替手段をさらに具備したことを特徴とする記憶装置システム。

【請求項１０】
データ処理装置が直接アクセスを行う論理的記憶装置が配置された物理的記憶装置を含む複数の物理的記憶装置と、前記データ処理装置と前記複数の物理的記憶装置の間のデータ転送を制御する記憶制御装置とを有する記憶装置システムにおいて、
前記記憶制御装置は、前記データ処理装置によるアクセス状況を取得し、前記アクセス状況に基づいて前記論理的記憶装置のデータを、前記複数の物理的記憶装置のうち、該論理的記憶装置に配置されている物理的記憶装置から、該論理的記憶装置が配置された物理的記憶装置とは異なる物理的記憶装置に移動させることを特徴とする記憶装置システム。

【請求項１１】
キャッシュメモリを有し、
前記記憶制御装置は、前記第一の物理的記憶装置群から前記キャッシュメモリへデータを読み出す手段と、前記読み出されたデータを前記第二の物理的記憶装置群に書き込む手段とを有することを特徴とする請求項８記載の記憶装置システム。

【請求項１２】
前記第二の物理的記憶装置群は、前記第一の物理的記憶装置群よりもデータの読み出し速度が高速であることを特徴とする請求項８又は請求項１１記載の記憶装置システム。

【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は、記憶装置システムに関し、さらに詳しくは、シーケンシャルアクセスの場合やランダムアクセルでヒット率が低い場合でもアクセス性能を向上することが出来る記憶装置システムおよびデータの信頼性を向上することが出来る記憶装置システムに関する。
特に、本発明は、ディスクアレイ向きの高機能ディスク装置、その高機能ディスク装置とディスク制御装置とにより構成される記憶装置サブシステム、およびその記憶装置サブシステムとデータ処理装置とにより構成される情報処理システムに有用である。
【０００２】
【従来の技術】
シカゴのイリノイ大学で開かれた「ACM SIGMOD」会議において発表された論文「D.Patterson,G.gibson,and R.H.Kartz;A Case for Redundant Arrays of Inexpensive Disks (RAID),ACM SIGMOD Conference,Chicago,IL,(June 1988),pp.109-116」は、ディスクアレイ上のデータ配置に関する技術を開示している。
【０００３】
また、特開平７－８４７３２号公報では、ディスク装置の一部をディスクキャッシュの如く用いる技術が開示されている。具体的には、ディスク装置を一時的にデータを格納するテンポラリ領域と最終的にデータを書き込む領域とに分け、更新データはパリティを生成せずに一旦テンポラリ領域に二重書きし、非同期にパリティ生成し、最終領域に書き込む。
【０００４】
一方、電気情報通信学会技術研究報告「ＤＥ９５－６８（茂木他：Hot Mirroring を用いたディスクアレイのディスク故障時の性能評価、１９９５年１２月、電気情報通信学会技報 Vol.95-No.407、pp.19-24）」には、アクセス頻度の違いにより、データを保持するＲＡＩＤレベルを動的に変更する技術が開示されている。具体的には、ディスク装置をＲＡＩＤ１構成の部分とＲＡＩＤ５構成の部分に分け、ライトアクセスのあったデータを優先的にＲＡＩＤ１構成の部分に格納するようにデータの格納位置を動的に変更することにより、アクセス頻度の高いデータはＲＡＩＤ１構成の部分に格納し、アクセス頻度の低いものはＲＡＩＤ５構成の部分に格納するように出来る。
この技術によれば、記憶容量の異なる物理ディスク装置やＲＡＩＤレベルの異なる物理ディスク装置を記憶装置サブシステム内で混在させることが可能であり、論理ディスク装置内のデータを、そのアクセス頻度やアクセスパターンなどの指標に基づいて、任意の物理ディスク装置に格納することが出来る。また、アクセス頻度の高いデータを、より高速な物理ディスク装置に格納するように、動的に格納位置を変更することも出来る。
なお、ＲＡＩＤ１のディスクアレイは、データ処理装置からの書き込みデータに対して、その複製をミラーと呼ばれる副ディスク装置に書き込み、データの信頼性を確保する。冗長データが元のデータの複製であるため、冗長データ作成のオーバヘッドが小さく、アクセス性能が良い。但し、物理的記憶装置の使用効率は、５０％と低い。一方、ＲＡＩＤ５のディスクアレイは、データ処理装置からの複数の書き込みデータに対して、パリティと呼ばれる冗長データを作成する。パリティ作成時に更新前データと更新前パリティのリードが必要なため、冗長データ作成のオーバヘッドが大きく、アクセス性能は悪い。但し、複数のデータに対して１つのパリティを作成するため、記憶装置の使用効率はＲＡＩＤ１に比べ高い。
【０００５】
【発明が解決しようとする課題】
上記従来技術では、アクセスするデータ単位でデータの格納位置の変更を行うため、データ処理装置が直接アクセスを行う論理ディスク装置上では連続なデータが、実際にデータを記憶する物理ディスク装置上では非連続となってしまう。このため、一連のデータをリード／ライトするシーケンシャルアクセスの場合、実際には複数データをまとめてリード

／ライトできなくなり、アクセス性能の低下を招く問題点がある。

【０００６】

一方、上記報告「ＤＥ９５－６８」の従来技術では、ライトの度に、アクセス頻度が低いと判断したデータをＲＡＩＤ１構成の部分からＲＡＩＤ５構成の部分に移し、空いたＲＡＩＤ１構成の部分にライトデータを書き込むため、アクセスパターンがランダムアクセルでヒット率が低い場合には、ＲＡＩＤ１構成の部分に移したデータの多くは再びＲＡＩＤ５構成の部分に戻されることになる。このため、ヒット率が低い場合、アクセス性能の向上は期待できず、逆にデータを移す処理のオーバヘッドがアクセス性能の低下を引き起こす問題点がある。

【０００７】

また、上記の従来技術では、データの信頼性の向上については全く考慮されていない問題点がある。

【０００８】

そこで、本発明の第１の目的は、シーケンシャルアクセスの場合やランダムアクセルでヒット率が低い場合でも、アクセス性能を向上することが出来る記憶装置システムを提供することにある。

また、本発明の第２の目的は、データの信頼性を向上することが出来る記憶装置システムを提供することにある。

【０００９】

【課題を解決するための手段】

第１の観点では、本発明は、データ処理装置が直接アクセスを行う論理的記憶装置が配置された物理的記憶装置を含む複数の物理的記憶装置と、前記データ処理装置と前記複数の物理的記憶装置の間のデータ転送を制御する記憶制御装置とを有する記憶装置システムにおいて、予め定めた指標に基づいて、前記論理的記憶装置を、前記複数の物理的記憶装置のうち、該論理的記憶装置が配置された物理的記憶装置とは異なる物理的記憶装置に再配置する際に、再配置先の物理的記憶装置に、前記論理的記憶装置に格納されたデータ全体を連続的に格納する論理的記憶装置再配置手段を有することを特徴とする記憶装置システムを提供する。上記第１の観点による記憶装置システムでは、アクセスするデータ単位でデータの格納位置の変更を行うのではなく、論理的記憶装置を単位として物理的記憶装置への再配置を行い、且つ、再配置先の物理的記憶装置にデータを連続的に格納する。従って、シーケンシャルアクセスの場合でも、アクセス性能を向上することが出来る。また、ライトの度にデータの格納位置の変更を行うのではなく、予め定めた指標に基づいて前記再配置を行うから、ランダムアクセルでヒット率が低い場合でも、アクセス性能を向上することが出来る。

【００１０】

第２の観点では、本発明は、データ処理装置が直接アクセスを行う論理的記憶装置が配置された物理的記憶装置を含む複数の物理的記憶装置と、前記データ処理装置と前記複数の物理的記憶装置の間のデータ転送を制御する記憶制御装置とを有する記憶装置システムにおいて、前記データ転送の制御の運用中にデータ処理装置の論理的記憶装置へのアクセス情報を指標として採取するアクセス情報採取手段と、前記指標に基づいて前記論理的記憶装置を、前記複数の物理的記憶装置のうち、該論理的記憶装置が配置された物理的記憶装置とは異なる物理的記憶装置に再配置する際に、再配置先の物理的記憶装置に、前記論理的記憶装置に格納されたデータ全体を連続的に格納する論理的記憶装置再配置手段とを有することを特徴とする記憶装置システムを提供する。

上記第２の観点による記憶装置システムでは、アクセスするデータ単位でデータの格納位置の変更を行うのではなく、論理的記憶装置を単位として物理的記憶装置への再配置を行い、且つ、再配置先の物理的記憶装置にデータを連続的に格納する。従って、シーケンシャルアクセスの場合でも、アクセス性能を向上することが出来る。また、ライトの度にデータの格納位置の変更を行うのではなく、アクセス情報を採取し、それを統計的に利用して前記再配置を行うから、ランダムアクセルでヒット率が低い場合でも、アクセス性能を

向上することが出来る。

【００１１】

第３の観点では、本発明は、上記構成の記憶装置システムにおいて、前記アクセス情報が、前記データ処理装置から前記論理的記憶装置へのアクセス頻度情報を含むことを特徴とする記憶装置システムを提供する。

上記第３の観点による記憶装置システムでは、アクセス頻度の高い論理的記憶装置をより高速な物理的記憶装置へ再配置することが出来る。従って、アクセス性能を向上することが出来る。

【００１２】

第４の観点では、本発明は、上記構成の記憶装置システムにおいて、前記アクセス情報が、前記データ処理装置から前記論理的記憶装置へのアクセスパターン情報を含むことを特徴とする記憶装置システムを提供する。

上記第４の観点による記憶装置システムでは、シーケンシャルアクセスの比率の高い論理的記憶装置をよりシーケンシャルアクセス性能の高い物理的記憶装置へ再配置することが出来る。従って、アクセス性能を向上することが出来る。

【００１３】

第５の観点では、本発明は、上記構成の記憶装置システムにおいて、前記指標が、前記論理的記憶装置に求められる信頼性であることを特徴とする記憶装置システムを提供する。

上記第５の観点による記憶装置システムでは、信頼性が高いことが求められる論理的記憶装置をより信頼性の高い物理的記憶装置へ再配置することが出来る。従って、データの信頼性を向上することが出来る。

【００１４】

第６の観点では、本発明は、上記構成の記憶装置システムにおいて、前記指標を保守員に提示する指標提示手段と、保守員からの再配置指示を受け付ける再配置指示受付手段とを具備したことを特徴とする記憶装置システムを提供する。

上記第６の観点による記憶装置システムでは、保守員が再配置指示を入力できるため、非常に柔軟に前記再配置を行うことが出来る。

【００１５】

第７の観点では、本発明は、上記構成の記憶装置システムにおいて、データ処理装置からの再配置指示を受け付ける再配置指示受付手段を具備したことを特徴とする記憶装置システムを提供する。

上記第７の観点による記憶装置システムでは、データ処理装置が再配置指示を入力できるため、保守員では判断不可能な高度の条件下で前記再配置を行うことが出来る。

【００１６】

第８の観点では、本発明は、複数の物理的記憶装置と、データ処理装置と前記複数の物理的記憶装置の間のデータ転送を制御する記憶制御装置とを有する記憶装置システムにおいて、前記複数の物理的記憶装置のうちの第一の物理的記憶装置群にまたがって設定され、前記データ処理装置が直接アクセスを行う論理的記憶装置を有し、前記記憶制御装置は、予め定めた指標に基づいて、前記第一の物理的記憶装置群とは異なる第二の物理的記憶装置群を選択すると共に、前記第二の物理的記憶装置群に、前記論理的記憶装置に格納されたデータ全体を連続的に格納する論理的記憶装置再配置手段を有することを特徴とする記憶装置システムを提供する。

上記第８の観点による記憶装置システムでは、再配置により、アクセス性能を向上することが出来る。

【００１７】

第９の観点では、本発明は、上記構成の記憶装置システムにおいて、再配置中の論理的記憶装置にデータ処理装置からのアクセスがあったとき、再配置中の論理的記憶装置の再配置完了領域と再配置未完領域とを識別し、前記アクセス位置が前記再配置完了領域ならば再配置先の物理的記憶装置にアクセスさせ、前記アクセス位置が前記再配置未完領域ならば再配置前の物理的記憶装置にアクセスさせるアクセス位置切替手段をさらに具備したこ

とを特徴とする記憶装置システムを提供する。
上記第９の観点による記憶装置システムでは、再配置中の論理的記憶装置の再配置完了領域と再配置未完領域とを識別し、データ処理装置からのアクセス位置を切り替えるから、データ処理装置と物理的記憶装置の間のデータ転送を運用中に再配置を行うことが出来る。
第１０の観点では、本発明は、データ処理装置が直接アクセスを行う論理的記憶装置が配置された物理的記憶装置を含む複数の物理的記憶装置と、前記データ処理装置と前記複数の物理的記憶装置の間のデータ転送を制御する記憶制御装置とを有する記憶装置システムにおいて、前記記憶制御装置は、前記データ処理装置によるアクセス状況を取得し、前記アクセス状況に基づいて前記論理的記憶装置のデータを、前記複数の物理的記憶装置のうち、該論理的記憶装置に配置されている物理的記憶装置から、該論理的記憶装置が配置された物理的記憶装置とは異なる物理的記憶装置に移動させることを特徴とする記憶装置システムを提供する。
上記第１０の観点による記憶装置システムでは、データ処理装置によるアクセス状況に応じて論理的記憶装置のデータを、前記複数の物理的記憶装置のうち、該論理的記憶装置に配置されている物理的記憶装置から、該論理的記憶装置が配置された物理的記憶装置とは異なる物理的記憶装置へと移動させるから、アクセス性能を向上することが出来る。
第１１の観点では、本発明は、キャッシュメモリを有し、前記記憶制御装置は、前記第一の物理的記憶装置群から前記キャッシュメモリへデータを読み出す手段と、前記読み出されたデータを前記第二の物理的記憶装置群に書き込む手段とを有することを特徴とする上記第８の観点の記憶装置システムを提供する。
第１２の観点では、本発明は、前記第二の物理的記憶装置群は、前記第一の物理的記憶装置群よりもデータの読み出し速度が高速であることを特徴とする上記第８又は第１１の観点の記憶装置システムを提供する。
【００１８】
【発明の実施の形態】
以下、本発明の実施形態を説明する。なお、これにより本発明が限定されるものではない。
【００１９】
－第１の実施形態－
第１の実施形態は、各論理ディスク装置のアクセス情報を記憶制御装置で採取し、ＳＶＰ（サービスプロセッサ）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、論理ディスク装置の物理ディスク装置への再配置を行うものである。
【００２０】
図１は、本発明の第１の実施形態にかかる記憶制御装置を含む情報処理システムのブロック図である。
この情報処理システム１は、データ処理装置１００と、記憶制御装置１０４と、１台以上の物理ディスク装置１０５と、ＳＶＰ１１１とを接続してなっている。
【００２１】
前記データ処理装置１００は、ＣＰＵ１０１と、主記憶１０２と、チャネル１０３とを有している。
【００２２】
前記記憶制御装置１０４は、１つ以上のディレクタ１０６と、キャッシュメモリ１０７と、ディレクトリ１０８と、不揮発性メモリ１０９と、不揮発性メモリ管理情報１１０と、論理物理対応情報３００と、論理ディスク装置情報４００と、アクセス情報５００を有している。
前記ディレクタ１０６は、データ処理装置１００のチャネル１０３と物理ディスク装置１０５の間のデータ転送、データ処理装置１００のチャネル１０３と前記キャッシュメモリ１０７の間のデータ転送および前記キャッシュメモリ１０７と物理ディスク装置１０５の間のデータ転送を行う。

前記キャッシュメモリ１０７には、物理ディスク装置１０５の中のアクセス頻度の高いデータをロードしておく。このロード処理は、前記ディレクタ１０６が実行する。ロードするデータの具体例は、データ処理装置１００のＣＰＵ１０１のアクセス対象データや，このアクセス対象データと物理ディスク装置１０５上の格納位置が近いデータ等である。
前記ディレクトリ１０８は、前記キャッシュメモリ１０７の管理情報を格納する。
前記不揮発性メモリ１０９は、前記キャッシュメモリ１０７と同様に、物理ディスク装置１０５の中のアクセス頻度の高いデータをロードしておく。
前記不揮発性メモリ管理情報１１０は、前記不揮発性メモリ１０９の管理情報を格納する。
前記論理物理対応情報３００は、各論理ディスク装置（図２の２００）が配置されている物理ディスク装置１０５上の位置および各物理ディスク装置１０５に配置されている論理ディスク装置（図２の２００）を示す情報である。この情報を用いて、データ処理装置１００のＣＰＵ１０１のアクセス対象データの物理ディスク装置１０５上の格納領域の算出などを行う。
前記論理ディスク装置情報４００は、各論理ディスク装置（図２の２００）のアクセス可否等の状態を示す。
前記アクセス情報５００は、各論理ディスク装置（図２の２００）のアクセス頻度やアクセスパターンなどの情報である。

【００２３】

論理物理対応情報３００と論理ディスク情報４００は、電源断などによる消失を防ぐために不揮発の媒体に記録する。

【００２４】

前記物理ディスク装置１０５は、データを記録する媒体と、記録されたデータを読み書きする装置とから構成される。

【００２５】

前記ＳＶＰ１１１は、アクセス情報５００の保守員への提示や，保守員からの再配置指示６２０の入力の受け付けを行う。また、保守員からの情報処理システム１への指示の発信や，情報処理システム１の障害状態等の保守員への提示を行う。

【００２６】

図２は、論理ディスク装置２００と物理ディスク装置１０５の関連を表わした図である。
論理ディスク装置２００は、データ処理装置１００のＣＰＵ１０１が直接アクセスする見掛け上のディスク装置で、アクセス対象データが実際に格納される物理ディスク装置１０５と対応している。論理ディスク装置２００上のデータは、シーケンシャルアクセスを考慮して、物理ディスク装置１０５上に連続的に配置されている。論理ディスク装置２００のデータが配置されている物理ディスク装置１０５がディスクアレイ構成の場合、該論理ディスク装置２００は複数の物理ディスク装置１０５と対応する。また、物理ディスク装置１０５の容量が論理ディスク装置２００より大きく、複数の論理ディスク装置のデータを１台の物理ディスク装置１０５に格納できる場合には、該物理ディスク装置１０５は複数の論理ディスク装置２００と対応する。この論理ディスク装置２００と物理ディスク装置１０５の対応は前記論理物理対応情報３００で管理される。例えば、データ処理装置１００のＣＰＵ１０１が論理ディスク装置２００のデータ２０１をリードする時、記憶制御装置１０４で論理物理対応情報３００に基づき論理ディスク装置２００に対応する物理ディスク装置１０５を求め、その物理ディスク装置１０５の領域内のデータ格納位置２０２を求め、データ転送を行う。

【００２７】

図３は、論理物理対応情報３００を表わした図である。
論理物理対応情報３００は、論理ディスク構成情報３１０と、物理ディスク構成情報３２０とから構成される。前記論理ディスク構成情報３１０は、各論理ディスク装置２００が配置されている物理ディスク装置１０５上の領域に関する情報であり、論理ディスク装置２００から対応する物理ディスク装置１０５を求める時に用いる。一方、前記物理ディス

ク構成情報３２０は、各物理ディスク装置１０５に配置されている論理ディスク装置２００に関する情報で、物理ディスク装置１０５から対応する論理ディスク装置２００を求める時に用いる。

【００２８】

前記論理ディスク構成情報３１０は、物理ディスク装置グループ３１１，ＲＡＩＤ構成３１２および開始位置３１３の組を、論理ディスク装置２００の数だけ有している。

前記物理ディスク装置グループ３１１は、当該論理ディスク装置２００が配置されている物理ディスク装置１０５を示す情報である。

前記ＲＡＩＤ構成３１２は、前記物理ディスク装置グループ３１１のＲＡＩＤレベルを示す。

前記開始位置３１３は、当該論理ディスク装置２００が物理ディスク装置１０５上で配置されている先頭位置を示す。

【００２９】

前記物理ディスク構成情報３２０は、論理ディスク装置グループ３２１を、物理ディスク装置１０５の数だけ有している。

前記論理ディスク装置グループ３２１は、当該物理ディスク装置１０５に配置されている論理ディスク装置２００を示す。

【００３０】

図４は、論理ディスク情報４００を表わした図である。

論理ディスク情報４００は、論理ディスク状態４０１と再配置完了ポインタ４０２とを、論理ディスク装置２００の数だけ有している。

前記論理ディスク状態４０１は、「正常」「閉塞」「フォーマット中」「再配置中」などの論理ディスク装置２００の状態を表わす。

前記再配置完了ポインタ４０２は、前記論理ディスク状態４０１が「再配置中」の時のみ有効な情報で、当該論理ディスク装置２００の再配置処理を完了している領域の次の位置すなわち当該論理ディスク装置２００が未だ再配置処理を終えていない領域の先頭位置を示す。「再配置中」におけるデータアクセス時、再配置完了ポインタ４０２よりも前の領域へのアクセスの場合には、再配置後の物理ディスク装置１０５へアクセスしなければならない。一方、再配置完了ポインタ４０２以後の領域へのアクセスの場合には、再配置前の物理ディスク装置１０５へアクセスしなければならない。

【００３１】

図５は、アクセス情報５００を表わしている。

アクセス情報５００は、アクセス頻度情報５０１とアクセスパターン情報５０２とを、論理ディスク装置２００の数だけ有している。このアクセス情報５００は、記憶制御装置１０４，データ処理装置１００，ＳＶＰ１１１のいずれからも参照することが出来る。

前記アクセス頻度情報５０１は、単位時間あたりの当該論理ディスク装置２００へのアクセス回数を管理する。このアクセス頻度情報５０１は、各論理ディスク装置２００の中でアクセス頻度の高いもの又は低いものを求める指標として用いる。

前記アクセスパターン情報５０２は、当該論理ディスク装置２００へのシーケンシャルアクセスとランダムアクセスの割合を管理する。このアクセスパターン情報５０２は、シーケンシャルアクセスが多く、よりシーケンシャル性能の高い物理ディスク装置１０５に再配置するのが望ましい論理ディスク装置２００を求める指標として用いる。

【００３２】

次に、記憶制御装置１０４の動作を説明する。

図６は、記憶制御装置１０４の動作を詳細に表わした図である。

まず、リード／ライト処理時の動作について説明する。

ディレクタ１０６は、通常リード／ライト処理を実行する際、ＣＰＵ１０１からチャネル１０３を経由してＣＰＵからの指示６００を受け取る。このＣＰＵからの指示６００は、リード（またはライト）対象のレコードが記憶されている論理ディスク装置２００を指定する指定情報１と、リード（またはライト）対象のレコードが記憶されている論理ディス

ク装置２００内の位置（トラック，セクタ，レコード）を指定する指定情報２とを含んでいる。
ディレクタ１０６は、物理ディスク装置上のアクセス位置算出処理（６１０）で、前記ＣＰＵからの指示６００と論理物理対応情報３００とを用いて、物理ディスク装置１０５上でのアクセス位置を算出する。この物理ディスク装置アクセス位置算出処理（６１０）については図８を参照して後で詳述する。
その後、たとえばリード処理では、算出した物理ディスク装置１０５上のデータ格納位置２０２のデータをキャッシュメモリ１０７上に読み上げてデータ２０１とし、その読み上げたデータ２０１をチャネル１０３を通じて主記憶１０２に転送する。
【００３３】
次に、アクセス情報５００の採取処理について説明する。
ＣＰＵ１０１からのリード／ライト処理のアクセス時に、ディレクタ１０６は、アクセス対象論理ディスク装置２００のアクセス情報５００を更新する。
アクセス頻度情報５０１の採取は、例えば、アクセスの度に内部カウンタをカウントアップしていき、一定時間または一定回数のアクセス経過後のアクセス時に、前記内部ウンタからアクセス頻度を判定する。
アクセスパターン情報５０２の採取は、例えば、アクセスの度に内部カウンタにシーケンシャルアクセス回数をカウントアップしていき、一定時間または一定回数のアクセス経過後のアクセス時に、前記内部カウンタからアクセスパターンを判定する。
【００３４】
次に、再配置指示６２０を説明する。
保守員は、ＳＶＰ１１１を通じて提示されたアクセス情報５００を参照して、各論理ディスク装置２００の再配置の必要性を検討する。この検討の結果、再配置を決定した論理ディスク装置２００があれば、ＳＶＰ１１１を通じて記憶制御装置１０４に対して再配置指示６２０を出す。
この再配置指示６２０は、再配置対象の論理ディスク装置２００を２つ指定する指示情報１－２からなる。
保守員が行う検討の内容は、後述する第３の実施形態で図１０を参照して説明する論理ディスク装置再配置要否決定処理（９１０）と同様である。
【００３５】
次に、論理ディスク装置再配置処理（６３０）を説明する。
ディレクタ１０６は、前記再配置指示６２０を受けて、指定された２つの論理ディスク装置２００の間で論理ディスク装置再配置処理（６３０）を行う。
図７は、論理ディスク装置再配置処理部６３０の処理フロー図である。
ステップ７００では、論理ディスク情報４００のうちの指定された２つの論理ディスク装置２００の論理ディスク状態４０１を「再配置中」に設定する。
ステップ７０１では、論理ディスク情報４００のうちの指定された２つの論理ディスク装置２００の再配置完了ポインタ４０２を各論理ディスク装置２００の先頭位置に初期化する。
ステップ７０２では、論理ディスク情報４００のうちの指定された２つの論理ディスク装置２００の再配置完了ポインタ４０２をチェックし、全領域の再配置が完了していなければステップ７０３へ進み、完了していればステップ７０７へ進む。
【００３６】
ステップ７０３では、再配置完了ポインタ４０２が示す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に対して物理ディスク装置１０５からキャッシュメモリ１０７上へのデータ転送を行う。ここで、１回の処理単位分のデータ量は、再配置対象の２つの論理ディスク装置２００の冗長データ１つに対応する各データ量の最小公倍数に決定される。
たとえば、再配置をＲＡＩＤ５の論理ディスク装置２００とＲＡＩＤ１の論理ディスク装置２００の間で行うならば、ＲＡＩＤ１の論理ディスク装置２００の冗長データ１つに対応するデータ量は“１”であるから、１回の処理単位分のデータ量は、ＲＡＩＤ５の論理

ディスク装置２００の冗長データ１つに対応するデータ量すなわちパリティ１つに対応するデータ量に決定される。

【００３７】

ステップ７０４では、再配置対象の各論理ディスク装置２００の再配置先論理ディスク装置２００がパリティを有するＲＡＩＤレベルのものである場合、キャッシュメモリ１０７上の再配置対象の１回の処理単位分のデータ２０１に対してパリティを生成する。

ステップ７０５では、キャッシュメモリ１０７上の再配置対象の１回の処理単位分のデータ２０１および前記ステップ７０４で作成したパリティを、再配置先の物理ディスク装置１０５へ書き込む。

ステップ７０６では、１回の処理単位分だけ再配置完了ポインタ４０２を進める。そして、前記ステップ７０２に戻る。

【００３８】

なお、上記ステップ７０３，７０４において、データおよびパリティは、不揮発性メモリ１０９にも転送して二重化し、キャッシュ障害によるデータ消失を防ぐ。この理由は、上記ステップ７０５での書き込み時に、例えば、第１の論理ディスク装置２００と第２の論理ディスク装置２００のデータのうち、第１の論理ディスク装置２００のデータを物理ディスク装置１０５（元は第２の論理ディスク装置２００に配置されていた物理ディスク装置１０５）へ書き込んだ段階で障害によりキャッシュメモリ１０７上のデータがアクセス不能になったとすると、書き込みが終了してない第２の論理ディスク装置２００のデータが消失するからである（元は第２の論理ディスク装置２００に配置されていた物理ディスク装置１０５には、上記のように第１の論理ディスク装置２００のデータが上書きされてしまっている）。

【００３９】

ステップ７０７では、論理物理対応情報３００を更新する。すなわち、論理ディスク構成情報３１０と物理ディスク構成情報３２１を変更する。

ステップ７０８では、論理ディスク情報４００の論理ディスク状態４０１を元の状態に戻し、再配置処理（６３０）を終了する。

【００４０】

次に、物理ディスク装置アクセス位置算出処理（６１０）を説明する。

図８は、物理ディスク装置アクセス位置算出処理部６１０の処理フロー図である。

ステップ８００では、論理ディスク情報４００のうちのアクセス対象論理ディスク装置２００の論理ディスク状態４０１が「再配置中」であるか否かをチェックし、「再配置中」ならばステップ８０１に進み、「再配置中で」なければステップ８０３に進む。

【００４１】

ステップ８０１では、論理ディスク情報４００のうちのアクセス対象論理ディスク装置２００の再配置完了ポインタ４０２とアクセスデータ位置とを比較し、アクセスデータ位置が再配置完了ポインタ４０２の指す位置以後ならばステップ８０２に進み、アクセスデータ位置が再配置完了ポインタ４０２の指す位置より前ならばステップ８０３に進む。

【００４２】

ステップ８０２では、当該論理ディスク装置２００の再配置先の論理ディスク装置２００をアクセス対象にする。そして、ステップ８０４へ進む。

【００４３】

ステップ８０３では、当該論理ディスク装置２００をアクセス対象とする。

【００４４】

ステップ８０４では、アクセス対象の論理ディスク装置２００に対応した物理ディスク装置１０５上でのアクセス位置を、論理物理対応情報３００を用いて算出する。

【００４５】

以上の第１の実施形態にかかる情報処理システム１および記憶制御装置１０４によれば、アクセス情報５００に基づく保守員の判断により、アクセス頻度の高い論理ディスク装置をより高速な物理ディスク装置へ再配置することが出来る。また、シーケンシャルアクセ

スの比率の高い論理ディスク装置をよりシーケンシャルアクセス性能の高い物理ディスク装置へ再配置することが出来る。従って、アクセス性能を向上することが出来る。

【００４６】

－第２の実施形態－

上記第１の実施形態を変形して、記憶制御装置１０４からアクセス情報５００をデータ処理装置１００に提示し、データ処理装置１００が再配置要否を決定し記憶制御装置１０４に再配置指示（６２０相当）を出すようにしてもよい。

【００４７】

－第３の実施形態－

第３の実施形態は、再配置指示をＳＶＰ１１１やデータ処理装置１００から受けるのではなく、記憶制御装置１０４が自己決定するものである。

【００４８】

図９は、記憶制御装置１０４の動作を詳細に表わした図である。

第１の実施形態（図６）との違いは、論理ディスク再配置要否決定処理部９１０が再配置指示６２０を出すことである。

【００４９】

図１０は、上記論理ディスク再配置要否決定処理部９１０の処理フロー図である。

この論理ディスク再配置要否決定処理（９１０）は、ディレクタ１０６が一定周期で各論理ディスク装置２００のアクセス情報５００を検査して行う。

ステップ１０００では、アクセス情報５００のアクセス頻度情報５０１を参照し、アクセス頻度が規定値を超え且つ配置されている物理ディスク装置１０５が比較的低速なものである論理ディスク装置（以下、これを第１候補論理ディスク装置という）２００があるか否かをチェックし、該当する論理ディスク装置２００があればステップ１００１へ進み、なければステップ１００５へ進む。

【００５０】

ステップ１００１では、前記第１候補論理ディスク装置２００のアクセスパターン情報５０２を参照し、シーケンシャルアクセスの比率が規定値以上であるか否かをチェックし、規定値以上でなければステップ１００２へ進み、規定値以上であればステップ１００４へ進む。

【００５１】

ステップ１００２では、前記第１候補論理ディスク装置２００より高速な物理ディスク装置１０５に配置されている論理ディスク装置２００のアクセス頻度情報５０１を参照し、アクセス頻度が規定値以下の論理ディスク装置（以下、これを第２候補論理ディスク装置という）２００があるか否かをチェックし、あればステップ１００３へ進み、なければステップ１００５へ進む。

【００５２】

ステップ１００３では、前記第１候補論理ディスク装置２００と前記第２候補論理ディスク装置２００の間で再配置処理（６３０）が必要であると決定し、再配置指示６２０を出す。そして、処理を終了する。

【００５３】

ステップ１００４では、前記第１候補論理ディスク装置２００よりシーケンシャル性能の高い物理ディスク装置１０５に配置されている論理ディスク装置２００のアクセスパターン情報５０２を参照し、シーケンシャルアクセスの比率が規定値以下の論理ディスク装置（以下、これを第２候補論理ディスク装置という）２００があるか否かをチェックし、あれば前記ステップ１００３へ進み、なければ前記ステップ１００２へ進む。

【００５４】

ステップ１００５では、論理ディスク装置２００の再配置処理（６３０）は不要であると決定する。そして、処理を終了する。

【００５５】

以上の第３の実施形態にかかる情報処理システム１および記憶制御装置１０４によれば、

アクセス情報５００に基づいて自動的に、アクセス頻度の高い論理ディスク装置をより高速な物理ディスク装置へ再配置することが出来る。また、シーケンシャルアクセスの比率の高い論理ディスク装置をよりシーケンシャルアクセス性能の高い物理ディスク装置へ再配置することが出来る。従って、アクセス性能を向上することが出来る。

【００５６】

－第４の実施形態－

上記第１～第３の実施形態を変形して、アクセス情報５００に代えて又は加えて、論理ディスク装置２００に要求される信頼性を再配置処理要否決定の指標に用いてもよい。

信頼性を指標に用いれば、論理ディスク装置２００上のデータの信頼性を向上させることが出来る。

【００５７】

【発明の効果】

本発明の記憶装置システムによれば、シーケンシャルアクセスの場合やランダムアクセルでヒット率が低い場合でも、アクセス性能を向上することが出来る。また、本発明の記憶装置システムによれば、データの信頼性を向上することが出来る。

【図面の簡単な説明】

【図１】本発明の第１の実施形態にかかる記憶制御装置を含む情報処理システムのブロック図である。

【図２】論理ディスク装置と物理ディスク装置との対応関係の説明図である。

【図３】論理物理対応情報の構成例示図である。

【図４】論理ディスク情報の構成例示図である。

【図５】アクセス情報の構成例示図である。

【図６】本発明の第１の実施形態における記憶制御装置の動作を示すブロック図である。

【図７】論理ディスク装置再配置処理部の処理フロー図である。

【図８】物理ディスク装置アクセス位置算出処理部の処理フロー図である。

【図９】本発明の第３の実施形態における記憶制御装置の動作を示すブロック図である。

【図１０】論理ディスク装置再配置要否決定処理部の処理フロー図である。

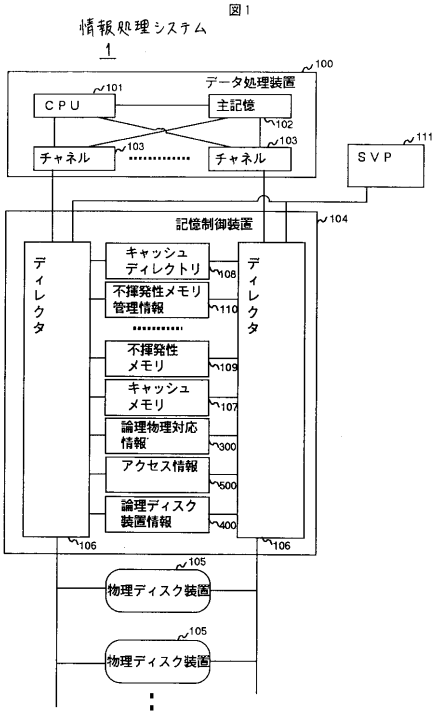
【符号の説明】

１　…情報処理システム
１００…データ処理装置
１０１…ＣＰＵ
１０２…主記憶
１０３…チャネル
１０４…記憶制御装置
１０５…物理ディスク装置
１０６…ディレクタ
１０７…キャッシュメモリ
１０８…キャッシュディレクトリ
１０９…不揮発性メモリ
１１０…不揮発性メモリ管理情報
１１１…ＳＶＰ
２００…論理ディスク装置
２０１…データ
２０２…データ格納位置
３００…論理物理対応情報
４００…論理ディスク情報
５００…アクセス情報
６００…ＣＰＵからの指示
６１０…物理ディスク装置上のアクセス位置算出処理部
６２０…指示情報

６３０…論理ディスク装置再配置処理部
９１０…論理ディスク再配置要否決定処理部

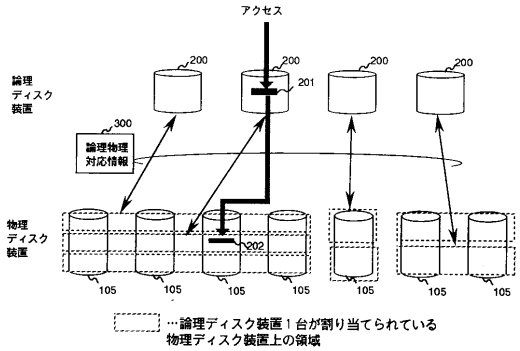
【図１】

図１



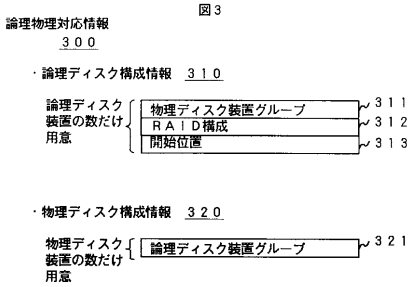
【図２】

図２



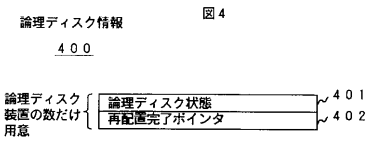
【図３】

図３



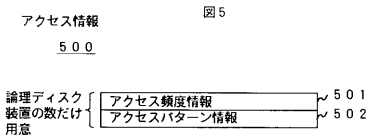
【図４】

図4



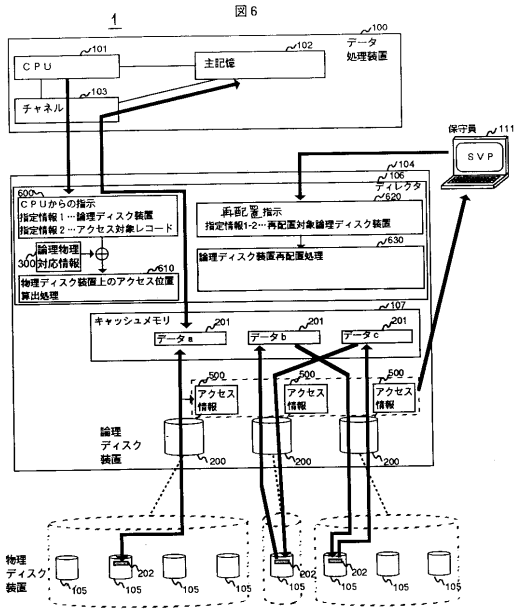
【図５】

図5



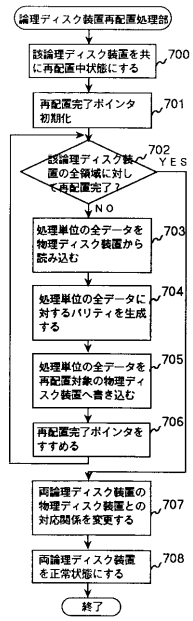
【図６】

図6



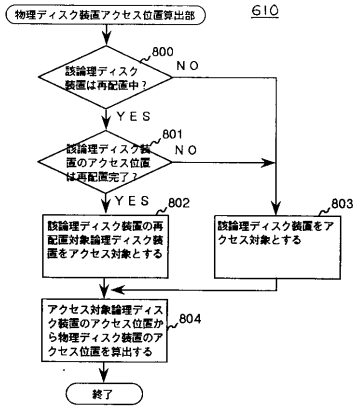
【図７】

図7



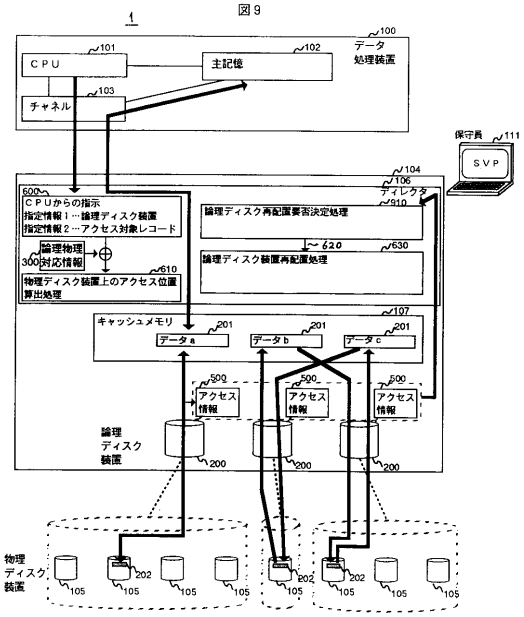
【図８】

図8



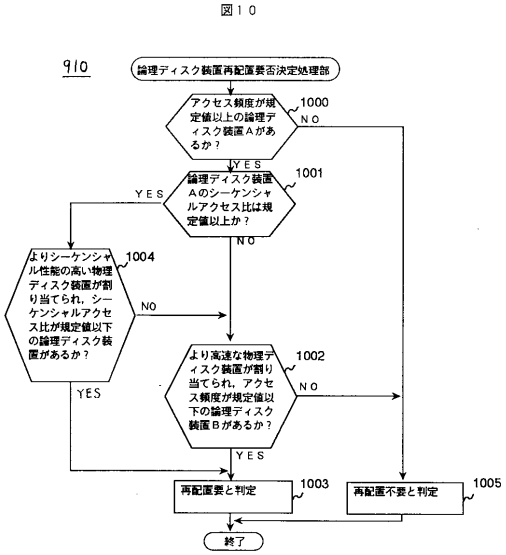
【図９】

図９


1
100
データ処理装置
101
CPU
102
主記憶
103
チャネル
保守員
111
SVP
104
106
ディレクタ
500
CPUからの指示
指定情報1…論理ディスク装置
指定情報2…アクセス対象レコード
論理物理対応情報
300
610
物理ディスク装置上のアクセス位置算出処理
910
論理ディスク再配置要否決定処理
620
630
論理ディスク装置再配置処理
107
キャッシュメモリ
201
データa
データb
データc
500
アクセス情報
論理ディスク装置
200
物理ディスク装置
202
105

【図１０】

図１０


910
論理ディスク装置再配置要否決定処理部
アクセス頻度が規定値以上の論理ディスク装置Aがあるか？
1000
NO
YES
論理ディスク装置Aのシーケンシャルアクセス比は規定値以上か？
1001
YES
NO
よりシーケンシャル性能の高い物理ディスク装置が割り当てられ，シーケンシャルアクセス比が規定値以下の論理ディスク装置があるか？
1004
NO
YES
より高速な物理ディスク装置が割り当てられ，アクセス頻度が規定値以下の論理ディスク装置Bがあるか？
1002
NO
YES
再配置要と判定
1003
再配置不要と判定
1005
終了

フロントページの続き

(72)発明者　佐藤　孝夫
神奈川県小田原市国府津２８８０番地　　株式会社日立製作所　ストレージシステム事業部内

合議体
審判長　大日方　和幸
審判官　大野　克人
審判官　矢島　伸一

(56)参考文献　特開平５－１００８０１（ＪＰ，Ａ）
特開平６－１３９０２７（ＪＰ，Ａ）
特開平７－１４６７５７（ＪＰ，Ａ）
特開平３－１０２４１８（ＪＰ，Ａ）
特開平８－６３２９８（ＪＰ，Ａ）
特開昭６２－６７６２９（ＪＰ，Ａ）
特開平７－１４１１２１（ＪＰ，Ａ）
月刊アスキー　第１６巻９号通巻１８３号ｐ１６６（株式会社アスキー）