(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3715000号
(P3715000)
(45)発行日　平成17年11月9日(2005.11.9)　(24)登録日　平成17年9月2日(2005.9.2)

(51)Int.Cl.[7]	F I		
G06F 3/06	G06F	3/06	302J
G06F 12/08	G06F	3/06	540
G06F 12/16	G06F	12/08	541A
	G06F	12/16	320L

請求項の数 9　(全 24 頁)

(21)出願番号　特願平7-217105
(22)出願日　平成7年8月25日(1995.8.25)
(65)公開番号　特開平8-137632
(43)公開日　平成8年5月31日(1996.5.31)
審査請求日　平成14年2月6日(2002.2.6)
(31)優先権主張番号　329,349
(32)優先日　平成6年10月26日(1994.10.26)
(33)優先権主張国　米国(US)

(73)特許権者　398038580
ヒューレット・パッカード・カンパニー
HEWLETT-PACKARD COMPANY
アメリカ合衆国カリフォルニア州パロアルト　ハノーバー・ストリート　3000
(74)代理人　100081721
弁理士　岡田　次生
(72)発明者　テレサ・エイ・バークス
アメリカ合衆国83707アイダホ州メリディアン、サン・レイモン・ドライブ　3100

最終頁に続く

(54)【発明の名称】データ記憶装置のデータを選択する方法

(57)【特許請求の範囲】

【請求項１】

複数の仮想ブロックからなる母集団としての仮想記憶空間に物理的記憶空間がマッピングされている、複数の記憶ディスクを持つ階層データ記憶装置において、データを移行のために選択する方法であって、

前記データ記憶装置に記憶されたデータを所定の分布基準にしたがって特徴づける、選ばれた数の区間を持つ、境界づけられた度数分布表を作成するステップを有し、該区間のそれぞれは、該区間での分布基準を満たす仮想ブロックの数のカウントを持ち、

前記度数分布表をセグメント化し、前記母集団のサブセットである移行対象となる仮想ブロックのための選択基準を特定するステップと、

仮想記憶空間内の仮想ブロックを探索して、前記選択基準を満たす１つまたは複数の仮想ブロックを見つけるステップと、

前記見つけられた仮想ブロックを前記階層データ記憶装置における一つのRAIDレベルから他のRAIDレベルに移行するステップと、

を有する方法。

【請求項２】

データが仮想ブロックに記憶された時を示すタイムスタンプを仮想ブロックにつけるステップをさらに有し、前記所定の分布基準はデータアクセスの最近性であり、前記区間は時間の範囲であり、時間区間に関連するカウントはその時間区間の範囲のタイムスタンプを持つ仮想ブロックの数を表す、請求項１に記載の方法。

【請求項３】
仮想ブロックの中のデータがアクセスされた回数を示す頻度スタンプを仮想ブロックに付けるステップをさらに有し、前記所定の分布基準はデータアクセス頻度であり、前記区間は頻度値に関連し、区間に関連するカウントはそれぞれの区間の範囲の頻度値を持つ仮想ブロックの数を表す、請求項１に記載の方法。
【請求項４】
新しい仮想ブロックが使われるときに前記度数分布表を更新するステップをさらに有し、前記更新ステップは、（１）新しい仮想ブロックが既存の区間に対する分布基準を満たす場合、その既存の区間に関連するカウントを増やすステップ、（２）新しい仮想ブロックが既存の区間に対する分布基準を満たさない場合、新しい区間を作り該新しい区間に最初のカウントを入れるステップのいずれかを有する、請求項１に記載の方法。
【請求項５】
新しい区間を作る前記ステップは、
前記度数分布表の中の２つ以上の区間を接合して、該接合される複数の区間のカウントの和に等しいカウントを持つ１つの接合区間を規定するステップと、
前記分布基準を延ばす新しい区間を規定して前記新しい仮想ブロックをカバーするステップと、
を有する請求項４に記載の方法。
【請求項６】
既存の仮想ブロックがアクセスされるときに度数分布表を更新するステップをさらに有し、
該更新ステップは、アクセスされる前に既存の仮想ブロックが分布基準を満たした区間のカウントを減らし、アクセスされた後に既存の仮想ブロックが分布基準を満たした別の区間のカウントを増やす、請求項１に記載の方法。
【請求項７】
物理的記憶空間を規定する複数の記憶ディスクを持つ階層データ記憶装置の中で移行するデータを選択する方法であって、前記物理的記憶空間は、異なるＲＡＩＤレベルにしたがってデータを記憶する少なくとも２つのＲＡＩＤ区域として物理的記憶空間を表すＲＡＩＤレベル仮想記憶空間にマッピングされ、該ＲＡＩＤレベル仮想記憶空間は、複数の仮想ブロックからなる母集団として物理的記憶空間を表すアプリケーション・レベル仮想記憶空間にマッピングされ、
前記方法は、
個々の区間が関連する区間に対する分布基準を満たす仮想ブロックの数のカウントを持つ選択された数の区間を有する、第１のＲＡＩＤレベルの下で第１のＲＡＩＤ区域に記憶されたデータを所定の分布基準にしたがって特徴づける第１の度数分布表を作成するステップと、
個々の区間が関連する区間に対する分布基準を満たす仮想ブロックの数のカウントを持つ選択された数の区間を有する、第２のＲＡＩＤレベルの下で第２のＲＡＩＤ区域に記憶されたデータを所定の分布基準に従って特徴づける第２の度数分布表を作成するステップと、
個々のセグメントが前記第１の度数分布表の１つまたは複数の区間を持つ、第１の母集団の複数のセグメントを規定するステップと、
前記複数のセグメントから少なくとも１つのセグメントを選択するステップと、
前記アプリケーション・レベル仮想記憶空間の中の仮想ブロックを探索して、選択したセグメントに含まれる前記１つまたは複数の区間に対する分布基準を満たし、前記第１ＲＡＩＤレベルの下の前記第１ＲＡＩＤ区域に記憶されたデータを表す１つまたは複数の仮想ブロックを見つけるステップと、
見つかった仮想ブロックを前記第１ＲＡＩＤ区域から前記第２ＲＡＩＤ区域へと移行するステップと、
を有する方法。

【請求項８】
仮想ブロックに単調増加する値を持つスタンプをつけて、より新しくスタンプをつけられた仮想ブロックが前記第１および第２の度数分布表の最後の区間にカウントされるようにするステップをさらに有する、請求項７に記載の方法。
【請求項９】
物理的記憶空間を規定する複数の記憶ディスクを持つディスクアレーと、
該ディスクアレーに結合され、ディスクへまたはディスクからのデータ転送を調整するディスクアレー制御装置と、
該ディスクアレー制御装置に接合され、様々なＲＡＩＤレベルにしたがってデータを記憶する複数のＲＡＩＤ区域として物理的記憶空間を表すＲＡＩＤレベル仮想記憶空間と、複数の仮想ブロックからなる母集団として物理的記憶空間を表すアプリケーション・レベル仮想記憶空間とを、記憶ディスクの物理的記憶空間にマッピングするＲＡＩＤ管理装置と、
個々の区間が関連する区間に対する分布基準を満たす仮想ブロックの数のカウントを持つ選択された数の区間を有する、第１および第２のＲＡＩＤ区域に記憶されたデータを所定の分布基準にしたがって特徴付ける第１および第２の度数分布表と、を有し、
前記ＲＡＩＤ管理装置は記憶ディスク上の第１および第２のＲＡＩＤ区域の間で仮想ブロックを移行し、
前記ＲＡＩＤ管理装置は前記第１または第２の度数分布表のどちらかを区分することによって、移行する仮想ブロックを選択して、それぞれの母集団のサブセットを構成する仮想ブロックの選択基準を特定する、データ記憶装置。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は、階層データ記憶装置における移行のために、度数分布表を使ってデータを選択する方法、およびそのような度数分布表を具体化した階層データ記憶装置に関する。
【０００２】
【従来の技術】
コンピュータ・システムは、速度、信頼性、および処理能力の点で絶え間なく改良が続いている。その結果、コンピュータはより複雑で高度なアプリケーションが処理できる。コンピュータが改良されるのに応じて、大容量記憶装置および入出力装置に対する性能要求が増大している。発展するコンピュータ・システムに性能面で歩調がとれる大容量記憶装置を設計することが常に求められている。
【０００３】
ディスクアレー・データ記憶装置は、単一の大容量記憶装置を形成するように配列され調整された複数のディスク駆動装置を持つ。大容量記憶装置には、コスト、性能、可用性の３つの主要な設計基準がある。メガバイトあたりの低いコスト、高い入出力性能、および、高いデータ可用性を持つ記憶装置を作ることが望ましい。"可用性"（availability）とは、記憶装置に記憶されているデータをアクセスする能力、および、ある程度の故障が起きても動作が続けられることを保証できる能力のことである。典型的には、データの可用性は、複数の場所に記憶されたデータまたはデータ間の関係を持つ冗長度の使用を通して提供される。
【０００４】
冗長データを記憶するには２つの一般的な方法がある。第１の方法すなわち"ミラー"（鏡像：mirror）方法では、データが複製され、記憶装置の２つの別々の区域に記憶される。たとえば、ディスクアレーでは、同一のデータがディスクアレーの２つの別々のディスクの上に提供される。ミラー方法は、二重に記憶する手法により、高性能および高データ可用性の利点がある。しかし、ミラー方法は、データを記憶するコストが実質的に２倍になるので、相対的に費用の高いものになる。
【０００５】

第２の方法すなわち"パリティ"（奇偶：parity）方法では、記憶区域の一部を冗長データを記憶するのに使うが、冗長データを記憶する区域は、元のデータを記憶するのに使う記憶空間より小さい。たとえば、５台のディスクを持つディスクアレーでは、４台をデータを記憶するのに使い、５台目をもっぱら冗長データを記憶するのに使う。パリティ方法の利点は、ミラー方法より低費用であることであるが、ミラー方法にくらべると、性能および可用性が低いという特徴がある。

【０００６】

【発明が解決しようとする課題】

本発明は、従来の２つの方法の調和をはかり、性能および信頼性を最適化するデータ記憶装置を提供することを目的とする。

【０００７】

【課題を解決するための手段】

本発明のデータ記憶装置は、複数の記憶ディスクを持つディスクアレー、並びに、ディスクにおよびディスクからデータ転送を調整するディスクアレー制御装置を含む。記憶ディスクは、物理的記憶空間を規定する。データ記憶装置は、また、２つの仮想記憶空間を記憶ディスクの物理的記憶空間にマッピングするために、ディスクアレー制御装置を動作させるように結合されたＲＡＩＤ（Redundant Array of Indepedent Disks）管理装置を含む。ＲＡＩＤレベル仮想記憶空間は、物理的記憶空間を、ＲＡＩＤレベル１（ミラー冗長性）およびＲＡＩＤレベル５（パリティ冗長性）にしたがってデータを記憶するミラーＲＡＩＤ区域およびパリティＲＡＩＤ区域として表す。アプリケーション・レベル仮想記憶空間は、物理的記憶空間を、複数の仮想ブロックとして表す。すべての仮想ブロックの組が母集団（population）を規定する。ＲＡＩＤ管理装置は、仮想ブロックをミラーＲＡＩＤ区域とパリティＲＡＩＤ区域との間で移行させ、その結果、データはＲＡＩＤレベル１からＲＡＩＤレベル５への冗長性の変更を受ける。

【０００８】

データ記憶装置は、２つの移行方針、すなわち、データアクセスの最近性またはデータアクセスの頻度のどちらかに基づいて移行を実行する。データアクセスの最近性の方針の下では、データは、データがユーザによってどれくらい最近にアクセスされたかに基づいて移行される。データアクセスの頻度の方針の下では、データは、データがユーザによってどれくらい頻繁にアクセスされたかに基づいて移行される。より古いデータ、または、あまり頻繁にアクセスされなかったデータは、典型的には、性能の高いＲＡＩＤレベル１から性能のより低いＲＡＩＤレベル５への移行の候補になる。

【０００９】

本発明によるデータ記憶装置は、移行の対象となる仮想ブロックを識別するために度数分布表を使う。ＲＡＩＤレベルのそれぞれに対して１つの度数分布表を用意することが望ましい。度数分布表は、ＲＡＩＤ区域に記憶されたデータを、予定の分布基準にしたがって特徴づける。度数分布表は、一定の数の区間によって区切られている。各区間に関連づけられた数（カウント）が、それぞれの区間の分布基準を満たす仮想ブロックの数を表す。各度数分布表の中の仮想ブロックの全数が母集団の数である。

【００１０】

ＲＡＩＤ管理装置は、仮想ブロックの選択基準を決めるために、移行源ＲＡＩＤ区域の度数分布表を区分することによって、移行源ＲＡＩＤ区域から移行先ＲＡＩＤ区域に移行する仮想ブロックを選択する。アプリケーション・レベル仮想記憶空間は、選択基準を満たす仮想ブロックが見つかるまで探索される。見つかった仮想ブロックは、移行源ＲＡＩＤ区域から移行先ＲＡＩＤ区域に移行される。

【００１１】

度数分布表を使用することは、データ記憶装置が、最も古いまたは最もアクセス頻度が少ない仮想ブロックを見つけるために、すべての仮想ブロックをソートする必要がないので、利点がある。むしろ、データ記憶装置は、移行する候補を選択するための選択基準を特定し、十分な数の移行候補を持つ母集団のサブセットを選ぶためにその選択基準を使うだ

けである。この処理は、記憶要求を満たす十分な数の仮想ブロックが見つかり次第、早めに終了することができる。この手法により、かなりの効率が得られる。

【００１２】

【発明の実施の形態】

図１に、本発明にしたがって構成されたデータ記憶装置１０を示す。データ記憶装置１０は、複数の記憶ディスク１２を持つ階層ディスクアレー１１、記憶ディスク１２への、および記憶ディスク１２からのデータ転送を調整するためにディスクアレー１１に結合されたディスクアレー制御装置１４、およびＲＡＩＤ管理装置１６を含む、ディスクアレー・データ記憶装置であることが望ましい。

【００１３】

本明細書の説明の目的のために、"ディスク"とは、不揮発性で、ランダム・アクセスが可能で、自身の記憶故障を検出する能力のある、書き換え可能な大容量記憶装置を指す。それは、回転する磁気および光ディスク、並びに、固体ディスク、または不揮発性の電子的記憶要素、たとえばＰＲＯＭ、ＥＰＲＯＭ、およびＥＥＰＲＯＭを含む。"ディスクアレー"という用語は、複数のディスク、ディスクを１つまたは複数のホスト・コンピュータに接続するのに必要なハードウェア、および物理的ディスクの動作を制御し、物理的ディスクをホストの動作環境に１つまたは複数の仮想ディスクとして提示するのに使われる管理ソフトウェアの集まりを指す。"仮想ディスク"とは、管理ソフトウェアによってディスクアレーの中に実現される抽象的実体である。

【００１４】

"ＲＡＩＤ"（Random Array of Independent Disks）と言う用語は、記憶容量の残りの部分に記憶されたユーザデータについての冗長情報を記憶するために、その物理的記憶容量の１部が使用されているディスクアレーを意味する。冗長情報は、ディスクアレーを構成するディスクの１つ、または、そのディスクへのアクセス経路が故障したときにユーザデータの再生を可能にする。ＲＡＩＤ装置に関する詳しい記述は、「ＲＡＩＤ手法の資料集」（The RAIDBook: A Source Book for RAID Technology, published June 6, 1993, by the RAID Advisory Board, Lino Lakes, Minnesota）を参照されたい。

【００１５】

ディスクアレー制御装置１４は、たとえばＳＣＳＩ（small computer system interface）のような１つまたは複数のインターフェース・バス１３を介してディスクアレー１１に結合されている。ＲＡＩＤ管理装置１６は、インターフェース・プロトコル１５を介して、ディスクアレー制御装置１４を動作させるために結合されている。データ記憶装置１０はホスト・コンピュータ（図示せず）にも入出力インターフェース・バス１７を介して結合されている。ＲＡＩＤ管理装置１６は、ディスクの記憶および信頼性レベルを制御し、多様な信頼性記憶レベル間のデータ転送を行うためのデータマネージャ手段を提供するために、別個の装置要素として具体化してもよいし、ディスクアレー制御装置１４内またはホスト・コンピュータ内に構成してもよい。これらの信頼性記憶レベルは、以下に述べるように、ミラーまたはパリティ冗長レベルであることが望ましいが、全く冗長性を持たない信頼性記憶レベルを含むこともできる。

【００１６】

ディスクアレー制御装置１４は、ディスクアレー制御装置Ａ　１４ａ、およびディスクアレー制御装置Ｂ　１４ｂからなる二重の制御装置として具体化することが望ましい。二重制御装置１４ａおよび１４ｂは、どちらか１つが動作できなくなった場合にも、継続したバックアップおよび冗長性を提供することによって信頼性を向上させる。しかし、本発明は、１台の制御装置またはその他の構成でも実施することができる。

【００１７】

階層ディスクアレー１１は、物理的記憶空間および１つまたは複数の仮想記憶空間を含む異なる記憶空間として特徴づけられる。記憶に関するこれらの多様なビューはマッピング手法に関連している。たとえば、ディスクアレーの物理的記憶空間は、多様なデータ信頼性レベルにしたがって記憶区域を描く仮想記憶空間にマッピングすることができる。仮想

記憶空間内のある区域を第１信頼性記憶レベル（たとえば、ミラーすなわちＲＡＩＤレベル１）のために割り当て、その他の区域を第２信頼性記憶レベル（たとえば、パリティすなわちＲＡＩＤレベル５）のために割り当てることができる。ＲＡＩＤレベルに関しての多様なマッピング手法および仮想空間については以下に詳しく述べる。

【００１８】

データ記憶装置１０は、異なる記憶空間を互いにマッピングするために用いる仮想マッピング情報を持続的に記憶するメモリマップ記憶装置２１を含む。メモリマップ記憶装置はディスクアレーの外にあり、ディスクアレー制御装置１４内にあることが望ましい。メモリマップ情報は、異なるビューの中で多様なマッピング構成が変わるにしたがい、制御装置またはＲＡＩＤ管理装置によって絶えずまたは定期的に更新される。

【００１９】

メモリマップ記憶装置２１は、それぞれ、制御装置１４ａおよび１４ｂに置かれる２つの不揮発性ＲＡＭ（Random Access Memory）２１ａおよび２１ｂとして具体化されるのが望ましい。不揮発性ＲＡＭ（non-volatile RAM: NVRAM）の１つの例は、バッテリ付きのＲＡＭである。バッテリ付きＲＡＭは、データをメモリに維持するために、データ記憶装置１０への電源が切られたときに独立したバッテリ電源からのエネルギーを使う。１つの望ましい構成は、自身でリフレッシュできるバッテリ付きＤＲＡＭ（Dynamic RAM）である。

【００２０】

二重のＮＶＲＡＭ２１ａおよび２１ｂは、メモリマップ情報の冗長記憶を供給する。仮想マッピング情報はミラー冗長方法にしたがって複製され、ＮＶＲＡＭ２１ａおよび２１ｂの両方に記憶される。このようにして、ＮＶＲＡＭ２１ａはもっぱら元のマッピング情報を記憶するのに用い、ＮＶＲＡＭ２１ｂはもっぱら冗長マッピング情報を記憶するのに用いることができる。別の構成として、データを二重に記憶するための十分な空間を備えた１つの不揮発性ＲＡＭを用いて、ミラーにしたメモリマップ記憶装置を構成することもできる。

【００２１】

図１に示すように、ディスクアレー１１は複数のディスク駆動装置１２を持つ。これらの記憶ディスクの大きさの例は、１ないし３ギガバイトである。記憶ディスクは、ＳＣＳＩバス１３とのインターフェースを提供する機械的ベイに、それぞれ独立して接続したりはずしたりできる。１つの具体化では、データ記憶装置は、作動する機械的ベイを１２個、および追加の４個の機械的ベイを持つように設計される。これらのベイをディスクアレー制御装置１４にインターフェースさせるために４つのＳＣＳＩバス（すなわち、３ないし４個の機械的ベイに対して１つのバス）が使われる。これらのベイがいっぱいになるまでロードされた場合、データ記憶装置は全体で、たとえば、１２ないし３６ギガバイトの容量を持つ。ディスクアレー制御装置１４は、記憶ディスク１２がどのベイに差し込まれていても、認識することができる。データ記憶装置１０は、ディスクアレーの動作中に、ディスクアレーの使用可能なベイに追加のディスクを差し込む（"ホットプラグ"：hot plug）ことができるように設計されている。

【００２２】

本発明の説明の目的として、ディスクアレー１１の記憶ディスク１２は、複数のディスク２０を持つミラー・グループ１８、および、複数のディスク２４を持つパリティ・グループ２２に配列されたものとして概念化することができる。ミラー・グループ１８は、第１のレベルすなわちミラー冗長レベルにしたがってデータを記憶するディスクアレーの第１の記憶位置すなわちＲＡＩＤ区域を表す。このミラー冗長レベルは、またＲＡＩＤレベル１と考えることができる。ＲＡＩＤレベル１、すなわちディスクのミラー化は、データの各ビットが複製されてデータ記憶装置内に記憶されるという意味で１対１の保護を提供することにより、最高の信頼性を提供する。ミラー冗長は、図１で、３対のディスク２０で表されている。元のデータは第１の組のディスク２６に記憶され、複製された冗長データは、対になった第２の組のディスク２８に記憶される。

【００２３】
図２に、ＲＡＩＤレベル１にしたがったデータの記憶をより詳しく示す。垂直の列は個々のディスク０、１、２、３を表す。４個のディスクからなるこのディスクアレーに含まれる物理的記憶空間は、水平の行で表される複数のストライプに構成される。１つの"ストライプ"は複数の記憶ディスクにわたって延び、多数の同じ大きさの記憶空間のセグメントで構成され、１つのセグメントはアレーの中の各ディスクに関連している。すなわち、１つのセグメントは１つのディスクの上に存在する１つのストライプの部分である。各ストライプは、記憶ディスク全体に分配されているデータの予定の量を保持する。ストライプのうちのあるセグメントは元のデータのために使われ、その他のセグメントは冗長データのために使われる。
【００２４】
ミラー冗長（ＲＡＩＤレベル１）のこの例では、ディスク０上のストライプ０のセグメント０に記憶されているデータは、複製されて、ディスク１上のストライプ０のセグメント０'に記憶される。同様に、ディスク２のストライプ２のセグメント５に記憶されているデータは、ディスク３のストライプ２のセグメント５'に記憶される。このようにして、各データは複製されディスク上に記憶される。図２の冗長のレイアウトは説明の目的で図示したものである。冗長データを同じストライプに、図示したように、きちんと並べる必要はない。たとえば、ディスク０のストライプ１のセグメント２のデータは複製されて、ディスク３のストライプＳのセグメントＴ'に置くことができる。
【００２５】
図１を再び参照すると、ディスク２４のパリティ・グループ２２は、第２の記憶位置すなわちＲＡＩＤ区域を表し、データは第２の冗長レベル、たとえばＲＡＩＤレベル５にしたがって記憶される。図に例として示した６個のディスクでは、元のデータは５個のディスク３０に記憶され、冗長"パリティ"データは６番目のディスク３２に記憶される。
【００２６】
図３にパリティＲＡＩＤ区域のレイアウトを詳しく示す。図２のミラーＲＡＩＤ区域のレイアウトと同様に、ディスク０、１、２、３の物理的記憶空間は、複数の同じ大きさのストライプに構成される。例として示したこの図では、データはＲＡＩＤレベル５にしたがって記憶され、冗長データはＰで表したセグメントに記憶される。冗長Ｐセグメントは、ストライプの中のその他のセグメントのパリティを記憶する。たとえば、ストライプ０において、ディスク３上の冗長Ｐセグメントは、ディスク０、１、２のパリティを記憶する。各ストライプのパリティは、たとえば、排他的ＯＲ関数のような、ある関数によって計算される。最初の４つのストライプのパリティは次のように表され、下付き数字は対応するストライプを表す。
【００２７】
【数１】

Ｐ０　＝セグメント０ ⊕ セグメント１ ⊕ セグメント２

　　＝ディスク０ ⊕ ディスク１ ⊕ ディスク２

Ｐ１　＝セグメント３ ⊕ セグメント４ ⊕ セグメント５

　　＝ディスク０ ⊕ ディスク１ ⊕ ディスク３

Ｐ２　＝セグメント６ ⊕ セグメント７ ⊕ セグメント８

　　＝ディスク０ ⊕ ディスク２ ⊕ ディスク３

Ｐ３　＝セグメント９ ⊕ セグメント１０ ⊕ セグメント１１

　　＝ディスク１ ⊕ ディスク２ ⊕ ディスク３

（記号 ⊕ は排他的ＯＲ関数を表す）

パリティ冗長は、ディスクの１つで使用不可能になったデータの再生を可能にする。たとえば、セグメント５のデータが使用できなくなったとき、その内容は、セグメント３および４、および、セグメントＰのパリティ・データから確かめることができる。パリティ記憶はミラー記憶よりも費用はかからないが、信頼性はより低く性能も低い。

【００２８】

図１に示したディスク配列は概念を示す目的のものである。実際には、ディスクアレー１１は、ミラーおよびパリティ冗長にしたがってデータを記憶することができる複数のディスク１２を単に持つだけである。全てのディスク１２によって提供される使用可能な記憶空間のうち、その記憶空間の１部がミラー冗長に割り当てられ、別の１部がパリティ冗長に割り当てられる。ディスク１２は複数の同じ大きさの記憶領域（図４の番号３５で表すような）を含むように構成され、個々の領域が複数のセグメントを持つようにするのが望ましい。領域は、記憶空間の１つの仮想ビューの中にＲＡＩＤ区域を形成するようにまとめられる。さらに、ＲＡＩＤ区域およびデータ冗長記憶手法がユーザまたはホストに透明（トランスペアレント：transparent）になるように、記憶空間の別の（ホストによって定義された）ビューがユーザまたはホストに提示される。これらの特徴については、図４を参照して以下に詳しく述べる。

【００２９】

データ記憶装置１０は、ミラー記憶方法およびパリティ記憶方法の間のデータの"移行"（migration）を管理する。両タイプの冗長性の管理はＲＡＩＤ管理装置１６（図１）によって調整される。ＲＡＩＤ管理装置１６は、ディスクアレーの中の２つの異なった型のＲＡＩＤ区域を、ミラーＲＡＩＤ区域がパリティＲＡＩＤ区域に対してキャッシュとしてはたらくのと同様なメモリ階層として管理する。ＲＡＩＤ管理装置１６は、定義された性能プロトコルにしたがって、ミラーＲＡＩＤ区域とパリティＲＡＩＤ区域との間のデータをシフトし、組織し、その他の方法で管理する。ミラーＲＡＩＤ区域とパリティＲＡＩＤ区域との間のデータ移動の処理は"移行"と呼ばれる。

【００３０】

データ記憶装置１０は、最高の性能および信頼性がもたらされるように、性能により厳しいデータをミラーＲＡＩＤ区域に置くように努める。ＲＡＩＤ管理装置１６によって具体化される性能プロトコルは、２つの望ましい移行方針のうちのどちらかを含む。第１の移行方針は"アクセス頻度"と呼ばれ、この方針にしたがって、階層ディスクアレー上の最も

頻繁にアクセスされるデータがミラーＲＡＩＤ区域１８に維持される。アクセス頻度の低いデータはパリティＲＡＩＤ区域２２に維持される。"アクセスの最近性"と呼ばれる第２の移行方針にしたがって、最も最近にアクセスされたデータはミラーＲＡＩＤ区域１８に維持され、アクセスの最近性の低いデータはパリティＲＡＩＤ区域２２に記憶される。この他の性能プロトコルを使用することもできる。そのようなプロトコルは、具体的なコンピュータ・アプリケーション、およびユーザのニーズに基づいて定義することが理想的である。

【００３１】

さらに、ＲＡＩＤ管理装置１６は、２つのパラメータの関数、すなわち物理的記憶容量の大きさ、およびデータ記憶装置に記憶されているユーザデータの現在量にしたがって、データ記憶装置の記憶資源を自動的に"調整"する。最初は全てのデータがミラーＲＡＩＤ区域に記憶される。これによって最高の性能および信頼性が得られるからである。より多くのデータがデータ記憶装置に加えられるにつれて、性能および信頼性を最適化するために、ミラーＲＡＩＤ区域とパリティＲＡＩＤ区域との間でデータの移行が行われる。データ記憶装置が満杯に近づくにしたがい、冗長性を通して信頼性を提供しながらユーザの要求をすべて満たすようにするために、いっそう多くのデータがパリティＲＡＩＤ区域に移行される。したがって、本発明によるデータ記憶装置は最大の融通性と適応をもたらす。この装置は、特定の記憶方法をユーザが選択する必要はないが、ユーザからのどんな要求をも満たすように適応できる。

【００３２】

図４に、データ記憶装置１０の使用可能な記憶空間のメモリ・マッピングを、マッピングされた仮想記憶空間の複数の層として示す。図の垂直に延びた矩形のそれぞれは、物理的記憶空間のビューを表す。この図において、物理的記憶空間３４は２つの仮想記憶ビュー４０および５０で参照されている。物理的記憶空間３４は４個のディスク（図１のディスク１２のような）で表され、数字０、１、２、３で参照される。これらのディスクに関連している４つの矩形は物理的記憶空間のビューを表し、ディスク１、２、３は大体同じ記憶容量を持っているが、ディスク０はやや少ない記憶容量を持っている。これらのディスクの記憶容量の例は、１ないし３ギガバイトである。記憶空間３４は区域Ａ０、Ａ１、Ａ２等に分割されている。個々の区域は、たとえば区域Ａ０の中のストライプ０～Ｑのように、多くのストライプを持つ。個々の区域には、また、多くの領域３５がある。領域３５は、それぞれの領域が全ディスクアレーにわたって同じ大きさになるように、各記憶ディスク上で、均一な大きさの、ある選択された数のセグメントで構成されるのが望ましい。１つの領域３５の大きさの例は１メガバイトである。

【００３３】

ディスクの記憶空間は、物理的記憶空間３４の第１の中間的な、ＲＡＩＤレベル仮想ビュー４０にマッピングされる。この第１仮想ビューは、概念的には、連続的にアドレス可能な記憶空間を表す第２のアプリケーション・ビューにマッピングすることができる１組のＲＡＩＤ区域である。記憶空間の物理的な構成およびＲＡＩＤビューはアプリケーション・ビューからは隠れている。

【００３４】

ＲＡＩＤ区域の記憶空間４０は、ミラーおよびパリティ記憶空間を区別する記憶のビューである。たとえば、ＲＡＩＤ区域４２はＭ個の割り当てブロック４３を持つミラーＲＡＩＤ区域を表し、ＲＡＩＤ区域４４はＮ個の割り当て区域４５を持つパリティＲＡＩＤ区域を表す。割り当てブロック４３および４５は同じ大きさ（たとえば６４ｋバイト）であることが望ましい。これらのＲＡＩＤ区域は、物理的記憶空間３４上の対応する物理的区域Ａ０、Ａ１、Ａ２等に関連する。たとえば、ＲＡＩＤ仮想ビューの１６個の６４ｋ割り当てブロック４３または４５は、１個の１Ｍ（メガ）領域３５の上にマッピングすることができる。

【００３５】

ミラーおよびパリティＲＡＩＤ区域は、ディスクアレーの記憶空間３４の全体を使いきる

場合とそうでない場合とがある。したがって、あるアプリケーションの間、特定のＲＡＩＤ区域に対応しない、未使用で未指定の記憶空間がある場合がある。しかし、そのような記憶空間は、ミラーまたはパリティＲＡＩＤ区域のどちらにも変換することができる。また、ＲＡＩＤ区域は、ディスクアレー上の連続した区域にマッピングされ、ＲＡＩＤ区域に関連する領域のそれぞれは各記憶ディスク上の同じ物理アドレスに置かれるものとして図示してあることに留意されたい。代替方法として、ＲＡＩＤ区域はディスクアレー上の不連続な区域にマッピングすることもできる。

【００３６】

ＲＡＩＤ区域で使用可能な記憶空間は、ユーザまたはホスト・アプリケーション・プログラムによって定義され、ユーザまたはホスト・アプリケーション・プログラムに提示される記憶のビューである第２の前面のアプリケーション・レベルの仮想ビュー５０にマッピングされる。ユーザまたはホスト・アプリケーション・プログラムから見た場合、アプリケーション・レベルの仮想ビュー５０は、記憶ディスク１２上で使用可能な記憶空間を示す１つの大きな記憶容量を表す。仮想記憶空間５０は、それぞれが、０、１、２、.....、Ｊ－１、Ｊ、Ｊ＋１、.....、Ｌ－１、Ｌ、Ｌ＋１、.....、等として参照される同じ大きさの記憶仮想ブロック５２および５３からなる、線形の１組のビューを提示する。仮想ブロック５２および５３は、ＲＡＩＤ区域４０の割り当てブロックと同じ大きさ（たとえば６４ｋバイト）であることが望ましい。仮想ブロック記憶空間５０は、ＲＡＩＤ区域４０で提示されているビューの割り当てブロックを指す参照テーブルまたはポインタ（たとえば図４の矢印５４で示す）によって表される。したがって、アプリケーション仮想ビュー５０の仮想ブロック５２および５３は、仮想ブロック・テーブルで維持されているポインタを介して、ＲＡＩＤ仮想ビュー４０の割り当てブロック４３および４５に関連づけられている。仮想ブロック・テーブルから参照できるＲＡＩＤ区域の型には少なくとも２つ（すなわちミラーおよびパリティ）の型がある。

【００３７】

ＲＡＩＤ管理装置１６は、物理的記憶空間全面にわたるＲＡＩＤ区域の構成を動的に変更できる。各型のＲＡＩＤ区域の数は、装置に記憶されつつあるユーザデータの量および物理的なディスク記憶空間の大きさに応じて増減される。その結果、ＲＡＩＤレベル仮想ビュー４０の中のＲＡＩＤ区域のディスク上へのマッピング、および、前面仮想ビュー５０のＲＡＩＤビュー４０へのマッピングは常にその状態が変化している。ＮＶＲＡＭ２１ａおよび２１ｂ（図１）のメモリマップ記憶装置は、ＲＡＩＤ管理装置１６がＲＡＩＤ区域をディスク上にマッピングするために使う最新のマッピング情報、および、２つの仮想ビュー間のマッピングに用いられる情報を維持している。ＲＡＩＤ管理装置は、ＲＡＩＤレベルのマッピングを動的に変更するのに応じて、その変更を反映するためにメモリマップ記憶装置のマッピング情報も更新する。

【００３８】

図１ないし４を参照して、データ記憶装置１０の移行動作について以下に述べる。

【００３９】

説明を続ける目的のために、アプリケーション・レベル仮想記憶空間５０の仮想ブロック５３は、物理的記憶空間３４の区域Ａ１に記憶されているパリティＲＡＩＤ区域４４の中の関連する割り当てブロック４５を参照する。そのような仮想ブロック５３は"パリティ仮想ブロック"と呼ばれ、関連する割り当てブロック４５は"パリティ割り当てブロック"と呼ばれる。同様に、仮想ブロック５２は、物理的記憶空間３４の区域Ａ０に記憶されているミラーＲＡＩＤ区域４２の中の関連する割り当てブロック４３を参照する。そのような仮想ブロック５２は、"ミラー仮想ブロック"と呼ばれ、関連する割り当てブロック４３は"ミラー割り当てブロック"と呼ばれる。

【００４０】

一般的に、データを１つのＲＡＩＤ区域から別のＲＡＩＤ区域に移行するためには、第１のＲＡＩＤレベル型（たとえば、ミラーすなわちレベル１）の割り当てブロックに関連した仮想ブロックが選択される。次に、第２のＲＡＩＤレベル型（たとえば、パリティすな

わちレベル５）を表す、未使用の割り当てブロックが探される。未使用の割り当てブロックが見つからなければ、割り当てブロックが作られる。次に、データが、選択された仮想ブロックに以前関連していた割り当てブロックから、未使用の割り当てブロックに転送され、データは冗長レベルの変更を受ける。たとえば、以前ミラー冗長にしたがって記憶されていたデータは、あらたにパリティ冗長にしたがって（またはその逆）記憶される。最後のステップとして、アプリケーション・レベル仮想記憶空間５０のＲＡＩＤレベル仮想記憶空間４０へのマッピング５４は、データのシフトを反映するために修正され更新される。第１のＲＡＩＤレベル型の割り当てブロックに以前関連していた選択された仮想ブロックは、こんどは、更新されたポインタを介して、移行されたデータを含む第２のＲＡＩＤレベル型の割り当てブロックを参照する。この処理中に起こるマッピング変更はメモリマップ記憶装置２１の中で更新される。

【００４１】

以下に、本発明の方法およびシーケンスにしたがった、ミラーおよびパリティ記憶区域の間のデータの移行について説明する。パリティ記憶区域からミラー記憶区域へデータを移行するために、次のシーケンスが使われる。

１．ＲＡＩＤ管理装置が、ミラーＲＡＩＤ区域４２の中に未使用のミラー割り当てブロック４３を探す。

２．見つからなければ、ＲＡＩＤ管理装置はミラー割り当てブロックを作る（以下に述べる）。

３．ＲＡＩＤ管理装置は、移行される仮想ブロックへの新しい記憶要求を中断する。

４．ＲＡＩＤ管理装置は、その仮想ブロックに対する実行中のデータ記憶要求がすべて完了するまで待つ。

５．仮想ブロック５３に関連しているパリティ割り当てブロック４５からのデータが臨時のメモリバッファに読み込まれる。

６．データが、ステップ２で選ばれたミラー割り当てブロック４３に書かれる。

７．仮想ブロック・テーブルが、ミラー割り当てブロック４３にあるデータの新しい位置を参照するように修正される。

８．中断された記憶要求が再開される。

上述の手続きにより、アプリケーション・レベル仮想ビュー５０の仮想ブロック５３は、パリティからミラー記憶に移行する。中間的仮想ビュー４０に関しては、データは、パリティＲＡＩＤ区域４４のパリティ割り当てブロック４５からミラーＲＡＩＤ区域４２のミラー割り当てブロック４３に移行されている。物理的記憶空間では、データは区域Ａ１から区域Ａ０に移動している。

【００４２】

未使用のミラー割り当てブロックが見つからなかった場合（上記のステップ１）、ＲＡＩＤ管理装置は、３つの手法からなる次のシーケンスを試みる。先ず、ＲＡＩＤ管理装置は、未使用（したがって未指定）のＲＡＩＤ区域を見つけることを試みる。このＲＡＩＤ区域は、移行が常に進行できることを保証するのに必要な未使用のＲＡＩＤレベル記憶のシステムしきい値を犯すことなく、ミラーＲＡＩＤ区域に変換できるものである。これが失敗し、装置が未使用のＲＡＩＤレベル記憶の予備量以上を持っている場合、装置は、未使用のパリティ割り当てブロックを未使用のＲＡＩＤ区域に集めるために、パリティ記憶内のデータを移行する。この移行により、上記のようにミラーＲＡＩＤ区域に変換できる未使用のＲＡＩＤ区域が生じる場合、装置は未使用のＲＡＩＤ区域をミラーＲＡＩＤ区域に変換する。そうでない場合、代替方法として、未使用のミラー割り当てブロックを求めるため、または、未使用のＲＡＩＤ区域をミラーＲＡＩＤ区域に変換するのに十分な未使用のＲＡＩＤレベル記憶を増やすまで、装置はデータをミラーからパリティ記憶に移行し、ミラー記憶を詰め、未使用のＲＡＩＤレベル記憶をパリティに変換する。ミラー割り当てブロックはパリティ割り当てブロックよりも多くの物理的記憶空間を占めるので、この最後の方法は、未使用のＲＡＩＤレベル記憶装置の純増につながる。

【００４３】

未使用のミラー割り当てブロックを求めて確立するために使われる作成／変換のプロトコルは、ユーザデータの量および物理的記憶空間の大きさに応じて、ＲＡＩＤ管理装置がパリティ区域とミラー区域との間のメモリ割り当てを選択的に調節できるので、利点がある。ＲＡＩＤ管理装置は、データ使用および記憶容量が変化するにしたがい、ミラー記憶に保持されるデータ量が最大になるように、上記の１つまたは複数の手法を用いる。

【００４４】

ＲＡＩＤ管理装置は、空き時間（アイドルタイム）中に未使用のＲＡＩＤ区域を作ることにより、空間作成シーケンスが未使用のミラー割り当てブロックを生じる間、記憶要求が待たなければならない状況を避けるように試みる。しかし、ある状況においては、空間作成シーケンスの間、記憶要求が中断される場合がある。ＲＡＩＤ管理装置は、仮想空間がＲＡＩＤビューより小さくなるような方法で仮想ブロック記憶空間を構成する。これにより、少なくとも１つのＲＡＩＤ区域に等しい空き空間が、移行またはその他の目的のためにとって置かれることが保証される。このようにして、この手法のシーケンスにより未使用のミラー割り当てブロックが常に生じる。

【００４５】

ミラーからパリティ記憶にデータを移行するために、次のシーケンスが使われる。

１．ＲＡＩＤ管理装置は、アクセス最近性またはアクセス頻度のような移行方針にしたがってミラーからパリティ記憶に移行するために、５２から仮想ブロックを選ぶ。

２．ＲＡＩＤ管理装置は、パリティＲＡＩＤ区域４４の中に未使用のパリティ割り当てブロック４５を探す。

３．そのようなブロックが見つからなかった場合、上述の作成方法にしたがって、移行のために確保しておいた空間がパリティＲＡＩＤ区域に変換される。

４．移行される仮想ブロックに対する新しい記憶要求が中断される。

５．ＲＡＩＤ管理装置は、その仮想ブロックへの実行中の記憶要求がすべて完了するまで待つ。

６．データが、仮想ブロック５２に関連しているミラー割り当てブロック４３から臨時のメモリバッファに読み込まれる。

７．データが、選ばれたパリティ割り当てブロック４５に書かれる。

８．仮想ブロック・テーブルが、パリティ割り当てブロック４５の中のデータの新しい位置を参照するために修正される。

９．仮想ブロックへのデータ要求が再開される。

上述の９ステップからなる２つのシーケンスは、２つの異なる冗長レベルの間のデータの移行を行うために、本発明による記憶装置がどのように動作するかの例を提供している。

【００４６】

データは、データアクセスの最近性またはデータアクセスの頻度の２つの移行方針のどちらかを使って、パリティＲＡＩＤ区域とミラーＲＡＩＤ区域との間で移行される。一般的な動作において、本発明によるデータ記憶装置は、最も古い、または最もアクセス頻度が少ない仮想ブロックを、ミラー記憶からパリティ記憶に移行する。逆に、最も新しい、または最もアクセス頻度が高い仮想ブロックを、パリティ記憶からミラー記憶に移行することは有用である。データ記憶装置の仕事の１つは、移行するのにふさわしい仮想ブロックを選択することである。

【００４７】

本発明では、選択処理を助けるために、仮想ブロックに単調に増加するマークがつけられる。移行方針がアクセスの最近性である場合、仮想ブロックに、データがその仮想ブロックにいつ記憶されたかを示すタイムスタンプがつけられる。タイムスタンプは、ディスクアレー制御装置１４の中の、約１分毎（ある具体化では65.5秒毎）に増加的に時間刻み（チック：tick）を出力するカウンタによって生成されるのが望ましい。それによって、カウンタは、そのような記憶ブロックにマークをつける適切な手段を形成する。その他の手段も、この代わりに使うことができる。アクセス頻度の場合、仮想ブロックに、その仮想ブロックのデータがアクセスされた回数を示す頻度スタンプがつけられる。頻度スタンプ

は、対応する仮想ブロックがアクセスされる度に１増える整数値である。

【００４８】

データ移行にふさわしい仮想ブロックを見つける１つの単純な手法は、第２仮想記憶空間５０全体をタイムスタンプまたは頻度スタンプでソートすることである。ソートすることによって、パリティＲＡＩＤ区域とミラーＲＡＩＤ区域との間の移行のための、最も古い／最も新しい、またはアクセス頻度が最も少ない／最も多い仮想ブロックが得られる。しかし、この手法は、仮想ブロック全体をソートするのにかなりのメモリ資源および処理時間がかかるので、あまり好ましくない。ソート手法の欠点は、ユーザデータが増えるにしたがい、いっそう顕著になる。処理時間は、ソートされる仮想ブロックの数の非線形関数であり、ブロックの数が増えるにしたがい処理時間はいっそう長くなる。

【００４９】

第２の手法すなわち本発明によるいっそう好ましい手法は、全面的なソートに頼らずに、適切な仮想ブロックを統計的に選択することである。この手法にしたがい、仮想ブロックをタイムスタンプまたは頻度スタンプの観点から統計的に特徴づけるために、ＲＡＩＤ管理装置１６によって１つまたは複数の度数分布表が作られる。次に、より古い／より新しい仮想ブロック、またはアクセス頻度がより少ない／より多い仮想ブロックを、仮想ブロックの残りの母集団から区分するスタンプしきい値を特定するために、度数分布表が処理される。ある区分スタンプしきい値が特定されると、その区分に入る十分な数の仮想ブロックが見つかるまで、アプリケーション・レベル仮想記憶空間がすばやく探索される。探索はそこで終わる。見つけられた仮想ブロックの中のデータは、たとえば他の割り当てブロックへの移行のような所望の機能を実行するために処理される。統計的な選択手法は、メモリの使用および処理時間を最適化するので、全体をソートする手法よりも好ましい手法である。

【００５０】

図５ないし１２に、本発明の階層データ記憶装置における度数分布表の種々な使用、作成および維持を示す。これらの側面および対応する新方法については、図１０ないし１２の流れ図を参照して説明する。本明細書で説明する方法は、ＲＡＩＤ管理装置１６のファームウェアとして具体化されるのが望ましい。

【００５１】

図１０は、初期化処理を示す。最初のステップ１００および１０２により、階層データ記憶装置のＲＡＩＤ環境が確立される。ステップ１００で、ＲＡＩＤ管理装置１６は、ディスクアレー１２によって提供される物理的記憶空間を、ＲＡＩＤレベル仮想記憶空間４０にマッピングする。ＲＡＩＤレベル仮想記憶空間は、物理的記憶空間を、種々なＲＡＩＤレベル、たとえばＲＡＩＤレベル１およびＲＡＩＤレベル５にしたがってデータを記憶する複数のＲＡＩＤ区域として表す。ステップ１０２で、ＲＡＩＤ管理装置は、ＲＡＩＤレベル仮想記憶空間を、物理的記憶空間を複数の仮想ブロックとして表すアプリケーション・レベル仮想記憶空間にマッピングする。仮想ブロックは、対応するＲＡＩＤ区域に関連する。データは、関連するＲＡＩＤ区域のＲＡＩＤレベルにしたがって仮想ブロックに記憶される。

【００５２】

ステップ１０４で、ＲＡＩＤ管理装置は、１つまたは複数の度数分布表を作る。ＲＡＩＤ区域のそれぞれの型（タイプ）に対して、１つの度数分布表が作られるのが望ましい。度数分布表は、ディスクアレー制御装置１４の揮発性ＲＡＭに形成され、装置の初期化の度に作りなおされるのが望ましい。代替方法として、度数分布表は、ディスクアレー制御装置１４のメモリマップ記憶装置２１のＮＶＲＡＭに常駐させることもできる。装置の初期化毎に表を作りなおすことに時間がかかることが予見される場合、ＮＶＲＡＭで具体化することが望ましい。これらの表の側面については、図５および６を参照して詳しく述べる。

【００５３】

図５は棒グラフの度数分布表６０を示し、図６は同じ表を概念化したメモリのかたちで示

す。度数分布表６０は、データ記憶装置に記憶されたデータを、予定の分布基準にしたがって統計的に特徴づける。２つの望ましい分布基準は、データアクセスの最近性またはデータアクセスの頻度である。度数分布表は、第２仮想記憶空間５０の仮想ブロックのレベルで特徴づけを行う。したがって、表は、ユーザデータを記憶するために使用される仮想ブロックについての統計的情報を含む。

【００５４】

度数分布表６０は、固定数の区間で区切られるのが望ましい。区間の数は、図５および６に示すように、たとえば１０である。区間は総称的に１～１０とラベルをつけてあるが、しかし実際には、そのような区間は、始めの境界値から終わりの境界値までの範囲の値で構成される。分布基準がアクセスの最近性である場合、区間は時間の範囲で表され、各区間は、仮想ブロックにつけられたタイムスタンプの範囲をカバーする。区間の１つの例は、１０２４個のタイムスタンプからなり、スタンプあたり65.5秒で、これで約１日の時間の中で使われる仮想ブロックのすべてがカバーできる。分布基準がアクセス頻度である場合、区間は頻度値で表され、各区間は、仮想ブロックにつけられた頻度スタンプの１つの範囲をカバーする。アクセス頻度方法の下での種々の区間およびカウントは、どのくらい活発にユーザがデータをアクセスするかに依存する。

【００５５】

度数分布表は、各区間内に起こる仮想ブロックの数を含む。これは、図５の垂直の棒および関連する数で表されている。たとえば、３２個の仮想ブロックが、区間１の範囲内に落ちるタイムスタンプまたは頻度スタンプを持っている。４８個の仮想ブロックが、区間２の範囲内に落ちるタイムスタンプまたは頻度スタンプを持っている。以下同様。表は、また、母集団全体を規定するために、仮想ブロックの累積度数すなわち全体数を作表する。この例の場合、母集団は５６６個の仮想ブロックからなる。

【００５６】

図６において、度数分布表６０は、母集団の数５６６を記憶する第１フィールド６２を含む。表は、また、母集団列（キュー：queue）６４、空き列６６、および１０個の区間フィールド６８を含む。母集団列６４は対象としている母集団のための区間要素を維持し、空き列６６は未使用の区間要素を追跡する。区間フィールド６８の各々は、関連する区間のカウント（区間１には３２、区間２には４８、等々）、始めの境界値、および区間の大きさを含む。始めの境界値および区間の大きさから、終わりの境界値は容易に計算できる。

【００５７】

度数分布表は、装置の初期化中に、アプリケーション・レベル仮想記憶空間に含まれる仮想ブロックを走査することによって作成される。最初、母集団列６４は空で、空き列６６は１０個の未使用の区間要素を持っている。１０個の区間フィールド６８のすべてのカウントおよび始めの境界値は省略時の値すなわちデフォルト値としてゼロに設定される。各区間フィールドの大きさは、あるデフォルト値、たとえば１０２４に初期化される。次に、仮想ブロックのスタンプ（たとえばタイムスタンプまたは頻度スタンプ）に基づいて、表が統計的に作られる（図１０のステップ１０６）。各区間を満たす仮想ブロックの数が数えられ、該当するフィールド６８に入力される。

【００５８】

各仮想ブロックのスタンプは、一致が見つかる（図１０のステップ１０７）まで、区間フィールド６８に対して順番に（図６の矢印でしめすように）調べられる。たとえば、仮想ブロックが、第３の区間フィールドの区間内に落ちるスタンプを持っていると仮定する。仮想ブロックのスタンプは、先ず、第１の区間フィールドの区間に対して評価される。仮想ブロックがこの区間フィールドの外に落ちるので、次に第２の区間フィールドと比較される。各区間フィールドは、分析を次の区間フィールドのアドレスに方向づけるためのポインタを持っていることに留意されたい。この処理は、仮想ブロックのスタンプが比較され該当する区間フィールドに一致する（この場合、第３の区間フィールド）までフィールド毎に続けられ、一致したとき、その区間フィールドのカウントが増やされる（図１０の

ステップ１０８）。未使用の仮想ブロックは関連するスタンプを持っていないので、度数分布表に表されないことに留意されたい。最初と最後の区間ポインタ値を維持することにより、母集団列６４は、最初と最後の区間フィールド６８を見つけてそれを指すことが容易にできる。

【００５９】

仮想ブロックが走査され、統計的情報がとられるにしたがい、母集団列６４は満ち始め、空き列６６は空になり始める。表が各区間にカウントを持つと、空き列は空になる（空き列のヌルまたは接地記号で表されている）。２つ以上の表が必要な場合（以下に詳しく述べるように）、同じ走査ステップ中に同時に作成することができることに留意されたい。

【００６０】

度数分布表が作成されると、それは直ぐ使うことができる。図１１および図１２の流れ図は、度数分布表を使用する望ましい動作モードを示す。度数分布表を使用するその他の動作も可能である。第１のモードすなわちホストアクセス動作モードは、ユーザが既存または新しい仮想ブロックをアクセスすると、表を更新するものである。第２のモードすなわち移行動作モードは、ＲＡＩＤ区域間の移行のために使用する仮想ブロックを識別するために表を区分し、度数分布表を更新することに関する。

【００６１】

図１１は、ユーザデータを記憶するために既存または新しい仮想ブロックが使われるときに表が絶え間なく更新される、ホストアクセス動作モードを示す。ステップ１０９で、先ず、ユーザが既存の仮想ブロックをアクセスしているのか、新しい仮想ブロックを作っているのかどうかが判断される。ユーザが既存の仮想ブロックをアクセスしている（ステップ１０９がＮＯ）場合、既存の仮想ブロックに新しいスタンプがつけられる。新しいスタンプを反映するために度数分布表が更新される。これは、古いスタンプを持っている区間のカウントを減らし（ステップ１１０）、ステップ１１１の後で、新しいスタンプを持っている区間のカウントを増やすことによって行われる。

【００６２】

図７は、既存の仮想ブロックがアクセスされるときの表の更新の処理を示す。区間６の範囲のタイムスタンプを持つ仮想ブロックがアクセスされたと仮定する。新しい、最も最近のタイムスタンプが仮想ブロックにつけられる。単調に増加するスタンプを持つことの１つの重要な利点は、最も最近／最も頻繁にアクセスされた仮想ブロック、および最も古く／最も少なくアクセスされた仮想ブロックが最初および最後の区間に置かれているので、装置がこれらの仮想ブロックを見つけて処理することが容易であることである。したがって、最も最近のタイムスタンプは、予想通り、区間１０の範囲内に落ちる。度数分布表６０は、区間６のカウントを１（８０から７９に）減らし、区間１０のカウントを１（７６から７７に）増やすことによって更新される。

【００６３】

図１１を参照すると、仮想ブロックが新しい（ステップ１０９がＹＥＳ）場合、分析は図１１の判断ステップ１１１に進み、新しい仮想ブロックが表の範囲内の既存の区間に一致するかどうかが決められる。仮想ブロックにスタンプをつけるのに単調な性質があることにより、仮想ブロックは、単に、表の最後の区間と比較される。新しい仮想ブロックが最後の区間に一致する（判断ステップ１１１がＹＥＳ）場合、最後の区間に関連するカウントが増やされる（ステップ１１２）。たとえば、仮想ブロックが区間１０の範囲に落ちるタイムスタンプを持っている場合、区間１０のカウントは７６から７７に増やされる。他方、仮想ブロックが、最後の区間のどれも満たさないスタンプを持っている（判断ステップ１１１がＮＯ）場合、ＲＡＩＤ管理装置はすばやく空きの区間があるかどうかを調べる（ステップ１１３）。この状況は、最後の仮想ブロックの度数分布を表すのに、設定した区間の数（この例では１０区間）より少ない数が使われた場合に起こる。空きの区間がある（ステップ１１３がＹＥＳ）場合、ステップ１１６で、新しい最後の区間が、新しい仮想ブロックを含むように規定される。

【００６４】

空きの区間がない（ステップ１１３がＮＯ）場合、増加する分布範囲を収めるために、新しい区間が作られなければならない。度数分布表は固定した数の区間で区切られているので、新しい区間を作るには、度数分布表を若干調整する必要がある。ＲＡＩＤ管理装置は、単に新しい区間を加えることはせず、先ず、新しい区間を収容するために、表の内部に余裕を作る。

【００６５】

図８は、新しい区間を生成する処理を示す。ＲＡＩＤ管理装置は、新しい接合された区間を形成するために、先ず、隣接する区間を組み合わせる（図１１のステップ１１４）。新しい接合区間は、接合された２つの区間のカウントの和に等しいカウントを持つ。隣接する２つの区間は、それぞれのカウントを合わせたときに最小のカウントになるようなものを選ぶのが望ましい。図８において、表６０'の新しい区間２を形成するために、区間２および３が接合される。新しい区間２のカウントは７６（４８＋２８＝７６）である。

【００６６】

この処理により、新しい最後の区間（たとえば、新しい最後の区間１０）を規定する（図１１のステップ１１６）ために、度数分布表の空間を空ける。次に、新しい仮想ブロックを反映するために、新しい最後の区間のカウントが増やされる（ステップ１１８）。表６０'において、新しい仮想ブロックが新しい最後の区間に一致し前の区間に一致しなかったことを統計的に示すために、新しい最後の区間１０がカウント１を持っている。この接合処理を通して、度数分布表は定められた区間の数で区切られたまま残り、扱い易く、その中の統計的情報は正確に維持される。

【００６７】

上述したように、本発明の具体化では、各ＲＡＩＤレベルに対して１つの度数分布表が作られる。階層ディスクアレーがＲＡＩＤレベル１および５を使う場合、ＲＡＩＤレベル１のミラーＲＡＩＤ区域に関連する仮想ブロックに対して第１の度数分布表が作られ、ＲＡＩＤレベル５のパリティＲＡＩＤ区域に関連する仮想ブロックに対して第２の度数分布表が作られる。３つ以上のＲＡＩＤレベルが使われる場合、それに対応する３つ以上の度数分布表が作られる。このメモリ配列の１つの使用例は、"中くらいに古い"（middle-aged）仮想ブロックを、その特定のデータに最もふさわしい性能特性を持つ記憶レベルに移行することである。

【００６８】

図９に、２つの度数分布表６０および７０を示す。上部に示した表６０は、ＲＡＩＤレベル１の記憶に関連した仮想ブロックを特徴づけるために使われる。下部に示した表７０は、ＲＡＩＤレベル５の記憶に関連した仮想ブロックを特徴づけるために使われる。

【００６９】

図９は、また、異なるＲＡＩＤレベルの移行が度数分布表に与える影響を示す。移行のために表を使う方法は、図１２の移行動作モードに示される。この例の目的のために、図９の表はデータアクセスの最近性の分布基準に基づいており、区間は時間で表されている。ＲＡＩＤ管理装置１６が、新しいデータにミラー記憶空間を空けるために、"より古い"仮想ブロックをミラー記憶からパリティ記憶に移行したいと仮定する。この場合、ミラーＲＡＩＤ区域はデータ移行のための移行源ＲＡＩＤ区域として動作し、パリティＲＡＩＤ区域は移行先ＲＡＩＤ区域として動作する（図１２のステップ１２０）。

【００７０】

ＲＡＩＤ管理装置は、先ず、表６０の中の仮想ブロックの母集団を、それぞれが大体同じ数の要素を持つ複数の同じ大きさのセグメントに分ける（ステップ１２２）。たとえば、図９の上部の表６０に示すように、ＲＡＩＤ管理装置は、母集団を５つのセグメントに分ける。母集団が５６６である場合、各セグメントは、理論的には、およそ１１３個の仮想ブロックを表す。第１のセグメントは区間１ないし４にわたる。というのは、区間１ないし３だけでは、各セグメントのしきい値１１３より少ない１０５個の仮想ブロックしかないからである。

【００７１】

次に、ＲＡＩＤ管理装置は、第１のセグメントがより古い仮想ブロックを含むので、第１のセグメントを選ぶ（図１２のステップ１２４）。総称的に１ないし４として参照した時間区間内に落ちるタイムスタンプを持つ仮想ブロックが、この基準を満たす。ＲＡＩＤ管理装置は、第１のセグメントの最も高い境界値を決めるが、これは、この例の場合、区間４の終わりの境界時間値である。終わりの境界時間値は、区間４の始めの境界時間値をとり、区間４の大きさを加えることによって計算できる。終わりの境界時間値は、比較のためのしきい値として使われる。

【００７２】

次に、ＲＡＩＤ管理装置は、実際に１つまたは複数の仮想ブロックを見つけるために、アプリケーション・レベル仮想記憶空間の仮想ブロックを探索する（ステップ１２６）。各仮想ブロックのタイムスタンプは、その仮想ブロックがしきい時間値よりも早いタイムスタンプを持ち、したがって第１のセグメントに入れるべきかどうかを調べるために、しきい時間値と比較される。十分な数の仮想ブロックが見つかったところで探索は止まる。次に、見つかった仮想ブロックは、ミラー記憶からパリティ記憶に移行される（ステップ１２８）。

【００７３】

移行が行われたことを反映するために、２つの度数分布表６０および７０が更新される（ステップ１３０）。ＲＡＩＤレベル１の表６０の区間２で以前表されていた３つの仮想ブロックが移行されたことによって、いま、ＲＡＩＤレベル５の表７０の区間１０に反映されていると仮定する。表６０の区間２に関連するカウントは、単純に、仮想ブロックがなくなったことを反映するために、４８から４５に３つ減らされる。同様に、表７０の区間１０に関連するカウントは、仮想ブロックが増えたことを反映するために、４７から５０に３つ増やされる。これらの３つの仮想ブロックのタイムスタンプは、移行中に変わらないことに留意されたい。データ移行の結果生じた分布の変更を反映するために、各ＲＡＩＤレベルの統計的情報だけが修正される。

【００７４】

移行は、バッチ的手続きではなく、連続した方法で行われるのが望ましいことに留意されたい。したがって、基準を満たすブロックが見つかるたびに、ブロック毎に移行が行われ、表が更新される。

【００７５】

度数分布表を使うことは、全体をソートする方法にくらべて、多くの利点をもたらす。統計的手法は、ソートによる方法よりもはるかに性能がよく、メモリ資源の使用もはるかに少ない。母集団を区分化し、仮想ブロックを識別するための１つのセグメントを選び、仮想ブロックをすばやく探索する処理方法は、全体をソートする手法とほぼ等しい結果を生じる。この方法は費用をかけずに、識別判断に基づいた選択を可能にするので、この仕事は、効率よく費用効果のよい方法で達成できる。さらに、この処理は、特定されたセグメントに十分な数の仮想ブロックが見つかり次第止めることができる。より早く止めることにより、途中で断ち切られた時間は、一般的に、全体をソートする時間よりもはるかに少ない。

【００７６】

もう１つの利点は、全体をソートする手法の非線形関数と異なり、処理時間は、探索する仮想ブロックの数の線形関数であることである。したがって、ディスクアレーに記憶されているユーザデータの量に関し、より古い／より新しい仮想ブロック、および最近アクセスされない／最近アクセスされた仮想ブロックがよりすばやく識別される。

【００７７】

また、統計的度数分布表は、異なる型の記憶装置の間を移動させるデータを選ぶのに使うこともできる。例として、データ記憶装置が、たとえばリールからリールのテープ記憶装置のようなバックアップ記憶機構をさらに備えている場合、ディスクアレーからバックアップ記憶機構にデータを移すために、"非常に古い"（ancient）記憶ブロックを選ぶのに統計的度数分布表を使うことができる。

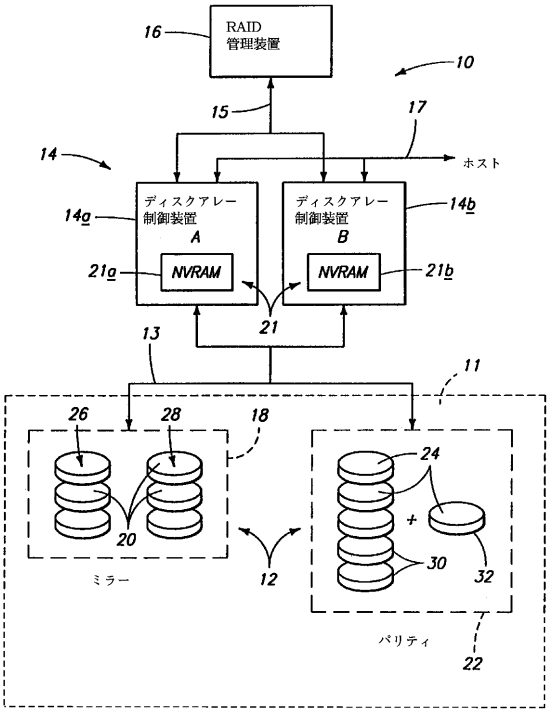
【００７８】
本発明には、例として次のような実施態様が含まれる。
（１）物理的記憶空間を複数の仮想ブロックからなる母集団として表す仮想記憶空間にマッピングされた物理的記憶空間（３４）を規定する複数の記憶ディスク（１２）を持つデータ記憶装置（１０）のデータを選択する方法であって、
個々の区間が、関連する区間に対する分布基準を満たす仮想ブロックの数のカウントを持つ選択された数の区間を有する、前記データ記憶装置（１０）上に記憶されたデータを予定の分布基準にしたがって特徴づける度数分布表（６０）を作成するステップと、
母集団のサブセットである仮想ブロックの選択基準を特定するために前記度数分布表（６０）を区分するステップと、
前記選択基準を満たす１つまたは複数の仮想ブロックを見つけるために、仮想記憶空間の中の仮想ブロックを探索するステップと、
を有する方法。
【００７９】
（２）仮想ブロックにデータがいつ記憶されたかを示すタイムスタンプを仮想ブロックにつける付加ステップをさらに有し、前記予定の分布基準がデータアクセスの最近性で、区間が時間の範囲であり、時間区間に関連するカウントがその時間区間の範囲のタイムスタンプを持つ仮想ブロックの数を表す、上記（１）に記載の方法。
（３）仮想ブロックの中のデータがアクセスされた回数を示す頻度スタンプを仮想ブロックにつける付加ステップをさらに有し、前記予定の分布基準がデータアクセス頻度で、区間が頻度値に関連し、区間に関連するカウントがそれぞれの区間の範囲の頻度値を持つ仮想ブロックの数を表す、上記（１）に記載の方法。
（４）新しい仮想ブロックが使われるときに前記度数分布表（６０）を更新するステップをさらに有し、前記更新ステップが、（１）新しい仮想ブロックが既存の区間に対する分布基準を満たす場合、その既存の区間に関連するカウントを増やすステップ、または（２）新しい仮想ブロックが既存の区間に対する分布基準を満たさない場合、新しい区間を作り、その新しい区間に最初のカウントを入れるステップのどちらかを有する、上記（１）に記載の方法。
（５）新しい区間を生成する前記ステップが、
接合される複数の区間のカウントの和に等しいカウントを持つ１つの接合区間を規定するために、前記度数分布表（６０）の中の２つ以上の区間を接合するステップと、
前記新しい仮想ブロックをカバーするために分布基準を延ばす新しい区間を規定するステップと、
を有する、上記（４）に記載の方法。
（６）既存の仮想ブロックがアクセスされるときに度数分布表（６０）を更新する付加ステップをさらに有し、前記更新ステップが、アクセスされる前に既存の仮想ブロックが分布基準を満たした区間のカウントを減らし、アクセスされた後に既存の仮想ブロックが分布基準を満たした別の区間のカウントを増やす、上記（１）に記載の方法。
【００８０】
（７）物理的記憶空間（３４）を規定する複数の記憶ディスクを持つ階層データ記憶装置の中で移行を行うためにデータを選択する方法であって、物理的記憶空間が、物理的記憶空間を異なるＲＡＩＤレベルにしたがってデータを記憶する少なくとも２つのＲＡＩＤ区域として表すＲＡＩＤレベル仮想記憶空間（４０）にマッピングされ、前記ＲＡＩＤレベル仮想記憶空間が、物理的記憶空間を複数の仮想ブロックからなる母集団として表すアプリケーション・レベル仮想記憶空間（５０）にマッピングされ、前記方法が、
個々の区間が、関連する区間に対する分布基準を満たす仮想ブロックの数のカウントを持つ選択された数の区間を有する、第１のＲＡＩＤレベルの下で第１のＲＡＩＤ区域に記憶されたデータを予定の分布基準にしたがって特徴づける第１の度数分布表（６０）を作るステップと、
個々の区間が、関連する区間に対する分布基準を満たす仮想ブロックの数のカウントを持

つ選択された数の区間を有する、第２のＲＡＩＤレベルの下で第２のＲＡＩＤ区域に記憶されたデータを予定の分布基準にしたがって特徴づける第２の度数分布表（７０）を作るステップと、
個々のセグメントが前記第１の度数分布表の１つまたは複数の区間を持つ、第１の母集団の複数のセグメントを規定するステップと、
前記複数のセグメントから少なくとも１つのセグメントを選択するステップと、前記選択されたセグメントに含まれる前記１つまたは複数の区間に対する分布基準を満たし、前記第１ＲＡＩＤレベルの下の前記第１ＲＡＩＤ区域に記憶されたデータを表す１つまたは複数の仮想ブロックを見つけるために、前記アプリケーション・レベル仮想記憶空間（５０）の中に仮想ブロックを探索するステップと、
見つかった仮想ブロックを前記第１ＲＡＩＤ区域から前記第２ＲＡＩＤ区域に移行するステップと、
を有する方法。
（８）仮想ブロックに単調に増える値を持つスタンプをつけて、より新しくスタンプをつけられた仮想ブロックが、前記第１および第２の度数分布表の最後の区間にカウントされるようにする付加ステップをさらに有する、上記（７）に記載の方法。
【００８１】
（９）物理的記憶空間（３４）を規定する複数の記憶ディスク（１２）を持つディスクアレー（１１）と、
ディスク（１２）におよびディスクからデータ転送を調整するようにディスクアレー（１１）に結合されたディスクアレー制御装置（１４）と、
物理的記憶空間を様々なＲＡＩＤレベルにしたがってデータを記憶する複数のＲＡＩＤ区域として表すＲＡＩＤレベル仮想記憶空間（４０）、および、物理的記憶空間を複数の仮想ブロックからなる母集団として表すアプリケーション・レベル仮想記憶空間（５０）を、記憶ディスク（１２）の物理的記憶空間（３４）にマッピングするために、ディスクアレー制御装置（１４）を動作させるように結合されたＲＡＩＤ管理装置（１６）と、
記憶ディスク（１２）上の第１および第２のＲＡＩＤ区域の間で仮想ブロックを移行させるＲＡＩＤ管理装置（１６）と、
個々の区間が、関連する区間に対する分布基準を満たす仮想ブロックの数のカウントを持つ選択された数の区間を有する、第１および第２のＲＡＩＤ区域に記憶されたデータを予定の分布基準にしたがって特徴づける第１および第２の度数分布表（６０、７０）と、
それぞれの母集団のサブセットを構成する仮想ブロックを選択する選択基準を特定するために、前記第１または第２の度数分布表（６０、７０）のどちらかを区分することによって、移行のための仮想ブロックを選択するＲＡＩＤ管理装置（１６）と、
を備えるデータ記憶装置。
（１０）単調に増える値を持つスタンプで仮想ブロックをマークする手段をさらに備える、上記（９）に記載のデータ記憶装置。
【００８２】
【発明の効果】
度数分布表を用いて、データアクセスの最近性または頻度によりデータを選択し移行することにより、階層データ記憶装置の性能および信頼性が最適化できる。
【図面の簡単な説明】
【図１】本発明によるディスクアレー・データ記憶装置を示すブロック図。
【図２】複数の記憶ディスク上の記憶空間、および、ＲＡＩＤレベル１によるデータ記憶を示す図。
【図３】複数の記憶ディスク上の記憶空間、および、ＲＡＩＤレベル５によるデータ記憶を示す図。
【図４】１つの物理的記憶空間に関して２つの仮想記憶空間がマッピングされている、本発明によるメモリマップ配列を示す図。
【図５】本発明による、棒グラフで表した度数分布表を示す図。

【図６】図５の度数分布表を、ポインタをつけた具体化の形で表し、表がどのように作成されるかを示す図。
【図７】既存の仮想ブロックをアクセスした結果、図５の度数分布表の更新を示す図。
【図８】一定の数の区間で区切られた度数分布表に新しい区間を生成する手法を示すために、２つの区間を接合する前の図５の表、および、２つの区間を接合した後の同じ表を示す図。
【図９】図５と同じＲＡＩＤレベル１の度数分布表、およびＲＡＩＤレベル５の度数分布表を示し、２つのＲＡＩＤレベルの間でデータを移行したことによる、表に与える影響を示す図。
【図１０】本発明の階層データ記憶装置で使用される度数分布表を初期化する方法を示す流れ図。
【図１１】本発明による、ホストアクセス動作モードで度数分布表を使用する方法を示す流れ図。
【図１２】本発明による、移行動作モードで度数分布表を使用する方法を示す流れ図。
【符号の説明】

１０	データ記憶装置
１１	階層ディスクアレー
１２	記憶ディスクまたはディスク駆動装置
１３	インターフェース・バス（ＳＣＳＩバス）
１４、14a、14b	ディスクアレー制御装置
１５	インターフェース・プロトコル
１６	ＲＡＩＤ管理装置
１７	入出力インターフェース・バス
１８	ミラー・グループ
２１	メモリマップ記憶装置
２１ａ、２１ｂ	不揮発性ＲＡＭ
２２	パリティ・グループ
２６	ディスクの第１の組
２８	ディスクの第２の組
３０	５個のディスク
３２	６番目のディスク
60、60'、70	度数分布表
６４	母集団列
６６	空き列
６８	区間フィールド

【図１】



【図２】

ディスク

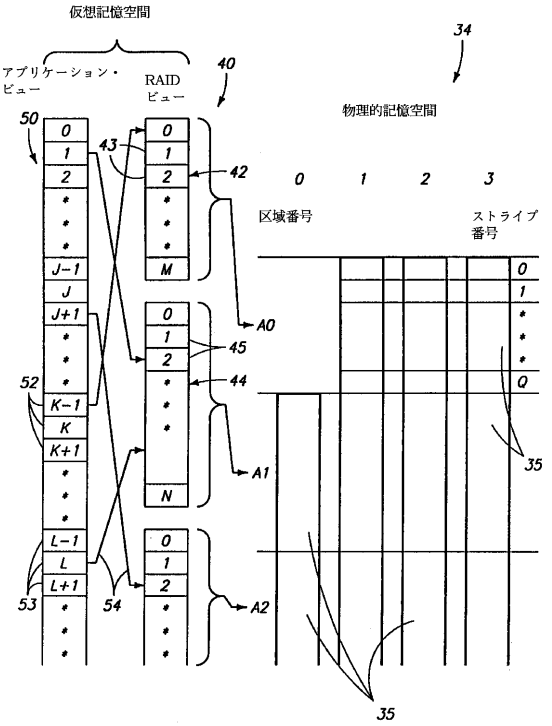
0	1	2	3	ストライプ番号
0	0'	1	1'	0
2	2'	3	3'	1
4	4'	5	5'	2
6	6'	7	7'	3
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
T-1	T-1'	T	T'	S

【図３】

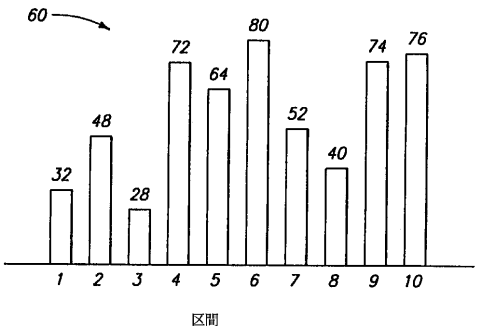
ディスク

0	1	2	3	ストライプ番号
0	1	2	P	0
3	4	P	5	1
6	P	7	8	2
P	9	10	11	3
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
P	R-2	R-1	R	Q

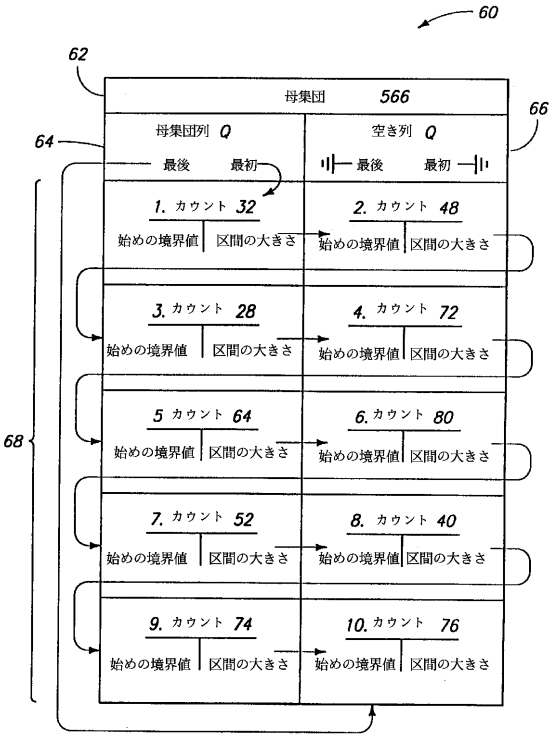
【図４】



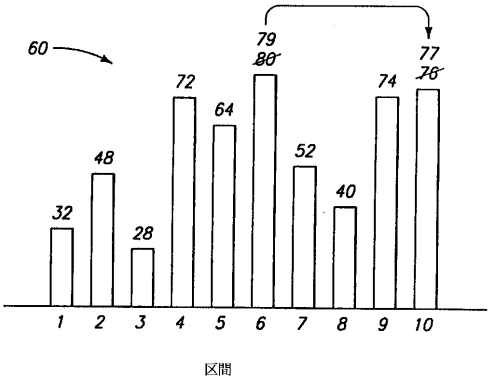
【図５】



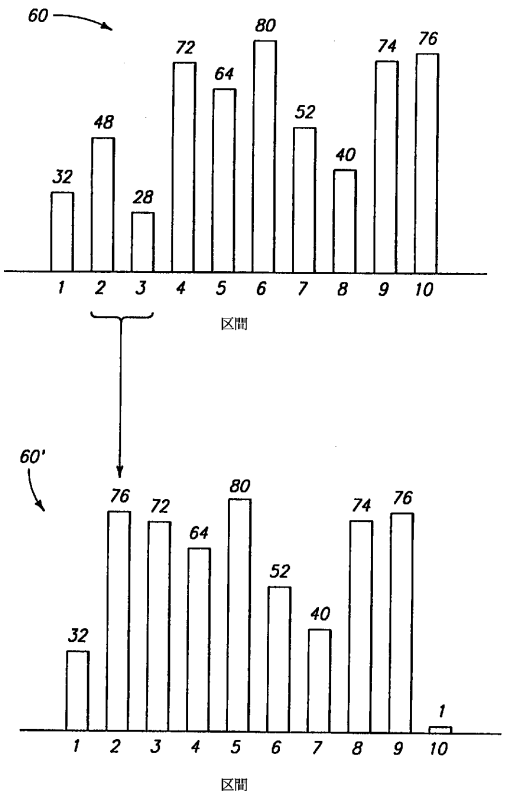
【図６】


60
62
母集団 566
66
64
母集団列 Q
空き列 Q
最後 最初
最後 最初
1. カウント 32
始めの境界値 区間の大きさ
2. カウント 48
始めの境界値 区間の大きさ
3. カウント 28
始めの境界値 区間の大きさ
4. カウント 72
始めの境界値 区間の大きさ
5 カウント 64
始めの境界値 区間の大きさ
6. カウント 80
始めの境界値 区間の大きさ
7. カウント 52
始めの境界値 区間の大きさ
8. カウント 40
始めの境界値 区間の大きさ
9. カウント 74
始めの境界値 区間の大きさ
10. カウント 76
始めの境界値 区間の大きさ
68

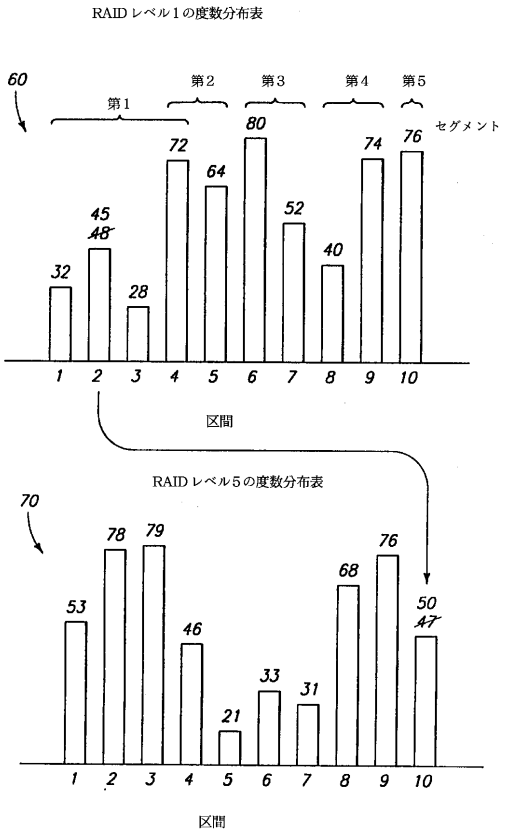
【図７】


60
32
48
28
72
64
79
80
52
40
74
77
76
1 2 3 4 5 6 7 8 9 10
区間

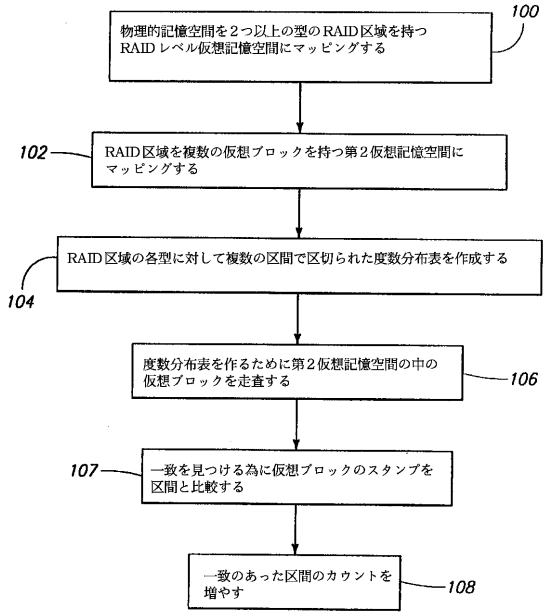
【図８】


60
32
48
28
72
64
80
52
40
74
76
1 2 3 4 5 6 7 8 9 10
区間
60'
32
76
72
64
80
52
40
74
76
1
1 2 3 4 5 6 7 8 9 10
区間

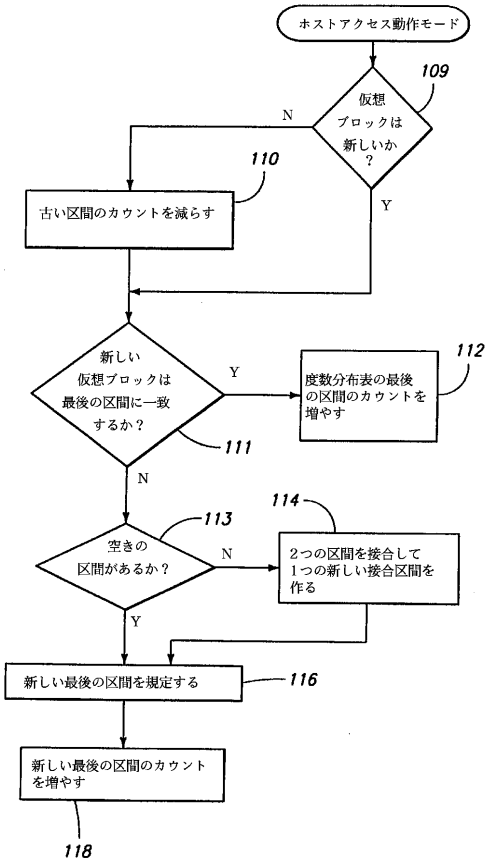
【図９】


RAID レベル１の度数分布表
60
第1
第2
第3
第4
第5
セグメント
32
45
48
28
72
64
80
52
40
74
76
1 2 3 4 5 6 7 8 9 10
区間
RAID レベル５の度数分布表
70
53
78
79
46
21
33
31
68
76
50
47
1 2 3 4 5 6 7 8 9 10
区間

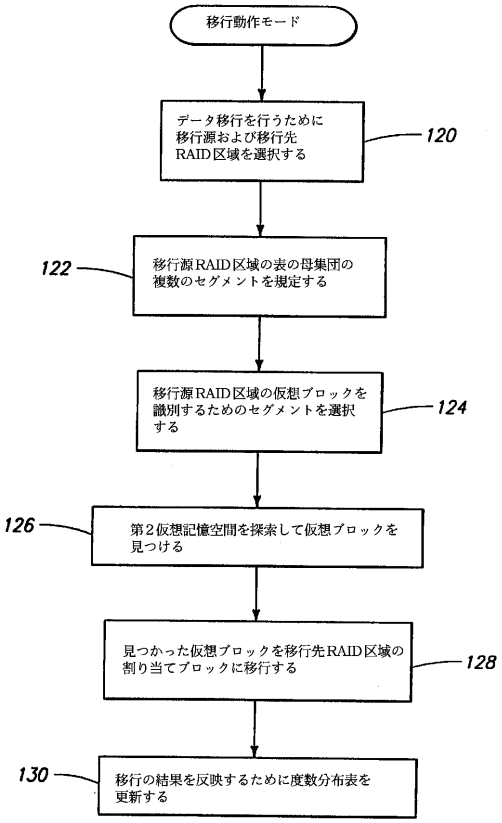
【図１０】


物理的記憶空間を２つ以上の型のRAID区域を持つRAIDレベル仮想記憶空間にマッピングする
100
RAID区域を複数の仮想ブロックを持つ第２仮想記憶空間にマッピングする
102
RAID区域の各型に対して複数の区間で区切られた度数分布表を作成する
104
度数分布表を作るために第２仮想記憶空間の中の仮想ブロックを走査する
106
一致を見つける為に仮想ブロックのスタンプを区間と比較する
107
一致のあった区間のカウントを増やす
108

【図１１】


ホストアクセス動作モード
仮想ブロックは新しいか？
109
N
Y
古い区間のカウントを減らす
110
新しい仮想ブロックは最後の区間に一致するか？
111
Y
度数分布表の最後の区間のカウントを増やす
112
N
空きの区間があるか？
113
N
2つの区間を接合して1つの新しい接合区間を作る
114
Y
新しい最後の区間を規定する
116
新しい最後の区間のカウントを増やす
118

【図１２】


移行動作モード
データ移行を行うために移行源および移行先RAID区域を選択する
120
移行源RAID区域の表の母集団の複数のセグメントを規定する
122
移行源RAID区域の仮想ブロックを識別するためのセグメントを選択する
124
第２仮想記憶空間を探索して仮想ブロックを見つける
126
見つかった仮想ブロックを移行先RAID区域の割り当てブロックに移行する
128
移行の結果を反映するために度数分布表を更新する
130

フロントページの続き

(72)発明者　ダグラス・エル・ヴォイト
アメリカ合衆国８３７０２アイダホ州ボイジー、エヌ・ツウェンティーフォース　３０３０

審査官　木村　貴俊

(56)参考文献　特開平０４－２５９０４８（ＪＰ，Ａ）
特開平０５－２５０２３７（ＪＰ，Ａ）
USING DUAL ACTUATOR SHARED DATA DIRECT ACCESS STORAGE DEVICES DRIVES IN A REDUNDANT ARRAY, IBM TECHNICAL DISCLOSURE BULLETIN, 米国, IBM Corp., １９９１年　1月, Vol.33 No.8, p.270-272

(58)調査した分野(Int.Cl.[7], ＤＢ名)
G06F 3/06- 3/08
G06F 12/00-12/16