

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6065695号  
(P6065695)

(45) 発行日 平成29年1月25日 (2017. 1. 25)

(24) 登録日 平成29年1月6日 (2017. 1. 6)

(51) Int. Cl.

F I

G 0 6 F 12/00 (2006.01)

G 0 6 F 12/00 5 0 1 B

G 0 6 F 3/06 (2006.01)

G 0 6 F 3/06 3 0 1 M

G 0 6 F 3/06 3 0 1 Z

請求項の数 10 (全 22 頁)

(21) 出願番号 特願2013-63854 (P2013-63854)  
 (22) 出願日 平成25年3月26日 (2013. 3. 26)  
 (65) 公開番号 特開2014-191398 (P2014-191398A)  
 (43) 公開日 平成26年10月6日 (2014. 10. 6)  
 審査請求日 平成27年11月11日 (2015. 11. 11)

(73) 特許権者 000005223  
 富士通株式会社  
 神奈川県川崎市中原区上小田中4丁目1番  
 1号  
 (74) 代理人 100092978  
 弁理士 真田 有  
 (74) 代理人 100112678  
 弁理士 山本 雅久  
 (72) 発明者 阿部 智明  
 神奈川県川崎市中原区上小田中4丁目1番  
 1号 富士通株式会社内

審査官 井上 宏一

最終頁に続く

(54) 【発明の名称】 ストレージ制御方法、ストレージシステム、およびストレージ制御プログラム

(57) 【特許請求の範囲】

【請求項 1】

特性の異なる複数のディスクを有する階層化ディスクを含むストレージ装置と、前記ストレージ装置にアクセスする情報処理装置とが接続されたストレージシステムであって、  
 前記ストレージ装置は、

前記階層化ディスクに保存された各ファイルに含まれるデータブロックの記憶位置を含むブロック情報と、前記データブロックに対するアクセス頻度を含むアクセス情報とを、  
 前記情報処理装置に通知するディスク制御部を有し、

前記情報処理装置は、

前記階層化ディスクに対する入出力要求を受けると前記入出力要求の対象ファイルと前記対象ファイルのアクセス態様とを前記ファイル毎のアクセス態様情報として特定するアクセス態様特定部と、

前記アクセス態様特定部によって特定された前記ファイル毎のアクセス態様情報と、前記ディスク制御部から通知された前記ブロック情報および前記アクセス情報に含まれる情報とを対応付けてファイルアクセス態様情報として管理し、前記ファイルアクセス態様情報に基づいて再配置対象データブロックを判断し、前記ファイルアクセス態様情報に基づいて前記再配置対象データブロックのアクセス態様を判断し、前記再配置対象データブロックのアクセス態様と前記ブロック情報とに基づいて、前記再配置対象データブロックを配置する前記階層化ディスクにおける記憶位置の再配置の内容を決定して前記ストレージ装置の前記ディスク制御部に指示する管理部と、を有し、

10

20

前記ストレージ装置の前記ディスク制御部は、前記情報処理装置の前記管理部から指示された前記再配置の内容に従って、前記再配置対象データブロックを配置する前記階層化ディスクにおける記憶位置の再配置を行なう、ストレージシステム。

【請求項 2】

前記アクセス態様特定部は、前記入出力要求された複数のファイルのうち、前記ファイルアクセス態様情報に登録されていないファイルについて前記アクセス態様を特定する、請求項 1 記載のストレージシステム。

【請求項 3】

前記アクセス態様特定部は、前記入出力要求が前記対象ファイルの入出力位置をインデックスでシークするか否かを測定し、シークしない場合に前記対象ファイルのアクセス態様をシーケンシャルアクセスと判断する一方、シークする場合に前記対象ファイルのアクセス態様をランダムアクセスと判断する、請求項 1 または請求項 2 に記載のストレージシステム。

【請求項 4】

前記情報処理装置の前記管理部は、前記再配置対象データブロックのアクセス態様がシーケンシャルアクセスである場合、且つ、前記再配置対象データブロックが再配置される前記複数のディスクのうちの一のディスク上に、連続する複数のデータブロックが存在する場合、前記一のディスク上に前記複数のデータブロックを連続的に再配置するように、前記再配置の内容を決定して前記ストレージ装置の前記ディスク制御部に指示する、請求項 3 記載のストレージシステム。

【請求項 5】

前記情報処理装置の前記管理部は、前記再配置対象データブロックのアクセス態様がシーケンシャルアクセスである場合、且つ、前記再配置対象データブロックが再配置される前記複数のディスクのうちの一のディスク上に、前記再配置対象データブロックの前側または後側に続く他のデータブロックが存在しない場合、前記一のディスク上に再配置された前記再配置対象データブロックの前側または後側に、前記他のデータブロック用の空き領域を設けるように、前記再配置の内容を決定して前記ストレージ装置の前記ディスク制御部に指示する、請求項 3 または請求項 4 に記載のストレージシステム。

【請求項 6】

前記情報処理装置の前記管理部は、前記空き領域を設けた後、所定期間が経過しても、または、前記再配置対象データブロックの再配置を所定回数行なっても、前記空き領域に前記他のデータブロックが再配置されない場合、前記空き領域を解放するように、前記再配置の内容を決定して前記ストレージ装置の前記ディスク制御部に指示する、請求項 5 記載のストレージシステム。

【請求項 7】

前記情報処理装置の前記管理部は、アクセス態様がランダムアクセスであるデータブロックを、アクセス態様がシーケンシャルアクセスであるデータブロックの配置状態に応じて再配置するように、前記再配置の内容を決定して前記ストレージ装置の前記ディスク制御部に指示する、請求項 3 ～ 請求項 6 に記載のストレージシステム。

【請求項 8】

特性の異なる複数のディスクを有する階層化ディスクを含むストレージ装置と、前記ストレージ装置にアクセスする情報処理装置とが接続されたストレージシステムにおいて、前記ストレージ装置は、

前記階層化ディスクに保存された各ファイルに含まれるデータブロックの記憶位置を含むブロック情報と、前記データブロックに対するアクセス頻度を含むアクセス情報とを、前記情報処理装置に通知し、

前記情報処理装置は、

前記階層化ディスクに対する入出力要求を受けると前記入出力要求の対象ファイルと前記対象ファイルのアクセス態様とを前記ファイル毎のアクセス態様情報として特定し、

特定された前記ファイル毎のアクセス態様情報と、前記ストレージ装置から通知された

10

20

30

40

50

前記ブロック情報および前記アクセス情報に含まれる情報とを対応付けてファイルアクセス態様情報として管理し、

前記ファイルアクセス態様情報に基づいて再配置対象データブロックを判断し、

前記ファイルアクセス態様情報に基づいて前記再配置対象データブロックのアクセス態様を判断し、

前記再配置対象データブロックのアクセス態様と前記ブロック情報とに基づいて、前記再配置対象データブロックを配置する前記階層化ディスクにおける記憶位置の再配置の内容を決定して前記ストレージ装置に指示し、

前記ストレージ装置は、前記情報処理装置から指示された前記再配置の内容に従って、前記再配置対象データブロックを配置する前記階層化ディスクにおける記憶位置の再配置を行なう、ストレージ制御方法。

10

【請求項 9】

特性の異なる複数のディスクを有する階層化ディスクを含むストレージ装置と、前記ストレージ装置にアクセスする情報処理装置とが接続されたストレージシステムにおける前記ストレージ装置に含まれるコンピュータに、

前記階層化ディスクに保存された各ファイルに含まれるデータブロックの記憶位置を含むブロック情報と、前記データブロックに対するアクセス頻度を含むアクセス情報とを、前記ストレージ装置から前記情報処理装置に通知し、

前記情報処理装置が、前記階層化ディスクに対する入出力要求を受けると前記入出力要求の対象ファイルと前記対象ファイルのアクセス態様とを前記ファイル毎のアクセス態様情報として特定し、特定された前記ファイル毎のアクセス態様情報と、前記ストレージ装置から通知された前記ブロック情報および前記アクセス情報とを対応付けてファイルアクセス態様情報として管理し、前記ファイルアクセス態様情報に基づいて再配置対象データブロックを判断し、前記ファイルアクセス態様情報に基づいて前記再配置対象データブロックのアクセス態様を判断し、前記再配置対象データブロックのアクセス態様と前記ブロック情報とに基づいて、前記再配置対象データブロックを配置する前記階層化ディスクにおける記憶位置の再配置の内容を決定して前記ストレージ装置に指示することによって、前記情報処理装置から指示された、前記再配置の内容に従って、前記再配置対象データブロックを配置する前記階層化ディスクにおける記憶位置の再配置を行なう、  
処理を実行させる、ストレージ制御プログラム。

20

30

【請求項 10】

特性の異なる複数のディスクを有する階層化ディスクを含むストレージ装置と、前記ストレージ装置にアクセスする情報処理装置とが接続されたストレージシステムにおける前記情報処理装置に含まれるコンピュータに、

前記階層化ディスクに対する入出力要求を受けると前記入出力要求の対象ファイルと前記対象ファイルのアクセス態様とを前記ファイル毎のアクセス態様情報として特定し、

特定された前記ファイル毎のアクセス態様情報と、前記ストレージ装置から通知された、前記階層化ディスクに保存された各ファイルに含まれるデータブロックの記憶位置を含むブロック情報、および、前記ストレージ装置から通知された、前記データブロックに対するアクセス頻度を含むアクセス情報に含まれる情報と、を対応付けてファイルアクセス態様情報として管理し、

40

前記ファイルアクセス態様情報に基づいて再配置対象データブロックを判断し、

前記ファイルアクセス態様情報に基づいて前記再配置対象データブロックのアクセス態様を判断し、

前記再配置対象データブロックのアクセス態様と前記ブロック情報とに基づいて、前記再配置対象データブロックを配置する前記階層化ディスクにおける位置情報の再配置を行なう、

処理を実行させる、ストレージ制御プログラム。

【発明の詳細な説明】

【技術分野】

50

## 【 0 0 0 1 】

本発明は、ストレージ制御方法、ストレージシステム、およびストレージ制御プログラムに関する。

## 【 背景技術 】

## 【 0 0 0 2 】

ディスクアレイシステム等のストレージシステムで使用されるディスクドライブとしては、例えば、SSD (Solid State Drive) , SAS (Serial Attached SCSI (Small Computer System Interface)) ディスク, SATA (Serial Advanced Technology Attachment) ディスクの3種類が挙げられる。図9に示すように、これら3種類のうち、最も信頼性および性能が高いのはSSDであり、2番目に信頼性および性能が高いのはSASディスクであり、3番目に信頼性および性能が高いのはSATAディスクである。

10

## 【 0 0 0 3 】

このように異なる特性をもつ3種類のディスクを有する階層化ディスク (図9参照) を含むストレージ装置を、サーバが用いる場合、以下のようなストレージ階層化制御が行なわれる。つまり、階層化ディスクを用いる構成では、ディスク側において、アクセス頻度に基づき、再配置すべきデータブロックと、そのデータブロックの再配置先のディスクとが決定され、図9に示すように、当該データブロックの再配置が行なわれる。

## 【 0 0 0 4 】

例えば、SSD上のファイルに含まれるデータブロックに対するアクセス頻度が低い場合、当該データブロックは、SSDからSASディスクへ移動されSASディスク上で再配置される。同様に、SASディスク上のデータブロックに対するアクセス頻度が低い場合、当該データブロックは、SASディスクからSATAディスクへ移動されSATAディスク上で再配置される。逆に、SATAディスク上のデータブロックに対するアクセス頻度が高くなった場合、当該データブロックは、SATAディスクからSASディスクへ移動されSASディスク上で再配置される。同様に、SASディスク上のデータブロックに対するアクセス頻度が高くなった場合、当該データブロックは、SASディスクからSSDへ移動されSSD上で再配置される。

20

## 【 先行技術文献 】

## 【 特許文献 】

## 【 0 0 0 5 】

30

【 特許文献 1 】 特開 2 0 0 6 - 4 0 1 1 号 公 報

## 【 発明の概要 】

## 【 発明が解決しようとする課題 】

## 【 0 0 0 6 】

ところで、SSD上におけるシーケンシャルアクセスのファイルに含まれるデータブロックの一部をSASディスク上で再配置した場合、当該ファイルを構成するデータブロックの配置状態は、シーケンシャルでなくなる。このため、SASディスク上のデータブロックに対するシーケンシャルアクセスの性能は、SSD上のデータブロックに対するアクセス性能と比べて劣化する。そして、当該ファイルの再配置の頻度が高くなると、当該ファイルのデータはより断片化し、SASディスク上のデータブロックに対するシーケンシャルアクセスの性能は、より劣化する。

40

## 【 0 0 0 7 】

データの断片化を抑止すべく、ファイルに含まれるデータブロックに対するアクセス態様がシーケンシャルアクセスであるか否かを判別してデータブロックの再配置を行なうためには、階層化ディスクにアクセスするアプリケーション側の仕様に係るアクセス(Read/Write)層での情報が必要である。つまり、階層化ディスクを用いるサーバは、階層化ディスクが実際にどのようなアクセス態様でデータ配置を行なっているかに関する情報を有していない。このため、現状の階層化ディスクでは、シーケンシャルアクセスにおける性能劣化や、再配置によるデータの断片化等の悪影響が生じる。

## 【 0 0 0 8 】

50

一つの側面で、本発明は、階層化ディスクに対するデータブロックの再配置によるデータの断片化を抑止することを目的とする。

なお、前記目的に限らず、後述する発明を実施するための最良の形態に示す各構成により導かれる作用効果であって、従来の技術によっては得られない作用効果を奏することも本件の他の目的の一つとして位置付けることができる。

【課題を解決するための手段】

【0009】

本件のストレージシステムは、特性の異なる複数のディスクを有する階層化ディスクを含むストレージ装置と、前記ストレージ装置にアクセスする情報処理装置とが接続されたストレージシステムであって、前記ストレージ装置は、前記階層化ディスクに保存された各ファイルに含まれるデータブロックの記憶位置を含むブロック情報と、前記データブロックに対するアクセス頻度を含むアクセス情報とを、前記情報処理装置に通知するディスク制御部を有し、前記情報処理装置は、前記階層化ディスクに対する入出力要求を受けると前記入出力要求の対象ファイルと前記対象ファイルのアクセス態様とを前記ファイル毎のアクセス態様情報として特定するアクセス態様特定部と、前記アクセス態様特定部によって特定された前記ファイル毎のアクセス態様情報と、前記ディスク制御部から通知された前記ブロック情報および前記アクセス情報に含まれる情報とを対応付けてファイルアクセス態様情報として管理し、前記ファイルアクセス態様情報に基づいて再配置対象データブロックを判断し、前記ファイルアクセス態様情報に基づいて前記再配置対象データブロックのアクセス態様を判断し、前記再配置対象データブロックのアクセス態様と前記ブロック情報とに基づいて、前記再配置対象データブロックを配置する前記階層化ディスクにおける記憶位置の再配置の内容を決定して前記ストレージ装置の前記ディスク制御部に指示する管理部と、を有し、前記ストレージ装置の前記ディスク制御部は、前記情報処理装置の前記管理部から指示された前記再配置の内容に従って、前記再配置対象データブロックを配置する前記階層化ディスクにおける記憶位置の再配置を行なう。

【発明の効果】

【0010】

一実施形態によれば、階層化ディスクに対するデータブロックの再配置によるデータの断片化が抑止される。

【図面の簡単な説明】

【0011】

【図1】本実施形態のストレージシステムのハードウェア構成および機能構成を示すブロック図である。

【図2】図1に示すストレージシステムの機能や動作を概略的に説明する図である。

【図3】本実施形態のファイルアクセス態様テーブルの例を示す図である。

【図4】図1に示すストレージシステムの動作を説明するフローチャートである。

【図5】本実施形態における対象ファイルのアクセス態様の判断手順を説明するフローチャートである。

【図6】本実施形態におけるSSDからSASディスクへの再配置手順を説明するフローチャートである。

【図7】図1に示すストレージシステムにおけるデータブロックの具体的な再配置例を示す図である。

【図8】図1に示すストレージシステムにおけるデータブロックの具体的な再配置例を示す図である。

【図9】階層化ディスクおよびストレージ階層化制御を説明する図である。

【発明を実施するための形態】

【0012】

以下、図面を参照して実施の形態を説明する。

〔1〕本実施形態のストレージシステムの構成

図1は、本実施形態のストレージシステム1のハードウェア構成および機能構成を示す

10

20

30

40

50

ブロック図である。図 1 に示すように、ストレージシステム 1 は、ストレージ装置 1 0 およびサーバ ( 情報処理装置 ) 2 0 を有している。

【 0 0 1 3 】

ストレージ装置 1 0 は、サーバ 2 0 ( アプリケーション 2 1 1 ) によってアクセスされるとともにサーバ 2 0 ( 管理部 2 1 3 ) によって管理され、階層化ディスク 1 1 , C P U ( Central Processing Unit ) 1 2 およびメモリ 1 3 を有している。なお、図 1 では、ストレージ装置 1 0 において論理ユニット 1 1 0 が図示されている。論理ユニット 1 1 0 は、L U N ( Logical Unit Number ) によって特定されるもので、後述するアプリケーション 2 1 1 から見える一つの論理ディスクである。

【 0 0 1 4 】

階層化ディスク 1 1 は、特性の異なる複数のディスクを有する。本実施形態の階層化ディスク 1 1 では、図 1 に示すように、2 種類のディスク、つまり S S D 1 1 1 および S A S ディスク 1 1 2 が備えられている。なお、本発明は、これに限定されるものでなく、例えば図 9 を参照しながら前述したように、階層化ディスク 1 1 が 3 種類のディスク、つまり S S D , S A S ディスクおよび S A T A ディスクを有する場合にも本実施形態と同様に適用される。

【 0 0 1 5 】

ここで、S S D は、ハードディスクのようにディスクをもたないので、読取装置 ( ヘッド ) をディスク上で移動させる時間 ( シークタイム ) や、目的のデータがヘッド位置まで回転してくるまでの待ち時間 ( サーチタイム ) がない。このため、S S D では、シーケン

【 0 0 1 6 】

これに対し、S A S ディスクや S A T A ディスクでは、シーケンシャルアクセスのファイルに含まれるデータブロックが、断片化してランダムに配置されていると、連続的に配置されている場合に比べ、シークタイムやサーチタイムを要し、アクセス性能が低下する。

【 0 0 1 7 】

そこで、本実施形態では、少なくともシーケンシャルアクセスのデータブロックを S A S ディスクや S A T A ディスク上で再配置する際に、再配置対象のデータブロックのアクセス態様がシーケンシャルアクセスであるか否かを考慮する。

【 0 0 1 8 】

このため、本実施形態では、S S D 1 1 1 から S A S ディスク 1 1 2 への再配置を行なう場合、再配置対象のデータブロックのアクセス態様が判断され、判断されたアクセス態様に基づく再配置が実行される。なお、階層化ディスク 1 1 が上述した 3 種類のディスクを有する場合、S S D から S A S ディスクへの再配置 , S A S ディスクから S A T A ディスクへの再配置 , S A T A ディスクから S A S ディスクへの再配置を行なう際に本実施形態のごとくアクセス態様に基づく再配置が実行される。また、S A S ディスクから S S D への再配置を行なう場合、再配置対象のデータブロックのアクセス態様に基づく再配置が実行されてもよいし実行されなくてもよい。

【 0 0 1 9 】

C P U ( 処理部 , コンピュータ ) 1 2 は、メモリ 1 3 に格納されたプログラムを読み出して実行することで、後述するディスク制御部 1 2 1 としての機能を果たす。なお、ディスク制御部 1 2 1 は、階層化ディスク 1 1 の制御を行なうもので、図 9 を参照しながら前述したストレージ階層化制御の機能も果たす。

【 0 0 2 0 】

メモリ 1 3 は、R A M ( Random Access Memory ) 等であり、C P U 1 2 が実行するプログラムを保存するほか、ストレージ装置 1 0 において C P U 1 2 が処理 ( ディスク制御部 1 2 1 による制御処理等 ) を行なう際に必要となる各種データ等を保存する。

【 0 0 2 1 】

サーバ 20 は、F C (Fibre Channel) / i S C S I を介しストレージ装置 10 に接続され、ストレージ装置 10 の階層化ディスク 11 にアクセスする。また、サーバ 20 は、L A N (Local Area Network) を介しストレージ装置 10 のディスク制御部 121 (C P U 12) に接続され、ストレージ装置 10 の管理を行なう。サーバ 20 は、C P U 21 およびメモリ 22 を有している。

【0022】

C P U 21 は、メモリ 22 に格納されたアプリケーション (アプリケーションプログラム) 211 を読み出して実行することで、ストレージ装置 10 の階層化ディスク 11 に対し R / W (Read/Write) 命令 (入出力要求) を発行する。なお、前述したように、アプリケーション 211 は、階層化ディスク 11 を論理ディスク 110 として認識し、論理ディスク 110 に対し R / W 命令を発行する。

10

【0023】

また、C P U 21 は、メモリ 22 に格納されたプログラムを読み出して実行することで、アプリケーション 211 とストレージ装置 10 との間に、後述するアクセス態様特定部として機能する仮想ドライバ (仮想デバイス) 212 を構築する。

【0024】

さらに、C P U 21 は、メモリ 22 に格納されたプログラムを読み出して実行することで、後述するファイルアクセス態様テーブル 221 を作成し、当該テーブル 221 を用いて、後述する管理部 213 としての機能を果たす。

【0025】

20

メモリ 22 は、R A M 等であり、C P U 21 が実行するプログラムや、テーブル 221 を保存するほか、サーバ 20 において C P U 21 が処理 (アプリケーション 211, 仮想ドライバ 212, 管理部 213 による処理等) を行なう際に必要となる各種データ等を保存する。

【0026】

なお、図 1 に示すストレージシステム 1 では、一つのサーバ 20 が、アプリケーション 211, 仮想ドライバ 212, 管理部 213 およびテーブル 221 を備えているが、二つのサーバが、アプリケーション 211, 仮想ドライバ 212, 管理部 213 およびテーブル 221 を分けて備えてもよい。例えば、ストレージ装置 10 にアクセスする業務サーバが、アプリケーション 211 および仮想ドライバ 212 を備え、管理サーバが、L A N を介して業務サーバおよびストレージ装置 10 に接続され、管理部 213 およびテーブル 221 を備えてもよい。

30

【0027】

以下、本実施形態のストレージシステム 1 におけるディスク制御部 121, 仮想ドライバ 212 および管理部 213 の具体的な機能について、図 2 を参照しながら説明する。併せて、サーバ 20 のメモリ 22 に保存されるファイルアクセス態様テーブル 221 の具体例について、図 3 を参照しながら説明する。なお、図 2 は、ストレージシステム 1 の機能や動作を概略的に説明する図、図 3 は、ファイルアクセス態様テーブル 221 の例を示す図である。

【0028】

40

仮想ドライバ (仮想デバイス) 212 は、アプリケーション 211 から、ストレージ装置 10 に対する R / W 命令 (入出力要求) を受けると、ファイル毎のアクセス態様情報を特定するアクセス態様特定部として機能する。アクセス態様情報は、ストレージ装置 10 を利用するアプリケーション 211 を特定する情報 (アプリケーション名等) と、R / W 命令の対象ファイルを特定する情報 (ファイル名等) と、当該対象ファイルのアクセス態様とを含む。仮想ドライバ 212 は、取得・特定したアクセス態様情報を管理部 213 に通知する (図 2 の矢印 A1 参照)。

【0029】

ここで、仮想ドライバ 212 は、以下のようにして、対象ファイルのアクセス態様がシーケンシャルアクセスであるかランダムアクセスであるかを特定する。つまり、仮想デバ

50

イス 2 1 2 は、R / W 命令が対象ファイルの R / W 位置（入出力位置）をインデックスでシークしているか否かを測定する。仮想デバイス 2 1 2 は、R / W 命令がインデックスでシークしていない場合、即ち対象ファイルの先頭から読み込んでいる場合、対象ファイルのアクセス態様をシーケンシャルアクセスと判断する。一方、仮想デバイス 2 1 2 は、R / W 命令がインデックスでシークしている場合に対象ファイルのアクセス態様をランダムアクセスと判断する。

#### 【 0 0 3 0 】

なお、仮想ドライバ 2 1 2 は、アプリケーション 2 1 1 が用いるファイルのうち、テーブル 2 2 1 に登録されていない対象ファイルについてアクセス態様を特定する。また、仮想ドライバ 2 1 2 は、アプリケーション 2 1 1 によるアクセス開始後、所定時間の間、アクセス態様の特定を行なう。

10

#### 【 0 0 3 1 】

ディスク制御部 1 2 1 は、階層化ディスク 1 1 に保存されている各ファイルに含まれるデータブロックの記憶位置（物理位置）を含むブロック情報や、各データブロックに対するアクセス情報を取得し（図 2 の矢印 A 2 参照）、取得したブロック情報やアクセス情報をサーバ 2 0 の管理部 2 1 3 に通知する（図 2 の矢印 A 3 参照）。ブロック情報には、アプリケーション 2 1 1 を特定する情報（アプリケーション名等）と、アプリケーション 2 1 1 が用いる対象ファイルを特定する情報（ファイル名等）と、当該対象ファイルに含まれるデータブロックが実際に存在する物理位置（つまりサーバ 2 0 上のアプリケーション 2 1 1 が認識している論理ディスク 1 1 0 が存在する物理ストレージ）に関する情報とが含まれる。また、アクセス情報は、アプリケーション 2 1 1 による、各データブロックに対するアクセス頻度を含む。

20

#### 【 0 0 3 2 】

管理部 2 1 3 は、仮想デバイス 2 1 2 から通知されたアクセス態様情報と、ディスク制御部 1 2 1 から通知されたブロック情報やアクセス情報とを、メモリ 2 2 上のファイルアクセス態様テーブル 2 2 1 に登録する（図 2 の矢印 A 4 参照）。

#### 【 0 0 3 3 】

ファイルアクセス態様テーブル 2 2 1 には、図 3 に示すように、仮想デバイス 2 1 2 から通知されたアクセス態様情報に基づき、アプリケーション 2 1 1 のアプリケーション名と、当該アプリケーションが用いる対象ファイルのファイル名と、仮想ドライバ 2 1 2 によって特定された対象ファイルのアクセス態様（シーケンシャルアクセスかランダムアクセスか）とが登録される。また、ファイルアクセス態様テーブル 2 2 1 には、図 3 に示すように、ディスク制御部 1 2 1 から通知されたブロック情報やアクセス情報が、アプリケーション名やファイル名によってアクセス態様情報に対応付けられて登録される。即ち、対象ファイルに含まれる各データブロックの記憶位置と、各データブロックに対するアクセス頻度とが登録される。図 3 では、一つのデータブロックの記憶位置およびアクセス頻度が示されているが、実際には、データブロック毎に記憶位置およびアクセス頻度が登録される。なお、図 3 において、アクセス頻度の単位は、1 秒間当たりの読み込み / 書き込み回数を示す I O P S（Input/Output Per Second）である。

30

#### 【 0 0 3 4 】

そして、管理部 2 1 3 は、テーブル 2 2 1 を参照し（図 2 の矢印 A 5 参照）、再配置すべきデータブロックを判断する。このとき、管理部 2 1 3 は、SSD 1 1 1 に配置され且つアクセス頻度の低いデータブロックを、SAS ディスク 1 1 2 に再配置すべきデータブロックであると判断する。一方、管理部 2 1 3 は、SAS ディスク 1 1 2 に配置され且つアクセス頻度の高いデータブロックを、SSD 1 1 1 に再配置すべきデータブロックであると判断する。また、管理部 2 1 3 は、テーブル 2 2 1 を参照し（図 2 の矢印 A 5 参照）、再配置対象のデータブロックのアクセス態様（シーケンシャルアクセスかランダムアクセスか）を判断する。つまり、管理部 2 1 3 は、アプリケーション 2 1 1 が用いるファイル毎のアクセス態様情報、および、ディスク制御部 1 2 1 から通知された各ファイルのブロック情報に基づき、データブロックのアクセス態様を判断する。

40

50



## 【 0 0 3 5 】

また、管理部 2 1 3 は、アクセス頻度に基づき判断された再配置対象のデータブロックについてのブロック情報（記憶位置）と、管理部 2 1 3 によって判断された再配置対象のデータブロックのアクセス態様とに基づき、後述するごとく、階層化ディスク 1 1 に対する再配置の内容を決定する。管理部 2 1 3 は、決定された再配置の内容に従ったデータブロックの再配置を行なうように、ストレージ装置 1 0 のディスク制御部 1 2 1 に対し指示する（図 2 の矢印 A 6 参照）。

## 【 0 0 3 6 】

ストレージ装置 1 0 のディスク制御部 1 2 1 は、管理部 2 1 3 から指示されたデータブロックの再配置を階層化ディスク 1 1 に対し行なう。

10

## 【 0 0 3 7 】

このとき、本実施形態のストレージシステム 1 では、SSD 1 1 1 から SAS ディスク 1 1 2 へデータブロックを再配置する場合、項目(A1)～(A4)を参照しながら後述するように、管理部 2 1 3 によって再配置の内容が決定され、ディスク制御部 1 2 1 によって再配置対象のデータブロックの再配置が実行される。なお、管理部 2 1 3 による再配置の内容の具体的な決定手法等については後述する。また、ディスク制御部 1 2 1 による具体的な再配置例については、図 7 および図 8 を参照しながら後述する。

## 【 0 0 3 8 】

## 〔 2 〕本実施形態のストレージシステムの動作

ストレージ装置 1 0 においては、ディスク制御部 1 2 1 のストレージ階層化制御の機能によって、データブロック単位でデータ領域へのアクセス情報（アクセス頻度）が測定される。ディスク制御部 1 2 1 のストレージ階層化制御の機能によってデータブロック単位で測定されたデータでは、各データブロック内のデータがどのようなアクセス態様（シーケンシャルアクセスかランダムアクセスか）をもつかは不明であった。本実施形態では、ディスク制御部 1 2 1 によって得られるブロック情報と、サーバ 2 0 上で仮想ドライバ 2 1 2 によって得られるアクセス態様情報とを対応付けることによって、各ファイルを構成するデータブロックのアクセス態様の判断が可能になる。

20

## 【 0 0 3 9 】

そして、シーケンシャルアクセスと判断されたデータブロックは、ファイルを構成するデータブロックの連続性が保たれるように再配置される。シーケンシャルアクセスのファイルを構成するデータブロックが SSD 1 1 1 および SAS ディスク 1 1 2 の両方に分かれて存在する場合、連続的な再配置を容易に行なえるようデータブロックの前後に空き領域が設けられる。しかし、SAS ディスク 1 1 2 に再配置されたデータブロックの前後を単純に空き領域とした場合、SAS ディスク 1 1 2 の利用効率が低下する。そこで、管理部 2 1 3 は、前後のデータブロックが再配置されるか否かを再配置の頻度から予測し、再配置される可能性が高い場合に、再配置されたデータブロックの前後を空き領域とするようにしてもよい。

30

## 【 0 0 4 0 】

以下、本実施形態のストレージシステム 1 の動作について、図 4 ～図 8 を参照しながら説明する。

40

## 【 0 0 4 1 】

## 〔 2 - 1 〕ストレージシステムの動作

図 4 に示すフローチャート（ステップ S 1 ～ S 9 ）に従って、ストレージシステム 1 の動作について説明する。

## 【 0 0 4 2 】

上述したように、ディスク制御部 1 2 1 のストレージ階層化制御機能では、各データブロックに対するアクセス頻度を測定することは可能であるが、各データブロックのアクセス態様を判断することはできなかった。そこで、本実施形態のストレージシステム 1 においては、サーバ 2 0 上に仮想ドライバ 2 1 2 が構築され、仮想ドライバ 2 1 2 とディスク制御部 1 2 1 のストレージ階層化制御機能とを管理部 2 1 3 により連携させることで、各

50

データブロックのアクセス態様の判断が可能になっている。

【0043】

サーバ20において、アプリケーション211が対象ファイルへR/W命令を発行し、仮想ドライバ212がR/W命令を受けると(ステップS1のYESルート)、仮想ドライバ212は、管理部213を介してメモリ22のファイルアクセス態様テーブル221を参照する。そして、仮想ドライバ212は、今回R/W命令を発行したアプリケーション211のアプリケーション名、および、当該アプリケーション211によってアクセスされる対象ファイルのファイル名が、ファイルアクセス態様テーブル221に登録されているか否かを判断する(ステップS2)。

【0044】

アプリケーション名およびファイル名の両方が登録されている場合(ステップS2のYESルート)、ストレージシステム1は、ステップS3、S4の処理をスキップし、ステップS5の処理に移行する。一方、アプリケーション名およびファイル名の少なくとも一方が登録されていない場合(ステップS2のNORルート)、仮想ドライバ212は、アプリケーション211のアプリケーション名と、対象ファイルのファイル名と、当該対象ファイルのアクセス態様とを測定・取得する(ステップS3)。ステップS3におけるアクセス態様の判断手順については、図5を参照しながら後述する。

【0045】

そして、仮想ドライバ212は、取得したアプリケーション名、ファイル名およびアクセス態様を、管理部213に通知し(図2の矢印A1参照)、管理部213経由でファイルアクセス態様テーブル221に登録する(ステップS4;図2の矢印A4参照)。なお、テーブル221において、アプリケーション名が登録されているが、ファイル名が登録されていない場合、仮想ドライバ212は、ファイル名とアクセス態様とを管理部213に通知する。

【0046】

この後、ステップS5において、ディスク制御部121は、アプリケーション211がアクセスしている対象ファイルのデータブロックの記憶位置(物理位置)を含むブロック情報や、当該データブロックに対するアクセス情報(アクセス情報)を取得する(ステップS5;図2の矢印A2参照)。

【0047】

ディスク制御部121は、取得したブロック情報やアクセス情報をサーバ20の管理部213に通知し(図2の矢印A3参照)、管理部213経由でファイルアクセス態様テーブル221に登録する(ステップS6;図2の矢印A4参照)。このとき、前述した通り、ディスク制御部121から通知された記憶位置やアクセス頻度が、アプリケーション名やファイル名によってアクセス態様情報に対応付けられて登録される。メモリ22に保存されるファイルアクセス態様テーブル221は、アクセス頻度等の評価を繰り返すことで更新され、管理部213は、以下のように、更新されたテーブル221に基づき、ディスク制御部に対し、データブロックの再配置の指示を行なう。

【0048】

つまり、管理部213は、ファイルアクセス態様テーブル221を参照し(図2の矢印A5参照)、再配置すべきデータブロックを判断する(ステップS7)。このとき、管理部213は、SSD111に配置され且つアクセス頻度の低いデータブロックを、SASディスク112に再配置すべきデータブロックであると判断する。一方、管理部213は、SASディスク112に配置され且つアクセス頻度の高いデータブロックを、SSD111に再配置すべきデータブロックであると判断する。また、管理部213は、テーブル221を参照し(図2の矢印A5参照)、再配置対象のデータブロックのアクセス態様(シーケンシャルアクセスかランダムアクセスか)を判断する。

【0049】

また、管理部213は、アクセス頻度に基づき判断された再配置対象のデータブロックについてのブロック情報(記憶位置)と、管理部213によって判断された再配置対象の

10

20

30

40

50

データブロックのアクセス態様とに基づき、階層化ディスク 11 に対する再配置の内容を決定する。つまり、管理部 213 は、テーブル 221 を参照して再配置対象のデータブロックの記憶位置を把握する。記憶位置を把握した結果、SAS ディスク 112 に配置すべきデータブロックが SSD 111 上に配置されている場合、管理部 213 は、SAS ディスク 112 に再配置対象のデータブロックを再配置するよう再配置の内容を決定する。また、記憶位置を把握した結果、SSD 111 に配置すべきデータブロックが SAS ディスク 112 上に配置されている場合、管理部 213 は、SSD 111 に再配置対象のデータブロックを再配置するよう再配置の内容を決定する。一方、記憶位置を把握した結果、SSD 111 に再配置すべきデータブロックが既に SSD 111 に配置されている場合、管理部 213 は、再配置対象のデータブロックについては現状の配置状態を維持する。同様に、記憶位置を把握した結果、SAS ディスク 112 に再配置すべきデータブロックが既に SAS ディスク 112 に配置されている場合、管理部 213 は、再配置対象のデータブロックについては現状の配置状態を維持する。

10

#### 【0050】

管理部 213 は、上述のごとく決定された再配置の内容に従ったデータブロックの再配置を行なうように、ストレージ装置 10 のディスク制御部 121 に対し指示する（ステップ S8；図 2 の矢印 A6 参照）。

#### 【0051】

そして、ストレージ装置 10 のディスク制御部 121 は、管理部 213 から指示されたデータブロックの再配置の内容に従って、階層化ディスク 11 におけるデータブロックの再配置を行なう（ステップ S9）。

20

#### 【0052】

##### 〔2-2〕アクセス態様の判断手順

図 2 のステップ S3 において、仮想ドライバ 212 は、以下のように、図 5 に示す手順で対象ファイルのアクセス態様を判断する。図 5 は、本実施形態における対象ファイルのアクセス態様（アクセス特性）の判断手順を説明するフローチャート（ステップ S11～S13）である。

#### 【0053】

つまり、仮想デバイス 212 は、R/W 命令が対象ファイルの R/W 位置をインデックスでシークしているか否かを測定する（ステップ S11）。そして、仮想デバイス 212 は、R/W 命令がインデックスでシークしていない場合、即ち対象ファイルの先頭から読み込んでいる場合（ステップ S11 の NO ルート）、対象ファイルのアクセス態様をシーケンシャルアクセスと判断する（ステップ S12）。一方、仮想デバイス 212 は、R/W 命令がインデックスでシークしている場合（ステップ S11 の YES ルート）、対象ファイルのアクセス態様をランダムアクセスと判断する（ステップ S13）。

30

#### 【0054】

##### 〔2-3〕再配置手順

次に、図 6 に示すフローチャート（ステップ S21～S26）に従って、本実施形態における SSD 111 から SAS ディスク 112 への再配置手順（図 4 のステップ S7～S9 の処理）について、より具体的に説明する。

40

#### 【0055】

まず、管理部 213 は、ファイルアクセス態様テーブル 221 を参照し、アプリケーション 211 がアクセスした対象ファイルのデータブロックの I/O 特性（アクセス情報 / アクセス頻度）を評価する（ステップ S21）。これにより、管理部 213 は、SSD 111 に配置され且つアクセス頻度の低いデータブロックを、SAS ディスク 112 に再配置すべきデータブロックであると判断する。

#### 【0056】

また、管理部 213 は、ファイルアクセス態様テーブル 221 を参照し、再配置対象のデータブロックのアクセス態様がシーケンシャルアクセスであるかランダムアクセスであるかについて判断する（ステップ S22）。

50

## 【 0 0 5 7 】

さらに、管理部 2 1 3 は、再配置対象のデータブロックに関する情報（データブロックの記憶位置）とデータブロックのアクセス態様とに基づき、図 4 のステップ S 8 と同様にして、階層化ディスク 1 1 に対する再配置の内容を決定する。そして、管理部 2 1 3 は、決定した再配置の内容に従って、再配置対象のデータブロックを S S D 1 1 1 から S A S ディスク 1 1 2 へ移動（再配置）する際のデータ配置を、マッピング情報としてメモリ 2 2 に保持させる（ステップ S 2 3 ）。

## 【 0 0 5 8 】

この後、管理部 2 1 3 は、予め定めた評価期間が経過したか否かを判断し（ステップ S 2 4 ）、評価期間が経過していない場合（ステップ S 2 4 の N O ルート）、管理部 2 1 3 は、上述したステップ S 2 1 ～ S 2 4 の処理を繰り返し実行する。その際、管理部 2 1 3 は、図 4 のステップ S 1 ～ S 4 の処理によって更新されたファイルアクセス態様テーブル 2 2 1 を参照する。

## 【 0 0 5 9 】

一方、評価期間が経過した場合（ステップ S 2 4 の Y E S ルート）、管理部 2 1 3 は、メモリ 2 2 に保持されているマッピング情報に基づき、ディスク制御部 1 2 1 に対しデータブロックの再配置を指示する（ステップ S 2 5 ）。そして、ディスク制御部 1 2 1 は、管理部 2 1 3 から指示された再配置の内容に従って、再配置対象のデータブロックを、S S D 1 1 1 から S A S ディスク 1 1 2 へ移行（再配置）する（ステップ S 2 6 ）。

## 【 0 0 6 0 】

なお、管理部 2 1 3 は、S A S ディスク 1 1 2 に配置され且つアクセス頻度の高いデータブロックを、S S D 1 1 1 に再配置すべきデータブロックであると判断する。上述した通り、S S D 1 1 1 では、シーケンシャルアクセスのファイルに含まれるデータブロックが連続的に配置されていても断片化してランダムに配置されていても、アクセス性能に差異はない。このため、S A S ディスク 1 1 2 から S S D 1 1 1 への再配置を行なう場合、管理部 2 1 3 は、再配置対象のデータブロックのアクセス態様を判断することなく、再配置対象のデータブロックを、S S D 1 1 1 の適当な空き領域に再配置するよう、ディスク制御部 1 2 1 に指示する。ディスク制御部 1 2 1 は、このような管理部 2 1 3 からの指示に従い、S A S ディスク 1 1 2 上の再配置対象のデータブロックを S S D 1 1 1 の適当な空き領域に再配置する。

## 【 0 0 6 1 】

## 〔 2 - 4 〕 データブロックの再配置内容の決定手法

次に、管理部 2 1 3 による、図 4 のステップ S 8 や図 6 のステップ S 2 3 における再配置内容の決定手法について説明する。ここでは、特に、シーケンシャルアクセスと判断されたデータブロックを、連続性を保ちながら再配置するための再配置内容の決定手法が説明される。なお、以下では、アクセス態様がシーケンシャルアクセスであるファイルをシーケンシャルファイルといい、アクセス態様がランダムアクセスであるファイルをランダムファイルという。

## 【 0 0 6 2 】

まず、本実施形態のストレージシステム 1 における再配置内容の決定手法を概略的に説明すると、以下の通りである。シーケンシャルファイルを構成するデータブロックの再配置に際し、管理部 2 1 3 は、仮想ドライバ 2 1 2 で取得した情報に基づき、データブロックがシーケンシャルな並びになるように再配置内容を決める。ただし、ファイルを構成する一部のデータブロックが移動し、ファイル全体が連続的に再配置されない場合、管理部 2 1 3 は、連続的に再配置された一部のデータブロックの前後に空き領域を配置する。これにより、その後、再配置が実行された際に、管理部 2 1 3 は、空き領域を用いて前後のデータブロックを連続的に配置することが可能になる。一方、ランダムファイルを構成するデータブロックの再配置に際し、管理部 2 1 3 は、シーケンシャルファイルのデータブロックの領域や、前記空き領域を避けた領域に、ランダムファイルのデータブロックを再配置するように再配置内容を決める。また、管理部 2 1 3 は、シーケンシャルファイ

10

20

30

40

50

ルの再配置に際し部分的データブロックの前後に設けた空き領域を、以下のような場合に解放してもよい。つまり、管理部 2 1 3 は、ランダムファイルの再配置の契機でディスク制御部 1 2 1 のストレージ階層化制御機能によりアクセス頻度評価を再度実施する。そして、その評価結果が変わらない場合、管理部 2 1 3 は、空き領域に対しシーケンシャルファイルのデータブロックが再配置されないと判断して、空き領域を解放し、ランダムファイルのデータブロックの再配置を可能にする。なお、前述したように、再配置されたデータブロックの前後を単純に空き領域とした場合、ディスクの利用効率が低下する。そこで、管理部 2 1 3 は、前後のデータブロックが再配置されるか否かを再配置の頻度から予測し、再配置される可能性が高い場合に、再配置されたデータブロックの前後を空き領域とするようにしてもよい。

10

#### 【 0 0 6 3 】

本実施形態のストレージシステム 1 では、SSD 1 1 1 から SAS ディスク 1 1 2 ヘデータブロックを再配置する場合、以下の項目(A1)～(A4)のように、管理部 2 1 3 によって再配置の内容が決定され、ディスク制御部 1 2 1 によって再配置対象のデータブロックの再配置が実行される。

#### 【 0 0 6 4 】

(A1)再配置対象のデータブロックのアクセス態様がシーケンシャルアクセスである場合、且つ、データブロックが再配置される SAS ディスク 1 1 2 上に、連続する複数のデータブロックが存在する場合： 管理部 2 1 3 は、SAS ディスク 1 1 2 上に前記連続する複数のデータブロックを連続的に再配置するよう再配置内容を決定しディスク制御部 1 2 1 に対し指示する。これにより、ディスク制御部 1 2 1 は、ストレージ装置 1 0 の階層化ディスク 1 1 において、SAS ディスク 1 1 2 上に前記連続する複数のデータブロックを連続的に再配置する。

20

#### 【 0 0 6 5 】

(A2)再配置対象のデータブロックのアクセス態様がシーケンシャルアクセスである場合、且つ、データブロックが再配置される SAS ディスク 1 1 2 上に、当該データブロックの前側または後側に続く他のデータブロックが存在しない場合： 管理部 2 1 3 は、SAS ディスク 1 1 2 上に再配置された当該データブロックの前側または後側に、前記他のデータブロック用の空き領域を設けるよう再配置内容を決定しディスク制御部 1 2 1 に対し指示する。これにより、ディスク制御部 1 2 1 は、ストレージ装置 1 0 の階層化ディスク 1 1 において、SAS ディスク 1 1 2 上に再配置された当該データブロックの前側または後側に、他のデータブロック用の空き領域を設ける。

30

#### 【 0 0 6 6 】

(A3)空き領域を設けた後、所定期間が経過しても、または、データブロックの再配置処理を所定回数行なっても、空き領域に前記他のデータブロックが再配置されない場合：

管理部 2 1 3 は、当該空き領域を解放し、当該空き領域に前記他のデータブロック以外のデータブロックを再配置するよう再配置内容を決定しディスク制御部 1 2 1 に対し指示する。これにより、ディスク制御部 1 2 1 は、当該空き領域に前記他のデータブロック以外のデータブロックを再配置する。

#### 【 0 0 6 7 】

(A4)再配置対象のデータブロックのアクセス態様がランダムアクセスである場合： 管理部 2 1 3 は、当該データブロックを、アクセス態様がシーケンシャルアクセスであるデータブロックの配置状態に応じて再配置するよう再配置内容を決定しディスク制御部 1 2 1 に対し指示する。これにより、ディスク制御部 1 2 1 は、シーケンシャルアクセスのデータブロックの領域や、前記空き領域を考慮し、これらのシーケンシャルアクセスのデータブロックの領域や、前記空き領域を避けた領域において、ランダムアクセスのデータブロックを再配置する。

40

#### 【 0 0 6 8 】

次に、図 7 および図 8 を参照しながら、ディスク制御部 1 2 1 による具体的な再配置例について説明する。図 7 および図 8 は、ストレージシステム 1 におけるデータブロックの

50

具体的な再配置例を示す図である。図7および図8において、4個のブロックS1～S4は、シーケンシャルファイルAを構成する、シーケンシャルなデータブロックを示し、この順で連続的に配置されるべき4個のデータブロックS1～S4によって一つのシーケンシャルファイルAが構成されている。また、「空」を付されたブロックは、データブロック1個分の容量を有する空き領域を示している。

【0069】

図7では、SSD111上に配置されたシーケンシャルファイルに含まれる4個のデータブロックS1～S4と、SSD111上に配置されたランダムファイルに含まれる4個のデータブロックRとが、SASディスク112で再配置される場合が示されている。この場合、管理部213において各データブロックのアクセス態様がシーケンシャルアクセスかランダムアクセスかが判断された上で、SSD111からSASディスク112へのデータブロックのデフラグ再配置が実行される。つまり、シーケンシャルファイルに含まれる4個のデータブロックS1～S4は、SASディスク112上において、シーケンシャルな並びになるように再配置される。そして、ランダムファイルに含まれる4個のデータブロックRは、SASディスク112上において、データブロックS1～S4の配置領域以外の適当な空き領域に再配置される。

【0070】

図8の上段のSASディスク112では、図7の例と同様、SSD111上のファイルAおよびファイルBに含まれる全てのデータブロックが、移動対象（再配置対象）であり、SASディスク112で再配置される場合が示されている。この場合は、図7と同様の再配置が実行される。

【0071】

図8の下段のSASディスク112では、SSD111上のファイルAおよびファイルBに含まれるデータブロックの一部が、移動対象（再配置対象）であり、SASディスク112で再配置される場合が示されている。ここでは、シーケンシャルファイルAに含まれる2個のデータブロックS3、S4と、ランダムファイルBに含まれる2個のデータブロックRとが、移動対象（再配置対象）であり、SASディスク112で再配置される場合が示されている。この場合も、管理部213において各データブロックのアクセス態様がシーケンシャルアクセスかランダムアクセスかが判断された上で、SSD111からSASディスク112へのデータブロックのデフラグ再配置が実行される。つまり、シーケンシャルファイルに含まれる2個のデータブロックS3、S4は、SASディスク112上において、シーケンシャルな並びになるように再配置される。さらに、データブロックS3、S4の前側には、以降の再配置処理で連続するデータブロックS1、S2が再配置される可能性があるので、データブロック2個分の空き領域が設けられる。そして、ランダムファイルに含まれる2個のデータブロックRは、SASディスク112上において、データブロックS3、S4の配置領域と、データブロックS3、S4の前側に設けたデータブロック2個分の空き領域とを除く、適当な空き領域に再配置される。

【0072】

〔3〕本実施形態のストレージシステム1によれば、以下のような作用効果を得ることができる。

【0073】

ディスク制御部121によって得られるブロック情報と、サーバ20上で仮想ドライバ212によって得られるアクセス態様情報とを対応付けることによって、各ファイルを構成するデータブロックのアクセス態様の判断が可能になる。これにより、データブロックのアクセス態様の判断結果に基づきデータブロックの再配置を行なうことができるので、SSD111からSASディスク112へ移動したデータについて、シーケンシャルアクセスにおける性能劣化を回避することができる。また、階層化ディスクに対するデータブロックの再配置によるデータの断片化が抑止され、断片化等の悪影響を回避することができる。

【0074】

また、アプリケーション 2 1 1 が用いる対象ファイルごとのアクセス態様が、仮想ドライバ 2 1 2 によって判断される。これにより、OS (Operating System) に近い部分で、ファイルごとに、アクセス態様が、ランダムアクセスであるのか、シーケンシャルアクセスであるのかを確実に判別することができる。ただし、アクセス態様を判断するために仮想ドライバ 2 1 2 を単純に用いると、アクセス速度の低下を招く虞がある。そこで、仮想ドライバ 2 1 2 は、アプリケーション 2 1 1 によるアクセス開始後、所定時間の間だけ、アクセス態様の測定を行ない、所定時間経過後はアクセス態様の測定を行なわないようにする。これにより、アクセス速度の低下を回避することができる。

【0075】

さらに、仮想ドライバ 2 1 2 は、ファイルアクセス態様テーブル 2 2 1 に登録されていないファイルのアクセス態様についてのみ、アクセス態様情報を収集する。これにより、ディスクアクセスに対する負荷を増やすことなく、階層化ディスク 1 1 でのブロック配置を最適化することができる。

【0076】

また、シーケンシャルファイルを構成するデータブロックが SSD 1 1 1 および SAS ディスク 1 1 2 の両方に分かれて存在する場合、データブロックの前後に空き領域を設けることで、連続的な再配置を容易に行なうことができる。

【0077】

このとき、管理部 2 1 3 は、前後のデータブロックが再配置されるか否かを再配置の頻度から予測し、再配置される可能性が高い場合に、再配置されたデータブロックの前後に空き領域を設定する。これにより、SAS ディスク 1 1 2 の利用効率の低下を招くことなく、SAS ディスク 1 1 2 に再配置されたデータブロックの前後に空き領域を設定することができる。

【0078】

再配置されたデータブロックの前後に空き領域を設けた後、所定期間が経過しても、または、データブロックの再配置処理を所定回数行なっても、空き領域に、連続する他のデータブロックが再配置されない場合、管理部 2 1 3 は、当該空き領域を解放する。これにより、管理部 2 1 3 は、当該空き領域にデータブロックを再配置することができ、SAS ディスク 1 1 2 の領域を有効に利用することが可能になる。

【0079】

〔4〕その他

以上、本発明の好ましい実施形態について詳述したが、本発明は、係る特定の実施形態に限定されるものではなく、本発明の趣旨を逸脱しない範囲内において、種々の変形、変更して実施することができる。

【0080】

なお、本実施形態では、一つのストレージ装置 1 0 に対し一つのサーバ 2 0 を備えた場合について説明したが、ストレージシステム 1 は、ストレージ装置 1 0 に対し複数のサーバがアクセスするように構成されてもよい。この場合、複数のサーバのうちの少なくとも一つが上述したサーバ 2 0 として機能するように構成し、サーバ 2 0 によって得られる各種情報を前記複数のサーバの間で連携させることで、一つのサーバ 2 0 によって複数のサーバにおいて上述した実施形態と同様の作用効果を得ることができる。

【0081】

上述したディスク制御部 1 2 1、仮想ドライバ 2 1 2 および管理部 2 1 3 としての機能の全部もしくは一部は、コンピュータ (CPU、情報処理装置、各種端末を含む) が所定のプログラム (ストレージ制御プログラム) を実行することによって実現される。

【0082】

また、上記アプリケーションプログラムは、例えばフレキシブルディスク、CD (CD-ROM, CD-R, CD-RW など)、DVD (DVD-ROM, DVD-RAM, DVD-R, DVD-RW, DVD+R, DVD+RW など)、ブルーレイディスク等のコンピュータ読取可能な記録媒体に記録された形態で提供される。この場合、コンピュータ

はその記録媒体からプログラムを読み取って内部記憶装置または外部記憶装置に転送し格納して用いる。

【 0 0 8 3 】

ここで、コンピュータとは、ハードウェアとOSとを含む概念であり、OSの制御の下で動作するハードウェアを意味している。また、OSが不要でプログラム単独でハードウェアを動作させるような場合には、そのハードウェア自体がコンピュータに相当する。ハードウェアは、少なくとも、CPU等のマイクロプロセッサと、記録媒体に記録されたコンピュータプログラムを読み取る手段とをそなえている。上記プログラムは、上述のようなコンピュータに、ディスク制御部 1 2 1 , 仮想ドライバ 2 1 2 および管理部 2 1 3 としての機能を実現させるプログラムコードを含んでいる。また、その機能の一部は、プログラムではなくOSによって実現されてもよい。

10

【 0 0 8 4 】

( 5 ) 付記

以上の実施形態に関し、さらに以下の付記を開示する。

【 0 0 8 5 】

( 付記 1 )

複数のディスクを有する階層化ディスクを含むストレージ装置と、前記ストレージ装置にアクセスする情報処理装置とを備えたストレージシステムにおいて、

前記ストレージ装置は、前記階層化ディスクに保存された各ファイルに含まれるデータブロックの記憶位置を含むブロック情報を、前記情報処理装置に通知し、

20

前記情報処理装置は、前記階層化ディスクへのアクセスを行なうアプリケーションが用いるファイル毎のアクセス態様情報、および、前記ストレージ装置から通知された前記各ファイルの前記ブロック情報に基づき、前記データブロックのアクセス態様を判断し、

前記ストレージ装置は、前記情報処理装置において判断された前記データブロックのアクセス態様に基づく前記階層化ディスクに対する前記データブロックの再配置を行なう、ストレージ制御方法。

【 0 0 8 6 】

( 付記 2 )

前記情報処理装置は、前記データブロックに対するアクセス頻度を前記ストレージ装置から取得し、取得した前記アクセス頻度と、前記情報処理装置において判断された前記データブロックのアクセス態様とに基づき、前記階層化ディスクに対する前記データブロックの再配置の内容を決定し、当該内容に従った前記データブロックの再配置を前記ストレージ装置に指示し、

30

前記ストレージ装置は、前記情報処理装置から指示された前記データブロックの再配置を前記階層化ディスクに対し行なう、付記 1 記載のストレージ制御方法。

【 0 0 8 7 】

( 付記 3 )

前記情報処理装置は、仮想デバイスによって、前記アプリケーションから入出力要求を受けると前記入出力要求の対象ファイルと前記対象ファイルのアクセス態様とを前記ファイル毎のアクセス態様情報として特定する、付記 1 または付記 2 に記載のストレージ制御方法。

40

【 0 0 8 8 】

( 付記 4 )

前記情報処理装置は、前記仮想デバイスによって特定された前記対象ファイルと前記対象ファイルのアクセス態様とを前記ファイル毎のアクセス態様情報としてテーブルに登録し、前記テーブルに登録された前記ファイル毎のアクセス態様情報と、前記ストレージ装置から通知された前記各ファイルの前記ブロック情報とに基づき、前記データブロックのアクセス態様を判断する、付記 3 記載のストレージ制御方法。

【 0 0 8 9 】

( 付記 5 )

50



前記仮想デバイスは、前記アプリケーションが用いる前記ファイルのうち、前記テーブルに登録されていない前記対象ファイルについて前記アクセス態様を特定する、付記 4 記載のストレージ制御方法。

【 0 0 9 0 】

( 付記 6 )

前記仮想デバイスは、前記アプリケーションによるアクセス開始後、所定期間だけ前記アクセス態様の特定を行なう、付記 3 ~ 付記 5 のいずれか一項に記載のストレージ制御方法。

【 0 0 9 1 】

( 付記 7 )

前記仮想デバイスは、前記入出力要求が前記対象ファイルの入出力位置をインデックスでシークするか否かを測定し、シークしない場合に前記対象ファイルのアクセス態様をシークエンシャルアクセスと判断する一方、シークする場合に前記対象ファイルのアクセス態様をランダムアクセスと判断する、付記 3 ~ 付記 6 のいずれか一項に記載のストレージ制御方法。

【 0 0 9 2 】

( 付記 8 )

前記ストレージ装置は、前記データブロックのアクセス態様がシークエンシャルアクセスである場合、且つ、前記データブロックが再配置される前記複数のディスクのうちの一のディスク上に、連続する複数の前記データブロックが存在する場合、前記一のディスク上に複数の前記データブロックを連続的に再配置する、付記 7 記載のストレージ制御方法。

【 0 0 9 3 】

( 付記 9 )

前記ストレージ装置は、前記データブロックのアクセス態様がシークエンシャルアクセスである場合、且つ、前記データブロックが再配置される前記複数のディスクのうちの一のディスク上に、前記データブロックの前側または後側に続く他のデータブロックが存在しない場合、前記一のディスク上に再配置された前記データブロックの前側または後側に、前記他のデータブロック用の空き領域を設ける、付記 7 または付記 8 に記載のストレージ制御方法。

【 0 0 9 4 】

( 付記 1 0 )

前記ストレージ装置は、前記空き領域を設けた後、所定期間が経過しても、または、前記データブロックの再配置を所定回数行なっても、前記空き領域に前記他のデータブロックが再配置されない場合、前記空き領域を解放する、付記 9 記載のストレージ制御方法。

【 0 0 9 5 】

( 付記 1 1 )

前記ストレージ装置は、アクセス態様がランダムアクセスであるデータブロックを、アクセス態様がシークエンシャルアクセスであるデータブロックの配置状態に応じて再配置する、付記 8 ~ 付記 1 0 のいずれか一項に記載のストレージ制御方法。

【 0 0 9 6 】

( 付記 1 2 )

複数の情報処理装置が前記ストレージ装置にアクセスする場合、前記複数の情報処理装置のうちの少なくとも一つを前記情報処理装置として機能させる、付記 1 ~ 付記 1 1 のいずれか一項に記載のストレージ制御方法。

【 0 0 9 7 】

( 付記 1 3 )

複数のディスクを有する階層化ディスクを含むストレージ装置と、前記ストレージ装置にアクセスする情報処理装置とを備えたストレージシステムであって、

前記ストレージ装置は、

前記階層化ディスクに保存された各ファイルに含まれるデータブロックの記憶位置を含

10

20

30

40

50

むブロック情報を、前記情報処理装置に通知するディスク制御部を有し、

前記情報処理装置は、

前記階層化ディスクへのアクセスを行なうアプリケーションが用いるファイル毎のアクセス態様情報、および、前記ディスク制御部から通知された前記各ファイルの前記ブロック情報に基づき、前記データブロックのアクセス態様を判断する管理部を有し、

前記ストレージ装置の前記ディスク制御部は、前記管理部によって判断された前記データブロックのアクセス態様に基づき、前記階層化ディスクに対する前記データブロックの再配置を行なう、ストレージシステム。

【0098】

(付記14)

前記情報処理装置の前記管理部は、前記データブロックに対するアクセス頻度を前記ストレージ装置から取得し、取得した前記アクセス頻度と、前記情報処理装置において判断された前記データブロックのアクセス態様とに基づき、前記階層化ディスクに対する前記データブロックの再配置の内容を決定し、当該内容に従った前記データブロックの再配置を前記ストレージ装置に指示し、

前記ストレージ装置の前記ディスク制御部は、前記管理部から指示された前記データブロックの再配置を前記階層化ディスクに対し行なう、付記13記載のストレージシステム。

【0099】

(付記15)

前記情報処理装置は、

前記アプリケーションから入出力要求を受けると前記入出力要求の対象ファイルと前記対象ファイルのアクセス態様とを前記ファイル毎のアクセス態様情報として特定するように構築された仮想デバイスを有する、付記14記載のストレージシステム。

【0100】

(付記16)

前記情報処理装置は、

前記仮想デバイスによって特定された前記対象ファイルに関する情報と前記対象ファイルのアクセス態様とを前記ファイル毎のアクセス態様情報として登録するテーブルを有し、

前記情報処理装置の前記管理部は、

前記テーブルに登録された前記ファイル毎のアクセス態様情報と前記ディスク制御部から通知された前記各ファイルの前記ブロック情報とに基づき、前記データブロックのアクセス態様を判断する、付記15記載のストレージシステム。

【0101】

(付記17)

前記仮想デバイスは、前記アプリケーションが用いる前記ファイルのうち、前記テーブルに登録されていない前記対象ファイルについて前記アクセス態様を特定する、付記16記載のストレージシステム。

【0102】

(付記18)

前記仮想デバイスは、前記アプリケーションによるアクセス開始後、所定時間の間、前記アクセス態様の特定を行なう、付記15～付記17のいずれか一項に記載のストレージシステム。

【0103】

(付記19)

前記仮想デバイスは、前記入出力要求が前記対象ファイルの入出力位置をインデックスでシークするか否かを測定し、シークしない場合に前記対象ファイルのアクセス態様をシーケンシャルアクセスと判断する一方、シークする場合に前記対象ファイルのアクセス態様をランダムアクセスと判断する、付記15～付記18のいずれか一項に記載のストレ

10

20

30

40

50

ジシステム。

【 0 1 0 4 】

( 付記 2 0 )

複数のディスクを有する階層化ディスクを含むストレージ装置と、前記ストレージ装置にアクセスする情報処理装置とを備えたストレージシステムに含まれるコンピュータに、

前記階層化ディスクに保存された各ファイルに含まれるデータブロックの記憶位置を含むブロック情報を、前記ストレージ装置から前記情報処理装置に通知し、

前記情報処理装置において、前記階層化ディスクへのアクセスを行なうアプリケーションが用いるファイル毎のアクセス態様情報、および、前記ストレージ装置から通知された前記各ファイルの前記ブロック情報に基づき、前記データブロックのアクセス態様を判断し、

前記ストレージ装置において、前記情報処理装置において判断された前記データブロックのアクセス態様に基づく前記階層化ディスクに対する前記データブロックの再配置を行なう、

処理を実行させる、ストレージ制御プログラム。

【 0 1 0 5 】

( 付記 2 1 )

複数のディスクを有する階層化ディスクを含むストレージ装置と、前記ストレージ装置にアクセスする情報処理装置とを備えたストレージシステムに含まれるコンピュータに、

前記階層化ディスクに保存された各ファイルに含まれるデータブロックの記憶位置を含むブロック情報を、前記ストレージ装置から前記情報処理装置に通知し、

前記情報処理装置において、前記階層化ディスクへのアクセスを行なうアプリケーションが用いるファイル毎のアクセス態様情報、および、前記ストレージ装置から通知された前記各ファイルの前記ブロック情報に基づき、前記データブロックのアクセス態様を判断し、

前記ストレージ装置において、前記情報処理装置において判断された前記データブロックのアクセス態様に基づく前記階層化ディスクに対する前記データブロックの再配置を行なう、

処理を実行させる、ストレージ制御プログラムを記録したコンピュータ読取可能な記録媒体。

【 符号の説明 】

【 0 1 0 6 】

- 1        ストレージシステム
- 1 0      ストレージ装置
- 1 1      階層化ディスク
- 1 1 0    論理ディスク ( 論理ユニット )
- 1 1 1    S S D ( ディスク )
- 1 1 2    S A S ディスク ( ディスク )
- 1 2      C P U ( 処理部 , コンピュータ )
- 1 2 1    ディスク制御部
- 1 3      メモリ
- 2 0      サーバ ( 情報処理装置 )
- 2 1      C P U ( 処理部 , コンピュータ )
- 2 1 1    アプリケーション ( アプリケーションプログラム )
- 2 1 2    仮想ドライバ ( 仮想デバイス , アクセス態様特定部 )
- 2 1 3    管理部
- 2 2      メモリ
- 2 2 1    ファイル・アクセス態様テーブル ( テーブル )

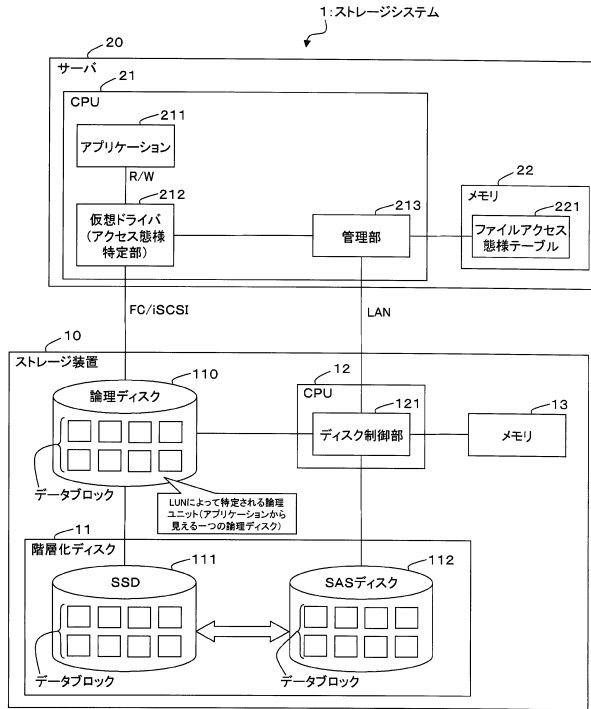
10

20

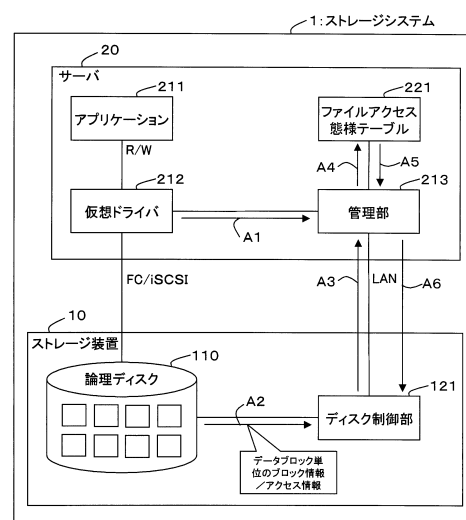
30

40

【図 1】



【図 2】

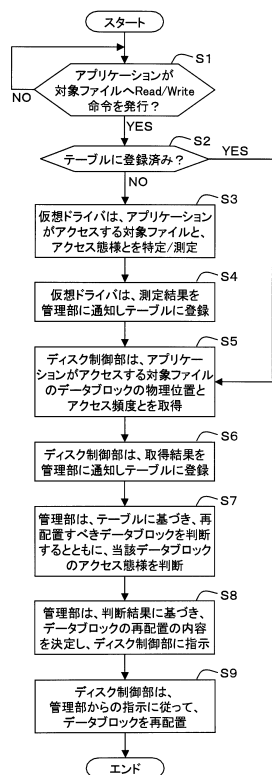


【図 3】

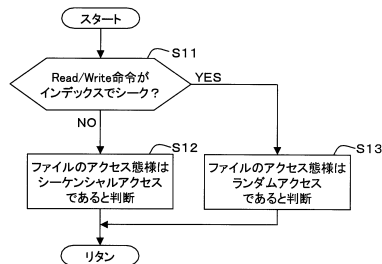
221: ファイルアクセス態様テーブル

アプリケーション	ファイル	アクセス態様	物理ストレージ(記憶位置)	アクセス頻度
Oracle	DB	ランダム	storage1-lun0-block10	200 IOPS
	ログ	シーケンシャル	storage1-lun0-block20	100 IOPS
...	...	...	...	...

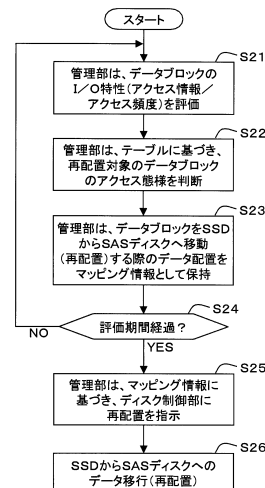
【図 4】



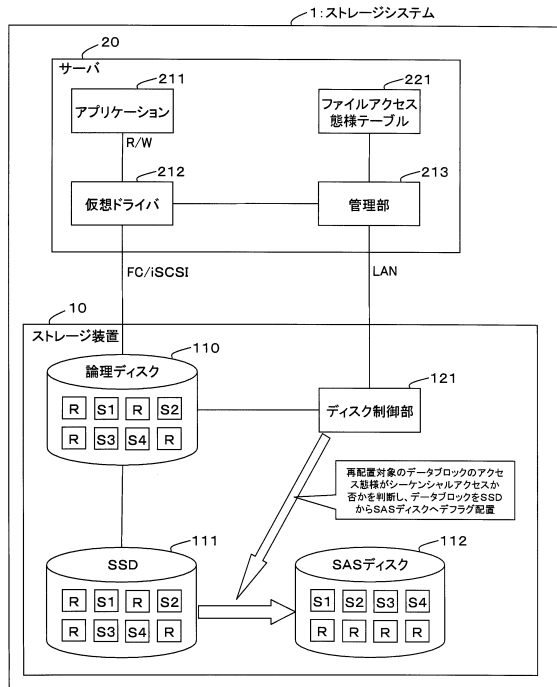
【図 5】



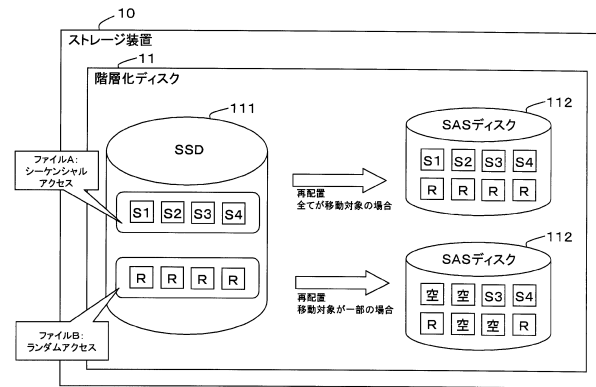
【図 6】



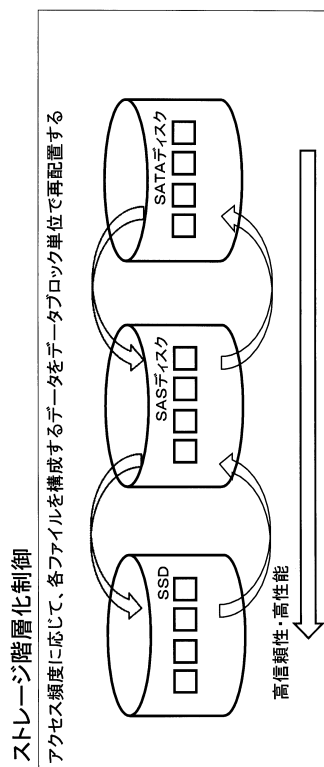
【図 7】



【図 8】



【図 9】



---

フロントページの続き

(56)参考文献 特開平 9 - 44381 (JP, A)

特開2004-227594 (JP, A)

米国特許出願公開第2010/0037002 (US, A1)

干場 一彦, 「ストレージ最新技術動向 仮想化・クラウド対応へ機能強化 階層化やスケールアウトで性能高まる」, 日経SYSTEMS, 日経BP社, 2011年 3月26日, 第216号, pp.36-39, ISSN 1881-1620

ハーバート タム, シャイヤー ロバート L., 「ストレージ環境のROIを高める最強タッグ! [ILM] + [SSD] 相乗効果が生み出すメリット!」, COMPUTERWORLD, 株式会社IDGインタラクティブ, 2010年 8月 1日, 第7巻, 第8号, pp.20-27

(58)調査した分野(Int.Cl., DB名)

G06F 12/00

G06F 3/06