

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4144727号
(P4144727)

(45) 発行日 平成20年9月3日(2008.9.3)

(24) 登録日 平成20年6月27日(2008.6.27)

(51) Int.Cl.

F I

G 0 6 F 12/00 (2006.01)

G 0 6 F 12/00 5 4 5 A

G 0 6 F 12/00 5 2 0 E

G 0 6 F 12/00 5 2 0 J

請求項の数 16 (全 26 頁)

(21) 出願番号 特願2001-200455 (P2001-200455)
 (22) 出願日 平成13年7月2日(2001.7.2)
 (65) 公開番号 特開2003-15931 (P2003-15931A)
 (43) 公開日 平成15年1月17日(2003.1.17)
 審査請求日 平成16年10月21日(2004.10.21)

(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 100096954
 弁理士 矢島 保夫
 (72) 発明者 兼田 泰典
 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】 情報処理システム、記憶領域提供方法、およびデータ保持管理装置

(57) 【特許請求の範囲】

【請求項1】

複数のストレージデバイスと、
 前記複数のストレージデバイスのうち一つ以上のストレージデバイスに記憶される第一データを送信する第一コンピュータと、
 複数のストレージデバイスの各々の属性を示す属性情報を記憶することが可能な第二コンピュータと、
 を有し、
 前記第二コンピュータは、
 前記第一データがどのように保持されるべきであることを示すコントロール情報を受信し

10

、
 前記コントロール情報が前記第一データが高信頼に保持されるべきであることを示す場合には、前記属性情報に基づいて前記複数のストレージデバイスの中の高信頼なストレージデバイスを特定し、前記高信頼なストレージデバイスに前記第一データを保持するように制御し、

前記コントロール情報が前記第一データが高信頼に保持されるべきであることと、他の要求(「他の要求」とは、前記第一データを保持する制御態様を指示する要求であり、「低価格に」若しくは「長期に」のうちの一つ以上の要求をいう)と、の両方を示す場合には、前記属性情報に基づいて前記複数のストレージデバイスの中の前記他の要求を満たすストレージデバイスを複数特定し、前記特定された複数のストレージデバイスの各々に前

20

記第一データを保持するように制御する、
ことを特徴とする情報処理システム。

【請求項 2】

請求項 1 に記載の情報処理システムにおいて、
前記複数のストレージデバイスは、各々、自己の属性情報を保持する属性保持手段を備え、
前記第二コンピュータは、前記複数のストレージデバイスの属性保持手段からそれぞれ属性情報を取得する
ことを特徴とする情報処理システム。

【請求項 3】

請求項 1 または 2 のいずれか 1 つに記載の情報処理システムにおいて、
前記属性情報は、前記ストレージデバイスについての、種別に関する情報、性能に関する情報、およびデータの保持にかかるコストに関するコスト情報を含むことを特徴とする情報処理システム。

【請求項 4】

請求項 1 から 3 のいずれか 1 つに記載の情報処理システムにおいて、
前記第二コンピュータは、
前記コントロール情報を保持する手段を有し、前記コントロール情報が変更された場合、前記変更されたコントロール情報と前記属性情報に基づいて前記複数のストレージデバイスの中からストレージデバイスを選択し、前記コントロール情報が変更される前に前記第一データを保持していたストレージデバイスから前記選択したストレージデバイスに前記第一データを複製または移動するように制御することを特徴とする情報処理システム。

【請求項 5】

請求項 1 から 4 のいずれか 1 つに記載の情報処理システムにおいて、
前記第二コンピュータは、
前記各ストレージデバイスから稼働率を取得する手段を備え、前記ストレージデバイスを特定または選択する際には、取得した稼働率を前記特定または選択の指標として用いることを特徴とする情報処理システム。

【請求項 6】

請求項 1 から 5 のいずれか 1 つに記載の情報処理システムにおいて、
前記第一データには、その第一データが属するグループを特定するグループ識別子が付随しており、
前記第二コンピュータは、前記第一データを複製または移動する際、前記グループ識別子に基づいて同じグループに属する第一データが一つのメディアに記憶されるように制御することを特徴とする情報処理システム。

【請求項 7】

複数のストレージデバイスと、第一コンピュータと、第二コンピュータとを有する情報処理システムにおいて、データを保持する記憶領域を提供する記憶領域提供方法であって、

前記第一コンピュータが、前記複数のストレージデバイスのうち一つ以上のストレージデバイスに記憶される第一データを送信するステップと、

前記第二コンピュータが、前記複数のストレージデバイスの各々の属性を示す属性情報を記憶するステップと、

前記第二コンピュータが、前記第一データがどのように保持されるべきであることを示すコントロール情報を受信するステップと、

前記第二コンピュータが、前記コントロール情報が前記第一データが高信頼に保持されるべきであることを示す場合には、前記属性情報に基づいて前記複数のストレージデバイスの中の高信頼なストレージデバイスを特定し、前記高信頼なストレージデバイスに前記第一データを保持するように制御するステップと、

前記第二コンピュータが、前記コントロール情報が前記第一データが高信頼に保持され

10

20

30

40

50

るべきであることと、他の要求（「他の要求」とは、前記第一データを保持する制御態様を指示する要求であり、「低価格に」若しくは「長期に」のうちの一つ以上の要求をいう）と、の両方を示す場合には、前記属性情報に基づいて前記複数のストレージデバイスの中の前記他の要求を満たすストレージデバイスを複数特定し、前記特定された複数のストレージデバイスの各々に前記第一データを保持するように制御するステップと、
を備えることを特徴とする記憶領域提供方法。

【請求項 8】

請求項 7 に記載の情報処理システムの記憶領域提供方法において、
前記複数のストレージデバイスは、各々、自己の属性情報を保持するものとし、
前記第二コンピュータは、前記複数のストレージデバイスからそれぞれ属性情報を取得する
ことを特徴とする記憶領域提供方法。 10

【請求項 9】

請求項 7 または 8 のいずれか 1 つに記載の情報処理システムの記憶領域提供方法において、
前記属性情報は、前記ストレージデバイスについての、種別に関する情報、性能に関する情報、およびデータの保持にかかるコストに関するコスト情報を含むことを特徴とする記憶領域提供方法。

【請求項 10】

請求項 7 から 9 のいずれか 1 つに記載の情報処理システムの記憶領域提供方法において
、
前記第二コンピュータは、
前記コントロール情報を保持し、前記コントロール情報が変更された場合、前記変更されたコントロール情報と前記属性情報に基づいて前記複数のストレージデバイスの中からストレージデバイスを選択し、前記コントロール情報が変更される前に前記第一データを保持していたストレージデバイスから前記選択したストレージデバイスに前記第一データを複製または移動するように制御することを特徴とする記憶領域提供方法。 20

【請求項 11】

請求項 7 から 10 のいずれか 1 つに記載の情報処理システムの記憶領域提供方法において、
前記第二コンピュータは、
前記各ストレージデバイスから稼働率を取得し、前記ストレージデバイスを特定または選択する際には、取得した稼働率を前記特定または選択の指標として用いることを特徴とする記憶領域提供方法。 30

【請求項 12】

請求項 7 から 11 のいずれか 1 つに記載の情報処理システムの記憶領域提供方法において、
前記第一データには、その第一データが属するグループを特定するグループ識別子が付随しており、
前記第二コンピュータは、前記第一データを複製または移動する際、前記グループ識別子に基づいて同じグループに属する第一データが一つのメディアに記憶されるように制御することを特徴とする記憶領域提供方法。 40

【請求項 13】

複数のストレージデバイスおよび前記複数のストレージデバイスのうち一つ以上のストレージデバイスに記憶される第一データを送信する第一コンピュータに接続される第二コンピュータであるデータ保持管理装置であって、
複数のストレージデバイスの各々の属性を示す属性情報を記憶する手段を有し、
前記第一データがどのように保持されるべきであることを示すコントロール情報を受信し、
前記コントロール情報が前記第一データが高信頼に保持されるべきであることを示す場 50

合には、前記属性情報に基づいて前記複数のストレージデバイスの中の高信頼なストレージデバイスを特定し、前記高信頼なストレージデバイスに前記第一データを保持するように制御し、

前記コントロール情報が前記第一データが高信頼に保持されるべきであることと、他の要求（「他の要求」とは、前記第一データを保持する制御態様を指示する要求であり、「低価格に」若しくは「長期に」のうちの一つ以上の要求をいう）と、の両方を示す場合には、前記属性情報に基づいて前記複数のストレージデバイスの中の前記他の要求を満たすストレージデバイスを複数特定し、前記特定された複数のストレージデバイスの各々に前記第一データを保持するように制御する、

ことを特徴とするデータ保持管理装置。

10

【請求項 14】

請求項 13 に記載のデータ保持管理装置において、

前記複数のストレージデバイスは、各々、自己の属性情報を保持する属性保持手段を備えるものであり、

前記データ保持管理装置は、前記複数のストレージデバイスの属性保持手段からそれぞれ属性情報を取得する

ことを特徴とするデータ保持管理装置。

【請求項 15】

請求項 13 または 14 のいずれか 1 つに記載のデータ保持管理装置において、

前記属性情報は、前記ストレージデバイスについての、種別に関する情報、性能に関する情報、およびデータの保持にかかるコストに関するコスト情報を含むことを特徴とするデータ保持管理装置。

20

【請求項 16】

請求項 13 から 15 のいずれか 1 つに記載のデータ保持管理装置において、

前記コントロール情報を保持する手段を有し、前記コントロール情報が変更された場合、前記変更されたコントロール情報と前記属性情報に基づいて前記複数のストレージデバイスの中からストレージデバイスを選択し、前記コントロール情報が変更される前に前記第一データを保持していたストレージデバイスから前記選択したストレージデバイスに前記第一データを複製または移動するように制御することを特徴とするデータ保持管理装置。

30

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、情報処理システムにおける記憶装置（ディスクアレイ装置やハードディスク装置、およびテープライブラリ装置や光ディスクライブラリ装置など）の利用技術に関し、特に、複数の異なる特性をもつ記憶装置を複数の制御装置（コンピュータ）から利用する技術に関する。

【0002】

【従来の技術】

複数のコンピュータ（制御装置）がストレージ（記憶装置）を共有する技術として、ストレージエリアネットワークが広く使われるようになってきた。ストレージエリアネットワークが利用される以前は、コンピュータシステムのファイルシステムソフトウェアは、複数のコンピュータで一つのストレージを共有して使用することが考慮されておらず、ストレージエリアネットワーク環境で共有して利用すると、ストレージに記録されているデータ（情報）が失われてしまう危険性があった。この点に関しては、米国特許第5,950,203号にその記載がある。

40

【0003】

上記特許では、複数のコンピュータからストレージを共有する技術が開示されているが、ストレージが異なる属性（性能、信頼性、保存性、容量あたりのコストなど）を有する場合のストレージの利用方式についての記載がない。

50

【 0 0 0 4 】

この点については、Tivoli Systems Inc.のホワイトペーパー「Vision, Tivoli Storage Management Solutions for the Information Grid」に、ストレージエリアネットワークにおいて、属性の異なるストレージを使い分けてデータを保持する技術が紹介されている。しかし、このホワイトペーパーにおいても、ストレージにおける属性管理の方法や位置管理手段におけるデータ管理の方式、およびデータの保持に関する課金方式などについては記載がない。

【 0 0 0 5 】

一方、ストレージサービスプロバイダ（SSP:Storage Service Provider）と呼ばれるビジネスがある。ストレージサービスプロバイダは、顧客の要求にあったストレージをアウトソーシングにより提供する。顧客はストレージサービスプロバイダにストレージの管理を委託することで、ストレージの維持管理コストを削減できる。サービスプロバイダは容量の管理を行えばよいだけでなく、顧客の求める性能や信頼性、コスト、耐障害性などに答えていかなければならない。また、ストレージサービスプロバイダ自身も、性能や信頼性を維持しつつ、ストレージにかかる維持管理コストを削減しなければならない。

【 0 0 0 6 】

【発明が解決しようとする課題】

本発明の目的は、たとえばストレージサービスプロバイダなどのビジネス形態で実施される、複数の異なる属性を有するストレージ（記憶装置）で構成されるストレージエリアネットワークにおいて、ストレージの属性（性能やコスト）や稼働率、およびデータの利用頻度を考慮したデータの保持方式を提供することにある。

【 0 0 0 7 】

【課題を解決するための手段】

上記目的を達成するために、請求項1に係る発明は、複数のストレージデバイスと、前記複数のストレージデバイスのうち一つ以上のストレージデバイスに記憶される第一データを送信する第一コンピュータと、複数のストレージデバイスの各々の属性を示す属性情報を記憶することが可能な第二コンピュータと、を有し、前記第二コンピュータは、前記第一データがどのように保持されるべきであることを示すコントロール情報を受信し、前記コントロール情報が前記第一データが高信頼に保持されるべきであることを示す場合には、前記属性情報に基づいて前記複数のストレージデバイスの中の高信頼なストレージデバイスを特定し、前記高信頼なストレージデバイスに前記第一データを保持するように制御し、前記コントロール情報が前記第一データが高信頼に保持されるべきであることと、他の要求（「他の要求」とは、前記第一データを保持する制御態様を指示する要求であり、「低価格に」若しくは「長期に」のうちの一つ以上の要求をいう）と、の両方を示す場合には、前記属性情報に基づいて前記複数のストレージデバイスの中の前記他の要求を満たすストレージデバイスを複数特定し、前記特定された複数のストレージデバイスの各々に前記第一データを保持するように制御する、ことを特徴とする。

【 0 0 0 8 】

請求項7に係る発明は、複数のストレージデバイスと、第一コンピュータと、第二コンピュータとを有する情報処理システムにおいて、データを保持する記憶領域を提供する記憶領域提供方法であって、前記第一コンピュータが、前記複数のストレージデバイスのうち一つ以上のストレージデバイスに記憶される第一データを送信するステップと、前記第二コンピュータが、前記複数のストレージデバイスの各々の属性を示す属性情報を記憶するステップと、前記第二コンピュータが、前記第一データがどのように保持されるべきであることを示すコントロール情報を受信するステップと、前記第二コンピュータが、前記コントロール情報が前記第一データが高信頼に保持されるべきであることを示す場合には、前記属性情報に基づいて前記複数のストレージデバイスの中の高信頼なストレージデバイスを特定し、前記高信頼なストレージデバイスに前記第一データを保持するように制御するステップと、前記第二コンピュータが、前記コントロール情報が前記第一データが高信頼に保持されるべきであることと、他の要求（「他の要求」とは、前記第一データを保持

する制御態様を指示する要求であり、「低価格に」若しくは「長期に」のうちの一つ以上の要求をいう」と、の両方を示す場合には、前記属性情報に基づいて前記複数のストレージデバイスの中の前記他の要求を満たすストレージデバイスを複数特定し、前記特定された複数のストレージデバイスの各々に前記第一データを保持するように制御するステップと、を備えることを特徴とする。

【 0 0 0 9 】

請求項 1 3 に係る発明は、複数のストレージデバイスおよび前記複数のストレージデバイスのうち一つ以上のストレージデバイスに記憶される第一データを送信する第一コンピュータに接続される第二コンピュータであるデータ保持管理装置であって、複数のストレージデバイスの各々の属性を示す属性情報を記憶する手段を有し、前記第一データがどのように保持されるべきであることを示すコントロール情報を受信し、前記コントロール情報が前記第一データが高信頼に保持されるべきであることを示す場合には、前記属性情報に基づいて前記複数のストレージデバイスの中の高信頼なストレージデバイスを特定し、前記高信頼なストレージデバイスに前記第一データを保持するように制御し、前記コントロール情報が前記第一データが高信頼に保持されるべきであることと、他の要求（「他の要求」とは、前記第一データを保持する制御態様を指示する要求であり、「低価格に」若しくは「長期に」のうちの一つ以上の要求をいう」と、の両方を示す場合には、前記属性情報に基づいて前記複数のストレージデバイスの中の前記他の要求を満たすストレージデバイスを複数特定し、前記特定された複数のストレージデバイスの各々に前記第一データを保持するように制御する、ことを特徴とする。

【 0 0 1 0 】

複数のストレージデバイスは、各々、自己の属性情報を保持し、第二コンピュータは、前記複数のストレージデバイスのそれぞれから属性情報を取得するようにするとよい。また、属性情報は、前記ストレージデバイスについての、種別に関する情報、性能に関する情報、およびデータの保持にかかるコストに関するコスト情報を含むようにするとよい。

【 0 0 1 1 】

また、前記コントロール情報を保持するようにし、前記コントロール情報が変更された場合、前記変更されたコントロール情報と前記属性情報に基づいて前記複数のストレージデバイスの中からストレージデバイスを選択し、前記コントロール情報が変更される前に前記第一データを保持していたストレージデバイスから前記選択したストレージデバイスに前記第一データを複製または移動するように制御してもよい。

【 0 0 1 2 】

また、前記各ストレージデバイスから稼働率を取得し、前記ストレージデバイスを特定または選択する際には、取得した稼働率を前記特定または選択の指標として用いるようにしてもよい。

【 0 0 1 3 】

また、前記第一データに、その第一データが属するグループを特定するグループ識別子を付随させ、前記第二コンピュータは、前記第一データを複製または移動する際、前記グループ識別子に基づいて同じグループに属する第一データが一つのメディアに記憶されるように制御してもよい。

【 0 0 1 4 】

【発明の実施の形態】

以下、図面を用いて本発明の第 1 の実施形態（キャッシュなしの場合）を説明する。

【 0 0 1 5 】

図 1 は、第 1 の実施形態のシステム構成図を示す。本システムは、3 台のコンピュータ 1 0 1 , 1 0 2 , 1 5 1 と、3 台のディスクアレイ装置 2 0 1 , 2 0 2 , 2 0 3 と、ライブラリ装置 3 0 1 と、ファイバチャネルスイッチ 5 0 1 とを備えている。コンピュータ 1 0 1 , 1 0 2 , 1 5 1 と、ディスクアレイ装置 2 0 1 , 2 0 2 , 2 0 3 と、ライブラリ装置 3 0 1 は、ファイバチャネル 5 0 を介して、ファイバチャネルスイッチ 5 0 1 に接続する。ファイバチャネルスイッチ 5 0 1 は、接続された各機器間のパスの確立と切り替えを行

う。また、コンピュータ 101, 102, 151 は、イーサネット 60 を用いて相互に接続する。

【0016】

なお、第 1 の実施形態の後に第 2 および第 3 の実施形態について説明するが、第 2 および第 3 の実施形態のシステムも第 1 の実施形態のシステムと同様の構成であるので図 1 を用いて説明する。図 1 において、コンピュータ 101, 102 中のキャッシュディスク 90 と情報受領手段 91、並びに、コンピュータ 101, 102、ディスクアレイ装置 201, 202, 203、およびライブラリ装置 301 中の稼働率計測手段 82 については、第 2 の実施形態のシステムに固有の構成であるので第 1 の実施形態では言及しない。同様に、クライアント 70 についても、第 3 の実施形態のシステムに固有の構成であるので第 1 および第 2 の実施形態では言及しない。

10

【0017】

図 1 において、ファイバチャネル 50 を用いて構成したネットワークは一般的に SAN (Storage Area Network) と呼ばれ、イーサネット 60 を用いて構成したネットワークは一般的に LAN (Local Area Network) と呼ばれる。図 1 では、ファイバチャネル 50 とイーサネット 60 の二つのネットワークを用いて構成したが、これらは、他のネットワークでもよいし、イーサネット 60 の機能をファイバチャネル 50 で肩代わりすることでも実現可能である。

【0018】

3 台のコンピュータのうち、コンピュータ 151 は、データの位置管理手段 801 と情報複製手段 890 を備える。位置管理手段 801 と情報複製手段 890 の詳細については後述する。

20

【0019】

次に、ディスクアレイ装置と属性保持手段について説明する。ディスクアレイ装置は、一般的に複数のハードディスク装置を有し、RAID (Redundant Array of Inexpensive Disks) 技術により、性能および信頼性の改善を図った記憶装置である。ディスクアレイ装置は、ディスクアレイ装置を構成するハードディスク装置の性能、台数、および RAID レベルによりその性能と信頼性が大きく異なってくる。たとえば、RAID 0 と呼ばれるレベルは、データを複数のハードディスクに分散することで性能は高くなるが、RAID 0 を構成する複数のハードディスクのうち 1 台でも故障するとデータを失う危険性がある。また、RAID 5 と呼ばれるレベルは、データを複数のハードディスクに分散することで性能を上げ、さらにパリティによる冗長データを付加することで、RAID 5 を構成する複数のハードディスクのうち 1 台が故障しても、データを復元可能な信頼性を提供している。RAID 5 は、RAID 0 に比べてパリティによる信頼性向上が見込まれるが、その分、書き込み時の性能が低下する。また、RAID 0 および RAID 5 とともに、構成するハードディスク装置の台数によりその性能が異なる。

30

【0020】

本実施形態では、ディスクアレイ装置 201 と 202 を RAID 0 で構成し、ディスクアレイ装置 203 を RAID 5 で構成するものとする。それぞれのディスクアレイ装置には、属性保持手段 80 を設けた。ディスクアレイ装置 201 と 202 の属性保持手段 80 は、ハードディスク装置で構成される RAID 0 であること、単位容量あたりの保持コスト、構成するハードディスク装置台数、ハードディスク装置の性能（回転数、シーク時間、転送速度など）を保持する。ディスクアレイ装置 203 の属性保持手段 80 は、ハードディスク装置で構成される RAID 5 であること、単位容量あたりの保持コスト、構成するハードディスク装置台数、ハードディスク装置の性能（回転数、シーク時間、転送速度など）を保持する。よって、位置管理手段 801 は、属性保持手段 80 からこれらの属性情報を取り出せば、そのディスクアレイ装置がどのような性能と信頼性を有するかを容易に把握することができる。属性保持手段 80 の実現手段としては、ディスクアレイ装置内にある不揮発性の記憶手段が望ましい。たとえば、装置識別子や装置情報が格納されているメモリを用いて属性保持手段 80 を実現すればよい。また、属性保持手段 80 に格納され

40

50

ている情報の取出しには、S C S I (Small Computer System Interface)で定義されているINQUIRYコマンドやMODE SENSEコマンドを利用する。

【0021】

次に、ライブラリ装置と属性保持手段について説明する。本実施形態において、ライブラリ装置301は、2台のDVD-RAMドライブ30と、メディアを格納可能な4個の棚31と、メディアをドライブ30と棚31との間で搬送する1つの搬送機構32を有する。ただし本発明は、ライブラリ装置におけるドライブ台数、棚の数、搬送機構の数などに依存しない。また、本発明は、複数のライブラリ装置を用いても構成できる。本実施形態では、4個の棚31それぞれに1枚ずつDVD-RAMメディアを格納するものとする。ライブラリ装置301は、搬送機構32を用いて、コンピュータからの指示に基づいて、指定されたメディアを、ドライブ-ドライブ間、ドライブ-棚間、および棚-棚間で搬送することができる。

10

【0022】

ライブラリ装置301に格納されたメディアに記録されているデータを読み出すためには、コンピュータは、まずライブラリ装置に、読みたいデータの記録されているメディアを棚からドライブに搬送するように指示する。コンピュータは、ライブラリ装置301からの搬送終了の報告を待って、ドライブ30にデータの読み出しを指示する。このように、目的のデータを読み出すためには、搬送機構32によるメディアの搬送時間と、ドライブにおいてメディアが読み出し可能になるまでの時間（主にメディアが読み出し可能な回転数に達するまでのスピニング時間）がかかることになる。この時間を約10秒とすると、ディスクアレイ装置201、202、203などのデータ読み出しにかかる時間が数msであることと比べると、その差は3桁以上になる。

20

【0023】

本実施形態では、ライブラリ装置301にも属性保持手段80を設けた。ライブラリ装置301の属性保持手段80は、DVD-RAMドライブで構成されるライブラリ装置であること、単位容量あたりの保持コスト、搬送時間、ライブラリ装置301に搭載されているドライブ30の台数および種類、メディアの枚数、種類および寿命、並びに、ドライブの性能（回転数、シーク時間、転送速度など）を保持する。属性保持手段80は、ライブラリ装置301内にある不揮発性の記憶手段を用いて実現するのが望ましい。

【0024】

次に、位置管理手段と変換表について説明する。

30

【0025】

コンピュータ151は、位置管理手段801を備える。具体的には、位置管理手段801は、コンピュータ151上で実行されるソフトウェアとして実装してある。位置管理手段801は、コンピュータ101や102とはLAN60を介して通信する。通信のためのプロトコルに関しては特に限定されることはないが、たとえばTCPやUDPなどが利用できる。位置管理手段801は、コンピュータ101や102からファイル名を受け取ると、そのファイル名のファイルのデータを必要に応じてロックし、そのデータの位置をコンピュータ151に報告する。このため、ファイル名とデータの位置情報とを対応付けた変換表811を用意する。

40

【0026】

図2は、変換表811の構成を示す。変換表811は、ホストコンピュータ151のメモリに格納される。変換表811は、1つのファイル名に対してライトビット、リードカウント、制御情報、複数の位置情報、およびファイル履歴の各フィールドを備える（グループIDについては後述する）。また、各位置情報は、有効ビット、記憶装置番号、メディア番号、LBA(Logical Block Address)、およびデータ長を持つことができる。本実施形態の場合、記憶装置番号には、WWN(World Wide Name)を使用した。WWNは、ファイバチャネルアダプタ固有に割り当てられる番号である。本実施形態では、ディスクアレイ装置201のWWNを「201」、ディスクアレイ装置202のWWNを「202」、ディスクアレイ装置203のWWNを「203」、ライブラリ装置301のWWNを

50

「301」とした。メディア番号は、対象となる記憶装置がライブラリ装置の場合にのみ有効で、そのデータが格納されているメディアの番号（または棚の番号）を保持する。LBAには、そのデータの格納されている論理ブロックアドレスを使用する。論理ブロックとは、記憶装置が管理するデータを記憶する最小単位である。ディスクアレイの場合には一般的に512バイトのデータを格納でき、DVD-RAMメディアの場合には一般的に2048バイトのデータを格納できる。論理ブロックアドレスとは、この論理ブロックに割り当てられた番号のことである。位置情報の有効ビットがセットされている場合に限り、これら位置情報が有効になる。

【0027】

次に、位置管理手段と領域管理表について説明する。位置管理手段801には、領域管理表812を設けた。図3に領域管理表812の一例を示す。領域管理表812は、ディスクアレイ装置201, 202, 203の場合にはディスクアレイ装置毎に、ライブラリ装置301の場合にはメディア毎に用意する。領域管理表812は、その装置またはメディアにおいてどの論理ブロックが使われているかを示した表で、1つの論理ブロックあたり1ビットの情報で構成している。すなわち、論理ブロックに対応したビットが「1」の場合にはすでにその論理ブロックにデータがあり、「0」の場合には未使用であることを示す。領域管理表812は、コンピュータ151のメモリに格納される。

【0028】

次に、新規書き込み処理について説明する。図5は、本システムにおいて、記憶装置にデータを新規に書き込む処理手順を示すフローチャートである。コンピュータ101を使用するユーザが、ワープロなどのアプリケーションを用いて新規に文書を作成し初めて保存する場合などに相当する(5000)。

【0029】

コンピュータ101のファイルシステム10は、イーサネット60を介して、ファイルネーム「AAA」というデータに対する新規領域割り当てを、コンピュータ151の位置管理手段801に要求する(5001)。このとき、コンピュータ101は、データ「AAA」(データ「AAA」とは、ファイルネーム「AAA」のファイルの実体データである)のデータ長に加えて、データ「AAA」をどのように保存したいかを示す制御情報を数値化して付加する。制御情報の例としては、「高速に」、「高信頼に」、「長期に」、「低価格に」、「災害対応に」、「一時的に」、「並列に」、および「可搬に」などがある。これらは同時に設定可能としてもよいし、排他的な設定とすることもできる。ここでは、「高信頼に」を指定したとする。

【0030】

位置管理手段801は、まず変換表811から空いている一行を探し出し(5002)、ライトビットを「1」にして(5003)書き込みの開始を宣言し、ファイルネームを書き込む(5004)。ここでは変換表811の1番目(#1)の位置情報を確保したとする。領域管理表812から、データ「AAA」を格納するのに必要な領域を検索する(5005)。ここでは、すべての記憶装置に十分な空き領域があるとする(5006, 5007)。位置管理手段801は、すべての記憶装置から割り当て可能位置を獲得することができる。位置管理手段801は、それぞれの記憶装置の属性保持手段80から、その記憶装置の属性情報を獲得する(5008)。位置管理手段801は、コンピュータ101が指示した「高信頼に」という制御情報をもとに、3台のディスクアレイ装置の中から、RAID5で構成されたディスクアレイ装置203を選択し(5009)、対応する領域管理表812のLBAを割り当て済みに変更し(5010)、位置情報をコンピュータ101に報告する(5011)。

【0031】

コンピュータ101のファイルシステム10は、空き領域の報告を受けると、指定された空き領域にデータを書き込むためのライトコマンドをディスクアレイ203に発行し(5012)、データを書き込む(5014)。このライトコマンドの発行と書き込むデータの転送は、ファイバチャネル50を介して行う。コンピュータ101のファイルシステム

10

20

30

40

50

10は、ディスクアレイ装置203から書き込み完了の報告(5015)を受け取ると、コンピュータ151の位置管理手段801に書き込みが完了したことを報告する(5016)。

【0032】

位置管理手段801は、書き込み完了の報告を受け取ると、変換表811の当該位置情報、ここでは1番目の位置情報に、ディスクアレイ装置203のWWN、LBA、およびデータ長を書き込み(5017)、有効ビットを「1」にセットする(5018)ことで、1番目の位置情報が有効であることを示す。この場合、データをディスクアレイ装置に書き込んだので、メディア番号には値を入れる必要はない。また、ファイル履歴には、新規作成とその日時を書き込む(5019)。最後に、ライトビットを「0」にして(5020)、書き込みを完了する(5021)。ファイルシステム10は、位置管理手段801から書き込み完了の報告を受け取ると、アプリケーションに書き込み完了の報告をする(5022)。これにより、アプリケーションからの新規文書保存が完了する(5023)。

10

【0033】

図5に示すフローチャートの処理ステップ5006において、割り当て可能な領域が無かった場合には、書き込み処理は失敗となる(5098, 5099)。また処理ステップ5006で割り当て可能な領域はあるが、処理ステップ5007で複数の記憶装置に領域が無かったときは、処理ステップ5010に進む。

【0034】

次に、読み出し処理について説明する。図6は、本システムにおいて記憶装置からデータを読み出す処理手順を示すフローチャートである。コンピュータ101を使用するユーザが、ワープロなどのアプリケーションを用いて、前回保存しておいた文書を再読み込みする場合などに相当する(6000)。

20

【0035】

コンピュータ101のファイルシステム10は、イーサネット60を介して、ファイルネーム「AAA」のデータの読み出しを位置管理手段801に要求する(6001)。位置管理手段801は、変換表811からファイルネーム「AAA」に対応する行を選択し(6002)、ライトビットが「0」であることを確認する(6003)。もしライトビットが「1」の場合には、他のアプリケーション(他のコンピュータのアプリケーションでもよい)がデータ「AAA」を利用中であるので読み出しをすることができない(6098, 6099)。

30

【0036】

次に、リードカウントをインクリメント(1加算)し(6004)、データ「AAA」の入っている位置情報をコンピュータ101に報告する(6007)。データ「AAA」に対応する位置情報を1つしかもっていない場合には、その位置情報を報告するしかないが、もし、2つ以上の位置情報を有する場合には、コンピュータ101に最適な(たとえば、どこから読み出すのが最も速いかなど)位置情報を選択し(6006)、報告する(6007)。ここでは、上記新規書き込みで書き込んだデータ「AAA」の位置情報であるとする。

40

【0037】

コンピュータ101のファイルシステム10は、位置情報(WWN、LBA、データ長)の報告を受けると、ディスクアレイ装置203にリードコマンドを発行し(6008)、データを読み出す(6009)。このリードコマンドの発行とリードデータの転送は、ファイバチャネル50を介して行う。コンピュータ101のファイルシステム10は、ディスクアレイ装置203から読み出し完了の報告(6010)を受け取ると、コンピュータ151の位置管理手段801に読み出しが完了したことを報告する(6011)。

【0038】

位置管理手段801は、読み出し完了の報告を受け取ると、ファイル履歴に最終アクセス日時を書き込み(6012)、リードカウントをデクリメント(1減算)する(6013)

50

）。リードカウントが0であれば、誰もデータ「AAA」を参照していないことを示し、0でなければまだ読み出し中であることがわかる。ファイルシステム10は、位置管理手段801から読み出し完了の報告(6014)を受け取ると、アプリケーションに読み出しの完了を報告する(6015)。これにより、アプリケーションからの文書読み出しが完了する(6016)。

【0039】

次に、制御情報の変更について説明する。ユーザは、コンピュータ101や102上に用意したユーティリティを用いて、設定した制御情報の参照および再設定が可能である。ここでは、先に書き込んだデータ「AAA」の制御情報を変更する例を示す。ユーザが、制御情報の再設定をユーティリティに要求すると、ファイルシステム10は、位置管理手段801に制御情報の読み出しを要求する。位置管理手段801は、変換表811よりデータ「AAA」の制御情報を読み出し、ファイルシステム10に報告する。ファイルシステム10は、制御情報を受け取ると、ユーティリティに対して制御情報を報告する。ここでは、位置管理手段801が、先に設定した「高信頼に」の制御情報を読み出したとする。

【0040】

次に、ユーザが「高信頼に」の制御情報を「低価格に」に変更すると、ファイルシステム10は、位置管理手段801に制御情報の書き込みを要求する。位置管理手段801は、変換表811のデータ「AAA」に対応する制御情報を「低価格に」に変更し、変更完了をファイルシステム10に報告する。ファイルシステム10は、ユーティリティに変更の完了を報告し、制御情報の変更が終了する。

【0041】

次に、位置管理手段801による監視と移動について説明する。図7と図8は、位置管理手段801におけるデータの監視と移動の処理手順を示すフローチャートである。ここでは、データ「AAA」をディスクアレイ装置203からライブラリ装置301に移動する例について説明する。位置管理手段801は、一定の周期で変換表811を繰り返し検索している(7000)。先ほど、制御情報を「高信頼に」から「低価格に」へ変更したデータ「AAA」についても、この一定周期の巡回で変更が発見される(7001)。位置管理手段801は、このデータの最終アクセス日時と現在の日時とを比較し(7002)、一定期間(たとえば90日とか)経っているか否かをチェックする(7003)。一定期間経っていた場合には、移動対象であるものとして、データ「AAA」をよりビットコストの安い記憶装置に移動する。本実施形態では、ディスクアレイ装置201と202の保持コストを「10」、ディスクアレイ装置203の保持コストを「15」、ライブラリ装置301の保持コストを「1」とすると、位置管理手段801は、ライブラリ装置301にデータ「AAA」を移動することでデータの保持コストの低減を図る。

【0042】

そこで、位置管理手段801は、変換表811で、ライトビットが「0」で(7004)、リードカウントが「0」であることを確認し(7005)、ライトビットを「1」に設定し(7006)、次に領域管理表812を参照しライブラリ装置301のメディアからデータ「AAA」を格納するのに必要な領域を検索する。本実施形態の場合は、属性保持手段80の属性情報を獲得し(7007)、その情報をもとにデータ保持コストのより低い記憶装置を選択し(7008)、新規領域を割り当てる(7009)。

【0043】

データ「AAA」を格納できる空き領域が見つかり(7010)、位置管理手段801は、対応する領域管理表812のLBAを割り当て済みに変更し(7011)、情報複製手段890にデータ「AAA」の複製を要求する(8000)。本実施形態では、情報複製手段890を、コンピュータ151に設けているが、ファイバチャネルスイッチ501に設けてもよいし、ファイバチャネルスイッチ501に接続される他の機器に設けてもよい。

【0044】

図8に、情報複製手段890による複製処理の手順を示す。情報複製手段890は、まず

10

20

30

40

50

、ライブラリ装置 301 に対して、空き領域が見つかったメディアをドライブ 30 に搬送するように搬送命令を発行する (8002)。すでに対象のメディアがドライブにある場合には、メディアを搬送する必要はない (8001)。搬送機構 32 は、指定されたメディアを、それが格納された棚から指定された目的のドライブへ搬送する (8003, 8004)。情報複製手段 890 は、搬送したドライブで読み書きができる状態になるまで待つ (8005)。次に、情報複製手段 890 は、すでにデータ「AAA」が書き込まれているディスクアレイ装置 203 に対してリードコマンドを発行し (8006)、データを読み出し (8007, 8008)、搬送されたメディアが装着されたドライブにライトコマンドを発行し (8009)、データを書き込む (8010, 8011)。情報複製手段 890 は、ライトコマンドを完了すると、位置管理手段 801 に複製が完了したことを報告する (8012)。

10

【0045】

再び図 7 に戻って、位置管理手段 801 は、複製完了の報告を受け取ると、変換表 811 の空いている行、ここでは 2 番目の位置情報に、ライブラリ装置 301 の WWN、メディア番号、LBA、およびデータ長を書き込み (7013)、有効ビットを「1」にセットする (7014)。次に、位置管理手段 801 は、1 番目の位置情報の有効ビットを「0」にすることで (7015)、1 番目の位置情報を無効にし、1 番目の位置情報に対応する領域管理表 812 の LBA を未使用にする。最後に、位置管理手段 801 は、ファイル履歴に移動日時を書き込み、ライトビットを「0」にして移動を完了する (7016)。

【0046】

20

このデータ「AAA」の移動の後、たとえば、コンピュータ 102 からデータ「AAA」の読み出しが要求されたとしても、位置管理情報 801 は、ライブラリ 301 に格納されたデータ「AAA」の位置を正しく報告することができる。これは、位置管理手段 801 が、変換表 811 の位置情報を一元的に管理しているためである。データをディスクアレイ装置 203 からライブラリ装置 301 に移動したことを、コンピュータ 101 と 102 に報告する必要はない。

【0047】

また、もしデータ「AAA」が、「低価格に」に加えて「高信頼に」の制御情報をもっていた場合、位置管理手段 801 は、2 枚のメディアに対して空き領域を検索し、各々にデータを移動する。これにより、1 枚のメディアが障害を発生して読めなくなったり、1 枚のメディアを紛失した場合でもデータの読み出しが可能になり、「高信頼に」の制御ができる。

30

【0048】

次に、ライブラリ装置 301 に格納されたデータの参照について説明する。図 15 と図 16 は、ライブラリ装置 301 に格納されたデータ「AAA」の読み出し処理手順を示すフローチャートである。たとえば、上記ディスクアレイ装置 203 からライブラリ装置 301 へのデータ「AAA」の移動後、コンピュータ 102 が、データ「AAA」の読み出しを要求したとする (15000)。コンピュータ 102 のファイルシステム 10 は、イーサネット 60 を介して、ファイル名「AAA」というデータの読み出しを要求する (15001)。

40

【0049】

位置管理手段 801 は、変換表 811 からファイル名「AAA」に対応する行を選択し (15002)、ライトビットが「0」であることを確認し (15003)、リードカウントをインクリメントし読み出し処理を開始する (15004)。位置管理手段 801 は、ディスクアレイ装置 201 にデータ「AAA」を格納可能な空き領域を検索する (15005)。空き領域が見つかり、対応する領域管理表 812 の LBA を割り当て済みに変更し (15006)、情報複製手段 890 に対しデータ「AAA」をライブラリ装置 301 からディスクアレイ装置 201 に複製するように要求する (16000)。

【0050】

図 16 に、情報複製手段 890 による複製処理の手順を示す。情報複製手段 890 は、デ

50

ータ「AAA」の格納されたメディアをドライブ30に搬送するようにライブラリ装置301に指示する(16002)。もし、データ「AAA」の格納されているメディアがすでにドライブ内にあるのであれば搬送する必要はない(16001)。搬送機構32は、指定されたメディアの格納された棚から指定された目的のドライブへメディアを搬送する(16003, 16004)。ドライブに搬送したメディアが読み出し可能になるのを待って(16005)、ライブラリ装置301に対してリードコマンドを発行してデータを読み出し(16007, 16008)、ディスクアレイ装置201にライトコマンドを発行し(16009)、データを書き込む(16010, 16011)。情報複製手段890は、ライトコマンドを完了すると、位置管理手段801に複製が完了したことを報告する(16012)。

10

【0051】

再び図15に戻って、位置管理手段801は、複製完了の報告を受け取ると、変換表811の空いている位置情報にディスクアレイ装置201のWWN、LBA、およびデータ長を書き込み(15007)、有効ビットを「1」にセットし(15008)、この位置情報をコンピュータ102のファイルシステム10に報告する(15009)。

【0052】

コンピュータ102のファイルシステム10は、位置情報(WWN、メディア番号、LBA、データ長)の報告を受けると、ディスクアレイ装置201にリードコマンドを発行し(15010)、データを読み出す(15011, 15012)。コンピュータ102のファイルシステム10は、ディスクアレイ装置201から読み出し完了の報告を受け取ると、コンピュータ151の位置管理手段801に読み出しが完了したことを報告する(15013)。位置管理手段801は、読み出し完了の報告を受け取ると、変換表811のファイル履歴に最終アクセス日時と複製日時を書き込み(15014)、リードカウントをデクリメントし(15015)、読み出しが完了する(15016, 15017, 15018)。

20

【0053】

この場合、たとえばデータ「AAA」に「高信頼に」の制御情報が与えられていたとしても、RAID5で構成されているディスクアレイ203にファイル「AAA」を複製する必要性は必ずしもない。なぜなら、ライブラリ装置301とRAID0のディスクアレイ装置201の2箇所にデータが保持されているためである。

30

【0054】

ここで、位置管理手段801は、複製したディスクアレイ装置201上のデータが一定期間アクセスされていないことを確認すると、変換表811の位置情報の有効ビットを「0」にリセットする。この際、この位置情報に対応する領域管理表812のLBAは未使用に戻される。

【0055】

次に、制御情報のバリエーションについて説明する。これまでに「低価格に」と「高信頼に」の2つの制御情報について記載してきたが、その他の制御情報が設定された場合の動作について記載する。

【0056】

「高速に」の制御情報が与えられた場合には、2台のRAID0ディスクアレイ装置201と202の2箇所にデータを記録する。RAID0は信頼性がRAID5に比べて劣るため、RAID0のディスクアレイ装置2台にデータを記録することで信頼性を維持する。「高速に」の制御情報に加えて、「一時的に」の制御情報が付加されていた場合には、1台のRAID0ディスクアレイ装置にのみデータを記録する。

40

【0057】

「長期に」の制御情報が与えられた場合には、DVD-RAMメディアなど、取り外し可能な光記録メディアにデータを記録する。DVD-RAMメディアは、30年の保存が可能であるといわれており、ハードディスク装置に比べ長期間に渡りデータを保管するのに向いている。「長期に」の制御情報に加えて、「高信頼に」の制御情報が付加されている

50

場合には、2枚のメディアにデータを記録する。また、「長期に」の制御情報に加えて、「高速に」の制御情報が付加されている場合には、ディスクアレイ装置にもデータを記録しておく。ディスクアレイ装置に記録したデータは、アクセスの遅いライブラリ装置に対するキャッシュとして働く。

【0058】

「災害対応に」の制御情報が与えられた場合には、遠隔地に設置されている記憶装置にデータを複写する。この場合、属性情報には記憶装置の設置場所を保持し、位置管理手段801は、異なる設置場所の2台以上の記憶装置にデータを記憶する。たとえば、ディスクアレイ装置201の設置場所が「東京」で、ディスクアレイ装置202の設置場所が「横浜」で、ディスクアレイ装置203の設置場所が「大阪」の場合、位置管理手段801は、設置場所の地理関係を把握し、「東京」のディスクアレイ装置201と「大阪」のディスクアレイ装置203の2箇所にデータを記録する。地震などの大災害時に、東京と横浜の距離では災害対応にならないことを位置管理手段801に設定しておく。

【0059】

ファイバチャネル50の接続距離は最大10kmとなっており、実際に東京と横浜および大阪を接続するためには、ファイバチャネル50をATM(Asynchronous Transfer Mode)51など他のネットワークに変換して接続する必要があるのはいうまでもない。

【0060】

「並列に」の制御情報が与えられた場合には、ディスクアレイ装置201と202の2箇所にデータを記録し、ディスクアレイ装置201に記録したデータをコンピュータ101専用とし、ディスクアレイ装置202に記録したデータをコンピュータ102専用とする。位置管理手段801は、読み出しを要求したコンピュータを判断して、対応する位置情報を報告する。

【0061】

「可搬に」の制御情報が与えられた場合には、媒体の取り外しが可能なライブラリ装置301にデータを記録する。このとき、制御情報にグループID(図2)を付加して位置管理手段801に与えることで、位置管理手段801は、同じグループIDをもつデータをひとつのメディアに集めるようにデータを記録する。「可搬に」の制御情報を扱うためには、変換表811にグループIDを保持する領域を設ける必要がある。

【0062】

また、これまで記載してきた実施形態では、データを新規に書き込むときに制御情報を与えることが必要であった。もちろん、ユーティリティを用いて途中で制御情報を変更することもできる。以下では、別の制御情報の与え方について説明する。

【0063】

図4は、ディレクトリ構造の例を示す図である。コンピュータシステムでは、ディレクトリ構造によるファイル管理を用いることが多い。これは、ファイルをフォルダに分類することにより、ユーザによるファイルの検索を容易にするためである。たとえば図4のような構成では、出発点になるルートフォルダ1には2つのサブフォルダ2と3があり、フォルダ3にはさらに2つのサブフォルダ4と5がある。データ「AAA」が、このフォルダ4に収められているとすると、データ「AAA」は、「/1/3/4/AAA」となる。このようなディレクトリ構成では、フォルダに基本的な制御情報を与えておき、そのフォルダに格納されるファイルやサブフォルダには、そのフォルダの制御情報が継承されるように構成する。たとえば、フォルダ1には「制御情報なし」、フォルダ2には「高速に」、フォルダ3には「高信頼に」の制御情報を与えておけば、これらのフォルダに格納されるファイルについてはそれぞれの制御情報が継承される。また、フォルダ3にフォルダ4と5を作成すると、自動的にフォルダ3の制御情報である「高信頼に」がフォルダ4と5にも与えられる。継承させずに新規に制御情報を与えなおすこともできる。フォルダ4に「長期に」の制御情報を追加すると、フォルダ4に保持されるデータ「AAA」には、デフォルトの制御情報として、「高信頼に」と「長期に」の制御情報が与えられることになる。

10

20

30

40

50

【 0 0 6 4 】

また、フォルダ 3 の制御情報を変更すると、それ以下のフォルダとファイルすべての制御情報を一括して変更することも可能である。また、フォルダ 4 からフォルダ 2 にデータ「A A A」を移動したときに、すでにデータ「A A A」の持っている制御情報を維持することもできるし、フォルダ 2 の制御情報を与えなおすこともできる。本実施形態の場合、このような制御情報の継承に関するすべての設定は、位置管理手段 8 0 1 に対して設定できる。

【 0 0 6 5 】

上記実施形態のシステムによれば、位置管理手段は、属性保持手段に保持された属性情報と、データに設定された制御情報によって、複数のコンピュータによって共有された複数の記憶装置の中から適切な記憶装置を選択し、データを保持することができる。位置管理手段は、データを記録する複数の位置情報をもつことができるので、データを複数の記憶装置に記録することができ、ユーザの設定したさまざまな制御情報にあわせたデータの記録が可能になる。またデータの位置は、すべて位置管理手段によって一元的に管理されるので、データの移動や複写を行うたびに位置管理手段が複数のコンピュータにその処理を報告する必要がない。

【 0 0 6 6 】

次に、本発明の第 2 の実施形態（キャッシュありの場合）を説明する。

【 0 0 6 7 】

図 1 は、第 2 の実施形態のシステム構成図を示す。第 1 の実施形態のシステムとの違いは、コンピュータ 1 0 1 と 1 0 2 にキャッシュディスク 9 0 と情報受領手段 9 1 を設けている点と、コンピュータ 1 0 1 , 1 0 2 、ディスクアレイ装置 2 0 1 , 2 0 2 , 2 0 3 、およびライブラリ装置 3 0 1 に稼働率計測手段 8 2 を設けている点である。コンピュータ 1 0 1 と 1 0 2 に設けたキャッシュディスク 9 0 は、ファイルシステム 1 0 とファイバチャネルスイッチ 5 0 1 から 1 つの WWN を持つハードディスク装置として認識されるように構成されている。

【 0 0 6 8 】

次に、稼働率測定手段 8 2 について説明する。コンピュータに設けた稼働率測定手段 8 2 は、キャッシュディスク 9 0 の処理性能を測定する。また、ディスクアレイ装置 2 0 1 , 2 0 2 , 2 0 3 に設けた稼働率測定手段 8 2 は、ディスクアレイ装置の処理性能を測定する。本実施形態における処理性能とは、リードまたはライトにかかった平均処理時間である。ある記憶装置の平均処理時間が長くなると、その記憶装置にアクセスが集中していることが判る。ライブラリ装置 3 0 1 に設けた稼働率測定手段 8 2 は、メディアごとのアクセス頻度を測定する。

【 0 0 6 9 】

次に、位置管理手段 8 0 1 について説明する。第 2 の実施形態の位置管理手段 8 0 1 は、第 1 の実施形態と同様に図 2 の変換表 8 1 1 を用いる。領域管理表 8 1 2 は、ディスクアレイ装置 2 0 1 , 2 0 2 , 2 0 3 とライブラリ装置 3 0 1 のメディアに加えて、コンピュータ 1 0 1 と 1 0 2 に設けたキャッシュディスク 9 0 の空き領域も管理する。

【 0 0 7 0 】

次に、新規書き込み処理について説明する。図 9 と図 1 0 は、記憶装置にデータを新規に書き込む処理手順を示すフローチャートである。コンピュータ 1 0 1 のファイルシステム 1 0 は、イーサネット 6 0 を介して、ファイル名「A A A」というデータに対する新規領域割り当てをコンピュータ 1 5 1 の位置管理手段 8 0 1 に要求する（9 0 0 1）。位置管理手段 8 0 1 は、まず、変換表 8 1 1 の空いている一行を探し出し（9 0 0 2）、ライトビットを「1」にして書き込みの開始を宣言し（9 0 0 3）、ファイル名を書き込む（9 0 0 4）。ここでは、変換表 8 1 1 の 1 番目の位置情報に書き込んだとする。コンピュータ 1 0 1 のキャッシュディスク 9 0 から、データ「A A A」を格納するのに必要な領域を領域管理表 8 1 2 より検索する（9 0 0 5）。ここでは、十分な空き領域があるとすると、割り当て可能位置を獲得することができる。位置管理手段 8 0 1 は、対応する

領域管理表 8 1 2 の L B A を割り当て済みに変更し (9 0 0 6)、位置情報をコンピュータ 1 0 1 に報告する (9 0 0 7)。

【 0 0 7 1 】

コンピュータ 1 0 1 のファイルシステム 1 0 は、空き領域の報告を受けると、指定された空き領域にデータを書き込むためのライトコマンドをコンピュータ 1 0 1 のキャッシュディスク 9 0 に発行する (9 0 0 8)。この書き込みは同一コンピュータ内で処理されるため、きわめて高速である (9 0 0 9)。ファイルシステム 1 0 は、キャッシュディスク 9 0 から書き込み完了の報告 (9 0 1 0) を受け取ると、位置管理手段 8 0 1 に書き込みが完了したことを報告する (9 0 1 1)。位置管理手段 8 0 1 は、書き込み完了の報告を受けると、変換表 8 1 1 の 1 番目の位置情報に、コンピュータ 1 0 1 のキャッシュディスク 9 0 の WWN、L B A、およびデータ長を書き込み (9 0 1 2)、有効ビットを「 1 」にセットする (9 0 1 3)。

10

【 0 0 7 2 】

次に、位置管理手段 8 0 1 は、コンピュータ 1 0 1 以外のコンピュータ上のキャッシュディスク 9 0 とディスクアレイ装置 2 0 1、2 0 2、2 0 3 の稼働率測定手段 8 2 から、それぞれ稼働率を獲得し (9 0 1 4)、一番稼働率の低い記憶装置を選択し (9 0 1 5)、空き領域を検索する。ここでは、ディスクアレイ装置 2 0 1 の稼働率が一番低いとすると、ディスクアレイ装置 2 0 1 に対しデータ「 A A A 」を格納するのに必要な領域を検索する。データ「 A A A 」を格納できる空き領域が見つかり (9 0 1 6)、対応する領域管理表 8 1 2 の L B A を割り当て済みに変更し (9 0 1 7)、情報複製手段 8 9 0 にデータ「 A A A 」の複製を要求する (1 0 0 0 0)。

20

【 0 0 7 3 】

図 1 0 に、情報複製手段 8 9 0 による複製処理の手順を示す。情報複製手段 8 9 0 は、すでにデータ「 A A A 」が書き込まれているコンピュータ 1 0 1 のキャッシュディスク 9 0 に対して情報受領手段 9 1 を介してリードコマンドを発行し (1 0 0 0 1)、データ「 A A A 」を読み出し (1 0 0 0 2、1 0 0 0 3)、ディスクアレイ装置 2 0 1 にライトコマンドを発行し (1 0 0 0 4)、データを書き込む (1 0 0 0 5、1 0 0 0 6)。情報複製手段 8 9 0 は、ライトコマンドを完了すると、位置管理手段 8 0 1 に複製が完了したことを報告する (1 0 0 0 7)。

【 0 0 7 4 】

30

再び図 9 に戻って、位置管理手段 8 0 1 は、複製完了の報告を受け取ると、変換表 8 1 1 の空いている行、ここでは 2 番目の位置情報に、ディスクアレイ装置 2 0 1 の WWN、L B A、およびデータ長などの情報を書き込み (9 0 1 9)、有効ビットを「 1 」にセットする (9 0 2 0)。そして、ファイル履歴には、新規作成とその日時、複製日時を書き込む (9 0 2 1)。最後に、ライトビットを「 0 」にして (9 0 2 2)、書き込みを完了する (9 0 2 3、9 0 2 4、9 0 2 5)。

【 0 0 7 5 】

次に、読み出し処理について説明する。図 1 1 と図 1 2 は、記憶装置からデータを読み出す処理手順を示すフローチャートである。上記書き込み処理において、データ「 A A A 」は、ホストコンピュータ 1 0 1 のキャッシュディスク 9 0 と、ディスクアレイ装置 2 0 1 上にある。ここで、ホストコンピュータ 1 0 2 がデータ「 A A A 」を読み出す場合 (1 1 0 0 0)、ホストコンピュータ 1 0 2 のファイルシステム 1 0 が、イーサネット 6 0 を介してファイルネーム「 A A A 」のデータの読み出しを位置管理手段 8 0 1 に要求する (1 1 0 0 1)。位置管理手段 8 0 1 は、変換表 8 1 1 からファイルネーム「 A A A 」に対応する行を選択し (1 1 0 0 2)、ライトビットが立っていないことを確認し (1 1 0 0 3)、リードカウントをインクリメントする (1 1 0 0 4)。

40

【 0 0 7 6 】

次に、位置管理手段 8 0 1 は、データ「 A A A 」の格納されているホストコンピュータ 1 0 1 のキャッシュディスク 9 0 とディスクアレイ装置 2 0 1 の稼働率計測手段 8 2 から稼働率を獲得し (1 1 0 0 6)、一番稼働率の低い記憶装置を選択する (1 1 0 0 7)。こ

50

ここでは、コンピュータ101のキャッシュディスク90の稼働率が低かったとする。位置管理手段801は、領域管理表812のコンピュータ102のキャッシュディスク90からデータ「AAA」を格納するのに必要な領域を検索する(11008)。ここでは、十分な空き領域があるとすると、割り当て可能位置を獲得することができる。位置管理手段801は、対応する領域管理表812のLBAを割り当て済みに変更し(11009)、情報複製手段890に対し、コンピュータ101のキャッシュディスク90からコンピュータ102のキャッシュディスク90にデータ「AAA」をコピーするように要求する(12000)。

【0077】

図12に、情報複製手段890による複製処理の手順を示す。情報複製手段890は、すでにデータ「AAA」が書き込まれているコンピュータ101のキャッシュディスク90に対してリードコマンドを発行し(12001)、データを読み出し(12002, 12003)、コンピュータ102のキャッシュディスク90にライトコマンドを発行し(12004)、データを書き込む(12005, 12006)。情報複製手段890は、ライトコマンドを完了すると、位置管理手段801に複製が完了したことを報告する(12007)。

【0078】

再び図11に戻って、位置管理手段801は、情報複製手段890から複製の完了を受け取ると、変換表811の空いている行、ここでは3番目の位置情報にホストコンピュータ102のキャッシュディスク90のWWN、LBA、およびデータ長などの情報を書き込み(11011)、有効ビットを「1」にセットする(11012)。また、位置管理手段801は、ファイル履歴に複製日時を書き込む(11013)。そしてデータ「AAA」の入っているホストコンピュータ102のキャッシュディスク90の位置情報をコンピュータ102に報告する(11014)。

【0079】

コンピュータ102のファイルシステム10は、位置情報(WWN、LBA、データ長)の報告を受けると、コンピュータ102のキャッシュディスク90にリードコマンドを発行し(11015)、データを読み出す(11016)。コンピュータ102のファイルシステム10は、キャッシュディスク90から読み出し完了の報告(11017)を受け取ると、コンピュータ151の位置管理手段801に読み出しが完了したことを報告する(11018)。位置管理手段801は、読み出し完了の報告を受け取ると、ファイル履歴に最終アクセス日時を書き込み(11019)、リードカウントをデクリメントし(11020)、読み出しが完了する(11021, 11022, 11023)。

【0080】

次に、データの更新処理について説明する。図13と図14は、データを上書きし更新する処理手順を示すフローチャートである。上記書き込み処理と読み出し処理において、データ「AAA」は、ホストコンピュータ101と102のキャッシュディスク90と、ディスクアレイ装置201上の3箇所にある。ここで、ホストコンピュータ102がデータ「AAA」を更新する場合(13000)、ホストコンピュータ102のファイルシステム10は、イーサネット60を介して、ファイル名「AAA」のデータの更新を位置管理手段801に要求する(13001)。位置管理手段801は、変換表811からファイル名「AAA」に対応する行を選択し(13002)、ライトビットが「0」でリードカウントが「0」であることを確認する(13003, 13004)。ライトビットが「0」でない場合、またはリードカウントが「0」でない場合は、更新失敗となる(13090, 13091, 13098, 13099)。

【0081】

次に、位置管理手段801は、ライトビットを「1」にして更新の開始を宣言する(13005)。もしデータ「AAA」のデータ長が大きくなっていた場合には、領域管理表812を用いて再度空き領域を検索する必要があるが(13006, 13007)、ここではデータ長が不変とする。位置管理手段801は、コンピュータ102のファイルシステ

10

20

30

40

50

ム 1 0 に、変換表 8 1 1 の 3 番目の位置情報 (コンピュータ 1 0 2 のキャッシュディスク 9 0 内のデータ「 A A A 」を示す) を報告する (1 3 0 0 8) 。

【 0 0 8 2 】

コンピュータ 1 0 2 のファイルシステム 1 0 は、位置情報の報告を受けると、指定された位置にデータを書き込むためのライトコマンドをコンピュータ 1 0 2 のキャッシュディスク 9 0 に発行する (1 3 0 0 9) 。コンピュータ 1 0 2 のキャッシュディスク 9 0 は、データを書き込み (1 3 0 1 0) 、書き込み完了を報告する (1 3 0 1 1) 。ファイルシステム 1 0 は、キャッシュディスク 9 0 から書き込み完了の報告を受け取ると、コンピュータ 1 5 1 の位置管理手段 8 0 1 に書き込みが完了したことを報告する (1 3 0 1 2) 。

【 0 0 8 3 】

位置管理手段 8 0 1 は、書き込み完了の報告を受け取ると、変換表 8 1 1 の 1 番目 (コンピュータ 1 0 1 のキャッシュディスク 9 0 内のデータ「 A A A 」を示す) と 2 番目 (ディスクアレイ 2 0 1 内のデータ「 A A A 」を示す) の位置情報に、更新したデータ「 A A A 」を複製するように、情報複製手段 8 9 0 に要求する (1 4 0 0 0) 。

【 0 0 8 4 】

図 1 4 に、情報複製手段 8 9 0 による複製処理の手順を示す。情報複製手段 8 9 0 は、すでにデータ「 A A A 」が書き込まれているコンピュータ 1 0 2 のキャッシュディスク 9 0 に対してリードコマンドを発行し (1 4 0 0 1) 、データを読み出し (1 4 0 0 2 , 1 4 0 0 3) 、コンピュータ 1 0 1 のキャッシュディスク 9 0 とディスクアレイ装置 2 0 1 にライトコマンドを同時に発行し (1 4 0 0 4) 、データを書き込む (1 4 0 0 5 , 1 4 0 0 6 , 1 4 0 0 5 ' , 1 4 0 0 6 ') 。情報複製手段 8 9 0 は、二つのライトコマンドを完了すると、位置管理手段 8 0 1 に複製が完了したことを報告する (1 4 0 0 7) 。本実施形態における情報複製手段 8 9 0 は、ファイバチャネルスイッチ 5 0 1 と連携し、1つのデータを複数の記憶装置に同時に複製することができる。もし、情報複製手段 8 9 0 が、1つの記憶装置にしかデータを複製できない場合には2回に分けて処理する。この場合、処理時間が長くなることが予想されるので、コピー先が多くなればなるほど不利になる。本発明の実施には、複数の記憶装置にデータを複製可能な情報複製手段 8 9 0 を用いることが望ましい。

【 0 0 8 5 】

再び図 1 3 に戻って、位置管理手段 8 0 1 は、情報複製手段 8 9 0 から複製完了の報告を受け取ると、ファイル履歴に最終更新日時を書き込み (1 3 0 1 4) 、ライトビットを「 0 」にして (1 3 0 1 5) 、書き込みを完了する (1 3 0 1 6 , 1 3 0 1 7 , 1 3 0 1 8) 。

【 0 0 8 6 】

上記第 2 の実施の形態のシステムによれば、ホストコンピュータにディスクキャッシュを設けることで、ホストコンピュータ上に書き込みして間もないデータを保持できるので、高速にアクセスできるようになる。位置管理手段は、各記憶装置に設けた稼働率測定手段からアクセス頻度の少ない記憶装置を選択してデータを 2 箇所以上に保存するので、ホストコンピュータが停止するようなことがあってもデータを失うことはない。

【 0 0 8 7 】

次に、本発明の第 3 の実施形態 (ストレージプロバイダにおける保存データの課金方式) について説明する。

【 0 0 8 8 】

図 1 は、第 3 の実施形態のシステム構成図を示す。第 3 の実施形態では、イーサネット 6 1 に複数のクライアント 7 0 を接続している。コンピュータ 1 0 1 と 1 0 2 は、これらクライアント 7 0 からのデータの書き込みおよび読み出し要求を受け付けるファイルサーバとして機能する。本実施形態ではクライアント 7 0 との接続にイーサネット 6 1 を用いて説明を行うが、本発明はこの接続手段によらず実現可能である。

【 0 0 8 9 】

本実施形態ではコンピュータ 1 0 1 と 1 0 2 はストレージサービスプロバイダ側に設置さ

10

20

30

40

50

れているとして説明を行うが、もちろんコンピュータ101や102を各事業所や支店に設置することもできる。もし、コンピュータ102とファイバチャネルスイッチ501との距離が10km以上離れる場合には、図1に記載のように、エクステンダ52を介し、ATM (Asynchronous Transfer Mode) 51を介して接続する。もちろんATM 51に代えて他の接続手段を用いても問題ない。

【0090】

次に、運用形態について説明する。クライアント70はたとえば各事業所や支店などに用意された端末であり、ユーザは、これらの端末から、保存を希望するデータをイーサネット61を介してコンピュータ101または102に送りつける。この際、データをどのように保存するかを指定する（データ毎に指定する保存の仕方をデータポリシーと称す）。第1の実施形態に記載した制御情報がこれにあたる。ユーザは、アクセス頻度の高いデータはより高速な記憶装置に、アクセス頻度の低いデータはアクセスが遅くても低コストな記憶装置に、また、大切なデータは、コストがかかっても災害でデータを失うことのないように遠隔地の記憶装置にも複製を保持することを希望する。よって、データの保持される記憶装置の属性（性能や信頼性）が、データの保持コストに反映されることが望ましい。

【0091】

次に、課金について説明する。本実施形態のシステムでは、記憶装置毎の属性保持手段80に保持した単位容量あたりの保持コストと、位置管理情報の変換表811に保持したデータ長、ファイル履歴、および位置情報をもとに、保持データに対して課金する。たとえば、データ「AAA」がディスクアレイ装置201と202に10日間保持され、次にディスクアレイ装置203に移動されて10日間保存され、次にライブラリ装置301に移動されて10日間保存された場合、この30日間のデータ保持コストは、

$$\begin{aligned} \text{データ保持コスト} = & \{ (\text{RAID0の保持コスト10}) \times 10\text{日} \times 2 \\ & + (\text{RAID5の保持コスト15}) \times 10\text{日} \\ & + (\text{ライブラリの保持コスト1}) \times 10\text{日} \} \times \text{データ長} \end{aligned}$$

で求めることができる。ここで、RAID0の保持コストが2倍されているのは、2台のディスクアレイにデータが二重に保持されていたことを反映するためである。位置管理手段801は、一ヶ月ごとにすべてのデータのコストを算出し、ユーザに請求する。コスト算出後、位置管理手段801は、ファイル履歴のうち、データの移動と複製に関する情報は消去する。上記コストの算出に必要な情報は、第1の実施形態および第2の実施形態に含まれている。

【0092】

また、別の例として、データ「BBB」が、2枚のDVD-RAMメディアに30日間保持され、一度参照したために、ディスクアレイ装置201に3日間キャッシュされた場合、この30日間のデータ保持コストは、

$$\begin{aligned} \text{データ保持コスト} = & \{ (\text{RAID0の保持コスト10}) \times 3\text{日} \\ & + (\text{ライブラリの保持コスト1}) \times 30\text{日} \times 2 \} \times \text{データ長} \end{aligned}$$

で求めることができる。

【0093】

ユーザ毎のデータコストを算出するには、WWNを用いる。たとえば、ユーザAがコンピュータ101を介してデータを保存し、ユーザBがホストコンピュータ102を介してデータを保持した場合、ホストコンピュータ101と102は異なるWWNを有するので、目的のホストコンピュータのWWNで変換表811を検索すれば、そのホストコンピュータを使用するユーザの総データ保持コスト算出することができる。

【0094】

次に、グローバルポリシーの適用について説明する。上述したように本実施形態のシステ

ムでは、データ単位でのコストを容易に算出することが出来る。そこで、ユーザからたとえば一ヶ月あたりのデータ保持コストが指定されていた場合には、そのコストを満足できるように、データを移動させる。当然ながらデータ量が多くなると性能が犠牲になる。

【0095】

このような指定を、データポリシーに対して、グローバルポリシーと称す。コスト以外にも性能最優先のグローバルポリシーなどが考えられる。また、データポリシーとグローバルポリシーのどちらを優先するかも指定可能である。もし、データポリシーが優先されていて、ユーザから指定されたデータ保持コスト（グローバルポリシー）を満たすことが出来ない場合には、その旨をユーザに報告する。通常、グローバルポリシーの設定は、ユーザ側のクライアント70からホストコンピュータ101または102を介してコンピュータ151に通知するが、ユーザとストレージサービスプロバイダが書面により交換したグローバルポリシーの設定基準に基づき、ストレージサービスプロバイダ側の管理者がコンピュータ151に設定することもできる。

10

【0096】

上記第3の実施形態のシステムによれば、ユーザから保持要求のあったデータを、ユーザの指定した制御情報と、属性保持手段、位置管理手段、および情報複製手段とにより、よりきめ細かく位置管理することができる。これにより、データのアクティビティや信頼性に対し、データ保持にかかるコストを関連付けることができ、ユーザは、高いアクティビティや高い信頼性を必要とするデータに対しては高い保持コストを、低いアクティビティや低い信頼性のデータには安い保持コストを割り当てることができる。一般的に、高いアクティビティを必要とするデータは少ないと言えるが、本実施形態のシステムでは、上述したようにデータのアクティビティや信頼性に応じたコストを関連付けることができるので、全体としてユーザはデータを保持しつづけるコストを低減できる。

20

【0097】

【発明の効果】

以上説明したように、本発明によれば、位置管理手段が、ストレージの稼働率、データの利用状況、制御情報、および課金情報などに基づいてデータを移動および複製することで、ユーザの求める性能や信頼性、保存性、およびコストに対応したデータの保存が可能になる。また、複数のコンピュータ、複数のストレージのデータの位置管理を位置管理手段で集中して行うことで、データを移動したことを各コンピュータに通知する必要がない。また、情報複製手段により、同時に複数のデータの複製を可能とすることで、コンピュータの台数が多い場合にも、複製にかかる時間を低減することができる。

30

【図面の簡単な説明】

【図1】本発明の実施形態を示すシステム構成図である。

【図2】変換表を示す図である。

【図3】領域管理表を示す図である。

【図4】ディレクトリ構成を示す図である。

【図5】第1の実施形態における新規書き込み処理を示すフローチャート図である。

【図6】第1の実施形態における読み出し処理を示すフローチャート図である。

【図7】第1の実施形態における位置管理手段の監視と移動処理を示すフローチャート図である。

40

【図8】複製処理を示すフローチャート図である。

【図9】第2の実施形態における新規書き込み処理を示すフローチャート図である。

【図10】複製処理を示すフローチャート図である。

【図11】第2の実施形態における読み出し処理を示すフローチャート図である。

【図12】複製処理を示すフローチャート図である。

【図13】第2の実施形態におけるデータの更新処理を示すフローチャート図である。

【図14】複製処理を示すフローチャート図である。

【図15】第1の実施形態においてライブラリ装置に格納されたデータを読み出す処理を示すフローチャート図である。

50

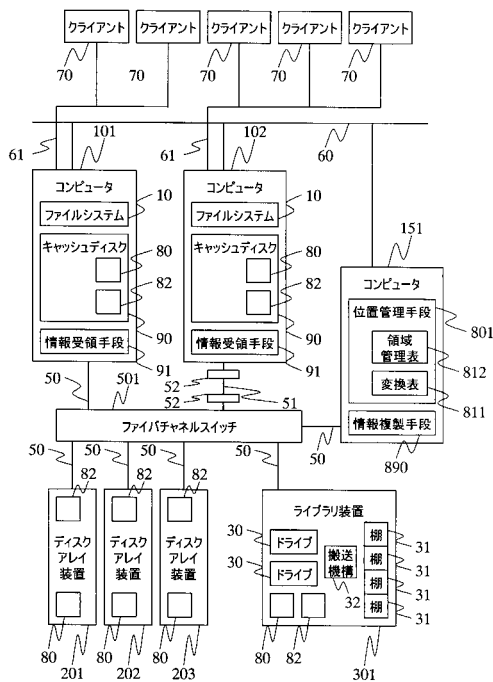
【図 1 6】複製処理を示すフローチャート図である。

【符号の説明】

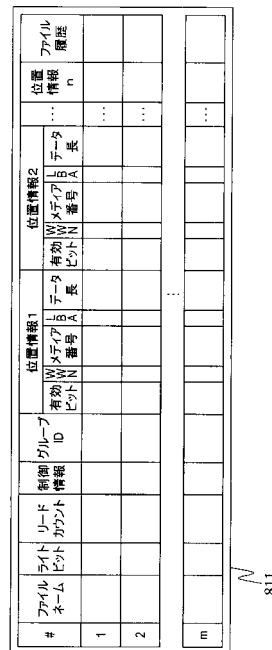
1 0 ... ファイルシステム、3 0 ... ドライブ、3 1 ... 棚、3 2 ... 搬送機構、5 0 ... ファイバ
チャネル、5 1 ... A T M、5 2 ... エクステンダ、6 0 ... イーサネット、6 1 ... イーサネッ
ト、7 0 ... クライアント、8 0 ... 属性保持手段、8 2 ... 稼働率測定手段、9 0 ... キャッシ
ュディスク、9 1 ... 情報受領手段、1 0 1 ... コンピュータ、1 0 2 ... コンピュータ、1 5
1 ... コンピュータ、2 0 1 ... ディスクアレイ装置、2 0 2 ... ディスクアレイ装置、2 0 3
... ディスクアレイ装置、3 0 1 ... ライブラリ装置、5 0 1 ... ファイバチャネルスイッチ、
8 0 1 ... 位置管理手段、8 1 1 ... 変換表、8 1 2 ... 領域管理手段、8 9 0 ... 情報複製手段
。

10

【図 1】



【図 2】

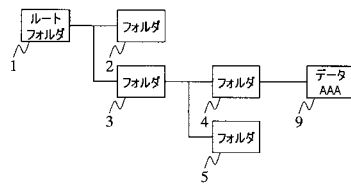


【 図 3 】

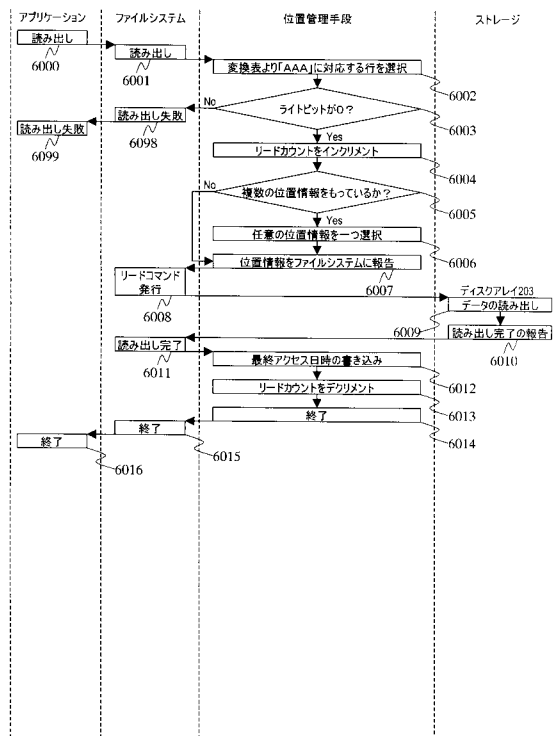
[illegible]

812

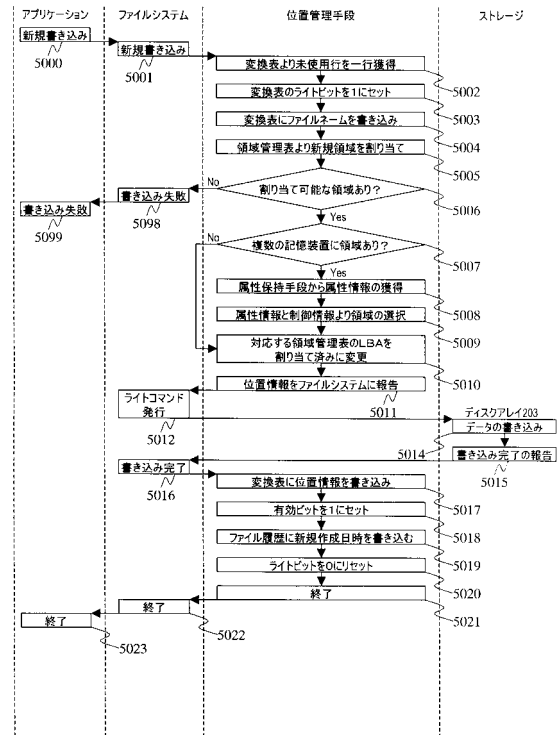
【 図 4 】



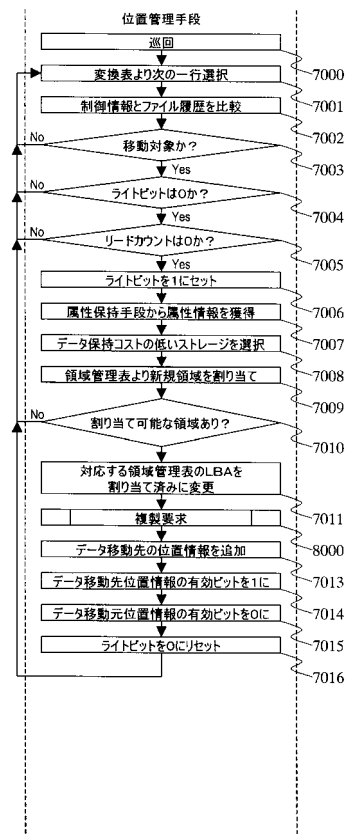
【 図 6 】



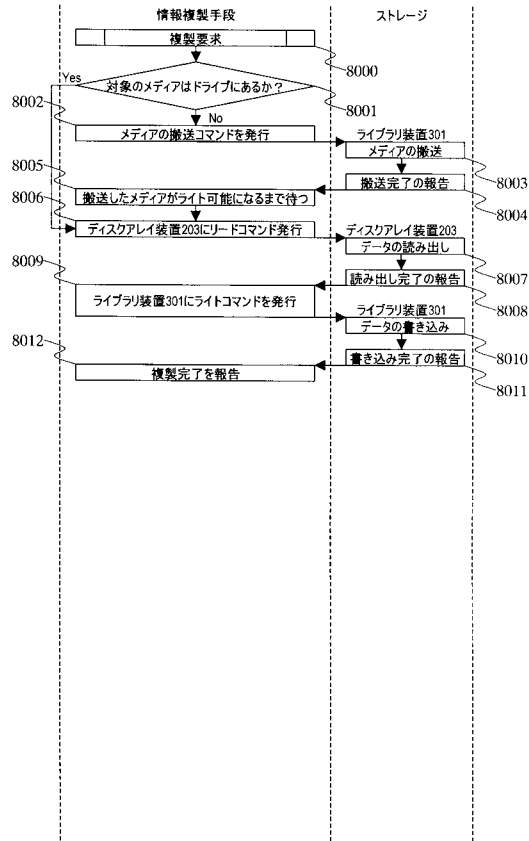
【 図 5 】



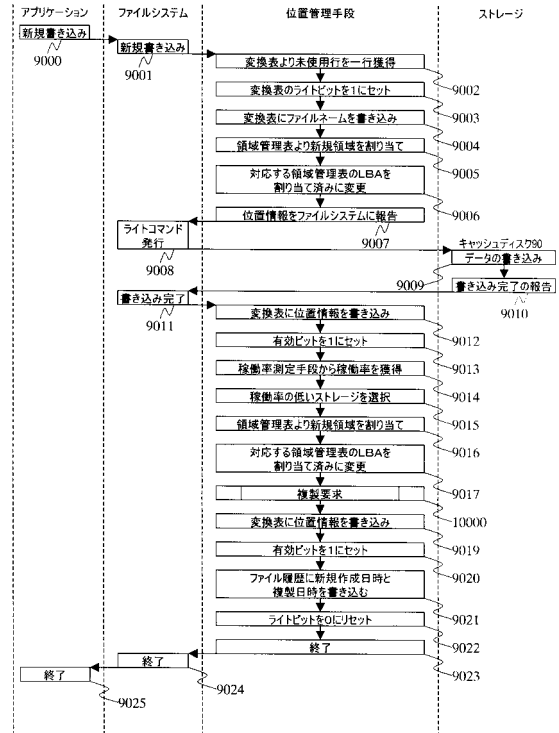
【圖 7】



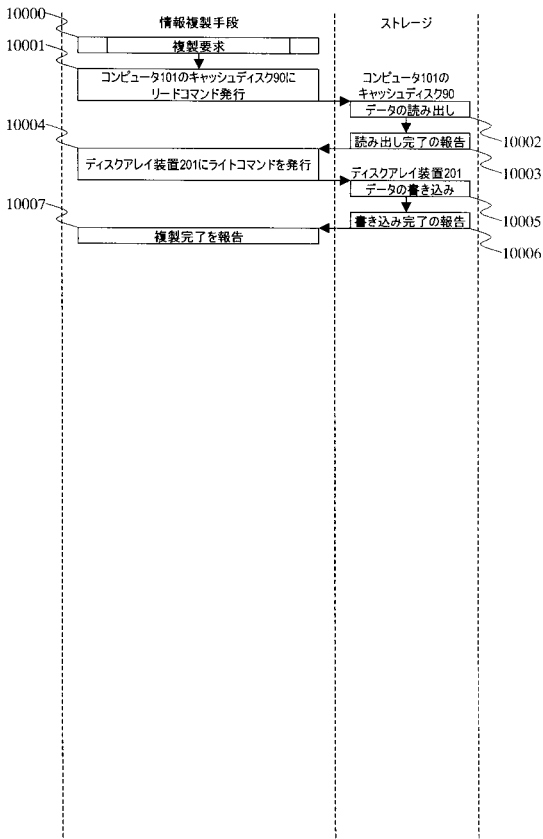
【図 8】



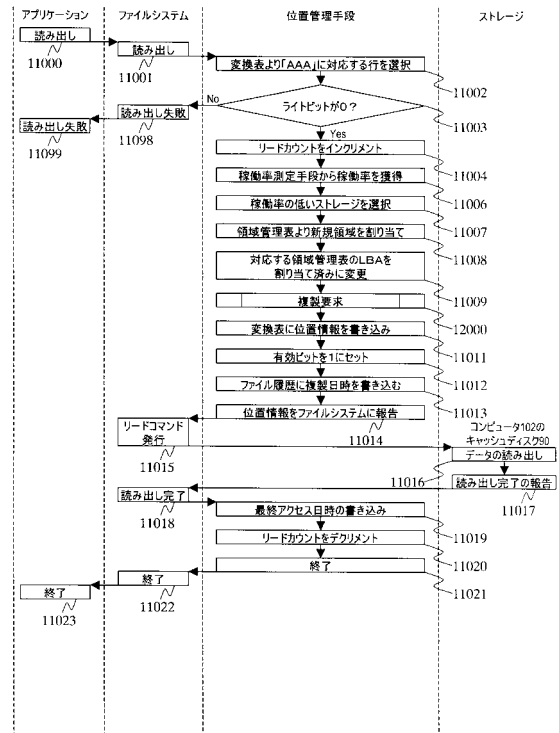
【図 9】



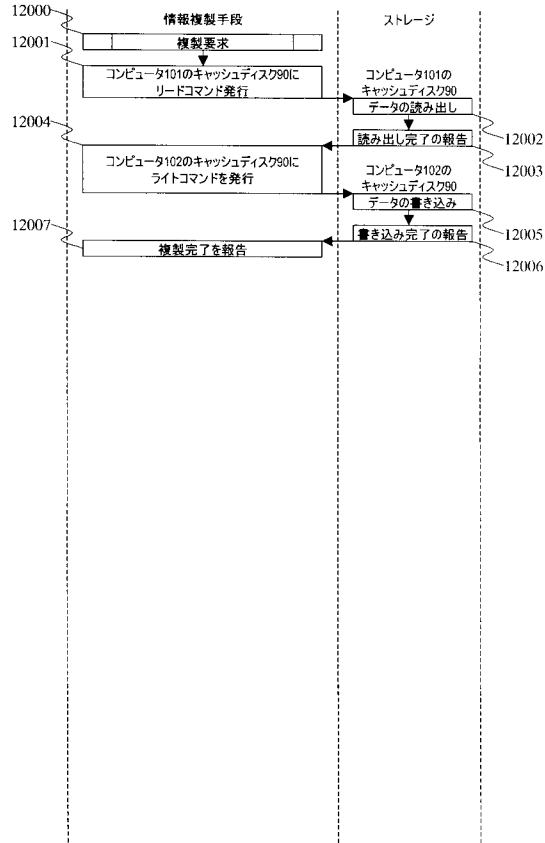
【図 10】



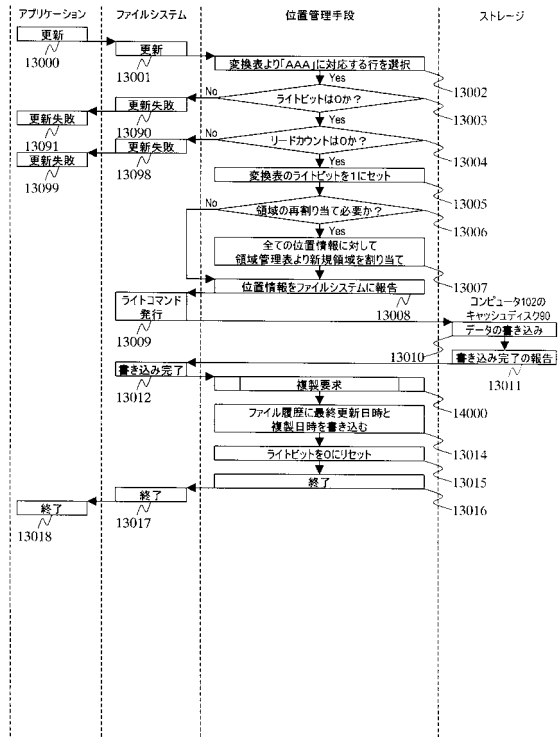
【図 11】



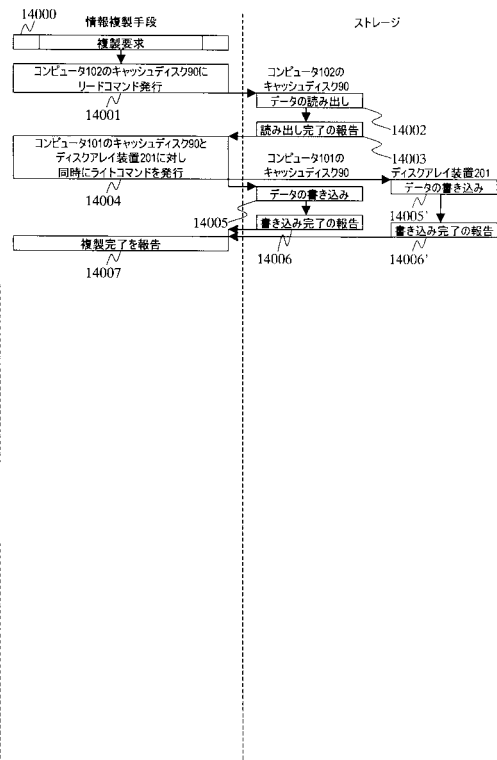
【図 12】



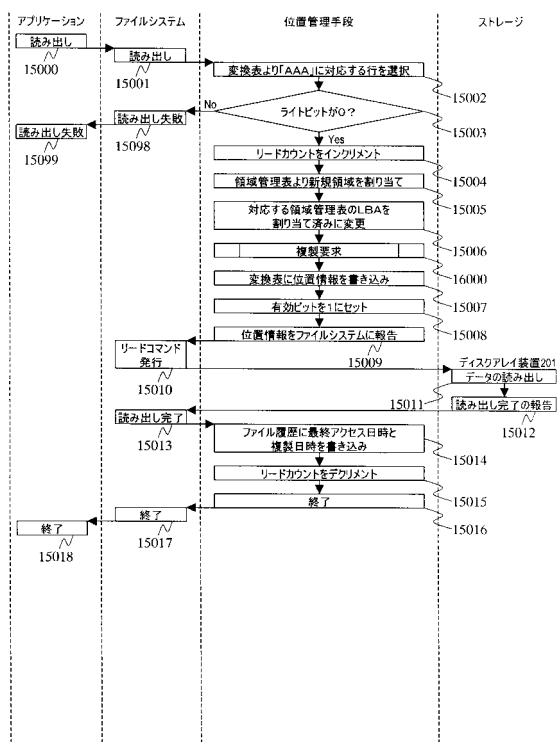
【図 13】



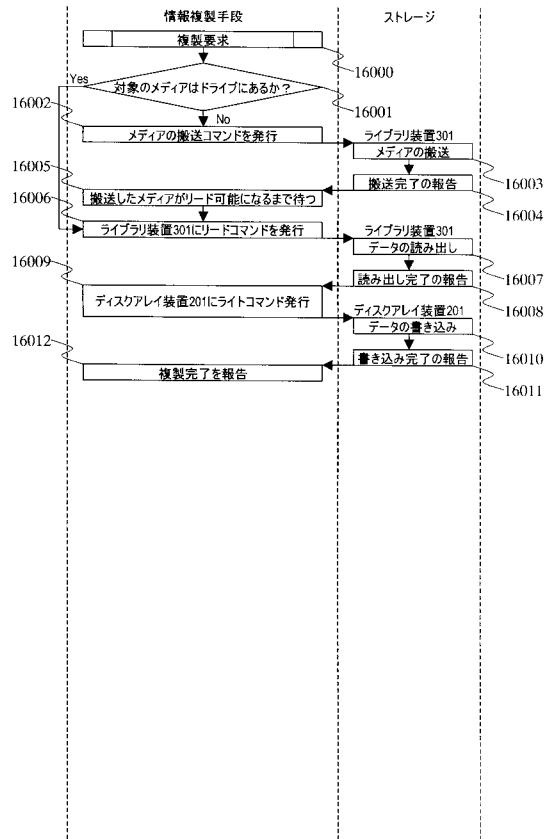
【図 14】



【図 15】



【図 16】



フロントページの続き

- (72)発明者 江口 賢哲
神奈川県川崎市麻生区王禅寺 1 0 9 9 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 茂木 和彦
神奈川県川崎市麻生区王禅寺 1 0 9 9 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
神奈川県小田原市国府津 2 8 8 0 番地 株式会社 日立製作所 ストレージシステム事業部内

審査官 浜岸 広明

- (56)参考文献 特開平 0 3 - 1 1 1 9 4 8 (J P , A)
特開平 0 6 - 0 4 4 1 1 1 (J P , A)
特開平 0 1 - 3 0 5 4 1 9 (J P , A)

- (58)調査した分野(Int.Cl. , D B 名)
G06F 12/00