

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4080227号  
(P4080227)

(45) 発行日 平成20年4月23日 (2008. 4. 23)

(24) 登録日 平成20年2月15日 (2008. 2. 15)

(51) Int. Cl.

F 1

G 0 6 F 12/14 (2006. 01)

G 0 6 F 12/14 5 1 0 D

G 0 6 F 3/06 (2006. 01)

G 0 6 F 3/06 3 0 5 A

G 0 6 F 13/00 (2006. 01)

G 0 6 F 13/00 3 0 1 E

請求項の数 13 (全 16 頁)

(21) 出願番号 特願2002-92685 (P2002-92685)  
 (22) 出願日 平成14年3月28日 (2002. 3. 28)  
 (65) 公開番号 特開2003-288270 (P2003-288270A)  
 (43) 公開日 平成15年10月10日 (2003. 10. 10)  
 審査請求日 平成17年2月4日 (2005. 2. 4)

(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 110000176  
 一色国際特許業務法人  
 (74) 代理人 100094042  
 弁理士 鈴木 知  
 (72) 発明者 大橋 和伸  
 神奈川県小田原市中里322番地2号 株  
 式会社日立製作所 R A I Dシステム事業  
 部内  
 (72) 発明者 佐藤 孝夫  
 神奈川県小田原市中里322番地2号 株  
 式会社日立製作所 R A I Dシステム事業  
 部内

最終頁に続く

(54) 【発明の名称】 データ検証方法およびディスクアレイ装置

(57) 【特許請求の範囲】

【請求項 1】

ホストコンピュータと、このホストコンピュータと通信手段を介して接続し、前記ホストコンピュータから送信された入出力要求を受信してこの要求に対応するデータ入出力処理を記憶手段に対して実行する記憶制御装置と、を含んで構成されるストレージシステムにおけるデータ検証方法であって、

ホストコンピュータが、記憶制御装置に記憶するデータを規定の仕様に従って編成し管理するアプリケーションプログラムを実行する工程と、

ホストコンピュータが、前記アプリケーションプログラムが使用する前記記憶手段の記憶領域の範囲を特定する領域管理データを記憶制御装置に送信する工程と、

記憶制御装置が、前記領域管理データを受信する工程と、

記憶制御装置が、前記アプリケーションプログラムの処理に起因してホストコンピュータから送信されるデータ入出力要求を受信して、そのデータ入出力要求の処理対象となるデータのうち前記領域管理データに基づいて特定される前記アプリケーションプログラムが使用する記憶領域に対して入出力されるデータについて、そのデータが前記規定の仕様に従って編成されているかどうかを検証する工程と、

を有することを特徴とするストレージシステムにおけるデータ検証方法。

【請求項 2】

前記記憶制御装置が、前記ホストコンピュータで稼働する複数の前記アプリケーションプログラムのそれぞれについての前記検証を行う複数のプログラムを記憶する工程と、

10

20

前記記憶制御装置が、前記ホストコンピュータから送信されてくる前記領域管理データに基づいて前記各アプリケーションプログラムのそれぞれが使用する記憶領域を特定する工程と、

前記記憶制御装置が、前記各アプリケーションプログラムの処理に起因してホストコンピュータから送信されるデータ入出力要求を受信して、前記領域管理データに基づいて前記データ入出力要求に関わる前記アプリケーションプログラムを特定する工程と、

前記記憶制御装置が、前記データ入出力要求の処理対象となるデータのうち前記領域管理データで特定される記憶領域に入出力されるデータについて、そのデータが前記規定の仕様に従って編成されているかを、前記特定したアプリケーションプログラムに対応する前記プログラムにより検証する工程と、

10

を有することを特徴とする請求項 1 に記載のストレージシステムにおけるデータ検証方法。

【請求項 3】

前記ホストコンピュータが、前記プログラムを前記ホストコンピュータから前記記憶制御装置に送信する工程と、

前記記憶制御装置が前記プログラムを受信して記憶する工程と、

を有することを特徴とする請求項 2 に記載のストレージシステムにおけるデータ検証方法。

【請求項 4】

前記ホストコンピュータが、前記範囲が変更された場合に、その変更内容が反映された前記領域管理データを前記記憶制御装置に送信する工程を有することを特徴とする請求項 1 に記載のストレージシステムにおけるデータ検証方法。

20

【請求項 5】

前記アプリケーションプログラムが使用する前記記憶手段の記憶領域は、前記ホストコンピュータ上で稼働する前記アプリケーションプログラムとは異なるソフトウェアにより付加される制御情報の使用領域を除いた領域であることを特徴とする請求項 1 に記載のストレージシステムにおけるデータ検証方法。

【請求項 6】

前記領域管理データには、前記記憶手段の前記記憶領域を用いて論理的に編成されたボリュームを単位として前記アプリケーションプログラムに提供される前記記憶領域の範囲を指定するデータが含まれることを特徴とする請求項 1 に記載のストレージシステムにおけるデータ検証方法。

30

【請求項 7】

前記ボリュームは、前記記憶手段の前記記憶領域をRAID方式で運用することにより編成されることを特徴とする請求項 6 に記載のストレージシステムにおけるデータ検証方法。

【請求項 8】

前記記憶手段は、前記記憶制御装置に一体に設けられることを特徴とする請求項 1 に記載のストレージシステムにおけるデータ検証方法。

【請求項 9】

前記記憶領域は、ディスクドライブなどの物理的な記憶デバイスにより提供される記憶領域上に編成される論理的な記憶デバイスとして提供されることを特徴とする請求項 1 に記載のストレージシステムにおけるデータ検証方法。

40

【請求項 10】

前記通信手段は、SCSI規格に準拠した通信方式によるものであることを特徴とする請求項 1 に記載のストレージシステムにおけるデータ検証方法。

【請求項 11】

前記通信手段は、SANであることを特徴とする請求項 1 に記載のストレージシステムにおけるデータ検証方法。

【請求項 12】

前記通信手段は、LANであることを特徴とする請求項 1 に記載のストレージシステム

50

におけるデータ検証方法。

【請求項 13】

ホストコンピュータと通信手段を介して接続し、前記ホストコンピュータから送信される入出力要求を受信して、前記入出力要求に対応するデータ入出力処理をディスクドライブに対して実行するディスクアレイ装置であって、

前記ディスクドライブが提供する記憶領域上に論理デバイスが編成され、

前記ホストコンピュータで動作するアプリケーションプログラムのIDであるアプリケーションIDと、前記アプリケーションプログラムが使用する前記論理デバイスのIDである論理デバイスIDとを対応付ける領域管理テーブルを前記ホストコンピュータから受信し、受信した前記領域管理テーブルを前記ディスクドライブに記憶し、

前記アプリケーションIDに対応付けて、前記論理デバイスに記憶されているデータが、前記アプリケーションプログラムに規定の仕様に従って編成されているかどうかを検証するデータ検証用プログラムを前記ディスクドライブに記憶し、

前記アプリケーションプログラムの処理に起因して前記ホストコンピュータから送信されるデータ入出力要求を受信し、

受信した前記データ入出力要求にセットされている前記論理デバイスIDに対応づけられているアプリケーションIDを調べ、調べた前記アプリケーションIDに対応する前記データ検証用プログラムを起動すること、

を特徴とするディスクアレイ装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

この発明は、データ検証方法およびディスクアレイ装置に関する。

【0002】

【従来の技術】

ストレージの集約的運用などを目的として、ホストコンピュータとディスクアレイ装置などの記憶制御装置とを、SAN (Storage Area Network) などのネットワークで結んだ形態のストレージシステムが注目されており、IDC (Internet Data Center) などの多くの場所でこのような形態での運用が行われている。また、最近ではネットワークにLAN (Local Area Network) を利用し、ファイルシステムを搭載したNAS (Network Attached Storage) と呼ばれる記憶制御装置を用いるものも注目されている。

【0003】

このような形態で運用されるストレージシステムにおいて、ホストコンピュータ上で稼働する、例えば、データベースソフトウェアなどのアプリケーションプログラムなどは、データが規定の仕様を具備していること（以下、「適合性」と呼ぶ）を要求するものが少なくない。そこで、アプリケーションプログラムの中には、処理の対象となるデータの適合性を検証するアルゴリズムを備えるものがある。

【0004】

【発明が解決しようとする課題】

ところで、アプリケーションプログラムが前記のようなアルゴリズムを備えていたとしても、ホストコンピュータと記憶制御装置とを結ぶ、例えば、前述したSANなどの通信路や、記憶制御装置内部の処理によりデータの適合性が失われてしまうことがある。また、記憶制御装置は、複数のホストコンピュータに共用されていることも多く、この場合、他のホストコンピュータ上で動作するアプリケーションプログラムが誤って他のホストコンピュータ上で動作するアプリケーションプログラムのデータにアクセスしてしまい、データの適合性が損なわれてしまうということもある。

【0005】

ここで、例えば、記憶制御装置への書き込み対象のデータについての適合性が、このようにアプリケーションプログラムが関知しないところで損なわれてしまったような場合には、記憶制御装置側では適合性が失われているかどうかを認知することができず、そのデー

10

20

30

40

50

タはそのまま記憶制御装置側に記憶されてしまうことになる。そして、この場合、アプリケーションプログラムはそのデータを読み込んだ際に初めて適合性が失われていることを事後的に認知することになり、そのデータの書き込み時から読み出し時までのタイムラグが長い場合には、修復が困難となることもある。

【 0 0 0 6 】

また、記憶制御装置側でバックアップ運用が行われている場合でも、バックアップされているデータが、既に適合性が損なわれているデータであった場合には、データを修復することもできない。さらに、このようにアプリケーションプログラムが関知しない範囲でデータの適合性が失われてしまった場合には、事後的な原因追求が難しいことが多く、障害復旧等の作業を困難にさせる要因ともなる。

10

【 0 0 0 7 】

そこで、記憶制御装置側でデータ検証を行うようにすることが考えられるが、一般にストレージシステムにおいては、アプリケーションプログラムが記憶制御装置に記憶するデータには、ホストコンピュータ上で稼働するオペレーティングシステム（以下、「OS」と記載する）が付加する情報やホストコンピュータ上で論理的に編成されたボリュームに関する情報などが付帯されている。また、アプリケーションプログラムが記憶制御装置に記憶するデータは、OSやボリューム管理プログラムの都合などにより、ホストコンピュータ上でアプリケーションプログラムが参照する配置とは異なった配置で記憶されているが、記憶制御装置側では、前述のような情報や配置を知ることができない。つまり、現状のストレージシステムの仕組みでは、アプリケーションプログラムが行っているデータ検証を、記憶制御装置側では行うことができず、実施する場合には記憶制御装置は必要な情報を、アプリケーションプログラム及びOSやボリューム管理プログラムから取得する仕組みが必要である。

20

【 0 0 0 8 】

ところで、例えば、特開平 8 - 2 6 3 2 2 3 号公報には、アプリケーションプログラムから磁気ディスクに対するデータの読み出し書き込みの際に行う R A S (Reliability Availability Serviceability)、すなわちデータの信頼性向上のためのデータチェックを容易に行えるようにする仕組みとして、アプリケーションプログラムからの I / O を契機として C P U が R A S データを作成し、これを磁気ディスクに書き込もうとするデータに付帯させ、これにより磁気ディスク入出力チャネルおよび磁気ディスクコントローラがデータチェックを行うようにした仕組みが開示されている。

30

【 0 0 0 9 】

しかしながら、この公報の仕組みでは、前記の R A S データはあくまで C P U によって独自に作成されてデータに付加されるものであるし、また、R A S データはアプリケーションプログラムが要求するデータの適合性に関するものではない。また、前記の公報の仕組みは、このような情報や配置を認知するものでもなく、ホストコンピュータ上で稼働する個々のアプリケーションプログラムが要求する前記の適合性を検証するものでも無い。

【 0 0 1 0 】

この発明は、以上の観点からなされたもので、ストレージシステムにおいて、データの適合性を、より確実に検証し確保することができる、データ検証方法およびディスクレイ装置を提供することを目的とする。

40

【 0 0 1 1 】

【課題を解決するための手段】

この目的を達成する、本発明の主たる発明は、ホストコンピュータと、このホストコンピュータと通信手段を介して接続し、前記ホストコンピュータから送信された入出力要求を受信してこの要求に対応するデータ入出力処理を記憶手段に対して実行する記憶制御装置と、を含んで構成されるストレージシステムにおけるデータ検証方法であって、ホストコンピュータが、記憶制御装置に記憶するデータを規定の仕様に従って編成し管理するアプリケーションプログラムを実行する工程と、ホストコンピュータが、前記アプリケーションプログラムが使用する前記記憶手段の記憶

50

領域の範囲を特定する領域管理データを記憶制御装置に送信する工程と、  
記憶制御装置が、前記領域管理データを受信する工程と、  
記憶制御装置が、前記アプリケーションプログラムの処理に起因してホストコンピュータから送信されるデータ入出力要求を受信して、そのデータ入出力要求の処理対象となるデータのうち前記領域指定データに基づいて特定される前記アプリケーションプログラムが使用する記憶領域に対して入出力されるデータについて、そのデータが前記規定の仕様に従って編成されているかどうかを検証する工程と、  
を有することを特徴とする。

【 0 0 1 2 】

【 発明の実施の形態 】

= = = 開示の概要 = = =

以下の開示により、少なくとも次のことが明らかにされる。

ホストコンピュータと、このホストコンピュータと通信手段を介して接続し、前記ホストコンピュータから送信された入出力要求を受信してこの要求に対応するデータ入出力処理を記憶手段に対して実行する記憶制御装置と、を含んで構成されるストレージシステムにおけるデータ検証方法であって、ホストコンピュータが、記憶制御装置に記憶するデータを規定の仕様に従って編成し管理するアプリケーションプログラムを実行する工程と、ホストコンピュータが、前記アプリケーションプログラムが使用する前記記憶手段の記憶領域の範囲を特定する領域管理データを記憶制御装置に送信する工程と、記憶制御装置が、前記領域管理データを受信する工程と、記憶制御装置が、前記アプリケーションプログラムの処理に起因してホストコンピュータから送信されるデータ入出力要求を受信して、そのデータ入出力要求の処理対象となるデータのうち前記領域指定データに基づいて特定される前記アプリケーションプログラムが使用する記憶領域に対して入出力されるデータについて、そのデータが前記規定の仕様に従って編成されているかどうかを検証する工程と、を有することを特徴とするストレージシステムにおけるデータ検証方法。

【 0 0 1 3 】

ここで前記ホストコンピュータから送信された入出力要求とは、例えば、記憶制御装置に含まれ、もしくは、外部接続するディスクドライブなどの前記記憶手段に対する、データ書き込み要求やデータ読み出し要求である。また、前記領域管理データとは、具体的には、後述する領域管理テーブルに登録されているデータである。このデータ検証方法では、アプリケーションプログラムが取り扱うデータが規定の仕様に従って編成されているかどうかを、記憶制御装置においても検証する。これにより、データの適合性を、より確実に検証し確保することができる。

【 0 0 1 4 】

また、前記データ検証方法は、前記記憶制御装置が、前記ホストコンピュータで稼働する複数の前記アプリケーションプログラムのそれぞれについての前記検証を行う複数の前記アルゴリズムを記憶する工程と、前記ホストコンピュータから送信されてくる前記領域管理データに基づいて前記各アプリケーションプログラムのそれぞれが使用する記憶領域を特定する工程と、前記各アプリケーションプログラムの処理に起因してホストコンピュータから送信されるデータ入出力要求を受信して、前記領域指定データに基づいて前記データ入出力要求に関わる前記アプリケーションプログラムを特定する工程と、前記データ入出力要求の処理対象となるデータのうち前記領域指定データで特定される記憶領域に入出力されるデータについて、そのデータが前記規定の仕様に従って編成されているかどうかを、前記特定したアプリケーションプログラムに対応する前記アルゴリズムにより検証する工程と、を有していてもよい。

【 0 0 1 5 】

これにより、複数のアプリケーションプログラムがホストコンピュータ上で稼働する場合にも、それぞれのアプリケーションプログラムが取り扱うデータについて記憶制御装置側でデータ検証を行うことができる。

【 0 0 1 6 】

10

20

30

40

50

また、前記データ検証方法は、前記ホストコンピュータが、前記アルゴリズムを前記ホストコンピュータから前記記憶制御装置に送信する工程と、前記記憶制御装置が前記アルゴリズムを受信して記憶する工程と、を有していてもよい。これによりデータ検証用のアルゴリズムの記憶制御装置への登録を簡便に行うことができる。

【0017】

また、前記データ検証方法は、前記ホストコンピュータが前記範囲が変更された場合にその変更内容が反映された前記領域管理データを前記記憶制御装置に送信する工程を有していてもよい。これにより記憶制御装置側においても、アプリケーションプログラムが使用する前記記憶手段の記憶領域の範囲についての最新の情報を把握することができる。

【0018】

また、前記アプリケーションプログラムが使用する前記記憶手段の記憶領域は、例えば、前記ホストコンピュータ上で稼働する前記アプリケーションプログラムとは異なるソフトウェアにより付加される制御情報の使用領域を除いた領域であってもよい。

【0019】

また、前記領域管理データには、前記記憶手段の前記記憶領域を用いて論理的に編成されたボリュームを単位として前記アプリケーションプログラムに提供される前記記憶領域の範囲を指定するデータが含まれていてもよい。

【0020】

また、前記ボリュームは、前記記憶手段の前記記憶領域をRAID方式で運用することにより編成されていてもよい。

【0021】

また、前記記憶手段は、前記記憶制御装置に一体に設けられていてもよい。また、前記検証の結果記載したデータを前記ホストコンピュータに送信するようにしてもよい。また、例えば、前記記憶領域は、ディスクドライブなどの物理的な記憶デバイスにより提供される記憶領域上に編成される論理的な記憶デバイスとして提供される。また、規定の仕様は例えば、前記データの規定位置に規定のデータが挿入されていることである。また、前記ホストコンピュータと前記記憶制御装置とを通信可能に接続する手段は、例えば、SCSI規格に準拠した通信方式によるも、SAN、LANなどである。

【0022】

また、前記ホストコンピュータによる前記領域管理データの前記記憶制御装置への前記送信は、前記通信手段を提供する通信路とは別の通信路を介して行われる構成としてもよい。これにより、例えば、通信負荷の分散が図られる。

【0023】

=== 実施例 ===

図1に本発明の一実施例として説明するストレージシステムの構成を示す。この図の記憶制御装置10は、例えばディスクアレイ装置であり、ホストコンピュータ20は、記憶制御装置10を記憶資源として利用する汎用機やパソコンなどである。記憶制御装置10とホストコンピュータ20とは、通信手段40を介して接続する。通信手段40は、例えば、SCSI規格のインタフェースに準拠した通信線やLAN (Local Area Network)、もしくは、SAN (Storage Area Network) である。

【0024】

記憶制御装置10は、当該装置10内の各部の制御や各種処理プログラムの実行制御に用いられるCPU11、各種の情報記憶用として利用される制御メモリ12、記憶手段であるディスクドライブなどの物理デバイス(不図示)、ホストコンピュータ20と接続するためのホストインタフェース13、ホストコンピュータ20から受信したデータ入出力要求に従い、物理デバイスに対するデータ入出力を制御するデータコントローラ14、キャッシュメモリ15、キャッシュメモリ15に登録するデータを一時的に記憶しておくためのデータバッファ16、データコントローラ14からの命令に応じて物理デバイスを制御するディスクインタフェース17などを備える。なお、ホストインタフェース13は、一以上の接続ポート131を備える。また、ホストコンピュータ20側の外部インタフェー

10

20

30

40

50

ス 2 6 は、例えば、チャネルインタフェースとしてのホストベースアダプタである。

【 0 0 2 5 】

記憶制御装置 1 0 は、物理デバイスが提供する記憶領域上に、一以上の論理デバイス 1 8 を編成している。編成した論理デバイスのそれぞれには固有の論理デバイス ID が付与され、ホストコンピュータ 2 0 から記憶制御装置 1 0 の記憶領域の指定を、この論理デバイス ID を用いてすることができる。なお、この実施例では、記憶制御装置 1 0 にそれぞれ A , A ' , B , C , X の論理デバイス ID で指定される 5 つ論理デバイス 1 8 が編成されているものとする。

【 0 0 2 6 】

ホストコンピュータ 2 0 では、OS 2 1 が稼働する。そして、この OS 2 1 上で、論理ボリューム管理プログラム 2 2、アプリケーションプログラム 2 3、領域管理プログラム 2 4 が稼働する。

10

【 0 0 2 7 】

このうち論理ボリューム管理プログラム 2 2 は、記憶制御装置 1 0 の論理デバイス 1 8 の記憶領域上に一以上の論理ボリューム 2 7 を編成し、この論理ボリューム 2 7 で指定される記憶領域をアプリケーションプログラム 2 3 に提供している。論理ボリューム管理プログラム 2 2 の具体例としては、記憶制御装置 1 0 の論理デバイスが提供する記憶領域をソフトウェア的に RAID ( Redundant Array of Inexpensive Disks ) 方式で運用し、これにより編成される論理ボリューム 2 7 をアプリケーションプログラム 2 3 に提供するいわゆるソフトレイドプログラムがあげられる。以下では、論理ボリューム管理プログラム 2 2 はソフトレイドプログラムであるとして説明する。

20

【 0 0 2 8 】

アプリケーションプログラム 2 3 は、例えば、データベースソフトウェアであり、記憶制御装置 1 0 に記憶するデータを規定の仕様に従って編成し管理する仕組み、すなわち、データの適合性を検証する仕組みを備えたプログラムである。またアプリケーションプログラム 2 3 は、処理対象となるデータが規定の仕様に従って編成されているか、すなわち、データが適合性を具備しているかどうかを検証するアルゴリズムを備える。

【 0 0 2 9 】

領域管理プログラム 2 4 は、ホストコンピュータ 2 0 のメモリ 2 8 上に記憶される後述の領域管理テーブル 2 5 を管理する。また、後述するように、領域管理プログラム 2 4 は、領域管理テーブル 2 5 を、適宜なタイミングで記憶制御装置 1 0 に送信する機能を備える。

30

【 0 0 3 0 】

＝ ＝ 領域管理テーブル ＝ ＝

図 2 は領域管理テーブル 2 5 の一例である。領域管理テーブル 2 5 は、論理ボリューム管理プログラム 2 2 が編成する論理ボリューム 2 7 ごとに生成される。領域管理プログラム 2 4 は、例えば、論理ボリューム管理プログラム 2 2 が論理デバイス 1 8 と論理ボリューム 2 7 との対応を変更した場合などにおいて、論理ボリューム管理プログラム 2 2 やアプリケーションプログラム 2 3 などから取得した情報に基づいて、適宜に領域管理テーブル 2 5 を更新する。

40

【 0 0 3 1 】

図 2 において、論理ボリューム ID 2 1 0 は、各論理ボリューム 2 7 に固有に割り当てられる識別子であり、その領域管理テーブル 2 5 がどの論理ボリューム 2 7 についてのものであるかを示す。OS 制御情報オフセット 2 1 1 および OS 制御情報オフセットサイズ 2 1 2 は、論理デバイス 1 8 や論理ボリューム 2 7 の管理のために OS 2 1 が付与する OS 制御情報の論理デバイス 1 8 上の格納位置を示す。OS 制御情報の具体例としては、ファイル制御情報などがあげられる。

【 0 0 3 2 】

論理ボリューム制御情報オフセット 2 1 3 および論理ボリューム制御情報オフセットサイズ 2 1 4 は、論理ボリューム 2 7 の管理のために論理ボリューム管理プログラム 2 2 が生

50

成する論理ボリューム制御情報の論理デバイス 18 上の格納位置を示す。論理ボリューム制御情報とは、論理ボリューム管理プログラム 22 が論理デバイス 18 の管理に用いる情報であり、その具体例としては、論理ボリューム管理プログラム 22 が論理デバイス 18 をRAID方式で運用している場合に必要となるストライプ構成情報などがあげられる。

【0033】

論理デバイスID 215 は、その論理ボリューム 27 を構成している論理デバイス 18 のIDである。アプリケーションID 216 は、その論理ボリューム 27 を使用しているアプリケーションプログラム 23 のIDである。アプリケーションIDは、例えば、ユーザが、ホストコンピュータ 20 のユーザインタフェースを操作して登録する。

【0034】

アプリケーション取扱データサイズ 217 は、この論理ボリューム 27 を使用するアプリケーションプログラム 23 が、論理ボリューム 27 に対してデータ入出力を行う際のデータサイズである。ストライプサイズ 218 は、論理ボリューム管理プログラム 22 が論理デバイス 18 にRAID方式でデータをストライピングして書き込む際のデータサイズである。また、全データサイズ 219 は、その論理ボリューム 27 の総記憶容量を示す。

【0035】

== 領域管理テーブルの送信 ==

領域管理テーブル 25 は、領域管理プログラム 24 により適宜なタイミングでホストコンピュータ 20 から記憶制御装置 10 に送信される。ここで適宜なタイミングとは、例えば、ホストコンピュータ 20 が起動された時や、領域管理テーブル 25 が更新された時、領域管理プログラム 24 のスケジュール機能に設定された時刻などである。

【0036】

記憶制御装置 10 は、ホストコンピュータ 20 から送られてくる領域管理テーブル 25 を受信すると、これを論理デバイス 18 に記憶する。なお、この実施例では、領域管理テーブル 25 は、論理デバイスIDが「X」の論理デバイス 18 に記憶されるものとするが、制御メモリ 12 などに記憶してもよい。

【0037】

図 3 は、ホストコンピュータ 20 から記憶制御装置 10 に領域管理テーブル 25 を送信する際に行われる処理を説明するフローチャートである。この処理には、大別して、領域管理プログラム 24 が最新ポジションの領域管理テーブル 25 を取得する処理と、取得した領域管理テーブル 25 を送信する処理とが含まれる。

【0038】

領域管理プログラム 24 は、まず、処理中に使用するフラグ等の内容を初期化し (S310)、最新のOS制御情報オフセット 211 およびOS制御情報サイズ 212 をOS 21 に要求して取得する (S311)。つぎに、ホストコンピュータ 20 のメモリ 28 上に領域管理テーブル 25 が既に存在しているかどうかを調べる (S312)。

【0039】

ここで領域管理テーブル 25 がメモリ上に存在しない場合には、メモリ 28 上に領域管理テーブル 25 を生成する (S313)。一方、既に領域管理テーブル 25 がメモリ 28 上に存在する場合には、領域管理テーブル 25 のOS制御情報オフセット 211 およびOS制御情報サイズ 212 と、取得したOS制御情報オフセット 211 および前記OS制御情報サイズ 212 とを比較し (S314)、領域管理テーブル 25 のそれぞれの内容と異なっていれば、前記取得した内容に更新する (S315)。そして、更新を行った場合には、領域管理プログラム 24 は更新フラグをオンにする (S316)。

【0040】

つぎに、領域管理プログラム 24 は、論理ボリューム制御情報オフセット 213 および論理ボリューム制御情報サイズ 214 についてもOS 21 から情報を取得してOS制御情報の場合と同様の比較処理を行い、メモリ上の領域管理テーブル 25 の内容と比較を行い (S317, S318)、その内容が異なっていればその内容に領域管理テーブル 25 を更新し (S319)、更新を行った場合は更新フラグをオンにする (S320)。

10

20

30

40

50



## 【 0 0 4 1 】

以上により、メモリ 2 8 上に最新の状態の領域管理テーブル 2 5 が生成される。つぎに領域管理プログラム 2 4 は、更新フラグを調べ (S321)、オンである場合には、メモリ 2 8 に記憶している領域管理テーブル 2 5 の内容を記憶制御装置 1 0 に送信する (S322)。記憶制御装置 1 0 は、領域管理テーブル 2 5 を受信して、これを論理デバイス X に記憶する。

## 【 0 0 4 2 】

以上に説明した、ホストコンピュータ 2 0 から記憶制御装置 1 0 への領域管理テーブル 2 5 の送信に関する処理は、OS 2 1 の起動時などにおいて、OS 制御情報または論理ボリューム制御情報に変更があった場合や、論理ボリューム 2 7 を構成する論理デバイス 1 8 の増減されるなど論理ボリューム 2 7 の構成要素が変更された場合に適宜に実行される。これにより記憶制御装置 1 0 には、常に最新の状態の領域管理テーブル 2 5 がセットアップされることになる。

## 【 0 0 4 3 】

=== 論理デバイス管理テーブル ===

記憶制御装置 1 0 の制御メモリ 1 2 には、論理デバイス管理テーブル 1 2 1 が記憶されている。図 4 はその一例である。このテーブルには、論理デバイス ID 4 1 1 に対応づけて、これに一意に対応する L U N (Logical Unit Number) 4 1 2 と、各論理デバイス 1 8 の記憶容量 4 1 3、各論理デバイス 1 8 が接続するホストインタフェース 1 3 のポート 1 3 1 の ID であるポート ID 4 1 4 などが管理されている。論理デバイス管理テーブル 1 2 1 の内容は、例えば、記憶制御装置 1 0 に接続する管理端末 (不図示) に対するオペレータ操作により更新されたり、記憶制御装置 1 0 が記憶している情報もしくはホストコンピュータ 2 0 から送信されてくる情報などによる自動更新などにより最新の状態に維持される。

## 【 0 0 4 4 】

=== データの状態 ===

つぎに、ホストコンピュータ 2 0 上で稼働するアプリケーションプログラム 2 3 から出力されたデータが、記憶制御装置 1 0 の論理デバイス 1 8 にどのようにして記憶されるかについて、データの構成に着目して説明する。

## 【 0 0 4 5 】

図 5 ( a ) は、アプリケーションプログラム 2 3 が論理ボリューム 2 7 にデータを入出力する際のデータ単位を示している。データ単位 5 1 のデータサイズは、図 2 におけるアプリケーションプログラム取扱データサイズに一致する。

図 5 ( b ) は、論理ボリューム管理プログラム 2 2 が、記憶制御装置 1 0 に対して入出力に際し、図 5 ( a ) のデータ単位 5 1 をどのようにして分割するかを示しており、この図の例では、データ単位 5 1 は、データ 5 2、データ 5 3、データ 5 4 に 3 分割されている。

図 5 ( c ) は、データ 5 2、データ 5 3、データ 5 4 が、それぞれ論理デバイス 1 8 上に記憶されている様子を示している。この図では、アプリケーションプログラム 2 3 が指定する論理ボリューム ID 2 1 5 が「001h」である論理ボリューム 2 7 が、論理デバイス ID 2 1 5 が「A」および「A'」の論理デバイス 1 8 により編成されている場合である。分割されたデータのうち、データ 5 2、5 3 は、論理デバイス ID 2 1 5 が「A」の論理デバイス 1 8 に、データ 5 4 は、論理デバイス ID 2 1 5 が「A'」の論理デバイス 1 8 にそれぞれ記憶されている。また、これら各論理デバイス 1 8 の固定領域には、前述した OS 制御情報 5 5、5 6 や、論理ボリューム制御情報 5 7、5 8 が記憶されている。

## 【 0 0 4 6 】

一方、図 6 ( a ) は、図 5 ( a ) のデータ単位を S C S I 規格におけるデータフォーマットとして示したものであり、図 6 ( b ) は、図 5 ( b ) に示すデータについての S C S I 規格レベルでのデータフォーマットを示したものである。

## 【 0 0 4 7 】

図 6 ( b ) に示すように、コマンドフレームのオペコード欄 6 1 1 ~ 6 1 3 には、この処

10

20

30

40

50

理要求の種類を示すコマンドがセットされる。この例では書き込みコマンドがセットされる。LUN 6 2 1 ~ 6 2 3 の欄には、処理対象となる論理デバイスID 2 1 5 もしくはLUN (Logical Unit Number) がセットされる。この例では、LUN 6 2 1, 6 2 2 には「0001h」が、LUN 6 2 3 には「0002h」がそれぞれセットされている。論理アドレス 6 3 1 ~ 6 3 3 には、各データ 5 2 ~ 5 4 が書き込まれる論理デバイス 1 8 上の格納開始位置に対応するアドレスがセットされる。データ長 6 4 1 ~ 6 4 3 には、各コマンドフレームに付帯する書き込みデータ 5 2 ~ 5 4 のそれぞれに対応するデータ長がセットされる。

【0048】

なお、以上はアプリケーションプログラム 2 3 からのデータ書き込み要求が出力された場合であるが、データ読み出し要求が出力された場合には、前述のオペコード欄 6 1 1 ~ 6 1 3 に読み出しコマンドがセットされる。また論理アドレス 6 3 1 ~ 6 3 3 には、論理デバイス 1 8 上のデータ読み出し開始位置を指定するアドレスがセットされ、データ長に読み出し対象となるデータのサイズがセットされることになる。

【0049】

=== データ検証処理 ===

つぎに、ホストコンピュータ 2 0 から記憶制御装置 1 0 に SCSI データが送信された場合に記憶制御装置 1 0 において行われる、データ検証処理について説明する。

【0050】

データ検証処理は、記憶制御装置 1 0 が制御メモリ 1 2 に記憶している、データ検証用のプログラム 1 2 2 (アルゴリズム) を実行することで行われる。データ検証用プログラムは、ホストコンピュータ 2 0 で動作するアプリケーションプログラム 2 3 ごとに用意される。データ検証用プログラム 1 2 2 は、処理対象となるデータについての処理の実行に際し、事前に処理対象となるデータが規定の仕様を具備しているかどうかを検証する、ホストコンピュータ 2 0 で稼働するアプリケーションプログラム 2 3 が備える前述のアルゴリズムと同等以上の処理を実行する機能を備える。

【0051】

データ検証用プログラム 1 2 2 は、例えば、ホストコンピュータ 2 0 から送信されたり、記憶制御装置 1 0 の管理端末 (不図示) に対するオペレータ操作などにより、制御メモリ 1 2 に記憶される。記憶される各データ検証用プログラム 1 2 2 には、それぞれ対応するアプリケーションプログラム 2 3 のIDが付与される。

【0052】

なお、データ検証処理は、処理対象となるデータが、アプリケーションプログラム 2 3 が要求する規定の仕様を満たしているかどうかを検証するものであるから、論理デバイス 1 8 の記憶領域に記憶されるデータのうち、アプリケーションプログラム 2 3 から入出力されるデータが格納される領域のみを対象として行う。

【0053】

つぎに、ホストコンピュータ 2 0 上で稼働するアプリケーションプログラム 2 3 から図 6 (a) に示すデータ 5 1 を書き込みデータとするデータ書き込み要求が送信され、このデータ書き込み要求に対応する図 6 (b) に示す書き込みコマンドの一つが記憶制御装置 1 0 に送信された場合を例として、図 7 のフローチャートを用い、記憶制御装置 1 0 において行われるデータ検証処理について説明する。なお、この処理は、例えば、記憶制御装置 1 0 の制御メモリ 1 2 などに記憶されているマイクロプログラムなどにより実行される。

【0054】

記憶制御装置 1 0 は、ホストコンピュータ 2 0 から送られてくる書き込みコマンドのコマンドフレームを受信すると (S711)、そのコマンドフレームをデータバッファ 1 6 に記憶する (S712)。つぎに、記憶制御装置 1 0 は、論理デバイスIDが「X」である論理デバイス 1 8 に記憶している領域管理テーブル 2 5 を参照し、受信した前記データ書き込み要求にセットされている論理デバイスIDのOS制御情報オフセット 2 1 1 およびOS制御情報サイズ 2 1 2、論理ボリューム制御情報オフセット 2 1 3 および論理ボリューム制御情報サイズ 2 1 4 を取得する (S713)。また、記憶制御装置 1 0 は、論理デバイス管理テーブ

10

20

30

40

50

ル 1 2 1 を参照し、前記論理デバイス ID に対応する論理デバイス 1 8 の記憶容量 4 1 3 を取得する (S714)。

【 0 0 5 5 】

つぎに記憶制御装置 1 0 は、前記論理デバイス ID に対応する論理デバイス 1 8 の記憶領域のうち、前記 OS 制御情報オフセット 2 1 1 および OS 制御情報サイズ 2 1 2、および論理ボリューム制御情報オフセット 2 1 3、論理ボリューム制御情報サイズ 2 1 4 で指定される領域を除いた領域を指定するアドレス (例えば、一以上のスタートアドレスとエンドアドレス) を記憶する (S715)。なお、以下、このアドレスで指定される記憶領域を検証対象候補領域と称することとする。

【 0 0 5 6 】

つぎに、記憶制御装置 1 0 は、前記検証対象候補領域を指定するアドレスと、受信したコマンドフレームの論理アドレス 6 3 1 ~ 6 3 3 の欄にセットされているアドレスとを比較する (S716)。ここでこのアドレスが前記検証対象候補領域に含まれない場合には、データ検証処理を終了する (S717)。一方、前記アドレスが検証対象候補領域内に含まれる場合には、記憶制御装置 1 0 は、さらに、このアドレスと前記コマンドフレームのデータ長 6 4 1 ~ 6 4 3 にセットされているデータ長とにより定まる、当該書き込みデータの書き込み先として指定される記憶領域の全体が、前記検証対象候補領域に含まれるかどうかを調べる (S718)。ここで含まれる場合には、前記書き込み先として指定される記憶領域の全体をデータ検証の対象領域に設定し、この領域を指定するアドレス (例えば、この領域を指定するスタートアドレスとエンドアドレス) を記憶する (S719)。

【 0 0 5 7 】

一方、検証対象候補領域内に収まらない場合には、前記論理アドレスから前記検証対象候補の領域のエンドアドレスまでの領域をデータ検証の対象領域に設定し、この領域を指定するアドレス (例えば、この領域を指定するスタートアドレスとエンドアドレス) を記憶する (S720)。

【 0 0 5 8 】

以上のようにしてデータ検証の対象領域が設定されると、つぎに記憶制御装置 1 0 は、書き込み対象となる論理デバイス 1 8 に対応づけられているアプリケーション ID を調べる。ここで記憶制御装置 1 0 は、アプリケーション ID と、各アプリケーションプログラム 2 3 についてのデータ検証用プログラム 1 2 2 との対応付けを記憶しており、調べた前記ア

プリケーション ID に対応するデータ検証用プログラム 1 2 2 を起動する (S721)。

【 0 0 5 9 】

記憶制御装置 1 0 は、データ検証用プログラム 1 2 2 の起動に伴い、データ検証用プログラム 1 2 2 に前記データ検証の対象領域を指定するアドレスを引き渡し、データ検証用プログラム 1 2 2 は、引き渡された前記アドレスで指定される論理デバイス 1 8 上の記憶領域を対象として、データ検証処理を開始する (S722)。

【 0 0 6 0 】

ここでデータ検証用プログラムによるデータ検証処理において異常が検知された場合 (S723) には、記憶制御装置 1 0 は、その旨を記載したメッセージ (例えば、イリーガルリクエストやライトエラー等が記載されたもの) をホストコンピュータ 2 0 に送信する (S724)。一方、異常が検知されない場合には、記憶制御装置 1 0 は、データバッファ 1 6 に記憶されているコマンドフレームの書き込み対象データをキャッシュメモリ 1 5 に送信し、このデータを論理デバイス 1 8 に書き込む (S725)。

【 0 0 6 1 】

ところで、以上の説明は、記憶制御装置 1 0 が、ホストコンピュータ 2 0 からデータ書き込み要求を受信した場合であったが、データ読み出し要求を受信した場合には、コマンドフレームのデータを検証する代わりに、読み出し対象のデータを記憶手段から読み出してデータバッファ 1 6 に記憶し、書き込みコマンドの場合と同様の処理にてデータ検証を行う。ここでこのようにデータ読み出し要求が送られてきた場合にもデータ検証処理を行って、これにより異常が検知された場合にはその旨のメッセージがホストコンピュータ 2 0

に通知されるようにすることで、例えば、読み出されたデータがアプリケーションプログラム 23 に引き渡される前に、データの適合性が確保されていないことをアプリケーションプログラム 23 に通知することができ、アプリケーションプログラム 23 の誤動作や、データ喪失、破壊などを防ぐことができる。

#### 【0062】

以上のように、本発明によれば、データ検証を行うために必要な情報を記憶制御装置 10 側でも取得する。従って、アプリケーションプログラム 23 が行っているデータ検証処理を、記憶制御装置 10 側でも実施することが可能となる。そして、記憶制御装置 10 側でもデータ検証処理を実行することで、アプリケーションプログラム 23 が要求するデータの適合性を満たすよう、より確実にデータが管理されることとなる。

10

#### 【0063】

== その他 ==

ところで、以上の説明において、論理ボリューム管理プログラム 22 は必ずしも必須の構成要素では無い。

#### 【0064】

領域管理テーブル 25 は、ホストコンピュータ 20 から記憶制御装置 10 に常にその全ての内容を送信する必要は無く、変更差分データのみを送信するようにしてもよい。

#### 【0065】

また、領域管理プログラム 24 が初めて実行される際は、ホストコンピュータ 20 上の領域管理テーブル 25 は何も書き込まれていないが、この場合には、例えば、領域管理テーブル 25 に対する変更があったものとして領域管理テーブル 25 を記憶制御装置 10 に送信するようにする。

20

#### 【0066】

また、ホストコンピュータ 20 と記憶制御装置 10 とを、例えば、LANなどで接続し、領域管理テーブル 25 の送信をLAN経由で行うことで、ホストコンピュータ 20 から記憶制御装置 10 に高速に送信できる。

#### 【0067】

#### 【発明の効果】

本発明によれば、ストレージシステムにおいて、データの適合性を、より確実に検証し確保することができる、データ検証方法およびホストコンピュータおよび記憶制御装置を提供することができる。

30

#### 【図面の簡単な説明】

【図1】本発明の一実施例として説明するストレージシステムの構成を示す図である。

【図2】本発明の一実施例として説明する領域管理テーブルを示す図である。

【図3】本発明の一実施例による、ホストコンピュータから記憶制御装置に領域管理テーブルを送信する際に行われる処理を説明するフローチャートを示す図である。

【図4】本発明の一実施例として説明する論理デバイス管理テーブルを示す図である。

【図5】(a)は、本発明の一実施例による、アプリケーションプログラムが論理ボリュームにデータを入出力する際のデータ単位を示す図であり、(b)は、論理ボリューム管理プログラムが、記憶制御装置に対して入出力に際し(a)のデータ単位をどのようにして分割するかを示す図であり、(c)は、分割されたデータが論理デバイス上に記憶されている様子を示す図である。

40

【図6】(a)は、図5(a)のデータ単位をSCSI規格におけるデータフォーマットとして示した図であり、(b)は図5(b)のデータをSCSI規格におけるデータフォーマットとして示した図である。

【図7】本発明の一実施例による、記憶制御装置におけるデータ検証処理を説明するフローチャートである。

#### 【符号の説明】

10 記憶制御装置

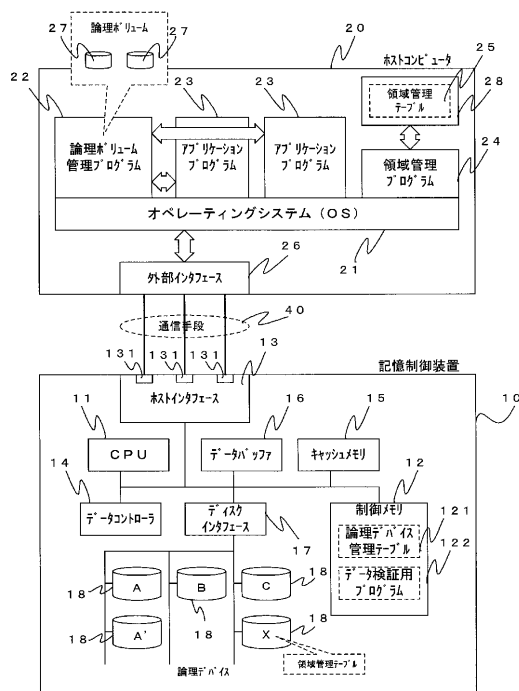
12 制御メモリ

50

- 15 キャッシュメモリ
- 16 データバッファ
- 18 論理デバイス
- 20 ホストコンピュータ
- 23 アプリケーションプログラム
- 24 領域管理プログラム
- 25 領域管理テーブル
- 27 論理ボリューム
- 40 通信手段
- 121 論理デバイス管理テーブル
- 122 データ検証用プログラム
- 211 OS制御情報オフセット
- 212 OS制御情報サイズ
- 213 論理ボリューム制御情報オフセット
- 214 論理ボリューム制御情報サイズ

10

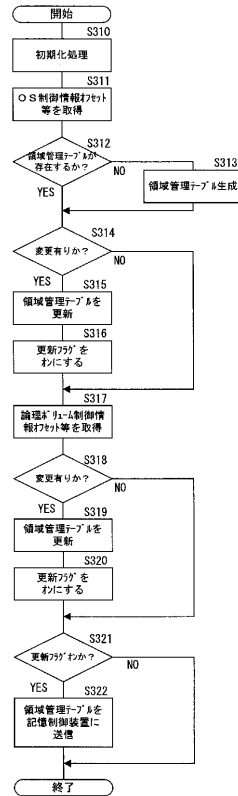
【図1】



【図2】

論理ボリュームID	001h	210
OS制御情報オフセット	0000h	211
OS制御情報サイズ	1024byte	212
論理ボリューム制御情報オフセット	1024byte	213
論理ボリューム制御情報サイズ	512byte	214
論理デバイスID	A, A'	215
アプリケーションID	$\alpha$	216
アプリケーション取扱データサイズ	1024	217
ストライプサイズ	4096	218
全データサイズ	100GB	219

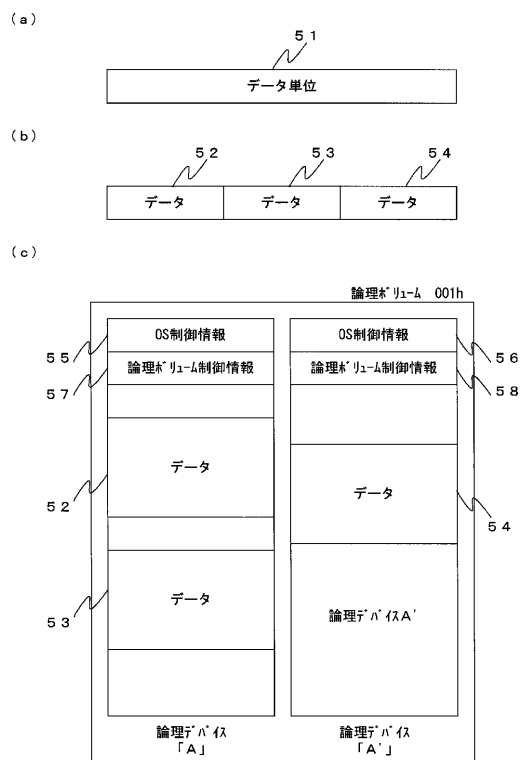
【 図 3 】



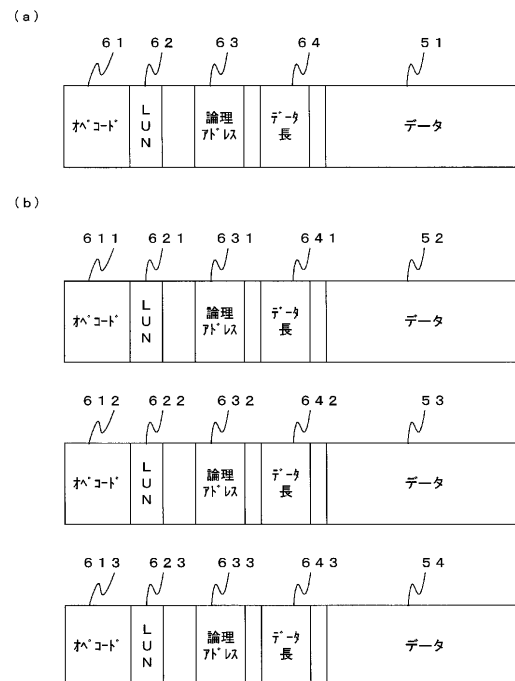
【 図 4 】

論理デバイスID	LUN	記憶容量	ポートID	・
A	0001h	500Gb	01h	・
A'	0002h	300Gb	01h	・
B	0003h	1Tb	02h	・
C	0004h	1Tb	03h	・
・	・	・	・	・
・	・	・	・	・
・	・	・	・	・

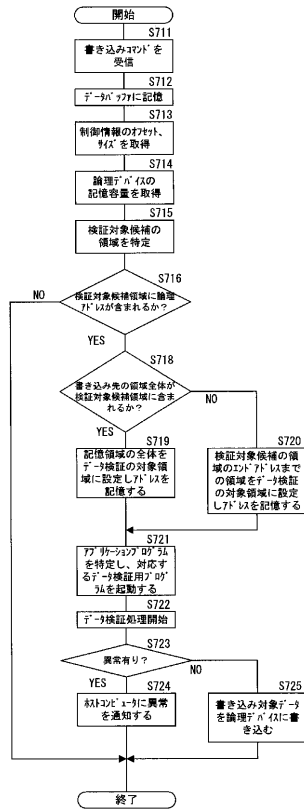
【 図 5 】



【 図 6 】



【図 7】



---

フロントページの続き

(72)発明者 占部 喜一郎

神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R A I Dシステム事業部内

(72)発明者 中野 俊夫

神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R A I Dシステム事業部内

(72)発明者 横畑 静生

神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R A I Dシステム事業部内

審査官 平井 誠

(56)参考文献 特開 2 0 0 1 - 3 3 1 3 2 3 ( J P , A )

特開 2 0 0 1 - 1 4 2 6 4 8 ( J P , A )

特開平 0 8 - 2 7 2 6 2 5 ( J P , A )

特開平 0 8 - 2 6 3 2 2 3 ( J P , A )

(58)調査した分野(Int.Cl. , D B 名)

G06F 12/14

G06F 3/06

G06F 13/00