

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6039149号
(P6039149)

(45) 発行日 平成28年12月7日 (2016. 12. 7)

(24) 登録日 平成28年11月11日 (2016. 11. 11)

(51) Int. Cl.	F I
G 0 6 F 3/06 (2006.01)	G 0 6 F 3/06 3 0 2 J
G 0 6 F 13/10 (2006.01)	G 0 6 F 3/06 3 0 1 K
	G 0 6 F 13/10 3 4 0 A

請求項の数 12 (全 42 頁)

(21) 出願番号	特願2015-127702 (P2015-127702)	(73) 特許権者	000005108
(22) 出願日	平成27年6月25日 (2015. 6. 25)		株式会社日立製作所
(62) 分割の表示	特願2014-92147 (P2014-92147)		東京都千代田区丸の内一丁目6番6号
	の分割	(74) 代理人	100093861
原出願日	平成22年8月11日 (2010. 8. 11)		弁理士 大賀 真司
(65) 公開番号	特開2015-201233 (P2015-201233A)	(74) 代理人	100129218
(43) 公開日	平成27年11月12日 (2015. 11. 12)		弁理士 百本 宏之
審査請求日	平成27年7月24日 (2015. 7. 24)	(72) 発明者	鎌 講平
			神奈川県横浜市戸塚区吉田町292番地
			株式会社日立製作所 横浜研究所内
		(72) 発明者	江口 賢哲
			神奈川県横浜市戸塚区吉田町292番地
			株式会社日立製作所 横浜研究所内

最終頁に続く

(54) 【発明の名称】 ストレージ装置及びその制御方法

(57) 【特許請求の範囲】

【請求項 1】

複数の一定サイズのページにて形成される仮想的な記憶領域である仮想ボリュームと、前記仮想ボリューム上にデータを格納するデータ部を纏めたデータ領域と、前記データ部へのアクセスに用いられる制御情報を格納する制御情報部を纏めた制御情報領域と、を形成するストレージコントローラと、

前記仮想ボリュームに実記憶領域を提供する記憶媒体と、
を備え、

前記ストレージコントローラは、

前記仮想ボリュームを設定後、一定の第1のサイズの前記データ領域と一定の第2のサイズの前記制御情報領域とを備え、前記第1のサイズの前記データ領域と前記第2のサイズの前記制御情報領域とを交互に配置するグループの単位で、前記仮想ボリュームの容量を拡張し、

前記仮想ボリューム上の第1のデータ部を対象とするライト要求があった場合に、前記第1のデータ部を含む第1のページへの実記憶領域の割り当ての有無を判断し、前記第1のページに実記憶領域が割り当てられていない場合は前記第1のページに実記憶領域を割り当て、前記第1のデータ部へのアクセスに用いられる制御情報を格納する第1の制御情報部を含む第2のページへの実記憶領域の割り当ての有無を判断し、前記第2のページに実記憶領域が割り当てられていない場合は前記第2のページに実記憶領域を割り当てるストレージシステム。

10

20

【請求項 2】

請求項 1 に記載のストレージシステムであって、

前記ストレージコントローラは、前記第 1 のデータ部を対象とする前記ライト要求があった場合に、前記第 1 のサイズと前記第 2 のサイズとに基づいて、前記第 1 のデータ部のアドレスと前記第 1 の制御情報部のアドレスを算出するストレージシステム。

【請求項 3】

請求項 2 に記載のストレージシステムであって、

前記ストレージコントローラは、

前記算出した前記第 1 のデータ部のアドレスに基づいて、前記第 1 のデータ部を含む前記第 1 のページを特定し、特定した前記第 1 のページへの実記憶領域の割り当ての有無を判断し、

10

前記算出した前記第 1 の制御情報部のアドレスに基づいて、前記第 1 の制御情報部を含む前記第 2 のページを特定し、特定した前記第 2 のページへの実記憶領域の割り当ての有無を判断するストレージシステム。

【請求項 4】

請求項 1 に記載のストレージシステムであって、

前記ストレージコントローラは、

前記第 1 のページに割り当てられた前記実記憶領域上、前記第 1 のデータ部のアドレスに対応する領域に、前記ライト要求に基づくデータをライトするストレージシステム。

【請求項 5】

20

請求項 1 に記載のストレージシステムであって、

前記ストレージコントローラは、

前記第 2 のページに割り当てられた前記実記憶領域上、前記第 1 の制御情報部のアドレスに対応する領域に、前記第 1 のデータ部へのアクセスに用いられる制御情報を格納するストレージシステム。

【請求項 6】

請求項 1 に記載のストレージシステムであって、

前記記憶媒体が提供する複数の前記実記憶領域を含むプールをさらに有し、

前記第 1 のページ及び前記第 2 のページに、前記プールの前記実記憶領域を割り当てるストレージシステム。

30

【請求項 7】

複数の一定サイズのページにて形成される仮想的な記憶領域である仮想ボリューム上に、データを格納するデータ部を纏めたデータ領域と、前記データ部へのアクセスに用いられる制御情報を格納する制御情報部を纏めた制御情報領域と、を形成し、

前記仮想ボリュームに実記憶領域を提供する、

ストレージシステムの制御方法であって、

前記ストレージシステムは、

前記仮想ボリュームを設定後、一定の第 1 のサイズの前記データ領域と一定の第 2 のサイズの前記制御情報領域とを備え、前記第 1 のサイズの前記データ領域と前記第 2 のサイズの前記制御情報領域とを交互に配置するグループの単位で、前記仮想ボリュームの容量を拡張し、

40

前記仮想ボリューム上の第 1 のデータ部を対象とするライト要求があった場合に、前記第 1 のデータ部を含む第 1 のページへの実記憶領域の割り当ての有無を判断し、前記第 1 のページに実記憶領域が割り当てられていない場合は前記第 1 のページに実記憶領域を割り当て、前記第 1 のデータ部へのアクセスに用いられる制御情報を格納する第 1 の制御情報部を含む第 2 のページへの実記憶領域の割り当ての有無を判断し、前記第 2 のページに実記憶領域が割り当てられていない場合は前記第 2 のページに実記憶領域を割り当てるストレージシステムの制御方法。

【請求項 8】

請求項 7 に記載のストレージシステムの制御方法であって、

50

前記ストレージシステムは、前記第1のデータ部を対象とする前記ライト要求があった場合に、前記第1のサイズと前記第2のサイズとに基づいて、前記第1のデータ部のアドレスと前記第1の制御情報部のアドレスを算出するストレージシステムの制御方法。

【請求項9】

請求項8に記載のストレージシステムの制御方法であって、

前記ストレージシステムは、

前記算出した前記第1のデータ部のアドレスに基づいて、前記第1のデータ部を含む前記第1のページを特定し、特定した前記第1のページへの実記憶領域の割り当ての有無を判断し、

前記算出した前記第1の制御情報部のアドレスに基づいて、前記第1の制御情報部を含む前記第2のページを特定し、特定した前記第2のページへの実記憶領域の割り当ての有無を判断するストレージシステムの制御方法。

10

【請求項10】

複数の一定サイズのページにて形成される仮想的な記憶領域である仮想ボリュームを管理するストレージコントローラと、実記憶領域を提供する記憶媒体と、を備え、

前記仮想ボリュームは、データを格納するデータ部を纏めたデータ領域と、前記データ部へのアクセスに用いられる制御情報を格納する制御情報部を纏めた制御情報領域と、を有し、

前記ストレージコントローラは、

前記仮想ボリュームを設定後、一定の第1のサイズの前記データ領域と一定の第2のサイズの前記制御情報領域とを備え、前記第1のサイズの前記データ領域と前記第2のサイズの前記制御情報領域とを交互に配置するグループの単位で、前記仮想ボリュームの容量を拡張し、

20

前記仮想ボリューム上の第1のデータ部を対象とするライト要求があった場合に、前記第1のデータ部を含む第1のページへの実記憶領域の割り当ての有無を判断し、前記第1のページに実記憶領域が割り当てられていない場合は前記第1のページに実記憶領域を割り当て、前記第1のデータ部へのアクセスに用いられる制御情報を格納する第1の制御情報部を含む第2のページへの実記憶領域の割り当ての有無を判断し、前記第2のページに実記憶領域が割り当てられていない場合は前記第2のページに実記憶領域を割り当てるストレージ装置。

30

【請求項11】

請求項10に記載のストレージ装置であって、

前記ストレージコントローラは、前記第1のデータ部を対象とする前記ライト要求があった場合に、前記第1のサイズと前記第2のサイズとに基づいて、前記第1のデータ部のアドレスと前記第1の制御情報部のアドレスを算出するストレージ装置。

【請求項12】

請求項11に記載のストレージ装置であって、

前記ストレージコントローラは、

前記算出した前記第1のデータ部のアドレスに基づいて、前記第1のデータ部を含む前記第1のページを特定し、特定した前記第1のページへの実記憶領域の割り当ての有無を判断し、

40

前記算出した前記第1の制御情報部のアドレスに基づいて、前記第1の制御情報部を含む前記第2のページを特定し、特定した前記第2のページへの実記憶領域の割り当ての有無を判断するストレージ装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、記憶デバイスで構成されたボリュームのデータ配置を管理するためのストレージ装置及びその制御方法に関する。

【背景技術】

50

【0002】

ストレージシステムあるいはストレージ装置として、記憶デバイスに、例えば、ディスクドライブを用いたものが知られている。記憶デバイスに対するデータの入出力処理を制御する場合、ディスクのヘッドを位置決めすることが行われている。

【0003】

例えば、特許文献1によると、記憶デバイスで構成されたボリュームは一つまたは複数のトラックを備えている。記憶デバイスの実記憶領域は、トラック内の先頭に制御情報部（ホームアドレス、レコード0）があり、その後ろにデータ部（レコード1、レコード2、・・・）が配置されている。メインフレーム（MF）ホスト計算機が、トラックのリードまたはライトをストレージコントローラに要求した場合、ストレージコントローラは、リードコマンドまたはライトコマンドに従ってディスクから、トラックを読み出し、トラックの制御情報部の値（ホームアドレス）を読んで、トラックの位置情報を確認し、トラックに対するヘッドの位置決めが正しく行われたことを確認した後に、データ部に対してリードまたはライトアクセスを実施する。

10

【先行技術文献】

【特許文献】

【0004】

【特許文献1】 米国特許公報第5506979号

【発明の概要】

【発明が解決しようとする課題】

20

【0005】

近年、オンライン商取引の活発化に伴い、ネットワークに接続されたストレージ装置に対して発行されるリードまたはライトアクセスの性能向上が望まれる。こうした環境下では、前述したストレージコントローラが、トラックの位置情報を確認する処理を省き、データ部のみをディスクに対して読み書きすることで、リードアクセスまたはライトアクセス処理の高速化を行う。

【0006】

例えば、ストレージコントローラが、ホスト計算機から、あるトラックに対するリードアクセスまたはライトアクセスを受領した場合に、このトラック内の制御情報の値を参照せず、トラックの先頭にディスクのヘッドを位置づける処理が正しく行われたことを確認する処理を省くときには、ディスクからストレージコントローラへの制御情報部の転送は不要となり、リードアクセスまたはライトアクセス処理を高速化できる。

30

【0007】

しかし、特許文献1に記載されている、ボリュームのデータ配置を採用したのでは、ストレージコントローラが、ディスクからデータ部のみを読み出すようにしたとしても、ディスクのヘッドを、トラックの先頭にある制御情報部の位置をスキップして、データ部に位置付ける処理がオーバーヘッドとなる。

【0008】

本発明の目的は、リード／ライトアクセスの性能を向上させることができるストレージ装置及びその制御方法を提供することにある。

40

【課題を解決するための手段】

【0009】

前記目的を達成するために、本発明は、複数の一定サイズのページにて形成される仮想的な記憶領域である仮想ボリュームと、前記仮想ボリューム上にデータを格納するデータ部を纏めたデータ領域と、前記データ部へのアクセスに用いられる制御情報を格納する制御情報部を纏めた制御情報領域と、を形成するストレージコントローラと、前記仮想ボリュームに実記憶領域を提供する記憶媒体と、を備え、前記ストレージコントローラは、前記仮想ボリュームを設定後、当該仮想ボリュームの容量を一定の第1のサイズの前記データ領域と一定の第2のサイズの前記制御情報領域とを備えたグループの単位で拡張し、前記仮想ボリューム上の第1のデータ部を対象とするライト要求があった場合に、前記第1

50

のサイズと前記第2のサイズとに基づいて、前記第1のデータ部を含む第1のページへの実記憶領域の割り当ての有無を判断し、当該割り当てが行われていない場合は前記第1のページに実記憶領域を割り当て、前記第1のデータ部へのアクセスに用いられる制御情報を格納する第1の制御情報部を含む第2のページへの実記憶領域の割り当ての有無を判断し、当該割り当てが行われていない場合は前記第2のページに実記憶領域を割り当て、前記仮想ボリュームと前記実記憶領域のそれぞれにおける前記データ領域と前記制御情報領域の配置の相違に基づいて、前記仮想ボリュームの前記データ領域に対する上位装置からのアクセス先アドレスに対応する、前記実記憶領域における前記データ領域の対応アドレスを算出することを特徴とするストレージ装置。

【発明の効果】

10

【0010】

本発明によれば、制御情報部を参照する必要がない場合でも、リード／ライトアクセスの性能を向上させることができる。

【図面の簡単な説明】

【0011】

【図1】ストレージシステムの全体構成を示すブロック構成図である。

【図2】ストレージコントローラ内で実行するマイクロプログラムの構成を示す構成図である。

【図3】ストレージ装置内の論理構成を示す構成図である。

【図4】HDEV番号とLDEV番号との対応関係を管理するための管理テーブルの構成図である。

20

【図5】LDEV番号とボリューム属性及びボリューム容量の対応関係を管理するための管理テーブルの構成図である。

【図6】LDEV番号とディスク番号との対応関係を管理するための管理テーブルの構成図である。

【図7】LDEV番号とプール番号との対応関係を管理するための管理テーブルの構成図である。

【図8】プール番号とディスク番号との対応関係を管理するための管理テーブルの構成図である。

【図9】3390形式のLDEV内のデータ配置を説明するための構成図である。

30

【図10】3390形式のLDEV内のデータ配置と実記憶領域上のデータ配置を説明するための構成図である。

【図11】分散ユーザデータ領域と分散制御情報領域のサイズを管理するための管理テーブルの構成図である。

【図12】分散ユーザデータ領域と分散制御情報領域のアドレスを管理するための管理テーブルの構成図である。

【図13】第1実施例におけるアドレス変換処理を説明するためのフローチャートである。

【図14】3390-A形式のLDEV内のデータ配置を説明するための構成図である。

【図15】シリンダグループ内の構成を説明するための構成図である。

40

【図16】3390形式のLDEV上のアドレスと3390-A形式のLDEV上のアドレスとの対応関係を説明するための構成図である。

【図17】分散ユーザデータ領域と分散制御情報領域のサイズを管理するための管理テーブルの構成図である。

【図18】分散ユーザデータ領域と分散制御情報領域のアドレスを管理するための管理テーブルの構成図である。

【図19】3390-A形式のLDEVの容量拡張処理を示すフローチャートである。

【図20】容量拡張後における3390形式のLDEV上のアドレスと3390-A形式のLDEV上のアドレスとの対応関係を説明するための構成図である。

【図21】第2実施例におけるアドレス変換処理を説明するためのフローチャートである

50

。

【図 2 2】第 3 実施例における H D E V と L D E V との関係を説明するための構成図である。

【図 2 3】第 3 実施例における分散ユーザデータ領域と分散制御情報領域のアドレスを管理するための管理テーブルの構成図である。

【図 2 4】第 4 実施例におけるコピー処理を説明するためのフローチャートである。

【図 2 5】ページ管理テーブルの構成図である。

【図 2 6】ページ管理ディレクトリの構成図である。

【図 2 7】第 5 実施例における M F ホストライトコマンド処理を説明するためのフローチャートである。

10

【図 2 8】第 6 実施例におけるコピーデータ受信処理を説明するためのフローチャートである。

【発明を実施するための形態】

【0012】

以下、本発明の実施例を説明する。

【0013】

なお、以下の説明では、「管理テーブル」等の表現にて各種情報を説明することがあるが、各種情報は、テーブル以外のデータ構造で表現されていてもよい。また、データ構造に依存しないことを示すために「管理テーブル」を「管理情報」と呼ぶことができる。

【0014】

20

また、以後の説明では、「プログラム」を主語として処理を説明する場合があるが、プログラムは、プロセッサ例えば、C P U (C e n t r a l P r o c e s s i n g U n i t) によって実行されることで、定められた処理を、適宜に記憶資源（例えばメモリ）及び通信インタフェース装置（例えば通信ポート）を用いながら行うため、処理の主語がプロセッサとされてもよい。プロセッサは、C P U の他に専用ハードウェアを有していても良い。コンピュータプログラムは、プログラムソースから各計算機にインストールされても良い。プログラムソースは、例えば、プログラム配布サーバ又は記憶メディアであっても良い。

【0015】

また、各要素例えば、L D E V (L o g i c a l D E V i c e) は、番号などで識別可能であるが、識別可能な情報であれば、名前など他種の識別情報が用いられても良い。

30

【0016】

以下、図を用いて、本発明の実施例を説明する。以下の図中、同一の部分には同一の符号を付加する。ただし、本発明が本実施例に制限されることは無く、本発明の思想に合致するあらゆる応用例が本発明の技術的範囲に含まれる。また、特に限定しない限り、各構成要素は複数でも単数でも構わない。

【0017】

<ストレージシステムの全体構成>

【0018】

図 1 に、ストレージシステムの全体構成を示す。

40

【0019】

図 1 において、ストレージシステムは、例えば、複数のストレージ装置 1 0 1 と、複数のメインフレーム（M F）ホスト計算機 1 0 2 と、管理ホスト計算機 1 0 3 とを備える計算機システムとして構成される。

【0020】

各ストレージ装置 1 0 1 と各 M F ホスト計算機 1 0 2 は、例えば、ネットワーク 1 1 1 を介して接続され、複数のストレージ装置のうち一つのストレージ装置（以下、第一のストレージ装置と称することもある。）1 0 1 と管理ホスト計算機 1 0 3 は、例えば、ネットワーク 1 1 3 を介して接続される。

【0021】

50

第1のストレージ装置101は、例えば、ネットワーク111を介して、異なるストレージ装置（以下、第二のストレージ装置と称することもある。）101に接続される。

【0022】

各MFホスト計算機102は、通信線114を介してネットワーク111に接続される。

【0023】

各ストレージ装置101は、通信線115を介してネットワーク111に接続される。

【0024】

管理ホスト計算機103は、通信線116を介してネットワーク113に接続される。

【0025】

第1のストレージ装置101は、通信線117を介してネットワーク113に接続される。

【0026】

なお、上述の通信線114～117は、例えば、メタルケーブルや光ファイバケーブル等の有線として構成される。しかし、各MFホスト計算機102と各ストレージ装置101、第1のストレージ装置101と管理計算機103、各MFホスト計算機102と管理計算機103をそれぞれ無線で接続することも可能である。この場合は、通信線114～117は省略される。

【0027】

また、ネットワーク111とネットワーク113は、共通のネットワークであっても良い。各ネットワーク111、113は、通信ネットワークであり、例えば、SAN（Storage Area Network）又はLAN（Local Area Network）である。

【0028】

次に、各ストレージ装置101の構成を説明する。

【0029】

各ストレージ装置101は、例えば、一つまたは複数のストレージコントローラ131と、一つまたは複数のディスク132を備える。なお、各ストレージ装置101は同一の構成であるので、以下、第一のストレージ装置101の構成について説明する。

【0030】

一つまたは複数のディスク132には、データを格納する記憶デバイスとして、少なくとも、SSD（Solid State Drive）、SAS（Serial Attached SCSI）－HDD（Hard Disk Drive）、SATA（Serial Advanced Technology Attachment）－HDDが含まれている。なお、ディスク132は、それらのうちの少なくとも1つに代えて、またはそれらに加えて、他の種類の物理的な記憶デバイスであっても良い。

【0031】

一つまたは複数のディスク132は、例えば、ファイバチャネルケーブル等の通信線112を介して、ストレージコントローラ131に接続される。なお、複数のディスク132で、一つまたは複数のRAID（Redundant Array of Independent Disks）グループを構成することができる。

【0032】

次に、ストレージコントローラ131の構成を説明する。

【0033】

ストレージコントローラ131は、各MFホスト計算機102から受信したコマンドに従って、ディスク132に対するデータの入出力処理、即ち、ディスク132へのデータのライト（書き込み）やリード（読み出し）を制御する。

【0034】

ストレージコントローラ131は、アクセス要求元である各MFホスト計算機102に対して、実記憶領域が既に割り当てられている状態の論理デバイス、または後述するシン

10

20

30

40

50

・プロビジョニング機能で用いられる仮想記憶領域で構成される論理デバイスを、アクセス対象の論理デバイスあるいは論理ボリュームとして提供する。

【0035】

仮想記憶領域からなる論理デバイスは、実記憶領域が既に割り当て済みであってもよいし、割り当て済みでなくともよい。

【0036】

この際、ストレージコントローラ131は、実記憶領域、または仮想記憶領域を、例えば、シリンダ・ヘッド番号（以下、トラック番号）によって参照したり、識別したりすることができる。

【0037】

ストレージコントローラ131は、例えば、記憶資源と、通信インタフェース装置（以下、インタフェース装置を「I/F」と略記）と、それらに接続されたCPU122とを有する。

【0038】

記憶資源は、例えば、主メモリ123、及びキャッシュメモリ124である。

【0039】

通信I/Fとしては、例えば、ホストI/F121と、管理ホストI/F127と、ディスクI/F125とを備える。これら主メモリ123、キャッシュメモリ124、CPU122、ホストI/F121、管理ホストI/F127、ディスクI/F125は、バス等の通信線からなるネットワーク126を介して相互に接続される。管理ホストI/F127は、例えば、NIC（Network Interface Card）である。

【0040】

各MFホスト計算機102や管理ホスト計算機103のハードウェア構成は、一般的な計算機と同じである。すなわち、各MFホスト計算機102や管理ホスト計算機103は、通信インタフェース装置と、記憶資源と、それらに接続されたCPUとを有する。通信インタフェース装置としては、例えば、ネットワーク111を介した通信を行うためのホストバスアダプタ（HBA）と、ネットワーク113を介した通信を行うためのNICである。記憶資源は、例えば、メモリや内蔵HDDで構成される。

【0041】

図2に、ストレージコントローラ131内で実行するマイクロプログラム231の構成を示す。

【0042】

図2において、主メモリ123には、一つまたは複数のマイクロプログラム231がCPU122によって読み込まれる。このマイクロプログラム231には、コマンド制御部251、RAID制御部252、構成制御部253、アドレス変換部254が含まれる。CPU122が、主メモリ123に読み込まれたマイクロプログラム231内の各制御部251～254を実行することにより、後述する種々の処理が行われる。

【0043】

キャッシュメモリ124は、例えば、各MFホスト計算機102から受信したライトデータおよびマイクロプログラム231内のRAID制御部252がディスク132から読み出したデータを一時的に記憶するバッファを構成する。

【0044】

ホストI/F121は、ネットワーク111を介して、各MFホスト計算機102に接続され、各MFホスト計算機102から、アクセス要求としてアクセスコマンド（ライトコマンド又はリードコマンド）を受信し、受信したアクセスコマンドをコマンド制御部251に転送する。

【0045】

管理ホストI/F127は、ネットワーク113を介して、管理ホスト計算機103に接続され、管理ホスト計算機103から、例えば、後述するLDEVの容量拡張を行う指示や、LDEV間のデータコピー指示を受信した場合、受信した指示をコマンド制御部2

10

20

30

40

50

5 1 または構成制御部 2 5 3 に転送する。

【0046】

ディスク I / F 1 2 5 は、各ディスク 1 3 2 と記憶資源（主メモリ 1 2 3 とキャッシュメモリ 1 2 4）との間のデータ送受信を行う。ディスク I / F 1 2 5 は、通信路 1 1 2 を介して各ディスク 1 3 2 にそれぞれ接続される。

【0047】

次に、ストレージ装置 1 0 1 の基本的動作を簡単に説明する。

【0048】

ストレージコントローラ 1 3 1 は、ホスト I / F 1 2 1 を介して、いずれかの MF ホスト計算機 1 0 2 からライトコマンドを受信した場合、MF ホスト計算機 1 0 2 から受信したライトデータをキャッシュメモリ 1 2 4 に記憶する。

10

【0049】

ストレージコントローラ 1 3 1 は、キャッシュメモリ 1 2 4 に記憶しているライトデータを、ディスク I / F 1 2 5 を介してディスク 1 3 2 に書き込む。なお、ストレージコントローラ 1 3 1 は、ライトデータをキャッシュメモリ 1 2 4 に記憶した時点で、ライトコマンドの処理が完了した旨を MF ホスト計算機 1 0 2 に通知する構成でもよいし、或いは、ライトデータをディスク 1 3 2 に書き込んだ時点で、ライトコマンドの処理が完了した旨を MF ホスト計算機 1 0 2 に通知する構成でもよい。

【0050】

MF ホスト計算機 1 0 2 からリードコマンドを受信した場合、ストレージコントローラ 1 3 1 は、そのリードコマンドにパラメータで指定されるデータ（リード対象データ）がキャッシュメモリ 1 2 4 に記憶されているか否かを調べる。

20

【0051】

リード対象データがキャッシュメモリ 1 2 4 に記憶されている場合、ストレージコントローラ 1 3 1 は、リード対象データをキャッシュメモリ 1 2 4 から読み出し、読み出したリード対象データを、ホスト I / F 1 2 1 を介して MF ホスト計算機 1 0 2 に送信する。一方、リード対象データがキャッシュメモリ 1 2 4 に記憶されていない場合は、ストレージコントローラ 1 3 1 は、ディスク I / F 1 2 5 を介して、一つまたは複数のディスク 1 3 2 からリード対象データを読み出し、読み出したリード対象データをキャッシュメモリ 1 2 4 に記憶する。その後、ストレージコントローラ 1 3 1 は、キャッシュメモリ 1 2 4 に記憶されているリード対象データを、ホスト I / F 1 2 1 を介して MF ホスト計算機 1 0 2 に送信する。

30

【0052】

<ストレージ装置内の論理構成>

【0053】

図 3 に、ストレージ装置 1 0 1 内の論理構成を示す。

【0054】

図 3 において、ストレージ装置 1 0 1 は、各 MF ホスト計算機 1 0 2 や管理ホスト計算機 1 0 3 から参照される記憶領域を構成するホスト用論理デバイスとして、一つまたは複数のホスト用論理デバイス（以下、HDEV と称することがある。）3 2 1 を有する。

40

【0055】

HDEV 3 2 1 には、ストレージ装置 1 0 1 内で固有の HDEV 番号が付けられ、MF ホスト計算機 1 0 2 や管理ホスト計算機 1 0 3 は、HDEV 番号により、HDEV 3 2 1 を識別する。例えば、各 MF ホスト計算機 1 0 2 上の OS 3 3 1 は、HDEV 3 2 1 に対してリードアクセスまたはライトアクセスを行う。

【0056】

HDEV 3 2 1 は、例えば、トラックの集合（トラック 1、トラック 2、・・・）で構成されており、各トラックに付与されたシリンダ・ヘッド番号（トラック番号）は、各 MF ホスト計算機 1 0 2 や管理ホスト計算機 1 0 3 が、各トラックを識別するときの参照対象となる。

50

【0057】

HDEV321には、一つまたは複数の論理デバイス（以下、LDEVまたはボリュームと称することがある。）322が対応付けられている。

【0058】

LDEV322には、ストレージ装置101内で固有のLDEV番号が付けられ、マイクロプログラム231は、LDEV番号により、LDEV322を識別する。

【0059】

LDEV322は、例えば、一つまたは複数のディスク132内の記憶領域として定義される。また、LDEV322は、複数のRAIDグループで構成される記憶領域として定義されていてもよい。

10

【0060】

LDEV322は、トラックに対応した複数のデータ部325と複数の制御情報部326から構成される。各データ部325は、トラック（トラック番号）に対応した1または複数のレコード（図示せず）で構成され、各レコードには、MFホスト計算機102がリードまたはライトするデータが格納される。各制御情報部326は、各トラックのデータ部325をアクセスするための制御情報あるいは、各トラックのデータ部325を参照または更新するための制御情報で構成される。

【0061】

この際、ストレージ装置101内で、全てのトラックサイズは一定の値である。ストレージ装置101内で、全てのデータ部325と全ての制御情報部326のサイズは一定の値である。

20

【0062】

各ディスク132には、ストレージ装置101内で固有のディスク番号がつけられ、マイクロプログラム231は、ディスク番号により、各ディスク132を識別する。

【0063】

また、LDEV322は、後述するシン・プロビジョニング機能で用いられる仮想ボリューム上に保存されていてもよい。

【0064】

LDEV322のデータが保存される仮想ボリュームは、この仮想ボリュームに対して実記憶領域を提供するプール323と関連付けられる。

30

【0065】

プール323には、ストレージ装置101内で固有のプール番号がつけられ、マイクロプログラム231は、プール番号により、プール323を識別する。

【0066】

プール323は、一つまたは複数のディスク132から構成される。また、プール323は、一つまたは複数のRAIDグループから構成されていてもよい。一つまたは複数のディスク132は、ストレージ装置の内部にあってもよいし、外部にあってもよい。

【0067】

図4に、HDEV番号とLDEV番号との対応関係を管理するための管理テーブル400の構成を示す。

40

【0068】

管理テーブル400は、HDEV番号フィールド401とLDEV番号フィールド402から構成される。HDEV番号は、ストレージ装置101内でHDEV321を一意に識別するための番号であって、HDEV番号フィールド401の各エントリには、HDEV321に対応した番号が格納される。LDEV番号は、ストレージ装置101内でLDEV322を一意に識別するための番号であって、LDEV番号フィールド402の各エントリには、LDEV322に対応した番号が格納される。

【0069】

図5に、LDEV番号とボリューム属性との対応関係を管理するための管理テーブル500の構成を示す。

50

【0070】

管理テーブル500は、LDEV番号フィールド501とボリューム属性フィールド502及びボリューム容量フィールド503から構成される。LDEV番号は、ストレージ装置101内でLDEV322を一意に識別するための番号であって、LDEV番号フィールド501の各エントリには、LDEV322に対応した番号が格納される。ボリューム属性は、LDEV322のボリューム属性が、通常ボリュームまたは仮想ボリュームであることを特定するものであり、ボリューム属性フィールド502の各エントリには、通常ボリュームまたは仮想ボリュームの名称が格納される。

【0071】

LDEV322のボリューム属性が通常ボリュームであるとき、このLDEV322は、一つまたは複数のディスク132内の実記憶領域で構成されたボリュームであることを示す。

10

【0072】

ボリューム容量は、通常ボリュームまたは仮想ボリュームにデータを格納することができる最大容量であり、ボリューム容量フィールド503の各エントリには、通常ボリュームまたは仮想ボリュームの容量が数値で格納される。

【0073】

図6に、LDEV番号とディスク番号との対応関係を管理するための管理テーブル600の構成を示す。

【0074】

20

管理テーブル600は、LDEV322のボリューム属性が通常ボリュームであるときの管理テーブルであって、LDEV番号フィールド601とディスク番号フィールド602から構成される。LDEV番号は、ストレージ装置101内でLDEV322を一意に識別するための番号であって、LDEV番号フィールド601の各エントリには、LDEV322に対応した番号が格納される。ディスク番号は、ストレージ装置101内でディスク132を一意に識別するための番号であって、ディスク番号フィールド602の各エントリには、LDEV322を構成するディスク132の番号が格納される。

【0075】

図7に、LDEV番号とプール番号との対応関係を管理するための管理テーブル700の構成を示す。

30

【0076】

管理テーブル700は、LDEV322のボリューム属性が仮想ボリュームであるときの管理テーブルであって、LDEV番号フィールド701とプール番号フィールド702から構成される。LDEV番号は、ストレージ装置101内でLDEV322を一意に識別するための番号であって、LDEV番号フィールド701の各エントリには、LDEV322に対応した番号が格納される。プール番号は、ストレージ装置101内でプール323を一意に識別するための番号であって、プール番号フィールド702の各エントリには、LDEV322に対応したプール323の番号が格納される。

【0077】

図8に、プール番号とディスク番号との対応関係を管理するための管理テーブル800の構成を示す。

40

【0078】

管理テーブル800は、LDEV322のボリューム属性が仮想ボリュームであるときの管理テーブルであって、プール番号フィールド801とディスク番号フィールド802から構成される。プール番号は、ストレージ装置101内でプール323を一意に識別するための番号であって、プール番号フィールド801の各エントリには、プール323に対応した番号が格納される。ディスク番号は、ストレージ装置101内でディスク132を一意に識別するための番号であって、ディスク番号フィールド602の各エントリには、プール323に実記憶領域が割り当てられるディスク132の番号が格納される。

【実施例1】

50

【0079】

本実施例は、データ部と制御情報部を異なる領域に格納、ストレージコントローラが、ディスクからデータ部のみを読み書きする場合に、前述した制御情報部の位置をスキップする処理を省き、ディスクのヘッドを直接データ部に位置付けできるようにしたものである。

【0080】

この際、データ部を配置するユーザデータ領域と、このユーザデータ領域に連なる記憶領域であって、制御情報部を配置する制御情報領域を、それぞれ容量変更の対象となる記憶領域、例えば、容量を拡張または縮小する場合の記憶領域として管理し、記憶領域として、ユーザデータ領域と制御情報領域が並ぶデータ配置（第1データ配置）を採用するボリウム（LDEV）を3390形式のボリウムとして管理する。

10

【0081】

＜3390形式のLDEV322内のデータ配置＞

図9に、HDEV321を構成する3390形式のLDEV322内のデータ配置を示す。

【0082】

図9において、LDEV322の記憶領域901は、ユーザデータ領域902と制御情報領域903から構成される。ユーザデータ領域902は、複数のデータ部（トラックに対応した1または複数のレコードを含む）325を格納する第1データ領域として構成され、制御情報領域903は、ユーザデータ領域902に連なり、かつユーザデータ領域902に格納された各データ部325をそれぞれアクセスするための制御情報で構成された制御情報部326を格納する第1制御情報領域として構成される。この際、ストレージコントローラ131は、ユーザデータ領域902に連続して制御情報領域903が配置された記憶領域901のデータ配置を第1データ配置とする第1ボリウム形式（以下、3390形式と称することもある。）でLDEV322を管理する。マイクロプログラム231は、各トラック内のデータ部325をトラック番号により識別し、各トラック内の制御情報部326をトラック番号により識別する。

20

【0083】

3390形式のLDEV322は、HDEV321内のトラック数が大きくなると、ユーザデータ領域902および制御情報領域903内のトラック数も大きくなる。即ち、3390形式のLDEV322は、容量の拡張に伴ってHDEV321内のトラック数が増加した場合、HDEV321内のトラック数の増加に応じて、ユーザデータ領域902および制御情報領域903内のトラック数も増加する。

30

【0084】

3390形式のLDEV322のデータ配置としては、ユーザデータ領域902の手前に、制御情報領域903を配置することもできる。

【0085】

＜3390形式のLDEV322に対するライトコマンドまたはリードコマンド処理の一例＞

【0086】

以下、MFホスト計算機102が、ホストI/F121を介して、ストレージコントローラ131に対して、ライトコマンドを要求した場合の処理を説明する。ライトコマンドのパラメータは、HDEV番号とHDEV321内のライト対象トラックのトラック番号と、ライト対象データである。

40

【0087】

ストレージコントローラ131内のコマンド制御部251は、MFホスト計算機102からのライトコマンドを受信した場合、ライトコマンドのパラメータであるHDEV番号を基に管理テーブル400を参照し、HDEV番号に対応したLDEV番号を取得する。

【0088】

コマンド制御部251は、取得したLDEV番号からLDEV322を特定し、ライト

50

対象トラックのトラック番号を基に、特定したLDEV322上のトラック番号(HDEV321内のライト対象トラックのトラック番号と同一のトラック番号)のデータ部325のアドレスと、特定したLDEV322上のトラック番号(HDEV321内のライト対象トラックのトラック番号と同一のトラック番号)の制御情報部326のアドレスを計算する。

【0089】

次に、コマンド制御部251は、前述したライトコマンドが、トラック番号の制御情報部326を参照する必要があるかどうかを判定する。トラック番号の制御情報部326を参照する必要がある場合、例えば、MFホスト計算機102が、制御情報部326の参照結果として、ディスクに対するヘッドの位置決め状態に関する情報の返送を求めている場合、RAID制御部252は、LDEV番号を基に管理テーブル600を参照し、制御情報部326のアドレスに基づき、一つまたは複数のディスク132から、トラック番号の制御情報部326を、キャッシュメモリ124に読み込む。コマンド制御部251は、キャッシュメモリ124に読み込まれたトラック番号の制御情報部326を参照する。

10

【0090】

コマンド制御部251が、トラック番号の制御情報部326を参照する必要がなければ、前述した制御情報部326のリード処理は省略してもよい。

【0091】

即ち、データ部325と制御情報部326をそれぞれ異なる領域に格納し、データ部325を格納するトラックのサイズを設定した値とした場合、コマンド制御部251が、制御情報部326の制御情報を参照せずに、ディスク側でディスクのヘッドをデータ部325に直接位置づける。この場合、ディスクのヘッド位置付けにおいて、制御情報部326の領域をスキップする処理が不要となるので、ストレージコントローラ131は、制御情報部326を参照するときよりも、リードアクセスまたはライトアクセス処理を高速化できる。

20

【0092】

次に、コマンド制御部251は、ライト対象データを、キャッシュメモリ124に書き込む。

【0093】

この後、RAID制御部252は、LDEV番号を基に管理テーブル600を参照し、LDEV番号に対応したディスク番号を取得し、LDEV322を構成するディスク132を特定し、データ部325のアドレスに基づき、キャッシュメモリ124上のライト対象データを特定した一つまたは複数のディスク132に書き込む。即ち、3390形式のデータ配置が実記憶領域上のデータ配置と同じである場合、アドレス変換部254によるアドレス変換で得られたアドレスを用いることなく、RAID制御部252は、コマンド制御部251の計算結果によるアドレスを用いることで、キャッシュメモリ124上のライト対象データを、特定した一つまたは複数のディスク132に書き込むことができる。

30

【0094】

次に、ストレージコントローラ131は、ホストI/F121を介して、ライトコマンドの処理が完了した旨をMFホスト計算機102に通知する。

40

【0095】

以上が、MFホスト計算機102が、ホストI/F121を介して、ストレージコントローラ131に対して、ライトコマンドを要求した場合の処理である。

【0096】

次に、MFホスト計算機102が、ホストI/F121を介して、ストレージコントローラ131に対して、リードコマンドを要求した場合の処理を説明する。リードコマンドのパラメータは、HDEV番号とHDEV321内のリード対象トラックのトラック番号である。

【0097】

ストレージコントローラ131内のコマンド制御部251は、MFホスト計算機102

50

からのリードコマンドを受信した場合、リードコマンドのパラメータであるHDEV番号を基に管理テーブル400を参照し、HDEV番号に対応したLDEV番号を取得する。

【0098】

コマンド制御部251は、取得したLDEV番号でLDEV322を特定し、リード対象トラックのトラック番号（HDEV321内のリード対象トラックのトラック番号と同一のトラック番号）を基に、特定したLDEV322上のトラック番号のデータ部325のアドレスと、リード対象トラックのトラック番号（HDEV321内のリード対象トラックのトラック番号と同一のトラック番号）の制御情報部326のアドレスを計算する。

【0099】

次に、コマンド制御部251は、前述したリードコマンドが、トラック番号の制御情報部326を参照する必要があるかどうかを判定する。トラック番号の制御情報部326を参照する必要がある場合、RAID制御部252は、LDEV番号を基に管理テーブル600を参照し、制御情報部326のアドレスに基づき、一つまたは複数のディスク132から、トラック番号の制御情報部326を、キャッシュメモリ124に読み込む。コマンド制御部251は、キャッシュメモリ124に読み込まれたトラック番号の制御情報部326を参照する。トラック番号942の制御情報部326を参照する必要がなければ、前述した制御情報部326のリード処理は省略してもよい。

10

【0100】

次に、RAID制御部252は、LDEV番号を基に管理テーブル600を参照し、LDEV番号に対応したディスク番号を取得してディスク132を特定し、データ部325のアドレスに基づき、リード対象トラックのトラック番号のデータ部325を、特定した一つまたは複数のディスク132から、キャッシュメモリ124に読み込む。

20

【0101】

ストレージコントローラ131は、キャッシュメモリ124にある、リード対象トラックのトラック番号のデータ部325を、ホストI/F121を介して、MFホスト計算機102に送信する。

【0102】

以上が、MFホスト計算機102が、ホストI/F121を介して、ストレージコントローラ131に対して、リードコマンドを要求した場合の処理である。

【0103】

30

前述したリードコマンドに対する処理では、リード対象トラックのトラック番号のデータ部325または制御情報部326が、キャッシュメモリ124に既に存在していた場合、一つまたは複数のディスク132からキャッシュメモリ124に、リード対象トラックのトラック番号のデータ部325または制御情報部326を読み込む処理を省略することができる。

【0104】

また、リードコマンドに対する処理でも、コマンド制御部251が、制御情報部326の制御情報を参照せずに、ディスクのヘッドを直接データ部325に位置付けることができるので、ストレージコントローラ131は、制御情報部326を参照しない場合は、制御情報部326をスキップする処理が不要であるため、リードアクセスまたはライトアクセス処理を高速化できる。

40

【0105】

<3390形式のLDEV322の容量拡張>

【0106】

図10に、3390形式のLDEV322の容量拡張を実施する際の一例を示す。

【0107】

3390形式のLDEV322の容量を変更する必要がある場合、例えば、図9に示す3390形式のLDEV322の容量を拡張する必要がある場合、アクセス要求元が、アクセス対象のボリュームを3390形式のボリュームとして認識できるとともに、コマンド制御部251が、LDEV322を、容量拡張後でも3390形式のLDEV32

50

2としてアクセスできるように、ユーザデータ領域に連続して制御情報領域を形成するデータ配置が採用される。

【0108】

具体的には、容量拡張後における3390形式のLDENV322の記憶領域910は、分散ユーザデータ領域1001、1002と、分散制御情報領域1011、1012から構成される。

【0109】

即ち、容量拡張前のLDENV322の記憶領域が、分散ユーザデータ領域1001と分散制御情報領域1011で構成され、この記憶領域に、分散ユーザデータ領域1002と分散制御情報領域1012を追加する場合、分散ユーザデータ領域1001に連続して、分散ユーザデータ領域1002を、新たに確保された記憶領域として配置し、分散ユーザデータ領域1002に連続して分散制御情報領域1011を配置し、分散制御情報領域1011に連続して、新たに確保された記憶領域として分散制御情報領域1012を配置し、記憶領域910全体として、ユーザデータ領域に連続して制御情報領域を形成するデータ配置（3390形式によるデータ配置）としている。

10

【0110】

一方、ディスク132上に構築される実記憶領域920は、分散ユーザデータ領域1001、分散制御情報領域1011、分散ユーザデータ領域1002、分散制御情報領域1012の順で構成される。

【0111】

即ち、容量拡張前の実記憶領域の記憶領域が、分散ユーザデータ領域1001と分散制御情報領域1011で構成され、この記憶領域に、分散ユーザデータ領域1002と分散制御情報領域1012を追加する場合、分散制御情報領域1011に連続して、新たに確保された記憶領域である分散ユーザデータ領域1002を配置し、分散ユーザデータ領域1002に連続して、新たに確保された記憶領域である分散制御情報領域1012を配置するデータ配置としている。

20

【0112】

この際、マイクロプログラム231内の構成制御部253は、管理ホストI/F127を介して、管理ホスト計算機103から容量拡張指示を受領した場合、分散制御情報領域1011に連なる実記憶領域として、新たに確保した実記憶領域を形成し、この実記憶領域に、分散ユーザデータ領域1002と、分散制御情報領域1012とを連続して配置する。

30

【0113】

各分散ユーザデータ領域1001、1002と分散制御情報領域1011、1012は、図9の記憶領域901に構成されるユーザデータ領域902と制御情報領域903と同一のもので構成することができる。

【0114】

実記憶領域920上において、分散ユーザデータ領域1001と分散制御情報領域1011という2つの領域からなる実記憶領域のデータ配置は、3390形式のデータ配置となり、分散ユーザデータ領域1002と分散制御情報領域1012という2つの領域からなる実記憶領域のデータ配置は、3390形式のデータ配置となる。即ち、実記憶領域920上において、3390形式のデータ配置が2組構成されたことになる。

40

【0115】

分散ユーザデータ領域1002の大きさは、管理ホスト計算機103上の管理画面において指定される拡張容量によって決まり、例えば、前述したトラックサイズの整数倍となる。

【0116】

分散ユーザデータ領域1001のサイズと分散制御情報領域1011のサイズの比は、分散ユーザデータ領域1002のサイズと分散制御情報領域1012のサイズの比に等しい。

50

【0117】

また、記憶領域910に配置された分散ユーザデータ領域1002におけるアドレス1021が示すデータ部325のトラック番号と、実記憶領域920に配置された分散ユーザデータ領域1002におけるアドレス1022が指すデータ部325のトラック番号は等しい。

【0118】

図10では、容量拡張後の実記憶領域920には、分散ユーザデータ領域1001、分散制御情報領域1011、分散ユーザデータ領域1002、分散制御情報領域1012の順番で、データが配置されることになる（以下、このデータ配置を、実記憶領域上のデータ配置と称することもある）。

10

【0119】

コマンド制御部251が参照する記憶領域910のデータ配置と、実記憶領域920のデータ配置が異なる構成の場合、3390形式のデータ配置である記憶領域910におけるトラック番号のデータ部325を指すアドレス（以下、データアドレスと称することもある）1021と、実記憶領域上のデータ配置である実記憶領域920におけるトラック番号のデータ部325を指すアドレス（以下、データアドレスと称することもある）1022は、トラック番号が同一であっても、各分散ユーザデータ領域1002におけるアドレスの値が異なる。このため、マイクロプログラム231内のアドレス変換部254は、例えば、記憶領域910の分散ユーザデータ領域1002におけるトラック番号のデータ部325を指すアドレス1021を、実記憶領域920におけるトラック番号のデータ部325を指すアドレス1022に変換するためのアドレス変換を実行する。

20

【0120】

同様に、マイクロプログラム231内のアドレス変換部254は、3390形式のデータ配置上である記憶領域910におけるトラック番号の制御情報部326を指すアドレス（以下、制御情報アドレスと称することもある。）を、実記憶領域上のデータ配置である実記憶領域920におけるトラック番号の制御情報部326を指すアドレス（以下、制御情報アドレスと称することもある。）に変換するためのアドレス変換を行う。

【0121】

コマンド制御部251は、LDENV322の容量拡張後も、アドレス変換部254のアドレス変換で得られたアドレスを用いることで、実記憶領域上のデータ配置に従うボリュームを、3390形式のボリュームとしてアクセスすることができる。

30

【0122】

<3390形式のデータ配置上のアドレス1021と、実記憶領域上のデータ配置におけるアドレス1022のアドレス変換処理>

【0123】

図11に、分散ユーザデータ領域と分散制御情報領域のサイズを管理するための管理テーブル1100の構成を示す。

【0124】

図11において、管理テーブル1100は、主メモリ123またはキャッシュメモリ124に格納されるテーブルであって、iフィールド1101と、分散ユーザデータ領域iのサイズフィールド1102と、分散制御情報領域iのサイズフィールド1103から構成される。iフィールド1101の各エントリには、分散ユーザデータ領域と分散制御情報領域を識別するための番号が格納される。分散ユーザデータ領域iサイズフィールド1102の各エントリには、例えば、分散ユーザデータ領域1001、1002のサイズが数値（150、100）で格納される。分散制御情報領域iのサイズフィールド1103の各エントリには、例えば、分散制御情報領域1011、1012のサイズが数値（15、10）で格納される。

40

【0125】

分散ユーザデータ領域と分散制御情報領域のサイズに関する情報は、前述した3390形式のボリュームの容量拡張を実施する度に、管理テーブル1100のエントリに追加さ

50

れる。

【0126】

図11における管理テーブル1100の各エントリにおいて、分散ユーザデータ領域のサイズと、分散制御情報領域のサイズと、は一定ではない。但し、管理テーブル1100の各エントリにおいて、分散ユーザデータ領域のサイズと分散制御情報領域のサイズとの比は一定である。

【0127】

図12に、分散ユーザデータ領域と分散制御情報領域のアドレスを管理するための管理テーブル1200を示す。

【0128】

図12において、管理テーブル1200は、iフィールド1201と、分散ユーザデータ領域iの先頭アドレスフィールド(3390形式)1202と、分散ユーザデータ領域iの先頭アドレスフィールド(実記憶領域上のデータ配置)1203、分散制御情報領域iの先頭アドレスフィールド1204と、分散制御情報領域iの先頭アドレスフィールド(実記憶領域上のデータ配置)1205から構成され、例えば、主メモリ123またはキャッシュメモリ124に格納される。

【0129】

iフィールド1201の各エントリには、分散ユーザデータ領域と分散制御情報領域を識別するための番号が格納される。分散ユーザデータ領域iの先頭アドレスフィールド1202の各エントリには、例えば、記憶領域910における分散ユーザデータ領域1001、1002の先頭アドレスが、数値(0、150)で格納される。分散ユーザデータ領域iの先頭アドレスフィールド1203の各エントリには、例えば、実記憶領域920における分散ユーザデータ領域1001、1002の先頭アドレスが、数値(0、165)で格納される。

【0130】

分散制御情報領域iの先頭アドレスフィールド1204の各エントリには、例えば、記憶領域910における分散制御情報領域1011、1012の先頭アドレスが、数値(250、265)で格納される。分散制御情報領域iの先頭アドレスフィールド1205の各エントリには、例えば、実記憶領域920における分散制御情報領域1011、1012の先頭アドレスが、数値(150、165)で格納される。

【0131】

分散ユーザデータ領域と分散制御情報領域の先頭アドレスに関する情報は、前述した3390形式のボリュームの容量拡張を実施する度に、管理テーブル1200のエントリに追加される。

【0132】

次に、前述した3390形式のデータ配置上のトラック番号のデータ部325を指すアドレスから、実記憶領域920上のデータ配置のトラック番号のデータ部325を指すアドレスを計算する処理を図13のフローチャートに従って説明する。

【0133】

マイクロプログラム231内のアドレス変換部254は、まず、iを0とし(S11)、管理テーブル1100、1200を参照して、3390形式上の0番目(i番目)の分散ユーザデータ領域の先頭アドレスとサイズに関する情報を取得し(S12)、次に、例えば、記憶領域910における分散ユーザデータ領域1002におけるアドレス1021が、0番目(i番目)の分散ユーザデータ領域に含まれるか否かを判定し(S13)、このステップS13で否定の判定結果を得たときには、iをインクリメント(例えば、iを1に)し(S14)、ステップS13で肯定の判定結果が得られるまで、ステップS12～S14の処理を繰り返す。

【0134】

ステップS13で肯定の判定結果が得られた場合、例えば、記憶領域910における分散ユーザデータ領域1002におけるアドレス1021は、2番目の分散ユーザデータ領

10

20

30

40

50

域 1 0 0 2 に含まれる場合、アドレス変換部 2 5 4 は、管理テーブル 1 1 0 0 に格納された各分散ユーザデータ領域と分散制御領域のサイズと、管理テーブル 1 2 0 0 に格納された各分散ユーザデータ領域と分散制御情報領域のアドレスと、3 3 9 0 形式のデータ配置上でトラック番号のデータ部 3 2 5 を指すアドレス 1 0 2 1 と、を基に、アドレス 1 0 2 1 が含まれる分散ユーザデータ領域 1 0 0 2 の番号 (2) と、分散ユーザデータ領域 1 0 0 2 の先頭アドレスからアドレス 1 0 2 1 までのオフセットアドレスと、を計算する (S 1 5)。

【0 1 3 5】

次に、アドレス変換部 2 5 4 は、分散ユーザデータ領域と分散制御情報領域のアドレスと、分散ユーザデータ領域 1 0 0 2 の番号 (2) と、を基に、管理テーブル 1 2 0 0 を参照して、分散ユーザデータ領域 1 0 0 2 の番号 (2) が表す分散ユーザデータ領域 1 0 0 2 の先頭アドレス (実記憶領域 9 2 0 における分散ユーザデータ領域 1 0 0 2 の先頭アドレス) を取得する (S 1 6)。

10

【0 1 3 6】

最後に、アドレス変換部 2 5 4 は、前述した実記憶領域 9 2 0 における分散ユーザデータ領域 1 0 0 2 の先頭アドレスと、記憶領域 9 1 0 における先頭アドレスからアドレス 1 0 2 1 までのオフセットアドレスと、を基に、実記憶領域 9 2 0 上のトラック番号のデータ部 3 2 5 を指すアドレス 1 0 2 2 を計算する (S 1 7)。

【0 1 3 7】

上記処理を実行することで、アドレス変換部 2 5 4 は、3 3 9 0 形式のデータ配置上のトラック番号のデータ部 3 2 5 を指すアドレス 1 0 2 1 を、実記憶領域 9 2 0 上のデータ配置のトラック番号のデータ部 3 2 5 を指すアドレス 1 0 2 2 に変換することができる。

20

【0 1 3 8】

また、アドレス変換部 2 5 4 は、アドレス 1 0 2 1 をアドレス 1 0 2 2 に変換する処理と同様の処理を実行することで、前述した 3 3 9 0 形式のデータ配置上のトラック番号 1 0 3 1 の制御情報部 3 2 6 を指すアドレスから、実記憶領域 9 2 0 上のトラック番号の制御情報部 3 2 6 を指すアドレスを計算することができる。

【0 1 3 9】

本実施例によれば、トラックの制御情報部 3 2 6 をスキップして、ディスクのヘッドをデータ部 3 2 5 に位置付ける処理のオーバーヘッドを減らせるので、ストレージコントローラ 1 3 1 は、トラックの制御情報部 3 2 6 を参照するときよりも、ディスクに対するリードアクセスまたはライトアクセスの性能を向上させることができる。

30

【0 1 4 0】

また、本実施例によれば、3 3 9 0 形式の記憶領域 9 1 0 の容量が拡張された場合でも、3 3 9 0 形式のデータ配置上のトラック番号のデータ部 3 2 5 を指すアドレス 1 0 2 1 を、実記憶領域 9 2 0 上のデータ配置のトラック番号のデータ部 3 2 5 を指すアドレス 1 0 2 2 に変換できるとともに、3 3 9 0 形式のデータ配置上のトラック番号の制御情報部 3 2 6 を指すアドレスを、実記憶領域 9 2 0 上のトラック番号の制御情報部 3 2 6 を指すアドレスに変換することができる。

【0 1 4 1】

40

このため、コマンド制御部 2 5 1 は、3 3 9 0 形式のデータ配置である、L D E V 3 2 の容量拡張後も、アドレス変換部 2 5 4 のアドレス変換で得られたアドレスを用いることで、実記憶領域上のデータ配置に従うボリュームを、3 3 9 0 形式のボリュームとしてアクセスすることができる。

【実施例 2】

【0 1 4 2】

本実施例は、一つの分散ユーザデータ領域 (第 2 データ領域) と一つの分散制御情報領域を一つのシリンダグループ (C G) として、シリンダグループを複数グループ有し、各シリンダグループを容量拡張単位の記憶領域として管理し、記憶領域として、複数のシリンダグループが並ぶデータ配置 (第 2 データ配置) を採用するボリュームを第 2 ボリューム

50

ム形式（以下、３３９０－Ａ形式と称することもある。）のボリューム（ＬＤＥＶ３２２）として管理し、３３９０形式のボリュームに関するアドレスを３３９０－Ａ形式のボリュームに関するアドレスに変換するものである。

【０１４３】

< ３３９０－Ａ形式のＬＤＥＶ３２２内のデータ配置 >

【０１４４】

図１４に、３３９０－Ａ形式のＬＤＥＶ３２２内のデータ配置を示す。

【０１４５】

図１４において、ＬＤＥＶ３２２の実記憶領域９３０は、分散ユーザデータ領域１４０１と、分散ユーザデータ領域１４０１に連なる分散制御情報領域１４１１を含むシリンダグループ（ＣＧ）１４２１と、分散ユーザデータ領域１４０２と、分散ユーザデータ領域１４０２に連なる分散制御情報領域１４１２を含むシリンダグループ（ＣＧ）１４２２から構成される。各分散ユーザデータ領域１４０１、１４０２は、それぞれ同一のサイズで構成され、各分散制御情報領域１４１１、１４１２もそれぞれ同一のサイズで構成される。即ち、各シリンダグループ（ＣＧ）１４１１、１４２２はそれぞれ同一のサイズで構成され、シリンダグループの容量が、容量拡張の単位になっている。

10

【０１４６】

この際、ストレージコントローラ１３１は、例えば、シリンダグループ（ＣＧ）１４２１を構成する記憶領域であって、分散ユーザデータ領域（第２データ領域）１４０１に連続して分散制御情報領域（第２制御情報領域）１４１１が配置された実記憶領域９３０のデータ配置を第２データ配置とする第２ボリューム形式（以下、３３９０－Ａ形式と称することもある。）でＬＤＥＶ３２２を管理する。

20

【０１４７】

３３９０－Ａ形式のＬＤＥＶ３２２を構成する一つまたは複数のＣＧ１４２１は、ＣＧ番号によって参照されてもよい。また、当該ＬＤＥＶ３２２を構成する一つまたは複数の分散ユーザデータ領域１４０１は、後述する分散ユーザデータ領域の番号によって参照されてもよい。また、当該ＬＤＥＶ３２２を構成する一つまたは複数の分散制御情報領域１４１１は、後述する分散制御情報領域の番号によって参照されてもよい。

【０１４８】

図１５に、シリンダグループ１４２１内の構成を示す。

30

【０１４９】

シリンダグループ（ＣＧ）１４２１は、分散ユーザデータ領域１４０１と分散制御情報領域１４１１から構成される。分散ユーザデータ領域１４０１は、複数（Ｄ個）のデータ部（トラック１～Ｄに対応した１または複数のレコードを含む）３２５を格納する第２データ領域として構成される。

【０１５０】

分散制御情報領域１４１１は、分散ユーザデータ領域１４０１に連なり、かつ分散ユーザデータ領域１４０１に格納された各データ部３２５をそれぞれアクセスするための制御情報で構成された複数（Ｄ個）の制御情報部３２６と、各データ部３２５の各トラックにレコードが存在するか否かを示すレコード有無情報で構成された複数（Ｄ個）のレコード有無情報部３２７と、ＬＤＥＶ３２２をストレージコントローラ１３１が管理するためのストレージ機能管理情報３２８と、ストレージコントローラ１３１がシリンダグループを拡張するためのモードか否かを判定するための拡張モード判定情報３２９を格納する第２制御情報領域として構成される。

40

【０１５１】

この際、マイクロプログラム２３１は、各トラック内のデータ部３２５をトラック番号により識別し、各トラック内の制御情報部３２６をトラック番号により識別する。

【０１５２】

各制御情報部３２６と各レコード有無情報部３２７は、例えば、ビットマップで構成することもできる。レコードは、各トラックのデータ部３２５に一つまたは複数保存される

50

。このレコードは、例えば、MFホスト計算機102からのライトデータなどである。

【0153】

また、各シリンダグループ1421、1422において、分散ユーザデータ領域1401、1402内のトラックのデータ部325の数(D個)と、分散制御情報領域1411、1412内のトラックの制御情報部326の数(D個)及び分散制御情報領域1411、1412内のトラックのレコード有無情報部327の数(D個)は等しい。

【0154】

図16に、3390形式のLDEV322上のアドレスと3390-A形式のLDEV322上のアドレスの対応を示す。

【0155】

図16において、3390形式のLDEV322の記憶領域910は、分散ユーザデータ領域1401、1402、1403と、分散制御情報領域1411、1412、1413から構成され、3390-A形式のLDEVの実記憶領域930は、分散ユーザデータ領域1401と分散制御情報領域1411を含むシリンダグループ1421と、分散ユーザデータ領域1402と分散制御情報領域1412を含むシリンダグループ1422と、分散ユーザデータ領域1403と分散制御情報領域1413を含むシリンダグループ1423から構成される。

【0156】

3390形式のボリュームのデータ配置は、分散ユーザデータ領域1401、分散ユーザデータ領域1402、分散ユーザデータ領域1403、分散制御情報領域1411、分散制御情報領域1412、分散制御情報領域1413の順番であり、3390-A形式のボリュームのデータ配置は、シリンダグループ1421、シリンダグループ1422、シリンダグループ1423の順番である。

【0157】

3390形式のボリュームのデータ配置と3390-A形式のボリュームのデータ配置が異なる構成を採用すると、3390形式のデータ配置である記憶領域910におけるトラック番号のデータ部325を指すアドレスと、3390-A形式のデータ配置である実記憶領域930におけるトラック番号のデータ部325を指すアドレスは、トラック番号が同一であっても、各分散ユーザデータ領域や各分散制御情報領域におけるアドレスの値が異なる。このため、マイクロプログラム231内のアドレス変換部254は、例えば、記憶領域910の分散ユーザデータ領域1403におけるトラック番号のデータ部325を指すアドレス1031を、実記憶領域930におけるトラック番号のデータ部325を指すアドレス1032に変換するためのアドレス変換を実行する。

【0158】

同様に、マイクロプログラム231内のアドレス変換部254は、3390形式のデータ配置上である記憶領域910におけるトラック番号の制御情報部326を指すアドレスを、3390-A形式のデータ配置である実記憶領域930におけるトラック番号の制御情報部326を指すアドレスに変換するためのアドレス変換を行う。

【0159】

コマンド制御部251は、アドレス変換部254のアドレス変換により得られたアドレスを用いることで、実記憶領域上は3390-A形式のデータ配置に従うLDEV322を、3390形式のLDEV322としてアクセスすることができる。即ち、コマンド制御部251は、3390形式のデータ配置のLDEV322にアクセスしても、3390形式のデータ配置のLDEV322にアドレスが、アドレス変換部254のアドレス変換によって、3390-A形式のデータ配置に従うLDEV322のアドレスが変換されるので、結果として、実記憶領域上のデータ部325をアクセスすることができる。

【0160】

<3390-A形式のLDEV322の容量拡張>

【0161】

図17に、分散ユーザデータ領域と分散制御情報領域のサイズを管理するための管理テ

10

20

30

40

50

ーブル1700の構成を示す。

【0162】

図17において、管理テーブル1700は、3390-A形式のLDENV322を管理するためのテーブルとして、分散ユーザデータ領域長フィールド1701と、分散制御情報領域長フィールド1702から構成され、例えば、主メモリ123またはキャッシュメモリ124に格納される。分散ユーザデータ領域長フィールド1701のエントリには、各分散ユーザデータ領域共通の長さを示す情報が数値で格納される。分散制御情報領域長フィールド1702のエントリには、各分散制御情報領域共通の長さを示す情報が数値で格納される。

【0163】

図18に、分散ユーザデータ領域と分散制御情報領域のアドレスを管理するための管理テーブル1800の構成を示す。

【0164】

図18において、管理テーブル1800は、iフィールド1801と、分散ユーザデータ領域iの先頭アドレスフィールド(3390形式)1802と、分散ユーザデータ領域iの先頭アドレスフィールド(3390-A形式のデータ配置)1803、分散制御情報領域iの先頭アドレスフィールド(3390形式)1804と、分散制御情報領域iの先頭アドレスフィールド(3390-A形式のデータ配置)1805から構成され、例えば、主メモリ123またはキャッシュメモリ124に格納される。

【0165】

iフィールド1201の各エントリには、分散ユーザデータ領域と分散制御情報領域を識別するための番号が格納される。分散ユーザデータ領域iの先頭アドレスフィールド1802の各エントリには、例えば、記憶領域910における分散ユーザデータ領域1401、1402の先頭アドレスが数値(0、100)で格納される。分散ユーザデータ領域iの先頭アドレスフィールド1803の各エントリには、例えば、実記憶領域930における分散ユーザデータ領域1401、1402の先頭アドレスが数値(0、150)で格納される。

【0166】

分散制御情報領域iの先頭アドレスフィールド1804の各エントリには、例えば、記憶領域910における分散制御情報領域1411、1412の先頭アドレスが数値(500、550)で格納される。分散制御情報領域iの先頭アドレスフィールド1805の各エントリには、例えば、実記憶領域930における分散制御情報領域1411、1412の先頭アドレスが数値(100、250)で格納される。

【0167】

各分散ユーザデータ領域と各分散制御情報領域の先頭アドレスに関する情報は、3390形式のボリュームの容量拡張を実施する度に、管理テーブル1800のエントリに追加される。

【0168】

次に、3390-A形式のLDENV322の容量拡張処理を図19のフローチャートに従って説明する。

【0169】

管理ホスト計算機103が、管理ホストI/F127を介して、構成制御部253にLDENV322の容量拡張指示を行うと、構成制御部253は、管理テーブル1700を参照し、分散ユーザデータ領域のサイズ(100)と分散制御情報領域のサイズ(50)を取得する(S21)。

【0170】

次に、構成制御部253は、管理ホスト103において指定された拡張容量が、取得した分散ユーザデータ領域のサイズ(100)の整数倍であるかどうかをチェックする(S22)。

【0171】

10

20

30

40

50

この管理画面で指定される拡張容量が、分散ユーザデータ領域のサイズ（１００）の整数倍でない場合、構成制御部２５３は、このＬＤＥＶ３２２の容量拡張は失敗したことを、管理ホストＩ／Ｆ１２７を介して、管理ホスト計算機１０３に通知して、容量拡張処理を終了する。

【０１７２】

一方、拡張容量が、分散ユーザデータ領域のサイズ（１００）の整数倍である場合、構成制御部２５３は、拡張容量を分散ユーザデータ領域のサイズ（１００）で割った商を、拡張シリンダグループ数として、主メモリ１２３に保存する。

【０１７３】

次に、構成制御部２５３は、前述した拡張シリンダグループ数の数だけ、シリンダグループ（ＣＧ）を保存できる記憶領域（以下、拡張ＣＧ領域と称することもある。）を確保する（Ｓ２３）。この場合、容量拡張において、確保する拡張ＣＧ領域のサイズは、拡張前領域の容量単位であるＣＧのサイズの整数倍である。

【０１７４】

この後、構成制御部２５３は、拡張する容量分のＣＧ領域確保に成功したか否かを判定し（Ｓ２４）、拡張ＣＧ領域の確保に失敗した場合、管理ホストＩ／Ｆ１２７を介して、ＬＤＥＶ３２２の容量拡張に失敗したことを管理ホスト計算機１０３に通知して、容量拡張処理を終了する。

【０１７５】

拡張ＣＧ領域の確保に成功した場合、構成制御部２５３は、ボリューム管理情報内のボリューム容量に変更後のサイズを設定し（Ｓ２５）、ホストＩ／Ｆ１２１を介して、ＭＦホスト計算機１０２に対して容量変更が行われたことを通知するとともに、容量拡張が成功したことを通知し、このルーチンでの処理を終了する。

【０１７６】

なお、拡張ＣＧ領域上に存在する、一つまたは複数の分散制御情報領域は予め初期化していてもいいし、していなくても良い。

【０１７７】

また、容量拡張指示は、前述したＭＦホスト計算機１０２およびストレージ装置１０１からなるストレージシステムがオンライン中に行ってもよい。

【０１７８】

図２０に、容量拡張後における３３９０形式のＬＤＥＶ３２２上のアドレスと３３９０－Ａ形式のＬＤＥＶ３２２上のアドレスの対応を示す。

【０１７９】

マイクロプログラム２３１内の構成制御部２５３が、管理ホストＩ／Ｆ１２７を介して、管理ホスト計算機１０３から容量拡張指示を受領すると、構成制御部２５３は、３３９０形式のＬＤＥＶ３２２上のデータ配置として、分散ユーザデータ領域１４０３の後に、分散ユーザデータ領域１４０４、１４０５を配置し、３３９０－Ａ形式のＬＤＥＶ３２２上のデータ配置として、シリンダグループ１４２３の後ろに、新たに確保した実記憶領域として、拡張シリンダグループ１４２４、１４１５を配置する。

【０１８０】

図２０において、３３９０形式のＬＤＥＶ３２２の記憶領域９１０は、分散ユーザデータ領域１４０１、１４０２、１４０３、１４０４、１４０５と、分散制御情報領域１４１１、１４１２、１４１３、１４１４、１４１５から構成され、３３９０－Ａ形式のＬＤＥＶの実記憶領域９３０は、分散ユーザデータ領域１４０１と分散制御情報領域１４１１を含むシリンダグループ１４２１と、分散ユーザデータ領域１４０２と分散制御情報領域１４１２を含むシリンダグループ１４２２と、分散ユーザデータ領域１４０３と分散制御情報領域１４１３を含むシリンダグループ１４２３と、分散ユーザデータ領域１４０４と分散制御情報領域１４１４を含む拡張シリンダグループ１４２４と、分散ユーザデータ領域１４０５と分散制御情報領域１４１５を含む拡張シリンダグループ１４２５から構成される。

10

20

30

40

50

【0181】

3390形式のボリュームのデータ配置は、分散ユーザデータ領域1401、分散ユーザデータ領域1402、分散ユーザデータ領域1403、分散ユーザデータ領域1404、分散ユーザデータ領域1405、分散制御情報領域1411、分散制御情報領域1412、分散制御情報領域1413、分散制御情報領域1414、分散制御情報領域1415の順番であり、実記憶領域となる3390-A形式のボリュームのデータ配置は、シリンダグループ(CG)1421、シリンダグループ1422、シリンダグループ1423、拡張シリンダグループ1424、拡張シリンダグループ1425の順番である。

【0182】

3390形式のボリュームのデータ配置と3390-A形式のボリュームのデータ配置が異なる構成を採用すると、3390形式のデータ配置である記憶領域910におけるトラック番号のデータ部325を指すアドレスと、3390-A形式のデータ配置である実記憶領域930におけるトラック番号のデータ部325を指すアドレスは、トラック番号が同一であっても、各分散ユーザデータ領域や各分散制御情報領域におけるアドレスの値が異なる。このため、マイクロプログラム231内のアドレス変換部254は、例えば、記憶領域910の分散ユーザデータ領域1405におけるトラック番号のデータ部325を指すアドレス1051を、実記憶領域930におけるトラック番号のデータ部325を指すアドレス1052に変換するためのアドレス変換を実行する。

10

【0183】

同様に、マイクロプログラム231内のアドレス変換部254は、3390形式のデータ配置上である記憶領域910におけるトラック番号の制御情報部326を指すアドレスを、3390-A形式のデータ配置である実記憶領域930におけるトラック番号の制御情報部326を指すアドレスに変換するためのアドレス変換を行う。

20

【0184】

コマンド制御部251は、3390-A形式のLDEV322の容量拡張後も、アドレス変換部254のアドレス変換により得られたアドレスを用いることで、実記憶領域上は3390-A形式のデータ配置に従うLDEV322を、3390形式のLDEV322としてアクセスすることができる。

【0185】

次に、3390形式のボリューム上のトラック番号のデータ部を指すアドレス1051から、3390-A形式のボリューム上のトラック番号のデータ部を指すアドレス1052を計算する処理を図21のフローチャートに従って説明する。

30

【0186】

マイクロプログラム231内のアドレス変換部254は、アドレス変換処理の開始とともに、3390形式のデータ配置上でトラック番号のデータ部325を指すアドレス1051を基に、管理テーブル1700と管理テーブル1800を参照し、アドレス1051が含まれる、3390形式の分散ユーザデータ領域1405の番号(5)と、3390形式の分散ユーザデータ領域1405の先頭アドレス(400)からアドレス1051までのオフセットアドレスと、を計算する(S31)。

【0187】

次に、アドレス変換部254は、アドレス1051が含まれる、3390形式の分散ユーザデータ領域1405の番号(5)と、3390形式の分散ユーザデータ領域1405の先頭アドレス(400)からアドレス1051までのオフセットアドレスと、を基に、管理テーブル1800を参照して、分散ユーザデータ領域1405の番号(5)が表す、3390-A形式の分散ユーザデータ領域1405の先頭アドレス(600)を取得する(S32)。

40

【0188】

最後に、アドレス変換部254は、取得した3390-A形式上の分散ユーザデータ領域1405の先頭アドレス(600)と、3390形式の分散ユーザデータ領域1405の先頭アドレス(400)からアドレス1051までのオフセットアドレスと、を基に、

50

3390 - A形式上のトラック番号のデータ部325を指すアドレス1052を計算し（S33）、このルーチンでの処理を終了する。

【0189】

同様に、アドレス変換部254は、前述した3390形式のデータ配置上のトラック番号の制御情報部326を指すアドレスから、3390 - A形式上のトラック番号の制御情報部326を指すアドレスを計算することもできる。

【0190】

＜3390 - A形式のLDEV322に対するリードコマンド処理の一例＞

【0191】

次に、ストレージコントローラ131が、ホストI/F121を介して、MFホスト計算機102からリードコマンドを受信した場合の処理を説明する。

【0192】

マイクロプログラム231内のコマンド制御部251は、前述した3390形式のLDEV322内のリード対象トラックのトラック番号のデータ部325のアドレスを計算し、計算結果をアドレス変換部254に通知する。

【0193】

次に、アドレス変換部254は、計算で得られたアドレスを基に実記憶領域上のトラックのデータ部325のアドレスと制御情報部326のアドレスを計算し、これらの値をコマンド制御部251に通知する。この際、LDEV322の実記憶領域上のデータ配置は3390 - A形式に従って構成されるものとする。

【0194】

コマンド制御部251が受信したアドレスが指すトラック番号のデータ部325と、コマンド制御部251が受信したアドレスが指すトラック番号の制御情報部326がキャッシュメモリ124に記憶されていない場合、コマンド制御部251は、RAID制御部252にデータの読み出しを指示する。RAID制御部252は、ディスクI/F125を介して、一つまたは複数のディスク132から、各アドレスに対応するデータを読み出し、それぞれ読み出したデータをキャッシュメモリ124に記憶する。

【0195】

コマンド制御部251が受信したアドレスが指すトラック番号のデータ部325と、コマンド制御部251が受信したアドレスが指すトラック番号の制御情報部326がキャッシュメモリ124に記憶されている場合、コマンド制御部251は、アドレスが指すトラックのデータ部325をキャッシュメモリ124から読み出し、アドレスが指すトラックの制御情報の値を確認した後に、キャッシュメモリ124から読み出したデータ部325をホストI/F121を介して、MFホスト計算機102に送信する。

【0196】

前述したリードコマンド処理のアドレス変換に関わる処理は、ライトコマンド処理においても同様に実行することができる。

【0197】

本実施例によれば、コマンド制御部251は、3390 - A形式のLDEV322の容量拡張後も、アドレス変換部254のアドレス変換により得られたアドレスを用いることで、実記憶領域上は3390 - A形式のデータ配置に従うLDEV322を、3390形式のLDEV322としてアクセスすることができる。

【実施例3】

【0198】

本実施例は、実記憶領域上のボリュームであって、3390 - A形式のボリュームを複数ボリューム用意し、一方のボリュームに分散ユーザデータ領域を配置し、他方のボリュームに分散制御情報領域を配置して、実記憶領域上のボリュームを管理するものである。

【0199】

＜ユーザデータ領域と制御情報領域を異なるボリュームに保存して管理する方式＞

【0200】

図22に、シリンダグループ1421内の分散ユーザデータ領域1401と分散制御情報領域1411を違うボリュームに保存して管理する一例を示す。

【0201】

図22において、HDEV2201は、HDEV321と同一の構成であり、このHDEV2201には、複数(m個)のトラックが配置され、2つのLDEV2202、2203が対応付けられている。LDEV2202は、複数(n個)の分散ユーザデータ領域1401、・・・を格納するボリュームとして構成され、LDEV2203は、複数(n個)の分散制御情報領域1411、・・・を格納するボリュームとして構成される。

【0202】

この際、シリンダグループ1421を分散ユーザデータ領域1401と分散制御情報領域1411で構成した場合、分散ユーザデータ領域1401がLDEV2202に配置され、分散制御情報領域1411が、LDEV2202とは異なるLDEV2203に配置される。このため、LDEV2202を大容量のもので構成し、LDEV2203を小容量のもので構成しても、ボリュームの容量拡張に対応することができる。

【0203】

また、例えば、LDEV2202に保存した分散ユーザデータ領域1401と、LDEV2203に保存した分散ユーザデータ領域1401に対応する分散制御情報領域1411を組み合わせ、一つのシリンダグループと見なしてもよい。さらに、LDEV2202の各分散ユーザデータ領域と、各分散ユーザデータ領域に対応するLDEV2203の制御情報領域を組み合わせ、一つまたは複数のシリンダグループを纏めて、3390-A形式のLDEVと見なしてもよい。

【0204】

図23に、分散ユーザデータ領域と分散制御情報領域のアドレスを管理するための管理テーブル2300の構成を示す。

【0205】

図23において、管理テーブル2300は、分散ユーザデータ領域番号フィールド2301と、LDEV内のアドレスフィールド2302と、分散制御情報領域番号フィールド2303と、LDEV内のアドレスフィールド2304から構成される。

【0206】

分散ユーザデータ領域番号フィールド2301の各エントリには、LDEV2202内に配置される分散ユーザデータ領域の番号が格納され、LDEV内のアドレスフィールド2302の各エントリには、LDEV2202内に配置される分散ユーザデータ領域のアドレスが格納される。分散制御情報領域番号フィールド2303の各エントリには、LDEV2203内に配置される分散制御情報領域の番号が格納され、LDEV内のアドレスフィールド2304の各エントリには、LDEV2203内に配置される分散制御情報領域のアドレスが格納される。

【0207】

次に、MFホスト計算機102が、前述したHDEV2201に対してリードアクセスを行う場合の処理を説明する。

【0208】

MFホスト計算機102によるリードアクセスのパラメータは、例えば、HDEV番号とトラック番号である。

【0209】

MFホスト計算機102が、リードコマンドをストレージコントローラ131に送信すると、マイクロプログラム231内のコマンド制御部251は、MFホスト計算機102からのリードコマンドのパラメータに従ってHDEV2201を参照する。この後、マイクロプログラム231内のコマンド制御部251は、前述したHDEV番号とトラック番号を基に、LDEV番号と、3390形式のLDEV上におけるトラック内のデータ部325のアドレスと、制御情報部326のアドレスと、を計算し、アドレス変換部254に通知する。

10

20

30

40

50

【0210】

アドレス変換部254は、管理テーブル1700と管理テーブル1800を参照して、分散ユーザデータ領域と分散制御情報領域のサイズと、3390形式の分散ユーザデータ領域と分散制御情報領域の先頭アドレスを取得し、取得した情報と、トラック内のデータ部325のアドレスと、制御情報部326のアドレスと、に基づき、3390-A形式のLDEV内のLDEV番号と、トラック番号と、を計算し、コマンド制御部251に通知する。

【0211】

コマンド制御部251は、ホストI/F121を介して、MFホスト計算機102からリードアクセスを受領すると、受信したトラック番号から3390-A形式のLDEV内の分散ユーザデータ領域の番号と、分散ユーザデータ領域の先頭アドレスからトラック番号のデータ部325までのオフセットアドレスと、3390-A形式のLDEV内の分散制御情報領域の番号と、分散制御情報領域の先頭アドレスからトラック番号の制御情報部326までのオフセットアドレスと、を計算する。

10

【0212】

コマンド制御部251は、管理テーブル2300を参照して、分散ユーザデータ領域のアドレスを取得し、取得したアドレスと、分散ユーザデータ領域の番号と、オフセットアドレスと、に基づき、LDEV2202における分散ユーザデータ領域内のトラック番号のデータ部325を指すアドレスを取得する。

20

【0213】

また、コマンド制御部251は、管理テーブル2300を参照して、分散制御情報領域のアドレスを取得し、取得したアドレスと、分散制御情報領域の番号と、オフセットアドレスと、に基づき、LDEV2203における分散制御情報領域内のトラック番号の制御情報部326を指すアドレスを取得する。

【0214】

次に、コマンド制御部251は、取得したアドレスに基づき、分散ユーザデータ領域内のリード対象トラックのデータ部325に対してアクセスを行う。

【0215】

また、コマンド制御部251は、取得したアドレスに基づき、分散制御情報領域内のリード対象トラックの制御情報部326を参照し、または更新してもよい。なお、リード対象トラックの制御情報部326を参照する必要がある場合には、コマンド制御部251は、リード対象トラックの制御情報部326を参照する処理を省略することができる。

30

【0216】

コマンド制御部251は、リード対象トラック内のデータ部325をホストI/F121を介して、MFホスト計算機102に送信する。

【0217】

以上が、分散ユーザデータ領域と分散制御情報領域を異なるボリュームに保存して管理した場合の、リードアクセスの一例である。

【0218】

なお、MFホスト計算機102からのアクセスの種類がライトアクセスであった場合も、前述したトラックに対するアクセスの処理は同様に実行することができる。

40

【0219】

本実施例によれば、実記憶領域上のボリュームであって、3390-A形式のボリュームを複数ボリューム用意し、一方のボリュームにユーザデータ領域を配置し、他方のボリュームに制御情報領域を配置して、実記憶領域上のボリュームを管理するようにしたため、分散制御情報領域を配置するボリューム(LDEV2203)を小容量のもので構成し、分散ユーザデータ領域を配置するボリューム(LDEV2202)を大容量のもので構成することができる。この場合、LDEV2203を小容量のもので構成しても、LDEV2202を大容量のもので構成することで、ボリュームの容量拡張に対応することができる。

50

【実施例 4】

【0220】

本実施例は、3390形式のLDEV322から3390-A形式のLDEV322へのデータコピーを行うものである。

【0221】

<3390形式のLDEV322から3390-A形式のLDEV322へのデータコピー処理>

【0222】

次に、3390形式のLDEV322から、3390-A形式のLDEV322への、トラックのデータ部をコピーする処理を図24のフローチャートに従って説明する。

10

【0223】

まず、コピー元の3390形式のLDEV322は、アドレス(s)によって参照され、コピー先は、3390-A形式のLDEV322であって、3390-A形式のLDEV322は、一つまたは複数のシリンダグループ1421から構成されることを前提する。この際、3390-A形式のLDEV内のトラックのデータ部325は、例えば、分散ユーザデータ領域の番号(n)と、n番目の分散ユーザデータ領域の先頭アドレスからのオフセットアドレス(d)によってアクセスされる。

【0224】

管理ホスト計算機103から、管理ホストI/F127を介してコピー開始指示をコマンド制御部251が受領すると、マイクロプログラム231内のコマンド制御部251は、コピー処理を開始する。

20

【0225】

コマンド制御部251は、コピー元のLDEV322の先頭アドレスを、コピー元アドレス(s)に設定し、コピー元アドレス(s)を0にする(S41)。

【0226】

次に、コマンド制御部251は、コピー元アドレス(s)がコピー元LDEV322の終端アドレスか否かを判定し(S42)、コピー元アドレス(s)がコピー元LDEV322の終端アドレスを指していた場合、コピー処理が完了したことを、管理ホストI/F127を介して、管理ホスト計算機103に通知して、コピー処理を終了する。

【0227】

30

一方、コマンド制御部251は、コピー元アドレス(s)がコピー元LDEV322の終端アドレスを指していないと判定した場合、アドレス変換部254に、コピー元アドレス(s)を通知する。

【0228】

次に、アドレス変換部254は、管理テーブル1700を参照して、分散ユーザデータ領域のサイズを取得し、管理テーブル1800を参照して、分散ユーザデータ領域と分散制御情報領域の先頭アドレスを取得し、取得したサイズ及び先頭アドレス、コピー元アドレス(s)と、を基に、分散ユーザデータ領域の番号(n)と、コピー先の分散ユーザデータ領域(n)の先頭アドレスからのオフセットアドレス(d)の値を計算し、計算結果をコマンド制御部251に通知する(S43)。

40

【0229】

コマンド制御部251は、分散ユーザデータ領域(n)が、コピー先LDEVの終端分散ユーザデータ領域であるか否かを判定し(S44)、分散ユーザデータ領域(n)が、コピー先LDEVの終端分散ユーザデータ領域である場合、コピー元LDEVの容量が、コピー先LDEV322の容量より大きいため、コピーに失敗したことを、管理ホストI/F127を介して、管理ホスト計算機103に通知し、コピー処理を終了する。

【0230】

一方、コマンド制御部251は、分散ユーザデータ領域(n)が、コピー先LDEVの終端分散ユーザデータ領域でないと判定した場合、コピー先の分散ユーザデータ領域(n)内の先頭アドレスからのオフセットアドレス(d)の指す記憶領域に、コピー元アドレ

50

ス (s) の指す記憶領域からデータ (データ部 325) をコピーする (S45)。

【0231】

この後、コマンド制御部 251 は、コピー元アドレス (s) をインクリメントし ($s = s + 1$)、ステップ S32 の処理に戻る (S46)。この後、ステップ S32 で肯定の判定結果が得られるまで、ステップ S32 ~ S36 の処理が繰り返えされる。

【0232】

ステップ S41 ~ S46 の処理を実行することで、例えば、3390 形式の記憶領域 910 における分散ユーザデータ領域 1401 内のデータ部 325 を、3390 - A 形式の実記憶領域 930 における分散ユーザデータ領域 1401 内にコピーすることができる。

【0233】

また、3390 形式の LDEV322 から 3390 - A 形式の LDEV322 への、トラックの制御情報部 326 をコピーする処理も同様に実行することができる。この場合、例えば、3390 形式の記憶領域 910 における分散制御情報領域 1411 内の制御情報部 326 を、3390 - A 形式の実記憶領域 930 における分散制御情報領域 1411 内にコピーすることができる。

【0234】

コピーするデータ量は、コピー元 LDEV322 内の全部のデータ、または一部のデータのどちらであってもよい。

【0235】

前述した 3390 形式の LDEV322 から 3390 - A 形式の LDEV322 へのデータコピー処理では、分散ユーザデータ領域のコピーを終了した後、分散制御情報領域のコピーを行うか、その逆の順番で行ってもよい。また、一部の分散ユーザデータ領域のコピー終了後、一部の分散制御情報領域をコピーすることを交互に行っても良い。

【0236】

また、3390 - A 形式の LDEV322 から、3390 形式の LDEV322 へのデータコピーを行う処理も、3390 - A 形式の LDEV322 のアドレスを 3390 形式の LDEV322 のアドレスに変更する処理を行うことで、同様に実行することができる。

【0237】

本実施例によれば、3390 形式の LDEV322 から 3390 - A 形式の LDEV322 へのデータコピーを行うことで、3390 形式の LDEV322 に属するデータ部 325 と制御情報部 326 から、実記憶領域上のデータ部 325 と制御情報部 326 を構築することができる。

【実施例 5】

【0238】

本実施例は、LDEV322 を、仮想記憶領域を有する仮想ボリュームとして管理し、仮想ボリュームに対してアクセスがある場合、仮想ボリュームに、プールから実記憶領域を割り当てるようにしたものである。

【0239】

<仮想ボリュームへの実記憶領域割り当てに関わる処理>

【0240】

仮想ボリュームへの実記憶領域割り当てに関わる処理として、前述したアドレス変換処理と、後述のページ割り当て処理とがある。ページ割り当て処理では、後述するページ管理テーブルとページ管理ディレクトリを基に、仮想ボリュームに対するホストライトを契機とした仮想記憶領域への実記憶領域の割り当てが行われる。以下、仮想ボリュームへの実記憶領域割り当てに関わる処理を行う機能を、シン・プロビジョニング機能と呼ぶ。

【0241】

図 25 に、ページ割り当て処理で用いるページ管理テーブル 25000 の構成を示す。

【0242】

図 25 において、ページ管理テーブル 25000 は、仮想ボリュームへの実記憶領域割

10

20

30

40

50

り当て処理を実行するときに用いるテーブルであって、ページ番号フィールド2501と、LDEV内のページ先頭アドレスフィールド2502と、ディスク番号フィールド2503と、ディスク内のページ先頭アドレスフィールド2504と、割り当て済み判定情報フィールド2505から構成され、LDEV322毎に存在し、例えば、主メモリ123またはキャッシュメモリ124に格納される。

【0243】

ページ番号フィールド2501の各エントリには、LDEV322内に割り当てられるページのページ番号が格納される。LDEV内のページ先頭アドレスフィールド2502の各エントリには、LDEV322内に割り当てられるページの先頭アドレスが格納される。ディスク番号フィールド2503の各エントリには、LDEV322を構成するディスク132の番号が格納される。

10

【0244】

ディスク内のページ先頭アドレスは、ディスク内の実記憶領域の先頭アドレスであって、ディスク内のページ先頭アドレスフィールド2504の各エントリには、LDEV322を構成するディスク132の先頭アドレスが格納される。割り当て済み判定情報フィールド2505の各エントリには、LDEV322にページが割り当てられている場合には、「割り当て済み」の情報が格納され、LDEV322にページが割り当てられていない場合には、「未割り当て」の情報が格納される。即ち、この割り当て済み判定情報は、ページを単位として、仮想記憶領域への実記憶領域の割り当て状態を管理するために用いられる。

20

【0245】

ページは、プール323に格納される、一定のサイズの記憶領域であり、LDEV322の記憶領域を分割管理するための単位である。また、ページは、例えば、一つまたは複数のトラックの集まりである。トラックは、ストレージ装置101内において、一定サイズで、例えば、59392バイトで構成することができる。

【0246】

図26に、ページ割り当て処理で用いるページ管理ディレクトリ2600の構成を示す。

【0247】

図26において、ページ管理ディレクトリ2600は、LDEV322とページ管理テーブル25000との対応関係を管理するためのテーブルであって、LDEV番号フィールド2601と、ページ管理テーブルのアドレスフィールド2602から構成され、例えば、主メモリ123またはキャッシュメモリ124あるいはディスク132上に保存される。

30

【0248】

LDEV番号フィールド2601の各エントリには、LDEV322の番号が格納される。ページ管理テーブルのアドレスフィールド2602の各エントリには、LDEV322のページを管理するためのページ管理テーブル25000のアドレスが格納される。

【0249】

次に、コマンド制御部251が、MFホスト計算機102からのライトアクセスを受領した際の処理の一例を説明する。ライトアクセス時のパラメータは、例えば、LDEV番号とLDEV内のトラック番号である。

40

【0250】

コマンド制御部251は、MFホスト計算機102からライトアクセスのパラメータを受信した場合、LDEV322のトラック番号とトラックサイズを基に、LDEV322内のアドレスを計算する。

【0251】

この後、コマンド制御部251は、ページ管理ディレクトリ2600を参照して、LDEV番号に対応するページ管理テーブル25000のアドレスを取得する。

【0252】

50

次に、コマンド制御部 251 は、取得したアドレスに対応したページ管理テーブル 25000 を参照して、ライトアクセス先のアドレスが示すトラックを含むページを検索し、この検索で得られたページ番号と、ページ先頭アドレスからライトアクセス先トラックまでのオフセットアドレスを取得する。

【0253】

この後、コマンド制御部 251 は、取得したページ番号のページエントリ内の割り当て済み判定情報を参照し、割り当て済み判定情報が割り当て済み状態を示す場合は、このページエントリに対応したディスク内のページ先頭アドレスで特定された、ライトアクセス先トラック番号のデータ部 325 に対してライトアクセスを行う。

【0254】

割り当て済み判定情報が未割り当て状態の場合、コマンド制御部 251 は、管理テーブル 700 を参照して、LDEV 番号に対応したプール番号を取得し、取得したプール番号を基に管理テーブル 800 を参照して、プール番号に対応したディスク番号を取得し、取得したディスク番号を基に、LDEV 番号 322 に対応するプール 323 に属する一つまたは複数のディスク 132 から、ページ単位で実記憶領域の割り当てを行う。

【0255】

この際、コマンド制御部 251 は、ページ管理テーブル 25000 を参照して、ライトアクセス先のアドレスが示すトラックを含むページエントリを探し、新たに割り当てたページの先頭アドレスをページ管理テーブル 25000 に登録し、割り当て済み判定情報フィールド 2502 の「未割り当て」を「割り当て済み」として、ページ管理テーブル 25000 を更新する。

【0256】

以上が、仮想ボリュームに対してプールからページ割り当てを行う処理である。

【0257】

<3390 - A 形式の LDEV 322 が仮想ボリューム上で構成されている場合のページ割り当て処理>

【0258】

次に、3390 - A 形式の LDEV 322 を仮想ボリューム上で構成する場合において、MF ホスト計算機 102 のライトアクセス契機に、ユーザデータ領域と、このユーザデータ領域に対応する制御情報領域のそれぞれにページを割り当てる処理を図 27 のフローチャートに従って説明する。

【0259】

コマンド制御部 251 が、MF ホスト I/F 121 を介して、MF ホスト計算機 102 からのライトアクセスを受領した際の処理は、以下の通りとなる。

【0260】

ライトアクセス時のパラメータは、例えば、3390 形式の LDEV 322 における LDEV 番号と LDEV 内のトラック番号である。

【0261】

まず、コマンド制御部 251 は、MF ホストライトコマンド処理を開始すると、図 5 の管理テーブル 500 を参照して、LDEV 322 のボリューム容量を取得し、取得した LDEV 322 のボリューム容量と予め設定されているトラックサイズを基に、LDEV 322 のトラック番号に対応する LDEV 322 内のトラックのデータ部 325 のアドレスと、同じくトラックの制御情報部 326 のアドレスを計算する。

【0262】

次に、コマンド制御部 251 は、管理テーブル 1700 を参照して、分散ユーザデータと分散制御情報のサイズを取得し、管理テーブル 1800 を参照して、分散ユーザデータ領域と分散制御情報領域の先頭アドレスを取得し、取得した分散ユーザデータと分散制御情報のサイズと、分散ユーザデータ領域と分散制御情報領域の先頭アドレスと、計算で得られたデータ部 325 のアドレス（データアドレス）および制御情報部 326 のアドレス（制御情報アドレス）と、を基に、3390 - A 形式の LDEV 上におけるトラックのデ

10

20

30

40

50

ータ部 3 2 5 のアドレス（データアドレス）と、制御情報部 3 2 6 のアドレス（制御情報アドレス）と、を計算する（S 5 1）。

【0 2 6 3】

この後、コマンド制御部 2 5 1 は、L D E V 番号を基にページ管理ディレクトリ 2 6 0 0 を参照して、L D E V 番号に対応するページ管理テーブル 2 5 0 0 0 のアドレスを取得する。

【0 2 6 4】

コマンド制御部 2 5 1 は、取得したアドレスに対応したページ管理テーブル 2 5 0 0 0 を参照して、計算で得られたアドレス（データアドレス）およびアドレス（制御情報アドレス）が指すデータを含むページの、ページ番号をそれぞれ取得する。

10

【0 2 6 5】

コマンド制御部 2 5 1 は、ページ管理テーブル 2 5 0 0 0 のうち、取得したページ番号に対応したページエントリ内の割り当て済み判定情報を参照し、アドレス（データアドレス）が含まれるページは、実記憶領域が割り当て済みか否かを判定し（S 5 2）、割り当て済み判定情報が割り当て済み状態を示す場合、即ち、実記憶領域が割り当て済みと判定した場合は、ステップ S 5 6 の処理に移行する。

【0 2 6 6】

一方、割り当て済み判定情報が未割り当て状態の場合、即ち、アドレス（データアドレス）が含まれるページは、実記憶領域が割り当て済みでないと判定した場合、コマンド制御部 2 5 1 は、管理テーブル 7 0 0 を参照して、L D E V 番号に対応したプール番号を取得し、取得したプール番号を基に管理テーブル 8 0 0 を参照して、プール番号に対応したディスク番号を取得し、取得したディスク番号を基に、L D E V 番号 3 2 2 に対応するプール 3 2 3 に属する一つまたは複数のディスク 1 3 2 から、ページ単位で実記憶領域の割り当てを行う（S 5 3）。即ち、アドレス（データアドレス）が含まれるページに対してページ割り当てを実行する。

20

【0 2 6 7】

この後、コマンド制御部 2 5 1 は、アドレス（制御情報アドレス）が含まれるページは、実記憶領域が割り当て済みか否かを判定し（S 5 4）、割り当て済み判定情報が割り当て済み状態を示す場合、即ち、実記憶領域が割り当て済みと判定した場合は、ステップ S 5 6 の処理に移行する。

30

【0 2 6 8】

ステップ S 5 4 で、アドレス（制御情報アドレス）が含まれるページは、実記憶領域が割り当て済みでないと判定した場合、コマンド制御部 2 5 1 は、管理テーブル 7 0 0 を参照して、L D E V 番号に対応したプール番号を取得し、取得したプール番号を基に管理テーブル 8 0 0 を参照して、プール番号に対応したディスク番号を取得し、取得したディスク番号を基に、L D E V 番号 3 2 2 に対応するプール 3 2 3 に属する一つまたは複数のディスク 1 3 2 から、ページ単位で実記憶領域の割り当てを行い（S 5 5）、ステップ S 5 6 の処理に移行する。即ち、ステップ S 5 5 では、アドレス（制御情報アドレス）が含まれるページに対してページ割り当てを実行する。

【0 2 6 9】

ステップ S 5 6 において、コマンド制御部 2 5 1 は、アドレス（制御情報アドレス）が示す領域に、トラックのデータ部 3 2 5 に対応する制御情報を生成する。

40

【0 2 7 0】

次に、アドレス（データアドレス）が含まれるページは、実記憶領域が割り当て済みである場合、コマンド制御部 2 5 1 は、アドレス（データアドレス）が示す領域に、データをライトする。即ち、コマンド制御部 2 5 1 は、ページ内のライトアクセス先トラック番号のデータ部 3 2 5 に対してライトアクセスを行う（S 5 7）。

【0 2 7 1】

この際、コマンド制御部 2 5 1 は、ページ管理テーブル 2 5 0 0 0 を参照して、ライトアクセス先のアドレスが示すトラックを含むページエントリを探し、新たに割り当てたペ

50

ージの先頭アドレスをページ管理テーブル25000に登録し、割り当て済み判定情報フィールド2502の「未割り当て」を「割り当て済み」として、ページ管理テーブル25000を更新し、このルーチンでの処理を終了する。

【0272】

コマンド制御部251は、前述したページ割り当て処理の中で、アドレス（制御情報アドレス）が属する制御情報領域を初期化後に、ライトアクセスを行ってもよい。

【0273】

本実施例によれば、仮想ボリュームに実記憶領域を割り当ててる場合、データ部325のページと制御情報部326のページをプールから割り当てることができる。

【実施例6】

【0274】

本実施例は、コピー元のストレージコントローラからコピー先のストレージコントローラに、リモートコピー処理に伴うデータとしてコピー対象データを送信し、このコピー対象データをコピー先ストレージコントローラが受信した場合、コピー先ストレージコントローラが、受信したコピー対象データの中にレコードの有無に関する情報を判定し、この判定結果に従った処理を実行するものである。

【0275】

<3390-A形式のLDEV322が仮想ボリューム上で構成されている場合のリモートコピー処理に同期する0データページ破棄処理>

【0276】

図28は、第二のストレージ装置において、仮想ボリューム上に保存された3390-A形式のLDEV322に対して、第一のストレージ装置101からデータを送信するリモートコピー処理に同期して、受信側のLDEV322内のページ状態を未割り当てにする処理を示すフローチャートである。

【0277】

以下、ページ状態を未割り当てに戻すことを、ページ破棄と呼ぶ。

【0278】

第一のストレージ装置101と第二のストレージ装置101において、リモートコピー処理を以下のように行う。

【0279】

第一のストレージ装置101内の管理ホストI/F127を介して、管理ホスト103から、第一のストレージ装置101内の第一のコマンド制御部251が、リモートコピーの実施要求を受け取ると、第一のコマンド制御部251は、第一のストレージ装置101内の第一のLDEV322（3390形式、または、3390-A形式）の一部または全部のトラックのデータ部325と、一部または全部のトラックの制御情報部326と、をコピー対象データとして、一つまたは複数のディスク132から、第一のキャッシュメモリ124に読み込む。

【0280】

第一のコマンド制御部251は、第一のキャッシュメモリ上のLDEV322の一部または全部のトラックを、第一のストレージ装置101内の第一のホストI/F121と、ネットワーク111と、を介して、一定のデータ単位で送信する。データ単位は、例えば、一つまたは複数のトラックのデータ部325であったり、一つまたは複数のトラックの制御情報部326であったりしてもよい。

【0281】

一方、第二のストレージ装置101内の第二のホストI/F121を介して、第二のストレージ装置101内の第二のコマンド制御部251が、第一のコマンド制御部251からコピー対象データを受信すると、第二のコマンド制御部251は、コピーデータ受信処理を開始し、第二のストレージ装置101内の第二のキャッシュメモリ124に、受信したコピー対象データを書き込む（S61）。

【0282】

10

20

30

40

50

この際、第二のストレージ装置101内の第二のLDEV322は、3390-A形式に従い、前述したシン・プロビジョニング機能で用いられる仮想ボリューム上で構成される。

【0283】

第二のコマンド制御部251は、第二のキャッシュメモリ124に保存されているコピー対象データの中の各データ部325に対して、トラック内に、前述したレコードが存在するか否かを判定し(S62)、キャッシュメモリ124上のトラックのデータ部325はレコード無し、と判定した場合は、ステップS66の処理に移行し、キャッシュメモリ124上のトラックのデータ部325はレコード有り、と判定した場合は、ステップS63の処理に移行する。このとき、第二のコマンド制御部251は、トラック毎にレコード有無を調べた結果を、第二のLDEV322の各トラックのデータ部325に対応するレコード有無情報部327に保存する。

10

【0284】

ステップS63において、第二のコマンド制御部251は、データ部325または制御情報部26を含むページは、実記憶領域が割り当て済みか否かを判定する。

【0285】

具体的には、第二のコマンド制御部251は、LDEV322のLDEV番号を基にページ管理テーブル25000とページ管理ディレクトリ2600を参照し、第二のキャッシュメモリ124上にある書き込み先仮想領域であって、コピー対象データに属するデータ部325と制御情報部326の書き込み先仮想領域に、プール323からライト先ページが割り当てられているかどうかをチェックする。

20

【0286】

データ部325または制御情報部26を含むページは、実記憶領域が割り当て済みでないと判定した場合、第二のコマンド制御部251は、データ部325または制御情報部26を含むページに対して、プール323内の実記憶領域を割り当てる(S64)。

【0287】

具体的には、第二のキャッシュメモリ124上の、一つまたは複数のトラックのデータ部325のライト先ページが、未割り当て状態であり、かつ、これらのトラックのデータ部325にレコードがある場合、第二のコマンド制御部251は、LDEV322のLDEV番号を基に管理テーブル700を参照して、LDEV番号に対応したプール番号を取得し、取得したプール番号を基に管理テーブル800を参照して、プール番号に対応したディスク番号を取得し、取得したディスク番号を基に、LDEV322のLDEV番号に対応するプール323に属する一つまたは複数のディスクから、ページ単位で実記憶領域の割り当てを行う。

30

【0288】

同様に、第二のキャッシュメモリ124上の、前述した一つまたは複数のトラックの制御情報部326のライト先ページが、未割り当て状態である場合、第二のコマンド制御部251は、LDEV322のLDEV番号を基に管理テーブル700を参照して、LDEV番号に対応したプール番号を取得し、取得したプール番号を基に管理テーブル800を参照して、プール番号に対応したディスク番号を取得し、取得したディスク番号を基に、LDEV322のLDEV番号に対応するプールに属する一つまたは複数のディスクから、ページ単位で実記憶領域の割り当てを行う。

40

【0289】

次に、第二のコマンド制御部251は、第二のキャッシュメモリ124上のデータ部325と制御情報部326を、割り当て済み実記憶領域に対して書き込む(S65)。

【0290】

具体的には、第二のキャッシュメモリ124上の書き込み先仮想領域に、コピー対象データに属するデータ部325のライト先ページが、割り当て済み状態である場合、第二のコマンド制御部251は、第二のキャッシュメモリ124上の、コピー対象データに属するデータ部325を、プール323内の実記憶領域を構成する第二の一つまたは複数のデ

50

ィスク 1 3 2 に書き込む。

【0 2 9 1】

同様に、第二のキャッシュメモリ 1 2 4 上の書き込み先仮想領域に、コピー対象データに属する制御情報部 3 2 6 のライト先ページが、割り当て済み状態である場合、第二のコマンド制御部 2 5 1 は、第二のキャッシュメモリ 1 2 4 上の、コピー対象データに属する制御情報部 3 2 6 を、プール 3 2 3 内の実記憶領域を構成する第二の一つまたは複数のディスク 1 3 2 に書き込む。

【0 2 9 2】

この後、第二のコマンド制御部 2 5 1 は、ページ管理テーブル 2 5 0 0 0 を基に、ライトアクセス先のトラックのデータ部 3 2 5 を含むページエントリを探し、新たに割り当てたページの先頭アドレスを登録し、割り当て済み判定情報を割り当て済み状態に更新する。同様に、第二のコマンド制御部 2 5 1 は、ページ管理テーブル 2 5 0 0 を基に、ライトアクセス先のトラックの制御情報部 3 2 6 を含むページエントリを探し、新たに割り当てたページの先頭アドレスを登録し、割り当て済み判定情報を割り当て済み状態に更新する。

10

【0 2 9 3】

次に、第二のコマンド制御部 2 5 1 は、前述した一つまたは複数のトラックのデータ部 3 2 5 をプール 3 2 3 内の実記憶領域を構成する第二の一つまたは複数のディスク 1 3 2 に書き込み、前述した一つまたは複数のトラックの制御情報部 3 2 6 をプール 3 2 3 内の実記憶領域を構成する第二の一つまたは複数のディスク 1 3 2 に書き込む。

20

【0 2 9 4】

この後、ステップ 6 6 において、第二のコマンド制御部 2 5 1 は、第二の L D E V 3 2 2 の前述した一つまたは複数のトラックのデータ部 3 2 5 に対応するレコード有無情報部 3 2 7 に、レコードが有ることを示す情報を保存して、レコード有無情報部 3 2 7 を更新する。

【0 2 9 5】

第二のキャッシュメモリ 1 2 4 上の、前述した一つまたは複数のトラックのデータ部のライト先ページが未割り当て状態であり、かつ、これらのトラックのデータ部にレコードが無い場合、第二のコマンド制御部 2 5 1 は、第二のキャッシュメモリ 1 2 4 の書き込み先仮想領域に、プール 3 2 3 からページを割り当てない。この場合、第二のコマンド制御部 2 5 1 は、前述した一つまたは複数のトラックのデータ部 3 2 5 を、第二の一つまたは複数のディスク 1 3 2 に書き込まない。さらに、第二のコマンド制御部 2 5 1 は、第二の L D E V 3 2 2 のトラックのデータ部 3 2 5 に対応するレコード有無情報部 3 2 7 には、レコード無し状態を示す情報を保存する。

30

【0 2 9 6】

第二のキャッシュメモリ 1 2 4 上の、前述した一つまたは複数のトラックのデータ部のライト先ページが未割り当て状態であり、かつ、これらのトラックのデータ部にレコードが無い場合、第二のコマンド制御部 2 5 1 は、プール 3 2 3 からページを割り当てた後に、レコードが無いことを示す初期値を書き込んでもよい。

【0 2 9 7】

また、ステップ S 6 6 では、第二のコマンド制御部 2 5 1 は、第二の L D E V 3 2 2 の割り当て済みページ内のトラックのデータ部 3 2 5 のレコード有無をチェックし、レコードが存在しないトラックのレコード有無情報部 3 2 7 に、レコードが無いことを示す情報を保存して、レコード有無情報部 3 2 7 を更新する。

40

【0 2 9 8】

次に、第二のコマンド制御部 2 5 1 は、第二の L D E V 3 2 2 の全てのトラックのレコード有無情報部 3 2 7 の値に基づき、第二の L D E V 3 2 2 の割り当て済みページ内の、トラックのデータ部 3 2 5 のレコード有無をチェックし、レコードが存在しないページを未割り当て状態、即ち、ページ破棄を実行し（S 6 7）、このルーチンでの処理を終了する。

50

【0299】

本実施例においては、第二のストレージ装置における第二のストレージコントローラ131は、第一のストレージ装置101における第一のストレージコントローラ131を他のコントローラとして、第一のストレージコントローラ131とネットワーク111を介して情報の授受を行い、第一のストレージコントローラ131から受信した情報の中に、第1データ部325または第2データ部325に関する情報として、トラックに対応したレコードに関する情報が存在するか否かを判定し、この判定の結果を、仮想ボリューム(3390-A形式のLDEV322)の第2制御情報領域に格納し、仮想ボリュームの第2制御情報領域に格納された判定結果が、トラックに対応したレコードが存在しないことを示す場合、トラックに対応した実記憶領域として、プールから仮想ボリュームに割り当てられているページを開放する処理を、リモートコピー処理に同期して実行することになる。

10

【0300】

前述したページ破棄処理は、第二のキャッシュメモリ124上の、一つまたは複数のトラックに対応したコピー対象データを、一つまたは複数のディスク132に書き込む前に行ってよい。

【0301】

以上が、第一のストレージ装置101内の第一のLDEV322から、第二のストレージ装置101内の第二のLDEV322に対して、リモートデータコピーを実施する処理の一例である。

20

【0302】

本実施例によれば、第二のストレージ装置101内の第二のストレージコントローラ131は、第一のストレージ装置101内の第一のストレージコントローラ131からコピー対象データを受信した場合、トラックのデータ部325のレコード有無をチェックし、レコードが存在しないページを未割り当て状態、即ち、ページ破棄の処理を実行することができる。

【0303】

本実施例では、第一のコマンド制御部251が、送信側の第一のLDEV322の一つまたは複数のトラックのデータ部325と、一つまたは複数のトラックの制御情報部326と、を含むコピー対象データを、一つまたは複数のディスク132から、第一のキャッシュメモリ124に読み込む処理において、第一のコマンド制御部251は、第一のキャッシュメモリ124上の、一つまたは複数のトラックのデータ部325に対応するレコード有無情報部327を参照し、コピー対象データに属するデータ部325にレコードが無い場合には、これらのトラックのデータ部325を第二のストレージ装置に送信しなくてもよい。また、これらのトラックのデータ部325にレコードが無いことを、第二のストレージ装置にメッセージで通知してもよい。

30

【0304】

3390形式のLDEV322により、MFホスト計算機102は、トラックの制御情報部326の値を確認しない処理を行う場合には、トラックの制御情報部326をスキップして、ディスクのヘッドをトラックのデータ部325に位置付ける処理のオーバーヘッドを減らすことができる。

40

【0305】

以上、本発明の実施例を説明したが、本発明は、各実施例に限定されるものでなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。例えば、シリンダグループの削除などで、ボリュームの容量を縮小する場合、ボリュームの容量を拡張するときのアドレス変換処理を、ボリュームの容量を縮小するときのアドレス変換処理に適用することもできる。

【符号の説明】

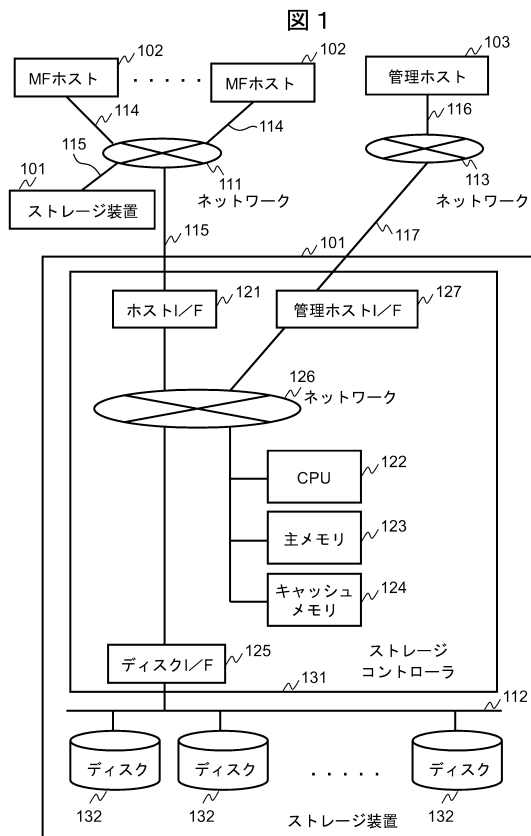
【0306】

101 ストレージ装置、102 MFホスト計算機、103 管理ホスト計算機、1

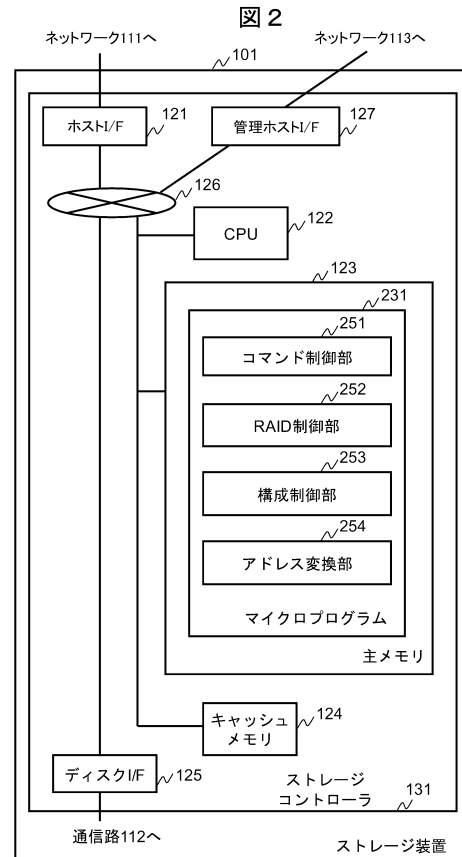
50

11、113、126 ネットワーク、122 CPU、123 主メモリ、231 マイクロプログラム、124 キャッシュメモリ、131 ストレージコントローラ、132 ディスク、322 LDEV、323 プール、325 データ部、326 制御情報部、327 レコード有無情報部、901 記憶領域 920、930 実記憶領域、1401～1405 分散ユーザデータ領域、1411～1415 分散制御情報領域、1421～1425 シリンダグループ (CG)、2202、2203 LDEV。

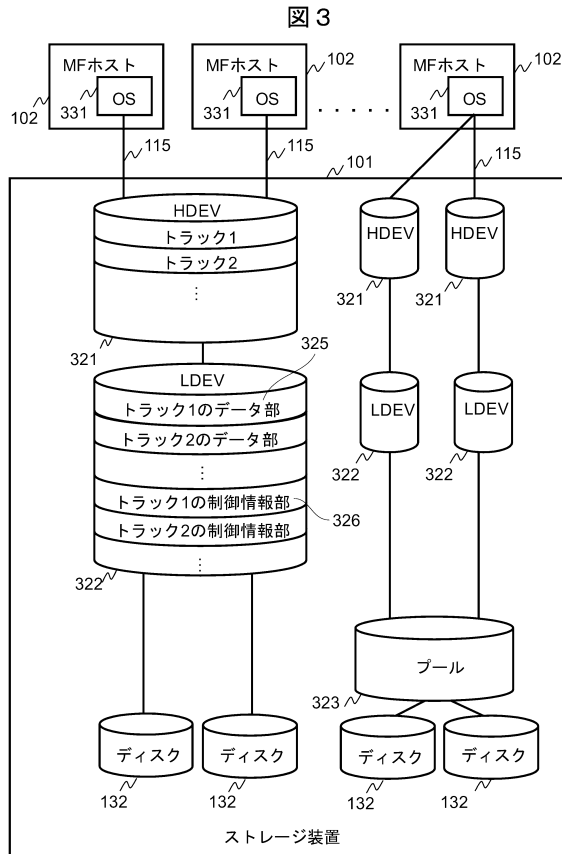
【図 1】



【図 2】



【図 3】



【図 4】

図 4

HDEV番号	LDEV番号
1	1
2	2
3	3
...	...

【図 5】

図 5

LDEV番号	ボリューム属性	ボリューム容量
1	通常ボリューム	XGB
2	仮想ボリューム	YGB
3	仮想ボリューム	ZGB
...	...	...

【図 6】

図 6

LDEV番号	ディスク番号
1	1
2	2
...	...

【図 7】

図 7

LDEV番号	プール番号
2	1
3	2
...	...

【図 8】

図 8

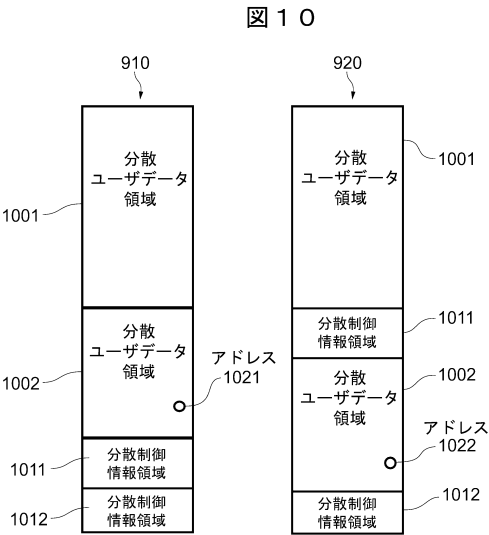
プール番号	ディスク番号
1	1
2	2
3	3
4	4
...	...

【図 9】

図 9

901	
902 ユーザデータ領域	325 トラック1のデータ部
	トラック2のデータ部
	...
903 制御情報領域	326 トラック1の制御情報
	トラック2の制御情報
	...

【図 1 0】



【図 1 2】

図 1 2

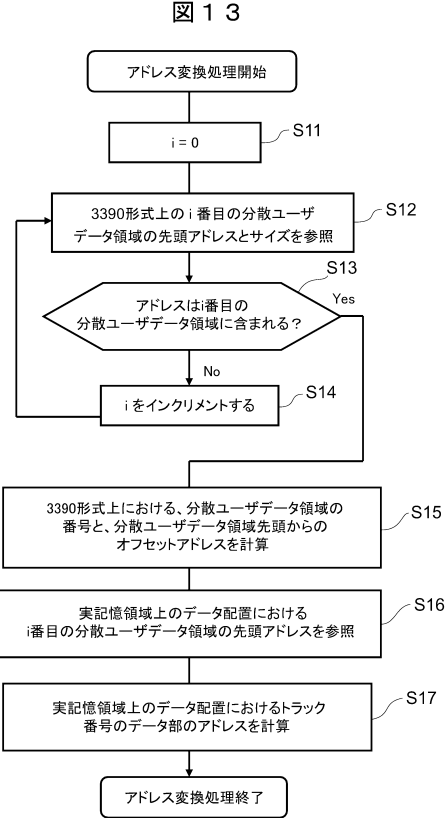
	1201	1202	1203	1204	1205
i	分散ユーザデータ領域 i の先頭アドレス (3390 形式)	分散ユーザデータ領域 i の先頭アドレス (実記憶領域上のデータ配置 (1))	分散制御情報領域 i の先頭アドレス (3390 形式)	分散制御情報領域 i の先頭アドレス (実記憶領域上のデータ配置 (1))	
1	0	0	250	150	
2	150	165	265	265	
...	...	...	...	...	

【図 1 1】

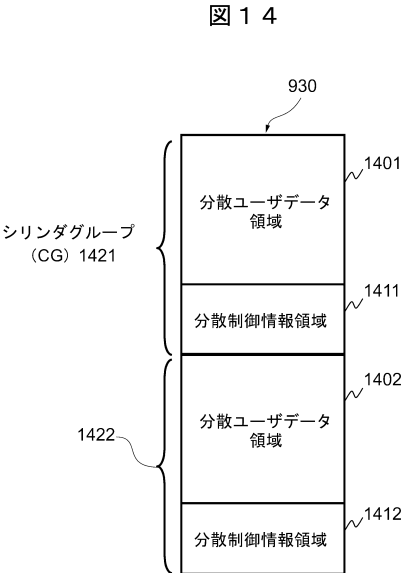
図 1 1

i	1101	1102	1103
i	分散ユーザデータ領域 i のサイズ	分散制御情報領域 i のサイズ	
1	150	15	
2	100	10	
...	...	...	

【図 1 3】

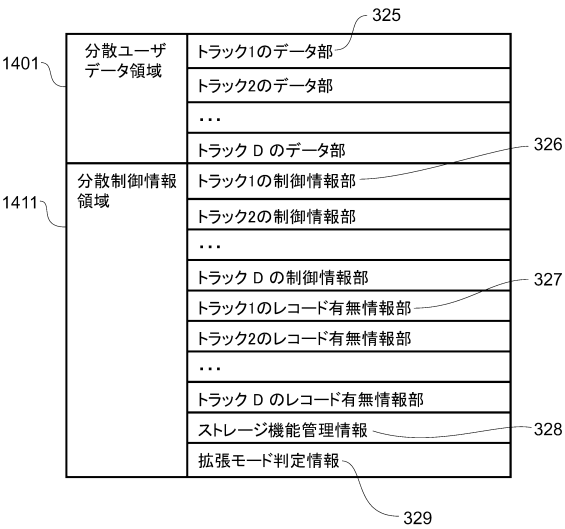


【図 1 4】



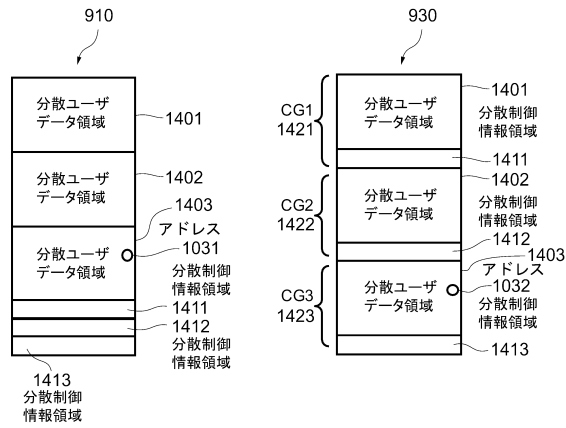
【図 1 5】

図 1 5



【図 1 6】

図 1 6



【図 1 7】

図 1 7

分散ユーザデータ領域長	分散制御情報領域長
100	50

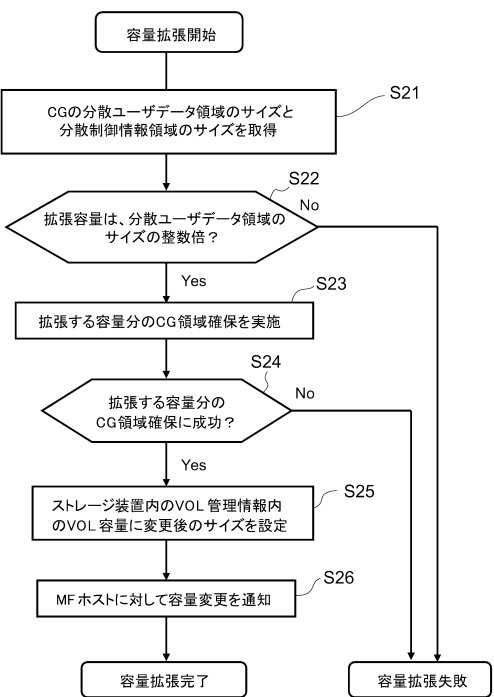
【図 1 8】

図 1 8

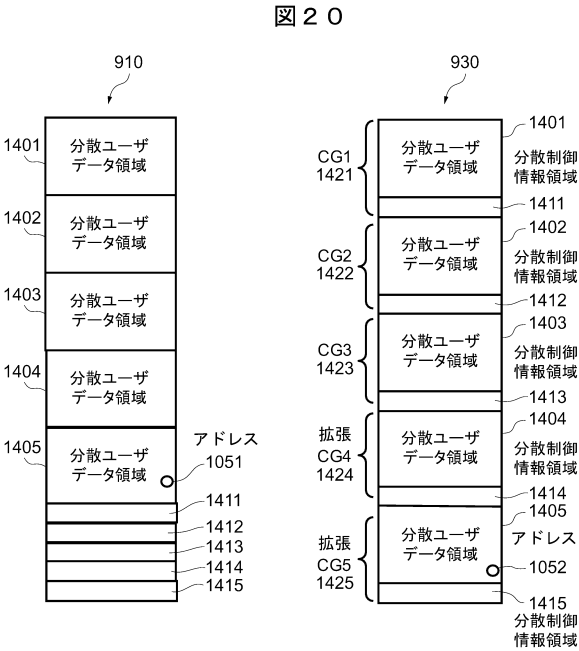
i	分散ユーザデータ領域 i の先頭アドレス (3390形式)	分散ユーザデータ領域 i の先頭アドレス (3390-A形式)	分散制御情報領域 i の先頭アドレス (3390形式)	分散制御情報領域 i の先頭アドレス (3390-A形式)
1	0	0	500	100
2	100	150	550	250
3	200	300	600	400
4	300	450	650	550
5	400	600	700	700
...	...	...	...	...

【図 1 9】

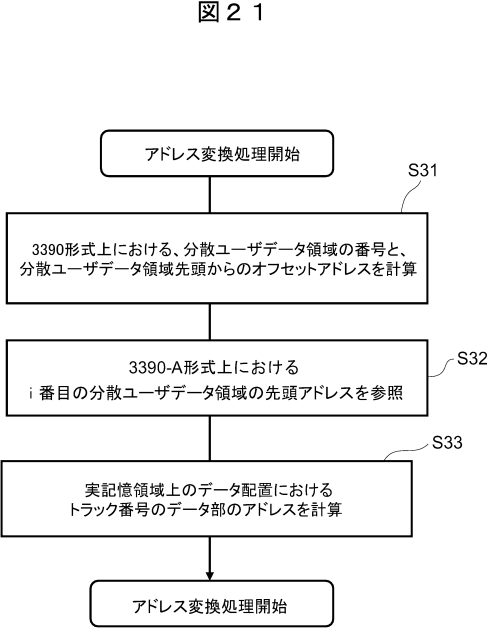
図 1 9



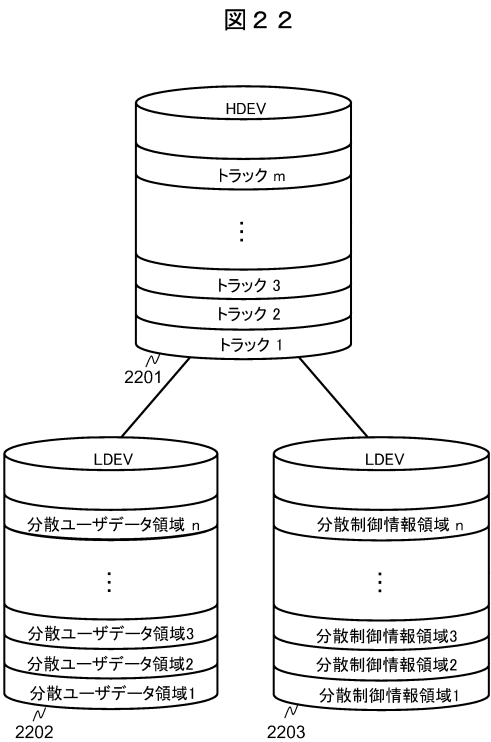
【図 2 0】



【図 2 1】



【図 2 2】

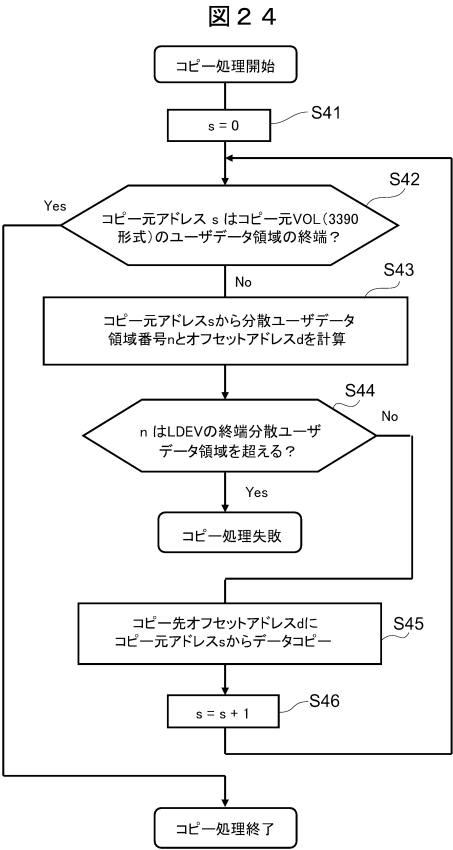


【図 2 3】

図 2 3

分散ユーザデータ領域番号	LDEV2202内のアドレス	分散制御情報領域番号	LDEV2203内のアドレス
1	0	1	0
2	100	2	50
3	200	3	100
...	...	...	...

【図 2 4】



【図 2 5】

図 2 5

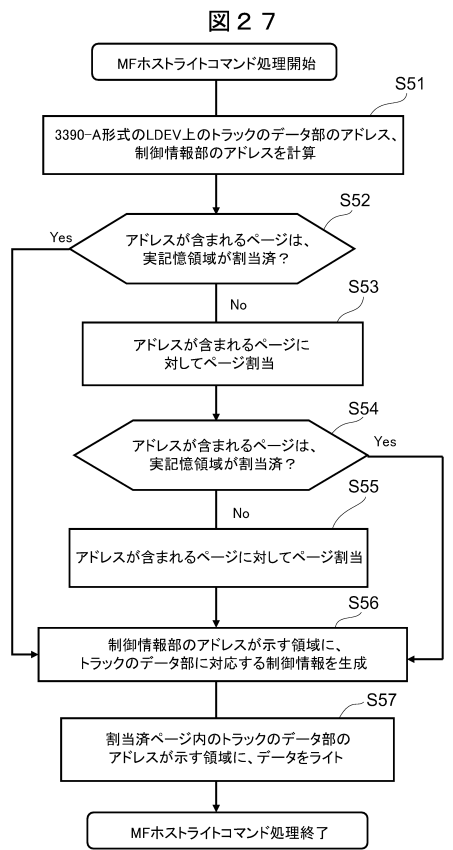
2501 ページ番号	2502 LDEV内のページ先頭アドレス	2503 ディスク番号	2500 ディスク内のページ先頭アドレス	2504 割り当て済み判定情報
1	0	1	128	割り当て済み
2	128	2	128	未割り当て
3	256	3	128	割り当て済み
4	384	1	256	割り当て済み
5	512	2	256	未割り当て
...	...	...	...	...

【図 2 6】

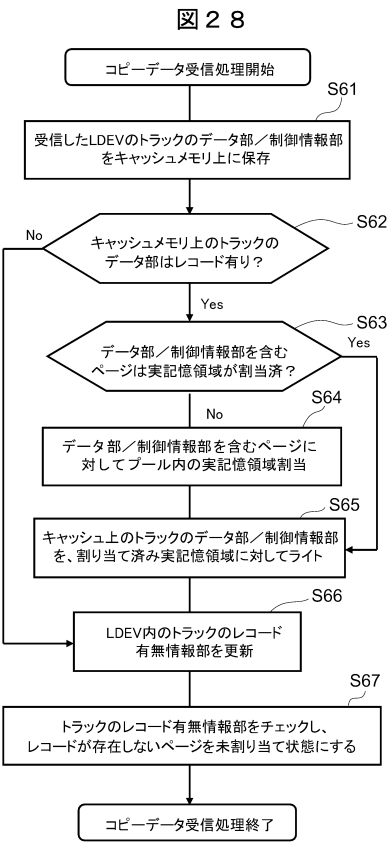
図 2 6

2601 LDEV番号	2600 ページ管理テーブルのアドレス	2602
2	0xC0000000	
3	0xC0000100	
...	...	

【図 2 7】



【図 2 8】



フロントページの続き

(72)発明者 竹内 久治

神奈川県小田原市中里322番地2号 株式会社日立製作所 ITプラットフォーム事業本部内

審査官 田上 隆一

(56)参考文献 特開2008-084053 (JP, A)

特開2010-020431 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G06F 3/06

G06F 13/10