

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5199464号

(P5199464)

(45) 発行日 平成25年5月15日(2013.5.15)

(24) 登録日 平成25年2月15日(2013.2.15)

(51) Int.Cl.

F I

G 0 6 F 3/06 (2006.01)

G 0 6 F 3/06 3 0 4 F

請求項の数 8 (全 52 頁)

(21) 出願番号	特願2011-514950 (P2011-514950)	(73) 特許権者	000005108
(86) (22) 出願日	平成21年1月20日 (2009.1.20)		株式会社日立製作所
(65) 公表番号	特表2012-504793 (P2012-504793A)		東京都千代田区丸の内一丁目6番6号
(43) 公表日	平成24年2月23日 (2012.2.23)	(74) 代理人	110000279
(86) 国際出願番号	PCT/JP2009/000182		特許業務法人ウィルフォート国際特許事務所
(87) 国際公開番号	W02010/084522	(72) 発明者	佐藤 貴仁
(87) 国際公開日	平成22年7月29日 (2010.7.29)		神奈川県小田原市中里322番2号 株式会社日立製作所 RAIDシステム事業部内
審査請求日	平成23年4月6日 (2011.4.6)		
		審査官	横山 佳弘

最終頁に続く

(54) 【発明の名称】 ストレージシステム及びストレージシステムの制御方法

(57) 【特許請求の範囲】

【請求項1】

ホストコンピュータ（5）と、そのホストコンピュータにより利用される複数の記憶制御装置（1-3）と、前記各記憶制御装置を管理するための管理装置（4）とが通信可能に接続されているストレージシステムであって、

前記複数の記憶制御装置には、第1記憶制御装置（1）と第2記憶制御装置（2）及び第3記憶制御装置（3）が含まれており、

前記第1記憶制御装置の有する第1ボリューム（1A）と前記第2記憶制御装置の有する第2ボリューム（2A）とをペアとして設定することにより、一つの仮想ボリューム（6）を生成して前記ホストコンピュータに提供させる仮想ボリューム設定部（4A）と、

前記第3記憶制御装置の有する第3ボリューム（3A）を、前記仮想ボリュームの使用を制御するための使用制御情報（L10）を記憶させる制御用ボリュームとして設定する、制御用ボリューム設定部（4B）と、

前記第1記憶制御装置内に設けられる仮想的な第4ボリュームを前記第3ボリュームに対応付け、かつ、前記第2記憶制御装置内に設けられる仮想的な第5ボリュームを前記第3ボリュームに対応付ける、対応付け設定部（4C）と、

を備えており、

前記第1記憶制御装置は前記第4ボリュームを介して前記第3ボリュームを使用し、前記第2記憶制御装置は前記第5ボリュームを介して前記第3ボリュームを使用するようになり、

10

20

前記第 3 ボリュームは、前記第 1 記憶制御装置及び前記第 2 記憶制御装置のみが使用でき、前記使用制御情報に含まれている識別情報以外の他の識別情報を有する他の記憶制御装置は、前記第 3 ボリュームを使用できないようになっており、

前記第 3 ボリュームに記憶される前記使用制御情報には、前記第 3 ボリュームを識別するための識別子（L 1 1 1）と、前記第 1 記憶制御装置及び前記第 2 記憶制御装置をそれぞれ特定するための識別情報（L 1 1 2，L 1 1 3）と、前記第 1 記憶制御装置が前記第 3 ボリュームを使用しているかを示す第 1 制御情報（L 1 2）および前記第 2 記憶制御装置が前記第 3 ボリュームを使用しているかを示す第 2 制御情報（L 1 3）とが含まれており、

前記制御用ボリューム設定部は、前記第 1 記憶制御装置または前記第 2 記憶制御装置のいずれかを介して前記第 3 ボリュームを前記制御用ボリュームとして設定することができ、前記第 3 ボリュームを識別する第 3 ボリューム識別子を示して前記第 3 ボリュームの作成が指示されると、前記第 3 記憶制御装置から前記第 3 ボリュームに記憶されている使用制御情報を読み出し、前記指示された第 3 ボリューム識別子が前記読み出した使用制御情報に使用されていない場合は前記指示された第 3 ボリューム識別子を用い、前記指示された第 3 ボリューム識別子が使用されている場合は前記読み出した使用制御情報内の第 3 ボリューム識別子を用い、前記第 1 制御情報または前記第 2 制御情報のうち作成される第 3 ボリュームを使用する記憶制御装置に対応する制御情報を更新することで、前記使用制御情報を更新するためのライトデータを作成して前記第 3 ボリュームに書き込む、
ストレージシステム。

【請求項 2】

前記仮想ボリューム設定部と前記制御用ボリューム設定部とは、前記管理装置に設けられている、請求項 1 に記載のストレージシステム。

【請求項 3】

前記仮想ボリューム設定部と前記制御用ボリューム設定部及び前記対応付け設定部とは、前記管理装置に設けられている、請求項 1 に記載のストレージシステム。

【請求項 4】

前記使用制御情報は、前記第 1 記憶制御装置のみが更新可能な領域と前記第 2 記憶制御装置のみが更新可能な領域とが含まれている、請求項 1 に記載のストレージシステム。

【請求項 5】

前記第 1 記憶制御装置は、前記使用制御情報に対応する第 1 管理テーブルを有しており、前記第 2 記憶制御装置は、前記使用制御情報に対応する第 2 管理テーブルを有しており、前記第 1 管理テーブル及び前記第 2 管理テーブルは、前記使用制御情報の更新に対応して更新される、請求項 1 に記載のストレージシステム。

【請求項 6】

前記仮想ボリューム設定部は、前記第 1 ボリュームと前記第 2 ボリュームとの間に差分が発生した場合に、所定の契機に基づいて、前記差分が解消するように前記第 1 ボリュームの記憶内容と前記第 2 ボリュームの記憶内容とを再同期させる、請求項 1 に記載のストレージシステム。

【請求項 7】

前記仮想ボリュームに関する前記ペアを削除する場合、前記仮想ボリューム設定部が前記ペアを削除した後で、前記制御用ボリューム設定部は、前記仮想ボリュームに関する前記使用制御情報を削除させる、請求項 1 に記載のストレージシステム。

【請求項 8】

ホストコンピュータと、そのホストコンピュータにより利用される複数の記憶制御装置と、前記各記憶制御装置を管理するための管理装置とが通信可能に接続されているストレージシステムを制御するための制御方法であって、

前記複数の記憶制御装置には、第 1 記憶制御装置と第 2 記憶制御装置及び第 3 記憶制御装置が含まれており、

前記第 1 記憶制御装置の有する第 1 ボリュームと前記第 2 記憶制御装置の有する第 2 ボ

10

20

30

40

50

リウムとをペアとして設定することにより、一つの仮想ボリュームを生成して前記ホストコンピュータに提供させるステップと、

前記第3記憶制御装置の有する第3ボリュームを、前記仮想ボリュームの使用を制御するための使用制御情報を記憶させる制御用ボリュームとして設定するステップであって、前記使用制御情報を前記第3ボリュームに書き込む場合、前記使用制御情報を更新するためのライトデータを作成して前記第3ボリュームに書き込み、その後に前記第3ボリュームから前記ライトデータを読み出し、前記第3ボリュームに書き込んだライトデータと前記第3ボリュームから読み出したライトデータとが一致するときは、前記使用制御情報の前記第3ボリュームへの書き込み処理が正常に完了したことを確認するステップと、

前記第1記憶制御装置内に設けられる仮想的な第4ボリュームを前記第3ボリュームに対応付け、かつ、前記第2記憶制御装置内に設けられる仮想的な第5ボリュームを前記第3ボリュームに対応付けるステップと、
を備えており、

前記第1記憶制御装置は前記第4ボリュームを介して前記第3ボリュームを使用し、前記第2記憶制御装置は前記第5ボリュームを介して前記第3ボリュームを使用するように
なっており、

前記第3ボリュームは、前記第1記憶制御装置及び前記第2記憶制御装置のみが使用でき、前記使用制御情報に含まれている識別情報以外の他の識別情報を有する他の記憶制御装置は、前記第3ボリュームを使用できないようになっており、

前記第3ボリュームに記憶される前記使用制御情報には、前記第1記憶制御装置及び前記第2記憶制御装置をそれぞれ特定するための識別情報と、前記第1記憶制御装置が前記第3ボリュームを使用しているかを示す制御情報および前記第2記憶制御装置が前記第3ボリュームを使用しているかを示す制御情報とが含まれており、

前記各ステップは、前記管理装置から前記第1記憶制御装置及び前記第2記憶制御装置に送られる指示に基づいて実行される、ストレージシステムの制御方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ストレージシステム及びストレージシステムの制御方法に関する。

【背景技術】

【0002】

例えば、企業等では、多種多量のデータを取り扱うために、比較的大規模なストレージシステムを用いてデータを管理する。このストレージシステムは、少なくとも一つの記憶制御装置を備える。記憶制御装置は、例えば、多数の記憶装置を備えており、RAID (Redundant Array of Inexpensive Disks) に基づく記憶領域を提供する。記憶装置群が提供する物理的な記憶領域上には、少なくとも1つ以上の論理デバイス（論理ボリュームとも呼ばれる）が形成される。ホストコンピュータ（以下、「ホスト」）は、論理デバイスに対してライトコマンドやリードコマンドを発行することにより、データの書き込みやデータの読出しを行う。

【0003】

ストレージシステムは、データの安全性等を改善するために、複数の論理デバイスに同一のデータをそれぞれ記憶させることができる。例えば、第1の従来技術として、ストレージシステムは、同一の記憶制御装置内のそれぞれ異なる論理デバイスに同一のデータをそれぞれ記憶させることができ、または、別々の記憶制御装置内の論理デバイスに同一のデータをそれぞれ記憶させることができる（JP-A-2007-150409）。

【0004】

さらに、第2の従来技術として、一方の記憶制御装置内の正ボリュームと他方の記憶制御装置内の副ボリュームとでリモートコピーのペアを形成し、そのリモートコピーペアを形成する2つの論理ボリュームを同一のデバイスとしてホストに認識させる、技術も知られている（JP-A-2008-134988）。

【発明の開示】

【発明が解決しようとする課題】

【0005】

前記第1の従来技術では、同一筐体内の複数の論理デバイスにデータをそれぞれ記憶させたり、あるいは、それぞれ別々の筐体内に位置する複数の論理デバイスにデータを記憶させることにより、プライマリの論理デバイスが使用不能な場合でも、セカンダリの論理デバイスを用いて業務処理を続行可能である。しかし、プライマリの論理デバイスからセカンダリの論理デバイスに切り替えるためには、ホストのアクセス先デバイスをプライマリの論理デバイスからセカンダリの論理デバイスに意図的に切り替える必要があり、切替作業に手間がかかる。

10

【0006】

前記第2の従来技術では、リモートコピーペアを形成する正ボリューム及び副ボリュームを同一の論理ボリュームとしてホストに認識させることにより、データを二重化して管理することができ、さらに、障害発生時に、ホストは副ボリュームに切り替えて情報処理を継続することができる。しかし、第2の従来技術では、各記憶制御装置の障害の有無をホスト側で管理する必要がある。

【0007】

本発明は、上記の問題点に鑑みてなされたもので、その目的は、異なる記憶制御装置内に存在する別々の論理ボリュームデバイスを一つの仮想ボリュームとして仮想化することができ、さらに、その仮想ボリューム設定、及び、その仮想ボリュームの使用を制御するための情報を別の論理ボリュームに格納して、データアクセスの整合性を確保できるようにしたストレージシステム、及びストレージシステムの制御方法を提供することにある。本発明の更なる目的は、後述する実施の形態の記載から明らかになるであろう。

20

【課題を解決するための手段】

【0008】

上記課題を解決すべく、本発明の第1観点に従うストレージシステムは、ホストコンピュータと、そのホストコンピュータにより利用される複数の記憶制御装置と、各記憶制御装置を管理するための管理装置とが通信可能に接続されているストレージシステムであって、複数の記憶制御装置には、第1記憶制御装置と第2記憶制御装置及び第3記憶制御装置が含まれており、第1記憶制御装置の有する第1ボリュームと第2記憶制御装置の有する第2ボリュームとをペアとして設定することにより、一つの仮想ボリュームを生成してホストコンピュータに提供させる仮想ボリューム設定部と、第3記憶制御装置の有する第3ボリュームを、仮想ボリュームの使用を制御するための使用制御情報を記憶させる制御用ボリュームとして設定する、制御用ボリューム設定部と、を備えており、第3ボリュームに記憶される使用制御情報には、第1記憶制御装置及び第2記憶制御装置をそれぞれ特定するための識別情報が含まれている。

30

【0009】

第2観点では、ホストコンピュータと第1記憶制御装置及び第2記憶制御装置とは、第1通信経路を介して接続されており、第1記憶制御装置と第2記憶制御装置とは第2通信経路を介して接続されており、第3記憶制御装置は、第3通信経路を介して第1記憶制御装置及び第2記憶制御装置に接続されており、管理装置は、ホストコンピュータと第1記憶制御装置と第2記憶制御装置及び第3記憶制御装置とに第4通信経路を介して接続されており、第1記憶制御装置は、第1管理部と、第1ボリューム及び仮想的に形成される第4ボリュームを備えており、第2記憶制御装置は、第2管理部と、第2ボリューム及び仮想的に形成される第5ボリュームを備えており、管理装置は、(1)第1管理部及び第2管理部に所定の指示を与えることにより、仮想ボリュームを生成してホストコンピュータに提供させる、仮想ボリューム設定部と、(2)第1管理部及び第2管理部に他の所定の指示を与えることにより、第3ボリュームを制御用ボリュームとして設定させる、制御用ボリューム設定部と、(3)第1管理部及び第2管理部にさらに別の指示を与えることにより、第4ボリューム及び第5ボリュームを第3ボリュームにそれぞれ対応付ける、対応

40

50

付け設定部と、を備えており、使用制御情報は、第3ボリュームを特定するための第3ボリューム識別情報と、第1記憶制御装置を特定するための第1識別情報と、第2記憶制御装置を特定するための第2識別情報と、第1記憶制御装置が第3ボリュームを使用しているか否かを示す第1使用情報と、第2記憶制御装置が第3ボリュームを使用しているか否かを示す第2使用情報と、ペアの解消後に第1ボリュームに差分データが生じていることを示す第1差分発生情報と、ペアの解消後に第2ボリュームに差分データが生じていることを示す第2差分発生情報と、を含み、第1識別情報と、第1使用情報と、第1差分発生情報とは、第1記憶制御装置のみが更新可能であり、第2識別情報と、第2使用情報と、第2差分発生情報とは、第2記憶制御装置のみが更新可能であり、第3ボリュームは、使用制御情報に対応付けられている第1記憶制御装置及び第2記憶制御装置のみが使用でき、使用制御情報に含まれている識別情報以外の他の識別情報を有する他の記憶制御装置は、第3ボリュームを使用できないようになっている。

10

【0010】

第3観点では、第1観点において、第1記憶制御装置内に設けられる仮想的な第4ボリュームを第3ボリュームに対応付け、かつ、第2記憶制御装置内に設けられる仮想的な第5ボリュームを第3ボリュームに対応付ける、対応付け設定部をさらに備え、第1記憶制御装置は第4ボリュームを介して第3ボリュームを使用し、第2記憶制御装置は第5ボリュームを介して第3ボリュームを使用する。

【0011】

第4観点では、第3観点において、第3ボリュームは、第1記憶制御装置及び第2記憶制御装置のみが使用でき、使用制御情報に含まれている識別情報以外の他の識別情報を有する他の記憶制御装置は、第3ボリュームを使用できないようになっている。

20

【0012】

第5観点では、第1観点において、仮想ボリューム設定部と制御用ボリューム設定部とは、管理装置に設けられている。

【0013】

第6観点では、第3観点において、仮想ボリューム設定部と制御用ボリューム設定部及び対応付け設定部とは、管理装置に設けられている。

【0014】

第7観点では、第1観点において、使用制御情報は、第1記憶制御装置のみが更新可能な領域と第2記憶制御装置のみが更新可能な領域とが含まれる。

30

【0015】

第8観点では、第1観点において、使用制御情報は、第3ボリュームを特定するための第3ボリューム識別情報と、第1記憶制御装置を特定するための第1識別情報と、第2記憶制御装置を特定するための第2識別情報と、第1記憶制御装置が第3ボリュームを使用しているか否かを示す第1使用情報と、第2記憶制御装置が第3ボリュームを使用しているか否かを示す第2使用情報と、ペアの解消後に第1ボリュームに差分データが生じていることを示す第1差分発生情報と、ペアの解消後に第2ボリュームに差分データが生じていることを示す第2差分発生情報と、を含んでいる。

【0016】

40

第9観点では、第8観点において、第1識別情報と、第1使用情報と、第1差分発生情報とは、第1記憶制御装置のみが更新可能であり、第2識別情報と、第2使用情報と、第2差分発生情報とは、第2記憶制御装置のみが更新可能である。

【0017】

第10観点では、第1観点において、使用制御情報を更新する場合には、更新後に、第3ボリュームから使用制御情報を読み出して正しく更新されたか否かを確認するようになっている。

【0018】

第11観点では、第1観点において、第1記憶制御装置は、使用制御情報に対応する第1管理テーブルを有しており、第2記憶制御装置は、使用制御情報に対応する第2管理テ

50

ーブルを有しており、第1管理テーブル及び第2管理テーブルは、使用制御情報の更新に対応して更新される。

【0019】

第12観点では、第1観点において、仮想ボリューム設定部は、第1ボリュームと第2ボリュームとの間に差分が発生した場合に、所定の契機に基づいて、差分が解消するように第1ボリュームの記憶内容と第2ボリュームの記憶内容とを再同期させる。

【0020】

第13観点では、第1観点において、仮想ボリュームに関するペアを削除する場合、仮想ボリューム設定部がペアを削除した後で、制御用ボリューム設定部は、仮想ボリュームに関する使用制御情報を削除させる。

10

【0021】

本発明の第14観点に従うストレージシステムの制御方法は、ホストコンピュータと、そのホストコンピュータにより利用される複数の記憶制御装置と、各記憶制御装置を管理するための管理装置とが通信可能に接続されているストレージシステムを制御するための制御方法であって、複数の記憶制御装置には、第1記憶制御装置と第2記憶制御装置及び第3記憶制御装置が含まれており、第1記憶制御装置の有する第1ボリュームと第2記憶制御装置の有する第2ボリュームとをペアとして設定することにより、一つの仮想ボリュームを生成してホストコンピュータに提供させるステップと、第3記憶制御装置の有する第3ボリュームを、仮想ボリュームの使用を制御するための使用制御情報を記憶させる制御用ボリュームとして設定するステップと、第3ボリュームに記憶される使用制御情報に、第1記憶制御装置及び第2記憶制御装置をそれぞれ特定するための識別情報を含ませるステップと、を備え、各ステップは、管理装置から第1記憶制御装置及び第2記憶制御装置に送られる指示に基づいて実行される。

20

【0022】

本発明の手段、機能、ステップの全部または一部は、コンピュータシステムにより実行されるコンピュータプログラムとして構成可能な場合がある。本発明の構成の全部または一部がコンピュータプログラムから構成された場合、このコンピュータプログラムは、例えば、各種記憶媒体に固定して配布等することができ、あるいは、通信ネットワークを介して送信することもできる。

【0023】

30

さらに、本発明の観点は明示された以外に種々組み合わせることも可能であり、そのような組合せは本発明の範囲に含まれる。

【図面の簡単な説明】

【0024】

【図1】本発明の実施形態を示す概念図である。

【図2】本発明の実施例に係るストレージシステムのハードウェア構成図である。

【図3】ホスト及び管理サーバのソフトウェア構成を模式的に示す説明図である。

【図4】ストレージ装置の記憶階層構造を示す説明図である。

【図5】仮想ボリュームの構成例を示す説明図である。

【図6】ロックディスクを管理するテーブルを示す説明図である。

40

【図7】ロック情報ビットマップの構成を模式的に示す説明図である。

【図8】使用制御情報の構成を示す説明図である。

【図9】仮想ボリュームを構成するリモートコピーペアを管理するためのテーブルを示す説明図である。

【図10】論理ボリュームを管理するテーブルを示す説明図である。

【図11】外部ボリュームを管理するテーブルを示す説明図である。

【図12】ロックディスク管理画面の説明図である。

【図13】第1ストレージ装置により実行される、ロックディスクを作成する処理を示すフローチャートである。

【図14】第2ストレージ装置により実行される、ロックディスクを作成する処理を示す

50

フローチャートである。

【図 1 5】ロックディスク作成時のロックディスク管理画面を示す説明図である。

【図 1 6】ロックディスク作成時のロックディスク管理テーブルを示す説明図である。

【図 1 7】リモートコピー管理画面の説明図である。

【図 1 8】リモートコピーペアに関するメニューの内容を示す説明図である。

【図 1 9】リモートコピーペアを作成するための画面を示す説明図である。

【図 2 0】リモートコピーペアに基づく仮想ボリュームを作成させるための処理を示すフローチャートである。

【図 2 1】仮想ボリューム作成時のリモートコピー管理画面の説明図である。

【図 2 2】仮想ボリューム作成時のペア管理テーブル T 2 0 を示す説明図である。

10

【図 2 3】ロックディスクを複数作成する場合のロックディスク管理画面を示す説明図である。

【図 2 4】一つのロックディスクに複数の仮想ボリュームを対応付ける場合のリモートコピー管理画面を示す説明図である。

【図 2 5】複数のロックディスクを作成する場合の、ロックディスク管理テーブルを示す説明図である。

【図 2 6】ペア管理テーブルを示す説明図である。

【図 2 7】ロックディスクを更新する場合の処理を示すフローチャートである。

【図 2 8】第 1 ストレージ装置の正ボリュームからデータを読み出す場合のリード処理を示すフローチャートである。

20

【図 2 9】第 2 ストレージ装置の副ボリュームからデータを読み出す場合のリード処理を示すフローチャートである。

【図 3 0】第 1 ストレージ装置の正ボリュームにデータを書き込む場合のライト処理を示すフローチャートである。

【図 3 1】第 2 ストレージ装置の副ボリュームにデータを書き込む場合のライト処理を示すフローチャートである。

【図 3 2】第 2 ストレージ装置の副ボリュームにデータを書き込む処理が失敗する場合を示すフローチャートである。

【図 3 3】仮想ボリュームの削除を行う場合の処理を示すフローチャートである。

【図 3 4】ロックディスクの削除を行う場合の処理を示すフローチャートである。

30

【図 3 5】ロックディスクの削除に問題が生じる場合を示すフローチャートである。

【図 3 6】リザーブコマンドを用いてロックディスクを削除する処理のフローチャートである。

【図 3 7】ロックディスクの削除と仮想ボリュームの削除とを連動して実行する処理を示すフローチャートである。

【図 3 8】サスペンド状態に移行する場合の処理を示すフローチャートである。

【図 3 9】リシンク処理を示すフローチャートである。

【図 4 0】スワップサスペンド状態に移行する場合の処理を示すフローチャートである。

【図 4 1】リバースリシンク処理を示すフローチャートである。

【図 4 2】自動的リバースリシンク処理を示すフローチャートである。

40

【符号の説明】

【0 0 2 5】

1, 2, 3 : ストレージ装置

1 A, 2 A : 論理ボリューム

1 B, 2 B : 論理ボリューム

3 A : ロックディスク

4 : 管理サーバ

4 A : 仮想ボリューム設定部

4 B : ロックディスク設定部

4 C : 外部接続設定部

50

5：ホスト
6：仮想ボリューム
10, 20, 30：ストレージ装置
70：ホスト
80：管理サーバ
100：コントローラ
140：共有メモリ
160：サービスプロセッサ
231：仮想ボリューム
232：ロックディスク
L10：使用制御情報。

10

【発明を実施するための最良の形態】

【0026】

図1は、本発明の実施形態の全体概要を示す構成説明図である。本実施形態では、後述のように、それぞれ異なるストレージ装置1, 2内の論理ボリューム1A, 2Aによって一つの仮想ボリューム6を形成する構成と、仮想的に形成される論理ボリューム1B, 2Bを別のストレージ装置3内の論理ボリューム3Aに接続させる構成と、論理ボリューム3Aを、仮想ボリューム6の使用を制御するための情報を記憶するロックディスクとして使用する構成とが、開示されている。

20

【0027】

ストレージシステムは、別々のストレージ装置1, 2内に存在する論理ボリューム1A, 2Aを仮想化して仮想ボリューム6を生成し、この仮想ボリューム6をホスト5に提供する。各論理ボリューム1A, 2Aには、同一のデバイス識別情報(LUN: Logical Unit Number)が設定されているため、ホスト5からは区別がつかない。正確には、副ボリューム2Aには、正ボリューム1Aのデバイス識別情報が設定されている。

【0028】

論理ボリューム1A, 2Aはリモートコピーのペアを形成しており、例えば、論理ボリューム1Aが正ボリュームに、論理ボリューム2Aが副ボリュームに、なっている。正ボリューム1Aに書き込まれたデータは、副ボリューム2Aに転送されて書き込まれる。正ボリューム1Aまたは副ボリューム2Aのいずれか一つに障害が発生した場合でも、正常なボリュームを用いてデータ入出力を行うことができる。

30

【0029】

ロックディスク3Aには、正ボリューム1Aと副ボリューム2Aとのどちらに差分が発生したか等を示す情報が記憶されている。各ストレージ装置1, 2は、ロックディスク3Aを共有しており、ロックディスク3Aに記憶された情報(使用制御情報)に基づいて、仮想ボリューム6を運用する。

【0030】

従って、本実施形態では、ロックディスク3Aを用いることにより、障害発生時等に、ホスト5が古いデータにアクセスするのを防止する。さらに、本実施形態では、管理サーバ4からの操作によって、仮想ボリューム6の設定及びロックディスク3Aの設定等を行うことができるようになっている。

40

【0031】

図1に示すストレージシステムを説明する。ストレージシステムは、例えば、「記憶制御装置」としてのストレージ装置1, 2, 3と、「管理装置」としての管理サーバ4と、「ホストコンピュータ」としてのホスト5とを備えて構成される。

【0032】

先に接続構成について説明する。第1ストレージ装置1及び第2ストレージ装置2は、「第1通信経路」としての第1通信ネットワークCN1を介して、ホスト5に接続されている。さらに、第1ストレージ装置1と第2ストレージ装置2とは、第2通信経路CN2を介して接続されている。

50

【0033】

第1ストレージ装置1と第2ストレージ装置2とは、「第3通信経路」としての第3通信ネットワークCN3を介して、第3ストレージ装置3に接続されている。管理サーバ4は、「第4通信経路」としての第4通信ネットワークCN4を介して、各ストレージ装置1～3及びホスト5と接続されている。

【0034】

例えば、通信ネットワークCN1、CN3は、FC__SAN (Fibre Channel Storage Area Network)、または、IP__SAN (Internet Protocol SAN) 等を用いて構成することができる。第4通信ネットワークCN4は、例えば、LAN (Local Area Network)、または、WAN (Wide Area Network) 等を用いて構成可能である。第2通信経路CN2は、例えば、各ストレージ装置1、2間を直接結ぶファイバケーブルまたはメタルケーブルとFCプロトコルとから構成可能である。なお、以上は例示であって、本発明は、上記の接続構成に限定されない。また、有線接続に限らず、無線によってデータ通信可能に接続する構成でもよい。

10

【0035】

各ストレージ装置1～3は、それぞれ物理的に異なる装置として構成されており、論理ボリューム1A～3Aをそれぞれ備えている。各ストレージ装置1～3は、複数の記憶装置を備えることができ、これら記憶装置の有する物理的記憶領域上に、論理デバイスとしての論理ボリュームが形成される。論理ボリューム1A～3Aは、例えば、RAID5やRAID6のような冗長化された物理的記憶領域上に設けることができる。なお、以下の説明では、「論理ボリューム」を「ボリューム」と略記する場合がある。図中では、論理デバイスとしての論理ボリュームを「LDEV」と表示している。

20

【0036】

記憶装置としては、例えば、ハードディスクデバイス、半導体メモリデバイス、光ディスクデバイス、光磁気ディスクデバイス、磁気テープデバイス、フレキシブルディスクデバイス等のデータを読み書き可能な種々のデバイスを利用可能である。

【0037】

記憶装置としてハードディスクデバイスを用いる場合、例えば、FC (Fibre Channel) ディスク、SCSI (Small Computer System Interface) ディスク、SATA ディスク、ATA (AT Attachment) ディスク、SAS (Serial Attached SCSI) ディスク等を用いることができる。

30

【0038】

記憶装置として半導体メモリデバイスを用いる場合、例えば、フラッシュメモリ、FeRAM (Ferroelectric Random Access Memory)、MRAM (Magnetoresistive Random Access Memory)、相変化メモリ (Ovonic Unified Memory)、RRAM (Resistance RAM) 等の種々のメモリデバイスを利用可能である。なお、記憶装置の種類は、上記のものに限定されず、将来製品化される他の種類の記憶装置を利用できるであろう。

【0039】

図1では、各ストレージ装置1～3が、それぞれ実論理ボリューム1A～3Aを有する場合を示している。実論理ボリュームとは、記憶装置の物理的記憶領域に直接的に対応付けられているボリュームである。第1ストレージ装置1及び第2ストレージ装置2は、外部の第3ストレージ装置3が有する論理ボリューム3Aを取り込んで、使用することもできる。外部のストレージ装置3が有する論理ボリューム3Aを自装置内に取り込み、それがあたかも自分自身の実論理ボリュームであるかのようにして使用する技術は、特開2005-107645号公報に記載されている。この公報に記載の技術を、本実施形態において援用することができる。

40

【0040】

従って、第1ストレージ装置1及び第2ストレージ装置2は、ハードディスクドライブ等の記憶装置を備えない構成でもよい。この場合、各ストレージ装置1、2は、例えば、スイッチ装置や仮想化装置のようなコンピュータ装置として構成可能である。

50

【0041】

管理サーバ4は、各ストレージ装置1～3の構成を管理したり、ホスト5に指示を与えるための装置である。管理サーバ4は、ストレージシステムを管理するための基本的機能に加えて、仮想ボリューム設定部4Aと、「制御用ボリューム設定部」としてのロックディスク設定部4Bと、「対応付け設定部」としての外部接続設定部4Cを備える。

【0042】

仮想ボリューム設定部4Aは、それぞれ異なるストレージ装置1, 2内に存在する別々の論理ボリューム1A, 2Aを、一つの仮想ボリューム6に仮想化してホスト5に提供させるための機能である。その仮想ボリューム6を、例えば、リモートコピーペア型仮想ボリュームと呼ぶこともできる。

10

【0043】

ロックディスク設定部4Bは、第3ストレージ装置3内の論理ボリューム3Aをロックディスクとして使用するための設定を行う機能である。以下、便宜上、論理ボリューム3Aをロックディスク3Aと呼ぶ場合がある。ロックディスク3Aには、仮想ボリューム6を使用するために参照される使用制御情報が記憶される。

【0044】

使用制御情報には、図8で後述するように、例えば、ロックディスク3Aを特定するための識別情報と、第1ストレージ装置1を特定するための識別情報と、第2ストレージ装置2を特定するための識別情報と、第1ストレージ装置1がロックディスク3Aを使用しているか否かを示す情報と、第2ストレージ装置2がロックディスク3Aを使用しているか否かを示す使用情報と、リモートコピーペアの解消後に第1ボリューム1Aに差分データが生じていることを示す情報と、リモートコピーペアの解消後に第2ボリューム2Aに差分データが生じていることを示す情報と、が含まれている。

20

【0045】

外部接続設定部4Cは、第1ストレージ装置1内のボリューム1Bと第3ストレージ装置3内のロックディスク3Aとを対応付け、さらに、第2ストレージ装置2内のボリューム2Bと第3ストレージ装置3内のロックディスク3Aとを対応付ける。第1ストレージ装置1は、自装置内のボリューム1Bを介して、ロックディスク3Aにアクセスする。同様に、第2ストレージ装置2は、自装置内のボリューム2Bを介して、ロックディスク3Aにアクセスする。ボリューム1Bに関するコマンドは、外部のロックディスク3Aへのコマンドに変換されて、第1ストレージ装置1から第3ストレージ装置3に転送される。同様に、ボリューム2Bへのコマンドは、外部のロックディスク3Aへのコマンドに変換されて、第2ストレージ装置2から第3ストレージ装置3に転送される。

30

【0046】

ホスト5は、例えば、メインフレームコンピュータ、サーバコンピュータ、エンジニアリングワークステーション、パーソナルコンピュータ等のコンピュータ装置として構成される。ホスト5がメインフレームコンピュータである場合、例えば、FICON（Fibre Connection：登録商標）、ESCON（Enterprise System Connection：登録商標）、ACONARC（Advanced Connection Architecture：登録商標）、FIBARC（Fibre Connection Architecture：登録商標）のような通信プロトコルが用いられる。ホスト5がサーバコンピュータやパーソナルコンピュータ等である場合、例えば、TCP/IP（Transmission Control Protocol/Internet Protocol）、FCP（Fibre Channel Protocol）、iSCSI（internet Small Computer System Interface）のような通信プロトコルが用いられる。

40

【0047】

ホスト5は、例えば、アプリケーションプログラム（以下、「アプリケーション」と略記する場合がある）5Aと、パス制御部5Bと、通信部5Cとを備えている。なお、各ストレージ装置1～3、管理サーバ4、ホスト5のハードウェア構成については、後述の実施例で説明する。アプリケーションプログラム5Aは、例えば、電子メール管理ソフトウェア、顧客管理ソフトウェア、文書作成ソフトウェア等の各種業務を遂行するための一つ

50

または複数のソフトウェア製品である。

【0048】

パス制御部5Bは、ホスト5が使用するアクセスパス（以下、「パス」と略記する場合がある）を切り替えて使用するソフトウェアである。ホスト5は、一方のパスP1を介して第1ストレージ装置1の論理ボリューム1Aに接続されている。ホスト5は、他方のパスP2を介して第2ストレージ装置2の論理ボリューム2Aに接続されている。

【0049】

通常の場合、一方のパスP1がアクティブパスとなり、他方のパスP2がパッシブパスとなる。パス制御部5Bは、アクティブパスP1を用いて仮想ボリューム6にアクセスできない場合、パッシブパスP2に切り替えて仮想ボリューム6にアクセスする。

10

【0050】

ホスト5は、例えば、Inquiryコマンドのような問合せコマンドを各ストレージ装置1, 2に送信することにより、各ストレージ装置1, 2内に設けられている各論理ボリューム1A, 2Aの識別子、装置番号、LU番号、パス情報等を得ることができる。パス制御部5Bは、同一の識別子を有するパスが複数検出された場合、その複数のパスを交替パスとして認識する。

【0051】

つまり、同一の仮想ボリューム6へアクセスするためのパスが複数検出された場合、一方のパスP1を通常時に使用するアクティブパス（プライマリパスとも呼ぶ）として認識し、他方のパスP2を異常時に使用するパッシブパス（セカンダリパスとも呼ぶ）として認識する。

20

【0052】

仮想ボリューム6は、それぞれ異なるストレージ装置1, 2内の論理ボリューム1A, 2Aを仮想化することにより構成される。仮想ボリューム6は、仮想ボリューム設定部4Aが各ストレージ装置1, 2に指示を与えることにより、生成される。仮想ボリューム6を構成する各論理ボリューム1A, 2Aは、例えば、構成要素ボリュームと呼ぶことができる。

【0053】

論理ボリューム1Aは、仮想ボリューム6内の正ボリュームとして設定されており、論理ボリューム2Aは、仮想ボリューム6内の副ボリュームとして設定される。但し、後述の実施例から明らかなように、正ボリュームであるか副ボリュームであるかは、必要に応じて変更される。論理ボリューム1Aにアクセス障害が発生した場合、論理ボリューム2Aの属性が「副ボリューム」から「正ボリューム」に変更される。論理ボリューム2Aの属性を正ボリュームに変更する場合、論理ボリューム2Aに設定されているデバイス識別情報は、そのまま保持される。論理ボリューム2Aのデバイス識別情報を、論理ボリューム1Aのデバイス識別情報とは別の値に変更してしまうと、ホスト5は、別の論理ボリュームとして認識してしまうためである。

30

【0054】

正ボリュームは、通常時にホスト5からアクセスされるボリュームであり、副ボリュームは、障害発生時にホスト5からアクセスされるボリュームである。従って、正ボリュームはアクティブボリュームと、副ボリュームはパッシブボリュームと呼び変えることもできる。仮想ボリューム6を構成する正ボリュームと副ボリュームとがコピーペアを形成する場合、正ボリュームをコピー元ボリュームと、副ボリュームをコピー先ボリュームと呼ぶことができる。

40

【0055】

仮想ボリューム6には、ストレージシステム内で仮想ボリューム6を一意に特定するための識別子が設定される。図1に示す例では、識別子として「#12」が仮想ボリューム6に設定されている。

【0056】

仮想ボリューム6に設定される識別子は、例えば、仮想ボリューム6を構成する各論理

50

ボリューム 1 A, 2 A の元の識別子に基づいて生成される。図 1 に示す例では、一方の論理ボリューム 1 A の元の識別子「# 1」、他方の論理ボリューム 2 A の元の識別子は「# 2」である。そこで、仮想ボリューム 6 には、一方の論理ボリューム 1 A の識別子「# 1」と他方の論理ボリューム 2 A の識別子「# 2」とを合体させて得られる識別子「# 1 2」が設定されている。仮想ボリューム 6 に設定される識別子は、ストレージシステム内に存在する他の各論理ボリュームの識別子と重複しないように生成される。

【0057】

仮想ボリューム 6 が設定されると、各ストレージ装置 1, 2 は、仮想ボリューム 6 を構成する論理ボリューム 1 A, 2 A に、仮想ボリューム 6 の識別子「# 1 2」と同一の識別子を設定する。つまり、第 1 ストレージ装置 1 は、論理ボリューム 1 A の識別子として識別子「# 1 2」を設定する。ストレージ装置 2 も、論理ボリューム 2 A の識別子として識別子「# 1 2」を設定する。「# 1 2」を仮想ボリューム 6 を特定するための仮想識別子と呼ぶことができる。

10

【0058】

仮想識別子「# 1 2」は、仮想ボリューム 6 を構成する各論理ボリューム 1 A, 2 A の元々の識別子「# 1」, 「# 2」に優先して設定される。従って、ホスト 5 からの問合せに対し、第 1 ストレージ装置 1 は論理ボリューム 1 A の識別子として仮想識別子「# 1 2」を返し、第 2 ストレージ装置 2 も論理ボリューム 2 A の識別子として仮想識別子「# 1 2」を返す。これにより、バス制御部 5 B は、論理ボリューム 1 A 及び論理ボリューム 2 A を、同一のボリューム（仮想ボリューム 6）として認識する。

20

【0059】

各論理ボリューム 1 A, 2 A に元々設定されている識別子「# 1」, 「# 2」は、各ストレージ装置 1, 2 内で論理ボリューム 1 A, 2 A を管理するために使用される、内部的な識別情報である。これに対し、仮想識別子「# 1 2」は、仮想ボリューム 6 をホスト 5 に認識させるための、対外的な識別情報である。

【0060】

論理ボリューム 1 A にアクセスするためのパス P 1 と、論理ボリューム 2 A にアクセスするためのパス P 2 とは、仮想ボリューム 6 にアクセスするためのパスとして、バス制御部 5 B に認識される。

【0061】

本ストレージシステムの動作について説明する。最初に、ユーザは、外部接続設定部 4 C を用いて、第 3 ストレージ装置 3 内の論理ボリューム 3 A と、第 1 ストレージ装置 1 内の仮想的な論理ボリューム 1 B 及び第 2 ストレージ装置 2 内の仮想的な論理ボリューム 2 B とを対応付けておく。

30

【0062】

次に、ユーザは、ロックディスク設定部 4 B を用いて、第 3 ストレージ装置 3 内の論理ボリューム 3 A を、仮想ボリューム 6 の使用を制御するためのロックディスク 3 A として設定する。

【0063】

さらに、ユーザは、仮想ボリューム設定部 4 A を用いて、仮想ボリューム 6 を構成させる各論理ボリューム 1 A, 2 A をそれぞれ指定し、各論理ボリューム 1 A, 2 A とロックディスク 3 A との対応関係を設定する。

40

【0064】

アプリケーションプログラム 5 A が、仮想ボリューム 6 にデータを書き込む場合、バス制御部 5 B は、アクティブパス P 1 を用いて、論理ボリューム 1 A にライトコマンドを発行する。

【0065】

第 1 ストレージ装置 1 は、ホスト 5 から受信したライトデータを論理ボリューム 1 A に書き込む。さらに、第 1 ストレージ装置 1 は、論理ボリューム 1 A と共に仮想ボリューム 6 を構成する論理ボリューム 2 A に、通信経路 C N 2 を介して、そのライトデータを送信

50

する。

【0066】

第2ストレージ装置2は、第1ストレージ装置1から受信したライトデータを論理ボリューム2Aに書込む。このように、仮想ボリューム6を提供する各ストレージ装置1, 2は、それぞれの論理ボリューム1A, 2Aにライトデータを書き込む。従って、通常の場合、仮想ボリューム6を構成する各論理ボリューム1A, 2Aは、それぞれの記憶内容が一致している。

【0067】

もしも第2ストレージ装置2に障害が発生したり、あるいは、第1ストレージ装置1と第2ストレージ装置2とを結ぶ通信経路CN2が切断された場合、ストレージシステムは運用を停止せずに、第1ストレージ装置1を用いて、仮想ボリューム6をホスト5に提供する。

10

【0068】

ストレージシステムの運用を継続すると、第1ストレージ装置1内の論理ボリューム1Aに新たなデータが蓄積されていき、論理ボリューム2Aの記憶内容と論理ボリューム1Aの記憶内容とに差分が生じる。論理ボリューム1Aに差分が発生した事は、第1ストレージ装置1によって、ロックディスク3A内の使用制御情報に書き込まれる。

【0069】

例えば、第2ストレージ装置2が障害から復旧した場合、または、第1ストレージ装置1と第2ストレージ装置2とを接続する通信経路CN2とが正常に戻った場合、論理ボリューム1A（正ボリューム）に蓄積された差分データは、論理ボリューム2A（副ボリューム）に転送される。これにより、正ボリューム1Aの記憶内容と副ボリューム2Aの記憶内容とは同期する。

20

【0070】

ホスト5が論理ボリューム2Aにアクセスしようとした場合、第2ストレージ装置2は、ロックディスク3A内の使用制御情報を参照する。使用制御情報には、各ボリューム1A, 2Aが同期していないこと、及び、論理ボリューム1Aを用いて仮想ボリューム6が運用されていること等が記憶されている。従って、第2ストレージ装置2は、ホスト5からのアクセスに応答せず、エラーを返す。これにより、ホスト5が古いデータにアクセスするのを未然に防止することができる。

30

【0071】

上記例とは逆に、第1ストレージ装置1に障害等が発生し、第2ストレージ装置2が論理ボリューム2Aを用いて仮想ボリューム6を運用する場合も同様である。この場合は、論理ボリューム2Aに差分データが蓄積される。使用制御情報には、論理ボリューム2Aに差分データが蓄積されていること、及び、論理ボリューム2Aを用いて仮想ボリューム6が運用されていること等が記憶される。仮想ボリューム6に関する主導権を得ていない第1ストレージ装置1は、ホスト5からのアクセスに対応しない。従って、ホスト5が古いデータ（論理ボリューム1A内のデータ）にアクセスするのを未然に防止できる。

【0072】

このように構成される本実施形態によれば、以下の効果を奏する。本実施形態では、第1ストレージ装置1及び第2ストレージ装置2とは別の第3ストレージ装置3内にロックディスク3Aを設け、そのロックディスク3A内に、論理ボリューム1Aと論理ボリューム2Aとから構成される仮想ボリューム6の使用を制御するための使用制御情報を格納させる。従って、本実施形態では、各ストレージ装置1, 2は、ロックディスク3Aを共有することにより、ストレージ装置1, 2間での切替を適切に行うことができ、ホスト5が各ストレージ装置1, 2の切替を意識する必要は無い。

40

【0073】

本実施形態では、使用制御情報には、第1ストレージ装置1及び第2ストレージ装置2をそれぞれ特定するための識別情報が含まれている。これにより、他のストレージ装置にロックディスク3Aを関連づける等の誤りを未然に防止できる。

50

【0074】

本実施形態では、各ストレージ装置1, 2内に仮想的に設けられる論理ボリューム1B, 2Bにロックディスク3Aを対応付け、論理ボリューム1B, 2Bを介してロックディスク3Aを使用する。従って、第1ストレージ装置1, 2の有するキャッシュメモリ量や機能を利用して、ロックディスク3Aにアクセスできる。

【0075】

本実施形態では、管理サーバ4内に、仮想ボリューム設定部4A, ロックディスク設定部4B, 外部接続設定部4Cを設ける。従って、ユーザは、管理サーバ4の各設定部4A~4Cを用いて、例えば、仮想ボリューム6の生成及び削除、ロックディスク3Aの生成及び対応付け、論理ボリューム1B, 2Bとロックディスク3Aとの接続等を行うことが

10

【0076】

後述の実施例に記載のように、使用制御情報に含まれる各情報のうち、第1ストレージ装置1を識別するための情報と、第1ストレージ装置1がロックディスク3Aを使用していることを示す情報と、論理ボリューム1Aに差分データが発生していることを示す情報とは、第1ストレージ装置1のみが更新可能である。同様に、使用制御情報に含まれる各情報のうち、第2ストレージ装置2を識別するための情報と、第2ストレージ装置2がロックディスク3Aを使用していることを示す情報と、論理ボリューム2Aに差分データが発生していることを示す情報とは、第2ストレージ装置2のみが更新可能である。従って、第1ストレージ装置1が誤って第2ストレージ装置2に関する情報を書き換えたり、その逆に、第2ストレージ装置2が誤って第1ストレージ装置1に関する情報を書き換えたりするのを未然に防止でき、信頼性が向上する。

20

【0077】

さらに、後述の実施例で明らかとなるように、使用制御情報を更新する場合には、その更新後に、ロックディスク3Aから使用制御情報を読み出して、正しく更新されたか否かを確認する。従って、それぞれ異なるストレージ装置1, 2が一つのロックディスク3Aを共有する場合でも、使用制御情報が適切に更新されることを保証でき、ストレージシステムの信頼性を高めることができる。

【0078】

さらに、後述の実施例で明らかにされるように、ロックディスクの削除時に、一つの指示で、仮想ボリュームも削除させることができる。これにより、使い勝手を高めることができる。以下、本実施形態を詳細に説明する。

30

【実施例1】

【0079】

図2は、本実施例によるストレージシステムの全体概要を示す説明図である。先に図1との対応関係について説明すると、図2中のストレージ装置10, 20, 30は、図1中のストレージ装置1, 2, 3にそれぞれ対応する。図2中のホスト70は図1中のホスト5に、図2中の管理サーバ80は図1中の管理サーバ4に、それぞれ対応する。

【0080】

図5に示す仮想ボリューム231は、図1中の仮想ボリューム6に対応する。図5に示すロックディスク232は、図1中のロックディスク3Aに対応する。図4に示す論理ボリューム230は、図1中の論理ボリューム1A, 2Aに対応する。第1通信ネットワークCN10は第1通信ネットワークCN1に、第2通信経路CN20は第2通信経路CN2に、第3通信ネットワークCN30は第3通信ネットワークCN3に、第4通信経路CN40は第4通信経路CN4に、対応する。以下、図1で述べた説明と重複する部分は簡単に説明する。

40

【0081】

ストレージシステムは、複数のストレージ装置10, 20, 30と、ホスト70と、管理サーバ80とを備えている。ストレージ装置10, 20とホスト70とは、通信ネットワークCN10を介して接続される。ストレージ装置10とストレージ装置20とは、通

50

信経路CN20を介して接続される。管理サーバ80は、通信ネットワークCN40を介して、各ストレージ装置10～30及びホスト70と接続される。ストレージ装置10, 20とストレージ装置30とは、通信経路CN30を介して接続される。

【0082】

但し、これに限らず、例えば、通信ネットワークCN10とCN30とを同一の通信ネットワークとして構成してもよい。また、管理用ネットワークCN40を廃止し、通信ネットワークCN10を用いて管理用の情報を流通させる構成でもよい。

【0083】

図2に示す構成は、ストレージ装置10, 20が外部接続の接続元となり、ストレージ装置30が外部接続の接続先となっている例を示している。外部接続とは、上述の通り、
10
自装置の外部に存在する論理ボリュームを自装置の内部に取り込む技術である。外部接続の接続元となるストレージ装置10, 20は、ストレージ装置30内の論理ボリューム230を利用可能である。従って、ある程度の量のキャッシュメモリを備えているのであれば、ストレージ装置10, 20は、実ボリュームを備える必要はない。各ストレージ装置10, 20は、スイッチ装置または仮想化専用装置等のように構成できる。

【0084】

ストレージ装置10～30の構成を説明する。各ストレージ装置10～30は、同一構成であってもよい。そこで、ストレージ装置10を例に挙げて説明する。

【0085】

ストレージ装置10は、例えば、コントローラ100と記憶装置搭載部（以下、HDU
20
とも呼ぶ）200とを備える。コントローラ100は、ストレージ装置10の動作を制御するものである。コントローラ100は、例えば、チャネルアダプタ110（以下、「CHA110」と）、ディスクアダプタ120（以下、「DKA120」と）、キャッシュメモリ130（図中、「CM」と）、共有メモリ140（図中、「SM」と）、接続制御部150（図中、「SW」と）、サービスプロセッサ160（図中、「SVP」と）を備えている。

【0086】

第1通信制御部と表現可能なCHA110は、ホスト70または他のストレージ装置との間でデータ通信を行うものである。各CHA110は、図4に示すように、通信ポート111（111A及び111Bの総称として符号111を用いる）をそれぞれ少なくとも
30
一つ以上備える。各CHA110は、それぞれCPUやメモリ等を備えたマイクロコンピュータシステムとして構成される。各CHA110は、ホスト70から受信したリードコマンドやライトコマンド等の各種コマンドを解釈して実行する。

【0087】

なお、通信機能とコマンド解釈及び実行機能とを分ける構成としてもよい。例えば、ホスト70または他のストレージ装置と通信を行うための通信制御基板と、コマンドを解釈して実行するための実行制御基板とを分ける構成でもよい。

【0088】

各CHA110には、それぞれを識別するためのネットワークアドレス（例えば、IP
40
アドレスやWWN（World Wide Name））が割り当てられている。各CHA110は、それぞれが個別にNAS（Network Attached Storage）として振る舞うこともできる。複数のホスト70が存在する場合、各CHA110は、各ホスト70からの要求をそれぞれ個別に受け付けて処理する。

【0089】

第2通信制御部と表現可能なDKA120は、HDU200が有するディスクドライブ210との間でデータ授受を行う。各DKA120は、CHA110と同様に、CPUやメモリ等を備えたマイクロコンピュータシステムとして構成される。前記同様に、通信機能とコマンド解釈及び実行機能とを分離する構成でもよい。

【0090】

各DKA120は、例えば、CHA110がホスト70から受信したデータや、他のス
50

ストレージ装置からのデータ等を、所定のディスクドライブ210に書込む。また、各DKA120は、所定のディスクドライブ210からデータを読み出し、ホスト70または外部のストレージ装置に送信する。ディスクドライブ210との間でデータ入出力を行う場合、各DKA120は、論理的なアドレスを物理的なアドレスに変換する。

【0091】

各DKA120は、ディスクドライブ210がRAIDに従って管理されている場合、RAID構成に応じたデータアクセスを行う。例えば、各DKA120は、同一のデータを別々のディスクドライブ群(RAIDグループ)にそれぞれ書き込んだり(RAID1)、あるいは、パリティ計算を実行し、データ及びパリティをディスクドライブ群に分散させて書き込む(RAID5、RAID6等)。

10

【0092】

キャッシュメモリ130は、ホスト70または外部のストレージ装置から受信したデータを記憶する。また、キャッシュメモリ130は、ディスクドライブ210から読み出されたデータを記憶する。後述のように、キャッシュメモリ130の記憶空間を利用して、仮想的な中間記憶装置(VDEV)が構築される。

【0093】

共有メモリ(制御メモリとも呼ばれる場合もある)140には、ストレージ装置10の作動に使用するための各種制御情報等が格納される。また、共有メモリ140には、ワーク領域が設定されるほか、後述する各種テーブル類も格納される。

【0094】

20

なお、ディスクドライブ210のいずれか一つあるいは複数を、キャッシュ用のディスクとして使用してもよい。また、キャッシュメモリ130と共有メモリ140とは、それぞれ別々のメモリとして構成することもできるし、同一のメモリの一部の記憶領域をキャッシュ領域として使用し、他の記憶領域を制御領域として使用することもできる。

【0095】

接続制御部150は、各CHA110、各DKA120、キャッシュメモリ130及び共有メモリ140を相互に接続させる。接続制御部150は、例えば、クロスバスイッチ等として構成可能である。

【0096】

HDU200は、複数のディスクドライブ210を備えている。ディスクドライブ210としては、例えば、ハードディスクドライブ、フラッシュメモリデバイス、磁気テープドライブ、半導体メモリドライブ、光ディスクドライブ等のような各種記憶装置及びこれらの均等物を用いることができる。

30

【0097】

例えば、複数のディスクドライブ210の物理的記憶領域を一つにまとめてRAIDグループ220を構成することができる。このRAIDグループ220の物理的記憶領域には、論理ボリューム230を少なくとも一つ以上設けることができる。

【0098】

SVP160は、LAN等の内部ネットワークを介して、各CHA110とそれぞれ接続されている。SVP160は、CHA110を介して、共有メモリ140やDKA120とデータ授受を行うことができる。SVP160は、ストレージ装置10内の各種情報を収集し、管理サーバ80に提供する。

40

【0099】

他のストレージ装置20、30もストレージ装置10と同様に構成可能である。しかし、ストレージ装置10～30の構成がそれぞれ異なってもよい。例えば、各ストレージ装置10～30の機種、ベンダ、型式、世代等が異なっても、本発明を適用することができる。

【0100】

ホスト70のハードウェア構成を説明する。ホスト70は、例えば、CPU71と、メモリ72と、HBA(Host Bus Adapter)73と、LANインターフェース74及び内蔵

50

ディスク 75 を備えて構成される。

【0101】

HBA 73 は、通信ネットワーク CN10 を介して、ストレージ装置 10, 20 にアクセスするための通信部であり、図 1 中の通信部 5C に対応する。LAN インターフェース 74 は、管理用の通信ネットワーク CN40 を介して、管理サーバ 80 と通信するための回路である。

【0102】

管理サーバ 80 の構成を説明する。管理サーバ 80 は、ストレージシステムの構成等を管理するためのコンピュータ装置であり、例えば、システム管理者や保守員等のユーザにより操作される。管理サーバ 80 は、例えば、CPU 81 と、メモリ 82 と、ユーザインターフェース 83 (図中、「UI」) と、LAN インターフェース 84 及び内蔵ディスク 85 を備えて構成される。LAN インターフェース 84 は、管理用の通信ネットワーク CN40 を介して、各ストレージ装置 10 ~ 30 及びホスト 70 と通信を行う。

10

【0103】

ユーザインターフェース 83 は、ユーザに後述の管理画面を提供し、かつ、ユーザからの入力を受け取るためのものである。ユーザインターフェース 83 は、例えば、ディスプレイ装置、キーボードスイッチ、ポインティングデバイス等から構成される。なお、例えば、音声入力によって各種の入力を行うことができる構成でもよい。

【0104】

図 3 は、ホスト 70 及び管理サーバ 80 のソフトウェア構成を示す説明図である。図 3 (a) に示すように、ホスト 70 は、例えば、オペレーティングシステム 76 と、HBA ドライバ 77 と、パス制御ソフトウェア 78 と、アプリケーションプログラム 79 とを備えている。

20

【0105】

HBA ドライバ 77 は、HBA 73 を制御するためのソフトウェアである。パス制御ソフトウェア 78 は、図 1 中のパス制御部 5B に対応する。パス制御ソフトウェア 78 は、アプリケーションプログラム 79 からのアクセス要求に応じて、どのアクセスパスを用いてアクセスすべきかを決定する。パス制御ソフトウェア 78 は、アクセス先のボリュームに繋がるアクセスパスが複数存在する場合、プライマリに設定されているアクセスパス (アクティブパス) とセカンダリに設定されているパス (パッシブパス) とを切り替えて使用する。

30

【0106】

以下の説明では、パス制御ソフトウェア 78 をパス制御部 78 と呼ぶ場合がある。アプリケーションプログラム 79 は、図 1 中のアプリケーションプログラム 5A に対応するソフトウェアである。

【0107】

図 3 (b) に示すように、管理サーバ 80 は、オペレーティングシステム 86 と、LAN カードドライバ 87 と、管理プログラム 88 とを備えている。管理プログラム 88 は、仮想ボリューム 231 の設定をストレージ装置に指示する機能と、ロックディスク 232 の作成等をストレージ装置に指示する機能と、ストレージ装置 30 の有する実ボリューム 230 をストレージ装置 10, 20 内の仮想的なボリューム (外部接続ボリューム) として設定するための機能とを備える。管理プログラム 88 は、図 1 中の仮想ボリューム設定部 4A、ロックディスク設定部 4B 及び外部接続設定部 4C に対応する。

40

【0108】

図 4 は、ストレージシステムの記憶構造を示す説明図である。図 4 には、上述した外部接続に関する構成等が示されている。

【0109】

ストレージ装置 10, 20 の記憶構造は、例えば、物理的記憶階層と論理的記憶階層とに大別できる。物理的記憶階層は、物理的なディスクである PDEV (Physical Device) 210 により構成される。PDEV は、ディスクドライブ 210 に該当する。

50

【0110】

論理的記憶階層は、複数の（例えば2種類の）階層から構成可能である。一つの論理的階層は、VDEV (Virtual Device) 220、または、VDEV220のように扱われる仮想的なVDEV221のいずれ一つから構成可能である。他の一つの論理的階層は、LDEV (Logical Device) 230から構成することができる。

【0111】

VDEV220は、例えば、4個1組(3D+1P)、8個1組(7D+1P)等のような所定数のPDEV210をグループ化して構成される。グループに属する各PDEV210によりそれぞれ提供する記憶領域を集合させて、一つのRAID記憶領域が形成される。このRAID記憶領域がVDEV220となる。

10

【0112】

VDEV220が物理的な記憶領域上に構築されるのとは対照的に、VDEV221は、物理的な記憶領域を直接には必要としない仮想的な中間記憶装置である。VDEV221は、物理的な記憶領域に直接関係づけられるものではなく、外部ストレージ装置としての第3ストレージ装置30のLU (Logical Unit) をマッピングするための受け皿となっている。ここでは、接続元であるストレージ装置10、20から見て、接続先のストレージ装置30はストレージ装置10、20の外部に存在するため、外部ストレージ装置30と呼ぶ。

【0113】

LDEV230は、VDEV220またはVDEV221上に、それぞれ少なくとも一つ以上設けることができる。LDEV230とは、上述の論理ボリューム230のことである。LDEV230は、例えば、VDEV220を所定サイズで分割することにより構成される。

20

【0114】

オープン系ホストの場合、LDEV230がLU240にマッピングされることにより、ホスト70は、LDEV230を一つの物理的なディスクとして認識する。オープン系のホストは、LUN (Logical Unit Number) や論理ブロックアドレスを指定することにより、所望のLDEV230にアクセスする。なお、メインフレーム系ホストの場合は、LDEV230を直接認識する。

【0115】

LU240は、SCSIの論理ユニットとして認識可能なデバイスである。各LU240は、ターゲットポート111Aを介してホスト70に接続される。各LU240には、少なくとも一つ以上のLDEV230をそれぞれ関連付けることができる。なお、一つのLU240に複数のLDEV230を関連付けることにより、LUサイズを仮想的に拡張することもできる。

30

【0116】

CMD (Command Device) 250は、ホスト70とストレージ装置10、20との間で、コマンドやステータスを受け渡すために使用される専用のLUである。ホスト70からのコマンドは、CMD250に書き込まれる。ストレージ装置10、20は、CMD250に書き込まれたコマンドに応じた処理を実行し、その実行結果をステータスとしてCMD250に書き込む。ホスト70は、CMD250に書き込まれたステータスを読み出して確認し、次に実行すべき処理内容をCMD250に書き込む。このようにして、ホスト70は、CMD250を介して、ストレージ装置10、20に各種の指示を与えることができる。

40

【0117】

なお、ホスト70から受信したコマンドを、CMD250に格納せずに、ストレージ装置10、20は、直接的に処理することもできる。また、実体のデバイス(LU)を定義せずに、CMDを仮想的なデバイスとして生成し、ホスト70からのコマンドを受け付けて処理するように構成してもよい。例えば、CHA110は、ホスト70から受信したコマンドを共有メモリ140に書き込み、この共有メモリ140に記憶されたコマンドを、

50

CHA110又はDKA120が処理する。その処理結果は共有メモリ140に書き込まれ、CHA110からホスト70に送信される。

【0118】

ストレージ装置10、20のイニシエータポート(External Port)111Bには、通信経路CN30を介して、外部ストレージ装置30が接続されている。この通信ポート111Bは、外部接続用の通信ポートである。

【0119】

外部ストレージ装置30は、複数のPDEV210と、PDEV210の提供する記憶領域上に設定されたVDEV220と、VDEV220上に少なくとも一つ以上設定可能なLDEV230とを備えている。そして、各LDEV230は、LU240にそれぞれ関連付けられている。

10

【0120】

外部ストレージ装置30のLU240は、VDEV221にマッピングされている。LDEV230Aは、仮想的なVDEV221に対応付けられている。ストレージ装置10、20は、LDEV230Aを介して、外部ストレージ装置30内の論理ボリューム(ロックディスク)を使用する。

【0121】

図5は、ストレージシステムの構成を模式的に示す説明図である。図5に示すように、ホスト70とストレージ装置10とは複数の通信パスP11(1)、P11(2)で接続されている。ホスト70とストレージ装置20も、複数の通信パスP12(1)、P12(2)に接続されている。通常の場合、各通信パスP11(1)、P11(2)はアクティブパスであり、各通信パスP12(1)、P12(2)はパッシブパスである。パス制御部78は、複数のアクティブパスP11(1)、P11(2)のいずれも使用できない場合に、パッシブパスP12(1)、P12(2)に切り替える。パス制御部78は、ラウンドロビン方式に基づいて、2つのアクティブパスP11(1)、P11(2)を切り替えて使用する。同様に、パス制御部78は、2つのパッシブパスP12(1)、P12(2)を切り替えて使用する。

20

【0122】

ストレージ装置10内の論理ボリューム230(正ボリューム)とストレージ装置20内の論理ボリューム230(副ボリューム)とにより、一つの仮想ボリューム231が形成される。正ボリュームと副ボリュームとは、リモートコピーペアを形成する。

30

【0123】

通常の場合、ホスト70は、ストレージ装置10内の正ボリュームにアクセスする。ホスト70が正ボリュームに記憶されているデータを更新させると、更新データはストレージ装置10からストレージ装置20に送信され、ストレージ装置20内の副ボリュームに反映される。仮想ボリューム231を構成する各論理ボリューム230には、それぞれ同一の識別子が設定されているため、パス制御部78は、各論理ボリューム230の見分けがつかず、同一のデバイスとして認識する。

【0124】

図6は、ロックディスクを管理するためのテーブルT10を示す。そのロックディスク管理テーブルT10は、各ストレージ装置10、20内の共有メモリ140に記憶されている。

40

【0125】

ロックディスク管理テーブルT10は、例えば、ロックディスク識別子C11(以下、識別子をIDと略す場合がある)と、管理フラグC12と、ロックディスクのLDEV番号C13と、自装置の製造番号C14と、相手装置の製造番号C15と、コントロール識別子C16と、ロックディスク情報ビットマップC17とを備えている。

【0126】

ロックディスクID_C11とは、ロックディスク232をストレージシステム内で一意に識別するための情報である。管理フラグC12は、ロックディスク232の状態等を

50

管理するための情報である。管理フラグC 1 2には、例えば、有効／無効フラグC 1 2 1と、ロックディスク作成中状態フラグC 1 2 2と、ロックディスク削除中状態フラグC 1 2 3とが含まれる。

【0 1 2 7】

有効／無効フラグC 1 2 1は、そのロックディスク2 3 2が有効であるか無効であるかを示すためのフラグである。ロックディスク作成中状態フラグC 1 2 2は、ロックディスク2 3 2が作成されている最中であることを示すフラグである。ロックディスク2 3 2の作成をストレージ装置に指示してから、ロックディスクの作成完了が報告されるまでの間、ロックディスクの状態は「作成中」に設定される。

【0 1 2 8】

ロックディスク削除中状態フラグC 1 2 3は、ロックディスク2 3 2が削除されている最中であることを示すフラグである。ロックディスクの削除を指示してから、ロックディスクの削除完了が報告されるまでの間、そのロックディスクの状態は「削除中」に設定される。

【0 1 2 9】

LDEV番号C 1 3は、ロックディスク2 3 2として使用される論理ボリューム2 3 0の番号を示す。第3ストレージ装置3 0内の論理ボリューム2 3 0が、ロックディスク2 3 2として使用される。

【0 1 3 0】

自装置の製造番号C 1 4には、そのロックディスク管理テーブルT 1 0がストレージ装置1 0に記憶されている場合には、ストレージ装置1 0の製造番号が設定される。これに対し、ストレージ装置2 0内のロックディスク管理テーブルT 1 0には、自装置の製造番号C 1 4として、ストレージ装置2 0の製造番号が設定される。

【0 1 3 1】

相手装置の製造番号C 1 5には、そのロックディスク管理テーブルT 1 0がストレージ装置1 0に記憶されている場合には、ストレージ装置2 0の製造番号が設定される。これに対し、ストレージ装置2 0内のロックディスク管理テーブルT 1 0の場合、相手装置の製造番号C 1 5にはストレージ装置1 0の製造番号が設定される。

【0 1 3 2】

コントロールID__C 1 6には、ストレージ装置の世代を示す番号が設定される。ストレージシステム内に異なる世代のストレージ装置が混在する場合でも、各ストレージ装置を正確に識別するためにストレージ装置の世代の情報も管理する。コントロールIDと製造番号との組合せにより、各ストレージ装置を一意に特定することができる。

【0 1 3 3】

ロック情報ビットマップc 1 7には、そのロックディスク2 3 2に対応付けられている仮想ボリューム2 3 1のロック情報（つまり、その仮想ボリューム2 3 1を構成するリモートコピーペアに関するロック情報）が、ビットマップ形式で設定される。

【0 1 3 4】

図7は、ロックディスク情報ビットマップC 1 7を模式的に示す説明図である。ロック情報ビットマップC 1 7では、そのロックディスク2 3 2によって管理されている、一つまたは複数の仮想ボリューム（図7中「ペア」と表示）毎に（図7（b））、それぞれ1ビットを割り当てている（図7（a））。

【0 1 3 5】

図7（c）、（d）に示すように、仮想ボリューム2 3 1に関するリモートコピーペアを構成する各ボリューム（正ボリュームと副ボリューム）がペア状態である場合、そのペアに対応するビットには「0」が設定される。

【0 1 3 6】

これに対し、障害等によってペアが解除された場合、正ボリュームまたは副ボリュームのいずれか一方がホスト7 0により更新され、正ボリュームの記憶内容と副ボリュームの記憶内容とが一致しなくなる。そこで、リモートコピーペアが解除された場合、その仮想

10

20

30

40

50

ボリュームに対応するビットには「1」が設定される。

【0137】

正ボリュームを用いて仮想ボリューム231の運用を継続する場合、ストレージ装置10内のロックディスク管理テーブルT10において、ロック情報ビットマップC17の対応ビットに「1」が設定される。ストレージ装置20内のロックディスク管理テーブルT10では、ロック情報ビットマップC17内の対応ビットの値は「0」である。

【0138】

これとは逆に、もしも副ボリュームを用いて仮想ボリューム231の運用が継続される場合、ストレージ装置20内のロックディスク管理テーブルT10において、ロック情報ビットマップC17の対応ビットに「1」が設定される。ストレージ装置10内のロックディスク管理テーブルT10では、ロック情報ビットマップC17内の対応ビットの値は「1」である。

【0139】

つまり、ロック情報ビットマップC17は、仮想ボリューム231を構成する複数のボリュームのうち、いずれのボリュームを用いて仮想ボリューム231が運用されているかを示す。換言すれば、複数のストレージ装置10、20のうち、いずれのストレージ装置が仮想ボリューム231の運用を担当しているかを示す。

【0140】

図8は、ロックディスク232に記憶される使用制御情報L10の構成例を示す説明図である。使用制御情報L10は、例えば、管理情報L11と、第1ストレージ装置10の制御情報L12と、第2ストレージ装置20の制御情報L13と、第1ストレージ装置のロック情報ビットマップL14と、第2ストレージ装置20のロック情報ビットマップL15とを備えている。

【0141】

管理情報L11は、例えば、ロックディスクID__L111と、第1ストレージ装置10の製造番号L112と、第2ストレージ装置20の製造番号L113とを含む。ロックディスクID__L111は、上述の通り、ロックディスク232をストレージシステム内で一意に特定するための識別情報である。

【0142】

第1ストレージ装置10の制御情報L12は、第1ストレージ装置10がロックディスク232を使用しているか否かを示す情報である。使用中の場合には「1」が設定され、未使用の場合には「0」が設定される。同様に、第2ストレージ装置20の制御情報L13は、第2ストレージ装置20がロックディスク232を使用しているか否かを示す情報である。

【0143】

第1ストレージ装置10のロック情報ビットマップL14及び第2ストレージ装置20のロック情報ビットマップL15は、上述の通り、そのロックディスク232で管理されている仮想ボリューム231をどちらのストレージ装置が使用しているか否か、換言すれば、正副の両ボリュームのうちいずれの論理ボリュームに差分データが蓄積されているかを示す情報である。

【0144】

ここで、第1ストレージ装置10は、ロックディスク232にアクセスすることにより、第1ストレージ装置10の製造番号L112と、第1ストレージ装置10の制御情報L12と、第1ストレージ装置10のロック情報ビットマップL14とに、それぞれ値を書き込むことができる。第1ストレージ装置10は、第2ストレージ装置20の製造番号L113と、第2ストレージ装置20の制御情報L13と、第2ストレージ装置20のロック情報ビットマップL15とを書き換えることはできない。

【0145】

同様に、第2ストレージ装置20も、自分自身に関する項目L113、L13、L15のみ更新できる。なお、ロックディスクID__L111は、ロックディスク232を作成

10

20

30

40

50

した方のストレージ装置が書き込む。

【0146】

図9は、ペア管理テーブルT20を示す説明図である。ペア管理テーブルT20は、仮想ボリューム231を構成するリモートコピーペアを管理する。ペア管理テーブルT20は、例えば、正ボリューム（図中、「PVOL」）に関する項目C21と、副ボリューム（図中、「SVOL」）に関する項目C22と、ロックディスクID_C23とを備えている。

【0147】

正ボリュームに関する項目C21は、例えば、正ボリュームの存在するストレージ装置の製造番号C211と、正ボリュームとして使用されている論理ボリュームのLDEV番号C212と、ペア状態C213とを含む。

10

【0148】

同様に、副ボリュームに関する項目C22は、副ボリュームの存在するストレージ装置の製造番号C221と、副ボリュームとして使用されている論理ボリュームのLDEV番号C222と、ペア状態C223とを含む。

【0149】

ペア状態としては、例えば、「ペア」、「SMP L（シンプレックス）」、「PSUS（サスペンド：PVOLの単独運用）」、「SSWS（スワップサスペンド：SVOLの単独運用）」、「ペアリシンク」、「リバースリシンク」等がある。

【0150】

20

「ペア」とは、正ボリュームと副ボリュームとがリモートコピーペアを形成しており、正ボリュームの記憶内容と副ボリュームの記憶内容とが一致している状態である。「SMP L」とは、通常の論理ボリュームであることを示す状態である。「PSUS」とは、正ボリュームがサスペンド状態であり、正ボリューム単独で仮想ボリューム231を運用している状態を示す。「SSWS」とは、副ボリュームに切り替えられて、副ボリューム単独で仮想ボリューム231を運用している状態を示す。「ペアリシンク」とは、正ボリュームの記憶内容と副ボリュームの記憶内容とを再同期させている状態を示す。「リバースリシンク」とは、副ボリュームに蓄積された差分を正ボリュームに書き込んで、正ボリュームと副ボリュームとを同期させる状態を示す。

【0151】

30

図10は、各ストレージ装置が論理ボリュームを管理するためのテーブルT30を示す説明図である。各ストレージ装置10、20の共有メモリ140には、LDEV管理テーブルT30がそれぞれ記憶されている。

【0152】

LDEV管理テーブルT30は、例えば、LDEV番号C31と、ボリューム種別C32と、VDEV番号C33と、開始アドレスC34と、サイズC35とを含む。LDEV番号C31とは、各ストレージ装置内で論理ボリューム230を管理するための識別情報である。

【0153】

ボリューム種別C32とは、内部ボリュームとして構成されているか外部ボリュームを利用して構成されているかの区別を示す。内部ボリュームとして構成されるボリュームとは、ストレージ装置内の物理的記憶領域を利用している実ボリュームである。外部ボリュームを利用して構成されるボリュームとは、外部ストレージ装置30内のボリューム（外部ボリューム）を利用しているボリューム（外部接続ボリューム）である。

40

【0154】

VDEV番号C33とは、そのボリュームの属するVDEVを特定する情報である。開始アドレスC34とは、VDEVの物理的記憶領域のどの部分からボリュームが開始するかを示す。サイズC35とは、ボリュームの記憶容量である。

【0155】

図11は、外部ボリュームを管理するためのテーブルT40を示す。この外部ボリュー

50

ム管理テーブルT 4 0 は、各ストレージ装置 1 0 , 2 0 の共有メモリ 1 4 0 にそれぞれ記憶されている。

【0 1 5 6】

外部ボリューム管理テーブルT 4 0 は、例えば、V D E V 番号C 4 1 と、接続ポートC 4 2 と、外部ストレージ情報C 4 3 とを含む。V D E V 番号C 4 1 とは、V D E V を特定する情報である。接続ポートC 4 2 とは、外部ストレージ装置が接続されている通信ポート 1 1 1 B を特定する情報である。

【0 1 5 7】

外部ストレージ情報C 4 3 は、外部ストレージ装置 3 0 の構成を示す。外部ストレージ情報C 4 3 は、例えば、L U N _ C 4 4 と、ベンダ名C 4 5 と、装置名C 4 6 と、ボリューム識別子C 4 7 とを含む。L U N _ C 4 4 とは、外部ボリュームに対応付けられているL U N を示す。ベンダ名C 4 5 とは、外部ストレージ装置の提供者名を示す。装置名C 4 6 とは、外部ストレージ装置を特定するための番号（製造番号）を示す。ボリューム識別子C 4 7 とは、外部ストレージ装置 3 0 内の外部ボリュームを外部ストレージ装置 3 0 が識別するための、識別子である。

10

【0 1 5 8】

図 1 2 は、ロックディスク管理画面G 1 0 を示す説明図である。管理サーバ 8 0 は、S V P 1 6 0 にアクセスして、図 1 2 に示すような設定画面を管理サーバ 8 0 の表示装置に表示させることができる。

【0 1 5 9】

20

ロックディスク管理画面G 1 0 は、例えば、ストレージシステムのツリー構成を示すツリー表示部G 1 1 と、L D E V に関する情報を表示するL D E V 情報表示部G 1 2 と、プレビュー表示部G 1 3 とを含む。

【0 1 6 0】

ツリー表示部G 1 1 は、ストレージシステムの構成を、例えば、ストレージ装置単位（D K C 単位）、ストレージ装置内に仮想的に設けられる仮想ストレージ装置単位（L D K C 単位）、使用中のロックディスク単位、未使用のロックディスク単位のようなレベルで、表示させる。

【0 1 6 1】

L D E V 情報表示部G 1 2 は、例えば、ロックディスク I D を表示するロックディスク I D 表示部G 1 2 1 と、ロックディスクとして使用されるL D E V（論理ボリューム 2 3 0）を特定するためのL D E V 特定情報を表示するL D E V 特定部G 1 2 2 と、仮想ボリューム 2 3 1 を構成するための、他方のボリュームを有する装置（つまり、相手装置）の製造番号を表示する製造番号表示部G 1 2 3 と、その相手方のストレージ装置の世代を示すコントロール I D を表示するコントロール I D 表示部G 1 2 4 と、を備えている。

30

【0 1 6 2】

L D E V 情報表示部G 1 2 において、例えば、マウスポインタを所望の行に合わせて右クリックすると、いわゆるコンテキストメニューM 1 0 が出現する。そのメニューM 1 0 には、例えば、「ロックディスクの作成」と「ロックディスクの削除」という項目が含まれる。ユーザは、メニューM 1 0 を使用することにより、ロックディスク 2 3 2 の作成または削除を行うことができる。

40

【0 1 6 3】

プレビュー表示部G 1 3 には、ユーザがL D E V 情報表示部G 1 2 で設定した値が表示される。その設定値の使用をユーザが希望する場合、ユーザは、「A p p l y」ボタンB 1 1 を操作する。これにより、後述のロックディスク作成処理またはロックディスク削除処理が実行される。

【0 1 6 4】

図 1 3 , 図 1 4 は、ロックディスク作成処理を示すフローチャートである。以下に述べる各フローチャートは、いわゆる当業者が理解及び実施可能な程度に、各処理の概要をそれぞれ示しており、実際のコンピュータプログラムとは相違する場合がある。いわゆる当

50

業者であれば、図示されたステップの変更または削除、新たなステップの追加を行うことができるであろう。なお、以下の説明では、第1ストレージ装置10内のSV P160を第1SV Pと呼び、第2ストレージ装置20内のSV P160を第2SV Pと呼ぶ。

【0165】

図13は、第1ストレージ装置10によるロックディスク作成処理を示し、図14は、第2ストレージ装置20によるロックディスク作成処理を示す。両方のロックディスク作成処理は同一のため、第1ストレージ装置10によるロックディスク作成処理を中心に説明する。

【0166】

ユーザは、管理サーバ80を介して第1SV Pにアクセスし、そして図12で述べたロックディスク管理画面G10を用いて、第1ストレージ装置10に、ロックディスクの作成を指示する(S10)。

10

【0167】

そのロックディスク作成指示には、ロックディスクID(G121)と、LDEV特定情報(G122)と、相手方のストレージ装置の製造番号(G123)及びコントロールID(G124)が含まれている。

【0168】

第1ストレージ装置10は、第1ストレージ装置10内の共有メモリ140に記憶されているロックディスク管理テーブルT10を参照し、第1SV Pから指定されたロックディスクIDが、使用中では無いことを確認する。

20

【0169】

そして、第1ストレージ装置10は、第3ストレージ装置30にリードコマンドを発行し、ロックディスク232内の使用制御情報L10を読み出す(S11)。第3ストレージ装置30は、要求された使用制御情報L10を第1ストレージ装置10に送信する(S12)。

【0170】

第1ストレージ装置10は、使用制御情報L10に基づいて、S10で指定されたロックディスクIDが他のストレージ装置(不図示)で使用されていないことを確認する。第1ストレージ装置10は、S10で指定されたロックディスクIDが未使用の場合、あるいは、そのロックディスクIDは使用中であるが、指定されたロックディスクIDと自装置の製造番号及び相手装置の製造番号とが3つとも一致する場合、のいずれかの場合に、使用制御情報L10を更新するためのライトデータを生成する(S13)。

30

【0171】

ライトデータは、例えば、以下のように作成される。第1SV Pから指定されたロックディスクIDが未使用の場合、第1ストレージ装置10は、その指定されたロックディスクIDをロックディスクID__L111として、使用する。

【0172】

指定されたロックディスクIDは既に使用中であるが、指定されたロックディスクIDと、そのロックディスクIDに対応する第1ストレージ装置10の製造番号L112と第2ストレージ装置20の製造番号L113とが、管理情報L11の値に一致する場合、第1ストレージ装置10は、管理情報L11内のロックディスクID__L111の値をそのまま使用する。

40

【0173】

さらに、第1ストレージ装置10は、第1ストレージ装置の制御情報L12に使用中であることを示すフラグ(=1)を設定する。さらに、第1ストレージ装置10は、第1ストレージ装置のロック情報ビットマップL14を0で初期化する。

【0174】

第1ストレージ装置10は、上記のように作成されたライトデータを、ロックディスク232に書き込む(S14)。第3ストレージ装置30は、書込みが完了した旨を第1ストレージ装置10に通知する(S15)。

50

【0175】

ここで、ライトデータをロックディスク232に書き込んでいる最中に、図外の他のストレージ装置から第3ストレージ装置30に別のライトコマンドが発行されて、ロックディスク232の内容が書き換えられてしまう可能性がある。そこで、本実施例では、第1ストレージ装置10は、第3ストレージ装置30にリードコマンドを発行し、ロックディスク232の使用制御情報L10を再び読み出す(S16)。第3ストレージ装置30は、リードコマンドに応じて、使用制御情報L10を第1ストレージ装置10に転送する(S17)。

【0176】

第1ストレージ装置10は、ロックディスク232から取得した使用制御情報L10に基づいて、S14の書込み処理(更新処理)が正常に完了していることを確認する。もしも、S16、S17で再度取得した使用制御情報L10とS14、S15で更新した使用制御情報L10とが不一致の場合、第1ストレージ装置10は、S14以下の処理を再度実行する。

10

【0177】

使用制御情報L10が正しく更新されていた場合、第1ストレージ装置10は、その使用制御情報L10に基づいて、共有メモリ140内のロックディスク管理テーブルT10を作成(更新)する(S18)。

【0178】

第1ストレージ装置10は、ロックディスク管理テーブルT10における、管理フラグC12と、LDEV番号C13と、自装置製造番号C14と、相手装置製造番号C15と、コントロールID__C16と、ロック情報ビットマップC17の、各値をそれぞれ更新させる(S18)。

20

【0179】

管理サーバ80は、第1SVPを介して第1ストレージ装置10に、ロックディスクの作成が完了したか否かを定期的に問い合わせている。ロックディスクの作成が完了したことを確認した場合、管理サーバ80は、ロックディスクの作成が完了したことをコンピュータ画面に表示して、ユーザに通知する。

【0180】

図14は、第2ストレージ装置20によるロックディスク作成処理を示すフローチャートである。この処理は、図13で述べた処理と同様のステップを備えている。図14中のS20～S28は、図13中のS10～S18に対応する。従って、重複した説明を割愛する。

30

【0181】

図15は、ロックディスクを作成する場合のロックディスク管理画面G10の例を示す説明図である。例えば、ユーザは、ロックディスクID(G121)として「00」を選択し、「00:40:00」で特定される論理ボリュームをロックディスク232として選択し、仮想ボリューム231に関する相手装置の製造番号として「64016」を選択し、コントロールIDとして「6」を選択する。

【0182】

図16は、ロックディスクの作成後におけるロックディスク管理テーブルT10を示す説明図である。例えば、ロックディスクID__C11には「0x00(=00)」が設定され、有効/無効フラグC12には「有効」が設定され、LDEV番号C13には「0x0040(=00:40:00)」が設定され、自装置製造番号C14には「64036」が設定され、相手装置製造番号には「64016」が設定され、コントロールID__C16には「0x0006(=6)」が設定される。ロック情報ビットマップC17には、0の列が設定される。

40

【0183】

図17は、リモートコピーを管理するための画面G20を示す説明図である。リモートコピー管理画面G20は、例えば、ツリー表示部G21と、LDEV情報表示部G22と

50

、プレビュー表示部G 2 3とを備える。

【0184】

ツリー表示部G 2 1は、例えば、ストレージ装置全体、ストレージ装置内の仮想ストレージ装置毎に、あるいは、ポート毎に、L D E V情報を表示させる。

【0185】

L D E V情報表示部G 2 2は、例えば、L D E V（論理ボリューム）を特定するためのL D E V特定部G 2 2 1と、そのL D E VのステータスG 2 2 2と、相手装置の製造番号C 2 2 3と、コントロールI D _ G 2 2 4と、ロックディスクI D _ G 2 2 5とを備えている。

【0186】

プレビュー表示部G 2 3は、例えば、L D E V特定部G 2 3 1と、ステータスG 2 3 2と、相手装置の製造番号C 2 3 3と、コントロールI D _ G 2 3 4と、ロックディスクI D _ G 2 2 5とを備える。

【0187】

ユーザが、例えば、L D E V情報表示部G 2 2でマウスを右クリックすると、コンテキストメニューM 2 0が表示される。図18は、メニューM 2 0の構成例を模式的に示す説明図である。

【0188】

このメニューM 2 0は、例えば、ペア作成M 2 1、ペア削除M 2 2、サスペンドM 2 3、スワップサスペンドM 2 4、リシンクM 2 5及びリバースリシンクM 2 6という複数のサブメニューを備える。

【0189】

ペア作成M 2 1は、仮想ボリューム231を構成するリモートコピーペアを作成するためのサブメニューである。ペア削除M 2 2は、仮想ボリューム231を構成するリモートコピーペアを削除するためのサブメニューである。サスペンドM 2 3は、リモートコピーペアをサスペンド状態にするためのサブメニューである。スワップサスペンドM 2 4は、リモートコピーペアをサスペンド状態とし、副ボリュームを用いて仮想ボリューム231の運用を継続させるためのサブメニューである。つまり、スワップサスペンドは、正ボリュームから副ボリュームへのフェイルオーバーを指示する。リシンクM 2 5は、正ボリュームに生じた差分を副ボリュームに転送し、両方のボリューム内容を同期させるためのサブメニューである。リバースリシンクM 2 6は、副ボリュームに生じた差分を正ボリュームに転送し、両方のボリューム内容を同期させるためのサブメニューである。

【0190】

ユーザは、シンプレックス状態の論理ボリュームを2つ選択して、ペア作成M 2 1を指定することにより、仮想ボリューム231を構成するリモートコピーペアを作成することができる。さらに、ユーザは、リモートコピーペアを形成する正ボリュームまたは副ボリュームのいずれかを選択して、ペア削除M 2 2を指定することにより、そのリモートコピーペアを削除させることができる。

【0191】

ユーザは、プレビュー表示部G 2 3に表示された内容を承認する場合、「A p p l y」ボタンB 2 1を操作する。取り消す場合は、「C a n c e l」ボタンを操作する。

【0192】

図19は、ペア作成M 2 1を操作した場合に管理サーバ80のコンピュータ画面に表示される、ペア作成画面G 3 0を示す。ペア作成画面G 3 0は、例えば、正ボリューム設定部G 3 1 A、G 3 1 Bと、副ボリューム設定部G 3 2 A、G 3 2 Bと、ストレージ装置間のパス設定部G 3 3 A、G 3 3 Bと、正ボリュームのフェンスレベル設定部G 3 4 A、G 3 4 Bと、ロックディスクI D設定部G 3 5 A、G 3 5 Bとを備える。

【0193】

正ボリューム設定部G 3 1 A、G 3 1 Bでは、正ボリュームとして使用する論理ボリュームを特定するための情報と、その論理ボリュームに接続されている通信ポートを特定す

10

20

30

40

50

るための情報とを、設定する。

【0194】

副ボリューム設定部G32A、G32Bでは、副ボリュームとして使用する論理ボリュームを特定するための情報と、その論理ボリュームに接続されている通信ポートを特定するための情報とを、設定する。

【0195】

ストレージ装置間のパス設定部G33A、G33Bでは、正ボリュームを有するストレージ装置と副ボリュームを有するストレージ装置との間でリモートコピーを行うための通信経路CN20を設定する。

【0196】

正ボリュームのフェンスレベル設定部G34A、G34Bでは、フェンスレベルを設定する。フェンスレベルの値としては、「データ」と「Never」とがある。フェンスレベルの値に「データ」が設定されると、障害発生時に、正ボリュームの記憶内容と副ボリュームの記憶内容とが同期していることが保証される。つまり、障害発生時に、仮想ボリューム231へのデータ更新は停止される。これに対し、フェンスレベルの値に「Never」が設定されると、障害発生時でも、正ボリュームまたは副ボリュームのいずれかを用いて、仮想ボリューム231のデータ更新が行われる。仮想ボリューム231を構成するリモートコピーペアにおいて、フェンスレベルの値を「データ」とすると、障害発生時の業務継続という仮想ボリューム231の特徴が失われる。そこで、本実施例では、フェンスレベルの初期値を「Never」に設定する。

【0197】

ロックディスクID設定部G35A、G35Bでは、仮想ボリューム231の使用を管理するためのロックディスク232のIDを設定する。各設定値を承認する場合、ユーザは「Set」ボタンB31を操作する。取り消す場合、ユーザは「Cancel」ボタンB32を操作する。

【0198】

図20は、リモートコピーペアを設定するための処理を示すフローチャートである。管理サーバ80は、第1SVPを介して第1ストレージ装置10に、リモートコピーペアに基づく仮想ボリューム231の作成を指示する(S30)。

【0199】

その作成指示には、G30に含まれる各値(G31B-G35B)が含まれており、第1ストレージ装置10は、それらの各値に基づいてペア管理テーブルT20を作成する(G31)。

【0200】

第1ストレージ装置10は、装置間通信経路CN20を介して、ペア管理テーブルT20の内容を第2ストレージ装置20に転送する(S32)。第2ストレージ装置20は、第1ストレージ装置10から受け取った情報を、第2ストレージ装置20内のペア管理テーブルT20に登録する(S33)。

【0201】

第2ストレージ装置20は、ロックディスク管理テーブルT10を参照して、第3ストレージ装置30内のロックディスク232を更新する(S34)。第3ストレージ装置30は、第2ストレージ装置20からの要求に基づいてロックディスク232内の使用制御情報L10を更新し(S35)、更新が完了した旨を第2ストレージ装置20に報告する(S36)。

【0202】

その更新処理では、上述のように、第2ストレージ装置20は、更新直後の使用制御情報L10をロックディスク232から読み出して検査することにより、更新が正常に完了したか否かを確認する。そして、使用制御情報L10の更新が完了すると、第2ストレージ装置20は、第1ストレージ装置10にロックディスク232の更新が完了した旨を通知する(S37)。

【0203】

ここで、第2ストレージ装置20は、使用制御情報L10のうち第2ストレージ装置20に関する項目L113、L13、L15のみを更新可能であり、第1ストレージ装置10に関する項目L112、L12、L14を更新させることはできない（ロックディスクID__L111は、第2ストレージ装置20が設定可能である。）。

【0204】

そこで、次に、第1ストレージ装置10は、使用制御情報L10の未設定項目を設定する（S38）。第3ストレージ装置30は、第1ストレージ装置10からの要求に基づいて、ロックディスク232内の使用制御情報L10を更新し（S39）、更新が完了した旨を第1ストレージ装置10に通知する（S40）。 10

【0205】

第1ストレージ装置10は、使用制御情報L10が作成されたことを確認すると、第1SVPを介して管理サーバ80に、リモートコピーペアに基づく仮想ボリューム231が作成された旨を報告する（S41）。

【0206】

その後、別のタイミングで、リモートコピーペアの初期コピー（形成コピー）が実行される（S42～S44）。第1ストレージ装置10は、第2ストレージ装置20に形成コピーの開始を通知し（S42）、正ボリュームの記憶内容を副ボリュームに転送させる（S43）。第2ストレージ装置20は、正ボリュームの記憶内容を副ボリュームに書込み、書込み完了を第1ストレージ装置10に報告する（S44）。形成コピーが完了することにより、正ボリュームの記憶内容と副ボリュームの記憶内容とが同期する。 20

【0207】

図21は、仮想ボリューム231を構成するリモートコピーペアが作成された後の、リモートコピー管理画面G20を示す。図22は、仮想ボリューム231を構成するリモートコピーペアが作成された後の、ペア管理テーブルT20を示す。リモートコピーペアに関するボリュームのステータスは、「シンプレックス」から「ペア」に変化する。

【0208】

図23～図26は、一つのロックディスク232に複数の仮想ボリューム231を関連づける場合を示している。まず、図23のロックディスク管理テーブルT10に示すように、例えば、ロックディスクID「00」、「01」の2つのロックディスク232を作成する。 30

【0209】

図24のリモートコピー管理画面G20に示すように、一方のロックディスク「00」には、正ボリューム「00：01：00」と、正ボリューム「00：01：01」との2つのリモートコピーペアが関連づけられる。他方のロックディスク「01」には、正ボリューム「00：01：0A」のリモートコピーペアが関連づけられる。

【0210】

図25に示すロックディスク管理テーブルT10には、複数のロックディスク「00」、「01」が登録される。図26に示すペア管理テーブルT20では、一方のロックディスク「00」に2つのリモートコピーペアが関連づけられ、他方のロックディスク「01」に1つのリモートコピーペアが関連づけられる。 40

【0211】

このように本実施例では、1つのロックディスク232に、リモートコピーペアに基づく仮想ボリュームを複数対応付けて管理することができる。

【0212】

図27は、ロックディスク232の使用制御情報L10を更新するための処理を示すフローチャートである。更新の契機としては、例えば、ロックディスクを作成する場合、ロックディスクを削除する場合、リモートコピーペア（仮想ボリューム。以下同様）を設定する場合、リモートコピーペアを削除する場合、仮想ボリュームにサスペンドを指示する場合、仮想ボリュームにリシンクを指示する場合、仮想ボリュームにスワップサスペンド 50

を指示する場合、仮想ボリュームにリバーシリンクを指示する場合、を挙げることができる。

【0213】

上記の更新契機が発生すると、管理サーバ80から第1ストレージ装置10に、更新契機に対応する所定の指示が入力される(S50)。第1ストレージ装置10は、ロックディスク232から読み出した使用制御情報L10がキャッシュメモリ130に残っているか否かを確認する。使用制御情報L10がキャッシュメモリ130に記憶されている場合、第1ストレージ装置10は、使用制御情報L10を破棄する。キャッシュメモリ上の使用制御情報L10は、古い情報となっている可能性があるためである。

【0214】

そして、第1ストレージ装置10は、第3ストレージ装置30に、最新の使用制御情報L10を要求する(S51)。第3ストレージ装置30は、ロックディスク232から読み出した使用制御情報L10を第1ストレージ装置10に転送する(S52)。

【0215】

第1ストレージ装置10は、前記更新契機に応じたライトデータ(使用制御情報L10を更新するためのデータ)を作成し(S53)、そのライトデータを第3ストレージ装置30に転送する(S54)。これにより、第3ストレージ装置30は、ロックディスク232内の使用制御情報L10を更新し、更新が完了した旨を第1ストレージ装置10に報告する(S55)。

【0216】

第1ストレージ装置10は、更新が正常に完了したことを確かめるべく、第3ストレージ装置30に使用制御情報L10の転送を再び要求する(S56)。第3ストレージ装置30は、ロックディスク232から読み出した使用制御情報L10を第1ストレージ装置10に転送する(S57)。

【0217】

第1ストレージ装置10は、使用制御情報L10が正しく更新されたことを確認すると、ロックディスク管理テーブルT10を更新する(S58)。なお、上述の通り、第1ストレージ装置10は、使用制御情報L10のうち第1ストレージ装置10に関する項目のみを更新可能である。従って、第2ストレージ装置20が図27に示す処理を実施することにより、使用制御情報L10の全体が適切に更新される。

【0218】

図28は、ホスト70が正ボリュームからデータを読み出す場合の処理を示すフローチャートである。ホスト70は、アクティブパスを用いて、第1ストレージ装置10にリードコマンドを発行する(S60)。

【0219】

第1ストレージ装置10は、仮想ボリューム231を構成する正ボリュームから、要求されたデータを読み出し(S61)、そのデータをホスト70に送信する(S62)。そして、第1ストレージ装置10は、リードコマンドの処理が完了した旨をホスト70に通知する(S63)。

【0220】

図29は、ホスト70が副ボリュームからデータを読み出す場合の処理を示すフローチャートである。まず最初に、ホスト70は、アクティブパスを用いて、第1ストレージ装置10にリードコマンドを発行する(S70)。

【0221】

もしも、ホスト70と第1ストレージ装置10とを接続するアクティブパスに障害が発生していたり、あるいは、第1ストレージ装置10が停止していた場合、第1ストレージ装置10は、そのリードコマンドを処理することができない(S71)。ホスト70のパス制御部78は、第1ストレージ装置10からのエラー応答によって、あるいは、所定時間内に応答が無かった事実によって、第1ストレージ装置10がリードコマンドを処理できないことを検出する(S72)。

10

20

30

40

50

【0222】

そこで、ホスト70のバス制御部78は、アクティブパスからパッシブパスに切り替えて（S73）、第2ストレージ装置20にリードコマンドを発行する（S74）。

【0223】

第2ストレージ装置20は、ホスト70からのリードコマンドを受信すると、第3ストレージ装置30に、ロックディスク232内の使用制御情報L10の転送を要求する（S75）。第3ストレージ装置30は、第2ストレージ装置20からの要求に応じて、ロックディスク232から読み出した使用制御情報L10を、第2ストレージ装置10に転送する（S76）。

【0224】

第2ストレージ装置20は、使用制御情報L10内の第1ストレージ装置のロック情報ビットマップL14を参照し、仮想ボリューム231に対応するビットの値が「1」であるか「0」であるかを判定する（S77）。

【0225】

仮想ボリューム231に対応するビットの値が「0」の場合は、正ボリュームと副ボリュームとは同期しているため、ホスト70から要求されたデータを副ボリュームから読み出してホスト70に送信する（S78）。そして、第2ストレージ装置20は、リードコマンドの処理が完了した旨をホスト70に通知する（S79）。

【0226】

これに対し、ロック情報ビットマップL14において、リードコマンドに対応する仮想ボリューム231に関連づけられているビットの値が「1」である場合は、正ボリュームと副ボリュームとは同期しておらず、正ボリュームに最新データが記憶されている場合である。つまり、副ボリュームに記憶されているデータは古い可能性がある。そこで、ホスト70が古いデータを誤って読み出したりしないように、第2ストレージ装置20は、チェック応答を返す（S80）。

【0227】

図30は、ホスト70が正ボリュームにデータを書き込む場合の処理を示すフローチャートである。ホスト70は、第1ストレージ装置10にライトコマンドを発行する（S90）。第1ストレージ装置10は、ライトコマンドを受領すると、ライトデータを記憶するための領域をキャッシュメモリ上に確保し、ホスト70にライトデータの受信準備が完了した旨を通知する。その通知を受信したホスト70は、アクティブパスを用いて、第1ストレージ装置10にライトデータを送信する（S90）。そのライトデータは、第1ストレージ装置10内のキャッシュメモリ130に記憶される。

【0228】

第1ストレージ装置10は、自身が正ボリュームを有する正ストレージ装置であることを確認する（S92）。そして、第1ストレージ装置10は、装置間通信経路CN20を介して、副ボリュームを有する第2ストレージ装置20にライトコマンドを発行する（S93）。

【0229】

第2ストレージ装置20は、ライトデータの受信準備が完了すると、第1ストレージ装置10にライトデータの転送を要求する。その要求を受信した第1ストレージ装置10は、S91で受信したライトデータを第2ストレージ装置20に送信する（S94）。第2ストレージ装置20は、第1ストレージ装置10から受信したライトデータを、第2ストレージ装置20内のキャッシュメモリ130に記憶し、第1ストレージ装置10に処理完了を報告する（S95）。

【0230】

第1ストレージ装置10は、ホスト70からのライトデータが副ボリュームに書き込まれたことを確認すると、ホスト70にS90で受信したライトコマンドの処理が完了した旨をホスト70に報告する（S96）。

【0231】

10

20

30

40

50

なお、所定のタイミングを見計らって、キャッシュメモリ130に記憶されたライトデータは、対応するディスクドライブ210に書き込まれる。キャッシュメモリ上のデータをディスクドライブに書き込んで保存する処理をデステージ処理と呼ぶ。デステージ処理は、ライトデータの受信直後に行ってもよいし（同期式）、ライトデータの受信とは別のタイミングで行ってもよい（非同期式）。

【0232】

図31は、ホスト70が副ボリュームにデータを書き込む場合の処理を示すフローチャートである。まず最初に、ホスト70は、正ボリュームを有する第1ストレージ装置10にライトコマンドを発行する（S100）。

【0233】

しかし、アクティブパスに障害が発生していたり、あるいは、第1ストレージ装置10が停止している場合、第1ストレージ装置10はライトコマンドを処理することができない（S101）。この場合、ホスト70は、第1ストレージ装置10からのエラー応答またはタイムアウトエラーによって、障害発生を検出する（S102）。パス制御部78は、アクティブパスからパッシブパスに切り替える（S103）。

【0234】

ホスト70は、パッシブパスを用いて、第2ストレージ装置20にライトコマンドを発行する（S104）。第2ストレージ装置20は、ライトデータの受け入れ準備が完了すると、ホスト70に通知する。その通知を受信したホスト70は、ライトデータを第2ストレージ装置20に送信する。第2ストレージ装置20は、ホスト70から受信したライトデータをキャッシュメモリ130に格納する。

【0235】

第2ストレージ装置20は、ロックディスク232にアクセスして、使用制御情報L10を更新する（S105）。第2ストレージ装置20は、使用制御情報L10内の第2ストレージ装置の制御情報を「1」に設定する。これにより、第2ストレージ装置20がロックディスク232を使用中であることが、設定される。さらに、第2ストレージ装置20は、第2ストレージ装置のロック情報ビットマップL15中、ライトデータが書き込まれた仮想ボリューム231に対応するビットを「1」に設定する。これにより、副ボリュームの記憶内容が最新であると設定される。

【0236】

S105に示すロックディスク更新処理（使用制御情報L10の更新処理）では、上述の通り、更新直後に使用制御情報L10を読み出すことによって、使用制御情報L10が正しく更新されたことを確認する。「ロックディスク更新」を行う各ステップは、同様の確認作業を行うため、以下は説明を省略する。

【0237】

第2ストレージ装置20は、第1ストレージ装置10にペア状態の変更を指示する（S106）。正ボリュームの状態は「ペア」から「サスペンド（PSUS）」に変更され、副ボリュームの状態は「ペア」から「スワップサスペンド（SSWS）」に変更される（S106）。

【0238】

第1ストレージ装置10は、ペア状態の変更を終えると、第2ストレージ装置20に処理完了を通知する（S107）。その通知を受信した第2ストレージ装置20は、ホスト70にライトコマンドの処理が完了した旨を通知する（S108）。

【0239】

図31では、副ボリュームへの書込みが成功した場合を示す。次に、副ボリュームへの書込みが失敗する場合を、図32のフローチャートを参照して説明する。図32の処理では、正ボリュームが単独で運用されているものとする。まず最初に、正ボリュームへの書込みが正常に行われる（S120-S124）。

【0240】

ホスト70は、正ボリュームを有する第1ストレージ装置10にライトコマンドを送信

10

20

30

40

50

し（S120）、受信準備を確認した後でライトデータを送信する（S121）。第1ストレージ装置10は、正ボリュームが単独で運用されていることを確認し（S122）、ロックディスク232内の使用制御情報L10を更新する（S123）。ここでは、例えば、第1ストレージ装置のロック情報ビットマップL14内の、S120のライトコマンドに対応する仮想ボリューム231に関連するビットの値が「1」に設定される。ロックディスクの更新が完了すると、第1ストレージ装置10は、ホスト70にライトコマンドの処理が完了した旨を通知する（S124）。

【0241】

別のタイミングで、ホスト70は、別のライトコマンドを第1ストレージ装置10に発行する（S130）。S124とS130との間で、アクティブパスに障害が発生したり、あるいは、第1ストレージ装置10が作動を停止したとする。

10

【0242】

その場合、第1ストレージ装置10は、ライトコマンドを処理することができない（S131）。ホスト70は、エラー応答等により、第1ストレージ装置10を使用できないことを知る（S132）。そこで、パス制御部78は、アクティブパスからパッシブパスに切り替える（S133）。

【0243】

ホスト70は、副ボリュームを有する第2ストレージ装置20にライトコマンドを発行する（S134）。第2ストレージ装置20は、副ボリュームの状態が「スワップサスペンド（SSWS）」以外の場合、ロックディスク232内の使用制御情報L10の更新処理を試みる（S135）。

20

【0244】

第2ストレージ装置20は、使用制御情報L10に記憶されている第1ストレージ装置のロック情報ビットマップL14によって、第1ストレージ装置10がロックディスクを使用する権利（ロック権）を有していることを知る（S136）。その場合、正ボリュームの記憶内容の方が副ボリュームの記憶内容よりも新しいため、副ボリュームを用いてホスト70からの要求に応答することはできない。従って、第2ストレージ装置20は、ホスト70にチェック応答を送信する（S137）。

【0245】

図33は、仮想ボリューム231を構成するリモートコピーペアを削除する処理を示すフローチャートである。管理サーバ80は、第1SVPを介して第1ストレージ装置10に、仮想ボリュームを構成するリモートコピーペアの削除を指示する（S140）。

30

【0246】

第1ストレージ装置10は、ペア管理テーブルT20を参照し、削除を指示されたりリモートコピーペアが存在しているか否か、及び、削除を指示されたりリモートコピーペアを削除可能であるか否かを確認する。

【0247】

例えば、指示されたりリモートコピーペアに基づく仮想ボリューム231がホスト70によって使用されている場合、そのリモートコピーペアを削除することはできない。なお、指示されたりリモートコピーペアが存在しない場合、または、指示されたりリモートコピーペアを削除できない場合のいずれかの場合は、本処理は中断される。

40

【0248】

指定されたりリモートコピーペアが存在し、かつ、削除可能である場合、第1ストレージ装置10は、リモートコピーペアの削除指示を第2ストレージ装置20に転送する（S141）。その指示を受信した第2ストレージ装置20は、ロックディスク232内の使用制御情報L10を更新する（S142）。第2ストレージ装置は、第2ストレージ装置のロック情報ビットマップL15内の、削除指定されたりリモートコピーペア（仮想ボリューム）に対応するビットを「0」に設定する。

【0249】

さらに、第2ストレージ装置20は、副ボリュームの状態を「ペア」から「シンプレッ

50

クス」に変更し（S 1 4 3）、リモートコピーペアに関する情報をペア管理テーブル T 2 0 から削除する（S 1 4 4）。そして、第 2 ストレージ装置 2 0 は、第 1 ストレージ装置 1 0 にリモートコピーペアの削除が完了した旨を通知する（S 1 4 5）。

【0 2 5 0】

その通知を受信した第 1 ストレージ装置 1 0 は、第 3 ストレージ装置 3 0 のロックディスク 2 3 2 にアクセスし、使用制御情報 L 1 0 を更新する（S 1 4 6）。第 1 ストレージ装置 1 0 は、第 1 ストレージ装置のロック情報ビットマップ L 1 4 内の、削除指定されたリモートコピーペアに対応するビットを「0」に設定する。

【0 2 5 1】

さらに、第 1 ストレージ装置 1 0 は、正ボリュームの状態を「ペア」から「シンプレックス」に変更し（S 1 4 7）、第 1 ストレージ装置 1 0 内のペア管理テーブル T 2 0 から、削除指示されたリモートコピーペアに関する情報を削除する（S 1 4 8）。第 1 ストレージ装置 1 0 は、管理サーバ 8 0 に、リモートコピーペアの削除が完了した旨を通知する（S 1 4 9）。

【0 2 5 2】

図 3 4 は、ロックディスク 2 3 2 を削除させる処理を示すフローチャートである。本処理では、第 1 ストレージ装置 1 0 から第 3 ストレージ装置 3 0 への指示と、第 2 ストレージ装置 2 0 から第 3 ストレージ装置 3 0 への指示とが競合しない場合を述べる。

【0 2 5 3】

管理サーバ 8 0 は、第 1 S V P を介して第 1 ストレージ装置 1 0 に、ロックディスクの削除を指示する（S 1 6 0）。第 1 ストレージ装置 1 0 は、ペア管理テーブル T 2 0 を参照し、削除指示されたロックディスクがいずれかの仮想ボリューム 2 3 1 で使用されているか否かを確認する（S 1 6 1）。もしも、いずれかの仮想ボリューム 2 3 1 で使用されているロックディスクの場合、本処理は中断される。

【0 2 5 4】

ロックディスクが使用されていない場合、第 1 ストレージ装置 1 0 は、キャッシュメモリ 1 3 0 に使用制御情報 L 1 0 が記憶されているか否かを確認する。もしも、キャッシュメモリ 1 3 0 に使用制御情報 L 1 0 が既に記憶されている場合、その内容は古くなっている可能性があるため、第 1 ストレージ装置 1 0 は、キャッシュメモリ上の使用制御情報 L 1 0 を破棄する（S 1 6 1）。S 1 6 1 では、ペア管理テーブル T 2 0 の参照及び古い使用制御情報 L 1 0 の破棄が実行される。

【0 2 5 5】

第 1 ストレージ装置 1 0 は、第 3 ストレージ装置 3 0 に使用制御情報 L 1 0 の読み出しを要求する（S 1 6 2）。第 3 ストレージ装置 3 0 は、ロックディスクから使用制御情報 L 1 0 を読み出して第 1 ストレージ装置 1 0 に送信する（S 1 6 3）。

【0 2 5 6】

第 1 ストレージ装置 1 0 は、使用制御情報 L 1 0 内の管理情報 L 1 1 とロックディスク管理テーブル T 1 0 の内容とが一致するか否かを確認した後、使用制御情報 L 1 0 を更新させるためのライトデータを作成する（S 1 6 4）。

【0 2 5 7】

そのライトデータにおいて、第 1 ストレージ装置 1 0 は、第 1 ストレージ装置の制御情報を「1」から「0」に変更し、第 1 ストレージ装置 1 0 がロックディスクを使用していない状態に戻す。さらに、第 1 ストレージ装置 1 0 は、第 1 ストレージ装置のロック情報ビットマップ L 1 4 を 0 で初期化する。

【0 2 5 8】

そして、第 1 ストレージ装置 1 0 は、上記のように作成されたライトデータを第 3 ストレージ装置 3 0 に送信し、ロックディスク 2 3 2 内の使用制御情報 L 1 0 を更新させる（S 1 6 5）。第 1 ストレージ装置 1 0 は、削除されたロックディスクに関する情報を、第 1 ストレージ装置 1 0 内のロックディスク管理テーブル T 1 0 から消去させる。

【0 2 5 9】

10

20

30

40

50

続いて、管理サーバ 80 は、第 2 SVP を介して、第 2 ストレージ装置 20 に、ロックディスクの削除を指示する (S166)。第 2 ストレージ装置 20 は、ペア管理テーブル T20 を参照し、削除指示されたロックディスクがいずれかの仮想ボリューム 231 で使用されているか否かを確認する (S167)。さらに、第 2 ストレージ装置 20 は、キャッシュメモリ 130 に使用制御情報 L10 が記憶されている場合、その使用制御情報 L10 を破棄する (S167)。

【0260】

第 2 ストレージ装置 20 は、第 3 ストレージ装置 30 に使用制御情報 L10 の読み出しを要求する (S168)。第 3 ストレージ装置 30 は、使用制御情報 L10 を第 2 ストレージ装置 20 に送信する (S169)。

10

【0261】

第 2 ストレージ装置 20 は、以下のように、使用制御情報 L10 を更新させるためのライトデータを作成する (S170)。そのライトデータでは、管理情報 L11 が消去される。第 1 ストレージ装置 10 はロックディスクを使用していないため、管理情報 L11 を消去できる。そのライトデータでは、第 2 ストレージ装置の制御情報は「1」から「0」に変更され、さらに、第 2 ストレージ装置のロック情報ビットマップ L15 は 0 で初期化される。

【0262】

第 2 ストレージ装置 20 は、そのライトデータを第 3 ストレージ装置 30 に送信し、使用制御情報 L10 を更新させる (S171)。これにより、そのロックディスクは削除される。

20

【0263】

図 35 は、ロックディスクの削除について、第 1 ストレージ装置から第 3 ストレージ装置 30 への指示と、第 2 ストレージ装置 20 から第 3 ストレージ装置 30 への指示とが競合する場合を示す。通信ネットワークの混雑の程度、または、ストレージ装置の応答の遅れ等により、適切な実行順序が得られない場合があり得る。なお、以下の説明では、指示が競合する点を中心に説明し、テーブルの更新内容等の詳細を割愛する。

【0264】

管理サーバ 80 は、第 1 SVP を介して第 1 ストレージ装置 10 に、ロックディスクの削除を指示する (S180)。続いて、管理サーバ 80 は、第 2 SVP を介して第 2 ストレージ装置 20 に、ロックディスクの削除を指示する (S181)。

30

【0265】

第 1 ストレージ装置 10 は、第 3 ストレージ装置 30 に使用制御情報 L10 の転送を要求する (S182)。第 3 ストレージ装置 30 は、第 1 ストレージ装置 10 に使用制御情報 L10 を転送する (S183)。第 1 ストレージ装置 10 は、読み出された使用制御情報 L10 を利用してライトデータを作成する (S188)。

【0266】

図 35 に示す例では、第 1 ストレージ装置 10 がライトデータを作成して使用制御情報 L10 を更新させるよりも前に、第 2 ストレージ装置 20 が、第 3 ストレージ装置 30 から使用制御情報 L10 を取得し (S184, S185)、ライトデータを作成する (S186)。そして、第 2 ストレージ装置 20 は、作成されたライトデータを第 3 ストレージ装置 30 に転送し、使用制御情報 L10 を更新させる (S187)。

40

【0267】

第 2 ストレージ装置 20 による使用制御情報 L10 の更新が完了した後で、第 1 ストレージ装置 10 は、ライトデータ (S188) を第 3 ストレージ装置 30 に送信し、ロックディスク内の使用制御情報 L10 を更新させる (S189)。

【0268】

第 1 ストレージ装置 10 は、予定通り更新されたか否かを確認するために、ロックディスクから使用制御情報 L10 を読み出し、ライトデータの内容と比較する。しかし、第 2 ストレージ装置 20 による更新処理が先に完了しているため、S182 で取得された使用

50

制御情報 L 1 0 に基づくライトデータと、S 1 8 9 の処理で再度取得された使用制御情報 L 1 0 とは一致しない (S 1 9 0)。

【0 2 6 9】

そこで、第 1 ストレージ装置 1 0 は、改めてライトデータを作り直し (S 1 8 8)、その新たなライトデータを用いてロックディスク内の使用制御情報 L 1 0 を更新させる (S 1 9 1)。そのライトデータでは、管理情報 L 1 1 は消去される。

【0 2 7 0】

図 3 6 は、図 3 5 で示した問題をリザーブコマンドの採用によって解消する例を示すフローチャートである。リザーブコマンドとは、処理の実行について予約するためのコマンドである。

10

【0 2 7 1】

管理サーバ 8 0 は、第 1 ストレージ装置 1 0 にロックディスクの削除を指示する (S 2 0 0)。続けて、管理サーバ 8 0 は、第 2 ストレージ装置 2 0 にもロックディスクの削除を指示する (S 2 0 1)。

【0 2 7 2】

第 1 ストレージ装置 1 0 は、第 3 ストレージ装置 3 0 にリザーブコマンドを発行する (S 2 0 2)。第 3 ストレージ装置 3 0 は、リザーブコマンドを受け付けた旨を第 1 ストレージ装置 1 0 に通知する (S 2 0 3)。これにより、削除対象のロックディスクについて、第 1 ストレージ装置 1 0 以外の他のストレージ装置によるリードアクセス及びライトアクセスが禁止される。

20

【0 2 7 3】

第 1 ストレージ装置 1 0 は、第 3 ストレージ装置 3 0 に使用制御情報 L 1 0 の転送を要求する (S 2 0 4)。第 3 ストレージ装置 3 0 は、第 1 ストレージ装置 1 0 に使用制御情報 L 1 0 を送信する (S 2 0 5)。

【0 2 7 4】

第 1 ストレージ装置 1 0 は、読み出された使用制御情報 L 1 0 に基づいて、ロックディスクを削除するためのライトデータを作成する (S 2 0 8)。

【0 2 7 5】

第 1 ストレージ装置 1 0 によってロックディスク内の使用制御情報 L 1 0 が更新されるよりも前に、第 2 ストレージ装置 2 0 が、第 3 ストレージ装置 3 0 にリザーブコマンドを発行したとする (S 2 0 6)。既に、削除対象のロックディスクについては、第 1 ストレージ装置 1 0 からリザーブコマンドが発行されている (S 2 0 2)。従って、第 3 ストレージ装置 3 0 は、第 2 ストレージ装置 2 0 にエラーを返す。リザーブコマンドは、リリースコマンドによって明示的に解除される必要がある。

30

【0 2 7 6】

第 1 ストレージ装置 1 0 は、ライトデータ (S 2 0 8) を第 3 ストレージ装置 3 0 に送信し、ロックディスク内の使用制御情報 L 1 0 を更新させる (S 2 0 9)。更新完了後に、第 1 ストレージ装置 1 0 は、第 3 ストレージ装置 3 0 にリリースコマンドを発行する (S 2 1 0)。第 3 ストレージ装置 3 0 は、リリースコマンドを受信すると、S 2 0 2 で受け付けたリザーブコマンドによるリザーブ状態を解除する (S 2 1 1)。

40

【0 2 7 7】

その後で、第 2 ストレージ装置 2 0 は、ロックディスク内の使用制御情報 L 1 0 を更新させる (S 2 0 2 - S 2 0 5, S 2 0 8 - S 2 1 0)。これにより、ロックディスクが削除される。

【0 2 7 8】

図 3 7 は、一回の指示で、ロックディスクの削除及び仮想ボリュームの削除を行う場合の例を示す。管理サーバ 8 0 は、第 1 SVP を介して第 1 ストレージ装置 1 0 に、ロックディスクの削除を指示する (S 2 2 0)。

【0 2 7 9】

第 1 ストレージ装置 1 0 は、ロックディスクの削除指示を受信すると、まず最初に、削

50

除を指示されたロックディスクに関連する全てのリモートコピーペア（仮想ボリューム）の削除を、第2ストレージ装置20に指示する（S221）。

【0280】

第2ストレージ装置20は、仮想ボリュームを削除するためのライトデータを作成して第3ストレージ装置30に送信し、使用制御情報L10を更新させる（S222）。第2ストレージ装置20は、副ボリュームの状態を「ペア」から「シンプレックス」に変更し、さらに、削除対象の仮想ボリュームに関する情報をペア管理テーブルT20から消去する（S223）。第2ストレージ装置20は、第2ストレージ装置側における仮想ボリュームの削除が完了した旨を、第1ストレージ装置10に報告する（S224）。

【0281】

第1ストレージ装置10は、第2ストレージ装置20からの通知を受信すると、削除対象のロックディスクに対応付けられている仮想ボリュームを削除するべく、ライトデータを作成して第3ストレージ装置30に送信し、ロックディスク内の使用制御情報L10を更新させる（S225）。第1ストレージ装置10は、正ボリュームの状態を「ペア」から「シンプレックス」に変更し、さらに、削除対象の仮想ボリュームに関する情報をペア管理テーブルT20から消去する（S226）。

【0282】

続いて、第1ストレージ装置10は、ロックディスクを削除するためのライトデータを作成し、そのライトデータを第3ストレージ装置30に送信し、使用制御情報L10を更新させる（S227）。第1ストレージ装置10は、ロックディスク管理テーブルT10から、削除対象のロックディスクに関する情報を消去する（S228）。そして、第1ストレージ装置10は、ロックディスクの削除が完了した旨をホスト70に通知する（S229）。

【0283】

図38は、正ボリューム単独で運用する場合を示すフローチャートである。例えば、第2ストレージ装置の保守作業のために、第1ストレージ装置10のみを作動させる必要が生じる。

【0284】

管理サーバ80は、第1SVPを介して第1ストレージ装置10に、サスペンドを指示する（S240）。第1ストレージ装置10は、ペア管理テーブルT10を参照し、サスペンド可能か否かを判定する。サスペンドできない場合、本処理は中断する。

【0285】

サスペンド可能な場合、第1ストレージ装置10は、使用制御情報L10を更新させる（S241）。具体的には、第1ストレージ装置のロック情報ビットマップL14において、正ボリュームに関する仮想ボリュームに対応するビットを「1」に設定する。

【0286】

第1ストレージ装置10は、ロックディスク管理テーブルT10を更新し（S242）、第2ストレージ装置20にサスペンドへの移行を指示する（S243）。第2ストレージ装置20は、その指示を受けて、ペア状態を「PSUS」に変更し（S244）、ステータス変更が完了した旨を第1ストレージ装置10に通知する（S245）。

【0287】

第1ストレージ装置10は、第2ストレージ装置20からの通知を受信すると、ペア管理テーブルT20に記憶されているペア状態を「PSUS」に変更する（S246）。第1ストレージ装置10は、サスペンドへの移行が完了した旨を管理サーバ80に通知する（S247）。

【0288】

図39は、正ボリューム単独で運用する状態から通常の状態に戻すためのペアリシンク処理を示すフローチャートである。管理サーバ80は、第1ストレージ装置10に、ペアリシンクを指示する（S250）。第1ストレージ装置10は、ペア管理テーブルT20を参照し、ペアリシンクが可能か否かを判定する。ペアリシンクが不可能な場合、本処理

10

20

30

40

50

は中断する。

【0289】

ペアリシンクが可能な場合、第1ストレージ装置10は、ロックディスク内の使用制御情報L10を更新させる(S251)。具体的には、第1ストレージ装置のロック情報ビットマップL14内の、対応するビットの値を「1」から「0」に変更させる。第1ストレージ装置10は、第1ストレージ装置10内のロックディスク管理テーブルT10を更新する(S252)。

【0290】

そして、第1ストレージ装置10は、第2ストレージ装置20にペアリシンクを指示する(S253)。第2ストレージ装置20は、第2ストレージ装置内のペア管理テーブルT20において、リシンク対象のリモートコピーペアの状態を「ペア」に変更させる(S254)。第2ストレージ装置20は、ペア状態を変更した旨を第1ストレージ装置10に通知する(S255)。

10

【0291】

第1ストレージ装置10は、第1ストレージ装置内のペア管理テーブルT20において、リシンク対象のリモートコピーペアの状態を「ペア」に変更させる(S256)。第1ストレージ装置10は、管理サーバ80に、ペアリシンクが完了した旨を報告する(S257)。

【0292】

その後、ペア状態の変更とは別のタイミングで、正ボリュームの記憶内容と副ボリュームの記憶内容との再同期が行われる。正ボリューム単独で運用されている間に、ホスト70によって更新されたデータの位置は、差分ビットマップで管理されている。差分ビットマップは、正ボリュームの記憶内容と副ボリュームの記憶内容との間で生じた差分を管理するための情報である。

20

【0293】

第1ストレージ装置10は、第2ストレージ装置20に、差分コピーの開始を指示する(S260)。第1ストレージ装置10は、差分ビットマップを用いて、第2ストレージ装置20に差分データを転送する(S261)。第2ストレージ装置20は、第1ストレージ装置10から受信した差分データを副ボリュームに書き込む。差分コピーが完了すると、第2ストレージ装置20は、第1ストレージ装置10に差分コピーの完了を通知する(S262)。

30

【0294】

図40は、副ボリューム単独で運用する場合の処理を示すフローチャートである。例えば、保守作業等のために、副ボリュームのみを稼働させる場合がある。まず最初に、管理サーバ80は、第2SVPを介して第2ストレージ装置20に、スワップサスペンドへの移行を指示する(S270)。

【0295】

第2ストレージ装置20は、ペア管理テーブルT20を参照してスワップサスペンドが可能か否かを判定する。スワップサスペンドが可能な場合、第2ストレージ装置20は、第3ストレージ装置30内のロックディスク232にアクセスし、使用制御情報L10を更新させる(S271)。具体的には、第2ストレージ装置20は、第2ストレージ装置のロック情報ビットマップL15内の、スワップサスペンド対象の仮想ボリュームに対応するビットの値を「1」に設定させる。

40

【0296】

第2ストレージ装置20は、項目C17についてロックディスク管理テーブルT10を更新し(S272)、さらに、第1ストレージ装置10にスワップサスペンドへの移行を通知する(S273)。

【0297】

第1ストレージ装置10は、第1ストレージ装置の有するペア管理テーブルT20内の正ボリュームのペア状態を、「PSUS(サスペンド)」に変更し(S274)、状態変

50

更が完了した旨を第2ストレージ装置20に通知する。

【0298】

第1ストレージ装置10からの通知を受信した第2ストレージ装置20は、第2ストレージ装置の有するペア管理テーブルT20内の副ボリュームのペア状態を、「SSWS（スワップサスペンド）」に変更する（S275）。第2ストレージ装置20は、管理サーバ80にスワップサスペンドへの移行が完了した旨を通知する（S277）。

【0299】

図41は、副ボリューム単独での運用状態から、通常のリモートコピーペア状態に戻すための処理を示すフローチャートである。

【0300】

管理サーバ80は、第2ストレージ装置20にリバースリシンの実行を指示する（S280）。第2ストレージ装置20は、ペア管理テーブルT20を参照し、リバースリシンが可能か否かを判定する。リバースリシンが可能な場合、第2ストレージ装置20は、ロックディスク232内の使用制御情報L10を更新させる（S281）。第2ストレージ装置20は、第2ストレージ装置のロック情報ビットマップL15内の、リバースリシン対象のボリュームに対応するビットの値を「0」に変更させる。

【0301】

第2ストレージ装置20は、ロックディスク管理テーブルT10を更新し（S282）、第1ストレージ装置10にリバースリシンの実行を通知する（S283）。第1ストレージ装置10は、ペア管理テーブルT20において、正ボリュームを副ボリュームに変更し、さらに、ペア状態を「PAIR」に変更する（S284）。第1ストレージ装置10は、第2ストレージ装置20に変更完了を通知する（S285）。

【0302】

第2ストレージ装置20は、ペア管理テーブルT20において、副ボリュームを正ボリュームに変更し、ペア状態を「PAIR」に変更する（S286）。つまり、正ボリュームを副ボリュームに変更し（S284）、かつ、副ボリュームを正ボリュームに変更することにより（S286）、正ボリュームと副ボリュームとを入れ替える。

【0303】

第2ストレージ装置20は、管理サーバ80に、リバースリシンの完了を通知する（S287）。そして、別のタイミングで、正ボリューム（以前の副ボリューム）から副ボリューム（以前の正ボリューム）に、差分データのコピーが行われる。

【0304】

正ストレージ装置となった第2ストレージ装置20は、副ストレージ装置となった第1ストレージ装置10に、差分コピーの実行を通知する（S290）。第2ストレージ装置20は、第1ストレージ装置10に差分データを転送する（S291）。第1ストレージ装置10は、差分データをキャッシュメモリ130に保存した後で、副ボリュームに書き込む。第2ストレージ装置20は、第1ストレージ装置10に差分コピーの完了を通知する（S292）。

【0305】

図42は、所定の契機が到来した場合に、自動的にリバースリシンを行うための処理を示すフローチャートである。図41の場合は、ユーザが管理サーバ80から、リバースリシンの実行を手動で指示している。これに対し、図42に示す処理では、例えば、スワップサスペンドの移行後に、リバースリシンを自動的に実行する。

【0306】

ホスト70は、第1ストレージ装置10内の正ボリュームにライトコマンドを発行する（S301）。しかし、障害等により、第1ストレージ装置10はライトコマンドを処理することができず、エラー応答が返される（S302）。

【0307】

そこで、ホスト70のパス制御部78は、アクティブパスからパッシブパスに切り替えて（S303）、第2ストレージ装置20内の副ボリュームにライトコマンドを発行する

10

20

30

40

50

(S304)。

【0308】

第2ストレージ装置20は、ロックディスク内の使用制御情報L10を更新し、スワップサスペンド状態に移行する(S304)。ライトデータは、副ボリュームにのみ書き込まれる。第2ストレージ装置20は、副ボリュームにライトデータを書き込んだ後、ホスト70に処理完了を通知する(不図示)。

【0309】

スワップサスペンドに移行した後で、第2ストレージ装置20は、リバーシリシンの実行契機が到来したか否かを判定し、実行契機の到来を検出した場合には(S305)、リバーシリシンクを実行する(S306-S322)。

10

【0310】

リバーシリシンの実行契機としては、例えば、スワップサスペンド状態への移行直後、あるいは、スワップサスペンド状態への移行から所定時間経過後、あるいは、スワップサスペンド状態への移行からハートビート通信の再開後等、を挙げることができる。

【0311】

第2ストレージ装置20は、第1ストレージ装置10にリバーシリシンの実行を通知する(S306)。この通知を受信した第1ストレージ装置10は、ペア管理テーブルT20において、正ボリュームを副ボリュームに変更し、かつ、ペア状態を「PAIR」に変更する(S307)。

【0312】

20

第2ストレージ装置20は、第1ストレージ装置10側の変更完了を確認すると、ペア管理テーブルT20において、副ボリュームを正ボリュームに変更し、かつ、ペア状態を「PAIR」に変更する(S308)。第2ストレージ装置20は、ロックディスク内の使用制御情報L10を更新し、ロック情報ビットマップL15内の対応ビットを「0」に変更させる(S309)。第2ストレージ装置20は、ホスト70に、リバーシリシンクが完了した旨を通知する(S310)。

【0313】

そして、別のタイミングで、第2ストレージ装置20は、第1ストレージ装置10に差分コピーの実行を通知し(S320)、差分データを第1ストレージ装置10に転送する(S321)。第1ストレージ装置10は、差分データをキャッシュメモリ130に格納し、差分コピーが完了した旨を第2ストレージ装置20に報告する(S322)。

30

【0314】

本実施例は、上述のように構成されるため、以下の効果を奏する。本実施例では、第1ストレージ装置10及び第2ストレージ装置20とは別の第3ストレージ装置30内にロックディスク232を設け、そのロックディスク232内に、正ボリュームと副ボリュームとから構成される仮想ボリューム231の使用を制御するための使用制御情報L10を格納させる。従って、本実施形態では、各ストレージ装置10、20は、ロックディスク232を共有することにより、ストレージ装置10、20間での切替を適切に行うことができ、ホスト70が各ストレージ装置10、20の切替を意識する必要は無い。

【0315】

40

本実施例では、使用制御情報L10の管理情報L11に、ロックディスクID_L111と、第1ストレージ装置10及び第2ストレージ装置20をそれぞれ特定するための識別情報L112、L113とが含まれている。つまり、本実施例では、ロックディスクIDと各ストレージ装置の製造番号との合計3つの情報を対応付けて管理することができ、他のストレージ装置にロックディスク232が関連づけられてしまう等の誤りを未然に防止できる。

【0316】

本実施例では、各ストレージ装置10、20内に仮想的に設けられる外部接続ボリュームに、外部ボリュームとして構成されるロックディスク232を対応付ける。従って、第3ストレージ装置30の記憶資源を活用することができる。

50

【0317】

本実施例では、ユーザは、管理サーバ80から、仮想ボリュームの設定と、ロックディスクの設定と、外部接続の設定とを、ストレージ装置に指示することができる。従って、使い勝手が高まる。

【0318】

本実施例では、第1ストレージ装置10は、使用制御情報L10のうち第1ストレージ装置に関する情報のみを更新可能であり、第2ストレージ装置20は、使用制御情報L10のうち第2ストレージ装置に関する情報のみを更新可能である。従って、第1ストレージ装置10が誤って第2ストレージ装置20に関する情報を書き換えたり、その逆に、第2ストレージ装置20が誤って第1ストレージ装置10に関する情報を書き換えたりするのを未然に防止でき、信頼性が向上する。

10

【0319】

本実施例では、使用制御情報L10を更新する場合、その更新直後に、ロックディスク232から使用制御情報L10を読み出して、正しく更新されたか否かを確認する。従って、それぞれ異なるストレージ装置10、20が一つのロックディスク232を共有する場合でも、使用制御情報L10が適切に更新されることを保証でき、ストレージシステムの信頼性を高めることができる。

【0320】

本実施例では、ロックディスク232の削除を一回指示するだけで、そのロックディスク232に関連する仮想ボリューム231も削除することができる。従って、ユーザの使い勝手が向上する。

20

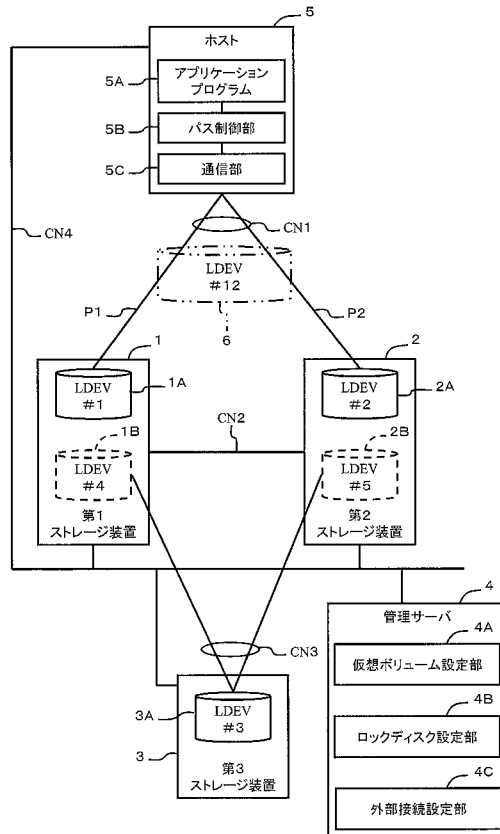
【0321】

本実施例では、スワップサスペンドへの移行後に、所定の実行契機が検出された場合には、自動的にリバースリシンクを行うこともできる。従って、ユーザの使い勝手を高めることができる。

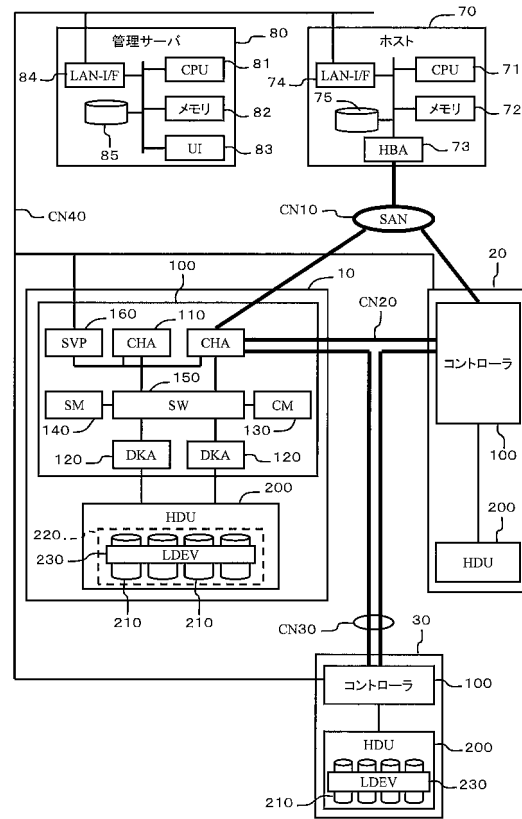
【0322】

なお、本発明は、上述した実施例に限定されない。当業者であれば、本発明の範囲内で、種々の追加や変更等を行うことができる。

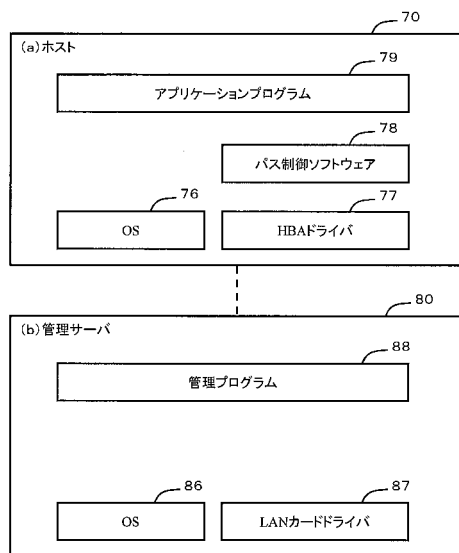
【図 1】



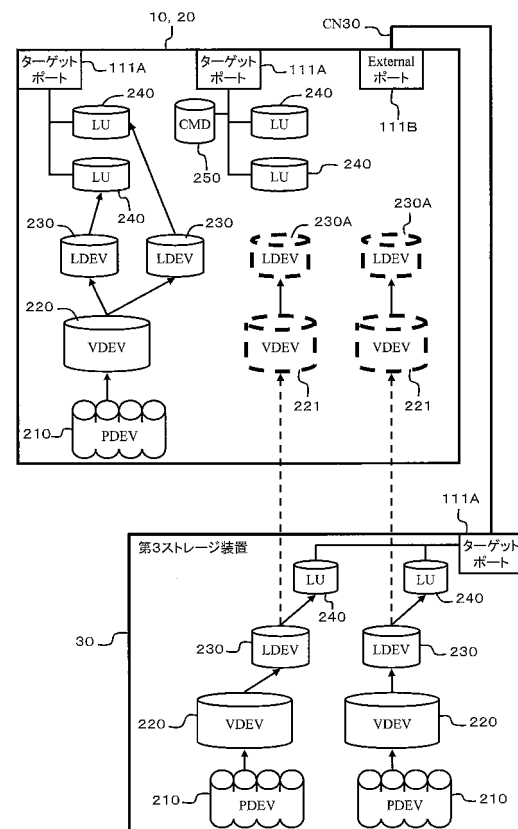
【図 2】



【図 3】



【図 4】



【図 9】

[illegible]

【図 10】

LDEV管理テーブル						
C31	C32	C33	C34	C35	T30	
LDEV #	内部VOL／外部VOL	VDEV #	開始アドレス	サイズ		
0x0000	内部VOL	0x0000	0x00000000	1000		
0x0001	内部VOL	0x0000	0x00010000	1000		
0x0002	内部VOL	0x0001	0x00010000	2000		
0x0003	外部VOL	0x0003	0x00010000	1000		
...	...	...	...	...		

【図 1 1】

外部ボリューム管理テーブル					
VDEV#	接続 ポート	外部ストレージ情報			
		LUN	ベンダ名	装置名	ボリュームID
0x0000					
0x0001					
0x0002					
0x0003	0x01	0x02	H	RA	123456789
0x0004					
0x0005					
...	...	...	...	...	...

【图 1 2】

G11

Subsystem	
--LDKC#00	Used
--	
--LDKC#01	Not Used
--	
--LDKC#02	

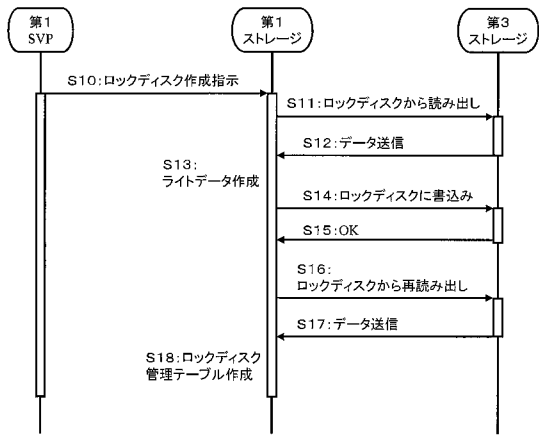
G12

Lock ID	LDKC: CU: LDEV	相手装置S/N	Control ID
00	-	-	-
01	-	-	-
02	-	-	-
03	-	-	-
...	...	...	...

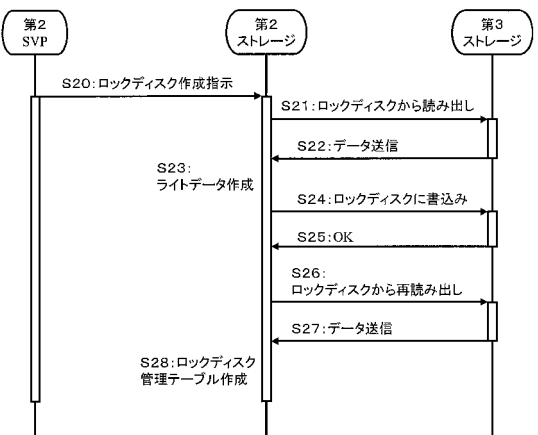
Callout box pointing to LDKC#00 in G11:
ロックデバイス生成
ロックデバイス削除

B11: Apply, Cancel

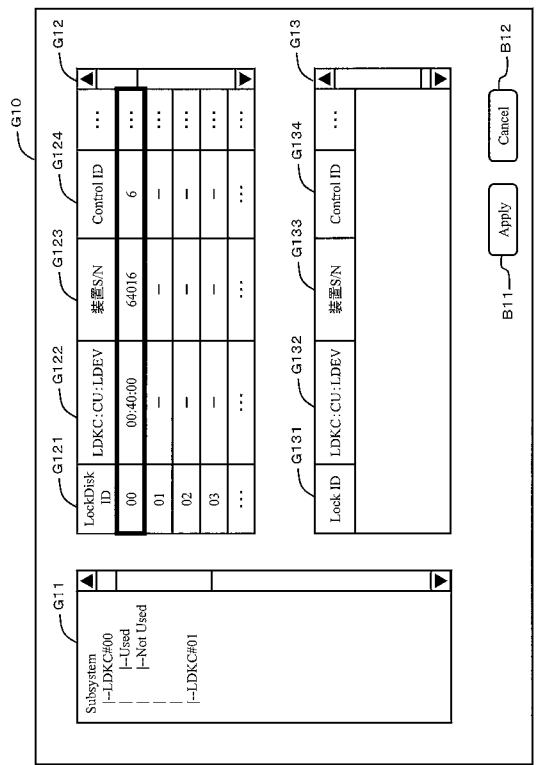
【図 1 3】



【図 1 4】



【図 1 5】



【図 1 6】

ロックディスク管理テーブル						
ロックディスク ID	管理フラグ		自装置 S/N	相手装置 S/N	Control ID	ロック情報BMP
	C121	C122				
	有効 / 無効	ロックディスク作成中 / 削除中				
0x00	有効	OFF	64036	64016	0x0066	0x000000000000...
0x01	無効	OFF	0	0	0	0

【図 18】

Figure 1 illustrates the sequence of operations for the first program. The operations are listed vertically in a column:

- ベア作成 (Bare metal creation) - M21
- ベア削除 (Bare metal deletion) - M22
- サスペンド (Suspend) - M23
- スワップサスペンド (Swap suspend) - M24
- リシンク (Resync) - M25
- リバースリシンク (Reverse resync) - M26

An arrow labeled M20 points to the start of the first operation (ベア作成).

【図 20】

```

sequenceDiagram
    participant SVP as 第1 SVP
    participant ST1 as 第1 ストレージ
    participant ST2 as 第2 ストレージ
    participant ST3 as 第3 ストレージ

    SVP->>ST1: S30: 仮バックボリューム作成指示
    ST1->>ST2: S31: ベア管理情報を作成、  
テーブルに登録
    ST2->>ST3: S32: ベア管理情報を送信
    ST2->>ST2: S33: ベア管理情報登録
    ST2->>ST3: S34: ロックディスクを更新
    ST2->>ST2: S35: 更新
    ST2->>ST3: S36: OK
    ST2->>ST2: S37: OK
    ST2->>ST3: S38: ロックディスクを更新
    ST2->>ST2: S39: 更新
    ST2->>ST3: S40: OK
    ST2->>ST1: S41: OK
    ST2->>ST3: S42: 形成コピー指示
    ST2->>ST3: S43: データ転送
    ST2->>ST3: S44: OK
  
```

【図 2 1】

Subsystem
|--LDKC#00
|--CL1A
|--CL3A
|--CL5A
|--CL7A
|--CL1B
|--CL3B
|--CL5B
|--CL7B
|--CL2A
|--CL4A
|--CL2B
|--CL4B
|--CL6B

LDKC:CU:LDEV

00:01:00

00:01:01

...

相手装置S/N

PAIR

SMPL

02

...

ステータス

PAIR

03

...

Control ID

6

...

LockDisk ID

0x00

...

...

...

LDKC:CU:LDEV

00:01:00

00:01:01

...

相手装置S/N

PAIR

SMPL

02

...

ステータス

PAIR

03

...

Control ID

6

...

LockDisk ID

0x00

...

...

...

LDKC:CU:LDEV

00:01:00

00:01:01

...

相手装置S/N

PAIR

SMPL

02

...

ステータス

PAIR

03

...

Control ID

6

...

LockDisk ID

0x00

...

...

...

【図 2 2】

ペア管理テーブル									
PVOL情報				SVOL情報			ロックディスクID		
C211	C212	C213	C221	C222	C223	C224	C225	C23	C24
S/N	LDEV#	ペア状態	S/N	LDEV#	ペア状態	S/N	LDEV#	ペア状態	ロックディスクID
64036	0x0100	PAIR	64016	0x0200	PAIR	64016	0x0200	PAIR	0x00

【図 2 3】

Subsystem
|--LDKC#00
|--Used
|--Not Used
|--LDKC#01

LDKC:CU:LDEV

00:40:00

00:04:01

相手装置S/N

64016

64016

Control ID

6

6

LockDisk ID

00

01

02

03

LDKC:CU:LDEV

00:40:00

00:04:01

相手装置S/N

64016

64016

Control ID

6

6

LockDisk ID

00

01

02

03

LDKC:CU:LDEV

00:40:00

00:04:01

相手装置S/N

64016

64016

Control ID

6

6

LockDisk ID

00

01

02

03

【図 2 4】

Subsystem
|--LDKC#00
|--CL1A
|--CL3A
|--CL5A
|--CL7A
|--CL1B
|--CL3B
|--CL5B
|--CL7B
|--CL2A
|--CL4A
|--CL2B
|--CL4B
|--CL6B

LDKC:CU:LDEV

00:01:00

00:01:01

00:01:0A

相手装置S/N

64016

64016

64016

ステータス

PAIR

PAIR

PAIR

03

Control ID

6

6

6

LockDisk ID

0x00

0x00

0x01

LDKC:CU:LDEV

00:01:00

00:01:01

00:01:0A

相手装置S/N

64016

64016

64016

ステータス

PAIR

PAIR

PAIR

03

Control ID

6

6

6

LockDisk ID

0x00

0x00

0x01

LDKC:CU:LDEV

00:01:00

00:01:01

00:01:0A

相手装置S/N

64016

64016

64016

ステータス

PAIR

PAIR

PAIR

03

Control ID

6

6

6

LockDisk ID

0x00

0x00

0x01

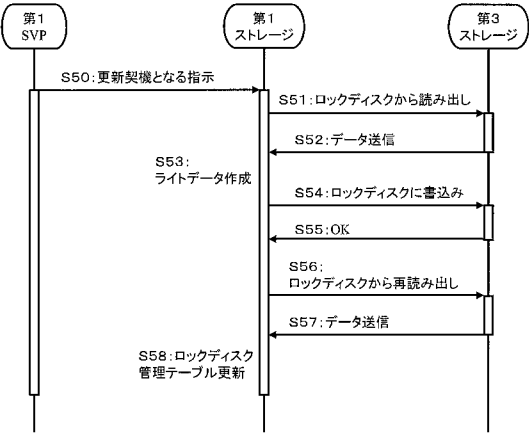
【図 25】

ロックディスク管理テーブル										T10
C11	C12			C13	C14	C15	C16	C17		
	管理フラグ									
ロックディスク ID	C121	C122	C123	LDEV #	自装置 S/N	相手装置 S/N	Control ID	ロック情報BMP		
	有効 /無効	ロックディスク 作成中	ロックディスク 削除中							
0x00	有効	OFF	OFF	0x0400	64036	64016	0x0006	0x000000000000...		
0x01	有効	OFF	OFF	0x0401	64036	64016	0x0006	0x000000000000...		

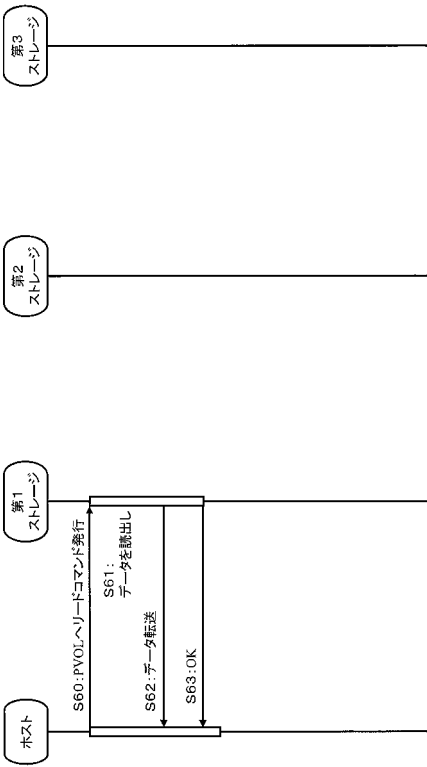
【図 26】

ペア管理テーブル										T20
C21			C22			C23				
PVOL情報			SVOL情報			ロックディスクID				
C211	C212	C213	C221	C222	C223					
S/N	LDEV#	ペア状態	S/N	LDEV#	ペア状態					
64036	00:01:00	PAIR	64016	00:07:50	PAIR	0x00				
64036	00:01:01	PAIR	64016	00:07:51	PAIR	0x00				
64036	00:01:0A	PAIR	64016	00:07:5A	PAIR	0x01				

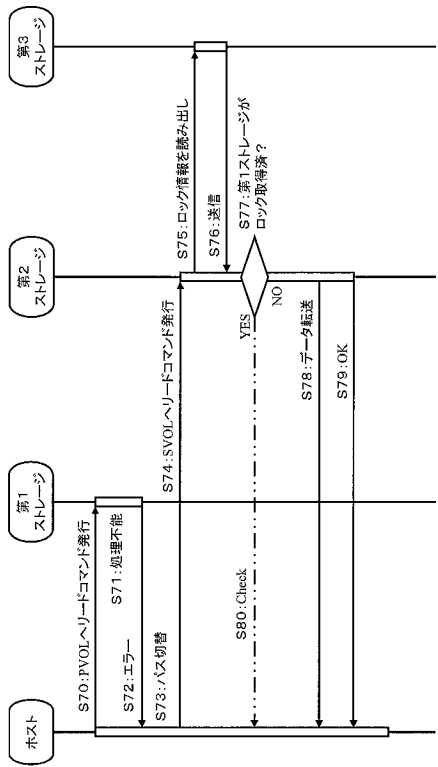
【図 27】



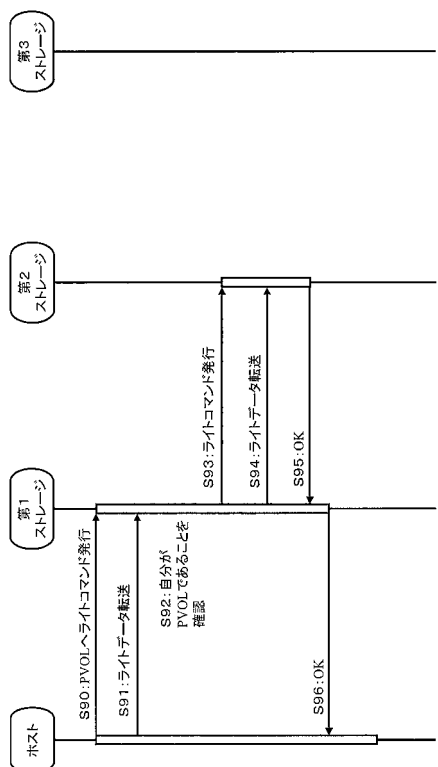
【図 28】



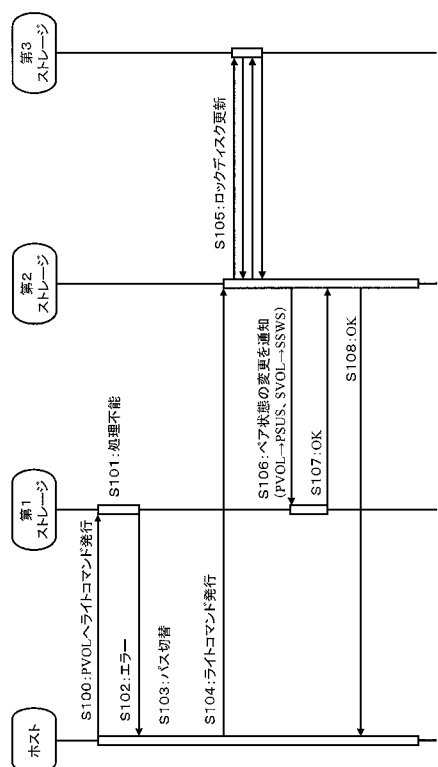
【図 29】



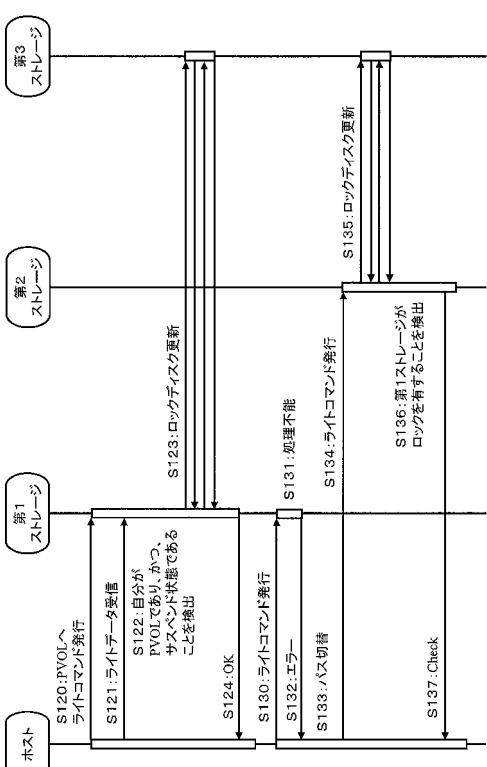
【図 30】



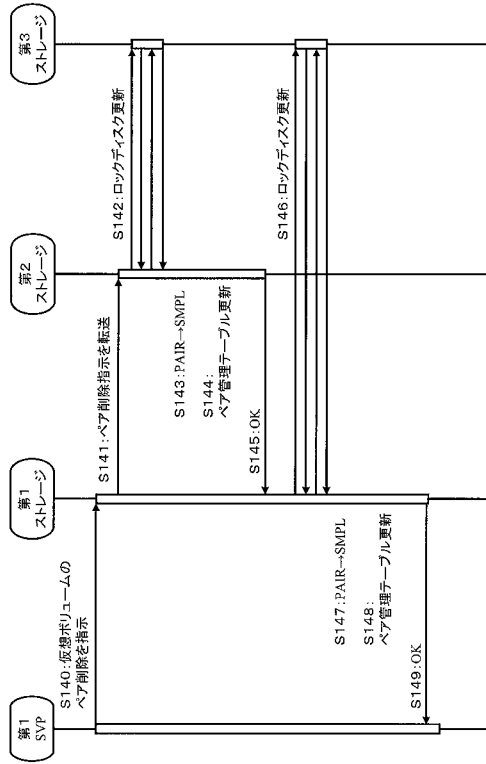
【図 31】



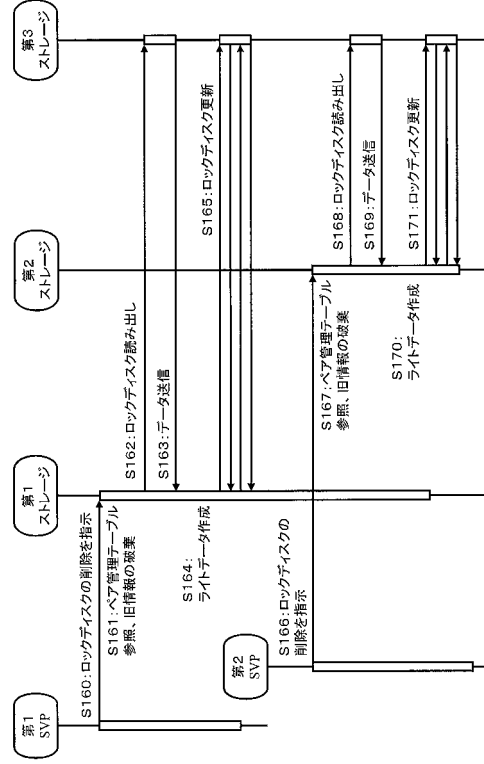
【図 32】



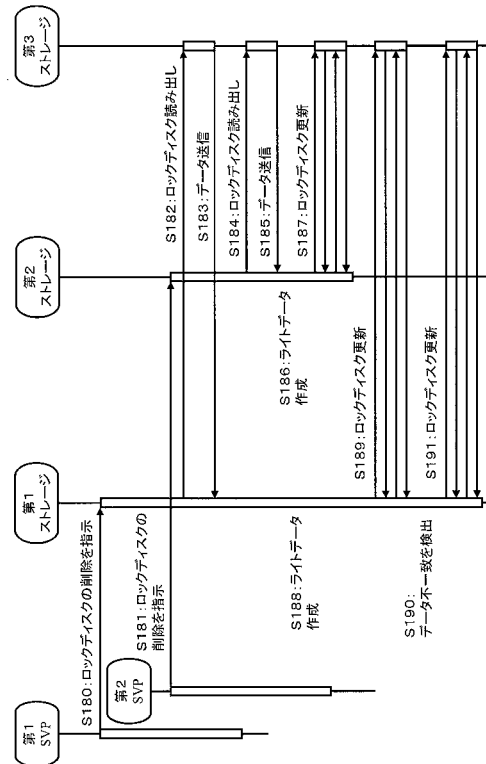
【図 3 3】



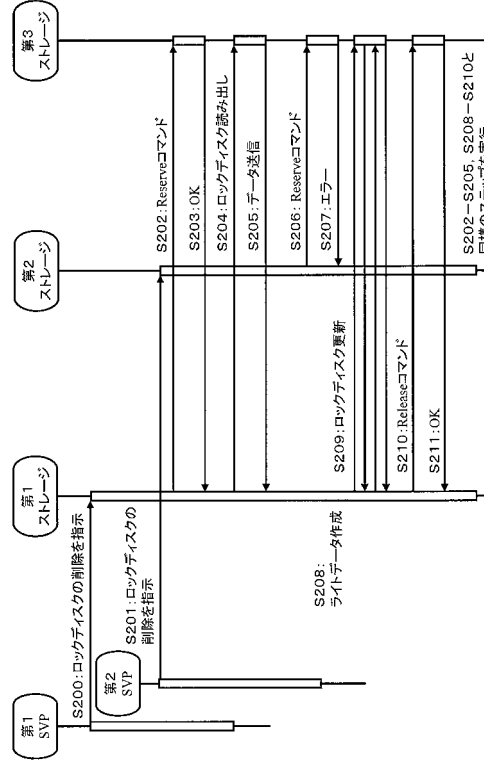
【図 3 4】



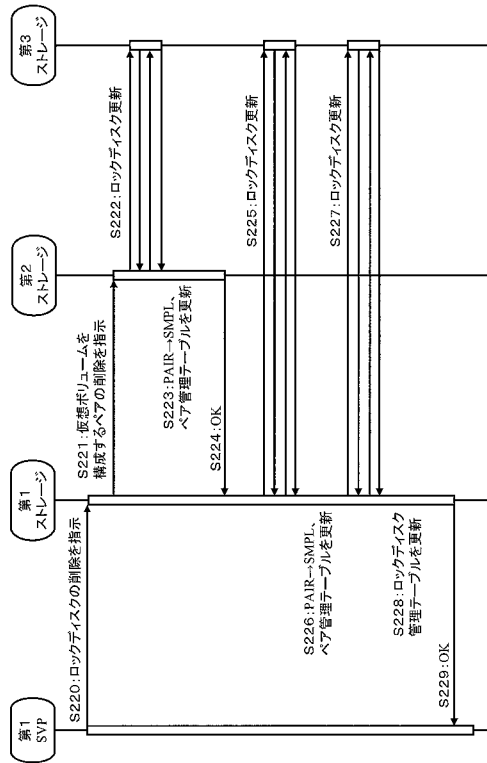
【図 3 5】



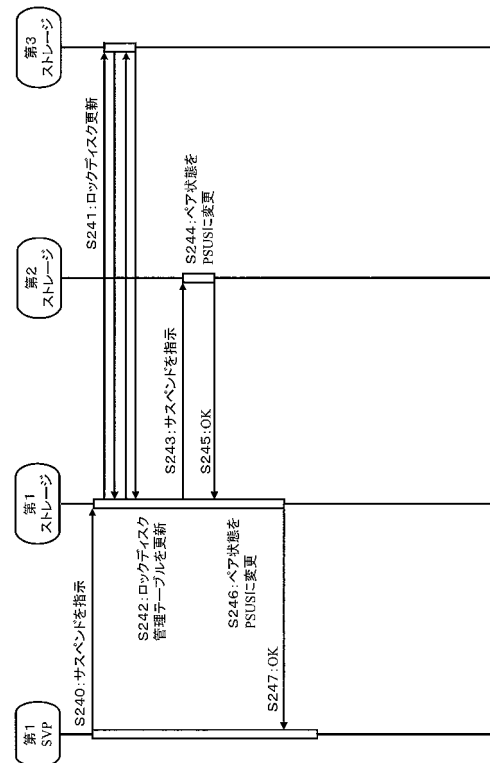
【図 3 6】



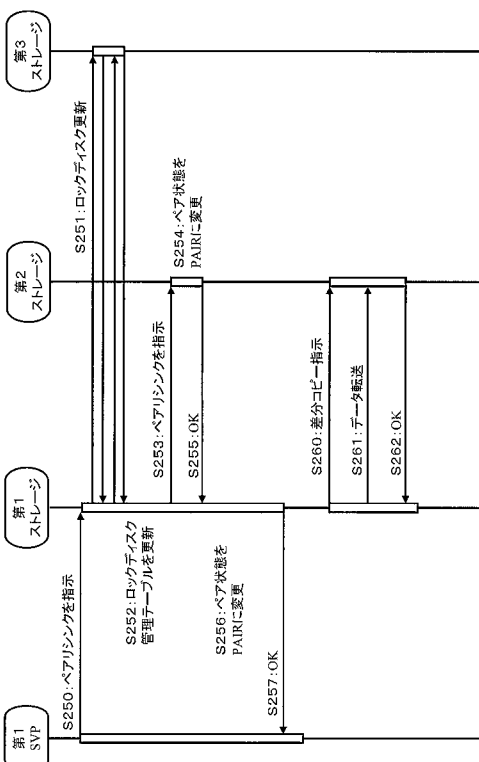
【図 37】



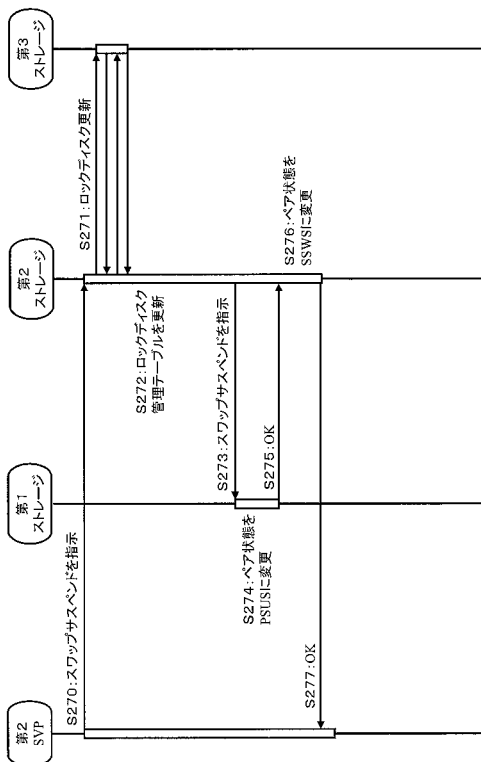
【図 38】



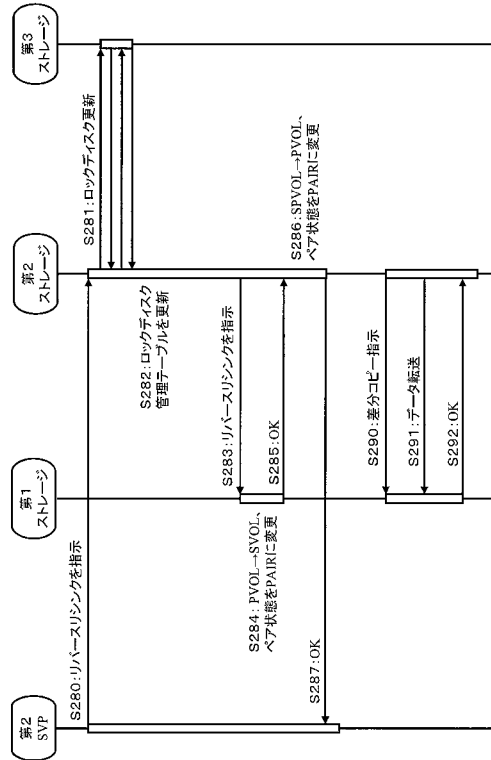
【図 39】



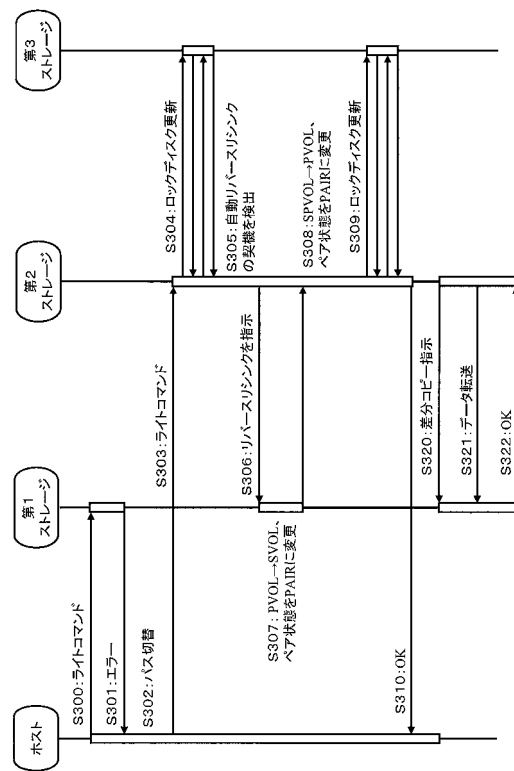
【図 40】



【図 4 1】



【図 4 2】



フロントページの続き

- (56)参考文献 特開2008-134987 (JP, A)
国際公開第2001/063424 (WO, A1)
特開2008-181271 (JP, A)
特開2008-134988 (JP, A)
特開2003-131917 (JP, A)
特開2007-213611 (JP, A)
特開2008-033704 (JP, A)
特開2006-285336 (JP, A)
特開2006-134021 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G06F 3/06

G06F 13/10