

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2007-140601

(P2007-140601A)

(43) 公開日 平成19年6月7日(2007.6.7)

(51) Int. Cl.

G06F 3/06 (2006.01)

F I

G06F 3/06 301Z

テーマコード(参考)

5B065

審査請求 未請求 請求項の数 15 O L (全 29 頁)

(21) 出願番号 特願2005-329265 (P2005-329265)  
 (22) 出願日 平成17年11月14日(2005.11.14)

(71) 出願人 000005108  
 株式会社日立製作所  
 東京都千代田区丸の内一丁目6番6号  
 (74) 代理人 100079108  
 弁理士 稲葉 良幸  
 (74) 代理人 100093861  
 弁理士 大賀 眞司  
 (72) 発明者 尾崎 聡  
 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  
 (72) 発明者 本間 久雄  
 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

最終頁に続く

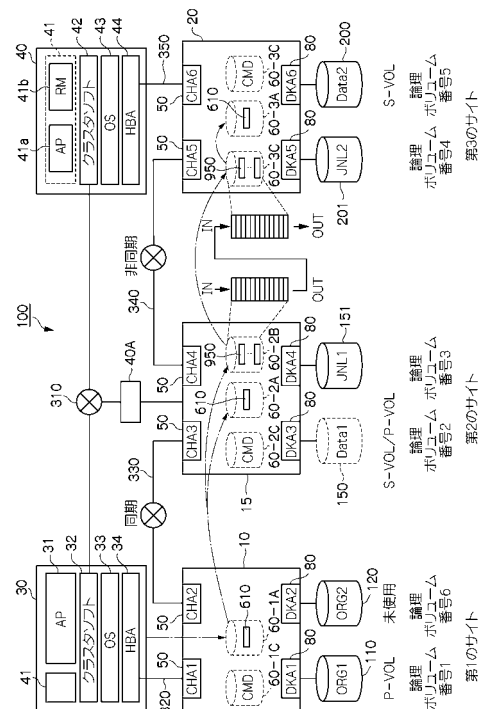
(54) 【発明の名称】 記憶制御システム

(57) 【要約】

【課題】 特定の記憶制御装置に接続される他の記憶制御装置にホスト装置を設けることなく、特定の記憶制御装置に接続する上位装置から他の記憶制御装置に対して、当該他の記憶制御装置が処理すべき制御コマンドを発行できる、記憶制御システムを提供する。

【解決手段】 第1の記憶制御装置は、仮想ボリュームと、前記仮想ボリュームを前記論理デバイスにマッピングする情報と、前記論理デバイスの属性情報と、が記録された記憶部と、を備え、上位装置は、前記仮想ボリュームを第2の記憶制御装置の論理デバイスとして認識し、次いで、当該論理デバイスに対して、前記仮想ボリュームを介して、制御コマンドを発行できるように構成した。

【選択図】 図1



**【特許請求の範囲】****【請求項 1】**

上位装置と、  
当該上位装置と通信可能に接続された第 1 の記憶制御装置と、  
当該第 1 の記憶制御装置と通信可能に接続され、制御コマンドが書き込まれる論理デバイスを備えるとともに、当該前記制御コマンドを実行するように構成された、第 2 の記憶制御サイトと、  
を備える記憶制御システムであって、  
前記第 1 の記憶制御装置は、  
仮想ボリュームと、  
前記仮想ボリュームを前記論理デバイスにマッピングする情報と、前記論理デバイスの属性情報と、が記録された記憶部と、  
を備え、前記上位装置は、前記仮想ボリュームを第 2 の記憶制御装置の前記論理デバイスとして認識し、次いで、当該論理デバイスに対して、前記仮想ボリュームを介して、前記制御コマンドを発行できるように構成した、記憶制御システム。

10

**【請求項 2】**

前記上位装置が前記第 1 の記憶制御装置に照会コマンドを送信すると、前記第 1 の記憶制御装置は前記上位装置に前記論理デバイスの属性情報を返信して、前記上位装置が前記第 2 の記憶制御装置の前記論理デバイスを認識し、次いで、当該論理デバイスに対して前記制御コマンドを発行できるように構成した、請求項 1 記載の記憶制御システム。

20

**【請求項 3】**

前記第 2 の記憶制御装置に第 2 の上位装置が接続され、この第 2 の上位装置と前記第 1 の記憶制御装置に接続する前記第 1 の上位装置との間に、前記第 1 の上位装置と前記第 2 の上位装置とを直接接続する通信経路が形成されている請求項 1 記載の記憶制御システム。

**【請求項 4】**

前記第 1 の上位装置は、前記通信経路を介して、前記第 2 の上位装置にアクセスし、前記論理デバイスに前記制御コマンドを発行できるように構成された、請求項 2 記載の記憶制御システム。

**【請求項 5】**

前記第 1 の上位装置と前記第 2 の上位装置との間の前記通信経路に通信障害が生じた際、前記第 1 の上位装置は、前記仮想ボリュームを介して、前記制御コマンドを前記論理デバイスに発行するように構成された、請求項 4 記載の記憶制御システム。

30

**【請求項 6】**

前記上位装置からの制御コマンドが書き込まれる第 2 の論理デバイスを備え、当該第 2 の論理デバイスの前記制御コマンドを実行するように構成された、第 3 の記憶制御装置が前記第 2 の記憶制御装置に通信可能に接続されている、請求項 1 記載の記憶制御システム。

**【請求項 7】**

前記第 1 の記憶制御装置は前記第 3 の記憶制御装置と接続してなり、当該第 1 の記憶制御装置は、前記第 2 の論理デバイスにマッピングされた第 2 の仮想ボリュームを備え、この第 2 の仮想ボリュームを介して前記第 2 の論理デバイスに、当該第 1 の記憶制御装置に接続される、前記上位装置の前記制御コマンドを発行するように構成された、請求項 6 記載の記憶制御システム。

40

**【請求項 8】**

前記第 3 の記憶制御装置は、前記第 2 の記憶制御装置の前記論理デバイスにマッピングされた第 3 の仮想ボリュームを備え、第 1 の記憶制御装置は前記第 2 の仮想ボリュームと前記第 3 の仮想ボリュームを介して、前記第 2 の記憶制御装置の論理デバイスに前記制御コマンドを発行する、請求項 7 記載の記憶制御システム。

**【請求項 9】**

50

前記第 2 の記憶制御装置に第 3 の記憶制御装置が接続され、前記第 2 の記憶制御装置には前記第 3 の記憶制御装置にマッピングされる第 2 の仮想ボリュームを備え、この第 2 の仮想ボリュームは前記第 1 の仮想ボリュームにマッピングされており、前記上位装置が前記第 1 の仮想ボリュームに対して前記制御コマンドを発行すると、当該制御コマンドは前記第 2 の仮想ボリュームを介して前記第 3 の記憶制御装置の論理デバイスに前記制御コマンドを発行できるようにした請求項 1 記載の記憶制御システム。

【請求項 10】

第 1 の記憶制御装置と、  
当該第 1 の記憶制御装置に接続する第 1 の上位装置と、  
前記第 1 の記憶制御装置に接続する第 2 の記憶制御装置と、  
当該第 2 の記憶制御装置に接続する第 2 の上位装置と、  
を備える記憶制御システムであって、  
前記第 2 の記憶制御装置は、制御コマンドが書き込まれる論理デバイスを備えるとともに、当該制御コマンドを実行するように構成され、

10

前記第 1 の記憶制御装置は、前記論理デバイスにマッピングされた仮想ボリュームを備え、前記第 1 の上位装置が前記仮想ボリュームを前記第 2 の記憶制御装置の前記論理デバイスとして認識し、次いで、前記仮想ボリュームを介して当該論理デバイスに対して前記制御コマンドを発行できるように構成した、記憶制御システム。

【請求項 11】

前記第 1 の上位装置と前記第 2 の上位装置とを接続する通信経路が形成されている請求項 10 記載の記憶制御システム。

20

【請求項 12】

前記第 1 の上位装置は、前記通信経路又は前記仮想ボリュームを介した前記第 2 の記憶制御装置との間に形成された通信経路によって、前記制御コマンドを前記論理デバイスに発行するようにした、請求項 11 記載の記憶制御システム。

【請求項 13】

複数の記憶制御装置間を接続した記憶制御システムの記憶制御方法であって、  
一つの記憶制御装置に接続された上位装置が、前記一つの記憶制御装置に接続された他の記憶制御装置の論理デバイスを認識するステップと、  
前記上位装置が前記論理デバイスにマッピングされた、前記第一つの記憶制御装置の仮想ボリュームに、前記他の記憶制御装置が実行する制御コマンドを発行するステップと、  
前記仮想ボリュームから前記制御コマンドを前記論理デバイスに送信するステップと、  
前記他の記憶制御装置が前記論理デバイスの前記制御コマンドを実行するステップと、  
を備える記憶制御方法。

30

【請求項 14】

上位装置と、  
この上位装置と通信可能に接続された第 1 の記憶サイトと、  
当該第 1 の記憶サイトと通信可能に接続される第 2 の記憶サイトと、  
を備えてなる記憶制御システムであって、  
前記第 2 の記憶サイトは、前記上位装置からの制御コマンドが書き込まれる論理デバイスを備え、当該制御コマンドを実行するように構成され、

40

前記第 1 の記憶サイトは前記論理デバイスにマッピングされた仮想ボリュームを備え、前記上位装置は前記仮想ボリュームを前記第 2 の記憶制御装置の前記論理デバイスとして認識し、ユーザに対して前記論理デバイスが前記第 2 の記憶制御装置に属していることを表示するように構成された記憶制御システム。

【請求項 15】

前記第 2 の記憶サイトが複数の記憶制御装置から構成され、各記憶制御装置は前記論理デバイスを備え、前記上位装置は複数の前記論理デバイスがそれぞれどの記憶制御装置に属しているかを表示させるように構成された、請求項 14 記載の記憶制御システム。

【発明の詳細な説明】

50

## 【技術分野】

## 【0001】

本発明は、ある記憶制御装置に他の記憶制御装置が接続されている、記憶制御システムに関するものである。

## 【背景技術】

## 【0002】

この種の記憶制御システムの一つにディザスタリカバリを目的とした、記憶制御装置（ストレージ装置）間ボリュームレプリケーション機能を利用したシステムが知られている。このシステムは、リモートコピーシステムと呼ばれている。リモートコピーシステムは、一つの記憶制御システムとは離れた地理的位置に他の記憶制御システムを備え、一つの記憶制御システムに格納されたデータの複製を、他の記憶制御システムに格納するようにしている。

10

## 【0003】

このリモートコピーシステムに関する従来例として、例えば、特開2005-157521号公報に記載のものが存在する。このものは、ホストコンピュータに直結していない遠隔地の記憶サブシステムのリモートコピー状態情報を取得する事を目的として、各記憶サブシステムに、ホストコンピュータからの状態情報取得命令を受信する手段と、受信した命令を解析し、自身が対象となっているか否か判断する手段と、自身が対象でなければ、自身に接続する下位の記憶サブシステムに当該命令を送出する手段と、下位の記憶サブシステムから状態情報を受信した場合、自身に接続する上位の記憶サブシステムに状態情報を送出手段とを備えさせる構成となっている。

20

## 【0004】

このリモートコピーシステムでは、特定の記憶制御装置にリモートサイトの記憶制御装置が通信手段を介して接続する構成を有している。この記憶制御装置同士の接続は、例えば、特開2005-107645号公報に記載されている。なお、この公報には、外部の記憶資源を内部記憶資源として仮想化することにより、記憶資源を有効利用する記憶制御装置が記載されている。

## 【0005】

この種のリモートコピーシステムにおいて、中間サイトを設け、特定の記憶制御装置とリモートサイトの記憶制御装置とが中間サイトの記憶制御装置を介して接続するシステムも知られている。このシステムでは、中間サイトの記憶制御装置の論理ボリュームとリモートサイトの記憶制御装置の論理ボリュームとをコピーペアの関係に設定するとともに、特定サイトの記憶制御装置の論理ボリュームをリモートサイトの論理ボリュームに同期コピーを行っている。中間サイトの論理ボリュームはリモートサイトの論理ボリュームに非同期コピーされている。

30

## 【0006】

この種のリモートコピーシステムにおいて、記憶制御装置はホスト計算機からの制御コマンドを処理するために、コマンドデバイスを備えている。コマンドデバイスとは、記憶制御装置が上位装置であるホスト計算機との制御コマンドを通信するための専用の論理デバイス（或いは論理ボリューム）である。ホスト計算機は、記憶制御装置に形成されたコマンドデバイスに対して制御コマンドをデータとして書き込み、記憶制御装置は、コマンドデバイスに記憶された制御コマンドを処理する。なお、特開2005-115898号公報には、リモートサイトの記憶制御装置のコマンドデバイスを特定記憶制御装置に対して仮想化する技術が記載されている。

40

【特許文献1】特開2005-157521号公報

【特許文献2】特開2005-107645号公報

【特許文献3】特開2005-115898号公報

## 【発明の開示】

## 【発明が解決しようとする課題】

## 【0007】

50

メインサイトにある特定の記憶制御装置に接続するホスト装置からの、論理ボリュームのペア設定などの制御コマンドを、前記中間サイトやリモートサイトの記憶制御措置に実行させるためには、これらサイトにもホスト装置を接続し、ホスト装置同士間で制御コマンドを通信するためのネットワークが必要である。したがって、ホスト装置同士の通信が障害を受けた場合などにおいて、メインサイトのホスト装置から他のサイトのホスト装置に制御コマンドを発行できないという課題がある。

【0008】

そこで、本発明は、特定の記憶制御装置に接続される他の記憶制御装置にホスト装置を設けることなく、特定の記憶制御装置に接続する上位装置から他の記憶制御装置に対して、当該他の記憶制御装置が処理すべき制御コマンドを発行できる、記憶制御システムを提供する事を目的とする。本発明の他の目的は、特定の記憶制御装置に接続される他の記憶制御装置にホスト装置を設けたシステムにおいて、特定の記憶制御装置に接続する第1のホスト装置と他の記憶制御装置に接続する第2のホスト装置間での通信に障害が発生しても、第1のホスト装置から他の記憶制御装置に制御コマンドを発行できる、記憶制御システムを提供することにある。

10

【課題を解決するための手段】

【0009】

本発明に係る制御システムは、前記目的を達成するために、特定の記憶制御装置に、他の記憶制御装置のコマンドデバイスに割り当てられた仮想ボリュームを設け、特定の記憶制御装置に接続する上位装置は、当該仮想ボリュームを、前記他の記憶制御装置が処理すべき制御コマンドが設定されるべきコマンドデバイスとして認識することを特徴とするものである。上位装置が前記仮想ボリュームに対して制御コマンドを発行すると、この仮想ボリュームを介して、上位装置 - 上位装置間の接続を介する事無く、前記特定の記憶制御装置に接続する前記他の記憶制御装置のコマンドデバイスに、前記制御コマンドを発行することができる。

20

【0010】

前記本発明の一つの態様は、上位装置と、当該上位装置と通信可能に接続された第1の記憶制御装置と、当該第1の記憶制御装置と通信可能に接続され、制御コマンドが書き込まれる論理デバイスを備えるとともに、当該制御コマンドを実行するように構成された、第2の記憶制御装置と、を備える記憶制御システムであって、前記第1の記憶制御装置は、仮想ボリュームと、前記仮想ボリュームを前記論理デバイスにマッピングする情報と、前記論理デバイスの属性情報と、が記録された記憶部と、を備え、前記上位装置は、前記仮想ボリュームを第2の記憶制御装置の前記論理デバイスとして認識し、次いで、当該論理デバイスに対して、前記仮想ボリュームを介して、制御コマンドを発行できるように構成した、ことを特徴とするものである。

30

【0011】

さらに、本発明の他の態様は、第1の記憶制御装置と、当該第1の記憶制御装置に接続する第1の上位装置と、前記第1の記憶制御装置に接続する第2の記憶制御装置と、当該第2の記憶制御装置に接続する第2の上位装置と、を備える記憶制御システムであって、前記第2の記憶制御装置は、制御コマンドが書き込まれる論理デバイスを備えるとともに、当該制御コマンドを実行するように構成され、前記第1の記憶制御装置は、前記論理デバイスにマッピングされた仮想ボリュームを備え、前記第1の上位装置が前記仮想ボリュームを前記第2の記憶制御装置の前記論理デバイスとして認識し、次いで、前記仮想ボリュームを介して当該論理デバイスに対して前記制御コマンドを発行できるように構成した、ことを特徴とするものである。

40

【0012】

さらに、本発明の他の態様は、複数の記憶制御装置間を接続した記憶制御システムの記憶制御方法であって、一つの記憶制御装置に接続された上位装置が、前記一つの記憶制御装置に接続された他の記憶制御装置の論理デバイスを認識するステップと、前記上位装置が前記論理デバイスにマッピングされた、前記第一の記憶制御装置の仮想ボリュームに

50

、前記他の記憶制御装置が実行する制御コマンドを発行するステップと、前記仮想ボリュームから前記制御コマンドを前記論理デバイスに送信するステップと、前記他の記憶制御装置が前記論理デバイスの前記制御コマンドを実行するステップと、を備えることを特徴とするものである。

【0013】

さらに、本発明の他の態様は、上位装置と、この上位装置と通信可能に接続された第1の記憶サイトと、当該第1の記憶サイトと通信可能に接続される第2の記憶サイトと、を備えてなる記憶制御システムであって、前記第2の記憶サイトは、前記上位装置からの制御コマンドが書き込まれる論理デバイスを備え、当該制御コマンドを実行するように構成され、前記第1の記憶サイトは前記論理デバイスにマッピングされた仮想ボリュームを備え、前記上位装置は前記仮想ボリュームを前記第2の記憶制御装置の前記論理デバイスとして認識し、ユーザに対して前記論理デバイスが前記第2の記憶制御装置に属していることを表示するように構成されたことを特徴とするものである。

10

【発明の効果】

【0014】

以上説明したように本発明によれば、特定の記憶制御装置の上位装置と他の記憶制御装置の上位装置間での接続経路を介することなく、特定の記憶制御装置に接続する上位装置から他の記憶制御装置に対して、当該他の記憶制御装置が処理すべき内容を持った制御コマンドを発行できる。

【発明を実施するための最良の形態】

20

【0015】

以下、各図を参照して本発明の実施形態について説明する。以下に示す実施形態は本発明の一例であって、本発明を限定解釈するものではない。以下の実施形態は変更・改良され得るとともに、本発明にはその等価物も含まれる。図1はリモートコピーシステム100の構成図である。リモートコピーシステム100は、第1のサイト(「プライマリサイト」又は「メインサイト」ともいう。)に設置された第1の記憶システム10と、第2のサイト(「セカンダリサイト」又は「ローカルサイト」)に設置された第2の記憶システム15と、第3のサイト(リモートサイト)に設置された第3の記憶システム20とを備えて構成されている。各サイトの記憶システムは、記憶制御装置から構成されている。

【0016】

30

第2のサイトは第1のサイトから近距離又は中距離の範囲内に位置し、第3のサイトは第1のサイトから遠距離に位置している。第1の記憶システム10は、ホストコンピュータ(第1の上位計算機システム)30と接続することにより、現用系(稼働系)のデータ処理システムを構築している。一方、第3の記憶システム20は、ホストコンピュータ(第2の上位計算機システム)40と接続することにより、交替系(待機系)のデータ処理システムを構築している。これらのデータ処理システムはクラスタを構成しており、現用系のデータ処理システムに障害が発生した場合に、交替系のデータ処理システムにフェイルオーバーできるように構成されている。

【0017】

ホストコンピュータ30はホストバスアダプタ34を備えており、通信回線320を通じて第1の記憶システム10のチャンネルアダプタ(CHA1)80と接続している。ホストコンピュータ30にはオペレーティングシステム33、クラスタソフトウェア32、及びアプリケーションプログラム31が実装されている。クラスタソフトウェア32はアプリケーションプログラム31が正常に動作しているか否かをチェックする。

40

【0018】

一方、ホストコンピュータ40はホストバスアダプタ44を備えており、通信回線350を通じて第3の記憶システム20のチャンネルアダプタ(CHA6)80と接続している。ホストコンピュータ40にはオペレーティングシステム43、クラスタソフトウェア42、及びリソースグループ41が実装されている。リソースグループ41には、アプリケーションプログラム41a、記憶装置管理ソフトウェア(RAIDマネージャ)41bが

50

含まれている。リソースグループ 41 はホストコンピュータ 30 にも含まれている。記憶装置管理ソフトウェアは、後述の論理ボリュームのペア設定、ペア解除、ペア状態参照など記憶制御装置の制御を実行する。

#### 【0019】

ホストコンピュータ 30, 40 は通信回線 310 を通じて接続されており、第 1 のサイトに障害が発生してアプリケーションプログラム 31 が正常に動作できなくなると、クラスタソフトウェア 32 は障害発生を検出して交替系のホストコンピュータ 40 に起動指示を送る。これにより、現用系のデータ処理システムから交替系のデータ処理システムへのフェイルオーバーが可能になる。アプリケーションプログラム 31, 41 a には、銀行の預貯金管理などの動作を制御するプログラムが含まれている。

10

#### 【0020】

図 1 と図 2 を参照しながら第 1 の記憶システム 10 の構成について説明する。第 1 の記憶システム 10 は、チャンネルアダプタ 50、キャッシュメモリ 60、共有メモリ 70、ディスクアダプタ 80、接続部 90、及び物理ボリューム 900 を備えている。チャンネルアダプタ 50 はホストコンピュータ 30 からの入出力要求を受け付けるインターフェースである。キャッシュメモリ 60 と共有メモリ 70 は、チャンネルアダプタ 50 とディスクアダプタ 80 が共用するメモリである。共有メモリ 70 は、主として、制御情報やコマンド等を記憶するために利用され、例えば、ボリューム情報テーブル 400、ペア設定情報テーブル 500、及びジャーナルグループ設定情報テーブル 600 が格納される。キャッシュメモリ 60 は、主として、データを一時記憶するために利用される。

20

#### 【0021】

例えば、あるチャンネルアダプタ 50 がホストコンピュータ 30 から受け取ったデータ入出力コマンドが書き込みコマンドであった場合に、チャンネルアダプタ 50 はその書き込みコマンドを共有メモリ 70 に書き込むとともに、ホストコンピュータ 30 から受け取ったライトデータをキャッシュメモリ 60 に書き込む。一方、ディスクアダプタ 80 は共有メモリ 70 を監視しており、共有メモリ 70 に書き込みコマンドが書き込まれたことを検出すると、その書き込みコマンドに従ってキャッシュメモリ 60 からライトデータを読み出して、これを物理ボリューム 900 に書き込む。

#### 【0022】

一方、あるチャンネルアダプタ 50 がホストコンピュータ 30 から受け取ったデータ入出力コマンドが読み出しコマンドであった場合に、チャンネルアダプタ 50 はその読み出しコマンドを共有メモリ 70 に書き込むとともに、読み出し対象となるデータがキャッシュメモリ 60 に存在しているか否かをチェックする。ここで、読み出し対象となるデータがキャッシュメモリ 60 に存在している場合には、チャンネルアダプタ 50 はそのデータをキャッシュメモリ 60 から読み取ってホストコンピュータ 30 に送信する。読み出し対象となるデータがキャッシュメモリ 60 に存在していない場合には、共有メモリ 70 に読み出しコマンドが書き込まれたことを検出したディスクアダプタ 80 は読み出し対象となるデータを物理ボリューム 900 から読み出してこれをキャッシュメモリ 60 に書き込むとともに、その旨を共有メモリ 70 に書き込む。チャンネルアダプタ 50 は共有メモリ 70 を監視することにより、読み出し対象となるデータがキャッシュメモリ 60 に書き込まれたことを検出すると、そのデータをキャッシュメモリ 60 から読み出してホストコンピュータ 30 に送信する。

30

40

#### 【0023】

ディスクアダプタ 80 はチャンネルアダプタ 50 から送信された論理アドレス指定によるデータアクセス要求を、物理アドレス指定によるデータアクセス要求に変換し、物理ボリューム 900 へのデータの書き込み又は読み取りを行う。物理ボリューム 900 が RAID によって構成されている場合には、ディスクアダプタ 80 は RAID 構成に従ったデータのアクセスを行う。この他にも、ディスクアダプタ 80 は物理ボリューム 900 に記憶されたデータの複製管理、バックアップ制御、災害発生時のデータ消失防止（ディザスタリカバリ）等を目的としてレプリケーション制御又はリモートコピー制御等を行う。

50

## 【 0 0 2 4 】

接続部 90 は、チャンネルアダプタ 50、キャッシュメモリ 60、共有メモリ 70、ディスクアダプタ 80 を相互に接続する。接続部 90 は例えば高速スイッチングによりデータ伝送を行う超高速クロスバススイッチ等の高速バスで構成される。これにより、チャンネルアダプタ 50 相互間の通信パフォーマンスが大幅に向上するとともに、高速なファイル共有機能や高速フェイルオーバー等が可能になる。尚、キャッシュメモリ 60 と共有メモリ 70 とは上述のように異なる記憶資源によって構成してもよく、或いはキャッシュメモリ 60 の記憶領域の一部を共有メモリ 70 として割り当ててもよい。

## 【 0 0 2 5 】

第 1 の記憶システム 10 は単一又は複数の物理ボリューム 900 を備えており、ホストコンピュータ 30 からアクセス可能な記憶領域を提供する。第 1 の記憶システム 10 が提供する記憶領域には、論理ボリューム (ORG 1) 110 と、論理ボリューム (ORG 2) 120 とが単一又は複数の物理ボリューム 900 の記憶空間上に定義されている。物理ボリューム 900 として、例えば、ハードディスク装置やフレキシブルディスク装置等を用いることができる。物理ボリューム 900 のストレージ構成として、例えば、複数のディスクドライブにより RAID 方式のディスクアレイを構成してもよい。また、物理ボリューム 900 と記憶システム 10 は直接接続してもよく、或いはネットワークを介して接続してもよい。更に、物理ボリューム 900 は第 1 の記憶システム 10 と一体的に構成してもよい。

10

## 【 0 0 2 6 】

以下の説明においては、論理ボリューム (ORG 1) 110 に複製対象となるオリジナルデータが格納されるものとする。また、複製対象のデータと複製データとの区別を容易にするために、複製対象のデータが蓄積される論理ボリュームを正論理ボリューム (P-VOL) と称し、複製データが蓄積される論理ボリュームを副論理ボリューム (S-VOL) と称する。また、一对の正論理ボリュームと副論理ボリュームをペアと称する。

20

## 【 0 0 2 7 】

次に、図 1 と図 3 を参照しながら第 2 の記憶システム 15 の構成について説明する。図中、図 2 に示した符号と同一符号の装置等は同一の装置等を示すものとして、詳細な説明を省略する。第 2 の記憶システム 15 は単一又は複数の物理ボリューム 900 を備えており、論理ボリューム (Data 1) 150 と、論理ボリューム (JNL 1) 151 とが単一又は複数の物理ボリューム 900 の記憶空間上に定義されている。論理ボリューム (Data 1) 150 は第 2 の記憶システム 15 が提供する記憶領域を第 1 の記憶システム 10 から指定させるために仮想的に設定された、実ボリュームを持たない仮想なものである。論理ボリューム (Data 1) 150 は論理ボリューム (ORG 1) 110 の複製を保持する。つまり、論理ボリューム (ORG 1) 110 と論理ボリューム (Data 1) 150 との関係においては、前者が正論理ボリュームであり、後者が副論理ボリュームである。

30

## 【 0 0 2 8 】

次に、図 1 と図 4 を参照しながら第 3 の記憶システム 20 の構成について説明する。図中、図 2 に示した符号と同一符号の装置等は同一の装置等を示すものとして、詳細な説明を省略する。第 3 の記憶システム 20 は単一又は複数の物理ボリューム 900 を備えており、論理ボリューム (Data 2) 200 と、論理ボリューム (JNL 2) 201 とが単一又は複数の物理ボリューム 900 の記憶空間上に定義されている。論理ボリューム (Data 2) 200 は論理ボリューム (Data 1) 150 の複製を保持する。つまり、論理ボリューム (Data 1) 150 と論理ボリューム (Data 2) 200 との関係においては、前者が正論理ボリュームであり、後者が副論理ボリュームである。

40

## 【 0 0 2 9 】

図 5 はボリューム情報テーブル 400 を示している。ボリューム情報テーブル 400 には、各論理ボリュームの物理ボリューム 900 上の物理アドレスを定義付けるとともに、各論理ボリュームの容量やフォーマット形式等の属性情報、及びペア情報が定義されてい

50

る。ここでは説明の便宜上、論理ボリューム番号がリモートコピーシステム100内の各論理ボリュームにユニークなものとして取り扱うが、論理ボリューム番号を各記憶システム単位でユニークに定義し、記憶システム自体の識別子と併せて識別するように設定してもよい。

#### 【0030】

同テーブル400において、論理ボリューム番号1は論理ボリューム(ORG1)110を、論理ボリューム番号2は論理ボリューム(Data1)150を、論理ボリューム番号3は論理ボリューム(JNL1)151を、論理ボリューム番号4は論理ボリューム(JNL2)201を、論理ボリューム番号5は論理ボリューム(Data2)200を、論理ボリューム番号6は論理ボリューム(ORG2)120を、それぞれ示している。論理ボリューム(ORG1)110は論理ボリューム(Data1)150との間にペア番号1のペアが定義されている。また、論理ボリューム(ORG2)120は未使用として定義されている。

10

#### 【0031】

また、同テーブル400において、ボリューム状態の「正」は正論理ボリュームとして正常に動作できる状態にあることを示し、「副」は副論理ボリュームとして正常に動作できる状態にあることを示している。「正常」とあるのは、他の論理ボリュームとはペアが設定されていないが、正常に動作できる状態にあることを示している。また、ディスクアダプタ80は同テーブル400に定義された物理アドレスに基づいて、キャッシュメモリ60から読み出したデータの物理ボリューム900への書き込みを制御し、或いは物理ボリューム900から読み出したデータのキャッシュメモリ60への書き込みを制御する。

20

#### 【0032】

図6はペア設定情報テーブル500を示している。同テーブル500は、論理ボリューム(ORG1)110と論理ボリューム(Data1)150との間にペア番号1のペア関係を定義している。更に、論理ボリューム(Data1)150と論理ボリューム(Data2)200との間にペア番号2のペア関係を定義している。同テーブル500の仮想化「ON」は、ペア関係にある一对の論理ボリュームのうち副論理ボリュームが仮想化されていることを示している。ペア関係が設定されると、ペアのステータスに応じて、正論理ボリュームに生じた書き込み処理が、副論理ボリュームに対する各種の処理を行う契機となる。例えば、ペアのステータスには、ペア状態、サスペンド状態、初期コピー状態等がある。ペアのステータスがペア状態の場合には、正論理ボリュームに書き込まれたデータが副論理ボリュームにも書き込まれるという処理がなされる。ペアのステータスがサスペンド状態の場合には、正論理ボリュームに書き込まれたデータが副論理ボリュームには反映されず、両者の同期がとれていた時点に基づいて正論理ボリュームに対してどのようなデータ更新が行われたかを示す差分情報ビットマップが生成される。

30

#### 【0033】

次に、ジャーナルデータについて説明する。説明の便宜上、データが更新される更新元の論理ボリュームをソース論理ボリュームと称し、更新元論理ボリュームのコピーを保持するボリュームをコピー論理ボリュームと称する。ジャーナルデータは、あるソース論理ボリュームにデータ更新があった場合に、少なくとも更新データそのものと、その更新がソース論理ボリュームのどの位置にあったのかを示す更新情報(例えば、ソース論理ボリュームの論理アドレス)とからなる。ソース論理ボリュームにデータ更新があった場合にジャーナルデータを保持しておけば、そのジャーナルデータからソース論理ボリュームを再現できる。

40

#### 【0034】

つまり、ある時点でソース論理ボリュームとコピー論理ボリュームが同期化しており、両者のデータイメージが同一であることを前提として、その時点以降のソース論理ボリュームに対するデータの更新がある都度、ジャーナルデータを保持しておけば、そのジャーナルデータを用いて、その時点以降のソース論理ボリュームのデータイメージをコピー論理ボリュームに再現可能である。ジャーナルデータを用いれば、ソース論理ボリュームと

50

同じ容量を必要とすることなく、ソース論理ボリュームのデータイメージをコピー論理ボリュームに再現させることができる。ジャーナルデータを保持する論理ボリュームをジャーナル論理ボリュームと称する。上述した論理ボリューム ( J N L 1 ) 1 5 1 と論理ボリューム ( J N L 2 ) 2 0 1 はジャーナル論理ボリュームである。

【 0 0 3 5 】

図 7 はジャーナルグループ設定情報テーブル 6 0 0 を示している。ジャーナルグループとは、論理ボリューム同士のペアのことをいい、ある論理ボリュームと、その論理ボリュームにデータ更新があった場合に、そのライトコマンドをライト先アドレス等の更新情報 6 2 0 とライトデータ 6 1 0 に区分して蓄積するジャーナルボリュームからなる。同テーブル 6 0 0 に示す例では、論理ボリューム ( D a t a 1 ) 1 5 0 と論理ボリューム ( J N L 1 ) 1 5 1 とがジャーナルグループ番号 1 で定義されるジャーナルグループであり、論理ボリューム ( D a t a 2 ) 2 0 0 と論理ボリューム ( J N L 2 ) 2 0 1 とがジャーナルグループ番号 2 で定義されるジャーナルグループである。ジャーナルグループはジャーナルペアと称される場合もある。

10

【 0 0 3 6 】

図 8 を参照しながら、ジャーナルデータについて詳述する。同図は、あるソース論理ボリュームのアドレス 7 0 0 番地から 1 0 0 0 番地までのデータが更新データ 6 3 0 によって更新された状態を示している。この論理ボリュームと対をなすジャーナル論理ボリュームは更新情報領域 9 0 0 0 とライトデータ領域 9 1 0 0 を含む。更新データ 6 3 0 はライトデータ 6 1 0 としてライトデータ領域 9 1 0 0 に書き込まれる。つまり、更新データ 6 3 0 とライトデータ 6 1 0 は同一データである。また、ソース論理ボリュームのどの位置が更新されたかというような、更新に係わる情報 ( 例えば、ソース論理ボリュームのアドレス 7 0 0 番地から 1 0 0 0 番地までのデータが更新されたことを示す情報 ) は更新情報 6 2 0 として、更新情報領域 9 0 0 0 に書き込まれる。

20

【 0 0 3 7 】

ジャーナルデータ 9 5 0 はライトデータ 6 1 0 と更新情報 6 2 0 とからなる。更新情報領域 9 0 0 0 は先頭位置から更新情報 6 2 0 を更新された順に格納していき、更新情報 6 2 0 の格納位置が更新情報領域 9 0 0 0 の終端に達すると、再び、更新情報領域 9 0 0 0 の先頭位置から更新情報 6 2 0 を格納していく。同様に、ライトデータ領域 9 1 0 0 は先頭位置からライトデータ 6 1 0 を更新された順に格納していき、ライトデータ 6 1 0 の格納位置がライトデータ領域 9 1 0 0 の終端に達すると、再び、ライトデータ領域 9 1 0 0 の先頭位置からライトデータ 6 1 0 を格納していく。更新情報領域 9 0 0 0 とライトデータ領域 9 1 0 0 との容量比は固定値でもよく、或いは適宜変更できるように設定してもよい。

30

【 0 0 3 8 】

ここで、図 1 を参照しながら、第 1 の記憶システム 1 0 の論理ボリューム ( O R G 1 ) 1 1 0 へのデータ更新を、第 2 の記憶システム 1 5 を介して、第 3 の記憶システム 2 0 の論理ボリューム ( D a t a 2 ) 2 0 0 に反映する動作について説明する。ホストコンピュータ 3 0 が第 1 の記憶システム 1 0 へのライトアクセスを実行するとき、そのライトコマンドはターゲットチャネルアダプタ ( C H A 1 ) 5 0 に対して発行される。ターゲットチャネルアダプタ ( C H A 1 ) 5 0 はライトコマンドを受信すると、ライトデータ 6 1 0 をキャッシュメモリ 6 0 の記憶領域 6 0 - 1 A に書き込む。

40

【 0 0 3 9 】

このライトデータ 6 1 0 はディスクアダプタ 8 0 によって読み出されて論理ボリューム ( O R G 1 ) 1 1 0 に書き込まれる。一方、チャネルアダプタ ( C H A 2 ) 5 0 はイニシエータとして機能し、記憶領域 6 0 - 1 A に書き込まれたライトデータ 6 1 0 を論理ボリューム ( D a t a 1 ) 1 5 0 に書き込むことを指示するライトコマンドを、通信回線 3 3 0 を介して第 2 の記憶システム 1 5 のターゲットチャネルアダプタ ( C H A 3 ) 5 0 に対して発行する。ターゲットチャネルアダプタ ( C H A 3 ) 5 0 はライトコマンドを受信すると、ライトデータ 6 1 0 をキャッシュメモリ 6 0 の記憶領域 6 0 - 2 A に書き込む。

50

## 【 0 0 4 0 】

更に、ターゲットチャネルアダプタ ( C H A 3 ) 5 0 はジャーナルデータ 9 5 0 をキャッシュメモリ 6 0 の記憶領域 6 0 - 2 B に書き込む。記憶領域 6 0 - 2 B は F I F O ( F i r s t I n F i r s t O u t ) 構造を備えており、ジャーナルデータ 9 5 0 を時系列的に順次蓄積する。このジャーナルデータはディスクアダプタ ( D K A 4 ) 8 0 によって論理ボリューム ( J N L 1 ) 1 5 1 に書き込まれる。尚、本実施例では、論理ボリューム ( D a t a 1 ) 1 5 0 は仮想ボリュームであるので、ディスクアダプタ ( D K A 3 ) 8 0 による論理ボリューム ( D a t a 1 ) 1 5 0 への書き込み処理は行われない。

## 【 0 0 4 1 】

第 3 の記憶システム 2 0 のチャネルアダプタ ( C H A 5 ) 5 0 はイニシエータとして機能し、ジャーナルデータの転送を要求するジャーナルリードコマンドを、通信回線 3 4 0 を介して第 2 の記憶システム 1 5 のターゲットチャネルアダプタ ( C H A 4 ) 5 0 に対して適宜のタイミングで発行する ( P U L L 方式)。ジャーナルリードコマンドを受信したターゲットチャネルアダプタ ( C H A 4 ) 5 0 は記憶領域 6 0 - 2 B に蓄積されたジャーナルデータ 9 5 0 を古い順から読み出し、チャネルアダプタ ( C H A 5 ) 5 0 に転送する。記憶領域 6 0 - 2 B からのジャーナルデータの読み出し位置はポインタによって指定される。

## 【 0 0 4 2 】

チャネルアダプタ ( C H A 5 ) 5 0 はジャーナルデータを受信すると、これをキャッシュメモリ 6 0 の記憶領域 6 0 - 3 B に書き込む。記憶領域 6 0 - 3 B は F I F O 構造を備えており、ジャーナルデータ 9 5 0 を時系列的に順次蓄積する。このジャーナルデータはディスクアダプタ ( D K A 5 ) 8 0 によって論理ボリューム ( J N L 2 ) 2 0 1 に書き込まれる。ディスクアダプタ ( D K A 5 ) 8 0 は論理ボリューム ( J N L 2 ) 2 0 1 に書き込まれたジャーナルデータを読み出し、ライトデータ 6 1 0 をキャッシュメモリ 6 0 の記憶領域 6 0 - 3 A に書き込む。記憶領域 6 0 - 3 A に書き込まれたライトデータ 6 1 0 はディスクアダプタ ( D K A 5 ) 8 0 によって読み出されて、論理ボリューム ( D a t a 2 ) 2 0 0 に書き込まれる。ジャーナルデータ 9 5 0 を論理ボリューム ( J N L 2 ) 2 0 1 に保持しておくことにより、例えば、第 2 の記憶システム 2 0 の負荷が大きい場合に、ジャーナルデータ 9 5 0 の正規化処理を行わず、第 2 の記憶システム 2 0 の負荷が小さくなってからジャーナルデータ 9 5 0 の正規化処理を行ってもよい。また、第 2 の記憶システム 1 5 から第 3 の記憶システム 2 0 にジャーナルデータ 9 5 0 を転送する上で、第 2 の記憶システム 1 5 から自発的に第 3 の記憶システム 2 0 にジャーナルデータ 9 5 0 を転送してもよい ( P U S H 方式)。

## 【 0 0 4 3 】

尚、上述の説明において、第 1 の記憶システム 1 0 と第 2 の記憶システム 1 5 との間は同期転送によるリモートコピー ( 同期コピー ) が行われており、第 2 の記憶システム 1 5 と第 3 の記憶システム 2 0 との間は非同期転送によるリモートコピー ( 非同期コピー ) が行われる。同期コピーとは、本実施例で例えると、ホストコンピュータ 3 0 から第 1 の記憶システム 1 0 にデータ更新要求があったときに、当該データを第 1 の記憶システム 1 0 から第 2 の記憶システム 1 5 に転送し、第 2 の記憶システム 1 5 でのデータ更新完了をもって第 1 の記憶システム 1 0 のデータ更新完了を保証する処理手順をいう。

## 【 0 0 4 4 】

第 1 の記憶システム 1 0 と第 2 の記憶システム 1 5 との間で同期コピーを行うことで、論理ボリューム ( O R G 1 ) 1 1 0 と論理ボリューム ( D a t a 1 ) 1 5 0 のデータイメージは巨視的に視ると常に一致している。巨視的に視て常に一致とは、データの同期転送を行っている間は、データの転送時間や各記憶システム 1 0 , 1 5 の処理時間の単位 (  $\mu$  sec ) で一致してない状態があり得るが、データ更新処理完了の時点ではデータイメージは常に一致していることを示している。

## 【 0 0 4 5 】

これに対し、非同期コピーとは、本実施例で例えると、第 1 の記憶システム 1 0 から第

2の記憶システム15に対するデータ更新要求の延長では、当該データを第3の記憶システム20には転送せず、第2の記憶システム15でのデータ更新完了後、非同期に第3の記憶システム20にデータを転送する処理手順をいう。つまり、第2の記憶システム15は自己のスケジュールに応じて(例えば、処理負荷が小さいときを選択して)、第1の記憶システム15からのデータ更新要求とは非同期に、第3の記憶システム20にデータを転送することをいう。第2の記憶システム15と第3の記憶システム20との間で非同期コピーを行うことで、論理ボリューム(Data2)200のデータイメージは、過去のある時点における論理ボリューム(Data1)150のデータイメージに一致しているが、現時点における論理ボリューム(Data1)150のデータイメージと一致しているとは限らない。

10

## 【0046】

図9はリモートコピーシステム100の初期設定手順を記述したフローチャートである。これから述べる各設定は、ユーザがホストコンピュータ30,40又は保守端末のGUI(Graphical User Interface)を通じて所望の入力操作を行うことで設定することができる。まず、ユーザは、第3の記憶システム20のジャーナルグループを設定する(S101)。具体的には、ジャーナルグループ設定情報テーブル600に、論理ボリューム(Data2)200と論理ボリューム(JNL2)201とから成るジャーナルグループを設定する。

## 【0047】

次に、論理ボリューム(ORG1)110と論理ボリューム(Data2)200との間にペア関係を設定し、初期コピーを行う(S102)。これにより、論理ボリューム(ORG1)110と論理ボリューム(Data2)200に同一のデータイメージを持たせることができる。そして、初期コピー完了後に、論理ボリューム(ORG1)110と論理ボリューム(Data2)200とのペア関係は解除される(S103)。次に、論理ボリューム(ORG1)110と論理ボリューム(Data1)150との間にペア関係を設定し(S104)、論理ボリューム(Data1)150と論理ボリューム(JNL1)151とをジャーナルグループとして登録する(S105)。この初期設定処理後、第2の記憶システム20におけるライトデータの正規化処理が可能になる。

20

## 【0048】

図10は第2の記憶システム15が行うアクセス受信処理の説明図である。図1に示した符号と同一符号の装置等は同一の装置等を示すものとして詳細な説明は省略する。第1の記憶システム10はホストコンピュータ30からライトコマンドを受信すると、指定された論理ボリューム(ORG1)110にデータの書き込みを行う(処理A1)。ここで、第1の記憶システム10の論理ボリューム(ORG1)110は、第2の記憶システム15の論理ボリューム(Data1)150とペアになっているので、第1の記憶システム10は、ホストコンピュータ30から受信したライトコマンドと同一のライトコマンドを第2の記憶システム15に発行する(処理A2)。

30

## 【0049】

このライトコマンドはターゲットチャネルアダプタ(CHA3)50によって受信される。ターゲットチャネルアダプタ(CHA3)50は、ライトコマンドによって指定された書き込み先の論理ボリューム(Data1)150が実ボリュームであるか、或いは仮想ボリュームであるかの判定を、ペア設定情報テーブル500を参照して行う。本実施例では論理ボリューム(Data1)150は仮想ボリュームとして設定されているので、ターゲットチャネルアダプタ(CHA3)50は、論理ボリューム(Data1)150が実体を有さないものとして取り扱い、ライトデータ610を論理ボリューム(JNL1)151のライトデータ領域9100に対応するキャッシュメモリ60上の記憶領域に書き込み(処理A3)、更に、そのライトコマンドが論理ボリューム(Data1)150のどの位置に対して行われるものであったかを更新情報620として、論理ボリューム(JNL1)151の更新情報領域9000に対応するキャッシュメモリ60上の記憶領域に書き込む(処理A4)。ディスクアダプタ(DKA4)80は、適宜のタイミングで、

40

50

キャッシュメモリ 60 上のライトデータ 610、及び更新情報 620 を論理ボリューム (JNL1) 151 に書き込む (処理 A5, A6)。

【0050】

図 11 は第 2 の記憶システム 15 が行うアクセス受信処理のフローチャートである。同フローチャートを参照しながら、第 2 の記憶システム 15 が行うアクセス受信処理について再述する。第 2 の記憶システム 15 のターゲットチャネルアダプタ (CHA3) 50 はアクセスコマンドを受信すると、そのアクセスコマンドがライトコマンドであるか否かを判定する (S201)。アクセスコマンドがライトコマンドではなく (S201; NO)、ジャーナルリードコマンドである場合には (S202; YES)、ジャーナルリードコマンド受信処理を行う (S203)。ジャーナルリードコマンド受信処理の詳細については後述する。一方、アクセスコマンドがライトコマンドである場合には (S201; YES)、書き込み先のボリュームが正常であるか否かを判定する (S204)。ボリューム状態が正常でない場合には (S204; NO)、保守端末又は上位装置 (第 1 の記憶装置 10) に異常報告を行って (S205)、処理を終了する。

10

【0051】

一方、ボリューム状態が正常である場合には (S204; YES)、書き込み先の論理ボリュームが仮想ボリュームであるか否かを、ペア設定情報テーブル 500 を参照して判定する (S206)。書き込み先の論理ボリュームが仮想ボリュームである場合には (S206; YES)、論理ボリューム (JNL1) 151 へのジャーナルデータ 950 の書き込み処理を行い (S207)、上位装置に終了報告を通知する (S208)。一方、書き込み先の論理ボリュームが仮想ボリュームでない場合には (S206; NO)、キャッシュメモリ 60 の記憶領域上にデータを書き込んで (S209)、上位装置に終了報告を通知する (S210)。次に、書き込み先の論理ボリュームがジャーナルグループを有する論理ボリュームであるか否かを判定する (S211)。書き込み先の論理ボリュームがジャーナルグループを有する場合には (S211; YES)、論理ボリューム (JNL1) 151 へのジャーナルデータ 950 の書き込み処理を行う (S212)。このように、論理ボリューム (Data1) 150 を仮想化することによって、副論理ボリュームに実体的な記憶容量を持たずことなく、論理ボリューム (ORG1) 110 のリモートコピーの相手先として定義させることができる。

20

【0052】

図 12 はジャーナルリードコマンドを受信した第 2 の記憶システム 15 のターゲットチャネルアダプタ (CHA4) 50 の動作説明図である。第 2 の記憶システム 15 のターゲットチャネルアダプタ (CHA4) 50 は第 3 の記憶システム 20 からジャーナルリードコマンドを受信する (処理 B1)。論理ボリューム (JNL1) 151 に未送信のジャーナルデータ 950 が存在する場合、ターゲットチャネルアダプタ (CHA4) 50 はディスクアダプタ (DKA4) 80 に対して、更新情報 620 とライトデータ 610 をキャッシュメモリ 60 に書き込むことを命令する (処理 B2)。ディスクアダプタ (DKA4) 80 は論理ボリューム (JNL1) 151 から更新情報 620 とライトデータ 610 を読み込み、これらをキャッシュメモリ 60 に書き込んで、ターゲットチャネルアダプタ (CHA4) 50 にリード終了を通知する (処理 B3, B4)。ターゲットチャネルアダプタ (CHA4) 50 はリード終了の通知を受けて、更新情報 620 とライトデータ 610 をキャッシュメモリ 60 から読み取って、第 3 の記憶システム 20 に送信する (処理 B5)。これにより、ジャーナルデータ 950 が書き込まれていたキャッシュメモリ 60 は開放される。

30

40

【0053】

尚、上述したジャーナルリードコマンド受信処理では、論理ボリューム (JNL1) 151 から読み出したジャーナルデータ 950 をキャッシュメモリ 60 に書き込む例を示したが、キャッシュメモリ 60 上にジャーナルデータ 950 が既に存在する場合は、論理ボリューム (JNL1) 151 からのジャーナルデータ 950 の読み取りは不要である。また、第 2 の記憶システム 15 は単一のジャーナルデータ 950 を第 3 の記憶システム 20

50

に個別に送信していたが、複数のジャーナルデータ 950 を同時に第 3 の記憶システム 20 に送信してもよい。

【0054】

また、ジャーナルリードコマンドによって送信されるジャーナルデータ数は、第 3 の記憶システム 20 がジャーナルリードコマンド内に指定してもよく、又はジャーナルグループ登録の際にユーザが第 2 の記憶システム 15 或いは第 3 の記憶システム 20 内に設定してもよい。また、第 2 の記憶システム 15 から第 3 の記憶システム 20 に転送されるジャーナルデータ数は通信回線 340 の転送能力又は伝送負荷に応じて動的に変更してもよい。また、第 2 の記憶システム 15 によるジャーナルデータ 950 の記憶領域の開放処理は、第 3 の記憶システムがジャーナルリードコマンド内に開放してよい更新番号を含ませることにより、第 2 の記憶システム 15 はその指示に従ってジャーナルデータ 950 の記憶領域を開放してもよい。

10

【0055】

図 13 はジャーナルリードコマンドを受信した第 2 の記憶システム 15 のターゲットチャンネルアダプタ (CHA4) 50 の動作を記述したフローチャートである。第 2 の記憶システム 15 のターゲットチャンネルアダプタ (CHA4) 50 は第 3 の記憶システム 20 からアクセスコマンドを受信すると、そのアクセスコマンドがジャーナルリードコマンドである場合に、ジャーナルグループ設定情報テーブル 600 を参照して、ジャーナルグループ状態が正常であるか否かを確認する (S301)。ジャーナルグループ状態に障害が生じており、正常でない場合には (S301; NO)、第 3 の記憶システム 20 にジャーナルグループ状態を通知し、処理を終了する。

20

【0056】

ジャーナルグループ状態が正常である場合は (S301; YES)、ターゲットチャンネルアダプタ (CHA4) 50 は論理ボリューム (JNL1) 151 の状態が正常であるか否かを判定する (S302)。論理ボリューム (JNL1) 151 の状態が正常でない場合には (S302; NO)、ターゲットチャンネルアダプタ (CHA4) 50 はジャーナルグループ設定情報テーブル 600 のペア状態を「障害」に変更し、第 3 の記憶システム 20 にその旨通知して処理を終了する。一方、論理ボリューム (JNL1) 151 の状態が正常である場合には (S302; YES)、ターゲットチャンネルアダプタ (CHA4) 50 は未送信のジャーナルデータ 950 が論理ボリューム (JNL1) 151 に存在しているか否かを判定する (S303)。

30

【0057】

未送信のジャーナルデータ 950 が論理ボリューム (JNL1) 151 に存在していれば (S303; YES)、ターゲットチャンネルアダプタ (CHA4) 50 は第 3 の記憶システム 20 にジャーナルデータ 950 を送信する (S304)。ジャーナルデータ 950 を受信した第 3 の記憶システム 20 は正規化処理を行うことで、論理ボリューム (ORG1) 110 に対するデータ更新を論理ボリューム (Data2) 200 に反映させることができる。一方、未送信のジャーナルデータ 950 が論理ボリューム (JNL1) 151 に存在していなければ (S303; NO)、ターゲットチャンネルアダプタ (CHA4) 50 は第 3 の記憶システム 20 にその旨を通知する (S305)。その後、ジャーナルデータ 950 が書き込まれていた論理ボリューム (JNL1) 151 の記憶領域を開放する (S306)。即ち、第 1 の記憶システム 10 と第 3 の記憶システム 20 においてデータを二重化した後、第 2 の記憶システム 15 はデータを開放することができる。これにより、第 2 の記憶システム 15 の記憶資源を新たな用途に使用することができる。

40

【0058】

図 14 は第 3 の記憶システム 20 のチャンネルアダプタ (CHA6) 50 がジャーナルデータ 950 を利用して論理ボリューム (Data2) 200 のデータ更新を行う動作説明図である。論理ボリューム (JNL2) 201 に正規化対象のジャーナルデータ 950 が存在する場合、最も古いジャーナルデータ 950 に対して正規化処理を行う。ジャーナルデータ 950 に更新番号を連続的に付与し、更新番号が最小 (最古) のジャーナルデータ

50

950から正規化処理を行えばよい。チャンネルアダプタ（CHA6）50はキャッシュメモリ60を確保し、ディスクアダプタ（DKA5）80に対して、更新情報最古のものから更新情報620とライトデータ610を読み出すように指示を送る（処理C1）。

【0059】

ディスクアダプタ（DKA5）80は論理ボリューム（JNL2）201から読み出した更新情報620とライトデータ610をキャッシュメモリ60に書き込む（処理C2，C3）。そして、ディスクアダプタ（DKA5）80はキャッシュメモリ60からライトデータ610を読み取って、論理ボリューム（Data2）200に書き込む（処理C4）。その後、論理ボリューム（Data2）200のデータ更新に反映された更新情報620とライトデータ610とが存在していた記憶領域は開放される。尚、正規化処理はディスクアダプタ（DKA5）80が行ってもよい。

10

【0060】

尚、未送信のジャーナルデータ量がある閾値を超えた場合に、ホストコンピュータ30から第1の記憶システム10へのアクセスを制限し（例えば、第1の記憶システム10のレスポンスを遅くする）、第2の記憶システム15から第3の記憶システム20へのジャーナルデータ950の転送を優先するのが好ましい。

【0061】

図15は第3の記憶システム20のチャンネルアダプタ（CHA6）50による正規化処理の動作手順を記述したフローチャートである。チャンネルアダプタ（CHA6）50は正規化対象となるジャーナルデータ950が論理ボリューム（JNL2）201に存在するか否かを判定する（S401）。正規化対象となるジャーナルデータ950が存在しない場合（S401；NO）、正規化処理は一旦終了し、一定時間経過後に正規化処理を再開する（S401）。正規化対象となるジャーナルデータ950が存在する場合（S401；YES）、更新情報620とライトデータ610を論理ボリューム（JNL2）201からキャッシュメモリ60に読み出すように、ディスクアダプタ（DKA5）80に指示を送る（S402）。次いで、ディスクアダプタ（DKA5）80はキャッシュメモリ60に読み出されたライトデータ610を論理ボリューム（Data2）200に書き込むことにより、論理ボリューム（Data2）200のデータ更新を行う（S403）。次いで、論理ボリューム（Data2）200のデータ更新に反映された更新情報620とライトデータ610とが存在していた記憶領域は開放される（S404）。チャンネルアダプタ（CHA6）50は引き続き正規化処理を行うか否かを判定し（S405）、処理を継続する場合には（S405；YES）、S401に戻る。

20

30

【0062】

図16はフェイルオーバー発生時に第1の記憶システム10と第3の記憶システム20のデータイメージを同期化させるための処理手順を記述したフローチャートである。例えば、第1の記憶システム10に障害が発生すると、第1の記憶システム10はアプリケーションプログラム31からの入出力要求に応答できなくなる。アプリケーションプログラム31はリトライを試みるが、入出力要求に失敗し、ダウンする。すると、クラスタソフトウェア32は障害の発生を検出し、交替系の起動指示を送る。交替系のクラスタソフトウェア42は、現用系のクラスタソフトウェア32から起動指示を受信すると、リソースグループ41を起動させる（S501）。これにより、起動スクリプトが実行される（S502）。起動スクリプトが実行されると、まず、正副切り替え処理（horctakeoverコマンド）が実行される（S503）。

40

【0063】

正副切り替え処理においては、正論理ボリュームとしての論理ボリューム（Data1）150と、副論理ボリュームとしての論理ボリューム（Data2）200との間のペアのステータスが一時的にサスペンド状態にされる。この状態下において、未送信のジャーナルデータ950が第2の記憶システム15から第3の記憶システム20に送信され、論理ボリューム（Data2）200のデータ更新が行われる。未送信のジャーナルデータ950が第2の記憶システム15にどの程度残存しているかは、第3の記憶システム2

50

0 から第 2 の記憶システム 15 に問い合わせを行うことで把握できる。より詳細には、記憶装置管理ソフトウェア 41 b が第 3 の記憶システム 20 のコマンドデバイス 60 - 3 C にコマンド (ジャーナルデータ 950 の残存量を第 2 の記憶システム 15 に問い合わせるためのコマンド) を書き込むと、チャンネルアダプタ (CHA5) 50 が第 2 の記憶システム 15 に問い合わせを行う。

#### 【0064】

このようにして、論理ボリューム (Data1) 150 のデータイメージと、論理ボリューム (Data2) 200 のデータイメージが同期化されると (P-S 同期化)、論理ボリューム (Data2) 200 を正論理ボリュームに切り替え、論理ボリューム (Data1) 150 を副論理ボリュームに切り替える処理が行われる (P-S スワップ処理)。通常、副論理ボリュームに対するライトアクセスは禁止されているので、論理ボリューム (Data2) 200 を正論理ボリュームに切り替えることによって、ホストコンピュータ 40 から論理ボリューム (Data2) 200 へのライトアクセスが可能になる。このようにして、正副切り替え処理が完了すると、記憶装置管理ソフトウェア 41 b はファイルシステムが破損していないか否かをチェックし (S504)、ファイルシステムが正常であることを確認して、ファイルシステムをマウントし (S505)、アプリケーションプログラム 41 a を起動させる (S506)。これによりホストコンピュータ 40 は第 3 の記憶システム 20 を使用して、フェイルオーバー時にホストコンピュータ 30 が行っていた処理を引き継ぐことができる。

#### 【0065】

次に、第 3 の記憶システム 20 に障害が発生した場合のデータの二重化について説明する。本実施例では、第 2 の記憶システム 15 の論理ボリューム (Data1) 150 は実ボリュームを持たない仮想ボリュームである。第 3 の記憶システム 20 に障害が発生すると、実データは第 1 の記憶システム 10 のみに残されるので、データの二重化を行って耐障害性を高めるように構成するのが好ましい。第 3 の記憶システム 20 に障害が発生すると、図 17 に示すように、第 2 の記憶システム 15 は自動的に、或いはユーザの手動設定によって、物理ボリューム 900 上に論理ボリューム (Data1') をアサインする。論理ボリューム (Data1') は第 2 の記憶システム 15 が提供する記憶領域を第 1 の記憶システム 10 から指定させるためのアドレスを有する実ボリュームである。論理ボリューム (Data1') と論理ボリューム (ORG1) 110 を同期化させるために、まず、論理ボリューム (ORG1) 110 と論理ボリューム (Data1) 150 との間のペアのステータスをサスペンド状態にし、論理ボリューム (ORG1) 110 から論理ボリューム (Data1') にイニシャルコピーを行う。

#### 【0066】

この間にホストコンピュータ 30 から論理ボリューム (ORG1) 110 に対して加えられたデータ更新は差分情報ビットマップとして蓄積される。論理ボリューム (ORG1) 110 から論理ボリューム (Data1') へのイニシャルコピーが完了したならば、差分情報ビットマップに基づいて論理ボリューム (Data1') のデータ更新を行う。このようにして、論理ボリューム (ORG1) 110 と論理ボリューム (Data1') とが同期化すると、両者のペアのステータスをペア状態に設定する。すると、論理ボリューム (ORG1) 110 に加えられたデータ更新は論理ボリューム (Data1') にも反映され、データの二重化を行うことができる。

#### 【0067】

尚、第 3 の記憶システム 20 に障害が発生したか否かを判定する上で、例えば、第 1 の記憶システム 10 にあるコマンドデバイス 60 - 1 C と、第 2 の記憶システム 15 にあるコマンドデバイス 60 - 2 C を用いることができる。ホストコンピュータ 30 は、第 2 の記憶システム 15 が正常に作動しているか否かを第 1 の記憶システム 10 に確認させるためのコマンドをコマンドデバイス 60 - 1 C に書き込む。このコマンドデバイス 60 - 1 C への書き込みを受けて、第 1 の記憶システム 10 は第 2 の記憶システム 15 が正常に作動しているか否かを、相互通信を通じて確認する。また、第 1 の記憶システム 10 は、第

10

20

30

40

50

3の記憶システム20が正常に作動しているか否かを第2の記憶システム15に確認させるためのコマンドをコマンドデバイス60-2Cに書き込む。このコマンドデバイス60-2Cへの書き込みを受けて、第2の記憶システム15は第3の記憶システム20が正常に作動しているか否かを、相互通信を通じて確認する。

【0068】

次に、既述のコマンドデバイスについてさらに説明する。第1乃至第3の前記コマンドデバイス60-1C, 60-2C, 60-3Cには、既述の論理ボリュームのペア設定やペア解除などの制御コマンドも設定される。制御コマンドは、第1のサイトのホスト計算機30によって作成され、第2のサイトのコマンドデバイス60-2Cと、第3のサイトのコマンドデバイス60-3Cに対して発行される。このコマンドデバイスが通信されるルートには二通りあり、第1がIPネットワーク『通信回線』310を介して、第1のサイトのホストコンピュータ30から第2のサイトに接続するホストコンピュータ40A(ホストコンピュータ30とほぼ同じ構成を有する。)へ至るルートと、ホストコンピュータ30から第3のサイトのホストコンピュータ40へ至るルートである。第2が、ホストコンピュータ30から、第1の記憶制御装置10の記憶領域に形成された仮想ボリュームと記憶制御装置間の直接接続330を介して、制御コマンドをコマンドデバイス60-2C及び/または60-3C発行するルートである。

【0069】

図17は、コマンドデバイスの通信ルートを示した、記憶制御システムのブロック図である。矢印1000はIPリンクを介したホスト装置間の制御コマンドの通信を示す。矢印2は、第1の記憶制御装置10に形成された仮想ボリューム600を介する制御コマンドの通信を示す。この仮想ボリュームに第2の記憶制御装置Bのコマンドデバイス60-2Cが割り当てられている。なお、仮想ボリュームには既述のようにデータボリュームがマッピングされたものと、ここで説明した、コマンドデバイスがマッピングされた仮想ボリュームがある。後者の仮想ボリュームを前者の仮想ボリュームから区別するために、以後「リモートコマンドデバイス」と便宜上称することがある。なお、再度説明するが、仮想ボリュームとは、ホストコンピュータによって認識される論理ボリュームであるが、記憶の実体は記憶制御装置に無く、この記憶制御装置に接続する外部記憶制御装置に存在する。ホストコンピュータ30が仮想ボリューム600にアクセスすると、ホストコンピュータは、第2サイトの記憶制御装置15の論理デバイス(コマンドデバイ)60-2Cを認識する。

【0070】

ホストコンピュータ30は、第2サイトの記憶制御装置15のコマンドデバイスに対する制御コマンドを形成し、これを仮想ボリューム600に対して発行すると(矢印1002)、第1の記憶制御装置と第2の記憶制御装置とを直結する通信ルート330を介して、制御コマンドが第2の記憶制御装置のコマンドデバイス60-2Cに設定される。第2の記憶制御装置15のCHAのコントローラは、この制御コマンドを処理して、ボリューム150が第3の記憶制御装置20のボリューム200とペアを形成するように、共有メモリ内の制御テーブルを更新する。仮想ボリュームに対して発行された制御コマンドはキャッシュメモリを介して、記憶制御装置間で伝達される。

【0071】

ホストコンピュータが制御コマンドを第2の記憶制御装置のコマンドデバイスに設定するルートには既述のとおり二通りあるが、通常はどちらかのルートを主とし、他方を交代パスとすれば良い。図17に基づいて、一例を述べると、ホストコンピュータ30と同40間ルート310を主とし、仮想ボリューム600を介するルートを第2のルートを交代パスとする。その逆であっても良い。ホストコンピュータ間の通信異常を、ホストコンピュータ30が検出すると、ホストコンピュータは制御コマンドを、仮想ボリューム600を介してコマンドデバイス60-2Cに送信する。なお、制御コマンドには論理ボリュームのペア設定の他、論理ボリュームのペア解除、或いはペア状態の参照などがある。

【0072】

10

20

30

40

50

図18は、第1の記憶制御装置10の共有メモリ70に設定される、コマンドデバイスマッピングテーブルを示したものである。LUNの覧には、記憶制御装置の仮想ボリュームのLUNが設定される。装置の覧には、コマンドデバイスがどの記憶制御装置のものであるかが登録される。コマンドデバイスLUNの覧には、コマンドデバイスの実体を表す論理デバイスのLUNが設定されている。コマンドデバイスLUNには、第2サイトのコマンドデバイス及び/又は第3の記憶制御装置のコマンドデバイスのLUNが登録される。図18においては、第1の記憶制御装置の仮想ボリュームLUN番号0に第2の記憶制御装置15のコマンドデバイス(60-2C)LUN番号0が対応していること、第1の記憶制御装置の仮想ボリュームLUN番号1に第3の記憶制御装置20のコマンドデバイス(60-3C)LUN番号1が対応している事を示している。

10

**【0073】**

さらに、共有メモリには、仮想ボリュームが外部記憶制御装置のコマンドデバイスにマッピングされた、「リモートコマンドデバイス」であることを、ホストコンピュータに認識させるための、リモートコマンドデバイス管理テーブルも登録されている。このテーブルには、リモートコマンドデバイス(仮想ボリューム)のLUN番号と、これにマッピングされる、外部コマンドデバイスの属性とが記憶されている。図19はこのテーブルの概要を示したものである。コマンドデバイスの属性情報は、コマンドデバイスへのパス情報とコマンドデバイスの記憶容量を含み、ビット情報は、仮想ボリュームがコマンドデバイスへマッピングされたリモートコマンドデバイスであるか否かを示す第1の属性ビット、コマンドデバイスの属性情報として必要な情報がテーブルに取り込まれたか否かを示す第2の属性ビットを含んでいる。また、パス情報は、各通信ポートに固有な識別情報(WWN)と、コマンドデバイスを識別するためのLUN番号とを含んで構成できる。

20

**【0074】**

図20は、コマンドデバイスを第1の記憶制御装置10の仮想ボリュームにマッピングする際に必要な、前記マッピングテーブルを作成するための、第1の記憶制御装置のチャネルアダプタのコントローラが実行する、タイミングフローである。まず、図1に示すように、第1の記憶制御装置10は、チャネルアダプタCHA2のイニシエータポートを介して、第2の記憶制御装置15及び/又は第3の記憶制御装置20にログインする(S1)。外部の記憶制御装置が、第1の記憶制御装置10のログインに対して応答を返すことにより、ログインが完了する(S2)。次に、第1の記憶制御装置10は、例えば、SCSI (Small Computer System Interface)規格で定められている照会コマンド(inquiryコマンド)を、外部の記憶制御装置に送信し、外部の記憶制御装置の有する論理デバイスの詳細について応答を求める(S3)。

30

**【0075】**

照会コマンドは、照会先の装置の種類及び構成を明らかにするために用いられるもので、照会先装置の有する階層を透過して外部装置の論理デバイスの物理的構造を把握することができる。照会コマンドを使用することにより、第1の記憶制御装置100は、例えば、装置名、デバイスタイプ、製造番号(プロダクトID)、論理ボリューム番号、各種バージョン情報、ベンダID等の情報を外部の記憶制御装置から取得できる(S4)。外部の記憶制御装置40は、問合せされた情報を第1の記憶制御装置に送信し、応答する(S5)。次いで、第1の記憶制御装置10は、仮想ボリュームを各コマンドデバイスに自動的に割り当て、既述のリモートコマンド制御情報テーブルと前記マッピングテーブルに必要な情報を登録して(S6)、さらにリモートコマンド制御情報テーブルに対して、仮想ボリュームがリモートコマンドデバイスとして登録されたことを示すビットをオンする。これにより、ホストコンピュータ30は仮想ボリュームにアクセスすることによって、外部記憶制御装置のコマンドデバイスを特定して認識することができ、記憶制御装置間のバスを介して、外部記憶制御装置を制御するための制御コマンドを当該外部記憶制御装置を制御することができる。第1の記憶制御装置10がリモートコマンドデバイス600(図17)にマッピングされた外部の記憶制御装置のコマンドデバイス60-2Cとの間でデータの通信を行う場合は必要なアドレス変換を行う。

40

50

## 【 0 0 7 6 】

次に、ホストコンピュータが外部の記憶制御装置のコマンドデバイスへリモートコマンドデバイス経由で制御コマンドを発行する際のフローチャートを図 17 に基づいて説明する。図 21 はこのフローチャートを示したものである。ホストコンピュータ 30 第 1 の記憶装置 10 に照会コマンドを発行する ( 2 1 0 0 ) と、チャンネルアダプタのコントローラは、共有メモリ 70 ( 図 2 ) の既述のリモートコマンドデバイス属性情報テーブル ( 図 19 ) の内容を読み込み、これをホストコンピュータに返信する ( 2 1 0 2 )。

## 【 0 0 7 7 】

ホストコンピュータは、リモートコマンドデバイス 600 を外部の記憶制御装置 15 のコマンドデバイス 60 - 2 C として認識するための必要な情報を自身の記憶領域に登録する ( 2 1 0 4 )。これによって、ホストコンピュータはユーザ ( 例えば、ホストコンピュータに接続するクライアント装置 ) に対して、外部記憶制御装置のコマンドデバイスへのアクセスを可能にする。ホストコンピュータ 30 がコマンドデバイス 60 - 2 C を特定して、このコマンドデバイスに制御コマンドを発行すると、第 1 の記憶制御装置 10 は、前記リモートコマンドデバイス制御情報テーブルを参照して、ホストコンピュータを仮想ボリューム 600 にアクセスさせ ( 2 1 0 6 )、制御コマンドを仮想ボリュームに対応するキャッシュメモリの所定領域に書き込む。

## 【 0 0 7 8 】

第 1 の記憶制御装置 10 は、前記マッピング情報テーブルを参照して、リモートコマンドデバイスに発行された制御コマンドを対応するコマンドデバイス 60 - 2 C に、記憶制御装置同士の筐体間通信を介して送信する ( 2 1 0 8 )。次いで、外部記憶制御装置 15 の C H A はコマンドデバイス 60 - 2 C に設定された制御コマンドを処理し、結果を第 1 の記憶制御装置を経由してホスト装置に送信して一連の動作を終了する。なお、第 1 の記憶制御装置は、前記リモートコマンド制御情報テーブルやマッピングテーブルの更新を適宜行なうことにより、最新の情報を共有メモリに登録することができる。ホストコンピュータは、前記照会コマンドの発行を適時行なうことにより、この最新情報を得ることができる。また、第 1 の記憶制御装置がホストコンピュータからコマンドが発行されると、第 1 の記憶制御装置の C H A は、前記外部記憶制御装置のコマンドデバイスに対するコマンドか否かを判定し、これが肯定された場合に図 21 のフローチャートが実行される。データボリュームに対する I O コマンドの場合は、通常の I O 処理を実行する。また、リモートコマンドデバイスが登録されていない場合や記憶制御装置の筐体間通信に障害がある場合には、制御コマンドを目的とするコマンドデバイスに発行できないこと、第 1 の記憶制御装置がホストコンピュータに通知する。この場合、ホストコンピュータ 30 は、ホストコンピュータ間の I P リンクを経由して、制御コマンドを外部記憶制御装置に接続するホストコンピュータに送信することができる。また、ホストコンピュータ同士の通信が障害を受けている場合に、図 21 のフローチャートが実行されても良い。

## 【 0 0 7 9 】

図 22 は、図 17 の変形例であって、ホストコンピュータが第 1 の記憶制御装置 10 のみに接続されている例である。ホストコンピュータ - ホストコンピュータ間で制御コマンドを通信できないが、ホストコンピュータ 30 から第 2 の記憶制御装置 15 へ、第 1 の記憶制御装置のリモートコマンドデバイス 600 を経由することにより可能となる ( 矢印 1 0 0 4 )。この例では、ホスト装置 - ホスト装置間の I P リンクをホスト装置間のバスに統合することにより、I P リンクを削減することができる。

## 【 0 0 8 0 】

図 23 は、さらに他の変形例であって、第 1 の記憶制御装置 10 のホストコンピュータ 30 と第 2 の記憶制御装置 15 のホストコンピュータ 40 A との間に I P リンクが形成され、第 1 の記憶制御装置のホストコンピュータ 30 と第 3 の記憶制御装置 20 のホストコンピュータ 40 との間にも I P リンクが形成されている。さらに、第 1 の記憶制御装置 10 と第 2 の記憶制御装置 15 間はバス接続 B 1 され、第 2 の記憶制御装置 15 と第 3 の記憶制御装置 20 もバス接続 B 2 されている。第 1 の記憶制御装置 10 には、第 2 の記憶

10

20

30

40

50

制御装置のコマンドデバイスAに対応する仮想ボリュームA'と、第3の記憶制御装置のコマンドデバイスBに対する仮想ボリュームB'が形成されている。第3の記憶制御装置には第2記憶制御装置のコマンドデバイスBにマッピングされた仮想ボリュームB'が形成されている。

【0081】

今、図24に示すように、第1の記憶制御装置10と第2の記憶制御装置間15での、ホスト装置間のIPリンク及びバス接続の通信が損なわれた場合(240)、第2の記憶制御装置15のSvolをサスペンドする必要があるが、第1の記憶制御装置10及びホストコンピュータ30から第2の記憶制御装置へ直接サスペンドのためのコマンド発行できない。この場合、図25に示されるように、ホストコンピュータ30は第1の記憶制御装置の仮想ボリュームB'を介して第3の記憶制御装置20のコマンドデバイスBにサスペンドコマンドを発行する。又は、ホストコンピュータ30-ホストコンピュータ40間のIPリンクを介して、コマンドデバイスBにサスペンドコマンドを発行することができる。第3の記憶制御装置20は、コマンドデバイスBの制御コマンドを第2の記憶制御装置15の仮想ボリュームA'を介してコマンドデバイスAに対して発行する。第3の記憶制御装置20は、自身の論理ボリュームを正論理ボリューム(Pvol)として第2の記憶制御装置15の副ボリューム(Svol)との間でペアを形成して、リモートコピーシステムの運用を再開する(図26)。

10

【0082】

図27はさらに他の形態を示すものであり、第2の記憶制御装置15に第1の記憶制御装置10の仮想ボリュームA''にマッピングされた仮想ボリュームA'を備えているシステムである。第2の記憶制御装置15の仮想ボリュームA'は第3の記憶制御装置20のコマンドデバイスAにマッピングされている。したがって、ホストコンピュータはこれら仮想ボリュームを介して第3の記憶制御装置20のコマンドデバイスAにデータボリューム270と272とのペア操作要求などの制御コマンドを発行できる。このように仮想ボリュームをカスケード接続することにより、2つ以上離れた外部記憶制御装置にホスト装置からの要求を発行することができる。

20

【0083】

図28は、複数の記憶制御装置同士A-DをN×M(2×2)にバス接続することにより、ホスト装置を設ける事無く、或る記憶制御装置から所望の記憶制御装置に制御コマンドを発行できる例を示したものである。各記憶制御装置はそれぞれコマンドデバイスと、他の全ての記憶制御装置のコマンドデバイスにそれぞれマッピングされた仮想ボリュームを複数備えている。「'」がアルファベットに付されている符号が、このアルファベット(コマンドデバイス)に対する仮想ボリュームである。

30

【0084】

たとえば、記憶制御装置Aには、この装置のコマンドデバイスと他の全ての記憶制御装置毎のコマンドデバイスに対応する仮想ボリュームが構成されている。今装置Aを装置Cに対するホスト装置とすると、装置AのCHAのコントローラは、コマンドデバイスAに制御コマンドを登録する。CHAはコマンドデバイスAの制御コマンドを仮想ボリュームC'を介して装置CのコマンドデバイスCに発行することができる。装置Cは第1の記憶制御装置として、第2の記憶制御装置である装置BのコマンドデバイスB又は装置DのコマンドデバイスDに、仮想ボリュームB'又はD'を介して、制御コマンドを発行することができる。

40

【0085】

次に、第1の記憶制御装置10に接続するホストコンピュータ30がユーザに、外部記憶制御装置のコマンドデバイスの存在を表示する実施形態について説明する。図29は記憶制御システムのブロック図であり、図30はホストコンピュータ又はホストコンピュータに接続するクライアントの画面表示例である。図29においてAは、第1の記憶制御装置10のコマンドデバイス、Baは第2の記憶制御装置のコマンドデバイス、Bbは第2の記憶制御装置15のデータボリュームである。Cは第3の記憶制御装置20のコマンド

50

デバイスである。C´は第1の記憶制御装置10内に設定された、コマンドデバイスCにマッピングされた仮想ボリュームであり、Ba´はコマンドデバイスBaにマッピングされた仮想ボリュームであり、Bb´はデータボリュームBbにマッピングされた仮想ボリュームである。第1の記憶制御装置10の各ボリュームにはLUN番号(LU )が設定されている。

【0086】

図30は、ホストコンピュータ30がアプリケーションプログラムによって、から、自ホスト配下に接続されているボリュームを確認するコマンド(たとえばOSがHP-UXであればioscanというコマンド)を実行した場合に、ホストコンピュータ30に表示されるイメージ図である。ボリュームAは、自ホスト内のデータボリュームであるので、LU 0(ボリュームA)には、第1の記憶制御装置のシリアル番号が表示される。LU 1(ボリュームBa´)は、第2の記憶制御装置15のコマンドデバイスBaにマッピングされた仮想ボリュームであるので、第2の記憶制御装置10のシリアル番号が表示される。LU 2についても同様に、第3の記憶制御装置15のシリアル番号が表示される。LU 3は、第2の記憶制御装置10のデータボリュームBbに対する仮想ボリュームBb´であるので、ホストコンピュータ30は、この仮想ボリュームを自ホストのボリュームとして認識する。

10

【0087】

以上説明した実施形態では、主として3基の記憶制御装置を接続したリモートコピーシステムに関するものだったが、2基の記憶制御装置同士を接続するシステムについても本発明を適用することができる。

20

【図面の簡単な説明】

【0088】

【図1】実施例1に係わるリモートコピーシステムの構成図である。

【図2】第1の記憶システムの構成図である。

【図3】第2の記憶システムの構成図である。

【図4】第3の記憶システムの構成図である。

【図5】ボリューム情報テーブルの説明図である。

【図6】ペア設定情報テーブルの説明図である。

【図7】ジャーナルグループ設定情報テーブルの説明図である。

30

【図8】ジャーナルデータの説明図である。

【図9】初期化設定処理のフローチャートである。

【図10】アクセスコマンド受信処理の説明図である。

【図11】アクセスコマンド受信処理のフローチャートである。

【図12】ジャーナルリードコマンド受信処理の説明図である。

【図13】ジャーナルリードコマンド受信処理のフローチャートである。

【図14】正規化処理の説明図である。

【図15】正規化処理のフローチャートである。

【図16】第2の記憶システムの構成図である。

【図17】第1の記憶制御装置の仮想ボリュームに第2の記憶制御装置のコマンドデバイスをマッピングさせた状態を説明する、記憶制御システムのブロック図である。

40

【図18】仮想ボリュームとコマンドデバイスとのマッピングテーブルである。

【図19】コマンドデバイスにマッピングされ仮想ボリュームの制御情報テーブルの一例である。

【図20】テーブル作成のための、第1の記憶制御装置の動作を示すタイミングフローである。

【図21】ホストコンピュータが制御コマンドを発行する際のフローチャートである。

【図22】図17の変形例を示すブロック図である。

【図23】さらに他の変形例を示すブロック図である。

【図24】図23のシステムにおける制御コマンドの送信処理動作の一過程を示すブロッ

50

ク図である。

【図25】さらに他の過程を示すブロック図である。

【図26】さらに他の過程を示すブロック図である。

【図27】図17の変形例を示すブロック図である。

【図28】さらに他の変形例を示すブロック図である。

【図29】第1の記憶制御装置に接続するホストコンピュータがユーザに、外部記憶制御装置のコマンドデバイスの存在を表示する実施形態に係るブロック図である。

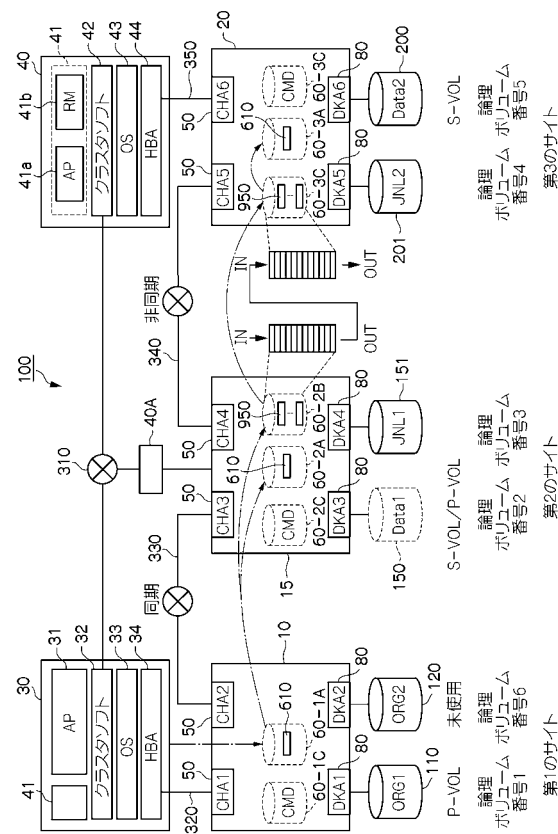
【図30】ホストコンピュータ又はホストコンピュータに接続するクライアントの画面表示例である。

【符号の説明】

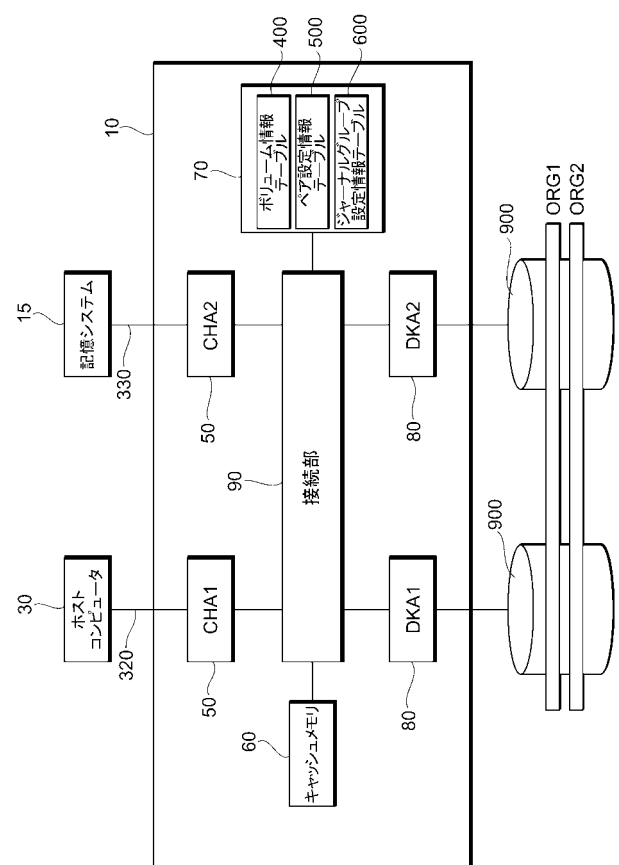
【0089】

10 ... 第1の記憶システム 15 ... 第2の記憶システム 20 ... 第3の記憶システム 30 ... ホストコンピュータ 40 ... ホストコンピュータ 100 ... リモートコピーシステム

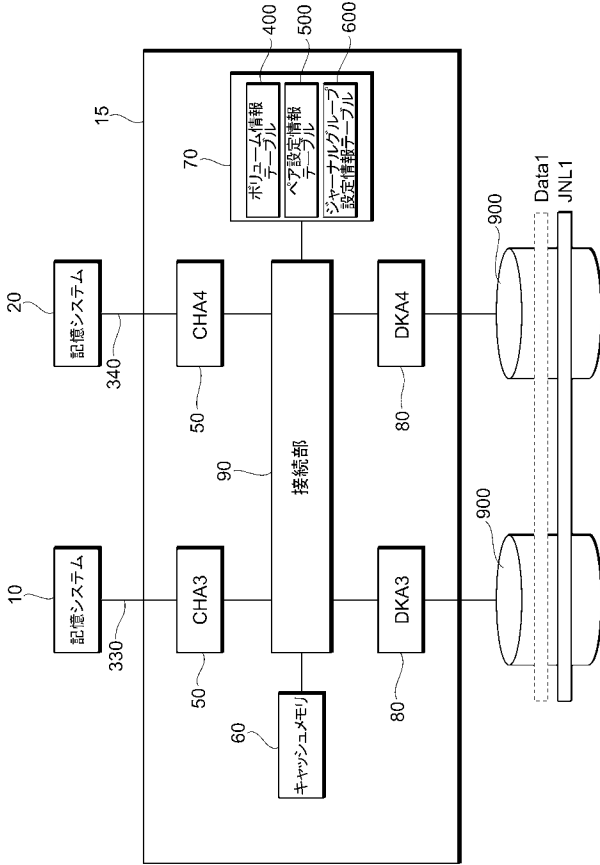
【図1】



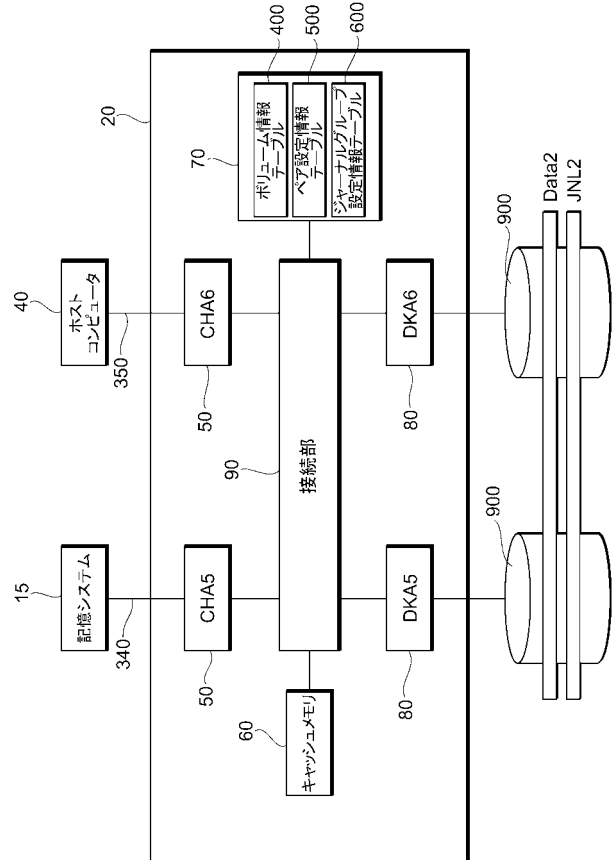
【図2】



【 図 3 】



【 図 4 】

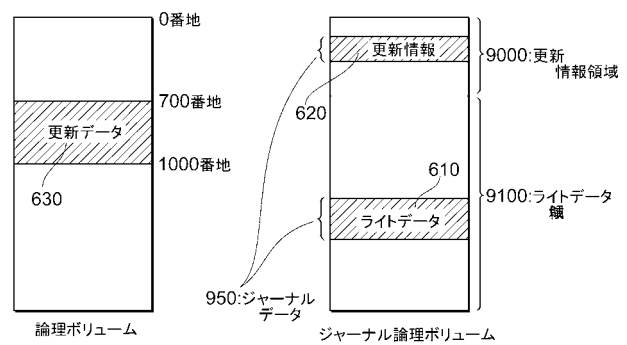


【 図 5 】

ボリューム情報テーブル

論理 ボリューム番号	ボリューム 状態	フォーマット 形式	容量	ペア 番号	物理アドレス	
					記憶 システム番号	先頭からの 位置
1	正	OPEN3	6	1	1	0
2	副	OPEN3	6	1	2	3
3	正常	OPEN3	3	-	2	9
4	正常	OPEN3	3	-	3	0
5	正常	OPEN3	6	-	3	9
6	未使用	OPEN6	6	-	1	12

【 図 8 】

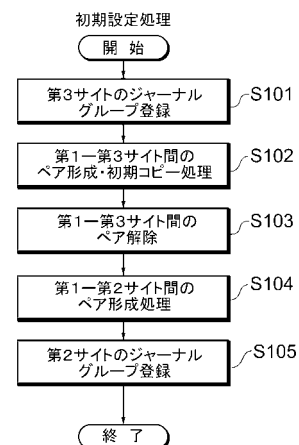


【 図 6 】

ペア設定情報テーブル

ペア番号	ペア状態	正論理 ボリューム 番号	副論理 ボリューム 番号	仮想化
1	正常	1	2	ON
2	正常	2	5	
3	未使用	0	0	
4	未使用	0	0	
5	未使用	0	0	

【 図 9 】

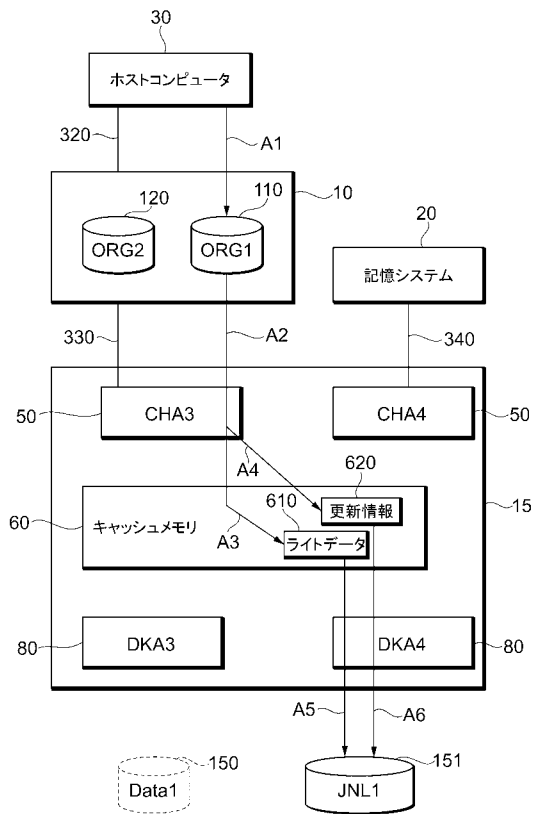


【 図 7 】

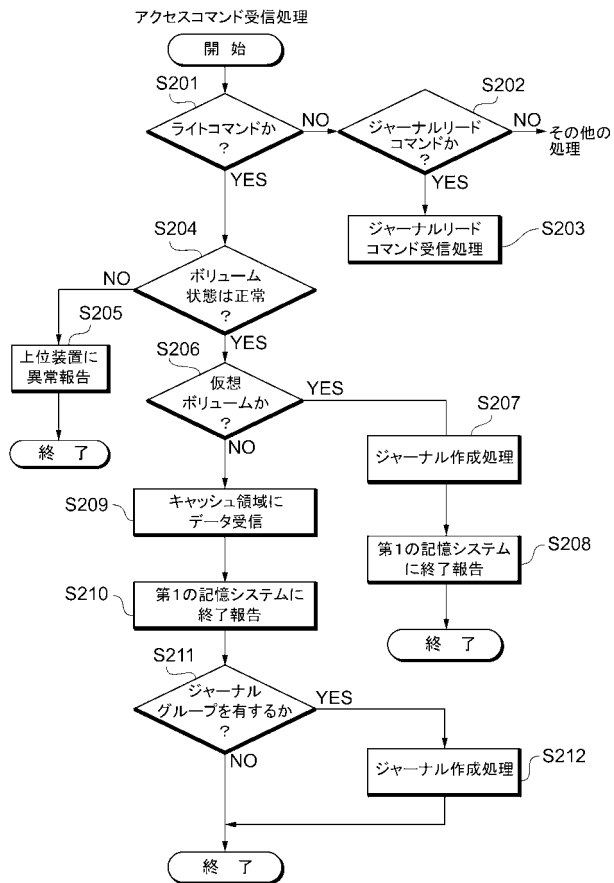
ジャーナルグループ設定情報テーブル

ジャーナル グループ 番号	ペア状態	ソース ボリューム 番号	ジャーナル ボリューム 番号
1	正常	2(Data1)	3(JNL1)
2	正常	5(Data2)	4(JNL2)

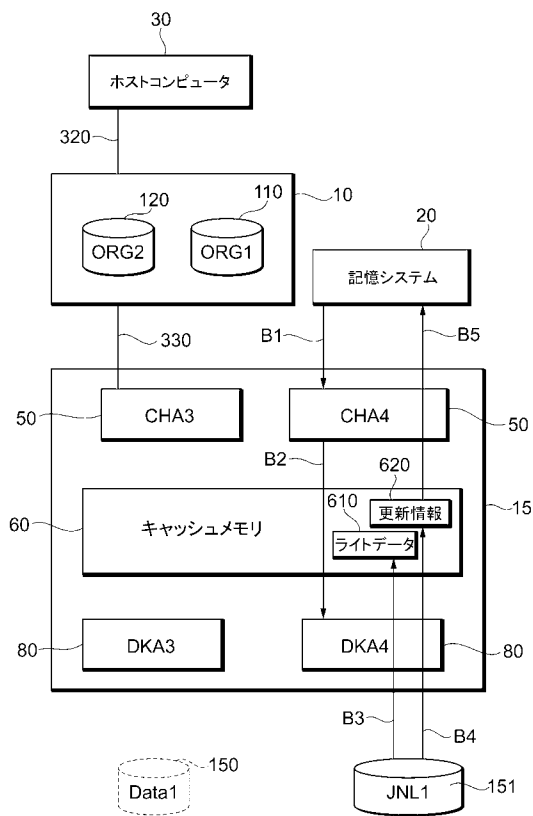
【図10】



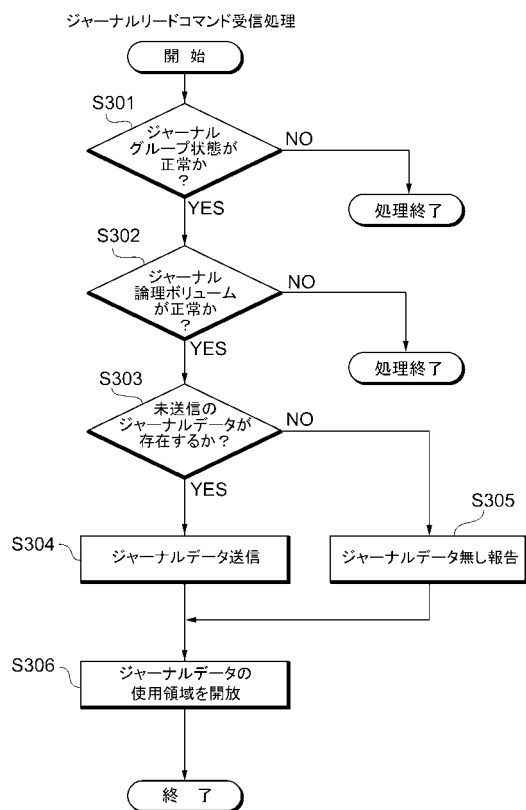
【図11】



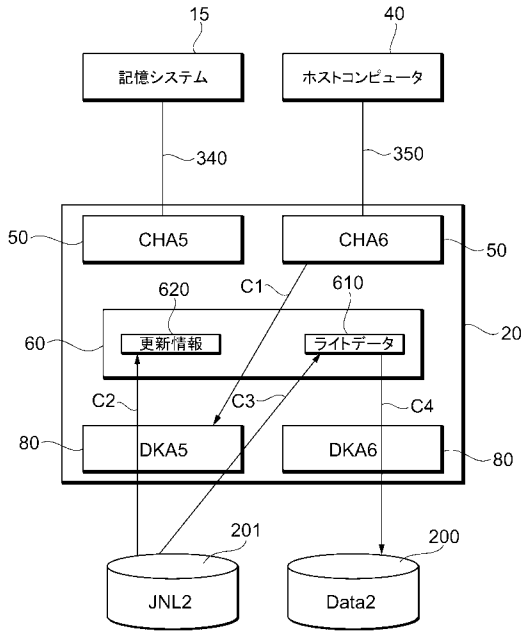
【図12】



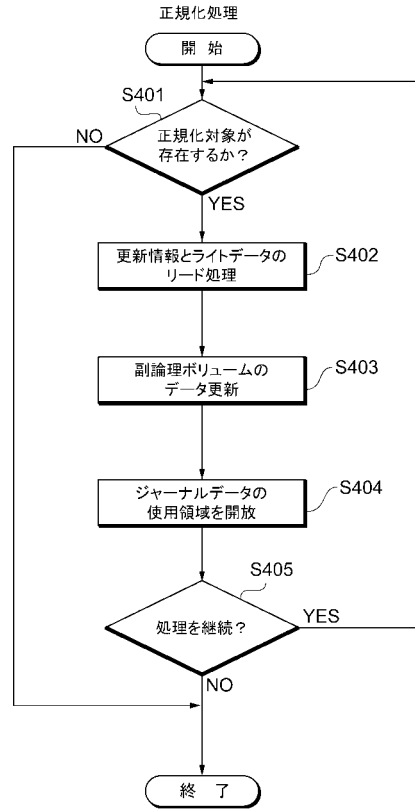
【図13】



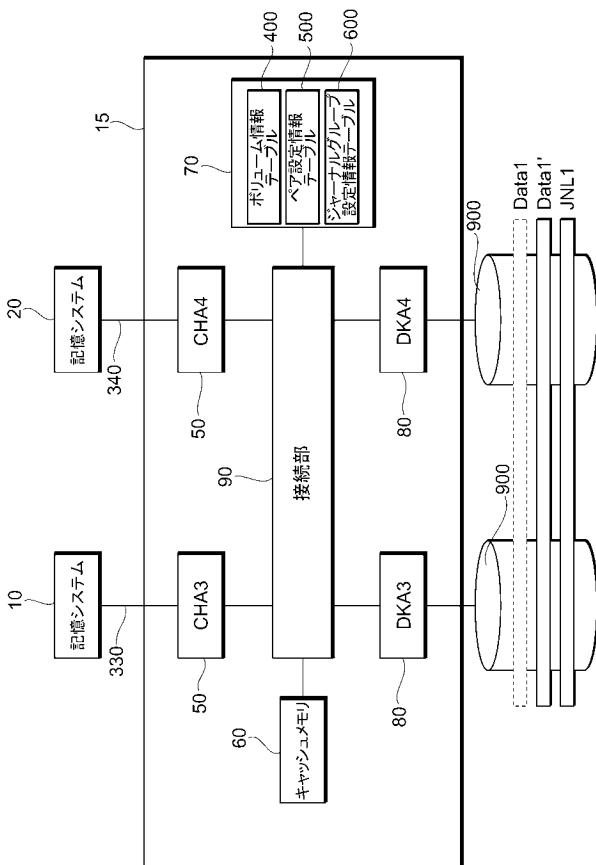
【 図 1 4 】



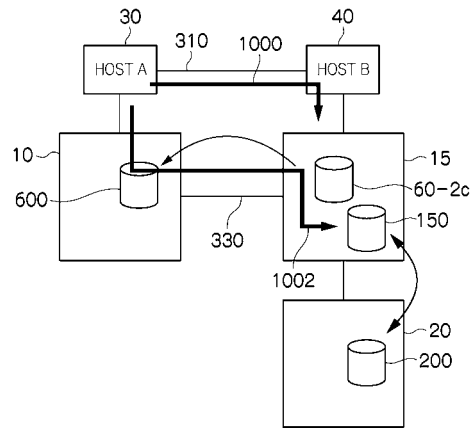
【 図 1 5 】



【 図 1 6 】



【 図 1 7 】



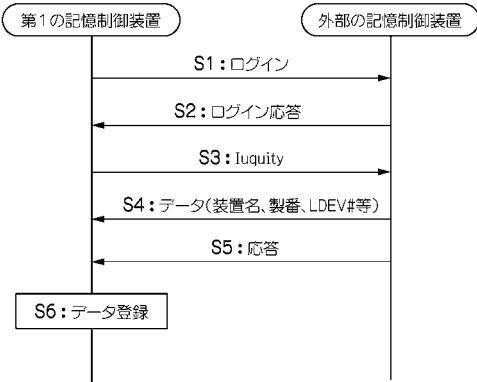
【 図 1 8 】

LUN	装置	コマンドデバイスLUN
0	第2	0
1	第3	1
...	...	...

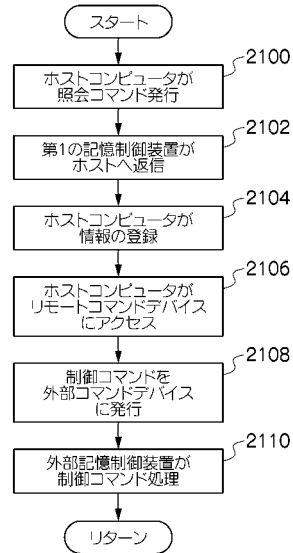
【 図 1 9 】

LUN	コマンドデバイスの属性情報	bit情報
0	* * * *	* * *
⋮	⋮	⋮

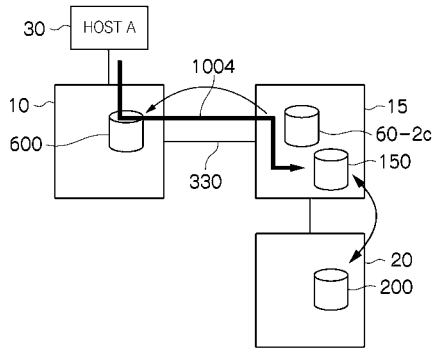
【 図 2 0 】



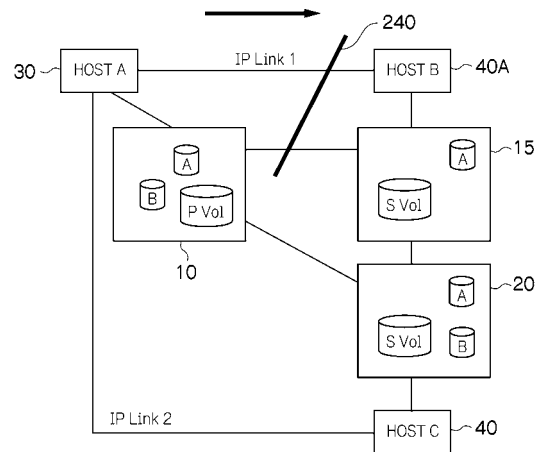
【 図 2 1 】



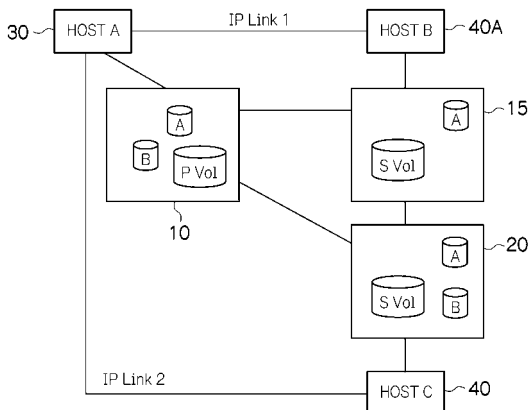
【 図 2 2 】



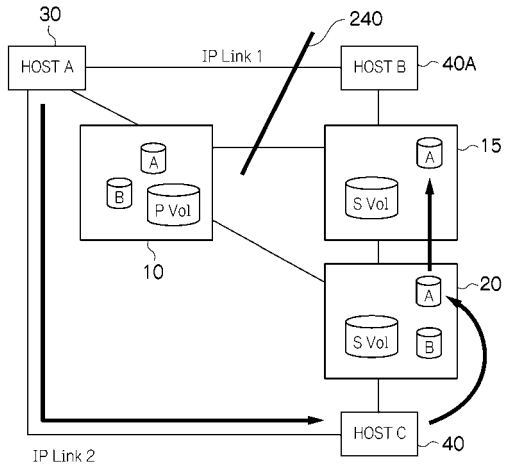
【 図 2 4 】



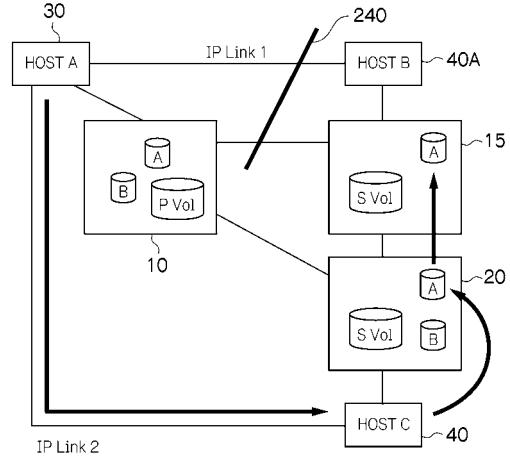
【 図 2 3 】



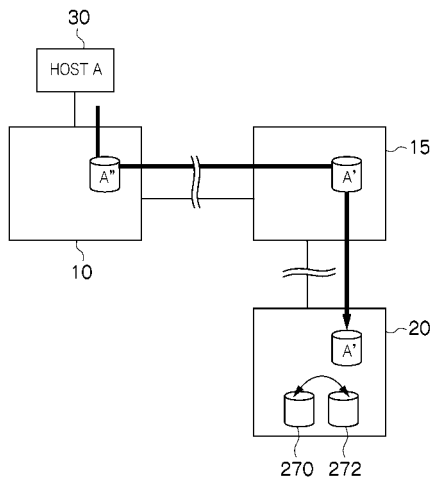
【 図 2 5 】



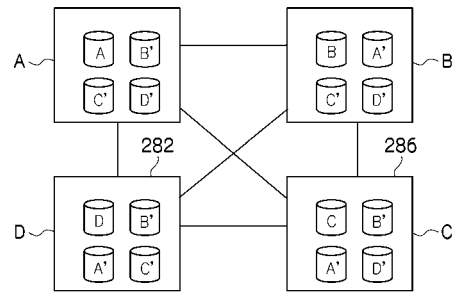
【 図 2 6 】



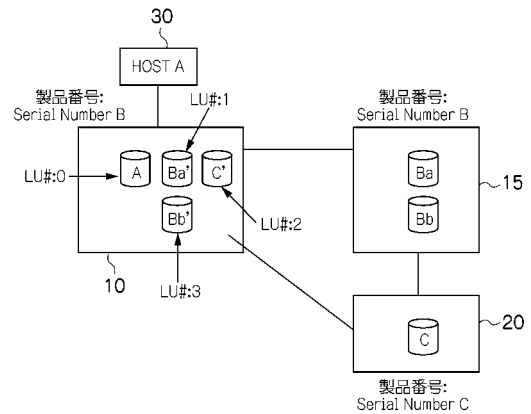
【 図 2 7 】



【 図 2 8 】



【 図 2 9 】



【 図 3 0 】

LU#	製品番号
0	Serial Number A
1	Serial Number B
2	Serial Number C
3	Serial Number A

---

フロントページの続き

- (72)発明者 占部 喜一郎  
神奈川県小田原市中里3 2 2番2号 株式会社日立製作所RAIDシステム事業部内
- (72)発明者 中川 義仁  
神奈川県小田原市中里3 2 2番2号 株式会社日立製作所RAIDシステム事業部内
- (72)発明者 田村 圭史  
神奈川県小田原市中里3 2 2番2号 株式会社日立製作所RAIDシステム事業部内
- Fターム(参考) 5B065 BA01 CA11 EA33 ZA01