(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5057366号
(P5057366)
(45)発行日　平成24年10月24日(2012.10.24)　(24)登録日　平成24年8月10日(2012.8.10)

(51)Int.Cl.　F I
G06F 3/06 (2006.01)　G06F 3/06 305F
G06F 3/06 301Z

請求項の数 4　(全 72 頁)

(21)出願番号　特願2007-85680(P2007-85680)
(22)出願日　平成19年3月28日(2007.3.28)
(65)公開番号　特開2008-134987(P2008-134987A)
(43)公開日　平成20年6月12日(2008.6.12)
審査請求日　平成21年6月26日(2009.6.26)
(31)優先権主張番号　特願2006-293485(P2006-293485)
(32)優先日　平成18年10月30日(2006.10.30)
(33)優先権主張国　日本国(JP)

(73)特許権者　000005108
株式会社日立製作所
東京都千代田区丸の内一丁目６番６号
(74)代理人　100093861
弁理士　大賀　眞司
(74)代理人　100129218
弁理士　百本　宏之
(72)発明者　岩村　卓成
神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所システム開発研究所内
(72)発明者　二瀬　健太
神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所システム開発研究所内

最終頁に続く

(54)【発明の名称】情報システム及び情報システムのデータ転送方法

(57)【特許請求の範囲】
【請求項１】
ファイルシステム、ＨＢＡデバイスドライバーを備えるオペレーティングシステム、及び前記ファイルシステムを通じてライトリクエストを送信するアプリケーションを備えるホストコンピューターと、前記ホストコンピューターに接続される第１のストレージ装置と、前記第１のストレージ装置及び前記ホストコンピューターに接続される第２のストレージ装置と、前記第１のストレージ装置及び前記第２のストレージ装置に接続され、複数の第１のＨＤＤとコントローラーとキャッシュメモリから構成され、前記複数の第１のＨＤＤの一部の領域から構成される第１のボリュームを有する第３のストレージ装置と、を有する情報システムのデータ転送方法であって、
前記第１のストレージ装置は、前記第３のストレージ装置の前記第１のボリュームに実体が存在する第１の仮想ボリュームを定義し、
前記第２のストレージ装置は、前記第３のストレージ装置の前記第１のボリュームに実体が存在し、当該第１のボリュームを前記第１の仮想ボリュームと共有する第２の仮想ボリュームを定義し、
前記第１のストレージ装置及び前記第２のストレージ装置は、それぞれ複数の第２のＨＤＤとコントローラーとキャッシュメモリとを備え、それぞれ前記複数の第２のＨＤＤの一部の領域から構成される第２のボリュームも有し、
前記第１のストレージ装置及び前記第２のストレージ装置は、前記第１の仮想ボリュームのデータを前記第２の仮想ボリュームへコピーするリモートコピーを設定し、

前記ホストコンピューターは、前記ファイルシステムがライトリクエスト処理を行う際、前記リモートコピーに関する情報を管理するＩ／Ｏパスマネージャーが前記ファイルシステムを通じた前記アプリケーションからのライトリクエストを、前記リモートコピーに関する情報に基づいて、前記ＨＢＡデバイスドライバーを通じて前記第１の仮想ボリュームが存在する前記第１のストレージ装置へ送信し、

前記第１のストレージ装置は、受信した前記ライトリクエストが前記第１の仮想ボリューム宛なのか前記第１のストレージ装置の前記第２のボリューム宛なのかを判断し、

前記第１の仮想ボリューム宛の場合、前記ライトリクエストのデータを前記第１のストレージ装置及び前記第２のストレージ装置の前記キャッシュメモリに記憶し、

前記第１のストレージ装置の前記キャッシュメモリからデステージするデータを特定し、当該特定したデータをデステージしてから前記第２のストレージ装置の前記キャッシュメモリに記憶された当該データの破棄を指示する

ことを特徴とする情報システムのデータ転送方法。

【請求項２】

請求項１記載の方法であって、

前記第１のストレージ装置と、前記第２のストレージ装置との間の通信に障害が発生した場合、前記第１のストレージ装置は、前記通信障害を前記ホストコンピューターに通知し、前記ホストコンピューターは、前記第１のストレージ装置と、前記第２のストレージ装置とにライトリクエストを発行する

ことを特徴とする情報システムのデータ転送方法。

【請求項３】

請求項１記載の方法であって、

前記第１のストレージ装置の障害にかかわらずアプリケーションが継続処理可能である

ことを特徴とする情報システムのデータ転送方法。

【請求項４】

ファイルシステム、ＨＢＡデバイスドライバーを備えるオペレーティングシステム、及び前記ファイルシステムを通じてライトリクエストを送信するアプリケーションを備えるホストコンピューターと、前記ホストコンピューターに接続される第１のストレージ装置と、前記第１のストレージ装置及び前記ホストコンピューターに接続される第２のストレージ装置と、前記第１のストレージ装置及び前記第２のストレージ装置に接続され、複数の第１のＨＤＤとコントローラーとキャッシュメモリから構成され、前記複数の第１のＨＤＤの一部の領域から構成される第１のボリュームを有する第３のストレージ装置と、を有する情報システムのデータ転送方法であって、

前記第１のストレージ装置は、前記第３のストレージ装置の前記第１のボリュームに実体が存在する第１の仮想ボリュームを定義し、

前記第２のストレージ装置は、前記第３のストレージ装置の前記第１のボリュームに実体が存在し、当該第１のボリュームを前記第１の仮想ボリュームと共有する第２の仮想ボリュームを定義し、

前記第１のストレージ装置及び前記第２のストレージ装置は、それぞれ複数の第２のＨＤＤとコントローラーとキャッシュメモリとを備え、それぞれ前記複数の第２のＨＤＤの一部の領域から構成される第２のボリュームも有し、

前記第１のストレージ装置及び前記第２のストレージ装置は、前記第１の仮想ボリュームのデータを前記第２の仮想ボリュームへコピーするリモートコピーを設定し、

前記ホストコンピューターは、前記ファイルシステムがライトリクエスト処理を行う際、前記リモートコピーに関する情報を管理するＩ／Ｏパスマネージャーが前記ファイルシステムを通じた前記アプリケーションからのライトリクエストを、前記リモートコピーに関する情報に基づいて、前記ＨＢＡデバイスドライバーを通じて前記第１の仮想ボリュームが存在する前記第１のストレージ装置へ送信し、

前記第１のストレージ装置は、受信した前記ライトリクエストが前記第１の仮想ボリューム宛なのか前記第１のストレージ装置の前記第２のボリューム宛なのかを判断し、

前記第１の仮想ボリューム宛の場合、前記ライトリクエストのデータを前記第１のストレージ装置及び前記第２のストレージ装置の前記キャッシュメモリに記憶し、
前記第１のストレージ装置の前記キャッシュメモリからデステージするデータを特定し、当該特定したデータをデステージする一方、
前記第１のストレージ装置と、前記第２のストレージ装置との間の通信に障害が発生した場合、前記第１のストレージ装置は、前記通信障害を前記ホストコンピューターに通知し、前記ホストコンピューターは、前記第１のストレージ装置と、前記第２のストレージ装置とにライトリクエストを発行し、
前記第１のストレージ装置及び前記第２のストレージ装置は、自らを正系として前記ライトリクエストを処理する場合には、前記第３のストレージ装置の前記第１のボリュームに対して排他制御を行った後に、当該第１のストレージ装置又は当該第２のストレージ装置の前記キャッシュメモリに記憶した前記データを前記第１のボリュームにデステージする
ことを特徴とする情報システムのデータ転送方法。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、複数の記憶領域を有するストレージシステムと、ストレージシステムに接続されたホストコンピューターと、を備えた情報システムのデータ転送方法に関する。

【背景技術】

【０００２】

一般に、情報システムでは、記憶デバイスとしてＨＤＤ（ハードディスクドライブ）を用いるストレージ装置が備えられ、そのストレージ装置を含むストレージシステムが、ストレージエリアネットワーク（ＳＡＮ：Storage Area Network）経由で、複数の上位装置（例えばホスト）からアクセスされる。一般的にストレージ装置では、ＲＡＩＤ（Redundant Array of Independent (or Inexpensive) Disks）技術に従う高信頼化方法が採用されることでＨＤＤ単体の信頼性を超えた信頼性をストレージ装置として提供している。しかし、近年の情報化社会の進化によって上記ＲＡＩＤによる信頼性が貢献する情報システムの可用性（サービス継続性）では不足してきた。

【０００３】

このような状況に対応する高可用化技術として、特許文献１に開示された技術がある。当該技術では、ホストコンピューター（以後ホストと省略する）とストレージ装置をそれぞれ有するプロダクションサイトとバックアップサイトを用意し、プロダクションサイトのストレージ装置が保存するデータをバックアップサイトのストレージ装置にミラーリングする。もし、プロダクションサイトのストレージ装置が障害停止した場合は、バックアップサイトのストレージ装置とホストを用いて装置停止の結果停止していたアプリケーション処理を再開させる。本技術は一般的にリモートコピー又はリモートミラーリングと呼ばれる。

【０００４】

【特許文献１】特開平７‐２４４５９７号
【特許文献２】米国特許　７，０８０，１９７号

【発明の開示】

【発明が解決しようとする課題】

【０００５】

特許文献１の技術ではストレージ装置の障害停止の結果、異なるホストでアプリケーションを再開させるため、アプリケーションの再起動処理が必要になる。当然ながらアプリケーション停止から再起動完了まではアプリケーションは通常動作ができないため、可用性については問題がある。本発明は、２台以上のストレージ装置間でリモートコピーを行うストレージシステムと当該ストレージシステムを利用するホストを含めた情報システムの可用性を向上させることを目的とする。

【課題を解決するための手段】
【０００６】
かかる課題を解決するため本発明においては、ファイルシステム、ＨＢＡデバイスドライバーを備えるオペレーティングシステム、及び前記ファイルシステムを通じてライトリクエストを送信するアプリケーションを備えるホストコンピューターと、前記ホストコンピューターに接続される第１のストレージ装置と、前記第１のストレージ装置及び前記ホストコンピューターに接続される第２のストレージ装置と、前記第１のストレージ装置及び前記第２のストレージ装置に接続され、複数の第１のＨＤＤとコントローラーとキャッシュメモリから構成され、前記複数の第１のＨＤＤの一部の領域から構成される第１のボリュームを有する第３のストレージ装置と、を有する情報システムのデータ転送方法であって、前記第１のストレージ装置は、前記第３のストレージ装置の前記第１のボリュームに実体が存在する第１の仮想ボリュームを定義し、前記第２のストレージ装置は、前記第３のストレージ装置の前記第１のボリュームに実体が存在し、当該第１のボリュームを前記第１の仮想ボリュームと共有する第２の仮想ボリュームを定義し、前記第１のストレージ装置及び前記第２のストレージ装置は、それぞれ複数の第２のＨＤＤとコントローラーとキャッシュメモリとを備え、それぞれ前記複数の第２のＨＤＤの一部の領域から構成される第２のボリュームも有し、前記第１のストレージ装置及び前記第２のストレージ装置は、前記第１の仮想ボリュームのデータを前記第２の仮想ボリュームへコピーするリモートコピーを設定し、前記ホストコンピューターは、前記ファイルシステムがライトリクエスト処理を行う際、前記リモートコピーに関する情報を管理するＩ／Ｏパスマネージャーが前記ファイルシステムを通じた前記アプリケーションからのライトリクエストを、前記リモートコピーに関する情報に基づいて、前記ＨＢＡデバイスドライバーを通じて前記第１の仮想ボリュームが存在する前記第１のストレージ装置へ送信し、前記第１のストレージ装置は、受信した前記ライトリクエストが前記第１の仮想ボリューム宛なのか前記第１のストレージ装置の前記第２のボリューム宛なのかを判断し、前記第１の仮想ボリューム宛の場合、前記ライトリクエストのデータを前記第１のストレージ装置及び前記第２のストレージ装置の前記キャッシュメモリに記憶し、前記第１のストレージ装置の前記キャッシュメモリからデステージするデータを特定し、当該特定したデータをデステージしてから前記第２のストレージ装置の前記キャッシュメモリに記憶された当該データの破棄を指示することを特徴とする。
【０００７】
また本発明においては、ファイルシステム、ＨＢＡデバイスドライバーを備えるオペレーティングシステム、及び前記ファイルシステムを通じてライトリクエストを送信するアプリケーションを備えるホストコンピューターと、前記ホストコンピューターに接続される第１のストレージ装置と、前記第１のストレージ装置及び前記ホストコンピューターに接続される第２のストレージ装置と、前記第１のストレージ装置及び前記第２のストレージ装置に接続され、複数の第１のＨＤＤとコントローラーとキャッシュメモリから構成され、前記複数の第１のＨＤＤの一部の領域から構成される第１のボリュームを有する第３のストレージ装置と、を有する情報システムのデータ転送方法であって、前記第１のストレージ装置は、前記第３のストレージ装置の前記第１のボリュームに実体が存在する第１の仮想ボリュームを定義し、前記第２のストレージ装置は、前記第３のストレージ装置の前記第１のボリュームに実体が存在し、前記第１のボリュームを前記第１の仮想ボリュームと共有する第２の仮想ボリュームを定義し、前記第１のストレージ装置及び前記第２のストレージ装置は、それぞれ複数の第２のＨＤＤとコントローラーとキャッシュメモリとを備え、それぞれ前記複数の第２のＨＤＤの一部の領域から構成される第２のボリュームも有し、前記第１のストレージ装置及び前記第２のストレージ装置は、前記第１の仮想ボリュームのデータを前記第２の仮想ボリュームへコピーするリモートコピーを設定し、前記ホストコンピューターは、前記ファイルシステムがライトリクエスト処理を行う際、前記リモートコピーに関する情報を管理するＩ／Ｏパスマネージャーが前記ファイルシステムを通じた前記アプリケーションからのライトリクエストを、前記リモートコピーに関す

る情報に基づいて、前記ＨＢＡデバイスドライバーを通じて前記第１の仮想ボリュームが存在する前記第１のストレージ装置へ送信し、前記第１のストレージ装置は、受信した前記ライトリクエストが前記第１の仮想ボリューム宛なのか前記第１のストレージ装置の前記第２のボリューム宛なのかを判断し、前記第１の仮想ボリューム宛の場合、前記ライトリクエストのデータを前記第１のストレージ装置及び前記第２のストレージ装置の前記キャッシュメモリに記憶し、前記第１のストレージ装置の前記キャッシュメモリからデステージするデータを特定し、当該特定したデータをデステージする一方、前記第１のストレージ装置と、前記第２のストレージ装置との間の通信に障害が発生した場合、前記第１のストレージ装置は、前記通信障害を前記ホストコンピューターに通知し、前記ホストコンピューターは、前記第１のストレージ装置と、前記第２のストレージ装置とにライトリクエストを発行し、前記第１のストレージ装置及び前記第２のストレージ装置は、自らを正系として前記ライトリクエストを処理する場合には、前記第３のストレージ装置の前記第１のボリュームに対して排他制御を行った後に、当該第１のストレージ装置又は当該第２のストレージ装置の前記キャッシュメモリに記憶した前記データを前記第１のボリュームにデステージすることを特徴とする。

【０００８】

さらに、本発明においては、実施の形態として、前記第１のストレージ装置と、前記２のストレージ装置との間の通信に障害が発生した場合、前記第１のストレージ装置は、前記通信障害を前記ホストコンピューターに通知し、前記ホストコンピューターは、前記第１のストレージ装置と、前記第２のストレージ装置とに書き込み要求を発行する。また、前記第１のストレージ装置の障害にかかわらずアプリケーションが継続処理可能である。

【発明の効果】

【０００９】

本発明によれば、２台以上のストレージ装置間でリモートコピーを行う情報システムと当該情報システムを利用するデータ転送の可用性を向上させることができる。

【発明を実施するための最良の形態】

【００１０】

以下、図面を参照して、本発明の実施の形態を説明する。

【００１１】

（１）第１の実施の形態

＜１．情報システムの構成＞

図１は、本発明の一実施の形態に係る情報システムのハードウェア構成の一例を示す図である。

【００１２】

情報システムは、例えば、ストレージ装置１５００、ホストコンピューター（以後ホストと省略する）１１００、管理ホスト１２００と、２台以上の仮想化ストレージ装置１０００とから構成される。ストレージ装置１５００、ホストコンピューター（以後ホストと省略する）１１００、管理ホスト１２００の数は、それぞれ、１以上とすることができる。仮想化ストレージ装置１０００とホスト１１００は、Ｉ／Ｏネットワーク１３００を介して、相互に接続される。仮想化ストレージ装置１０００とストレージ装置１５００と管理ホスト１２００は、管理ネットワーク（図示せず）又はＩ／Ｏネットワーク１３００を介して相互に接続される。

【００１３】

ホスト１１００には、ホスト内部ネットワーク１１０４があり、そのネットワーク１１０４に、プロセッサ（図中ではProcと略記）１１０１と、メモリ（図中ではMemと略記）１１０２と、Ｉ／Ｏポート（図中ではI/O Pと略記）１１０３とが接続されている。管理ホスト１２００も、ホスト１１００と同じハードウェア構成を有することができる。なお、Ｉ／ＯポートをホストＨ１１００に追加する拡張カードをＨＢＡ（Host Bas Adapter)と呼ぶことがある。

【００１４】

管理ホスト１２００は、表示装置を有し、その表示装置に、仮想化ストレージ装置１０００とストレージ装置１５００の管理用の画面を表示することができる。また、管理ホスト１２００は、管理操作リクエストを、ユーザー（例えば管理ホスト１２００のオペレーター）から受付け、その受け付けた管理操作リクエストを、仮想化ストレージ装置１０００やストレージ装置１５００に送信することができる。管理操作リクエストは、仮想化ストレージ装置１０００やストレージ装置１５００の操作のためのリクエストであり、例えば、パリティグループ作成リクエスト、内部ＬＵ（Logical Unit）作成リクエスト、パス定義リクエスト、及び仮想化機能に関する操作がある。

【００１５】

Ｉ／Ｏネットワーク１３００は、ファイバーチャネルによる接続が第一に考えられるが、それ以外でも、ＦＩＣＯＮ（FIbre CONnection：登録商標）やＥｔｈｅｒｎｅｔ（登録商標）とＴＣＰ／ＩＰ（Transmission Control Protocol/Internet Protocol）とｉＳＣＳＩ（internet ＳＣＳＩ（Small Computer System Interface））の組み合わせや、Ｅｔｈｅｒｎｅｔ（登録商標）とＮＦＳ(Network File System)やＣＩＦＳ(Common Internet File System)等のネットワークファイルシステムの組み合わせ等が考えられる。さらに、Ｉ／Ｏネットワーク１３００は、Ｉ／Ｏリクエストを転送可能な通信装置であればこれ以外でもよい。また、仮想化ストレージ装置１０００とストレージ装置５００を接続するネットワークについてもＩ／Ｏネットワーク１３００と同様である。

【００１６】

仮想化ストレージ装置１０００は、コントローラー（図中はCTLと表記）１０１０と、キャッシュメモリ（図中はCMと表記）１０２０と、複数のＨＤＤ１０３０とを備える。好ましい形態としては、コントローラー１０１０及びキャッシュメモリ１０２０は、それぞれ複数のコンポーネントから構成することが考えられる。なぜなら、コンポーネント単体に障害が発生して閉塞した場合でも、残りのコンポーネントを用いてリードやライトに代表されるＩ／Ｏリクエストを引き続き受けることができるためである。

【００１７】

コントローラー１０１０は、仮想化ストレージ装置１０００の動作を制御する装置（例えば回路基盤）である。コントローラー１０１０には、内部ネットワーク１０１７があり、その内部ネットワーク１０１７に、Ｉ／Ｏポート１０１３、キャッシュポート（図中ではC Pと表記）１０１５、管理ポート（図中ではM Pと表記）１０１６、バックエンドポート（図中ではB/E Pと表記）１０１４、プロセッサ（例えばＣＰＵ（Central Processing Unit））１０１１及びメモリ１０１２が接続されている。コントローラー１０１０同士とキャッシュメモリ１０２０は、ストレージ内部ネットワーク１０５０にて相互に接続される。また、コントローラー１０１０と各ＨＤＤ１０３０は、複数のバックエンドネットワーク１０４０にて相互接続される。

【００１８】

ストレージ装置１５００のハードウェア構成は仮想化ストレージ装置１０００と同種の部品から構成される。なお、仮想化ストレージ装置１０００がＨＤＤを持たない仮想化専用装置またスイッチの場合は、ストレージ装置１５００は仮想化ストレージ装置１０００と同種の部品から構成されなくてもいい。さらに、ホスト１１００及び仮想化ストレージ装置１０００の内部のネットワークは、好ましくは、Ｉ／Ｏポート１０１３の有する転送帯域より広帯域であり、また、バスやスイッチ型のネットワークによって全てまた一部が代替されてもよい。また、図１では、Ｉ／Ｏポート１０１３は、コントローラー１０１０に一つ存在することになっているが、実際には、複数のＩ／Ｏポート１０１３がコントローラー１０１０に存在してもよい。

【００１９】

以上のハードウェア構成によって、仮想化ストレージ装置１０００やストレージ装置１５００のＨＤＤに保存された全て又は一部のデータを、ホスト１１００が読出したり書き込んだりすることができるようになる。なお、以後の説明では、データ保存を担当するシステムをストレージクラスタと呼ぶ。また、ストレージクラスタ内部に当該システムを２

系統含むことで高可用化を実現するサブシステムで、仮想化ストレージ装置１０００とストレージ装置１５００の片方又は両方を含むサブシステムをストレージサブシステムと呼ぶ。

【００２０】

＜２．本実施の形態の概要＞

本実施の形態では、他のストレージ装置内のボリューム等の記憶領域を仮想化する仮想化機能を有する仮想化ストレージ装置１０００を含むストレージシステムの可用性を向上させるため、もう一台の仮想化ストレージ装置１０００を用いた二重化構成を採用する。図２はその概要を示した図である。

【００２１】

本概要では、ストレージシステムに仮想化ストレージ装置１０００Ｌ、仮想化ストレージ装置１０００Ｒ、ストレージ装置１５００Ｌ、ストレージ装置１５００Ｒが含まれる。なお、以下においては、説明を容易にするため、仮想化ストレージ装置１０００Ｌ及びストレージ装置１５００Ｌを正系（プロダクション系）、仮想化ストレージ装置１０００Ｒ及びストレージ装置１５００Ｒを副系（バックアップ系）の役割をもっているものとする。しかし、それぞれの仮想化ストレージ装置１０００Ｌ，１０００Ｒがホスト１１００へ提供するボリュームが二つ以上の場合は、仮想化ストレージ装置単位で正系・副系を担当する代わりにボリューム単位で正系を担当する仮想化ストレージ装置１０００Ｌ，１０００Ｒが定まっていればよい。

【００２２】

それぞれの仮想化ストレージ装置１０００Ｌ，１０００Ｒは自身が有するＨＤＤ１０３０を構成要素とするパリティグループ（ＲＡＩＤ技術によって構成される）の一部又は全ての領域をボリューム３０００ＬＡやボリューム３０００ＲＡとしてホスト１１００に提供する（図中の円柱内に'Ａ'と記された部分が対応）。また、仮想化ストレージ装置１０００はオプションとして仮想化機能による仮想ボリューム３０００ＬＢ，３０００ＲＢ（対応するＨＤＤ等の不揮発記憶領域が仮想化ストレージ装置１０００Ｌ，１０００Ｒの外部に存在するボリュームのこと）を提供することができる。本概要ではストレージ装置１５００Ｌ、１５００Ｒが提供するボリューム３５００ＬＢ、３５００ＲＢの一部又は全てを対応する不揮発記憶領域として用いている。なお、以後の説明では「ボリュームのデータ」と書いた場合は、ＨＤＤ１０３０に保存されたデータに加えてキャッシュメモリ１０２０に一時保存されたデータも含む。また、後ほど述べる「仮想ボリュームのデータ」に関してはストレージ装置１５００Ｌ，１５００Ｒのボリューム３５００ＬＢ、３５００ＲＢに保存されたデータに加えて仮想化ストレージ装置１０００Ｌ，１０００Ｒのキャッシュメモリ１０２０に一時保存されたデータを含む。

【００２３】

一方、ホスト１１００上ではアプリケーションプログラム（以後、アプリケーションと略すことがある）２０１０と、ＯＳと、ＯＳの設定・処理を補佐するデーモンや管理プログラムに代表されるシステムプログラムとが動作している。ＯＳはアプリケーション２０１０に対して仮想化ストレージ装置１０００Ｌ，１０００Ｒが提供するボリューム３０００ＬＡ，３０００ＬＢ，３０００ＲＡ，３０００ＲＢ内に存在するデータに対するＩ／Ｏリクエスト用インターフェースを提供し、アプリケーション２０１０からの要求に応じて適切な仮想化ストレージ装置１０００Ｌ，１０００Ｒ及びボリューム３０００ＬＡ，３０００ＬＢ，３０００ＲＡ，３０００ＲＢに対するＩ／Ｏリクエストを送信する。通常状態ではホスト１１００は仮想化ストレージ装置１０００Ｌのボリューム３０００ＬＡ，３０００ＬＢに対してリードやライトに代表されるＩ／Ｏリクエストを発行し、データの送受信を行う。つまり、リードリクエストを受け取った場合、仮想化ストレージ装置１０００Ｌは、リクエスト対象のボリューム３０００ＬＡ，３０００ＬＢ，３５００ＬＢが仮想化ストレージ装置１０００Ｌ内部のＨＤＤ１０３０に対応している場合は当該ＨＤＤ１０３０からデータを読み上げてこれをホスト１１００に返したり、ストレージ装置１５００Ｌに対してリードリクエストを発行することで必要なデータを取得し、そのデータ（の全て

又は一部）をホスト１１００に返す。

【００２４】

ライトリクエストの場合は、データの冗長化のために、ライトデータを受け取った仮想化ストレージ装置１０００Ｌは副系である仮想化ストレージ装置１０００Ｒへライトデータを送信し、仮想化ストレージ装置１０００Ｌがライトデータの受け取り完了メッセージを仮想化ストレージ装置１０００Ｒから受け取った後にホスト１１００に対してライト完了メッセージを返す。なお、仮想化ストレージ装置１０００Ｌに対するライトデータも仮想化ストレージ装置１０００Ｒが仮想化ストレージ装置１０００Ｌを経由して受け取ったライトデータも、各仮想化ストレージ装置１０００Ｌ，１０００Ｒ内のキャッシュメモリ１０２０Ｌ，１０２０Ｒに一時保持されてもよい。なお、本実施の形態の一つとして、このライトデータの転送はストレージリモートコピーによって行われる。

【００２５】

図３は通常状態下で仮想化ストレージ装置１０００Ｌに障害が発生した後の情報システムの処理概要を示している。

【００２６】

正系の仮想化ストレージ装置１０００Ｌが障害によって停止した場合、ホスト１１００上のシステムプログラムはその障害を検知し、Ｉ／Ｏリクエストの発行先を正系の仮想化ストレージ装置１０００Ｌから副系の仮想化ストレージ装置１０００Ｒへ切り替える。ただし、その場合もアプリケーション２０１０はＩ／Ｏリクエストの発行先が切り替わったことを認識せずにＩ／Ｏを継続することができる。そのために、通常時からシステムプログラムはＯＳレイヤ（より具体的にはファイルシステムより下位のレイヤ）にて、アプリケーション２０１０やファイルシステムからＩ／Ｏリクエスト時に指定されるボリューム識別子として仮想的なボリューム識別子（又はデバイスファイル）を指定させるようにしておき、ＯＳの下位レイヤは当該識別子と実際のボリュームに対して割り当てられた識別子（又はデバイスファイル）の対応を管理しておく。Ｉ／Ｏリクエストの発行先を切り替える場合は、その対応関係をこれまでの仮想化ストレージ装置１０００Ｌのボリューム３０００ＬＡ、ボリューム３０００ＬＢ宛であったものを仮想化ストレージ装置１０００Ｒのボリューム３０００ＲＡとボリューム３０００ＲＢ宛に切り替えることでアプリケーション２０１０に対して透過に切り替えを実現する。

【００２７】

さらに、仮想化ストレージ装置１０００Ｒもホスト１１００からの当該ボリューム３０００ＲＡ，３０００ＲＢに対するライトリクエストの到着やその他明示的なフェイルオーバー要求に応じて、ライトリクエストを処理できるようにする。この変更処理の一例としては、仮想化ストレージ装置１０００Ｌから仮想化ストレージ装置１０００Ｒに対するデータコピーに伴い、仮想化ストレージ装置１０００Ｒのボリューム３０００ＲＡ，３０００ＲＢに対するホスト１１００からのライトリクエストの拒否が設定されている場合はそれを解除する。また、リモートコピーを用いてライトデータの転送を行っている場合はリモートコピーのコピー状態の変更を行うことも考えられる。

【００２８】

図４は仮想化ストレージ装置１０００Ｌ，１０００Ｒ間のネットワークに障害が発生した後の情報システムの処理概要を示している。

【００２９】

ネットワーク障害を検知した仮想化ストレージ装置１０００Ｌはホスト１１００に当該障害を通知する。障害通知を受けたホスト１１００は、副系の仮想化ストレージ装置１０００Ｒに対してライトリクエストを処理できるように要求し、以後のライトリクエストは正系の仮想化ストレージ装置１０００Ｌ及び副系の仮想化ストレージ装置１０００Ｒの両方に発行することで、正系と副系のデータを同一にする。

【００３０】

＜３．ホスト１１００で実行されるプログラム及び情報＞

図５はホスト１１００上で実行されるソフトウェアプログラムと、当該ソフトウェアプ

ログラムが用いる情報とに加えて、各ソフトウェアプログラムが提供する概念について記した図である。なお、当該ソフトウェアプログラムはメモリ１１０２（図１）とプロセッサ１１０１（図１）とによって保持と実行がされるが、その一部をハードウェア化して実行してもよい。

【００３１】

ホスト１１００上ではアプリケーション２０１０、リモートコピーマネージャー５０３０に加えて、ＯＳ又はＫｅｒｎｅｌ内部のプログラムモジュールとしてファイルシステム５０２０、Ｉ／Ｏパスマネージャー５０００及びＨＢＡデバイスドライバー５０１０が実行される（ファイルシステム５０２０、Ｉ／Ｏパスマネージャー５０００又はＨＢＡデバイスドライバー５０１０は、全ての処理がＫｅｒｎｅｌ内部で実行される必要はない）。

【００３２】

ＨＢＡデバイスドライバー５０１０はＨＢＡに搭載されたＩ／Ｏポート１１０３（図１）を通じてＩ／Ｏリクエストやそれに伴うデータを送受信したり、その他の仮想化ストレージ装置１０００Ｌ，１０００Ｒやストレージ装置１５００Ｌ，１５００Ｒ等との通信を制御するプログラムである。ＨＢＡデバイスドライバー５０１０は、また、上位レイヤに対して仮想化ストレージ装置１０００Ｌ，１０００Ｒが提供するボリューム３０００ＬＡ，３０００ＬＢ，３０００ＲＡ，３０００ＲＢに対応する識別子を提供し、その識別子を伴ったＩ／Ｏリクエストを受け付けることができる。ボリューム５０４０はその概念を示したもので、仮想ストレージ装置１０００Ｌ，１０００Ｒが提供するボリューム３０００ＬＡ，３０００ＬＢ，３０００ＲＡ，３０００ＲＢにそれぞれ対応している。

【００３３】

Ｉ／Ｏパスマネージャー５０００は、アプリケーション２０１０のＩ／Ｏリクエスト発信先を切り替えるためのモジュールである。当該モジュールはＨＢＡデバイスドライバー５０１０が提供するボリューム５０４０に対応する識別子と同種のホスト１１００内での仮想的なボリュームに対応する識別子及びＩ／Ｏリクエスト用インターフェースをファイルシステム５０２０に対して提供する。このホスト１１００内での仮想的なボリュームに対応する識別子は当該モジュール内でＨＢＡデバイスドライバー５０１０が提供するボリューム５０４０に対応する識別子と対応しており、デバイス関係テーブル５００１がその対応関係を保持している。ボリューム５０５０はこのホスト１１００内での仮想的なボリュームの概念を示したもので、本図ではその対応関係の一例として仮想化ストレージ装置１０００Ｌのボリューム３０００ＬＡ，３０００ＬＢに対応する識別子と対応している（他の言い方をすると、ホスト１１００内での仮想的なボリューム５０５０の実体は仮想化ストレージ装置１０００Ｌのボリューム３０００ＬＡ，３０００ＬＢであるともいえる）。

【００３４】

ここまでのレイヤでのＩ／Ｏリクエストは通常固定長ブロックアクセス形式で指定する。ただし、ホスト１１００がメインフレームの場合はこれに限定されず、ＣＫＤ（Count Key Data）形式で指定してもよい。

【００３５】

ファイルシステム５０２０は、ＨＢＡデバイスドライバー５０１０が提供するボリューム５０４０に対応する識別子及びＩ／Ｏインターフェースと、Ｉ／Ｏパスマネージャー５０００が提供するホスト１１００内での仮想的なボリューム５０５０に対応する識別子及びＩ／Ｏインターフェースとを通じて、仮想化ストレージ装置１０００Ｌ，１０００ＲへのＩ／Ｏリクエストを送信したり、データの送受信を行うモジュールである。図５では例としてファイルシステム５０２０内部にディレクトリツリーの構造を示し、そのツリー構造の一部５０５２が、Ｉ／Ｏパスマネージャー５０００がホスト１１００内での仮想化で提供したボリューム５０５０に保存されている状態を示している（これまで説明した通り、より正確にはＩ／Ｏパスマネージャー５０００のホスト１１００内での仮想的なボリューム５０５０の提供は識別子を通じたものであり、さらに、そのボリューム５０５０に保存されていると書いたデータは実際にはデバイス関係テーブル５００１にて示される仮想

化ストレージ装置１０００Ｌ，１０００Ｒが提供するボリューム３０００ＬＡ，３０００ＬＢ，３０００ＲＡ，３０００ＰＢに保存されている）。ファイルシステム５０２０はアプリケーション２０１０に対してファイルＩ／Ｏのインターフェースを提供する。ファイルＩ／Ｏインターフェースを通じてアプリケーション２０１０から呼び出されたファイルシステム５０２０は、ファイル名とファイル内でのデータオフセットを伴ったリード又はライトリクエストをディレクトリファイルやｉｎｏｄｅといったファイルシステム５０２０内の構造化情報を参照しつつ、ブロック形式のリード又はライトリクエストに変換し、Ｉ／Ｏパスマネージャー５０００又はＨＢＡデバイスドライバー５０１０へリード又はライトリクエストを渡す。

【００３６】

なお、Ｕｎｉｔ系やＷｉｎｄｏｗｓ（登録商標）系のＯＳではファイルＩ／Ｏのインターフェースを用いて直接ボリュームのデータを操作するためのインターフェースとしてデバイスファイルシステムと呼ばれる機能を提供している。通常、デバイスファイルシステムはファイル空間の'/dev'ディレクトリ配下に展開されており、当該ディレクトリ以下のファイル（図中の例では、rsda等）のファイル名はファイルシステム５０２０の下位レイヤ（ＨＢＡデバイスドライバー５０１０やＩ／Ｏパスマネージャー５０００）が提供するボリューム５０４０，５０５０に対応する。そして、当該ボリューム５０４０，５０５０に保存されたデータはデバイスファイル５０７０，５０８０に保存されたデータであるかのようにファイルＩ／Ｏ用インターフェースで読み書き可能となる。なお、図５では例としてデバイスファイル５０７０（rsda，rsdb，rsdc，rsdd）はＨＢＡデバイスドライバー５０１０が認識し、提供しているボリューム５０４０に対応し、デバイスファイル５０８０（vsda，vsdb）はＩ／Ｏパスマネージャー５０００が提供しているボリューム５０５０に対応している。このデバイスファイル５０７０，５０８０は、アプリケーション２０１０がデータベースである場合に、独自のデータ編成やバッファ管理を実現する目的で使われることがある。

【００３７】

リモートコピーマネージャー５０３０は仮想化ストレージ装置１０００Ｌ，１０００Ｒとの間のデータ転送を実現するリモートコピーの状態を取得したり、ホスト１１００やＩ／Ｏパスマネージャー５０００がリモートコピーの操作を行うためのプログラムで、当該プログラムを使用するプログラム、ユーザー又はＩ／Ｏパスマネージャー５０００の要求に応じて仮想化ストレージ装置１０００Ｌ，１０００Ｒと通信を行う。

【００３８】

なお、これまで説明した通りＨＢＡデバイスドライバー５０１０やＩ／Ｏパスマネージャー５０００は一部又は全ての機能がＫｅｒｎｅｌ内部のモジュールとしてインストールやアンインストールすることができることが望ましい。なぜならば、ＨＢＡデバイスドライバー５０２０はＨＢＡを制御するプログラムであるが故、ＨＢＡの製造会社が提供することが多い。同様にＩ／Ｏパスマネージャー５０００は仮想化ストレージ装置１０００Ｌ，１０００Ｒの処理を前提として処理が決定されるため、一部又は全てのモジュールが仮想化ストレージ装置１０００Ｌ，１０００Ｒの製造会社が提供することが考えられる。したがって、当該プログラムがインストール・アンインストールできることによって幅広いＨＢＡと仮想化ストレージ装置１０００Ｌ，１０００Ｒの組み合わせによる情報システムを構築することができる。また、本発明ではアプリケーション２０１０に対して透過に正系と副系の切り替えを行うためにＫｅｒｎｅｌ内部で処理を実行することでアプリケーション２０１０の再コンパイル等が不要な透過的な切り替えが可能である。さらに、Ｉ／Ｏパスマネージャー５０００がファイルシステム５０２０とＨＢＡデバイスドライバー５０１０の中間レイヤに存在することで、ファイルシステム５０２０に対する再コンパイル等を不要とし、さらにファイルシステム透過性も確保している。そして、Ｉ／Ｏパスマネージャー５０００がＨＢＡデバイスドライバー５０１０の機能を利用することができるようになっている。

【００３９】

また、Ｋｅｒｎｅｌ内部にいるＩ／Ｏパスマネージャー５０００がリモートコピーマネージャー５０３０を呼び出す場合やその逆の通信方法として以下の二通りが考えられる。

【００４０】

（Ａ）Ｉ／Ｏパスマネージャー５０００は通信用の仮想的なボリュームを作成し、ファイルシステム５０２０はこの通信用ボリュームをデバイスファイルとしてファイル空間に作成する。リモートコピーマネージャー５０３０は定期的にデバイスファイルに対してリードシステムコールを実行した状態で待つ。Ｉ／Ｏパスマネージャー５０００はリモートコピーマネージャー５０３０からのＩ／Ｏリクエストを受信するが、内部で保留する。そして、当該モジュールがリモートコピーマネージャー５０３０に対するメッセージ送信をする必要が出てきたらＩ／Ｏリクエストの返り値として定められたメッセージを含むデータをファイルシステム５０２０を通じてリモートコピーマネージャー５０３０に返す。なおこの際リモートコピーマネージャーが発行するリードシステムコールは長時間Ｋｅｒｎｅｌ内部で待たされることになる。それが好ましくない場合は、Ｉ／Ｏパスマネージャー５０００が、一定時間経過後に何もメッセージがない旨のデータをファイルシステム５０２０を通じてリモートコピーマネージャー５０３０へ返し、それを受信したリモートコピーマネージャー５０３０が再度リードシステムコールを実行すればよい。

【００４１】

（Ｂ）Ｕｎｉｘ（登録商標）ドメインソケットを用いて仮想的なネットワーク通信として扱う。具体的には、ソケットの一方のエンドをリモートコピーマネージャー５０３０が操作し、残りのエンドをＩ／Ｏパスマネージャー５０００が操作する。

【００４２】

なお、以後の説明ではＩ／Ｏパスマネージャー５０００がリモートコピーの操作や状態参照を行う場合はこのような通信によってリモートコピーマネージャー５０３０を呼び出すことで操作を行っているものとする。

【００４３】

＜４．仮想ストレージ装置１０００で実行されるプログラム及び情報＞

図６は、仮想化ストレージ装置１０００（１０００Ｌ，１０００Ｒ）とストレージ装置１５００（１５００Ｌ，１５００Ｒ）とで実行されるプログラムと、当該プログラムにより管理される情報とについて示した図である。なお、当該プログラムはメモリ１０１２（図１）と、プロセッサ１０１１（図１）と、キャッシュメモリ１０２０とによって保持と実行がされるが、その一部をハードウェア化して実行してもよい。

【００４４】

＜４．１．Ｉ／Ｏ処理プログラム６０２０、パリティグループ情報６０６０及びボリューム情報６０５０＞

パリティグループ情報６０６０には、パリティグループ毎の以下の構成に関連する情報が含まれる。

（１）パリティグループを構成するＨＤＤ１０３０の識別子。パリティグループには複数のＨＤＤ１０３０が参加しているため、当該情報はパリティグループ毎に複数存在する。

（２）ＲＡＩＤレベル

【００４５】

また、ボリューム情報６０５０には、ボリューム毎の以下の構成に関連する情報が含まれる。

（１）ボリューム容量

（２）ボリュームに対応するデータが保存されるパリティグループの識別子とパリティグループ内の領域（開始アドレスと終了アドレスの片方又は両方）。

【００４６】

Ｉ／Ｏ処理プログラム６０２０は、ボリューム情報６０５０やパリティグループ情報６０６０を参照してホスト１１００から受信したＩ／Ｏリクエストに関する以下の処理を実行する。

【００４７】

（Ａ）ステージング： ＨＤＤ１０３０に保存されたデータをキャッシュメモリ１０２０上にコピーする。
（Ｂ）デステージング：キャッシュメモリ１０２０に保存されたデータをＨＤＤ１０３０へコピーする。なお、その前の処理としてＲＡＩＤ技術による冗長データを作成してもよい。
【００４８】
（Ｃ）リード処理：ホスト１１００から受信したリードリクエストに対して、当該リクエストに対応するデータがキャッシュメモリ１０２０上に存在するかどうか判定する。そして、当該リクエストに対応するデータがキャッシュメモリ１０２０上に存在しない場合は、ステージング処理を実行して当該データをキャッシュメモリ１０２０上にコピーした後に、そのデータをホスト１１００に対して送信する。なお、キャッシュメモリ１０２０上にかかるデータが存在する場合は、当該データをホスト１１００に対して送信する。
【００４９】
（Ｄ）ライト処理：ホスト１１００から受信したライトデータをキャッシュメモリ１０２０上に保存する。なお、当該処理時にキャッシュメモリ１０２０上に十分な空き領域が無い場合はデステージング処理を実行して適切なデータをＨＤＤ１０３０上にコピーした後にキャッシュメモリ１０２０上の当該領域を流用する。また既にキャッシュメモリ１０２０上に保存された領域がライトリクエストに含まれる場合は、そのまま既存のキャッシュメモリ１０２０上の領域へ上書きすることもある。
【００５０】
（Ｅ）キャッシュアルゴリズム：キャッシュメモリ１０２０上のデータの参照頻度や参照時期等を元にＬＲＵ等のアルゴリズムによってステージングすべきＨＤＤ１０３０上のデータやデステージングすべきキャッシュメモリ１０２０上のデータを決定する。
【００５１】
＜４．２．仮想化プログラム６０３０と仮想化情報６０７０＞
仮想化情報６０７０には、仮想化ボリューム毎の以下の構成に関連する情報が含まれる。
【００５２】
（１）ストレージ装置１５００内のボリューム内の領域とその領域が仮想ボリューム上のアドレス空間のどの領域としてホスト１１００に提供するかに関する以下の情報。仮想ボリュームが複数で構成される場合は下記情報も複数存在する。
（１－１）仮想ボリュームを構成する、ストレージ装置１５００の識別子（又はポートの識別子）と、ボリュームの識別子と、ボリューム内の領域（開始アドレスと終了アドレス）
（１－２）仮想ボリュームにおける領域（開始アドレスと終了アドレス）
【００５３】
（２）仮想ボリュームの容量
仮想化プログラム６０３０は、仮想化ストレージ装置１０００が、ストレージ装置１５００が提供するボリュームを用いてホスト１１００にボリュームを提供するためのプログラムである。なお、仮想化プログラム６０３０が提供する仮想ボリュームと、それに対応するストレージ装置１５００上のボリュームとの対応関係として、以下のパターンがある。
【００５４】
（Ａ）ストレージ装置１５００上のボリューム全体を仮想ボリュームの記憶領域として用いる場合。この場合、仮想ボリュームの容量は選択したボリュームとおおよそ同容量となる（制御情報や冗長情報をストレージ装置１５００上のボリュームに保存する場合。当該情報等がない場合は同一容量）。
（Ｂ）ストレージ装置１５００上のボリュームの一部の領域を仮想化ボリュームに対応する保存領域として用いる場合。この場合、仮想ボリュームの容量は当該利用対象の領域容量と大体同じとなる。

【００５５】
（Ｃ）複数のストレージ装置１５００上の複数のボリュームを仮想ボリュームの記憶領域として結合して用いる場合。この場合、仮想ボリュームの容量は各ボリューム容量の合計値とおおよそ同容量となる。なお、この結合方式としてはストライピングやＣｏｎｃａｔｅｎａｔｅ（複数ボリュームを連結して一つのボリュームとして扱う方法）等がある。
（Ｄ）（Ｃ）のパターンに付随してパリティ情報やミラーデータを保存する場合。この場合、仮想ボリュームの容量はミラーデータを保存する場合は（Ｃ）の半分で、パリティを保存する場合はパリティ計算方式に依存する。ストレージ装置１５００内部でＲＡＩＤによる高信頼化と組み合わせることによって仮想ボリュームに保存されたデータについての信頼性がより向上する。
【００５６】
なお、いずれのパターンについても、Ｉ／Ｏリクエストで指定するストレージ装置識別子（又はポート識別子）とボリューム識別子（Ｉ／Ｏリクエストで用いる、仮想化ストレージ装置内又はポート配下のボリュームを識別する情報で、ＬＵＮ（Logical Unit Number）や、ＣＫＤ形式のＣＵ番号とＬＤＥＶ（Logical DEVice）番号等がある）が元々のボリュームと異なる。
【００５７】
仮想化プログラム６０３０は、ステージングやデステージング対象となるデータが仮想ボリュームに対応する場合にＩ／Ｏ処理プログラム６０２０により呼び出され、仮想化情報６０７０を用いて以下の処理を実行する。
【００５８】
（Ａ）ステージング： 仮想化ボリュームとストレージ装置１５００のボリュームの対応関係を元に、どのストレージ装置１５００のボリュームに保存されたデータをキャッシュメモリ１０２０上にコピーすべきかを決定した後に、キャッシュメモリ１０２０上へデータコピーする。
【００５９】
（Ｂ）デステージング：仮想化ボリュームとストレージ装置１５００のボリュームの対応関係を元に、どのストレージ装置１５００のボリュームへキャッシュメモリ１０２０上のデータをコピーすべきかを決定した後に、ストレージ装置１５００のボリュームへデータコピーする。なお、その前の処理としてＲＡＩＤ技術による冗長データを作成してもよい。
【００６０】
＜４．３．リモートコピープログラム６０１０とコピーペア情報６０４０＞
コピーペア情報６０４０はリモートコピーのコピー元ボリュームとコピー先ボリュームのコピーペア（ペアと省略することがある）毎に以下の情報を持つ。なお、本実施の形態では、コピー元ボリューム及びコピー先ボリュームは高可用性を実現する対象ボリュームが指定されることになる：
【００６１】
（１）コピー元ボリュームを持つ仮想化ストレージ装置１０００の識別子及びボリュームの識別子
（２）コピー先ボリュームを持つ仮想化ストレージ装置１０００の識別子とボリュームの識別子
（３）コピーペアの状態（詳細は後ほど述べる）
【００６２】
リモートコピープログラム６０１０は、コピー元ボリュームに保存されたデータをコピー先ボリュームにミラーリングするプログラムであり、コピーペア情報６０４０を参照して処理を行う。以下にリモートコピー（特に同期リモートコピー）の処理概要とペア状態について説明する。
【００６３】
＜４．３．１．同期リモートコピーのコピー処理動作＞

同期リモートコピーとは、前述の様に、コピー元の仮想化ストレージ装置１０００がホスト１１００からコピー元ボリュームに対するライトリクエストを受け付けた場合、ライトデータをコピー先の仮想化ストレージ装置１０００に送信した後に、ホスト１１００に対してライトリクエスト完了を返すリモートコピー方法である。

【００６４】

同期リモートコピーが実行される際、コピー元ボリュームとコピー先ボリュームとのペア間におけるリモートコピーの状況を管理１２００に表示したり、リモートコピーの状態を操作するために、仮想化ストレージ装置１０００のコントローラー１０１０は、コピーペア状態（Ｓｉｍｐｌｅｘ、Ｉｎｉｔｉａｌ‐Ｃｏｐｙｉｎｇ、Ｄｕｐｌｅｘ、Ｓｕｓｐｅｎｄ及びＤｕｐｌｅｘ‐Ｐｅｎｄｉｎｇ）と呼ばれる情報を管理する。図７に同期リモートコピーのペア状態に関する状態遷移図を示す。以下、各ペア状態について説明する。

【００６５】

＜４．３．１．１．Ｓｉｍｐｌｅｘ状態＞

Ｓｉｍｐｌｅｘ状態は、ペアを構成するコピー元ボリュームとコピー先ボリュームとの間でコピーが開始されていない状態である。

【００６６】

＜４．３．１．２．Ｄｕｐｌｅｘ状態＞

Ｄｕｐｌｅｘ状態は、同期リモートコピーが開始され、後述する初期化コピーも完了してペアを構成するコピー元ボリューム及びコピー先ボリュームのデータ内容が同一となった状態である。本状態では、書き込み途中の領域を除けば、コピー元ボリュームのデータ及びコピー先ボリュームのデータの内容は同じとなる。なお、Ｄｕｐｌｅｘ中及びＤｕｐｌｅｘ‐Ｐｅｎｄｉｎｇ及びＩｎｉｔｉａｌ‐Ｃｏｐｙｉｎｇ状態ではホスト１１００からコピー先ボリュームへのライトリクエストは拒否される。

【００６７】

＜４．３．１．３．Ｉｎｉｔｉａｌ‐Ｃｏｐｙｉｎｇ状態＞

Ｉｎｉｔｉａｌ‐Ｃｏｐｙｉｎｇ状態は、Ｓｉｍｐｌｅｘ状態からＤｕｐｌｅｘ状態へ遷移するまでの中間状態であり、この期間中に、必要ならばコピー元ボリュームからコピー先ボリュームへの初期化コピー（コピー元ボリュームに既に格納されていたデータのコピー先ボリュームへのコピー）が行われる。初期化コピーが完了し、Ｄｕｐｌｅｘ状態へ遷移するために必要な処理が終わったら、ペア状態はＤｕｐｌｅｘとなる。

【００６８】

＜４．３．１．４．Ｓｕｓｐｅｎｄ状態＞

Ｓｕｓｐｅｎｄ状態は、コピー元ボリュームに対する書き込みの内容をコピー先ボリュームに反映させない状態である。この状態では、ペアを構成しているコピー元ボリューム及びコピー先ボリュームのデータの内容は同じでない。ユーザーやホスト１１００からの指示を契機に、ペア状態は他の状態からＳｕｓｐｅｎｄ状態へ遷移する。それ以外に、仮想化ストレージ装置１０００間のネットワーク障害等が原因で同期リモートコピーを行うことが出来なくなった場合に自動的にペア状態がＳｕｓｐｅｎｄ状態に遷移することが考えられる。

【００６９】

以後の説明では、後者の場合、即ち障害により生じたＳｕｓｐｅｎｄ状態を障害Ｓｕｓｐｅｎｄ状態と呼ぶことにする。障害Ｓｕｓｐｅｎｄ状態となる代表的な原因としては、ネットワーク障害のほかに、コピー元ボリュームやコピー先ボリュームの障害、コントローラー１０１０の障害が考えられる。

【００７０】

Ｓｕｓｐｅｎｄ状態となった場合、コピー元ストレージ１０００は、Ｓｕｓｐｅｎｄ状態となった時点以降にコピー元ボリュームに対するライトリクエストがあると、ライトリクエストに従ってライトデータを受信し、コピー元ボリュームに保存するが、コピー先の仮想化ストレージ装置１０００にはライトデータを送信しない。またコピー元の仮想化ストレージ装置１０００は、書き込まれたライトデータのコピー元ボリューム上での書き込

み位置を差分ビットマップ等として記憶する。
【００７１】
なおＳｕｓｐｅｎｄ状態となった時点以降にコピー先ボリュームに対してライトリクエストがあった場合には、コピー先の仮想化ストレージ装置１０００も上記の動作を行う。また、ペアが障害Ｓｕｓｐｅｎｄ状態となるより前に、当該ペアに対してフェンスと呼ばれる設定を行った場合、ペア状態が障害Ｓｕｓｐｅｎｄに遷移するとコピー元ボリュームに対するライトを拒否する。なお、コピー先の仮想化ストレージ装置１０００は障害Ｓｕｓｐｅｎｄ状態中のコピー先ボリュームに対するライトリクエストを拒否してもよい。
【００７２】
＜４．３．１．５．Ｄｕｐｌｅｘ‐Ｐｅｎｄｉｎｇ状態＞
Ｄｕｐｌｅｘ‐Ｐｅｎｄｉｎｇ状態は、Ｓｕｓｐｅｎｄ状態からＤｕｐｌｅｘ状態に遷移するまでの中間状態である。この状態では、コピー元ボリューム及びコピー先ボリュームのデータの内容を一致させるために、コピー元ボリュームからコピー先ボリュームへのデータのコピーが実行される。コピー元ボリューム及びコピー先ボリュームのデータの内容が同一になった後、ペア状態はＤｕｐｌｅｘとなる。
【００７３】
なお、Ｄｕｐｌｅｘ‐Ｐｅｎｄｉｎｇ状態におけるデータのコピーは、Ｓｕｓｐｅｎｄ状態の間、コピー元の仮想化ストレージ装置１０００又はコピー先の仮想化ストレージ装置１０００が記録した書き込み位置（例えば上述の差分ビットマップ等）を利用して、更新が必要な部分（即ちコピー元ボリュームとコピー先ボリュームとのデータの不一致部分）だけをコピーする差分コピーによって実行される。
【００７４】
また、以上の説明ではＩｎｉｔｉａｌ‐Ｃｏｐｙｉｎｇ状態とＤｕｐｌｅｘ‐Ｐｅｎｄｉｎｇ状態は別々な状態としたが、これらをまとめて一つの状態として管理ホスト１２００の画面に表示したり、状態を遷移させても良い。
【００７５】
＜４．３．１．６．ペア操作指示＞
ペア状態はホスト１１００や管理ホスト１２００からの以下の指示によって他の状態へ遷移する。
【００７６】
（Ａ）初期化指示：Ｓｉｍｐｌｅｘ状態にて本指示を受信するとＩｎｉｔｉａｌ‐Ｃｏｐｙｉｎｇ状態へ遷移する。
（Ｂ）再同期指示：Ｓｕｐｅｎｄ状態又は障害Ｓｕｓｐｅｎｄ状態にて本指示を受信するとＤｕｐｌｅｘ‐Ｐｅｎｄｉｎｇ状態へ遷移する。
（Ｃ）分割指示：Ｄｕｐｌｅｘ状態にて本指示を受信するとＳｕｓｐｅｎｄ状態へ遷移する。
（Ｄ）コピー方向反転指示：Ｄｕｐｌｅｘ状態、Ｓｕｓｐｅｎｄ状態又は障害Ｓｕｓｐｅｎｄ状態にて本指示を受信すると、コピー元とコピー先との関係が反転する。Ｄｕｐｌｅｘ状態の場合は、本指示を受信することでコピー方向も反転する。
【００７７】
なお、初期化指示はコピー元の仮想化ストレージ装置１０００及びコピー元ボリュームと、コピー先の仮想化ストレージ装置１０００及びコピー先ボリュームとを指定することが考えられ、その他の指示については既にペア関係が出来上がっているため当該関係を示す識別子（コピー元の仮想化ストレージ装置１０００及びコピー元ボリュームと、コピー先の仮想化ストレージ装置１０００及びコピー先ボリュームとの組み合わせもその識別子の一つである）を指示すればよい。
【００７８】
＜５．ストレージ装置１５００で実行されるプログラム及び情報＞
図６にはストレージ装置１５００にて実行されるプログラム及び情報について記されているが、それぞれのプログラム及び情報は仮想化ストレージ装置１０００と同様の動作を

行う。
【００７９】
＜６．デバイス関係テーブル５００１＞
図８はデバイス関係テーブル５００１が有する情報を示した図である。デバイス関係テーブル５００１は、Ｉ／Ｏパスマネージャー５０００が提供するホスト１１００内で仮想的なボリューム（より正確には当該ボリュームに対応する識別子）毎に以下の情報を管理する。
【００８０】
（Ａ）ホスト１１００内で仮想的なボリュームの識別子
（Ｂ）関係ボリューム識別子リスト：上記ホスト１１００で仮想的なボリュームの実体となりうるストレージ装置１５００のボリュームの識別子が入る。なお、個々の識別子はＩ／Ｏパスマネージャー５０００の下位レイヤであるＨＢＡデバイスドライバー５０１０が割り当てた識別子を用いる。本実施の形態においては、正系の仮想化ストレージ装置１０００（１０００Ｌ）が有するボリュームと副系の仮想化ストレージ装置１０００（１０００Ｒ）が有するボリュームの識別子がリストアップされる（通常状態ならば）。
【００８１】
（Ｃ）正系ボリューム：（Ｂ）でリストアップしたどちらのボリュームが正系かを示す。
（Ｄ）障害状態
（Ｅ）ペア状態
【００８２】
なお、ファイルシステム５０２０の視点からは（Ａ）の識別子も（Ｂ）の識別子も同様の扱いとするため、（Ａ）や（Ｂ）の識別子はそれぞれ重複が許されない。また（Ａ）と（Ｂ）をあわせた場合にも重複が許されないため、Ｉ／Ｏパスマネージャー５０００はその点を考慮して（Ａ）の識別子を生成する必要がある。
【００８３】
＜７．初期化処理＞
図９は、Ｉ／Ｏパスマネージャー５０００の初期化処理について記したフローチャートである。以下、このフローチャートを参照して、かかる初期化処理について説明する。なお、以下においては各種処理の処理主体を「Ｉ／Ｏパスマネージャー５０００」として説明する場合があるが、実際上は、ホスト１１００のプロセッサ１１０１（図１）が「Ｉ／Ｏパスマネージャー５０００」というプログラムに基づいて対応する処理を実行することは言うまでもない。
【００８４】
（Ｓ９００１）Ｉ／Ｏパスマネージャー５０００は、管理ホスト１２００やホスト１１００のユーザーからの以下の情報を含んだ初期化指示を受信する。尚、二重化システムの初期化処理として、ＨＡ（ハイ　アベイラビリティ）初期化指示ともいう。
（Ａ）正系の仮想化ストレージ装置１０００とその中のボリューム
（Ｂ）副系の仮想化ストレージ装置１０００とその中のボリューム
【００８５】
（Ｓ９００２）Ｉ／Ｏパスマネージャー５０００は、Ｓ９００１で指示された仮想化ストレージ装置１０００の両方と通信をしてボリュームの存在の有無及び容量を取得する。
（Ｓ９００３）Ｉ／Ｏパスマネージャー５０００は、Ｓ９００１で指定されたボリュームが存在し、同容量であることを確認する。確認できない場合は、Ｉ／Ｏパスマネージャー５０００は指示発信元へエラーを返す。
【００８６】
（Ｓ９００４）Ｉ／Ｏパスマネージャー５０００は、仮想化ストレージ装置１０００の一つ又は両方に対して、リモートコピー初期化指示を送信する。この初期化指示には正系のボリュームをコピー元ボリューム、副系のボリュームをコピー先ボリュームとして指示を出す。本指示によって仮想化ストレージ装置１０００はリモートコピーを開始する。
【００８７】

（Ｓ９００５）Ｉ／Ｏパスマネージャー５０００は、デバイス関係テーブル５００１に以下の情報を登録し、その後初期化指示の発信元へ初期化開始応答を返す。
（Ａ）ホスト１１００内で仮想的なボリュームの識別子（＝Ｉ／Ｏパスマネージャー５０００が作成した値）
（Ｂ）関係ボリューム識別子リスト（＝Ｓ９００１で指定された仮想化ストレージ装置１０００とボリュームに対応する識別子が二つ（正系及び副系の両方））。
（Ｃ）正系ボリューム（＝Ｓ９００１で指定された正系ボリューム）の識別子
（Ｄ）障害状態（＝副系準備中）
（Ｅ）ペア状態（＝Ｉｎｉｔｉａｌ－Ｃｏｐｙｉｎｇ）

【００８８】

（Ｓ９００６）Ｉ／Ｏパスマネージャー５０００は、リモートコピーのペア状態を監視し、Ｄｕｐｌｅｘ状態に遷移したらデバイス関係テーブル５００１を以下の情報に更新する。
（Ｄ）障害状態（＝通常状態）
（Ｅ）ペア状態（＝Ｄｕｐｌｅｘ）

【００８９】

以上の処理によって、Ｉ／Ｏパスマネージャー５０００は、ユーザー指示に応じてリモートコピーの設定を含めた高可用化のための準備を開始することができる。なお、実際にはＳ９００５の直後にＩ／Ｏパスマネージャー５０００がホスト１１００内で仮想的なボリュームを提供できるため、ファイル形式でアクセスしたいユーザーは当該ボリュームに対するマウント指示等を出して、ファイルＩ／Ｏを開始することができる。また、別な方法としてＩ／Ｏパスマネージャー５０００はリモートコピー設定前に既に高可用化すべきボリュームに対応するホスト１１００内で仮想的なボリュームを定義し、ファイルシステム５０２０も当該ボリュームをマウントした状態から、ユーザーが副系となるボリュームを指定することによって上記の処理を開始してもよい。

【００９０】

＜８．ライトリクエスト処理フロー＞

図１０は、Ｉ／Ｏパスマネージャー５０００がファイルシステム５０２０からライトリクエストを受信した時の処理フローを示した図である。

【００９１】

（Ｓ１０００１）Ｉ／Ｏパスマネージャー５０００は、ファイルシステム５０２０より、ライト先となるホスト１１００内の仮想的なボリュームの識別子と、当該ボリュームのライト位置と、ライト長とを含むライトリクエスト関数を呼び出される（又はメッセージを受信する）。
（Ｓ１０００２）Ｉ／Ｏパスマネージャー５０００は、当該仮想的なボリュームの障害状態を確認し、リモートコピー失敗状態ならばＳ１００２０の両書き処理に制御を移し、それ以外ならばＳ１０００３を実行する。

【００９２】

（Ｓ１０００３）Ｉ／Ｏパスマネージャー５０００は、正系ボリュームに対してライトリクエストを発行する。なお、当該ライトリクエストの発行は実際は下位レイヤのＨＢＡデバイスドライバー５０１０を呼び出すことで実現する。
（Ｓ１０００４）Ｉ／Ｏパスマネージャー５０００は、ライトリクエストの応答を確認し、正常終了ならばファイルシステム５０２０に対して完了応答を返し、リモートコピー失敗ならＳ１００２０の両書き処理に制御を移し、無応答など、これ以外の場合はＳ１００１０の切り替え処理に制御を移す。

【００９３】

なお、Ｓ１００２０の両書き処理は以下のステップで実行される。
（Ｓ１００２１）リモートコピーの設定によって、正系又は副系のボリュームに対するライトが拒否されている場合は、Ｉ／Ｏパスマネージャー５０００はこの設定を解除する。
（Ｓ１００２２）Ｉ／Ｏパスマネージャー５０００は、正系ボリュームに対してライトリ

クエストを発行する。
（Ｓ１００２３）Ｉ／Ｏパスマネージャー５０００は、副系ボリュームに対してライトリクエストを発行する。Ｉ／Ｏパスマネージャー５０００は、正系と副系の両方からのライトリクエスト応答の到着を待って、ファイルシステム５０２０に対して完了応答を返す。
【００９４】
＜８．１．切り替え処理のフロー＞
以下、引き続き切り替え処理にて実現される処理を説明する。
【００９５】
（Ｓ１００１１）Ｉ／Ｏパスマネージャー５０００は、まず、デバイス関係テーブル５００１の障害状態を参照することで副系ボリュームが使用可能であるか確認し、使用不可能だと判断した場合はファイルシステム５０２０に対してエラー応答を返し、利用可能であればＳ１００１２を実行する。なお、使用不可能と判断できる状態としては、副系なし（障害によって副系の仮想化ストレージ装置１０００が機能してない場合や、初めから副系の仮想化ストレージ装置１０００を設定していないボリュームの場合）の状態や、前述の初期化準備中の状態がある。
【００９６】
（Ｓ１００１２）Ｉ／Ｏパスマネージャー５０００は、副系の仮想化ストレージ装置１０００に対してリモートコピーの停止指示を発行し、コピー状態がＳｕｓｐｅｎｄ状態となったことを確認後、コピー方向反転指示を指示する。
（Ｓ１００１３）Ｉ／Ｏパスマネージャー５０００は、副系の仮想化ストレージ装置１０００に対してリモートコピーの再同期指示を発行する。なお、実際に再同期が完了してペア状態がＤｕｐｌｅｘ状態に遷移するまで待つ必要はない。
【００９７】
（Ｓ１００１４）Ｉ／Ｏパスマネージャー５０００は、デバイス関係テーブル５００１の正系ボリューム識別子をこれまで副系であったボリューム識別子に更新し、正系と副系を入れ替える。そして新たに正系となったボリュームに対してライトリクエストを、ＨＢＡデバイスドライバー５０１０を通じて送信する。
（Ｓ１００１５）Ｉ／Ｏパスマネージャー５０００は、ライトリクエストの応答を確認し、正常終了ならばファイルシステム５０２０に対して完了応答を返し、エラーならばエラー応答を返して終了する。
【００９８】
＜８．１．１．両書き処理中のライトリクエスト失敗への対策＞
Ｓ１００２０の両書き処理中にＳ１００２２の正系ボリュームに対するライトリクエストが失敗に終わった場合は、Ｓ１００１０の切り替え処理に制御を移すことが考えられる。また、Ｓ１００２３の副系ボリュームに対するライトリクエストが失敗に終わった場合は、デバイス関係テーブル５００１の障害状態を'副系なし'に変更し、ライト完了とする。
【００９９】
また、両書き処理中はペア状態が障害Ｓｕｓｐｅｎｄ状態であるため、仮想化ストレージ装置１０００のボリュームにはリモートコピーの差分ビットマップによってライト位置が記される。しかし、両書き処理によって両ボリュームに書かれるライトデータは同一であるため、両書き処理が正常に行われている間はこの差分ビットマップへの記録を回避し、通信障害回復後の再同期処理では差分データだけコピーできるようにすることが望ましい。その解決策として、両書き処理が正常に行われている間は正系と副系両方の仮想化ストレージ装置１０００の当該ボリュームの差分ビットマップを一定時間ごとに繰り返しクリアすることが考えられる。この方式ではクリア指示をライトリクエスト毎に発行する必要がなく、かつリモートコピーの再同期では対象ボリュームの全領域コピーは回避できる。なぜならば、直近に実施したクリア以後に行われた両書きのライトリクエストは両書きが失敗したライトリクエストと共にライト位置が差分ビットマップに記録されるが、両書きにて記録されたデータ領域が再同期でコピーされた場合もコピー先のデータ内容が変わ

らないため、データ不整合やコピー漏れ領域が発生しないからである。

【０１００】

なお、上記解決策では正系と副系両方の差分ビットマップをクリアするために一時的にライトリクエストの処理を停止してもよい。その停止方法としてはＩ／Ｏパスマネージャー５０００がファイルシステム５０２０から受け取ったライトリクエストを、両方の差分ビットマップのクリアが完了するまで、仮想化ストレージ装置１０００へ転送しない方法が考えられるし、正系の仮想化ストレージ装置１０００にて、両方の差分ビットマップのクリアが完了するまでライトリクエストの処理を保留する方法も考えられる。

【０１０１】

第２の回避策としては、正系と副系のボリュームに対してそれぞれ２面の差分ビットマップを割り当てる方式がある。以下にその処理内容を示す。

【０１０２】

（初期状態）正系と副系の仮想化ストレージ装置１０００は、それぞれ２面の差分ビットマップの片面に対してライトリクエストの位置を記録する。そのために、両仮想化ストレージ装置１０００は、アクティブ面（ライトリクエスト到着時にライト位置を記録する面を指し、もう一面の差分ビットマップは非アクティブ面と呼ぶ）に関する情報を保持・管理する。また、非アクティブ面の差分ビットマップは何も記録されていない状態が望ましい。

【０１０３】

（Ｓｔｅｐ１）正系の仮想化ストレージ装置１０００は、アクティブ面の管理情報を非アクティブ面になっていたもう一つの差分ビットマップへ更新することで、ライトリクエストの位置の記録先となる差分ビットマップを切り替え、以後のライトリクエストは切り替え後の差分ビットマップへ記録する。副系の仮想化ストレージ装置１０００も同様に切り替える。なお、当該切り替え処理開始の契機はＩ／Ｏパスマネージャー５０００が両仮想化ストレージ装置１０００へ与える。なお、正系と副系の切り替え処理はどちらが先に実行してもよく、並列に実行してもよい。

【０１０４】

（Ｓｔｅｐ２）Ｉ／Ｏパスマネージャー５０００は、両仮想化ストレージ装置１０００からの切り替え完了の応答を待ってから、両仮想化ストレージ装置１０００に対して差分ビットマップのクリア指示を出す。クリア指示を受信した仮想化ストレージ装置１０００は、非アクティブ面となっている差分ビットマップのライト位置をクリアし、Ｉ／Ｏパスマネージャー５０００へ応答を返す。切り替え処理と同様に、正系と副系のクリア処理はどちらが先に実行してもよく、並列に実行してもよい。

【０１０５】

（Ｓｔｅｐ３）Ｉ／Ｏパスマネージャー５０００は、両仮想化ストレージ装置１０００からのクリア完了の応答を待ち、時間経過後にＳｔｅｐ１から再度実行する。

【０１０６】

本解決策の場合、通信障害回復後の再同期処理では、正系と副系のビットマップ４面の論理和を計算することで、Ｄｕｐｌｅｘ‐Ｐｅｎｄｉｎｇ状態中に差分コピーを行う領域を決定することができる。また本方式ではビットマップの面数が多いものの、ライトリクエストの保留は必要ない。

【０１０７】

第３の解決策としては、上記第２の解決策の変形の以下の方式がある。

（初期状態）正系及び副系の仮想化ストレージ装置１０００は、それぞれ２面の差分ビットマップの両面に対してライトリクエストの位置を記録する。また、両仮想化ストレージ装置１０００は前回クリアを行った差分ビットマップ面に関する情報を保持・管理しておく。

【０１０８】

（Ｓｔｅｐ１）Ｉ／Ｏパスマネージャー５０００は、両仮想化ストレージ装置１０００に対して差分ビットマップのクリア指示を出す。クリア指示を受信した仮想化ストレージ装

置１０００は、前回クリアした差分ビットマップでないもう一つの差分ビットマップのライト位置をクリアし、Ｉ／Ｏパスマネージャーへ応答を返す。
（Ｓｔｅｐ３）Ｉ／Ｏパスマネージャー５０００は、両仮想化ストレージ装置１０００からのクリア完了の応答を待ち、時間経過後にＳｔｅｐ１から再度実行する。
【０１０９】
＜９．リードリクエスト処理フロー＞
図１１はＩ／Ｏパスマネージャー５０００がファイルシステム５０２０からリードリクエストを受信したときの処理内容を示すフローチャートである。
【０１１０】
（Ｓ１１００１）Ｉ／Ｏパスマネージャー５０００は、ファイルシステム５０２０より、リード先となるホスト内の仮想的なボリュームの識別子と、当該ボリュームのライト位置と、ライト長とを含むライトリードリクエスト関数を呼び出される（又はメッセージを受信する）。
【０１１１】
（Ｓ１１００２）Ｉ／Ｏパスマネージャー５０００は、当該仮想的なボリュームの障害状態を確認し、通常状態でかつ正系ボリュームに対するＩ／Ｏ負荷が高い場合（たとえば、一定ＩＯＰＳを超える場合や一定帯域を超える場合等）と判断したときにはＳ１１０２１を実行し、それ以外の状態（副系なし、副系準備中、通常状態等）のときにはＳ１１００３を実行する。
【０１１２】
（Ｓ１１００３）Ｉ／Ｏパスマネージャー５０００は、正系ボリュームに対してリードリクエストを発行する。
（Ｓ１１００４）Ｉ／Ｏパスマネージャー５０００は、リードリクエストの応答を確認し、正常終了ならばファイルシステム５０２０に対して完了応答を返し、それ以外ならばＳ１１０１０の切り替え処理に制御を移す。
【０１１３】
（Ｓ１１０２１）Ｉ／Ｏパスマネージャー５０００は、副系ボリュームに対してリードリクエストを発行する。
（Ｓ１１０２２）Ｉ／Ｏパスマネージャー５０００は、リードリクエストの応答を確認し、正常終了ならばファイルシステム５０２０に対して完了応答を返し、それ以外ならばＳ１１０２３を実行する。
（Ｓ１１０２３）Ｉ／Ｏパスマネージャー５０００は、デバイス関係テーブル５００１の障害状態を'副系なし'に更新し、Ｓ１１００３を実行する。
【０１１４】
＜９．１．切り替え処理のフロー＞
以下、引き続き切り替え処理にて実現される処理を説明する。
（Ｓ１１０１１）Ｉ／Ｏパスマネージャー５０００は、まず、デバイス関係テーブル５００１の障害状態を参照することで副系ボリュームが使用可能であるか確認し、使用不可能だと判断した場合はファイルシステム５０２０に対してエラー応答を返し、利用可能だと判断した場合はＳ１１０１２を実行する。なお、使用不可能と判断できる状態としては、副系なし（障害によって副系の仮想化ストレージ装置１０００が機能してない場合や、初めから副系の仮想化ストレージ装置１０００を設定していないボリュームの場合）の状態や、前述の初期化準備中の状態がある。
【０１１５】
（Ｓ１００１２）Ｉ／Ｏパスマネージャー５０００は、副系の仮想化ストレージ装置１０００に対してリモートコピーの停止指示を発行し、コピー状態がＳｕｓｐｅｎｄ状態となったことを確認後、コピー方向反転指示を指示する。
（Ｓ１００１３）Ｉ／Ｏパスマネージャー５０００は、副系の仮想化ストレージ装置１０００に対してリモートコピーの再同期指示を発行する。なお、実際に再同期が完了してペア状態がＤｕｐｌｅｘ状態に遷移するまで待つ必要はない。

【０１１６】
（Ｓ１００１４）Ｉ／Ｏパスマネージャー５０００は、デバイス関係テーブル５００１の正系ボリューム識別子をこれまで副系であったボリュームの識別子に更新し、正系と副系を入れ替える。そして新たに正系となったボリュームに対してリードリクエストを、ＨＢＡデバイスドライバー５０１０を通じて送信する。
（Ｓ１００１５）Ｉ／Ｏパスマネージャー５０００は、リードリクエストの応答を確認し、正常終了ならばファイルシステム５０２０に対して完了応答を返し、エラーならばエラー応答を返して終了する。
【０１１７】
＜１０．障害対策処理フロー＞
本章では、Ｉ／Ｏパスマネージャー５０００が障害を検知してから回復を完了するまでの処理の流れを説明する。なお、本処理は定期的にバックグラウンドで実行される。
【０１１８】
＜１０．１．仮想化ストレージ装置１０００間のネットワーク障害＞
（Ｓｔｅｐ１）Ｉ／Ｏパスマネージャー５０００は、リモートコピーのペア状態を監視し、障害Ｓｕｓｐｅｎｄ状態を発見することで何らかの障害発生を検知する。
【０１１９】
（Ｓｔｅｐ２）Ｉ／Ｏパスマネージャー５０００は、副系の仮想化ストレージ装置１０００に対してリモートコピーの停止指示を発行し、コピー状態がＳｕｓｐｅｎｄ状態となった事を確認後、コピー方向を反転し、各仮想化ストレージ装置１０００に対して状態問い合わせを行い、仮想化ストレージ装置１０００自体に障害が発生しておらず、ネットワーク障害が原因であることを確認したら、デバイス関係テーブル５００１の障害状態を'リモートコピー失敗'に更新する。なお、本処理はストレージ管理者が行った作業結果を利用してもよい。
【０１２０】
（Ｓｔｅｐ３）当該ネットワークが回復するまで待つ。
（Ｓｔｅｐ４）Ｉ／Ｏパスマネージャー５０００は、正系の仮想化ストレージ装置１０００に対してペアの再同期指示を発行する。
（Ｓｔｅｐ５）Ｉ／Ｏパスマネージャー５０００は、デバイス関係テーブル５００１の障害状態を'副系準備中'に更新する。
（Ｓｔｅｐ６）Ｉ／Ｏパスマネージャー５０００は、ペア状態がＤｕｐｌｅｘになるまで待った後に、デバイス関係テーブル５００１の障害状態を'通常状態'に更新する。
【０１２１】
＜１０．２．正系仮想化ストレージ装置１０００の障害停止＞
（Ｓｔｅｐ１）Ｉ／Ｏパスマネージャー５０００は、正系の仮想化ストレージ装置１０００の状態を監視することで障害発生を検知する。
（Ｓｔｅｐ２）Ｉ／Ｏパスマネージャー５０００は、デバイス関係テーブル５００１の正系ボリュームの識別子を副系ボリュームの識別子に変更することで以後のＩ／Ｏリクエスト先を副系の仮想化ストレージ装置１０００に切り替え、さらに障害状態を'副系なし'に更新する。
（Ｓｔｅｐ３）Ｉ／Ｏパスマネージャー５０００は、旧正系（Ｓｔｅｐ２にて切り替えたので現副系）の仮想化ストレージ装置１０００が回復するまで待つ。
【０１２２】
（Ｓｔｅｐ４）Ｉ／Ｏパスマネージャー５０００は、正系の仮想化ストレージ装置１０００に対してペアの再同期指示又は初期化指示を発行する。
（Ｓｔｅｐ５）Ｉ／Ｏパスマネージャー５０００は、デバイス関係テーブル５００１の障害状態を'副系準備中'に更新する。
（Ｓｔｅｐ６）Ｉ／Ｏパスマネージャー５０００は、ペア状態がＤｕｐｌｅｘになるまで待った後に、デバイス関係テーブル５００１の障害状態を'通常状態'に更新する。
【０１２３】

＜１０．３．副系仮想化ストレージ装置１０００の障害停止＞
（Ｓｔｅｐ１）Ｉ／Ｏパスマネージャー５０００は、副系の仮想化ストレージ装置１０００の状態を監視することで障害発生を検知する。
（Ｓｔｅｐ２）Ｉ／Ｏパスマネージャー５０００は、デバイス関係テーブル５００１の障害状態を'副系なし'に更新する。
（Ｓｔｅｐ３）Ｉ／Ｏパスマネージャー５０００は、副系の仮想化ストレージ装置１０００が回復するまで待つ。
【０１２４】
（Ｓｔｅｐ４）Ｉ／Ｏパスマネージャー５０００は、正系の仮想化ストレージ装置１０００に対してペアの再同期指示又は初期化指示を発行する。
（Ｓｔｅｐ５）Ｉ／Ｏパスマネージャー５０００は、デバイス関係テーブル５００１の障害状態を'副系準備中'に更新する。
（Ｓｔｅｐ６）Ｉ／Ｏパスマネージャー５０００は、ペア状態がＤｕｐｌｅｘになるまで待った後に、デバイス関係テーブル５００１の障害状態を'通常状態'に更新する。
【０１２５】
＜１１．もう一つの初期化方法＞
これまでの説明では、Ｉ／Ｏパスマネージャー５０００に出された初期化要求に応じて仮想化ストレージ装置１０００にリモートコピーの設定を行ったが、以下に示す逆の方法も考えられる。
【０１２６】
（Ｓｔｅｐ１）管理ホスト１２００は、仮想化ストレージ装置１０００に対してリモートコピーのペア初期化指示を出すことで、リモートコピーを開始する。
（Ｓｔｅｐ２）Ｉ／Ｏパスマネージャー５０００は、スキャニング要求を受信する。
（Ｓｔｅｐ３）Ｉ／Ｏパスマネージャー５０００は、ＨＢＡデバイスドライバー５０１０を通じて各ボリュームに対するリモートコピーの設定（リモートコピー設定の有無やコピー元かコピー先か、ペアの相手となる仮想化ストレージ装置１０００とボリューム）を取得する。なお、この取得方法としてＩ／Ｏネットワーク上でＳＣＳＩコマンドを使うことも考えられるし、それ以外の通信ネットワークを用いて情報を取得してもよい。
【０１２７】
（Ｓｔｅｐ４）Ｉ／Ｏパスマネージャー５０００は、前ステップで取得した情報を元に、デバイス関係テーブル５００１を作成し、これまで説明してきた処理を開始する。なお、当該デバイス関係テーブル５００１の作成例としては以下がある。
（Ａ）ホスト１１００内で仮想的なボリュームの識別子＝Ｉ／Ｏパスマネージャー５０００が作成した値
（Ｂ）関係ボリューム識別子リスト＝リモートコピーのコピー元ボリュームとコピー先ボリュームの識別子
（Ｃ）正系ボリューム＝リモートコピーのコピー元ボリューム
（Ｄ）障害状態＝仮想化ストレージ装置１０００から取得したペア状態がＤｕｐｌｅｘ状態ならば'通常状態'、Ｉｎｉｔｉａｌ‐Ｃｏｐｙｉｎｇ又はＤｕｐｌｅｘ‐Ｐｅｎｄｉｎｇ状態ならば'副系準備中'、Ｓｕｓｐｅｎｄ又は障害Ｓｕｓｐｅｎｄ状態ならば'リモートコピー失敗'
（Ｅ）ペア状態＝仮想化ストレージ装置１０００から取得したペア状態
【０１２８】
以上、これまで説明したハードウェア及びプログラムの動作によって本実施の形態では高可用性を実現する。なお、図１０と図１１等に記した切り替え処理に長時間要する場合の対策として、Ｉ／Ｏパスマネージャー５０００がＩ／Ｏリクエストを再送信する必要が出てきた場合に、予備処理として前記切り替え処理の一部を実行してもよい。この場合、再送信したＩ／Ｏリクエストが正常応答で返ってきた場合は先行して行った切り替え処理を元に戻せば良く、一方で再送信したＩ／Ｏリクエストがエラー応答で返ってきたり、まったく応答がなければ前記切り替え処理の残り部分を実行すればよい。また、本実施の形

態は全てのボリュームが仮想化ストレージ装置１０００によって仮想化され、実体がストレージ装置１５００にある仮想ボリュームで、仮想化ストレージ装置１０００は仮想化専用のアプライアンスであってもよく、またその逆に全てのボリュームの実体が仮想化ストレージ装置１０００の内部にある構成であってもよい。また、仮想化ストレージ装置１０００が提供するボリュームには容量以外にもさまざまな属性が設定されることがある（たとえば、エミュレーションタイプやＳＣＳＩ規格で定められたＩｎｑｕｉｒｙコマンドで取得可能なボリューム識別番号がある）。

【０１２９】

こうした属性情報や属性変更もリモートコピーによって正系の仮想化ストレージ装置から副系の仮想化ストレージ装置へ転送し、両方の仮想化ストレージ装置にて管理することも考えられる。

【０１３０】

＜１２．もう一つのリード・ライト処理＞

図１０や図１１に記したライト・リード処理では、Ｉ／Ｏパスマネージャー５０００が明示的にリモートコピーの操作を仮想化ストレージ装置１０００へ転送する。しかし、当該リモートコピーの操作が仮想化ストレージ装置１０００のベンダー毎に異なる場合があるため、Ｉ／Ｏパスマネージャー５０００のライト処理やリード処理に含めないほうが好ましい場合がある。図２５～図２７にこうした形態での処理内容を示す。なお、以下においては各種処理の処理主体を「仮想化ストレージ装置１０００」として説明する場合があるが、実際上は、その仮想化ストレージ装置１０００内のプロセッサ１０１１（図１）がメモリ１０１２（図１）に格納されたプログラムに基づいて対応する処理を実行することは言うまでもない。

【０１３１】

＜１２．１．Ｉ／Ｏパスマネージャーのライト処理＞

図２５は、Ｉ／Ｏパスマネージャー５０００で実行される図１０の大体処理を示したフローチャートである。以下の点が図１０と異なる。

（相違点１）リモートコピーの操作Ｓ１００１２、Ｓ１００１３，Ｓ１００２１がスキップされる。

（相違点２）リモートコピー失敗時のフローＳ１００２０に到達しない。ただし、本相違点は通常のリード／ライト処理ではリモートコピー失敗を意味するエラーメッセージを識別できない場合に限った話である。

【０１３２】

＜１２．２．ストレージ装置１０００の処理＞

図２７は、仮想化ストレージ装置１０００がライトリクエストを受信した時に行うリモートコピーの操作について示した図である。

【０１３３】

（Ｓ２７００１）仮想化ストレージ装置１０００は、ライトリクエストを受信する。

（Ｓ２７００２）仮想化ストレージ装置１０００は、ライトリクエストが対象とするボリュームがリモートコピーに関係するかどうか判断し、無関係の場合はＳ２７００３を実行し、関係する場合はＳ２７００４を実行する。

（Ｓ２７００３）仮想化ストレージ装置１０００は、通常のライト処理を行い、ホスト１１００へ応答を返して終了する。

【０１３４】

（Ｓ２７００４）仮想化ストレージ装置１０００は、ライトリクエストが対象とするボリュームのリモートコピーの属性を判断し、コピー元属性の場合はＳ２７００５を実行し、コピー先属性の場合はＳ２７０１１を実行する。

（Ｓ２７００５）仮想化ストレージ装置１０００は、同期リモートコピー処理を実行し、副系ストレージへライトデータを転送し、応答を待つ。

（Ｓ２７００６）仮想化ストレージ装置１０００は、コピーが成功したかどうか判断し、成功ならばＳ２７００８を実行し、失敗ならばＳ２７００７を実行する。

（Ｓ２７００７）仮想化ストレージ装置１０００は、対象ボリュームがコピー元となるリモートコピーペアの状態を障害Ｓｕｓｐｅｎｄ状態に遷移する。ただし、当該ボリュームに対するライトは禁止しない。

【０１３５】

（Ｓ２７００８）仮想化ストレージ装置１０００は、通常のライト処理を行い、ホスト１１００へ応答を返して終了する。

（Ｓ２７０１１）仮想化ストレージ装置１０００は、リモートコピーを停止し、コピー元とコピー先の関係を反転する。

（Ｓ２７０１２）仮想化ストレージ装置１０００は、再同期処理を開始する。

（Ｓ２７０１３）仮想化ストレージ装置１０００は、通常のライト処理を行い、ホスト１１００へ応答を返して終了する。

【０１３６】

なお、Ｓ２７０１２の再同期処理は完了まで待たなくても良い。なぜならば、Ｓ２７０１２を実行する仮想化ストレージ装置１０００は副系であり、正系の仮想化ストレージ装置１０００が正常動作しているとは限らないこと、及び再同期処理が完了するまでの時間が長いことが考えられるからである。なお、こうしたケースは＜１０．障害対策処理フロー＞で述べた処理によって回復される点はこれまでと同じである。

【０１３７】

＜１２．３．Ｉ／Ｏパスマネージャーのリード処理＞

図２６は、Ｉ／Ｏパスマネージャー５０００で実行される図１１の大体処理を示したフローチャートである。以下の点が図１１と異なる。

（相違点１）リモートコピーの操作Ｓ１１０１２、Ｓ１１０１３がスキップされる。

【０１３８】

なお、図１１ではリード処理に応じてリモートコピーの向きが反転したが、本処理では反転させない。なぜならば、副系の仮想化ストレージ装置１０００に対するリードリクエストは正系の仮想化ストレージ装置１０００が（ホスト＝仮想化ストレージ装置間の通信障害による原因を含めて）応答を返さない場合に加えて、正系の仮想化ストレージ装置１０００の過負荷が原因の場合もあるからである。そのため、副系の仮想化ストレージ装置１０００がコピー先ボリュームに対するリードリクエストを契機としてリモートコピーのペア反転を行うと、たまたま副系の仮想化ストレージ装置１０００に出されたリードリクエストでペアが反転し、その次のリードリクエストで再びペアが反転してしまうため、リード性能が悪化する結果となるからである。

【０１３９】

ただし、Ｓ１１０２１の実行が抑制される場合は、仮想化ストレージ装置１０００はリード処理に際して以下の処理を行うことでリモートコピーのペア反転を行っても良い。

（Ｓｔｅｐ１）仮想化ストレージ装置１０００は、リードリクエストを受信する。

（Ｓｔｅｐ２）仮想化ストレージ装置１０００は、通常のリード処理を行う。

（Ｓｔｅｐ３）仮想化ストレージ装置１０００は、リード対象のボリュームがリモートコピーのコピー先ボリュームであるかどうかを判断し、該当する場合は次のＳｔｅｐ４を実行し、そうでない場合は終了する。

（Ｓｔｅｐ４）仮想化ストレージ装置１０００は、リモートコピーを停止し、コピー元とコピー先の関係を反転する。

【０１４０】

（２）第２の実施の形態

次に第２の実施の形態について図１２を用いて説明する。第１の実施の形態と異なる点は、ストレージ装置１５００Ｌが複数の仮想化ストレージ装置１０００Ｌ，１０００Ｒに接続され、これら仮想化ストレージ装置１０００Ｌ，１０００Ｒがストレージ装置１５００Ｌ内のボリュームを共有することによって、仮想化ストレージ装置１０００Ｌ，１０００Ｒの片方が停止した場合でも第１の実施の形態よりも低コストでサービスが継続できるようになる点である。

【０１４１】
ただし、仮想化ストレージ装置１０００Ｌ，１０００Ｒはキャッシュメモリ１０２０Ｌ，１０２０Ｒを有するため、仮想化ボリュームに対してライトデータを書き込んだ直後に正系の仮想化ストレージ装置１０００Ｌが災害停止した場合に備えて、ライトデータを副系の仮想化ストレージ装置１０００Ｒのキャッシュメモリ１０２０Ｒにも保存する必要があり、また両方の仮想化ストレージ装置１０００Ｌ，１０００Ｒのデステージングやステージングに対して工夫が必要となる。
【０１４２】
通常状態におけるライトリクエストは以下のステップにて処理される。
【０１４３】
（Ｓｔｅｐ１）ホスト１１００からライトリクエストを受信した正系の仮想化ストレージ装置１０００Ｌは当該ライトリクエストが当該仮想化ストレージ装置１０００Ｌ内部のＨＤＤ１０３０に対応するボリューム３０００ＬＡ宛なのか、両方の仮想化ストレージ装置１０００Ｌ，１０００Ｒがストレージ装置１５００Ｌのボリューム３５００Ｌを共有して提供する仮想化ボリューム（以後、共有仮想化ボリュームと呼ぶ）３０００ＬＢ宛なのか、通常の仮想化ボリューム宛なのかを判断する。なお、共有仮想化ボリューム３０００ＬＢ以外の処理については第１の実施の形態と同様の処理を行う。
【０１４４】
（Ｓｔｅｐ２）正系の仮想化ストレージ装置１０００Ｌは自身のキャッシュメモリ１０２０Ｌに当該ライトデータを保存すると共に、当該ライトデータをリモートコピープログラムによって副系の仮想化ストレージ装置１０００Ｒのキャッシュメモリ１０２０Ｒに保存した後に、ホスト１１００に対して正常応答を返す。
【０１４５】
（Ｓｔｅｐ３）正系の仮想化ストレージ装置１０００Ｌのキャッシングアルゴリズムがデステージすべきキャッシュメモリ１０２０Ｌ上のデータを決定し、当該データをストレージ装置１５００Ｌのボリュームにデステージする。
【０１４６】
（Ｓｔｅｐ４）デステージ完了後、正系の仮想化ストレージ装置１０００Ｌはデステージしたキャッシュメモリ１０２０Ｌ上のデータのアドレスを破棄するように副系の仮想化ストレージ装置１０００Ｒに指示する。なお、指示を受けた副系の仮想化ストレージ装置１０００Ｒは指示を受けたデータをキャッシュメモリ１０２０Ｒから破棄する。
【０１４７】
なお、本構成では仮想化ストレージ装置１０００Ｌ，１０００Ｒ間のネットワークが切断された状態で副系の仮想化ストレージ装置１０００ＲにＩ／Ｏリクエストの切り替えを行った場合、仮想化ストレージ装置１０００Ｌ，１０００Ｒの両方が正系として自立的にデステージングを行う場合がある。そういった状況を回避するため、両仮想化ストレージ装置１０００Ｌ，１０００Ｒは自らを正系として処理する場合は先にストレージ装置１５００Ｌ内のかかる共有化されたボリューム３５００Ｌに対してＳＣＳＩ　Ｒｅｓｅｒｖｅ等の機能を用いて排他制御を行ってもよい。また、これ以外の方式として共有仮想化ボリューム３０００ＬＢについては仮想化ストレージ装置１０００Ｌのキャッシングを無効化してもよく、この場合は当該共有仮想ボリューム３０００ＬＢのアクセス権限がリードオンリーのアクセス権限へ変更された場合は当該変更に応じてキャッシングを有効にすることが考えられる。
【０１４８】
（３）第３の実施の形態
次に第３の実施の形態について図１３を用いて説明する。本実施の形態はこれまでの実施の形態に記した情報システムをこれまでのプロダクションサイトと異なる遠隔地（バックアップサイト）に別途用意し、リモートコピーを行うもので、これによりプロダクションサイト被災時にバックアップサイトでサービスを再開することができる。
【０１４９】

なお、これ以後の説明では、上述の「仮想化ストレージ装置」をストレージ装置と、「コピー元ボリューム」を正ボリュームと、「コピー先ボリューム」を副ボリュームと、「正系」をアクティブ側と、「副系」をスタンバイ側と呼ぶことがある。また、プロダクションサイトとバックアップサイトの情報システムをあわせてリモートコピーシステムと呼ぶことがある。

【０１５０】

＜１．リモートコピーシステムの構成＞

本実施の形態では、各サイトはホスト１３０１０，１３０２０と複数のストレージサブシステム１３００１，１３００２，１３００３，１３００４とから構成されている。そしてプロダクションサイトでは、ストレージサブシステム１３００１，１３００２同士でこれまで説明してきた高可用化構成を採用している。またバックアップサイトでも同様に、ストレージサブシステム１３００３，１３００４同士でかかる高可用化構成を採用している。

【０１５１】

さらに本実施の形態では、プロダクションサイトのアクティブ側のストレージサブシステム（コピー元ボリュームを持つ）１３００１からバックアップサイトのアクティブ側のストレージサブシステム（コピー先ボリュームを持つ）１３００３に対して同期又は非同期リモートコピーを行う。そしてプロダクションサイト被災時にはバックアップサイトのホスト１３０１０が高可用構成のストレージサブシステム１３００３，１３００４のいずれかアクティブな側に対してＩ／Ｏリクエストを発行することで、再起動したアプリケーション２０１０が処理を再開する。

【０１５２】

なお、前述の通り、ストレージサブシステムとは仮想化ストレージ装置１０００（図１）の仮想化機能を用いない設定の構成や、仮想化ストレージ装置１０００とストレージ装置１５００（図１）の組み合わせで仮想化ストレージ装置１０００が仮想化機能を用いて仮想化ボリュームを提供している構成のどちらの概念も含んだものとして呼んでいる。また、本実施の形態では個々のストレージサブシステム１３００１、１３００２、１３００３、１３００４が別々な内部構成（例えば、ストレージサブシステム１３００１だけ仮想化ストレージ装置１０００のみで構成し、仮想化機能を用いない場合や、バックアップサイトのストレージサブシステム１３００３と１３００４でストレージ装置１５００（図１）を共有し、プロダクションサイト側では共有しない場合）を採用してもよい。

【０１５３】

なお、以下においては各種処理の処理主体を「ストレージサブシステム」として説明する場合があるが、実際上は、そのストレージサブシステム内のプロセッサが当該ストレージサブシステム内のメモリに格納されたプログラムに基づいて対応する処理を実行することは言うまでもない。

【０１５４】

＜２．処理＞

プロダクションサイトのホスト１３０１０のアプリケーション２０１０がライトリクエストを発行すると、ＯＳによってプロダクションサイト内のアクティブ側のストレージサブシステムを判断し、そちらにライトリクエストを転送する。なお、本図ではストレージサブシステム１３００１がこれに対応する。

【０１５５】

プロダクションサイトのアクティブ側のストレージサブシステム１３００１は同期リモートコピーによってライトデータをプロダクションサイト内のスタンバイ側のストレージサブシステム（本図では１３００２が対応する）へ転送する。また、アクティブ側のストレージサブシステム１３００１はバックアップサイトのアクティブ側のストレージサブシステム（本図では１３００３が対応する）へ向けて同期又は非同期のリモートコピーとしてライトデータを転送する（本実施の形態による高可用構成ではアクティブ側にのみライトリクエストを処理するようにしているため、リモートコピーであっても同様にアクティ

ブ側にて処理を行う）。ライトデータを受信したバックアップサイト内のアクティブ側のストレージサブシステム１３００３は受け取ったライトデータをサイト内のスタンバイ側のストレージサブシステム１３００４へ同期リモートコピーによって転送する。

【０１５６】

そのため、プロダクションサイトのストレージサブシステム１３００１，１３００２はバックアップサイトのアクティブ側のストレージサブシステムを把握しており、バックアップサイトのストレージサブシステム１３００３，１３００４も想定外のストレージサブシステムからのリモートコピーを受け付けないために、プロダクションサイトのアクティブなストレージサブシステム（ストレージサブシステム１３０１）を把握している。

【０１５７】

以上の処理によってプロダクションサイト、バックアップサイト共にサイト内の高い可用性を実現している。ただしバックアップサイト側では、コスト削減のために高可用構成をとらない構成であってもよい。

【０１５８】

＜３．非同期リモートコピー＞

これまで説明してきた同期リモートコピーとは異なり、非同期リモートコピーはホスト１３０１０からのライトリクエストが到着した時点でライトデータを転送するのではなく、当該リクエスト完了応答後に転送する（言い方を変えると、非同期リモートコピーはホスト１３０１０へのリクエスト応答とは独立なタイミングでライトデータを転送する）。そのため、非同期リモートコピーはサイト間の距離が長く通信遅延が大きな場合でもライトリクエストの応答時間を低下させずにリモートコピーを行うことができる。しかし、非同期リモートコピーではプロダクションサイト側のストレージサブシステム１３００１にてライトデータをバッファリングする必要がある。このライトデータのバッファリング方式としては以下が考えられる。

【０１５９】

（１）プロダクションサイトのストレージサブシステム１３００１は、コピー元ボリュームへのライトデータとライトデータの順序情報を含むジャーナルを作成し、これを自身のキャッシュメモリ又は専用ボリュームに保存すると共に、このジャーナルをバックアップサイトのストレージサブシステム１３００３へ転送し、バックアップサイトのストレージサブシステム１３００３はジャーナルの順序情報を参考にコピー先ボリュームへライトデータを保存する。これにより、プロダクションサイト災害時にはライト順序が守られた（より正確には依存関係のあるライトデータ）データをバックアップサイト側で提供できる。

【０１６０】

（２）プロダクションサイトのストレージサブシステム１３００１は、ある期間毎のコピー元ボリュームへライトされたデータをグループ化して自身のキャッシュメモリ又は専用ボリュームへ保存し、非同期にバックアップサイトのストレージサブシステム１３００３へ転送し、当該グループ単位でバックアップサイトのストレージサブシステム１３００３が有するコピー先ボリュームへデータを保存する。

【０１６１】

そのため、これら非同期リモートコピーのためにバッファリングされるライトデータもスタンバイ側のストレージサブシステム１３００２で保持しなければ、アクティブ側ストレージサブシステム１３００１が停止したときに非同期リモートコピーを引き継ぐことができない。よって、プロダクションサイトのアクティブ側のストレージサブシステム１３００１はライトデータだけではなく、コピー先ボリュームの情報や、前述の順序情報や、グループ化するタイミング等をスタンバイ側のストレージサブシステム１３００２へ伝え、スタンバイ側のストレージサブシステム１３００２はそれに従ってアクティブ側と同じ非同期リモートコピーのためのバッファリングデータを作成する。

【０１６２】

なお、バックアップサイトのストレージサブシステム１３００３もプロダクションサイ

トから受け取ったライトデータを直ぐにコピー先ボリュームへ保存せずに、バッファリングを行うため、プロダクションサイト側と同様にアクティブ側の指示に従ってスタンバイ側も同様のバッファリングデータを作成し、また同様のタイミングでコピー先ボリュームにライトデータを保存する必要がある。

【０１６３】

（４）第４の実施の形態

次に第４の実施の形態について図１４を用いて説明する。本実施の形態では、２台のストレージ装置により先に説明した同期リモートコピーを用いて冗長構成された情報システムにおいて、ストレージ装置が提供する機能を制御するインターフェース（機能Ｉ／Ｆ）の構成について述べる。

【０１６４】

なお、本実施の形態から第１４の実施の形態までは、これまで仮想化ストレージ装置１０００Ｌ，１０００Ｒ、ストレージ装置１５００Ｌ，１５００Ｒと呼んでいたコンポーネントを、それぞれストレージ装置１５０００Ａ，１５０００Ｂ及び外部ストレージ装置１６０００Ａ，１６０００Ｂと呼ぶ。また、以下においては各種処理の処理主体を「ストレージ装置１５０００Ａ，１５０００Ｂ」や「外部ストレージ装置１６０００Ａ，１６０００Ｂ」として説明する場合があるが、実際上は、そのストレージ装置１５０００Ａ，１５０００Ｂ内の図示しないプロセッサやその外部ストレージ装置１６０００Ａ，１６０００Ｂ内のプロセッサが当該ストレージ装置１５０００Ａ，１５０００Ｂ又は外部ストレージ装置１６０００Ａ，１６０００Ｂ内のメモリに格納されたプログラムに基づいて対応する処理を実行することは言うまでもない。

【０１６５】

本実施の形態は、ホスト１４０００からの機能制御要求が、ストレージ装置１５０００Ａに送信された後、ストレージ装置１５０００Ａが機能制御要求をストレージ装置１５０００Ｂに転送し、ストレージ装置１５０００Ａ，１５０００Ｂの双方が当該機能制御要求を解釈し実行する例を示している。

【０１６６】

コマンドデバイス１５００２Ａ，コマンドデバイス１５００２Ｂはそれぞれストレージ装置１５０００Ａ、ストレージ装置１５０００Ｂが提供する論理ボリュームであり、機能を制御するホスト１４０００とのインターフェースとなる。なお、本実施の形態ではコマンドデバイス１５００２Ａがアクティブ側と仮定している。

【０１６７】

また、同期リモートコピーにより、コマンドデバイス１５００２Ａの内容はコマンドデバイス１５００２Ｂの内容と常に一致している。コマンドデバイス１５００２Ａ、コマンドデバイス１５００２Ｂはオペレーティングシステム１４００１が提供するパス管理機能（Ｉ／Ｏパスマネージャー５０００（図１）が提供する機能に相当する）によりひとつのボリューム１４００４として機能管理プログラム１４００３に提供される。

【０１６８】

論理ボリューム１５００１Ａ、論理ボリューム１５００１Ｂはそれぞれストレージ装置１５０００Ａ、ストレージ装置１５０００Ｂが提供する論理ボリュームであり、機能制御対象の論理ボリュームである。なお、本実施の形態では論理ボリューム１５００１Ａがアクティブ側と仮定している。

【０１６９】

また、同期リモートコピーにより、論理ボリューム１５００１Ａの内容は、論理ボリューム１５００１Ｂの内容と常に一致している。論理ボリューム１５００１Ａ、論理ボリューム１５００１Ｂはオペレーティングシステム１４００１が提供するパス管理機能によりひとつのボリューム１４００５としてアプリケーションプログラム１４００２に提供される。

【０１７０】

なお、ここで説明した機能制御対象の論理ボリュームは複数あってもよい。

【０１７１】
機能管理プログラム１４００３の機能制御要求処理部１４００５は、ユーザーあるいはホスト１４０００内の他のプログラムあるいはホスト１４０００とは別のホスト（管理ホストなど）内のプログラムから、機能制御要求を受け付ける。機能制御要求を受け付けた機能制御要求処理部１４００５はボリューム１４００４に対する制御要求の内容をボリューム１４００４に対してライト／リードする。本実施の形態ではコマンドデバイス１５００２Ａがアクティブ側であるため、ライト／リードはコマンドデバイス１５００２Ａに対して発行される。
【０１７２】
コマンドデバイス１５００２Ａに対するライトは機能制御を起動するときに用いられ、コマンドデバイス１５００２Ａに対するリードは機能制御の結果の出力値を得るために用いられる。
【０１７３】
機能制御要求処理部１４００５が受け付ける制御要求には制御対象のストレージ装置１５０００Ａ，１５０００Ｂを一意に識別する情報（装置情報とも呼ぶ）と、制御対象の論理ボリューム１５００１Ａ，１５００１Ｂを一意に識別する情報（ボリューム情報とも呼ぶ）と、機能制御に付随する情報とが含まれる。
【０１７４】
ストレージ装置１５０００Ａの制御I／Ｆ処理部１５００３Ａはコマンドデバイス１５００２Ａに制御要求がライトされたことを検出する。制御Ｉ／Ｆ 処理部１５００３Ａは制御要求の装置情報が自ストレージ装置（ストレージ装置１５０００Ａ）に一致するか判定する（判定１００）。本実施の形態ではコマンドデバイス１５００２Ａがアクティブ側なので、判定の結果は「一致する」となる。一致した場合、制御Ｉ／Ｆ 処理部１５００３Ａはボリューム情報に対応する論理ボリューム１５００１Ａに対して所定の機能制御を実行するよう機能処理部１５００４Ａを呼び出す。具体的な例としては、ストレージ装置１５０００Ａが提供する機能のひとつであるローカルコピー機能（後で説明）のペア状態の参照操作がある。当該操作が論理ボリューム１５００１Ａに対して呼び出された場合、機能処理部１５００４Ａは、ローカルコピー機能の管理情報を参照し、ペア状態を取得した後、制御Ｉ／Ｆ処理部１５００３Ａ、コマンドデバイス１５００２Ａ及びボリューム１４００４を介して、機能制御要求処理部１４００５に対して、ペア状態を送信する。
【０１７５】
一方、ストレージ装置１５０００Ｂの制御Ｉ／Ｆ 処理部１５００３Ｂも同様の処理を行うが、本実施の形態では、コマンドデバイス１５００２Ｂはスタンバイ側なので、判定１００の結果は「一致しない」となる。この場合、制御Ｉ／Ｆ 処理部１５００３Ｂは同期リモートコピーのペアの管理情報を参照し、ボリューム情報（論理ボリューム１５００１Ａに対応）に対応する自ストレージ装置（ストレージ装置１５０００Ｂ）内の論理ボリューム（論理ボリューム１５００１Ｂに対応）を特定する。そして、制御Ｉ／Ｆ処理部１５００３Ｂは論理ボリューム１５００１Ｂに対して所定の機能制御を実行するよう機能処理部１５００４Ｂを呼び出す。
【０１７６】
以上により、ストレージ装置１５０００Ａの論理ボリューム１５００１Ａ、ストレージ装置１５０００Ｂの論理ボリューム１５００１Ｂに対して、所定の機能の制御が実行される。
【０１７７】
本実施の形態では、ストレージ装置１５０００Ａ，１５０００Ｂが提供するローカルコピー機能のペア状態の参照操作を例にとって説明したが、（１）ローカルコピー機能のその他のペア操作（ペアの作成、ペアの分割等）、（２）ストレージ装置１５０００Ａ，１５０００Ｂが提供するローカルコピー機能の各種ペア操作、（３）ストレージ装置１５０００Ａ，１５０００Ｂが提供する論理ボリューム１５００１Ａ，１５００１Ｂに対するセキュリティ機能（後で説明するＬＤＥＶガード機能）の操作、（４）ストレージ装置１５

０００Ａ，１５０００Ｂが提供する論理スナップショット機能（後で説明）の操作、等、ストレージ装置１５０００Ａ，１５０００Ｂが提供する各種機能の操作について適用できる。

【０１７８】

なお、別な実行形態としては、アクティブ側とスタンバイ側両方のストレージ装置１５０００Ａ，１５０００Ｂに発行すべきコマンドを受けた場合は、アクティブ側のストレージ装置１５０００Ａは受取ったコマンドを処理すると共に、スタンバイ側のストレージ装置１５０００Ｂへ転送してコマンド処理をしてもらうことで、ホスト１４０００からは１回のコマンドで両方のストレージ処理を開始することも考えられる。また、プログラムの状態取得に関するコマンドの場合は、コマンドを受取ったアクティブ側のストレージ装置１５０００Ａがスタンバイ側のストレージ装置１５０００Ｂに同じコマンドを転送して状態を取得し、アクティブ側のストレージ装置１５０００Ａが両方の状態を比較した後にコマンド発信元へ状態を返すことも考えられる。

【０１７９】

（５）第５の実施の形態

本実施の形態では機能Ｉ／Ｆの別の構成について述べる。図１５を用いて本実施の形態の構成を説明する。

【０１８０】

本実施の形態の構成は図１４とほぼ同様である。図１４との違いは、

（１）コマンドデバイス１５００２Ａ、コマンドデバイス１５００２Ｂが同期リモートコピーのペアでない。

（２）機能管理プログラム１４００３からはコマンドデバイス１５００２Ａ及びコマンドデバイス１５００２Ｂが別々のボリューム１４００４Ａ、１４００４Ｂとして認識されている。

（３）機能制御要求処理部１４００５は機能制御要求をコマンドデバイス１５００２Ａ及びコマンドデバイス１５００２Ｂに送信する。

という３点である。

【０１８１】

本実施の形態では、第４の実施の形態と同様に、機能制御要求処理部１４００５が受け付ける制御要求には制御対象のストレージ装置１５０００Ａ，１５０００Ｂを一意に識別する情報（装置情報とも呼ぶ）と、制御対象の論理ボリューム１５００１Ａ，１５００１Ｂを一意に識別する情報（ボリューム情報とも呼ぶ）と、機能制御に付随する情報とが含まれる。

【０１８２】

本実施の形態では、第４の実施の形態と異なり、前述のように、ユーザーあるいはホスト１４０００内の他のプログラムあるいはホスト１４０００とは別のホスト内のプログラムから機能制御要求を受け付けた機能制御要求処理部１４００５は、両方のコマンドデバイス１５００２Ａ、１５００２Ｂに制御要求を送信する。

【０１８３】

なお、機能制御要求処理部１４００５が装置情報を判定し、コマンドデバイス１５００２Ａに対しては、ボリューム情報として論理ボリューム１５００１Ａを指定し、コマンドデバイス１５００２Ｂに対しては、ボリューム情報として論理ボリューム１５００１Ｂを指定するように制御要求を書き換えてもよい。

【０１８４】

さらにまた、ユーザーあるいはホスト１４０００内の他のプログラムあるいはホスト１４０００とは別のホスト内のプログラムがストレージ装置１５０００Ａ，１５０００Ｂを識別し、ストレージ装置１５０００Ａ、１５０００Ｂに対して二重に異なる制御要求を出してもよい。即ち、コマンドデバイス１５００２Ａに対して、論理ボリューム１５００１Ａの制御要求を出し、コマンドデバイス１５００２Ｂに対して、論理ボリューム１５００１Ｂの制御要求を出す。

【０１８５】
（６）第６の実施の形態
本実施の形態では機能Ｉ／Ｆの更に別の構成について述べる。図１６を用いて本実施の形態の構成を説明する。
【０１８６】
第６の実施の形態は第４の実施の形態とほぼ同様である。第４の実施の形態との違いは以下の点である。
（１）ホスト１４０００、ストレージ装置１５０００Ａ、ストレージ装置１５０００Ｂは互いにＬＡＮ（Local Area Network）のような相互結合網により接続されている。なお、これらはＬＡＮにより直結されていてもよいし、スイッチを経由して接続されていてもよい。
【０１８７】
（２）コマンドデバイスがない構成であり、３者（ホスト１４０００、ストレージ装置１５０００Ａ、ストレージ装置１５０００Ｂ）間の通信はＬＡＮを介して行なわれる。
（３）機能制御要求処理部１４００５はＬＡＮを介して、制御要求を制御Ｉ／Ｆ 処理部１５００３Ａに送信する。
（４）制御要求を受け取った制御Ｉ／Ｆ 処理部１５００３ＡはＬＡＮを介して、制御要求を制御Ｉ／Ｆ 処理部１５００３Ｂに送信する。
【０１８８】
制御Ｉ／Ｆ 処理部１５００３Ａ、１５００３Ｂが受け取った制御要求を処理する点は第４の実施の形態と同様であり、第６の実施の形態は第４の実施の形態と同等の機能Ｉ／Ｆを提供することができる。
【０１８９】
（７）第７の実施の形態
本実施の形態では機能Ｉ／Ｆの更に別の構成について述べる。図１７を用いて本実施の形態の構成を説明する。
【０１９０】
第７の実施の形態は第６の実施の形態とほぼ同様である。第６の実施の形態との違いは以下の点である。
（１）機能制御要求処理部１４００５はＬＡＮを介して、制御要求を両方の制御Ｉ／Ｆ処理部１５００３Ａ、１５００３Ｂに送信する。
（２）制御Ｉ／Ｆ 処理部１５００３Ａは制御Ｉ／Ｆ 処理部１５００３Ｂに対して、制御要求を送信しない。
【０１９１】
制御Ｉ／Ｆ 処理部１５００３Ａ、１５００３Ｂが受け取った制御要求を処理する点は第６の実施の形態と同様であり、第７の実施の形態は第６の実施の形態と同等の機能Ｉ／Ｆを提供することができる。
【０１９２】
（８）第８の実施の形態
本実施の形態では、ストレージ装置内の論理ボリュームに対してセキュリティ機能（ＬＤＥＶセキュリティ機能）を適用する場合の例を説明する。
【０１９３】
図１８はＬＤＥＶセキュリティ機能の一実施の形態を示したものである。本実施の形態の構成は第４の実施の形態の図１４とほぼ同一である。図１４と異なる点は論理ボリュームセキュリティ情報１５００５Ａ、１５００５Ｂが追加された点である。論理ボリュームセキュリティ情報１５００５Ａ、１５００５Ｂは、ホスト１４０００からストレージ装置１５０００Ａ，１５０００Ｂ内の論理ボリューム１５００１Ａ，１５００１Ｂに対するアクセス制御を行うために用いられる。アクセス制御の例としては、論理ボリューム１５００１Ａ，１５００１Ｂ内のデータの改ざんを抑止するために論理ボリューム１５００１Ａ，１５００１Ｂに対するライトアクセスを一切禁止する制御がある。また、別の例として

は、法令等により一定期間の保存を義務付けられたデータに対して、所定の期間ライトを禁止する機能がある。さらに、別の例としては、機密情報の保護の観点等から特定のホストからのリード／ライトアクセスを禁止する機能がある。

【０１９４】

図１８のように２台のストレージ構成１５０００Ａ，１５０００Ｂを用いて同期リモートコピーにより冗長化を図った構成においてもＬＤＥＶセキュリティ機能を適用したい場合が考えられる。この場合においても第４の実施の形態で説明した機能Ｉ／Ｆを用いてＬＤＥＶセキュリティ機能を制御することができる。具体的には、機能処理部１５００４において、対象ボリュームに対するセキュリティ情報を格納する論理ボリュームセキュリティ情報１５００５Ａ、１５００５Ｂに、ＬＤＥＶセキュリティに関するパラメータを設定したり、参照したりすればよい。

【０１９５】

（９）第９の実施の形態

本実施の形態では、ストレージ装置内の論理ボリュームにローカルコピー機能を適用した場合の例を説明する。

【０１９６】

ローカルコピー機能とは、ユーザーから指定されたボリュームの複製を、コピー元ボリュームと同じストレージ装置内において作成する機能である。本機能を用いて作成されたボリュームの複製はデータマイニングやテープバックアップのためにホストがアクセスを行ったり、あるいはバックアップデータとして長時間保存される。ローカルコピー機能は複製を作成したいボリュームと複製先ボリュームをペア（コピーペア）として指定し、そのペアに対してユーザーが操作を行うことでユーザーは複製を作成する。以後の説明では複製対象ボリュームを正ボリュームと呼び、複製先ボリュームを副ボリュームと呼ぶことがある。本実施の形態ではこのローカルコピー機能についてもアクティブ側のストレージとスタンバイ側のストレージで連携することで可用性を向上させる。

【０１９７】

図１９はローカルコピー機能の一実施の形態を示したものである。図１９においては、ホスト１４０００はストレージ装置１５０００Ａとストレージ装置１５０００Ｂに接続されている。また、ストレージ装置１５０００Ａは外部ストレージ装置１６０００Ａと接続され、ストレージ装置１５０００Ｂは外部ストレージ装置１６０００Ｂと接続されている。また、ローカルコピー機能及び差分ビットマップ（正ボリューム１５００６Ａ，１５００６Ｂと副ボリューム１５００７Ａ，１５００７Ｂの間の差分の有無を示す情報）がストレージ装置１５０００Ａとストレージ装置１５０００Ｂにて実行及び管理される。

【０１９８】

本実施の形態は正ボリューム１５００６Ａ，１５００６Ｂがストレージ装置１５０００Ａ，１５０００Ｂ内にあり、副ボリューム１５００７Ａ，１５００７Ｂが外部ストレージ装置１６０００Ａ，１６０００Ｂ内にある構成例を示している。正ボリューム１５００６Ａと副ボリューム１５００７Ａはペアであり、副ボリューム１５００７Ａの実体は外部ボリューム１６００１Ａ内にある。同様に、正ボリューム１５００６Ｂと副ボリューム１５００７Ｂはペアであり、副ボリューム１５００７Ｂの実体は外部ボリューム１６００１Ｂ内にある。

【０１９９】

＜Ｄｕｐｌｅｘ状態における動作＞

Ｄｕｐｌｅｘ状態とはペア状態のひとつで正ボリューム１５００６Ａ，１５００６Ｂから副ボリューム１５００７Ａ，１５００７Ｂへ後述するバックグラウンドコピーが行われている状態である。

【０２００】

以下ではＤｕｐｌｅｘ状態におけるリード／ライト処理について述べる。なお、以下のリード／ライト処理の説明は、リード／ライト処理の対象ボリューム（正ボリューム１５００６Ａ，１５００６Ｂ）のアクティブ側がストレージ装置１５０００Ａであるという前

提である。

【０２０１】

まずリード処理について説明する。アプリケーションプログラム１４００２からリード要求を受け付けたオペレーティングシステム１４００１はパス管理機能により、（リード対象の正ボリュームに関して）アクティブ側のストレージがストレージ装置１５０００Ａとストレージ装置１５０００Ｂのどちらかを判断し、アクティブ側のストレージ装置１５０００Ａにリード要求を発行する。リード要求を受信したストレージ装置１５０００Ａはリード対象データをホスト１４０００に送信する。アプリケーション１４００２はオペレーティングシステム１４００１を介してリード対象データを受信する。以上によりリード処理は完了する。

【０２０２】

次にライト処理について説明する。アプリケーションプログラム１４００２からライトリクエストを受け付けたオペレーティングシステム１４００１はパス管理機能により、（リード対象の正ボリュームに関して）アクティブ側のストレージ装置がストレージ装置１５０００Ａとストレージ装置１５０００Ｂのどちらかを判断し、アクティブ側のストレージ装置１５０００Ａにライトリクエストを発行する。ライトリクエストを受信したストレージ装置１５０００Ａは、ライトデータを受信し、図示しないキャッシュメモリにライトデータを格納すると共にライトデータに対応する差分ビットマップのビットを１（オン）に設定する。

【０２０３】

その後、当該ライトデータはリモートコピー機能により、ストレージ装置１５０００Ａ内のキャッシュメモリからストレージ装置１５０００Ｂ内の正ボリューム１５００６Ｂにコピー（同期リモートコピー）する。なお、同期リモートコピーの方法はこれまで説明した通りである。同期リモートコピーによりストレージ装置１５０００Ａからライトデータを受信したストレージ装置１５０００Ｂは、図示しないキャッシュメモリにライトデータを格納すると共にライトデータに対応する差分ビットマップのビットを１（オン）に設定する。その後、ストレージ装置１５０００Ｂはストレージ装置１５０００Ａに対してライト完了報告に送信し、ライト完了報告を受信したストレージ装置１５０００Ａはホスト１４０００に対してライト完了報告を送信する。

【０２０４】

なお、ストレージ装置１５０００Ａの正ボリューム１５００６Ａ、ストレージ装置１５０００Ｂの正ボリューム１５００６Ｂにライトされたデータは、正ボリューム１５００６Ａ，１５００６Ｂへのライトとは非同期に副ボリューム１５００７Ａ，１５００７Ｂへコピーされる（以後、本処理をバックグラウンドコピー処理と呼ぶ）。バックグラウンドコピーは、差分ビットマップ定期的に監視し、差分あり（すなわちビットがオン）と記録された領域のデータを正ボリューム１５００６Ａ，１５００６Ｂから副ボリューム１５００７Ａ，１５００７Ｂへコピーし、コピーが終了したらビットをクリア（オフ又は０に）することにより行なわれる。

【０２０５】

一方、スタンバイ側のストレージ装置１５０００Ｂも同期リモートコピーによってライトデータが到着した時点を契機として同様の処理を行う。

【０２０６】

なお、本発明は上記例以外の構成、たとえば正ボリューム１５００６Ａ，１５００６Ｂは外部ストレージ装置１６０００Ａ内にあってもよいし、ストレージ装置１５０００Ａ，１５０００Ｂ内にあってもよい。副ボリューム１５００７Ａ，１５００７Ｂもまた、外部ストレージ装置１６０００Ａ内にあってもよいし、ストレージ装置１５０００Ａ，１５０００Ｂ内にあってもよい。

【０２０７】

何らかの障害が発生し、アクティブ側の正ボリューム１５００６Ａに対するＩ／Ｏリクエストが処理できなくなった場合には、すでに説明した通り、オペレーティングシステム

１４００１は、Ｉ／Ｏリクエストの対象を正ボリューム１５００６Ｂに切り替えてアクセスを継続する。この場合でも、ストレージ装置１５０００Ｂ内にはローカルコピー機能のペアが存在するため、副ボリューム１５００７Ｂを用いて先に述べたバックアップ等の処理を行なうことができる。

【０２０８】

＜ペアＳｐｌｉｔとＳｐｌｉｔ状態の動作＞

Ｓｐｌｉｔ状態とはペア状態のひとつで、副ボリュームのイメージが確定した状態のことを指す。この状態では、正ボリュームと副ボリュームの内容が一致しておらず、正ボリュームと副ボリュームの間の差分が差分ビットマップで管理されている。また、この状態においては、副ボリュームが静止した状態になるため、ユーザーは先に述べたバックアップ等の処理を行なうことができる。

【０２０９】

ホスト１４０００はローカルコピーのＤｕｐｌｅｘ状態のペアをＳｐｌｉｔ状態にする場合、これまで説明してきたバックグラウンドコピーの動作を停止させる（これをペアＳｐｌｉｔと呼ぶ）。ペアＳｐｌｉｔは第４～第７の実施の形態で説明した機能Ｉ／Ｆを介して実施する。

【０２１０】

（１）ホスト１４０００は機能Ｉ／Ｆを介してストレージ装置１５０００Ａ，１５０００Ｂにローカルコピーの停止命令を出す。通常、ホスト側ではこの停止命令直前にＩ／Ｏリクエストの発行を停止する。

（２）アクティブ側及びスタンバイ側のストレージ装置１５０００Ａ，１５０００Ｂはそれぞれ差分ビットマップ上でオンとなった領域のバックグラウンドコピーを完了させる。ホスト１４０００は両ストレージ装置１５０００Ａ，１５０００Ｂにおけるバックグラウンドコピーが完了ことを認識するメッセージをアクティブ側のストレージ装置１５０００Ａ、もしくは両ストレージ装置１５０００Ａ，１５０００Ｂから受領する。

（３）ホスト１４０００は当該メッセージを受領した後、Ｉ／Ｏ発行を再開する。

【０２１１】

（２）までの処理により、アクティブ側及びスタンバイ側のストレージ装置１５０００Ａ，１５０００Ｂ内のペアはＳｐｌｉｔ状態になったことが確定する。この時点で両ストレージ装置１５０００Ａ，１５０００Ｂ内のペア状態はＳｐｌｉｔ状態となっている。なお、Ｓｐｌｉｔ中に行われた正ボリュームまたは副ボリュームへ行われたライトリクエストのライト位置は、後ほど説明するペア再同期のために差分ビットマップに記録される。

【０２１２】

その後のＩ／Ｏリクエストの処理はＤｕｐｌｅｘ状態とほぼ同様である。Ｄｕｐｌｅｘ状態との違いは、バックグラウンドコピー処理が動作しない点である。

【０２１３】

＜ペア作成＞

正ボリュームと副ボリュームがペア関係にない状態をＳｉｍｐｌｅｘ状態と呼ぶ。Ｓｉｍｐｌｅｘ状態からＤｕｐｌｅｘ状態に遷移させるための処理をペア作成と呼ぶ。ペア状態がＳｉｍｐｌｅｘ状態からＤｕｐｌｅｘ状態に遷移している過渡状態をＩｎｉｔｉａｌ－Ｃｏｐｙｉｎｇ状態と呼ぶ。

【０２１４】

ペア作成の指示は、第４～第７の実施の形態で説明した機能Ｉ／Ｆを介して実施する。

（１）ホスト１４０００は機能Ｉ／Ｆを介して、ストレージ装置１５０００Ａに対してペア作成指示を出す。この結果アクティブ側及びスタンバイ側の両ストレージ装置１５０００Ａ，１５０００Ｂでペア作成処理が開始される。

（２）両ストレージ装置１５０００Ａ，１５０００Ｂはペア状態をＩｎｉｔｉａｌ－Ｃｏｐｙｉｎｇ状態に設定し、差分ビットマップ上のビットを全てオンにし、バックグラウンドコピーを開始する。

（３）バックグラウンドコピーが差分ビットマップの最後まで完了したら、ストレージ装

置１５０００Ａ，１５０００Ｂはペア状態をＤｕｐｌｅｘ状態に設定する。
【０２１５】
Ｉｎｉｔｉａｌ‐Ｃｏｐｙｉｎｇ状態におけるリード／ライト処理はＤｕｐｌｅｘ状態におけるリード／ライト処理と同様である。
【０２１６】
＜ペア再同期＞
ペア状態をＳｕｓｕｐｅｎｄ状態からＤｕｐｌｅｘ状態に遷移させる操作をペア再同期と呼ぶ。ペア状態がＳｕｓｕｐｅｎｄ状態からＤｕｐｌｅｘ状態に遷移している過渡状態をＤｕｐｌｅｘ‐Ｐｅｎｄｉｎｇ状態と呼ぶ。
【０２１７】
ペア再同期の指示は、第４～第７の実施の形態で説明した機能Ｉ／Ｆを介して実施する。
（１）ホスト１４０００は機能Ｉ／Ｆを介して、ストレージ装置１５０００Ａに対してペア再同期指示を出す。この結果アクティブ側及びスタンバイ側の両ストレージ装置１５０００Ａ，１５０００Ｂでペア再同期処理が開始される。
（２）両ストレージ装置１５０００Ａ，１５０００Ｂはペア状態をＤｕｐｌｅｘ‐Ｐｅｎｄｉｎｇに設定し、バックグラウンドコピーを開始する。
（３）バックグラウンドコピーが差分ビットマップの最後まで完了したら、ストレージ装置１５０００Ａ，１５０００Ｂはペア状態をＤｕｐｌｅｘ状態に設定する。
【０２１８】
Ｄｕｐｌｅｘ‐Ｐｅｎｄｉｎｇ状態におけるリード／ライト処理はＤｕｐｌｅｘ状態におけるリード／ライト処理と同様である。
【０２１９】
（１０）第１０の実施の形態
本実施の形態では第９の実施の形態とは異なるローカルコピー機能の実施の形態を説明する。本実施の形態の一構成例を図２０に示す。
【０２２０】
まず、本実施の形態と第９の実施の形態との構成の違いは、外部ストレージ装置１６０００Ｂが存在せず、副ボリューム１５００７Ａ，１５００７Ｂの実体がいずれも外部ストレージ装置１６０００Ａ内の外部ボリューム１６００１Ａとなるようにマッピングされている点である。その他の構成は第９の実施の形態と同様である。
【０２２１】
このように構成することにより、副ボリューム１５００７Ａ，１５００７Ｂに必要とされる物理的な記憶装置を削減することができる。
【０２２２】
本実施の形態と第９の実施の形態の処理動作との大きな違いはスタンバイ側のストレージ装置１５０００Ｂが外部ボリューム１６００１Ａに対するバックグラウンドコピーを行なわず、ストレージ装置１５０００Ａとの通信により、ペアに関する制御情報であるペア状態と差分ビットマップ１５０１０Ｂのみを操作する点である。
【０２２３】
以下では処理動作を詳細に説明する。
【０２２４】
＜Ｄｕｐｌｅｘ状態における動作＞
以下ではＤｕｐｌｅｘ状態におけるリード／ライト処理について述べる。
【０２２５】
まず、リード処理は第９の実施の形態でのリード処理と同様である。
【０２２６】
次にライト処理について説明する。アプリケーションプログラム１４００２からライトリクエストを受け付けたオペレーティングシステム１４００１はパス管理機能により、（リード対象の正ボリューム１５００６Ａに関して）アクティブ側のストレージ装置がスト

レージ装置１５０００Ａ及びストレージ装置１５０００Ｂのどちらかを判断し、アクティブ側のストレージ装置１５０００Ａにライトリクエストを発行する。ライトリクエストを受信したストレージ装置１５０００Ａは、ライトデータを受信し、図示しないキャッシュメモリにライトデータを格納すると共にライトデータに対応する差分ビットマップ１５０１０Ａのビットを１（オン）に設定する。

【０２２７】

その後、当該ライトデータは同期リモートコピー機能により、ストレージ装置１５０００Ａ内の正ボリューム１５００６Ａからストレージ装置１５０００Ｂ内の正ボリューム１５００６Ｂにコピーされる。なお、同期リモートコピーの方法はこれまで説明した通りである。同期リモートコピー機能によりストレージ装置１５０００Ａからライトデータを受信したストレージ装置１５０００Ｂは、図示しないキャッシュメモリにライトデータを格納すると共にライトデータに対応する差分ビットマップ１５０１０Ｂのビットを１（オン）に設定する。その後、ストレージ装置１５０００Ｂはストレージ装置１５０００Ａに対してライト完了報告に送信し、ライト完了報告を受信したストレージ装置１５０００Ａはホスト１４０００に対してライト完了報告を送信する。

【０２２８】

なお、ストレージ装置１５０００Ａの正ボリューム１５００６Ａにライトされたデータは、正ボリューム１５００６Ａへのライトとは非同期に副ボリューム１５００７Ａへバックグラウンドコピーされる。第９の実施の形態でのライト処理と異なり、ストレージ装置１５０００Ｂの正ボリューム１５００６Ｂにライトされたデータはバックグラウンドコピーされない。

【０２２９】

ストレージ装置１５０００Ａにおけるバックグラウンドコピーは、差分ビットマップ１５０１０Ａを定期的に監視し、差分あり（すなわちビットがオン）と記録された領域のデータを正ボリューム１５００６Ａから副ボリューム１５００７Ａへコピーし、コピーが終了したらビットをクリア（オフ又は０に）することにより行なわれる。なお、本実施の形態では、第９の実施の形態でのライト処理と異なり、ストレージ装置１５０００Ｂにおいてはバックグラウンドコピーが行なわれない。

【０２３０】

その後、第９の実施の形態でのライト処理と異なり、ストレージ装置１５０００Ａはクリアした差分ビットマップ１５０１０Ａ上のビットの位置情報をストレージ装置１５０００Ｂに通知する。通知を受信したストレージ装置１５０００Ｂは当該ビットに対応するストレージ装置１５０００Ｂ内の差分ビットマップ１５０１０Ｂ上のビット（差分ビット）をクリアする。なお、差分ビットの位置情報の通知はストレージ装置１５０００Ｂ内のコマンドデバイスを介して行なわれる。また、本実施の形態における構成では、コマンドデバイスを介して通知を行なったが、ストレージ装置１５０００Ａ，１５０００Ｂ間がＬＡＮで接続された構成である場合は、ＬＡＮを介した通信により通知を行なってもよい。以後、ストレージ装置１５０００Ａとストレージ装置１５０００Ｂとの間における、差分ビットやペア状態等といった機能の制御情報に関する通信はコマンドデバイスやＬＡＮを介して行うものとする。

【０２３１】

何らかの障害が発生し、アクティブ側の正ボリューム１５００６Ａに対するＩ／Ｏリクエストが処理できなくなった場合、オペレーティングシステム１４００１は、第９の実施の形態と同様に、Ｉ／Ｏリクエストの対象を正ボリューム１５００６Ｂに切り替えてアクセスを継続する。

【０２３２】

＜ペアＳｐｌｉｔとＳｐｌｉｔ状態の動作＞

ホスト１４０００はローカルコピーのＤｕｐｌｅｘ状態のペアをＳｐｌｉｔ状態にする場合、第９の実施の形態と同様にペアＳｐｌｉｔを行なう。なお、ペアＳｐｌｉｔにおいては、バックグラウンドコピーの終了処理が行なわれるが、本実施の形態ではストレージ

装置１５０００Ｂにおいては、バックグラウンドコピーは動作していないため、実際には終了処理は行なわれない。

【０２３３】

その後のＩ／Ｏリクエストの処理はＤｕｐｌｅｘ状態とほぼ同様である。Ｄｕｐｌｅｘ状態との違いは、ストレージ装置１５０００Ｂにおいてバックグラウンドコピー処理が動作しない点である。

【０２３４】

＜ペア作成＞

ペア作成の指示は、第４～第７の実施の形態で説明した機能Ｉ／Ｆを介して実施されるのは第９の実施の形態と同様である。

【０２３５】

（１）ホスト１４０００は機能Ｉ／Ｆを介して、ストレージ装置１５０００Ａに対してペア作成指示を出す。この結果アクティブ側及びスタンバイ側の両ストレージ装置１５０００Ａ，１５０００Ｂでペア作成処理が開始される。

（２）両ストレージ装置１５０００Ａ，１５０００Ｂはペア状態をＩｎｉｔｉａｌ－Ｃｏｐｙｉｎｇ状態に設定する。ストレージ装置１５０００Ａは差分ビットマップ１５０１０Ａ上のビットを全てオンにし、バックグラウンドコピーを開始する。第９の実施の形態と異なり、ストレージ装置１５０００Ｂは差分ビットマップ１５０１０Ｂ上のビットを全てオンにするが、バックグラウンドコピーを行なわない。

【０２３６】

（３）ストレージ装置１５０００Ａはバックグラウンドコピーが完了した領域に対応する差分ビットをクリアする処理とそれに付随する動作（差分ビットの位置情報の通知と差分ビットのクリア）はＤｕｐｌｅｘ状態における動作と同様である。

（４）第９の実施の形態と異なり、ストレージ装置１５０００Ａは、バックグラウンドコピーが差分ビットマップ１５０１０Ａの最後まで完了したら、ペア状態をＤｕｐｌｅｘ状態に設定し、ペア状態がＤｕｐｌｅｘ状態に変わったことをストレージ装置１５０００Ｂに通知する。通知を受信したストレージ装置１５０００Ｂはペア状態をＤｕｐｌｅｘ状態に設定する。

【０２３７】

Ｉｎｉｔｉａｌ‐Ｃｏｐｙｉｎｇ状態におけるリード／ライト処理はＤｕｐｌｅｘ状態におけるリード／ライト処理と同様である。

【０２３８】

＜ペア再同期＞

ペア再同期の指示は、第４～第７の実施の形態で説明した機能Ｉ／Ｆを介して実施されるのは第９の実施の形態と同様である。

【０２３９】

（１）ホスト１４０００は機能Ｉ／Ｆを介して、ストレージ装置１５０００Ａに対してペア再同期指示を出す。この結果アクティブ側及びスタンバイ側の両ストレージ装置１５０００Ａ，１５０００Ｂでペア再同期処理が開始される。

（２）ストレージ装置１５０００Ａはペア状態をＤｕｐｌｅｘ‐Ｐｅｎｄｉｎｇに設定し、バックグラウンドコピーを開始する。第９の実施の形態と異なり、ストレージ装置１５０００Ｂにおいては、バックグラウンドコピーは行なわない。

【０２４０】

（３）ストレージ装置１５０００Ａは、バックグラウンドコピーが差分ビットマップ１５０１０Ａの最後まで完了したら、ペア状態をＤｕｐｌｅｘ状態に設定する。ただし、第９の実施の形態と異なり、この処理を行なうのはストレージ装置１５０００Ａのみである。その後、ストレージ装置１５０００Ａは、ペア状態がＤｕｐｌｅｘ状態に変わったことをストレージ装置１５０００Ｂに通知する。通知を受信したストレージ装置１５０００Ｂはペア状態をＤｕｐｌｅｘ状態に設定する。

【０２４１】

Ｄｕｐｌｅｘ‐Ｐｅｎｄｉｎｇ状態におけるリード／ライト処理はＤｕｐｌｅｘ状態におけるリード／ライト処理と同様である。

【０２４２】

（１１）第１１の実施の形態

ＡＯＵ（Allocation On Use）機能の構成について述べる。ＡＯＵ機能はホストから使用された（ライトされた）領域に関してのみ実記憶領域を割り当てる機能である。

【０２４３】

ＡＯＵ機能はデータが実際に格納される実ボリュームの集合体であるプールと、ホストに見せるボリュームである仮想ボリュームから構成される。本実施の形態における仮想ボリュームはライトが行われた部分のみ実データが割り当てられるという意味で仮想的である。ホストに見せているボリュームの全アドレス空間に実データが割り当てられている訳ではない。なお、実ボリュームは外部ストレージ装置内にあってもよいし、仮想ボリュームと同じストレージ装置内にあってもよい。

【０２４４】

図２１はＡＯＵ機能の一実施の形態を示したものである。図２１においては、ホスト１４０００はストレージ装置１５０００Ａとストレージ装置１５０００Ｂに接続されている。また、ストレージ装置１５０００Ａは外部ストレージ装置１６０００Ａと接続され、ストレージ装置１５０００Ｂは外部ストレージ装置１６０００Ｂと接続されている。

【０２４５】

本実施の形態は実ボリューム１６００２Ａが外部ストレージ装置１６０００Ａ，１６０００Ｂ内にある構成例を示している。仮想ボリューム１５００８Ａ内のデータはプール１６００３Ａの実ボリューム１６００２Ａ内のデータと対応付けられる。同様に仮想ボリューム１５００８Ｂ内のデータはプール１６００３Ｂの実ボリューム１６００２Ｂ内のデータと対応付けられる。また、仮想ボリューム１５００８Ａと仮想ボリューム１５００８Ｂは同期リモートコピー機能により内容が一致するように構成される。同期リモートコピーの方法はこれまで説明したとおりである。

【０２４６】

次に本構成におけるリード／ライト処理について述べる。なお、以下のリード／ライト処理の説明は、リード／ライト処理の対象ボリュームのアクティブ側がストレージ装置１５０００Ａであるという前提である。

【０２４７】

まずリード処理について説明する。アプリケーションプログラム１４００２からリードリクエストを受け付けたオペレーティングシステム１４００１はパス管理機能によりアクティブ側のストレージがストレージ装置１５０００Ａ及びストレージ装置１５０００Ｂのどちらかを判断し、アクティブ側のストレージ装置１５０００Ａにリードリクエストを発行する。リードリクエストを受け付けたストレージ装置１５０００Ａは、仮想アドレス実アドレス変換テーブル１５００９Ａを参照し、リードデータにプール１６００３Ａ内の実領域が割り当てられているか判定する。

【０２４８】

前述の判定で実領域が割り当てられている場合、ストレージ装置１５０００Ａは、当該実領域からリードデータを読み出してホスト１４０００に送信する。アプリケーション１４００２はオペレーティングシステム１４００１を介してリードデータを受信する。以上によりリード処理は完了する。

【０２４９】

次にライト処理について説明する。アプリケーションプログラム１４００２からライトリクエストを受け付けたオペレーティングシステム１４００１はパス管理機能によりアクティブ側のストレージ装置がストレージ装置１５０００Ａとストレージ装置１５０００Ｂのどちらかを判断し、アクティブ側のストレージ装置１５０００Ａにライトリクエストを発行する。ライトリクエストを受け取ったストレージ装置１５０００Ａは仮想アドレス実アドレス変換テーブル１５００９Ａを参照し、ライト対象データにプール１６００３Ａ内

の実領域が割り当てられているか判定する（判定２００）。

【０２５０】

前述の判定で実領域が割り当てられている場合、ストレージ装置１５０００Ａは、ホスト１４０００からライトデータを受信し、当該実領域に対応する図示しないキャッシュメモリ内の領域にライトデータを格納する。そして、同期リモートコピー機能によりライトデータをストレージ装置１５０００Ｂにライトリクエストを送信する。ストレージ装置１５０００Ａからライトリクエストを受信したストレージ装置１５０００Ｂは、ライトデータにプール１６００３Ａ内の実領域が割り当てられているか判定する。ここで、仮想ボリューム１５００８Ａの内容と仮想ボリューム１５００８Ｂの内容は同期リモートコピー機能により一致しているため、実領域は割り当てられていると判定される。その後、ストレージ装置１５０００Ｂは、ストレージ装置１５０００Ａからライトデータを受信し、当該実領域に対応する図示しないキャッシュメモリ内の領域にライトデータを格納し、ストレージ装置１５０００Ａにライト完了報告を行う。

【０２５１】

前述の判定（判定２００）で実領域が割り当てられていない場合、ストレージ装置１５０００Ａは、仮想アドレス実アドレス変換テーブル１５００９Ａにライトデータのアドレスを登録し、実領域を確保する。その後、ストレージ装置１５０００Ａは、ホスト１４０００からライトデータを受信し、当該実領域に対応する図示しないキャッシュメモリ内の領域にライトデータを格納する。そして、同期リモートコピー機能によりライトデータをストレージ装置１５０００Ｂにライトリクエストを送信する。

【０２５２】

ストレージ装置１５０００Ａからライトリクエストを受信したストレージ装置１５０００Ｂは、ライトデータにプール１６００３Ｂ内の実領域が割り当てられているか判定する。ここで、仮想ボリューム１５００８Ａの内容と仮想ボリューム１５００８Ｂの内容は同期リモートコピー機能により一致しているため、実領域は割り当てられていないと判定される。その後、ストレージ装置１５０００Ｂは、仮想アドレス実アドレス変換テーブル１５００９Ｂにライトデータのアドレスを登録し、実領域を確保する。そして、ストレージ装置１５０００Ｂは、ストレージ装置１５０００Ｂからライトデータを受信し、当該実領域に対応する図示しないキャッシュメモリ内の領域にライトデータを格納した後、ストレージ装置１５０００Ａにライト完了報告を行う。ライト完了報告を受信したストレージ装置１５０００Ａはホスト１４０００にライト完了報告を行う。ホスト１４０００がライト完了報告を受信し、ライト処理は完了する。

【０２５３】

なお、キャッシュメモリに格納されたデータはキャッシュメモリへの格納とは非同期に実ボリューム１６００２Ａ，１６００２Ｂへライトされる。

【０２５４】

何らかの障害により、アプリケーション１４００２がストレージ装置１５０００Ａ内の仮想ボリューム１５００８Ａ経由でのリード／ライト処理が不可能になった場合、オペレーティングシステム１４００１の提供するパス管理機能は障害を検出し、リード／ライト処理のアクセス経路をストレージ装置１５０００Ｂ内の仮想ボリューム１５００８Ｂ経由に切り替える。仮想ボリューム１５００８Ａの内容と仮想ボリューム１５００８Ｂの内容は同期リモート機能により一致しているため、アクセス経路が切り替わっても、継続して正常にリード／ライト処理を行うことができる。

【０２５５】

（１２）第１２の実施の形態

本実施の形態ではＡＯＵ機能の第１１の実施の形態とは異なる実施の形態について述べる。本実施の形態の一構成例を図２２に示す。

【０２５６】

まず、本実施の形態と第１１の実施の形態との構成の違いは、外部ストレージ装置１６０００Ｂが存在せず、仮想ボリューム１５００８Ａ、１５００８Ｂの実領域がいずれも外

部ストレージ装置１６０００Ａ内のプール１６０００３Ａ内の領域に割り当てられている点である。その他の構成は第１１の実施の形態と同様である。

【０２５７】

なお、本実施の形態はストレージ装置１５０００Ａ及びストレージ装置１５０００Ｂが共通のプールとして、共通の外部ストレージ装置１６０００Ａ内の実ボリューム１６０００２Ａを用いるため、第１１の実施の形態と異なり、実ボリューム１６０００２Ａが外部ストレージ装置１６０００Ａ内にある構成に限定される。

【０２５８】

このように構成することにより、プールに必要とされる物理的な記憶装置（ＨＤＤなど）の容量を削減することができる。

【０２５９】

本実施の形態と第１１の実施の形態の処理動作の大きな違いは、スタンバイ側のストレージ装置１５０００Ｂがキャッシュメモリから外部ストレージ装置１６０００Ａ内の実ボリューム１６０００２Ａに対してライトを行なわない点と、ストレージ装置１５０００Ａが仮想アドレス実アドレス変換テーブル１５００９Ａへの更新をストレージ装置１５０００Ｂに通知し、通知を受けたストレージ装置１５０００Ｂが仮想アドレス実アドレス変換テーブル１５００９Ｂを更新する点である。

【０２６０】

以下では処理動作を詳細に説明する。

【０２６１】

まずリード処理は第１１の実施の形態におけるリード処理と同様である。

【０２６２】

次にライト処理について説明する。アプリケーションプログラム１４００２からライトリクエストを受け付けたオペレーティングシステム１４００１はパス管理機能によりアクティブ側のストレージがストレージ装置１５０００Ａ及びストレージ装置１５０００Ｂのどちらかを判断し、アクティブ側のストレージ装置１５０００Ａにライトリクエストを発行する。ライトリクエストを受け取ったストレージ装置１５０００Ａは仮想アドレス実アドレス変換テーブル１５００９Ａを参照し、ライトデータにプール１６０００３Ａ内の実領域が割り当てられているか判定する（判定３００）。

【０２６３】

前述の判定で実領域が割り当てられている場合、ストレージ装置１５０００Ａは、ホスト１４０００からライトデータを受信し、当該実領域に対応するキャッシュメモリ内の領域にライトデータを格納する。そして、同期リモートコピー機能によりライトデータをストレージ装置１５０００Ｂにライトリクエストを送信する。次に、本実施の形態では、第１１の実施の形態と異なり、ストレージ装置１５０００Ａからライトリクエストを受信したストレージ装置１５０００Ｂは、即座にストレージ装置１５０００Ａからライトデータを受信し、キャッシュメモリに当該データを格納した後、ストレージ装置１５０００Ａにライト完了報告を行なう。ストレージ装置１５０００Ｂからライト完了報告を受信したストレージ装置１５０００Ａはホスト１４０００に対してライト完了報告を送信する。

【０２６４】

前述の判定（判定３００）で実領域が割り当てられていない場合、ストレージ装置１５０００Ａは、仮想アドレス実アドレス変換テーブル１５００９Ａにライトデータのアドレスを登録し、実領域を確保する。その後、ストレージ装置１５０００Ａは、ホスト１４０００からライトデータを受信し、当該実領域に対応するキャッシュメモリ内の領域にライトデータを格納する。そして、ストレージ装置１５０００Ａは、同期リモートコピー機能によりライトデータをストレージ装置１５０００Ｂにライトリクエストを送信する。

【０２６５】

次に、本実施の形態では、第１１の実施の形態と異なり、ストレージ装置１５０００Ａからライトリクエストを受信したストレージ装置１５０００Ｂは、即座にストレージ装置１５０００Ａからライト対象データを受信し、キャッシュメモリに当該データを格納した

後、ストレージ装置１５０００Ａにライト完了報告を行なう。ストレージ装置１５０００Ａは、ストレージ装置１５０００Ｂからライト完了報告を受信した後、仮想アドレス実アドレス変換テーブル１５００９Ａへの変更内容をストレージ装置１５０００Ｂに送信する。

【０２６６】

仮想アドレス実アドレス変換テーブル１５００９Ａへの変更内容を受信したストレージ装置１５０００Ｂは、同様の変更を仮想アドレス実アドレス変換テーブル１５００９Ｂに対して行なう。これによりストレージ装置１５０００Ｂ内の仮想ボリューム１５００８Ｂ内の当該ライト領域の実領域が共通の外部ストレージ装置１６０００Ａの実ボリューム１６００２Ａ内の（ストレージ装置１５０００Ａにより割り当てられた）実領域にマッピングされることになる。ストレージ装置１５０００Ｂは仮想アドレス実アドレス変換テーブル１５００９Ｂを更新した旨をストレージ装置１５０００Ａに通知する。その後、通知を受信したストレージ装置１５０００Ａはホスト１４０００に対してライト完了報告を行なう。なお、ストレージ装置１５０００Ａは（１）同期リモートコピーのデータ送信と、（２）仮想アドレス実アドレス変換テーブル１５００９Ａへの変更内容の送信を同時に行い、（１）及び（２）の処理の完了報告を受信した後ホスト１４０００に対してライト完了報告を行なってもよい。その後、ホスト１４０００がライト完了報告を受信し、ライト処理は完了する。

【０２６７】

なお、ストレージ装置１５０００Ａ内のキャッシュメモリに格納されたデータはキャッシュメモリへの格納とは非同期に、ストレージ装置１５０００Ａにより実ボリューム１６００２Ａへライト（デステージ）される。デステージが完了した後、ストレージ装置１５０００Ａはストレージ装置１５０００Ｂにその旨を通知する。通知を受けたストレージ装置１５０００Ｂは当該ライトに対応するキャッシュメモリの領域を破棄する。なお、破棄せずに当該ライトに対応するキャッシュメモリの領域の属性をクリーン（キャッシュメモリの内容と記憶装置（ＨＤＤなど）内のデータの内容が一致している状態）としてもよい。

【０２６８】

何らかの障害により、アプリケーション１４００２がストレージ装置１５０００Ａ内の仮想ボリューム１５００８Ａ経由でのリード／ライト処理が不可能になった場合、オペレーティングシステム１４００１の提供するパス管理機能は障害を検出し、リード／ライト処理のアクセス経路をストレージ装置１５０００Ｂ内の仮想ボリューム１５００８Ｂ経由に切り替える。仮想ボリューム１５００８Ａの内容と仮想ボリューム１５００８Ｂの内容は同期リモート機能により一致しているため、アクセス経路が切り替わっても、継続して正常にリード／ライト処理を行うことができる。

【０２６９】

（１３）第１３の実施の形態

本実施の形態ではストレージ装置内のボリュームに論理スナップショット機能を適用した場合の例を説明する。

【０２７０】

論理スナップショット機能とは、ローカルレプリケーションと類似した機能であり、ユーザーの指示時点の複製データをホストに提供する機能である。しかし、複製データを有する副ボリュームは、プールに属する実ボリュームの領域に保存された複製作成指示以後のライトデータと、正ボリュームのデータを用いて提供される仮想的な存在である。仮想的な副ボリュームの実体は実ボリュームの集合体であるプールに保持される。正ボリュームと副ボリュームの関係をスナップショットペアもしくは単にペアと呼ぶこともある。論理スナップショット機能においては、静止化ポイントにおける正ボリュームの内容と同一の内容の論理ボリュームが実際に作成される訳ではないという意味で、副ボリュームは仮想的である。論理スナップショット機能は先に説明したローカルコピー機能とは異なり、正ボリュームのサイズと同一のサイズの副ボリュームが不要である。これにより、副ボリ

ュームの内容を保持するために必要な記憶装置（ＨＤＤなど）の容量を削減することが可能である。

【０２７１】

本実施の形態ではこの論理スナップショット機能についてもアクティブ側のストレージとスタンバイ側のストレージで連携することで可用性を向上させることができる。

【０２７２】

図２３はスナップショット機能の一実施の形態を示したものである。図２３においては、ホスト１４０００はストレージ装置１５０００Ａとストレージ装置１５０００Ｂに接続されている。また、ストレージ装置１５０００Ａは外部ストレージ装置１６０００Ａと接続され、ストレージ装置１５０００Ｂは外部ストレージ装置１６０００Ｂと接続されている。また、スナップショット機能及び差分ビットマップ（静止化ポイントにおける正ボリューム１５００６Ａ，１５００６Ｂと現時点における正ボリューム１５００６Ａ，１５００６Ｂの間の差分の有無を示す情報）１５０１０Ａ，１５０１０Ｂと仮想アドレス実アドレス変換テーブル（仮想的な副ボリューム１５００７Ａ，１５００７Ｂの実体の位置を管理するテーブル）１５００９Ａ，１５００９Ｂがストレージ装置１５０００Ａとストレージ装置１５０００Ｂにて実行及び管理される。更に、ストレージ装置１５０００Ａ内の正ボリューム１５００６Ａとストレージ装置１５０００Ｂ内の正ボリューム１５００６Ｂはリモートコピーのペアとなるように構成される。

【０２７３】

本実施の形態は正ボリューム１５００６Ａ，１５００６Ｂがストレージ装置１５０００Ａ，１５０００Ｂ内にあり、プール１６００３Ａ，１６００３Ｂが外部ストレージ装置１６０００Ａ，１６０００Ｂ内にある構成例を示している。なお、プール１６００３Ａ，１６００３Ｂはストレージ装置１５０００Ａ，１５０００Ｂ内にあってもよい。

【０２７４】

＜論理スナップショット作成指示＞

ホスト１４０００を利用するユーザーが論理スナップショット作成を指示すると、前記実施の形態に記載の方式によって、アクティブ側のストレージ装置１５０００Ａとスタンバイ側のストレージ装置１５０００Ｂに作成指示を発行する。作成指示を受信したストレージ装置１５０００Ａ，１５０００Ｂは当該指示を受けて、仮想的な副ボリューム１５００７Ａ，１５００７Ｂを準備し、この副ボリューム１５００７Ａ，１５００７Ｂに全て０（差分なしの意味）の差分ビットマップ１５０１０Ａ，１５０１０Ｂと仮想アドレス実アドレス変換テーブル１５００９Ａ，１５００９Ｂとを割り当てる。

【０２７５】

＜正ボリュームに対するリード処理＞

これまで述べた実施の形態と同じである。

【０２７６】

＜正ボリュームに対するライト処理＞

アプリケーションプログラム１４００２からライトリクエストを受け付けたオペレーティングシステム１４００１はパス管理機能により、（ライト対象の正ボリュームに関して）アクティブ側のストレージがストレージ装置１５０００Ａ及びストレージ装置１５０００Ｂのどちらかを判断し、アクティブ側のストレージ装置１５０００Ａにライトリクエストを発行する。ライトリクエストを受信したストレージ装置１５０００Ａは、ライト対象アドレスの差分ビットマップ１５０１０Ａをチェックする。結果、１であればキャッシュメモリに正ボリューム１５００６Ａのライトデータとして、格納する。一方、０の場合は正ボリューム１５００６Ａの更新前のデータを副ボリューム１５００７Ａ用のデータとして用いるための以下に示すＣｏｐｙ‐Ｏｎ‐Ｗｒｉｔｅ処理を行う。

【０２７７】

（Ｓｔｅｐ１）プール１６００３Ａに属する実ボリューム１６００２Ａの記憶領域を確保する。

（Ｓｔｅｐ２）正ボリューム１５００６Ａから当該記憶領域へ更新前データをキャッシュ

メモリを利用しつつコピーする。
（Ｓｔｅｐ３）退避する更新前データの保存先を管理するプール管理情報を更新し、当該データがプール１６０００３Ａ内の実ボリューム１６０００２Ａのどの領域に保存されたかわかるようにする。
（Ｓｔｅｐ４）受信したライトデータをキャッシュメモリに正ボリューム１５０００６Ａの当該アドレス宛のデータとして保存し、ライト完了応答を返す。
【０２７８】
これと並行して、当該ライトデータはリモートコピー機能により、ストレージ装置１５０００Ａ内の正ボリューム１５０００６Ａからストレージ装置１５０００Ｂ内の正ボリューム１５０００６Ｂにコピーされ、同様の処理がなされる。そのため、各ストレージ装置１５０００Ａ，１５０００Ｂはそれぞれで仮想アドレス実アドレス変換テーブル１５００９Ａ，１５００９Ｂや差分ビットマップ１５０１０Ａ，１５０１０Ｂの管理を行う。
【０２７９】
＜副ボリュームに対するリード処理＞
アプリケーションプログラム１４０００２からライトリクエストを受け付けたオペレーティングシステム１４０００１はパス管理機能により、（リード対象の副ボリュームに関して）アクティブ側のストレージがストレージ装置１５０００Ａ及びストレージ装置１５０００Ｂのどちらかを判断し、アクティブ側のストレージ装置１５０００Ａにリードリクエストを発行する。リードリクエストを受信したストレージ装置１５０００Ａは、正ボリューム１５０００６Ａに対して記録していた差分ビットマップ１５０１０Ａをチェックする。結果、リード対象アドレスのビットが０であれば正ボリューム１５０００６Ａの同じアドレスに保存されたデータをホスト１４０００へ返し、オペレーティングシステム１４０００１は当該データをアプリケーション１４０００２へ返す。一方、リード対象アドレスのビットが１の場合は仮想アドレス実アドレス変換テーブル１５００９Ａを参照して、正ボリューム１５０００６Ａのリード対象アドレスに関する更新前のデータの場所を決定し、プール１６０００３Ａに属する実ボリューム１６０００２Ａからデータをホスト１４０００(アプリケーション１４０００２)へ返す。
【０２８０】
＜副ボリュームに対するライト処理＞
アプリケーションプログラム１４０００２からライトリクエストを受け付けたオペレーティングシステム１４０００１はパス管理機能により、（ライト対象の副ボリュームに関して）アクティブ側のストレージがストレージ装置１５０００Ａとストレージ装置１５０００Ｂのどちらかを判断し、アクティブ側のストレージ装置１５０００Ａにライトリクエストを発行する。ライトリクエストを受信したストレージ装置１５０００Ａは、正ボリューム１５０００６Ａに割り当てられたライト対象アドレスの差分ビットマップ１５０１０Ａをチェックする。結果、１であれば仮想アドレス実アドレス変換テーブル１５００９Ａを参照することで、正ボリューム１５０００６Ａの当該アドレスの更新前データが保存されたプール１６０００３Ａ内の実ボリューム１６０００２Ａの記憶領域を探し、当該領域へライトデータを保存する。一方、０の場合は以下の処理を行う。
【０２８１】
（Ａ）プール１６０００３Ａに属する実ボリューム１６０００２Ａの領域を確保する。
（Ｂ）確保した領域にライトデータを保存し、仮想アドレス実アドレス変換テーブル１５００９Ａを更新することで当該ライトデータがプール１６０００３Ａ内の実ボリューム１６０００２Ａのどの領域に保存されたかわかるようにする。
（Ｃ）差分ビットマップ１５０１０Ａの当該アドレスに対応するビットを１に更新する。
【０２８２】
これと並行して、当該ライトデータはリモートコピー機能により、ストレージ装置１５０００Ａ内の正ボリューム１５０００６Ａからストレージ装置１５０００Ｂ内の正ボリューム１５０００６Ｂにコピーされ、同様の処理がなされる。そのため、各ストレージ装置１５０００Ａ，１５０００Ｂはそれぞれで仮想アドレス実アドレス変換テーブル１５００９Ａ

，１５００９Ｂや差分ビットマップ１５０１０Ａ，１５０１０Ｂの管理を行う。
【０２８３】
＜Ｃｏｐｙ‐Ａｆｔｅｒ‐Ｗｒｉｔｅ処理＞
ストレージ装置１５０００Ａ，１５０００Ｂは、正ボリューム１５００６Ａ，１５００６Ｂに対するライト時に実行されるＣｏｐｙ‐Ｏｎ‐Ｗｒｉｔｅ処理の代わりとして以下に示すＣｏｐｙ‐Ａｆｔｅｒ‐Ｗｒｉｔｅ処理を実行してもよい。
【０２８４】
（Ｓｔｅｐ１）受信したライトデータをキャッシュメモリに正ボリューム１５００６Ａ，１５００６Ｂの当該アドレス宛のデータとして保存し、ライト完了応答を返す。ただし、当該ライトデータのデステージングは抑制する。
（Ｓｔｅｐ２）プール１６０００３Ａ，１６０００３Ｂに属する実ボリューム１６０００２Ａ，１６０００２Ｂの記憶領域を確保する。
（Ｓｔｅｐ３）正ボリューム１５００６Ａ，１５００６Ｂから当該記憶領域に更新前データをキャッシュメモリを利用しつつコピーする。
【０２８５】
（Ｓｔｅｐ４）退避した更新前データの保存先を管理するプール管理情報を更新し、当該データがプール１６０００３Ａ，１６０００３Ｂ内の実ボリューム１６０００２Ａ，１６０００２Ｂのどの領域に保存されたかわかるようにする。
（Ｓｔｅｐ５）抑制していたライトデータのデステージを許可する。
【０２８６】
＜障害＞
何らかの障害が発生し、アクティブ側の正ボリューム１５００６Ａ及び副ボリューム１５００７Ａに対するＩ／Ｏリクエストが処理できなくなった場合には、すでに説明した通り、オペレーティングシステム１４００１は、Ｉ／Ｏリクエストの対象を正ボリューム１５００６Ｂ及び副ボリューム１５００７Ｂに切り替えてアクセスを継続することができる。なお、前述の通り、好ましくはスナップショット機能の正ボリューム１５００６Ａ，１５００６Ｂ及び副ボリューム１５００７Ａ，１５００７Ｂは同一のストレージ装置１５０００Ａ，１５０００Ｂに対してライトリクエストを発行したいため、正ボリューム１５００６Ａ，１５００６Ｂに対する切り替えが必要な場合は副ボリューム１５００７Ａ，１５００７Ｂも同時に行い、逆に副ボリューム１５００７Ａ，１５００７Ｂに対する切り替えが必要な場合は正ボリューム１５００６Ａ，１５００６Ｂに対しても切り替えを行う連携を行うことがこのましい。
【０２８７】
（１４）第１４の実施の形態
本実施の形態では、第１３の実施の形態とは異なる論理スナップショット機能の実施の形態を説明する。図２４に本実施の形態の一構成例を示す。
【０２８８】
まず、本実施の形態と第１３の実施の形態との構成の違いは、外部ストレージ装置１６０００Ｂが存在せず、仮想的な副ボリューム１５００７Ａ，１５００７Ｂの実領域がいずれも外部ストレージ装置１６０００Ａ内のプール１６０００３Ａ内の領域に割り当てられている点である。その他の構成は第１３の実施の形態と同様である。
【０２８９】
なお、本実施の形態はストレージ装置１５０００Ａとストレージ装置１５０００Ｂが共通のプール１６０００３Ａとして、共通の外部ストレージ装置１６０００Ａ内の実ボリューム１６０００２Ａを用いるため、第１３の実施の形態と異なり、実ボリューム１６０００２Ａが外部ストレージ装置１６０００Ａ内にある構成に限定される。
【０２９０】
このように構成することにより、プール１６０００３Ａに必要とされる物理的な記憶装置（ＨＤＤなど）の容量を削減することができる。
【０２９１】

本実施の形態と第１３の実施の形態の処理動作の大きな違いは以下の通りである。
（Ａ）通常時はスタンバイ側のストレージ装置１５０００Ｂがキャッシュメモリから外部ストレージ装置１６０００Ａ内の実ボリューム１６００２Ａに対してライトを行なわない代わりに、アクティブ側のストレージ装置１５０００Ａが正ボリューム１５００６Ａ、副ボリューム１５００７Ａ、プール１６００３Ａ内の実ボリューム１６００２Ａに対応するデータをデステージングする時にスタンバイ側のストレージ装置１５０００Ｂにこれを伝え、スタンバイ側のストレージ装置１５０００Ｂではこれによってキャッシュメモリ上のデータを破棄する。
【０２９２】
（Ｂ）ストレージ装置１５０００Ａが仮想アドレス実アドレス変換テーブル１５００９Ａへの更新をストレージ装置１５０００Ｂに通知し、通知を受けたストレージ装置１５０００Ｂが仮想アドレス実アドレス変換テーブル１５００９Ｂを更新する。
【０２９３】
また、（Ａ）の処理に変えて、副ボリューム１５００７Ａ，１５００７Ｂ又はプール１６００３Ａ内の実ボリューム１６００２Ａに対応するデータについてはキャッシングを無効化してもよい。この場合、前述のＣｏｐｙ‐Ｏｎ‐ライト処理による更新前データ退避では正ボリューム１５００６Ａ，１５００６Ｂに対するライト完了までにプール１６００３Ａ内の実ボリューム１６００２Ａへの退避データ保存が含まれるため、性能が悪化するが、Ｃｏｐｙ‐Ａｆｔｅｒ‐ライト方式ではそれが無いため、好適である。
【０２９４】
以上、本発明の幾つかの実施態様を説明したが、これらの実施の形態は本発明の説明のための例示にすぎず、本発明の範囲をそれらの実施の形態にのみ限定する趣旨ではない。本発明は、その要旨を逸脱することなく、その他の様々な態様でも実施することができる。例えば、ＨＤＤ１０３０やキャッシュメモリ１０２０の代わりに不揮発性メモリを使用することができる。不揮発性メモリとしては、例えば、フラッシュメモリ（具体的には、例えば、ＮＡＮＤ型フラッシュメモリ）、ＭＲＡＭ（Magnetoresistive Random Access Memory）、ＰＲＡＭ（Parameter Random Access Memory）など、種々の不揮発性メモリを採用することができる。
【０２９５】
（１５）第１５の実施の形態
図１との対応部分に同一符号を付して示す図２８は、仮想化機能を持つネットワークスイッチ（仮想化スイッチ）２８０００Ｌ，２８０００Ｒを適用した場合の実施の形態を示した概要図である。
【０２９６】
＜１．本実施の形態のハードウェア構成＞
仮想化スイッチ２８０００Ｌ，２８０００Ｒは、複数のネットワークポートを持ち、ネットワークポート制御用のプロセッサが各ポートの転送制御や障害の検知や後術する仮想化を行う。なお、本概要図には図示されていないが、図１について上述した第１の実施の形態と同様に仮想化スイッチ２８０００Ｌ，２８０００Ｒには管理ホストが接続され、この管理ホストを介して仮想化スイッチ２８０００Ｌ，２８０００Ｒに対する設定を行ったり、仮想化スイッチ２８０００Ｌ，２８０００Ｒ間の設定コピーを行うことができる。なお、その他コンポーネントについては第１～第１４の実施の形態と同じであるため、説明を省略する。
【０２９７】
＜２．仮想化スイッチを用いた本実施の形態の特徴＞
仮想化スイッチ２８０００Ｌ，２８０００Ｒが提供する仮想化は第１～第１４の実施の形態と異なる以下の特徴を持つ。
【０２９８】
（特徴１）仮想的なＷＷＮ（又はポートネーム）を提供できる。ファイバーチャネルスイッチのポートはＦポート又はＥポートと呼ばれ、通常のホストやストレージが持つＮポー

ト（通信の始点又は終点となることを意味する）とは異なる属性を持つ。そのため、仮想化スイッチ２８０００Ｌ，２８０００Ｒにおいて仮想化を行う場合、内部で実際に接続されていない仮想的なＷＷＮを仮想化スイッチ２８０００Ｌ，２８０００Ｒの両方で作成・提供すれば、ホスト１１００上のソフトウェアが明示的にＩ／Ｏパスを切り替える必要がなくなる。なお、より具体的にはファイバーチャネルの通信は前述のポートネームによって行われるが、これはファイバーチャネルスイッチが割り当てる識別子であり、識別子内部にはルーティング用に用いられるスイッチを識別するための情報が含まれている。そのため、両仮想化スイッチ２８０００Ｌ，２８０００Ｒは、ホスト１１００に対して、仮想的なＷＷＮを持つＮポートが仮想的なスイッチを経由して両仮想化スイッチ２８０００Ｌ，２８０００Ｒに接続されているかのごとく模擬できるように、ポートネームを割り当てて、ルーティングを行う。

【０２９９】

（特徴２）スイッチでキャッシングを行わない。ファイバーチャネルスイッチは、通常、制御ヘッダだけ参照して転送先を決定し、データのバッファリングを行わない、いわゆるカットスルー方式で転送制御を行うため、仮想化機能を提供する場合もキャッシングを行わないことが多い。なお、キャッシングを行う場合は、本特徴が関係する処理についてはこれまで説明してきた実施の形態と同様の処理によって実現される。また、キャッシングを行わない場合の仮想化スイッチ２８０００Ｌ，２８０００Ｒのリード／ライト処理は、Ｉ／Ｏリクエストを受け付けてからストレージ装置１５００Ｌへのリクエスト処理転送及び処理完了を待ってホスト１１００に処理完了が返る、ライトスルー型の制御に類似したものと考えることができる。

【０３００】

（特徴３）本実施の形態での高可用化は両方の仮想化スイッチ２８０００Ｌ，２８０００Ｒに同じ仮想化の設定を行うだけでよい。これは仮想化スイッチ２８０００Ｌ，２８０００Ｒにてキャッシングを行っていないために可能となる。なお、仮想化スイッチ２８０００Ｌ，２８０００Ｒがリモートコピーやローカルコピーを行う場合、差分ビットマップ等スイッチ内部に存在する情報がある場合は、これまでの実施の形態と同じく、正系と副系の両方で内部情報を保持する必要がある。

【０３０１】

なお、ここまで仮想化スイッチ２８０００Ｌ，２８０００Ｒはファイバーチャネルスイッチであるものとして説明を行っているが、仮想化スイッチ２８０００Ｌ，２８０００ＲがＥｔｈｅｒｎｅｔ（登録商標）や、ｉＳＣＳＩ又はＴＣＰ／ＩＰを用いたものでもよい。この場合、ＷＷＮがＭＡＣアドレスで、ポートネームがＩＰアドレスに対応するものとすることができるが、Ｅｔｈｅｒｎｅｔ（登録商標）やＴＣＰ／ＩＰの場合は仮想的なスイッチを提供せずに、直接仮想的なポートとそれに割り当てたＩＰアドレスを外部へ提供し、当該ＩＰアドレスに対するルーティングを行えばよい。

【０３０２】

（１６）第１６の実施の形態

次に、第１６の実施の形態について説明する。本実施の形態は第１１の実施の形態及び第１２の実施の形態にて説明したＡＯＵ機能を高可用化構成の仮想化ストレージ装置が提供することに関する発明である。なお、以下において説明していない機能等については、第１～第１５の実施の形態による情報システムと同様の構成を有する。

【０３０３】

前述の通り、ＡＯＵ機能とは仮想化ストレージ装置が当該機能によってホスト１１００へ提供するボリューム（以後ＡＯＵボリュームと呼ぶ）について、使用開始時からＡＯＵボリュームの全アドレスに対してＨＤＤの記憶領域を割り当てるのではなく、ホスト１１００がライトを行ったアドレスに対してＨＤＤの記憶領域（より正確にはＨＤＤから構成されるボリューム（プールボリュームと呼ぶ）の記憶領域の一部又は全て）を割り当てる機能である。ＡＯＵ機能はＨＤＤを有効利用することができる。なぜならば、ホスト１１００で動作するファイルシステムの一部の種類ではアクセス継続中に動的なデータ容量拡

張ができないため、ホスト１１００の管理者は将来利用する可能性のあるデータ容量も含めてボリュームの容量設定を行う。そのため、従来技術ではボリュームの容量設定を行った時点では使用せず、また将来も確実に使うとは限らないＨＤＤを搭載していなければならなかったからである。

【０３０４】

なお、ＨＤＤ容量の有効利用の観点から考えた場合、ＡＯＵボリュームにライトが発生する前の領域に対してプールボリュームの領域が未割り当ての状態であったほうがよいが、他の目的（高性能化等）がある場合はこれに限られない。

【０３０５】

＜１．本実施の形態の概要＞

図１との対応部分に同一符号を付した図２９に本実施の形態の概要を示す。本実施の形態による情報システムは、２台の仮想化ストレージ装置１０００Ｌ，１０００Ｒが共通のストレージ装置１５００Ｌと接続されている。そして、高可用化された２台の仮想化ストレージ装置１０００Ｌ，１０００ＲがＡＯＵ機能を有することで、情報システムのサービス停止時間を短縮している。なお、特に記載しない限り、ストレージ装置１５００Ｌは両方の仮想化ストレージ装置１０００Ｌ，１０００Ｒからアクセス可能な状態、言い換えれば共有された状態にあるものとするが、共有されていないストレージ装置が存在して当該ストレージ装置内のボリュームをＡＯＵの記憶領域として用いてもよい。また図２９には図示していないが、本実施の形態の場合、第１の実施の形態と同様に、仮想化ストレージ装置１０００Ｌ，１０００Ｒには管理ホスト１２００（図１）が接続されている。

【０３０６】

ここでは、これまでに説明してきた実施の形態と異なる部分を中心に述べていく。２台の仮想化ストレージ装置１０００Ｌ，１０００Ｒは、ＡＯＵアドレス変換情報３１０３０Ｌ，３１０３０Ｒを用いてＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒを生成し、これをホスト１１００に提供する。ＡＯＵアドレス変換情報３１０３０Ｌ，３１０３０Ｒには、ＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒのアドレス空間と仮想化ストレージ装置１０００Ｌ，１０００Ｒ内のプールボリュームの領域又はストレージ装置１５００Ｌ内のプールボリュームの領域との対応関係が含まれる。

【０３０７】

ホスト１１００からＡＯＵボリューム２９０１０Ｌに対してライトリクエストが発行されると、正系の仮想化ストレージ装置１０００Ｌは、リクエスト対象のアドレス範囲にプールボリュームの領域が割り当てられているか判断し、割り当てられていない場合は仮想化ストレージ装置１０００Ｌ又はストレージ装置１５００Ｌが有するプールボリュームの領域を割り当てる。そしてライトリクエストが処理されることで、正系の仮想化ストレージ装置１０００Ｌのキャッシュ領域にライトデータが保存される。また、ＡＯＵボリューム２９０１０Ｌに対するライトデータは同期リモートコピーによって副系の仮想化ストレージ装置１０００Ｒへ転送され、正系と同様にキャッシュ領域にライトデータが保存される。

【０３０８】

その後、両方の仮想化ストレージ装置１０００Ｌ，１０００Ｒはデステージング処理を行うが、ストレージ装置１５００Ｌに対応したライトデータに対しては仮想化ストレージ装置１０００Ｌ，１０００Ｒの片方だけがデステージングを行う。なぜならば、両方の仮想化ストレージ装置Ｌ，１０００Ｒが独立にライトデータのデステージングを行うと、ストレージ装置１５００Ｌに保存されるデータが不整合な状態（例えば、最後にライトしたデータが消えて、前回のライトデータに戻ってしまう等のデータ消失やライト順序の不整合がある）になってしまうからである。そのため、デステージングが必要になる前に予めどちらの仮想化ストレージ装置１０００Ｌ，１０００Ｒがデステージングを行うか決めておく必要がある。本実施の形態ではその一例として正系の仮想化ストレージ装置１０００Ｌがデステージングを行う場合について説明を行うが、副系の仮想化ストレージ装置１０００Ｒが行ってもよく、又はデステージング対象のアドレス空間を元にどちらの仮想化ス

トレージ装置１０００Ｌ，１０００Ｒが当該役割を担うか決定してもよい。
【０３０９】
リードリクエストの場合も、正系の仮想化ストレージ装置１０００Ｌは、まずはリクエスト対象のアドレス範囲にプールボリュームの領域が割り当てられているかどうかを判断する。判断の結果、割り当てられている領域に対しては、仮想化ストレージ装置１０００Ｌは、該当するプールボリュームの領域（図示しないキャッシュメモリ上のデータを含む）からデータを読み出してホスト１１００へ転送し、割り当てられていない場合は予め定められた値（例えばゼロ）を返す。
【０３１０】
図３０は、正系の仮想化ストレージ装置１０００Ｌの機能停止後に副系の仮想化ストレージ装置１０００ＲへＩ／Ｏリクエスト処理を切り替えた後の概要図である。本図にあるとおり、副系の仮想化ストレージ装置１０００ＲはＡＯＵアドレス変換情報３１０３０Ｒを元にストレージ装置１５００Ｌや仮想化ストレージ装置１０００Ｌ内部のＡＯＵアドレス変換情報３１０３０Ｒを用いてＩ／Ｏリクエストを処理する。そのために、正系と副系の仮想化ストレージ装置１０００Ｌ，１０００Ｒは、通常時から通信を行うことで、ＡＯＵアドレス変換情報３１０３０Ｌ，３１０３０Ｒのストレージ装置１５００Ｌが関係する部分を同じ内容に維持する。これにより副系の仮想化ストレージ装置１０００Ｒは、ストレージ装置１５００Ｌの割り当て状況を引き継ぐことができる。また、副系の仮想化ストレージ装置１０００Ｒは、正系の仮想化ストレージ装置１０００Ｌ内のキャッシュメモリからデステージングされたデータでない限り、キャッシュメモリに格納されているデータを当該キャッシュメモリから削除しないようにする。これにより、機能停止時に正系の仮想化ストレージ装置１０００Ｌ内のキャッシュメモリからデータが揮発した場合においてもデータ消失が発生しないようにすることができる。
【０３１１】
＜２．仮想化ストレージ装置で実行されるプログラム及び情報＞
図６との対応部分に同一符号を付した図３１は、仮想化ストレージ装置１０００Ｌ，１０００Ｒ上で実行されるソフトウェアプログラムと当該プログラムが用いる情報について示している。
【０３１２】
この図３１において、ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、仮想化ストレージ装置１０００Ｌ，１０００Ｒが受信したＩ／Ｏリクエストを処理するプログラムで、第１～第１４の実施の形態におけるＩ／Ｏ処理プログラム６０２０（図６）の機能を一部に含む。
【０３１３】
ＡＯＵ管理プログラム３１０２０は、ＡＯＵ機能に関する設定や後術する重複削除（Deduplication）処理を実行するためのプログラムである。またＡＯＵアドレス変換情報３１０３０は、ＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒとプールボリュームの領域との対応関係に関する情報である。さらにＡＯＵプール管理情報３１０４０は、ＡＯＵ機能が用いるプールボリュームの集合（プール）を管理するための情報である。
【０３１４】
＜２．１．ＡＯＵアドレス変換情報＞
図３５はＡＯＵアドレス変換情報３１０１０の具体的な内容を示している。仮想化ストレージ装置１０００Ｌ，１０００Ｒは、ホスト１１００へ提供するボリュームの識別子と当該ボリューム内のアドレス空間を先頭から決められた大きさ（セグメントサイズ）に区切った領域（セグメント）のアドレスとでデータの保存領域等を管理する。なお、このセグメントサイズはプール定義時に設定される値である。
【０３１５】
図３５において、「ＡＯＵボリューム識別子」及び「アドレス空間」は、対応するセグメントを含むＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒの識別子と当該ＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒ内における当該セグメントのアドレスとをそれぞれ示す

。またプールＩＤは、そのＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒに領域を割り当てるプールの識別子を示す。

【０３１６】

「ＣＯＷ（Copy On Write）フラグ」は、そのセグメントに対するライトリクエストが到着したときに、対応するライトデータを別途確保したプールボリューム領域に保存する必要があるかどうかを示すフラグである。本フラグは、異なるセグメントが同じプールボリュームの領域に対応付けられていた場合に、ライトデータを他のプールボリュームに保存する必要があることを意味する「ＯＮ」となることがある。

【０３１７】

「プールボリューム領域識別子」は、そのセグメントが保存すべきデータを実際に保存しているプールボリューム領域の識別子を示す情報（識別情報）である。この識別情報は例えば以下の情報から構成される。

【０３１８】

（１）仮想化ストレージ装置１０００Ｌ，１０００Ｒの内部のボリュームの領域を用いている場合は、内部ボリュームの識別子及びアドレス範囲。

（２）ストレージ装置１５００Ｌ内のボリュームの領域をもちいている場合は、ポートネーム等の装置又は通信先を識別する情報、ＬＵＮ等の装置内のボリュームを識別する情報、及びアドレス範囲。

（３）未割り当て領域の場合はＮＵＬＬ

【０３１９】

「引継ぎ領域」は、対応する「プールボリューム領域識別子」の欄に識別子が記載されたプールボリュームが正系と副系の仮想化ストレージ装置１０００Ｌ，１０００Ｒの両方から管理されるかどうかを示す情報（両方から管理される場合は「Ｙｅｓ」、両方から管理されない場合は「Ｎｏ」）である。

【０３２０】

「対ＡＯＵボリューム識別子」は、対応するＡＯＵボリューム識別子で特定されるボリュームと対を形成するＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒの識別子が保持される。この識別子としては、対応する仮想化ストレージ装置１０００Ｌ，１０００Ｒの識別子と、対応するＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒの識別子とを組み合せたものが用いられる。

【０３２１】

なお、上述のようにＡＯＵの領域管理をセグメントを単位として行うのは、ブロック単位で管理するとＡＯＵアドレス変換情報３１０３０等の管理情報が大きくなりすぎてしまうためにＩ／Ｏ性能が悪化してしまうことが理由の一つである。

【０３２２】

＜２．２．ＡＯＵプール管理情報＞

図３６は、ＡＯＵプール管理情報３１０４０の具体的な構成を示している。ＡＯＵプール管理情報３１０４０はプール毎に以下の情報を保持する。

【０３２３】

（１）セグメントサイズ

（２）プールに割り当てられたボリューム（プールボリューム）のリスト

（３）プールボリュームの領域で割り当てられていない領域のリスト

（４）空き容量

（５）容量が不足してきたことを警告するアラートを出すスレッショルド値

（６）プール対の相手が設定された仮想化ストレージ装置の識別子と当該装置内のプールＩＤ。なお、「プール対」については後ほど説明する。

【０３２４】

＜３．初期化＞

本実施の形態の初期化は以下の手順で行われる。

【０３２５】

１．プールの初期化
２．ＡＯＵボリュームの生成
３．ＡＯＵボリューム同士の関連付け
４．同期リモートコピーの設定
【０３２６】
以下に、その詳細について説明する。なお、以下においては、一部処理の処理主体を「管理ホスト」や「プログラム」として説明しているが、「管理ホスト」の部分については、その管理ホスト内のプロセッサが当該管理ホスト内のメモリに格納された対応するプログラムに基づいてその処理を実行し、「プログラム」の部分については、対応する仮想化ストレージ装置１０００Ｌ，１０００Ｒ内のプロセッサ１０１１がそのプログラムに基づいて処理を実行することは言うまでもない。
【０３２７】
＜３．１．プールの初期化＞
以下の手順で初期化を行う。
【０３２８】
（Ｓｔｅｐ１）管理ホスト１２００からの指示によって、仮想化ストレージ装置１０００Ｌ，１０００Ｒの片方で実行されるＡＯＵ管理プログラム３１０２０は、プールを作成する。この時、当該指示にはセグメントサイズが含まれる。また、プール作成の過程で、ＡＯＵ管理プログラム３１０２０はプールＩＤを含むＡＯＵプール管理情報３１０４０の該当エントリを作成する。
（Ｓｔｅｐ２）Ｓｔｅｐ１と同様の処理によって、仮想化ストレージ装置１０００Ｒ，１０００Ｌのもう片方にもプールを作成する。
【０３２９】
（Ｓｔｅｐ３）管理ホスト１２００は、仮想化ストレージ装置１０００Ｌ，１０００Ｒの両方に対して、Ｓｔｅｐ１とＳｔｅｐ２で作成したプールをプール対に設定すべき指示を発行する。当該指示にはプール対となるプールのＩＤと、そのプールを提供する仮想化ストレージ装置１０００Ｌ，１０００Ｒの識別子との組が含まれる。当該指示を受信したＡＯＵ管理プログラム３１０２０は、プール対となる相手の仮想化ストレージ装置１０００Ｌ，１０００ＲのＡＯＵ管理プログラム３１０２０と通信を行い、両プールに設定されたセグメントサイズが等しく、両プールとも既にプール対になっていないことが確認できたときに、それらプールをプール対に設定する。なお、ＡＯＵ管理プログラム３１０２０は、かかるプールをプール対に設定するに際して、相手のプールＩＤの識別子をＡＯＵプール管理情報３１０４０に登録する。
【０３３０】
（Ｓｔｅｐ４）管理ホスト１２００は、プールボリューム作成の指示を仮想化ストレージ１０００Ｌ，１０００Ｒの片方に発行する。なお、当該指示には仮想化ストレージ装置１０００Ｌ，１０００Ｒ内部に定義されたボリュームの識別子が含まれている。当該指示を受信した仮想化ストレージ装置１０００Ｌ，１０００ＲのＡＯＵ管理プログラム３１０２０は、指定されたボリュームの属性をプールボリュームに変更し、指定されたボリュームの識別子をＡＯＵプール管理情報３１０４０のプールボリュームリストに追加する。
【０３３１】
（Ｓｔｅｐ５）管理ホスト１２００は、Ｓｔｅｐ３と同様の指示を仮想化ストレージ装置１０００Ｒ，１０００Ｌのもう片方へ発行する。指示を受け取った仮想化ストレージ装置１０００Ｒ，１０００Ｌのもう片方はＳｔｅｐ３と同様の処理を行う。
【０３３２】
なお、管理者が仮想化ストレージ装置１０００内部のボリュームをＡＯＵに用いないと判断した場合はＳｔｅｐ４及びＳｔｅｐ５は省略することができる。
【０３３３】
（Ｓｔｅｐ６）管理ホスト１２００は、ストレージ装置１５００Ｌのボリュームをプールボリュームに設定する指示を仮想化ストレージ装置１０００Ｌ，１０００Ｒのどちらかに

発行する。なお、理解を容易にするために、以後の説明では、指示発行先を仮想化ストレージ装置１０００Ｌ、支持発行先と対の仮想化ストレージ装置１０００を仮想化ストレージ装置１０００Ｒであるものとするが、逆の関係であってもよい。ここで、当該指示にはストレージ装置１５００Ｌと当該ボリュームを識別する情報のほかに、当該ボリュームがプール対になっている相手の仮想化ストレージ装置１０００Ｒが引き継ぐことを示す情報が含まれる。指示を受信した仮想化ストレージ装置１０００ＬのＡＯＵ管理プログラム３１０２０は、対の仮想化ストレージ装置１０００Ｒと以下に示す連携を行う。

【０３３４】

（Ａ）指示を受信した仮想化ストレージ装置１０００Ｌは、指示に含まれるストレージ装置１５００Ｌのボリュームに対してリードリクエスト（又はライトリクエスト）を発行することで、当該ストレージ装置１５００Ｌ及び当該ボリュームのいずれもが存在し、かつ当該ボリュームにアクセス可能であるかどうか確認する。ストレージ装置１５００Ｌやかかるボリュームが存在しなかったり、当該ボリュームにアクセスができなかった場合は管理ホスト１２００にエラーを返し、アクセス可能であった場合は次に進む。なお、当該エラーにはストレージ装置１５００Ｌに対するアクセスができなかったことを示す情報を添付するものとし、この情報を管理ホスト１２００において表示するようにしてもよい。

【０３３５】

（Ｂ）指示を受信した仮想化ストレージ装置１０００Ｌは、対の仮想化ストレージ装置１０００Ｒに対してプールボリューム作成指示を転送する。なお、当該指示には管理ホスト１２００からの指示に含まれていた対象ボリュームを識別する情報と、当該ボリュームがプール対に属する両方のプールで管理することを示す情報とが含まれている。なお、プールボリューム作成指示の転送先は、ＡＯＵプール管理情報３１０４０における「プール対の識別情報」を参照することで特定できる。

【０３３６】

（Ｃ）仮想化ストレージ装置１０００Ｒは、（Ｂ）の指示を受信すると、（Ａ）と同様の処理を行うことでストレージ装置１５００Ｌ内のかかるボリュームにアクセス可能であることを確認する。そして、かかるボリュームにアクセス可能ならば、ＡＯＵプール管理情報３１０４０のプールボリュームリストに当該ボリュームを共通管理であることを示す情報と共に追加し、前述の指示を受信した仮想化ストレージ装置１０００Ｌへ結果を返す。一方、確認の結果、かかるボリュームにアクセスが不可能だった場合は、対の仮想化ストレージ装置１０００Ｒからストレージ装置１５００Ｌへのアクセスができなかったことを示す情報を添付して失敗を意味する結果を返す。

【０３３７】

（Ｄ）結果を受け取った前述の指示を受信した仮想化ストレージ装置１０００Ｌは、かかるボリュームへのアクセス結果が失敗だった場合は、理由と共にその結果を管理ホスト１２００へ転送し、一連の処理を終了する。一方、かかるボリュームへのアクセス結果が成功だった場合は、この結果をＡＯＵプール管理情報３１０４０のプールボリュームリストに当該ボリュームが共通管理すべきボリュームであることを示す情報と共に追加し、管理ホスト１２００へ成功を意味する結果を転送し、終了する。

【０３３８】

なお、（Ｃ）及び（Ｄ）でボリュームをプールボリュームリストに追加した場合、ＡＯＵ管理プログラム２１０２０は、対応する「空き容量」の欄に格納されていた空き容量を、追加したボリュームの容量を加算した値に更新し、そのボリュームの領域を空き領域リストに追加する。また、Ｓｔｅｐ５の処理は管理ホスト１２００から仮想化ストレージ装置１０００Ｌ，１０００Ｒの両方に対して別々に指示を発行することで行っても良い。

【０３３９】

（Ｓｔｅｐ７）管理ホスト１２００は、仮想化ストレージ装置１０００Ｌ，１０００Ｒの両方に対して容量警告の設定値の設定指示を転送し、当該指示を受信した仮想化ストレージ装置１０００Ｌ，１０００Ｒの各々は指示に含まれる値をＡＯＵプール管理情報３１０４０に設定する。

【０３４０】
＜３．２．ＡＯＵボリュームの作成＞
ＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒの作成は、仮想化ストレージ装置１０００Ｌ，１０００Ｒの各々に対して指示を出すことで行われる。以下にその手順を示す。
【０３４１】
（Ｓｔｅｐ１）管理ホスト１２００は、ボリュームの容量とプールＩＤを伴ったＡＯＵボリューム作成指示を、仮想化ストレージ装置１０００Ｌ，１０００Ｒの各々へ転送する。
（Ｓｔｅｐ２）当該指示を受信した仮想化ストレージ装置１０００Ｌは、新しいＡＯＵボリューム２９０１０Ｌに関するＡＯＵアドレス変換情報３１０３０Ｌを作成する。このとき全てのセグメントについて、対応する「ＣＯＷフラグ」及び「引継ぎ領域」を「Ｎｏ」に設定し、「プールボリューム領域識別子」も「ＮＵＬＬ」に設定する。そして仮想化ストレージ装置１０００Ｌは作成完了応答を返す。
（Ｓｔｅｐ３）同様に、該指示を受信した仮想化ストレージ装置１０００Ｒは、新しいＡＯＵボリューム２９０１０Ｒに関するＡＯＵアドレス変換情報３１０３０Ｒを作成する。作成の詳細はＳｔｅｐ２と同様である。
【０３４２】
なお、ＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒの作成は、管理ホスト１２００から仮想化ストレージ装置１０００Ｌ，１０００Ｒのいずれか片方に指示を出し、指示を受けた仮想化ストレージ装置１０００Ｌ，１０００Ｒが対の仮想化ストレージ装置Ｒ，１０００Ｌに指示を出し直してもよい。なお、ボリューム作成指示にポートネームやＬＵＮを含めることでＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒに管理者が指定したポートネーム配下のＬＵＮを割り当ててもよい。また、ＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒの作成指示後にポートネームとＬＵＮを割り当ててもよい。
【０３４３】
＜３．３．ＡＯＵボリューム同士の関連付け＞
それぞれの仮想化ストレージ装置１０００Ｌ，１０００Ｒに作成したＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒ同士を関連付ける。そのために、管理ホスト１２００は、仮想化ストレージ装置１０００Ｌ，１０００Ｒにそれら２つのＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒの識別子を含む関連付け指示を転送する。当該指示を受けた仮想化ストレージ装置１０００Ｌ，１０００Ｒは、ＡＯＵアドレス変換情報３１０３０の該当する「ＡＯＵボリューム識別子」の欄に、対となるＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒを登録する。本指示は、それぞれの仮想化ストレージ装置１０００Ｌ，１０００Ｒに対して行うことでＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒ同士の関連付けが行われるが、他の実施の形態に開示されている通り、片方の仮想化ストレージ装置１０００Ｌ，１０００Ｒがもう片方の仮想化ストレージ装置１０００Ｒ，１０００Ｌに本指示を転送する形で実現してもよい。
【０３４４】
なお、上記関連付けの指示の際、指示に含まれるＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒの存在を確認すると共に、ＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒの片方がプール対の片方のプールから生成されており、もう片方のＡＯＵボリューム２９０１０Ｒ，２９０１０Ｌがプール対のもう片方のプールから生成されていることを確認することで、プール管理の実装を簡単にしてもよい。また、本関連付けはＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒの作成や同期リモートコピーの設定に伴って行われても良い。
【０３４５】
＜３．４．同期リモートコピーの設定＞
これまで説明した同期リモートコピーではＩｎｉｔｉａｌ‐Ｃｏｐｙｉｎｇ状態でボリュームの全ての領域をコピーする必要があったが、本実施の形態では、形成コピーは以下に示す手順で行う。また、理解を簡単にするために、以後の説明では正系の仮想化ストレージ装置１０００を仮想化ストレージ装置１０００Ｌとし、副系の仮想化ストレージ装置１０００を仮想化ストレージ装置１０００Ｒであるものとして説明する。

【０３４６】
（Ｓｔｅｐ１）コピー元となる（すなわち当該ボリュームに対しては正系となる）仮想化ストレージ装置１０００Ｌは、変数ｉにＡＯＵボリューム２９０１０Ｌの先頭セグメントを代入する。
【０３４７】
（Ｓｔｅｐ２）コピー元の仮想化ストレージ装置１０００Ｌは、ＡＯＵアドレス変換情報３１０３０におけるセグメントｉの「引継ぎ領域」及び「プールボリューム領域識別子」をそれぞれ確認し、それぞれの条件下で以下の処理をおこなう。
（Ａ）「引継ぎ領域」が「Ｎｏ」の場合には、通常の形成コピーに従ってセグメントｉのデータをコピーする。仮想化ストレージ装置１０００Ｌ内部のプールボリュームの領域のため、冗長性確保のためにコピーしなければならないからである。
（Ｂ）「引継ぎ領域」が「Ｙｅｓ」の場合には、セグメントｉに関するその仮想化ストレージ装置１０００Ｌ内の図示しないキャッシュメモリ上のダーティデータをデステージングするか、形成コピーでコピー先（すなわち当該ボリュームに対しては副系となる）仮想化ストレージ装置１０００Ｒのキャッシュ領域へコピーする。キャッシュメモリ上のデータを除けばデータは正系の仮想化ストレージ装置１０００Ｌの外部にあるため、キャッシュメモリ上のデータを当該仮想化ストレージ１０００Ｌの外部へ追い出せば正系の仮想化ストレージ装置１０００Ｌが機能停止しても失うデータはないからである。
（Ｃ）「プールボリューム領域識別子」が「ＮＵＬＬ」の場合には、セグメントｉには正系・副系共に領域が割り当てられていないため、コピーは行わない。
【０３４８】
（Ｓｔｅｐ３）コピー元の仮想化ストレージ装置１０００Ｌは、セグメントｉが最後の場合は形成コピーを終えてペア状態をＤｕｐｌｅｘ状態に遷移させ、そうでない場合は変数ｉに次のセグメントを設定してＳｔｅｐ１へ戻る。
【０３４９】
なお、上記処理は、仮想化ストレージ装置１０００Ｌ，１０００Ｒ間の再同期処理で用いてもよく、片方の仮想化ストレージ装置１０００Ｌ，１０００Ｒが機能停止し、復旧した後の処理で用いてもよい。
【０３５０】
＜４．Ｉ／Ｏリクエスト処理について＞
ここから、本実施の形態のＩ／Ｏリクエスト処理について説明する。
【０３５１】
＜４．１．ライトリクエスト処理＞
図３２は、ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０がライトリクエストを受信したときに実行する処理内容を示すフローチャートである。なお、これまでの説明では、ライトリクエストを構成するコマンドとライトデータの個々についてフローチャートを用いた説明はしなかったが、本処理はライトリクエスト対象の一部の領域が割り当て済みで他の領域が未割り当てである場合もあるため、フローチャートを参照しながら詳細に説明する。
【０３５２】
（Ｓ３２００１）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、ライトリクエストを構成するライトコマンドを受信する。このライトコマンドにはアドレス（位置）及びデータ長が含まれる。
【０３５３】
（Ｓ３２１００）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、受信したライトコマンドを元に割り当て処理を実行する。本処理を実行することで、ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、セグメント毎にプールボリュームの領域が割り当てられているか否かをチェックし、プールボリュームの領域が未割り当てのセグメントや、他のセグメントと共有の領域を割り当てられている場合で「ＣＯＷフラグ」が「ＯＮ」のセグメント（ライトの際には共有領域以外に書き込む必要があるセグメント）に対しては、プールボリュ

ームの領域を割り当てる。またＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、かかるプールボリュームの領域の割り当て結果をＡＯＵアドレス変換情報３１０３０に反映させる。

【０３５４】

（Ｓ３２００３）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、ＡＯＵボリューム２９０１０Ｒ，２９０１０Ｌの属性を確認し、当該ＡＯＵボリューム２９０１０Ｒ，２９０１０Ｌがコピー元ボリュームの場合にはＳ３２００４を実行し、そうでなければＳ３２００５を実行する。

（Ｓ３２００４）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、リモートコピープログラム６０１０を呼び出すことでコピー先ボリュームを有する仮想化ストレージ装置（副系の仮想化ストレージ装置）１０００Ｒに同期リモートコピーのコマンドを転送する。

（Ｓ３２００５）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、Ｓ３２００１に対応するライトリクエストを構成するライトデータ（の一部又は全て）を受信する。

【０３５５】

（Ｓ３２００６）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、ＡＯＵボリューム２９０１０Ｒ，２９０１０Ｌの属性を確認し、当該ＡＯＵボリューム２９０１０Ｒ，２９０１０Ｌがコピー元ボリュームの場合はＳ３２００７を実行し、そうでなければＳ３２００８を実行する。

（Ｓ３２００７）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、リモートコピープログラム６０１０を呼び出すことで、コピー先ボリュームを有する仮想化ストレージ装置（副系の仮想化ストレージ装置）１０００Ｒにライトデータを転送する。

【０３５６】

（Ｓ３２００８）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、ＡＯＵアドレス変換情報３１０３０に基づいてＡＯＵボリューム２９０１０Ｒ，２９０１０Ｌ上のアドレスから、実際にライトデータが保存されるプールボリュームの領域を求める。そして求めた領域に対するライトデータをキャッシュメモリ上で保存・管理する。

（Ｓ３２００９）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、ライトデータ受信の続きの有無を判断し、続きがある場合はＳ３２００５を再び実行する。

（Ｓ３２０１０）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、ライト完了の応答を正系の仮想化ストレージ装置１０００Ｌ又はホスト１１００に転送し、このライトリクエスト処理を完了する。

【０３５７】

なお、副系の仮想化ストレージ装置１０００Ｒは、同期リモートコピーのコマンドの受信をホスト１１００からのライトコマンドの受信と同様に扱う。同様に、仮想化ストレージ装置１０００Ｒは、同期リモートコピーのデータ転送によるデータ受信をホスト１１００からのライトデータの受信と同様に扱う。これによって副系の仮想化ストレージ装置１０００Ｒでのライトリクエスト処理が理解されるであろう。

【０３５８】

＜４．１．１．割り当て処理＞

以下に図３２の割り当て処理について説明する。

【０３５９】

（Ｓ３２１０１）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、ライトコマンドで指定されたライト範囲（即ちライトアドレスとデータ長）をセグメント毎に分割する。

（Ｓ３２１０２）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、分割で生成した複数のセグメントの最初のセグメントを変数ｉに代入する。

【０３６０】

（Ｓ３２１０３）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、セグメントｉの割り当て状態やＣＯＷ（Copy On Write）が必要かどうかを判断する。なお、このときの判断にはＡＯＵアドレス変換情報３１０３０を用いる。かかる判断の結果、プールボリュームの領域の割り当てが不要な場合はＳ３２１０５を実行し、プールボリュームの領域が未割り

当ての場合や、割り当て済みであってもＣＯＷフラグが立っている場合は（例えば、他のＡＯＵボリューム２９０１０Ｒ，２９０１０Ｌ上のセグメントと割り当て領域を共有している場合）Ｓ３２１０４を実行する。

【０３６１】

（Ｓ３２１０４）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、セグメントｉに割り当てるため、プールボリュームの領域から未使用の領域を探す。そして、探した領域をＡＯＵアドレス変換情報３１０３０の「プールボリューム領域識別子」へ登録する。なお、未使用領域が見つからない場合は、ライトコマンドが失敗したことを示す応答を転送して、この割り当て処理を終了する。

【０３６２】

なお、失敗応答を転送する際には、この失敗応答と共に何らかのエラーメッセージを返しても良く、当該失敗応答の原因としてプールの容量不足が原因であることを示す情報を含めても良い。さらに、「ＣＯＷフラグ」が立っている場合の領域割り当ての場合、ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、領域割り当てに際して旧領域（共有領域）から割り当て領域へデータコピーを行うようにしてもよい。ただし、セグメントｉ全体がライト対象の場合はこのデータコピーを省略することができる。また、領域の割り当てに伴って、ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、ＡＯＵプール管理情報の空き領域リストを編集し、空き容量の削減を行うようにしてもよい。

【０３６３】

さらに、ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、割り当てたプールボリューム上の領域と当該領域を割り当てたＡＯＵボリューム２９０１０Ｒ，２９０１０Ｌのセグメントの情報を副系の仮想化ストレージ装置１０００Ｒへ転送する。なお、当該割り当て情報は同期リモートコピーのコマンドと共に転送してもよい。

【０３６４】

（Ｓ３２１０５）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、次のセグメントが存在するかどうか確認し、存在する場合はＳ３２１０６を実行し、存在しない場合は本処理を終了し、ライトリクエスト処理へ戻る。

（Ｓ３２１０６）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、変数ｉに次のセグメントを代入する。

【０３６５】

以上の処理によって、仮想化ストレージ装置１０００Ｌは、セグメント毎の割り当て状況を確認し、必要ならばセグメントにプールボリュームの領域を割り当てる。

【０３６６】

＜４．１．２．副系のプールボリューム領域割り当て方法＞

副系の仮想化ストレージ装置１０００Ｒのプールボリューム領域割り当てステップ（Ｓ３２１０４）は、正系の仮想化ストレージ装置１０００Ｌから受信した割り当て情報を元に以下の方法によって、セグメントに対して領域を割り当てる。

【０３６７】

（Ａ）正系の仮想化ストレージ装置１０００Ｌが共有のストレージ装置（すなわちストレージ装置１５００Ｌ）のプールボリュームから領域を割り当てた場合には、副系の仮想化ストレージ装置１０００Ｒは、ＡＯＵアドレス変換情報３１０３０における対応するセグメントの「引継ぎ領域」を「Ｙｅｓ」に、「プールボリューム領域識別子」を受信した領域識別子に設定する。これによって、共有ストレージ装置１５００Ｌに関するプールボリューム領域の割り当ては正系と副系で同じ対応になる。

【０３６８】

（Ｂ）正系の仮想化ストレージ装置１０００Ｌが仮想化ストレージ装置１０００Ｒ内部のボリュームから領域を割り当てた場合には、副系の仮想化ストレージ装置１０００Ｒは、内部ボリュームの空き領域を探して該当するセグメントに割り当てる。その結果、ＡＯＵアドレス変換情報３１０３０における当該セグメントの「引継ぎ領域」は「Ｎｏ」に、「プールボリューム領域識別子」は内部ボリュームの領域が設定される。これによって、正

系の仮想化ストレージ装置１０００Ｌが内部ボリュームの領域を割り当てたセグメントは副系の仮想化ストレージ装置１０００Ｒでも内部ボリュームを割り当てることができる。

【０３６９】

＜４．２．リードリクエスト処理＞

図３３は、ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０が、リードリクエストを受信したときに実行する処理内容を示すフローチャートである。以下に当該フローチャートを参照して、かかる処理内容について説明する。

【０３７０】

（Ｓ３３００１）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、リードリクエストを構成するリードコマンドを受信する。なお、受信したリードコマンドにはアドレス（位置）及びデータ長が含まれる。

（Ｓ３３００２）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、リードコマンドで指定されたリード範囲（即ちライトアドレスとデータ長）をセグメント毎に分割する。

（Ｓ３３００３）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、分割で生成した複数のセグメントの最初のものを変数ｉに代入する。

【０３７１】

（Ｓ３３００４）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、セグメントｉにプールボリュームの領域が割り当てられているかどうかを判断する。なお、判断にはＡＯＵアドレス変換情報３１０３０を用いる。かかる判断の結果、プールボリュームの領域が割り当てられる場合にはＳ３３００６を実行し、プールボリュームの領域が未割り当ての場合にはＳ３３００５を実行する。

（Ｓ３３００５）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、その仮想化ストレージ装置１０００Ｌ，１０００Ｒ内のキャッシュメモリ上に当該セグメント向けのキャッシュ領域を確保し、確保したキャッシュ領域をゼロで初期化し、ホスト１１００へゼロデータを転送する。

【０３７２】

（Ｓ３３００６）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、割り当てられたプールボリュームの領域に保存されたデータを転送する。なお、当該プールボリュームの領域が既にキャッシュ領域に存在する場合（ステージング済みの場合）には、かかるデータをそのキャッシュ領域から転送し、キャッシュ領域に存在しない場合はステージング後に、当該データの転送を行う。

【０３７３】

（Ｓ３３００８）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、続きのセグメントがあるかどうかを判断し、ある場合はＳ３３００９を実行し、ない場合はＳ３３０１０を実行する。

（Ｓ３３００９）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、変数ｉに次のセグメントを代入し、再びＳ３３００４を実行する。

（Ｓ３３０１０）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、リード完了の応答をホスト１１００に転送し、完了する。

【０３７４】

なお、処理の単純化のために、仮想化ストレージ装置１０００Ｌはプールボリュームのある決められた領域に対して予め定められた値（ゼロ）を保存しておき、当該領域に保存されたデータをＡＯＵボリューム２９０１０Ｒ，２９０１０Ｌの未割り当て領域に対するリードで転送してもよい。

【０３７５】

＜４．３．ＡＯＵ向けデステージング処理＞

図３４はＡＯＵ向けＩ／Ｏ処理プログラム３１０１０が実行するデステージング処理の処理内容を示すフローチャートである。以下、かかるデステージング処理について、当該フローチャートを参照しながら説明する。

【０３７６】

（Ｓ３４００１）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、キャッシュアルゴリズムによってデステージ対象とするキャッシュメモリ上のデータを決定する。なお、キャッシュアルゴリズムはＬＲＵ（Less Recently Used）アルゴリズムを用いてダーティデータを対象として決定する方法が一般的であるが、これ以外のアルゴリズムを用いて決定してもよい。

【０３７７】

（Ｓ３４００２）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、デステージ対象のデータが共有ストレージ装置（すなわち、ストレージ装置１５００Ｌ）が有するボリュームに対応するものかどうかを判断し、対応する場合はＳ３４００３を実行し、対応しない場合はＳ３４００４を実行する。

【０３７８】

（Ｓ３４００３）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、デステージング処理を実行し、その後、この一連の処理を終了する。なお、デステージング処理は他の実施の形態と同様に行われる。

（Ｓ３４００４）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、デステージ対象のデータが格納されたボリュームのボリューム属性を判断し、当該ボリュームがコピー元ボリュームである場合にはＳ３４００５を実行し、当該ボリュームがコピー先ボリュームの場合にはＳ３４００７を実行し、それ以外の場合はＳ３４００３を実行する。

【０３７９】

（Ｓ３４００５）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、デステージング処理を実行する。

（Ｓ３４００６）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、デステージが終了したデータのＲＣデステージ許可指示を副系の仮想化ストレージ装置１０００Ｒへ転送し、処理を終了する。

【０３８０】

（Ｓ３４００７）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、ＲＣデステージ許可フラグがＯＮかどうかを確認し、ＯＦＦの場合はＳ３４００１を再び実行し、別なデステージ対象のデータを選択し直す。なお、ＲＣデステージ許可フラグは、同期リモートコピーによってキャッシュメモリ上にライトデータが保存又は更新された時点ではＯＦＦが設定され、Ｓ３４００６で送信された指示を受信するとＯＮが設定される。

【０３８１】

（Ｓ３４００８）ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０は、デステージング処理を実行し、処理を終了する。

【０３８２】

本アルゴリズムによって、以下のキャッシュ制御が実現される。

【０３８３】

（Ａ）共有ストレージ装置向けでなく、正系と副系の仮想化ストレージ装置１０００Ｌ，１０００Ｒでデステージを連携する必要のないキャッシュデータは両系独立にデステージを行う。

（Ｂ）正系の仮想化ストレージ装置１０００Ｌでのデステージ処理後に送信されるメッセージによって、副系の仮想化ストレージ装置１０００Ｒのキャッシュデータのデステージが行われる。

【０３８４】

なお、ステージング処理は、第１～第１４の実施の形態と同様に行なわれる。

【０３８５】

＜４．３．１．ＲＣデステージ許可指示＞

ＲＣデステージ許可指示の転送は、非同期に指示を送信してもよい。ただし、正系及び副系の仮想化ストレージ装置１０００Ｌ，１０００Ｒはリモートコピーを契機として、ＲＣデステージフラグに未反映の当該指示を無効化してもよい。

【０３８６】

＜４．４．プールの空き領域監視＞
ＡＯＵ管理プログラム３１０２０は、定期的に各プールの空き領域を監視し、ユーザーが設定したスレッショルド値を下回った場合は、管理ホスト１２００へメッセージを送信する。これによって、容量不足に伴うホスト１１００からのライトリクエストの失敗を回避することができる。さらに、ＡＯＵ管理プログラム３１０２０は、空き領域の監視を共有のストレージ装置１５００Ｌと共有でないストレージ装置とで分けて管理し、容量不足の際に転送するメッセージを使い分けても良い。
【０３８７】
＜５．正系の仮想化ストレージ装置障害時の切り替え＞
正系の仮想化ストレージ装置１０００Ｌが障害などで機能を停止した場合は、他の実施の形態と同様の処理を行うことでホスト１１００は引きつづきアプリケーションを動作させることができる。
【０３８８】
一方で、ホスト１１００は、コピー元ボリュームに対するライトリクエストが容量不足で失敗したことを契機として副系の仮想化ストレージ装置１０００ＲにＩ／Ｏリクエスト先を切り替える場合もある。副系の仮想化ストレージ装置１０００Ｒが有するプール容量が正系よりも多い場合は、当該切り替えによって、ホスト１１００においてＩ／Ｏリクエストを発行しているアプリケーション２０１０（図３０）の処理を継続することができるからである。
【０３８９】
なお、この場合はリクエスト先の切り替えによってリモートコピーの向きは反転するが、リモートコピーは停止する。なぜならば、旧正系の仮想化ストレージ装置１０００Ｌはライトリクエスト時のプール容量不足で当該リクエストが失敗しているので、同期リモートコピーによって新正系（旧副系）の仮想化ストレージ装置１０００Ｒに対してライトデータを書き込もうとしても失敗するからである。
【０３９０】
ただし、旧正系の仮想化ストレージ装置１０００Ｌに対するリクエスト（特にリード）は継続可能であるため、本障害は仮想化ストレージ装置１０００Ｌ，１０００Ｒ間の通信路障害と見分けがつかず、ホスト１１００が旧正系の仮想化ストレージ１０００Ｌの古いデータをリードする可能性がある。
【０３９１】
こうした状況を回避するため、リモートコピー失敗の理由がプール容量不足である場合には、ホスト１１００からの旧正系の仮想化ストレージ装置１０００Ｌへのリードリクエスト発行を抑制してもよい。あるいは、リモートコピー失敗の理由が絞れない間は、ホスト１１００からの副系の仮想化ストレージ装置１０００Ｒまたは１０００Ｌに対するリードを抑制し、通信路障害であることが判明した時点でかかる抑制を解除するようにしてもよい。
【０３９２】
以上の処理によって、本実施の形態によるストレージシステムがサービス継続性の高いＡＯＵ機能を持ったストレージサービスを提供することができる。また、ＡＯＵ機能は、Ｉ／Ｏリクエスト毎にＡＯＵアドレス変換情報３１０３０Ｌ，３１０３０Ｒを参照・変更する必要があり、通常のストレージＩ／Ｏよりもコントローラーの負荷が高い。したがって、ホスト１１００が必要とするボリュームの一部（又は半分）については片方の仮想化ストレージ装置１０００Ｌ，１０００Ｒが正系としてリードとライトを担当し、残りのボリュームについてはもう片方の仮想化ストレージ装置１０００Ｒ，１０００Ｌが正系としてリードとライトを担当するようにしてもよい。このような構成を採用することで、ストレージシステムの可用性を維持しつつ、仮想化ストレージ装置１０００Ｌ，１０００Ｒとの間でのＡＯＵ機能のコントローラー負荷の平準化を実現できる。
【０３９３】
＜６．プールボリューム領域の割り当てとデータ移行について＞

これまでに述べたとおり、本実施の形態では仮想化ストレージ装置１０００Ｌ，１０００Ｒ内部のボリュームとストレージ装置１５００Ｌのボリュームの両方をプールボリュームとすることができる。そのため、アクセス頻度の高いデータが格納される又は格納されたセグメントに対して仮想化ストレージ装置１０００Ｌ，１０００Ｒ内部のボリュームを割り当てることによって、アクセス性能の向上が図れるほかに、仮想化ストレージ装置１０００Ｌ，１０００Ｒとストレージ装置１５００Ｌとの間の通信ネットワークのボトルネック化を回避することもできる。

【０３９４】

しかし、ＡＯＵでは最初のライトリクエストによってセグメントにプールボリュームの領域を割り当てるため、仮想化ストレージ装置１０００Ｌ，１０００Ｒ単体でアクセス頻度を考慮した割り当てをすることは難しい。こうした課題を解決する方法として以下の方法が考えられる。

【０３９５】

＜６．１．ＡＯＵボリュームに属性を付加する方法＞

ＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒを作成する時点でアクセス頻度に関する属性を与え、ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０がセグメントにプールボリュームの領域の割り当てを行う際に、そのセグメントに書き込まれるデータのアクセス頻度がある程度分かっているときには、かかるアクセス頻度属性を参照して、アクセス頻度の高いデータが格納されるセグメントについてはその仮想化ストレージ装置１０００Ｌ，１０００Ｒ内部のボリュームを割り当て、アクセス頻度の低いデータ（例えばバックアップデータ）が格納されるセグメントについてはストレージ装置１５００Ｌ内のボリュームの領域を割り当てる。

【０３９６】

＜６．２．プールボリューム領域のデータ移行＞

ＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒに対するアクセス頻度をセグメント単位（又は複数セグメント単位）で測定し、アクセス頻度の高いセグメントに格納されているデータは仮想化ストレージ装置１０００Ｌ，１０００Ｒ内部のプールボリュームの領域に移動させる。この場合、データの移行に伴って、ＡＯＵボリューム２９０１０Ｌ，２９０１０Ｒにおける当該データの移行が行なわれたセグメントの対応先をストレージ装置１５００Ｌ内のボリューム内のセグメントから、仮想化ストレージ装置１０００Ｌ，１０００Ｒにおけるデータの移行先のセグメントに変更する必要があるが、ＡＯＵ機能では元々仮想化ストレージ装置１０００Ｌ，１０００Ｒ内においてアドレス変換を行っているため、ホスト１１００に対して透過的にデータ移行を行うことができる。

【０３９７】

なお、本実施の形態でこのようなデータ移行を行う場合、対象となるセグメントのデータは、正系と副系の両方の仮想ストレージ装置１０００Ｌ，１０００Ｒ内部のプールボリュームに保存されることが望ましい。しかし、他に効果がある場合（以下に列挙した）は片方のセグメントだけ仮想化ストレージ装置１０００Ｌ，１０００Ｒ内部のプールボリュームの領域が割り当てられた形態にデータ移行を行うことも考えられる。

【０３９８】

（例１）どちらかの仮想化ストレージ装置１０００Ｌ，１０００Ｒが先に内部のプールボリュームを使い果たし、共有のストレージ装置１５００Ｌしかない場合。

（例２）コピー元のＡＯＵボリューム２９０１０Ｌに対するリードリクエストの負荷が大きく、正系の仮想化ストレージ装置１０００Ｌとストレージ装置１５００Ｌの間のネットワーク性能を圧迫する場合。

【０３９９】

こうした場合、正系の仮想化ストレージ装置１０００Ｌは、ストレージ装置１５００Ｌ内部のプールボリュームの領域から自身のプールボリュームの領域にセグメントのデータをコピーし、コピー先の領域を用いてＡＯＵボリューム２９０１０Ｌを提供する。一方の副系の仮想化ストレージ装置１０００Ｒはコピー元のストレージ装置１５００Ｌのプール

ボリュームの領域を用いてＡＯＵボリューム２９０１０Ｒを提供することもできる。この場合、ストレージ装置１５００Ｌのプールボリュームの領域に対するライトデータの反映は副系の仮想化ストレージ装置１０００Ｒが行っても良い。

【０４００】

さらに、リードもライトも含めたアクセス性能向上のためのセグメントのデータ移行の中間状態として、前述の正系の仮想化ストレージ装置１０００Ｌだけ内部のプールボリュームの領域を用い、副系の仮想化ストレージ装置１０００Ｒがストレージ装置１５００Ｌのプールボリュームを用いる構成を採用してもよい。

【０４０１】

＜７．本実施の形態のバリエーション＞

＜７．１．ステージングやデステージング処理でアドレス変換を実施＞

これまで述べてきた本実施の形態では、リードリクエスト処理やライトリクエスト処理でアドレス変換を行っている。本方法は、ライトリクエストの受付の時点で、プールボリュームの容量不足を契機とした失敗応答を返すことができる反面、リクエスト毎にアドレス変換を行うため、性能上の課題がある。こうした課題を解決する方法としてステージングやデステージング処理でアドレス変換を行う方法が考えられる。ただし、この方法ではデステージングの時点でセグメントに対してプールボリュームの領域の割り当てを行うため、ＨＤＤ１０３０の二重閉塞等が原因のボリューム閉塞時と類似するデータ消失が発生する。そのため、後者の方式では空き容量に余裕が少なくなってきた時点からリクエストの処理を遅らせるか停止する等の処理を行っても良い。

【０４０２】

なお、具体的な処理はこれまで図３２及び図３３について説明してきた処理内容を以下に示すとおりに変更すればよい。

（ライトとデステージング）図３２のＳ３２１００の割り当て処理をデステージング処理のＳ３４００１の後に移動する。

（リードとステージング）図３３のＳ３３００４～Ｓ３３００６で行っているアドレス変換を伴った割り当て有無の判断と未割り当て時のゼロデータの転送とを、ステージングにて行う。

【０４０３】

さらに、両者の利点を併せ持つために、ＡＯＵ向けＩ／Ｏ処理プログラム３１０１０が、プールボリュームの空き容量がスレッショルド値以上の場合にはステージング／デステージング処理で変換を行い、かかる空き容量がスレッショルド値以下になった場合にはＩ／Ｏリクエスト処理で変換を行ってもよい。

【０４０４】

＜７．２．Ｄｅ‐Ｄｕｐｌｉｃａｔｉｏｎ＞

ＡＯＵ管理プログラム３１０１０は、Ｉ／Ｏリクエストとは独立にＤｅ‐ｄｕｐｌｉｃａｔｉｏｎと呼ばれる以下の処理を行っても良い。

【０４０５】

（Ｓｔｅｐ１）ＡＯＵ管理プログラム３１０１０は、各プールボリュームの領域のデータをスキャンし、重複するセグメントを探す。

（Ｓｔｅｐ２）ＡＯＵ管理プログラム３１０１０は、プールボリューム領域同士で保存するデータが重複している事を検知した場合、いずれか一つの領域だけを残し他の領域は空き領域として開放する。そして、ＡＯＵアドレス変換情報３１０３０における開放した領域に対応したセグメントの「プールボリューム領域識別子」は一つだけ残した領域に更新し、「ＣＯＷフラグ」を「ＯＮ」にする。

【０４０６】

ここで、重複検知の方法としては、プールボリュームの領域毎のハッシュ値を計算後に、領域毎に、そのハッシュ値を他の領域のハッシュ値と順次比較し、同じ値の場合はさらに実際のデータを比較する２段階方式を採用してもよい。さらに、ハッシュ値の計算とデータの比較は負荷の高い処理であるため、副系の仮想化ストレージ装置１０００Ｒにて処

理を行うことで負荷分散を行うことも考えられる。

【図面の簡単な説明】

【０４０７】

【図１】図１は、第１の実施の形態にかかる情報システムのハードウェア構成の一例を示すブロック図である。

【図２】図２は、第１の実施の形態の概要を示す第１の概念図である。

【図３】図３は、第１の実施の形態の概要を示す第２の概念図である。

【図４】図４は、第１の実施の形態の概要を示す第３の概念図である。

【図５】図５は、ホスト上のソフトウェア構成を表した概念図である。

【図６】図６は、仮想化ストレージ装置及びストレージ装置上のソフトウェア構成を表したブロック図である。

【図７】図７は、リモートコピーのペア状態とペア状態の遷移を表した概念図である。

【図８】図８は、Ｉ／Ｏパスマネージャーが管理するデバイス関係テーブルを示す概念図である。

【図９】図９は、Ｉ／Ｏパスマネージャーが初期化処理を行うときのフローを示したフローチャートである。

【図１０】図１０は、Ｉ／Ｏパスマネージャーがライト処理を行うときのフローを示したフローチャートである。

【図１１】図１１は、Ｉ／Ｏパスマネージャーがリード処理を行うときのフローを示したフローチャートである。

【図１２】図１２は、第２の実施の形態の概要を示す概念図である。

【図１３】図１３は、第３の実施の形態の概要を示す概念図である。

【図１４】図１４は、第４の実施の形態の概要を示す概念図である。

【図１５】図１５は、第５の実施の形態の概要を示す概念図である。

【図１６】図１６は、第６の実施の形態の概要を示す概念図である。

【図１７】図１７は、第７の実施の形態の概要を示す概念図である。

【図１８】図１８は、第８の実施の形態の概要を示す概念図である。

【図１９】図１９は、第９の実施の形態の概要を示す概念図である。

【図２０】図２０は、第１０の実施の形態の概要を示す概念図である。

【図２１】図２１は、第１１の実施の形態の概要を示す概念図である。

【図２２】図２２は、第１２の実施の形態の概要を示す概念図である。

【図２３】図２３は、第１３の実施の形態の概要を示す概念図である。

【図２４】図２４は、第１４の実施の形態の概要を示す概念図である。

【図２５】図２５は、Ｉ／Ｏパスマネージャーがライト処理を行うときの別なフローを示したフローチャートである。

【図２６】図２６は、Ｉ／Ｏパスマネージャーがリード処理を行うときの別なフローを示したフローチャートである。

【図２７】図２７は、Ｉ／Ｏパスマネージャーが図２５に記したライト処理を行うときに、ストレージ装置にて行うライトリクエストに応じたペア操作を示したフローチャートである。

【図２８】図２８は、第１５の実施の形態の概要を示す概念図である。

【図２９】図２９は、第１６の実施の形態の概要を示す概念図である。

【図３０】図３０は、第１６の実施の形態の概要を示す概念図である。

【図３１】図３１は、本実施の形態における仮想化ストレージ装置及びストレージ装置上のソフトウェア構成を表したブロック図である。

【図３２】図３２は、仮想化ストレージ装置がライト処理を行うときのフローを示したフローチャートである。

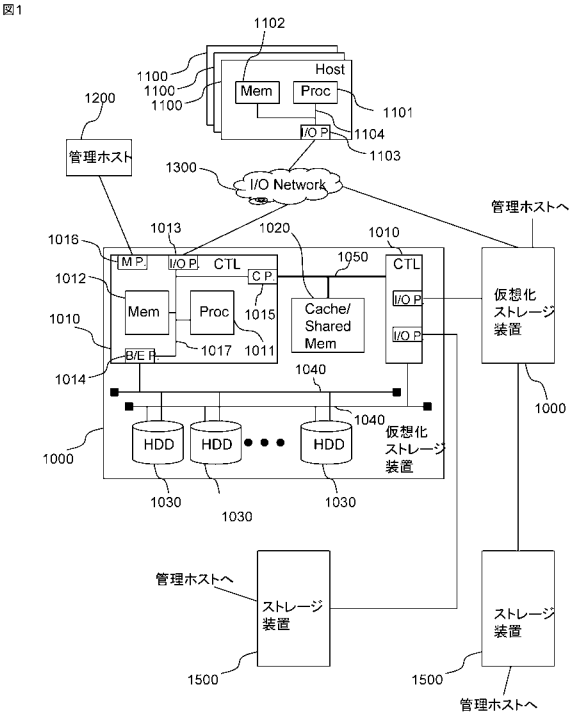
【図３３】図３３は、仮想化ストレージ装置がリード処理を行うときのフローを示したフローチャートである。

【図３４】図３４は、ＡＯＵ向けでステージング処理のフローを示したフローチャートで

ある。
【図３５】ＡＯＵアドレス変換情報の具体的内容の説明に供する概念図である。
【図３６】ＡＯＵプール管理情報の具体的内容の説明に供する概念図である。
【符号の説明】
【０４０８】
１０００，１０００Ｌ，１０００Ｒ……仮想化ストレージ装置、１０１０……コントローラー、１０１１，１１０１……プロセッサ、１０２０，１０２０Ｌ，１０２０Ｒ……キャッシュメモリ、１０３０……ＨＤＤ、１１００，１３０１０，１４０００……ホスト、１５００，１５００Ｌ，１５００Ｒ，１５０００，１５０００Ｌ，１５０００Ｒ……ストレージ装置，２８００Ｌ，２８００Ｒ……仮想化スイッチ、３５００ＬＢ，３５００ＲＢ，５０４０，５０５０……ボリューム、２０１０，１４００２……アプリケーションプログラム、５０００……Ｉ／Ｏパスマネージャー、５０１０……ＨＢＡデバイスドライバー、５０２０……ファイルシステム、１３００１，１３００２……ストレージサブシステム、１５００２Ａ，１５００２Ｂ……コマンドデバイス、１５０１０Ａ，１５０１０Ｂ……差分ビットマップ、１６０００……外部ストレージ装置。

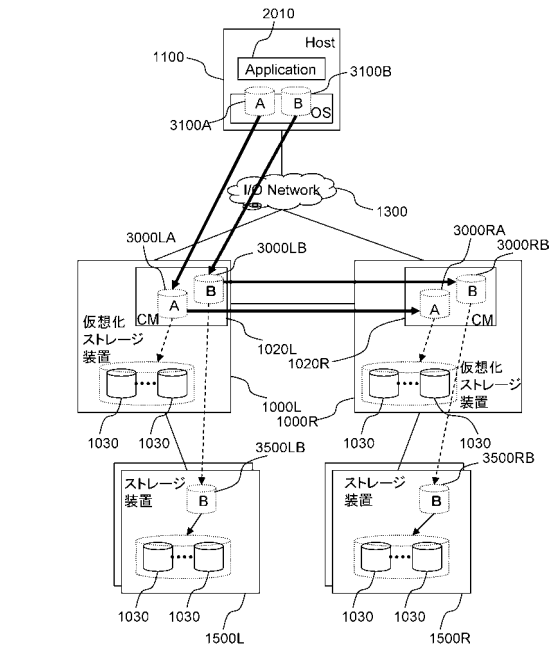
【図１】

図1



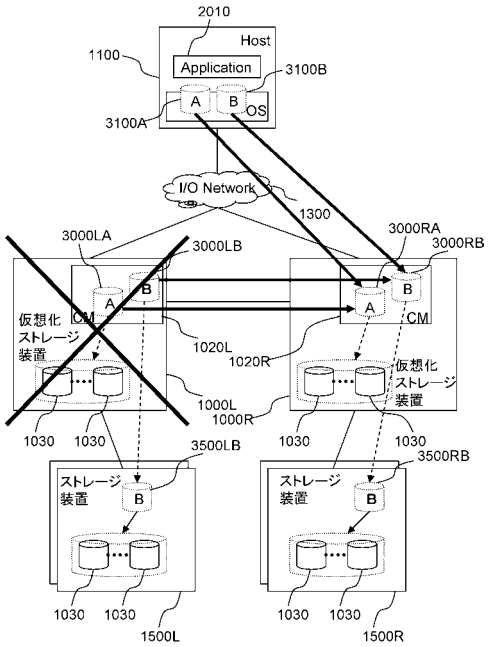
【図２】

図2



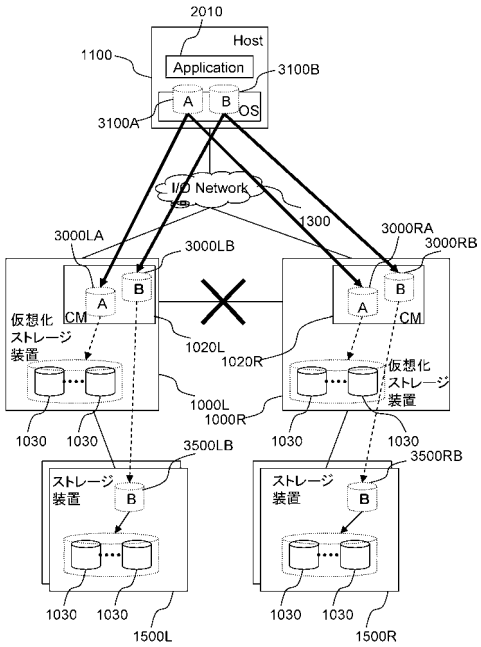
【図３】

図3



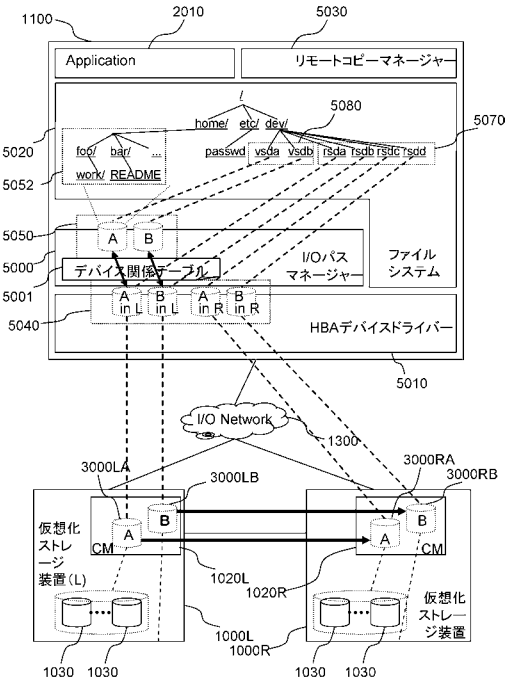
【図４】

図4



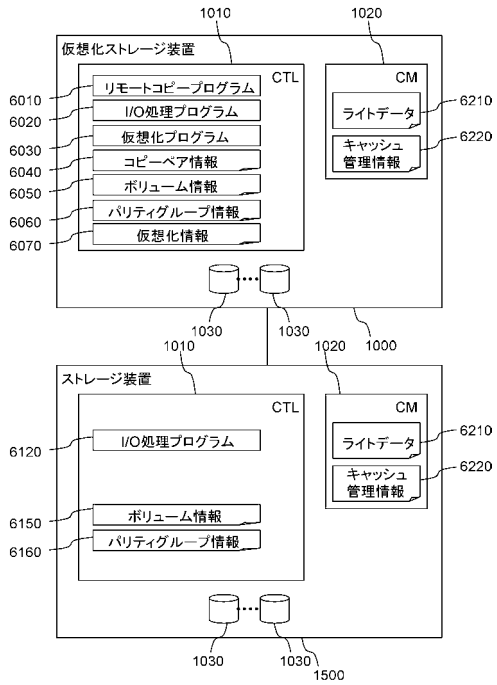
【図５】

図5



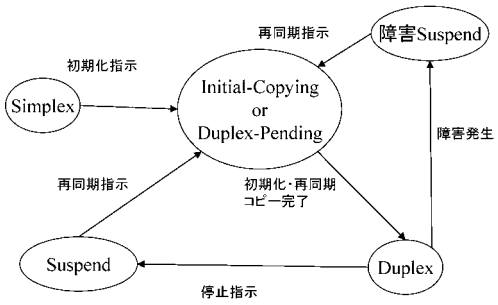
【図６】

図6



【図７】

図7



【図８】

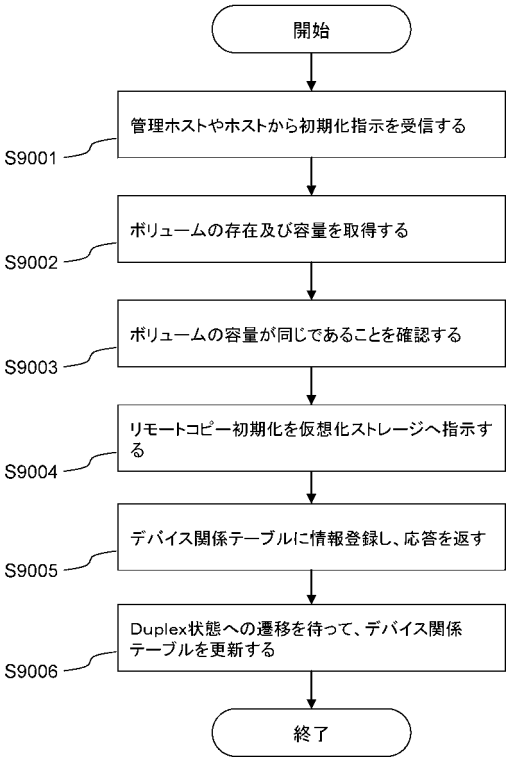
図8

5001

ホスト内で仮想的なボリュームの識別子	関係ボリューム識別子リスト	正系ボリューム識別子	障害状態	ペア状態
vsda	rsda, rsdb	rsda	通常状態	Duplex
vsdb	rsdc, rsdd	rsdc	副系準備中	Duplex-Pending
..	..		..	..

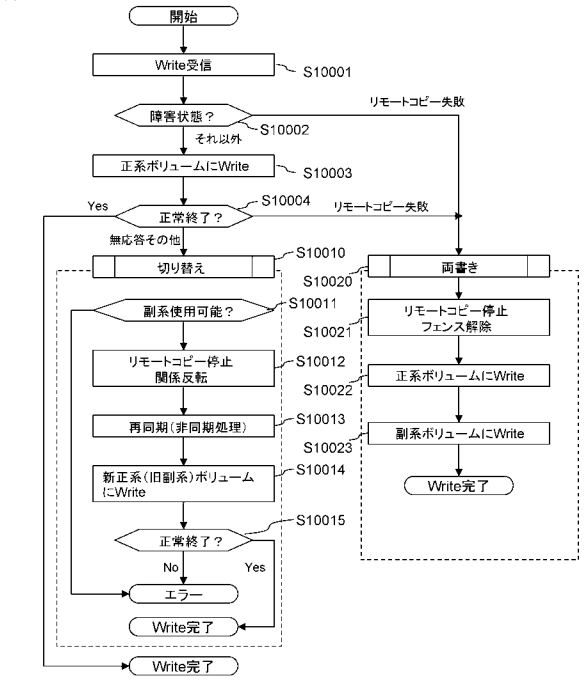
【図９】

図9



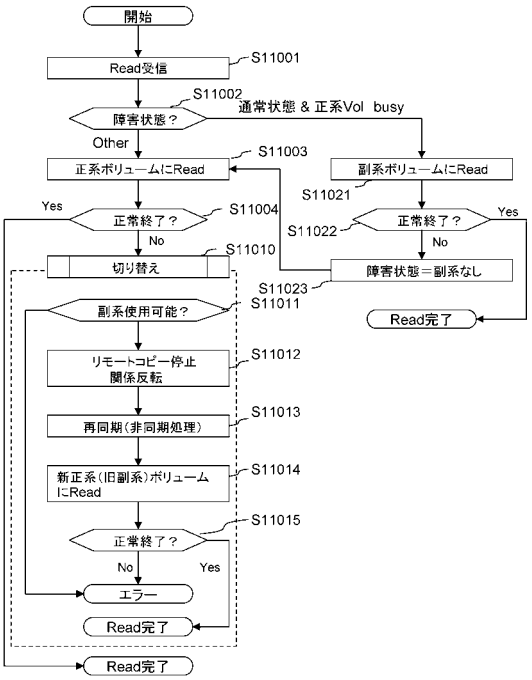
【図１０】

図10

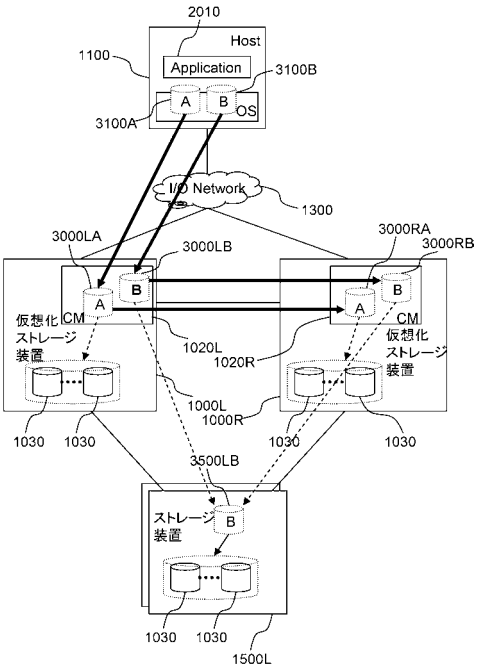


【図１１】

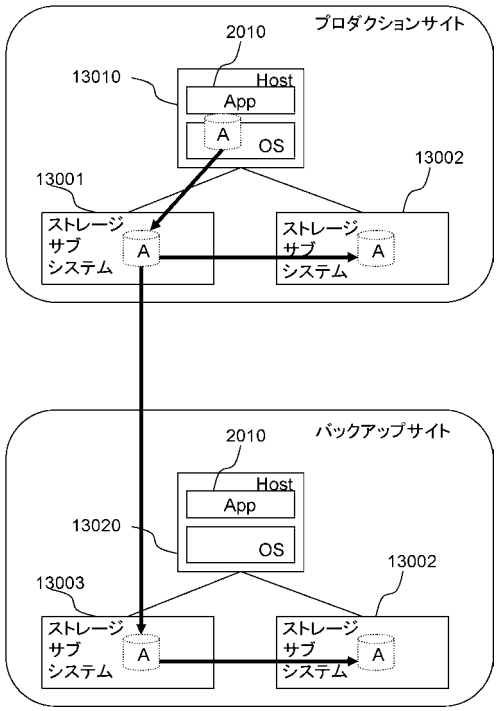
図11



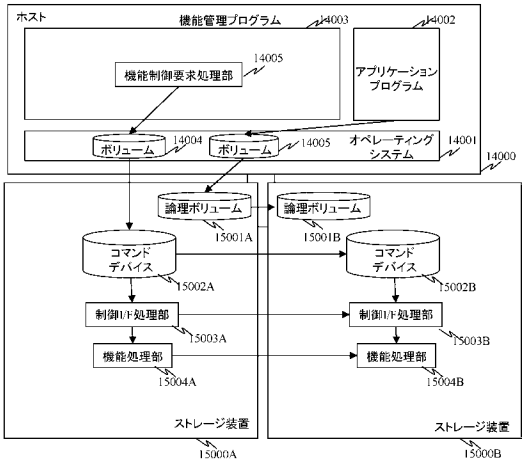
【図１２】
図12



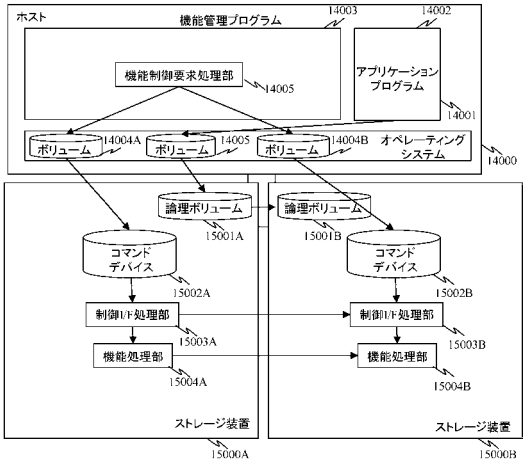
【図１３】
図13



【図１４】
図14

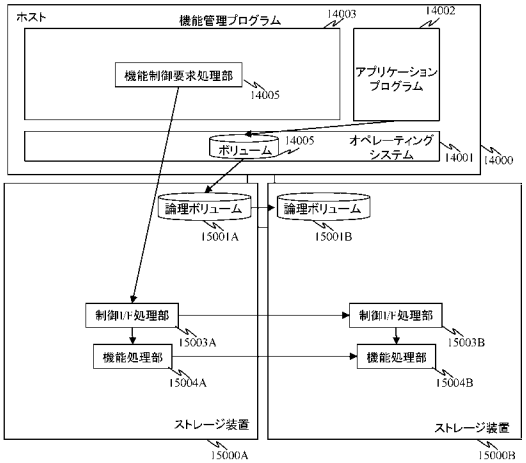


【図１５】
図15



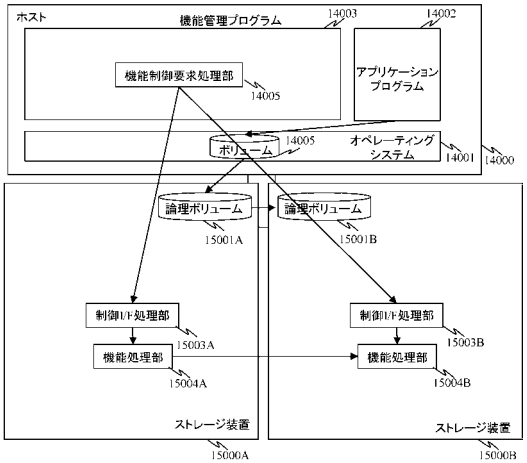
【図１６】

図16



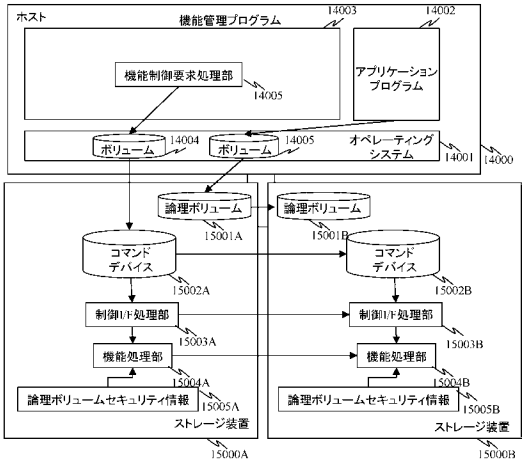
【図１７】

図17



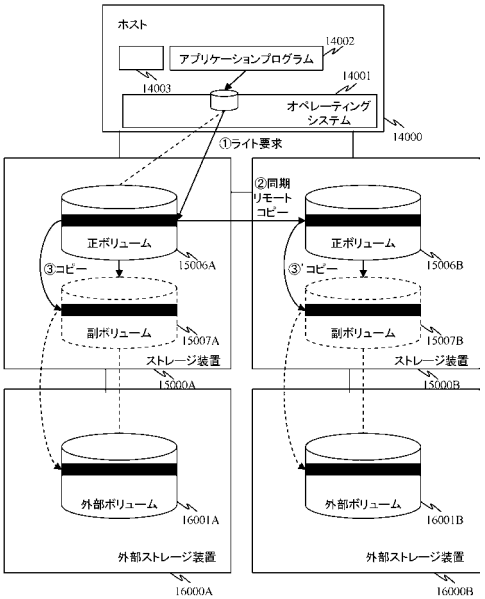
【図１８】

図18

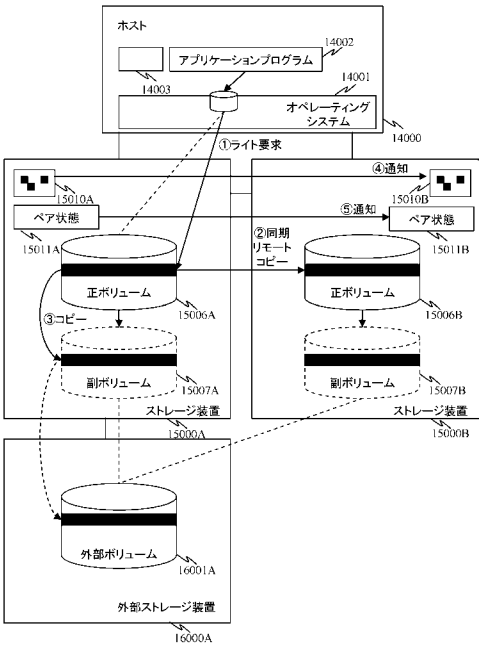


【図１９】

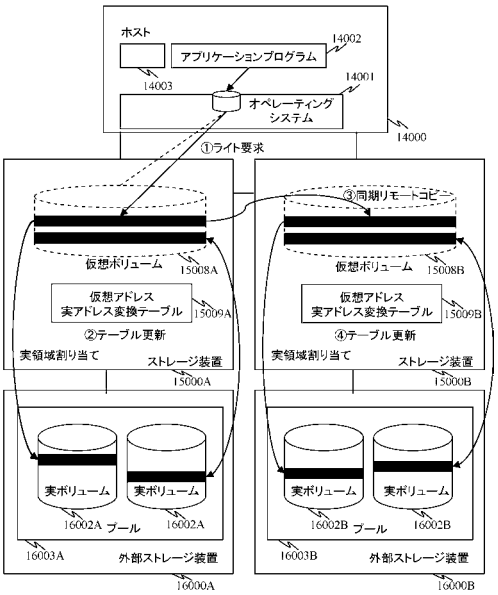
図19



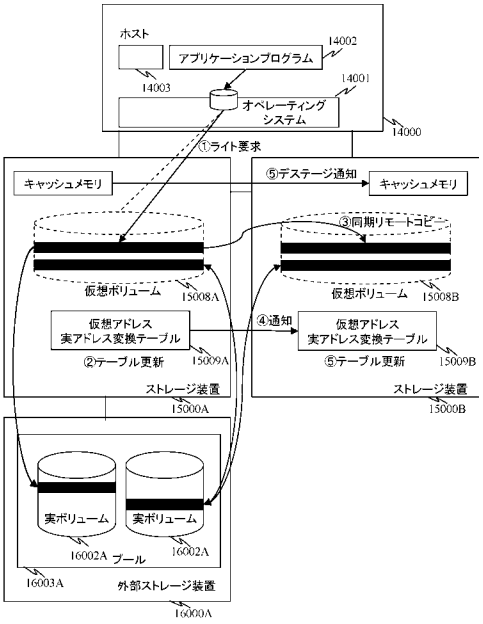
【図２０】
図20



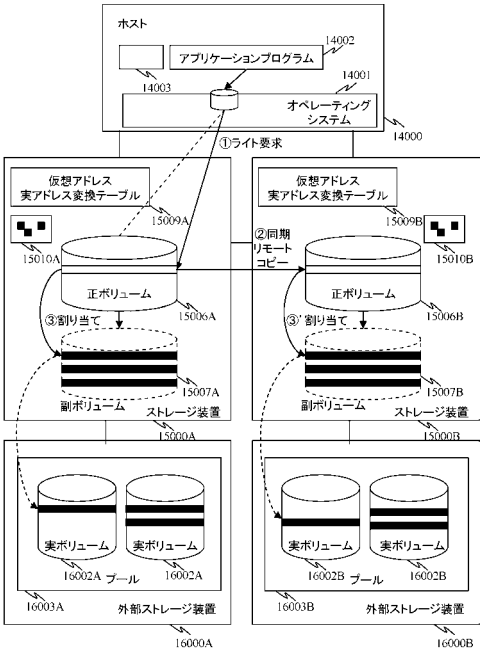
【図２１】
図21



【図２２】
図22

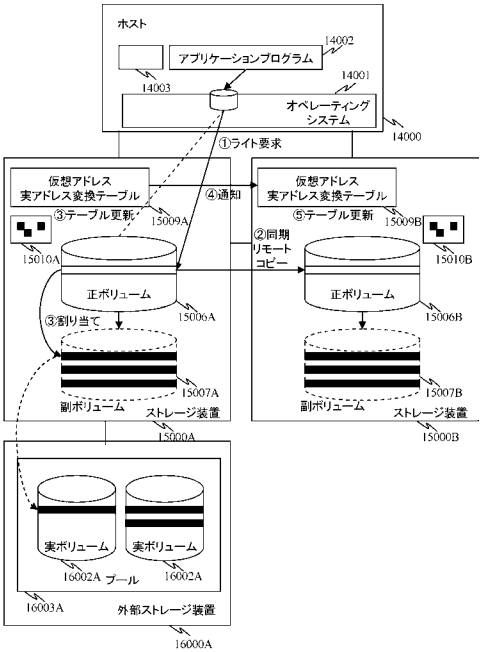


【図２３】
図23



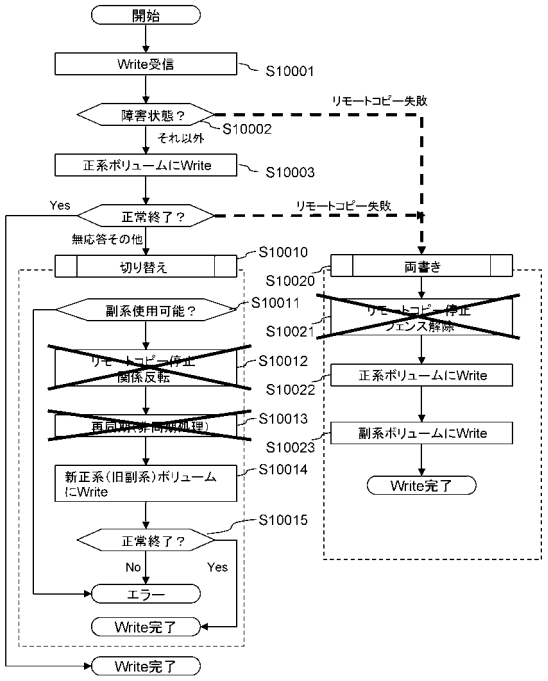
【図２４】

図24



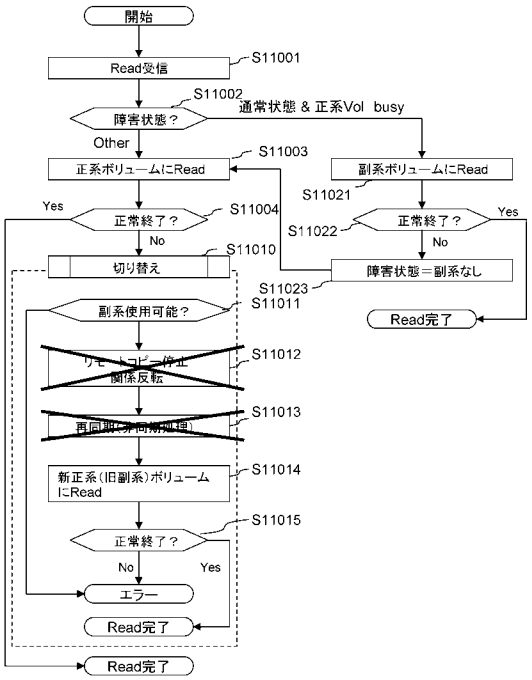
【図２５】

図25



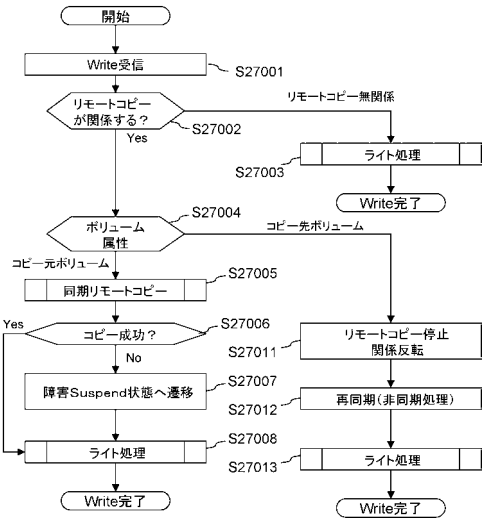
【図２６】

図26

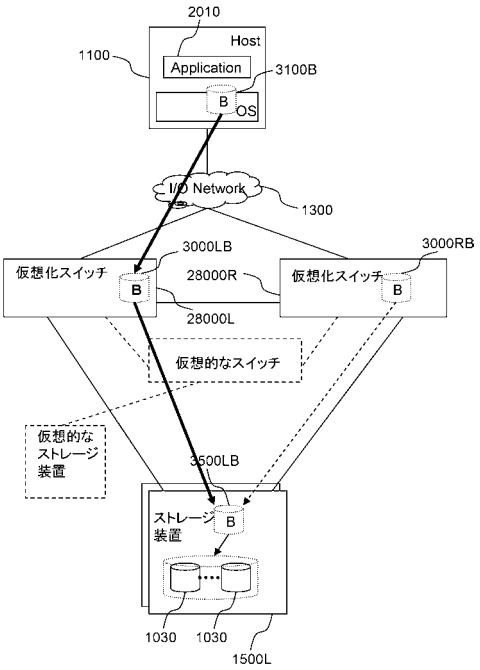


【図２７】

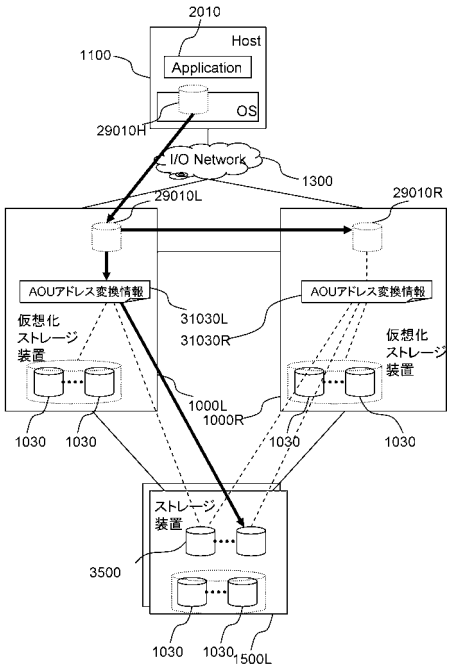
図27



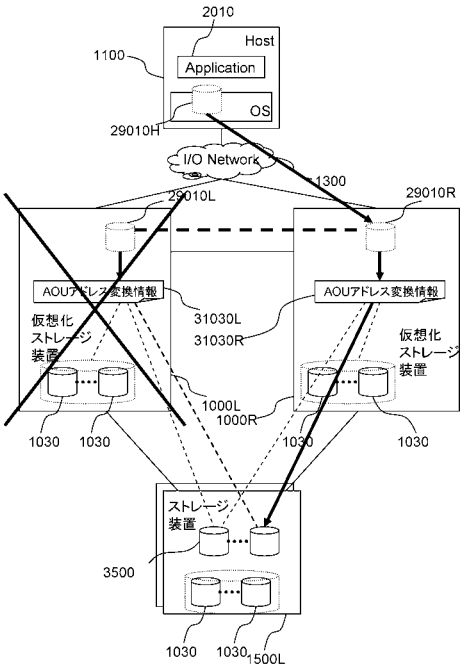
【図２８】
図28



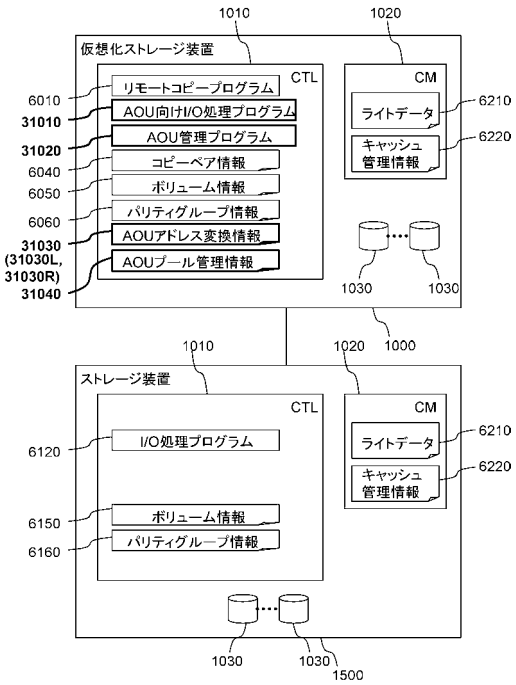
【図２９】
図29



【図３０】
図30

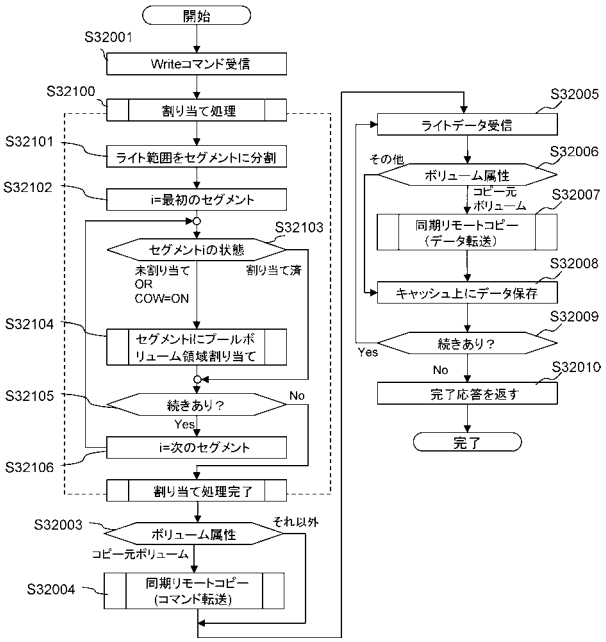


【図３１】
図31



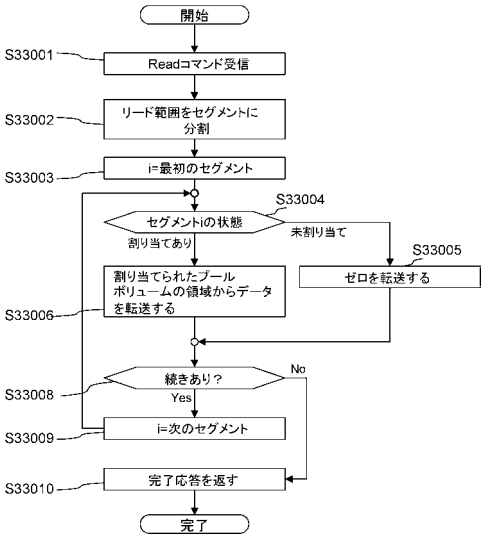
【図３２】

図32



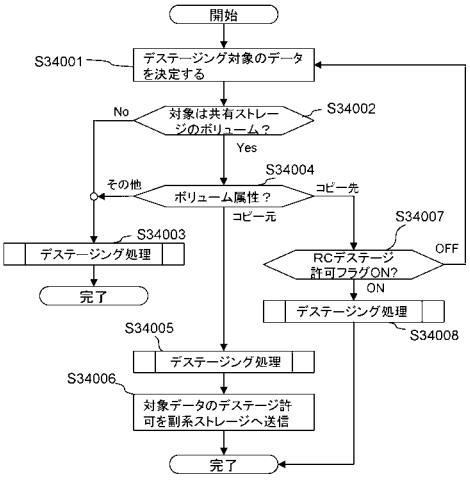
【図３３】

図33



【図３４】

図34



【図３５】

図35

31030

プールID	AOUボリューム識別子	対AOUボリューム識別子	アドレス空間	COWフラグ	引継ぎ領域	プールボリューム領域識別子
1	A-LDEV1	装置C A-LDEV5	0-1999	No	No	LDEV1(0-1999)
			2000-3999	No	Yes	Port=xxx, LUN=yyy (8000-9999)
			…	…	…	…
			18000-19999	No	No	NULL
1	A-LDEV2	装置C A-LDEV6	0-1999	No	No	LDEV2(0-1999)
			2000-3999	Yes	Yes	Port=xxx, LUN=yyy (10000-11999)
			…	…	…	…
			48000-49999	No	No	NULL

【図３６】
図36

31040

プールID	属性	値
1	セグメントサイズ	2000block
	プールボリューム リスト	LDEV1, LDEV2, Port=xxx&LUN=yyy, Port=zzz&LUN=aaa,....
	空き領域リスト	LDEV1(2000-40000), LDEV2(2000-10000), Port=xxx&LUN=yyy,...
	空き容量	3000000block
	容量警告	10000block
	プール対の識別子	装置L プールID=3
2	セグメントサイズ	1000block
	プールボリューム リスト	LDEV8, Port=aaa&LUN=bbb,....
	空き領域リスト	LDEV8(0000-40000), Port=aaa&LUN=bbb,...
	空き容量	100000block
	容量警告	10000block
	プール対の識別子	装置L プールID=3

フロントページの続き

(72)発明者　江口　賢哲
神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所システム開発研究所内
(72)発明者　渡辺　恭男
神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所システム開発研究所内
(72)発明者　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
(72)発明者　山本　康友
神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所システム開発研究所内

審査官　古河　雅輝

(56)参考文献　特開２００６－０４８６７６（ＪＰ，Ａ）
米国特許出願公開第２００４／０２６０７３６（ＵＳ，Ａ１）
国際公開第２００５／０７１５４４（WO，Ａ１）
特開２００５－２６７２１６（ＪＰ，Ａ）
特開２００６－０２４２１５（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　　３／０６－　３／０８
Ｇ０６Ｆ　１２／００
Ｇ０６Ｆ　１３／１０－１３／１４