

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4963892号  
(P4963892)

(45) 発行日 平成24年6月27日 (2012.6.27)

(24) 登録日 平成24年4月6日 (2012.4.6)

(51) Int. Cl. F I  
**G06F 13/10 (2006.01)** G O 6 F 13/10 3 4 O A  
**G06F 3/06 (2006.01)** G O 6 F 3/06 3 O 1 Z

請求項の数 13 (全 30 頁)

(21) 出願番号	特願2006-210464 (P2006-210464)	(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
(22) 出願日	平成18年8月2日 (2006.8.2)	(74) 代理人	110000279 特許業務法人ウィルフォート国際特許事務所
(65) 公開番号	特開2008-40571 (P2008-40571A)	(72) 発明者	里山 愛 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
(43) 公開日	平成20年2月21日 (2008.2.21)	(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
審査請求日	平成21年1月27日 (2009.1.27)		

最終頁に続く

(54) 【発明の名称】 仮想ストレージシステムの構成要素となることが可能なストレージシステムの制御装置

(57) 【特許請求の範囲】

【請求項1】

上位装置から送信された要求に従って第一のストレージシステムで実行される処理の対象である管理対象を管理し、該第一のストレージシステムの動作を制御し、該第一のストレージシステムを含む複数のストレージシステムを論理的に一つの記憶資源に仮想化する仮想ストレージシステムを構成するストレージシステムの外部の装置であって、

複数の管理対象の各々についての第一と第二の管理識別子を記録した識別子対応情報を記憶する識別子対応記憶部と、

第一の管理識別子を指定した要求を前記上位装置から受信する受信部と、

前記受信した要求で指定されている第一の管理識別子に対応した第二の管理識別子を前記識別子対応情報から特定する対応特定部と、

前記特定された第二の管理識別子を前記第一のストレージシステムの処理部に出力する出力部と、

を備え、

前記第一及び第二の管理識別子は、同種の管理識別子が階層化されたものであり、

前記第二の管理識別子は、前記第一のストレージシステムでの処理のために使用される種類の管理識別子であり、前記出力された第二の管理識別子から、前記第一のストレージシステムの処理部が、処理の対象を識別し、

前記第一の管理識別子は、前記上位装置で認識される種類の管理識別子であり、

前記第一のストレージシステムが、前記仮想ストレージシステムの構成要素ではない場

10

20

合、前記第一のストレージシステムが有する1つの管理対象について、前記第一の管理識別子と前記第二の管理識別子の値は同一であり、

前記第一のストレージシステムが前記仮想ストレージシステムの構成要素とされる際に、前記第一のストレージシステムが有する管理対象の第一の管理識別子が、該仮想ストレージシステムにおける他のストレージシステムが有する管理対象の第一の管理識別子と重複しない値に変更され、

前記識別子対応記憶部は、前記第一のストレージシステムの前記変更後の第一と第二の管理識別子の識別子対応情報と、前記他のストレージシステムの第一と第二の管理識別子の識別子対応情報とを統合する、

制御装置。

10

【請求項2】

前記第一と第二の管理識別子は、前記第一のストレージシステムに存在する複数の論理的な記憶デバイスの各々を識別するための識別子、及び/又は、複数の特定のリソースの各々を識別するための識別子である、

請求項1記載の制御装置。

【請求項3】

前記仮想ストレージシステムを構成するストレージシステムの台数の変更に応じて前記第一の管理識別子を変更する変更部、

を更に備える請求項1又は2記載の制御装置。

【請求項4】

20

前記台数の変更とは、既存の仮想ストレージシステムに前記第一のストレージシステムを増設することであり、

前記変更部は、前記第一のストレージシステムの複数の管理対象にそれぞれ対応した複数の第一の管理識別子を、それぞれ、前記既存の仮想ストレージシステムにおける各他のストレージシステムでの複数の第一の管理識別子と重複しない値に変更する、

請求項3記載の制御装置。

【請求項5】

前記台数の変更とは、前記第一のストレージシステムを含んだ既存の第一の仮想ストレージシステムに、第二の仮想ストレージシステムをマージすることで、第三の仮想ストレージシステムを構築することであり、

30

前記変更部は、前記第一のストレージシステムの複数の管理対象にそれぞれ対応した複数の第一の管理識別子を、それぞれ、前記第三の仮想ストレージシステムの各他のストレージシステムにおける複数の第一の管理識別子と重複しない値に変更する、

請求項3記載の制御装置。

【請求項6】

変更後の前記複数の第一の管理識別子を前記上位装置に通知する変更報告部、  
を更に備える請求項3乃至5のうちのいずれか1項に記載の制御装置。

【請求項7】

前記変更部は、前記仮想ストレージシステムから前記第一のストレージシステムが減設されたことにより該第一のストレージシステムが単独となった場合には、前記複数の第一の管理識別子を、それぞれ、前記複数の第二の管理識別子と同じ値に変更する、

40

請求項3乃至6のうちのいずれか1項に記載の制御装置。

【請求項8】

前記複数の管理対象の各々は、論理的な記憶デバイスであり、

前記仮想化ストレージシステムを構成する前記第一のストレージシステムと第二のストレージシステムは、前記仮想化ストレージシステムの構成要素ではない共通の外部のストレージシステムに接続されており、且つ、前記第一のストレージシステムにおける実体の無い第一の仮想的な論理記憶デバイスと、前記第二のストレージシステムにおける実体の無い第二の仮想的な論理記憶デバイスとが、前記外部ストレージシステムにおける実体の有る外部論理記憶デバイスに接続されており、

50

前記変更部は、前記第一の仮想的な論理記憶デバイスの第一の管理識別子を、前記第二の仮想的な論理記憶デバイスの第一の管理識別子と同じ値にする、  
請求項 3 乃至 7 のうちのいずれか 1 項に記載の制御装置。

【請求項 9】

上位装置から送信された要求に従って第一のストレージシステムで実行される処理の対象である管理対象を管理し、該第一のストレージシステムの動作を制御する制御装置と、  
前記第一のストレージシステムを含んだ複数のストレージシステムを論理的に一つの記憶資源に仮想化してなる仮想ストレージシステムと、

前記仮想ストレージシステムを前記一つの記憶資源として提供するための仮想化部とを備え、

10

前記制御装置は、前記仮想ストレージシステムを構成するストレージシステムの外部の装置であり、

複数の管理対象の各々についての第一と第二の管理識別子を記録した識別子対応情報を記憶する識別子対応記憶部と、

第一の管理識別子を指定した要求を前記上位装置から受信する受信部と、

前記受信した要求で指定されている第一の管理識別子に対応した第二の管理識別子を前記識別子対応情報から特定する対応特定部と、

前記特定された第二の管理識別子を前記第一のストレージシステムの処理部に出力する出力部と

を備え、

20

前記第二の管理識別子は、前記第一のストレージシステムでの処理のために使用される種類の管理識別子であり、前記出力された第二の管理識別子から、前記第一のストレージシステムの処理部が、処理の対象を識別し、

前記第一の管理識別子は、前記上位装置で認識される種類の管理識別子であり、

前記第一及び第二の管理識別子は、同種の管理識別子が階層化されたものであり、

前記第一のストレージシステムが、前記仮想ストレージシステムの構成要素ではない場合、前記第一のストレージシステムが有する 1 つの管理対象について、前記第一の管理識別子と前記第二の管理識別子の値は同一であり、

前記第一のストレージシステムが、前記仮想ストレージシステムの構成要素とされる際に、前記第一のストレージシステムが有する管理対象の第一の管理識別子が、該仮想ストレージシステムが有する他のストレージシステムでの管理対象の第一の管理識別子と重複しない値に変更され、

30

前記識別子対応記憶部は、前記第一のストレージシステムの前記変更後の第一と第二の管理識別子の識別子対応情報と、前記他のストレージシステムの第一と第二の管理識別子の識別子対応情報とを統合する、

計算機システム。

【請求項 10】

前記制御装置が、

前記仮想ストレージシステムを構成するストレージシステムの台数の変更に応じて前記第一の管理識別子を変更する変更部と、

40

変更後の前記複数の第一の管理識別子を前記仮想化部に通知する変更報告部とを更に備え、

前記仮想化部が、ホスト計算機が指定し得る複数の第三の管理識別子と前記複数の第一の管理識別子との対応を管理し、前記変更後の複数の第一の管理識別子を受信した場合、前記複数の第三の管理識別子にそれぞれ対応付けられている複数の第一の管理識別子を、前記受信した変更後の複数の第一の管理識別子に変更する、

請求項 9 記載の計算機システム。

【請求項 11】

前記上位装置が、前記ホスト計算機と前記仮想ストレージシステムとの間に介在するスイッチであり、

50

前記仮想化部が、前記スイッチに搭載されている、  
請求項 9 又は 10 記載の計算機システム。

【請求項 12】

前記上位装置が、前記ホスト計算機であり、  
前記仮想化部が、前記ホスト計算機に搭載されている、  
請求項 9 又は 10 記載の計算機システム。

【請求項 13】

請求項 1 乃至 8 のうちのいずれか 1 項に記載の制御装置を備えたストレージシステム。

【発明の詳細な説明】

【技術分野】

10

【0001】

本発明は、複数のストレージシステムを論理的に一つの記憶資源として仮想化した仮想ストレージシステムに関する。

【背景技術】

【0002】

例えば、ストレージ分野では、ネットワーク上に分散する複数のコンピューティングリソースをあたかも一つのコンピューティングリソースであるかのように運用し、処理性能や耐障害性の向上を図る「ストレージグリッド（広域広帯域共有ファイルシステム）」が注目されている。ストレージグリッドの運用方法として、複数のストレージシステムを論理的に一つの記憶資源として仮想化する方法が知られている。

20

【0003】

また、ボリュームを仮想化する技術として、第一のストレージシステムの外部に存在する第二のストレージシステムの論理ユニットを第一のストレージシステム内で管理する技術が知られている（例えば特開 2005 - 107645 号公報）。

【0004】

【特許文献 1】特開 2005 - 107645 号公報

【発明の開示】

【発明が解決しようとする課題】

【0005】

ストレージグリッドを構築する方法として、例えば、単体のストレージシステムとして稼動していたものを複数台集めて構築することが考えられる。具体例として、ストレージシステム A と B の 2 台をストレージグリッドとする場合について考える。

30

【0006】

単純に、それらのストレージシステム A と B の境界をなくして仮想的に 1 台のシステムとした場合、論理的な記憶デバイス（論理ボリュームとも呼ばれる、以下、「LDEV」と略記する）の番号（番号に限らず他種の識別子でも良い）は、ストレージシステム A、B それぞれにつけられているため、同じ LDEV 番号の LDEV が複数存在してしまうという問題が発生し得る。これは、LDEV 番号だけではなく、他種の管理識別子でも生じ得る。例えば、コピー系の制御情報、例としては、ペア番号やグループ番号（複数のペアから成るグループの番号）等においても同じ問題が発生し得る。このため、個別に稼動していたストレージシステムをスムーズに仮想ストレージシステムの構成要素とすることは難しい。

40

【0007】

さらに、ストレージシステム A、B は、それぞれ、少なくとも 1 つの管理装置（以下、SVP（Service Processor））を備えている場合がある。SVP は、例えば、管理対象のストレージシステムへの情報入力の指示を受けたり、情報を出力したり、ストレージシステムの管理情報を保持したりする。ストレージシステム A、B をストレージグリッドとして仮想ストレージシステムを構築し、ストレージシステム A の SVP で管理する情報を、ストレージシステム B が取得したい場合、ストレージシステム B は、ストレージシステム A の SVP への問い合わせが必要となる。このため、ストレージシステム B は、ストレ

50

ージシステムAのSVPから管理情報を取得するたびに、そのSVPに問い合わせをすることになるので、処理が煩雑になり、負荷も増加する。

【0008】

従って、本発明の目的は、個別に稼動していたストレージシステムをスムーズに仮想ストレージシステムの構成要素とすることができるようにすることにある。

【0009】

本発明の別の目的は、仮想ストレージシステムにおける各ストレージシステムが各他のストレージシステムの管理装置に管理情報を問い合わせなくても済む新規な管理方法を提供することにある。

【0010】

本発明の他の目的は、後の説明から明らかになるであろう。

【課題を解決するための手段】

【0011】

本発明の第一の側面に従う制御装置は、上位装置から送信された要求に従って第一のストレージシステムで実行される処理の対象である管理対象を管理し、該第一のストレージシステムの動作を制御する装置である。この制御装置は、複数の管理対象の各々についての第一と第二の管理識別子を記録した識別子対応情報を記憶する識別子対応記憶部と、第一の管理識別子を指定した要求を前記上位装置から受信する受信部と、前記受信した要求で指定されている第一の管理識別子に対応した第二の管理識別子を前記識別子対応情報から特定する対応特定部と、前記特定された第二の管理識別子を前記第一のストレージシステムの処理部に出力する出力部とを備える。前記第二の管理識別子は、前記第一のストレージシステムでの処理のために使用される種類の管理識別子である。前記出力された第二の管理識別子から、前記第一のストレージシステムの処理部が、処理の対象を識別することができる。前記第一の管理識別子は、前記上位装置で認識される種類の管理識別子であり、前記第一のストレージシステムが、複数のストレージシステムを論理的に一つの記憶資源に仮想化してなる仮想ストレージシステムの構成要素とされる際に、該仮想ストレージシステムにおける他のストレージシステムでの管理対象の第一の管理識別子と重複しない値に変更される。

【0012】

この制御装置は、第一のストレージシステムに搭載されていても良いし、該第一のストレージシステムとは別の装置であってもよい。

【0013】

この制御装置の第一の実施態様では、前記第一と第二の管理識別子は、前記第一のストレージシステムに存在する複数の論理的な記憶デバイスの各々を識別するための識別子、及び/又は、複数の特定のリソースの各々を識別するための識別子である。特定のリソースとは、例えば、論理的な記憶デバイスから別の論理的な記憶デバイスへのコピーのためのペアとすることができる。このペアは、単一のペアであっても良いし、複数のペアで構成されたペアグループの構成要素であるペアであってもよい。

【0014】

この制御装置の第二の実施態様では、前記仮想ストレージシステムを構成するストレージシステムの台数の変更に応じて前記第一の管理識別子を変更する変更部が更に備えられる。

【0015】

この制御装置の第三の実施態様では、前記第二の実施態様において、前記台数の変更とは、既存の仮想ストレージシステムに前記第一のストレージシステムを増設することである。前記変更部は、前記第一のストレージシステムの複数の管理対象にそれぞれ対応した複数の第一の管理識別子を、それぞれ、前記既存の仮想ストレージシステムにおける各他のストレージシステムでの複数の第一の管理識別子と重複しない値に変更する。

【0016】

この制御装置の第四の実施態様では、前記第二の実施態様において、前記台数の変更と

10

20

30

40

50

は、前記第一のストレージシステムを含んだ既存の第一の仮想ストレージシステムに、第二の仮想ストレージシステムをマージすることで、第三の仮想ストレージシステムを構築することである。前記変更部は、前記第一のストレージシステムの複数の管理対象にそれぞれ対応した複数の第一の管理識別子を、それぞれ、前記第三の仮想ストレージシステムの各他のストレージシステムにおける複数の第一の管理識別子と重複しない値に変更する。

【0017】

この制御装置の第五の実施態様では、前記第二の実施態様において、変更後の前記複数の第一の管理識別子を前記上位装置に通知する変更報告部が更に備えられる。

【0018】

この制御装置の第六の実施態様では、前記第二の実施態様において、前記変更部は、前記仮想ストレージシステムから前記第一のストレージシステムが減設されたことにより該第一のストレージシステムが単独となった場合には、前記複数の第一の管理識別子を、それぞれ、前記複数の第二の管理識別子と同じ値に変更する。

【0019】

この制御装置の第七の実施態様では、前記第二の実施態様において、前記複数の管理対象の各々は、論理的な記憶デバイスである。前記仮想化ストレージシステムを構成する前記第一のストレージシステムと第二のストレージシステムは、前記仮想化ストレージシステムを構成しない外部のストレージシステムに接続されており、且つ、前記第一のストレージシステムにおける実体の無い第一の仮想的な論理記憶デバイスと、前記第二のストレージシステムにおける実体の無い第二の仮想的な論理記憶デバイスとが、前記外部ストレージシステムにおける実体の有る外部論理記憶デバイスに接続されている。この場合、前記変更部は、前記第一の仮想的な論理記憶デバイスの第一の管理識別子を、前記第二の仮想的な論理記憶デバイスの第一の管理識別子と同じ値にする。

【0020】

本発明の第二の側面に従う管理方法は、複数のストレージシステムを論理的に一つの記憶資源を仮想化してなる仮想ストレージシステムでの管理方法である。この管理方法では、前記複数のストレージシステムのそれぞれが保持する複数の管理計算機をネットワークで接続し、前記複数の管理計算機でそれぞれ管理される複数の管理情報を一つのマスタの管理計算機で一括して管理する。

【0021】

この管理方法の第一の実施態様では、前記複数の管理計算機とは別の管理計算機を前記ネットワークに接続する。前記別の管理計算機を前記マスタの管理計算機とし、前記複数の管理計算機をスレイブの管理計算機とする。前記マスタの管理計算機が、前記複数のスレイブの管理計算機のそれぞれの識別子を記憶する。前記複数のスレイブの管理計算機の各々が、前記マスタの管理計算機の識別子を記憶する。前記複数のスレイブの管理計算機の各々が、前記マスタの管理計算機に、管理情報を送信する。前記マスタの管理計算機が、前記複数のスレイブの管理計算機からそれぞれ複数の管理情報を受信し、前記受信した複数の管理情報を、前記マスタの管理計算機の記憶資源に格納する。

【0022】

この管理方法の第二の実施態様では、前記複数の管理計算機のうちの一つを前記マスタの管理計算機とし、他の管理計算機をスレイブの管理計算機とする。前記マスタの管理計算機が、スレイブの管理計算機の識別子を記憶する。前記スレイブの管理計算機が、前記マスタの管理計算機の識別子を記憶する。前記スレイブの管理計算機が、前記マスタの管理計算機に、管理情報を送信する。前記マスタの管理計算機が、前記スレイブの管理計算機から管理情報を受信し、前記受信した管理情報を、前記マスタの管理計算機の記憶資源に格納する。前記マスタの管理計算機は、前記スレイブの管理計算機の管理情報の他に、自分が接続されているストレージシステムを管理するための管理情報も、前記マスタの管理計算機の前記記憶資源に格納する。

【0023】

この管理方法の第三の実施態様では、前記マスタの管理計算機を多重化し、多重化されたマスタの管理計算機のうちの一つをアクティブとし、他をパッシブとする。前記アクティブのマスタの管理計算機が停止した場合に、いずれかのパッシブのマスタの管理計算機がアクティブとなって、前記停止したマスタの管理計算機に代わって、マスタの管理計算機として稼動する。

【0024】

この管理方法の第四の実施態様では、前記マスタの管理計算機は、自分以外の管理計算機に、自分の識別子であるマスタ識別子をブロードキャストする。前記マスタ識別子を受信した管理計算機が、自分の記憶資源に、前記受信したマスタ識別子を格納し、それに  
10 応答して、自分の識別子であるスレイブ識別子を、前記マスタ識別子の送信元である前記マスタの管理計算機に送信する。前記マスタの管理計算機は、自分以外の管理計算機から前記スレイブ識別子を受信し、該受信したスレイブ識別子を、自分の記憶資源に格納する。

【0025】

本発明の第三の側面に従う計算機システムは、上位装置から送信された要求に従って第一のストレージシステムで実行される処理の対象である管理対象を管理し、該第一のストレージシステムの動作を制御する制御装置と、前記第一のストレージシステムを含んだ複数のストレージシステムを論理的に一つの記憶資源に仮想化してなる仮想ストレージシステムと、前記仮想ストレージシステムを前記一つの記憶資源として提供するための仮想化部とを備える。前記制御装置は、複数の管理対象の各々についての第一と第二の管理識別子を記録した識別子対応情報を記憶する識別子対応記憶部と、第一の管理識別子を指定した要求を前記上位装置から受信する受信部と、前記受信した要求で指定されている第一の管理識別子に対応した第二の管理識別子を前記識別子対応情報から特定する対応特定部と、前記特定された第二の管理識別子を前記第一のストレージシステムの処理部に出力する出力部とを備える。前記第二の管理識別子は、前記第一のストレージシステムでの処理のために使用される種類の管理識別子であり、前記出力された第二の管理識別子から、前記第一のストレージシステムの処理部が、処理の対象を識別する。前記第一の管理識別子は、前記上位装置で認識される種類の管理識別子であり、前記第一のストレージシステムが、複数のストレージシステムを論理的に一つの記憶資源に仮想化してなる仮想ストレージシステムの構成要素とされる際に、該仮想ストレージシステムにおける他のストレージシステムでの管理対象の第一の管理識別子と重複しない値に変更される。  
20  
30

【0026】

この計算機システムの第一の実施態様では、前記制御装置が、前記仮想ストレージシステムを構成するストレージシステムの台数の変更に応じて前記第一の管理識別子を変更する変更部と、変更後の前記複数の第一の管理識別子を前記仮想化部に通知する変更報告部とを更に備える。前記仮想化部が、ホスト計算機が指定し得る複数の第三の管理識別子と前記複数の第一の管理識別子との対応を管理し、前記変更後の複数の第一の管理識別子を受信した場合、前記複数の第三の管理識別子にそれぞれ対応付けられている複数の第一の管理識別子を、前記受信した変更後の複数の第一の管理識別子に変更する。

【0027】

この計算機システムの第二の実施態様では、前記上位装置が、前記ホスト計算機と前記仮想ストレージシステムとの間に介在するスイッチである。前記仮想化部が、前記スイッチに搭載されている。  
40

【0028】

この計算機システムの第三の実施態様では、前記上位装置が、前記ホスト計算機であり、前記仮想化部が、前記ホスト計算機に搭載されている。

【0029】

上述した制御装置或いは計算機システムの記憶部は、例えば、メモリなどの記憶資源により構築することができる。また、制御装置或いは計算機システムでの他の各部は、ハードウェア、コンピュータプログラム又はそれらの組み合わせ（例えば一部をコンピュータ  
50

プログラムにより実現し残りをハードウェアで実現すること)により構築することができる。コンピュータプログラムは、所定のプロセッサに読み込まれて実行される。また、コンピュータプログラムがプロセッサに読み込まれて行われる情報処理の際、適宜に、メモリ等のハードウェア資源上に存在する記憶域が使用されてもよい。また、コンピュータプログラムは、CD-ROM等の記録媒体から計算機にインストールされてもよいし、通信ネットワークを介して計算機にダウンロードされてもよい。

【発明の効果】

【0030】

本発明の制御装置や計算機システムによれば、個別に稼動していたストレージシステムをスムーズに仮想ストレージシステムの構成要素とすることができる。

10

【0031】

本発明の統合管理方法によれば、仮想ストレージシステムにおける各ストレージシステムが各他のストレージシステムの管理装置に管理情報を問い合わせなくても済む新規な管理方法を提供することができる。

【発明を実施するための最良の形態】

【0032】

以下、本発明の一実施形態を説明する。

【0033】

図1は、本発明の一実施形態に係るシステム全体の構成例を示す。以下の説明では、同種の要素を区別しないで説明する場合には、親番号のみ(例えば100)を用いて説明し、同種の要素を区別して説明する場合には、親番号と子符号(例えば100A、100B)を用いて説明することにする。

20

【0034】

仮想ストレージシステム100Aは、複数のストレージシステム20A、20Bと、外部ストレージシステム70とを備える。同様に仮想ストレージシステム100Bは、複数のストレージシステム20C、20Dと、外部ストレージシステム70とを備える。仮想化ストレージシステム100Aと同様な構成である仮想化ストレージシステム100Bが同じSAN41に接続されていてもよい。

【0035】

各ホスト計算機51は、図示しないCPUや記憶資源(例えばメモリ)を備える。記憶資源に、コンピュータプログラムが記憶されており、CPUが、そのコンピュータプログラムを実行することができる。以下、コンピュータプログラムが主語になる場合は、実際にはそのコンピュータプログラムを実行するCPUによって処理が行われるものとする。

30

【0036】

それぞれのストレージシステム20A、20Bには、ホスト計算機51Aが接続されている。ホスト計算機51Aは、SAN(Storage Area Network)41を介して、それぞれのストレージシステム20A、20Bに接続している。ホスト計算機51Aは、交替パスソフトウェア61の機能により、複数のストレージシステム20A、20Bを論理的に一つの記憶資源(仮想ストレージシステム)100Aとして認識する。

【0037】

40

交替パスソフトウェア61を有しない他のホスト計算機51Bは、SAN42、仮想化装置52、及びSAN41を介して、それぞれのストレージシステム20A、20Bに接続している。ホスト計算機51Bは、仮想化装置52内の交替パスソフトウェア61の機能により、複数のストレージシステム20A、20Bを一つの仮想ストレージシステム100Aとして認識する。

【0038】

仮想化装置52は、複数のストレージシステム20A、20Bのそれぞれが提供する記憶領域を論理的な一つの記憶領域に仮想化する。仮想化装置52は、例えば、仮想化スイッチ、インテリジェントスイッチ、又は仮想化専用装置などである。尚、仮想ストレージシステム100、及び仮想化装置52を含んで構成されるシステムをストレージネットワ

50

ークシステムと称する。

【0039】

各ホスト計算機51は、アプリケーションプログラム60を有する。アプリケーションプログラム60は、例えば、データベース管理ソフトウェア、Webアプリケーションソフトウェア、ストリーミングアプリケーションソフトウェアなどの業務プログラムである。

【0040】

交替パスソフトウェア61は、ホスト計算機51Aと仮想ストレージシステム100A内の論理的な記憶デバイス(LDEV)との間の論理パスを管理し、交替パス制御を実行する。交替パス制御とは、複数の論理パスの中からどの論理パスを使用してLDEVにアクセスするのかを制御することをいう。交替パスソフトウェア61は、交替パス管理プログラム(図示せず)、ストレージシステム間交替パスプログラム(図示せず)、優先制御プログラム(図示せず)、及び、図2の同一デバイス対応テーブル62を有する。尚、これらは仮想ストレージシステム100A、100Bごとに管理している。以下の説明でも同様である。ここでは、仮想ストレージシステム100Aについて説明を行う。

10

【0041】

交替パス管理プログラムは、交替パスソフトウェア61全体の管理及び制御を行う。

【0042】

ストレージシステム間交替パスプログラムは、複数のストレージシステム20をそれぞれ経由して同一の外部LDEV701に接続する複数の論理パスを交替パスとして認識する。なお、外部LDEVとは、外部ストレージシステム70に備えられるLDEVであり、仮想ストレージシステム100内のストレージシステムのLDEVにマッピングされているLDEVである。

20

【0043】

優先制御プログラムは、同一のLDEVに接続する複数の論理パスのそれぞれの使用率に基づいて、複数の論理パスの中から優先的に使用する論理パスを選択する。

【0044】

同一デバイス対応テーブル62は、異なる論理パスが同一の論理デバイスに接続しているネットワーク環境下において、論理パスと論理デバイスとの対応関係を示す。具体的には、例えば、同一デバイス対応テーブル62には、図2に示すように、各論理パス毎に、パス番号、仮想化ストレージシステムにおけるストレージシステムの識別子、LUN及びグローバルLDEV番号が記録される。ここで、パス番号0、2及び3の論理パスのグローバルLDEV番号が同じになっている理由は、同じグローバルLDEV番号に対応する各ローカルLDEV番号が同一の外部LDEVに接続されているためである。つまり、それら複数の論理パスのうちどの論理パスでアクセスが行われても、最終的に、同一の外部LDEVへのアクセスが行われるのである(グローバルLDEV番号及びローカルLDEV番号については、後述する)。このような複数の論理パスの各々を、以下、「交替パス」と呼ぶことにする。なお、この同一デバイス対応テーブル62は、例えば、ストレージシステム20毎に用意され、複数のストレージシステムにより仮想ストレージシステムが構築された場合には、仮想ストレージシステム100毎に用意される。また、同一デバイス対応テーブル62では、交替パスのみならず、単一の論理パスも管理できる。すなわち、同一デバイス対応テーブル62には、アプリケーション60に認識されているLUNと、ストレージシステム20で管理されているグローバルLDEV番号との対応関係が登録される。

30

40

【0045】

ストレージシステム20は、例えば、制御部と、記憶部とに大別することができる。制御部は、例えば、CPU(Central Processing Unit)21、メモリ22、ディスクインターフェース制御部23、FC(Fiber Channel)インターフェース制御部25、26、及びLAN(Local Area Network)インターフェース制御部27、28を備える。記憶部は、例えば、ディスクユニット24である。

【0046】

50

C P U 2 1 は、メモリ 2 2 に格納されている各種プログラム又はモジュールを実行することにより、ストレージシステム 2 0 における各種制御処理を実行する。メモリ 2 2 は、内部記憶装置と称されるものであり、各種モジュール等を記憶する不揮発性メモリと、C P U 2 1 の演算処理結果を一時的に保持する揮発性メモリとを含む。

【 0 0 4 7 】

C P U 2 1 は、ディスクインターフェース制御部 2 3 を介してディスクユニット 2 4 に接続されている。ディスクインターフェース制御部 2 3 は、C P U 2 1 から出力される論理アドレスを L B A ( Logical Block Address ) に変換し、C P U 2 1 による論理デバイスへのアクセスを可能とする。

【 0 0 4 8 】

ディスクユニット 2 4 は、R A I D ( Redundant Arrays of Independent Inexpensive Disks ) 構成された複数のディスクドライブ 2 4 0 を有する。ディスクドライブ 2 4 0 は、F C ( Fibre Channel ) ディスクドライブ、S A T A ( Serial Advanced Technology Attachment ) ディスクドライブ、P A T A ( Parallel Advanced Technology Attachment ) ディスクドライブ、F A T A ( Fibre Attached Technology Adapted ) ディスクドライブ、S A S ( Serial Attached SCSI ) ディスクドライブ或いは S C S I ( Small Computer System Interface ) ディスクドライブ等の物理デバイスである。物理デバイスとは、実記憶領域を有する実デバイスである。なお、物理デバイスは、ディスクドライブ 2 4 0 に限らず、他種の記憶デバイス ( 例えばフラッシュメモリ ) が採用されても良い。

【 0 0 4 9 】

L D E V 2 4 2 は、実記憶領域を有しない仮想的なデバイスである。データを記憶する実体的な記憶領域は、外部ストレージシステム 7 0 内の論理デバイス 7 0 1 に存在する。即ち、ストレージシステム 2 0 は、外部ストレージシステム 7 0 の外部 L D E V 7 0 1 を自己の内部デバイスとして取り込み、これを論理ユニットとしてホスト計算機 5 1 に提供している。ホスト計算機 5 1 が U N I X ( 登録商標 ) 系のシステムである場合には、論理ユニットは、デバイスファイル ( Device File ) に対応付けられる。ホスト計算機 5 1 が W i n d o w s ( 登録商標 ) 系のシステムである場合には、論理ユニットは、ドライブレター ( ドライブ名 ) に対応付けられる。論理ユニットには、L U N ( Logical Unit Number ) がアサインされる。外部 L D E V 7 0 1 があたかもストレージシステム 2 0 の内部デバイスであるかのように仮想化する方法の詳細については、特開 2 0 0 5 - 1 0 7 6 4 5 号公報に開示されている。すなわち、本実施形態における外部接続は、例えば、特開 2 0 0 5 - 1 0 7 6 4 5 号公報 ( U S 出願番号 1 0 / 7 6 9 8 0 5 号、U S 出願番号 1 1 / 4 7 1 5 5 6 号 ) に開示の技術を援用することができる。

【 0 0 5 0 】

コマンドデバイス 2 4 3 は、ホスト計算機 5 1 とストレージシステム 2 0 との間でコマンドやステータスを受け渡すための専用の論理ユニットである。ホスト計算機 5 1 からストレージシステム 2 0 に送信されるコマンドは、コマンドデバイス 2 4 3 に書き込まれる。ストレージシステム 2 0 は、コマンドデバイス 2 4 3 に書き込まれたコマンドに対応する処理を実行し、その実行結果をステータスとしてコマンドデバイス 2 4 3 に書き込む。コマンドデバイス 2 4 3 を用いたコマンド管理の詳細については、後述する。

【 0 0 5 1 】

ディスクインターフェース制御部 2 3 は、複数のディスクドライブ 2 4 0 を R A I D 方式に規定される R A I D レベル ( 例えば、0 , 1 , 5 ) で制御することができる。R A I D 方式においては、複数のディスクドライブ 2 4 0 が一つの R A I D グループとして管理される。R A I D グループは、例えば、4 つのディスクドライブ 2 4 0 を一組としてグループ化することにより ( 3 D + 1 P ) 、或いは、8 つのディスクドライブ 2 4 0 を一組としてグループ化することにより ( 7 D + 1 P ) 、構成される。即ち、複数のディスクドライブ 2 4 0 のそれぞれが提供する記憶領域が集合して一つの R A I D グループが構成される。R A I D グループ上には、ホスト計算機 5 1 からのアクセス単位である複数の論理デバイス 2 4 1 が定義される。ホスト計算機 5 1 が認識する論理的な記憶領域である論理ユ

10

20

30

40

50

ニットには、一つ以上の論理デバイス 2 4 1 がマッピングされる。ホスト計算機 5 1 は、L U N と L B A とを指定することにより、論理デバイス 2 4 1 にアクセスできる。

【 0 0 5 2 】

尚、ストレージシステム 2 0 は、外部ストレージシステム 7 0 に接続することにより、外部ストレージシステム 7 0 の記憶領域をホスト計算機 5 1 に提供できるならば、必ずしもディスクユニット 2 4 とディスクインターフェース制御部 2 3 とを備えている必要はない。

【 0 0 5 3 】

仮想ストレージシステム 1 0 0 A 内のストレージシステム 2 0 A と 2 0 B の間はコマンドや制御情報及びデータを転送するための専用の通信線 4 4 がある。通信線 4 4 がない構成の場合、ホスト計算機 5 1 を経由して転送することができる。

10

【 0 0 5 4 】

また、1つの仮想ストレージシステム 1 0 0 は、外部ストレージシステム 7 0 は1台以上と接続しており、外部ストレージシステム 7 0 は、複数の仮想ストレージシステム 1 0 0 から接続されていてもよい。

【 0 0 5 5 】

F C インターフェース制御部 2 5 は、ストレージシステム 2 0 と外部ストレージシステム 7 0 との間の S A N 4 3 を通じてのコマンドやデータの転送を制御する。F C インターフェース制御部 2 6 は、ストレージシステム 2 0 とホスト計算機 5 1 との間の S A N 4 1 を通じてのコマンドやデータの転送を制御する。S A N 4 1 , 4 3 を介するデータ通信プロトコルとしては、例えば、ファイバチャネルプロトコルや i S C S I プロトコルなどを採用することができる。尚、S A N 4 3 は必須ではなく、F C インターフェース制御部 2 5 と外部ストレージシステム 7 0 とが光ファイバケーブルなどで直接接続されていてもよい。

20

【 0 0 5 6 】

L A N インターフェース制御部 2 7 は、管理ネットワーク 4 0 を介して管理サーバ 5 0 に接続する。管理ネットワーク 4 0 は、イーサネット（登録商標）ケーブルなどで構成される L A N である。管理ネットワーク 4 0 におけるデータ通信プロトコルは、例えば、T C P / I P である。管理サーバ 5 0 は、仮想ストレージシステム 1 0 0 内の L D E V の作成、ホスト計算機 5 1 への L D E V の割り当て、ホスト計算機 5 1 と仮想ストレージシステム 1 0 0 との間のアクセスパスの設定（L U N マスキングやゾーニングなど）などを管理することができる。

30

【 0 0 5 7 】

外部ストレージシステム 7 0 は、R A I D 構成された一つ以上のディスクドライブ 7 0 0 を有する。ディスクドライブ 7 0 0 は、F C (Fibre Channel) ディスクドライブ、S A T A (Serial Advanced Technology Attachment) ディスクドライブ、P A T A (Parallel Advanced Technology Attachment) ディスクドライブ、F A T A (Fibre Attached Technology Adapted) ディスクドライブ、S A S (Serial Attached SCSI) ディスクドライブ或いは S C S I (Small Computer System Interface) ディスクドライブ等のストレージデバイスである。複数のディスクドライブ 7 0 0 上には、外部 L D E V 7 0 1 が形成されている。なお、ディスクドライブ 7 0 0 に代えて又は加えて、他種の記憶デバイス（例えばフラッシュメモリ）があってもよい。

40

【 0 0 5 8 】

尚、外部ストレージシステム 7 0 は、ストレージシステム 2 0 とは異なる機種であってもよく、或いはストレージシステム 2 0 と同一の機種であってもよい。

【 0 0 5 9 】

S V P (サービスプロセッサ) 8 1 は、ストレージシステム 2 0 の構成情報（後述するデバイス情報テーブル 1 0 6、マッピングテーブル 1 0 8 を含む）の設定や、稼働情報の取得などを行うコンピュータ端末である。S V P 8 1 は、ストレージシステム 2 0 内の L A N インターフェース制御部 2 8 に接続される。各ストレージシステム 2 0 に接続する各

50

SVP81は、一つのマスタSVPに接続される。マスタSVPは、各SVP81から収集した情報（ストレージシステム20の構成情報、稼働情報など）を一元管理する。尚、複数のSVP81のうちの一つがマスタSVPとして機能してもよく、或いは各ストレージシステム20がマスタSVP82に直接接続されていてもよい。マスタSVPを備えた管理方法については、後に詳述する。

#### 【0060】

仮想ストレージシステム100は、ストレージシステム20間で構成情報を送受信することにより、各ストレージシステム20間におけるコマンド転送やボリュームマイグレーションを実行できる範囲を示している。各ストレージシステム20に外部ストレージシステム70が接続されている場合には、仮想ストレージシステム100の範囲は、複数のストレージシステム20と外部ストレージシステム70とを含む。各ストレージシステム20に外部ストレージシステム70が接続されていない場合には、仮想ストレージシステム100の範囲は、複数のストレージシステム20を含み、外部ストレージシステム70を含まない。

10

#### 【0061】

各ストレージシステム20は、例えば、RAID構成された複数のディスクドライブ240を備えるディスクアレイシステムとして構成することもできるし、或いは各ストレージシステム20自体がSCSIターゲットになる仮想化スイッチとして構成することもできる。LDEV番号などの階層化された管理識別子を管理しストレージシステムの動作を制御する装置を、「制御装置」と総称することができる。制御装置は、ストレージシステム20の中にあっても良いし、外にあっても良い。

20

#### 【0062】

管理ネットワーク40は、一つ以上のホスト計算機51に接続している。また、管理ネットワーク40には、一つ以上のクライアント計算機30が接続されている。各クライアント計算機30は、管理ネットワーク40を介してホスト計算機51に接続し、アプリケーションプログラム60を介して仮想ストレージシステム100にデータ入出力を要求できる。

#### 【0063】

図3は、ストレージシステム20のメモリ22内に格納されている各種プログラム及びテーブルを示す。

30

#### 【0064】

メモリ22には、例えば、オペレーティングシステム101、番号管理プログラム102、番号管理テーブル105、デバイス情報テーブル106及びマッピングテーブル108が格納される。

#### 【0065】

番号管理テーブル105は、グローバルLDEV番号とローカルLDEV番号との対応表である。デバイス情報テーブル106は、論理デバイスの属性を保持するテーブルである。マッピングテーブル108は、外部LDEV701のLDEV番号とローカルLDEV番号（それに代えて又は加えて、グローバルLDEV番号であっても良い）との対応関係を表すテーブルである。

40

#### 【0066】

図4a、図4bは番号管理テーブル105の構成例を示す。

#### 【0067】

番号管理テーブル105は、仮想ストレージシステム100の上位の装置（ホスト計算機51及び/又は仮想化装置52）に見せる番号である「グローバルLDEV番号」と、ストレージシステム20内部で処理するときの番号である「ローカルLDEV番号」との対応表である。例えば、ホスト計算機51Aからは、グローバルLDEV番号で要求を受け、番号管理プログラム102によって、そのグローバルLDEV番号に対応するローカルLDEV番号が本テーブル105から取得され、取得されたローカルLDEV番号を用いて、処理（例えば、LDEVに対する書き込み又は読出し（I/O）の処理）が行われ

50

る。交替パスソフトウェア 6 1 には、前述したように、グローバル L D E V 番号の情報のみ保持される（図 2 参照）。なお、図 4 a と図 4 b の違いについては、後述する。

【 0 0 6 8 】

また、ここでは、L D E V 番号についてのテーブル例を示すが、L D E V 番号に限らず、個々のストレージシステムで利用されていたが 1 つの仮想ストレージシステムとなったために重複してしまうおそれのある他種の管理識別子（例えば、コピー系のペア番号やグループ番号など）についても、同様のテーブルを保持することができる。特に、その他種の管理識別子としては、上位の装置（例えばホスト計算機）からの何らかの要求で指定される管理識別子を挙げるができる。

【 0 0 6 9 】

図 5 は、デバイス情報テーブル 1 0 6 の構成例を示す。

【 0 0 7 0 】

デバイス情報テーブル 1 0 6 は、各 L D E V の属性を管理するためのテーブルである。デバイス情報テーブル 1 0 6 には、例えば、各 L D E V 毎に、「デバイス番号」、「ステータスフラグ」、及び「外部デバイスフラグ」が記録される。

【 0 0 7 1 】

「デバイス番号」は、ストレージシステム 2 0 内で L D E V 2 4 1 , 2 4 2 を一意に識別するためのローカル L D E V 番号である。

【 0 0 7 2 】

「ステータスフラグ」は、ストレージシステム 2 0 内の L D E V 2 4 1 , 2 4 2 を使用している状態にあるか、或いは使用していない状態（空き状態）にあるかを示す。図中の "ON" は、L D E V 2 4 1 , 2 4 2 が使用されている状態にあることを示す。

【 0 0 7 3 】

「外部デバイスフラグ」は、ストレージシステム 2 0 内の L D E V（仮想デバイス）2 4 2 が外部 L D E V 7 0 1 として使用されている状態にあるか否かを示す。図中の "ON" は、L D E V 2 4 2 が外部 L D E V として使用されていることを示す。"OFF" は、L D E V 2 4 1 が内部 L D E V として使用されていることを示す。外部デバイスフラグのデフォルト値は、"OFF" であり、ストレージシステム 2 0 内の L D E V 2 4 2 が外部ストレージシステム 7 0 内の論理デバイス 7 0 1 にマッピングされると、"OFF" が "ON" に変更される。

【 0 0 7 4 】

図 6 は、ストレージシステム 2 0 がアクセス要求を受けた場合に行われる処理の流れを示す。この図 6 と、図 1 及び図 3 を参照して、ホスト計算機 5 1 から仮想ストレージシステム 1 0 0 A に対するアクセスの際に行われる一連の処理流れを説明する。

【 0 0 7 5 】

アプリケーション 6 0 は、L U N を指定したアクセス要求を発行する。交替パスソフトウェア 6 1 が、そのアクセス要求で指定されている L U N に対応するストレージシステム識別子及びグローバル L D E V 番号を、同一デバイス対応テーブル 6 2 から特定する。交替パスソフトウェア 6 1 が、そのストレージシステム識別子から識別されるストレージシステム 2 0 に、そのグローバル L D E V 番号を指定したアクセス要求を発行する。なお、交替パスソフトウェア 6 1 は、指定された L U N に対応する交替パスが混んでいる等の特段の理由がある場合には、別の交替パスを使用してアクセス要求を発行しても良い。

【 0 0 7 6 】

図 6 に示すように、ストレージシステム 6 0 は、グローバル L D E V 番号が指定されたアクセス要求を受信する（ステップ S 1 0 1）。この場合、番号管理プログラム 1 0 2 によって、そのグローバル L D E V 番号に対応するローカル L D E V 番号が、番号管理テーブル 1 0 5 から特定される。ストレージシステム 6 0（例えば、図示しない I / O 処理プログラム）が、デバイス情報テーブル 1 0 6 を参照し、その特定されたローカル L D E V 番号が、内部 L D E V に対応したものか（外部デバイスフラグが "OFF" か）、外部 L D E V に対応したものか（外部デバイスフラグ "ON" か）を判定する（S 1 0 2）。つまり

10

20

30

40

50

、アクセス要求が、外部 L D E V 7 0 1 へのアクセス要求であるか、内部 L D E V 2 4 1 へのアクセス要求であるかが判定される。

【 0 0 7 7 】

ストレージシステム 2 0 は、外部 L D E V 7 0 1 へのアクセス要求であるならば ( S 1 0 2 で Y E S )、マッピングテーブル 1 0 8 を参照し、外部 L D E V 7 0 1 へのアドレス変換等を行って、外部 L D E V 7 0 1 へのアクセス要求を外部ストレージシステム 7 0 に送信する ( S 1 0 4 )。一方、ストレージシステム 2 0 は、内部 L D E V 2 4 1 へのアクセス要求であるならば ( S 1 0 2 で N O )、内部 L D E V 2 4 1 へのアクセスを行う ( S 1 0 3 )。

【 0 0 7 8 】

以上が、本実施形態のシステム全体についての説明である。

【 0 0 7 9 】

本実施形態では、仮想ストレージシステムが、以下に説明する手順で構築される。その手順を、図 2 0 に従って説明する。

【 0 0 8 0 】

1 . 仮想ストレージシステムを構築するストレージシステムの決定 ( S 6 0 1 )

例えば、管理サーバ 5 0 が、管理者から、仮想ストレージシステムの番号 ( 管理サーバ 5 0 内で管理している仮想ストレージシステムのユニークな番号 ) の指定と、その仮想ストレージシステムに含めるストレージシステムの番号 ( 管理サーバ 5 0 内で管理しているストレージシステムのユニークな番号 ) の指定とを受ける。指定された各種番号の登録が行われる。なお、新規に仮想ストレージシステムが作成される場合、所定の方法で生成された新たな仮想ストレージシステム番号 ( 例えば、最後の仮想ストレージシステム番号の最終桁のアルファベットを次のアルファベットに変えたもの ) が、その新規の仮想ストレージシステムに割り当てられる。

【 0 0 8 1 】

2 . S V P 統合 ( S 6 0 2 )

図 1 2 , 1 3 の M S V P 番号と S V P 番号一覧が登録される。S V P を統合せずに、複数 S V P のままで仮想ストレージシステム環境を作る場合は、本処理は行われない。

【 0 0 8 2 】

3 . Active/Passive なストレージシステムの決定 ( S 6 0 3 )

仮想ストレージシステム 1 0 0 の詳細が決められる。具体的には、外部 L D E V 7 0 1 をどのストレージシステム 2 0 で交替パスを張るかが決められる。これは、ストレージシステムの各 L D E V 毎に行われる。

【 0 0 8 3 】

4 . 外部 L D E V 決定処理 ( S 6 0 4 )

ストレージシステム 2 0 に認識させる外部 L D E V 7 0 1 が決定される ( 例えば追加される )。複数のストレージシステム 2 0 の各々の仮想デバイスに同一の外部 L D E V 7 0 1 がマッピングされる場合には、それら複数のストレージシステム 2 0 にその外部 L D E V 7 0 1 が認識される。

【 0 0 8 4 】

5 . ストレージシステム間での交替パスの設定 ( S 6 0 5 )

例えば、ストレージシステム 2 0 A と 2 0 B の両方に外部 L D E V 7 0 1 が認識された場合、異なるストレージシステム 2 0 A、2 0 B をそれぞれ経由する交替パスが設定される。なお、上記 3 . では、同一の外部 L D E V 7 0 1 にアクセス可能な複数のストレージシステム 2 0 のうちのひとつ ( 例えば制御権を有するストレージシステム ) が Active とされ、他が Passive とされる。

【 0 0 8 5 】

以下、本実施形態について更に説明を行う。以下、その説明は、( A ) 階層化された L D E V 番号、( B ) 統合 S V P、( C ) 仮想ストレージシステム 1 0 0 でのフェイルオーバ、の順で行う。また、ストレージシステム 2 0 の上位の装置として、ホスト計算機 5 1

10

20

30

40

50

や仮想化装置 5 2 などが考えられるが、以下の説明では、その上位装置はホスト計算機 5 1 A とする。

【 0 0 8 6 】

< ( A ) 階層化された L D E V 番号 >。

【 0 0 8 7 】

本実施形態の特徴の一つとして、L D E V 番号が階層化されている点にある。すなわち、ストレージシステム内での処理に使用される L D E V 番号をローカル L D E V 番号とし、ホスト計算機 5 1 A に提供される新規の概念の L D E V 番号、つまりグローバル L D E V 番号が用意される。各ストレージシステムは、仮想ストレージシステム 1 0 0 の一員であろうとなかろうと、ローカル L D E V 番号の他に、グローバル L D E V 番号を保持して  
10  
おく。グローバル L D E V 番号が格別必要とされなくても、ストレージシステムがグローバル L D E V 番号を保持することは、本実施形態では有用である。なぜなら、そのストレージシステムを仮想ストレージシステムの一員とすることをスムーズに行うことができるからである。

【 0 0 8 8 】

なお、本実施形態での L D E V 番号の階層化は、例えば上記外部接続の公報に開示されているような、L U N と L D E V 番号との階層とは、技術思想が全く異なる。本実施形態において、L U N とは、オープンホスト（オープン系のシステムでのホスト計算機）が一つの L U （論理ユニット）を識別するための識別子であり、ポート番号と S C S I - I D の組み合わせで決まる。このため、勝手に変更できる番号ではない。本実施形態では、この勝手に  
20  
変更できない L U N の下位に、階層化された L D E V 番号を用意し、その階層化された L D E V 番号を変更することで、上位にインパクトを与えずに、単独で稼動していたストレージシステムを仮想ストレージシステムに組み込んだり、そのストレージシステムを仮想ストレージシステムから減設したりすることができるのである。

【 0 0 8 9 】

以下、具体的に説明する。なお、以下、ストレージシステム 2 0 A 、 2 0 B のそれぞれの要素を説明に用いる際、その要素の親番号に、その要素を有するストレージシステムの子番号と同じ子番号を用いる場合がある。例えば、番号管理プログラムを説明に用いる場合、ストレージシステム 2 0 A の番号管理プログラムは「 1 0 2 A 」とし、ストレージシステム 2 0 B の番号管理プログラムは「 1 0 2 B 」とすることがある。  
30

【 0 0 9 0 】

図 7 a にあるように、ストレージシステム 2 0 A とストレージシステム 2 0 B がそれぞれ単体のシステムとして稼動しているとする。ストレージシステム 2 0 A 内の L D E V 2 4 1 、 2 4 2 は、ストレージシステム 2 0 A 内での L D E V 番号が割り振られ、この L D E V 番号に基づき処理を行っている。この L D E V 番号が、ローカル L D E V 番号である。仮想ストレージシステム 1 0 0 A の一員とならない場合、例えば、L D E V のグローバル L D E V 番号は、その L D E V のローカル L D E V 番号と同じ値とされる。この点は、ストレージシステム 2 0 B についても同様である。ストレージシステム 2 0 A 、 2 0 B では、それぞれ、ローカル L D E V 番号もグローバル L D E V 番号も 0 ~ 6 3 番であるとする。  
40

【 0 0 9 1 】

本実施形態では、ストレージシステム 2 0 が仮想ストレージシステム 1 0 0 の一員であるか否かに関わらず、交替パスソフトウェア 6 1 には、グローバル L D E V 番号が提供される。ホスト計算機 5 1 A からのアクセス要求では、グローバル L D E V 番号が指定されている。例えば、ストレージシステム 2 0 B は、単体で稼動している場合には、番号管理テーブル 1 0 5 B に登録されるグローバル L D E V 番号とローカル L D E V 番号は同じ値である。図 4 a が、単体で稼動しているストレージシステム 2 0 B 内の番号管理テーブル 1 0 5 B の一例を示している。

【 0 0 9 2 】

ここで、ストレージシステム 2 0 A と 2 0 B により仮想ストレージシステム 1 0 0 A が  
50

新規に構築されるとする。この場合、本実施形態では、上位の L D E V 番号であるグローバル L D E V 番号を付け直す処理が行われる。

【 0 0 9 3 】

具体的には、番号管理プログラム 1 0 2 B が、ストレージシステム 2 0 B におけるグローバル L D E V 番号を 0 ~ 6 3 番から 6 4 ~ 1 2 7 番に付け替える ( S 2 0 1 )。図 4 b が、グローバル L D E V 番号が付け替えられた後の、ストレージシステム 2 0 B 内の番号管理テーブル 1 0 5 を示す。なお、ストレージシステム 2 0 A と 2 0 B のグローバル L D E V 番号がそれぞれ 0 ~ 6 3 であることは、番号管理プログラム 1 0 2 同士が対話する、管理サーバ 5 0 から知らせを受ける等、種々の方法で特定することができる。

【 0 0 9 4 】

その後、番号管理プログラム 1 0 2 B が、交替パスソフトウェア 6 1 に、グローバル L D E V 番号の変更に関する番号変更情報を報告する ( S 2 0 2 )。この番号変更情報には、例えば、新たに構築された仮想ストレージシステム 1 0 0 A の番号と、仮想ストレージシステム 1 0 0 A を構成するストレージシステム 2 0 A、2 0 B のそれぞれの番号と、各ストレージシステム 2 0 A、2 0 B での変更前及び変更後のそれぞれのグローバル L D E V 番号とが含まれる。

【 0 0 9 5 】

交替パスソフトウェア 6 1 は、番号変更情報から、ストレージシステム 2 0 A、2 0 B が仮想ストレージシステム 1 0 0 A となりグローバル L D E V 番号が 0 ~ 1 2 7 に変わったことを認識する ( S 2 0 3 )。交替パスソフトウェア 6 1 は、受信した番号変更情報を基に、同一デバイス対応テーブル 6 2 を更新する。具体的には、例えば、ストレージシステム 2 0 A、2 0 B が個々に稼動していたので、交替パスソフトウェア 6 1 は、ストレージシステム 2 0 A、2 0 B にそれぞれ対応した同一デバイス対応テーブル 6 2 A、6 2 B を有していた。交替パスソフトウェア 6 1 は、受信した番号変更情報から、ストレージシステム 2 0 A、2 0 B によって一つの仮想ストレージシステム 1 0 0 A が構築されたことを知る。この場合、交替パスソフトウェア 6 1 は、それらのテーブル 6 2 A、6 2 B を、一つのテーブルとする (例えば、新たに一つの同一デバイス対応テーブルを準備し、その新規の同一デバイス対応テーブルに、同一デバイス対応テーブル 6 2 A、6 2 B に記録されている各情報を登録する)。そして、交替パスソフトウェア 6 1 は、新規の同一デバイス対応テーブルから、上記受信した番号変更情報に含まれている、各ストレージシステム 2 0 A、2 0 B での変更前の各グローバル L D E V 番号を特定する。交替パスソフトウェア 6 1 は、特定した各グローバル L D E V 番号を、上記受信した番号変更情報に含まれている、ストレージシステム 2 0 A、2 0 B での変更後の各グローバル L D E V 番号に更新する。

【 0 0 9 6 】

以上の一連の処理により、個々に稼動していたストレージシステム 2 0 A、2 0 B が一つの仮想ストレージシステム 1 0 0 A とされても、アプリケーション 6 0 に格別影響を与えないようにすることができる。交替パスソフトウェア 6 1 が、ホスト計算機 5 1 ではなく仮想化装置 5 2 に搭載されている場合には、ホスト計算機 5 1 全体に格別影響を与えないようにすることができる。なお、グローバル L D E V 番号の付け替え、番号変更情報の報告は、番号管理プログラム 1 0 2 B に代えて又は加えて、番号管理プログラム 1 0 2 A が行っても良い。

【 0 0 9 7 】

さて、以上のように、グローバル L D E V 番号の付け替えられて同一デバイス対応テーブルの更新が行われた場合、ストレージシステム 2 0 B は、ホスト計算機 5 1 A から、グローバル L D E V 番号 6 4 ~ 1 2 7 番でアクセス要求を受ける。ストレージシステム 2 0 B は、アクセス要求で指定された変更後のグローバル L D E V 番号に対応するローカル L D E V 番号を、更新後の番号管理テーブル 1 0 5 B から特定し、特定されたローカル L D E V 番号で、アクセス要求に従う処理 (特定されたローカル L D E V 番号に対応する L D E V へのアクセス) を実行することができる。

10

20

30

40

50

## 【 0 0 9 8 】

なお、S 2 0 1では、番号管理プログラム 1 0 2 B（及び/又は 1 0 2 A）が、マッピングテーブル 1 0 8 A、1 0 8 Bから、ストレージシステム 2 0 Bのローカル L D E V 番号に対応付けられている外部 L D E V が、ストレージシステム 2 0 Aにも対応付けられているかを調査し、対応付けられているならば、グローバル L D E V 番号の付け替えにおいて、同一の外部 L D E V に対応付けられている複数のローカル L D E V 番号に対し、同一のグローバル L D E V 番号を付与することができる。

## 【 0 0 9 9 】

また、上述した一連の処理は、複数のストレージシステムにより一つの仮想ストレージシステムを構築することに限らず、一つの仮想ストレージシステムに別の仮想ストレージシステムをマージすることにも適用することができる。図 8 a、図 8 bは、仮想ストレージシステム 1 0 0 Aに仮想ストレージシステム 1 0 0 Bをマージして一つの仮想ストレージシステム 1 0 0 Cを構築する場合の具体例を示す。この例によれば、図 8 aに示すように、仮想ストレージシステム 1 0 0 A、1 0 0 Bには、それぞれ、グローバル L D E V 番号として 0 ~ 1 2 7 番が付与されているが、一つの仮想ストレージシステム 1 0 0 Cが構築される場合、図 8 bに示すように、仮想ストレージシステム 1 0 0 Bのグローバル L D E V 番号が、0 ~ 1 2 7 番から 1 2 8 ~ 2 5 5 番に付け替えられる。なお、この番号の付け替えは、番号管理プログラム 1 0 2 A ~ 1 0 2 Dのうちの少なくとも一つが行うことができる。

## 【 0 1 0 0 】

このように、グローバル L D E V 番号の付け替えや、番号変更情報の報告は、新規に仮想ストレージシステムを構築する場合に限らず、仮想ストレージシステムに仮想ストレージシステム或いはストレージシステムをマージする場合にも適用することができる。

## 【 0 1 0 1 】

一つの仮想ストレージシステム 1 0 0 からストレージシステム 2 0（或いは仮想ストレージシステム）を減設する場合は、当然、その分、グローバル L D E V 番号の数も減る。例えば、図 8 bにおいて、ストレージシステム 2 0 Bを減設すると、仮想ストレージシステム 1 0 0 Cのグローバル L D E V 番号 0 ~ 2 5 5 番のうち、グローバル L D E V 番号 6 4 ~ 1 2 7 番がなくなる。この場合、この減設されたタイミングで、グローバル L D E V 番号の付け替えが行われても良いし（1 2 8 ~ 2 5 5 番が、6 4 ~ 1 9 1 番に変更されても良い）、次回にストレージシステム 2 0 がマージされた場合に、グローバル L D E V 番号の付け替えが行われても良い。なお、減設されたストレージシステム 2 0 Bは、単独で稼働させることも可能である。減設されたストレージシステム 2 0 Bの番号管理プログラム 1 0 2 Bは、仮想ストレージシステムから減設されて単独になったことを検知した場合に、グローバル L D E V 番号 6 4 ~ 1 2 7 番を、自分のローカル L D E V 番号（例えば 0 ~ 6 3 番）と同じ番号に変更する。

## 【 0 1 0 2 】

また、一つの仮想ストレージシステム内で、或るストレージシステムの L D E V を別のストレージシステムに移行することも可能である。以下、仮想ストレージシステム 1 0 0 Aのグローバル L D E V 番号 1 0 番の L D E V をストレージシステム 2 0 Aからストレージシステム 2 0 Bへ移行する例を説明する。

## 【 0 1 0 3 】

例えば、ストレージシステム 2 0 Aが、グローバル L D E V 番号 1 0 番の L D E V の移行指示を受けると、グローバル L D E V 番号 1 0 番の L D E V の制御権をストレージシステム 2 0 Aから 2 0 Bへ移行する。このとき、グローバル L D E V 番号 1 0 番の L D E V は、仮想ストレージシステム 1 0 0 Aの中でユニークであるため、ストレージシステム 2 0 Aからストレージシステム 2 0 Bに移行しても、L D E V のグローバル L D E V 番号は 1 0 番のまま変わらない。一方、グローバル L D E V 番号 1 0 番とローカル L D E V 番号との対応付けは、ストレージシステム 2 0 Aとストレージシステム 2 0 Bのそれぞれにおいて変更される。具体的には、例えば、番号管理プログラム 1 0 2 Aが、グローバル L D

10

20

30

40

50

E V 番号 10 番を番号管理テーブル 105 A から削除し、番号管理プログラム 102 B が、番号管理テーブル 105 B から、未使用のローカル L D E V 番号を特定し、特定した未使用のローカル L D E V 番号に、グローバル L D E V 番号 10 番を対応付ける。例として、移行前は、グローバル L D E V 番号 10 番は、ストレージシステム 20 A 内でのローカル L D E V 番号 10 番に対応付けられていたが、移行後は、ストレージシステム 20 B 内でのローカル L D E V 番号 30 番に対応付けられたとする。こうすることで、ホスト計算機 51 からは、移行したことは見えず、アクセス要求を出す際にも、グローバル L D E V 番号が変わらないため影響しない。ストレージシステム 20 A , 20 B 側ではボリュームの制御権が移行したことを認識し、ストレージシステム 20 B は、移行対象のグローバル L D E V 番号に未使用のローカル L D E V 番号を対応付ける。

10

## 【0104】

仮想ストレージシステム 100 間で生成するリモートコピーの場合、グローバル L D E V 番号をコピー元とコピー先で合わせる必要はない。従来のリモートコピーのペアを生成する要領でよい。また、仮想ストレージシステム 100 内で生成するローカルコピーの場合も同様である。

## 【0105】

仮想ストレージシステム 100 内のストレージシステム 20 間で移行できるボリュームを「仮想ストレージシステム適用ボリューム」と呼ぶ。仮想ストレージシステム 100 内のボリュームは、全てが仮想ストレージシステム適用ボリュームでもよいし、仮想ストレージシステム 100 内のボリュームの一部が仮想ストレージシステム適用ボリュームで、残りを各ストレージシステム 20 内のみで使用する従来のボリュームと混在してもよい。

20

## 【0106】

< ( B ) 統合 S V P >。

## 【0107】

次に、S V P の統合管理方法について、図を参照しながら説明する。

## 【0108】

S V P の統合管理とは、仮想ストレージシステム 100 に、一つのマスター S V P ( 以下、M S V P ) を設け、その M S V P によって、複数の他の S V P を管理することである。複数の他の S V P が取得する情報は、M S V P に能動的に送信され、M S V P にて、一括管理される。

30

## 【0109】

S V P の統合管理の第一の方法として、図 9 a に示すようなマスター S V P ( 以下、M S V P ) 8 2 を新規に設定する方法がある。手順について、図 10 を参照しながら説明する。

## 【0110】

各ストレージシステム 20 にある S V P 8 1 を、ネットワークで接続する ( S 3 0 1 ) 。そのネットワーク上に、新規な計算機を接続し、これを M S V P 8 2 とする ( S 3 0 2 ) 。M S V P 8 2 では、所定のイベントが発生した場合に ( 例えば、計算機がネットワークに接続されたことを検知したことを契機に ) 、M S V P としての設定を行うための M S V P 設定プログラム 8 2 1 が、M S V P 8 2 のプロセッサで実行される。

40

## 【0111】

M S V P 設定プログラム 8 2 1 は、仮想ストレージシステム 100 内の全 S V P 8 1 の識別子を認識する ( S 3 0 3 ) 。各 S V P 8 1 には、その S V P の識別子を設定しておく。方法としては、M S V P 設定プログラム 8 2 1 が、仮想ストレージシステム 100 内の全 S V P 8 1 の識別子の入力をユーザから受けるか、或いは、全 S V P 8 1 へブロードキャストでメッセージと M S V P 8 2 の識別子を送り、各 S V P 8 1 の S V P 間認識プログラム 8 1 1 より、その S V P 8 1 の識別子を返信してもらう。M S V P 設定プログラム 8 2 1 は、M S V P 8 2 の記憶資源 ( 例えばメモリ ) に、S V P 一覧表 ( 図示せず ) を記憶させる。その S V P 一覧表に、各 S V P 8 1 の S V P 識別子が記録される。

## 【0112】

50

各SVP81(具体的には、例えば、SVPのプロセッサで実行される制御プログラム)は、MSVP82から受信したMSVP識別子を、自分の記憶資源(例えばメモリ)に登録する(S304)。なお、MSVP識別子は、ユーザから入力されても良い。

【0113】

各SVP81は、SVP81上で管理する情報を、MSVP82に送り、MSVP82にて、各々のSVP(これらはスレーブSVP:SSVPと呼ぶ)81で管理されている情報が一括管理される(S305)。例えば、SSVP81は、自分が接続されているストレージシステムから、番号管理テーブル105を取得し、取得した番号管理テーブル105を、MSVP82に送信する。

【0114】

MSVP82と他のSSVP81が設定されたことを、例えば、MSVP設定プログラム821が、仮想ストレージシステムを構成する全てのストレージシステム20に報告する(S306)。例えば、MSVP識別子と各SSVP識別子が、各ストレージシステム20に報告され、そのストレージシステム20のメモリ22等に記録される。

【0115】

MSVP設定プログラム821は、各SSVP81から情報を受信した場合、受信した情報に、その情報の送信元のSSVP81の識別子を関連付ける。これにより、MSVP82は、どのストレージシステム20のSSVPの情報かを意識して管理することができる。MSVP82は交替系を持たせるため2重化されても良い。

【0116】

この第一の方法によれば、SSVP81の負荷を均等にすることができる。また、ストレージシステム20の増減設にも対応しやすい。

【0117】

SVP統合管理の第二の方法として、図9bに示すように、複数のSVP81のうち1台をMSVPと兼用する方法もある。以下、その手順について、図11を参照しながら説明する。

【0118】

各ストレージシステム20にあるSVP81をネットワークで接続する(S401)。

【0119】

次に、例えばユーザが、複数のSPV81の中から、MSVPに兼用するものを決定する(S402)。各SVP81は、どれもMSVPになる可能性があるため、全てのSVP81が、前述したMSVP設定プログラム821を予め持ち、MSVPに選択された場合だけ、そのプログラム821が実行される。また、全SVP81は、SVP間認識プログラム811を持ち、SVP81間の相互通信を行う。

【0120】

ユーザは、MSVPと決めたSVP81のMSVP設定プログラム821を実行させる(S403)。MSVP設定プログラム821が実行されたSVP82(81)は、自分がMSVPであることを認識し、MSVP82内の制御情報へ登録する(S404)。具体的には、例えば、MSVP設定プログラム821が、MSVPであることを意味する情報を、SVP82の記憶資源に登録する。

【0121】

MSVP設定プログラム821は、他のSVP81へ、MSVP識別子(自分の識別子)を、ブロードキャストで知らせる(S405)。各他のSVP81のSVP間認識プログラム811は、MSVP識別子を受信し、自分を有するSVP81の記憶資源に、そのMSVP識別子に登録する。また、各他のSVP81は、MSVP識別子を受信に回答して、自分の識別子を、MSVP82に返信する。また、各他のSVP81のSVP間認識プログラム811は、自分を有するSVP81の記憶資源で管理されている情報を、MSVP82に送る。これにより、全ての他のSVP81での管理情報がMSVP82にて一括管理される。他のSVP81からMSVP82に送信される情報として、例えば、該他のSVP81が接続されているストレージシステム内の番号管理テーブル105がある。

10

20

30

40

50

## 【 0 1 2 2 】

M S V P と他の S V P (以下、S S V P) が設定されたことを、例えば M S V P 設定プログラム 8 2 1 が、仮想ストレージシステムを構成する全てのストレージシステム 2 0 に報告する (S 4 0 6)。具体的には、例えば、各 S V P の識別子と、各識別子が M S V P の識別子であるか S S V P の識別子であるかを示す情報とを、各ストレージシステム 2 0 に報告する。各ストレージシステム 2 0 は、それらの情報から、自分の S V P が M S V P か S S V P か認識し登録することができる。

## 【 0 1 2 3 】

以上の第二の方法では、複数の S S V P の中から、M S V P の交替系となるものを選択しておいてもよい。すなわち、M S V P を 2 重化しておいてもよい。この場合、稼働している M S V P がダウンした場合、選択されていた他の S V P が M S V P として稼働することができる。M S V P がダウン (例えば故障) したか否かは、例えば、S V P 間認識プログラム 8 1 1 同士の通信から検出することができる。なお、この第二の方法でも、前述した第一の方法でも、同時期に稼働する M S V P の台数は、仮想ストレージシステム 1 0 0 において 1 台とする。その 1 台の M S V P がダウンした場合に、交替系の M S V P が稼働することができる。

10

## 【 0 1 2 4 】

第一の方法及び第二の方法のいずれでも、M S V P が設定されたら、S V P 8 1 からの指定はできなくなり、それ以降は M S V P から指定する。M S V P や S V P で、前述した番号管理プログラム 1 0 2 が実行されても良い。

20

## 【 0 1 2 5 】

また、第二の方法の場合、M S V P が設定されたストレージシステム 2 0 A が停止されたら、M S V P 機能も停止してしまう。このため、他のストレージシステム 2 0 の S V P 8 1 に、M S V P の役割を移す必要がある。そこで、ストレージシステム 2 0 を計画的に停止する際には (つまり意図的に停止する際には)、そのストレージシステム 2 0 が停止することを他のストレージシステム 2 0 へ通知し、他のストレージシステム 2 0 で、そのストレージシステム 2 0 が停止することを認識できるようにする。その時の通知方法は、ホスト経由でもストレージシステム間通信 4 4 を使用してもよい。M S V P のあるストレージシステム 2 0 A が計画的に停止される場合、そのストレージシステム 2 0 A は、他のストレージシステム 2 0 へ、M S V P になることのメッセージを送信し、且つ、M S V P で管理されていた全ての管理情報を送る。そのメッセージを受けたストレージシステム 2 0 は、自分に接続されている S V P 8 1 を M S V P に設定し、且つ、送られて来た管理情報を、その S V P 8 1 の記憶資源に格納する。また、障害などで M S V P のあるストレージシステム 2 0 が突然停止した場合も、他のストレージシステム 2 0 は、定期的にストレージシステム 2 0 間で通信し、停止したことを検知できるようにしておく。停止したと検知したストレージシステム 2 0 が、M S V P があるストレージシステム 2 0 A であると判断したら、自分に接続されている S V P が M S V P となるように設定する。そのストレージシステム 2 0 は、停止したストレージシステム 2 0 A の M S V P から管理情報を取れ出せれば、その管理情報を抽出し、抽出した管理情報を、M S V P としての S V P の記憶資源に格納する。以上のストレージシステム間のやり取りは、S V P 間で行われても良い。

30

40

## 【 0 1 2 6 】

また、各 S S V P は、自分を備えるストレージシステムから取得し記憶した情報を、自分の S S V P 識別子と共に、M S V P に送信する。M S V P 設定プログラムは、受信した情報を、該情報と共に受信した S S V P 識別子を関連付けて、M S V P の記憶資源に格納する。

## 【 0 1 2 7 】

また、各ストレージシステムへの M S V P 設定報告で報告することとしては、例えば、SVP を統合したことである。

## 【 0 1 2 8 】

さて、ストレージシステムの S V P が M S V P であるか等が書かれたストレージシステ

50

ム管理情報を、管理サーバ50にて管理することができる。ストレージシステム管理情報は、ストレージシステム毎に用意される。以下、それについて説明する。

【0129】

図12aは、ストレージシステム20Aについてのストレージシステム管理情報を示す。図12bは、ストレージシステム20Dについてのストレージシステム管理情報を示す。以下、図12aを代表的に参照して説明する。

【0130】

ストレージシステム管理情報951Aには、例えば、仮想ストレージシステム構成情報、ストレージシステム情報、仮想ストレージシステム番号、SVP番号及びMSVP番号が記録される。

10

【0131】

「仮想ストレージシステム構成情報」は、ストレージシステム20Aが仮想ストレージシステム100を構成しているか否かを示すビット情報である。図12aによれば、「Yes」という値が記録されているので、ストレージシステム20Aが仮想ストレージシステム100を構成していることがわかる。一方、図12bによれば、「No」という値が記録されているので、ストレージシステム20Dが仮想ストレージシステム100を構成していない（つまり単独で稼動している）ことがわかる。

【0132】

「ストレージシステム情報」は、その仮想ストレージシステム100を構成している他のストレージシステム20の番号である。この図によれば、ストレージシステム20Aと20Bによって一つの仮想ストレージシステムが構成されていることがわかる。

20

【0133】

「仮想ストレージシステム番号」は、ストレージシステム20Aを構成要素の一つとしている仮想ストレージシステム100の番号である。

【0134】

「SVP番号」は、ストレージシステム20AのSVPの識別子である。

【0135】

「MSVP番号」は、MSVPの識別子である。ストレージシステム20AのSVPがMSVPの場合には、SVP番号と同じ番号が、MSVP番号として記録される。なお、単独でストレージシステム20Dが稼動している場合にも、SVP番号と同じ番号が、MSVP番号として記録される（この場合、それに代えて、MSVP番号は空白であっても良い）。

30

【0136】

図13は、仮想ストレージシステム管理情報を示す。

【0137】

仮想ストレージシステム管理情報953は、仮想ストレージシステム100毎に1つ用意される。この情報も、管理サーバ50の記憶資源に記憶される。以下、この仮想ストレージシステム管理情報953は、仮想ストレージシステム100Aに対応した情報とする。

【0138】

仮想ストレージシステム管理情報953には、仮想ストレージシステム番号、ストレージシステム情報、MSVP番号及びSVP番号一覧が記録される。

40

【0139】

「仮想ストレージシステム番号」は、仮想ストレージシステム100Aの識別子（番号）である。

【0140】

「ストレージシステム情報」は、仮想ストレージシステム100Aを構成するストレージシステム20の識別子（番号）の一覧である。

【0141】

「MSVP番号」は、仮想ストレージシステム100AのMSVPの識別子である。

50

## 【 0 1 4 2 】

「SVP番号一覧」は、仮想ストレージシステム100A内のSVPの識別子の一覧である。ここでは、例えば、第二の方法でSVP統合管理が行われる場合、図示のように、MSVP識別子と同じ識別子(S20A)が記録される。一方、第一の方法でSVP統合管理が行われる場合、MSVP識別子と同じ識別子がSVP識別子として記録されない(第二の方法と同様に記録されても良い)。

## 【 0 1 4 3 】

図14は、仮想ストレージシステム全体情報を示す。

## 【 0 1 4 4 】

仮想ストレージシステム全体情報955は、管理サーバ50の記憶資源に記憶される。この情報955には、「仮想ストレージシステム番号一覧」という項目があり、これは、管理サーバ50の管理対象である全ての仮想ストレージシステムの番号である。「仮想ストレージシステム番号一覧」に含まれる番号は、例えば、図20のS601で追加される。

10

## 【 0 1 4 5 】

<(C)仮想ストレージシステムでのフェイルオーバー>。

## 【 0 1 4 6 】

仮想ストレージシステム100は、ホスト計算機51からは1つのシステムとして見える。仮想ストレージシステム100にて、例えばフェイルオーバー機能を実現する。このとき、ストレージシステム20Bをストレージシステム20Aの交替系とする。実際には、LDEV単位で交替系を設定するため、別のLDEVに関しては、ストレージシステム20Aがストレージシステム20Bの交替系となってもよい。

20

## 【 0 1 4 7 】

図15は、フェイルオーバー管理情報を示す。具体的には、図15aが、フェイルオーバーするLDEVのためのフェイルオーバー管理情報を示す。一方、図15bが、フェイルオーバーしないLDEVのフェイルオーバー管理情報を示す。

## 【 0 1 4 8 】

フェイルオーバー管理情報957は、例えば、各ストレージシステムのメモリ22に記憶される。フェイルオーバー管理情報957には、その情報が対応付けられたグローバルLDEV番号と、初期Activeストレージシステム番号(初期にActiveとするストレージシステムの番号)と、現Activeストレージシステム番号(現在Activeのストレージシステムの番号)と、現Passiveストレージシステム番号(現在Passiveのストレージシステムの番号)と、フェイルオーバー有無(フェイルオーバーを行うか否か)と、状態(例えば、フェイルオーバーを行っている最中か否か)とが記録される。フェイルオーバーしないグローバルLDEV番号に対応したフェイルオーバー管理情報は、図15bに示すように、全てがブランクとなっている。なお、フェイルオーバーしないLDEVについては、フェイルオーバー管理情報は無くても良い。

30

## 【 0 1 4 9 】

通常系のストレージシステム20Aを「Active」と呼び、交替系のストレージシステム20Bを「Passive」と呼ぶ。このようにすることで、ストレージシステム20と外部ストレージシステム70間のパス障害やストレージシステム20の計画停止時及びストレージシステム20の障害時にも、「Active」から「Passive」へ処理の主導権を切り替えることにより、外部ボリューム701に影響させずに処理を続けることができる。

40

## 【 0 1 5 0 】

また、ストレージシステム20のマイグレーションをする際に、新旧のストレージシステム20間で外部ボリューム701の制御権を切り替える移行機能においても、最初に移行元を「Active」(外部ボリュームへのパスが接続されている)、移行先を「Passive」とし、Active/Passiveを切り替え処理を行うことで実現できる。

## 【 0 1 5 1 】

ストレージシステム20Aとストレージシステム20B間でのフェイルオーバー処理につ

50

いて図 17 を参照しながら説明する。

【 0 1 5 2 】

図 17 は、図 1 をフェイルオーバ処理に関連のある部分を抽出して簡易化したものである。ここでは、外部ストレージシステム 70 内の外部 L D E V X が実体で、それに対応付けられた L D E V 240 を、仮想デバイス A、B とする。仮想デバイス A は、ストレージシステム 20 A 内にあり、仮想デバイス B は、ストレージシステム 20 B 内にある。仮想デバイス A、B は、同一の実体 X に接続されているので、同一のグローバル L D E V 番号（以下、グローバル L D E V 番号 X）が付与されているものとする。仮想デバイス A、B へホスト計算機 51 からパスを張り、それらを交替パスソフトウェア 61 により交替パスとして設定される。ここでは、ストレージシステム 20 A を Active、ストレージシステム 20 B を Passive とする。

10

【 0 1 5 3 】

ホスト計算機 51 から、グローバル L D E V 番号 X が指定されたアクセス要求が、ストレージシステム 20 A 又は 20 B に対して出される。ストレージシステム 20 A が、そのアクセス要求を受けた場合、グローバル L D E V 番号 X に対応したフェイルオーバ管理情報 957 では、ストレージシステム 20 A が Active となっているため、そのアクセス要求は、ストレージシステム 20 A で処理される。このため、もし、ストレージシステム 20 B で、そのアクセス要求を受けた場合は、そのアクセス要求を、ストレージシステム 20 B から 20 A へ専用線 44 を介して転送する。ストレージシステム 20 A は、ホスト計算機 51 から受けたアクセス要求であるかストレージシステム 20 B から受けたアクセス要求であるかを区別することなく処理することができる。

20

【 0 1 5 4 】

ここで、仮想デバイス A と外部 L D E V X との間のパスに障害が発生したとする。この場合、Active 側のパスが障害になるため、外部 L D E V X に対してのコントローラ部分であるストレージシステム 20 A と 20 B の Active と Passive の役割を入れ替えることを行う。

【 0 1 5 5 】

まず、図 19 に示すように、ストレージシステム 20 B が、仮想デバイス B と外部 L D E V X 間のパスを有効にする（S501）。

【 0 1 5 6 】

一方、ストレージシステム 20 A が、仮想デバイス A 側で既に受けつけ済みであるが、外部 L D E V X へのアクセスが未処理なアクセス要求を、仮想デバイス B 側に転送する（S502）。これにより、ストレージシステム 20 B が、転送されて来たアクセス要求に従って、外部 L D E V X へアクセスする（S503）。

30

【 0 1 5 7 】

その後、ストレージシステム 20 A と 20 B の Active/Passive が切り替えられる（S504）。すなわち、少なくとも、グローバル L D E V 番号 X に対応したフェイルオーバ管理情報において、現 Active ストレージ番号が、ストレージシステム 20 B の番号となり、現 Passive ストレージ番号が、ストレージシステム 20 A の番号となる。

【 0 1 5 8 】

以上により、ストレージシステム 20 A と 20 B の Active と Passive の切替が完了する。以後、グローバル L D E V 番号 X を指定したアクセス要求を、ストレージシステム 20 B が受けた場合は、図 18 に示すように、ストレージシステム 20 B が、そのアクセス要求を処理する。一方、ストレージシステム 20 A が、そのアクセス要求を受けた場合は、そのアクセス要求を、ストレージシステム 20 A から専用線 44 を介してストレージシステム 20 B へ転送する。

40

【 0 1 5 9 】

このようにして、一つの仮想ストレージシステム 100 において、ストレージシステム間のフェイルオーバを実現することができる。

【 0 1 6 0 】

50

なお、図1は、ホスト計算機51から各ストレージシステム20へコマンドを発行するためのネットワーク(SAN41, 42)と、管理サーバ50が各ストレージシステム20との間で管理情報を送受信するためのネットワーク(管理ネットワーク40)とが異なるアウトバンド方式となっているが、本実施形態は、インバンド方式にも適用できる。例えば、管理サーバ50がSAN41経由で各ストレージシステム20との間で管理情報を送受信するように構成すれば、インバンド方式となる。

#### 【0161】

以上、本発明の好適な実施形態を説明したが、これは本発明の説明のための例示であって、本発明の範囲をこの実施形態にのみ限定する趣旨ではない。本発明は、他の種々の形態でも実施することが可能である。

#### 【0162】

例えば、図16に示すように、外部ストレージシステム70は無くても良い。また、上述した実施形態に係るシステム全体構成では、メインフレームシステムとオープンシステムのいずれか一方或いはそれらが混在してもよい。

#### 【0163】

また、例えば、階層化して管理されるのは、LDEV番号に限らない。例えば、LDEV番号に代えて又は加えて、コピー系のペア番号やグループ番号などのリソース番号を階層化し、LDEV番号と同様の方法で付け替え等を行うことができる。今まで別々のストレージシステムにあったリソース番号が、1つの仮想ストレージシステムとなったために、ユニークな番号であることが必要となる。例えば、ストレージシステム20Aに、LDEV#0とそのレプリケーションであるLDEV#100をペア#0として管理しているとする。同じように、ストレージシステム20Bに、LDEV#0とそのレプリケーションであるLDEV#100をペア#0として管理しているとする。ここで、ストレージシステム20Aと20Bを仮想ストレージシステム100とした場合、LDEV#0と#100が重複するため、前記のようにLDEV番号を付け直す必要がでるが、さらに、ペア番号もペア#0で重複してしまう。このため、LDEV番号と同様に、グローバルペア番号とローカルペア番号を設けることができる。番号の付け替え方法は、LDEV番号の付け替え方法と同様にすることができる。コピー系のペア番号をグループにしたグループ番号についても、グローバルグループ番号とローカルグループ番号を設けて同様に処理を行うことができる。

#### 【図面の簡単な説明】

#### 【0164】

【図1】図1は、本発明の一実施形態に係るシステム全体の構成例を示す。

【図2】図2は、同一デバイス対応テーブル62の構成例を示す。

【図3】図3は、ストレージシステム20のメモリ22内に格納されている各種プログラム及びテーブルを示す。

【図4】図4aは、単体で稼動しているストレージシステム20B内の番号管理テーブル105Bの一例を示す。図4bが、グローバルLDEV番号が付け替えられた後の、ストレージシステム20B内の番号管理テーブル105を示す。

【図5】図5は、デバイス情報テーブル106の構成例を示す。

【図6】図6は、ストレージシステム20がアクセス要求を受けた場合に行われる処理の流れを示す。

【図7】図7aは、仮想化ストレージシステムを構成しない各ストレージシステム20A、20BのグローバルLDEV番号とローカルLDEV番号の関係を示す。図7bは、仮想化ストレージシステムを構成したストレージシステム20A、20BのグローバルLDEV番号とローカルLDEV番号の関係を示す。

【図8】図8aは、別々の仮想化ストレージシステム100A, 100BのグローバルLDEV番号とローカルLDEV番号の関係を示す。図8bは、仮想ストレージシステム100Aに別の仮想ストレージシステム100Bがマージされたことにより構成された仮想ストレージシステム100CのグローバルLDEV番号とローカルLDEV番号の関係を示す。

10

20

30

40

50

【図 9】図 9 a は、S V P 統合管理の第一の方法の説明図である。図 9 b は、S V P 統合管理の第二の方法の説明図である。

【図 10】図 10 は、第一の方法での S V P 統合管理を行うための手順を示すフローチャートである。

【図 11】図 11 は、第二の方法での S V P 統合管理を行うための手順を示すフローチャートである。

【図 12】図 12 a は、ストレージシステム 20 A についてのストレージシステム管理情報を示す。図 12 b は、ストレージシステム 20 D についてのストレージシステム管理情報を示す。

【図 13】図 13 は、仮想ストレージシステム管理情報を示す。

10

【図 14】図 14 は、仮想ストレージシステム全体情報を示す。

【図 15】図 15 a は、フェイルオーバーする L D E V のためのフェイルオーバー管理情報を示す。図 15 b は、フェイルオーバーしない L D E V のフェイルオーバー管理情報を示す。

【図 16】図 16 は、本実施形態の一変形例に係るシステム全体構成を示す。

【図 17】図 17 は、仮想ストレージシステムでのアクセスの流れの一例である。

【図 18】図 18 は、図 17 のシステムで、ストレージシステム 20 A と外部ストレージシステム 70 との間のパスに障害が生じた後の、アクセスの流れの一例である。

【図 19】図 19 は、ストレージシステム間での Active/Passive の切替の流れの一例である。

【図 20】図 20 は、仮想ストレージシステムを構築する手順の一例を示すフローチャートである。

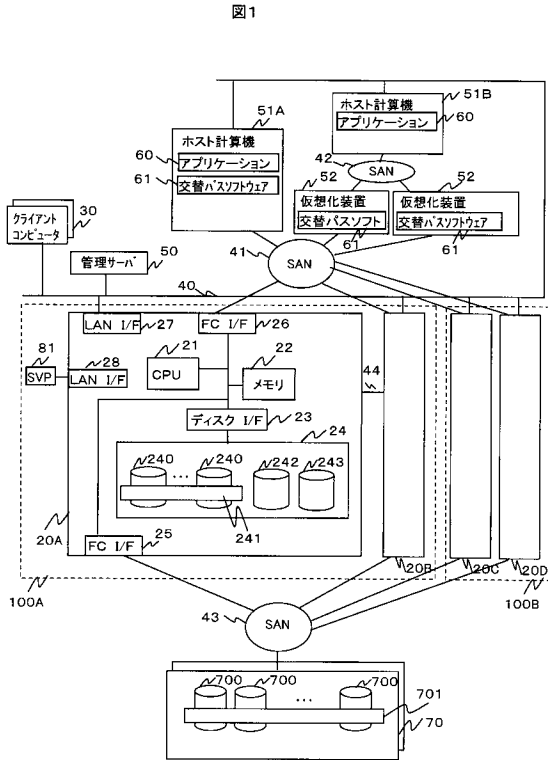
20

【符号の説明】

【0165】

20 ... ストレージシステム    51 ... ホスト計算機    52 ... 仮想化装置    70 ... 外部ストレージシステム  
100 ... 仮想ストレージシステム    105 ... 番号管理テーブル    108 ... マッピングテーブル

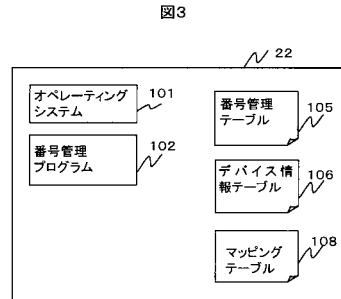
【図1】



【図2】

バス番号	ストレージシステム識別子	LUN	グローバルLD EV番号
0	HA	12	11
1	HA	10	9
2	HA	11	11
3	HB	9	11

【図3】



【図4】

図4 a

グローバルLDEV番号	ローカルLDEV番号
0	0
1	1
...	...

【図5】

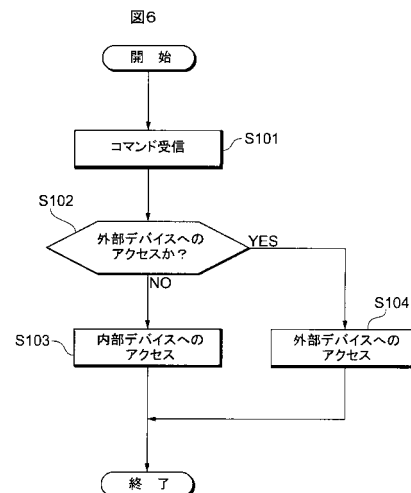
図5

デバイス番号	ステータスフラグ	外部デバイスフラグ
0	ON	OFF
1	ON	OFF
2	ON	OFF
3	ON	ON

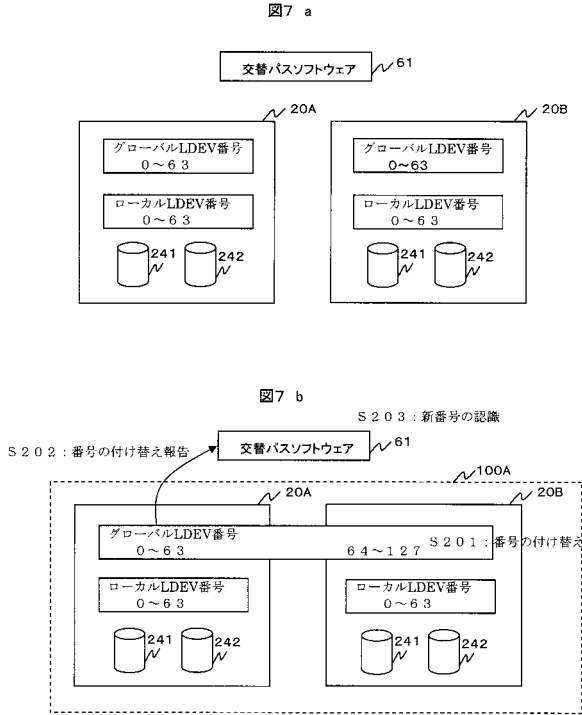
図4 b

グローバルLDEV番号	ローカルLDEV番号
64	0
65	1
66	2
...	...

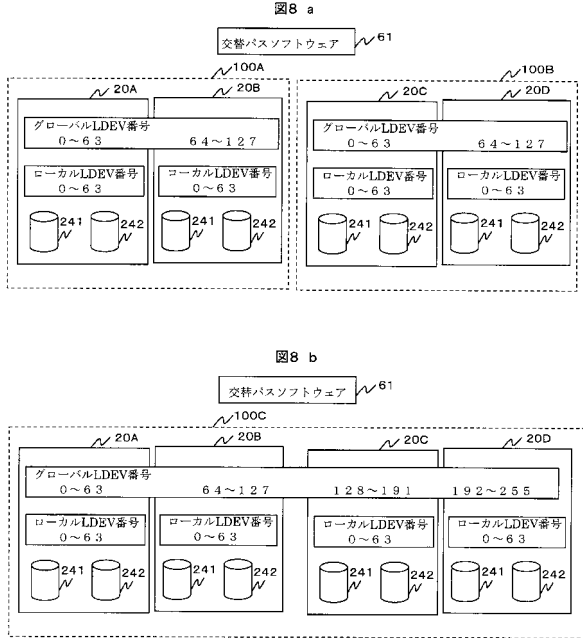
【図6】



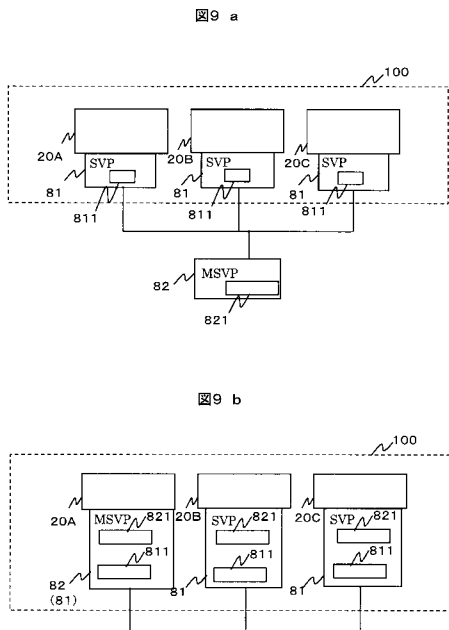
【 図 7 】



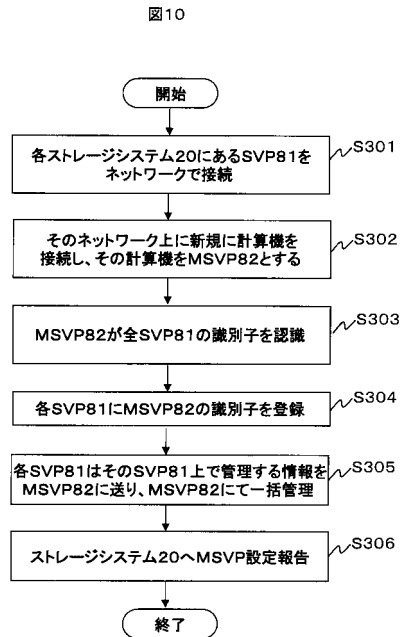
【 図 8 】



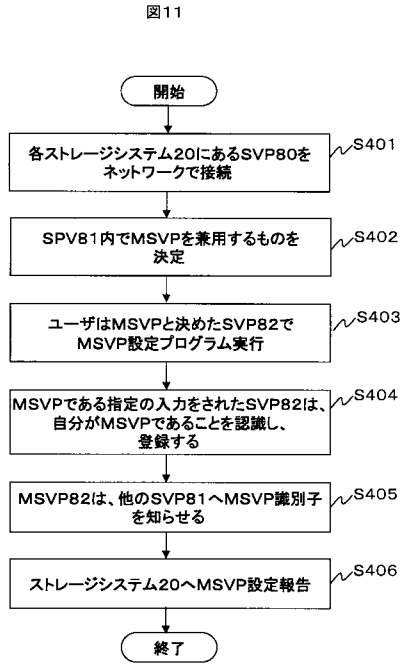
【 図 9 】



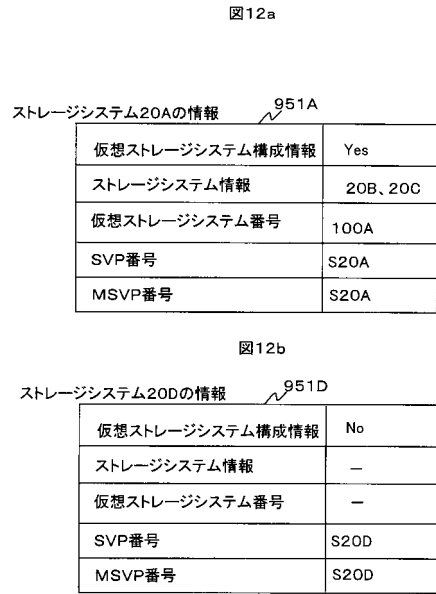
【 図 1 0 】



【 図 1 1 】



【 図 1 2 】



【 図 1 3 】

図13

953

仮想ストレージシステム構成番号	100A
ストレージシステム情報	20A、20B
MSVP番号	S20A
SVP番号一覧	S20A、S20B、

【 図 1 5 】

図15 a

957

グローバルLDEV番号	1
初期Activeストレージシステム番号	20A
現Activeストレージシステム番号	20B
現Passiveストレージシステム番号	20A
フェイルオーバー有無	する
状態(移行中/定常)	移行中

【 図 1 4 】

図14

955

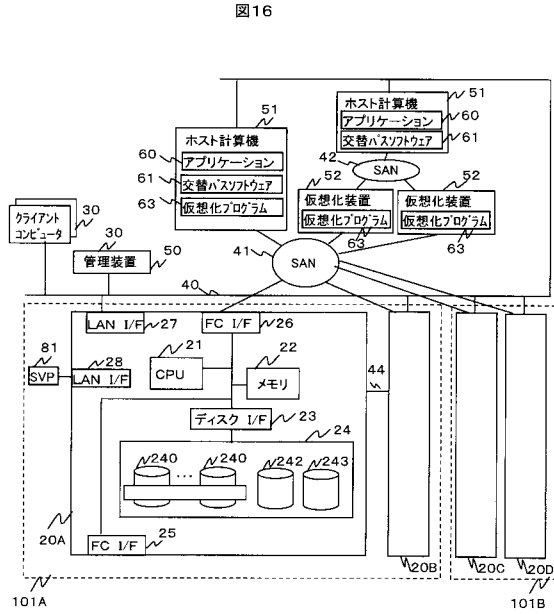
仮想ストレージシステム番号一覧	100A 100B
-----------------	-----------

図15 b

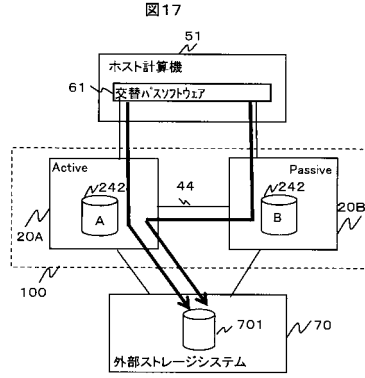
957

グローバルLDEV番号	-
初期Activeストレージシステム番号	-
現Activeストレージシステム番号	-
現Passiveストレージシステム番号	-
フェイルオーバー有無	-
状態(移行中/定常)	-

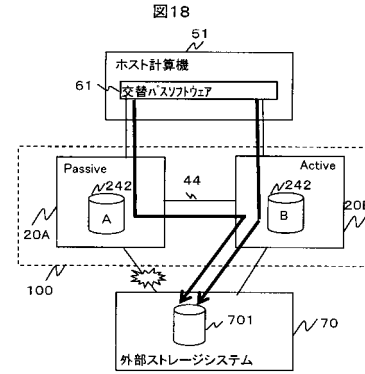
【図16】



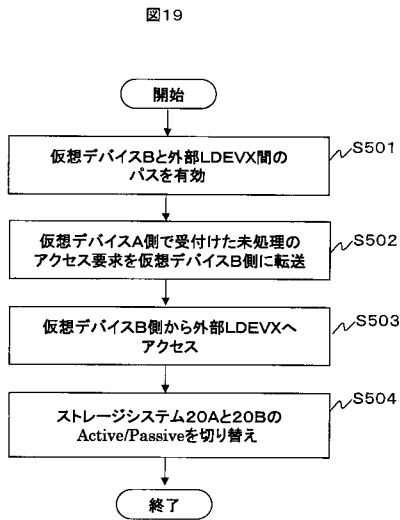
【図17】



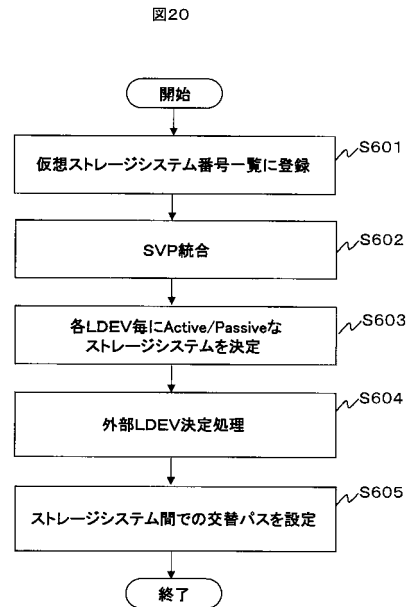
【図18】



【図19】



【図20】



---

フロントページの続き

審査官 横山 佳弘

- (56)参考文献 特開2005-107645(JP,A)  
特開2000-099278(JP,A)  
特開2005-301561(JP,A)  
特開2001-350707(JP,A)  
特開2006-178720(JP,A)  
特開平10-275056(JP,A)

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