

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6186787号
(P6186787)

(45) 発行日 平成29年8月30日 (2017. 8. 30)

(24) 登録日 平成29年8月10日 (2017. 8. 10)

(51) Int. Cl.

F I

G 0 6 F 3 / 0 6 (2006. 01)
G 0 6 F 1 2 / 0 0 (2006. 01)G 0 6 F 3 / 0 6 3 0 1 X
G 0 6 F 1 2 / 0 0 5 4 5 M

請求項の数 9 (全 24 頁)

(21) 出願番号 特願2013-62383 (P2013-62383)
 (22) 出願日 平成25年3月25日 (2013. 3. 25)
 (65) 公開番号 特開2014-186652 (P2014-186652A)
 (43) 公開日 平成26年10月2日 (2014. 10. 2)
 審査請求日 平成27年11月6日 (2015. 11. 6)

(73) 特許権者 000005223
 富士通株式会社
 神奈川県川崎市中原区上小田中4丁目1番
 1号
 (74) 代理人 100107766
 弁理士 伊東 忠重
 (74) 代理人 100070150
 弁理士 伊東 忠彦
 (74) 代理人 100146776
 弁理士 山口 昭則
 (72) 発明者 繁田 聡一
 神奈川県川崎市中原区上小田中4丁目1番
 1号 富士通株式会社内

審査官 桜井 茂行

最終頁に続く

(54) 【発明の名称】 データ転送装置、データ転送システム、データ転送方法及びプログラム

(57) 【特許請求の範囲】

【請求項 1】

記憶装置の間にてネットワークを介してデータの転送を行うデータ転送装置であって、
 前記記憶装置内の転送すべきデータの転送目的を判定する判定部と、
 前記データの転送目的が一の装置から他の装置へのV M移行目的であると判定された場
 合であって、かつ前記ネットワークの帯域の使用量が所定の閾値以上である場合、前記転
 送すべきデータの中から転送するデータを選択する選択部と、
 を有するデータ転送装置。

【請求項 2】

前記ネットワークの帯域を監視する帯域監視部を更に有し、
 前記選択部は、前記帯域監視部により監視される前記ネットワークの帯域の使用量に応
 じて、前記転送すべきデータの中から転送するデータを選択する、
 請求項 1 に記載のデータ転送装置。

【請求項 3】

前記記憶装置に記憶されたデータの更新頻度を記憶する更新頻度記憶部を更に有し、
 前記選択部は、特定の前記転送目的のデータに関して前記更新頻度記憶部が記憶する更
 新頻度に基づいて、前記転送すべきデータの中から転送するデータを選択する、
 請求項 1 又は 2 に記載のデータ転送装置。

【請求項 4】

前記記憶装置に記憶されたデータは、複数のブロックに分割され、

10

20

前記複数のブロックに分割されて前記記憶装置に記憶された各々のデータの読取頻度を記憶する読取頻度記憶部を更に有し、

前記選択部は、特定の前記転送目的のデータに関して前記読取頻度記憶部が記憶するブロック毎の読取頻度に基づいて、前記転送すべきデータの中から転送するデータのブロックを選択する、

請求項 3 に記載のデータ転送装置。

【請求項 5】

前記選択部は、前記データの転送目的が一の装置から他の装置への V M 移行目的であると判定されたであって、かつ前記ネットワークの帯域の使用量が所定の閾値以上である場合に、前記転送すべきデータの中から転送するデータを選択する、

10

請求項 1 ～ 4 のいずれか一項に記載のデータ転送装置。

【請求項 6】

複数の記憶装置と、該複数の記憶装置間にてネットワークを介してデータの転送を行うデータ転送装置と、を含むデータ転送システムであって、

前記データ転送装置は、

前記複数の記憶装置のいずれかの記憶装置内の転送すべきデータの転送目的を判定する判定部と、

前記データの転送目的が一の装置から他の装置への V M 移行目的であると判定された場合であって、かつ前記ネットワークの帯域の使用量が所定の閾値以上である場合、前記転送すべきデータの中から転送するデータを選択する選択部と、

20

を有するデータ転送システム。

【請求項 7】

記憶装置間にてネットワークを介してデータを転送する処理をコンピュータが実行するデータ転送方法であって、

転送すべきデータの転送目的を判定し、

前記データの転送目的が一の装置から他の装置への V M 移行目的であると判定された場合であって、かつ前記ネットワークの帯域の使用量が所定の閾値以上である場合、前記転送すべきデータの中から転送するデータを選択する、

処理をコンピュータが実行するデータ転送方法。

【請求項 8】

30

記憶装置間にてネットワークを介してデータを転送する処理をコンピュータに実行させるプログラムであって、

転送すべきデータの転送目的を判定し、

前記データの転送目的が一の装置から他の装置への V M 移行目的であると判定された場合であって、かつ前記ネットワークの帯域の使用量が所定の閾値以上である場合、前記転送すべきデータの中から転送するデータを選択する、

処理をコンピュータに実行させるプログラム。

【請求項 9】

前記データの転送目的がデータのバックアップ目的又はシステム管理に用いるデータの転送目的であると判定された場合、前記転送すべきデータを送信し、前記データの転送目的が一の装置から他の装置への V M 移行目的であると判定された場合、前記転送すべきデータの中から選択したデータを送信する送信部を有する、

40

請求項 1 ～ 5 のいずれか一項に記載のデータ転送装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、データ転送装置、データ転送システム、データ転送方法及びプログラムに関する。

【背景技術】

【0002】

50

データを複数のデータセンター間で遠隔的にミラーリングする技術がある（例えば、特許文献１、２参照）。遠隔的なミラーリングでは、データ転送先の候補となるデータセンターのうち、容量に余裕のあるデータセンターにデータがミラーリングされる。

【０００３】

ミラーリングの一の目的としては、災害対策ＤＲ（Disaster Recovery）のためのバックアップが挙げられる。ミラーリングの別の目的としては、仮想マシン（以下、単にＶＭ（Virtual Machine）という。）の移行準備目的が挙げられる。ＶＭ移行の準備目的では、広域負荷分散のための仮想マシンの移行に備えて、予め仮想マシンの仮想ディスクを他のデータセンターにミラーリングする。

【先行技術文献】

【特許文献】

【０００４】

【特許文献１】特開２００５－１４９２８５号公報

【特許文献２】特開２０１２－１６４１８５号公報

【発明の概要】

【発明が解決しようとする課題】

【０００５】

しかしながら、災害対策ＤＲ等のバックアップを目的としたミラーリングも、ＶＭ移行の準備を目的としたミラーリングも、データセンター間を接続するネットワークを介して行われる。よって、ミラーリングのデータ転送量が増加すると、ネットワークの帯域が不足する事態が起こり得る。ネットワークの帯域不足は、ミラーリングの目的が何であろうと、実行されているすべてのミラーリングの処理に影響を与え、ミラーリングが完了するまでの時間を長引かせるという課題が生じる。

【０００６】

そこで、一側面では、ネットワークの帯域不足の抑制を目的とする。

【課題を解決するための手段】

【０００７】

一つの場合では、記憶装置の間にてネットワークを介してデータの転送を行うデータ転送装置であって、前記記憶装置内の転送すべきデータの転送目的を判定する判定部と、前記データの転送目的が一の装置から他の装置へのＶＭ移行目的であると判定された場合であって、かつ前記ネットワークの帯域の使用量が所定の閾値以上である場合、前記転送すべきデータの中から転送するデータを選択する選択部と、を有するデータ転送装置が提供される。

【発明の効果】

【０００８】

一態様によれば、ネットワークの帯域不足の抑制を図ることができる。

【図面の簡単な説明】

【０００９】

【図１】データ転送システムの一実施形態を示した図である。

【図２】データ転送システムの一実施形態にかかる帯域の使用状況例を示した図である。

【図３】データ転送システムの一実施形態にかかる全体構成図である。

【図４】ＶＭ情報テーブルの一例を示した図である。

【図５】ディスク情報テーブルの一例を示した図である。

【図６】状態情報テーブルの一例を示した図である。

【図７】帯域情報テーブルの一例を示した図である。

【図８】サイト情報テーブルの一例を示した図である。

【図９】ストレージ制御装置の一実施形態にかかる機能構成図である。

【図１０】状態情報テーブルの一例を示した図である。

【図１１】状態情報テーブルの一例を示した図である。

【図１２】状態情報テーブルの一例を示した図である。

10

20

30

40

50

【図 1 3】状態情報テーブルの一例を示した図である。

【図 1 4】状態情報テーブルの一例を示した図である。

【図 1 5】転送目的の判定処理の一実施形態を示したフローチャートである。

【図 1 6】転送データの選択処理の一実施形態を示したフローチャートである。

【図 1 7】ブロック単位の転送データの選択処理の一実施形態を示したフローチャートである。

【図 1 8】ストレージ制御装置の一実施形態にかかるハードウェア構成図である。

【発明を実施するための形態】

【0010】

以下、本発明の実施形態について添付の図面を参照しながら説明する。なお、本明細書及び図面において、実質的に同一の機能構成を有する構成要素については、同一の符号を付することにより重複した説明を省く。

【0011】

[システムの全体説明]

まず、データ転送システムの一実施形態について、図 1 ~ 図 3 を参照しながら説明する。図 1 ~ 図 3 は、データ転送システムの一実施形態を説明するための図である。

【0012】

本実施形態にかかるデータ転送システム 1 は、サーバ、ストレージ、OS、ミドルウェアなどのリソースをサービスとしてネットワーク経由で提供するクラウドコンピューティングのシステム構成上で実現される。本実施形態にかかるデータ転送システム 1 では、サービス提供者側のデータセンターは、データセンターに予め用意されているサーバ、ストレージ、OS、ミドルウェアなどのリソースを必要に応じて利用者側のクライアント機器に提供する。ただし、本実施形態にかかるデータ転送システム 1 は、クラウドコンピューティングのシステム構成上で実現されるものに限られない。本実施形態にかかるデータ転送システム 1 は、ネットワークを介して接続された 2 つの記憶装置を含み、その 2 つの記憶装置の一方の記憶装置から他方の記憶装置へデータを転送するシステムであれば、どのようなシステム構成であってもよい。以下では、記憶装置の一例としてデータセンター内のストレージ装置を挙げて説明する。

【0013】

[ネットワークを介したストレージ装置間のミラーリング]

データ転送システム 1 の一実施形態にかかるミラーリング（データ転送）について、図 1 を参照しながら説明する。本実施形態にかかるデータ転送システム 1 では、データセンター A とデータセンター B とが、WAN (Wide Area Network) 10 を介して接続されている。なお、WAN 10 はストレージ装置間を接続するネットワークの一例であり、ネットワークはこれに限られず、LAN (Local Area Network) 等の専用線であってもよいし、インターネット等であってもよい。

【0014】

データセンター A にて仮想化して生成された仮想ディスクのデータは、VM 11 により使用される。本実施形態におけるミラーリングでは、WAN 10 の帯域を使用して仮想ディスクのデータを転送する。図 1 では、データセンター A の仮想ディスクのデータが、WAN 10 の帯域を使用してデータセンター B へ転送されている。このように、本実施形態におけるミラーリングは、複数のデータセンターにまったく同じデータを書き込む方式の遠隔的なミラーリングに好適である。

【0015】

ミラーリングの目的には、災害対策 DR 等のためのバックアップ目的や VM 移行の準備目的がある。

【0016】

災害対策 DR のためのバックアップ目的のミラーリングは、例えば、データセンター A が震災等で利用できなくなった場合に、データセンター B にバックアップされているデータを使用してデータセンター A のサービスを復旧させる場合を想定して行われる。

【 0 0 1 7 】

VM移行の準備目的のミラーリングについて説明する。データセンターの利用の拡大に伴い、例えば数百～数千のVM11を管理する必要が生じる。一方、1つのデータセンターにおけるサーバ装置、ストレージ装置等のリソースには物理的に限界がある。このため、管理すべきVM11の増大に伴い、利用者側からのサービス利用の要求に対応できなくなる可能性がある。そこで、複数のデータセンターを連携させて、一のデータセンターにて物理的なリソースの不足が発生した場合にも他のデータセンターの物理的なリソースを利用することで、サービス利用の要求に対応できるシステムが必要である。そのため、リソースの不足が発生した一のデータセンターからリソース供給能力に余力がある他のデータセンターにVM11を移行してVM11の処理の負荷を広域的に分散し、サービス利用の需要拡大に対応できるようにする。そこで、VM移行の準備目的のミラーリングでは、災害対策DRのためのミラーリングとは別に、広域負荷分散のためのVM11の移行に備えて、予めVM11の仮想ディスクを他のデータセンターにコピーする。

10

【 0 0 1 8 】

VM移行の準備目的のミラーリングでは、数10GB～数100GBの仮想ディスクを、WAN10を経由してコピーするため、ミラーリングが完了するまでに多くの時間を要する。よって、VM11の移行が必要になってから仮想ディスクのミラーリングを開始したのでは、実際にVM11を移行できるまでに相当の待ち時間が生じてしまう。そこで、VM11の移行要求が発生する前にVM移行の準備目的のミラーリングが行われることが好ましい。

20

【 0 0 1 9 】

ところが、図2に示されるように、災害対策DR等のバックアップを目的としたミラーリングも、VM移行の準備を目的としたミラーリングも、データセンター間のWAN10を経由して実行される。データセンターは数千のVM11を管理し、各VM11は、例えばシステム用とデータ用の2つ以上の仮想ディスクを使用することが多い。そのため、ミラーリングする仮想ディスクの数はVM11の数よりも多くなる。よって、図2に示されるように、WAN10の帯域が不足する事態が起こり得る。WAN10の帯域が不足すると、バックアップ目的及びVM移行の準備目的によらず、実行されているすべてのミラーリングの処理が遅くなる。これにより、ミラーリング元のデータセンターAにてデータが更新されてからミラーリング先のデータセンターBにそのデータの更新を反映し終えるまでの時間が増加する。

30

【 0 0 2 0 】

ミラーリングの即時性の要求は、ミラーリングの目的によって異なる。例えば、災害対策DRのバックアップを目的としたミラーリングでは、ミラーリング元のデータセンターAで生じたデータの更新は、即時的にミラーリング先のデータセンターBに反映されることが好ましい。

【 0 0 2 1 】

一方、VM移行の準備を目的としたミラーリングでは、ミラーリング元のデータセンターAで生じたデータの更新が即時的にミラーリング先のデータセンターBに反映される必要性は低い。つまり、実際にVM11の移行が実施されるまでに、ミラーリングが終了していればよい。このように、災害対策DR等のバックアップを目的としたミラーリングでは即時性の要求は高く、VM移行の準備を目的としたミラーリングでは、災害対策DR等のバックアップを目的としたミラーリングよりも即時性の要求は低いといえる。

40

【 0 0 2 2 】

そこで、本実施形態では、WANの帯域不足を考慮して、ミラーリングの目的を判定する。そして、判定されたミラーリングの目的に応じて即時性の要求が相対的に低いミラーリング目的（例えば、VM移行の準備を目的）のデータ転送については、すべてのデータを転送せず、転送すべきデータの中から選択されたデータを転送する。一方、即時性の要求が相対的に高いミラーリング目的（例えば、災害対策DR等のバックアップ目的）のデータ転送については、即時的にデータを転送できるようにWAN10の帯域の一部を確保

50

しておくことが好ましい。これにより、データセンター A がいつ災害により利用できなくなってもデータセンター B にデータセンター A の最新データが確保されているため、容易に復旧を行うことができる。

【 0 0 2 3 】

よって、本実施形態では、データセンター間でバックアップを目的としたミラーリングと、VM 移行の準備を目的としたミラーリングとが混在する環境で、バックアップを目的としたミラーリングは、即時性が要求されると判断し、即時的にミラーリングを行う。一方、VM 移行の準備を目的としたミラーリングは、即時性が要求されないと判断し、転送すべきデータの中からデータを選択し、選択されたデータをコピーする。これにより、帯域の不足を抑制し、バックアップを目的としたミラーリングの遅延発生を回避する。

10

【 0 0 2 4 】

以下、転送すべきデータについてミラーリングの目的を判定し、判定された結果が特定の転送目的を示す場合、転送すべきデータの中から転送するデータを選択することが可能なデータ転送システム 1 の一実施形態について説明する。

【 0 0 2 5 】

〔データ転送システムの全体構成〕

一実施形態にかかるデータ転送システム 1 の全体構成について、図 3 を参照しながら説明する。図 3 は、データ転送システムの一実施形態にかかる全体構成図である。本実施形態にかかるデータ転送システム 1 では、WAN 10 を介して接続されたデータセンター A からデータセンター B へ仮想ディスクのデータを転送する場合について説明する。この場合、データセンター A はミラーリング元のデータセンターであり、データセンター B はミラーリング先のデータセンターである。しかしながら、ミラーリングは、必ずしもこれに限られず、データセンター B からデータセンター A へデータ転送を行ってもよい。

20

【 0 0 2 6 】

データセンター A とデータセンター B との構成は同じであるため、以下ではデータセンター A の構成を説明する。ただし、識別のために、データセンター A の物理ストレージ装置 20 a、ストレージ制御装置 30 a 及びミラーリング・コントローラ 31 a と、データセンター B の物理ストレージ装置 20 b、ストレージ制御装置 30 b 及びミラーリング・コントローラ 31 b とを区別して記載する。

【 0 0 2 7 】

データセンター A は、物理サーバ 15、物理ストレージ装置 20 a 及びストレージ制御装置 30 a を有する。データセンター A が運用する物理サーバ 15 及び物理ストレージ装置 20 a の個数は 1 台以上であれば何台でもよい。本実施形態では、複数の物理サーバ 15 及び複数のストレージ装置 20 a を有している。以下、物理サーバを単にサーバといい、物理ストレージ装置を単にストレージ装置という。

30

【 0 0 2 8 】

サーバ 15 上には、複数の VM 11 (仮想サーバ) が構築されている。VM 11 は、仮想マシンソフトウェアによって生成される。本実施形態のデータセンター A には、2 つの VM 11 が保持されたサーバ 15 が 2 つ設置されている。サーバ 15 上の VM 11 は、ストレージ制御装置 30 a を経由してストレージ装置 20 a に接続されている。

40

【 0 0 2 9 】

サーバ 15 は、VM 情報テーブル 16 を有している。VM 情報テーブル 16 には、複数の VM 11 に関する情報が記憶される。図 4 には、VM 情報テーブル 16 の一例が示されている。VM 情報テーブル 16 には、VM 11 の ID 16 a、作成者 ID 16 b、OS 種別 16 c、DR (災害対策) 属性 16 d、割り当て済仮想ディスク ID 16 e が記憶される。例えば、DR 属性 16 d は、災害対策 DR の対象の場合には「TRUE」に設定され、災害対策 DR の対象外の場合には「FALSE」に設定される。

【 0 0 3 0 】

ストレージ制御装置 30 a は、ストレージ装置 20 a を制御し、ストレージ装置 20 a を仮想化して、物理的な構成に拘束されない単位の仮想ディスクをサーバ 15 に提供する

50

。これにより、ストレージ制御装置 30 a は、サーバ 15 に対して、複数のストレージ装置を 1 つのストレージ装置であるように見せたり、1 つのストレージ装置を複数のストレージ装置であるように見せたりすることができる。よって、サーバ 15 は、ストレージ制御装置 30 a に仮想ディスクへの入出力要求（読み取り要求又は書き込み要求）を行えばよく、どのストレージ装置 20 a を利用しているかを認識して動作する必要はない。ストレージ制御装置 30 a は、サーバ 15 からの要求に応じてストレージ装置 20 a への読み取り又は書き込みを実行する。

【0031】

ストレージ制御装置 30 a は、ミラーリング・コントローラ 31 a を有している。ミラーリング・コントローラ 31 a は、データセンター B へ仮想ディスクのデータを転送する（ミラーリング）。ミラーリングされたデータは、災害対策 DR 等に利用される。ミラーリング・コントローラ 31 a は、仮想ディスクの生成に応じて仮想ディスクに保存されたデータや、仮想ディスクへの更新データであって、ミラーリング先のミラーリング・コントローラ 31 b に転送されていないデータを蓄えるバッファ 317 を有する。ミラーリング・コントローラ 31 a は、データ転送装置に相当する。

【0032】

ストレージ制御装置 30 a は、閾値テーブル 35、ディスク情報テーブル 36、状態情報テーブル 37 及び帯域情報テーブル 38 を有する。閾値テーブル 35 には、WAN 10 が帯域不足かを判断する際に使用する閾値や、データの読取頻度を判断する際に使用する閾値が設定されている。

【0033】

ディスク情報テーブル 36 には、仮想ディスクに関する情報が記憶される。図 5 には、ディスク情報テーブル 36 の一例が示されている。ディスク情報テーブル 36 には、仮想ディスクの ID 36 a、容量 36 b、ミラーリング先のデータセンターの ID 36 c、ミラーリング先に作成される仮想ディスク ID 36 d、単位時間当たりの更新（書き込み）回数 36 e 及びミラーリング目的 36 f が記憶される。

【0034】

状態情報テーブル 37 には、仮想ディスクの各ブロックの情報が記憶される。すなわち、仮想ディスクは、複数のブロックに分割されている。図 6 には、状態情報テーブル 37 の一例が示されている。状態情報テーブル 37 には、例えば、仮想ディスクの ID 36 a で識別される仮想ディスクについてブロック毎に、ブロック番号 37 a、更新時刻 37 b、バッファフラグ 37 c 及びリードカウンタ 37 d が記憶される。バッファフラグ 37 c 及びリードカウンタ 37 d については、後述する。

【0035】

帯域情報テーブル 38 には、WAN 10 の帯域に関する情報が記憶される。図 7 には、帯域情報テーブル 38 の一例が示されている。帯域情報テーブル 38 には、WAN 10 を使用するソース提供側のデータセンター ID 38 a（ここではデータセンター A の ID）、データの転送先となるディスティネーションデータセンター ID（ここではデータセンター B の ID）38 b、帯域の状態 38 c、最大帯域 38 d 及び使用帯域 e が記憶される。

【0036】

サイト情報テーブル 42 は、データセンター全体の情報を記憶する。図 8 には、サイト情報テーブル 42 の一例が示されている。サイト情報テーブル 42 には、データセンター ID 42 a、管理用ポータル IP アドレス 42 b、管理者 ID 42 c、運用種別 42 d 及び状態 42 e が記憶される。例えば、運用種別 42 d は、データセンター ID 42 a で識別されるデータセンターがシステム運用を行う場合には「運用系」に設定され、データセンターが災害対策 DR の待機用である場合には「待機系」に設定される。サイト情報テーブル 42 の更新内容は、他のデータセンターが保持するサイト情報テーブル 42 に反映される。

【0037】

[ストレージ制御装置の機能構成]

10

20

30

40

50

次に、一実施形態にかかるストレージ制御装置 30 a の機能構成について、図 9 を参照しながら説明する。図 9 は、ストレージ制御装置の一実施形態にかかる機能構成図である。

【 0 0 3 8 】

本実施形態にかかるストレージ制御装置 30 a は、制御部 32、R/W 要求受付部 33、R/W 実行部 34、閾値テーブル 35、ディスク情報テーブル 36、状態情報テーブル 37 及び帯域情報テーブル 38 を有する。

【 0 0 3 9 】

制御部 32 は、ストレージ制御装置 30 a の全体を制御する。制御部 32 は、サーバ 15 からの入出力要求に応じてストレージ装置 20 a に仮想ディスクを動的に作成したり、削除したりする。制御部 32 は、サーバ 15 からの要求に応じて仮想ディスクへのデータの読み取り及び書き込みを実行する。

10

【 0 0 4 0 】

R/W 要求受付部 33 は、サーバ 15 からの入出力要求を受け付ける。R/W 実行部 34 は、R/W 要求受付部 33 からの入出力要求に応じて、ストレージ装置 20 a の所定の仮想ディスクにデータを書き込んだり、読み取ってサーバ 15 に送ったりする。

【 0 0 4 1 】

ストレージ制御装置 30 a は、閾値テーブル 35、ディスク情報テーブル 36、状態情報テーブル 37 及び帯域情報テーブル 38 を有する。しかし、これらのテーブルは、ストレージ制御装置 30 a 外の記憶領域に記憶されてもよい。その場合にも、閾値テーブル 35、ディスク情報テーブル 36、状態情報テーブル 37 及び帯域情報テーブル 38 は、記憶部 316 により管理される。

20

【 0 0 4 2 】

[ミラーリング・コントローラ 31 a の機能構成]

ミラーリング・コントローラ 31 a は、ストレージ制御装置 30 a に内蔵され、データ転送に関するストレージ制御装置 30 a の拡張機能を有する。しかしながら、これに限らず、ミラーリング・コントローラ 31 a は、ストレージ制御装置 30 a とは別体に構成することもできる。ミラーリング・コントローラ 31 a は、データセンター B のミラーリング・コントローラ 31 b との間でデータ転送を実行する。

【 0 0 4 3 】

30

ミラーリング・コントローラ 31 a は、判定部 311、選択部 312、帯域監視部 313、送信部 314、受信部 315、記憶部 316 及びバッファ 317 を有する。

【 0 0 4 4 】

判定部 311 は、ストレージ装置 20 a に記憶された仮想ディスクのデータの転送目的を判定する。仮想ディスクのデータとは、仮想ディスクが生成された際に書き込まれたデータや、該仮想ディスクを更新した際の更新データが含まれる。また、仮想ディスクのデータとは、仮想ディスクに含まれる全データであってもよいし、仮想ディスクに含まれるデータを分割したブロック毎のデータであってもよい。

【 0 0 4 5 】

選択部 312 は、判定部 311 により特定の転送目的と判定されたデータに関して、ストレージ装置 20 a に記憶された仮想ディスクからデータセンター B のストレージ装置 20 b へコピーするデータを選択する。

40

【 0 0 4 6 】

帯域監視部 313 は、WAN 10 の帯域の使用量等、WAN 10 の状態を監視する。帯域監視部 313 は、定期的（例えば、1 分毎）に WAN 10 の帯域を監視する。その結果は、随時、帯域情報テーブル 38 に記憶される。

【 0 0 4 7 】

送信部 314 は、選択部 312 により選択されたデータをデータセンター B へ送信する。受信部 315 は、WAN 10 を介してデータを受信する。

【 0 0 4 8 】

50

記憶部 316 は、閾値テーブル 35、ディスク情報テーブル 36、状態情報テーブル 37 及び帯域情報テーブル 38 にデータを記憶する。特に、更新頻度記憶部 316a は、ディスク情報テーブル 36 の更新回数（更新頻度に相当）36e を記憶する。読取頻度記憶部 316b は、状態情報テーブル 37 のリードカウンタ 37d（ブロック毎の読取頻度に相当）を記憶する。

【0049】

選択部 312 によりミラーリング対象に選択されなかったデータはデータセンター B に即時的に転送されない。そのため、バッファ 317 は、ミラーリング対象に選択されなかったデータを一時的に蓄える。バッファ 317 に蓄えられたデータに関して、記憶部 316 は、図 6 の状態情報テーブル 37 に示されるように、仮想ディスク ID のブロック単位で、バッファフラグ 37c を「true」に設定する。バッファ 317 に蓄えられていない仮想ディスク ID のブロックについては、記憶部 316 は、バッファフラグ 37c を「false」に設定する。

10

【0050】

[ストレージ制御装置の動作]

次に、ストレージ制御装置 30a の動作について説明する。ストレージ制御装置 30a は、入出力要求（読み取り要求又は書き込み要求）に対して次の処理を行う。

【0051】

（読み取り要求）

R/W要求受付部 33 は、サーバ 15 から所定の仮想ディスクへの読み取り要求を受け付けると、ストレージ装置 20a から所定の仮想ディスクのデータを読み出し、読み出したデータをサーバ 15 に送る。

20

【0052】

（書き込み要求）

R/W要求受付部 33 は、サーバ 15 から所定の仮想ディスクへの書き込み要求を受け付けると、ストレージ装置 20a の所定の仮想ディスクにデータを書き込み、サーバ 15 に更新完了通知を送る。また、仮想ディスクがミラーリングの対象となっている場合、R/W実行部 34 は、ミラーリング・コントローラ 31a に更新通知を送り、ミラーリング・コントローラ 31a からの処理完了通知（ACK Acknowledgement：確認応答）を待機する。

30

【0053】

（リモートのミラーリング・コントローラからの更新要求）

リモートのミラーリング・コントローラ（ここでは、ミラーリング・コントローラ 31b）からのデータの更新要求は、ミラーリング・コントローラ 31a の受信部 315 が受信する。R/W実行部 34 は、受信部 315 が受信した更新要求に含まれている更新データをストレージ装置 20a に書き込む。送信部 314 は、リモートのミラーリング・コントローラ 31b へ処理完了（ACK）を通知する。

【0054】

[ミラーリング・コントローラの動作]

次に、ミラーリング・コントローラ 31a の動作について説明する。ミラーリング・コントローラ 31a は、ストレージ制御装置 30a が実現している仮想ディスク内のデータの更新の通知を受け付ける。

40

【0055】

WAN10 の帯域が十分な場合、送信部 314 は、ミラーリング・コントローラ 31b へ更新されたデータを転送し、ミラーリング・コントローラ 31b からの更新完了通知（ACK）を待機する。ミラーリング・コントローラ 31a は、ミラーリング・コントローラ 31b から更新完了通知（ACK）を受信した場合、ストレージ制御装置 30a へ処理完了（ACK）を通知する。

【0056】

WAN10 の帯域が不十分な場合、VM移行の準備目的のデータに関しては、記憶部 3

50

16は、更新データをバッファ317に格納する。つまり、ミラーリング先のミラーリング・コントローラ31bに転送しない。記憶部316は、ストレージ制御装置20aへ処理完了(ACK)を通知する。

【0057】

記憶部316は、状態情報テーブル37のバッファフラグ37cに、データがバッファリングされていることを示すフラグとして「true」をセットし、リードカウンタ37dに「0」をセットする。

【0058】

例えば、図10に示したように、仮想ディスクID36aのブロック番号37a「00000002」に書き込みが発生された場合について具体的に説明する。ブロック番号37a「00000002」のデータが2012/10/02 13:31:45に更新(書き込み)されたとする。このとき、WAN10の帯域が不十分な状態であり、更新されたデータがVM移行の準備目的のデータであるため、ミラーリング先のミラーリング・コントローラ31bに転送されずにバッファ317に保存されたとする。そのとき、記憶部316は、ブロック番号37a「00000002」のバッファフラグ37cに「true」を設定し、リードカウンタ37dに「0」を設定し、更新時刻37bに「2012/10/02 13:31:45」を記憶する。なお、リードカウンタ37dは、予め定められた一定時間ごとに「0」にリセットされてもよい。

【0059】

なお、WAN10の帯域が不十分な場合であっても、バックアップ目的のデータに関しては、送信部314は、更新されたデータをミラーリング・コントローラ31bへ転送し、WAN10の帯域が十分な場合と同様の更新完了通知処理を行う。

【0060】

選択部312は、一定時間毎にWAN10の帯域が許す範囲内で、バッファ317に蓄積されている更新データの中から転送するデータを選択する。具体的には、選択部312は、バッファフラグ37cが「true」になっていて、リードカウンタ37dで示される読取回数が大きいデータから順に優先的に選択する。このようにリードカウンタ37dは、バッファ317に記憶されたデータの読取頻度を示す。送信部314は、選択されたデータをミラーリング・コントローラ31bに転送する。受信部315は、ミラーリング・コントローラ31bからの更新完了通知(ACK)を待機する。更新完了通知(ACK)を受信した場合、記憶部316は、転送データに対する状態情報テーブル37のブロック番号37aのバッファフラグ37cに「false」を設定する。また、リードカウンタ37dに「-(バッファされていないことを示す(例:NULL))」を設定する。

【0061】

例えば、WAN10の帯域の使用量に基づき、バッファ317に蓄積されている更新データのうちの1個のブロックの転送ができる状況と判断されたとする。この場合、図11に示したように、状態情報テーブル37のバッファフラグ37cが「true」のブロックのうちで、リードカウンタ37dの値が最大「12」のブロック番号37a「00000003」のデータが選択される。選択されたデータは、ミラーリング・コントローラ31bに転送される。転送されると、バッファ317内のデータは転送済であってバッファ317から消去してよい。よって、ブロック番号37a「00000003」のバッファフラグ37cが「false」に、リードカウンタ37dが「-」に設定される。

【0062】

(データが読み出された場合)

次に、バッファ317内のデータと同一データがストレージ装置20aから読み出された場合のミラーリング・コントローラ31aの動作について説明する。その場合、読み出されたデータに対する状態情報テーブル37のブロック番号37aのバッファフラグ37cが「true」のデータであれば、記憶部316は、リードカウンタ37dを「1」増加させる。図12に示したように、データが転送されていない状態で、バッファ317内の例えばブロック番号37a「00000002」のデータがR/W実行部34により

10

20

30

40

50

読み出された場合、記憶部 316 は、リードカウンタ 37d を「0」から「1」に一増加させる。読み出されたデータはサーバ 15 に送られる。

【0063】

更に、図 13 に示したように、データが転送されていない状態で、再度、ブロック番号 37a「0000 0002」のデータが R/W 実行部 34 により読み出された場合、リードカウンタ 37d を「1」から「2」に一増加させる。読み出されたデータはサーバ 15 に送られる。

【0064】

(データが更新された場合)

次に、バッファ 317 内のデータと同一データがストレージ装置 20a で更新された場合のミラーリング・コントローラ 31a の動作について説明する。ストレージ装置 20a に記憶された仮想ディスク内のデータが更新された場合、記憶部 316 は、図 14 のブロック番号 37a「0000 0005」に示したように、バッファ 317 に蓄積された同一のデータも更新する。このとき、更新されたデータは一度も読み出されていない。よって、記憶部 316 は、リードカウンタ 37d を「0」にリセットする。

【0065】

なお、一定時間経過後にリードカウンタ 37d が閾値を超えていたら、ミラーリング先のストレージ装置に反映すべきデータとして選択してもよい。この場合、リードカウンタ 37d が閾値を超えていなければ、データを転送せずバッファ 317 への蓄積を継続する。なお、本実施形態では、仮想ディスクの書き込み及び読み取りは仮想ディスクを複数に分割したブロック単位で行われるが、必ずしもこれに限らず、仮想ディスク単位で行われてもよい。

【0066】

このようにして、WAN の帯域が不十分な場合、VM 移行の準備目的のデータに関しては、更新データは即時的にストレージ制御装置 30b に転送されず、バッファ 317 に格納される。また、バッファ 317 内のデータの状態が状態情報テーブル 37 に蓄積される。

【0067】

以下では、ミラーリングの目的を判定する処理、バッファ 317 内に蓄積されたデータから転送データを選択する処理の順に、ミラーリング・コントローラ 31a の動作について説明する。

【0068】

[データ転送(ミラーリング)目的の判定]

本実施形態にかかるデータ転送目的の判定処理について、図 15 を参照しながら説明する。図 15 は、データ転送目的の判定処理の一実施形態を示したフローチャートである。

【0069】

ミラーリング・コントローラ 31a は、運用管理者又は利用者側のクライアント機器からの仮想ディスクのミラーリング指示に応じて、ミラーリングの目的を判定する。まず、R/W 要求受付部 33 は、VM 11 の要求を反映してサーバ 15 から要求されるミラーリング指示のための入力を待機する(ステップ S100)。R/W 要求受付部 33 は、ミラーリング指示のための入力情報として、ミラーリング元(ここではデータセンター A)の仮想ディスク ID 及びミラーリング先(ここではデータセンター B)のデータセンター ID を受け付ける。加えて、R/W 要求受付部 33 は、ミラーリング先の仮想ディスク ID 及びミラーリングの目的の情報を受け付けてもよい。ただし、ミラーリング先の仮想ディスク ID 及びミラーリングの目的の情報は必ずしも指定される必要はなく、入力情報に含まれない場合もある。

【0070】

次に、判定部 311 は、入力情報にミラーリング元の仮想ディスク ID 及びミラーリング先のデータセンター ID が指定されているかを判定する(ステップ S102)。ミラーリング元の仮想ディスク ID 及びミラーリング先のデータセンター ID が指定されてい

10

20

30

40

50

いと判定された場合、判定部 3 1 1 は、エラーを出力し、ストレージ制御装置 3 0 a に通知する（ステップ S 1 0 4 ）。

【 0 0 7 1 】

ステップ S 1 0 2 にて、ミラーリング元の仮想ディスク I D 及びミラーリング先のデータセンター I D が指定されていると判定された場合、判定部 3 1 1 は、入力情報にミラーリング先の仮想ディスク I D が指定されているかを判定する（ステップ S 1 0 6 ）。ミラーリング先の仮想ディスク I D が指定されていると判定された場合、制御部 3 2 は、そのままステップ S 1 1 0 に進む。一方、ミラーリング先の仮想ディスク I D が指定されていないと判定された場合、判定部 3 1 1 は、ミラーリング先のストレージ制御装置 3 0 b に対して、ミラーリング先となる仮想ディスクの作成を依頼する。ストレージ制御装置 3 0 b は、ミラーリング先のストレージ装置 2 0 b に、ミラーリング先となる仮想ディスクを作成し、その仮想ディスクの I D を通知する。判定部 3 1 1 は、この通知を受信することにより、ミラーリング先の仮想ディスクの I D を取得する（ステップ S 1 0 8 ）。

10

【 0 0 7 2 】

次に、判定部 3 1 1 は、入力情報にミラーリングの目的が指定されているかを判定する（ステップ S 1 1 0 ）。ミラーリングの目的が指定されていると判定された場合、判定部 3 1 1 は、対象となる仮想ディスクのミラーリング目的は指定された目的であると判定する。記憶部 3 1 6 は、指定されたミラーリングの目的を、ディスク情報テーブル 3 6 に記憶する（ステップ S 1 1 2 ）。対象となる仮想ディスクのミラーリング目的が、バックアップ目的の場合、記憶部 3 1 6 は、図 5 のミラーリング目的 3 6 f に「バックアップ」を記憶する。V M 1 1 の移行目的である場合、記憶部 3 1 6 は、図 5 のミラーリング目的 3 6 f に「V M 移行準備」を記憶し（ステップ S 1 1 2 ）、スタートに戻る。

20

【 0 0 7 3 】

ステップ S 1 1 0 にてミラーリングの目的が指定されていないと判定された場合、判定部 3 1 1 は、ミラーリング先のデータセンター I D で識別されるデータセンターが災害対策 D R の待機系であるかを判定する（ステップ S 1 1 4 ）。災害対策 D R の待機系であると判定された場合、判定部 3 1 1 は、対象となる仮想ディスクのミラーリング目的はバックアップ目的であると判定する。記憶部 3 1 6 は、判定されたミラーリングの目的がバックアップ目的であるとディスク情報テーブル 3 6 に記憶し（ステップ S 1 1 6 ）、スタートに戻る。

30

【 0 0 7 4 】

例えば、判定部 3 1 1 は、図 8 のサイト情報テーブル 4 2 に基づき、ミラーリング先のデータセンター I D 4 2 a が「B」の運用種別 4 2 d が「待機系」と判断し、データセンター B が災害対策 D R の待機系であると判定する。記憶部 3 1 6 は、図 5 のミラーリング目的 3 6 f に「バックアップ」を記憶する。

【 0 0 7 5 】

ステップ S 1 1 8 にてミラーリング先のストレージ装置 I D で識別されるデータセンターが災害対策 D R の待機系でないと判定された場合、判定部 3 1 1 は、仮想ディスクの割り当て先の V M 1 1 の D R 属性が「true」であるかを判定する（ステップ S 1 1 8 ）。仮想ディスクの割り当て先の V M 1 1 の D R 属性が「true」であると判定された場合、判定部 3 1 1 は、対象となる仮想ディスクのミラーリング目的はバックアップ目的であると判定する。記憶部 3 1 6 は、判定されたミラーリングの目的がバックアップ目的であるとディスク情報テーブル 3 6 に記憶し（ステップ S 1 2 0 ）、スタートに戻る。

40

【 0 0 7 6 】

ステップ S 1 1 8 にて仮想ディスクの割り当て先の V M 1 1 の D R 属性が「TRUE」でないと判定された場合、判定部 3 1 1 は、対象となる仮想ディスクのミラーリング目的は V M 移行の準備が目的であると判定する。記憶部 3 1 6 は、判定されたミラーリングの V M 移行の準備目的であるとディスク情報テーブル 3 6 に記憶し（ステップ S 1 2 2 ）、スタートに戻る。

【 0 0 7 7 】

50

例えば、判定部 3 1 1 は、図 4 の V M 情報テーブル 1 6 に基づき、仮想ディスクの割り当て先の V M 1 1 の I D 1 6 a の D R 属性 1 6 d が「T R U E」であるかを判定し、「T R U E」の場合、データセンター B が災害対策 D R の待機系であると判定する。その場合、記憶部 3 1 6 は、図 5 のミラーリング目的 3 6 f に「バックアップ」を記憶する。

【 0 0 7 8 】

以上に説明したように、本実施形態にかかるストレージ制御装置 3 0 によれば、ミラーリング目的が指定されていない場合であっても、判定部 3 1 1 は、ミラーリング先のデータセンターが災害対策 D R の待機用として機能するかどうかでミラーリング目的を判断する。例えば、データセンター A がシステム運用に使用され、データセンター B が災害時の待機用であること等がデータセンター毎にサイト情報テーブル 4 2 に予め設定されている。また、V M 情報テーブル 1 6 には V M 1 1 毎の D R 属性が設定されている。よって、本実施形態にかかるストレージ制御装置 3 0 では、これらのテーブルの情報に基づき、ミラーリング先のデータセンター B が災害対策 D R の待機系であるか、V M 1 1 毎の D R 属性が「T R U E」であるかによりミラーリング目的を判定することができる。

【 0 0 7 9 】

〔 転送データの選択 〕

次に、本実施形態にかかる転送データの選択処理について、図 1 6 を参照しながら説明する。図 1 6 は、転送データの選択処理の一実施形態を示したフローチャートである。

【 0 0 8 0 】

（ 仮想ディスクの選択 ）

帯域監視部 3 1 3 は、データセンター A、B 間の W A N 1 0 の帯域を定期的（例えば、1 分毎）に監視する（ステップ S 2 0 0）。次に、判定部 3 1 1 は、帯域の使用量が閾値を超えているかを判定する（ステップ S 2 0 2）。

【 0 0 8 1 】

例えば、帯域不足の発生が予測される場合として、帯域監視部 3 1 3 は、帯域の使用量が閾値テーブル 3 5 に記憶された閾値を超えているかを判定する。本実施形態では、帯域監視部 3 1 3 は、帯域の使用量が、図 7 の帯域情報テーブル 3 8 中のソースデータセンター I D が「A」の最大帯域 3 8 d の 8 0 % を超えているかを判定する。

【 0 0 8 2 】

帯域の使用量が閾値を超えていると判定された場合、判定部 3 1 1 は、ミラーリング対象の仮想ディスクが、V M 移行の準備目的のミラーリングであるかを判定する（ステップ S 2 0 4）。V M 移行の準備目的のミラーリングでないと判定された場合、判定部 3 1 1 は、災害対策 D R 等のバックアップ目的のミラーリングと判定し、スタートに戻る。

【 0 0 8 3 】

この結果、災害対策 D R 等のバックアップ目的のミラーリングは、即時性を考慮して継続して行われる。つまり、前述の W A N 1 0 の帯域が十分な場合に行われるミラーリングが実行される。これにより、災害対策 D R 等のバックアップ目的のミラーリングでは、仮想ディスクのデータを即時的にミラーリング先のストレージ装置 2 0 b にバックアップすることができる。

【 0 0 8 4 】

ステップ S 2 0 4 にて V M 1 1 の移行準備目的のミラーリングであると判定された場合、選択部 3 1 2 は、V M 移行の準備目的でミラーリングされている仮想ディスクで、且つ、転送すべきデータを選択してミラーリングする対象となっていないものの中で、単位時間当たりのデータの更新頻度が最大のものを選択する（ステップ S 2 0 6）。ここで、更新頻度が最大のものが複数存在する場合、選択部 3 1 2 は、任意の 1 つを選択する。

【 0 0 8 5 】

具体的には、選択部 3 1 2 は、ディスク情報テーブル 3 6 の単位時間当たりの更新回数 3 6 e に基づき、どの仮想ディスクをミラーリング対象とするかの選択を行う。選択部 3 1 2 は、更新頻度が低い仮想ディスクから順にミラーリング対象とする。例えば、図 5 のディスク情報テーブル 3 6 の場合、選択部 3 1 2 は、即時性が要求されない V M 移行準備

10

20

30

40

50

の目的の仮想ディスクのうち、更新回数 3 6 e が「1 6」の仮想ディスク（仮想ディスク ID v d i s k - A - 0 3 0 0）を選択する。

【0086】

更新頻度が高い仮想ディスクは、ミラーリング先のミラーリング・コントローラ 3 1 b への頻繁なデータ転送を引き起こす。これを回避するために、更新回数が高い仮想ディスク程ミラーリングの優先順位を下げることで、W A N 1 0 の帯域圧迫を抑制する。

【0087】

次に、選択部 3 1 2 は、転送するデータを選択してミラーリングする対象とする（ステップ S 2 0 8）。仮想ディスク全体を転送する場合、選択部 3 1 2 は、ステップ S 2 0 6 で選択された仮想ディスクの全データをミラーリング対象とする。

10

【0088】

（仮想ディスクの一部の選択）

一方、仮想ディスクの一部を転送する場合、選択された仮想ディスクのうち、仮想ディスクを分割した複数のブロックのうちから転送するブロックを選択する。

【0089】

ステップ S 2 0 8 の転送すべきデータの中から転送するデータを選択する処理の詳細について、図 1 7 を参照しながら説明する。図 1 7 は、ブロック単位の転送データの選択処理を示したフローチャートである。

【0090】

まず、ステップ S 3 0 0 にて、選択部 3 1 2 は、バッファ 3 1 7 に蓄えられている更新データがあるかを判定する。バッファ 3 1 7 に蓄えられている更新データがないと判定された場合、選択部 3 1 2 は、一定時間（例えば、1 8 0 秒）待ち（ステップ S 3 0 2）、スタートへ戻る。

20

【0091】

一方、バッファ 3 1 7 に蓄えられている更新データがあると判定された場合、選択部 3 1 2 は、W A N 1 0 の現在の利用可能帯域から転送可能な更新データの個数 N を計算する（ステップ S 3 0 4）。次に、選択部 3 1 2 は、バッファ 3 1 7 に蓄えられている更新データのうち、リードカウンタ 3 7 d の値が大きい仮想ディスクのブロックから順番に N 個のデータをブロック単位で選択する（ステップ S 3 0 6）。

【0092】

30

次に、送信部 3 1 4 は、選択されたデータをミラーリング先のミラーリング・コントローラ 3 1 b に送信し、（ステップ S 3 0 8）。ミラーリング先のストレージ装置 2 0 b のデータ更新を要求する。

【0093】

以上のように、仮想ディスクの一部をブロック単位で転送する場合、ミラーリング対象とするかの判定には読取回数を用いる。そして、読取回数が高いものから優先的にコピーする。これにより、よく利用されるデータは、ミラーリング先のストレージ装置に記憶されるタイミングが早くなり、あまり利用されないデータは、ミラーリング先のストレージ装置へ記憶されるタイミングが遅くなる。

【0094】

40

図 1 6 のステップ S 2 0 2 に戻って、帯域の使用量が閾値テーブル 3 5 に設定された閾値を超えていないと判定された場合、判定部 3 1 1 は、転送するデータを選択してミラーリングしている仮想ディスクがあるかを判定する（ステップ S 2 1 0）。なお、帯域の使用量が閾値を超えていないと判定された場合とは、W A N 1 0 の帯域が不足していない場合や、発生していた W A N 1 0 の帯域不足が解消した場合等を含む。

【0095】

ステップ S 2 1 0 にてデータを選択してミラーリングしている仮想ディスクがないと判定された場合、スタートに戻る。一方、ステップ S 2 1 0 にてデータを選択してミラーリングしている仮想ディスクがあると判定された場合、選択部 3 1 2 は、データを選択してミラーリングする対象となっているものの中で、単位時間当たりのデータの更新頻度が最

50

小のものを選択する（ステップS212）。ここで、更新頻度が最小のものが複数存在する場合、選択部312は、任意の1つを選択する。次に、VM移行の準備目的のミラーリングを、データの更新を即時的に行うミラーリングの対象に戻す（ステップS214）。これにより、次にステップS202にて帯域不足と判定されるまで、VM移行の準備目的のミラーリングもバックアップ目的のミラーリングと同様に、即時的に転送が実行される。

【0096】

以上に説明したように、本実施形態にかかる選択処理によれば、VM移行の準備目的のミラーリングでは、仮想ディスクのミラーリングを即時的に行わず、転送すべきデータの中から転送するデータを選択して、選択されたデータをミラーリングする。このようにして、VM移行の準備目的のミラーリングではデータの転送頻度を減らすことで、WAN10の帯域の消費を抑制することができる。

10

【0097】

また、上記選択処理では、ミラーリング先のストレージ装置への頻繁なデータの更新がWAN10の帯域を圧迫する。このため、データの更新頻度が低い仮想ディスクから順に転送されるように仮想ディスクのデータを選択する。この結果、よく書き込まれる仮想ディスクのデータ程ミラーリング先のミラーリング・コントローラへの転送が抑制される。これにより、ミラーリング先のストレージ装置への頻繁なデータ更新を抑制し、WAN10の帯域の消費を抑制する。

【0098】

20

加えて、選択された仮想ディスクのうちの一部をブロック単位で転送する場合、相対的に利用頻度の高いブロックのデータから優先してミラーリング先のストレージ装置へ反映される。これにより、利用価値が高いと想定されるブロックのデータ程ミラーリング先のストレージ装置へのデータの更新タイミングを早めることができる。よって、即時性が要求されないVM移行の準備目的のデータの中でも相対的に即時性が高いと想定されるデータを優先してミラーリングすることができる。このように、本実施形態にかかるデータ転送システム1によれば、データの更新頻度や読取頻度に応じて転送データが選択されるため、データが運用中に動的に変化した場合にもこれに応じて最適な転送データを選択することができる。

【0099】

30

従って、頻繁に更新されるけれどもあまり利用されないデータについては更新データの転送を自動的に抑制してネットワーク帯域の消費を抑え、利用されることが多いデータの更新データを自動的に選択して転送することができる。

【0100】

なお、上記実施形態にかかる判定処理では、判定部311により判定された特定の転送目的としてVM移行の準備目的を例に挙げたが、これに限られない。即時性が要求されない転送目的であれば、特定の転送目的となり得る。例えば、データセンター等の設備が古くなったのでそのデータセンター内で運用されていたストレージ装置を別に移行する場合、ストレージ装置内のデータを他のストレージ装置に移行する必要がある。この場合にも、即時性が要求されない転送目的として特定の転送目的となり得る。

40

【0101】

一方、例えば、システム管理に用いるリソースのステータス等のデータは、定期的にストレージ装置に書き込まれるとともに、管理者側の機器のディスプレイに表示するために定期的に読み出される可能性が高い。このようなデータは、即時性が要求される転送目的を有するデータとして特定の転送目的から除外され、バックアップ目的と同様に即時性が要求されるミラーリング目的とされる。

【0102】

また、上記本実施形態にかかる選択処理では、WANの帯域が不足した場合であって、VM移行の準備目的のミラーリングの場合には、仮想ディスクのミラーリングを即時的に行わず、ミラーリングすべき仮想ディスクから転送するデータを選択した。しかしながら

50

、WANの帯域が不足しているかにかかわらず、随時、本実施形態にかかる選択処理を実行してもよい。これにより、VM移行の準備目的のミラーリングではデータの転送頻度を減らすことで、WAN10の帯域の消費をさらに抑制することができる。

【0103】

また、上記実施形態におけるミラーリングでは、ネットワークの帯域を使用して2台のストレージ装置20a、20bのうち、ストレージ装置20aからストレージ装置20bへ仮想ディスク内のミラーリングが行われた。しかしながら、ストレージ装置20bからストレージ装置20aへ仮想ディスク内のミラーリングが行われてもよいし、図示しない他のストレージ装置からストレージ装置20aへ仮想ディスクのミラーリングが行われてもよい。

10

【0104】

(ハードウェア構成例)

最後に、ストレージ制御装置30aのハードウェア構成例について簡単に説明する。図18は、本実施形態にかかるストレージ制御装置のハードウェア構成例を示す図である。

【0105】

図18に示すように、ストレージ制御装置30aは、入力装置101、表示装置102、外部I/F103、RAM(Random Access Memory)104、ROM(Read Only Memory)105、CPU(Central Processing Unit)106、通信I/F107、HDD(Hard Disk Drive)108及び入出力I/F109を備え、それぞれがバスBで相互に接続されている。

20

【0106】

入力装置101は、キーボードやマウスなどを含み、ストレージ制御装置30aに各操作を入力するのに用いられる。表示装置102は、ディスプレイなどを含み、ストレージ制御装置30aの管理者にシステム運用結果等を表示する。

【0107】

通信I/F107は、ストレージ制御装置30aに内蔵されたミラーリング・コントローラ31aをWAN10に接続するインタフェースである。これにより、ミラーリング・コントローラ31aは、通信I/F107を介して、他のミラーリング・コントローラと遠隔的なミラーリングを行うことができる。

【0108】

HDD108は、プログラムやデータを格納している不揮発性の記憶装置である。格納されるプログラムやデータには、装置全体を制御する基本ソフトウェアであるOS(Operating System)、及びOS上において各種機能を提供するアプリケーションソフトウェアなどがある。また、HDD108は、上記実施形態のミラーリング目的の判定処理や転送データの選択処理を行うためにCPU106により実行されるプログラムを格納する。

30

【0109】

外部I/F103は、外部装置とのインタフェースである。外部装置には、記録媒体103aなどがある。ストレージ制御装置30aは、外部I/F103を介して、記録媒体103aの読み取り及び/又は書き込みを行うことができる。記録媒体103aとしては、CD(Compact Disk)、及びDVD(Digital Versatile Disk)、ならびに、SDメモリカード(SD Memory card)やUSBメモリ(Universal Serial Bus memory)等が挙げられる。

40

【0110】

ROM105は、不揮発性の半導体メモリ(記憶装置)であり、起動時に実行されるBIOS(Basic Input/Output System)、OS設定、及びネットワーク設定などのプログラムやデータが格納されている。RAM104は、プログラムやデータを一時保持する揮発性の半導体メモリ(記憶装置)である。CPU106は、上記記憶装置(例えば「HDD」や「ROM」など)から、プログラムやデータをRAM上に読み出し、処理を実行することで、装置全体の制御や搭載機能を実現する演算装置である。

【0111】

50

入出力 I / F 1 0 9 は、サーバ 1 5 とのインタフェースである。これにより、ストレージ制御装置 3 0 a は、入出力 I / F 1 0 9 を介して、サーバ 1 5 からのストレージ装置 2 0 a への読み取り要求や書き込み要求を受け付けることができる。

【 0 1 1 2 】

制御部 3 2、R / W 要求受付部 3 3、R / W 実行部 3 4 及びミラーリング・コントローラ 3 1 0 a の各部は、H D D 1 0 8 にインストールされたプログラムが C P U 1 0 6 に実行させる処理により実現される。閾値テーブル 3 5、ディスク情報テーブル 3 6、状態情報テーブル 3 7 及び帯域情報テーブル 3 8 は、例えば、R A M 1 0 4、H D D 1 0 8、又はストレージ制御装置 3 0 a に W A N 1 0 を介して接続される記憶装置を用いて実現可能である。

10

【 0 1 1 3 】

以上のように、本実施形態にかかるストレージ制御装置 3 0 a は、上記ハードウェア構成により上記実施形態のミラーリング目的の判定処理や転送データの選択処理を提供することができる。

【 0 1 1 4 】

以上、データ転送装置、データ転送システム、データ転送方法及びプログラムを実施形態により説明したが、本発明は上記実施形態に限定されるものではなく、本発明の範囲内で種々の変形及び改良が可能である。

【 0 1 1 5 】

例えば、以上の説明では、仮想ディスク内のデータの転送を例に挙げて説明したが、本発明にかかるデータ転送システムは、物理ディスク内のデータの転送にも適用可能である。

20

【 0 1 1 6 】

以上の説明に関し、更に以下の項を開示する。

(付記 1)

記憶装置間にてネットワークを介してデータの転送を行うデータ転送装置であって、前記記憶装置内の転送すべきデータの転送目的を判定する判定部と、前記判定部により判定された結果が特定の転送目的を示す場合、前記転送すべきデータの中から転送するデータを選択する選択部と、を有するデータ転送装置。

(付記 2)

前記ネットワークの帯域を監視する帯域監視部を更に有し、前記選択部は、前記帯域監視部により監視される前記ネットワークの帯域の使用量に応じて、前記転送すべきデータの中から転送するデータを選択する付記 1 に記載のデータ転送装置。

30

(付記 3)

前記記憶装置に記憶されたデータの更新頻度を記憶する更新頻度記憶部を更に有し、前記選択部は、前記特定の転送目的のデータに関して前記更新頻度記憶部が記憶する更新頻度に基づいて、前記転送すべきデータの中から転送するデータを選択する付記 1 又は 2 に記載のデータ転送装置。

(付記 4)

前記記憶装置に記憶されたデータは、複数のブロックに分割され、前記複数のブロックに分割されて前記記憶装置に記憶された各々のデータの読取頻度を記憶する読取頻度記憶部を更に有し、前記選択部は、前記特定の転送目的のデータに関して前記読取頻度記憶部が記憶するブロック毎の読取頻度に基づいて、前記転送すべきデータの中から転送するデータのブロックを選択する付記 3 に記載のデータ転送装置。

40

(付記 5)

前記選択部は、前記判定部により判定された結果が、即時性が要求されないデータ転送の目的を示す場合に、前記転送すべきデータの中から転送するデータを選択する付記 1 ~ 4 のいずれか一項に記載のデータ転送装置。

50

(付記 6)

複数の記憶装置と、該複数の記憶装置間にてネットワークを介してデータの転送を行うデータ転送装置と、を含むデータ転送システムであって、

前記データ転送装置は、

前記複数の記憶装置のいずれかの記憶装置内の転送すべきデータの転送目的を判定する判定部と、

前記判定部により判定された結果が特定の転送目的を示す場合、前記転送すべきデータの中から転送するデータを選択する選択部と、を有するデータ転送システム。

(付記 7)

前記データ転送装置は、

前記ネットワークの帯域を監視する帯域監視部を更に有し、

前記選択部は、前記帯域監視部により監視される前記ネットワークの帯域の使用量に応じて、前記転送すべきデータの中から転送するデータを選択する付記 6 に記載のデータシステム。

(付記 8)

前記データ転送装置は、

前記記憶装置に記憶されたデータの更新頻度を記憶する更新頻度記憶部を更に有し、

前記選択部は、前記特定の転送目的のデータに関して前記更新頻度記憶部が記憶する更新頻度に基づいて、前記転送すべきデータの中から転送するデータを選択する付記 6 又は 7 に記載のデータ転送システム。

(付記 9)

前記記憶装置に記憶されたデータは、複数のブロックに分割され、

前記データ転送装置は、

前記複数のブロックに分割されて前記記憶装置に記憶された各々のデータの読取頻度を記憶する読取頻度記憶部を更に有し、

前記選択部は、前記特定の転送目的のデータに関して前記読取頻度記憶部が記憶するブロック毎の読取頻度に基づいて、前記転送すべきデータの中から転送するデータのブロックを選択する付記 8 に記載のデータ転送システム。

(付記 10)

前記選択部は、前記判定部により判定された結果が、即時性が要求されないデータ転送の目的を示す場合に、前記転送すべきデータの中から転送するデータを選択する付記 6 ~ 9 のいずれか一項に記載のデータ転送システム。

(付記 11)

記憶装置間にてネットワークを介してデータを転送する処理をコンピュータが実行するデータ転送方法であって、

転送すべきデータの転送目的を判定し、

前記判定された結果が特定の転送目的を示す場合、前記転送すべきデータの中から転送するデータを選択する、

処理をコンピュータが実行するデータ転送方法。

(付記 12)

前記ネットワークの帯域を監視し、

前記監視される前記ネットワークの帯域の使用量に応じて、前記転送すべきデータの中から転送するデータを選択する、

処理をコンピュータが実行する付記 11 に記載のデータ転送方法。

(付記 13)

前記記憶装置に記憶されたデータの更新頻度を記憶し、

前記特定の転送目的のデータに関して前記記憶された更新頻度に基づいて、前記転送すべきデータの中から転送するデータを選択する、

処理をコンピュータが実行する付記 11 又は 12 に記載のデータ転送方法。

(付記 14)

前記記憶装置に記憶されたデータは、複数のブロックに分割され、
前記複数のブロックに分割されて前記記憶装置に記憶された各々のデータの読取頻度を記憶し、

前記特定の転送目的のデータに関して前記記憶されたブロック毎の読取頻度に基づいて、前記転送すべきデータの中から転送するデータのブロックを選択する、
処理をコンピュータが実行する付記 13 に記載のデータ転送方法。

(付記 15)

前記判定された結果が、即時性が要求されないデータ転送の目的を示す場合に、前記転送すべきデータの中から転送するデータを選択する、

処理をコンピュータが実行する付記 11 ~ 14 のいずれか一項に記載のデータ転送方法

10

(付記 16)

記憶装置間にてネットワークを介してデータを転送する処理をコンピュータに実行させるプログラムであって、

転送すべきデータの転送目的を判定し、

前記判定された結果が特定の転送目的を示す場合、前記転送すべきデータの中から転送するデータを選択する、

処理をコンピュータに実行させるプログラム。

(付記 17)

前記ネットワークの帯域を監視し、

前記監視される前記ネットワークの帯域の使用量に応じて、前記転送すべきデータの中から転送するデータを選択する、

処理をコンピュータに実行させる付記 16 に記載のプログラム。

20

(付記 18)

前記記憶装置に記憶されたデータの更新頻度を記憶し、

前記特定の転送目的のデータに関して前記記憶された更新頻度に基づいて、前記転送すべきデータの中から転送するデータを選択する、

処理をコンピュータに実行させる付記 16 又は 17 に記載のプログラム。

(付記 19)

前記記憶装置に記憶されたデータは、複数のブロックに分割され、
前記複数のブロックに分割されて前記記憶装置に記憶された各々のデータの読取頻度を記憶し、

30

前記特定の転送目的のデータに関して前記記憶されたブロック毎の読取頻度に基づいて、前記転送すべきデータの中から転送するデータのブロックを選択する、

処理をコンピュータに実行させる付記 18 に記載のプログラム。

(付記 20)

前記判定された結果が、即時性が要求されないデータ転送の目的を示す場合に、前記転送すべきデータの中から転送するデータを選択する、

処理をコンピュータに実行させる付記 16 ~ 19 のいずれか一項に記載のプログラム。

(付記 21)

付記 16 ~ 20 に記載のプログラムを格納したコンピュータ読み取り可能な記憶媒体。

40

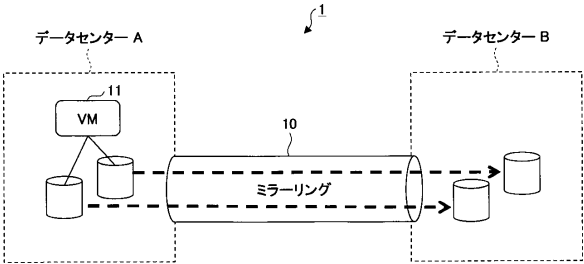
【符号の説明】

【0117】

1 : データ転送システム、10 : WAN、15 : 物理サーバ、16 : VM 情報テーブル、20a、20b : ストレージ装置、30a、30b : ストレージ制御装置、31a、31b : ミラーリング・コントローラ、36 : ディスク情報テーブル、36e : 更新回数、37 : 状態情報テーブル、37d : リードカウンタ、38 : 帯域情報テーブル、42 : サイト情報テーブル、311 : 判定部、312 : 選択部、313 : 帯域監視部、314 : 送信部、315 : 受信部、316 : 記憶部、317 : バッファ

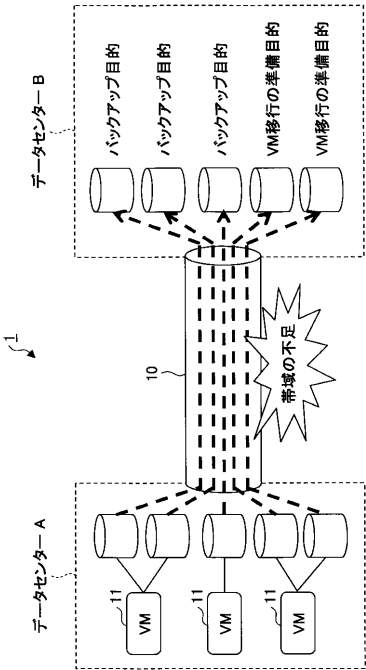
【図 1】

データ転送システムの一実施形態を示した図



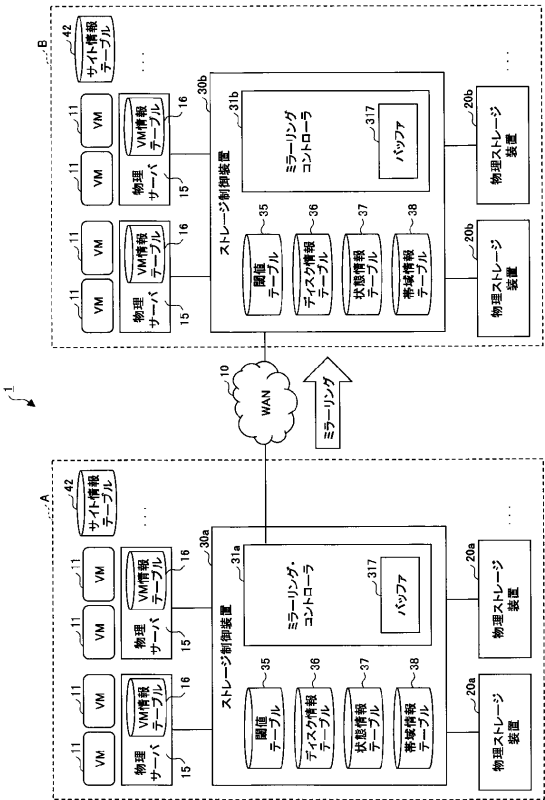
【図 2】

データ転送システムの一実施形態にかかる帯域の使用状況例を示した図



【図 3】

データ転送システムの一実施形態にかかる全体構成図



【図 4】

VM情報テーブルの一例を示した図

16a	16b	16c	16d	16e
VM ID	作成者ID	OS種別	DR属性	割り当て済仮想ディスクID
vm-0001	user-0011	Linux	FALSE	vdisk-A-0001
vm-0002	user-0011	Linux	FALSE	vdisk-A-0256
vm-0016	user-0501	Windows	TRUE	vdisk-A-0123

【図 5】

ディスク情報テーブルの一例を示した図

36a	36b	36c	36d	36e	36f
仮想ディスクID	容量	ミラー先のデータセンターID	ミラー先の仮想ディスクID	単位時間あたりの更新(write)回数	ミラーリング目的
vdisk-A-0001	500 GB	—	—	8	—
vdisk-A-0026	300 GB	B	vdisk-B-1226	6	バックアップ
vdisk-A-0123	700 GB	C	vdisk-C-2001	10	バックアップ
vdisk-A-0201	250 GB	—	—	25	—
vdisk-A-0256	500 GB	B	vdisk-B-1375	4	移行準備
vdisk-A-0300	750 GB	B	vdisk-B-1482	16	移行準備

【図 6】

状態情報テーブルの一例を示した図

36a 仮想ディスクID vdisk-A-0300			
37a ブロック番号	37b 更新時刻	37c バッファフラグ	37d リードカウンタ
0000 0001	2012/10/02 09:31:12	false	—
0000 0002	2012/09/27 22:54:33	false	—
0000 0003	2012/10/02 13:27:59	true	12
0000 0004	2012/09/30 02:11:45	false	—
0000 0005	2012/10/02 13:30:28	true	3

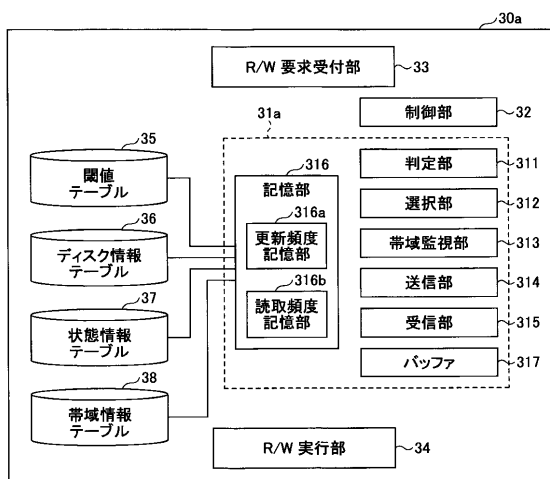
【図 7】

帯域情報テーブルの一例を示した図

38a ソース データセンターID	38b デスティネーション データセンターID	38c 状態	38d 最大帯域	38e 使用帯域
A	B	NORMAL	800 Kbps	660 Kbps
A	C	NORMAL	500 Kbps	480 Kbps

【図 9】

ストレージ制御装置の一実施形態にかかる機能構成図



【図 8】

サイト情報テーブルの一例を示した図

42a データセンター ID	42b 管理用ポータル IPアドレス	42c 管理者ID	42d 運用種別	42e 状態
A	10.123.45.1	Admin-A	運用系	サービス中
B	10.201.23.1	Admin-B	待機系	待機中
C	10.55.123.1	Admin-C	運用系	サービス中
D	10.20.3.127	Admin-D	運用系	メンテナンス中

【図 10】

状態情報テーブルの一例を示した図

36a 仮想ディスクID vdisk-A-0300			
37a ブロック番号	37b 更新時刻	37c バッファフラグ	37d リードカウンタ
0000 0001	2012/10/02 09:31:12	false	—
0000 0002	2012/09/27 22:54:33	false	—
	2012/10/02 13:31:45	true	0
0000 0003	2012/10/02 13:27:59	true	12
0000 0004	2012/09/30 02:11:45	false	—
0000 0005	2012/10/02 13:30:28	true	3

【図 11】

状態情報テーブルの一例を示した図

36a 仮想ディスクID vdisk-A-0300			
37a ブロック番号	37b 更新時刻	37c バッファフラグ	37d リードカウンタ
0000 0001	2012/10/02 09:31:12	false	—
0000 0002	2012/10/02 13:31:45	true	2
0000 0003	2012/10/02 13:27:59	true	12
		false	—
0000 0004	2012/09/30 02:11:45	false	—
0000 0005	2012/10/02 13:40:22	true	0

【図 1 2】

状態情報テーブルの一例を示した図

仮想ディスクID			
vdisk-A-0300			
ブロック番号	更新時刻	バッファフラグ	リードカウンタ
0000 0001	2012/10/02 09:31:12	false	—
0000 0002	2012/10/02 13:31:45	true	0 ↓ 1
0000 0003	2012/10/02 13:27:59	true	12
0000 0004	2012/09/30 02:11:45	false	—
0000 0005	2012/10/02 13:30:28	true	3

【図 1 4】

状態情報テーブルの一例を示した図

仮想ディスクID			
vdisk-A-0300			
ブロック番号	更新時刻	バッファフラグ	リードカウンタ
0000 0001	2012/10/02 09:31:12	false	—
0000 0002	2012/10/02 13:31:45	true	2
0000 0003	2012/10/02 13:27:59	true	12
0000 0004	2012/09/30 02:11:45	false	—
0000 0005	2012/10/02 13:30:28 ↓ 2012/10/02 13:40:22	true	3 ↓ 0

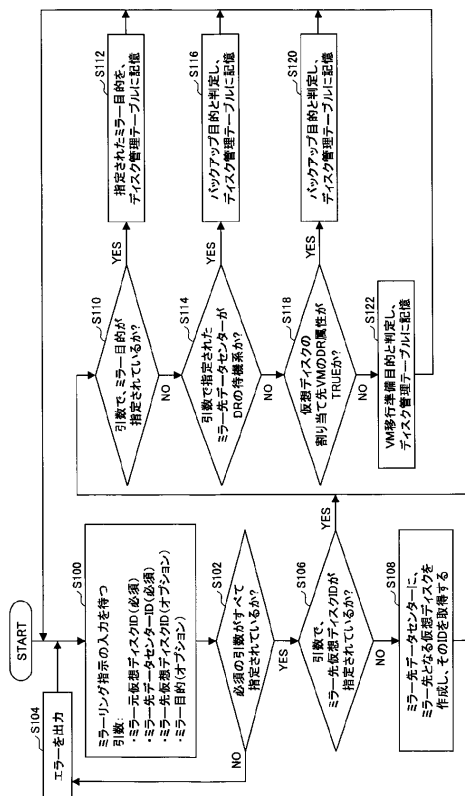
【図 1 3】

状態情報テーブルの一例を示した図

仮想ディスクID			
vdisk-A-0300			
ブロック番号	更新時刻	バッファフラグ	リードカウンタ
0000 0001	2012/10/02 09:31:12	false	—
0000 0002	2012/10/02 13:31:45	true	1 ↓ 2
0000 0003	2012/10/02 13:27:59	true	12
0000 0004	2012/09/30 02:11:45	false	—
0000 0005	2012/10/02 13:30:28	true	3

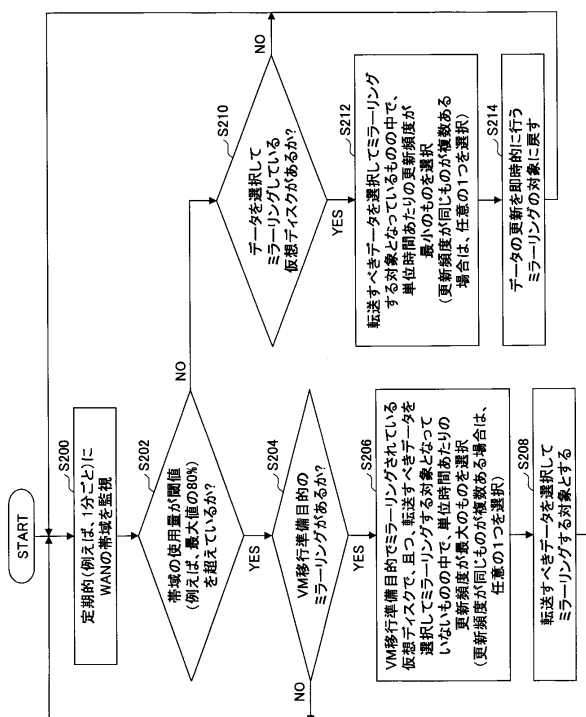
【図 1 5】

転送目的の判定処理の一実施形態を示したフローチャート



【図 1 6】

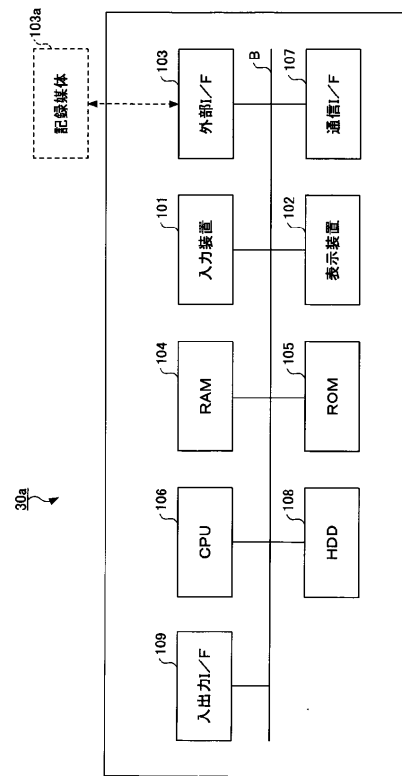
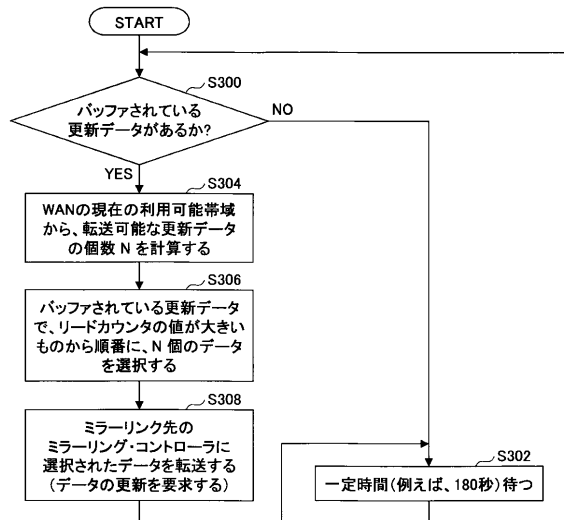
転送データの選択処理の一実施形態を示したフローチャート



【図 17】

【図 18】

ブロック単位の転送データの選択処理の一実施形態を示したフローチャート ストレージ制御装置の一実施形態にかかるハードウェア構成図



フロントページの続き

(56)参考文献 特開 2011 - 034164 (JP, A)
米国特許出願公開第 2011 / 0029748 (US, A1)
特開 2003 - 085032 (JP, A)
特開 2007 - 128335 (JP, A)
米国特許出願公開第 2007 / 0106712 (US, A1)
米国特許出願公開第 2006 / 0236056 (US, A1)
特開 2006 - 301820 (JP, A)
米国特許出願公開第 2011 / 0314183 (US, A1)

(58)調査した分野(Int.Cl., DB名)

G06F 3 / 06 - 3 / 08
G06F 12 / 00