

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2005-293469

(P2005-293469A)

(43) 公開日 平成17年10月20日(2005. 10. 20)

(51) Int.Cl.⁷

G06F 12/00

G06F 3/06

F I

G06F 12/00

5 3 1 D

G06F 3/06

3 0 4 F

テーマコード (参考)

5 B 0 6 5

5 B 0 8 2

審査請求 未請求 請求項の数 4 O L (全 17 頁)

(21) 出願番号 特願2004-111014 (P2004-111014)

(22) 出願日 平成16年4月5日 (2004. 4. 5)

(71) 出願人 000004226

日本電信電話株式会社

東京都千代田区大手町二丁目3番1号

(74) 代理人 100083552

弁理士 秋田 収喜

(74) 代理人 100103746

弁理士 近野 恵一

(72) 発明者 藤田 智成

東京都千代田区大手町二丁目3番1号 日

本電信電話株式会社内

Fターム(参考) 5B065 BA01 EA33

5B082 DE04 GA04

(54) 【発明の名称】 データコピーシステムおよびデータコピー方法

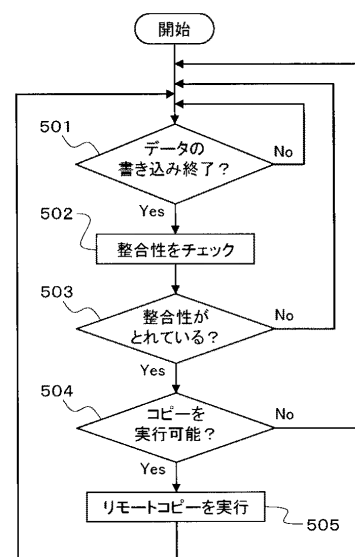
(57) 【要約】 (修正有)

【課題】 順序無保証式のリモートコピーを行う場合に、一方の記憶装置システムのデータが破損しても他の記憶装置システムにおいてデータを回復する。

【解決手段】 ホストコンピュータから正記憶装置システムの主論理ボリュームへのデータの書き込みが終了した時点で整合性を検査するステップ1と、整合性が保証された時点の主論理ボリュームの状態を記録したスナップショットを作成するステップ2と、スナップショットを用いて正記憶装置システムと副記憶装置システムの間でリモートコピーを実行するステップ3を有す。ステップ3は、正記憶装置システムの主論理ボリュームの状態を示すデータを副記憶装置システムに転送するステップ4と、リモートコピー実行前の副記憶装置システムの論理ボリュームの状態を記録したスナップショットを作成するステップ5と、ステップ4で転送されたデータを副記憶装置システムの論理ボリュームに適用するステップ6とを有する。

【選択図】 図2

図2



【特許請求の範囲】

【請求項 1】

ホストコンピュータと、前記ホストコンピュータと接続され、前記ホストコンピュータからのデータを記録保持する正記憶装置システムと、前記正記憶装置と接続され、前記正記憶装置システムが保持しているデータのコピーを記録保持する副記憶装置システムとを備え、前記正記憶装置システムは、ホストコンピュータからのデータを記録保持する主論理ボリュームを有する記憶装置と、前記ホストコンピュータから前記主論理ボリュームへのデータの書き込み、および前記副記憶装置システムの間で行うデータ（論理ボリューム）のコピーを制御する制御装置とを備え、前記副記憶装置システムは、前記正記憶装置システムの主論理ボリュームのコピーを記録保持する論理ボリュームを有する記憶装置と、前記正記憶装置システムとの間で行うデータ（論理ボリューム）のコピーを制御する制御装置とを備えるデータコピーシステムであって、

10

前記正記憶装置システムの記憶装置は、前記主論理ボリュームと、前記副記憶装置システムへデータをコピーするときに用いる前記主論理ボリュームの状態を記録したスナップショットを管理するスナップショット管理手段を備え、

前記正記憶装置システムの制御装置は、前記正記憶装置システムの主論理ボリュームのファイルシステムの整合性を検査する整合性検査手段と、前記副記憶装置システムとの間で行うデータのコピー（リモートコピー）を制御するリモートコピー制御手段と、前記正ホストコンピュータからのデータを前記主論理ボリュームに書き込む処理、前記整合性検査手段に整合性を検査させる処理、前記主論理ボリュームの状態を記録したスナップショットの作成処理、前記リモートコピー制御手段を介して前記副記憶装置システムに送信するコピー用データの作成処理の各処理を行うデータ処理手段とを備え、

20

前記副記憶装置システムの記憶装置は、前記論理ボリュームと、前記正記憶装置システムの主論理ボリュームを前記論理ボリュームにコピーするときに、コピー前の前記論理ボリュームの状態を記録したスナップショットを管理するスナップショット管理手段を備え、

前記副記憶装置システムの制御装置は、前記正記憶装置との間で行うデータのコピー（リモートコピー）を制御するリモートコピー制御手段と、前記リモートコピーを実行する前に、前記論理ボリュームの状態を記録したスナップショットを作成する処理、前記正記憶装置システムから送信された主論理ボリュームのデータを前記論理ボリュームに適用（コピー）する処理、前記リモートコピーに失敗したときに、前記スナップショットを用いて前記論理ボリュームを復元する処理の各処理を行うデータ処理手段とを備えることを特徴とするデータコピーシステム。

30

【請求項 2】

ホストコンピュータと、前記ホストコンピュータと接続され、前記ホストコンピュータからのデータを記録保持する正記憶装置システムと、前記正記憶装置と接続され、前記正記憶装置システムが保持しているデータのコピーを記録保持する副記憶装置システムとを備え、前記正記憶装置システムは、ホストコンピュータからのデータを記録保持する主論理ボリュームを有する記憶装置と、前記ホストコンピュータから前記主論理ボリュームへのデータの書き込み、および前記副記憶装置システムの間で行うデータ（論理ボリューム）のコピーを制御する制御装置とを備え、前記副記憶装置システムは、前記正記憶装置システムの主論理ボリュームのコピーを記録保持する論理ボリュームを有する記憶装置と、前記正記憶装置システムとの間で行うデータ（論理ボリューム）のコピーを制御する制御装置とを備えるデータコピーシステムによるデータコピー方法であって、

40

前記ホストコンピュータから前記正記憶装置システムの主論理ボリュームへのデータの書き込みが終了した時点で前記主論理ボリュームのファイルシステムとしての整合性を検査するステップ 1 と、

前記主論理ボリュームの整合性が保証された時点での前記主論理ボリュームの状態を記録したスナップショットを作成するステップ 2 と、

前記ステップ 2 で作成したスナップショットを用いて、前記正記憶装置システムと前記

50

副記憶装置システムの間でリモートコピーを実行し、前記副記憶装置システムの論理ボリュームの状態を前記主論理ボリュームと同じ状態にするステップ 3 を有し、

前記ステップ 3 は、前記ステップ 2 で作成されたスナップショットを用いて、前記正記憶装置システムの主論理ボリュームの状態を示すデータを前記副記憶装置システムに転送するステップ 4 と、前記リモートコピー実行前の前記副記憶装置システムの論理ボリュームの状態を記録したスナップショットを作成するステップ 5 と、前記ステップ 4 で転送されたデータを前記副記憶装置システムの論理ボリュームに適用するステップ 6 とを有することを特徴とするデータコピー方法。

【請求項 3】

前記ステップ 4 は、前回のリモートコピー実行時に作成された前記正記憶装置システムの主論理ボリュームの状態を記録したスナップショットと前記ステップ 2 で作成されたスナップショットの差分データを転送することを特徴とする請求項 2 に記載のデータコピー方法。 10

【請求項 4】

前記ステップ 3 は、前記ステップ 6 の処理が失敗した場合に、前記ステップ 5 で作成されたスナップショットを用いて、前記副記憶装置システムの論理ボリュームを前記リモートコピー実行前の状態に復元するステップ 7 を有することを特徴とする請求項 2 または請求項 3 に記載のデータコピー方法。

【発明の詳細な説明】

【技術分野】

20

【0001】

本発明は、データコピーシステムおよびデータコピー方法に関し、特に、正記憶装置から前記正記憶装置とは遠隔地にある副記憶装置へデータをコピー（ミラーリング）するデータコピーシステムおよびデータコピー方法に適用して有効な技術に関するものである。

【背景技術】

【0002】

従来、ホストコンピュータおよび複数の記憶装置を備える情報処理システムは、ユーザが長期的に保存することを望むデータを記憶装置に保存している。そして、ユーザからの要求に応じて、要求されたデータを前記記憶装置から読み出すことで、データの長期的な利用を可能としている。 30

【0003】

また、前記情報処理システムでは、前記記憶装置に書き込まれたデータの消失を防ぐために、前記ホストコンピュータと接続された 1 つの記憶装置にデータを書き込むときに、同一のデータを他の記憶装置にも書き込むのが一般的である。このように、1 つのデータを 2 つ以上の記憶装置のそれぞれに書き込む方法として、たとえば、リモートコピーと呼ばれる方法がある。

【0004】

前記リモートコピーは、前記ホストコンピュータを介在させずに、地理的に離れた場所にある複数の記憶装置システム同士でデータのコピーを行い、前記データを二カ所に保存する方法である。なお、前記記憶装置システムは、複数の記憶装置および前記記憶装置を制御するコントローラから構成されるシステムである。 40

【0005】

前記リモートコピーを行う情報処理システムでは、地理的に離れた場所に設置された記憶装置システム間が、専用回線あるいは公衆回線等のネットワーク回線で相互接続されている。そして、ある記憶装置システムが有する論理的な記憶領域（以下、論理ボリュームという）のうち、リモートコピーの対象となる論理ボリュームの容量と同容量の論理ボリュームが、コピー先の記憶装置システムに確保される。以下、前記リモートコピーの対象となる論理ボリュームをコピー元論理ボリューム、前記コピー先の記憶装置システムに確保される論理ボリュームをコピー先論理ボリュームという。

【0006】

50

前記コピー先論理ボリュームは、前記コピー元論理ボリュームと1対1の対応関係を持つように形成される。そして、前記コピー元論理ボリュームのデータは、前記ネットワーク回線を介して前記コピー先論理ボリュームにコピーされる。また、前記コピー元論理ボリュームに記録されているデータが更新された場合、更新されたデータは、前記ネットワーク回線を介して前記コピー先論理ボリュームを有する記憶装置システムに転送される。そして、転送されたデータ、すなわち前記コピー元論理ボリュームで更新されたデータが、前記コピー先論理ボリュームにコピーされる。このように、前記複数の記憶装置を有する情報処理システムでは、前記リモートコピーを行うことによって、複数の記憶装置で同一内容の論理ボリュームを保持することができる。

【0007】

10

また、前記リモートコピーには、順序無保証式と呼ばれる方法がある。前記順序無保証式のリモートコピーでは、コピー元論理ボリュームを備える記憶装置システムが、前記コピー先論理ボリュームを備える記憶装置システムにデータを送信したときに、前記コピー先論理ボリュームを備える記憶装置システムからのデータを受領したことを示す信号を待たずに、前記コピー先論理ボリュームに次のデータを送信する。

【0008】

また、前記順序無保証式のリモートコピーでは、ホストコンピュータがコピー元論理ボリュームにデータを書き込んだ順番と、それらのデータがコピー先論理ボリュームに転送される順番が異なる場合がある。そのため、前記順序無保証式のリモートコピーでは、このような性質に基づいて、ホストコンピュータが前記コピー元論理ボリュームの同じ位置に繰り返しデータを書き込んだ場合、一番最後に書き込まれたデータのみを、前記コピー先論理ボリュームを有する記憶装置システムに送ることができる。したがって、前記コピー元とコピー先の記憶装置システム間のネットワーク回線のトラフィック量を減らすことができる。

20

【0009】

また、前記ホストコンピュータが、前記記憶装置システムに書き込んだデータをファイルシステムとして管理している場合、前記ホストコンピュータのメモリを利用することで、前記ホストコンピュータから記憶装置システムへのデータ転送の指示と、実際の記憶装置システムへのデータ転送（書き込み処理）が非同期に行われる。

【0010】

30

しかしながら、前記ホストコンピュータから前記記憶装置システムへのデータ転送を非同期に実行すると、たとえば、停電等の障害が発生した場合、前記ファイルシステムに対してなされた変更の一部だけ、すなわち、実際に転送する全データのうちの一部だけが前記記憶装置システムのコピー元論理ボリュームに書き込まれた状態になり、前記ファイルシステムが破損するという問題が発生することがある。

【0011】

このようなファイルシステムの破損、言い換えると前記ファイルシステムの整合性が失われることを防ぐために、近年では、少なくとも、ファイルを管理するためのデータであるinode、ディレクトリ等のメタデータと呼ばれるデータが、突然の停電等でホストコンピュータから前記記憶装置システムへまだ転送されていない、メモリに保持されている非同期な書き込みのデータが失われても、前記ホストコンピュータが復帰したあとに、前記ファイルシステムを修復することができるような手順で、前記非同期の書き込みを管理している。

40

【0012】

前記非同期の書き込みの管理方法として代表的なのは、ジャーナリングと呼ばれる方法である。この方法では、ファイルシステムとは別のジャーナルと呼ばれる特殊なデータを利用することで、一部のファイルデータが失われることはあっても、ファイルシステムの破損は避けられる（たとえば、非特許文献1を参照。）。

【0013】

また、前記順序無保証式のリモートコピーでは、前記記憶装置システム間で転送（コピ

50

ー)が必要なデータの量を低減することができる。しかしながら、前記ホストコンピュータからコピー元論理ボリュームへのデータの書き込みと同期して、書き込んだデータが前記コピー先論理ボリュームに転送されない場合、前記コピー先論理ボリュームが、前記ホストコンピュータの意図した状態になっていないという状況が発生する。このような状況では、前記コピー先論理ボリュームに対してジャーナリングのような、ファイルシステムが持つ障害対策が有効に働かず、ファイルシステムが破損してしまう可能性があるという問題がある。

【0014】

前記ファイルシステムが持つ障害対策が有効に働かずに、ファイルシステムが破損してしまうという問題を解決する方法としては、たとえば、ホストコンピュータが前記コピー元論理ボリュームに保存されているファイルシステムの整合性が取れているある一時点を選択し、コピー元論理ボリュームを有する記憶装置システムに対して、順序無保証式のリモートコピーを開始する命令を送ってから、コピー先論理ボリュームを有する記憶装置システムとの間でデータの送受信(コピー)を開始する方法がある(たとえば、特許文献1を参照。)。この方法だと、コピー先論理ボリュームに保存されたデータがファイルシステムとして整合性が取れている状態であることを保証することができる。

【非特許文献1】Margo Seltzer and Greg Ganger and M. Kirk McKusick and Keith Smith and Craig Soules and Christopher Stein, "Journaling Versus Soft Updates: Asynchronous Meta-data Protection in File Systems," the 2000 USENIX Annual Technical Conference

【特許文献1】特開2002-259183号公報

【発明の開示】

【発明が解決しようとする課題】

【0015】

しかしながら、前記特許文献1に記載されたような方法の場合、前記記憶装置システムは、前記順序無保証式のリモートコピーを開始するタイミングを決定するために、前記ホストコンピュータからの情報(命令)を受け取る必要がある。そのため、前記順序無保証式のリモートコピーを想定していない情報処理システムでは、前記ホストコンピュータに必要なソフトウェアやハードウェアを新たに導入しなければならないという問題がある。

【0016】

本発明の目的は、順序無保証式のリモートコピーを行う場合に、記憶装置システムがホストコンピュータと情報をやりとりする必要が無く、ファイルシステムに固有の情報をを用いることで、複数の記憶装置システムの間でファイルシステムとして整合性が取れた状態を保持し、一方の記憶装置システムのデータが破損した場合にも、他の記憶装置システムにおいてデータを回復することが可能な技術を提供することにある。

【0017】

本発明の前記ならびにその他の目的と新規な特徴は、本明細書の記述及び添付図面によって明らかになるであろう。

【課題を解決するための手段】

【0018】

本願において開示される発明の概略を説明すれば、以下の通りである。

(1)ホストコンピュータと、前記ホストコンピュータと接続され、前記ホストコンピュータからのデータを記録保持する正記憶装置システムと、前記正記憶装置と接続され、前記正記憶装置システムが保持しているデータのコピーを記録保持する副記憶装置システムとを備え、前記正記憶装置システムは、ホストコンピュータからのデータを記録保持する主論理ボリュームを有する記憶装置と、前記ホストコンピュータから前記主論理ボリュームへのデータの書き込み、および前記副記憶装置システムの間で行うデータ(論理ボリューム)のコピーを制御する制御装置とを備え、前記副記憶装置システムは、前記正記憶装置システムの主論理ボリュームのコピーを記録保持する論理ボリュームを有する記憶装置と、前記正記憶装置システムとの間で行うデータ(論理ボリューム)のコピーを制御す

10

20

30

40

50

る制御装置とを備えるデータコピーシステムであって、前記正記憶装置システムの記憶装置は、前記主論理ボリュームと、前記副記憶装置システムへデータをコピーするときに用いる前記主論理ボリュームの状態を記録したスナップショットを管理するスナップショット管理手段を備え、前記正記憶装置システムの制御装置は、前記正記憶装置システムの主論理ボリュームのファイルシステムの整合性を検査する整合性検査手段と、前記副記憶装置システムとの間で行うデータのコピー（リモートコピー）を制御するリモートコピー制御手段と、前記正ホストコンピュータからのデータを前記主論理ボリュームに書き込む処理、前記整合性検査手段に整合性を検査させる処理、前記主論理ボリュームの状態を記録したスナップショットの作成処理、前記リモートコピー制御手段を介して前記副記憶装置システムに送信するコピー用データの作成処理の各処理を行うデータ処理手段とを備え、前記副記憶装置システムの記憶装置は、前記論理ボリュームと、前記正記憶装置システムの主論理ボリュームを前記論理ボリュームにコピーするときに、コピー前の前記論理ボリュームの状態を記録したスナップショットを管理するスナップショット管理手段を備え、前記副記憶装置システムの制御装置は、前記正記憶装置との間で行うデータのコピー（リモートコピー）を制御するリモートコピー制御手段と、前記リモートコピーを実行する前に、前記論理ボリュームの状態を記録したスナップショットを作成する処理、前記正記憶装置システムから送信された主論理ボリュームのデータを前記論理ボリュームに適用（コピー）する処理、前記リモートコピーに失敗したときに、前記スナップショットを用いて前記論理ボリュームを復元する処理の各処理を行うデータ処理手段とを備えるデータコピーシステムである。

10

20

【0019】

（２）ホストコンピュータと、前記ホストコンピュータと接続され、前記ホストコンピュータからのデータを記録保持する正記憶装置システムと、前記正記憶装置と接続され、前記正記憶装置システムが保持しているデータのコピーを記録保持する副記憶装置システムとを備え、前記正記憶装置システムは、ホストコンピュータからのデータを記録保持する主論理ボリュームを有する記憶装置と、前記ホストコンピュータから前記主論理ボリュームへのデータの書き込み、および前記副記憶装置システムの間で行うデータ（論理ボリューム）のコピーを制御する制御装置とを備え、前記副記憶装置システムは、前記正記憶装置システムの主論理ボリュームのコピーを記録保持する論理ボリュームを有する記憶装置と、前記正記憶装置システムとの間で行うデータ（論理ボリューム）のコピーを制御する制御装置とを備えるデータコピーシステムによるデータコピー方法であって、前記ホストコンピュータから前記正記憶装置システムの主論理ボリュームへのデータの書き込みが終了した時点で前記主論理ボリュームのファイルシステムとしての整合性を検査するステップ１と、前記主論理ボリュームの整合性が保証された時点での前記主論理ボリュームの状態を記録したスナップショットを作成するステップ２と、前記ステップ２で作成したスナップショットを用いて、前記正記憶装置システムと前記副記憶装置システムの間でリモートコピーを実行し、前記副記憶装置システムの論理ボリュームの状態を前記主論理ボリュームと同じ状態にするステップ３を有し、前記ステップ３は、前記ステップ２で作成されたスナップショットを用いて、前記正記憶装置システムの主論理ボリュームの状態を示すデータを前記副記憶装置システムに転送するステップ４と、前記リモートコピー実行前の前記副記憶装置システムの論理ボリュームの状態を記録したスナップショットを作成するステップ５と、前記ステップ４で転送されたデータを前記副記憶装置システムの論理ボリュームに適用するステップ６とを有するデータコピー方法である。

30

40

【発明の効果】

【0020】

本発明のデータコピー方法は、前記（１）に記載したように、前記正記憶装置システムの制御手段に、リモートコピーの実行時期を判断するデータ処理手段が備わっている。そのため、前記ホストコンピュータからリモートコピーの実行時期を指示することなく、前記正記憶装置システムと前記副記憶装置システムのみで前記リモートコピーを実行することができる。

50

【0021】

また、前記(1)に記載されたデータコピーシステムを用いてデータ(主論理ボリューム)のリモートコピーを実行するときには、たとえば、前記(2)に記載したように、前記正記憶装置システムの主論理ボリュームの整合性が取れているときに実行される。このとき、前記主論理ボリュームの整合性は、たとえば、前記ホストコンピュータから前記主論理ボリュームへのデータの書き込みが終了した時点で検査すればよい。このようにすると、前記リモートコピーを実行した後、前記副記憶装置システムの論理ボリュームは、ファイルシステムとして整合性が取れている状態になる。そのため、たとえば、前記正記憶装置システムの主論理ボリュームのデータが破損した場合にも、前記副記憶装置システムの論理ボリュームに記録保持されているデータを利用して、前記破損したデータを容易に回復することができる。

10

【0022】

また、前記(2)に記載したような方法でリモートコピーを実行する場合、前記ステップ4は、前回のリモートコピー実行時に作成された前記正記憶装置システムの主論理ボリュームの状態を記録したスナップショットと前記ステップ2で作成されたスナップショットの差分データを転送してもよい。このようにすると、リモートコピー実行時に前記正記憶装置システムから前記副記憶装置システムに転送するデータのデータ量を低減することができ、短時間でコピーできるとともに、ネットワーク等の回線にかかる負荷を低減することができる。

【0023】

20

また、前記(2)に記載したような方法でリモートコピーを実行する場合、実行中に前記正記憶装置システムと前記副記憶装置システムとを接続するネットワークに障害が発生し、リモートコピーが正常に完了しないことがある。その場合、前記副記憶装置システムの論理ボリュームがファイルシステムとして整合性の取れていない状態になる。そのため、前記ステップ3は、前記ステップ6の処理が失敗した場合に、前記ステップ5で作成されたスナップショットを用いて、前記副記憶装置システムの論理ボリュームを前記リモートコピー実行前の状態に復元するステップ7を有することが好ましい。このようにすることで、リモートコピーが失敗した場合でも、前記副記憶装置システムの論理ボリュームをファイルシステムとして整合性が取れている状態に復元することができる。そのため、たとえば、前記正記憶装置システムの主論理ボリュームのデータが破損した場合でも、容易にデータを回復することができる。

30

【0024】

また、前記(2)に記載したような方法でリモートコピーを実行する場合、前記正記憶装置システムの主論理ボリュームの整合性が保証されている時点のうち、最新の時点の前記主論理ボリューム、言い換えると、最後にリモートコピーに成功したときの主論理ボリュームの状態を、磁氣的または電氣的、あるいは光学的な記録媒体に記録してもよい。前記正記憶装置システムおよび前記副記憶装置システム以外の部分で障害が発生した場合、障害から回復した後、前記正記憶装置システムと前記副記憶装置システムの間で、再度前記論理ボリュームを同期させる必要がある。そのため、前記最後にリモートコピーに成功したときの主論理ボリュームの状態を前記記録媒体に記録しておけば、その時点から前記ホストコンピュータが変更したデータのみを前記副記憶装置システムに送るだけで、前記正記憶装置システムの主論理ボリュームと前記副記憶装置システムの論理ボリュームのデータを一致させることができ、障害発生時の回復作業を高速化させることができる。

40

【0025】

以下、本発明について、図面を参照して実施の形態(実施例)とともに詳細に説明する。

なお、実施例を説明するための全図において、同一機能を有するものは、同一符号を付け、その繰り返しの説明は省略する。

【発明を実施するための最良の形態】

【0026】

50

本発明のデータコピーシステムでは、正ホストコンピュータからのデータを保存する正記憶装置システムの論理ボリュームをファイルシステムとして管理する。そして、前記正記憶装置システムは、前記正ホストコンピュータから前記論理ボリュームへのデータの書き込みが終了した時点で、前記論理ボリュームのファイルシステムの整合性を検査する。このとき、前記論理ボリュームのファイルシステムの整合性が取れていると判断されたら、前記正記憶装置システムは、前記論理ボリュームのデータを副記憶装置システムに転送し、前記副記憶装置システムの論理ボリュームにコピーする。

【 0 0 2 7 】

[実施の形態]

図 1 は、本発明による一実施形態のデータコピーシステムの概略構成を示す模式図である。 10

図 1 において、1 A は正ホストコンピュータ、1 B は副ホストコンピュータ、2 は正記憶装置システム、2 0 1 は記憶装置、2 0 1 A は主論理ボリューム、2 0 1 B はスナップショット管理手段、2 0 2 は制御装置、2 0 2 A は整合性検査手段、2 0 2 B はリモートコピー制御手段、2 0 2 C はデータ処理手段、2 0 2 D はバッファ、3 はネットワーク、4 は副記憶装置システム、4 0 1 は記憶装置、4 0 1 A は論理ボリューム、4 0 1 B はスナップショット管理手段、4 0 2 は制御装置、4 0 2 A はリモートコピー制御手段、4 0 2 B はデータ処理手段、4 0 2 C はバッファである。

【 0 0 2 8 】

本実施形態のデータコピーシステムは、図 1 に示すように、正ホストコンピュータ 1 A および副ホストコンピュータ 1 B の 2 台のホストコンピュータと、前記正ホストコンピュータ 1 A とケーブルで接続されており、前記正ホストコンピュータ 1 A からのデータを記録保持する正記憶装置システム 2 と、前記正記憶装置システム 2 とインターネット等のネットワーク 3 を介して接続されており、前記正記憶装置システム 2 が保持しているデータのコピーを記録保持する副記憶装置システム 4 とにより構成されている。このとき、前記副記憶装置システム 4 は、前記正記憶装置システム 2 に障害が発生したときに、前記正記憶装置システム 2 の代わりをする記憶装置システムであり、前記正記憶装置システム 2 が保持しているデータと同じデータが保持されている。またこのとき、前記副ホストコンピュータ 1 B は、前記正ホストコンピュータ 1 A に障害が発生したときに前記正ホストコンピュータ 1 A の代わりとして稼働するホストコンピュータであり、たとえば、図 1 に示したように、前記副記憶装置システム 4 とケーブルで接続されている。 20 30

【 0 0 2 9 】

また、前記正記憶装置システム 2 は、図 1 に示したように、前記正ホストコンピュータ 1 A からのデータを記録保持する主論理ボリューム 2 0 1 A を有する記憶装置 2 0 1 と、前記正ホストコンピュータ 1 A から前記主論理ボリューム 2 0 1 A へのデータの書き込み、および前記副記憶装置システム 4 との間で行うデータ（論理ボリューム）のコピーを制御する制御装置 2 0 2 からなる。このとき、前記主論理ボリューム 2 0 1 A は、ファイルシステムとして管理されているとする。

【 0 0 3 0 】

またこのとき、前記正記憶装置システム 2 の記憶装置 2 0 1 は、前記副記憶装置システム 4 へデータをコピーするときに用いる前記主論理ボリューム 2 0 1 A の状態を記録したスナップショット SSM_n ($n = 1$) を管理するスナップショット管理手段 2 0 1 B を備える。 40

【 0 0 3 1 】

また、前記正記憶装置システム 2 の制御装置 2 0 2 は、前記正記憶装置システム 2 の主論理ボリューム 2 0 1 A のファイルシステムの整合性を検査する整合性検査手段 2 0 2 A と、前記副記憶装置システム 4 との間で行うデータのコピー（リモートコピー）を制御するリモートコピー制御手段 2 0 1 B と、前記正ホストコンピュータ 1 A からのデータを前記主論理ボリューム 2 0 1 A に書き込む処理、前記整合性検査手段 2 0 2 A に整合性を検査させる処理、前記主論理ボリューム 2 0 1 A の状態を記録したスナップショット SSM 50

の作成処理、前記リモートコピー制御手段 201B を介して前記副記憶装置システム 4 に送信するコピー用データの作成処理の各処理を行うデータ処理手段 202C とを備える。またこのとき、前記正記憶装置システム 2 の制御手段 202 は、前記各手段に加え、前記主論理ボリューム 201A の整合性の検査、および前記正記憶装置システム 2 と前記副記憶装置システム 4 との間でリモートコピーを実行しているときに前記正ホストコンピュータ 1A から送られてきたデータを一時的に保持するバッファ 202D を備える。

【0032】

また、前記データ処理手段 202C は、前記正ホストコンピュータ 1A から前記主論理ボリューム 201A へのデータの書き込みが終了した時点で、前記整合性検査手段 202A に整合性を検査させる命令をする。また、前記データ処理手段 202C は、前記整合性検査手段 202A で検査した結果、前記主論理ボリューム 201A の整合性が取れていると保証された時点で、前記リモートコピー制御手段 201B にリモートコピーを開始させる命令をする。

10

【0033】

一方、前記副記憶装置システム 4 は、図 1 に示したように、前記正記憶装置システム 2 の主論理ボリューム 201A のコピーを記録保持する論理ボリューム 401A を有する記憶装置 401 と、前記正記憶装置システム 2 との間で行うデータ（論理ボリューム）のコピーを制御する制御装置 402 からなる。

【0034】

またこのとき、前記副記憶装置システム 4 の記憶装置 401 は、前記正記憶装置システム 2 の主論理ボリューム 201A の状態を前記論理ボリューム 401A にコピーするときに、コピー前の前記論理ボリューム 401A の状態を記録したスナップショット SSC_m（m = 1）を管理するスナップショット管理手段 401B を備える。

20

【0035】

また、前記副記憶装置システム 4 の制御装置 402 は、前記正記憶装置システム 2 との間で行うデータのコピー（リモートコピー）を制御するリモートコピー制御手段 402A と、前記リモートコピーを実行する前に、前記論理ボリューム 401A の状態を記録したスナップショット SSC を作成する処理、前記正記憶装置システム 2 から送信された主論理ボリューム 201A のデータを前記論理ボリューム 401A に適用（コピー）する処理、前記リモートコピーに失敗したときに、前記スナップショット SSC を用いて前記論理ボリューム 401 を整合性が取れた状態に復元する処理の各処理を行うデータ処理手段 402B とを備える。またこのとき、前記副記憶装置システム 4 の制御装置 402 は、前記各手段に加え、前記論理ボリューム 401A のスナップショット SSC を作成しているときに前記正記憶装置システム 2 から送られてきたデータを一時的に保持するバッファ 402C を備える。

30

【0036】

図 2 乃至図 11 は、本実施形態のデータコピーシステムを用いたデータのコピー方法を説明するための模式図であり、図 2 は正記憶装置システムの制御手段の全体的な処理手順を示すフロー図、図 3 は整合性を検査するステップにおけるシステム上の処理を示す図、図 4 はリモートコピーを実行するステップにおける正記憶装置システムの処理手順を示すフロー図、図 5 はスナップショットを作成するステップにおけるシステム上の処理を示す図、図 6 は正記憶装置システムからデータを送信するステップにおけるシステム上の処理を示す図、図 7 はリモートコピーを実行するステップにおける副記憶装置システムの処理手順を示すフロー図、図 8 および図 9 は副記憶装置システムでデータを受信するステップにおけるシステム上の処理を示す図、図 10 はリモートコピーが完了した後の処理の一例を示す図、図 11 はリモートコピーが失敗した場合のデータの復元方法の一例を示す図である。

40

【0037】

本実施形態のデータコピーシステムでは、前記正記憶装置システム 2 の主論理ボリューム 201A のデータを前記副記憶装置システム 4 の論理ボリューム 401A にリモートコ

50

ピーする時期の判断は、前記正記憶装置システム2の制御装置202が決定する。前記正記憶装置システム2の制御装置202では、たとえば、図2に示すように、前記正ホストコンピュータ1Aから前記正記憶装置システム2の主論理ボリューム201Aへのデータの書き込みが終了したか否かを基準として、リモートコピーの実行時期を判断する(ステップ501)。前記ステップ501は、前記データ処理手段202Cで行う。このとき、たとえば、データの書き込み中であれば書き込みが終了するまで待機する。また、データの書き込みが行われていない時には、次の書き込みが行われ、終了するまで待機する。

【0038】

前記正ホストコンピュータ1Aから前記主論理ボリューム201Aへのデータの書き込みが終了したら、前記主論理ボリューム201Aの整合性を検査する(ステップ502)。前記ステップ502は、前記データ処理手段202Cが前記データの書き込み終了を確認した時点で、まず、たとえば、図3に示すように、前記データ処理手段202Cから前記整合性検査手段202Aに検査を命令する。前記データ処理手段202Cからの命令を受けた前記整合性検査手段202Aは、前記主論理ボリューム201Aのファイルシステムの整合性を検査し、検査した結果を前記データ処理手段202Cに返す。また、前記主論理ボリューム201Aの整合性を検査している間に、前記正ホストコンピュータ1Aから前記主論理ボリューム201Aへのデータの書き込みが行われた場合、たとえば、図3に示したように、前記正ホストコンピュータ1Aから転送されたデータを前記バッファ202Dに一時的に退避させておく。

【0039】

前記整合性検査手段202Aからの検査結果を受け取った前記データ処理手段202Cは、その結果から、前記主論理ボリュームのファイルシステムの整合性が取れているか否かの判断をする(ステップ503)。このとき、整合性が取れていないのであれば、何もせず、次の書き込みがあるまで待機する。

【0040】

また、前記ステップ503で整合性が取れていると判断した場合、次に、前記副記憶装置システム4との間でリモートコピーが可能であるか否かの判断をする(ステップ504)。前記ステップ504は、たとえば、図3に示すように、前記データ処理手段202Cから前記リモートコピー制御手段202Bに対して、リモートコピーが可能であるか確認させる命令を送る。前記確認命令を受け取った前記リモートコピー制御手段202Bは、たとえば、図3に示すように、前記副記憶装置システム4のリモートコピー制御手段402Aとの間で通信を試みてリモートコピーが可能であるか確認し、その結果を前記データ処理手段202Cに返す。前記データ処理手段202Cは、受け取った確認結果が、たとえば、前記ネットワーク3や前記副記憶装置システム4に障害が発生していてリモートコピーができないという内容であれば、いまはリモートコピーを実行しないと判断し、次の書き込みがあるまで待機する。

【0041】

また、前記ステップ504でリモートコピーを実行可能であると判断した場合、次に、リモートコピーを実行する(ステップ505)。

【0042】

前記リモートコピーを実行するステップ505において、前記正記憶装置システム2の制御装置202では、まず、図4に示すように、前記正ホストコンピュータ1Aから前記主論理ボリューム201Aへの書き込みを停止する(ステップ5052A)。そして次に、前記正記憶装置システム2の記憶装置201のスナップショット管理手段201Bに保持されているスナップショットSSM_n(n=1)をスナップショットSSM_{n+1}とする(ステップ5052B)。

【0043】

前記ステップ5052Aおよびステップ5052Bの処理が済んだら、次に、リモートコピーの対象となる前記主論理ボリューム201Aの現在の状態を記録したスナップショットSSM₁を作成する(ステップ5052C)。前記ステップ5052Cは、たとえば

10

20

30

40

50

、図 5 に示すように、前記データ処理手段 202C で行う。

【0044】

前記ステップ 5052C の処理が済んだら、次に、前記副記憶装置システム 4 に対してリモートコピー開始の指示を送信するとともに、前記正ホストコンピュータ 1A から前記正記憶装置システム 2 の主論理ボリューム 201A へのデータの書き込みを再開する（ステップ 5052D）。前記ステップ 5052D は、まず、たとえば、図 6 に示すように、前記データ処理手段 202C から前記正記憶装置システム 2 の制御装置 202 内のリモートコピー制御手段 202B にリモートコピーの開始指示を送る。そして、前記正記憶装置システム 2 内のリモートコピー制御手段 202B から、前記ネットワーク 3 を介して、前記副記憶装置システム 4 の制御装置 402 内のリモートコピー制御手段 402A に前記リモートコピーの開始命令を送信する。また、前記リモートコピーは、前記正記憶装置システム 2 内のスナップショット SSM₁ を用いて行うので、前記スナップショット SSM₁ を作成した後であれば、前記主論理ボリューム 201A にデータを書き込んでも構わない。そのため、図 6 に示すように、前記正ホストコンピュータ 1A から転送されたデータや前記バッファ 202D に退避させておいたデータを前記主論理ボリューム 201A に書き込む。

10

【0045】

前記ステップ 5052D の処理が済んだら、次に、前記スナップショット管理手段 201B で管理しているスナップショット SSM₁ と SSM₂ の差分データを作成し、前記副記憶装置システム 4 に送信する（ステップ 5052E）。前記スナップショット SSM₂ は、今回のリモートコピーが開始される前は SSM₁、すなわち、前回のリモートコピーを行ったときの前記主論理ボリューム 201A の状態を記録したスナップショットである。つまり、今回のリモートコピーを開始する前の、前記副記憶装置システム 4 の記憶装置 401 内の論理ボリューム 401A は、前記スナップショット SSM₂ と同じ状態になっている。そのため、前記スナップショット管理手段 201B で管理しているスナップショット SSM₁ と SSM₂ の差分データ、すなわち、前回のリモートコピー実行時から今回のリモートコピー実行時まで更新されたデータのみを前記副記憶装置システム 4 に送信することで、データ転送量を減らすことができ、前記ネットワーク 3 にかかる負荷を低減することができる。

20

【0046】

一方、前記副記憶装置システム 4 の制御装置 402 では、図 7 に示すように、前記正記憶装置システム 2 からリモートコピー開始指示を待っている（ステップ 5054A）。そして、図 6 に示したように、前記副記憶装置システム 4 の制御装置 402 内のリモートコピー制御手段 402A が前記リモートコピーの開始指示を受信すると、その指示が前記データ処理手段 402B に転送され、リモートコピーが開始される。

30

【0047】

このとき、前記副記憶装置システム 4 の記憶装置 401 のスナップショット管理手段 401B が、スナップショット SSC_m（m = 1）を保持しているのであれば、まず、前記スナップショット SSC_m を SSC_{m+1} とする（ステップ 5054B）。そして、前記ステップ 5054B の処理が済んだら、次に、前記副記憶装置システム 4 の記憶装置 401 内の論理ボリューム 401A の現在の状態を記録したスナップショット SSC₁ を作成する（ステップ 5054C）。前記ステップ 5054B およびステップ 5054C の処理は、たとえば、図 8 に示すように、前記データ処理手段 402B を介して行う。また、前記ステップ 5054B およびステップ 5054C の処理を実行している間に前記正記憶装置システム 2 から送信された前記差分データを受信した場合、その差分データは、図 8 に示すように、バッファ 402C に一時的に退避させておく。

40

【0048】

前記ステップ 5054B およびステップ 5054C の処理が済んだら、次に、前記正記憶装置システム 2 から送信された前記差分データを前記論理ボリューム 401A に適用し、ファイルの追加、削除等を行う（ステップ 5054D）。前記ステップ 5054D は、

50

図 9 に示すように、前記データ処理手段 4 0 2 B を介して行われる。またこのとき、前記差分データの一部をバッファ 4 0 2 C に退避させていたのであれば、そのデータもあわせて適用する。

【 0 0 4 9 】

前記ステップ 5 0 5 4 D の処理、すなわち、前記正記憶装置システム 2 から送信された前記差分データの適用（書き込み）が完了したら、次に、前記正記憶装置システム 2 に対して、書き込み完了通知を返信する（ステップ 5 0 5 4 E ）。

【 0 0 5 0 】

前記副記憶装置システム 4 から送信された前記書き込み完了通知を受信した前記正記憶装置システム 2 では、図 4 に示すように、前記副記憶装置システム 4 に対して、リモートコピーの完了命令を送信する（ステップ 5 0 5 2 F ）。前記ステップ 5 0 5 2 F は、たとえば、図 1 0 に示すように、前記正記憶装置システム 2 内のリモートコピー制御手段 2 0 2 B が行う。そして、前記正記憶装置システム 2 内のリモートコピー制御手段 2 0 2 B は、リモートコピーの完了命令を送信したら、前記データ処理手段 2 0 2 C にリモートコピーが完了したことを通知する。前記リモートコピーが完了した後、前記正記憶装置システム 2 内のスナップショット管理手段 2 0 1 B で保持しているスナップショット SSM_n （ $n = 1$ ）のうち、次回のリモートコピーで差分データを作成するために必要な SSM_1 以外のスナップショット SSM_n （ $n = 2$ ）は不要である。そのため、前記リモートコピーが完了した後、前記データ処理手段 2 0 2 C は、最後に、図 4 および図 1 0 に示したように、前記正記憶装置システム 2 内のスナップショット管理手段 2 0 1 B で保持しているスナップショット SSM_n （ $n = 2$ ）のうち、不要なスナップショットを削除してもよい（ステップ 5 0 5 2 G ）。

【 0 0 5 1 】

一方、前記副記憶装置システム 4 では、前記ステップ 5 0 5 4 D およびステップ 5 0 5 4 E の処理が済んだ後、図 7 に示すように、前記正記憶装置システムからのリモートコピー完了命令を待つ（ステップ 5 0 5 4 F ）。そして、前記リモートコピー完了命令を受信したらリモートコピーが終了し、前記副記憶装置システム 4 は、次のリモートコピー開始指示を待つ。また、リモートコピーが正常に完了した場合、前記副記憶装置システム 4 内のスナップショット管理手段 4 0 1 B で保持しているスナップショット SSC_m （ $m = 1$ ）は不要である。そのため、図 7 および図 1 0 に示すように、前記副記憶装置システム 4 内のスナップショット管理手段 4 0 1 B で保持しているスナップショット SSC_m を削除してもよい（ステップ 5 0 5 4 G ）。

【 0 0 5 2 】

また、前記ステップ 5 0 5 4 D およびステップ 5 0 5 4 E の処理が済んだ後、たとえば、ネットワーク 3 の障害等により、前記正記憶装置システム 2 からリモートコピーの完了命令を受信できない場合がある。この場合、前記副記憶装置システム 4 内の論理ボリューム 4 0 1 A は、ファイルシステムとしての整合性を保証できない状態である。そのため、前記リモートコピー完了命令を受信できなかった場合は、たとえば、図 7 および図 1 1 に示すように、前記スナップショット SSC_1 を用いて、前記副記憶装置システム 4 内の論理ボリューム 4 0 1 A をリモートコピー開始前の状態、すなわちファイルシステムとしての整合性が取れている状態に復元する（ステップ 5 0 5 4 H ）。そして、必要に応じて、たとえば、前記ステップ 5 0 5 4 G の処理をした後、前記副記憶装置システム 4 は、次のリモートコピー開始指示を受信するまで待機する。

【 0 0 5 3 】

以上説明したように、本実施形態のデータコピーシステムを用いたデータコピー方法によれば、前記正記憶装置システム 2 内でリモートコピーの実行時期を判断することにより、正ホストコンピュータ 1 A にリモートコピーの実行時期を指示するためのソフトウェアやハードウェアを導入しなくてもよい。

【 0 0 5 4 】

また、リモートコピーの実行時期の判断基準として、前記正記憶装置システム 2 内の主

10

20

30

40

50

論理ボリューム 201A がファイルシステムとして整合性の取れている状態を用いることで、前記副記憶装置システム 4 内の論理ボリューム 401A のファイルシステムの整合性も保証することができる。そのため、前記正記憶装置システム 2 内のデータが破損した場合にも、前記副記憶装置システム 4 においてデータを容易に回復することができる。

【0055】

以上、本発明を、前記実施例に基づき具体的に説明したが、本発明は、前記実施例に限定されるものではなく、その要旨を逸脱しない範囲において、種々変更可能であることはもちろんである。

【0056】

たとえば、前記実施形態で説明した方法でリモートコピーを実行する場合、前記正記憶装置システム 2 の主論理ボリューム 201A の整合性が保証されている時点のうち、最新の時点の前記主論理ボリューム、言い換えると、最後にリモートコピーに成功したときの主論理ボリュームの状態を、磁氣的または電氣的、あるいは光学的な記録媒体に記録してもよい。前記正記憶装置システム 2 および前記副記憶装置システム 4 以外の部分で障害が発生した場合、障害から回復した後、前記正記憶装置システム 2 と前記副記憶装置システム 4 の間で、再度前記論理ボリューム 201A、401A を同期させる必要がある。そのため、前記最後にリモートコピーに成功したときの前記主論理ボリューム 201A の状態を前記記録媒体に記録しておけば、その時点から前記ホストコンピュータ 1A が変更したデータのみを前記副記憶装置システム 4 に送るだけで、前記正記憶装置システム 2 の主論理ボリューム 201A と前記副記憶装置システム 4 の論理ボリューム 401A のデータを一致させることができ、障害発生時の回復作業を高速化させることができる。

【図面の簡単な説明】

【0057】

【図 1】本発明による一実施形態のデータコピーシステムの概略構成を示す模式図である。

【図 2】本実施形態のデータコピーシステムを用いたデータのコピー方法を説明するための模式図であり、正記憶装置システムの制御手段の全体的な処理手順を示すフロー図である。

【図 3】本実施形態のデータコピーシステムを用いたデータのコピー方法を説明するための模式図であり、整合性を検査するステップにおけるシステム上の処理を示す図である。

【図 4】本実施形態のデータコピーシステムを用いたデータのコピー方法を説明するための模式図であり、リモートコピーを実行するステップにおける正記憶装置システムの処理手順を示すフロー図である。

【図 5】本実施形態のデータコピーシステムを用いたデータのコピー方法を説明するための模式図であり、スナップショットを作成するステップにおけるシステム上の処理を示す図である。

【図 6】本実施形態のデータコピーシステムを用いたデータのコピー方法を説明するための模式図であり、正記憶装置システムからデータを送信するステップにおけるシステム上の処理を示す図である。

【図 7】本実施形態のデータコピーシステムを用いたデータのコピー方法を説明するための模式図であり、リモートコピーを実行するステップにおける副記憶装置システムの処理手順を示すフロー図である。

【図 8】本実施形態のデータコピーシステムを用いたデータのコピー方法を説明するための模式図であり、副記憶装置システムでデータを受信するステップにおけるシステム上の処理を示す図である。

【図 9】本実施形態のデータコピーシステムを用いたデータのコピー方法を説明するための模式図であり、副記憶装置システムでデータを受信するステップにおけるシステム上の処理を示す図である。

【図 10】本実施形態のデータコピーシステムを用いたデータのコピー方法を説明するための模式図であり、リモートコピーが完了した後の処理の一例を示す図である。

10

20

30

40

50

【図 1 1】本実施形態のデータコピーシステムを用いたデータのコピー方法を説明するための模式図であり、リモートコピーが失敗した場合のデータの復元方法の一例を示す図である。

【符号の説明】

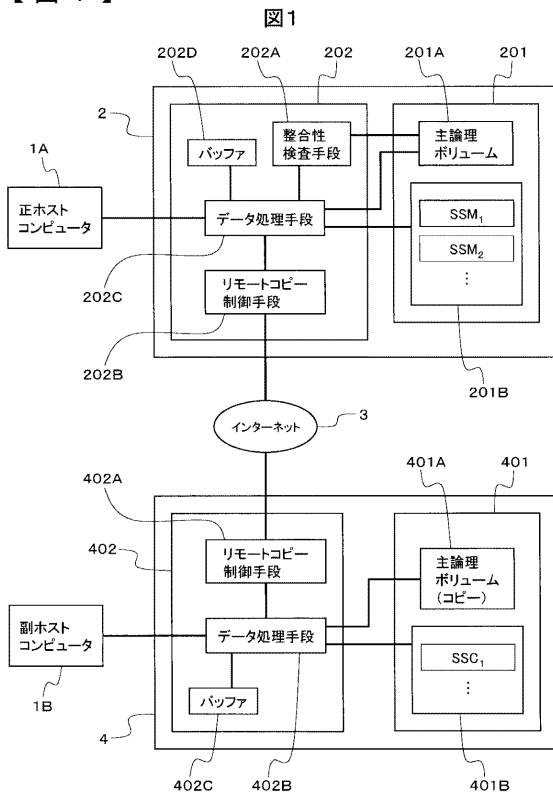
【0058】

- 1 A ... 正ホストコンピュータ
- 1 B ... 副ホストコンピュータ
- 2 ... 正記憶装置システム
- 2 0 1 ... 記憶装置
- 2 0 1 A ... 主論理ボリューム
- 2 0 1 B ... スナップショット管理手段
- 2 0 2 ... 制御装置
- 2 0 2 A ... 整合性検査手段
- 2 0 2 B ... リモートコピー制御手段
- 2 0 2 C ... データ処理手段
- 2 0 2 D ... バッファ
- 3 ... ネットワーク
- 4 ... 副記憶装置システム
- 4 0 1 ... 記憶装置
- 4 0 1 A ... 論理ボリューム
- 4 0 1 B ... スナップショット管理手段
- 4 0 2 ... 制御装置
- 4 0 2 A ... リモートコピー制御手段
- 4 0 2 B ... データ処理手段
- 4 0 2 C ... バッファ

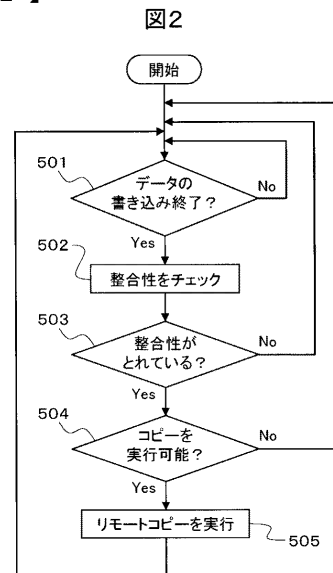
10

20

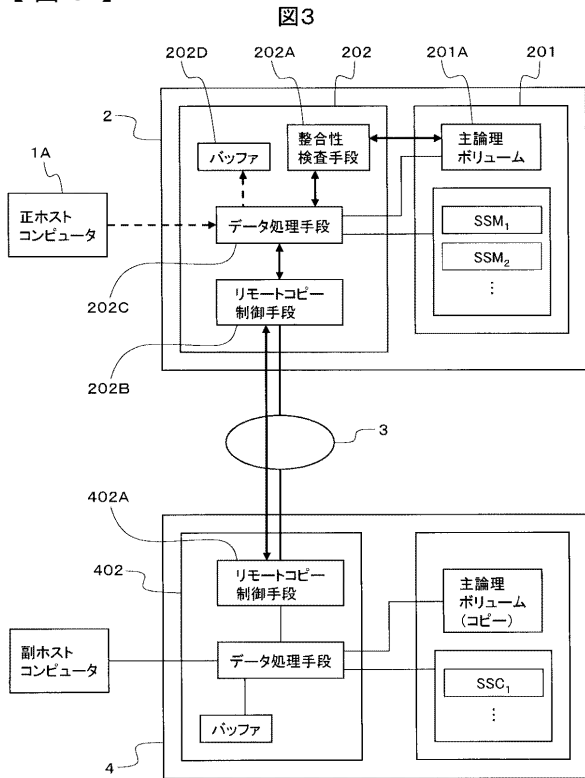
【図 1】



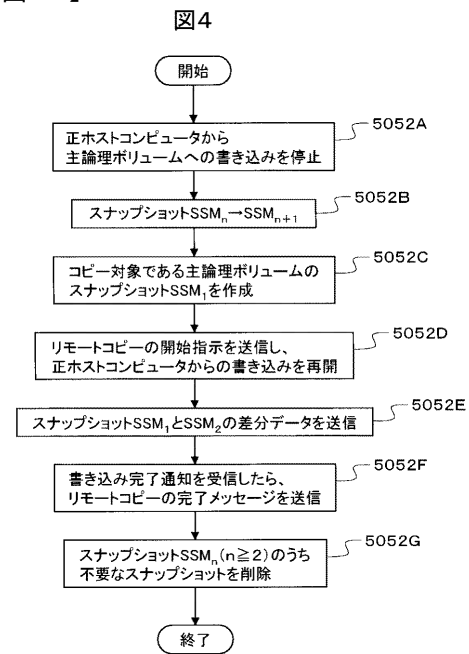
【図 2】



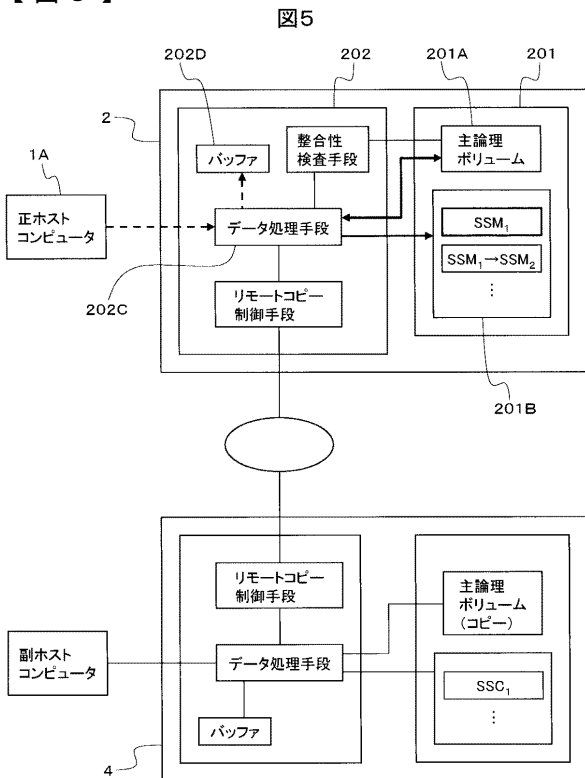
【図 3】



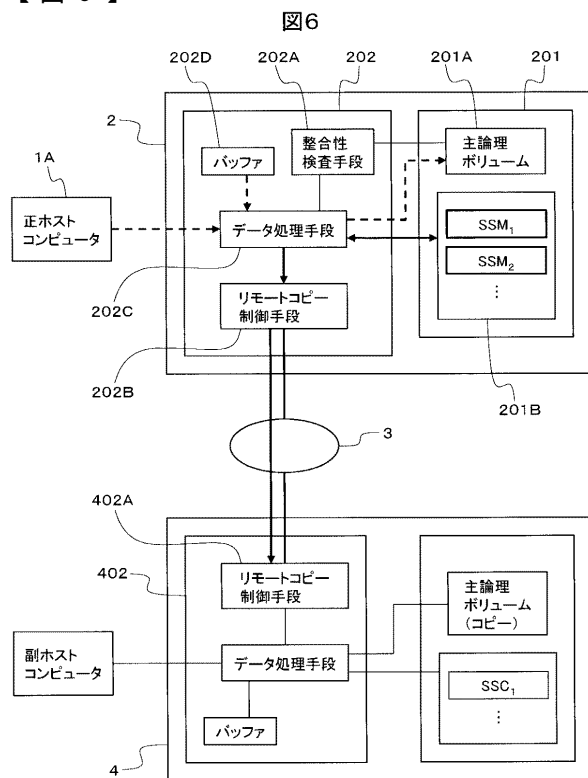
【図 4】



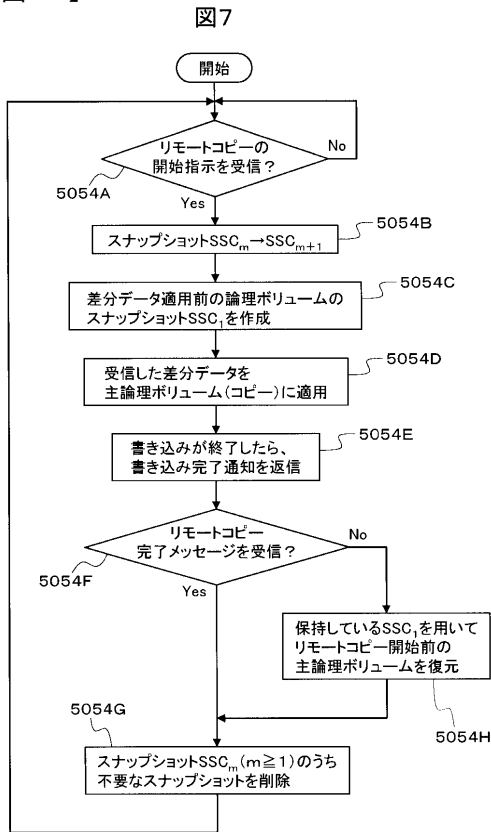
【図 5】



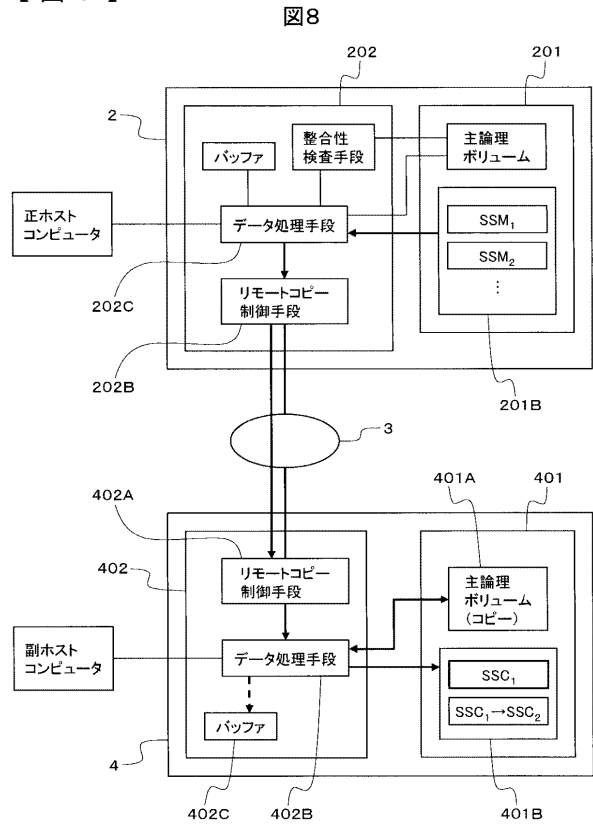
【図 6】



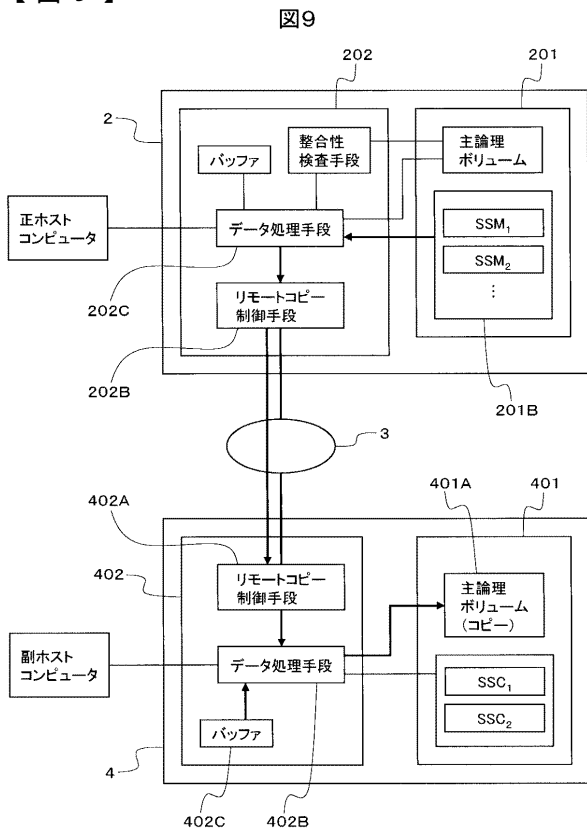
【図 7】



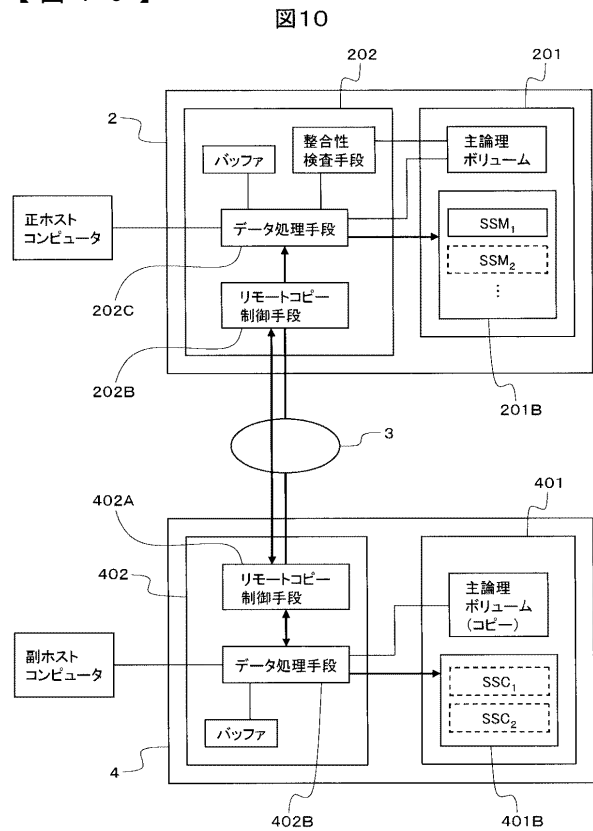
【図 8】



【図 9】



【図 10】



【図 11】

図 11

