(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第4294692号
(P4294692)

(45)発行日　平成21年7月15日(2009.7.15)　(24)登録日　平成21年4月17日(2009.4.17)

(51)Int.Cl.		F I		
G06F 12/00	(2006.01)	G06F	12/00	531R
G06F 3/06	(2006.01)	G06F	12/00	545A
G06F 13/10	(2006.01)	G06F	12/00	514E
		G06F	3/06	304Z
		G06F	3/06	540

請求項の数 1　(全 31 頁)　最終頁に続く

(21)出願番号　特願2007-2432(P2007-2432)
(22)出願日　平成19年1月10日(2007.1.10)
(62)分割の表示　特願2003-41986(P2003-41986)の分割
原出願日　平成15年2月20日(2003.2.20)
(65)公開番号　特開2007-179552(P2007-179552A)
(43)公開日　平成19年7月12日(2007.7.12)
審査請求日　平成19年2月6日(2007.2.6)

(73)特許権者　000005108
株式会社日立製作所
東京都千代田区丸の内一丁目６番６号
(74)代理人　110000062
特許業務法人第一国際特許事務所
(72)発明者　江口　賢哲
神奈川県川崎市麻生区王禅寺１０９９番地　株式会社　日立製作所　システム開発研究所内
(72)発明者　茂木　和彦
神奈川県川崎市麻生区王禅寺１０９９番地　株式会社　日立製作所　システム開発研究所内

最終頁に続く

(54)【発明の名称】情報処理システム

(57)【特許請求の範囲】
【請求項１】
計算機に接続し、制御部とキャッシュメモリと複数の記憶装置とを有する記憶装置システムにおけるジャーナルデータ生成方法であって、
前記複数の記憶装置から第一の論理記憶装置と第二の論理記憶装置とを構成する構成ステップと、
前記制御部にて、計算機から指示を取得し、ジャーナルモードとなるジャーナルモード開始ステップと、
前記制御部にて、前記計算機から複数のライト要求と前記複数のライト要求に対応する複数のライトデータを受信する、ライト要求受信ステップと、
前記複数のライト要求の対象がジャーナルモードの前記第一の論理記憶装置である場合の方法として、
前記制御部にて、前記複数のライト要求が前記第一の論理記憶装置の同じアドレスを指定した場合は前記複数のライトデータの各々を前記キャッシュメモリの異なる領域へ格納する、第一のライト要求格納ステップと、
前記制御部にて、前記複数のライト要求の各々に対応する複数のジャーナルデータを生成するためのジャーナルデータ生成領域を確保する、ジャーナルデータ生成確保ステップと、
前記制御部にて、前記キャッシュメモリの異なる領域に格納された前記複数のライトデータの各々を前記ジャーナルデータ生成領域に格納し、前記複数のライト要求が更新対象

とする従前データを前記キャッシュメモリ又は前記第一の論理記憶装置に対応する前記複数の記憶装置の一部から前記ジャーナルデータ生成領域へ格納することで、各々がライトデータ及び従前データを含む前記複数のジャーナルデータを前記キャッシュメモリに生成する、ジャーナルデータ生成ステップと、

前記ジャーナルデータ生成ステップとは非同期に、前記複数のジャーナルデータを前記キャッシュメモリから前記第二の論理記憶装置に対応する前記複数の記憶装置の一部へ書き込む、ジャーナルデータ格納ステップと、

を有し、

前記複数のライト要求の対象がジャーナルモードの前記第一の論理記憶装置でない場合、前記制御部にて、前記ライトデータの各々を前記キャッシュメモリに格納し、前記キャッシュメモリから前記ライトデータの各々を前記複数の記憶装置に格納する、第二のライト要求格納ステップと、

を有することを特徴とするジャーナルデータ生成方法。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、計算機や記憶装置システムを含む情報処理システムに関し、特に、障害などによって破壊された記憶装置システムに格納されたデータを復旧する情報処理システム及びそのデータ復旧方法に関する。

【背景技術】

【０００２】

情報処理システムで行われるオンライン処理やバッチ処理では、プログラムのバグや記憶装置システムの障害などによってこれらの処理が異常終了し、情報処理システムが有する記憶装置システムに格納されたデータが矛盾した状態になってしまうことがある。また、人為的ミスによって記憶装置システムに格納されたデータが消去されてしまうことも多い。

【０００３】

このような状態になった情報処理システムのデータを回復させる目的で、データの矛盾を解消して途中で止まった処理を再開させたり、あるいは、途中で止まった処理をもう一度やり直したりするための技術の一つとして、データのバックアップとリストアによるデータ回復技術がある。

【０００４】

バックアップおよびリストアに関する従来技術の一つが、特許文献１に開示されている。本文献には、ユーザが指定した時点における記憶装置システムに格納されたデータを、記憶装置システムに接続された計算機（以下「ホスト」）からのデータの入出力（以下「Ｉ／Ｏ」）を止めることなく磁気テープに複製し（以下データの複製を「データのバックアップ」と称する）、その複製されたデータ（以下、「バックアップデータ」）を用いてデータの回復（以下「リストア」）する技術が開示されている。

【０００５】

一方、特許文献２には、データのリストアにかかる時間を短縮するために、データのバックアップが実行された後、データが更新された個所についての情報を差分情報として保持し、記憶装置システムに格納されたデータをバックアップデータでリストアする際に、バックアップデータのうち、差分情報で示されるデータの部分のみをデータのリストアに用いる技術が記載されている。

【特許文献１】米国特許番号５，２６３，１５４号公報

【特許文献２】特開２００１－２１６１８５号公報

【発明の開示】

【発明が解決しようとする課題】

【０００６】

特許文献１に記載されたリストア処理では、磁気テープからバックアップデータを読み

出す際、バックアップデータを取得した時点から更新されていない部分（記憶装置システムのデータと磁気テープのデータの内容が一致している部分）も磁気テープから読み出され、記憶装置システムに書き込まれる。このようなデータの転送は、無駄が多く、リストアに要する時間を長びかせる。

【０００７】

一方、特許文献２に開示されている技術では、特許文献１に比べ、重複したデータの読み出しが発生しない分、リストアに係る時間は少なくなる。しかし、双方の技術をもってしても、データのバックアップの後から記憶装置システムが故障するまでの間に更新されたデータについては、データのリストアを行うことができない。データのバックアップ後に更新されたデータまでリストアしようとすると、そのデータの更新の内容等をホスト側がログ等で管理する必要があり、ホストへの負荷が大きく、かつ処理に長い時間がかかる。

【０００８】

本発明の目的は、障害発生前までの任意の時点におけるデータのリストア処理を高速に行う記憶装置システム並びに情報処理システムを提供することである。

【課題を解決するための手段】

【０００９】

上記目的を達成するために、本発明は以下の構成を有する。すなわち、計算機及び計算機に接続された記憶装置システムを有する情報処理システムであり、記憶装置システムは制御部及び複数の記憶装置を有する。そして、記憶装置システムは、所定の指示にしたがって、一つの記憶装置に格納されたデータを他の記憶装置に複製する。その後、記憶装置システムは、複製元となった記憶装置へのデータ更新を更新履歴として他の記憶装置に格納する。一方、計算機は、複製が作成された後の任意の時間において、ある識別情報を作成し、記憶装置システムへ送信する。識別情報を受信した記憶装置システムは、その識別情報を更新履歴と関連させて記憶装置へ格納する。

【００１０】

データを復元させたい場合、計算機は、記憶装置システムへ識別情報を送信する。識別情報を受信した記憶装置システムは、記録した識別情報から受信した識別情報と一致する識別情報を検索する。一致する識別情報を発見したら、記憶装置システムは、複製先の記憶装置に格納されたデータと、一致した識別情報と関連付けられる更新履歴より前に記録された更新履歴の内容を用いて、複製元の記憶装置にデータを復元する。

【００１１】

尚、本発明では、記憶装置システムへデータの更新を要求する計算機は、識別情報を作成する計算機と異なる構成も考えられる。

【００１２】

また、本発明では、識別情報を作成する計算機は、その識別情報を自計算機に格納する構成も考えられる。

【００１３】

更に、本発明では、計算機に格納された識別情報に関する情報をユーザに提示し、ユーザの指定した識別情報を記憶装置システムへ送信する構成も考えられる。

【００１４】

更に、本発明の構成として、以下が考えられる。すなわち、中央処理装置を備えた計算機と、記憶装置を備えた記憶装置システムとを有する構成とする。計算機は、記憶装置システムに対して記憶装置に格納されているデータの複製の作成保存を要求する手段、計算機の処理によるデータの更新部分の記録を要求する手段、及びシステムのある時点の状態を識別する識別情報を記憶装置システムに送信する手段とを保持する。記憶装置システムは、計算機の要求に応答して、記憶装置のデータの複製を作成保存する手段、記憶装置の内容が更新されたときに更新前後のデータ及び更新場所をジャーナルデータとして保存する手段、計算機より送信される識別情報を保持識別する手段、並びにジャーナルデータと識別情報を関連付ける手段を有する。更に、計算機は、記憶装置の内容をある時点の状態

に復旧する必要が生じた場合、状態識別情報を指定してデータの復旧要求を記憶装置システムに送信する手段を有し、記憶装置システムは送信された状態識別情報を識別し、前記データの複製とジャーナルデータを用いてデータをリストアする手段を有する。

【００１５】

更に本発明は、以下の構成を有する。即ち、計算機及び記憶装置システムで一つの識別情報を共有し、記憶装置システムではその識別情報と更新履歴を関連付けて管理し、計算機の指示に応じて、特定の識別情報で示される更新履歴まで、記憶装置に格納されたデータを復元するデータの復元方法である。

【発明の効果】

【００１６】

本発明によれば、記憶装置システムに格納されたデータを復旧する場合に、ホストに負担をかけず、短時間でデータを所定の状態までリストアすることができる。また、ユーザは、任意のシステム状態までデータをリストアすることができる。

【発明を実施するための最良の形態】

【００１７】

以下、図面を用いて、本発明の第一の実施形態について説明する。尚、これにより本発明が限定されるものではない。以下、「記憶装置システム」には、ディスク装置等の記憶装置、ディスクアレイ等のように複数の記憶装置を有するシステムが含まれるものとする。

【００１８】

図１は、本発明を適用した情報処理システムの第一の実施形態を示す図である。情報処理システムは、ホスト１、記憶装置システム２、管理端末３、ホスト１と記憶装置システム２とを接続するネットワーク４、並びにホスト１、記憶装置システム２及び管理端末３とを接続するネットワーク５を有する。

【００１９】

ホスト１は、パーソナルコンピュータ、ワークステーション、メインフレーム等の計算機である。ホスト１では、その計算機の種類に応じたオペレーティングシステム（以下「ＯＳ」）と様々な業務、用途に対応したアプリケーションプログラム（ＡＰ）、たとえばデータベース（ＤＢ）プログラム等、が動作する。本実施形態では、簡単のため、ホスト１を２つ記載しているが、ネットワーク４及び５に接続されるホスト１は幾つあってもよい。

【００２０】

管理端末３は、記憶装置システム２の障害、保守、構成、性能情報等の管理を行うために使用される計算機である。例えば、情報処理システムの管理者が、記憶装置システム２に論理的な記憶装置を設定する場合、データをバックアップするための記憶領域を設定する場合、又はデータを複製する際の記憶領域の対を設定する場合に、管理端末３が使用される。情報処理システムの管理者は、記憶装置システム２の保守・管理、記憶装置システム２が有する物理記憶装置１０の設定、及び記憶装置システム２と接続されるホスト１の設定等を行う場合に、管理端末３に設定したい内容を入力する。管理端末３は、ネットワーク５を介して記憶装置システム２及びホスト１に管理者が入力した内容を送信する。

【００２１】

ネットワーク４は、ホスト１が記憶装置システム２へＩ／Ｏの処理要求等を伝送するために使用される。ネットワーク４には、光ケーブルや銅線等が用いられる。又、ネットワーク４で使用される通信プロトコルには、イーサネット（登録商標）、ＦＤＤＩ、ファイバチャネル、ＳＣＳＩ、Ｉｎｆｉｎｉｂａｎｄ、ＴＣＰ／ＩＰ、ｉＳＣＳＩなどがある。

【００２２】

ネットワーク５は、記憶装置システム２が、自身の障害、保守、構成、性能等の管理情報を管理端末３やホスト１に送信したり、管理端末３やホスト１が、記憶装置システム２から管理情報を取得する際に使用される。ネットワーク５に使用されるケーブル及び通信プロトコルはネットワーク４と同一でも異なっていてもよい。

【００２３】
図２は、本実施形態における記憶装置システム２の構成を示す図である。記憶装置システム２は、ホスト１が使用するデータやプログラムを格納し、ホスト１のＩ／Ｏ処理要求を受信し、Ｉ／Ｏ処理要求に対応した処理を行い、その結果を所定のホスト１に送信する。

【００２４】
記憶装置システム２は、記憶装置制御装置１１、物理記憶装置１０、キャッシュメモリ１４、共有メモリ１９及びＬｏｃａｌ　Ｎｅｔｗｏｒｋ１８とを有する。

【００２５】
物理記憶装置１０には、ユーザが使用するデータが格納される。物理記憶装置１０は、電気的に不揮発な記憶媒体である磁気ディスクや不揮発性半導体メモリで構成される、シリコンディスク、光ディスク、光磁気ディスク又はハードディスク等である。尚、物理記憶装置１０は、物理記憶装置１０が有する記憶領域に障害がおきてもデータが損失しないように、冗長性を持つＲＡＩＤ（Ｒｅｄｕｎｄａｎｃｙ　Ａｒｒａｙ　Ｉｎｄｅｐｅｎｄｅｄ　Ｄｉｓｋ）構成になっていてもよい。

【００２６】
記憶装置制御装置１１は、ホスト１からのＩ／Ｏ要求の処理及び物理記憶装置１０の制御を行う装置である。記憶装置制御装置１１は、物理記憶装置１０と接続される物理記憶装置アダプタ１３、所定のプログラムを実行するプロセッサ１２、プロセッサ１２で実行されるプログラム、プログラムが動作する上で必要な情報、記憶装置システム２の設定情報及び構成情報等が格納される不揮発性メモリ１５、記憶装置システム２とネットワーク５とを接続するためのネットワークアダプタ１７、記憶装置システム２とネットワーク４とを接続するためのＩ／Ｏネットワークアダプタ１６とを有する。

【００２７】
尚、記憶装置制御装置１１は記憶装置システム２に複数存在しても良い。また記憶装置システム２の冗長性を確保するために、システム内の各装置、例えば、記憶装置制御装置１１内の各構成要素への電源供給のための回路、キャッシュメモリ１４、不揮発性メモリ１５、Ｌｏｃａｌ　Ｎｅｔｗｏｒｋ１８、物理記憶装置アダプタ１３等は、夫々２重化された冗長構成になっていても良い。

【００２８】
キャッシュメモリ１４は、記憶装置システム２にホスト１から入力されるデータ又は記憶装置システム２からホスト１へ転送されるデータが一時的に格納される記憶媒体である。

【００２９】
共有メモリ１９は、複数の記憶装置制御装置１１、複数のプロセッサ１２間で共有される情報を格納するための不揮発性メモリである。例えばＩ／Ｏ処理のためにキャッシュメモリ１４のある領域へアクセスを行うための排他処理用ビットや物理記憶装置１０とキャッシュメモリ１４との対応関係を示す情報等が格納される。Ｌｏｃａｌ　Ｎｅｔｗｏｒｋ１８は、記憶装置制御装置１１、キャッシュメモリ１４、及び物理記憶装置１０を相互に接続する。Ｌｏｃａｌ　Ｎｅｔｗｏｒｋ１８は、共有バス型の構成でもよいし、スター型等のネットワーク構成となっても良い。

【００３０】
図３は、ホスト１の構成を示す図である。ホスト１は、所定のプログラムを実行するプロセッサ２０、プロセッサ２０が実行するＯＳやＡＰ及びＡＰが使用するデータを格納するために使用されるメモリ２１、ＯＳやＡＰ、ＡＰが使用するデータが格納されるローカルディスク装置２２、ネットワーク４とホスト１とを接続するホストバスアダプタ２３、ネットワーク５とホスト１とを接続するためのネットワークアダプタ２４、フロッピー（登録商標）ディスク等の可搬記憶メディアからのデータの読み出し等を制御するリムーバブル記憶ドライブ装置２６、及びこれらの構成部品間を接続し、ＯＳやＡＰのデータや制御データの転送に用いられるＬｏｃａｌ　Ｉ／Ｏ　Ｎｅｔｗｏｒｋ２５とを有する。

【００３１】
リムーバブル記憶ドライブ装置２６で使用される可搬記憶媒体としては、ＣＤ－ＲＯＭ、ＣＤ－Ｒ、ＣＤ－ＲＷ、ＤＶＤやＭＯ等の光ディスク、光磁気ディスクや、ハードディスクやフロッピー（登録商標）ディスク等の磁気ディスク等がある。尚、以下に説明される各プログラムは、可搬記憶媒体からリムーバブル記憶ドライブ装置２６を介して読み出されることで、あるいはネットワーク４又は５を経由することで、ホスト１のローカルディスク装置２２にインストールされる。
【００３２】
ホスト１は、冗長性確保のために、プロセッサ２０等の構成部品を複数有していても良い。
【００３３】
図４は、記憶装置システム２が有するプログラムの構成及びシステムの論理的構成を示す図である。記憶装置システム２では、単数又は複数の物理記憶装置１０（図で点線で表示）が組み合わされ、冗長性を有するパリティグループ４０７が構成される。パリティグループ４０７は、データを格納する物理記憶装置１０及び格納されたデータから作成される冗長データが格納される物理記憶装置１０の組である。また、記憶装置システム２は、ホスト１に対して、パリティグループ４０７を構成する複数の物理記憶装置１０が作る記憶領域空間から、論理的な記憶領域を論理記憶装置４０８として提供する。したがって、ホスト１は、記憶装置システム２には、図４に示すような、記憶装置制御装置１１に接続された記憶装置（論理記憶装置４０８）が存在すると認識する。
【００３４】
記憶装置制御装置１１は、記憶装置システム２内の処理を制御するために、Ｉ／Ｏ処理プログラム４０３、レプリケーション制御処理プログラム４０４、ストレージサブシステム構成管理プログラム４０２、リストア制御処理プログラム４０６及びジャーナル制御部４０５の各プログラムを不揮発性メモリ１５に有する。記憶装置制御装置１１は、これらのプログラムをプロセッサ１２で実行することで、以下に説明する処理を制御する。
【００３５】
Ｉ／Ｏ処理プログラム４０３は、更に、コマンド処理プログラム４１５及びリードライト処理プログラム４１６からなる。記憶装置制御装置１１は、ホスト１からのＩ／Ｏ処理要求をネットワークインターフェース１７で受信すると、コマンド処理プログラム４１５を実行して、受信したＩ／Ｏ処理要求の内容を解析する。解析の結果、Ｉ／Ｏ処理要求の内容がデータの読み出しＩ／Ｏ（以下「リードＩ／Ｏ」）要求やデータの書き込みＩ／Ｏ（以下「ライトＩ／Ｏ」）処理要求であれば、記憶装置制御装置１１は、リードライト処理プログラム４１６を実行する。
【００３６】
ライトＩ／Ｏ処理要求の場合、記憶装置制御装置１１は、ホスト１からのライトＩ／Ｏ処理要求に対する応答処理（実際にホスト１から転送されるデータを受領できる状態にあるかどうかの応答）を行い、更に転送されてくる更新用のデータ（以下「ライトデータ」）をキャッシュメモリ１４又は物理記憶装置１０の所定の箇所への書き込み、またはキャッシュメモリ１４に格納されたライトデータを物理記憶装置１０に書き込む制御等を行う。リードＩ／Ｏ処理要求の場合、記憶装置制御装置１１は、リードＩ／Ｏ処理要求に対応するデータ（以下、「リードデータ」）を、キャッシュメモリ１４もしくは物理記憶装置１０の所定の箇所から読み出してホスト１に転送したり、物理記憶装置１０からリードデータを読み出してキャッシュメモリ１４に格納する処理を制御する。
【００３７】
その他の処理の場合、たとえばＳＣＳＩのＩｎｑｕｉｒｙコマンド（デバイスサーチを指示するコマンド）等の場合、記憶装置制御装置１１は、コマンド処理プログラム４１５を実行することによって、処理内容に対応した動作の制御を行う。
【００３８】
ストレージサブシステム構成管理プログラム４０２は、デバイス管理情報４１０及びデ

バイス管理プログラム４０９から構成される。デバイス管理情報４１０は、論理記憶装置４０８のアドレスと物理記憶装置１０のアドレスとの対応関係を示すマッピング情報、パリティグループ４０７を構成する物理記憶装置１０に関する情報、スナップショットペア４５０に関する情報、及びジャーナルデータ格納対象情報等とを保持するテーブルである。

【００３９】

デバイス管理プログラム４０９は、記憶装置制御装置１１がデバイス管理情報４１０を管理する際に実行されるプログラムである。記憶装置制御装置１１は、デバイス管理プログラム４０９を実行することによって、管理端末３等から入力される論理記憶装置４０８の定義やスナップショットが格納される対象となる論理記憶装置４０８の設定、ジャーナルデータ格納対象情報の登録等を行う。

【００４０】

記憶装置制御装置１１がデータのリードライトＩ／Ｏ処理を実行する際は、デバイス管理プログラム４０９を実行することによって、リードライトＩ／Ｏ処理要求が指定するリード又はライトデータが読み出され又は格納されるべき個所の論理記憶装置４０８のアドレスがどの物理記憶装置１０のアドレスに対応するかを計算し、その結果に基づいて、物理記憶装置１０へのアクセスを行う。

【００４１】

ジャーナル制御プログラム４０５は、記憶装置制御装置１１がジャーナルデータを作成する際に実行するジャーナル作成プログラム４１９、記憶装置制御装置１１が作成したジャーナルデータを読み出す際に実行するジャーナル読出しプログラム４２０、ジャーナル取得の対象となる論理記憶装置４０８についての情報が登録されたジャーナル管理情報４１８、及び記憶装置制御装置１１がジャーナル管理情報４１８の設定等を行う際に実行するジャーナル管理プログラム４１７から構成される。

【００４２】

記憶装置制御装置１１は、ジャーナルデータ取得を行うとき（以下、「ジャーナルモード時」）にホスト１からライトＩ／Ｏ処理要求を受信した場合、ジャーナル作成プログラム４１９を実行することで、ライトデータをキャッシュメモリ１４に書き込むとともに、ライトデータの格納される個所に存在している従前のデータ（以下「ライト対象データ」）及びライトデータを、キャッシュメモリ１４に確保されたジャーナルデータ作成用の所定の領域に書き込む。

【００４３】

尚、キャッシュメモリ１４に格納されたライト対象データ及びライトデータは、更新履歴であるジャーナルデータとして、ジャーナルデータを格納するための論理記憶装置４０８（以下「ジャーナル論理記憶装置」）に格納される。又、記憶装置制御装置１１は、リストアマネージャ４０６及びジャーナル読み込みプログラム４２０を実行することで、ホスト１からの指示に基づき、ジャーナル論理記憶装置に格納されたジャーナルデータを順次読み出し、読み出したジャーナルデータが有するアドレスで示される、複製先となる論理記憶装置４０８又は複製元である論理記憶装置４０８の記憶領域にデータを上書きする。

【００４４】

スナップショット制御プログラム４０４は、コピー処理プログラム４１３、差分情報４１４、ペア制御管理プログラム４１１及びペア管理情報４１２から構成される。記憶装置制御装置１１は、ペア制御管理プログラム４１１を実行することで、ホスト１からの指示に従って、ある論理記憶装置４０８（以下、「正論理記憶装置」）及び正論理記憶装置に格納されたデータの複製を格納する論理記憶装置４０８（以下、「副論理記憶装置」）について、ペア形成（Ｐａｉｒ　Ｃｒｅａｔｅ）、ペア分離（Ｐａｉｒ　Ｓｐｌｉｔ）、ペア再結合（Ｐａｉｒ　Ｒｅｓｙｎｃ）、ペア削除（Ｐａｉｒ　Ｄｅｌｅｔｅ）の処理を行う。ここで、「ペア」とは、正論理記憶装置と、正論理記憶装置に対応する副論理記憶装置の組（以下「スナップショットペア４５０」）を指す。

【００４５】
尚、１つの正論理記憶装置に対して、複数の副論理記憶装置を設定・作成することもできる。また、副論理記憶装置を新たな正論理記憶装置とし、新たな正論理記憶装置とペアになる副論理記憶装置を設定・作成することもできる。
【００４６】
ペア管理情報４１２には、ある論理記憶装置のスナップショットペア４５０がペア結合状態（Ｐａｉｒ　Ｄｕｐｌｅｘ）のペア同期状態（Ｐａｉｒ　Ｓｙｎｃｈｒｏｎｕｓ）、ペア結合状態（Ｐａｉｒ　Ｄｕｐｌｅｘ）のペア非同期状態（Ｐａｉｒ　Ａｓｙｎｃｈｒｏｎｕｓ）、ペア形成状態（Ｐａｉｒ　Ｃｒｅａｔｅ）、ペア分離状態（Ｐａｉｒ　Ｓｙｍｐｌｅｘ）にあるかどうかを示す情報が登録される。Ｐａｉｒ　Ｓｙｎｃｈｒｏｎｕｓ状態とは、ホスト１のライトＩ／Ｏによる正論理記憶装置の更新と副論理記憶装置の更新が同期して行われる状態を示す。Ｐａｉｒ　Ａｓｙｎｃｈｒｏｎｕｓ状態とは、ホスト１のライトＩ／Ｏによる正論理記憶装置の更新と副論理記憶装置の更新が非同期に行われる状態を示す。尚、Ｐａｉｒ　Ａｓｙｎｃｈｒｏｎｕｓ状態の場合は、副論理記憶装置に正論理記憶装置の更新が反映されるまで、ライトデータは、差分情報４１４で管理される。
【００４７】
差分情報４１４には、あるペアがペア非同期状態（Ｐａｉｒ　Ａｓｙｎｃｈｒｏｎｕｓ）又は分離状態（Ｐａｉｒ　Ｓｙｍｐｌｅｘ）の場合に、正論理記憶装置にデータの書き込みが発生することによって生ずる正論理記憶装置と副論理記憶装置との間の差異が有る部分を示すアドレス情報等が保持される。
【００４８】
記憶装置制御装置１１は、コピー処理プログラム４１３を実行することによって、ペア作成（Ｐａｉｒ　Ｃｒｅａｔｅ）時に正論理記憶装置の先頭アドレスから順次副論理記憶装置にデータを複写することで、正論理記憶装置に格納されたデータを副論理記憶装置にバックアップする。さらに記憶装置制御装置１１は、差分情報４１４を参照して、差異が有る部分のデータを正論理記憶装置から副論理記憶装置にコピーしたり、逆に、差分情報４１４を参照して、差異があるデータを副論理記憶装置から正論理記憶装置へコピーする。
【００４９】
バックアップ／リストア制御プログラム４０６は、リストアプログラム４２１とバックアッププログラム４２２から構成される。記憶装置制御装置１１は、リストアプログラム４２１を実行することで、ホスト１からのリストア要求に基づいて、指定された論理記憶装置４０８のデータをリストアする。尚、リストア処理の詳細は後述する。
バックアッププログラム４２２は、記憶装置制御装置１１が、ホスト１の指示等に従って、論理記憶装置４０８の複製を作成したり、記憶装置システム２のデータを他の記憶装置、例えばテープに転送したりする際に実行される。
【００５０】
図５は、ホスト１で動作するプログラム及び使用されるデータの例を示す図である。これらのプログラムは、ホスト１のローカルディスク装置２２又はメモリ２１に格納され、プロセッサ２０で実行される。ホスト１は、ＯＳ５００の下で動作するＡＰとして、データベースマネジメントソフトウエア（以下「ＤＢＭＳ」）５０１を有する。ＤＢＭＳ５０１は、ＯＳ５００、ファイルシステム（ＦＳ）５３０、ボリュームマネージャ（ＶＭ）５４０等を介して記憶装置システム２にアクセスする。また、ＤＢＭＳ５０１は、ユーザが使用する他のＡＰ５２０との間で、トランザクション処理等のＩ／Ｏ処理の遣り取りを行う。
【００５１】
ＤＢＭＳ５０１は、ＤＢファイル５０５、ＬＯＧファイル５０６、ＩＮＤＥＸファイル５０７、ＤＢバッファ５０９、ＬＯＧバッファ５１０、デバイス情報ファイル５１１、状態ファイル５０８、ＤＢ定義ファイル５１２、トランザクションマネージャ５０２、ログマネージャ５０３、バッファマネージャ５１３、及びリソースマネージャ５０４から構成

されている。

【００５２】

ＤＢバッファ５０９は、ＤＢＭＳ５０１の処理性能を上げる目的で、ホスト１のメモリ２１に確保されるＤＢＭＳ５０１専用の領域である。このバッファ５０９には、ＤＢＭＳ５０１によって良くアクセスされるデータが一時的に保持される。ログバッファ５１０もＤＢバッファ５０９と同様にメモリ２１上に確保された領域で、ＤＢＭＳ５０１の処理記録（以下「ログ」）が一時的に格納される。

【００５３】

ＤＢファイル５０５は、ＤＢのテーブル等ＤＢのデータそのものであり、実際には記憶装置システム２の物理記憶装置１０内に格納されている。そして、良く使用されるテーブル等のデータがＤＢバッファ５０９に一時格納され、ＤＢＭＳ５０１は、そのデータでトランザクション処理を行う。ＤＢバッファ５０９に要求されるデータが無い場合、ＤＢＭＳ５０１は、データを記憶装置システム２から読み上げる。

【００５４】

ログファイル５０６も実際には記憶装置システム２の物理記憶装置１０に格納されている。ログファイル５０６には、トランザクション処理等のＤＢＭＳ５０１がＤＢに対して行った処理のログ（処理を行ったＡＰの識別子、処理順序識別子、処理を行った時間や処理を行ったデータ及び処理対象前データ等を含む）が順次記録される。記録の際には、ログバッファ５１０を用いて順次追記される。ログファイル５０６には、ＡＰ５２０が一連の処理を行い整合性が取れた状態でコミットした際及びＤＢＭＳ５０１が一定時間間隔やトランザクション数等毎に物理記憶装置１０にバッファに格納されたダーティデータを格納するシンク処理を行った際にも、それを示す情報が記録される。

【００５５】

ホスト１は、トランザクションマネージャ５０２を実行することで、ＤＢに対するトランザクション処理や、ログファイル５０６に格納されたデータを読出してデータのリカバリを実行したり、チェックポイントの制御を行ったりする。又、ホスト１は、ログマネージャ５０３を実行することで、ＤＢに対するデータの入出力を制御する。

【００５６】

以下、本実施形態の動作概要について説明する。本実施形態の情報処理システムでは、まず、記憶装置システム２において、正論理記憶装置と正論理記憶装置のある時点に有するデータのバックアップデータ（以下「スナップショットデータ」）を有する副論理記憶装置を作成し保持する。スナップショットが作成された時点以降にホスト１からのライトＩ／Ｏ処理要求がある度に、記憶装置システム２は、ライトＩ／Ｏ処理前後のデータ（ライトデータ及びライト対象データ）をジャーナルデータ（「更新履歴」）として記録する。

【００５７】

さらに、ホスト１は、自身が作成する任意の識別情報であるチェックポイント情報（以下「ＣＰ情報」）を記憶装置システム２に対して通知する。具体的には、ホスト１は、任意の時点、例えば記憶装置システム２との間でのデータを一致させる処理（シンク処理）時に、ＣＰ情報を記憶装置システム２のジャーナルデータに書込む。これにより、記憶装置システム２は、ホスト１で作成されたものと同一のＣＰ情報を保持する。つまり、従来ホスト１でのみ管理されていたＣＰ情報をホスト１と記憶装置システム２の双方で管理する。これによって、ホスト１が指示するＣＰ情報及び記憶装置システム２内のジャーナルデータに格納されたＣＰ情報を利用して、記憶装置システム２は、ホスト１が意図した時（ＣＰ情報作成時）の記憶装置システム２が有していたデータの状態に高速にリストアを行う。

【００５８】

このような処理を実行するために、ホスト１は、あらかじめ、ジャーナルデータを取得する準備指示（ジャーナル取得開始準備指示）、及びジャーナル取得開始指示を記憶装置システム２に送信する。これにより、記憶装置システム２は、ジャーナルデータの取得を

開始し、ジャーナルモードとなる。その後、情報処理システムは、上述したＣＰ情報の遣り取りを行う。

【００５９】

以下、ホスト１がジャーナル取得開始準備指示を記憶装置システム２に発行した際に、記憶装置システム２で行われる処理について説明する。

【００６０】

ジャーナル取得開始準備指示指示には、ジャーナル論理記憶装置を指定する情報や、正論理記憶装置及び副論理記憶装置の作成指示等が含まれる。ジャーナル取得開始準備指示を受領した記憶装置システム２は、指示に従い、データ格納領域の割当等を実行する。正副論理記憶装置は、ジャーナル開始準備指示を受領する前からスナップショットペア４５０になっていても良いが、本実施形態では、記憶装置システム２が、ジャーナル取得開始準備指示に基づいて新たに論理記憶装置４０８をスナップショットペア４５０に設定する。

【００６１】

記憶装置システム２は、次に、正論理記憶装置のスナップショットデータを指定された副論理記憶装置に作成する。具体的には、記憶装置システム２がジャーナル取得開始準備指示を受取った時点で正論理記憶装置に格納されているデータを副論理記憶装置に複製し、正論理記憶装置と副論理記憶装置の状態を同期させる。尚、ジャーナル取得開始準備指示以前から正論理記憶装置とスナップショットペア４５０になっている副論理記憶装置が指定された場合は、記憶装置システム２は、副論理記憶装置と正論理記憶装置とを同期させた状態にするだけで良い。

【００６２】

更に、記憶装置システム２は、ホスト１の指示に基づいて、正論理記憶装置に対応するジャーナル論理記憶装置の設定も行う。

【００６３】

次に、ホスト１は、記憶装置システム２に、ジャーナル取得開始指示を出す。ジャーナル取得開始指示には、ジャーナルデータ取得開始を示す最初のＣＰ情報であるチェックポイント識別子（以下「ＣＰＩＤ」）が含まれている。記憶装置システム２は、受信した最初のＣＰＩＤを記録し、その後、ジャーナルデータの取得を開始する。尚、その後にホスト１から送信されるチェックポイントコマンドにも最初のＣＰＩＤとは別のＣＰＩＤが含まれている。ＣＰＩＤは、記憶装置システム２でジャーナルデータとして記録される。

【００６４】

図６は、ホスト１からジャーナル取得開始準備指示及びジャーナル取得開始指示を受領した記憶装置システム２における処理の詳細手順を示す図である。

【００６５】

ホスト１は、ＤＢＭＳ５０１を実行することで、記憶装置システム２に対して、ジャーナル取得開始準備指示を送信する。尚、本実施形態では、ＤＢＭＳ５０１が使用するＤＢのテーブルが格納された論理記憶装置４０８が正論理記憶装置として指定される。ジャーナル取得開始準備指示には、正論理記憶装置を示す識別子、ジャーナル取得開始準備指示を記憶装置システム２が受領した瞬間のある正論理記憶装置に格納されたデータのスナップショットデータを格納するための副論理記憶装置を示す識別子、ジャーナル論理記憶装置を示す識別子が含まれる（ステップ６０１）。

【００６６】

ジャーナルデータは、スナップショットデータが作成された後のライトＩ／Ｏ処理要求に基づくライト対象データ、ライトデータ及びこれらのデータの正論理記憶装置内における格納位置を示すアドレス情報等から構成される。構成の具体例は後述する。

【００６７】

尚、スナップショットデータが格納される副論理記憶装置やジャーナル論理記憶装置の設定は、ジャーナル取得開始準備指示とは別の指示に基づいて、予め行われていても良い。この場合、ジャーナル取得開始準備指示には、これらの論理記憶装置４０８を示す識別

子は含まれなくても良い。

【００６８】

ホスト１からジャーナル取得開始準備指示を受領した記憶装置制御装置１１は、指示に含まれていれている副論理記憶装置を示す識別子を用いてデバイス管理情報４１０を参照し、無効なデバイスの指定の有無、例えば、指定された副論理記憶装置の存在の有無や障害発生の有無、論理記憶装置の状態の確認、例えば指定された副論理記憶装置が、既に他の処理に使用されている等、の確認を行う。確認の結果、指定された副論理記憶装置が使用可能である場合、記憶装置制御装置１１は、指定された副論理記憶装置がジャーナル作成中であること示す情報をデバイス管理情報４１０に設定するとともに、指定された副論理記憶装置に関するジャーナル管理情報をジャーナル管理情報４１８に設定し、かつＰａｉｒ　Ｃｒｅａｔｅ状態と設定する（ステップ６０２）。

【００６９】

同様に、記憶装置制御装置１１は、ジャーナル論理記憶装置を示す識別子を用いてデバイス管理情報４１０を参照し、指定されたジャーナル論理記憶装置の無効なデバイスの指定の有無及び状態の確認を行う。指定されたジャーナル論理記憶装置が使用できる場合、指定されたジャーナル論理記憶装置がジャーナル作成中とする情報をデバイス管理情報４１０に登録する（ステップ６０３）。

【００７０】

次に、記憶装置制御装置１１は、副論理記憶装置に正論理記憶装置のスナップショットデータを作成する処理（以下「スナップショット作成処理」）を行う。スナップショット作成処理においては、ジャーナル取得開始準備処理指示のコマンド受領時に正論理記憶装置に格納されていたデータが、副論理記憶装置に順次転送される。尚、ジャーナル取得開始準備処理指示に副論理記憶装置の指示が含まれず、予めＤｕｐｌｅｘ状態のＰａｉｒである副論理記憶装置が管理端末３で指定されていた場合や、副論理記憶装置の指示が含まれていても、指定された副論理記憶装置が既に正論理記憶装置とＤｕｐｌｅｘ状態にある場合は、スナップショット作成処理は行わなくても良い。

【００７１】

尚、記憶装置システム２がスナップショット作成処理を実行している最中に、ホスト１から正論理記憶装置に格納されたデータに対するライトＩ／Ｏ処理要求があった場合、記憶装置制御装置１１は、要求時点でライト対象データが未だ副論理記憶装置にコピーされていなかったら正論理記憶装置にライトデータを書込み、要求時点で既にライト対象データが副論理記憶装置にコピーされていたら、ライトデータを正論理記憶装に書き込むとともに、副論理記憶装置にも書きこむ（ステップ６０４）。

【００７２】

スナップショット作成処理が終了したら、記憶装置制御装置１１は、ペア管理情報をＤｕｐｌｅｘ状態にし（ステップ６０５）、ジャーナル取得準備処理の完了を、ジャーナル取得開始準備指示を発行したホスト１に報告する。尚、Ｄｕｐｌｅｘ状態にあるスナップショットペア４５０では、正論理記憶装置に書き込まれたデータは、副論理記憶装置にも反映される（ステップ６０６）。

【００７３】

ジャーナル取得準備処理の完了報告を受領したホスト１は、任意のタイミング、例えば情報処理システムの状態が整合性が取れている時、指定時間又はあるトランザクション処理の前や後で、ジャーナル取得開始指示を記憶装置システム２に送信する（ステップ６０７）。

【００７４】

ジャーナル取得開始指示を受領した記憶装置制御装置１１は、先に準備したジャーナル論理記憶装置、正副論理記憶装置に障害が発生していないかを確認して、ジャーナル取得開始指示に対してＲｅａｄｙ応答を返す（ステップ６０８）。

【００７５】

その後、記憶装置制御装置１１は、正副論理記憶装置をＰａｉｒ　Ｓｐｌｉｔ状態にす

る。具体的には、ホスト１からライトＩ／Ｏ処理要求を受取っても、正論理記憶装置の更新が副論理記憶装置には一切反映されない状態にする（ステップ６０９）。

【００７６】

一方、Ｒｅａｄｙ応答を受領したホスト１は、チェックポイントコマンドを用いて、ＣＰＩＤを含むＣＰ情報を送信する（ステップ６１０）。

【００７７】

ＣＰ情報を受領した記憶装置システム２は、ジャーナル論理記憶装置に、受信したＣＰ情報、具体的には、ＣＰＩＤ、記憶装置システム２内の処理シーケンス番号及び処理時間をジャーナルデータとして格納する。もしくは、記憶装置制御装置１１にある不揮発性メモリ１５又は共有メモリ１９にＣＰ情報を格納する（ステップ６１１）。

【００７８】

チェックポイントコマンドを送信したホスト１は、ホスト１のメモリ２１に格納されているライトデータを記憶装置システム２に送信する（ステップ６１２）。

【００７９】

ライトデータを受領した記憶装置制御装置１１は、ライトデータを正論理記憶装置に書き込むと共に、ライト対象データ及びライトデータをジャーナル論理記憶装置に書きこむ（ステップ６１３）。

【００８０】

チェックポイントコマンド受領以降、記憶装置システム２はジャーナルデータの取得を継続するジャーナルモードとなる。また、これ以降、一定時間毎や一定トランザクション数毎等、ＤＢ管理者が設定した間隔で、ホスト１は、その時点にＤＢバッファ５０９上のデータ全てを記憶装置システム２に送信する。更に、記憶装置システム２とホスト１とでＣＰ情報を共有するタイミングである場合には、ホスト１は、ＣＰ情報を共有するタイミングであることを示すチェックポイントコマンドを送信する。

【００８１】

ジャーナルモード中にチェックポイントコマンドを受領した記憶装置制御装置１１は、ＣＰ情報をジャーナルデータとして、ジャーナル論理記憶装置、不揮発性メモリ１５又は共有メモリ１９に格納する。

【００８２】

図７は、ジャーナルモード中の記憶装置システム２が、ホスト１よりリードライトＩ／Ｏ処理要求を受信した場合の処理手順を示す図である。

【００８３】

ホスト１よりリードまたはライトＩ／Ｏ処理要求を受領した記憶装置システム２の記憶装置制御装置１１は（ステップ７０１）、受信した処理要求がライトＩ／Ｏ処理要求であるかどうかを判断する（ステップ７０２）。ライトＩ／Ｏ処理要求でない場合、記憶装置制御装置１１は、デバイス管理情報４１０を用いて、リードＩ／Ｏ処理要求の対象となっているリードデータを、対応する物理記憶装置１０又はキャッシュメモリ１４から読み出してＩ／Ｏインタフェース１６を介してホスト１に転送する（ステップ７０９）。

【００８４】

ステップ７０２でライトＩ／Ｏ処理要求と判断した場合は、記憶装置制御装置１１は、デバイス管理情報４１０を参照し、ライトＩ／Ｏ処理要求で指定される論理記憶装置４０８が、ジャーナルモードである正論理記憶装置であるかを判断する（ステップ７０３）。ジャーナルモードの正論理記憶装置でなければ、記憶装置制御装置１１は、キャッシュメモリ１４にライトＩ／Ｏ処理要求に伴うライトデータを格納する領域を確保する（ステップ７０７）。その後、記憶装置制御装置１１は、ライトデータをキャッシュメモリ１４の確保された領域に格納して、ライトＩ／Ｏ処理が終了したことをホスト１に通知する（ステップ７０８）。

【００８５】

尚、記憶装置制御装置１１は、キャッシュメモリ１４から物理記憶装置１０にデータを格納した後にライトＩ／Ｏ処理の終了をホスト１に報告してもよく、又ライトデータをキ

ャッシュメモリ１４を介さず直接物理記憶装置１０に格納してもよい。

【００８６】

一方、ステップ７０３でライトＩ／Ｏ処理の対象となる論理記憶装置４０８がジャーナルモードの正論理記憶装置であった場合、記憶装置制御装置１１は、ライトデータを格納するための領域をキャッシュメモリ１４に確保し、ホスト１から送信されるライトデータを当該領域に格納する。

【００８７】

尚、通常の論理記憶装置４０８へのライトデータの書き込みとは違い、記憶装置制御装置１１は、同じアドレスが指定される複数のライトデータの連続した書き込みの際は、各々のライトデータをキャッシュメモリ１４の異なる領域に格納しなければならない。これは、ライトＩ／Ｏ処理要求の対象となるライト対象データがキャッシュメモリ１４に存在するが物理記憶装置１０にそのライトデータが反映されていない場合、通常の書き込み処理の様にキャッシュメモリ１４に存在するライト対象データを更新してしまうと、更新前のライト対象データが失われ、ライト対象データをジャーナル論理記憶装置に格納することができなくなるからである（ステップ７０５）。その後、記憶装置制御装置１１は、ジャーナルデータの作成処理を行い、処理を終了する（ステップ７０６）。

【００８８】

図８は、図７のステップ７０６のジャーナルデータ作成処理の手順を示す図である。ライトデータをキャッシュメモリ１４に格納した記憶装置制御装置１１は、ジャーナルデータを一時的に格納するための領域をキャッシュメモリ１４に確保する（ステップ９０１）。

【００８９】

その後、記憶装置制御装置１１は、キャッシュメモリ１４に格納されているライトデータを、ＣＰ情報、処理シーケンス番号、処理時間とともに、キャッシュメモリ１４に確保されたジャーナルデータ格納用の領域にコピーする（ステップ９０２、９０３）。ただし、ＣＰ情報のＣＰＩＤ１００７エントリには、ホスト１からのチェックポイントコマンド受領時にのみＣＰＩＤが格納されるので、それ以外の場合は、ＣＰＩＤ１００７エントリには無効データが格納される。処理シーケンス番号は、プロセッサ１２が処理を行うごとに付ける処理通番号である。

【００９０】

同時に、記憶装置制御装置１１は、キャッシュメモリ１４に格納されたライトデータによって更新されるライト対象データを格納するための領域をキャッシュメモリ１４に確保し、そのライト対象データを物理記憶装置１０あるいはキャッシュメモリ１４から読みだして、キャッシュメモリ１４の確保された記憶領域に格納する（ステップ９０４、９０５）。これにより、ライトデータ、ライト対象データ、ＣＰ情報、処理シーケンス番号及び処理時間を含むジャーナルデータが作成される。

【００９１】

全ての処理が終了した後、記憶装置制御装置１１は、図７の処理に戻る。尚、キャッシュメモリ１４で作成されたジャーナルデータは、キャッシュメモリ１４にジャーナルデータが作成されるのとは非同期に、キャッシュメモリ１４から物理記憶装置１０に書き込まれる（ステップ９０６）。

【００９２】

図９は、ジャーナルデータのデータ形式を示す図である。

【００９３】

ジャーナルデータは、図６で説明したように、ジャーナル取得開始指示受信後、記憶装置システム２が正論理記憶装置に対するライトＩ／Ｏ処理要求を処理する毎にキャッシュメモリ１４上に作成され、その後物理記憶装置１０に格納される。ジャーナルデータは、ホスト１と記憶装置システム２でシステムの状態を一意に識別するＣＰ情報を格納するエントリ１００１、データが更新される箇所を示すブロックアドレスが格納されるエントリ１００２、更新に用いられるライトデータの長さが格納されるエントリ１００３、データ

が更新される個所に格納されていたライト対象データが格納されるエントリ１００４、及びライトデータが格納されるエントリ１００５とから構成される。ＣＰ情報エントリ１００１には更に、チェックポイントフラグエントリ１００６、ＣＰＩＤが格納されるエントリ１００７、処理順序番号エントリ１００８、及び時刻エントリ１００９が含まれている。

【００９４】

記憶装置システム２がホスト１よりチェックポイントコマンドを受領してＣＰ情報を受信した場合、記憶装置制御装置１１は、受信した際に作成されるジャーナルデータのＣＰ情報エントリ１００１に含まれるチェックポイントフラグエントリ１００６に「ＯＮ」を示す情報を登録し、ＣＰＩＤエントリ１００７に、送信されてきたＣＰＩＤを格納する。ＣＰＩＤエントリ１００７に格納されるＣＰＩＤは、ホスト１が管理するログファイルに記録されているＣＰ情報に含まれる特定のＣＰＩＤと対応する一意の値を持っている。したがって、ホスト１があるＣＰＩＤを指定すると、指定されたＣＰＩＤに対応する、ジャーナルデータに格納されたＣＰＩＤを指定することができる。

【００９５】

図１０は、ホスト１がＣＰ情報を記憶装置システム２に送信する処理手順を示す図である。ホスト１は、チェックポイントコマンドを発行し記憶装置システム２にＣＰ情報を送信することによって、ＤＢが有するデータの状態を確定しログファイルにチェックポイントを記録した（ＣＰＩＤ等の情報が記録される）ことを記憶装置システム２に通知することが出来る。

【００９６】

先ず、ホスト１は、ＤＢバッファ５０９及びログバッファ５１０等メモリ２１にあるバッファに格納されたデータを、記憶装置システム２へ強制的に書き込むためのライトＩ／Ｏ処理要求を記憶装置システム２に送信する。本処理によって、ホスト１は、これらのバッファにのみ格納されていて記憶装置システム２には格納されていないデータ（以下「ダーティデータ」）を記憶装置システム２に反映して、ＤＢのデータを確定することができる（ステップ１１０１）。

【００９７】

ライトＩ／Ｏ処理要求を受信した記憶装置制御装置１１は、ホスト１から送信されるデータをキャッシュメモリ１４に書き込む（ステップ１１０２）。転送されたデータを全てキャッシュメモリ１４に書き込んだら、記憶装置制御装置１１は、ライトＩ／Ｏ処理の終了をホスト１に通知する。この際、記憶装置制御装置１１は、これらのデータに対応するジャーナルデータの作成も行う（ステップ１１０３）。

【００９８】

尚、ライトＩ／Ｏ処理の終了の通知を受信したホスト１は、以下のステップで実行されるＣＰＩＤ書き込み処理の完了が記憶装置システム２から報告されるまでは、記憶装置システム２へのデータの書き込みを行わないが、データの読み出しは実行してもよい。

【００９９】

ライトＩ／Ｏ処理の終了が通知されたホスト１は、トランザクションマネージャ５０２を実行して、ＣＰ情報及びＣＰ処理に用いられるログを作成する。具体的には、ログファイル５０６にＣＰＩＤ等のＣＰ情報をログとして格納する。尚、ログのＣＰ情報には、ＣＰＩＤ、リソースマネージャの数、リソースマネージャの状態、動作中のトランザクションの数及び各々のトランザクション記述なども含まれる。尚、リソースマネージャに関しては、詳細を割愛する（ステップ１１０４～１１０５）。同時に、ホスト１は、チェックポイントコマンドを記憶装置システム２に対して発行する。チェックポイントコマンドにはＣＰＩＤが含まれている（ステップ１１０５）。

【０１００】

ホスト１からのチェックポイントコマンドを受信した記憶装置システム２は（ステップ１１０６）、受信したＣＰＩＤをジャーナルデータとしてジャーナル論理記憶装置に記録する。この場合、ジャーナルデータのエントリ１００４及び１００５に対応するライト対

象データ及びライトデータは存在しないので、これらのエントリには、データが格納されないか、無効データ（例えば－１）が格納される（ステップ１１０７）。記録が完了したら、記憶装置制御装置１１は、記録の完了をホスト１に通知する（ステップ１１０８）。

【０１０１】

ホスト１は、記憶装置システム２からＣＰＩＤ記録完了の報告を受領すると、ＣＰ情報に関する処理を終了する（ステップ１１０９）。

【０１０２】

図１１は、管理端末３やホスト１からリストア指示を受領した記憶装置システム２における処理手順を示す図である。尚、以下の処理は、記憶装置制御装置１１が、リストアプログラム４２１を実行することで行われる。

【０１０３】

本実施形態では、ＤＢを使用するＡＰ５４０のバグやユーザのオペレーションミス等により論理記憶装置４０８にホスト１にとって論理的不整合等の障害が起き、かつ障害が発生した論理記憶装置４０８がジャーナルモードの正論理記憶装置であった場合を考える。この場合、管理端末３又はホスト１からは、障害が発生した正論理記憶装置に対応する副論理記憶装置及びジャーナル論理記憶装置に格納されたデータを使用して記憶装置システム２内で正論理記憶装置に格納されたデータをリストアする指示が送信される。

【０１０４】

ホスト１は、ＡＰ５４０のログ情報等を参照し、オペミスや誤ったデータを送信したＡＰ等の誤った操作を起こした時点を解析し、その時点の直前のチェックポイントコマンド送信時を検索し、記憶装置システム２でリストアする際に使用されるＣＰＩＤを決定する。尚、ホスト１のユーザは、障害発生直前のＣＰＩＤではなく、ホスト１からＣＰ情報を記憶装置システム２に送信する際にホスト１に記録されるＣＰＩＤのリストから、任意のＣＰＩＤを選択することができる。これにより、本システムのユーザは、任意のＣＰＩＤを選択することで、選択されたＣＰＩＤが作成された時点に記憶装置システム２の正論理記憶装置が格納していたデータの状態まで、正論理記憶装置に格納されたデータをリストアすることができる（ステップ１２０１）。

【０１０５】

次に、ホスト１は、ステップ１２０１で選択したＣＰＩＤまでのデータのリストア処理要求を記憶装置システム２に発行する。リストア処理要求には、リストア処理の対象となる正論理記憶装置の識別子（例えばＷＷＮとＬＵＮ等）、正論理記憶装置に対応する副論理記憶装置を指定する識別子、ジャーナル論理記憶装置を指定する識別子、及び選択されたＣＰＩＤの情報等が含まれる。尚、正論理記憶装置に対応する副論理記憶装置が複数有る場合は、その内のいずれかを指定する情報もリストア処理要求に含まれる（ステップ１２０２）。

【０１０６】

ホスト１より発行されたリストア処理要求を受領した記憶装置制御装置１１は、リストアプログラム４２１を実行して、リストア処理要求に含まれる副論理記憶装置を示す識別子とペア管理情報４１２を比較参照し、指定された副論理記憶装置が正論理記憶装置に対する正しい副論理記憶装置であるかを確認する。また同様に、リストア処理要求に含まれるジャーナル論理記憶装置を示す識別子とジャーナル管理情報とを比較参照し、指定されたジャーナル論理記憶装置が正論理記憶装置に対応する正しいジャーナル論理記憶装置であるかを確認する（ステップ１２０３）。

【０１０７】

更に、記憶装置制御装置１１は、リストア処理要求の内容から、正論理記憶装置にリストア処理を行うのか、副論理記憶装置にリストア処理を行うのか、もしくは異なった未使用の論理記憶装置４０８にリストア処理を行うのかを確認する。尚、リストア処理対象に正論理記憶装置が指定されていても、正論理記憶装置が使用不可能であれば、論理記憶装置４０８の障害により処理続行が出来ない旨をホストに通知し、処理を中止する。また同様に副論理記憶装置やその他の論理記憶装置４０８にデータをリストアする指示であって

も、指定された論理記憶装置４０８に何らかの障害がある場合は障害により処理続行が出来ない旨をホストに通知し、処理を中止する（ステップ１２０４）。

【０１０８】

正論理記憶装置もしくはその他の空き論理記憶装置４０８にリストア処理を行う場合、記憶装置制御装置１１は、副論理記憶装置に格納されていたスナップショットデータを先頭から順次読み出して正論理記憶装置へコピーし、正論理記憶装置が有するディスクイメージを副論理記憶装置と同一にする。尚、副論理記憶装置にデータをリストアする場合は、本コピー処理は不要である（ステップ１２０６）。

【０１０９】

副論理記憶装置からのコピー処理が終了したら、あるいは副論理記憶装置へデータをリストアする場合、記憶装置制御装置１１は、キャッシュメモリ１４にデータ格納領域を確保する。その後、記憶装置制御装置１１は、正論理記憶装置に対応するジャーナル論理記憶装置の先頭から、具体的には、処理シーケンス番号順に、順次ジャーナルデータをキャッシュメモリ１４に確保された領域に読み出す。尚、ジャーナル論理記憶装置からのジャーナルデータの読み出しの先頭は、ホスト１から指定されても、記憶装置システム２が処理シーケンス番号で特定しても良い（ステップ１２０７）。

【０１１０】

その際、読み出されたジャーナルデータにＣＰ情報が含まれるかどうかを確認する。具体的には、ジャーナルデータのチェックポイントフラグ１００６がＯＮになっているかどうかを確認する（ステップ１２０８）。

【０１１１】

読み出されたジャーナルデータが、ＣＰ情報を含むジャーナルデータである場合、記憶装置制御装置１１は更に、読み出されたジャーナルデータのＣＰＩＤ１００７に含まれるＣＰＩＤがホスト１から指定されたＣＰＩＤかどうかを確認する（ステップ１２０９）。

【０１１２】

ＣＰＩＤ１００７に含まれるＣＰＩＤがホスト１から指定されたＣＰＩＤでない場合又はＣＰＩＤ１００７にＣＰＩＤが格納されていない場合（チェックポイントフラグがＯＮになっていない場合）、記憶装置制御装置１１は、読み出したジャーナルデータのアドレス１００２に格納された情報から、読み出されたジャーナルデータが、指定されたリストア対象である正論理記憶装置に関するジャーナルデータであるかどうかを確認する（ステップ１２１０）。

【０１１３】

読み出されたジャーナルデータがリストア対象の正論理記憶装置に関するジャーナルデータであれば、記憶装置制御装置１１は、読み出されたジャーナルデータに含まれるライトデータを、正論理記憶装置又は副論理記憶装置の対応するアドレスに書き込む。ただし、ＣＰＩＤに対応するジャーナルデータである場合には、ライトデータが存在しないので、データの書き込みは行われない（ステップ１２１１）。

【０１１４】

その後、記憶装置制御装置１１は、ステップ１２０７に戻り次のジャーナルデータの読出し処理を行う。また、ステップ１２１０で読み出されたジャーナルデータが指定された正論理記憶装置に対応するジャーナルデータでない場合、記憶装置制御装置１１は、ジャーナルデータをリストア先である論理記憶装置４０８に書き込まずに、ステップ１２０７の処理に戻る。以下、記憶装置制御装置１１は、ステップ１２０７～１２１１の処理を繰り返すことで、指示されたＣＰＩＤまでのジャーナルデータをリストアする。

【０１１５】

ステップ１２０９で、ＣＰＩＤ１００７のＣＰＩＤが指定されたＣＰＩＤと一致した場合、記憶装置制御装置１１は、リストアすべきデータをすべて正論理記憶装置、副論理記憶装置や他の論理記憶装置４０８に書き込んだと判断して、リストア処理の終了をホスト１に通知する。尚、正論理記憶装置以外にリストア処理を行う場合は、ホスト１への通知前に、論理物理マッピング情報を書き換えて、正論理記憶装置と副論理記憶装置またはそ

の他のリストア先となる論理記憶装置４０８とを交換し、ホスト１からアクセスする論理記憶装置４０８の識別子（たとえばＦＣのＷＷＮとＬＵ番号の組合せ）は変わらないようにする（ステップ１２１２）。

【０１１６】

尚、正論理記憶装置毎にジャーナル論理記憶装置が割り当てられている場合には、前記ステップ１２１０の処理、すなわち、読み出されたジャーナルデータと正論理記憶装置との対応関係の確認は不要である。

【０１１７】

ホスト１または管理端末３は、記憶装置システム２から終了報告を受領したら、ホスト１が指定したＣＰＩＤ時点までのデータが回復されたと判断して、他の処理を継続する（１２１３）。

【０１１８】

図１２は、デバイス管理情報４１０の一例を示した図である。

【０１１９】

デバイス管理情報４１０は、論理記憶装置４０８のアドレス情報を登録するエントリ１３０１及び物理記憶装置１０のアドレス情報を登録するエントリ１３０４とを有するテーブル１３００、ホスト１に提供される論理記憶装置番号を登録するエントリ１３３１、記憶装置システム２で論理記憶装置４０８を統一的に識別する記憶装置内論理記憶装置番号を登録するエントリ１３３２、記憶装置システム２内で管理するＰａｒｉｔｙ　Ｇｒｏｕｐの通し番号を登録するエントリ１３３３、論理記憶装置４０８のペア情報を登録するエントリ１３３４及びジャーナル情報を登録するエントリ１３３５を有するテーブル１３３０並びに、記憶装置システム２内の論理記憶装置番号が登録されるエントリ１３５１、空き／リザーブ情報が登録されるエントリ１３５２、Ｐａｔｈ定義情報が登録されるエントリ１３５３、Ｅｍｕｌａｔｉｏｎ　Ｔｙｐｅ／サイズが登録されるエントリ１３５４及び障害情報が登録されるエントリ１３５５とを有するテーブル１３５０とを保持する。

【０１２０】

テーブル１３００のエントリ１３０１は、更に、ホスト１に提供される論理記憶装置４０８の番号が登録されるエントリ１３１１、その論理記憶装置４０８に対応する内部アドレスが登録されるエントリ１３１２、記憶装置システム２内部で論理記憶装置４０８を統一的に識別する論理記憶装置番号が登録されるエントリ１３１３及びその内部論理記憶装置アドレスが登録されるエントリ１３１４を有する。また、テーブル１３００のエントリ１３０４は、更に、エントリ１３０１に登録された論理記憶装置４０８に対応する物理記憶装置１０のＰａｒｉｔｙＧｒｏｕｐ４０７の番号を登録するエントリ１３２１、物理記憶装置１０の番号を登録するエントリ１３２２及びその物理記憶装置１０のアドレス情報を登録するエントリ１３２３を有する。

【０１２１】

テーブル１３３０のペア情報エントリ１３３４には、論理記憶装置４０８がスナップショットペア状態にあるかどうかを示す情報が登録される。ジャーナル対象モードエントリ１３３５には、論理記憶装置４０８がジャーナル取得の対象、すなわちジャーナルモードの対象であるかどうかを示す情報が登録される。

【０１２２】

テーブル１３５０の空き／リザーブ情報エントリ１３５２には、論理記憶装置４０８が、副論理記憶装置やジャーナル論理記憶装置に用いるために予約されている状態にあるかを示す情報が登録される。リザーブ情報が登録されている論理記憶装置４０８は、その他の用途、例えば新たに業務用論理記憶装置として割り当てるなどが出来ない。Ｐａｔｈ定義情報エントリ１３５３には、論理記憶装置４０８がホスト１に提供されるために外部に公開されているかを示す情報が登録される。例えばＩ／Ｏ　ＮｅｔｗｏｒｋがＦＣだったら、論理記憶装置４０８とＦＣのＰｏｒｔとの関連付けに関する情報が登録される。

【０１２３】

Ｅｍｕｌａｔｉｏｎ　Ｔｙｐｅエントリ１３５４には、論理記憶装置４０８がＯＳが認

識できる記憶装置のいずれに擬似化されている（エミュレートされる）かを示す情報及びその記憶容量が登録される。例えば、具体的には、オープン系システムのＯＳが認識できる記憶装置であることを示す「ＯＰＥＮ」や、メインフレーム系のＯＳが認識できる記憶装置であることを示す「３９９０」等の情報が登録される。

【０１２４】

障害情報エントリ１３５５には、論理記憶装置４０８が何らかの障害になったかどうかを示す情報が登録される。ここで、障害とは、主に論理記憶装置４０８が存在する物理記憶装置１０の物理的障害や管理者が意識的に記憶装置システム２を閉塞状態にした場合等の論理的障害がある。

【０１２５】

図１３は、ペア管理情報情報４１２のテーブルの一例を示した図である。

【０１２６】

ペア管理情報４１２は、ホスト１に提供される論理記憶装置番号を登録するエントリ１４０１、記憶装置システム２内での論理記憶装置番号を登録するエントリ１４０２、Ｅｍｕｌａｔｉｏｎ　Ｔｙｐｅ／サイズを登録するエントリ１４０３、ペア状態を登録するエントリ１４０４、世代情報を登録するエントリ１４０５及びペア管理情報を登録するエントリ１４０６とを有する。

【０１２７】

ペア状態エントリ１４０４には、先に記したペア結合状態等のペアの状態を示す情報が登録される。ペア管理情報エントリ１４０６には、論理記憶装置４０８が正論理記憶装置か副論理記憶装置かを示す情報が登録される。論理記憶装置４０８が正論理記憶装置に指定されていれば、正側エントリ１４１１には０が登録され、対応する副側エントリ１４１２にはペアとなる副論理記憶装置の番号を示す値が登録される。一方、論理記憶装置４０８が副論理記憶装置に指定されていれば、副側エントリ１４１１には０の値が登録され、対応する正側エントリ１４１２にはペアとなる正論理記憶装置の番号を示す情報が登録される。

【０１２８】

また、論理記憶装置４０８が正副論理記憶装置として指定されていない場合には、正側エントリ１４１１及び副側エントリ１４１２の双方に無意味な値を示す「－１」が登録される。また、論理記憶装置４０８がスナップショットペア４５０のカスケード構成の真ん中、すなわち、一つのペアの副論理記憶装置でもあり、同時に他のペアの正論理記憶装置である場合は、正側エントリ１４１１、副側エントリ１４１２双方にペアを形成する他方の論理記憶装置４０８の番号を示す情報が登録される。また、正側エントリ１４１１、副側エントリ１４１２に複数の論理記憶装置番号が登録される場合もある。

【０１２９】

図１４は、ジャーナル管理情報４１８の一例を示した図である。

【０１３０】

ジャーナル管理情報４１８は、テーブル１５００及びＣＰ情報を管理するためのジャーナル管理テーブル１５２０を有する。テーブル１５００は、ＣＰＩＤを格納するエントリ１５０１、エントリ１５０１に格納されたＣＰＩＤが記録されたジャーナルデータが格納された位置を示すアドレスが登録される１５０２及びエントリ１５０１に格納されたＣＰＩＤがジャーナル論理記憶装置に記録された時間を示す時間情報１５０３とを有する。また、ジャーナル管理テーブル１５２０は、デバイス番号を登録するエントリ１５２１ごとに、ＣＰＩＤを登録するエントリ１５２２及びチェックポイント管理テーブルの格納アドレスを登録するエントリ１５２３を有する。

【０１３１】

次に、第二の実施形態として、ホスト１ではなく、管理端末３と記憶装置システム２との間でＣＰ情報を共有し、記憶装置システム２に障害が起きた場合のデータのリカバリを行う場合について述べる。

【０１３２】

本実施形態では、ホスト１が記憶装置システム２との間のログやチェックポイントを管理するプログラム、例えばＤＢＭＳ５０１を有しない場合に、ホスト１にエージェントというプログラムを導入する。以下エージェントが導入されたホストをホスト１’と称する。

【０１３３】

図２２は、ホスト１’が有するプログラムの構成を例示した図である。ホスト１と異なる点は、ＤＢＭＳ５０１が存在せず、替わりにＡｇｅｎｔプログラム２２００が含まれている点である。Ａｇｅｎｔプログラム２２００は、モード情報２２１０、ＦＳ　Ａｇｅｎｔ８２０、Ｉ／Ｏ制御プログラム２２３０、チェックポイントＡｇｅｎｔ２２５０、ＶＭ　Ａｇｅｎｔ２２４０、及び構成管理Ａｇｅｎｔ２２６０から構成されている。

【０１３４】

モード情報２２１０には、ホスト１’が管理端末３から受信した、スナップショットを取る時期やジャーナルデータを取る期間の状態が、モード情報として保持されている。ＦＳ　Ａｇｅｎｔ２２２０は、ＦＳ５３０に対してファイルの排他制御やファイルを閉じる処理を指示し、かつＦＳ５３０が管理するダーティデータをメモリ２１のアドレスとして管理する際に実行される。

【０１３５】

ＶＭ　Ａｇｅｎｔ２２４０は、ＶＭ５４０に対して、ＶＭ５４０で設定される論理記憶領域への読み出し／書き込みの可否を制御し、かつＶＭ５４０が管理するダーティデータをメモリ２１のアドレスとして管理するために実行される。

【０１３６】

Ｉ／Ｏ制御プログラム２２３０は、ホスト１’が、記憶装置システム２に強制的にダーティデータを転送する処理を行う際に実行される。構成管理Ａｇｅｎｔ２２６０は、記憶装置システム２がホスト１’に提供する論理記憶装置４０８とＶＭ５４０が構成する論理記憶領域との対応関係、及びＶＭ５４０が構成する論理記憶領域とＦＳが構成する論理記憶領域との関係を管理する際に実行される。

【０１３７】

チェックポイントＡｇｅｎｔ２２５０は、管理端末３からチェックポイントについて指示された際に、ホスト１’が、モード情報２２１０の設定、ＦＳ　Ａｇｅｎｔ２２２０、ＶＭ　Ａｇｅｎｔ２２４０、及びＩ／Ｏ制御プログラム２２３０等に所定の動作を指示する際に実行される。

【０１３８】

ホスト１’は、管理端末３からの指示により、エージェントプログラム２２００を実行して、ホスト１’のメモリ２１に存在するダーティデータを記憶装置システム２に送信する。一方、ホスト１’からのダーティデータの送信に合わせて、管理端末３は、チェックポイントコマンドを記憶装置システム２に送る。記憶装置システム２は、ホスト１’から送信されたダーティデータを処理する。記憶装置システム２は、また、管理端末３から送信されたＣＰ情報を、第一の実施形態で説明したホスト１から送信されたＣＰ情報と同様に扱って、自システム２内で管理する。このようにすることで、正論理記憶装置に論理的な障害が発生した際に、ホスト１にチェックポイント作成等の機能が無い場合でも、管理端末３からＣＰＩＤを指示することで記憶装置システム２側で管理しているチェックポイントまで高速にリストアを行い、システムの迅速な復旧を実現する。

【０１３９】

図１５は、管理端末３の詳細な構成を示した図である。尚、本構成は、他の実施形態で使用されてもよい。

【０１４０】

管理端末３は、プロセッサ１６０１、電気的に不揮発なメモリ１６０２、ネットワークＩ／Ｆ１６０５、入力部１６０４及び表示部１６０３とを有する。また、各々の構成部品は、データや制御命令等を伝送する伝送路１６１２で接続されている。

【０１４１】

プロセッサ１６０１は、管理端末３が有するプログラムを実行する。メモリ１６０２には、プロセッサ１６０１が実行するプログラムおよびそのプログラムが使用する情報等が格納される。例えば、表示部制御プログラム１６１０、入力部制御プログラム１６１１、記憶装置システム２の構成を管理する記憶装置制御情報１６０６、記憶装置制御情報１６０６に登録された情報を使用して記憶装置システム２を制御・管理するための記憶装置管理プログラム１６０７、記憶装置システム２に送信したＣＰ情報が含まれるシステム確定情報１６０８、及びシステム確定情報１６０８に登録された情報を用いて記憶装置システム２の状態を所定の時点に復旧する等の制御処理等を行うためのシステム状態管理プログラム１６０９等がメモリ１６０２に登録される。

【０１４２】

ネットワークＩ／Ｆ１６０５はネットワーク５に接続されている。管理端末３は、ネットワーク５を介して記憶装置システム２のシステム構成、例えばデバイス管理情報４１０、ペア管理情報４１２及びジャーナル管理情報４１８を取得する。又、管理端末３は、ネットワーク５を介して、構成定義処理（例えばＰａｒｉｔｙ　Ｇｒｏｕｐ４０７に論理記憶装置４０８を定義し、記憶装置システム２内部の論理記憶装置番号を割り振ることや論理記憶装置４０８をホスト１’に使用可能にするためにパスを定義してホスト１’が使用する論理記憶装置番号を割り振ること）をしたり、記憶装置システム２のリストア処理の実行を制御したりする。

【０１４３】

また、記憶装置システム２のユーザ又は管理者は、入力部１６０４及び表示部１６０３を使用して、記憶装置システム２の保守／管理やリストア処理の指示等を行う。

【０１４４】

図１６は、メモリ１６０２に格納されるシステム確定情報１６０８の一例を示す図である。管理端末３は、ホスト１’の状態が確定する時点を記憶装置システム２に指示する際に、管理端末３自身で記憶装置システム２に指示した内容をシステム確定情報１６０８としてメモリ１６０２に記録する。システム確定情報１６０８は、システムの状態が確定する時点のＣＰＩＤを登録するエントリ１７０１、論理記憶装置を示す番号が登録されるエントリ１７０２及びシステムの状態が確定する時点の時間を登録するエントリ１７０３を有する。

【０１４５】

図１７は、表示部１６０３における表示の一例を示す図である。本図では、表示部１６０３に、図１６に示したシステム確定情報１６０８の内容がＧＵＩを用いて表示されたものを例示している。このように、表示部１６０３は、システム状態が確定された時間を複数表示し、表示された複数の時間からユーザがある時間を選択したことを表示することができる。これにより、ユーザの利便性が向上する。

【０１４６】

具体的には、表示部１６０３は、管理情報を表示する領域１８０２を有する。その領域１８０２には、論理記憶装置番号を表示する領域１８０３及び領域１８０３に表示された論理記憶装置４０８の状態を確定した時間が表示される領域１８０４が含まれる。ユーザは、マウス等で操作可能なポインタ１８０５で、表示された論理記憶装置４０８について、チェックポイントコマンドによって状態が確定された時間を指定することができる。

【０１４７】

又、ユーザは、ある論理記憶装置４０８に障害が起きた場合、記憶装置システム２に対して、管理端末３のＧＵＩ１６０３を介してリストア処理の指示を行う。例えば、本図では、領域１８０３に表示された論理記憶装置４０８の内容を、領域１８０４で示された時刻中、２００２年５月５日１４：００の時点にリストアするための指示例を示している。ユーザは、ポインタ１８０５を用いて２００２年５月５日１４：００を示す領域１８０４を選択し、それを領域１８０３へＤｒａｇ＆Ｄｒｏｐ等を行うことで、論理記憶装置４０８のリストア時刻を指示する。

【０１４８】

管理端末３は、ユーザによって指定された論理記憶装置４０８及びリストア時間に基づいて、図１６に示されたシステム確定情報１６０８を検索し、リストアに使用するチェックポイントを特定する。その後、管理端末３は、記憶装置システム２に、検索の結果得られたＣＰ情報をリストアコマンドを用いて送信する。

【０１４９】

図１８は、ユーザが、管理端末３を介してジャーナルデータ開始準備処理を情報処理システムに指示する処理の流れを示した図である。

【０１５０】

まず、ユーザは、管理端末３の表示部１６０３及び入力部１６０４を用いて、ジャーナルデータを取得すべき対象となる正論理記憶装置や副論理記憶装置を指定する。管理端末３は、ユーザの指定に基づいて、ジャーナル取得準備指示コマンドを記憶装置システム２にネットワーク５を介し送信する。ジャーナル取得開始準備指示には、ユーザが指定した正論理記憶装置を示す識別子、当該正論理記憶装置と対になる副論理記憶装置を示す識別子、ジャーナル論理記憶装置を示す識別子が含まれる（ステップ１９０１）。

【０１５１】

ジャーナル準備処理指示を受領した記憶装置システム２は（ステップ１９６１）、ジャーナル準備処理を実行する。本処理は、図６のステップ６０２～ステップ６０６で説明された処理と同様の処理である（ステップ１９６２）。ジャーナル準備処理を終了した記憶装置システム２は、ネットワーク５を介して、管理端末３に終了報告を送信する（ステップ１９６３）。

【０１５２】

完了報告を受信した管理端末３は（ステップ１９０２）、ホスト１’にジャーナル開始モード指示のコマンドをネットワーク５を介して送信する（１９０３）。

【０１５３】

ジャーナル開始モード指示のコマンドを受領したホスト１’は、Ａｇｅｎｔ８００を実行することで、ジャーナルデータ取得の対象となる正論理記憶装置に対応するモード情報８１０をジャーナル開始モードに設定する。更に、ホスト１’は、ジャーナル開始モードに設定された正論理記憶装置に格納されるべきダーティデータを確定するために、ファイルの使用を終了する。尚、ジャーナル開始モード中は、ジャーナル開始モードを設定された正論理記憶装置に関連する記憶領域は書き込み禁止となる（ステップ１９２１）。

【０１５４】

次にホスト１’は、ＦＳが管理するメモリ２１に格納されたダーティデータをすべて記憶装置システム２に送信するため、記憶装置システム２にライトＩ／Ｏ処理要求を出す（ステップ１９２２）。

【０１５５】

ホスト１’からライトＩ／Ｏ処理要求を受け付けた記憶装置システム２は、ユーザが指定した正論理記憶装置への書き込み処理であれば、ジャーナル作成処理を行う。処理が終了すると、記憶装置システム２は、ホスト１’へ完了を報告する（ステップ１９６５）。

【０１５６】

完了の報告を受取ったホスト１’は、ＦＳが管理する全ダーティデータを記憶装置システム２に書き込んだかどうかを判断する（ステップ１９２３）。全ダーティデータの書き込みが完了していない場合、ホスト１’は、ステップ１９２２から処理を繰り返す。全ダーティデータの書き込みが終了した場合、ホスト１’は管理端末３に、完了報告をネットワーク５を介し送信する（ステップ１９２５）。

【０１５７】

ダーティデータのライト完了報告を受領した管理端末３は、記憶装置システム２に対しチェックポイントコマンドを発行するとともに、システム確定情報１６０８の更新を行う。具体的には、管理端末３は、ジャーナルデータを取得する論理記憶装置４０８を指定するデバイス番号に対応するエントリに、送信したＣＰＩＤと送信した時間を記録する（１９０５）。

【０１５８】
チェックポイントコマンドを受領した記憶装置システム２は（ステップ１９６６）、受領したチェックポイントコマンド中のＣＰＩＤをジャーナルデータとしてジャーナル論理記憶装置に記録する（ステップ１９６７）。記録が完了したら、記憶装置システム２は、完了報告を管理端末３にネットワーク５を介し送信する（ステップ１９６８）。
【０１５９】
完了報告を受領した管理端末３は（１９０６）、ホスト１’に対してジャーナル開始モード解除指示をネットワーク５を介し送信する（ステップ１９０７）。ジャーナル開始モード解除指示を受領したホスト１’は、ステップ１９２１で設定された、正論理記憶装置に対応するモード情報８１０のジャーナル開始モードを解除する。その後、ホスト１’は、正論理記憶装置に対応する記憶領域への書き込み禁止も解除する（ステップ１９２７）。
【０１６０】
その後、ユーザは、管理端末３を用いて、所定のタイミングでジャーナルモード開始指示をホスト１’及び記憶装置システム２に送信する。ジャーナルモード開始指示を受信したホスト１’は、指示で指定される正論理記憶装置に対応するモード情報８１０にジャーナルモードを設定する。一方、ジャーナルモード開始指示を受信した記憶装置システムは、先に指定されたジャーナル論理記憶装置にジャーナルデータの記録を開始する。
【０１６１】
図１９は、ジャーナルデータを取得している正論理記憶装置の内容を後にリストアできるように、ユーザの指示等に基づいて、ホスト１’の替わりに管理端末３がチェックポイントコマンドを記憶装置システム２に送信し、記憶装置システム２と管理端末３双方で一意のＣＰＩＤを格納する際の処理手順を示した図である。
【０１６２】
管理端末３は、ユーザの指示もしくは管理端末３自身のプログラムの実行に基づいて、チェックポイントモード指示をホスト１’にネットワーク５を介し送信する。チェックポイントモード指示には、チェックポイント取得の対象となる論理記憶装置４０８を示す番号が含まれている（ステップ２００１）。
【０１６３】
チェックポイントモード指示を受け取ったホスト１’は、Ａｇｅｎｔプログラム８００を実行して、指示に含まれる論理記憶装置の番号及びモード情報８１０に登録された情報とを参照し、指示された論理記憶装置４０８がジャーナルモードであることを確認する。指示された論理記憶装置４０８がジャーナルモードである場合、ホスト１’は、メモリ２１にあるダーティデータを記憶装置システム２へ強制的に転送する。
【０１６４】
具体的には、ホスト１’は、構成定義ＡｇｅｎｔＦＳ　Ａｇｅｎｔ２２２０を実行して、指定された論理記憶装置４０８を使用しているファイルが使用されているかを確認する。その後、ホスト１’は、ＦＳ　Ａｇｅｎｔ８２０を実行して、使用しているファイルを終了する又は使用しているファイルへの書き込み要求が実行されないようにする。その後、ホスト１’は、メモリ２１に格納されたダーティデータの転送を要求するライトＩ／Ｏ処理要求を記憶装置システム２に送信する。尚、ホスト１’がＶＭを使用している場合は、ホスト１’は、上述と同様の処理を、ＶＭ　Ａｇｅｎｔ２２４０を実行して行う（ステップ２０２２）。
【０１６５】
ライトＩ／Ｏ処理要求を受けた記憶装置システム２は、図７で説明したフローに従って、ジャーナルデータをジャーナル論理記憶装置に格納する処理を行う（ステップ２０６１、ステップ２０６２）。
【０１６６】
ジャーナル作成の完了を受信したホスト１’は、全てのダーティデータが記憶装置システム２に格納されたかどうかを確認する。全てのダーティデータが記憶装置システム２に

格納されていない場合、ホスト１’は、ステップ２０２２からの処理を繰り返す（ステップ２０２３）。

【０１６７】

全てのダーティデータが記憶装置システム２に格納されたと確認した場合、ホスト１’は、管理端末３に、チェックポイントモード指示に対する応答メッセージとしてダーティデータのライト完了報告を送信する（ステップ２０２５）。

【０１６８】

ライト完了報告を受領した管理端末３（ステップ２００２）は、記憶装置システム２に対し、チェックポイントコマンドを発行するとともに、システム確定情報１６０８の更新を行い、処理対象である論理記憶装置を示すデバイス番号に対応するエントリに、送信したＣＰＩＤと送信した時間を記録する（ステップ２００３）。

【０１６９】

チェックポイントコマンドを受領した記憶装置システム２は（ステップ２０６３）、受領したチェックポイントコマンドに含まれるＣＰＩＤをジャーナルデータとして記録する（ステップ２０６４）。その後、記憶装置システム２は、完了報告を管理端末３にネットワーク５を介し送信する（ステップ２０６５）。

【０１７０】

完了報告を受領した管理端末３は、ホスト１’に対して、チェックポイントモード解除指示をネットワーク５を介し送信する（ステップ２００４）。

【０１７１】

チェックポイントモード解除指示を受領したホスト１’は、ステップ２０２１でチェックポイントモードを設定された論理記憶装置４０８に対応するモード情報８１０に登録されたチェックポイントモードを解除する（ステップ２０２６）。その後、ホスト１’は、ファイルへの書き込みを再開するか、ファイルを使用可能状態とする（ステップ２０２７）。

【０１７２】

図２１は、ユーザが、管理端末３を介して、記憶装置システム２へリストア指示を出す際の処理手順を示す図である。本実施形態では、スナップショットペア４５０が既に形成され、副論理記憶装置に正論理記憶装置のスナップショットが取得されていて、ジャーナル論理記憶装置には、副論理記憶装置にスナップショットをとる時点より後もしくは前後のジャーナルデータが格納されているものとする。

【０１７３】

この場合において、正論理記憶装置を使用していたホスト１’のＡＰが、使用しているファイルに誤った編集をした等の理由で、バックアップデータに基づくリストアが必要になった場合を考える。

【０１７４】

まず、ユーザは、図１７で説明したように、管理端末３の入力部１６０４及び表示部１６０３を用いて、リストアの対象となる論理記憶装置４０８及びどの時点までリストアを行うかを指示する（ステップ２１０１）。

【０１７５】

ユーザの指示を受けた管理端末３は、ユーザが画面上で指定した入力情報が、システム確定情報１６０８のどのエントリに登録された情報と一致するかを判断し、一致したエントリに登録されているＣＰＩＤを決定する（ステップ２１０２）。その後、管理端末３は、リストアを行う論理記憶装置４０８を示す識別子（番号）及びＣＰＩＤを含むリストアコマンドを、記憶装置システム２に送信する（ステップ２１０３）。

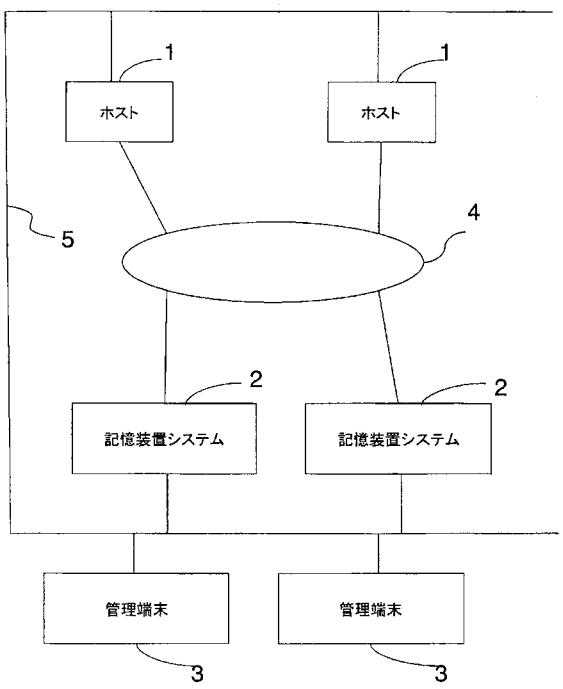
【０１７６】

管理端末３からリストアコマンドを受信した記憶装置システム２は、指定された論理記憶装置４０８について、図１１で説明したリストア処理を実行する。その後、リストア処理完了報告を管理端末３に送信する（ステップ２１０４）。完了報告を受領した管理端末３は、記憶装置制御情報１６０６を更新する（ステップ２１０５）。

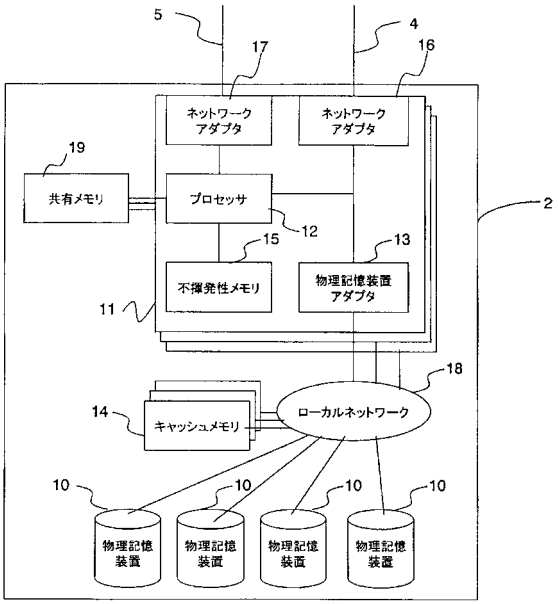
【０１７７】
本実施形態によれば、第一の実施形態と比較し、ホスト１’がＣＰ情報を管理することが無いので、その分ホスト１’の負荷を低減することが出来る。また、ホスト１等がチェックポイント作成機能を有さない場合でも、ＣＰ情報を用いたリストア処理を行うことが出来る。
【０１７８】
尚、本実施形態では、管理端末３から記憶装置システム２に対してチェックポイントコマンドを発行する前に、管理端末３は、ホスト１’に格納されているダーティデータを記憶装置システム２に反映させるために、ホスト１’に対してダーティデータをフラッシュさせる指示（ジャーナルモード開始指示、チェックポイントモード指示）を送信した。しかし、この場合、上述したように、ホスト１’にエージェントというプログラムを用意しなければならない。したがって、全てのホスト１’にエージェントを用意するのが困難な場合、本実施形態は採用しづらい。そこで、ホスト１’に存在するダーティデータを無視し、管理端末３と記憶装置システム２のみでジャーナルモードの設定、ＣＰ情報の遣り取り及びリストア処理の実行を行う第三の実施形態を考える。
【０１７９】
本実施形態は、第二の実施形態と以下の点で異なる。すなわち、図１８において、ジャーナル作成準備処理の完了を報告された管理端末３（ステップ１９０２）は、ステップ１９０３の処理を行わずに、直接記憶装置システム２に対してチェックポイントコマンドを送付するステップ１９０４の処理を行う。記憶装置システム２では、受信したチェックポイントコマンドに従って、ステップ１９６６以降の処理を行う。
【０１８０】
また、図１９においては、管理端末３は、ステップ２００１のチェックポイントモード指定をホスト１’に送信せず、直接チェックポイントコマンドを記憶装置システム２に送信する（ステップ２００３）。チェックポイントコマンドを受信した記憶装置システム２は、ステップ２０６３以降の処理を行う。
【０１８１】
尚、本実施形態で使用されるホストは、ホスト１のようにＤＢのログを有する計算機でも、ホスト１’のようにエージェントを有する計算機でも、あるいは、何ら特別なプログラムを有さない通常の計算機でも良い。他の構成及び処理、例えばリストア処理等は、第二の実施形態と同様である。
【０１８２】
本実施形態によれば、ホストの種別に関わらず、管理端末３及び記憶装置システム２との遣り取りだけで、記憶装置システム２の記憶装置を、任意のシステム状態までにリストアすることができる。
【図面の簡単な説明】
【０１８３】
【図１】本発明を適用した情報処理システムの構成を示す図である。
【図２】記憶装置システム２の構成を示す図である。
【図３】ホスト１の構成を示す図である。
【図４】記憶装置システム２が有するプログラム等の構成を示す図である。
【図５】ホスト１が有するプログラム等の構成を示す図である。
【図６】ジャーナル取得準備の処理手順を示す図である。
【図７】ジャーナルモード中のＩ／Ｏ処理要求の手順を示す図である。
【図８】ジャーナルデータ作成処理の手順を示す図である。
【図９】ジャーナルデータの形式を示す図である。
【図１０】ＣＰ情報の送信処理の手順を示す図である。
【図１１】リストア処理の手順を示す図である。
【図１２】デバイス管理情報の構成例を示す図である。
【図１３】ペア管理情報の構成例を示す図である。

【図１４】ジャーナル管理情報の構成例を示す図である。
【図１５】管理端末３の構成を示す図である。
【図１６】システム確定情報１６０８の構成例を示す図である。
【図１７】管理端末３の表示部の構成例を示す図である。
【図１８】第二の実施形態におけるジャーナルデータ取得指示の処理手順を示す図である。
【図１９】第二の実施形態におけるＣＰＩＤの送受信の処理手順を示す図である。
【図２０】第二の実施形態におけるリストア指示処理の手順を示す図である。
【図２１】第二の実施形態におけるホスト１’の論理的構成を示す図である。
【符号の説明】
【０１８４】
１　　ホスト
２　　記憶装置システム
３　　管理端末
４　　ネットワーク
５　　ネットワーク
１０　　物理記憶装置
１１　　記憶装置制御装置。

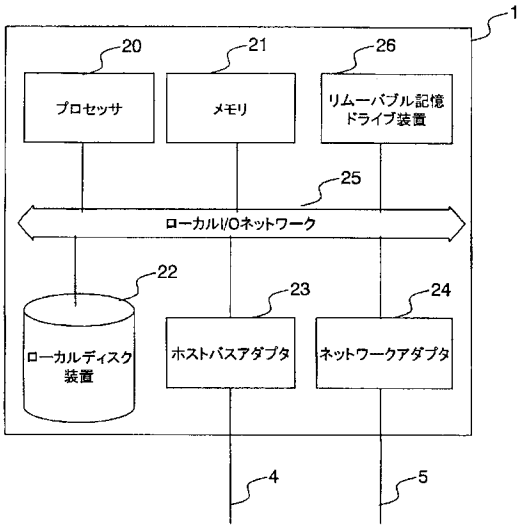
【図１】



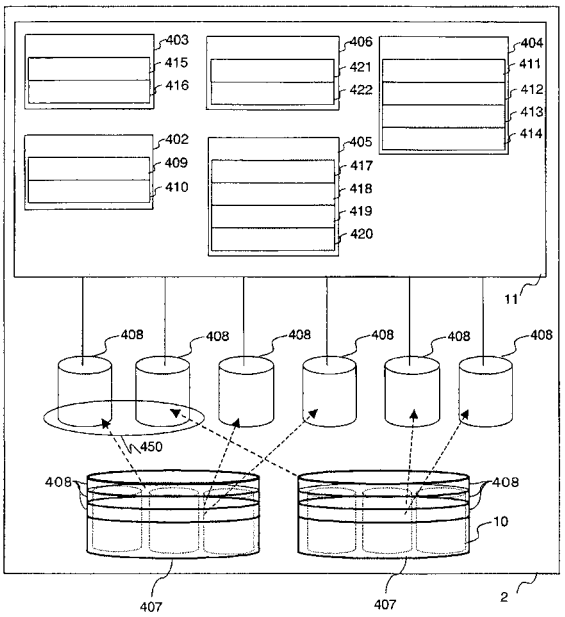
【図２】



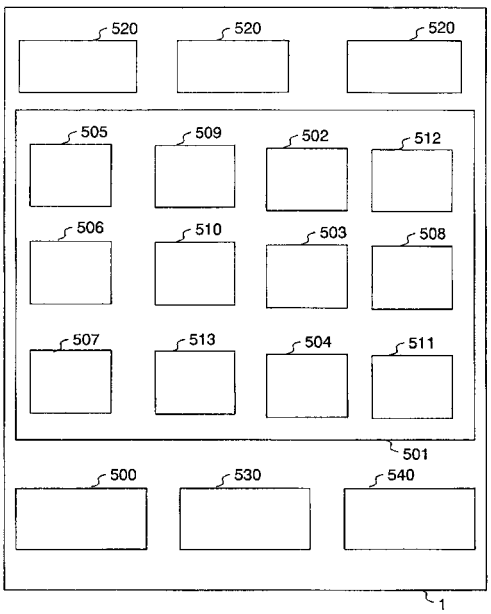
【図３】


1
20
21
26
プロセッサ
メモリ
リムーバブル記憶ドライブ装置
25
ローカルI/Oネットワーク
22
23
24
ローカルディスク装置
ホストバスアダプタ
ネットワークアダプタ
4
5

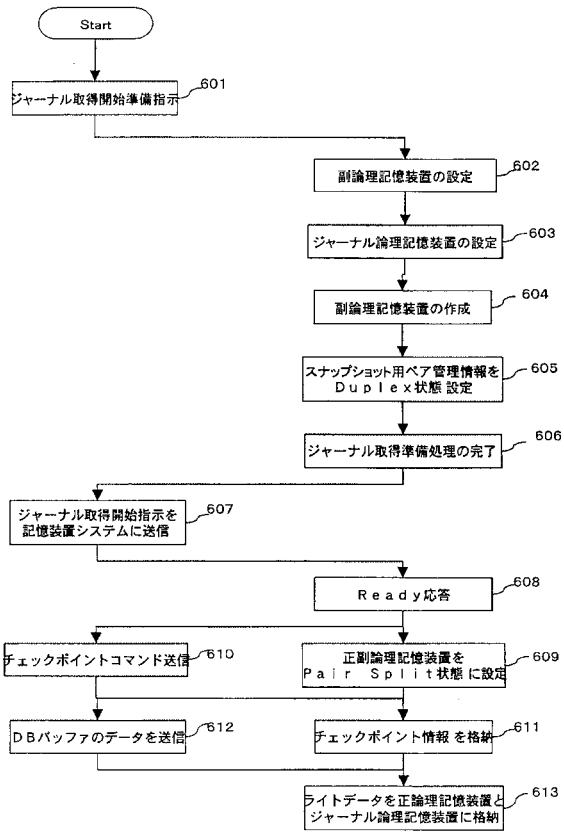
【図４】


403
415
416
406
421
422
404
411
412
413
414
402
409
410
405
417
418
419
420
11
408
450
10
407
2

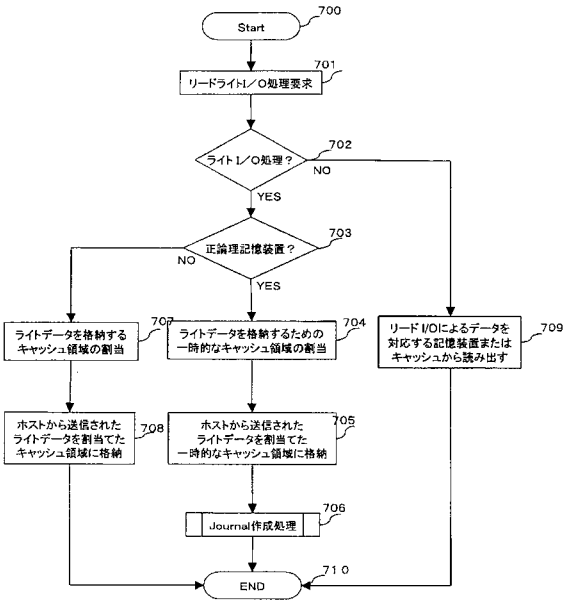
【図５】


520
505
509
502
512
506
510
503
508
507
513
504
511
501
500
530
540
1

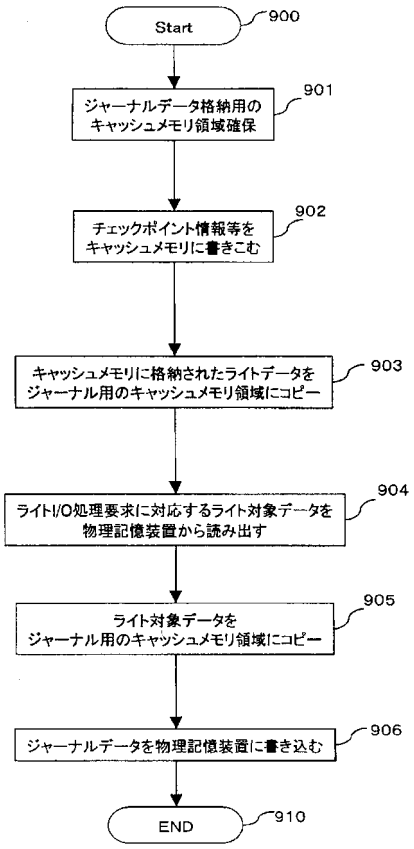
【図６】


Start
ジャーナル取得開始準備指示
601
副論理記憶装置の設定
602
ジャーナル論理記憶装置の設定
603
副論理記憶装置の作成
604
スナップショット用ペア管理情報をDuplex状態 設定
605
ジャーナル取得準備処理の完了
606
ジャーナル取得開始指示を記憶装置システムに送信
607
Ready応答
608
チェックポイントコマンド送信
610
正副論理記憶装置をPair Split状態 に設定
609
DBバッファのデータを送信
612
チェックポイント情報 を格納
611
ライトデータを正論理記憶装置とジャーナル論理記憶装置に格納
613

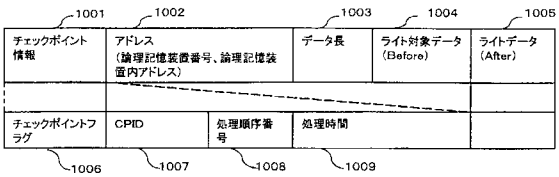
【図７】


Start
700
リードライトI／O処理要求
701
ライトI／O処理？
702
NO
YES
正論理記憶装置？
703
NO
YES
ライトデータを格納する
キャッシュ領域の割当
707
ライトデータを格納するための
一時的なキャッシュ領域の割当
704
リードI/Oによるデータを
対応する記憶装置または
キャッシュから読み出す
709
ホストから送信された
ライトデータを割当てた
キャッシュ領域に格納
708
ホストから送信された
ライトデータを割当てた
一時的なキャッシュ領域に格納
705
Journal作成処理
706
END
710

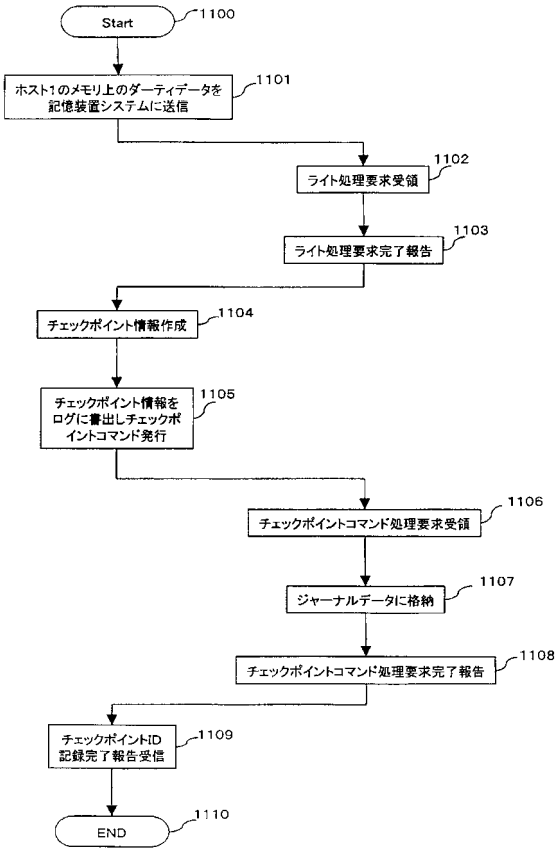
【図８】


Start
900
ジャーナルデータ格納用の
キャッシュメモリ領域確保
901
チェックポイント情報等を
キャッシュメモリに書きこむ
902
キャッシュメモリに格納されたライトデータを
ジャーナル用のキャッシュメモリ領域にコピー
903
ライトI/O処理要求に対応するライト対象データを
物理記憶装置から読み出す
904
ライト対象データを
ジャーナル用のキャッシュメモリ領域にコピー
905
ジャーナルデータを物理記憶装置に書き込む
906
END
910

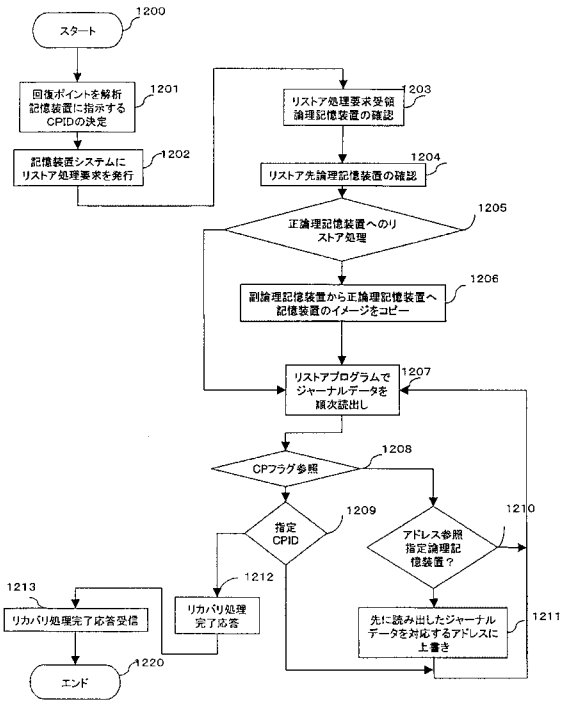
【図９】


1001
1002
1003
1004
1005
チェックポイント
情報
アドレス
（論理記憶装置番号、論理記憶装
置内アドレス）
データ長
ライト対象データ
（Before）
ライトデータ
（After）
チェックポイントフ
ラグ
CPID
処理順序番
号
処理時間
1006
1007
1008
1009

【図１０】


Start
1100
ホスト1のメモリ上のダーティデータを
記憶装置システムに送信
1101
ライト処理要求受領
1102
ライト処理要求完了報告
1103
チェックポイント情報作成
1104
チェックポイント情報を
ログに書出しチェックポ
イントコマンド発行
1105
チェックポイントコマンド処理要求受領
1106
ジャーナルデータに格納
1107
チェックポイントコマンド処理要求完了報告
1108
チェックポイントID
記録完了報告受信
1109
END
1110

【図１１】



【図１２】

1301

1311 ホスト提供論理記憶装置番号	1312 ホスト提供論理記憶装置内アドレス	1313 記憶装置システム内論理記憶装置番号	1314 論理記憶装置内アドレス	1304 / 1321 RAID Group番号	1322 物理記憶装置番号	1323 物理記憶装置内アドレス
Port#, LU#	0～512	Cu#, LDEV#	0～512	00	00	0～512
…	…	…	…	…	…	…

1300

1331 ホスト提供論理記憶装置番号	1332 記憶装置システム内論理記憶装置番号	1333 RAID Group番号	1334 Pair情報	1335 ジャーナル対象モード
01234567, 01	00, 01	00, 01	Pair Duplex	モードOFF
01234567, 02	00, 02	01, 01	No Pair	モードOFF
01234567, 03	00, 03	02, 01	Pair Sympl ex	モードON
…	…	…	…	…

1330

1350	1351 記憶装置システム内論理記憶装置番号	1352 空きノリザーブ情報	1353 Path定義情報	1354 Emulation Type、サイズ	1355 障害情報
	00, 01	無し	パス定義済 接続Port#	OPEN :72G	無し
	00, 02	無し	パス定義済 接続Port#	OPEN :72G	無し
	…	…	…	…	…
	0A, 01	スナップショット用リザーブ	無し	OPEN 36G	無し
	0A, 02	空き	無し	OPEN 36G	無し
	…	…	…		…

【図１３】

1401 ホスト提供論理記憶装置番号	1402 記憶装置システム内論理記憶装置番号	1403 Emulation Type サイズ	1404 ペア状態	1405 ペア管理情報 1411 正	1412 副
11234567, 01	01, 01	OPEN 9G	Pair Duplex	0	02, 01
11234567, 02	01, 02	OPEN 9G	Pair Sympl ex	02, 02	0
11234567, 03	01, 03	OPEN 16G	No Pair	−1	−1
11234567, 04	01, 04	OPEN 16G	Pair Resync	02, 04	0
…	…	…	…	…	…
11234568, 0A	0D, 0A	3390 3G	Pair Create	0A, 0A	0F, 0A
…	…	…	…	…	…

【図１４】

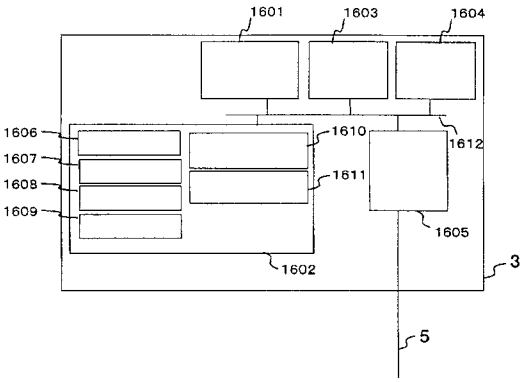
1501 チェックポイントID	1502 アドレス	1503 時間
ABCDEF12	12	2002.06.02 09:00:00
ABCDEF19	1A2B3C56	2002.06.02 21:00:00
…	…	…
0345FCA9	3ADE68B0	2002.06.12 09:00:00

1500

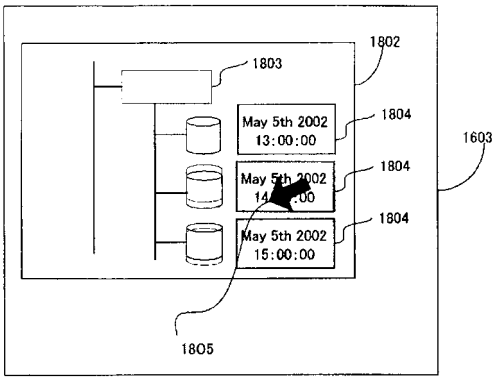
1521 ジャーナルデータ格納論理記憶装置番号	1522 チェックポイントID	1523 アドレス
01, 0F	ABCDEF12	0
01, 0F	ABCDEF19	FEDCBA12
…	…	…

1520

【図１５】



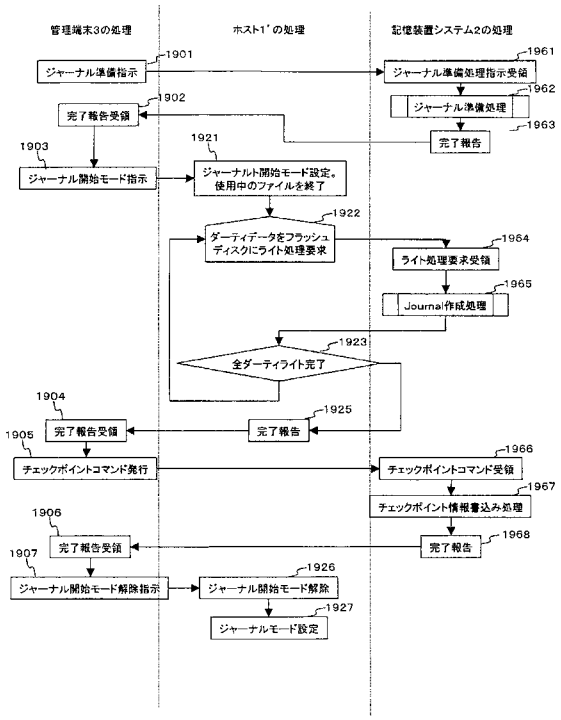
【図１７】



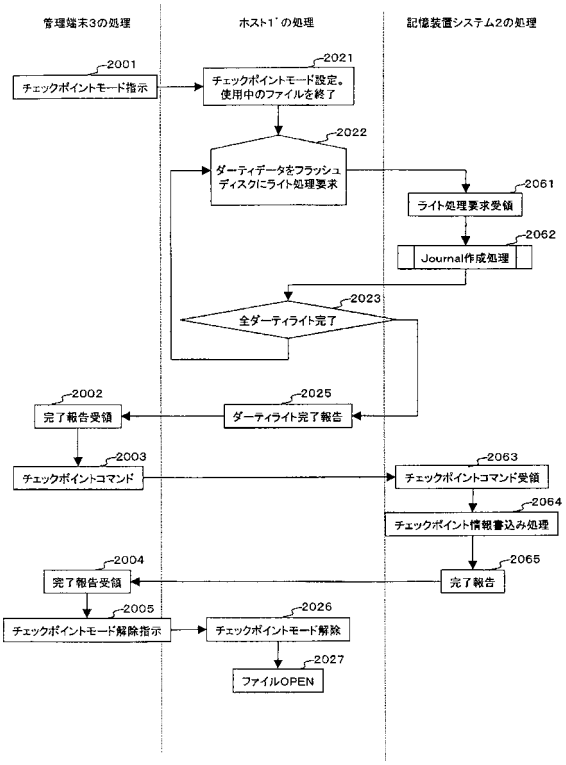
【図１６】

	1702	1701 / 1703
デバイス番号:010F	チェックポイント識別子	時間
	I	2002.05.05 14:00:00
	J	2002.05.05 15:00:00
	……	……
デバイス番号:0B03	チェックポイント識別子	時間
	I	2002.05.05 14:00:00
	J	2002.05.05 15:00:00
	……	……

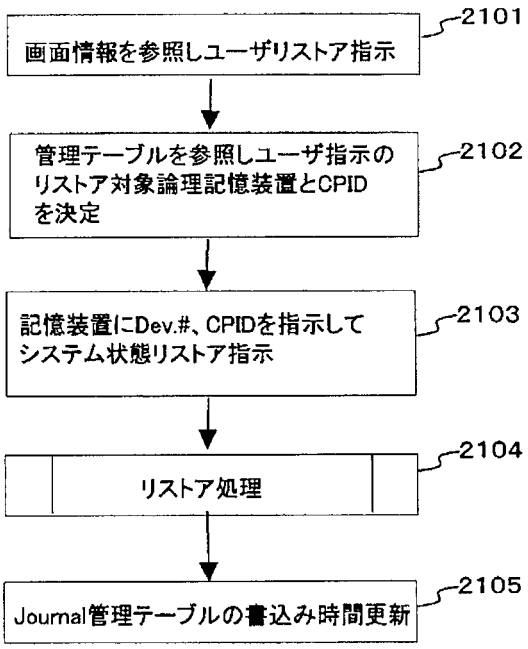
【図１８】



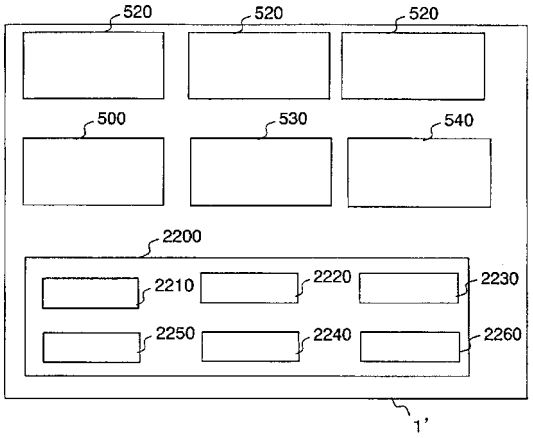
【図１９】



【図２０】


画面情報を参照しユーザリストア指示
2101
管理テーブルを参照しユーザ指示のリストア対象論理記憶装置とCPIDを決定
2102
記憶装置にDev.#、CPIDを指示してシステム状態リストア指示
2103
リストア処理
2104
Journal管理テーブルの書込み時間更新
2105

【図２１】


520
520
520
500
530
540
2200
2210
2220
2230
2250
2240
2260
1'

フロントページの続き

(51)Int.Cl.　　FI
　　G06F 13/10　340A

(72)発明者　山本　康友
神奈川県川崎市麻生区王禅寺１０９９番地　株式会社　日立製作所　システム開発研究所内
(72)発明者　大枝　高
神奈川県川崎市麻生区王禅寺１０９９番地　株式会社　日立製作所　システム開発研究所内
(72)発明者　荒井　弘治
神奈川県小田原市中里３３２番地２号　株式会社　日立製作所　ＲＡＩＤシステム事業部内

審査官　高瀬　勤

(56)参考文献　特開昭５８－１０１３５６（ＪＰ，Ａ）
特開平０２－１７６９４９（ＪＰ，Ａ）
特開２０００－３３０７２９（ＪＰ，Ａ）
特開２００１－２１６１８５（ＪＰ，Ａ）

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