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

特許第4124348号
(P4124348)

(45)発行日　平成20年7月23日(2008.7.23)　(24)登録日　平成20年5月16日(2008.5.16)

(51)Int.Cl.		F I		
G06F 12/00	(2006.01)	G06F	12/00	531J
G06F 3/06	(2006.01)	G06F	12/00	531D
G06F 12/08	(2006.01)	G06F	3/06	304F
		G06F	3/06	304P
		G06F	12/08	551Z

請求項の数 21　(全 33 頁)　最終頁に続く

(21)出願番号　特願2003-183734(P2003-183734)
(22)出願日　平成15年6月27日(2003.6.27)
(65)公開番号　特開2005-18506(P2005-18506A)
(43)公開日　平成17年1月20日(2005.1.20)
審査請求日　平成18年6月12日(2006.6.12)

早期審査対象出願

(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　110000279
特許業務法人ウィルフォート国際特許事務所
(72)発明者　平川　裕介
神奈川県小田原市中里３２２番地２号　株式会社日立製作所　ＲＡＩＤシステム事業部内
(72)発明者　荒川　敬史
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】記憶システム

(57)【特許請求の範囲】

【請求項１】

ホストコンピュータに接続される正ストレージシステムに接続可能であり、複数のディスクドライブを備える副ストレージシステムであって、

少なくとも一部の前記複数のディスクドライブに対応付けられており、前記ホストコンピュータから前記正ストレージシステムの複数の第１の正論理ボリュームへ送信される更新データと対応するデータを含み且つ前記正ストレージシステムから受信されるジャーナルデータを、複数記憶する少なくとも一つの第１のジャーナル論理ボリュームと、

少なくとも一部の前記複数のディスクドライブに対応付けられており、前記少なくとも一つの第１のジャーナル論理ボリュームに記憶される複数のジャーナルデータから取得される複数のデータを記憶する複数の第１の副論理ボリュームと、

少なくとも一部の前記複数のディスクドライブに対応付けられており、前記ホストコンピュータから前記正ストレージシステムの複数の第２の正論理ボリュームへ送信される更新データと対応するデータを含み且つ前記正ストレージシステムから受信されるジャーナルデータを、複数記憶する少なくとも一つの第２のジャーナル論理ボリュームと、

少なくとも一部の前記複数のディスクドライブに対応付けられており、前記少なくとも一つの第２のジャーナル論理ボリュームに記憶される複数のジャーナルデータから取得される複数のデータを記憶する複数の第２の副論理ボリュームと、

前記複数の第１の副論理ボリューム及び前記少なくとも一つの第１のジャーナル論理ボリュームが割当てられ、前記複数の第１の副論理ボリューム間でデータ整合性が維持され

る第１のグループと、前記複数の第２の副論理ボリューム及び前記少なくとも一つの第２のジャーナル論理ボリュームが割当てられ、前記複数の第２の副論理ボリューム間でデータ整合性が維持される第２のグループと、を管理する制御部と、
を有することを特徴とする副ストレージシステム。
【請求項２】
請求項１に記載の副ストレージシステムであって、
前記第１のジャーナル論理ボリュームに記憶される複数のジャーナルデータの各々は、前記第１のグループにおける更新順序に関する情報を含み、
前記第２のジャーナル論理ボリュームに記憶される複数のジャーナルデータの各々は、前記第２のグループにおける更新順序に関する情報を含む
ことを特徴とする副ストレージシステム。
【請求項３】
請求項１又は２に記載の副ストレージシステムであって、
前記少なくとも一つの第１のジャーナル論理ボリュームおよび前記少なくとも一つの第２のジャーナル論理ボリュームに記憶される複数のジャーナルデータの各々は、時刻情報を含む
ことを特徴とする副ストレージシステム。
【請求項４】
請求項１乃至３のいずれかに記載の副ストレージシステムであって、
前記第１のジャーナル論理ボリュームに記憶される複数のジャーナルデータは、前記第１のグループの識別子を含み、
前記第２のジャーナル論理ボリュームに記憶される複数のジャーナルデータは、前記第２のグループの識別子を含む
ことを特徴とする副ストレージシステム。
【請求項５】
請求項１乃至４のいずれかに記載の副ストレージシステムであって、
前記少なくとも一つの第１のジャーナル論理ボリュームに記憶される複数のジャーナルデータのうちの最古のジャーナルデータから取得されるデータが、前記複数の第１の副論理ボリュームのうちの少なくとも一つに書込まれ、
前記少なくとも一つの第２のジャーナル論理ボリュームに記憶される複数のジャーナルデータのうち最古のジャーナルデータから取得されるデータが、前記複数の第２の副論理ボリュームのうちの少なくとも一つに書込まれる
ことを特徴とする副ストレージシステム。
【請求項６】
請求項５に記載の副ストレージシステムであって、
前記少なくとも一つの第１のジャーナル論理ボリュームの一部は、前記少なくとも一つの第１のジャーナル論理ボリュームに記憶される最古のジャーナルデータから取得されるデータが前記複数の第１の副論理ボリュームのうちの少なくとも一つに書込まれた後に解放され、
前記少なくとも一つの第２のジャーナル論理ボリュームの一部は、前記少なくとも一つの第２のジャーナル論理ボリュームに記憶される最古のジャーナルデータから取得されるデータが前記複数の第２の副論理ボリュームのうちの少なくとも一つに書込まれた後に解放される
ことを特徴とする副ストレージシステム。
【請求項７】
請求項１乃至６のいずれかに記載の副ストレージシステムであって、
（１）前記正ストレージシステムから受信され前記制御部により前記少なくとも一つの第１のジャーナル論理ボリューム又は前記少なくとも一つの第２のジャーナル論理ボリュームに書き込まれるジャーナルデータ、及び、
（２）前記少なくとも一つの第１のジャーナル論理ボリューム又は前記少なくとも一つの

第２のジャーナル論理ボリュームに記憶されているジャーナルデータ内のデータを前記複数の第１の副論理ボリュームのうちの少なくとも一つ又は前記複数の第２の副論理ボリュームのうちの少なくとも一つに書き込むために、前記制御部により、前記少なくとも一つの第１のジャーナル論理ボリューム又は前記少なくとも一つの第２のジャーナル論理ボリュームから読み出されたジャーナルデータ、
を一時的に記憶するキャッシュメモリを更に備える、
ことを特徴とする副ストレージシステム。
【請求項８】
ホストコンピュータに接続される正ストレージシステムに接続可能であり、複数のディスクドライブを備える副ストレージシステムの制御部であって、
前記ホストコンピュータから前記正ストレージシステムの複数の第１の正論理ボリュームへ送信される更新データと対応するデータを含み且つ前記正ストレージシステムから受信されるジャーナルデータの複数を、少なくとも一部の前記複数のディスクドライブに対応付けられる少なくとも一つの第１のジャーナル論理ボリュームに記憶させるように制御するものであり、
前記少なくとも一つの第１のジャーナル論理ボリュームに記憶される複数のジャーナルデータから取得される複数のデータを、少なくとも一部の前記複数のディスクドライブに対応付けられる複数の第１の副論理ボリュームに記憶させるように制御するものであり、
前記ホストコンピュータから前記正ストレージシステムの複数の第２の正論理ボリュームへ送信される更新データと対応するデータを含み且つ前記正ストレージシステムから受信されるジャーナルデータの複数を、少なくとも一部の前記複数のディスクドライブに対応付けられる少なくとも一つの第２のジャーナル論理ボリュームに記憶させるように制御するものであり、
前記少なくとも一つの第２のジャーナル論理ボリュームに記憶される複数のジャーナルデータから取得される複数のデータを、少なくとも一部の前記複数のディスクドライブに対応付けられる複数の第２の副論理ボリュームに記憶させるように制御するものであり、
前記複数の第１の副論理ボリューム及び前記少なくとも一つの第１のジャーナル論理ボリュームが割当てられ、前記複数の第１の副論理ボリューム間でデータ整合性が維持される第１のグループと、前記複数の第２の副論理ボリューム及び前記少なくとも一つの第２のジャーナル論理ボリュームが割当てられ、前記複数の第２の副論理ボリューム間でデータ整合性が維持される第２のグループと、を管理するものである
ことを特徴とする副ストレージシステムの制御部。
【請求項９】
請求項８に記載の副ストレージシステムの制御部であって、
前記少なくとも一つの第１のジャーナル論理ボリュームに記憶される複数のジャーナルデータは、前記第１のグループの識別子を含み、
前記少なくとも一つの第２のジャーナル論理ボリュームに記憶される複数のジャーナルデータは、前記第２のグループの識別子を含む
ことを特徴とする副ストレージシステムの制御部。
【請求項１０】
請求項８又は９に記載の副ストレージシステムの制御部であって、
前記少なくとも一つの第１のジャーナル論理ボリュームに記憶される複数のジャーナルデータの各々は、前記第１のグループにおける更新順序に関する情報を含み、
前記少なくとも一つの第２のジャーナル論理ボリュームに記憶される複数のジャーナルデータの各々は、前記第２のグループにおける更新順序に関する情報を含む
ことを特徴とする副ストレージシステムの制御部。
【請求項１１】
請求項８乃至１０のいずれかに記載の副ストレージシステムの制御部であって、
前記少なくとも一つの第１のジャーナル論理ボリュームおよび前記少なくとも一つの第２のジャーナル論理ボリュームに記憶される複数のジャーナルデータの各々は、時刻情報

を含む
ことを特徴とする副ストレージシステムの制御部。
【請求項１２】
請求項８乃至１１のいずれかに記載の副ストレージシステムの制御部であって、
前記少なくとも一つの第１のジャーナル論理ボリュームに記憶される複数のジャーナルデータのうちの最古のジャーナルデータから取得されるデータを、前記複数の第１の副論理ボリュームのうちの少なくとも一つに記憶させるように制御するものであり、
前記少なくとも一つの第２のジャーナル論理ボリュームに記憶される複数のジャーナルデータのうち最古のジャーナルデータから取得されるデータを、第２の副論理ボリュームのうちの少なくとも一つに記憶させるように制御するものである
ことを特徴とする副ストレージシステムの制御部。
【請求項１３】
請求項１２に記載の副ストレージシステムの制御部であって、
前記少なくとも一つの第１のジャーナル論理ボリュームに記憶される複数のジャーナルデータのうちの最古のジャーナルデータから取得されるデータを前記複数の第１の副論理ボリュームのうちの少なくとも一つに記憶させた後に、前記少なくとも一つの第１のジャーナル論理ボリュームのうちの前記最古のジャーナルデータが記憶されていた記憶領域を解放するように制御するものであり、
前記少なくとも一つの第２のジャーナル論理ボリュームに記憶される複数のジャーナルデータのうちの最古のジャーナルデータから取得されるデータを前記複数の第２の副論理ボリュームのうちの少なくとも一つに記憶させた後に、前記少なくとも一つの第２のジャーナル論理ボリュームのうちの前記最古のジャーナルデータが記憶されていた記憶領域を解放するように制御するものである、
ことを特徴とする副ストレージシステムの制御部。
【請求項１４】
請求項８乃至１３のいずれかに記載の副ストレージシステムの制御部であって、
（１）前記正ストレージシステムから受信され前記少なくとも一つの第１のジャーナル論理ボリューム又は前記少なくとも一つの第２のジャーナル論理ボリュームに書き込むジャーナルデータ、及び、
（２）前記少なくとも一つの第１のジャーナル論理ボリューム又は前記少なくとも一つの第２のジャーナル論理ボリュームに記憶されているジャーナルデータ内のデータを前記複数の第１の副論理ボリュームのうちの少なくとも一つ又は前記複数の第２の副論理ボリュームのうちの少なくとも一つに書き込むために、前記少なくとも一つの第１のジャーナル論理ボリューム又は前記少なくとも一つの第２のジャーナル論理ボリュームから読み出したジャーナルデータ、
を一時的に記憶するキャッシュメモリを更に備える、
ことを特徴とする副ストレージシステムの制御部。
【請求項１５】
ホストコンピュータに接続される正ストレージシステムに接続可能であり、複数のディスクドライブを備える副ストレージシステムの制御部のプログラムであって、
前記副ストレージシステムの制御部に、
前記ホストコンピュータから前記正ストレージシステムの複数の第１の正論理ボリュームへ送信される更新データと対応するデータを含み且つ前記正ストレージシステムから受信されるジャーナルデータの複数を、少なくとも一部の前記複数のディスクドライブに対応付けられる少なくとも一つの第１のジャーナル論理ボリュームに記憶させるように制御する処理と、
前記少なくとも一つの第１のジャーナル論理ボリュームに記憶される複数のジャーナルデータから取得される複数のデータを、少なくとも一部の前記複数のディスクドライブに対応付けられる複数の第１の副論理ボリュームに記憶させるように制御する処理と、
前記ホストコンピュータから前記正ストレージシステムの複数の第２の正論理ボリュー

ムへ送信される更新データと対応するデータを含み且つ前記正ストレージシステムから受信されるジャーナルデータの複数を、少なくとも一部の前記複数のディスクドライブに対応付けられる少なくとも一つの第２のジャーナル論理ボリュームに記憶させるように制御する処理と、

前記少なくとも一つの第２のジャーナル論理ボリュームに記憶される複数のジャーナルデータから取得される複数のデータを、少なくとも一部の前記複数のディスクドライブに対応付けられる複数の第２の副論理ボリュームに記憶させるように制御する処理と、

前記複数の第１の副論理ボリューム及び前記少なくとも一つの第１のジャーナル論理ボリュームが割当てられ、前記複数の第１の副論理ボリューム間でデータ整合性が維持される第１のグループと、前記複数の第２の副論理ボリューム及び前記少なくとも一つの第２のジャーナル論理ボリュームが割当てられ、前記複数の第２の副論理ボリューム間でデータ整合性が維持される第２のグループとを管理する処理と、

を実行させることを特徴とする副ストレージシステムの制御部のプログラム。

【請求項１６】

請求項１５に記載の副ストレージシステムの制御部のプログラムであって、

前記少なくとも一つの第１のジャーナル論理ボリュームに記憶される複数のジャーナルデータの各々は、前記第１のグループの識別子を含み、

前記少なくとも一つの第２のジャーナル論理ボリュームに記憶される複数のジャーナルデータの各々は、前記第２のグループの識別子を含む

ことを特徴とする副ストレージシステムの制御部のプログラム。

【請求項１７】

請求項１５又は１６に記載の副ストレージシステムの制御部のプログラムであって、

前記少なくとも一つの第１のジャーナル論理ボリュームに記憶される複数のジャーナルデータの各々は、前記第１のグループにおける更新順序に関する情報を含み、

前記少なくとも一つの第２のジャーナル論理ボリュームに記憶される複数のジャーナルデータの各々は、前記第２のグループにおける更新順序に関する情報を含む

ことを特徴とする副ストレージシステムの制御部のプログラム。

【請求項１８】

請求項１５乃至１７のいずれかに記載の副ストレージシステムの制御部のプログラムであって、

前記少なくとも一つの第１のジャーナル論理ボリュームおよび前記少なくとも一つの第２のジャーナル論理ボリュームに記憶される複数のジャーナルデータの各々は、時刻情報を含む

ことを特徴とする副ストレージシステムの制御部のプログラム。

【請求項１９】

請求項１５乃至１８のいずれかに記載の副ストレージシステムの制御部のプログラムであって、

前記副ストレージシステムの制御部に、

前記少なくとも一つの第１のジャーナル論理ボリュームに記憶される複数のジャーナルデータのうちの最古のジャーナルデータから取得されるデータを前記複数の第１の副論理ボリュームのうちの少なくとも一つに記憶させるように制御する処理と、

前記少なくとも一つの第２のジャーナル論理ボリュームに記憶される複数のジャーナルデータのうち最古のジャーナルデータから取得されるデータを前記複数の第２の副論理ボリュームのうちの少なくとも一つに記憶させるように制御する処理とを、

実行させることを特徴とする副ストレージシステムの制御部のプログラム。

【請求項２０】

請求項１９に記載の副ストレージシステムの制御部のプログラムであって、

前記副ストレージシステムの制御部に、

前記少なくとも一つの第１のジャーナル論理ボリュームに記憶される複数のジャーナルデータのうちの最古のジャーナルデータから取得されるデータを前記複数の第１の副論理

ボリュームのうちの少なくとも一つに記憶させた後に、前記少なくとも一つの第１のジャーナル論理ボリュームに記憶される複数のジャーナルデータのうちの最古のジャーナルデータが記憶されていた記憶領域を解放する処理と、

前記少なくとも一つの第２のジャーナル論理ボリュームに記憶される複数のジャーナルデータのうちの前記最古のジャーナルデータから取得されるデータを前記複数の第２の副論理ボリュームのうちの少なくとも一に記憶させた後に、前記少なくとも一つの第２のジャーナル論理ボリュームに記憶される複数のジャーナルデータのうちの最古のジャーナルデータが記憶されていた記憶領域を解放する処理とを、

実行させることを特徴とする副ストレージシステムの制御部のプログラム。

【請求項２１】

請求項１５乃至２０のいずれかに記載の副ストレージシステムの制御部のプログラムであって、

前記制御部が、キャッシュメモリを備えており、

前記制御部に、以下の（１）及び（２）のジャーナルデータをキャッシュメモリに一時的に記憶させる処理、

（１）前記正ストレージシステムから受信され前記少なくとも一つの第１のジャーナル論理ボリューム又は前記少なくとも一つの第２のジャーナル論理ボリュームに書き込むジャーナルデータ、

（２）前記少なくとも一つの第１のジャーナル論理ボリューム又は前記少なくとも一つの第２のジャーナル論理ボリュームに記憶されているジャーナルデータ内のデータを前記複数の第１の副論理ボリュームのうちの少なくとも一つ又は前記複数の第２の副論理ボリュームのうちの少なくとも一つに書き込むために、前記少なくとも一つの第１のジャーナル論理ボリューム又は前記少なくとも一つの第２のジャーナル論理ボリュームから読み出したジャーナルデータ、

を実行させることを特徴とする副ストレージシステムの制御部のプログラム。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は記憶システムに関し、特に複数の記憶システム間でのデータの複製に関する。

【０００２】

【従来の技術】

近年、常に顧客に対して継続したサービスを提供するために、第一の記憶システムに障害が発生した場合でもデータ処理システムがサービスを提供できるよう、記憶システム間でのデータの複製に関する技術が重要になっている。第一の記憶システムに格納された情報を第二および第三の記憶システムに複製する技術として、以下の特許文献に開示された技術が存在する。

【０００３】

米国特許５１７０４８０号公報には、第一の記憶システムに接続された第一の計算機が、第一の記憶システムに格納されたデータを、第一の計算機と第二の計算機間の通信リンクを介し、第二の計算機に転送し、第二の計算機が第二の計算機に接続された第二の記憶システムに転送する技術が開示されている。

【０００４】

米国特許６２０９００２号公報には、第一の記憶システムが、第一の記憶システムに格納されたデータを、第二の記憶システムに転送し、さらに、第二の記憶システムが、第三の記憶システムに転送する技術が開示されている。計算機と第一の記憶システムとは通信リンクにより接続され、第一の記憶システムと第二の記憶システムとは通信リンクにより接続され、さらに、第二の記憶システムと第三の記憶システムとは通信リンクにより接続されている。第一の記憶システムは、複製対象の第一の論理ボリュームを保持する。第二の記憶システムは、第一の論理ボリュームの複製である第二の論理ボリューム、及び、第二の論理ボリュームの複製である第三の論理ボリュームを保持する。第三の記憶システムは

、第三の論理ボリュームの複製である第四の論理ボリュームを保持する。この特許文献において、第二の記憶システムは、第二の論理ボリュームから第三の論理ボリュームへのデータ複製処理と、第三の論理ボリュームから第四の論理ボリュームへのデータ複製処理とを排他的に実行する。

【０００５】

【特許文献１】

米国特許５１７０４８０号公報

【特許文献２】

米国特許６２０９００２号公報

【発明が解決しようとする課題】

米国特許５１７０４８０号公報に開示された技術は、データの複製のために、第一の計算機および第二の計算機を常に使用する。第一の計算機は、通常業務を行っており、第一の計算機にかかるデータ複製処理の負荷は無視できない。さらに、複製のためのデータは、第一の計算機と第一の記憶システム間の通信リンクを使用するため、通常業務のために必要なデータ転送と衝突し、通常業務に必要なデータ参照、データ更新時間が長くなるという課題がある。

【０００６】

米国特許６２０９００２号公報に開示された技術は、複製を行うデータ量の倍の記憶容量が第二の記憶システム及び第三の記憶システムに必要となる。また、複製対象のデータ量が多いことにより、データ複製処理に費やされる時間が長く、第三の記憶システムのデータは古いものとなってしまう。その結果、第三の記憶システムのデータを用いて業務が再開される場合、第三の記憶システムのデータを最新とするまでの時間が長くなり、業務再開までの時間が延びるという課題がある。さらに、かかる文献において、第一の記憶システムは、第一の記憶システム内のデータ更新処理に加えて第二の記憶システムとの間でのデータ行進処理が終了した時点で、上位の計算機にデータ更新完了報告を行う。したがって、計算機からのデータ更新に費やされる時間が長く、第一の記憶システムと第二の記憶システムとの間の距離が遠くなればなるほど、データ更新に費やされる時間は長くなる。その結果、かかる文献に開示された技術によれば、各記憶システム間の距離をあまり遠くすることができないという課題もある。

【０００７】

本発明の目的は、記憶システムの上位の計算機に影響を与えず、複数の記憶システム間でデータ転送又はデータの複製をするものである。さらに、記憶システムと計算機との間の通信にも影響を与えないものである。

【０００８】

さらに、本発明の目的は、複数の記憶システム内に保持するデータ格納領域を少なくすることができるものである。さらに、複数の記憶システムの上位の計算機の業務に影響を与えることのないように、高速かつ効率的に複数の記憶システム間でデータ転送又はデータの複製をするものである。

【０００９】

【課題を解決するための手段】

これらの課題を解決するために、本発明において、第一の記憶システムは、第一の記憶システムに格納されたデータの更新に関する情報をジャーナルとして格納する。ジャーナルは、具体的には、更新に用いられたデータのコピーと更新時のライト命令等の更新情報とによって構成される。さらに、第二の記憶システムは、第一の記憶システムと第二の記憶システム間の通信線を介して、前記ジャーナルを取得する。第二の記憶システムは、第一の記憶システムが保持するデータの複製を保持しており、前記ジャーナルを用いて、第一の記憶システムでのデータ更新順に、第一の記憶システムのデータと対応するデータを更新する。

【００１０】

本発明においては、第二の記憶システムは、第一の記憶システムが格納する第一の記憶領

域に格納されたデータの複製を格納する第二の記憶領域を保持し、第二の記憶領域のデータの更新に関する情報をジャーナルとしてジャーナル専用の第三の記憶領域に格納する。第三の記憶領域の記憶容量は、第二の記憶領域より少ない記憶容量とすることが可能である。さらに、第三の記憶システムは、第二の記憶システムと第三の記憶システム間の通信線を介して、前記ジャーナルを取得し、ジャーナル専用の第四の記憶領域に格納する。第四の記憶領域の記憶容量は、第二の記憶領域より少ない記憶容量とすることが可能である。第三の記憶システムは、第二の記憶領域に格納されたデータの複製を格納する第五の記憶領域を保持し、前記ジャーナルを用いて、第二の記記憶領域でのデータ更新順に、第二の記憶領域に対応する第五の記憶領域のデータを更新する。

【００１１】

【発明の実施の形態】

本発明によるデータ処理システムの実施形態を図面により詳細に説明する。

【００１２】

図１は、本発明の一実施形態の論理的な構成を示すブロック図である。

【００１３】

本発明の一実施形態は、ホストコンピュータ１８０と記憶システム１００Ａを接続パス１９０により接続し、記憶システム１００Ａと記憶システム１００Ａに保存されたデータの複製を保持する記憶システム１００Ｂを接続パス２００により接続した構成である。以下の説明において、複製対象のデータを保持する記憶システム１００と複製データを保持する記憶システム１００との区別を容易とするために、複製対象のデータを保持する記憶システム１００を正記憶システム１００Ａ、複製データを保持する記憶システム１００を副記憶システム１００Ｂとよぶこととする。なお、記憶システムの記憶領域は、分割して管理されており、分割された記憶領域を論理ボリュームと呼ぶこととする。

【００１４】

記憶システム１００の記憶領域は、分割して管理されており、分割した記憶領域を論理ボリューム２３０とよぶこととする。論理ボリューム２３０の容量および記憶システム１００内の物理的な格納位置（物理アドレス）は、記憶システム１００に接続したコンピュータ等の保守端末もしくはホストコンピュータ１８０を用いて指定できる。各論理ボリューム２３０の物理アドレスは、後述するボリューム情報４００に保存する。物理アドレスは、例えば、記憶システム１００内の記憶装置１５０を識別する番号（記憶装置番号）と記憶装置内の記憶領域を一意に示す数値、例えば、記憶装置の記憶領域の先頭からの位置である。以下の説明では、物理アドレスは、記憶装置番号と記憶装置の記憶領域の先頭からの位置の組とする。以下の説明では、論理ボリュームは、１つの記憶装置の記憶領域であるが、論理アドレスと物理アドレスの変換により、１つの論理ボリュームを複数の記憶装置の記憶領域に対応づけることも可能である。

【００１５】

記憶システム１００が保存しているデータの参照、更新は、論理ボリュームを識別する番号（論理ボリューム番号）と記憶領域を一意に示す数値、例えば、論理ボリュームの記憶領域の先頭からの位置により一意に指定することができ、以下、論理ボリューム番号と論理ボリュームの記憶領域の先頭からの位置（論理アドレス内位置）の組を論理アドレスとよぶ。

【００１６】

以下の説明において、複製対象のデータと複製データとの区別を容易とするために、複製対象の論理ボリューム２３０を正論理ボリューム、複製データである論理ボリューム２３０を副論理ボリュームとよぶこととする。一対の正論理ボリュームと副論理ボリュームをペアとよぶ。正論理ボリュームと副論理ボリュームの関係および状態等は後述するペア情報５００に保存する。

【００１７】

論理ボリューム間のデータの更新順序を守るために、グループという管理単位を設ける。例えば、ホストコンピュータ１８０が、正論理ボリューム１のデータ１を更新し、その後

、データ１を読み出し、データ１の数値を用いて、正論理ボリューム２のデータ２を更新する処理を行うとする。正論理ボリューム１から副論理ボリューム１へのデータ複製処理と、正論理ボリューム２から副論理ボリューム２へのデータ複製処理とが独立に行われる場合、副論理ボリューム１へのデータ１の複製処理より前に、副論理ボリューム２へのデータ２の複製処理が行われる場合がある。副論理ボリューム２へのデータ２の複製処理と副論理ボリューム１へのデータ１の複製処理との間に、故障等により副論理ボリューム１へのデータ１の複製処理が停止した場合、副論理ボリューム１と副論理ボリューム２のデータの整合性がなくなる。このような場合にも副論理ボリューム１と副論理ボリューム２のデータの整合性を保つために、データの更新順序を守る必要のある論理ボリュームは、同じグループに登録し、データの更新毎に、後述するグループ情報６００の更新番号を割り当て、更新番号順に、副論理ボリュームに複製処理を行う。例えば、図１では、正記憶システム１００Ａの論理ボリューム（ＤＡＴＡ１）と論理ボリューム（ＤＡＴＡ２）がグループ１を構成する。論理ボリューム（ＤＡＴＡ１）の副製である論理ボリューム（ＣＯＰＹ１）と論理ボリューム（ＤＡＴＡ２）の複製である論理ボリューム（ＣＯＰＹ２）は、副記憶システム内でグループ１を構成する。

【００１８】

データ複製対象である正論理ボリュームのデータを更新する場合、副論理ボリュームのデータを更新するために、後述するジャーナルを作成し、正記憶システム１００Ａ内の論理ボリュームに保存する。本実施例の説明では、グループ毎にジャーナルのみを保存する論理ボリューム（以下、ジャーナル論理ボリュームとよぶ）を割り当てる。図１では、グループ１に論理ボリューム（ＪＮＬ１）を割り当てている。

【００１９】

副記憶システム１００Ｂのグループにもジャーナル論理ボリュームを割り当てる。ジャーナル論理ボリュームは、正記憶システム１００Ａから副記憶システム１００Ｂに転送したジャーナルを保存するために使用する。ジャーナルをジャーナル論理ボリュームに保存することにより、例えば、副記憶システム１００Ｂの負荷が高い場合、ジャーナル受信時に副論理ボリュームのデータ更新を行わず、暫く後、副記憶システム１００Ｂの負荷が低い時に、副論理ボリュームのデータを更新することもできる。さらに、接続線２００が複数ある場合、正記憶システム１００Ａから副記憶システム１００Ｂへのジャーナルの転送を多重に行い、接続線２００の転送能力を有効に利用することができる。更新順番のため、副記憶システム１００Ｂに多くのジャーナルが溜まる可能性があるが、副論理ボリュームのデータ更新に直ぐに使用できないジャーナルは、ジャーナル論理ボリュームに退避することにより、キャッシュメモリを開放することができる。図１では、副記憶システム内のグループ１に論理ボリューム（ＪＮＬ２）を割り当てている。

【００２０】

ジャーナルは、ライトデータと更新情報とから構成する。更新情報は、ライトデータを管理するための情報で、ライト命令を受信した時刻、グループ番号、後述するグループ情報６００の更新番号、ライト命令の論理アドレス、ライトデータのデータサイズ、ライトデータを格納したジャーナル論理ボリュームの論理アドレス等からなる。更新情報は、ライト命令を受信した時刻と更新番号のどちらか一方のみを保持してもよい。ホストコンピュータ１８０からのライト命令内にライト命令の作成時刻が存在する場合は、ライト命令を受信した時刻の代わりに、当該ライト命令内の作成時刻を使用してもよい。図３と図２１を用いて、ジャーナルの更新情報の例を説明する。更新情報３１０は、１９９９年３月１７日の２２時２０分１０秒に受信したライト命令を記憶する。当該ライト命令は、ライトデータを論理ボリューム番号１の記憶領域の先頭から７００の位置に格納する命令であり、データサイズは３００である。ジャーナルのライトデータは、論理ボリューム番号４（ジャーナル論理ボリューム）の記憶領域の先頭から１５００の位置から格納されている。論理ボリューム番号１の論理ボリュームはグループ１に属し、グループ１のデータ複製開始から４番目のデータ更新であることがわかる。

【００２１】

ジャーナル論理ボリュームは、例えば、図３に示すように、更新情報を格納する記憶領域（更新情報領域）とライトデータを格納する記憶領域（ライトデータ領域）に分割して使用する。更新情報領域は、更新情報領域の先頭から、更新番号の順に格納し、更新情報領域の終端に達すると、更新情報領域の先頭から格納する。ライトデータ領域は、ライトデータ領域の先頭からライトデータを格納し、ライトデータ領域の終端に達すると、ライトデータ領域の先頭から格納する。更新情報領域およびライトデータ領域の比は、固定値でもよいし、保守端末あるいはホストコンピュータ１８０により設定可能としてもよい。これらの情報は、後述するポインタ情報７００内に保持する。以下の説明では、ジャーナル論理ボリュームを更新情報とライトデータの領域に分割して、ジャーナル論理ボリュームを使用するが、論理ボリュームの先頭から、ジャーナル、つまり、更新情報とライトデータを連続に格納する方式を採用してもよい。

【００２２】

図１を用いて、記憶システム１００Ａの正論理ボリュームへのデータ更新を副記憶システム１００Ｂの副論理ボリュームに反映する動作について概説する。

【００２３】

（１）記憶システム１００Ａは、ホストコンピュータ１８０から正論理ボリューム内のデータに対するライト命令を受信すると、後述する命令受信処理２１０およびリードライト処理２２０によって、正論理ボリューム（ＤＡＴＡ１）内のデータ更新と、ジャーナル論理ボリューム（ＪＮＬ１）にジャーナルの保存を行う(図１の２７０)。

【００２４】

（２）記憶システム１００Ｂは、後述するジャーナルリード処理２４０によって、記憶システム１００Ａからジャーナルをリードし、リードライト処理２２０によって、ジャーナル論理ボリューム（ＪＮＬ２）にジャーナルを保存する(図１の２８０)。

【００２５】

（３）記憶システム１００Ａは、記憶システム１００Ｂからジャーナルをリードする命令を受信すると、後述する命令受信処理２１０およびリードライト処理２２０によって、ジャーナル論理ボリューム（ＪＮＬ１）からジャーナルを読み出し、記憶システム１００Ｂに送信する(図１の２８０)。

【００２６】

（４）記憶システム１００Ｂは、後述するリストア処理２５０およびリードライト処理２２０によって、ポインタ情報７００を用いて、更新番号の昇順に、ジャーナル論理ボリューム（ＪＮＬ２）からジャーナルを読み出し、副論理ボリューム（ＣＯＰＹ１）のデータを更新する(図１の２９０)。

【００２７】

記憶システム１００の内部構造を図２に示す。記憶システム１００は、１つ以上のホストアダプタ１１０、１つ以上のディスクアダプタ１２０、１つ以上のキャッシュメモリ１３０、１つ以上の共有メモリ１４０、１つ以上の記憶装置１５０、１つ以上のコモンパス１６０、１つ以上の接続線１７０を備えて構成される。ホストアダプタ１１０、ディスクアダプタ１２０、キャッシュメモリ１３０、共有メモリ１４０はコモンパス１６０により相互間が接続されている。コモンパス１６０は、コモンパス１６０の障害時のために２重化されてもよい。ディスクアダプタ１２０と記憶装置１５０とは接続線１７０によって接続されている。また、図示していないが、記憶システム１００の設定、監視、保守等を行うための保守端末が全てのホストアダプタ１１０とディスクアダプタ１２０とに専用線を用いて接続されている。

【００２８】

ホストアダプタ１１０は、ホストコンピュータ１８０とキャッシュメモリ１３０間のデータ転送を制御する。ホストアダプタ１１０は、接続線１９０および接続線２００によりホストコンピュータ１８０もしくは、他の記憶システム１００と接続される。ディスクアダプタ１２０は、キャッシュメモリ１３０と記憶装置１５０との間のデータ転送を制御する。キャッシュメモリ１３０は、ホストコンピュータ１８０から受信したデータあるいは記

憶装置１５０から読み出したデータを一時的に保持するメモリである。共有メモリ１４０は、記憶システム１００内の全てのホストアダプタ１１０とディスクアダプタ１２０とが共有するメモリである。

【００２９】

ボリューム情報４００は、論理ボリュームを管理する情報であり、ボリューム状態、フォーマット形式、容量、ペア番号、物理アドレスを保持する。図４にボリューム情報４００の一例を示す。ボリューム情報４００は、ホストアダプタ１１０およびディスクアダプタ１２０から参照可能なメモリ、例えば管理メモリ１４０に保存される。ボリューム状態は、“正常”、“正”、“副”、“異常”、“未使用”のいずれかを保持する。ボリューム状態が“正常”もしくは“正”である論理ボリューム２３０は、ホストコンピュータ１８０から正常にアクセス可能な論理ボリューム２３０であることを示す。ボリューム状態が“副”である論理ボリューム２３０は、ホストコンピュータ１８０からのアクセスを許可してもよい。ボリューム状態が“正”である論理ボリューム２３０は、データの複製が行われている論理ボリューム２３０であることを示す。ボリューム状態が“副”である論理ボリューム２３０は、複製に使用されている論理ボリューム２３０であることを示す。ボリューム状態が“異常”の論理ボリューム２３０は、障害により正常にアクセスできない論理ボリューム２３０であることを示す。障害とは、例えば、論理ボリューム２３０を保持する記憶装置１５０の故障である。ボリューム状態が“未使用”の論理ボリューム２３０は、使用していない論理ボリューム２３０であることを示す。ペア番号は、ボリューム状態が“正”もしくは“副”の場合に有効であり、後述するペア情報５００を特定するためのペア番号を保持する。図４に示す例では、論理ボリューム１は、フォーマット形式がＯＰＥＮ３、容量が３ＧＢ、記憶装置番号１の記憶装置１５０の記憶領域の先頭からデータが格納されており、アクセス可能であり、データの複製対象であることを示す。

【００３０】

ペア情報５００は、ペアを管理する情報であり、ペア状態、正記憶システム番号、正論理ボリューム番号、副記憶システム番号、副論理ボリューム番号、グループ番号、コピー済みアドレスを保持する。図５にペア情報５００の一例を示す。ペア情報５００は、ホストアダプタ１１０およびディスクアダプタ１２０から参照可能なメモリ、例えば管理メモリ１４０に保存する。ペア状態は、“正常”、“異常”、“未使用”,“コピー未”、“コピー中”のいずれかを保持する。ペア状態が“正常”の場合は、正論理ボリューム２３０のデータ複製が正常に行われていることを示す。ペア状態が“異常”の場合は、障害により正論理ボリューム２３０の複製が行えないことを示す。障害とは、例えば、接続パス２００の断線などである。ペア状態が“未使用”の場合は、当該ペア番号の情報は有効でないことを示す。ペア状態が“コピー中”の場合は、後述する初期コピー処理中であることを示す。ペア状態が“コピー未”の場合は、後述する初期コピー処理が未だ行われていないことを示す。正記憶システム番号は、正論理ボリューム２３０を保持する正記憶システム１００Ａを特定する番号を保持する。副記憶システム番号は、副論理ボリューム２３０を保持する副記憶システム１００Ｂを特定する番号を保持する。グループ番号は、正記憶システムの場合は、正論理ボリュームが属するグループ番号を保持する。副記憶システムの場合は、副論理ボリュームが属するグループ番号を保持する。コピー済みアドレスは、後述する初期コピー処理にて説明する。図５のペア情報１は、データ複製対象が正記憶システム１の正論理ボリューム１であり、データ複製先が副記憶システム２の副論理ボリューム１であり、正常にデータ複製処理が行われていることを示す。

【００３１】

グループ情報６００は、グループ状態、ペア集合、ジャーナル論理ボリューム番号、更新番号を保持する。図６にグループ情報６００の一例を示す。グループ情報６００は、ホストアダプタ１１０およびディスクアダプタ１２０から参照可能なメモリ、例えば管理メモリ１４０に保存する。グループ状態は、“正常”、“異常”、“未使用”のいずれかを保持する。グループ状態が“正常”の場合は、ペア集合の少なくともひとつのペア状態が“正常”であることを示す。グループ状態が“異常”の場合は、ペア集合の全てのペア状態

が“異常”であることを示す。グループ状態が“未使用”の場合は、当該グループ番号の情報は有効でないことを示す。ペア集合は、正記憶システムの場合は、グループ番号が示すグループに属する全ての正論理ボリュームのペア番号を保持する。副記憶システムの場合は、グループ番号が示すグループに属する全ての副論理ボリュームのペア番号を保持する。ジャーナル論理ボリューム番号は、当該グループ番号のグループに属するジャーナル論理ボリューム番号を示す。更新番号は、初期値は１であり、グループ内の正論理ボリュームに対しデータの書き込みが行われると、１を加える。更新番号は、ジャーナルの更新情報に記憶し、副記憶システム１００Ｂにて、データの更新順を守るために使用する。例えば、図６のグループ情報１は、ペア情報１，２から、正論理ボリューム１，２と、ジャーナル論理ボリューム４から構成されており、正常にデータの複製処理が行われていることを示す。

【００３２】

ポインタ情報７００は、グループ毎に保持し、当該グループのジャーナル論理ボリュームを管理する情報であり、更新情報領域先頭アドレス、ライトデータ領域先頭アドレス、更新情報最新アドレス、更新情報最古アドレス、ライトデータ最新アドレス、ライトデータ最古アドレス、リード開始アドレス、リトライ開始アドレスを保持する。図７および図８にポインタ情報７００の一例を示す。更新情報領域先頭アドレスは、ジャーナル論理ボリュームの更新情報を格納する記憶領域（更新情報領域）の先頭の論理アドレスを保持する。ライトデータ領域先頭アドレスは、ジャーナル論理ボリュームのライトデータを格納する記憶領域（ライトデータ領域）の先頭の論理アドレスを保持する。更新情報最新アドレスは、次にジャーナルを格納する場合に、更新情報の保存に使用する先頭の論理アドレスを保持する。更新情報最古アドレスは、最古の（更新番号が小さい）ジャーナルの更新情報を保存する先頭の論理アドレスを保持する。ライトデータ最新アドレスは、次にジャーナルを格納する場合に、ライトデータの保存に使用する先頭の論理アドレスを保持する。ライトデータ最古アドレスは、最古の（更新番号が小さい）ジャーナルのライトデータを保存する先頭の論理アドレスを保持する。リード開始アドレスとリトライ開始アドレスは、正記憶システム１００Ａのみで使用し、後述するジャーナルリード受信処理にて使用する。図７および図８のポインタ情報７００の例では、ジャーナルの管理情報を保存する領域（更新情報領域）は、論理ボリューム４の記憶領域の先頭から６９９の位置までであり、ジャーナルのライトデータを保存する領域（ライトデータ領域）は、論理ボリューム４の記憶領域の７００の位置から２６９９の位置までである。ジャーナルの管理情報は、論理ボリューム４の記憶領域の２００の位置から４９９の位置まで保存されており、次のジャーナルの管理情報は、論理ボリューム４の記憶領域の５００の位置から保存する。ジャーナルのライトデータは論理ボリューム４の記憶領域の１３００の位置から２１９９の位置まで保存されており、次のジャーナルのライトデータは、論理ボリューム４の記憶領域の２２００の位置から保存する。

【００３３】

下記の説明では、１つのグループに１つのジャーナル論理ボリュームを割り当てる形態で説明しているが、１つのグループに複数のジャーナル論理ボリュームを割り当ててもよい。例えば、１つのグループに２つのジャーナル論理ボリュームを割り当て、ジャーナル論理ボリューム毎にポインタ情報７００を設け、交互にジャーナルを格納する。これにより、ジャーナルの記憶装置１５０への書き込みが分散でき、性能の向上が見込める。さらに、ジャーナルのリード性能も向上する。別の例としては、１つのグループに２つのジャーナル論理ボリュームを割り当て、通常は、１つのジャーナル論理ボリュームのみを使用する。もう一方のジャーナル論理ボリュームは、使用しているジャーナル論理ボリュームの性能が低下した場合に使用する。性能が低下する例は、ジャーナル論理ボリュームが、複数の記憶装置１５０から構成され、ＲＡＩＤ５の方式でデータを保持しており、構成する記憶装置１５０の一台が故障中の場合である。

【００３４】

なお、上述のボリューム情報４００、ペア情報５００、グループ情報６００、及びポイン

タ情報７００等は、共有メモリ１４０に格納されていることが好ましい。しかし、本実施例は、この場合に限られず、これらの情報を、キャッシュメモリ１３０、ホストアダプタ１１０、ディスクアダプタ１２０、その他記憶装置１５０に集中して格納または分散して格納することもよい。

【００３５】

次に、正記憶システム１００Ａから副記憶システム１００Ｂに対して、データ複製を開始する手順を図９、図１０を用いて説明する。

【００３６】

（１）グループ作成について説明する（ステップ９００）。ユーザは、保守端末あるいはホストコンピュータ１８０を使用して、正記憶システム１００Ａのグループ情報６００を参照し、グループ状態が“未使用”のグループ番号Ａを取得する。ユーザは、保守端末あるいはホストコンピュータ１８０を使用して、グループ番号Ａを指定し、グループ作成指示を正記憶システム１００Ａに行う。

【００３７】

グループ作成指示を受けて、正記憶システム１００Ａは、指定されたグループ番号Ａのグループ状態を“正常”に変更する。

【００３８】

同様に、ユーザは、副記憶システム１００Ｂのグループ情報６００を参照し、グループ状態が“未使用”のグループ番号Ｂを取得する。ユーザは、保守端末あるいはホストコンピュータ１８０を使用して、副記憶システム１００Ｂとグループ番号Ｂを指定し、グループ作成指示を正記憶システム１００Ａに行う。正記憶システム１００Ａは、受信したグループ作成指示を副記憶システム１００Ｂに転送する。副記憶システム１００Ｂは、指定されたグループ番号Ｂのグループ状態を“正常”に変更する。

【００３９】

ユーザは、副記憶システム１００Ｂの保守端末あるいは副記憶システム１００Ｂに接続したホストコンピュータ１８０を使用して、グループ番号Ｂを指定し、グループ作成指示を副記憶システム１００Ｂに行ってもよい。

【００４０】

（２）ペア登録について説明する（ステップ９１０）。ユーザは、保守端末あるいはホストコンピュータ１８０を使用して、データ複製対象を示す情報とデータ複製先を示す情報を指定し、ペア登録指示を正記憶システム１００Ａに行う。データ複製対象を示す情報は、データ複製対象のグループ番号Ａと正論理ボリューム番号である。データ複製先を示す情報は、複製データを保存する副記憶システム１００Ｂとグループ番号Ｂ、副論理ボリューム番号である。

【００４１】

前記ペア登録指示を受けて、正記憶システム１００Ａは、ペア情報５００からペア情報が“未使用”のペア番号を取得し、ペア状態を“コピー未”に、正記憶システム番号に、正記憶システム１００Ａを示す正記憶システム番号を、正論理ボリューム番号に、指示された正論理ボリューム番号を、副記憶システム番号に、指示された副記憶システム番号を、副論理ボリューム番号に、指示された副論理ボリューム番号を、グループ番号に、指示されたグループ番号Ａを設定する。正記憶システム１００Ａは、指示されたグループ番号Ａのグループ情報６００のペア集合に前記取得したペア番号を追加し、正論理ボリューム番号のボリューム状態を“正”に変更する。

【００４２】

正記憶システム１００Ａは、正記憶システム１００Ａを示す正記憶システム番号、ユーザから指定されたグループ番号Ｂ、正論理ボリューム番号、および副論理ボリューム番号を副記憶システム１００Ｂに指示する。副記憶システム１００Ｂは、ペア情報５００から未使用のペア番号を取得し、ペア状態を“コピー未”に、正記憶システム番号に、記憶システム１００Ａを示す正記憶システム番号を、正論理ボリューム番号に、指示された正論理ボリューム番号を、副記憶システム番号に、副記憶システムＢを示す副記憶システム番号

を、副論理ボリューム番号に、指示された副論理ボリューム番号を、グループ番号に、指示されたグループ番号Ｂを設定する。

【００４３】

副記憶システム１００Ｂは、指示されたグループ番号Ｂのグループ情報６００のペア集合に前記取得したペア番号を追加し、副論理ボリューム番号のボリューム状態を“副”に変更する。

【００４４】

以上の動作を全てのデータ複製対象のペアに対して行う。

【００４５】

前記の説明では、論理ボリュームのグループへの登録と、論理ボリュームのペアの設定を同時に行う処理を説明したが、それぞれ個別に行ってもよい。

【００４６】

（３）ジャーナル論理ボリューム登録について説明する（ステップ９２０）。ユーザは、保守端末あるいはホストコンピュータ１８０を使用して、ジャーナルの保存に使用する論理ボリューム（ジャーナル論理ボリューム）をグループに登録する指示（ジャーナル論理ボリューム登録指示）を正記憶システム１００Ａに行う。ジャーナル論理ボリューム登録指示は、グループ番号と論理ボリューム番号からなる。

【００４７】

正記憶システム１００Ａは、指示されたグループ番号のグループ情報６００のジャーナル論理ボリューム番号に指示された論理ボリューム番号を登録する。当該論理ボリュームのボリューム情報４００のボリューム状態を“正常”に設定する。

【００４８】

同様に、ユーザは、保守端末あるいはホストコンピュータ１８０を使用して、副記憶システム１００Ｂのボリューム情報４００を参照し、副記憶システム１００Ｂ、グループ番号Ｂ、ジャーナル論理ボリュームとして使用する論理ボリューム番号を指定し、ジャーナル論理ボリューム登録を正記憶システム１００Ａに行う。正記憶システム１００Ａは、ジャーナル論理ボリューム登録指示を副記憶システム１００Ｂに転送する。副記憶システム１００Ｂは、指示されたグループ番号Ｂのグループ情報６００のジャーナル論理ボリューム番号に指示された論理ボリューム番号を登録する。当該論理ボリュームのボリューム情報４００のボリューム状態を“正常”に設定する。

【００４９】

ユーザは、副記憶システム１００Ｂの保守端末あるいは副記憶システム１００Ｂに接続したホストコンピュータ１８０を使用して、グループ番号、ジャーナル論理ボリュームとして使用する論理ボリューム番号を指定し、ジャーナル論理ボリューム登録指示を副記憶システム１００Ｂに行ってもよい。

【００５０】

以上の動作を全てのジャーナル論理ボリュームとして使用する論理ボリュームに対して行う。ステップ９１０とステップ９２０の順は不同ではない。

【００５１】

（４）データ複製処理開始について説明する（ステップ９３０）。ユーザは、保守端末あるいはホストコンピュータ１８０を使用して、データ複製処理を開始するグループ番号を指定し、データ複製処理の開始を正記憶システム１００Ａに指示する。正記憶システム１００Ａは、指示されたグループに属する全てのペア情報４００のコピー済みアドレスを０に設定する。

【００５２】

正記憶システム１００Ａは、副記憶システム１００Ｂに、後述するジャーナルリード処理およびリストア処理の開始を指示する。

【００５３】

正記憶システム１００Ａは、後述する初期コピー処理を開始する。

【００５４】

（５）初期コピー終了について説明する（ステップ９４０）。
【００５５】
初期コピーが終了すると、正記憶システム１００Ａは、初期コピー処理の終了を副記憶システム１００Ｂに通知する。副記憶システム１００Ｂは、指示されたグループに属する全ての副論理ボリュームのペア状態を“正常”に変更する。
【００５６】
図１０は、初期コピー処理のフローチャートである。初期コピー処理は、データ複製対象の正論理ボリュームの全記憶領域に対し、ペア情報５００のコピー済みアドレスを用い、記憶領域の先頭から順に、単位サイズ毎にジャーナルを作成する。コピー済みアドレスは、初期値は0であり、ジャーナルの作成毎に、作成したデータ量を加算する。論理ボリュームの記憶領域の先頭から、コピー済みアドレスの一つ前までは、初期コピー処理にてジャーナルは作成済みである。初期コピー処理を行うことにより、正論理ボリュームの更新されていないデータを副論理ボリュームに転送することが可能となる。以下の説明では、正記憶システム１００Ａ内のホストアダプタＡが処理を行うように記載しているが、ディスクアダプタ１２０が行ってもよい。
【００５７】
（１）正記憶システム１００Ａ内のホストアダプタＡは、処理対象のグループに属するペアでペア状態が“コピー未”である正論理ボリュームＡを得、ペアの状態を“コピー中”に変更し、以下の処理を繰り返す（ステップ１０１０、１０２０）。正論理ボリュームＡが存在しない場合は、処理を終了する（ステップ１０３０）。
【００５８】
（２）ステップ１０２０にて、論理ボリュームＡが存在した場合、ホストアダプタＡは、単位サイズ（例えば、１ＭＢ）のデータを対象にジャーナルを作成する。ジャーナル作成処理は後述する（ステップ１０４０）。
【００５９】
（３）ホストアダプタＡは、コピー済みアドレスに作成したジャーナルのデータサイズを加算する（ステップ１０５０）。
【００６０】
（４）コピー済みアドレスが、正論理ボリュームＡの容量に達するまで、上記処理を繰り返す（ステップ１０６０）。コピー済みアドレスが、正論理ボリュームＡの容量と等しくなった場合、正論理ボリュームＡの全記憶領域に対してジャーナルを作成したため、ペア状態を“正常”に更新し、他の正論理ボリュームの処理を開始する（ステップ１０７０）。
【００６１】
前記のフローチャートでは、論理ボリュームを１つずつ対象とするように説明したが、複数の論理ボリュームを同時に処理してもよい。
【００６２】
図１１は命令受信処理２１０の処理を説明する図、図１２は命令受信処理２１０のフローチャート、図１３はジャーナル作成処理のフローチャートである。以下、これらを用いて、正記憶システム１００Ａが、ホストコンピュータ１８０からデータ複製対象の論理ボリューム２３０にライト命令を受信した場合の動作について説明する。
【００６３】
（１）記憶システム１００Ａ内のホストアダプタＡは、ホストコンピュータからアクセス命令を受信する。アクセス命令は、リード、ライト、後述するジャーナルリード等の命令、命令対象の論理アドレス、データ量等を含んでいる。以下、アクセス命令内の論理アドレスを論理アドレスＡ、論理ボリューム番号を論理ボリュームＡ、論理ボリューム内位置を論理ボリューム内位置Ａ、データ量をデータ量Ａとする（ステップ１２００）。
【００６４】
（２）ホストアダプタＡは、アクセス命令を調べる（ステップ１２１０、１２１５）。ステップ１２１５の調べで、アクセス命令がジャーナルリード命令の場合は、後述するジャ

ーナルリード受信処理を行う（ステップ１２２０）。アクセス命令がジャーナルリード命令およびライト命令以外、例えば、リード命令の場合は、従来技術と同じようにリード処理を行う（ステップ１２３０）。

【００６５】

（３）ステップ１２１０の調べで、アクセス命令がライト命令の場合は、論理ボリュームＡのボリューム情報４００を参照し、ボリューム状態を調べる（ステップ１２４０）。ステップ１２４０の調べで、論理ボリュームＡのボリューム状態が、“正常”もしくは“正”以外の場合は、論理ボリュームＡへのアクセスは不可能なため、ホストコンピュータ１８０に異常終了を報告する（ステップ１２４５）。

【００６６】

（４）ステップ１２４０の調べで、論理ボリュームＡのボリューム状態が、“正常”、“正”のいずれかの場合は、ホストアダプタＡは、キャッシュメモリ１３０を確保し、ホストコンピュータ１８０にデータ受信の準備ができたことを通知する。ホストコンピュータ１８０は、その通知を受け、ライトデータを正記憶システム１００Ａに送信する。ホストアダプタＡは、ライトデータを受信し、当該キャッシュメモリ１３０に保存する（ステップ１２５０、図１１の１１００）。

【００６７】

（５）ホストアダプタＡは、論理ボリュームＡのボリューム状態を参照し、論理ボリュームＡがデータ複製対象かどうかを調べる（ステップ１２６０）。ステップ１２６０の調べで、ボリューム状態が、“正”である場合は、論理ボリュームＡがデータ複製対象であるため、後述するジャーナル作成処理を行う（ステップ１２６５）。

【００６８】

（６）ステップ１２６０の調べで、ボリューム状態が、“正常”である場合、もしくはステップ１２６５のジャーナル作成処理の終了後、ホストアダプタＡは、ディスクアダプタ１２０にライトデータを記憶装置１５０に書き込むことを命令し（図１１の１１４０）、ホストコンピュータ１８０に終了報告する（ステップ１２７０、１２８０）。その後、当該ディスクアダプタ１２０は、リードライト処理により、記憶装置１５０にライトデータを保存する（図１１の１１１０）。

【００６９】

次に、ジャーナル作成処理について説明する。

【００７０】

（１）ホストアダプタＡは、ジャーナル論理ボリュームのボリューム状態を調べる（ステップ１３１０）。ステップ１３１０の調べで、ジャーナル論理ボリュームのボリューム状態が、“異常”の場合は、ジャーナル論理ボリュームにジャーナルの格納が不可能なため、グループ状態を“異常”に変更し、処理を終了する（ステップ１３１５）。この場合、ジャーナル論理ボリュームを正常な論理ボリュームに変更する等を行う。

【００７１】

（２）ステップ１３１０の調べで、ジャーナル論理ボリュームが正常である場合、ジャーナル作成処理を継続する。ジャーナル作成処理は、初期コピー処理内の処理であるか、命令受信処理内の処理であるかによって処理が異なる（ステップ１３２０）。ジャーナル作成処理が命令受信処理内の処理の場合は、ステップ１３３０からの処理を行う。ジャーナル作成処理が初期コピー処理内の場合は、ステップ１３７０からの処理を行う。

【００７２】

（３）ジャーナル作成処理が命令受信処理内の処理の場合、ホストアダプタＡは、ライト対象の論理アドレスＡが、初期コピー処理の処理対象となったかを調べる（ステップ１３３０）。論理ボリュームＡのペア状態が“コピー未”の場合は、後に初期コピー処理にてジャーナル作成処理が行われるため、ジャーナルを作成せずに処理を終了する（ステップ１３３５）。論理ボリュームＡのペア状態が“コピー中”の場合は、コピー済みアドレスが論理アドレス内位置Ａと等しいもしくは、小さい場合は、後に初期コピー処理にてジャーナル作成処理が行われるため、ジャーナルを作成せずに処理を終了する（ステップ１３

３５）。上記以外、つまり、論理ボリュームＡのペア状態が“コピー中”かつコピー済みアドレスが論理アドレス内位置Ａ以上の場合もしくは、論理ボリュームＡのペア状態が“正常”の場合は、既に初期コピー処理が完了しているため、ジャーナル作成処理を継続する。

【００７３】

（４）次に、ホストアダプタＡは、ジャーナルがジャーナル論理ボリュームに格納可能であるを調べる。ポインタ情報７００を用い、更新情報領域の未使用領域の有無を調べる（ステップ１３４０）。ポインタ情報７００の更新情報最新アドレスと更新情報最古アドレスが等しい場合は、更新情報領域に未使用領域が存在しないため、ジャーナル作成失敗として処理を終了する（ステップ１３９０）。

【００７４】

ステップ１３４０の調べで、更新情報領域に未使用領域が存在する場合は、ポインタ情報７００を用い、ライトデータ領域にライトデータが格納できるかを調べる（ステップ１３４５）。ライトデータ最新アドレスとデータ量Ａの和が、ライトデータ最古アドレスと等しいもしくは大きい場合は、ライトデータ領域に格納できないため、ジャーナル作成失敗として処理を終了する（ステップ１３９０）。

【００７５】

（５）ジャーナルが格納可能である場合、ホストアダプタＡは、更新番号と更新情報を格納する論理アドレスとライトデータを格納する論理アドレスを取得し、更新情報をキャッシュメモリ１３０内に作成する。更新番号は、対象グループのグループ情報６００から取得し、１を足した数値をグループ情報６００の更新番号に設定する。更新情報を格納する論理アドレスは、ポインタ情報７００の更新情報最新アドレスであり、更新情報のサイズを足した数値をポインタ情報７００の更新情報最新アドレスに設定する。ライトデータを格納する論理アドレスは、ポインタ情報７００のライトデータ最新アドレスであり、ライトデータ最新アドレスにデータ量Ａを足した数値をポインタ情報７００のライトデータ最新アドレスに設定する。

【００７６】

ホストアダプタＡは、上記取得した数値とグループ番号、ライト命令を受信した時刻、ライト命令内の論理アドレスＡ、データ量Ａを更新情報に設定する（ステップ１３５０、図１１の１１２０）。例えば、図６に示すグループ情報６００、図７に示すポインタ情報７００の状態で、グループ１に属する正論理ボリューム１の記憶領域の先頭から８００の位置にデータサイズ１００のライト命令を受信した場合、図２２に示す更新情報を作成する。グループ情報の更新番号は５、ポインタ情報の更新情報最新アドレスは６００（更新情報のサイズは１００とする）、ライトデータ最新アドレスは２３００となる。

【００７７】

（６）ホストアダプタＡは、ディスクアダプタ１２０に、ジャーナルの更新情報とライトデータを記憶装置１５０に書き込むことを命令し、正常終了する（ステップ１３６０、図１１の１１３０、１１４０、１１５０）。

【００７８】

（７）ジャーナル作成処理が、初期コピー処理内の処理の場合は、ステップ１３７０からの処理を行う。ホストアダプタＡは、ジャーナルが作成可能であるを調べる。ポインタ情報７００を用い、更新情報領域の未使用領域の有無を調べる（ステップ１３７０）。ポインタ情報７００の更新情報最新アドレスと更新情報最古アドレスが等しい場合は、更新情報領域に未使用領域が存在しないため、ジャーナル作成失敗として処理を終了する（ステップ１３９０）。本実施例で示した初期コピー処理の場合、ジャーナルのライトデータは、正論理ボリュームからリードし、ライトデータ領域は使用しないため、ライトデータ領域の未使用領域の確認は不要である。

【００７９】

（８）ステップ１３７０の調べで、ジャーナルが作成可能である場合、ホストアダプタＡは、更新情報に設定する数値を取得し、更新情報をキャッシュメモリ１３０内に作成する

。更新番号は、対象グループのグループ情報６００から取得し、１を足した数値をグループ情報６００の更新番号に設定する。更新情報を格納する論理アドレスは、ポインタ情報７００の更新情報最新アドレスの位置であり、更新情報のサイズを足した数値をポインタ情報７００の更新情報最新アドレスに設定する。

【００８０】

ホストアダプタＡは、上記取得した数値とグループ番号、本処理の開始時刻、初期コピー処理対象の論理アドレス、初期コピーの１回の処理量、ライトデータを格納したジャーナル論理ボリュームの論理アドレスに初期コピー処理対象の論理アドレスを設定する（ステップ１３８０、図１１の１１２０）。

【００８１】

（９）ホストアダプタＡは、ディスクアダプタ１２０に、更新情報を記憶装置１５０に書き込むことを命令し、正常終了する（ステップ１３８５、図１１の１１４０、１１６０）。

【００８２】

上記説明では、更新情報をキャッシュメモリ１３０内に存在するように記載しているが、共有メモリ１４０内等に格納してもよい。

【００８３】

ライトデータの記憶装置１５０への書き込みは、非同期、つまり、ステップ１３６０およびステップ１３８５の直後でなくともよい。ただし、ホストコンピュータ１８０が、論理アドレスＡにライト命令を再び行った場合、ジャーナルのライトデータが上書きされるため、ホストコンピュータ１８０からライトデータを受信する前に、ジャーナルのライトデータは、更新情報のジャーナル論理ボリュームの論理アドレスに対応する記憶装置１５０に書き込む必要がある。もしくは、別のキャッシュメモリに退避し、後に更新情報のジャーナル論理ボリュームの論理アドレスに対応する記憶装置１５０に書き込みを行ってもよい。

【００８４】

前述したジャーナル作成処理では、ジャーナルを記憶装置１５０に保存するとしていたが、ジャーナル用に予め一定量のキャッシュメモリ１３０を用意しておき、当該キャッシュメモリを全て使用してから、記憶装置１５０にジャーナルを保存してもよい。ジャーナル用のキャッシュメモリ量は、例えば、保守端末から指定する。

【００８５】

リードライト処理２２０は、ディスクアダプタ１２０が、ホストアダプタ１１０もしくはディスクアダプタ１２０から命令を受け、実施する処理である。実施する処理は、指定されたキャッシュメモリ１３０のデータを指定された論理アドレスに対応する記憶装置１５０内の記憶領域に書き込む処理、指定された論理アドレスに対応する記憶装置１５０内の記憶領域から指定されたキャッシュメモリ１３０にデータを読み込む処理等である。

【００８６】

図１４はジャーナルリード命令を受信した正記憶システム１００Ａのホストアダプタの動作（ジャーナルリード受信処理）を説明する図、図１５はフローチャートである。以下、これらを用いて、正記憶システム１００Ａが、副記憶システム１００Ｂからジャーナルリード命令を受信した場合の動作について説明する。

【００８７】

（１）正記憶システム１００Ａ内のホストアダプタＡは、副記憶システム１００Ｂからアクセス命令を受信する。アクセス命令は、ジャーナルリード命令であることを示す識別子、命令対象のグループ番号、リトライ指示の有無を含んでいる。以下、アクセス命令内のグループ番号をグループ番号Ａとする（ステップ１２２０、図１４の１４１０）。

【００８８】

（２）ホストアダプタＡは、グループ番号Ａのグループ状態が“正常”であるかを調べる（ステップ１５１０）。ステップ１５１０の調べで、グループ状態が“正常”以外、例えば、“障害”の場合は、副記憶システム１００Ｂにグループ状態を通知し、処理を終了す

る。副記憶システム１００Ｂは、受信したグループ状態に応じて処理を行う。例えば、グループ状態が“障害”の場合は、ジャーナルリード処理を終了する（ステップ１５１５）。

【００８９】

（３）ステップ１５１０の調べで、グループ番号Ａのグループ状態が“正常”の場合、ホストアダプタＡは、ジャーナル論理ボリュームの状態を調べる（ステップ１５２０）。ステップ１５２０の調べで、ジャーナル論理ボリュームのボリューム状態が“正常”でない場合は、例えば、“障害”の場合は、グループ状態を“障害”に変更し、副記憶システム１００Ｂにグループ状態を通知し、処理を終了する。副記憶システム１００Ｂは、受信したグループ状態に応じて処理を行う。例えば、グループ状態が“障害”の場合は、ジャーナルリード処理を終了する（ステップ１５２５）。

【００９０】

（４）ステップ１５２０の調べで、ジャーナル論理ボリュームのボリューム状態が“正常”の場合は、ジャーナルリード命令がリトライ指示かを調べる（ステップ１５３０）。

【００９１】

（５）ステップ１５３０の調べで、ジャーナルリード命令がリトライ指示の場合、ホストアダプタＡは、前回送信したジャーナルを再度、副記憶システム１００Ｂに送信する。ホストアダプタＡは、キャッシュメモリ１３０を確保し、ディスクアダプタに、ポインタ情報７００のリトライ開始アドレスから、更新情報のサイズの情報をキャッシュメモリに読み込むことを命令する（図１４の１４２０）。

【００９２】

ディスクアダプタのリードライト処理は、記憶装置１５０から更新情報を読み込み、キャッシュメモリ１３０に保存し、ホストアダプタＡに通知する（図１４の１４３０）。

【００９３】

ホストアダプタＡは、更新情報のリード終了の通知を受け、更新情報から、ライトデータの論理アドレスおよびライトデータのサイズを取得し、キャッシュメモリ１３０を確保し、ディスクアダプタにライトデータをキャッシュメモリに読み込むことを命令する（ステップ１５４０、図１４の１４４０）。

【００９４】

ディスクアダプタのリードライト処理は、記憶装置１５０からライトデータを読み込み、キャッシュメモリ１３０に保存し、ホストアダプタＡに通知する（図１４の１４５０）。

【００９５】

ホストアダプタＡは、ライトデータのリード終了の通知を受け、更新情報とライトデータを副記憶システム１００Ｂに送信し、ジャーナルを保持しているキャッシュメモリ１３０を開放し、処理を終了する（ステップ１５４５、図１４の１４６０）。

【００９６】

（６）ステップ１５３０の調べで、リトライ指示でない場合、ホストアダプタＡは、送信していないジャーナルが存在するかを調べ、存在すれば、ジャーナルを副記憶システム１００Ｂに送信する。ホストアダプタＡは、ポインタ情報７００のリード開始アドレスと更新情報最新アドレスを比較する（ステップ１５５０）。

【００９７】

リード開始アドレスが更新情報最新アドレスと等しい場合は、全てのジャーナルを副記憶システム１００Ｂに送信済みであるため、副記憶システム１００Ｂに“ジャーナル無“を送信し（ステップ１５６０）、前回のジャーナルリード命令の時に、副記憶システム１００Ｂに送信したジャーナルの記憶領域を開放する（ステップ１５９０）。

【００９８】

ジャーナルの記憶領域の開放処理は、ポインタ情報７００の更新情報最古アドレスに、リトライ開始アドレスを設定する。更新情報最古アドレスがライトデータ領域先頭アドレスとなった場合は、更新情報最古アドレスは０とする。ポインタ情報７００のライトデータ最古アドレスは、前回のリードジャーナル命令に応じて送信したライトデータのサイズを

足した数値に変更する。ライトデータ最古アドレスが、ジャーナル論理ボリュームの容量以上の論理アドレスとなった場合は、ライトデータ領域先頭アドレスを減じ、補正する。

【００９９】

（７）ステップ１５５０の調べで、未送信のジャーナルが存在する場合、ホストアダプタＡは、キャッシュメモリ１３０を確保し、ディスクアダプタにポインタ情報７００のリード開始アドレスから、更新情報のサイズの情報をキャッシュメモリに読み込むことを命令する（図１４の１４２０）。

【０１００】

ディスクアダプタＡのリードライト処理は、記憶装置１５０から更新情報を読み込み、キャッシュメモリ１３０に保存し、ホストアダプタに通知する（図１４の１４３０）。

【０１０１】

ホストアダプタＡは、更新情報のリード終了の通知を受け、更新情報から、ライトデータの論理アドレスおよびライトデータのサイズを取得し、キャッシュメモリ１３０を確保し、ディスクアダプタＡにライトデータをキャッシュメモリに読み込むことを命令する（ステップ１５７０、図１４の１４４０）。

【０１０２】

ディスクアダプタＡのリードライト処理は、記憶装置１５０からライトデータを読み込み、キャッシュメモリ１３０に保存し、ホストアダプタに通知する（図１４の１４５０）。

【０１０３】

ホストアダプタＡは、ライトデータのリード終了の通知を受け、更新情報とライトデータを副記憶システム１００Ｂに送信（ステップ１５８０）し、ジャーナルを保持しているキャッシュメモリ１３０を開放する（図１４の１４６０）。そして、ポインタ情報７００のリトライ開始アドレスにリード開始アドレスを設定し、リード開始アドレスに送信したジャーナルの更新情報サイズを足した数値を設定する。

【０１０４】

（８）ホストアダプタＡは、前回のジャーナルリード命令の処理時に、副記憶システム１００Ｂに送信したジャーナルの記憶領域を開放する（ステップ１５９０）。

【０１０５】

前述したジャーナルリード受信処理では、正記憶システム１００Ａは、ジャーナルを一つずつ副記憶システム１００Ｂに送信していたが、複数同時に副記憶システム１００Ｂに送信してもよい。１つのジャーナルリード命令で、送信するジャーナル数は、副記憶システム１００Ｂがジャーナルリード命令内に指定してもよいし、グループ登録の際等に、ユーザが正記憶システム１００Ａもしくは、副記憶システム１００Ｂに指定してもよい。さらに、正記憶システム１００Ａと副記憶システム１００Ｂの接続パス２００の転送能力もしくは、負荷等により、動的に１つのジャーナルリード命令で送信するジャーナル数を変更してもよい。また、ジャーナル数でなく、ジャーナルのライトデータのサイズを考慮し、ジャーナルの転送量を指定してもよい。

【０１０６】

前述したジャーナルリード受信処理では、ジャーナルを記憶装置１５０からキャッシュメモリ１３０に読み込んでいたが、キャッシュメモリ１３０に存在する場合は、当該処理は不要である。

【０１０７】

前述したジャーナルリード受信処理内のジャーナルの記憶領域の開放処理は、次のジャーナルリード命令の処理時としたが、副記憶システム１００Ｂにジャーナルを送信した直後に開放してもよい。また、副記憶システム１００Ｂが、ジャーナルリード命令内に開放してよい更新番号を設定し、正記憶システム１００Ａは、その指示に従って、ジャーナルの記憶領域を開放していもよい。

【０１０８】

図１６はジャーナルリード命令処理２４０を説明する図、図１７はフローチャート、図１８はジャーナル格納処理のフローチャートである。以下、これらを用いて、副記憶システ

ム１００Ｂのホストアダプタ Ｂが、正記憶システム１００Ａからジャーナルを読み出し、ジャーナル論理ボリュームに格納する動作について説明する。

【０１０９】

（１）副記憶システム１００Ｂ内のホストアダプタ Ｂは、ジャーナルを格納するキャッシュメモリ１３０を確保し、ジャーナルリード命令であることを示す識別子、命令対象の正記憶システム１００Ａのグループ番号、リトライ指示の有無を含むアクセス命令を正記憶システム１００Ａに送信する。以下、アクセス命令内のグループ番号をグループ番号Ａとする（ステップ１７００、図１６の１６１０）。

【０１１０】

（２）ホストアダプタ Ｂは、正記憶システム１００Ａの応答およびジャーナルを受信する（図１６の１６２０）。ホストアダプタ Ｂは応答を調べ、正記憶システム１００Ａからの応答が、“ジャーナル無”の場合は、正記憶システム１００Ａには、指定したグループのジャーナルが存在しないため、一定時間後、正記憶システム１００Ａにリードジャーナル命令を送信する（ステップ１７２０、１７２５）。

【０１１１】

（４）正記憶システム１００Ａの応答が、“グループ状態は障害”もしくは“グループ状態は未使用”の場合は、副記憶システム１００Ｂのグループ状態を受信した状態に変更し、ジャーナルリード処理を終了する（ステップ１７３０、１７３５）。

【０１１２】

（５）正記憶システム１００Ａの応答が、上記以外、つまり、正常終了の場合は、ジャーナル論理ボリュームのボリューム状態を調べる（ステップ１７４０）。ジャーナル論理ボリュームのボリューム状態が“異常”の場合は、ジャーナル論理ボリュームにジャーナルの格納が不可能なため、グループ状態を“異常”に変更し、処理を終了する（ステップ１７４５）。この場合、ジャーナル論理ボリュームを正常な論理ボリュームに変更する等を行い、グループの状態を正常に戻す。

【０１１３】

（６）ステップ１７４０の調べで、ジャーナル論理ボリュームのボリューム状態が“正常”の場合は、後述するジャーナル格納処理１８００を行う。ジャーナル格納処理１８００が正常に終了した場合は、次のジャーナルリード命令を送信する。もしくは、一定時間経過後、次のジャーナルリード命令を送信する（ステップ１７６０）。次のジャーナル命令を送信するタイミングは、一定の時間間隔で定期的に送信するものでもよく、受信したジャーナルの個数もしくは、接続線２００の通信量、副記憶システム１００Ｂが保持しているジャーナルの記憶容量、副記憶システム１００Ｂの負荷等によって決めてもよい。さらに、正記憶システム１００Ａが保持しているジャーナルの記憶容量、もしくは正記憶システム１００Ａのポインタ情報を副記憶システム１００Ｂから読み出し、その数値に基づいて決めてもよい。上記情報の転送は、専用のコマンドで行ってもよいし、ジャーナルリード命令の応答に含んでもよい。その後の処理は、ステップ１７１０以降と同じである。

【０１１４】

（７）ステップ１８００のジャーナル格納処理が正常に終了しない場合は、ジャーナル論理ボリュームの未使用領域が足りないため、受信したジャーナルを破棄し、一定時間後にリトライ指示のジャーナルリード命令を送信する（ステップ１７５５）。もしくは、ジャーナルをキャッシュメモリに保持しておき、一定時間後に、再度ジャーナル格納処理を行う。これは、後述するリストア処理２５０が行われることにより、一定時間後には、ジャーナル論理ボリュームに未使用領域が増える可能性があるためである。この方式の場合は、ジャーナルリード命令にリトライ指示の有無は不要である。

【０１１５】

次に、図１８に示すジャーナル格納処理１８００について説明する。

【０１１６】

（１）ホストアダプタ Ｂは、ジャーナルがジャーナル論理ボリュームに格納可能であるかを調べる。ポインタ情報７００を用い、更新情報領域に未使用領域の有無を調べる（ステ

ップ１８１０）。ポインタ情報７００の更新情報最新アドレスと更新情報最古アドレスが等しい場合は、更新情報領域に未使用領域が存在しないため、ジャーナル作成失敗として処理を終了する（ステップ１８２０）。

【０１１７】

（２）ステップ１８１０の調べで、更新情報領域に未使用領域が存在する場合は、ポインタ情報７００を用い、ライトデータ領域にライトデータが格納できるかを調べる（ステップ１８３０）。ライトデータ最新アドレスと受信したジャーナルのライトデータのデータ量の和が、ライトデータ最古アドレスと等しいもしくは大きい場合は、ライトデータ領域にライトデータを格納できないため、ジャーナル作成失敗として処理を終了する（ステップ１８２０）。

【０１１８】

（３）ジャーナルが格納可能である場合、ホストアダプタＢは、受信した更新情報のグループ番号とジャーナル論理ボリュームの論理アドレスを変更する。グループ番号は、副記憶システム１００Ｂのグループ番号に変更し、ジャーナル論理ボリュームの論理アドレスはポインタ情報７００のライトデータ最新アドレスに変更する。ホストアダプタＢは、ポインタ情報７００の更新情報最新アドレスを更新情報最新アドレスに更新情報のサイズを足した数値に変更する。ホストアダプタＢは、ポインタ情報７００のライトデータ最新アドレスを、ライトデータ最新アドレスにライトデータのサイズを足した数値に変更する（ステップ１８４０）。

【０１１９】

（４）ホストアダプタＢは、ディスクアダプタ１２０に、更新情報とライトデータを記憶装置１５０に書き込むことを命令し、ジャーナル作成成功として処理を終了する（ステップ１８５０、図１６の１６３０）。その後、ディスクアダプタ１２０は、リードライト処理により、記憶装置１５０に更新情報とライトデータを書き込み、キャッシュメモリ１３０を開放する（図１６の１６４０）。

【０１２０】

前述したジャーナル格納処理では、ジャーナルを記憶装置１５０に保存するとしていたが、ジャーナル用に予め一定量のキャッシュメモリ１３０を用意しておき、当該キャッシュメモリを全て使用してから、記憶装置１５０にジャーナルを保存してもよい。ジャーナル用のキャッシュメモリ量は、例えば、保守端末から指定する。

【０１２１】

図１９はリストア処理２５０を説明する図、図２０はフローチャートである。以下、これらを用いて、副記憶システム１００Ｂのホストアダプタが、ジャーナルを利用し、データの更新を行う動作について説明する。リストア処理２５０は副記憶システム１００Ｂのディスクアダプタ１２０が処理を行ってもよい。

【０１２２】

（１）ホストアダプタＢは、グループ番号Ｂのグループ状態が“正常”であるかを調べる（ステップ２０１０）。ステップ２０１０の調べで、グループ状態が“正常”以外、例えば、“障害”の場合は、リストア処理を終了する（ステップ２０１５）。

【０１２３】

（２）ステップ２０１０の調べで、グループ状態が“正常”の場合は、ジャーナル論理ボリュームのボリューム状態を調べる（ステップ２０２０）。ステップ２０２０の調べで、ジャーナル論理ボリュームのボリューム状態が、“異常”の場合は、アクセス不可能なため、グループ状態を“異常”に変更し、処理を終了する（ステップ２０２５）。

【０１２４】

（３）ステップ２０２０の調べで、ジャーナル論理ボリュームのボリューム状態が、“正常”の場合は、リストア対象のジャーナルが存在するかを調べる。ホストアダプタＢは、ポインタ情報７００の更新情報最古アドレスと更新情報最新アドレスを取得する。更新情報最古アドレスと更新情報最新アドレスが等しい場合、ジャーナルは存在しないため、リストア処理は一旦終了し、一定時間後、リストア処理を再開する（ステップ２０３０）。

【０１２５】
（４）ステップ２０３０の調べで、リストア対象のジャーナルが存在する場合、最古（最小）の更新番号を持つジャーナルに対して次の処理を行う。最古（最小）の更新番号を持つジャーナルの更新情報は、ポインタ情報７００の更新情報最古アドレスから保存されている。ホストアダプタＢは、キャッシュメモリ１３０を確保し、ディスクアダプタに更新情報最古アドレスから、更新情報のサイズの情報をキャッシュメモリ１３０に読み込むことを命令する（図１９の１９１０）。
【０１２６】
ディスクアダプタのリードライト処理は、記憶装置１５０から更新情報を読み込み、キャッシュメモリ１３０に保存し、ホストアダプタＢに通知する（図１９の１９２０）。
【０１２７】
ホストアダプタＢは、更新情報のリード終了の通知を受け、更新情報から、ライトデータの論理アドレスおよびライトデータのサイズを取得し、キャッシュメモリ１３０を確保し、ディスクアダプタにライトデータをキャッシュメモリに読み込むことを命令する（図１９の１９３０）。
【０１２８】
ディスクアダプタのリードライト処理は、記憶装置１５０からライトデータを読み込み、キャッシュメモリ１３０に保存し、ホストアダプタに通知する（ステップ２０４０、図１９の１９４０）。
【０１２９】
（５）ホストアダプタＢは、更新情報から更新する副論理ボリュームの論理アドレスを求め、ディスクアダプタに副論理ボリュームにライトデータを書き込むことを命令する（ステップ２０５０、図１９の１９５０）。ディスクアダプタのリードライト処理は、副論理ボリュームの論理アドレスに対応する記憶装置１５０にデータを書き込み、キャッシュメモリ１３０を開放し、ホストアダプタに通知する（図１９の１９６０）。
【０１３０】
（６）ホストアダプタＢは、ディスクアダプタのライト処理完了の通知を受け、ジャーナルの記憶領域を開放する。ジャーナルの記憶領域の開放処理は、ポインタ情報７００の更新情報最古アドレスを更新情報のサイズを足した数値に変更する。更新情報最古アドレスが、ライトデータ領域先頭アドレスとなった場合は、ライトデータ領域先頭アドレスは０とする。ポインタ情報７００のライトデータ最古アドレスは、ライトデータのサイズを足した数値に変更する。ライトデータ最古アドレスが、ジャーナル論理ボリュームの容量以上の論理アドレスとなった場合は、ライトデータ領域先頭アドレスを減じ、補正する。その後、ホストアダプタＢは、次のリストア処理を開始する（ステップ２０６０）。
【０１３１】
前述したリストア処理２５０では、記憶装置１５０からキャッシュメモリ１３０にジャーナルを読み込んでいたが、キャッシュメモリ１３０に存在する場合は、当該処理は不要である。
【０１３２】
前述したジャーナルリード受信処理とジャーナルリード命令処理２４０では、正記憶システム１００Ａが送信するジャーナルをポインタ情報７００により決めていたが、副記憶システム１００Ｂが送信するジャーナルを決めてもよい。例えば、ジャーナルリード命令に更新番号を追加する。この場合、ジャーナルリード受信処理にて、副記憶システム１００Ｂが指定した更新番号の更新情報の論理アドレスを求めるために、正記憶システム１００Ａの共有メモリ１４０内に、更新番号から更新情報を格納した論理アドレスを求めるテーブルもしくは検索方法を設ける。
【０１３３】
前述したジャーナルリード受信処理とジャーナルリード命令処理２４０では、ジャーナルリード命令を用いていたが、通常のリード命令を用いてもよい。例えば、正記憶システム１００Ａのグループ情報６００とポインタ情報７００を予め副記憶システム１００Ｂに転

送しておき、副記憶システム１００Ｂは、正記憶システム１００Ａのジャーナル論理ボリュームのデータ（つまり、ジャーナル）をリードする。

【０１３４】

前述したジャーナルリード受信処理では、更新番号の順に、正記憶システム１００Ａから副記憶システム１００Ｂにジャーナルを送信すると説明したが、更新番号の順に送信せずともよい。また、正記憶システム１００Ａから副記憶システム１００Ｂに複数のジャーナルリード命令を送信してもよい。この場合、リストア処理にて更新番号順にジャーナルを処理するために、副記憶システム１００Ｂに、更新番号から更新情報を格納した論理アドレスを求めるテーブルもしくは検索方法を設ける。

【０１３５】

前述した本発明のデータ処理システムでは、正記憶システムがジャーナルを取得し、副記憶システムがデータの複製を行う。これにより、正記憶システムに接続したホストコンピュータは、データの複製に関する負荷を負わない。さらに、正記憶システムと副記憶システム間でジャーナルを転送することにより、正記憶システムと正記憶システムに接続したホストコンピュータの通信線を使用しない。

【０１３６】

図２３は、第二の実施形態の論理的な構成を示す図である。

【０１３７】

ホストコンピュータ１８０と記憶システム１００Ｃを接続パス１９０により接続し、記憶システム１００Ｃと正記憶システム１００Ａを接続パス２００により接続し、正記憶システム１００Ａと副記憶システム１００Ｂを接続パス２００により接続した構成である。記憶システム１００Ｃは、記憶システム１００Ｃの論理ボリューム（ＯＲＧ１）へのデータ更新時、論理ボリューム（ＯＲＧ１）のデータ更新と正記憶システム１００Ａ内の論理ボリューム（ＤＡＴＡ１）のデータ更新を行う。

【０１３８】

正記憶システム１００Ａは、第１の実施例で説明した通り、正論理ボリューム（ＤＡＴＡ１）へのデータ更新時、前述した命令受信処理２１０およびリードライト処理２２０によって、ジャーナル論理ボリューム（ＪＮＬ１）にジャーナルの保存を行う(２３１０)。

【０１３９】

副記憶システム１００Ｂは、前述したジャーナルリード処理２４０によって、正記憶システム１００Ａからジャーナルをリードし、リードライト処理２２０によって、ジャーナル論理ボリューム（ＪＮＬ２）にジャーナルを保存する(２３２０)。

【０１４０】

正記憶システム１００Ａは、副記憶システム１００Ｂからジャーナルをリードする命令を受信すると、命令受信処理２１０およびリードライト処理２２０によって、ジャーナル論理ボリューム（ＪＮＬ１）からジャーナルを読み出し、副記憶システム１００Ｂに送信する(２３２０)。

【０１４１】

副記憶システム１００Ｂは、前述したリストア処理２５０およびリードライト処理２２０によって、更新番号に従い、ジャーナル論理ボリューム（ＪＮＬ２）からジャーナルを読み出し、正論理ボリューム（ＤＡＴＡ１）の複製である副論理ボリューム（ＣＯＰＹ１）のデータを更新する(２３３０)。このように、更新番号の順にデータを更新することにより、論理ボリューム間のデータの整合性を保つことが可能となる。

【０１４２】

前述した本発明のデータ処理システムでは、正記憶システムがジャーナルを取得し、ジャーナル専用の記憶領域に格納する。さらに、副記憶システムは、正記憶システムから受信したジャーナルをジャーナル専用の記憶領域に格納する。ジャーナル専用の記憶領域はデータ複製対象の記憶領域より少なくすることが可能であり、より少ない記憶容量で、副記憶システムに正記憶システムのデータの複製が可能となる。

【０１４３】

図２４は、第三の実施形態の論理的な構成を示す図である。

【０１４４】

ホストコンピュータ１８０と記憶システム１００Ｃを接続パス１９０により接続し、記憶システム１００Ｃと正記憶システム１００Ａを接続パス２００により接続し、正記憶システム１００Ａと副記憶システム１００Ｂを接続パス２００により接続した構成である。記憶システム１００Ｃは、従来技術で説明した通り、記憶システム１００Ｃの論理ボリューム（ＯＲＧ１）へのデータ更新時、論理ボリューム（ＯＲＧ１）のデータ更新と正記憶システム１００Ａ内の論理ボリューム（ＤＡＴＡ１）のデータ更新を行う。

【０１４５】

正記憶システム１００Ａは、記憶システム１００Ｃに対し、正論理ボリューム（ＤＡＴＡ１）があるように見せるが、実際の記憶領域、つまり記憶装置１５０は割り当てない。例えば、ボリューム情報４００の物理アドレスに記憶装置１５０を割り当てていないことを示す数値を設定する。正記憶システム１００Ａは、記憶システム１００Ｃから正論理ボリューム（ＤＡＴＡ１）のデータへのライト命令受信時、前述した命令受信処理２１０内のステップ１２７０の処理を行わず、ジャーナル論理ボリューム（ＪＮＬ１）にジャーナルの保存のみを行う(２４１０)。

【０１４６】

副記憶システム１００Ｂは、前述したジャーナルリード処理２４０によって、正記憶システム１００Ａからジャーナルをリードし、リードライト処理２２０によって、ジャーナル論理ボリューム（ＪＮＬ２）にジャーナルを保存する(２４２０)。

【０１４７】

正記憶システム１００Ａは、副記憶システム１００Ｂからジャーナルをリードする命令を受信すると、命令受信処理２１０およびリードライト処理２２０によって、ジャーナル論理ボリューム（ＪＮＬ１）からジャーナルを読み出し、記憶システム１００Ｂに送信する(２４２０)。

【０１４８】

副記憶システム１００Ｂは、前述したリストア処理２５０およびリードライト処理２２０によって、更新番号に従い、ジャーナル論理ボリューム（ＪＮＬ２）からジャーナルを読み出し、論理ボリューム（ＯＲＧ１）の複製である副論理ボリューム（ＣＯＰＹ１）のデータを更新する(２４３０)。このように、更新番号の順にデータを更新することにより、論理ボリューム間のデータの整合性を保つことが可能となる。

【０１４９】

前述した本発明のデータ処理システムでは、記憶システム１００Ｃもしくは、記憶システム１００Ｃに接続したホストコンピュータ１８０に障害が生じた場合、副記憶システム１００Ｂの論理ボリューム（ＣＯＰＹ１）に対し、正記憶システム１００Ａ内のジャーナル（ＪＮＬ１）を反映することにより、記憶システム１００Ｂに接続したホストコンピュータにより最新データの参照、更新が可能となる。さらに、正記憶システム１００Ａにデータの複製を保持せず、ジャーナルのみを格納することで、データ複製に必要な記憶容量が少なくすることが可能となる。

【０１５０】

以上、本発明者によってなされた発明を実施例の形態に基づき具体的に説明したが、本発明は前記実施の形態に限定されるものでなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【０１５１】

【発明の効果】

本発明によれば、記憶システムの上位の計算機に影響を与えず、複数の記憶システム間でデータ転送又はデータの複製をすることが可能な記憶システムを提供できる。さらに、記憶システムと計算機との間の通信にも影響を与えない記憶システムを提供できる。

【０１５２】

さらに、本発明によれば、複数の記憶システム内に保持するデータ格納領域を少なくする

ことができる。さらに、複数の記憶システムの上位の計算機の業務に影響を与えることのないように、高速かつ効率的に複数の記憶システム間でデータ転送又はデータの複製をすることができる。

【図面の簡単な説明】

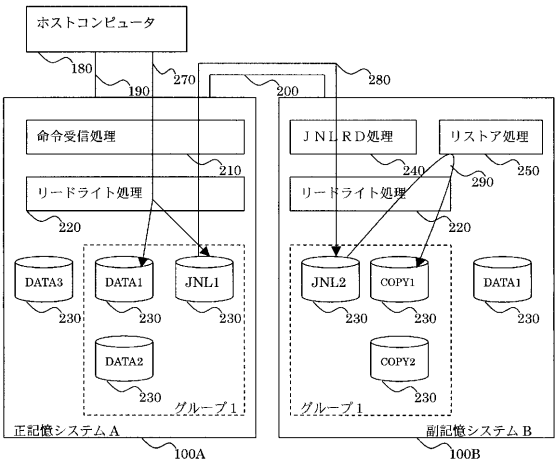
【図１】本発明の一実施形態の論理的な構成を示すブロック図である。
【図２】本発明の一実施形態の記憶システムのブロック図である。
【図３】本発明の一実施形態の更新情報とライトデータの関係を説明する図である。
【図４】本発明の一実施形態のボリューム情報の例を説明する図である。
【図５】本発明の一実施形態のペア情報の例を説明する図である。
【図６】本発明の一実施形態のグループ情報の例を説明する図である。
【図７】本発明の一実施形態のポインタ情報の例を説明する図である。
【図８】本発明の一実施形態のジャーナル論理ボリュームの構造を説明する図である。
【図９】本発明の一実施形態のデータの複製を開始する手順を説明するフローチャートである。
【図１０】本発明の一実施形態の初期コピー処理を説明するフローチャートである。
【図１１】本発明の一実施形態の命令受信処理を説明する図である。
【図１２】本発明の一実施形態の命令受信処理のフローチャートである。
【図１３】本発明の一実施形態のジャーナル作成処理のフローチャートである。
【図１４】本発明の一実施形態のジャーナルリード受信処理を説明する図である。
【図１５】本発明の一実施形態のジャーナルリード受信処理のフローチャートである。
【図１６】本発明の一実施形態のジャーナルリード命令処理を説明する図である。
【図１７】本発明の一実施形態のジャーナルリード命令処理のフローチャートである。
【図１８】本発明の一実施形態のジャーナル格納処理のフローチャートである。
【図１９】本発明の一実施形態のリストア処理を説明する図である。
【図２０】本発明の一実施形態のリストア処理のフローチャートである。
【図２１】本発明の一実施形態の更新情報の例を説明する図である。
【図２２】本発明の一実施形態のジャーナル作成処理時の更新情報の例を説明する図である。
【図２３】本発明の第２の実施形態を示す図である。
【図２４】本発明の第３の実施形態を示す図である。

【符号の説明】

１００　記憶システム
１１０　ホストアダプタ
１２０　ディスクアダプタ
１３０　キャッシュメモリ
１４０　管理メモリ
１５０　記憶装置
１６０　コモンパス
１７０　ディスクアダプタと記憶装置間の接続線
１８０　ホストコンピュータ

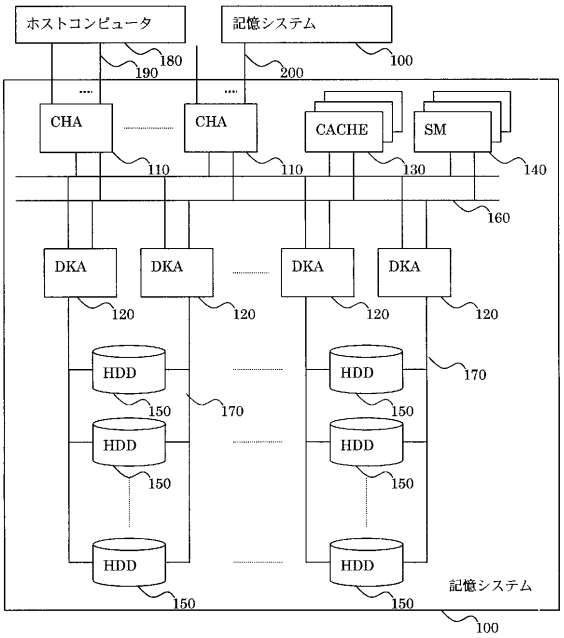
【図１】

図1



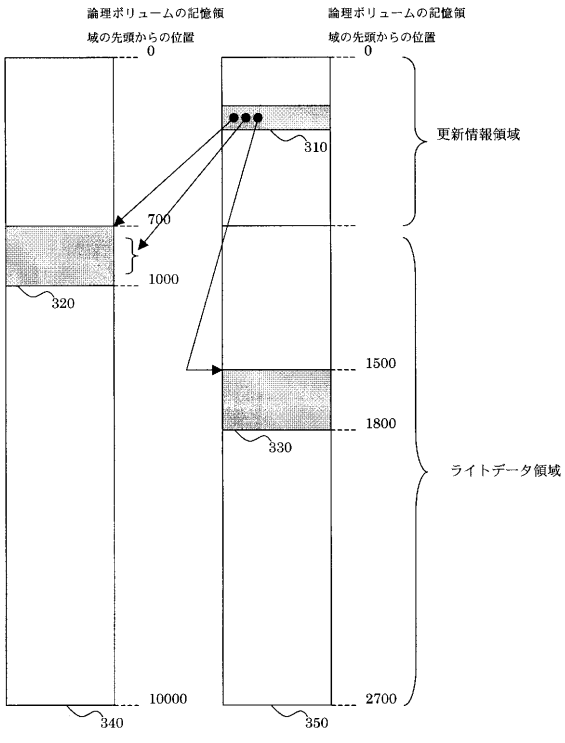
【図２】

図2



【図３】

図3



【図４】

図4

論理ボリューム番号	ボリューム状態	フォーマット形式	容量	ペア番号	物理アドレス	
					記憶装置番号	先頭から位置
1	正	OPEN3	3	1	1	0
2	正	OPEN6	6	2	1	3
3	未使用	OPEN6	6	0	1	9
4	正常	OPEN9	9	0	2	0
5	正常	OPEN3	3	0	2	9
6	未使用	OPEN6	6	0	2	12

400 ボリューム情報

【図５】

図5

ペア番号	ペア状態	正記憶システム番号	正論理ボリューム番号	副記憶システム番号	副論理ボリューム番号	グループ番号	コピー済みアドレス
1	正常	1	1	2	1	1	0
2	正常	1	2	2	3	1	0
3	未使用	0	0	0	0	0	0
4	未使用	0	0	0	0	0	0
5	未使用	0	0	0	0	0	0

500 ペア情報

【図６】

図6

グループ番号	グループ状態	ペア集合	ジャーナル論理ボリューム番号	更新番号
1	正常	1,2	4	1
2	未使用	0	0	0

600 グループ情報

【図７】

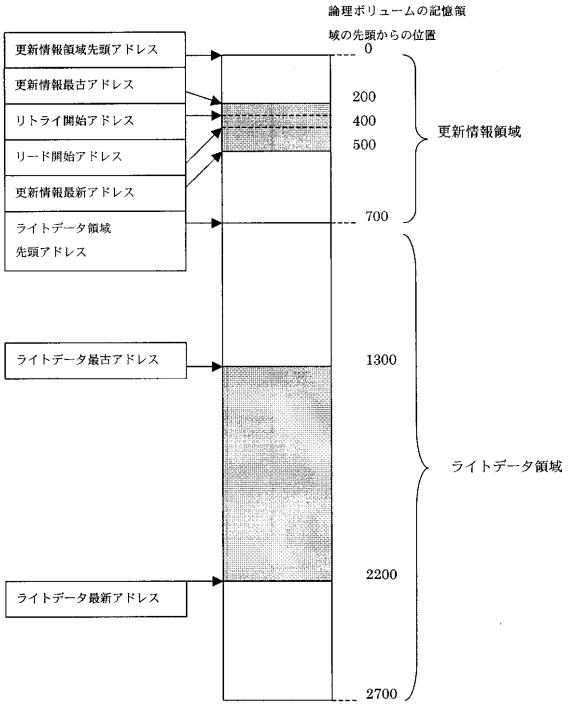
図7

	論理アドレス	
	論理ボリューム番号	論理ボリュームの記憶領域の先頭からの位置
更新情報領域先頭アドレス	4	0
ライトデータ領域先頭アドレス	4	700
更新情報最新アドレス	4	500
更新情報最古アドレス	4	200
ライトデータ最新アドレス	4	2200
ライトデータ最古アドレス	4	1300
リード開始アドレス	4	400
リトライ開始アドレス	4	300

700 ポインタ情報

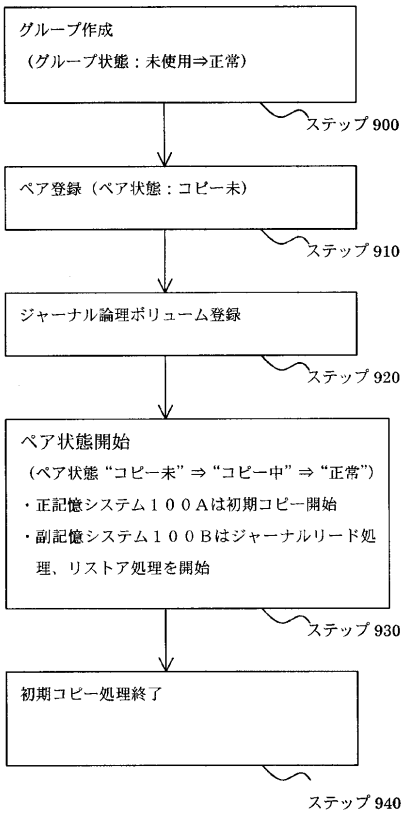
【図８】

図8



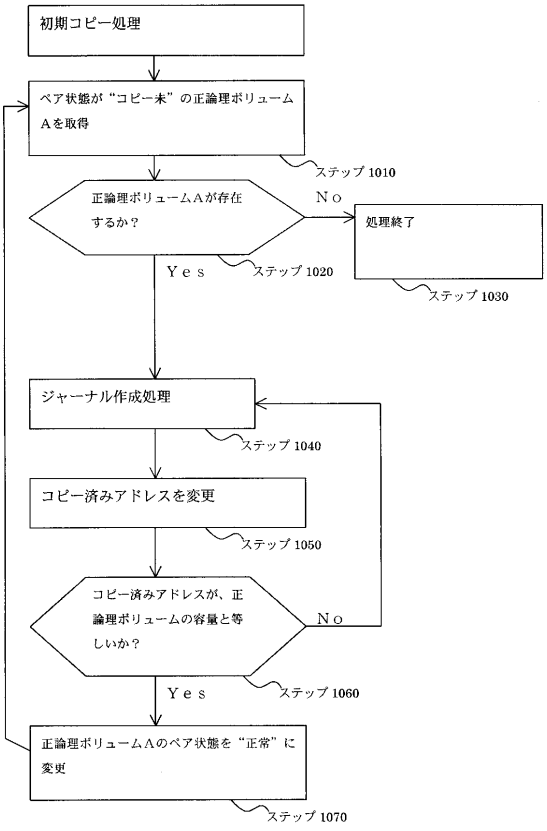
【図９】

図9



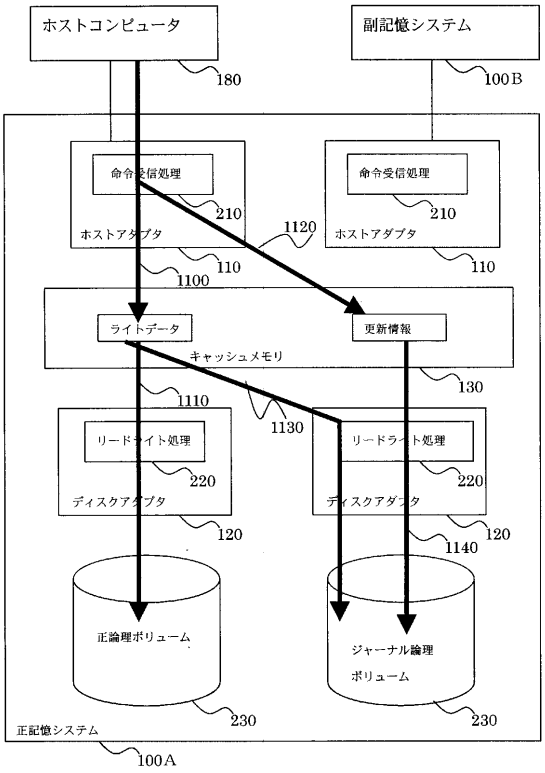
【図１０】

図１０



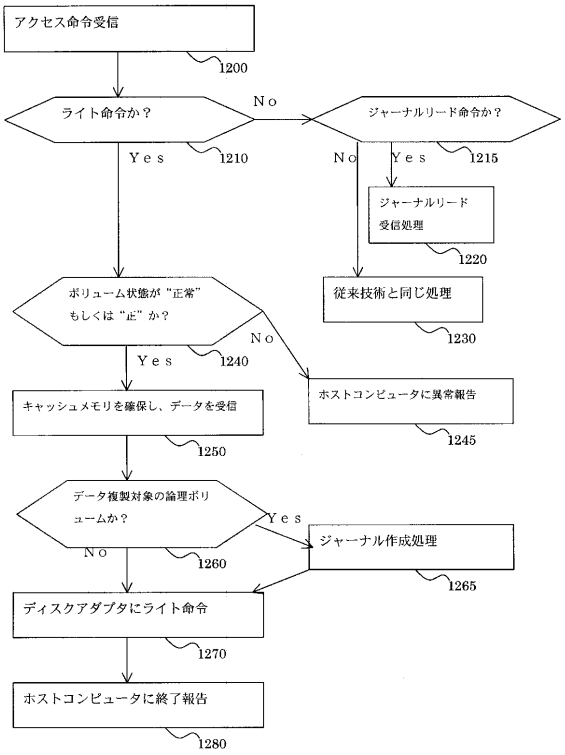
【図１１】

図１１



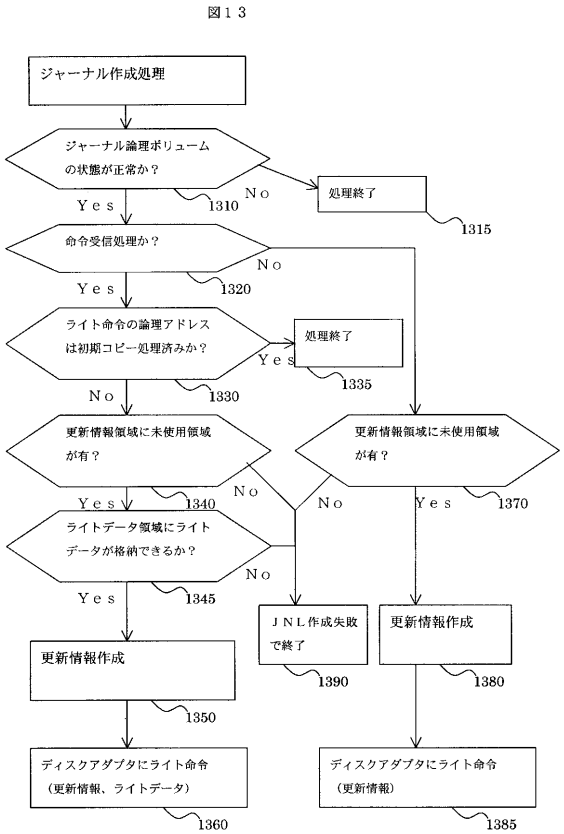
【図１２】

図１２



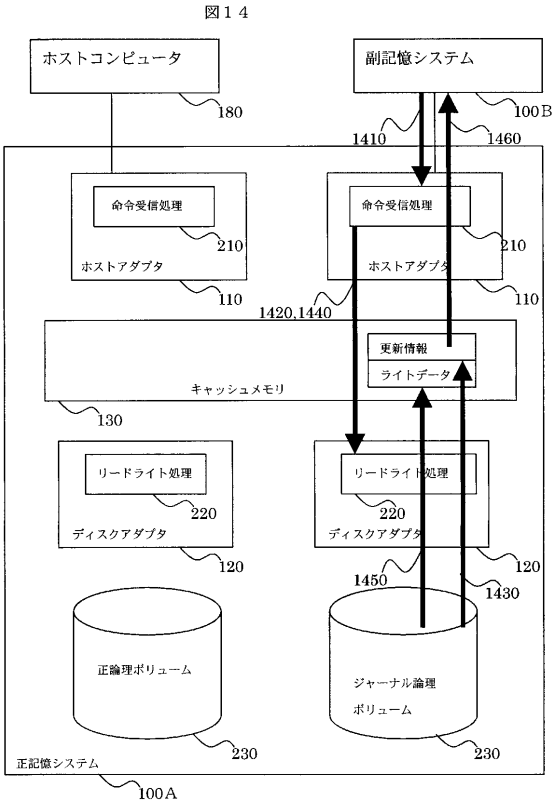
【図１３】

図１３



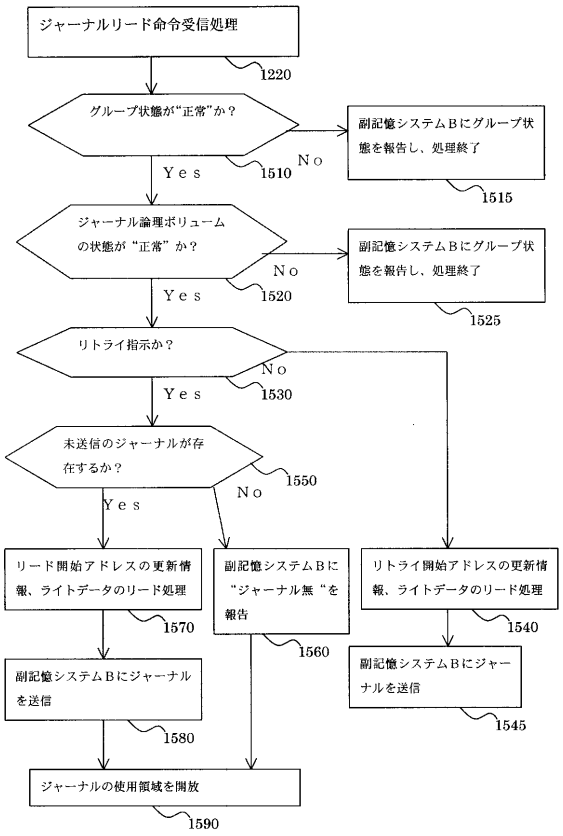
【図１４】

図１４



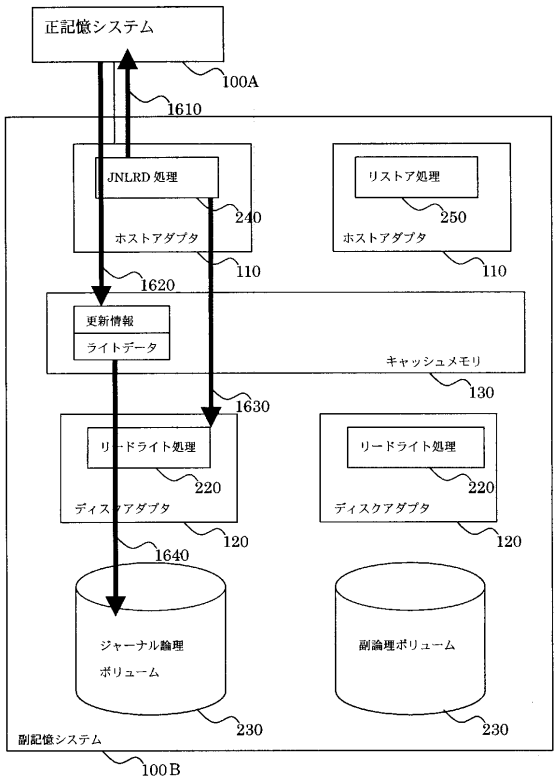
【図１５】

図１５



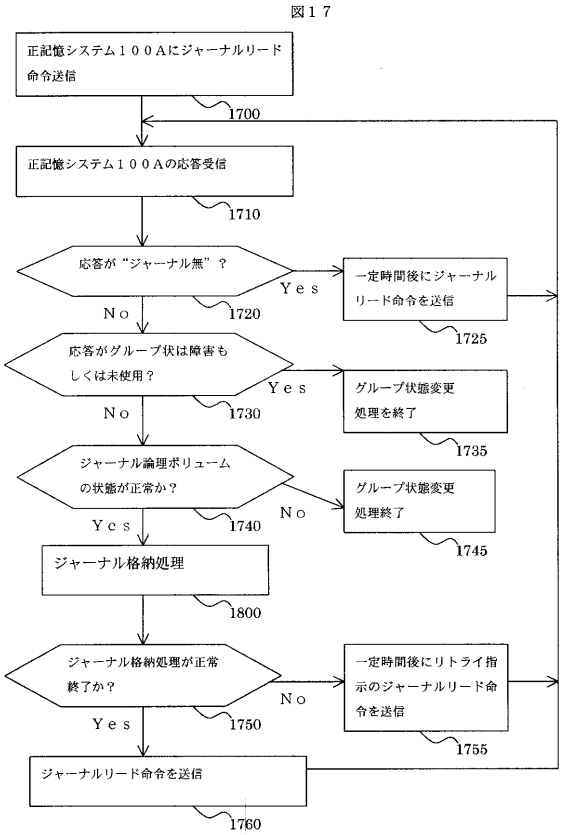
【図１６】

図１６



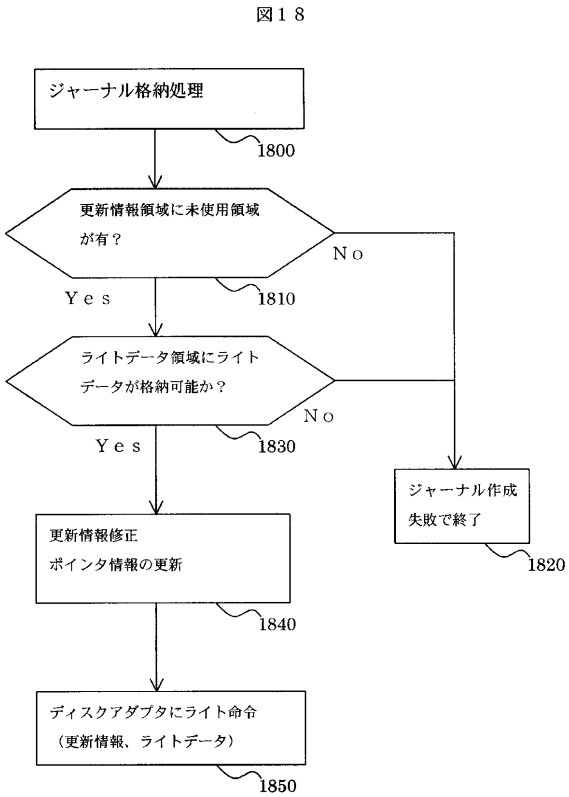
【図１７】

図１７



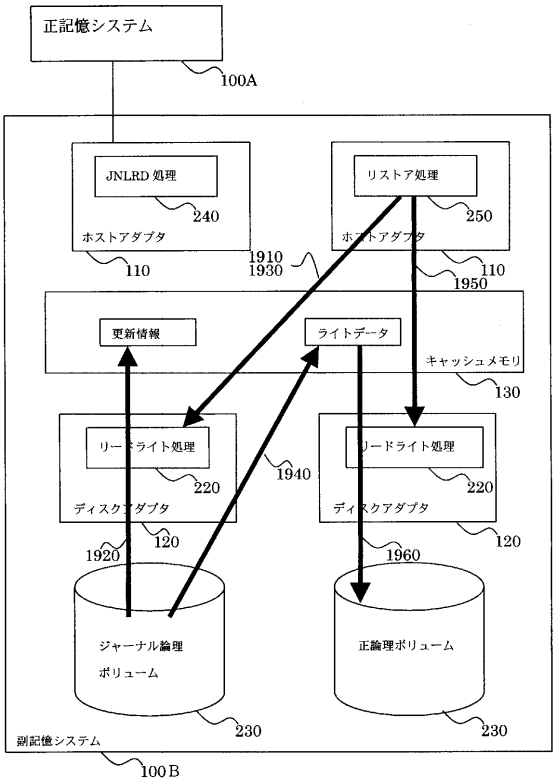
【図１８】

図１８



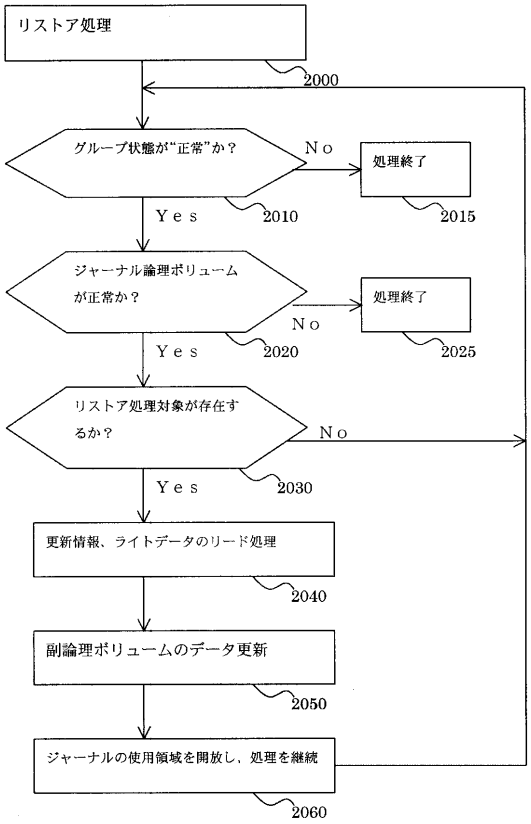
【図１９】

図１９



【図２０】

図２０



【図２１】

図２１

設定項目	設定値例
ライト命令を受信した時刻	1999/3/17 22:20:10
グループ番号	1
更新番号	4
ライト命令の論理アドレス	論理ボリューム番号：1 論理ボリュームの記憶領域の先頭からの位置：700
ライトデータのデータサイズ	300
ライトデータを格納したジャーナル論理ボリュームの論理アドレス	論理ボリューム番号：4 論理ボリュームの記憶領域の先頭からの位置：1500

310 更新情報

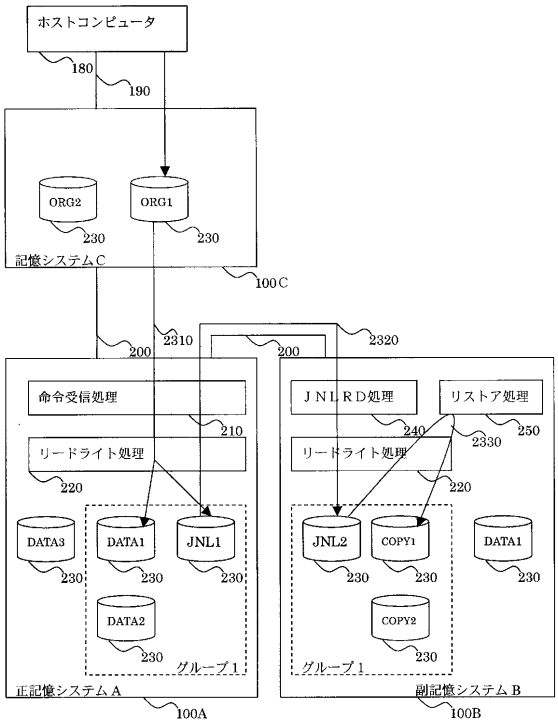
【図２２】

図２２

設定項目	設定値例
ライト命令を受信した時刻	1999/3/17 22:22:10
グループ番号	1
更新番号	4
ライト命令の論理アドレス	論理ボリューム番号：1 論理ボリュームの記憶領域の先頭からの位置：800
ライトデータのデータサイズ	100
ライトデータを格納したジャーナル論理ボリュームの論理アドレス	論理ボリューム番号：4 論理ボリュームの記憶領域の先頭からの位置：2200

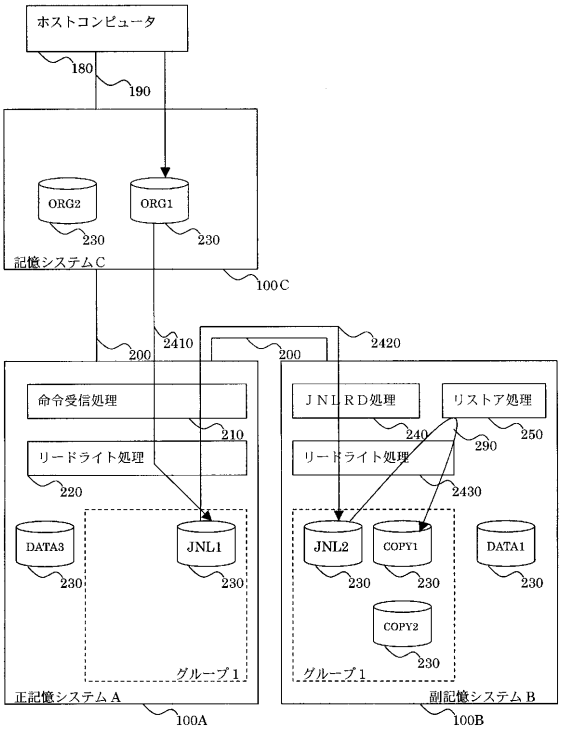
【図２３】

図２３



【図２４】

図２４


ホストコンピュータ
180
190
ORG2
ORG1
230
230
記憶システムC
100C
200
2410
200
2420
命令受信処理
210
リードライト処理
220
ＪＮＬＲＤ処理
240
リストア処理
250
290
リードライト処理
2430
DATA3
JNL1
JNL2
COPY1
DATA1
230
230
230
230
230
COPY2
230
グループ１
グループ１
正記憶システムA
副記憶システムB
100A
100B

フロントページの続き

(51)Int.Cl. FI
G06F 12/08 557

(72)発明者 武田 貴彦
神奈川県小田原市中里３２２番地２号 株式会社日立製作所 ＲＡＩＤシステム事業部内
(72)発明者 佐藤 孝夫
神奈川県小田原市中里３２２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２０００－１８１６３４（ＪＰ，Ａ）
国際公開第０２／０３１６９６（WO，Ａ１）
特表２００４－５１１８５４（ＪＰ，Ａ）
特開２００１－２８２６２８（ＪＰ，Ａ）
特開２００２－１８９５７０（ＪＰ，Ａ）
特開平１１－３０６０５８（ＪＰ，Ａ）
特開平０６－１９５２５０（ＪＰ，Ａ）
特開平０７－２４４５９７（ＪＰ，Ａ）
特開２００３－００６０１６（ＪＰ，Ａ）
特開２００３－０９９３０６（ＪＰ，Ａ）
特開昭５８－１７５０６６（ＪＰ，Ａ）
鈴木，中嶋，ストレージテクノロジー，ＦＵＪＩＴＳＵ，日本，富士通株式会社，１９９５年７月１０日，第46巻,第4号，p.389-397

(58)調査した分野(Int.Cl.，ＤＢ名)
G06F 12/00
G06F 3/06
G06F 12/08
JSTPlus(JDreamII)