

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4374953号
(P4374953)

(45) 発行日 平成21年12月2日(2009.12.2)

(24) 登録日 平成21年9月18日(2009.9.18)

(51) Int.Cl.

F I

G O 6 F 12/00 (2006.01)

G O 6 F 12/00 5 3 3 J

G O 6 F 3/06 (2006.01)

G O 6 F 12/00 5 3 1 J

G O 6 F 3/06 3 0 4 F

請求項の数 7 (全 39 頁)

(21) 出願番号 特願2003-316183 (P2003-316183)
 (22) 出願日 平成15年9月9日(2003.9.9)
 (65) 公開番号 特開2005-84953 (P2005-84953A)
 (43) 公開日 平成17年3月31日(2005.3.31)
 審査請求日 平成18年8月9日(2006.8.9)

(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 100100310
 弁理士 井上 学
 (72) 発明者 水主 和人
 神奈川県小田原市中里322番2号 株式
 会社日立製作所 R A I Dシステム事業部
 内
 (72) 発明者 平川 裕介
 神奈川県小田原市中里322番2号 株式
 会社日立製作所 R A I Dシステム事業部
 内

最終頁に続く

(54) 【発明の名称】 データ処理システム

(57) 【特許請求の範囲】

【請求項 1】

コンピュータから受信されるデータを格納するための正記憶領域としての第一の記憶領域と、当該データ及び当該データの受信順序に関する順序情報を格納する第二の記憶領域と、前記第一の正記憶領域及び前記第二の記憶領域への格納を制御する第一のアダプタ群と、を有する、第一の記憶システムと、

前記第一の記憶システムから受信されるデータを格納するための第一の副記憶領域としての第三の記憶領域と、当該データ及び当該データの順序情報を格納するための第四の記憶領域と、前記第三の記憶領域及び前記第四の記憶領域への格納を制御する第二のアダプタ群と、を有する、第二の記憶システムと、

前記第一の記憶システムから受信されるデータ及び当該データの順序情報を格納するための第五の記憶領域と、受信されるデータを前記第五の記憶領域に格納されている当該データの順序情報に基づいて格納するための第二の副記憶領域としての第六の記憶領域と、当該データ及び当該データの順序情報を送信するように要求する命令を前記第一の記憶システムに対して送信するように制御する第三のアダプタ群と、を有する、第三の記憶システムとを有し、

前記第一の記憶システムの前記第一のアダプタ群は、前記第一の記憶領域に対して書き込まれるデータが前記第一の記憶システムによって受信された場合に、当該データを前記第一の記憶領域に格納し、当該データを前記第二の記憶システムに送信すると共に前記第二の記憶領域に格納し、

10

20

前記第一の記憶システムの前記第一のアダプタ群は、前記第三の記憶システムからの前記命令に応じて、前記第二の記憶領域に格納されているデータ及び当該データの順序情報を、前記第一の記憶システムから前記第三の記憶システムへ送信するものであり、

前記第二の記憶システムの前記第二のアダプタ群は、前記第一の記憶システムから受信されるデータを前記第三の記憶領域へ格納し、前記第四の記憶領域内のより古いデータ及び当該古いデータの順序情報が格納された領域を開放した後に、当該データ及び当該データの順序情報を当該開放された領域に格納するものであり、

前記第三の記憶システムの前記第三のアダプタ群は、前記命令に応じて前記第一の記憶システムから受信されるデータ及び当該データの順序情報を前記第五の記憶領域に格納し、前記第五の記憶領域に格納されている当該データの順序情報に基づいて当該データを前記第六の記憶領域に格納するものであり、

10

前記第一の記憶システムに関する障害が発生した場合に、前記第三の記憶システムの前記第三のアダプタ群は、前記命令のあて先を、前記第一の記憶システムから前記第二の記憶システムへと変更するものであり、前記第二の記憶システムの前記第二のアダプタ群は、前記命令によって前記第三の記憶システムから要求されるデータ及び当該データの順序情報を前記第四の記憶領域に格納している場合には、前記第四の記憶領域に格納されている当該データ及び当該データの順序情報を前記第二の記憶システムから前記第三の記憶システムに送信するものである、ことを特徴とするリモートコピーシステム。

【請求項 2】

請求項 1 記載のリモートコピーシステムであって、

20

前記障害が発生し、前記命令によって前記第三の記憶システムから要求されるデータ及び当該データの順序情報を前記第四の記憶領域に格納していない場合には、前記第二の記憶システムの前記第二のアダプタ群は、前記第二の記憶システムの第三の記憶領域から前記第三の記憶システムの第六の記憶領域へ初期コピーを実施することを特徴とする、リモートコピーシステム。

【請求項 3】

請求項 1 記載のリモートコピーシステムであって、

前記障害が発生する前には、前記第一の記憶領域及び前記第六の記憶領域を有する非同期リモートコピーペアが構成されており、前記障害が発生した後には、前記第三の記憶領域及び前記第六の記憶領域を有する非同期リモートコピーペアが構成されることを特徴とするリモートコピーシステム。

30

【請求項 4】

請求項 1 記載のリモートコピーシステムであって、

前記第一の記憶システムに関する障害とは、前記コンピュータの障害であることを特徴とするリモートコピーシステム。

【請求項 5】

請求項 1 記載のリモートコピーシステムであって、

前記第一の記憶システムに関する障害とは、前記第一の記憶システムの障害であることを特徴とするリモートコピーシステム。

【請求項 6】

40

請求項 1 記載のリモートコピーシステムであって、

前記障害の後、前記第二の記憶システムはコンピュータからデータを受信すると、当該データを前記第三の記憶領域に格納し、当該データ及び当該データの順序情報を前記第四の記憶領域に格納し、

前記第三の記憶システムからの要求に応じて、前記第四の記憶領域に格納されているデータ及び当該データの順序情報を前記第三の記憶システムに送信することを特徴とするリモートコピーシステム。

【請求項 7】

請求項 6 記載のリモートコピーシステムであって、

前記障害の後、前記第二の記憶システムは更に、コンピュータから受信したデータと当

50

該データの順序情報とを前記第一の記憶システムに送信し、前記第一の記憶システムは前記第二の記憶システムから受信したデータを前記第一の記憶領域に格納し、前記第二の記憶システムから受信したデータ及び当該データの順序情報を前記第二の記憶領域に格納することを特徴とするリモートコピーシステム。

【発明の詳細な説明】

【技術分野】

【0001】

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複製および記憶システムの障害時のデータ複製処理の再開に関する。

【背景技術】

10

【0002】

近年、顧客に対して常に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】

従来技術では、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、第一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を保持する。第三の記憶システムは、第一の論理ボリュームの複製である第三の論理ボリュームを保持する。この特許文献において、第一の記憶システムは、第一の論理ボリュームの更新時に、第二の論理ボリュームへのデータ複製処理を行い、第三の論理ボリュームとのデータの差異を一定のデータサイズ毎に管理情報に記憶する。その後、当該管理情報を用いて、第三の論理ボリュームへのデータ複製処理を行う。

20

【0004】

30

【特許文献1】特開2003-122509号公報

【発明の開示】

【発明が解決しようとする課題】

【0005】

従来技術に開示された技術は、第一の論理ボリュームと第三の論理ボリュームのデータの差異を一定のデータサイズ毎に管理する。差異を管理する管理情報は、データ複製対象のデータ量に比例して増大するという課題がある。さらに、管理情報に基づいて、データ更新順と無関係な順序で、第三の論理ボリュームの更新を行うため、第三の論理ボリュームのデータの整合性は維持できない。

【0006】

40

したがって、本発明の目的は、データ処理システムを構成する第三の記憶システム内のデータの整合性を維持しつつ、第一の記憶システムから第三の記憶システムへのデータの複製処理を行うことである。さらに、本発明の目的は、データの複製に使用する管理情報を少なくすることである。

【0007】

さらに、本発明の目的は、データ処理システムを構成する第一の記憶システムが故障した場合に、第三の記憶システムのデータを最新とする間も第三の記憶システムのデータの整合性を維持することである。さらに、本発明の目的は、データを最新とするまでにかかる時間を短くすることである。

【課題を解決するための手段】

50

【 0 0 0 8 】

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、更新順序を示す更新番号等の更新情報とによって構成される。さらに、第三の記憶システムは、第一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、ジャーナル専用の記憶領域に格納する。第三の記憶システムは、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。

【 0 0 0 9 】

さらに、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、第一の記憶システムは、第二の記憶システムと第一の記憶システム間の通信線を介して、第一の記憶システムに格納されたデータの更新の際に、第二の記憶システムに格納されたデータの更新も行う。このデータ更新命令は、第一の記憶システムがジャーナルを作成するときを使用した更新番号もしくは更新時刻を含む。第二の記憶システムは、かかるデータ更新の際に、第一の記憶システムから受信した更新番号もしくは更新時刻を用いた更新情報を作成し、ジャーナルとしてジャーナル専用の記憶領域に格納する。

【 0 0 1 0 】

第三の記憶システムは、第一の記憶システムが故障した場合、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、第三の記憶システムが保持していないジャーナルのみを第二の記憶システムから取得し、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。

【 発明の効果 】

【 0 0 1 1 】

本発明によれば、複数の記憶システム間で、データの整合性を維持しつつ、データを複製する場合に必要な管理情報を減らすことができる。さらに、本発明によれば、データ処理システムを構成する記憶システムおよびホストコンピュータが故障した場合、データの整合性を維持しつつ、高速かつ効率的にデータの複製を継続することができる。

【 発明を実施するための最良の形態 】

【 0 0 1 2 】

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。

【 0 0 1 3 】

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。

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 を接続バス 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 と記憶システム 100C を接続バス 200 により接続した構成である。さらに、記憶システム 100B と記憶システム 100C は、接続バス 200 により接続されている。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 および副記憶システム 100C 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。

【 0 0 1 4 】

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

10

20

30

40

50

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。

10

【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。更新番号の代わりに更新時刻を用いてもよい。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。さらに、論理ボリューム（DATA1）の複製である論理ボリューム（data1）と論理ボリューム（DATA2）の複製である論理ボリューム（data2）は、副記憶システム100C内でグループ1を構成する。同様に、論理ボリューム（DATA1）の複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100B内でグループ1を構成する。

20

30

【0018】

データ複製対象である正論理ボリューム（DATA1、DATA2）のデータを更新する場合、副論理ボリューム（COPY1、COPY2）のデータを更新するために、正記憶システム100Aは、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のジャーナル論理ボリュームは、論理ボリューム（JNL1）である。

40

【0019】

同様に、副記憶システム100Cの副論理ボリューム（data1、data2）のデータを更新する場合、副記憶システム100Cは、後述するジャーナルを作成し、副記憶システム100C内のジャーナル論理ボリュームに保存する。図1では、グループ1のジャーナル論理ボリュームは、論理ボリューム（Jnl1）である。

【0020】

50

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 (COPY 1、COPY 2) 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 (COPY 1、COPY 2) 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格納することにより問題とならない。図 1 では、グループ 1 のジャーナル論理ボリュームは、論理ボリューム (JNL 2) である。

10

【0021】

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 (更新時刻)、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。

20

【0022】

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、後述するポイント情報 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。

30

40

【0023】

図 1 を用いて、正記憶システム 100A の正論理ボリューム (DATA 1) へのデータ更新を副記憶システム 100C の副論理ボリューム (data 1) および副記憶システム 100B の副論理ボリューム (COPY 1) に反映する動作について概説する。

【0024】

(1) 正記憶システム 100A は、ホストコンピュータ 180 から正論理ボリューム (DATA 1) 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 (DATA 1) 内のデータ更新と、ジャーナル論理ボリューム (JNL 1) へのジャーナルの保存と、副記憶システム 100C に副記憶システム 100C 内の副論理ボリューム (data 1) の対応するデー

50

タの更新（リモートライト命令）の命令を行う（図１の２７０）。

【００２５】

（２）副記憶システム１００Ｃは、正記憶システム１００Ａからのリモートライト命令を受け、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、副論理ボリューム（data1）内の対応するデータの更新と、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）。

【００２６】

（３）正記憶システム１００Ａは、リモートライト命令の応答を受けた後、ホストコンピュータにライト命令の終了を報告する。従って、正記憶システム１００Ａ内の正論理ボリューム（DATA1）と副記憶システム１００Ｃ内の副論理ボリューム（data1）のデータは完全に一致する。このようなデータ複製を同期のデータ複製とよぶ。

10

【００２７】

（４）副記憶システム１００Ｂは、後述するジャーナルリード処理２４０によって、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（図１の２８０）。

【００２８】

（５）正記憶システム１００Ａは、副記憶システム１００Ｂからジャーナルリード命令を受信すると、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、副記憶システム１００Ｂに送信する（図１の２８０）。

20

【００２９】

（６）副記憶システム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）。従って、正記憶システム１００Ａ内の正論理ボリューム（DATA1）と副記憶システム１００Ｂ内の副論理ボリューム（COPY1）のデータは、正論理ボリューム（DATA1）の更新の暫く後には、完全に一致する。このようなデータ複製を非同期のデータ複製とよぶ。

【００３０】

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つ以上のコモンバス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はコモンバス１６０により相互間が接続されている。コモンバス１６０は、コモンバス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。

30

【００３１】

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。

40

【００３２】

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

50

フォーマット形式、容量、同期ペア番号、非同期ペア番号、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、例えば共有メモリ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。同期ペア番号および非同期ペア番号は、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。保持するペア番号が存在しない場合は、無効値（例えば、“0”）を設定する。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。

【0033】

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、例えば共有メモリ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、接続バス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A内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B内の副論理ボリューム1であり、データ複製処理が行われていることを示す。

【0034】

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、複製種類、相手記憶システム番号、相手グループ番号を保持する。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、例えば共有メモリ140に保存する。グループ状態は、“正常”、“異常”、“未使用”、“停止”、“準備中”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。正記憶システムの場合、グループ状態の“停止”は、ジャーナル

10

20

30

40

50

の作成を一時的に行わないことを示す。これは、グループ状態が“正常”の場合に、ジャーナルの作成を一時的に停止したい場合に使用する。副記憶システムの場合、グループ状態の“停止”は、ジャーナルリード処理を一時的に行わないことを示す。これは、グループ状態が“正常”の場合に、正記憶システムからジャーナルをリードすることを一時的に停止したい場合に使用する。グループ状態が“準備中”の場合は、後述するデータ複製開始処理中であることを示す。正記憶システムの場合、ペア集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、ペア集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。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が存在しない場合は、無効値（例えば、“0”）を設定する。更新番号は、初期値は1であり、ジャーナルを作成した場合に変更する。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。複製種類は、“同期”、“非同期”のいずれかを保持する。複製種類が“同期”の場合、正論理ボリュームと副論理ボリュームの更新は同時に行われる。従って、正論理ボリュームのデータと副論理ボリュームのデータは完全に一致する。複製種類が“非同期”の場合、副論理ボリュームの更新は、正論理ボリュームの更新の後に行われる。従って、正論理ボリュームのデータと副論理ボリュームのデータは、一致していない（副論理ボリュームのデータは正論理ボリュームの古いデータ）場合があり、暫く後、副論理ボリュームのデータは正論理ボリュームのデータと完全に一致する。正記憶システムの場合、相手記憶システム番号は、グループに属するペアの副論理ボリュームを保持する副記憶システム番号を保持する。副記憶システムの場合、相手記憶システム番号は、グループに属するペアの正論理ボリュームを保持する正記憶システム番号を保持する。正記憶システムの場合、相手グループ番号は、グループに属するペアの副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、相手グループ番号は、グループに属するペアの正論理ボリュームが属するグループ番号を保持する。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（非同期）が行われていることを示す。

【0035】

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、ジャーナルの更新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。ジャーナルの更新情報は、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、次のジャーナルの更新情報は、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

10

20

30

40

50

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。

【0036】

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もしくは故障して使用できない場合に使用する。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、RAID方式でデータを保持しており、RAIDを構成する記憶装置150の一台が故障中の場合である。

10

【0037】

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。

20

【0038】

次に、正記憶システム100Aから副記憶システム100Bおよび副記憶システム100Cに対して、データ複製を開始する手順（データ複製開始処理）を図9、図10を用いて説明する。

【0039】

(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。同様に、副記憶システム100B（もしくは副記憶システム100C）のグループ番号Bを取得する。ユーザは、正記憶システム100Aに、保守端末あるいはホストコンピュータ180を使用して、グループ作成指示を行う。グループ作成指示は、指示対象グループ番号A、相手記憶システム番号B、相手グループ番号B、複製種類からなる。

30

【0040】

グループ作成指示を受けて、正記憶システム100Aは、グループ情報600を変更する。具体的には、正記憶システム100Aは、指示対象グループ番号Aのグループ情報600のグループ状態を“準備中”に、相手記憶システム番号を、指示された相手記憶システム番号Bに、相手グループ番号を、指示された相手グループ番号Bに、複製種別を指示された複製種別に変更する。グループ情報600の更新番号を1（初期値）に設定する。さらに、正記憶システム100Aは、相手記憶システム番号Bの記憶システムに対して、グループ作成指示を行う。グループ作成指示の指示対象グループ番号は相手グループ番号B、相手記憶システム番号は正記憶システム100Aの記憶システム番号、相手グループ番号は指示対象グループ番号A、複製種別は指示された複製種別とする。

40

【0041】

(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A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内の副論理ボリューム番号B

50

である。

【 0 0 4 2 】

前記ペア登録指示を受けて、正記憶システム 1 0 0 A は、ペア情報 5 0 0 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 1 0 0 A を示す正記憶システム番号 A 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 A を、副記憶システム番号にグループ情報 6 0 0 のグループ番号 A の相手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 B 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。指示されたグループ番号 A のグループ情報 6 0 0 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 A のボリューム状態を “ 正 ” に変更する。

10

【 0 0 4 3 】

正記憶システム 1 0 0 A は、指示されたグループ番号 A のグループ情報 6 0 0 の相手記憶システムに対して、正記憶システム 1 0 0 A を示す正記憶システム番号 A、グループ番号 A のグループ情報 6 0 0 の相手グループ番号 B、正論理ボリューム番号 A、および副論理ボリューム番号 B を通知し、ペア登録を命令する。副記憶システム 1 0 0 B は、ペア情報 5 0 0 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通知された正記憶システム番号 A を、正論理ボリューム番号に通知された正論理ボリューム番号 A を、副記憶システム番号に副記憶システム番号 B を、副論理ボリューム番号に通知された副論理ボリューム番号 B を、グループ番号に通知されたグループ番号 B を設定する。さらに、副記憶システム 1 0 0 B は、指示されたグループ番号 B のグループ情報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 B のボリューム状態を “ 副 ” に変更する。

20

【 0 0 4 4 】

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。

【 0 0 4 5 】

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。

【 0 0 4 6 】

(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 9 2 0) 。ユーザは、保守端末あるいはホストコンピュータ 1 8 0 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 1 0 0 A 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。

30

【 0 0 4 7 】

正記憶システム 1 0 0 A は、指示されたグループ番号のグループ情報 6 0 0 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。さらに、当該論理ボリュームのボリューム情報 4 0 0 のボリューム状態を “ 正常 ” に設定する。

【 0 0 4 8 】

同様に、ユーザは、保守端末あるいはホストコンピュータ 1 8 0 を使用して、副記憶システム 1 0 0 B のボリューム情報 4 0 0 を参照し、副記憶システム 1 0 0 B、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 1 0 0 A に行う。正記憶システム 1 0 0 A 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 1 0 0 B に転送する。副記憶システム 1 0 0 B は、指示されたグループ番号 B のグループ情報 6 0 0 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。さらに当該論理ボリュームのボリューム情報 4 0 0 のボリューム状態を “ 正常 ” に設定する。

40

【 0 0 4 9 】

ユーザは、副記憶システム 1 0 0 B の保守端末あるいは副記憶システム 1 0 0 B に接続したホストコンピュータ 1 8 0 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

50

システム 1 0 0 B に行ってもよい。副記憶システム 1 0 0 C に対しても同様に行う。

【 0 0 5 0 】

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 9 1 0 とステップ 9 2 0 の順は逆でもよい。

【 0 0 5 1 】

(4) データ複製処理開始について説明する (ステップ 9 3 0) 。ユーザは、保守端末あるいはホストコンピュータ 1 8 0 を使用して、データ複製処理を開始する複製種類が同期のグループ番号 C と複製種類が非同期のグループ番号 B を指定し、データ複製処理の開始を正記憶システム 1 0 0 A に指示する。正記憶システム 1 0 0 A は、グループ B に属する全てのペア情報 4 0 0 のコピー済みアドレスを 0 に設定する。

10

【 0 0 5 2 】

正記憶システム 1 0 0 A は、グループ番号 B のグループ情報 6 0 0 の相手記憶システム 1 0 0 B に対し、グループ番号 B のグループ情報 6 0 0 の相手グループ番号のグループ状態を“ 正常 ”に変更する事と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。正記憶システム 1 0 0 A は、グループ番号 C のグループ情報 6 0 0 の相手記憶システム 1 0 0 C に対し、グループ番号 C のグループ情報 6 0 0 の相手グループ番号のグループ状態を“ 正常 ”に変更する事を指示する。

【 0 0 5 3 】

正記憶システム 1 0 0 A は、グループ番号 C のグループ状態とグループ番号 B のグループ状態を“ 正常 ”に変更し、後述する初期コピー処理を開始する。

20

【 0 0 5 4 】

前記の説明では、同期のデータ複製処理と非同期のデータ複製処理の開始を同時に指示していたが、それぞれ個別に行ってもよい。

【 0 0 5 5 】

(5) 初期コピー終了について説明する (ステップ 9 4 0) 。

【 0 0 5 6 】

初期コピーが終了すると、正記憶システム 1 0 0 A は、初期コピー処理の終了を副記憶システム 1 0 0 B および副記憶システム 1 0 0 C に通知する。副記憶システム 1 0 0 B および副記憶システム 1 0 0 C は、グループ B およびグループ C 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。

30

【 0 0 5 7 】

図 1 0 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 5 0 0 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 1 0 0 A 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 1 2 0 が行ってもよい。

40

【 0 0 5 8 】

(1) 正記憶システム 1 0 0 A 内のホストアダプタ A は、処理対象の非同期複製のグループ B 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 1 0 1 0 、 1 0 2 0) 。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 1 0 3 0) 。

【 0 0 5 9 】

(2) ステップ 1 0 2 0 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、 1 M B) 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 1 0 4 0) 。

【 0 0 6 0 】

50

(3) ホストアダプタ A は、正論理ボリューム A の同期ペアである副論理ボリュームのデータを更新するために、同期ペアである副論理ボリュームを保持する副記憶システム C にリモートライト命令を送信する。リモートライト命令は、ライト命令、論理アドレス(論理ボリュームは同期ペア番号の副論理ボリューム C、論理ボリューム内位置はコピー済みアドレス)、データ量(単位サイズ)、ステップ 1040 で使用した更新番号を含んでいる。なお、更新番号の代わりにジャーナルを作成した時刻でもよい(ステップ 1045)。リモートライト命令を受信した場合の副記憶システム C の動作は後述する命令受信処理 210 にて説明する。

【0061】

(4) ホストアダプタ A は、リモートライト命令の応答を受信後、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)。

10

【0062】

(5) 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)。

【0063】

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。

20

【0064】

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、図 23 はリモートライト命令受信処理のフローチャート、図 24 はジャーナル複製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。

【0065】

(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)。

30

【0066】

(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、1228)。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、リモートライト命令の場合は、後述するリモートライト命令受信処理を行う(ステップ 2300)。それら以外の命令、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)。

【0067】

40

(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)。

【0068】

(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

50

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)。

【0069】

(5)ホストアダプタAは、論理ボリュームAのボリューム情報、ペア情報、グループ情報を参照し、論理ボリュームAが非同期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、論理ボリュームAのボリューム状態が“正”かつ、論理ボリュームAの非同期ペア番号のペアのペア状態が“正常”かつ、当該ペアが属するグループのグループ状態が“正常”である場合は、論理ボリュームAが非同期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)。

【0070】

(6)ホストアダプタAは、論理ボリュームAのボリューム情報、ペア情報、グループ情報を参照し、論理ボリュームAが同期複製対象かどうかを調べる(ステップ1267)。ステップ1267の調べで、論理ボリュームAのボリューム状態が“正”かつ、論理ボリュームAの同期ペア番号のペアのペア状態が“正常”かつ、当該ペアが属するグループのグループ状態が“正常”である場合は、論理ボリュームAが同期複製対象であるため、同期ペア番号のペアの副記憶システムCに、ホストコンピュータ180から受信したライトデータを保存するリモートライト命令を送信する(図11の1185)。リモートライト命令は、ライト命令、論理アドレス(論理ボリューム番号は、同期ペア番号のペアの副論理ボリュームC、論理ボリューム内位置は論理ボリューム内位置A)、データ量A、ステップ1265で使用した更新番号を含んでいる。なお、更新番号の代わりにホストコンピュータ180からライト命令を受信した時刻でもよい。ステップ1267の調べで、同期複製対象の論理ボリュームでない場合、もしくは、ステップ1268のジャーナル作成処理が成功しなかった場合は、更新番号に無効を示す数値“0”を設定する。

【0071】

(7)ステップ1267もしくは、ステップ1268のリモートライト命令の応答を受信した後、ホストアダプタAは、ディスクアダプタ120にライトデータを論理アドレスAに対応する記憶装置150の記憶領域に書き込むことを命令し(図11の1160)、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、前記記憶領域にライトデータを保存する(図11の1170)。

【0072】

次に、ジャーナル作成処理について説明する。

【0073】

(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。

【0074】

(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。

【0075】

(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

10

20

30

40

50

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 1335）。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 より大きい場合もしくは、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。

【0076】

（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かを調べる。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）。10

【0077】

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）。ライトデータ最古アドレスが、ライトデータ最新アドレスとライトデータ最新アドレスにデータ量 A を足した数値の範囲にある場合、当該ライトデータを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）。

【0078】

（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。20

【0079】

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 6、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。30

【0080】

（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）。40

【0081】

（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かを調べる。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。50

【 0 0 8 2 】

(8) ステップ 1 3 7 0 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、更新情報をキャッシュメモリ 1 3 0 内に作成する。更新情報のライト命令を受信した時刻は、更新番号を取得した時刻を設定する。グループ番号は、論理ボリュームの非同期ペア番号が属するグループ番号を設定する。更新番号は、グループ情報 6 0 0 から取得し、1 を足した数値をグループ情報 6 0 0 の更新番号に設定する。ライト命令の論理アドレスと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は、初期コピー処理対象の論理アドレス (ペア情報のコピー済みアドレス) を設定する。ライトデータのデータサイズは、初期コピー処理の単位サイズを設定する。更新情報を格納する論理アドレスは、ポインタ情報 7 0 0 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 7 0 0 の更新情報最新アドレスに設定する (ステップ 1 3 8 0 、図 1 1 の 1 1 2 0) 。

10

【 0 0 8 3 】

(9) ホストアダプタ A は、ディスクアダプタ 1 2 0 に、更新情報を記憶装置 1 5 0 に書き込むことを命令し、正常終了する (ステップ 1 3 8 5 、図 1 1 の 1 1 4 0 、 1 1 5 0) 。

【 0 0 8 4 】

上記説明では、更新情報をキャッシュメモリ 1 3 0 内に存在するように記載しているが、共有メモリ 1 4 0 内等に格納してもよい。

【 0 0 8 5 】

20

ライトデータの記憶装置 1 5 0 への書き込みは、非同期、つまり、ステップ 1 3 6 0 およびステップ 1 3 8 5 の直後でなくともよい。ただし、ホストコンピュータ 1 8 0 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 1 8 0 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 1 5 0 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 1 5 0 に書き込みを行ってもよい。

【 0 0 8 6 】

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 1 5 0 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 1 3 0 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 1 5 0 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。

30

【 0 0 8 7 】

次に、副記憶システム 1 0 0 C のホストアダプタ C が、正記憶システム 1 0 0 A からリモートライト命令を受信した場合の処理 (リモートライト命令受信処理) について説明する。リモートライト命令は、ライト命令、論理アドレス (副論理ボリューム C 、論理ボリューム内位置 A) 、データ量 A 、更新番号を含んでいる。

【 0 0 8 8 】

(1) 副記憶システム 1 0 0 C 内のホストアダプタ C は、論理ボリューム C のボリューム情報 4 0 0 を参照し、副論理ボリューム C のボリューム状態を調べる (ステップ 2 3 1 0) 。ステップ 2 3 1 0 の調べで、論理ボリューム C のボリューム状態が、“ 副 ” 以外の場合は、論理ボリューム C へのアクセスは不可能なため、正記憶システム 1 0 0 A に異常終了を報告する (ステップ 2 3 1 5) 。

40

【 0 0 8 9 】

(2) ステップ 2 3 1 0 の調べで、論理ボリューム C のボリューム状態が、“ 副 ” の場合は、ホストアダプタ C は、キャッシュメモリ 1 3 0 を確保し、正記憶システム 1 0 0 A にデータ受信の準備ができたことを通知する。正記憶システム 1 0 0 A は、その通知を受け、ライトデータを副記憶システム 1 0 0 C に送信する。ホストアダプタ C は、ライトデータを受信し、当該キャッシュメモリ 1 3 0 に保存する (ステップ 2 3 2 0) 。

50

【 0 0 9 0 】

(3) リモートライト命令に含まれる更新番号を調べ、無効値 “ 0 ” の場合は、正記憶システム 1 0 0 A にてジャーナルが作成されていないため、ステップ 2 4 0 0 のジャーナル複製処理を行わない (ステップ 2 3 3 0) 。

【 0 0 9 1 】

(4) リモートライト命令に含まれる更新番号を調べ、有効値 (“ 0 ” 以外) 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。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、ステップ 2 4 0 0 のジャーナル複製処理を行わない (ステップ 2 3 4 0) 。

【 0 0 9 2 】

(5) ステップ 2 3 4 0 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複製処理 2 4 0 0 を行う。

【 0 0 9 3 】

(6) ホストアダプタ C は、ディスクアダプタ 1 2 0 にライトデータをリモートライト命令の論理アドレスに対応する記憶装置 1 5 0 の記憶領域に書き込むことを命令し、正記憶システム A に終了報告する (ステップ 2 3 6 0 、 2 3 7 0) 。その後、当該ディスクアダプタ 1 2 0 は、リードライト処理により、前記憶領域にライトデータを保存する。

【 0 0 9 4 】

次に、ジャーナル複製処理 2 4 0 0 について説明する。

【 0 0 9 5 】

(1) ホストアダプタ C 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ことを調べる。ポインタ情報 7 0 0 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 2 4 1 0) 。ポインタ情報 7 0 0 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、最古のジャーナルの記憶領域を開放し、更新情報領域を確保する (ステップ 2 4 1 5) 。次に、ポインタ情報 7 0 0 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 2 4 2 0) 。ライトデータ最古アドレスが、ライトデータ最新アドレスとライトデータ最新アドレスにデータ量 A を足した数値の範囲にある場合、当該ライトデータをライトデータ領域に格納できないため、最古のジャーナルからジャーナルの記憶領域を開放し、ライトデータを格納できるようにする (ステップ 2 4 2 5) 。

【 0 0 9 6 】

(2) ホストアダプタ C は、更新情報をキャッシュメモリ 1 3 0 内に作成する。更新情報のライト命令を受信した時刻は、リモートライト命令内の更新時刻を設定する。グループ番号は、論理ボリューム C の同期ペア番号のペアが属するグループのグループ番号を設定する。更新番号は、リモートライト命令内の更新番号を設定する。ライト命令の論理アドレスは、リモートライト命令内の論理アドレスを設定する。ライトデータのデータサイズは、リモートライト命令内のデータ量 A を設定する。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は、ポインタ情報 7 0 0 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ライトデータのサイズを足した数値をポインタ情報 7 0 0 のライトデータ最新アドレスに設定する。更新情報を格納する論理アドレスは、ポインタ情報 7 0 0 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 7 0 0 の更新情報最新アドレスに設定する (ステップ 2 4 3 0) 。

【 0 0 9 7 】

(3) ホストアダプタ C は、ディスクアダプタ 1 2 0 に、更新情報とライトデータを記憶装置 1 5 0 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 2 4 4 0) 。その後、ディスクアダプタ 1 2 0 は、リードライト処理により、記憶装置 1 5 0 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 1 3 0 を開放する。

【 0 0 9 8 】

このように、副記憶システム C は、古いジャーナルの記憶領域を開放し、常に新しい複数のジャーナルを保持する。

10

20

30

40

50

【 0 0 9 9 】

リードライト処理 2 2 0 は、ディスクアダプタ 1 2 0 が、ホストアダプタ 1 1 0 もしくはディスクアダプタ 1 2 0 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 1 3 0 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 1 5 0 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 1 5 0 内の記憶領域から指定されたキャッシュメモリ 1 3 0 にデータを読み込む処理等である。

【 0 1 0 0 】

図 1 4 はジャーナルリード命令を受信した正記憶システム 1 0 0 A のホストアダプタ A の動作（ジャーナルリード受信処理）を説明する図、図 1 5 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 1 0 0 A が、副記憶システム 1 0 0 B 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。

10

【 0 1 0 1 】

(1) 正記憶システム 1 0 0 A 内のホストアダプタ A は、副記憶システム 1 0 0 B 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 1 2 2 0、図 1 4 の 1 4 1 0 ）。

【 0 1 0 2 】

(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 1 5 1 0 ）。ステップ 1 5 1 0 の調べで、グループ状態が“ 正常 ” 以外、例えば、“ 異常 ” の場合は、副記憶システム 1 0 0 B 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 1 0 0 B 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“ 異常 ” の場合は、ジャーナルリード処理を終了する（ステップ 1 5 1 5 ）。

20

【 0 1 0 3 】

(3) ステップ 1 5 1 0 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 1 5 2 0 ）。ステップ 1 5 2 0 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、“ 異常 ” の場合は、グループ状態を“ 異常 ” に変更し、副記憶システム 1 0 0 B 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 1 0 0 B 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“ 異常 ” の場合は、ジャーナルリード処理を終了する（ステップ 1 5 2 5 ）。

30

【 0 1 0 4 】

(4) ステップ 1 5 2 0 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 1 5 3 0 ）。

【 0 1 0 5 】

(5) ステップ 1 5 3 0 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 1 0 0 B に送信する。ホストアダプタ A は、キャッシュメモリ 1 3 0 を確保し、ディスクアダプタに、ポインタ情報 7 0 0 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 1 4 の 1 4 2 0 ）。

40

【 0 1 0 6 】

ディスクアダプタのリードライト処理は、記憶装置 1 5 0 から更新情報を読み込み、キャッシュメモリ 1 3 0 に保存し、ホストアダプタ A に通知する（図 1 4 の 1 4 3 0 ）。

【 0 1 0 7 】

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 1 3 0 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 1 5 4 0、図 1 4 の 1 4 4 0 ）。

【 0 1 0 8 】

ディスクアダプタのリードライト処理は、記憶装置 1 5 0 からライトデータを読み込み

50

、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）。

【0109】

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）。

【0110】

（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）。

10

【0111】

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）。

【0112】

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、前回のジャーナルリード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。

20

【0113】

（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）。

【0114】

30

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）。

【0115】

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）。

【0116】

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）。

40

【0117】

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。

【0118】

（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）。

50

【 0 1 1 9 】

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 1 0 0 A は、ジャーナルを一つずつ副記憶システム 1 0 0 B に送信していたが、複数同時に副記憶システム 1 0 0 B 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 1 0 0 B 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 1 0 0 A もしくは、副記憶システム 1 0 0 B に指定してもよい。さらに、正記憶システム 1 0 0 A と副記憶システム 1 0 0 B の接続パス 2 0 0 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。

10

【 0 1 2 0 】

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 1 5 0 からキャッシュメモリ 1 3 0 に読み込んでいたが、キャッシュメモリ 1 3 0 に存在する場合は、当該処理は不要である。

【 0 1 2 1 】

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 1 0 0 B 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 1 0 0 B 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 1 0 0 A は、その指示に従って、ジャーナルの記憶領域を開放してもよい。

20

【 0 1 2 2 】

図 1 6 はジャーナルリード処理 2 4 0 を説明する図、図 1 7 はフローチャート、図 1 8 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 1 0 0 B のホストアダプタ B が、正記憶システム 1 0 0 A 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。

【 0 1 2 3 】

(1) 副記憶システム 1 0 0 B 内のホストアダプタ B は、グループ状態が “ 正常 ” かつ複製種類が非同期であれば、ジャーナルを格納するキャッシュメモリ 1 3 0 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 1 0 0 A 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 1 0 0 A 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 1 7 0 0 、 図 1 6 の 1 6 1 0) 。

30

【 0 1 2 4 】

(2) ホストアダプタ B は、正記憶システム 1 0 0 A の応答およびジャーナルを受信する (図 1 6 の 1 6 2 0) 。ホストアダプタ B は応答を調べ、正記憶システム 1 0 0 A からの応答が、 “ ジャーナル無 ” の場合は、正記憶システム 1 0 0 A 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 1 0 0 A にジャーナルリード命令を送信する (ステップ 1 7 2 0 、 1 7 2 5) 。

【 0 1 2 5 】

(4) 正記憶システム 1 0 0 A の応答が、 “ グループ状態は異常 ” 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 1 0 0 B 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 1 7 3 0 、 1 7 3 5) 。

40

【 0 1 2 6 】

(5) 正記憶システム 1 0 0 A 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 1 7 4 0) 。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 1 7 4 5) 。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。

【 0 1 2 7 】

50

(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 1700)。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 1700 以降と同じである。

10

【0128】

(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。

【0129】

20

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。

(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ことを調べる。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)。

【0130】

(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)。ライトデータ最古アドレスが、ライトデータ最新アドレスとライトデータ最新アドレスにデータ量 A を足した数値の範囲にある場合、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)。

30

【0131】

(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。さらに、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、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。さらに、ホストアダプタ B は、グループ情報の更新番号を受信した更新情報の更新番号に変更する(ステップ 1840)。

40

【0132】

(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)。

【0133】

50

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。

【0134】

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 B 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。

【0135】

(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”もしくは“停止”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”および“停止”以外、例えば、“異常”の場合は、リストア処理を終了する(ステップ 2015)。

【0136】

(2) ステップ 2010 の調べで、グループ状態が“正常”もしくは“停止”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)。

【0137】

(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)。

【0138】

(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)。

【0139】

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)。

【0140】

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)。

【0141】

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 B に通知する(ステップ 2040、図 19 の 1940)。

【0142】

(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

10

20

30

40

50

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）。

【0143】

（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、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、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）。

10

【0144】

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。

【0145】

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。

20

【0146】

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。

【0147】

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。

30

【0148】

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 A は、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。記憶システム B は、記憶システム A が保持するデータの複製を保持しており、自律的に記憶システム A からジャーナルを取得し、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。これにより、記憶システム B は、データの整合性を維持しつつ、記憶システム A のデータの複製を保持できる。さらに、ジャーナルを管理する管理情報は、複製対象のデータ容量に依存しない。

40

【0149】

正記憶システム 100A が故障した場合に、ホストコンピュータ 180 にて行っている情報処理を、ホストコンピュータ 180C および記憶システム 100C を用いて再開し、記憶システム 100B でのデータ複製を再開する手順を図 25 に、論理的な構成を示すブロック図を図 42 に示す。ホストコンピュータ 180 とホストコンピュータ 180C は同じコンピュータであってもよい。

50

【 0 1 5 0 】

以下の説明において、正記憶システム 1 0 0 A が故障する前の正記憶システム 1 0 0 A のボリューム情報は図 4、ペア情報は図 5、グループ情報は図 6、ポインタ情報は図 7、ポインタ情報を説明する図は図 8 とする。正記憶システム 1 0 0 A が故障する前の副記憶システム 1 0 0 B（非同期複製）のボリューム情報は図 2 6、ペア情報は図 2 7、グループ情報は図 2 8、ポインタ情報は図 2 9、ポインタ情報を説明する図は図 3 0 とする。副記憶システム 1 0 0 B は、非同期のデータ複製を行っているため、正記憶システム 1 0 0 A が保持するジャーナル（更新番号 3 ～ 5）の全てを保持していない場合がある。本例では、副記憶システム 1 0 0 B は、更新番号 5 のジャーナルを保持していない。正記憶システム 1 0 0 A が故障する前の副記憶システム 1 0 0 C（同期複製）のボリューム情報は図 3 1、ペア情報は図 3 2、グループ情報は図 3 3、ポインタ情報は図 3 4、ポインタ情報を説明する図は図 3 5 とする。副記憶システム 1 0 0 C は、同期のデータ複製を行っているため、正記憶システム 1 0 0 A が保持する全てのジャーナル（更新番号 3 ～ 5）を保持する。

10

（ 1 ）正記憶システム 1 0 0 A に故障が発生し、正論理ボリューム（DATA 1、DATA 2 等）が使用不可能となる（ステップ 2 5 0 0）。

（ 2 ）ユーザは、記憶システム 1 0 0 C の保守端末もしくは、ホストコンピュータ 1 8 0 C を使用し、記憶システム 1 0 0 C に、非同期複製元変更を命令する。非同期複製元変更命令は、グループ単位で、非同期のデータ複製元（正論理ボリューム）を変更する命令であり、複製元情報（同期のデータ複製の副論理ボリューム（data 1、data 2）を保持する記憶システム番号 C とグループ番号 C）、複製先情報（非同期のデータ複製の副論理ボリューム（COPY 1、COPY 2）を保持する記憶システム番号 B とグループ番号 B）を含む（ステップ 2 5 1 0）。

20

（ 3 ）記憶システム 1 0 0 C は、非同期複製元変更命令を受信すると、記憶システム 1 0 0 B および記憶システム 1 0 0 C のボリューム情報、ペア情報、グループ情報を参照し、記憶システム 1 0 0 C 内のグループ状態が“未使用”のグループ番号 D を取得し、記憶システム 1 0 0 C 内のグループ C に属する論理ボリューム C（data 1、data 2）を正論理ボリューム、記憶システム 1 0 0 B 内のグループ B に属する論理ボリューム B（COPY 1、COPY 2）を副論理ボリュームとする非同期のデータ複製のペアとなるように、記憶システム 1 0 0 C のボリューム情報、ペア情報、グループ情報の変更を行う。ただし、論理ボリューム C と論理ボリューム B の組み合わせは、記憶システム 1 0 0 A 内の同一の論理ボリューム A とペアとなっていた論理ボリュームとする。さらに、記憶システム 1 0 0 C は、グループ C に属していたジャーナル論理ボリュームをグループ D にて継続して使用するようにグループ情報を変更する。具体的には、記憶システム 1 0 0 C は、グループ D の更新番号をグループ C の更新番号に、グループ D のジャーナル論理ボリューム番号をグループ C のジャーナル論理ボリューム番号に変更し、グループ D のポインタ情報の全ての項目を、グループ C のポインタ情報と同じにする。非同期複製元変更命令により、記憶システム 1 0 0 C は、図 3 2 に示す記憶システム 1 0 0 C のペア情報を図 3 9 に示すペア情報に、図 3 3 に示す記憶システム 1 0 0 C のグループ情報を図 4 0 に示すグループ情報に、図 3 1 に示す記憶システム 1 0 0 C のボリューム情報を図 3 8 に示すボリューム情報に変更する。

30

40

【 0 1 5 1 】

記憶システム 1 0 0 C は、記憶システム 1 0 0 B に対して、記憶システム 1 0 0 C 内のグループ C に属する論理ボリューム C（data 1、data 2）を正論理ボリューム、記憶システム 1 0 0 B 内のグループ B に属する論理ボリューム B（COPY 1、COPY 2）を副論理ボリュームとする非同期のデータ複製のペアとなるように、記憶システム 1 0 0 B のペア情報、グループ情報を変更するよう命ずる。ただし、論理ボリューム C と論理ボリューム B の組み合わせは、記憶システム 1 0 0 A 内の同一の論理ボリューム A とペアとなっていた論理ボリュームとする。

【 0 1 5 2 】

50

記憶システム 100B は、記憶システム 100B および記憶システム 100C のボリューム情報、ペア情報、グループ情報を参照し、記憶システム 100B のペア情報、グループ情報を変更する。記憶システム 100B は、図 27 に示すグループ B のペア情報を図 36 に示すペア情報に、図 28 に示すグループ情報を図 37 に示すグループ情報に、グループ B のグループ情報の状態を“停止”に変更することで、記憶システム 100A へのジャーナルリード処理を停止する。(ステップ 2530、2540)。

(4) 記憶システム 100C は、非同期複製元変更命令の応答をホストコンピュータ 180 もしくは保守端末に行う。ユーザは、ホストコンピュータ 180C もしくは保守端末から非同期複製元変更の終了を認識し、記憶システム 100C の使用を開始する(ステップ 2550、2560)。

10

(5) 記憶システム 100B は、記憶システム 100C にジャーナルリード位置指定命令を送信する(ステップ 2570)。ジャーナルリード位置指定命令は、記憶システム 100C のグループ D のポインタ情報を変更し、記憶システム 100B からのジャーナルリード命令により送信されるジャーナルを指定する命令であり、相手グループ番号 D、更新番号 B を含む。相手グループ番号 D は、グループ番号 B の相手グループ番号を指定する。更新番号は、グループ番号 B のグループ情報の更新番号に 1 を足した数値を指定する。図 37 に示す例では、グループ番号 2、更新番号 5 を指定する。

(6) 記憶システム 100C は、ジャーナルリード位置指定命令を受信した場合、ポインタ情報 700 を参照し、更新番号 B のジャーナルを保持しているかを調べる。記憶システム 100C は、ポインタ情報の更新情報最古アドレスの更新情報を記憶装置 150 から読み出し、最古(最小)の更新番号 C を得る。

20

【0153】

更新番号 C がジャーナルリード位置指定命令の更新番号 B と等しい、もしくは小さい場合は、記憶システム 100C が更新番号 B のジャーナルを保持しているので、記憶システム 100B は、非同期のデータの複製を継続することができる。この場合、記憶システム 100C は、更新番号 B より以前のジャーナルの記憶領域を開放し、リード開始アドレスおよびリトライ開始アドレスを更新番号 B の更新情報を格納しているアドレスに変更し、“再開可能”を記憶システム 100B に返す。これにより、図 34 に示すポインタ情報は、図 41 に示すポインタ情報に変更される(ステップ 2580)。

【0154】

30

他方、更新番号 C がジャーナルリード位置指定命令の更新番号 B より大きい場合は、記憶システム 100C が記憶システム 100B に必要なジャーナルを保持していないため、記憶システム 100B において非同期のデータの複製を継続することができない。この場合は、図 9、図 10 を用いて説明した手順により、正記憶システム 100C から副記憶システム 100B に対して、データ複製を開始する必要がある。

(7) 記憶システム 100B は、“再開可能”の応答を受信した場合、グループ B のグループ情報の状態を“正常”に変更することで、記憶システム 100C に対し、ジャーナルリード処理を再開する(ステップ 2590)。

【0155】

40

記憶システム 100B は、ジャーナルリード位置指定命令を行わなくてもよい。この場合、記憶システム 100B は、ジャーナルリード処理を開始し、記憶システム 100C から最古のジャーナルを受信する。受信したジャーナルの更新番号 C が、グループ番号 B のグループ情報の更新番号に 1 を足した数値(更新番号 B)より大きい場合は、記憶システム 100C が記憶システム 100B に必要なジャーナルを保持していないため、データ複製処理を中断する。受信したジャーナルの更新番号 C が、更新番号 B より小さい場合は、すでに記憶システム 100B が当該ジャーナルを保持しているため、当該ジャーナルを破棄し、ジャーナルリード処理を継続する。受信したジャーナルの更新番号 C が、更新番号 B と等しい場合は、受信したジャーナルをジャーナル論理ボリュームに格納し、ジャーナルリード処理を継続する。

【0156】

50

図42を用いて、ホストコンピュータ180Cが記憶システム100Cの使用を開始した後の、正記憶システム100Cの正論理ボリューム(d a t a 1)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(C O P Y 1)に反映する動作について概説する。

【0157】

(1) 正記憶システム100Cは、ホストコンピュータ180Cから正論理ボリューム(d a t a 1)内のデータに対するライト命令を受信すると、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、正論理ボリューム(d a t a 1)内のデータ更新と、ジャーナル論理ボリューム(j n l 1)へのジャーナルの保存を行い、ホストコンピュータにライト命令の終了を報告する(図42の4200)。

10

【0158】

(2) 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100C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(J N L 2)にジャーナルを保存する(図42の4210)。

【0159】

(3) 正記憶システム100Cは、副記憶システム100Bからジャーナルリード命令を受信すると、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(j n l 1)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(図42の4210)。

【0160】

20

(4) 副記憶システム100Bは、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(J N L 2)からジャーナルを読み出し、副論理ボリューム(C O P Y 1)のデータを更新する(図42の4220)。従って、正記憶システム100C内の正論理ボリューム(d a t a 1)と副記憶システム100B内の副論理ボリューム(C O P Y 1)のデータは、正論理ボリュームの更新の暫く後には、完全に一致する。

【0161】

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Cが記憶システムAからの更新番号および更新時刻を用いて、ジャーナルを作成する。データ複製対象の記憶システムAが故障し、記憶システムCを用いて情報処理を継続する場合に、記憶システムBは、記憶システムAから記憶システムCにジャーナルの取得先を変更する。これにより、記憶システムBは、データの整合性を維持しつつ、記憶システムAのデータの複製を継続できる。さらに、ジャーナルを管理する管理情報は、複製対象のデータ容量に依存しない。

30

【0162】

ホストコンピュータ180が故障した場合に、ホストコンピュータ180にて行っている情報処理を、ホストコンピュータ180Cおよび記憶システム100Cを用いて再開し、記憶システム100Bでのデータ複製を再開する手順について説明する。論理的な構成を示すブロック図を図48に示す。正記憶システム100Aが故障した場合との違いは、記憶システム100Aの使用が可能であるので、記憶システム100Bでの非同期のデータ複製に加え、同期のデータ複製を記憶システム100Aにて行う点である。

40

【0163】

以下の説明において、正記憶システム100Aが故障する前の正記憶システム100Aのボリューム情報は図4、ペア情報は図5、グループ情報は図6、ポインタ情報は図7、ポインタ情報を説明する図は図8とする。正記憶システム100Aが故障する前の副記憶システム100B(非同期複製)のボリューム情報は図26、ペア情報は図27、グループ情報は図28、ポインタ情報は図29、ポインタ情報を説明する図は図30とする。副記憶システム100Bは、非同期のデータ複製を行っているため、正記憶システム100Aが保持するジャーナル(更新番号3~5)の全てを保持していない場合がある。本例では、副記憶システム100Bは、更新番号5のジャーナルを保持していない。正記憶システム100Aが故障する前の副記憶システム100C(同期複製)のボリューム情報は図

50

31、ペア情報は図32、グループ情報は図33、ポインタ情報は図34、ポインタ情報を説明する図は図35とする。副記憶システム100Cは、同期のデータ複製を行っているため、正記憶システム100Aが保持する全てのジャーナル（更新番号3～5）を保持する。

（1）ホストコンピュータ180に故障が発生する。

（2）ユーザは、記憶システム100Cの保守端末もしくは、ホストコンピュータ180Cを使用し、記憶システム100Cに、前述した非同期複製元変更と同期複製交換を命令する。同期複製交換命令は、グループ単位で、同期のデータ複製の正論理ボリュームと副論理ボリュームの関係を逆転する命令であり、複製元情報（同期複製の正論理ボリューム（DATA1、DATA2）を保持する記憶システム番号Aとグループ番号A）、複製先情報（同期複製の副論理ボリューム（COPY1、COPY2）を保持する記憶システム番号Cとグループ番号C）を含む。

（3）記憶システム100Cは、同期複製交換命令を受信すると、記憶システム100Cのボリューム情報、ペア情報、グループ情報を参照し、記憶システム100A内のグループAに属する論理ボリュームA（DATA1、DATA2）を副論理ボリューム、記憶システム100C内のグループCに属する論理ボリュームC（COPY1、COPY2）を正論理ボリュームとする同期のデータ複製のペアとなるように、記憶システム100Cのボリューム情報、ペア情報、グループ情報を変更する。ただし、論理ボリュームAと論理ボリュームCの組み合わせは、既に同期のデータ複製のペアとなっていた論理ボリュームとする。

【0164】

記憶システム100Cは、記憶システム100Aに対して、記憶システム100A内のグループAに属する論理ボリュームA（DATA1、DATA2）を副論理ボリューム、記憶システム100C内のグループCに属する論理ボリュームC（COPY1、COPY2）を正論理ボリュームとする同期の複製のペアとなるように、記憶システム100Aのボリューム情報、ペア情報、グループ情報を変更するよう命ずる。ただし、論理ボリュームAと論理ボリュームCの組み合わせは、既に同期のデータ複製のペアとなっていた論理ボリュームとする。さらに、記憶システム100Cは、記憶システム100Aに同期のデータ複製のペアである副論理ボリュームへのデータ更新の際に、ジャーナルを作成することを命ずる。記憶システム100Aにて、ジャーナルを格納するジャーナル論理ボリュームは、例えば、ホストコンピュータの故障前に、記憶システム100A内の非同期のデータ複製の際に使用していたジャーナル論理ボリューム（JNL1）とする。

【0165】

記憶システム100Aは、記憶システム100Aのボリューム情報、ペア情報、グループ情報を参照し、記憶システム100Aのボリューム情報、ペア情報、グループ情報を変更する。記憶システム100Cは、記憶システム100Aと記憶システム100Cの情報の変更が終了後、同期複製交換命令の応答をホストコンピュータ180もしくは保守端末に返信する。

【0166】

非同期複製元変更命令と同期複製交換命令により、記憶システム100Cは、図32に示す記憶システム100Cのペア情報を図43に示すペア情報に、図33に示す記憶システム100Cのグループ情報を図44に示すグループ情報に、図31に示す記憶システム100Cのボリューム情報を、図38に示すボリューム情報に変更する。

【0167】

非同期複製元変更命令により、記憶システム100Bは、図27に示す記憶システム100Bのペア情報を図36に示すペア情報に、図28に示す記憶システム100Bのグループ情報を図37に示すグループ情報に変更する。

【0168】

同期複製交換命令により、記憶システム100Aは、図5に示す記憶システム100Cのペア情報を図46に示すペア情報に、図6に示す記憶システム100Cのグループ情報

10

20

30

40

50

を図４７に示すグループ情報に、図４に示す記憶システム１００Ｃのボリューム情報を図４５に示すボリューム情報に変更する。

【０１６９】

ユーザは、ホストコンピュータ１８０もしくは保守端末から非同期複製元変更と同期複製交換の終了を認識し、記憶システム１００Ｃの使用を開始する。以降の処理は、ステップ２５７０以降と同じである。

【０１７０】

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Ｃが記憶システムＡからの更新番号および更新時刻を用いて、ジャーナルを作成する。ホストコンピュータが故障し、別のホストコンピュータが記憶システムＣを用いて情報処理を継続する場合に、記憶システムＢは、記憶システムＡから記憶システムＣにジャーナルの取得先を変更する。これにより、記憶システムＢは、データの整合性を維持しつつ、記憶システムＡのデータの複製を継続できる。さらに、記憶システムＡが記憶システムＣの同期のデータの複製を保持するように設定を変更することで、ホストコンピュータが故障する前と同様に、同期と非同期の２つのデータ複製が可能となる。

【０１７１】

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。

【図面の簡単な説明】

【０１７２】

【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示す図である。

【図４】本発明の一実施形態のボリューム情報の例を示す図である。

【図５】本発明の一実施形態のペア情報の例を示す図である。

【図６】本発明の一実施形態のグループ情報の例を示す図である。

【図７】本発明の一実施形態のポイント情報の例を示す図である。

【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示す図である。

【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示す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示す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示す図である。

【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示す図である。

【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処理を示す図である。

【図１７】本発明の一実施形態のジャーナルリード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示す図である。

【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示す図である。

【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示す図である。

【図２３】本発明の一実施形態のリモートライト命令受信処理のフローチャートである。

【図２４】本発明の一実施形態のジャーナル複製処理のフローチャートである。

【図２５】本発明の一実施形態の正記憶システム１００Ａが故障した場合に記憶システム間のデータ複製を再開する手順を示すフローチャートである。

【図２６】本発明の一実施形態のボリューム情報の例を示す図である。

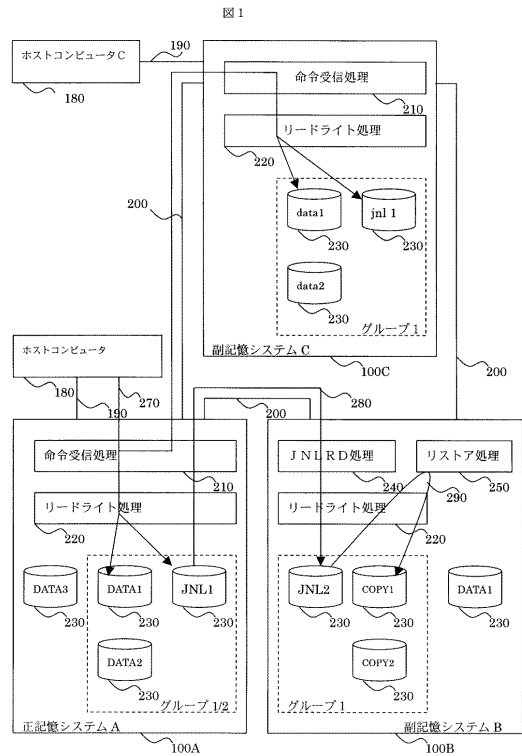
- 【図 27】本発明の一実施形態のペア情報の例を説明する図である。
- 【図 28】本発明の一実施形態のグループ情報の例を説明する図である。
- 【図 29】本発明の一実施形態のポインタ情報の例を説明する図である。
- 【図 30】本発明の一実施形態のジャーナル論理ボリュームの構造を説明する図である。
- 【図 31】本発明の一実施形態のボリューム情報の例を説明する図である。
- 【図 32】本発明の一実施形態のペア情報の例を説明する図である。
- 【図 33】本発明の一実施形態のグループ情報の例を説明する図である。
- 【図 34】本発明の一実施形態のポインタ情報の例を説明する図である。
- 【図 35】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 36】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。 10
- 【図 37】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 38】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 39】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 40】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 41】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 42】本発明の一実施形態の正記憶システム 100A が故障した場合の動作を説明するブロック図である。
- 【図 43】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 44】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 45】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。 20
- 【図 46】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 47】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 48】本発明の一実施形態のホストコンピュータ 180 が故障した場合の動作を説明するブロック図である。

【符号の説明】

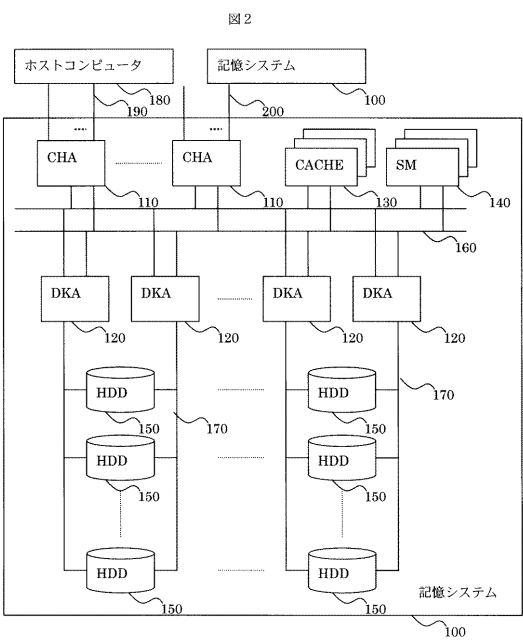
【0173】

- 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 30
- 140 共有メモリ
- 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

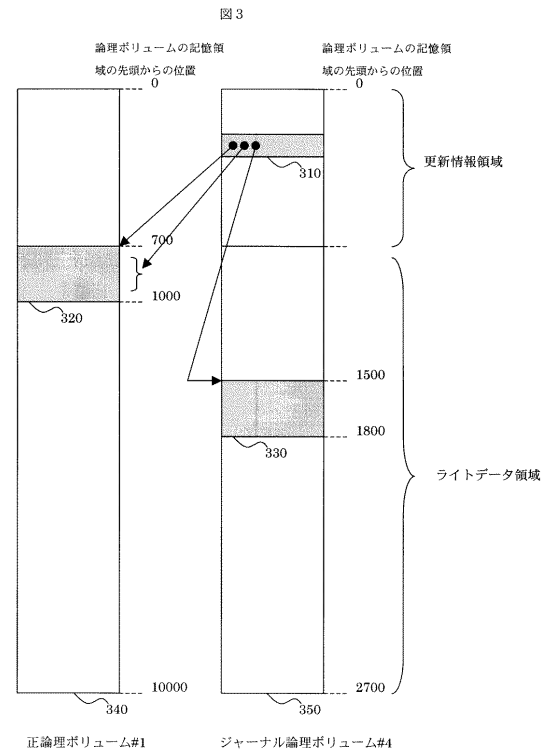
【図 1】



【図 2】



【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	同期ペア番号	非同期ペア番号	物理アドレス	
						記憶装置番号	先頭から位置
1	正	OPEN3	3	3	1	1	0
2	正	OPEN6	6	4	2	1	3
3	未使用	OPEN6	6	0	0	1	9
4	正常	OPEN9	9	0	0	2	0
5	正常	OPEN3	3	0	0	2	9
6	未使用	OPEN6	6	0	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	A	1	B	1	1	0
2	正常	A	2	B	2	1	0
3	正常	A	1	C	1	2	0
4	正常	A	2	C	2	2	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	複製種類	相手記憶システム番号	相手グループ番号
1	正常	1,2	4	5	非同期	B	1
2	正常	3,4	0	0	同期	C	1

600 グループ情報

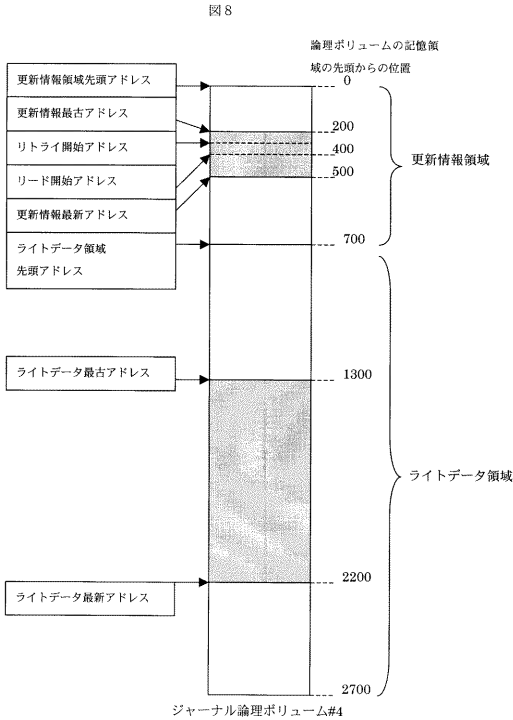
【図 7】

図 7

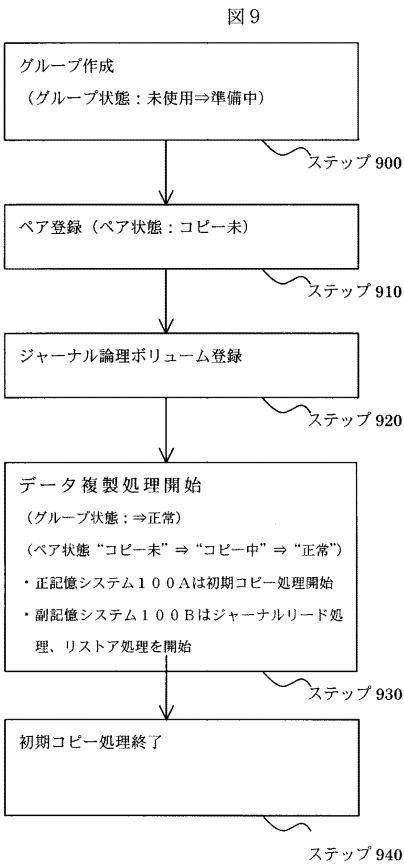
	論理アドレス	
	論理ボリューム番号	論理ボリュームの記憶領域の先頭からの位置
更新情報領域先頭アドレス	4	0
ライトデータ領域先頭アドレス	4	700
更新情報最新アドレス	4	500
更新情報最古アドレス	4	200
ライトデータ最新アドレス	4	2200
ライトデータ最古アドレス	4	1300
リード開始アドレス	4	400
リトライ開始アドレス	4	300

700 ポインタ情報

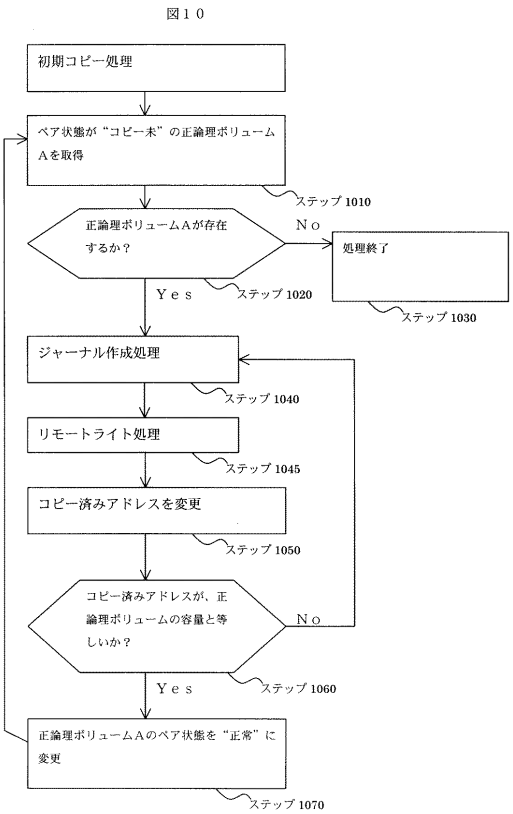
【図 8】



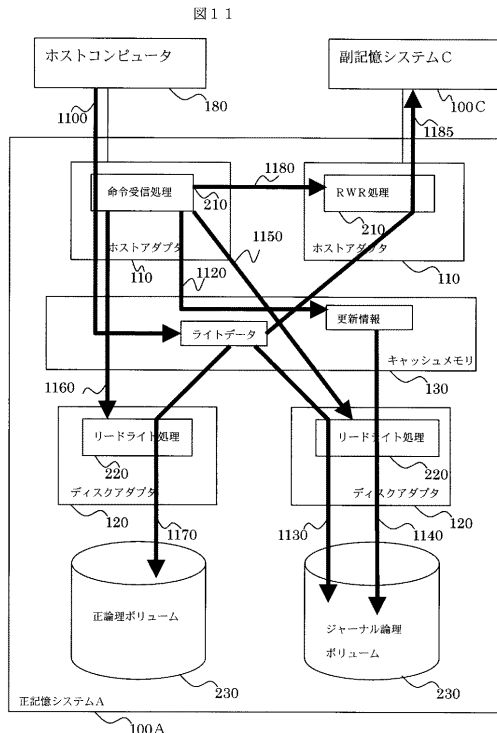
【図 9】



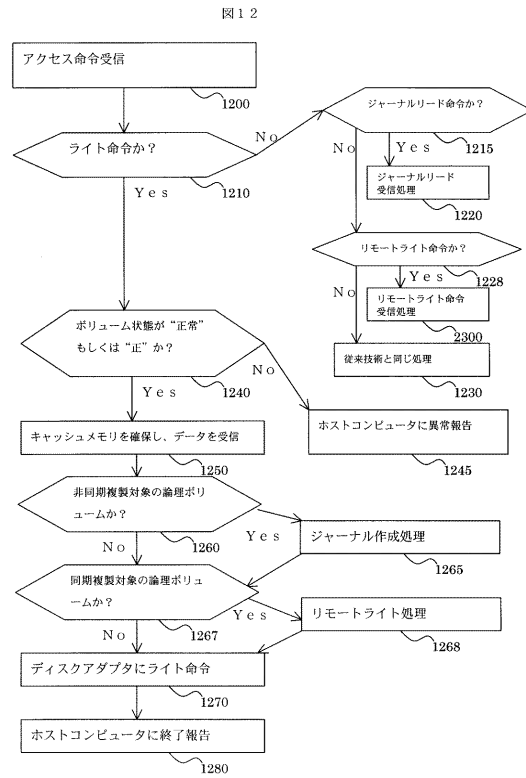
【図 10】



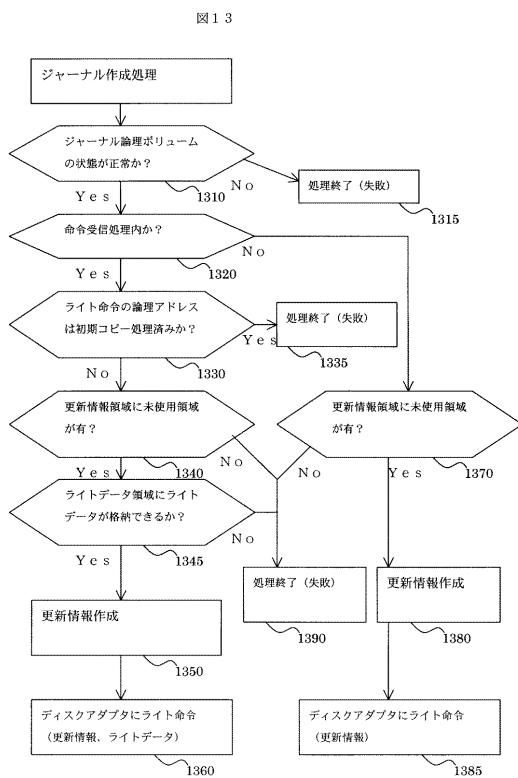
【図 1 1】



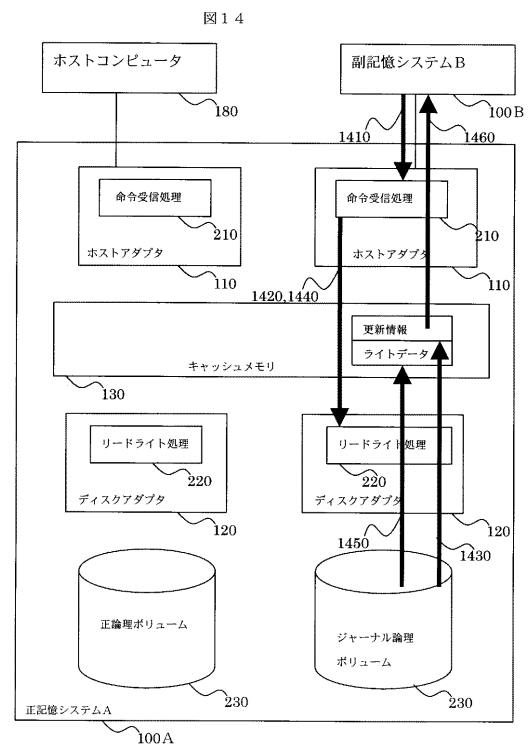
【図 1 2】



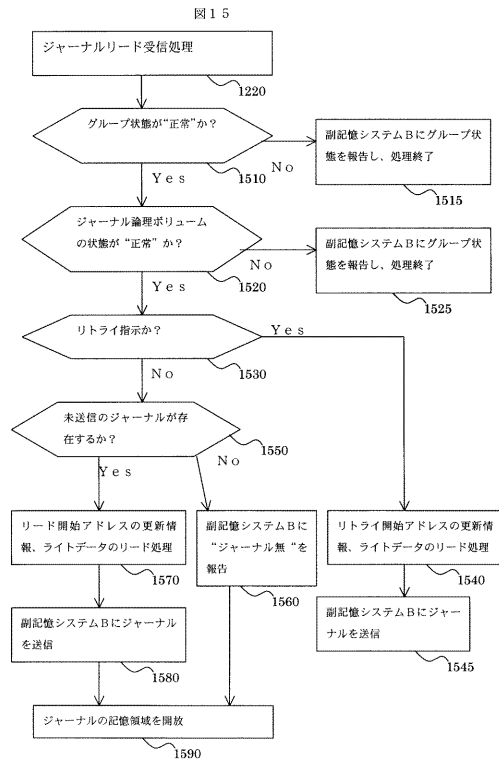
【図 1 3】



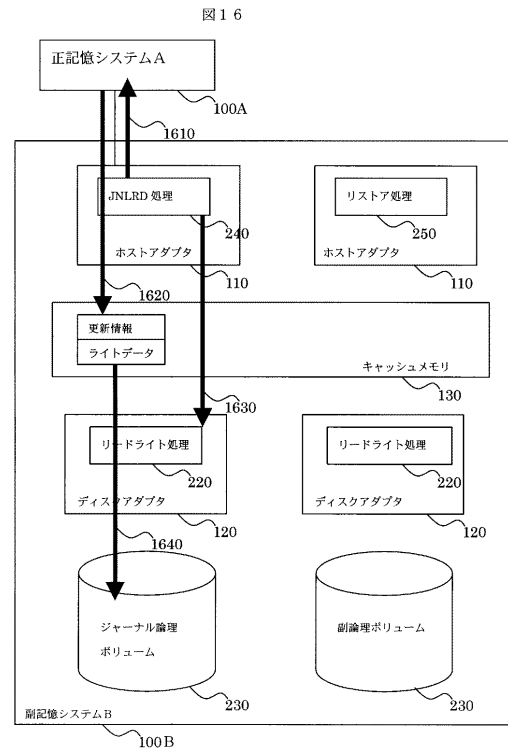
【図 1 4】



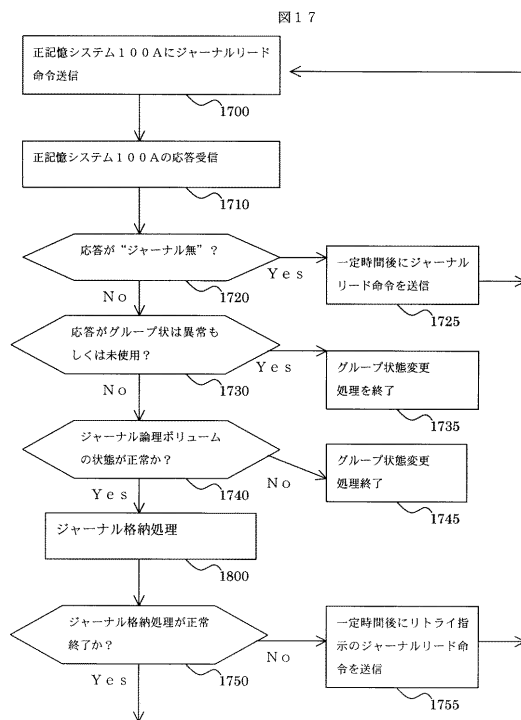
【 図 1 5 】



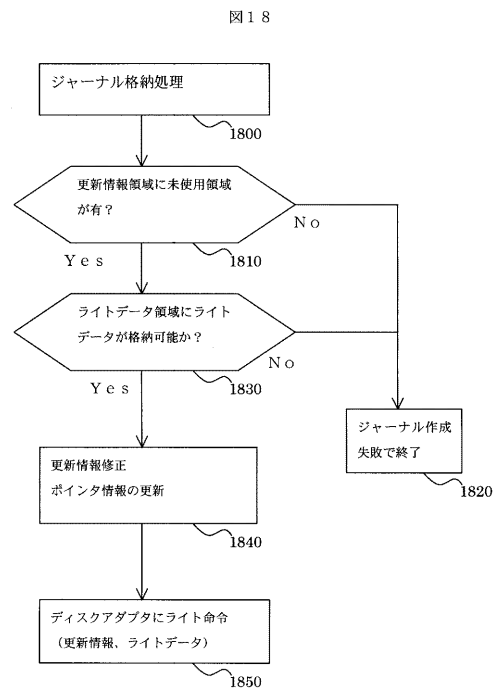
【 図 1 6 】



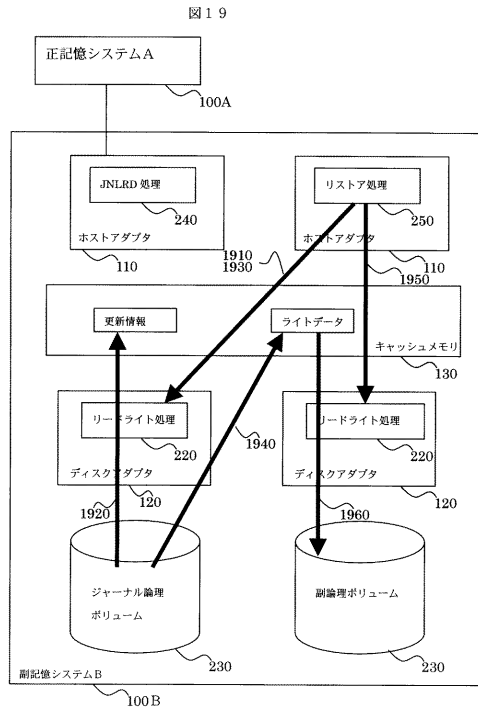
【 図 1 7 】



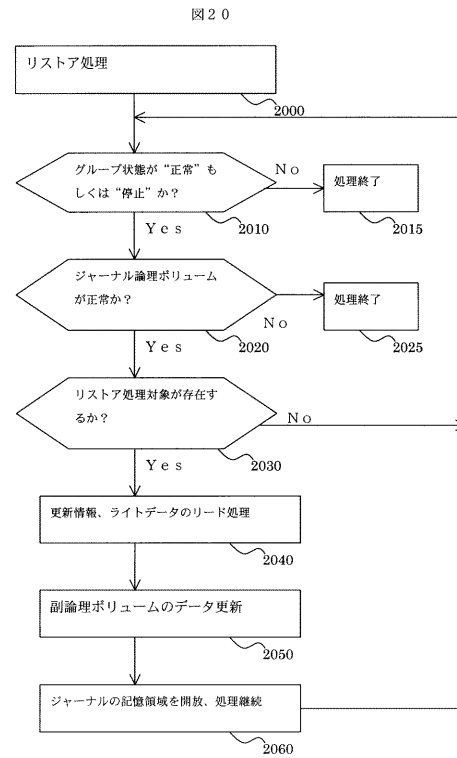
【 図 1 8 】



【図 19】



【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

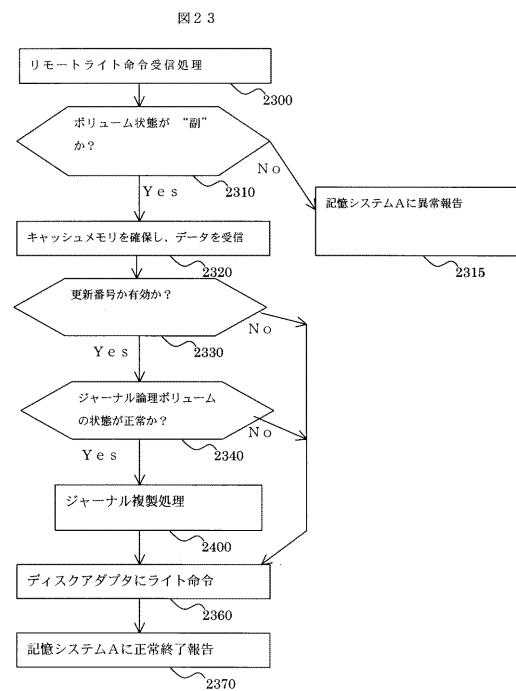
310 更新情報

【図 22】

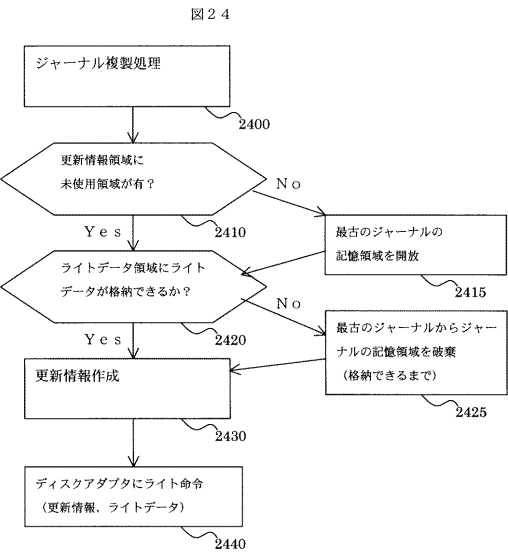
図 22

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	5
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

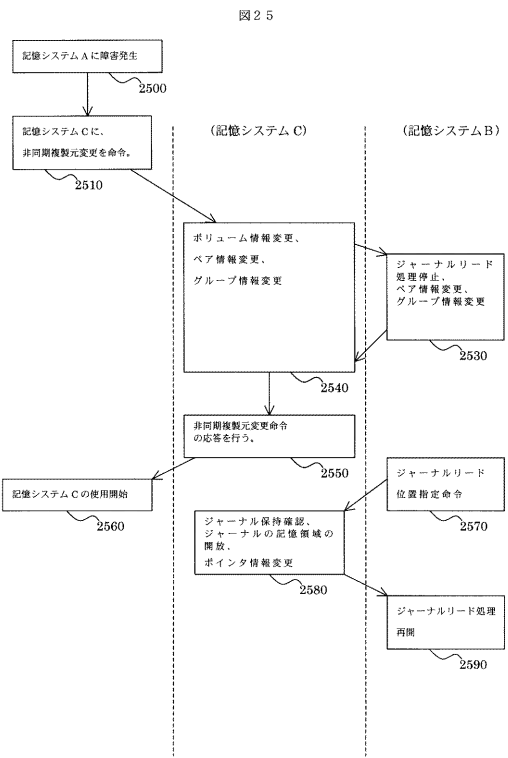
【図 23】



【図 2 4】



【図 2 5】



【図 2 6】

図 2 6

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	同期 ペア番号	非同期 ペア番号	物理アドレス	
						記憶装置 番号	先頭から位置
1	副	OPEN3	3	0	1	1	0
2	副	OPEN6	6	0	2	1	3
3	正常	OPEN9	9	0	0	2	0

【図 2 9】

図 2 9

	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	3	0
ライトデータ領域先頭アドレス	3	700
更新情報最新アドレス	3	400
更新情報最古アドレス	3	200
ライトデータ最新アドレス	3	2000
ライトデータ最古アドレス	3	1300
リード開始アドレス	3	0
リトライ開始アドレス	3	0

【図 2 7】

図 2 7

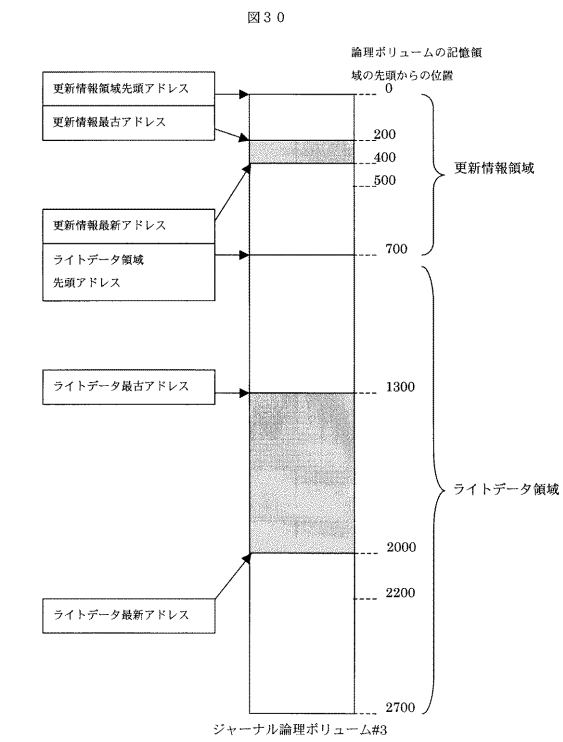
ペア 番号	ペア 状態	正記憶 システム 番号	正論理 ボリューム 番号	副記憶 システム 番号	副論理 ボリューム 番号	グループ 番号	コピー 済みア ドレス
1	正常	A	1	B	1	1	0
2	正常	A	2	B	2	1	0
3	未使用	0	0	0	0	0	0

【図 2 8】

図 2 8

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	複製種類	相手記憶 システム番 号	相手グループ 番号
1	正常	1,2	3	4	非同期	A	1
2	未使用	0	0	0	0	0	0

【図 3 0】



【図 3 1】

図 3 1

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	同期 ペア番号	非同期 ペア番号	物理アドレス	
						記憶装置 番号	先頭から位置
1	副	OPEN3	3	1	0	1	0
2	副	OPEN6	6	2	0	1	3
3	正常	OPEN9	9	0	0	2	0

【図 3 2】

図 3 2

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	A	1	C	1	1	0
2	正常	A	2	C	2	1	0
3	未使用	0	0	0	0	0	0
3	未使用	0	0	0	0	0	0

【図 3 3】

図 3 3

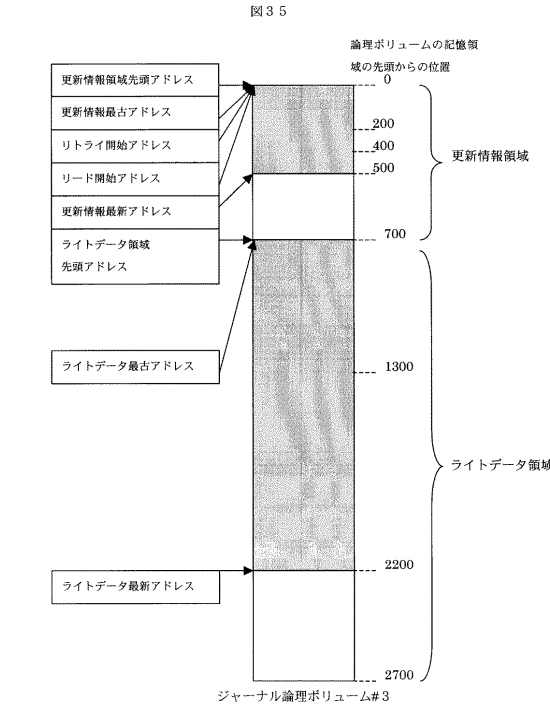
グループ 番号	グループ 状態	ペア集合	ジャーナル 論理ボリュ ーム番号	更新 番号	複製 種類	相手記憶 システム番号	相手グループ 番号
1	正常	1,2	3	5	同期	A	2
2	未使用	0	0	0	0	0	0

【図 3 4】

図 3 4

	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	3	0
ライトデータ領域先頭アドレス	3	700
更新情報最新アドレス	3	500
更新情報最古アドレス	3	0
ライトデータ最新アドレス	3	2200
ライトデータ最古アドレス	3	700
リード開始アドレス	3	0
リトライ開始アドレス	3	0

【図 3 5】



【図 36】

図 36

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	C	1	B	1	1	0
2	正常	C	2	B	2	1	0
3	未使用	0	0	0	0	0	0

【図 39】

図 39

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	異常	A	1	C	1	1	0
2	異常	A	2	C	2	1	0
3	正常	C	1	B	1	1	0
4	正常	C	2	B	2	1	0

【図 37】

図 37

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	複製 種類	相手記憶 システム番号	相手グループ 番号
1	停止	1,2	3	4	非同期	C	2
2	未使用	0	0	0	0	0	0

【図 40】

図 40

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	複製 種類	相手記憶 システム番号	相手グループ 番号
1	異常	1,2	0	5	同期	A	2
2	正常	3,4	3	5	非同期	B	1

【図 38】

図 38

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	同期 ペア番号	非同期 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	3	1	0
2	正	OPEN6	6	2	4	1	3
3	正常	OPEN9	9	0	0	2	0

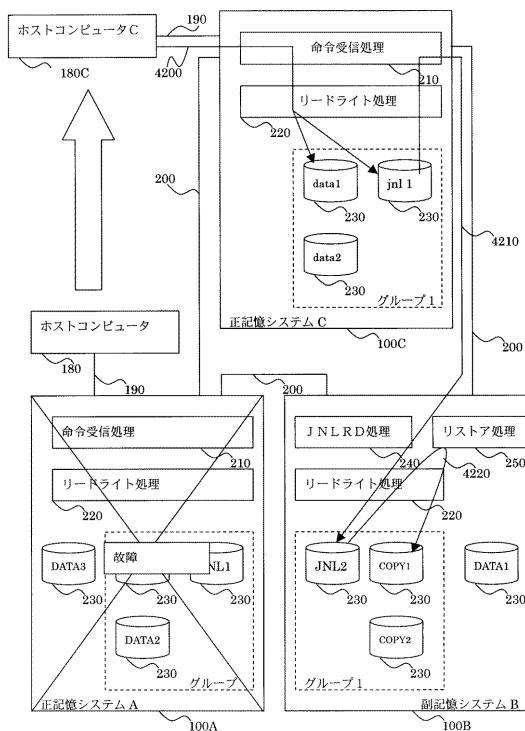
【図 41】

図 41

	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	3	0
ライトデータ領域先頭アドレス	3	700
更新情報最新アドレス	3	500
更新情報最古アドレス	3	400
ライトデータ最新アドレス	3	2200
ライトデータ最古アドレス	3	2000
リード開始アドレス	3	400
リトライ開始アドレス	3	400

【図 42】

図 42



【図 43】

図 43

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	C	1	A	1	1	0
2	正常	C	2	A	2	1	0
3	正常	C	1	B	1	1	0
4	正常	C	2	B	2	1	0

【図 44】

図 44

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	複製 種類	相手記憶 システム番号	相手グループ 番号
1	正常	1,2	0	5	同期	A	2
2	正常	3,4	3	5	非同期	B	1

【図 45】

図 45

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	同期 ペア番号	非同期 ペア番号	物理アドレス	
						記憶装置 番号	先頭から位置
1	副	OPEN3	3	3	0	1	0
2	副	OPEN6	6	4	0	1	3
3	未使用	OPEN6	6	0	0	1	9
4	正常	OPEN9	9	0	0	2	0
5	正常	OPEN3	3	0	0	2	9
6	未使用	OPEN6	6	0	0	2	12

【図 4 6】

図 4 6

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	未使用	A	1	B	1	1	0
2	未使用	A	2	B	2	1	0
3	正常	C	1	A	1	2	0
4	正常	C	2	A	2	2	0
5	未使用	0	0	0	0	0	0

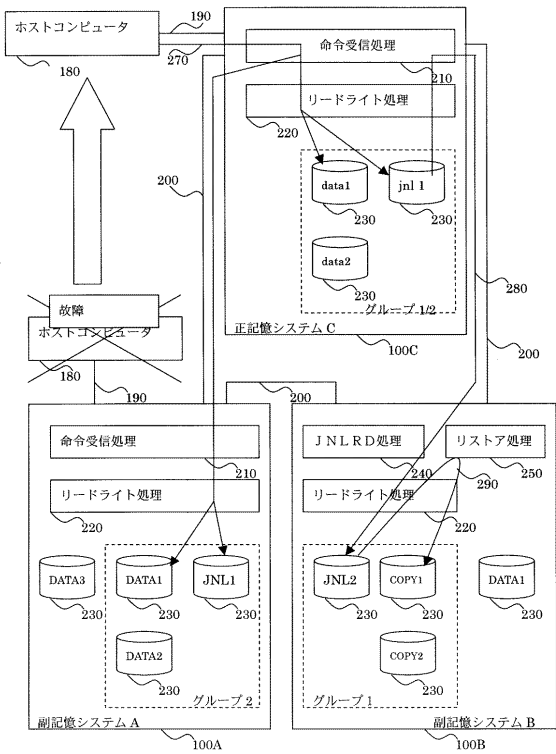
【図 4 7】

図 4 7

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	複製 種類	相手記憶 システム番号	相手グループ 番号
1	未使用	1,2	0	5	非同期	B	1
2	正常	3,4	4	5	同期	C	1

【図 4 8】

図 4 8



フロントページの続き

(72)発明者 安積 義弘

神奈川県小田原市中里322番2号 株式会社日立製作所 R A I Dシステム事業部内

審査官 野田 佳邦

(56)参考文献 特開2003-122509(JP,A)

特開2000-181634(JP,A)

特開2002-244907(JP,A)

特開2000-010845(JP,A)

特開2002-063058(JP,A)

特開平07-244597(JP,A)

特開2005-018506(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F 12/00

G06F 3/06