

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4783076号
(P4783076)

(45) 発行日 平成23年9月28日(2011.9.28)

(24) 登録日 平成23年7月15日(2011.7.15)

(51) Int.Cl.

F I

G 0 6 F 3/06 (2006.01)

G 0 6 F 3/06 3 0 4 F

請求項の数 10 (全 23 頁)

(21) 出願番号	特願2005-199811 (P2005-199811)	(73) 特許権者	000005108
(22) 出願日	平成17年7月8日(2005.7.8)		株式会社日立製作所
(65) 公開番号	特開2007-18312 (P2007-18312A)		東京都千代田区丸の内一丁目6番6号
(43) 公開日	平成19年1月25日(2007.1.25)	(74) 代理人	100093861
審査請求日	平成20年5月9日(2008.5.9)		弁理士 大賀 真司
		(72) 発明者	佐藤 元
			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	松井 佑光
			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	坂東 博司

最終頁に続く

(54) 【発明の名称】 ディスクアレイ装置及びその制御方法

(57) 【特許請求の範囲】

【請求項1】

指定された時点における第1の論理ボリュームの内容と、現在の前記第1の論理ボリュームの内容との差分を表す差分データ及びその管理情報である差分管理情報を保持し、保持した前記差分データ及び前記差分管理情報に基づいて、前記指定された時点における第1の論理ボリュームを復元する機能が搭載されたディスクアレイ装置において、

前記ディスクアレイ装置の稼動時に前記差分管理情報を保持するためのメモリと、

1又は複数の記憶デバイスと、

前記メモリ及び前記記憶デバイスに対するデータの読み書きを制御する制御部と、

を備え、

前記記憶デバイスは、

システムを起動及び制御するために用いられるシステム情報を格納するためのシステム領域及びユーザデータを格納するためのユーザデータ領域を記憶領域として提供し、

前記制御部は、

外部操作に応じて、前記第1の論理ボリューム及び前記差分管理情報を格納するための第2の論理ボリュームを前記ユーザデータ領域に設定し、

前記ディスクアレイ装置の稼動を停止させる際には、前記メモリに保持された前記差分管理情報のうち前記第1の論理ボリュームの内容変更の有無を管理するための情報である差分ビットマップを前記システム領域に退避させ、前記メモリに保持された前記差分管理情報のうちプールボリュームに格納された各差分データとそれぞれ1対1で対応付けられ

た複数の管理領域を有する差分情報テーブル及び前記第1の論理ボリュームの各ブロック領域と前記差分情報テーブル上の対応する管理領域とを対応付けたプライマリボリュームアドレステーブルを前記第2の論理ボリュームに退避させ、

前記ディスクアレイ装置を起動させる際には、当該差分管理情報を前記システム領域及び前記第2の論理ボリュームから前記メモリに読み出させ、

前記第2の論理ボリュームを設定する際には、前記メモリの記憶容量を検出し、次いで前記記憶デバイスが提供する論理ボリュームのうち、未使用の論理ボリュームを検出し、該検出した未使用の論理ボリュームのうち、前記検出したメモリの記憶容量のうち前記差分管理情報を格納するために使用する記憶容量よりも大きな記憶容量であり、かつ、予め定められた記憶容量以内の記憶容量の論理ボリュームを、前記差分管理情報を退避させる前記第2の論理ボリュームの候補としてユーザに提示する

10

ことを特徴とするディスクアレイ装置。

【請求項2】

前記制御部は、

前記外部操作に応じて、前記第2の論理ボリュームを複数設定し、

前記ディスクアレイ装置の稼動を停止させる際には、設定された各前記第2の論理ボリュームにそれぞれ前記差分管理情報を退避させる

ことを特徴とする請求項1に記載のディスクアレイ装置。

【請求項3】

前記制御部は、

前記第2の論理ボリュームを上位装置が認識できないように遮蔽する

ことを特徴とする請求項1に記載のディスクアレイ装置。

20

【請求項4】

前記差分管理情報は、

対応する前記差分データの世代を管理するための情報を含む

ことを特徴とする請求項1に記載のディスクアレイ装置。

【請求項5】

前記差分管理情報は、

対応する前記差分データと、当該差分データの前の前記世代及び又は次の前記世代の前記差分データとを関連付ける情報を含む

30

ことを特徴とする請求項1に記載のディスクアレイ装置。

【請求項6】

指定された時点における第1の論理ボリュームの内容と、現在の前記第1の論理ボリュームの内容との差分を表す差分データ及びその管理情報である差分管理情報を保持し、保持した前記差分データ及び前記差分管理情報に基づいて、前記指定された時点における第1の論理ボリュームを復元する機能が搭載されたディスクアレイ装置の制御方法において、

前記ディスクアレイ装置は、

前記ディスクアレイ装置の稼動時に前記差分管理情報を保持するためのメモリと、

1又は複数の記憶デバイスと、

40

前記メモリ及び前記記憶デバイスに対するデータの読み書きを制御する制御部と、を備え、

前記記憶デバイスが、

システムを起動及び制御するために用いられるシステム情報を格納するためのシステム領域及びユーザデータを格納するためのユーザデータ領域を記憶領域として提供する第1のステップと、

前記制御部が、

外部操作に応じて、前記第1の論理ボリューム及び前記差分管理情報を格納するための第2の論理ボリュームを前記ユーザデータ領域に設定する第2のステップと、

前記ディスクアレイ装置の稼動を停止させる際には、前記メモリに保持された前記差分

50

管理情報のうち前記第1の論理ボリュームの内容変更の有無を管理するための情報である差分ビットマップを前記システム領域に退避させ、前記メモリに保持された前記差分管理情報のうちプールボリュームに格納された各差分データとそれぞれ1対1で対応付けられた複数の管理領域を有する差分情報テーブル及び前記第1の論理ボリュームの各ブロック領域と前記差分情報テーブル上の対応する管理領域とを対応付けたプライマリボリュームアドレステーブルを前記第2の論理ボリュームに退避させる第3のステップと、

前記ディスクアレイ装置を起動させる際には、当該差分管理情報を前記システム領域及び前記第2の論理ボリュームから前記メモリに読み出させる第4のステップと、

前記第2の論理ボリュームを設定する際には、前記メモリの記憶容量を検出し、次いで前記記憶デバイスが提供する論理ボリュームのうち、未使用の論理ボリュームを検出し、該検出した未使用の論理ボリュームのうち、前記検出したメモリの記憶容量のうち前記差分管理情報を格納するために使用する記憶容量よりも大きな記憶容量であり、かつ、予め定められた記憶容量以内の記憶容量の論理ボリュームを、前記差分管理情報を退避させる前記第2の論理ボリュームの候補としてユーザに提示する第5のステップと

を備えることを特徴とするディスクアレイ装置の制御方法。

【請求項7】

前記第5のステップでは、

前記外部操作に応じて、前記第2の論理ボリュームを複数設定し、

前記第3のステップでは、

前記ディスクアレイ装置の稼働を停止させる際に、設定された各前記第2の論理ボリュームにそれぞれ前記差分管理情報を退避させる

ことを特徴とする請求項6に記載のディスクアレイ装置の制御方法。

【請求項8】

前記第2の論理ボリュームを上位装置が認識できないように遮蔽する

ことを特徴とする請求項6に記載のディスクアレイ装置の制御方法。

【請求項9】

前記差分管理情報は、

対応する前記差分データの世代を管理するための情報を含む

ことを特徴とする請求項6に記載のディスクアレイ装置の制御方法。

【請求項10】

前記差分管理情報は、

対応する前記差分データと、当該差分データの前の前記世代及び又は次の前記世代の前記差分データとを関連付ける情報を含む

ことを特徴とする請求項6に記載のディスクアレイ装置の制御方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ディスクアレイ装置及びその制御方法に関し、例えばスナップショット機能が搭載されたディスクアレイ装置に適用して好適なものである。

【背景技術】

【0002】

従来、下記特許文献1に記載のように、ディスクアレイ装置の機能の1つとして、スナップショットの作成指示を受けた時点における指定された論理ボリューム（以下、これをプライマリボリュームと呼ぶ）のイメージを保持する、いわゆるスナップショット機能がある。スナップショット機能は、人為的なミスによりデータが消失してしまったときや、所望時点におけるファイルシステムの状態を復元したいときなどに、その時点におけるプライマリボリュームを復元するために用いられる。

【0003】

かかるスナップショット機能により保持されるプライマリボリュームのイメージ（パーティションとも呼ばれる）は、スナップショットの作成指示を受けた時点における

10

20

30

40

50

プライマリボリューム全体のデータではなく、現在のプライマリボリュームのデータと、プールボリュームと呼ばれる専用の論理ボリュームに保持される差分データとから構成される。この差分データは、スナップショットの作成指示を受けた時点におけるプライマリボリュームと、現在のプライマリボリュームとの差分である。そしてこれらの差分データと現在のプライマリボリュームとに基づいて、かかるスナップショット作成指示された時点におけるプライマリボリュームの状態が復元される。従って、スナップショット機能によれば、プライマリボリュームの内容をそのまま記憶する場合に比べて、より小さい記憶容量でスナップショット作成指示された時点のプライマリボリュームを復元できるという利点がある。

【特許文献1】特開2003-242011号公報

10

【発明の開示】

【発明が解決しようとする課題】

【0004】

ところで、従来のディスクアレイ装置においては、OS (Operating System) などからなるマイクロプログラムや、各論理ボリュームの開始LBA (Logical Block Address)、終了LBA、容量及び属性などの構成情報、各外部ポートの設定情報などのディスクアレイ装置全体を起動及び制御するために必要な情報（以下、これらをまとめてシステム情報と呼ぶ）がディスクアレイ装置内の2つ又はそれ以上のディスクドライブにそれぞれミラー状に書き込まれて保持されている。このようにシステム情報が書き込まれる領域はシステム領域と呼ばれ、かかる各ディスクドライブに予め固定的に確保されている。

20

【0005】

そして、従来のディスクアレイ装置では、上述したスナップショット機能における各差分データの管理情報（以下、これらをスナップショット差分管理情報と呼ぶ）も、ユーザ指示により装置を停止させる計画停止の際に、このシステム領域に退避させて保持するように構成されていた。

【0006】

このため、かかる従来のディスクアレイ装置では、ユーザの希望によりスナップショット機能が搭載されていない場合や、未だスナップショット機能を利用していない場合においても、上述のようなスナップショット差分管理情報を保持するための記憶領域が、使用されないにもかかわらずシステム領域内に予め固定的に確保されるなど、記憶資源を有効かつ効率的に使用し難い問題があった。

30

【0007】

本発明は以上の点を考慮してなされたもので、記憶資源を有効かつ効率的に使用し得るディスクアレイ装置及びその制御方法を提案しようとするものである。

【課題を解決するための手段】

【0008】

かかる課題を解決するため、本発明は、指定された時点における第1の論理ボリュームの内容と、現在の前記第1の論理ボリュームの内容との差分を表す差分データ及びその管理情報である差分管理情報を保持し、保持した前記差分データ及び前記差分管理情報に基づいて、前記指定された時点における当該第1の論理ボリュームを復元する機能が搭載されたディスクアレイ装置において、前記ディスクアレイ装置の稼動時に前記差分管理情報を保持するためのメモリと、1又は複数の記憶デバイスと、前記メモリ及び前記記憶デバイスに対するデータの読み書きを制御する制御部とを備え、前記制御部は、外部操作に応じて、前記差分管理情報を格納するための第2の論理ボリュームを前記記憶デバイスが提供する記憶領域に設定し、前記ディスクアレイ装置の稼動を停止させる際には、前記メモリに保持された前記差分管理情報を前記第2の論理ボリュームに退避させ、前記ディスクアレイ装置を起動させる際には、当該差分管理情報を前記第2の論理ボリュームから前記メモリに読み出させることを特徴とする。

40

【0009】

また本発明は、指定された時点における第1の論理ボリュームの内容と、現在の前記第

50

1の論理ボリュームの内容との差分を表す差分データ及びその管理情報である差分管理情報を保持し、保持した前記差分データ及び前記差分管理情報に基づいて、前記指定された時点における当該第1の論理ボリュームを復元する機能が搭載されたディスクアレイ装置の制御方法において、外部操作に応じて、前記差分管理情報を格納するための第2の論理ボリュームを所定の記憶デバイスが提供する記憶領域に設定する第1のステップと、前記ディスクアレイ装置の稼動を停止させる際には、所定のメモリに保持された前記差分管理情報を前記第2の論理ボリュームに退避させ、前記ディスクアレイ装置を起動させる際には、当該差分管理情報を前記第2の論理ボリュームから前記メモリに読み出させる第2のステップとを備えることを特徴とする。

【発明の効果】

10

【0010】

本発明によれば、差分管理情報を格納するための第2の論理ボリュームを必要なときに設定できるようにすることができ、差分管理情報を格納するための記憶領域を予め固定的に確保しておく場合に比べて、記憶資源を有効かつ効率的に使用することができる。

【発明を実施するための最良の形態】

【0011】

以下図面について、本発明の一実施の形態を詳述する。

【0012】

(1) 第1の実施の形態

20

(1-1) 本実施の形態によるディスクアレイ装置1の構成

図1は、本実施例に係るディスクアレイ装置1の構成を示す。ディスクアレイ装置1は、通信ネットワーク2を介してホスト装置（上位ホスト装置）3に接続されている。

【0013】

ホスト装置3は、ディスクアレイ装置1の上位装置として機能するサーバ、パーソナルコンピュータ、ワークステーション、メインフレームなどから構成される。通信ネットワーク2としては、例えば、LAN(Local Area Network)、SAN(Storage Area Network)、インターネット、専用回線等を挙げることができる。

【0014】

ホスト装置3がオープン系の場合、通信ネットワーク2を介したホスト装置3及びディスクアレイ装置1間の通信は、例えばTCP/IP(Transmission Control Protocol/Internet Protocol)、FCP(Fibre Channel Protocol)、iSCSI(internet Small Computer System Interface)等のプロトコルに従って行われる。またホスト装置3がメインフレーム系の場合、通信ネットワーク2を介したホスト装置3及びディスクアレイ装置1間の通信は、例えばFICON(Fibre Connection:登録商標)、ESCON(Enterprise System Connection:登録商標)、ACONARC(Advanced Connection Architecture:登録商標)、FIBARC(Fibre Connection Architecture:登録商標)等の通信プロトコルに従って行われる。

30

【0015】

ディスクアレイ装置1は、二重化されたコントローラ10、20を備えるデュアルコントローラ構成を有する。コントローラ10は、主にCPU11、ローカルメモリ(LM)12、データコントローラ(DCTL)13、キャッシュメモリ(CM)14、ファイバチャネルプロトコルコントローラ(FCTL)15、PBC(Port Bypass Circuit)16、エンクロージャコントローラ(ECTL)17、及びFC-AL(Fibre Channel Arbitrated Loop)30、31を備える。

40

【0016】

CPU11は、ホスト装置3からのデータ入出力要求に応答して、複数のディスクドライバDRV0~DRV14へのデータ入出力処理(ライトアクセス又はリードアクセス)を制御するプロセッサであり、ローカルメモリ12に格納されたマイクロプログラムを実行することによりデータコントローラ13、ファイバチャネルプロトコルコントローラ1

50

5 及び P B C 1 6 等を制御する。

【0017】

データコントローラ 1 3 は、C P U 1 1 の制御に基づいて、ファイバチャネルプロトコルコントローラ 1 5 とキャッシュメモリ 1 4 との間のデータ転送を制御する。キャッシュメモリ 1 4 は、ファイバチャネルプロトコルコントローラ 1 5 を介してフロントインターフェース又はバックインターフェースとの間で授受されるデータを一時的に記憶する。

【0018】

ファイバチャネルプロトコルコントローラ 1 5 は、フロントインターフェース接続用ファイバチャネル 1 8 F を介してホスト装置 3 にインターフェース接続されており、ファイバチャネルプロトコルによるブロックアクセス要求を受信する機能を有する。またファイバチャネルプロトコルコントローラ 1 5 は、バックインターフェース接続用ファイバチャネル 1 8 B を介してディスクドライブ D R V 0 ～ D R V 1 4 にインターフェース接続されており、ディスクドライブ D R V 0 ～ D R V 1 4 を制御するコマンド等を規定するプロトコルに従って、ディスクドライブ D R V 0 ～ D R V 1 4 へのデータ入出力要求を送信する機能を有する。P B C 1 6 は、バックインターフェース接続用ファイバチャネル 1 8 B の接続先として、二つの F C - A L 3 0 , 3 1 のうち何れか一方又は両者を選択し、接続切り替えを行う。

【0019】

F C - A L 3 0 は、F C - A L 3 1 上に実装された P B C 3 5 を介して、偶数番号のディスクドライブ D R V 0 , D R V 2 , … , D R V 1 4 にループ接続されている。また F C - A L 3 1 は、F C - A L 3 1 上に実装された P B C 3 5 を介して、奇数番号のディスクドライブ D R V 1 , D R V 3 , … , D R V 1 3 にループ接続されている。

【0020】

P B C 3 5 は、システム運用上の障害を最小限に止めて、ノードの追加や削除が行えるように F C - A L 3 0 , 3 1 を自動的にオープン又はクローズする電子スイッチである。P B C 3 5 は、ファイバチャネルプロトコルコントローラ 1 5 やディスクドライブ D R V 0 ～ D R V 1 4 をバイパスして、これらを F C - A L 3 0 , 3 1 から電氣的に除外する機能を有する。例えば P B C 3 5 は、障害が発生したディスクドライブ D R V 0 ～ D R V 1 4 を F C - A L 3 0 , 3 1 から切り離して、他のディスクドライブ D R V 0 ～ D R V 1 4 とファイバチャネルプロトコルコントローラ 1 5 との間の通信を可能にする。また P B C 3 5 は、F C - A L 3 0 , 3 1 の動作を維持したまま、ディスクドライブ D R V 0 ～ D R V 1 4 の抜き差しを可能にする。例えば、新たなディスクドライブが装着された場合には、そのディスクドライブを F C - A L 3 0 , 3 1 に取り込み、ファイバチャネルプロトコルコントローラ 1 5 との間の通信を可能にする。

【0021】

エンクロージャコントローラ (E - C T L) 1 7 は、S E S (S C S I Enclosure Surcuit) ドライブを制御する。S E S ドライブは、S C S I 3 (Small Computer System Interface 3) 規格に規定される S E S (S C S I Enclosure Services) や E S I (Enclosure Service I/F) の機能を備えており、S E S (S C S I Enclosure Service) や E S I (Enclosure Service I/F) の機能を動作させることができる。本実施例の場合、S E S ドライブは、F C - A L 3 0 ～ 3 3 の他に、エンクロージャコントローラ 1 7 と通信可能なディスクドライブである。ここでは、各 F C - A L につき 2 台のディスクドライブ D R V 0 ～ D R V 3 を S E S ドライブとしているが、全てのディスクドライブ D R V 0 , D R V 2 , … , D R V 1 4 を S E S ドライブとしてもよい。

【0022】

一方、コントローラ 2 0 は、コントローラ 1 0 と同様に構成されており、C P U 2 1 、ローカルメモリ (L M) 2 2 、データコントローラ (D - C T L) 2 3 、キャッシュメモリ (C M) 2 4 、ファイバチャネルプロトコルコントローラ (F C - C T L) 2 5 、P B C 2 6 、エンクロージャコントローラ (E - C T L) 2 7 、フロントインターフェース接続用ファイバチャネル 2 8 F 、バックインターフェース接続用ファイバチャネル 2 8 B

、及びFC-AL32, 33を備えて構成される。

【0023】

PBC26は、バックインターフェース接続用ファイバチャネル28Bの接続先として、2つのFC-AL32, 33のうち何れか一方又は両方を選択し、接続切り替えを行う。FC-AL32は、PBC35を介して偶数番号のディスクドライブDRV0, DRV2, …, DRV14にループ接続している。FC-AL33は、PBC35を介して奇数番号のディスクドライブDRV1, DRV3, …, DRV13にループ接続されている。

【0024】

データコントローラ13, 23はバス36を介して相互に接続されており、一方のデータコントローラ13（又は23）がバス36を介して他方のデータコントローラ23（又は13）に対して、コマンド転送やデータ転送等を行うこともできる。例えば、両コントローラ10, 20が分担して同一の論理ボリュームに対するアクセスを行う場合に、両コントローラ間でライトデータ又はリードデータの転送を行うこともできる。

【0025】

コントローラ10, 20は、ディスクドライブDRV0, DRV2, …, DRV14をいわゆるRAID方式に規定されるRAIDレベル（例えば、0, 1, 5）で制御することができる。RAID方式では、複数のディスクドライブDRV0, DRV2, …, DRV14を一つのグループ（以下、RAIDグループと称する。）として管理する。RAIDグループ上には、ホスト装置3からのアクセス単位である論理ボリュームが形成される。各論理ボリュームには、LUN（Logical Unit Number）と呼ばれる識別子が付与されている。

【0026】

なお、PBC16とFC-AL30, 31との間には、それぞれPBC40, 41が実装されている。PBC40は、FC-AL30に障害が生じた場合には、バックインターフェース接続用ファイバチャネル18Bを、ファイバチャネル44（点線）を介してFC-AL32に接続する。PBC41は、FC-AL31に障害が生じた場合には、バックインターフェース接続用ファイバチャネル18Bを、ファイバチャネル45（点線）を介してFC-AL33に接続する。

【0027】

同様に、PBC26とFC-AL32, 33との間には、それぞれPBC42, 43が実装されている。PBC42は、FC-AL32に障害が生じた場合には、バックインターフェース接続用ファイバチャネル28Bを、ファイバチャネル46（点線）を介してFC-AL30に接続する。PBC43は、FC-AL33に障害が生じた場合には、バックインターフェース接続用ファイバチャネル28Bを、ファイバチャネル47（点線）を介してFC-AL31に接続する。

【0028】

仮に、各コントローラ10, 20に実装されている全てのPBC40～43がファイバチャネル44～47に接続することにより、バックインターフェース接続用ファイバチャネル18B, 28Bが相手コントローラ側のFC-AL30～33に接続された状態を想定すると、この状態は、例えば、特開2000-187561号の図1に示されているように、両コントローラ間を接続するファイバチャネルループに複数のディスクドライブが接続する状態と同じになる。

【0029】

以下の説明においては、FC-AL30をコントローラ20の稼働系FC-AL、FC-AL31をコントローラ20の待機系FC-AL、FC-AL32をコントローラ20の待機系FC-AL、FC-AL33をコントローラ20の稼働系FC-ALとした場合を想定して説明を進める。

【0030】

図2は複数のディスクドライブDRV10～DRV14に形成される論理ボリュームを示している。ここでは、説明の便宜上、二つの論理ボリュームLU1, LU2が示されて

10

20

30

40

50

いる。ホスト装置 3 に搭載されている OS (Operating System) は、論理ボリューム名を指定してディスクアレイ装置 10 にアクセスする。仮に、各論理ボリューム LU 1, LU 2 を構成する物理ディスクを一つにすると、その物理ディスクが故障したときに、その論理ボリューム LU 1, LU 2 は使用不能となる。これを防止するために、RAID 方式では、各論理ボリューム LU 1, LU 2 が異なる複数のディスクドライブ DRV 10 ~ DRV 14 上に形成される。その一方で、コントローラ 10 の稼働系 FC-AL 30 は偶数番号のディスクドライブ DRV 10、12、14 に接続され、コントローラ 20 の稼働系 FC-AL 33 は奇数番号のディスクドライブ DRV 11、13 に接続されているため、各々の論理ボリューム LU 1, LU 2 へのアクセスは、必ずしも何れか一方のコントローラ 10 又は 20 のみが行うものではなく、両コントローラ 10, 20 が協調制御の下、協力して行うように構成してもよい。

10

【0031】

ホスト装置 3 からディスクアレイ装置 1 の論理ボリューム LU 1 又は LU 2 に対して、ライトアクセスが行われると、このライトアクセスと共にホスト装置 3 から出力されて一方のコントローラ 10 内のキャッシュメモリ 14 に格納されたライトデータは、バス 36 を介して他方のコントローラ 20 内のキャッシュメモリ 14 にも格納される。このようにデータを二重化することで、一方のコントローラ 10 に障害が生じた場合でも、他方のコントローラ 20 にフェイルオーバーできる。キャッシュメモリ 14, 24 にライトデータが格納された時点でディスクアレイ装置 1 はホスト装置 3 に対してライトアクセス完了の通知を行う。次いで、コントローラ 10 は、FC-AL 30 を介してディスクドライブ DRV 0, DRV 2, ..., DRV 14 へのライトアクセスを行う。ライトアクセスが行われる論理ボリューム LU 1 又は LU 2 は、偶数番号のディスクドライブ DRV 0, DRV 1, ..., DRV 14 だけでなく、奇数番号のディスクドライブ DRV 1, DRV 3, ..., DRV 13 をも含むので、奇数番号のディスクドライブ DRV 1, DRV 3, ..., DRV 13 へのライトアクセスは、コントローラ 20 が FC-AL 33 を介して行う。

20

【0032】

一方、ホスト装置 3 からディスクアレイ装置 1 の論理ボリューム LU 1 又は LU 2 に対して、リードアクセスが行われると、コントローラ 10 はキャッシュメモリ 14 をチェックし、リードアクセスの対象となるデータがある場合には、該当データを読み取ってホスト装置 3 に送信する。キャッシュメモリ 14 に該当データがない場合には、コントローラ 10 は、FC-AL 61 を介してディスクドライブ DRV 0, DRV 2, ..., DRV 14 へのリードアクセスを行う。リードアクセスが行われる論理ボリューム LU 1 又は LU 2 は、偶数番号のディスクドライブ DRV 10、12、14 だけでなく、奇数番号のディスクドライブ DRV 11、13 をも含むので、奇数番号のディスクドライブ DRV 11、13 へのリードアクセスはコントローラ 30 が FC-AL 33 を介して行う。各コントローラ 10, 20 によって読み出されたリードデータはキャッシュメモリ 14, 24 に二重書きされる。

30

【0033】

このように、一つの論理ボリュームに対するアクセスを、両コントローラ 10, 20 が協調制御の下、協力して行うことにより、各コントローラ 10, 20 の処理負荷を均等化できる。本実施例では、偶数番号のディスクドライブ DRV 10、12、14 と、奇数番号のディスクドライブ DRV 11、13 とに二分して、各コントローラ 10, 20 の処理負荷を均等化する例を示したが、各コントローラ 10, 20 が担当するディスクドライブ数が略同程度になるように、例えば、ディスクドライブ DRV 11、12 と、ディスクドライブ DRV 13、14 とに二分して、各コントローラ 10, 20 の処理負荷を均等化してもよい。

40

【0034】

また、本実施例では、二つの PBC 16, 26 を制御して、バックインターフェース接続用ファイバチャネル 18B, 28B を稼働系 FC-AL 30, 33 に接続するだけで、全てのディスクドライブ DRV 0 ~ DRV 14 は、何れか一方のコントローラ 10 又は 2

50

0に接続されるため、FC-A L 30～33上に実装された個々のPBC 35を個別に制御する必要がなく、制御が容易である。また、各CPU 11, 21は、各コントローラ内の稼働系FC-A Lと待機系のFC-A Lを区別することなく、両者とも同一のデバイスとして認識しているので、制御が容易である。

【0035】

(1-2) スナップショット差分管理情報の管理方式

次に、かかるディスクアレイ装置1におけるスナップショット差分管理情報の管理方式について説明する。

【0036】

本ディスクアレイ装置1では、図3に示すように、FC-A L 30～33に接続された先頭から5つ分のディスクドライブDRV 0～4に固定的にシステム領域SAR 0～SAR 4が割り当てられ、これら5つのディスクドライブDRV 0～4の各システム領域SAR 0～SAR 4内にそれぞれマイクロプログラムなどの全てのユーザが共通に使用するシステム情報を保持(5重化)する一方で、オプション機能であるスナップショット機能の実行により得られる上述のスナップショット差分管理情報の一部を、予めユーザにより設定された専用の論理ボリューム(以下、これを差分管理ボリュームと呼ぶ)に保持するようになされた点を特徴の1つとしている。

【0037】

図4は、このような本ディスクアレイ装置1におけるスナップショット差分管理情報の管理方式を概念的に示したものである。本ディスクアレイ装置1の場合、スナップショット差分管理情報としては、差分ビットマップ50、プライマリボリュームアドレステーブル51及び差分情報テーブル52が存在する。

【0038】

差分ビットマップ50は、スナップショット作成対象の論理ボリュームLUであるプライマリボリュームLU_pの内容変更の有無を、64[Kbyte](ある範囲(例えば1～512[Kbyte]の範囲など)で可変とすることができる)のブロック領域BL単位で管理するためのビットマップであり、プライマリボリュームLU_p内の各ブロック領域BLとそれぞれ1対1で対応付けられた複数のビットを有する。この差分ビットマップ50は、スナップショット作成指示が与えられたタイミングでリセット(オール「0」)され、その後、プライマリボリューム内のいずれかのブロック領域BLの内容が最初に更新されたときに(すなわち、そのブロック領域BLに新たにデータが書き込まれ又はそのブロック領域BLに書き込まれていたデータが上書きされたときに)、そのブロック領域BLと対応付けられたビットが「0」から「1」に変更される。

【0039】

プライマリボリュームアドレステーブル51は、プライマリボリュームLU_pの各ブロック領域BLと差分情報テーブル52上の対応する管理領域52Aとを対応付けたテーブルであり、プライマリボリュームLU_pの各ブロック領域BLにそれぞれ対応させて、複数のアドレス領域51Aが設けられている。そして、これらアドレス領域51Aには、プライマリボリュームLU_pの対応するブロック領域BLの内容が更新されて、対応する差分データD_{DF}がプールボリュームLU_{POOL}に格納されたときに、後述のようにその差分データD_{DF}と対応付けて作成された差分情報テーブル52上の管理領域52Aのアドレスが格納される。

【0040】

差分情報テーブル52は、プールボリュームLU_{POOL}に格納された各差分データD_{DF}とそれぞれ1対1で対応付けられた複数の管理領域52Aを有している。各管理領域52Aには、例えば図5に示すように、4[Byte]の次差分管理情報アドレスフィールドFL₁、4[Byte]の前差分管理情報アドレスフィールドFL₂、同じく4[Byte]の世代管理ビットマップフィールドFL₃、1[Byte]のプライマリボリュームLUNフィールドFL₄、同じく1[Byte]のリザーブフィールドFL₅及び2[Byte]のプライマリボリューム通し番号フィールドFL₆がそれぞれ設けられている。

【0041】

このうち世代管理ビットマップフィールド FL_3 は、複数世代分のバーチャルボリューム（スナップショット）の作成を可能とするために世代管理用のビットマップを格納するフィールドであり、対応する差分データ D_{DF} が属するバーチャルボリュームの世代が格納される。また次差分管理情報アドレスフィールド FL_1 には、次世代のバーチャルボリュームにおける対応する差分データ D_{DF} と関連付けられた管理領域 52A の差分情報テーブル 52 上のアドレスが格納される。さらに前差分管理情報アドレスフィールド FL_2 には、前世代のバーチャルボリュームにおける対応する差分データ D_{DF} と関連付けられた管理領域 52A の差分情報テーブル 52 上のアドレスが格納される。この前差分管理情報アドレスフィールド FL_2 に格納されているアドレスを辿って行くことによって、より古い世

10

【0042】

プライマリボリューム LUN フィールド FL_4 には、対応するプライマリボリューム L_{Up} の LUN の下 1 [Byte] が格納される。これはチェックコードとして使用される。またプライマリボリューム内通し番号フィールド FL_6 には、プライマリボリューム L_{Up} 内での対応するブロック領域 BL の通し番号が格納される。これもチェックコードとして使用される。なお、リザーブフィールド FL_5 は、将来の拡張のために設けられたものであり、通常はダミーデータが格納される。

20

【0043】

このようなスナップショット差分管理情報のうち、差分ビットマップ 50 は、ディスクアレイ装置 1 の立ち下がり時にシステム情報の一部として図 3 について上述したシステム領域 $SAR_0 \sim SAR_4$ に格納されて保存され、プライマリボリュームアドレステーブル 51 及び差分情報テーブル 52 は、ディスクアレイ装置 1 の立下り時に差分管理ボリューム L_{UDM} に格納されて保存される。そしてこれら差分ビットマップ 50、プライマリボリュームアドレステーブル 51A 及び差分情報テーブル 52 は、ディスクアレイ装置 1 の立ち上がり時にそれぞれシステム領域 $SAR_0 \sim SAR_4$ や差分管理ボリューム L_{UDM} から読み出されてキャッシュメモリ 14、24 上に展開され、その後必要に応じて後述のように更新される。

30

【0044】

なお、差分ビットマップ 50、プライマリボリュームアドレステーブル 51 及び差分情報テーブル 52 は、コントローラ 10、20 毎にそれぞれ作成される。ただし、このとき各コントローラ 10、20 同士が通信を行いながら同期して差分情報テーブル 52 等を作成するため、同じものが作成されることとなる。このように差分ビットマップ 50、プライマリボリュームアドレステーブル 51 及び差分情報テーブル 52 をディスクアレイ装置 1 内において 2 重化してもつことにより、一方のデータが壊れた場合においても他方のデータによって対応することが可能となる。

【0045】

図 6 は、本ディスクアレイ装置 1 における差分管理ボリューム L_{UDM} の設定手順を示すフローチャートである。本ディスクアレイ装置 1 では、差分管理ボリューム L_{UDM} を設定する場合、通常の論理ボリューム LU を作成する場合と同様にして、まず複数のディスクドライブ $DRV_0 \sim DRV_4$ からなる $RAID$ グループを設定する（S1）。具体的には、専用のソフトウェアがインストールされたホスト装置 3（図 1）を操作して、そのディスプレイに所定の設定画面を表示させ、この設定画面を用いて、新たに作成しようとする $RAID$ グループを構成するディスクドライブ $DRV_0 \sim DRV_4$ 、 $RAID$ グループの容量、 $RAID$ 構成（ $4D+1P$ 等）及び $RAID$ レベル（1、 $1+0$ 、 $2 \sim 6$ ）などを設定する。

40

【0046】

続いて、ユーザは、このようにして設定した $RAID$ グループ内に論理ボリューム LU を作成する（S2）。具体的には、上述した設定画面を用いて、新たに作成しようとする

50

論理ボリューム L U の L U N や容量などをそれぞれ設定する。

【0047】

この後ユーザは、このようにして作成した 1 又は複数の論理ボリューム L U の中から所望の論理ボリューム L U を差分管理ボリューム L U_{DM} に設定する (S 3)。具体的には、ホスト装置 3 を操作して、そのディスプレイに例えば図 7 に示すような設定画面 6 0 を表示させ、その後この設定画面 6 0 の右下に表示された設定ボタン 6 1 をクリックするようにする。この結果、ホスト装置 3 のディスプレイ上に、図 8 に示すセレクトロジカルユニットダイアログ 6 2 が表示される。

【0048】

このセレクトロジカルユニットダイアログ 6 2 は、差分管理ボリューム L U_{DM} に設定しようとする論理ボリューム L U を選択するためのダイアログであり、そのディスクアレイ装置 1 内に形成された全ての論理ボリューム L U の中から対象となる各論理ボリューム L U の L U N や、容量及びその論理ボリューム L U が属する R A I D グループの識別番号などの各種情報が論理ボリューム情報表示欄 6 2 A 内に一覧表示される。

【0049】

かくしてユーザは、このセレクトロジカルユニットダイアログ 6 2 の論理ボリューム情報表示欄 6 2 A に表示された情報を目視確認しながら必要な条件（この実施の形態では容量が 5 [Gbyte] 以上）を満たす所望の論理ボリューム L U を選択し、その後 O K ボタン 6 3 をクリックするようにする。この結果、図 9 に示すような確認メッセージダイアログ 6 4 がホスト装置 3 のディスプレイに表示される。そしてユーザは、問題がなければこの確認メッセージダイアログ 6 4 の O K ボタン 6 5 をクリックするようにする。

【0050】

かくして、このとき選択された論理ボリューム L U が差分管理ボリューム L U_{DM} として設定され、図 1 0 に示すように、この選択された論理ボリューム L U （差分管理ボリューム L U_{DM}）の L U N や容量及び当該論理ボリューム L U が属する R A I D グループの識別番号などの各種情報が設定画面 6 0 の差分管理ボリューム表示欄 6 0 A に表示される。

【0051】

このように設定された差分管理ボリューム L U_{DM} は、ディスクアレイ装置 1 内において、他の通常の論理ボリューム L U と同様に扱われる。ただし、差分管理ボリューム L U_{DM} に設定された論理ボリューム L U は、ホスト装置 3 からは認識されないように遮蔽される。

【0052】

なお、本実施のディスクアレイ装置 1 の場合、セレクトロジカルユニットダイアログ 6 2 (図 8) を用いて、差分管理ボリューム L U_{DM} として最大 2 個の論理ボリューム L U を選択し設定することができる。この場合、2 個目の論理ボリューム L U は、1 個目の論理ボリューム L U のミラーリング用として用いられる。

【0053】

本ディスクアレイ装置 1 の場合、差分管理ボリューム L U_{DM} が設定されていなければ、スナップショット機能が実行されない。この場合における差分管理ボリューム L U_{DM} が設定されているか否かの判断は、オプションであるスナップショット機能用のプログラムを購入してインストールし、管理プログラムからのその開錠キーを入力した段階では行われず、実際にホスト装置 3 からディスクアレイ装置 1 に対してスナップショット作成指示を与えたときに当該ディスクアレイ装置 1 の各コントローラ 1 0, 2 0 内の C P U 1 1, 2 1 によって行われる。

【0054】

図 1 1 は、このような判断に関する C P U 1 1, 2 1 の処理手順を示すフローチャートである。C P U 1 1, 2 1 は、ホスト装置 3 からスナップショット作成指示が与えられると、まずキャッシュメモリ 1 4, 2 4 上に展開されているシステム情報に基づいて、差分管理ボリューム L U_{DM} が設定されているか否かを判定する (S 1 0)。

【0055】

10

20

30

40

50

そしてCPU 11, 21は、差分管理ボリュームLU_{DM}が設定されていないときには、ホスト装置3に対してエラーメッセージを送信する(S 11)。この結果、対応するホスト装置3のディスプレイに対応するエラーメッセージが表示される。

【0056】

一方、CPU 11, 21は、差分管理ボリュームLU_{DM}が設定されているときには、プライマリボリュームLU_pに対するデータのライト／リード処理と並行して、図12に示すフローチャートに従って、スナップショットの作成処理を実行する(S 12)。

【0057】

すなわち、CPU 11, 21は、スナップショット作成処理を開始すると、まず差分ビットマップ50(図4)及びプライマリボリュームアドレステーブル51(図4)をそれぞれクリアし(S 20)、その後プライマリボリュームLU_pへのライトアクセスを待ち受ける(S 21)。

【0058】

そしてCPU 11, 21は、この後ホスト装置3からプライマリボリュームLU_pへのライトアクセスがあると(S 21: YES)、まず差分ビットマップ50を参照して、そのときライトデータを書き込むべきプライマリボリュームLU_pのブロック領域BLに対するライトアクセスが、最後のスナップショット作成指示が与えられてから最初のものであるか否かを判定する(S 22)。

【0059】

CPU 11, 21は、このライトアクセスがそのブロック領域BLに対する最初のものでなかった場合には(S 22: NO)、再度、プライマリボリュームLU_pへのライトアクセスを待ち受ける待ち受け状態に戻る(S 21)。

【0060】

これに対して、CPU 11, 21は、このライトアクセスがそのブロック領域BLに対する最初のものであった場合には(S 22: YES)、プライマリボリュームLU_pにおけるライトアクセス直前の当該ブロック領域BLのデータを読み出し、これを差分データD_{DF}(図4)としてプールボリュームLU_{POOL}(図4)の対応するアドレス位置に書き込ませる(S 23)。

【0061】

またCPU 11, 21は、この後、キャッシュメモリ14, 24に展開されている差分ビットマップ50、プライマリボリュームアドレステーブル51及び差分情報テーブル52の更新処理を実行する(S 24)。

【0062】

具体的に、CPU 11, 21は、差分ビットマップ50の対応するビットを「0」から「1」に変更する。またCPU 11, 21は、差分情報テーブル52上にその差分データD_{DF}に対する管理領域52A(図4)を確保し、その管理領域52Aにおける世代管理ビットマップフィールドFL₃(図5)、プライマリボリュームLUNフィールドFL₄(図5)及びプライマリボリューム通し番号フィールドFL₆(図5)に必要な情報をそれぞれ格納する。この際CPU 11, 21は、前世代のバーチャルボリュームが存在する場合には、その管理領域52Aにおける前差分管理情報アドレスフィールドFL₂(図5)に、前世代の差分データD_{DF}と対応付けられた管理領域52Aの差分情報テーブル52上のアドレスを格納すると共に、当該前世代の差分データD_{DF}と対応付けられた管理領域52Aの次差分管理情報アドレスフィールドFL₁(図5)に、上述のようにそのときに確保した管理領域52Aの差分情報テーブル52上のアドレスを格納する。さらにCPU 11, 21は、この管理領域52Aのアドレスをプライマリボリュームアドレステーブル51の対応するアドレス領域51Aに格納する。

【0063】

そしてCPU 11, 21は、再度、プライマリボリュームLU_pへのライトアクセスを待ち受ける待ち受け状態に戻る(S 21)。また、この後CPU 11, 21は、同様の処理を繰り返すことで(S 21～S 24-S 21)、スナップショット作成指示時のプライ

10

20

30

40

50

マリボリューム LU_P のイメージ（バーチャルボリューム）を保持する。

【0064】

一方、図13は、ディスクアレイ装置1の計画停止時における差分情報の退避に関するCPU11, 21の処理手順を示すフローチャートである。

【0065】

CPU11, 21は、ユーザの操作に基づいてホスト装置3から稼働を停止すべき旨のコマンドが送信されてくると、まず、他方のコントローラ10, 20のCPU11, 21と通信して、自CPU11, 21及び他方のCPU21, 11の少なくとも一方がスナップショット差分管理情報を更新処理中であるか否かを判断する(S30)。

【0066】

CPU11, 21は、かかる判断において肯定結果を得ると(S30: YES)、双方のCPU11, 21がスナップショット差分管理情報の更新処理を終えるのを待ち受け、やがて双方のCPU11, 21がスナップショット差分管理情報の更新処理を終えると(S30: NO)、キャッシュメモリ14, 24に保持されているシステム情報に基づいて、差分管理ボリューム LU_{DM} のLUNを取得する(S31)。またCPU11, 21は、ミラーリング用の差分管理ボリューム(2個目の差分管理ボリューム) LU_{DM} が設定されている場合には、その差分管理ボリューム LU_{DM} のLUNも取得する(S31)。

【0067】

そしてCPU11, 21は、このようにして取得した差分管理ボリューム LU_{DM} のLUNとシステム情報とに基づいて、キャッシュメモリ14, 24に保持しているスナップショット差分管理情報のうちのプライマリボリュームアドレステーブル51及び差分情報テーブル52のデータを、差分管理ボリューム LU_{DM} に格納(退避)させる(S32)。この際CPU11, 21は、ミラーリング用の差分管理ボリューム LU_{DM} が設定されている場合には、かかるプライマリボリュームアドレステーブル51及び差分情報テーブル52のデータを、そのミラー側の差分管理ボリューム LU_{DM} にも退避させる(S32)。

【0068】

またCPU11, 21は、これと同様にしてシステムキャッシュメモリ14, 24に保持されているシステム情報に基づいてシステム領域 $SAR_0 \sim SAR_4$ が設けられたディスクドライブ $DRV_0 \sim DRV_4$ を検出し、キャッシュメモリ14, 24に展開しているスナップショット差分管理情報のうちの差分ビットマップ50を、キャッシュメモリ14, 24に展開している他のシステム情報と共にかかるシステム領域 $SAR_0 \sim SAR_4$ に格納(退避)させる(S32)。

【0069】

他方、図14は、ディスクアレイ装置1の立ち上がり時におけるスナップショット差分管理情報の読み出しに関するCPU11, 21の処理手順を示すフローチャートである。

【0070】

CPU11, 21は、ユーザの操作に基づいてホスト装置3からディスクアレイ装置1を立ち上げるべき旨のコマンドが送信されてくると、まず、上述のシステム領域 $SAR_0 \sim SAR_4$ が設けられた5つのディスクドライブ $DRV_0 \sim DRV_4$ のいずれかからシステム情報を読み出し、これをキャッシュメモリ14, 24に展開する(S40)。

【0071】

続いてCPU11, 21は、キャッシュメモリ14, 24に展開したシステム情報に基づいて差分管理ボリューム LU_{DM} のLUNを検出し(S41)、この後この検出したLUNの論理ボリュームLU、すなわち差分管理ボリューム LU_{DM} からスナップショット差分管理情報(プライマリボリュームアドレステーブル51及び差分情報テーブル52)を読み出して、これをキャッシュメモリ14, 24に展開する(S42)。

【0072】

この際、CPU11, 21は、差分管理ボリューム LU_{DM} が2つ設定されている場合には、図15に示すように、初期時には、これら2つの差分管理ボリューム LU_{DM} のうちの最初に設定された方の差分管理ボリューム LU_{DM} (以下、これをマスター側の差分管理ボ

10

20

30

40

50

リウム LU_{DM} と呼ぶ) からスナップショット差分管理情報を読み出すものの、退避時の書込みエラーによりデータが壊れているなど、なんらかの原因によりその読み出しに失敗したときには、対象を最後に設定された方の差分管理ボリューム LU_{DM} (以下、これをミラー側の差分管理ボリューム LU_{DM} と呼ぶ) に切り替えて、この差分管理ボリューム LU_{DM} からスナップショット差分管理情報を読み出すようにディスクドライブ $DRV0 \sim DRV4$ を制御する。また、このとき $CPU11, 21$ は、ミラー側の差分管理ボリューム LU_{DM} からのスナップショット差分管理情報の読み出しを、読み出しに失敗した位置から再開するように対応するディスクドライブ $DRV0 \sim DRV4$ を制御する。

【0073】

さらに $CPU11, 21$ は、ミラー側の差分管理ボリューム LU_{DM} からスナップショット差分管理情報を読み出ししている場合に、その読み出しに失敗したときには、対象をマスター側の差分管理ボリューム LU_{DM} に切り替えて、この差分管理ボリューム LU_{DM} からスナップショット差分管理情報を読み出すようにディスクドライブ 11 を制御する。この場合にも、 $CPU11, 21$ は、読み出しに失敗した位置から差分情報の読み出しを再開するように対応するディスクドライブ $DRV0 \sim DRV4$ を制御する。

【0074】

このように本ディスクアレイ装置 1 では、ユーザの都合に合わせてユーザが自由に差分管理ボリューム LU_{DM} を設定でき、この設定された差分管理ボリューム LU_{DM} に差分ビットマップ 50 を除くスナップショット差分管理情報を格納するようにしているため、例えば従来のディスクアレイ装置のように、スナップショット機能の購入の有無及びスナップショット機能の使用の有無にかかわらず、スナップショット差分管理情報を格納するための記憶容量をシステム領域 $SAR_0 \sim SAR_4$ 内に予め固定的に確保しておく場合に比べて、有効かつ効率的にディスクドライブ $DRV0 \sim DRV4$ の記憶容量 (記憶資源) を使用することができる。

【0075】

また本ディスクアレイ装置 1 では、このような差分管理ボリューム LU_{DM} を 2 つ設定でき、 2 つの差分管理ボリューム LU_{DM} が設定されたときには、キャッシュメモリ $14, 24$ に保持されたスナップショット差分管理情報を各差分管理ボリューム LU_{DM} のそれぞれに退避させるようにしているため、退避時の書込みエラーなどの原因により一方の差分管理ボリューム LU_{DM} に格納されたスナップショット差分管理情報に不具合が発生した場合においても、実用上十分に対応することができる。

【0076】

なお本ディスクアレイ装置 1 の場合、スナップショット差分管理情報を差分管理ボリューム LU_{DM} やシステム領域 $SAR_0 \sim SAR_4$ から読み出してキャッシュメモリ $14, 24$ 上に展開した後に、そのキャッシュメモリ $14, 24$ におけるスナップショット差分管理情報が格納された記憶領域にアクセスするための専用のウィンドが予め設けている。

【0077】

すなわち、通常、キャッシュメモリ $14, 24$ へのアクセス時には、キャッシュメモリ $14, 24$ の記憶領域全体を一度に参照することができないため、図 16 に示すように、キャッシュメモリ $14, 24$ におけるデータを読み出そうとする領域 M_{AR} のみを取り囲むようにウィンド W が設定され、その領域 M_{AR} にのみ排他的にアクセスできるようにする処理が行われる。

【0078】

この場合において、あるアクセスに対してウィンド W を設定する際には、先行する排他的な設定を削除後、ウィンド W を前のアクセスに対する位置から今回のアクセスに対する位置に切り替えるなどの処理が必要となる。このため、かかる処理によりオーバーヘッドが発生して、迅速なメモリアクセスを行い得ない場合がある。

【0079】

そこで本ディスクアレイ装置 1 では、キャッシュメモリ $14, 24$ 上の頻繁にアクセスされるスナップショット差分管理情報が格納される固定的な領域 $M_{AR(DM)}$ に対して専用の

10

20

30

40

50

ウインド W_{DM} を予め設定しておき、他の領域にウインド W が設定されている場合においても、専用のウインド W_{DM} を用いて当該スナップショット差分管理情報が格納された領域 $M_{AR(DM)}$ に対して排他的にアクセスできるようにすることで、ウインド切り替えに伴うオーバーヘッドを削減し、かかる領域 $M_{AR(DM)}$ に対する迅速なアクセスを可能としている。

【0080】

なお、このように専用のウインド W_{DM} を設定できるようにする領域 M_{AR} は、キャッシュメモリ14、24におけるスナップショット差分管理情報が格納される領域 $M_{AR(DM)}$ に限らず、頻繁にアクセスされる他の固定的な領域 M_{AR} についても同様に専用のウインド W を設定できるようにしても良い。

【0081】

(2) 第2の実施の形態

図1において70は、第2の実施の形態によるディスクアレイ装置を示す。このディスクアレイ装置70は、差分管理ボリューム LU_{DM} の設定方法が異なる点を除いて第1の実施の形態によるディスクアレイ装置1と同様に構成されている。

【0082】

すなわち、第1の実施の形態によるディスクアレイ装置1では、図17に示すように、1つのコントローラ10、20あたりのキャッシュメモリ14、24の容量に対して、スナップショット差分管理情報を格納するために使用する容量が予め定められている。この図17からも明らかなように、1つのコントローラ10、20あたりのキャッシュメモリ14、24の容量が1024～8192〔Mbyte〕である場合、スナップショット差分管理情報を格納するために使用できる容量は512～4096〔Mbyte〕である。

【0083】

このため第1の実施の形態では、図7～図10について上述した設定画面60等を用いて差分管理ボリューム LU_{DM} に設定する論理ボリューム LU を選択する際には、キャッシュメモリ14、24におけるスナップショット差分管理情報を格納するために使用する容量の最大値が4096〔MB〕であることを考慮して、その最低限度の記憶容量が5〔Gbyte〕であることが差分管理ボリューム LU_{DM} として選択できる論理ボリューム LU の条件となっていた。

【0084】

このような構成によれば、差分管理ボリューム LU_{DM} を設定する際に、ユーザがディスクアレイ装置1における各コントローラ10、20のキャッシュメモリ14、24の容量を気にすることなく、固定的に5〔Gbyte〕以上の容量を有する論理ボリューム LU を選択すればよい。そのため、差分管理ボリューム LU_{DM} の設定作業を容易化させ得るという利点がある。しかしながら、その反面、このような構成によると、差分管理ボリューム LU_{DM} として512〔Mbyte〕の容量しか必要がない場合にも固定的に5〔Gbyte〕以上の論理ボリューム LU を選択しなければならないため、その差分だけ容量が無駄になるという問題がある。

【0085】

そこで本実施の形態によるディスクアレイ装置70では、差分管理ボリューム LU_{DM} を設定する際、各コントローラ10、20のCPU71、72が当該ディスクアレイ装置70における各コントローラ10、20のキャッシュメモリ14、24の容量を調査し、その調査結果に基づいて最適な容量を有する論理ボリューム LU を差分管理ボリューム LU_{DM} の候補としてユーザに提示するようになされた点を特徴としている。

【0086】

図18は、このような差分管理ボリューム LU_{DM} の設定処理に関する本ディスクアレイ装置70内のCPU71、72の処理手順を示すフローチャートである。

【0087】

CPU71、72は、ユーザがホスト装置3を操作して差分管理ボリューム LU_{DM} を設定するための所定のコマンドを入力すると、まず、所定のディスクドライブDRV0～DRV4のシステム領域 SAR_0 ～ SAR_4 から読み出されてキャッシュメモリ14、24に

10

20

30

40

50

展開されたシステム情報に基づいて、各コントローラ10、20内のキャッシュメモリ14、24の容量を検出する(S50)。また、CPU71、72は、かかるシステム情報に基づいて、未使用の論理ボリュームLU及びその論理ボリュームLUの容量をそれぞれ検出する(S51)。

【0088】

続いてCPU71、72は、これらの検出結果に基づいて、差分管理ボリュームLU_{DM}の候補となり得る論理ボリュームLU、つまりコントローラ10、20におけるキャッシュメモリ14、24の容量よりも予め定められた所定量(例えば数Mbyte～数百Mbyte程度)以内の大きさの容量を有する未使用の論理ボリュームLUが存在するか否かを判定する(S52)。

10

【0089】

そしてCPU71、72は、この判定により差分管理ボリュームLU_{DM}の候補となり得る論理ボリュームLUが存在しないと判定したときには(S52:NO)、ホスト装置3に対してエラーメッセージを送信し、この後この一連の処理を終了する。かくして、このときそのホスト装置3のディスプレイに対応するエラーメッセージが表示される。

【0090】

これに対して、CPU71、72は、この判定により差分管理ボリュームLU_{DM}の候補となり得る論理ボリュームLUが存在すると判定したときには、対応する全ての論理ボリュームLUについての各種情報をホスト装置3に送信することにより、例えば図19に示すようなセレクトロジカルユニットダイアログ80をホスト装置3のディスプレイに表示させる(S54)。

20

【0091】

このセレクトロジカルユニットダイアログ80には、差分管理ボリュームLU_{DM}の候補となり得る全ての論理ボリュームLUについて、LUNや容量などの情報が一覧表示される。またこのセレクトロジカルユニットダイアログ80では、差分管理ボリュームLU_{DM}の候補となり得る全ての論理ボリュームLUのうち、その容量がコントローラ10、20内のキャッシュメモリ14、24の容量に対して必要となる容量よりも大きくかつ当該容量に最も近い容量の論理ボリュームLUが強調表示される。

【0092】

そしてCPU71、72は、セレクトロジカルユニットダイアログ80に掲載した差分管理ボリュームLU_{DM}の候補となり得る全ての論理ボリュームLUのうちのいずれかの論理ボリュームLUが選択されるのを待ち受け(S55)、やがてホスト装置3からセレクトロジカルユニットダイアログ80に一覧表示された論理ボリュームLUのうちの1つの論理ボリュームLUを選択する操作が入力された旨のコマンドが入力されると(S55:YES)、これに応動してその選択された論理ボリュームLUを差分管理ボリュームLU_{DM}に設定する(S56)。この後CPU71、72は、かかる一連の処理を終了する。

30

【0093】

これに対してCPU71、72は、セレクトロジカルユニットダイアログ80に掲載した論理ボリュームLUのいずれも選択されなかったときには(S55:NO)、その段階で管理ボリュームLU_{DM}を設定することなく、この一連の処理を終了する。従って、この場合には、第1の実施の形態において上述した手順に従ってユーザ自身で差分管理ボリュームLU_{DM}となり得る論理ボリュームLUを探すこととなる。

40

【0094】

このように本実施の形態によるディスクアレイ装置70では、差分管理ボリュームLU_{DM}を設定する際、各コントローラ10、20内のキャッシュメモリ14、24の容量を調査し、その調査結果に基づいて最適な容量を有する論理ボリュームLUを差分管理ボリュームLU_{DM}の候補としてユーザに提示するようになされているため、差分管理ボリュームLU_{DM}の設定が容易であり、例えばスナップショット機能の初期設定作業を簡易化させることができる。

【0095】

50

またこの場合において、本ディスクアレイ装置 70 は、差分管理ボリューム LU_{DM} の候補として、キャッシュメモリ 14, 24 の容量よりも予め定められた所定量以内の大きさの容量を有する未使用の論理ボリューム LU を提示するため、差分管理ボリューム LU_{DM} としてキャッシュメモリ 14, 24 の容量に応じた最適な容量の論理ボリューム LU を提示することができ、この結果キャッシュメモリ 14, 24 の容量に比して無駄に容量の大きい論理ボリュームが差分管理ボリューム LU_{DM} として設定されるという不具合を未然にかつ有効に防止することができる。

【0096】

(3) 他の実施の形態

なお上述の第 1 及び第 2 の実施の形態においては、ミラー側の差分管理ボリューム LU_{DM} を 1 つだけ設定（つまり差分管理ボリューム LU_{DM} を 2 つだけ設定）できるようにした場合について述べたが、本発明はこれに限らず、ミラー側の差分管理ボリューム LU_{DM} を 2 つ以上設定（つまり差分管理ボリューム LU_{DM} を 3 つ以上設定）できるようにしても良い。

【0097】

また上述の第 2 の実施の形態においては、差分管理ボリューム LU_{DM} の候補の論理ボリューム LU をユーザに提示する手法として、図 19 のようなセレクトロジカルユニットダイアログ 80 をホスト装置 3 のディスプレイに表示させるようにした場合について述べたが、本発明はこれに限らず、提示手法としては、この他種々の方法を広く適用することができる。

【産業上の利用可能性】

【0098】

本発明は、ディスクアレイ装置に適用することができる。

【図面の簡単な説明】

【0099】

【図 1】本実施の形態によるディスクアレイ装置の構成を示すブロック図である。

【図 2】論理ボリュームの説明に供する概念図である。

【図 3】本実施の形態によるスナップショット差分管理情報の管理方式の説明に供する概念図である。

【図 4】本実施の形態によるスナップショット差分管理情報の管理方式の説明に供する概念図である。

【図 5】差分情報テーブルの説明に供する概念図である。

【図 6】差分管理ボリュームの設定手順の説明に供するフローチャートである。

【図 7】設定画面を示す図である。

【図 8】セレクトロジカルユニットダイアログを示す図である。

【図 9】確認メッセージダイアログを示す図である。

【図 10】差分管理ボリューム設定後の設定画面を示す図である。

【図 11】スナップショット作成指示受信後の CPU の処理手順を示すフローチャートである。

【図 12】スナップショット作成処理に関する CPU の処理手順を示すフローチャートである。

【図 13】スナップショット差分管理情報の退避に関する CPU の処理手順を示すフローチャートである。

【図 14】スナップショット差分管理情報の読み出しに関する CPU の処理手順を示すフローチャートである。

【図 15】スナップショット差分情報の読出し障害時における CPU の処理内容の説明に供する概念図である。

【図 16】キャッシュメモリの固定領域に設定される専用のウインドの説明に供する概念図である。

【図 17】キャッシュメモリの容量とスナップショット差分管理情報のデータ量と関係の

10

20

30

40

50

説明に供する図表である。

【図 1 8】第 2 の実施の形態によるディスクアレイ装置における差分管理ボリュームの設定に関する C P U の処理手順を示すフローチャートである。

【図 1 9】差分管理ボリュームの候補の一覧表示の様子を示す図である。

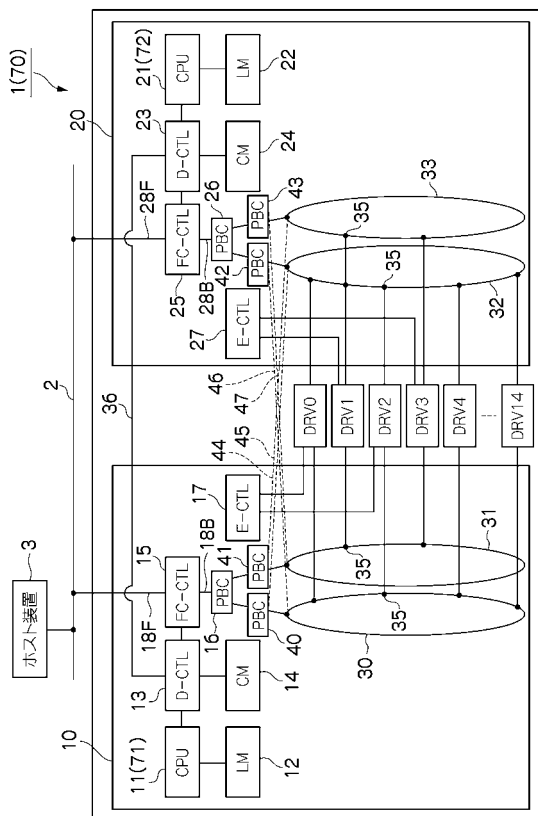
【符号の説明】

【 0 1 0 0 】

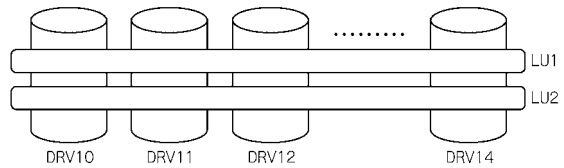
1, 70 …… ディスクアレイ装置、3 …… ホスト装置、10, 20 …… コントローラ、11, 21 …… C P U、12, 22 …… ローカルメモリ、14, 24 …… キャッシュメモリ、50 …… 差分ビットマップ、51 …… プライマリボリュームアドレステーブル、52 …… 差分情報テーブル、60 …… 設定画面、62, 80 …… セレクトロジカルユニットダイアログ、DRV0 ~ DRV14 …… ディスクドライブ、論理ボリューム …… L U, L U₁, L U₂ …… 論理ボリューム、L U_{DN} …… 差分管理ボリューム、L U_p …… プライマリボリューム、L U_{POOL} …… プールボリューム、M_{AR}, M_{AR(DM)} …… 領域、W, W_{DM} …… ウインド。

10

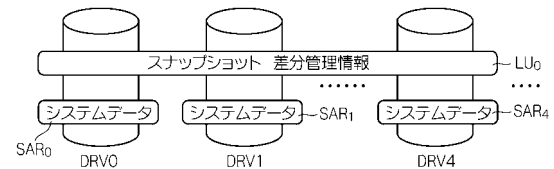
【図 1】



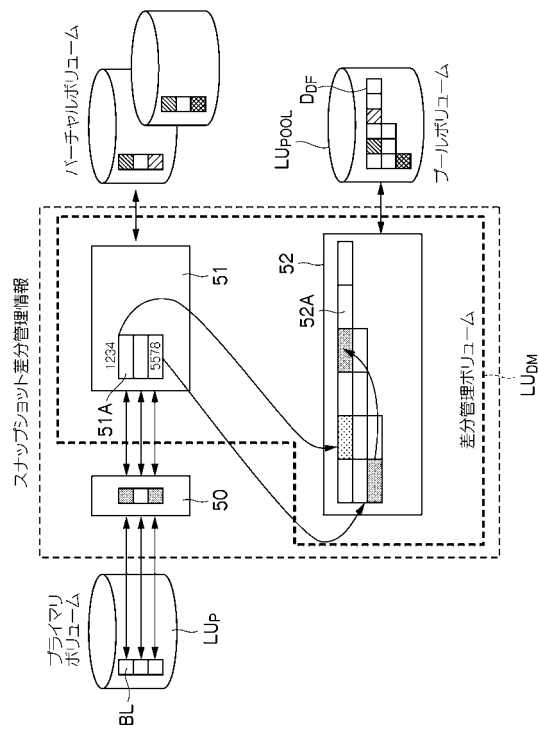
【図 2】



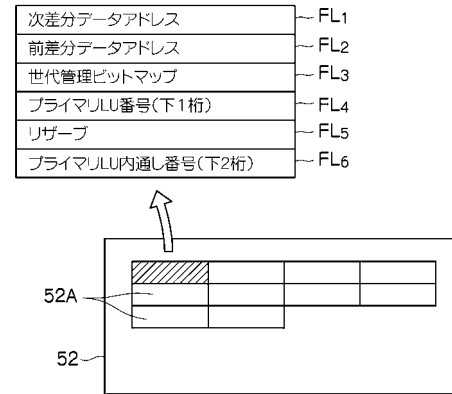
【図 3】



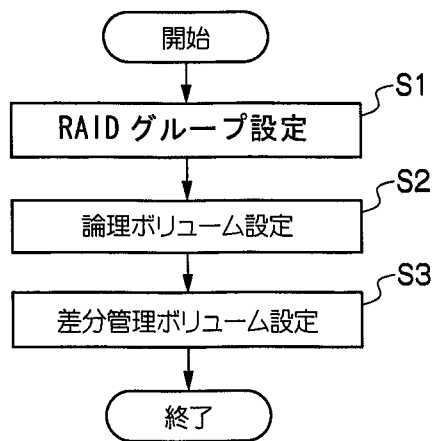
【図 4】



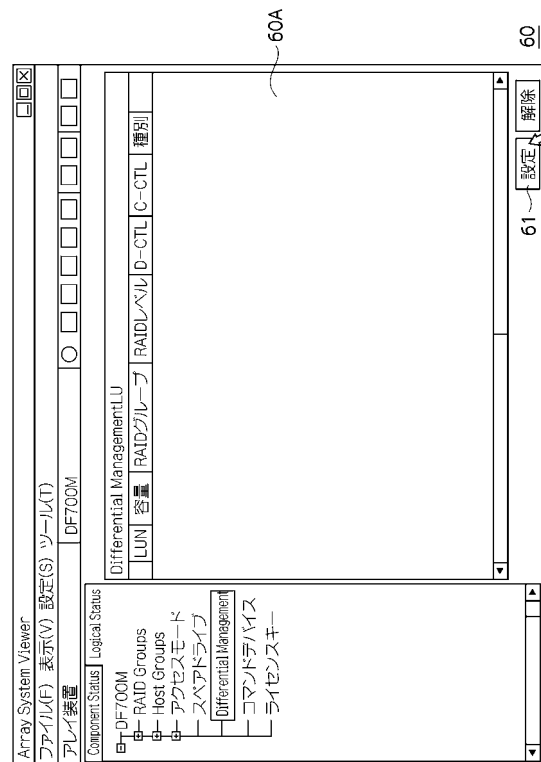
【図 5】



【図 6】



【図 7】



【図 8】

62

62A

63

OK

キャンセル

LUN	容量	RAIDグループ	RAIDレベル	D-CTL	C-CTL	種別
0000	2.0MB	00	RAID5(4D+1P)	0	0	FC Normal
0200	33.0MB	00	RAID5(4D+1P)	0	0	FC Normal
0300	5.0GB	00	RAID5(4D+1P)	0	0	FC Normal

【図 9】

65

64

DM-LUを設定します。よろしいですか?

OK

キャンセル

【図 10】

60A

60

Array System Viewer

ファイル(F) 表示(V) 設定(S) ツール(T)

アレキ装置

DF700M

Component Status

Logical Status

Differential Management LU

RAID Groups

Host Groups

アクセスマード

スベアドライブ

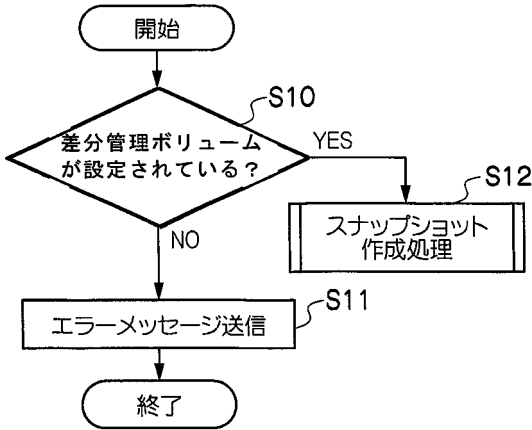
Differential Management

コマンドデバイス

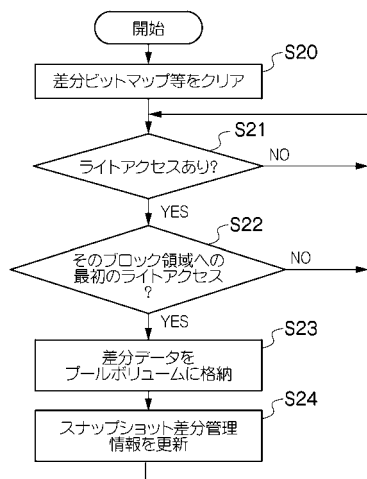
ライセンスキー

LUN	容量	RAIDグループ	RAIDレベル	D-CTL	C-CTL	種別
0300	5.0GB	00	RAID5(4D+1P)	0	0	FC Normal

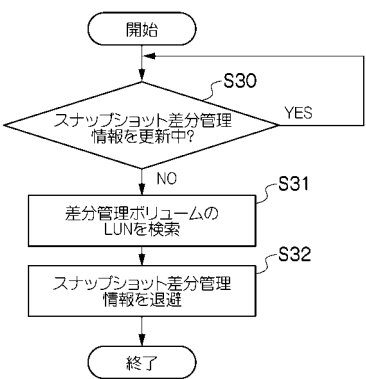
【図 11】



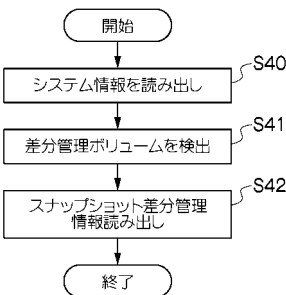
【図 1 2】



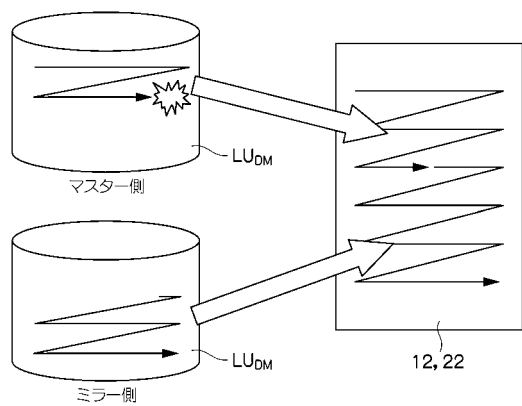
【図 1 3】



【図 1 4】



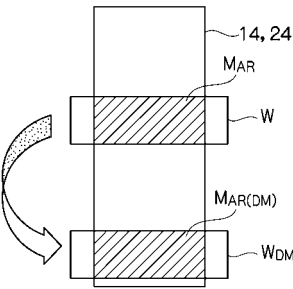
【図 1 5】



【図 1 7】

順番	機種	1CTLあたりの メモリ搭載量 (MB)	スナップショット差分 管理情報使用量 (MB)
1	A	1,024	512
2	B	1,024	512
3		2,048	1,024
4			
5		3,072	1,536
6			
7		4,096	2,048
8	C	2,048	1,024
9		4,096	2,048
10		6,144	3,072
11			
12		8,192	4,096

【図 1 6】



フロントページの続き

- (56)参考文献 特開2005-085117 (JP, A)
特開2005-174196 (JP, A)
特開2000-081921 (JP, A)
特開2005-141335 (JP, A)

- (58)調査した分野(Int.Cl., DB名)
G06F 3/06