

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6089527号
(P6089527)

(45) 発行日 平成29年3月8日(2017.3.8)

(24) 登録日 平成29年2月17日(2017.2.17)

(51) Int.Cl.	F I
G 0 6 F 3/06 (2006.01)	G 0 6 F 3/06 3 0 5 A
	G 0 6 F 3/06 3 0 1 H
	G 0 6 F 3/06 3 0 4 F
	G 0 6 F 3/06 3 0 6 K

請求項の数 7 (全 33 頁)

(21) 出願番号	特願2012-205012 (P2012-205012)	(73) 特許権者	000005223
(22) 出願日	平成24年9月18日 (2012. 9. 18)		富士通株式会社
(65) 公開番号	特開2014-59769 (P2014-59769A)		神奈川県川崎市中原区上小田中4丁目1番1号
(43) 公開日	平成26年4月3日 (2014. 4. 3)	(74) 代理人	100074099
審査請求日	平成27年5月12日 (2015. 5. 12)		弁理士 大菅 義之
		(74) 代理人	100133570
			弁理士 ▲徳▼永 民雄
		(72) 発明者	松村 忠
			神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
		(72) 発明者	矢須 紀之
			神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

最終頁に続く

(54) 【発明の名称】 ストレージ制御システム、リカバリプログラム、及び該方法

(57) 【特許請求の範囲】

【請求項 1】

所定のサイズ毎に区分された複数の区画で形成され且つ情報を格納する第1格納領域と、
前記複数の区画の各区画に格納される情報が有効であるか否かを示す情報とリカバリが
必要であるか否かを示す情報を含む属性情報を含む管理情報を格納する第2格納領域と、
を含む複数の格納部と、

前記複数の格納部から前記管理情報を取得する取得部と、

取得した前記管理情報を比較して異常な前記管理情報があるかを判定する判定部と、

前記異常な管理情報があると判定された場合、正常な管理情報の区画毎の属性情報に基づいて、該正常な管理情報が取得されたいずれかの前記格納部から、前記異常な管理情報が取得された前記格納部の前記第1格納領域に格納される情報を、区画単位で再構成するリカバリ部と、

を備えることを特徴とするストレージ制御システム。

【請求項 2】

前記ストレージ制御システムは、さらに、

前記格納部への対象情報の書き込み要求があった場合、前記複数の格納部において、前記第1格納領域に該対象情報を書き込み、該対象情報を書き込んだ区画に格納される情報に対応する属性情報に、該書き込んだ区画の情報が有効であるか否かを示す情報を書き込む書き込み部

を備え、

前記リカバリ部は、前記正常な管理情報が取得されたいずれかの前記格納部から、前記区画に格納された情報が有効であることを示す属性情報を有する該区画に格納された情報を抽出し、前記異常な管理情報が取得された格納部の前記第 1 格納領域における対応する区画に、該抽出した情報を格納する

ことを特徴とする請求項 1 に記載のストレージ制御システム。

【請求項 3】

前記書き込み部は、前記書き込みが失敗した格納部がある場合、該書き込みが成功した格納部の管理情報における、前記書き込み要求によるアクセス範囲に対応する前記区画に格納された情報の前記属性情報に、リカバリが必要であることを示す情報を書き込み、

前記リカバリ部は、前記書き込みが成功した前記格納部から、前記リカバリが必要であることを示す属性情報を有する該区画に格納された情報を抽出し、前記書き込みが失敗した格納部の前記第 1 格納領域における対応する区画に、該抽出した情報を格納する

ことを特徴とする請求項 2 に記載のストレージ制御システム。

【請求項 4】

前記リカバリが必要であることを示す属性情報は、前記書き込みが失敗した格納部を識別する情報である

ことを特徴とする請求項 3 に記載のストレージ制御システム。

【請求項 5】

前記属性情報は、前記区画に格納された情報の種別情報を含み、

前記書き込み部は、前記書き込み対象情報の書き込み要求があった場合であって、該書き込み要求によるアクセス範囲に対応する前記区画に格納された情報の属性情報に含まれる前記種別情報がメタデータを示す場合、前記書き込み対象情報を前記格納部の前記第 1 格納領域とバックアップ領域とに格納し、前記種別情報がメタデータを示さない場合、書き込み対象情報を前記第 1 格納領域に格納する

ことを特徴とする請求項 2 ～ 4 のうちいずれか 1 項に記載のストレージ制御システム。

【請求項 6】

コンピュータに、

所定のサイズ毎に区分された複数の区画で形成され且つ情報を格納する第 1 格納領域と、前記複数の区画の各区画に格納される情報が有効であるか否かを示す情報とリカバリが必要であるか否かを示す情報を含む属性情報を含む管理情報を格納する第 2 格納領域と、を含む、複数の格納部から前記管理情報を取得し、

取得した前記管理情報を比較して異常な前記管理情報があるかを判定し、

前記異常な管理情報があると判定された場合、正常な管理情報の区画毎の属性情報に基づいて、該正常な管理情報が取得されたいずれかの前記格納部から、前記異常な管理情報が取得された前記格納部の前記第 1 格納領域に格納される情報を、区画単位で再構成する処理を実行させることを特徴とするリカバリプログラム。

【請求項 7】

コンピュータが、

所定のサイズ毎に区分された複数の区画で形成され且つ情報を格納する第 1 格納領域と、前記複数の区画の各区画に格納される情報が有効であるか否かを示す情報とリカバリが必要であるか否かを示す情報を含む属性情報を含む管理情報を格納する第 2 格納領域と、を含む、複数の格納部から前記管理情報を取得し、

取得した前記管理情報を比較して異常な前記管理情報があるかを判定し、

前記異常な管理情報があると判定された場合、正常な管理情報の区画毎の属性情報に基づいて、該正常な管理情報が取得されたいずれかの前記格納部から、前記異常な管理情報が取得された前記格納部の前記第 1 格納領域に格納される情報を、区画単位で再構成する処理を実行することを特徴とするリカバリ方法。

【発明の詳細な説明】

【技術分野】

【0001】

10

20

30

40

50

本発明は、ストレージ装置の制御技術に関する。

【背景技術】

【0002】

ストレージ装置とは、複数のディスク装置を搭載したディスク記憶システムである。ストレージ装置は、ホストから送信されたデータをディスク装置に書き込むためのコントローラモジュール（ＣＭ：Controller Module、以下では「ＣＭ」と呼ぶ。）を有している。

【0003】

ストレージ装置には、ＣＭ内外にＢＵＤ（Back up Utility Device）を搭載している。ＢＵＤの一例として、ＨＤＤ（Hard Disk Drive）、ＳＳＤ（Solid State Drive）、ＵＳＢ（Universal Serial Bus）メモリ装置が使用されている。ＢＵＤには、装置ログやファームウェアアーカイブなど装置内のレコードが格納されている。

10

【0004】

ＢＵＤ上のデータは、利用目的毎（システム機能の要件）に必要な応じて、複数のＢＵＤによって冗長化されている。ＢＵＤ間のデータの整合性と冗長性の保持は、ＣＭの等価制御部によって実現される。その結果、ＢＵＤ間のデータは、相互に等価を保持することができる。

【0005】

記憶装置のリカバリ技術として次の技術がある。当該技術では、ＯＳ（Operating System）が起動した時に、サービスプログラムは、記憶装置の制御部に対してメタデータ記録・判断・修復部をロードする。制御部がデータロスエラーを検出した場合、メタデータ記録・判断・修復部は、次の処理を行う。すなわち、メタデータ記録・判断・修復部は、データ記憶部のデータロスエラーのエラー発生アドレスおよびメタデータバックアップ部から読み出したメタデータに基づいてエラー発生アドレスがメタデータ領域にあるかファイルデータ領域にあるかを判断する。メタデータ記録・判断・修復部は、エラー発生アドレスがメタデータ領域にあればメタデータバックアップ部のメタデータ領域バックアップデータをデータ記憶部のメタデータ領域に修復する。

20

【先行技術文献】

【特許文献】

【0006】

30

【特許文献１】特開２００５－１４９２４８号公報

【発明の概要】

【発明が解決しようとする課題】

【0007】

しかしながら、上記技術は、記憶装置単体に関するリカバリ技術であり、複数の記憶装置間のデータの整合性を考慮したリカバリシステムに適応したリカバリ技術ではない。そのため、冗長化したシステムにおいて複数の記憶装置間のデータで整合性がない場合、当該技術を用いてシステムを復旧させようすると、かなりの時間を要することになる。

【0008】

本発明の一側面では、記憶装置間のデータの整合性を保持しつつ、効率的なリカバリを行うストレージ制御技術を提供する。

40

【課題を解決するための手段】

【0009】

本実施形態の一側面に係るストレージ制御システムは、複数の格納部、取得部、判定部、リカバリ部を含む。格納部は、第１格納領域と、第２格納領域とを含む。第１格納領域は、情報を格納する。第２格納領域は、第１格納領域が区分された各区画に格納される情報を示す区画情報の属性に関する属性情報を含む管理情報を格納する。取得部は、複数の格納部から管理情報を取得する。判定部は、取得した管理情報を比較して異常な管理情報があるかを判定する。リカバリ部は、異常な管理情報があると判定された場合、正常な管理情報の区画毎の属性情報に基づいて、正常な管理情報が取得されたいずれかの格納部か

50

ら、異常な管理情報が取得された格納部の第 1 格納領域に格納される情報を、区画単位で再構成する。

【発明の効果】

【0010】

本発明によれば、記憶装置間のデータの整合性を保持しつつ、効率的なりカバリを行うことができる。

【図面の簡単な説明】

【0011】

【図 1】BUD間のデータの整合性を保持する等価制御について説明するための図である。

10

【図 2】本実施形態における、ストレージ制御システムの一例を示す。

【図 3】本実施形態におけるストレージシステムのハードウェア構成を示す。

【図 4】本実施形態におけるCMの機能ブロック図を示す。

【図 5】本実施形態におけるユーザ領域の一例を示す。

【図 6】本実施形態における等価管理データのデータ構成を示す。

【図 7】本実施形態におけるブロック属性マッピング情報を説明するための図である。

【図 8】本実施形態におけるブロック属性情報のデータ構造を示す。

【図 9】本実施形態におけるストレージ装置の初期起動時の等価制御部の動作フローを示す。

【図 10】本実施形態におけるストレージ装置の初期起動時のCM間の動作シーケンスの全体図である。

20

【図 11A】本実施形態におけるBUDの初期化処理(S34)の詳細なシーケンス図(その1)である。

【図 11B】本実施形態におけるBUDの初期化処理(S34)の詳細なシーケンス図(その2)である。

【図 12A】本実施形態におけるBUD間の等価処理(S36)の詳細なシーケンス図(その1)である。

【図 12B】本実施形態におけるBUD間の等価処理(S36)の詳細なシーケンス図(その2)である。

【図 12C】本実施形態におけるBUD間の等価処理(S36)の詳細なシーケンス図(その3)である。

30

【図 13】本実施形態における各機能制御部からBUDへ読み出し処理を行う場合の等価制御部の動作フローを示す。

【図 14】本実施形態における機能制御部からBUDへ読み出し処理を行う場合のCM間の動作シーケンスの全体図である。

【図 15】本実施形態における読み出し処理(S116)の詳細なシーケンス図である。

【図 16A】本実施形態における各機能制御部からBUDへ書込み処理を行う場合の等価制御部の動作フロー(その1)を示す。

【図 16B】本実施形態における各機能制御部からBUDへ書込み処理を行う場合の等価制御部の動作フロー(その2)を示す。

40

【図 17】本実施形態における書込み処理シーケンスの全体図である。

【図 18A】本実施形態における書き込み処理(S156)の詳細なシーケンス図(その1)である。

【図 18B】本実施形態における書き込み処理(S156)の詳細なシーケンス図(その2)である。

【図 19】本実施形態におけるリカバリ処理を行う場合の等価制御部の動作フローを示す。

【発明を実施するための形態】

【0012】

図 1 は、BUD間のデータの整合性を保持する等価制御について説明するための図であ

50

る。CM#0(1a)、CM#(1b)1は、BUD9a-BUD9b間のデータの整合性を保持する。BUD9a,9bは、記憶領域(ユーザ領域)12a,12bと、等価管理領域10a,10bと、バックアップ領域14a,14bとを有する。記憶領域(ユーザ領域)12a,12bは、CM1a,1bの各機能(機能#A、#B、#C・・・)がアクセスして使用する領域であり、実データ13a,13bを格納する。等価管理領域10a,10bと、バックアップ領域14a,14bは、等価制御部3a,3bが使用するための記憶領域である。等価管理領域10a,10bには、BUD9a-BUD9b間でデータの整合性を保持するための等価管理データ11a,11bが格納される。バックアップ領域14a,14bには、データの書き込み中の不良のファーム機能停止(電源断など)に考慮して、ユーザ領域、等価管理領域への書き込みのためにデータが一時的にバックアップされる。

10

【0013】

図1では、BUDアクセス機能#A(以下、機能#Aと称する)が使用する記憶領域については、符号「#A」が付与されている。CM#0(1a)、CM#1(1b)は同一の構成を有するので、以下ではCM#0(1a)について説明し、CM#1(1b)についてはその説明を省略する。

【0014】

CM#0(1a)は、機能制御部2a、等価制御部3a、BUDアクセス制御部7a、BUDドライバ8a、BUD#0(9a)を含む。機能制御部2aは、CMの機能(機能A~N)ごとに存在し、図1では、一例として、BUDアクセス機能A#の機能制御部が示されている。BUDアクセス機能#Aの機能制御部2aは、BUDアクセス機能#Aを実現させるために、BUD9aへのアクセスを実行させる制御部である。

20

【0015】

等価制御部3aは、等価管理領域10aにおいて、BUD9a-BUD9b間でデータの冗長性を管理する。具体的には、等価制御部3aは、等価管理データを用いて、各機能が使用するユーザ領域の実データ13aが、BUD9a-BUD9b間で同じになるように管理している。等価制御部3aは、バックアップ領域14aに実データ13aを一時的にバックアップし、その実データを用いて書き込み処理をリトライする。等価制御部3aは、ライト制御部4a、ライトサブ制御部5a、等価チェック制御部6aを含む。

【0016】

ライト(write)制御部4aは、BUD9a,9bに対する、機能#A~#Nからのデータの書き込み要求を制御するメインルーチンである。ライトサブ(write sub)制御部5aは、各制御部から指定されたBUD9aの記憶領域にデータを書き込むサブルーチンである。

30

【0017】

等価チェック制御部6aは、機能#A~#Nで使用する記憶領域に保持する実データ13a,13bがBUD9a-BUD9b間で等価状態にあるかチェックする。等価状態がないBUDが存在する場合は、等価チェック制御部6aは、等価状態となるようにリカバリ処理を動作させる。

【0018】

BUDアクセス制御部7aは、アクセス先のアドレスについて、論理アドレス-物理アドレス変換を行う。BUDドライバ8aは、BUD9a-BUD9b間の通信を行うドライバである。CM間通信ドライバ16aは、他CM1bとの通信を行うドライバである。

40

【0019】

BUD9aは、各機能に対応する、等価管理領域(10a)、ユーザ領域(12a)、バックアップ領域(14a)を有する。図1では、一例として、機能#Aに対応する等価管理領域#A(10a)、ユーザ領域#A(12a)、バックアップ領域#A(14a)が記載されている。

50

【 0 0 2 0 】

図 1 において、機能 # A (2 a) から B U D 9 a , 9 b のユーザ領域 # A への書き込み要求を受けると (S 1)、等価制御部 3 a は、B U D 9 a , 9 b において機能 # A に対応する等価管理領域 # A から等価管理データを読み出す。すなわち、機能 # A から等価制御部 3 a へ書き込み要求があった場合 (S 1)、ライト制御部 4 a は等価チェック制御部 6 a に等価チェック要求を行う (S 2)。等価チェック制御部 6 a は、B U D アクセス制御部 7 a、B U D ドライバ 8 a を介して、B U D # 0 (9 a) の等価管理領域 # A (1 0 a) から等価管理データ # A (1 1 a) を読み出す (S 3 , S 4)。また、等価チェック制御部 6 a は、B U D アクセス制御部 7 a、B U D ドライバ 8 a、C M 間通信ドライバ 1 6 a , 1 6 b、B U D ドライバ 8 b を介して、B U D # 1 (9 b) の等価管理領域 # A から等価管理データ # A (1 1 b) を読み出す (S 3 , S 4)。

10

【 0 0 2 1 】

次に、等価チェック制御部 6 a は、B U D 9 a , 9 b の等価管理データを比較して、異常がないかをチェックする。異常な等価管理データがある場合、等価チェック制御部 6 a は、そのチェック結果をライト制御部 4 a へ送信する (S 5)。ライト制御部 4 a は、リカバリ制御部 (不図示) に異常な BUD に対して、正常な BUD のデータ (リカバリデータ) を書き込むように要求を行う (S 6)。

【 0 0 2 2 】

書き込みを行う場合、等価制御部 3 a , 3 b (ライトサブ制御部 5 a , 5 b) は、停電時などの即断に備えて、リカバリデータをバックアップ用データとしてバックアップ領域に書き込む (S 7、S 8)。バックアップ用データは、書き込み情報 (書き込み位置と書き込みサイズ)、等価管理データ、実データを含む。

20

【 0 0 2 3 】

それから、等価制御部 3 a , 3 b (ライトサブ制御部 5 a , 5 b) は、リカバリのために、実データの書き込み前に、等価管理データを無効化する (S 9 , S 1 0)。その後、等価制御部 3 a , 3 b (ライトサブ制御部 5 a , 5 b) は、実データを、ユーザ領域 # A の指定の位置へ書き込む (S 1 1、S 1 2)。等価制御部 3 a , 3 b (ライトサブ制御部 5 a , 5 b) は、実データの書き込み後に等価管理データを有効にする (S 1 3 , S 1 4)。

【 0 0 2 4 】

このように、C M 内の機能から B U D の領域へのデータの書き込み要求が発生すると、B U D に対して複数の書き込み処理が行われる。さらに、実データは、ユーザ領域及びバックアップ領域にそれぞれ書き込まれ、書き込み処理時間は単純に 2 倍以上の時間を要する。このため、S S D のような高速アクセスデバイスを使用するとしても、書き込み処理のオーバーヘッドは無視することができない。

30

【 0 0 2 5 】

また、C M 保守 (C M の動作確認検査、C M の組み込み作業) 時には、管理しているユーザ領域内の全データをコピーしているので、非常に時間がかかり、組み込み時間が長くなっている。

【 0 0 2 6 】

さらに、B U D を初めて使用する場合、初回の読み出し処理において、B I D (Block ID) エラーなど発生しないように、使用する領域について、データ「 0 」を書き込みして初期化している。したがって、B U D の容量が大きくなるにつれて、B U D 初期化に時間がかかり、組み込み時間が長くなっている。

40

【 0 0 2 7 】

等価制御部を介して C M 内の各機能が B U D へアクセスする前後では、ユーザ領域へのアクセス競合を防止するために、B U D のユーザ領域に対してアクセスのロック / アンロックを行って、排他制御が行われる。これにより、ロックされていないユーザ領域を使用する機能制御部が B U D にアクセスすることが可能になる一方で、ロックされているユーザ領域を使用する機能制御部は B U D にアクセスすることができない。したがって、B U

50

Dに対するアクセス競合を防止して、書き込み処理においてBUD間でデータの整合性を保持することができる。

【0028】

等価制御部は、機能制御部からのアクセス要求によるアクセス先の領域に格納された情報がBUD間で等価状態（データとして等しい状態）となるように制御する。等価制御部によるBUD間の等価状態チェックは、等価管理領域に格納されている等価管理データを照合して行われる。

【0029】

等価管理データは、該当するユーザ領域への実データの書き込み要求で更新されるものであり、更新時のタイムスタンプ情報を含む。等価管理データは、ユーザ領域へのデータの書き込み前に一旦無効状態にされ、書き込み処理が完了した時点で有効状態に戻される。これにより、等価管理データが有効のBUDの領域、またはタイムスタンプが新しいBUDの領域から、等価管理データが無効のBUDの領域またはタイムスタンプが古いBUDの領域へデータが等価制御部によりコピーされ、等価状態を維持するように制御される。

【0030】

ところが、書き込み処理中に停電やプログラムが即時停止する事象が発生した場合、各領域の等価管理データは無効状態のままとなる。BUDへの書き込み処理は、各CMで同時に行われるため、復旧時に有効状態の等価管理データが存在しなくなる場合、データがロストすることになる。

【0031】

その問題を解決するために、ユーザ領域へのデータの書き込み前に、バックアップ領域へ一度データの書き込みを行う。停電等のプログラム即時停止が発生しても、復電時に書き込み処理中と判断した場合、バックアップ領域のデータから書き込み処理をリトライする。

【0032】

これにより、各領域内のレコードの整合性を維持することができる。バックアップ処理は、いつ発生するか分からない事象に備えて、書き込み要求時に毎回実行されるので、書き込み処理性能とトレードオフの仕組みとなっている。

【0033】

また、停電が発生した際にメモリ上のデータを不揮発性デバイスにバックアップする仕組みがある。しかしながら、各機能からの同時アクセスでかつ最大アクセスサイズを考慮したメモリ空間を用意しなければならず、メモリ空間を圧迫するので、当該仕組みを利用することは難しい。

【0034】

また、等価制御部は、管理する領域内のデータ構造を知らないので、ユーザ領域内のデータはすべて有効であると判断する。このため、CM保守またはBUD保守において、保守されたBUD側にデータを他の正常なBUDからコピーする際に全データをコピーするしかなく、組み込み処理に時間を要する。

【0035】

さらに、コピー処理中にコピー元データの読み出しができなかった場合、その読み出しができない領域がたとえ未使用領域だとしても、コピー処理していた領域を全て破棄するしかなく、データロストとなる。

【0036】

上記課題に鑑み、本実施形態では、BUDへの書き込みアクセス時に効率的な書き込み処理を行うことにより、書き込み処理の向上及びCM組み込み等でのリカバリ処理速度の向上を図る。

【0037】

図2は、本実施形態における、ストレージ制御システムの一例を示す。ストレージ制御システム21は、複数の格納部22、取得部23、判定部24、リカバリ部25を含む。

【0038】

複数の格納部 2 2 はそれぞれ、第 1 格納領域と、第 2 格納領域とを含む。第 1 格納領域は、情報を格納する。第 2 格納領域は、管理情報を格納する。管理情報は、第 1 格納領域が区分された各区画に格納される情報を示す区画情報の属性に関する属性情報を含む。格納部 2 2 の一例として、B U D 6 4 が挙げられる。第 1 格納領域の一例として、ユーザ領域 6 6 が挙げられる。第 2 格納領域の一例として、等価管理領域 6 5 が挙げられる。管理情報の一例として、等価管理データが挙げられる。属性情報の一例として、ブロック属性情報が挙げられる。

【 0 0 3 9 】

取得部 2 3 は、複数の格納部 2 から管理情報を取得する。取得部 2 3 の一例として、等価チェック制御部 5 9 が挙げられる。

10

【 0 0 4 0 】

判定部 2 4 は、取得した管理情報を比較して異常な管理情報があるかを判定する。判定部 4 の一例として、等価チェック制御部 5 9 が挙げられる。

【 0 0 4 1 】

リカバリ部 2 5 は、異常な管理情報があると判定された場合、次の処理を行う。すなわち、リカバリ部 2 5 は、正常な管理情報の区画毎の属性情報に基づいて、正常な管理情報が取得されたいずれかの格納部 2 2 から、異常な管理情報が取得された格納部 2 2 の第 1 格納領域に格納される情報を、区画単位で再構成する。リカバリ部 2 5 の一例として、リカバリ制御部 5 5 が挙げられる。

【 0 0 4 2 】

20

このように構成することにより、記憶装置間のデータの整合性を保持しつつ、効率的なリカバリを行うことができる。すなわち、区画単位でのリカバリが可能になるので、リカバリに要する時間を短縮することができる。

【 0 0 4 3 】

ストレージ制御システム 2 1 は、さらに、書き込み部 2 6 を含む。書き込み部 2 6 は、格納部 2 2 への対象情報の書き込み要求があった場合、複数の格納部 2 2 の第 1 格納領域に対象情報を書き込む。このとき、書き込み部 2 6 は複数の格納部 2 2 の第 2 格納領域に格納された管理情報における対象情報を書き込んだ領域の区画に対応する属性情報に、区画情報が有効であるか否かを示す情報を書き込む。書き込み部 2 6 の一例として、ライト制御部 5 3 が挙げられる。この場合、リカバリ部 2 5 は、正常な管理情報が取得されたいずれかの格納部 2 2 から、区画情報が有効であることを示す属性情報を有する区画情報を抽出する。そして、リカバリ部 2 5 は、異常な管理情報が取得された格納部 2 2 の第 1 格納領域における対応する区画に、抽出した区画情報を格納する。

30

【 0 0 4 4 】

このように構成することにより、正常な等価管理データが取得された B U D から、有効なデータをブロック単位で抽出して、リカバリ対象 B U D の、対応するブロックに書き込むことができる。これにより、無効なデータの抽出及び書き込みを行う必要がないので、リカバリに要する時間を短縮することができる。

【 0 0 4 5 】

書き込み部 2 6 は、書き込みが失敗した格納部 2 2 がある場合、書き込みが成功した格納部 2 2 の管理情報における、書き込み要求によるアクセス範囲に対応する区画に格納された前記区画情報の前記属性情報に、リカバリが必要であることを示す情報を書き込む。

40

【 0 0 4 6 】

リカバリ部 2 5 は、書き込みが成功した格納部 2 2 から、リカバリが必要であることを示す属性情報を有する区画情報を抽出する。リカバリ部 2 5 は、書き込みが失敗した格納部 2 2 の前記第 1 格納領域における対応する区画に、抽出した区画情報を格納する。

【 0 0 4 7 】

このように構成することにより、リカバリを行う場合、ユーザ領域全域ではなく、リカバリが必要であることを示す属性情報を有するブロックについて、リカバリを行えばよいので、リカバリに要する時間を短縮することができる。

50

【 0 0 4 8 】

また、リカバリが必要であることを示す属性情報は、書き込みが失敗した格納部 2 2 を識別する情報である。このように構成することにより、リカバリ対象の B U D を特定することができる。

【 0 0 4 9 】

属性情報は、区画情報の種別情報を含む。書き込み部 2 6 は、前記書き込み対象情報の書き込み要求があった場合であって、書き込み要求によるアクセス範囲に対応する区画に格納された区画情報の属性情報に含まれる種別情報がメタデータを示す場合、次の処理を行う。すなわち、書き込み部 2 6 は、書き込み対象情報を格納部 2 2 の第 1 格納領域とバックアップ領域とに格納する。書き込み部 2 6 は、その種別情報がメタデータを示さない場合、書き込み対象情報を第 1 格納領域に格納する。

10

【 0 0 5 0 】

このように構成することにより、メタデータの書き込みの場合にバックアップを行うようにし、実データの書き込みの場合にはバックアップを行わないので、書き込み処理の時間を短縮することができる。

【 0 0 5 1 】

図 3 は、本実施形態におけるストレージシステムのハードウェア構成を示す。ストレージシステム 3 1 は、複数のディスク装置 3 9 を搭載したディスク記憶システムである。図 3 の例では、ホストコンピュータ（以下、ホストと称する）4 1 からディスク装置 3 9 へのアクセスに関して、冗長化のために、各ディスク装置 3 9 に対するアクセスパスが存在している。またデータ自体についても、R A I D（Redundant Arrays of Independent Disks）を用いて複数台のディスクにデータが分散されて、冗長化した状態で保存されている。

20

【 0 0 5 2 】

ストレージシステム 3 1 は、ホスト 4 1、ファイバチャネル（F C）スイッチ 4 2、コントローラエンクロージャ（C E）3 2、ドライブエンクロージャ（D E）4 0を含む。ストレージシステム 3 1 は、ドライブエンクロージャ（D E）4 0と、ホスト 4 1とが、F C スwitch 4 2、コントローラモジュール（C M）3 3 を介して接続されている。

【 0 0 5 3 】

ホスト 4 1 は、F C スwitch 4 2 を介して C E 3 2 と通信をして、D E 4 0 が有するディスク装置 3 9 からデータを読み出したり、ディスク装置 3 9 にデータを書き込んだりする。なお、図 3 では、1 つのホスト 4 1 を図示しているが、複数のホスト 4 1 が、C E 3 2 に接続されていてもよい。

30

【 0 0 5 4 】

C E 3 2 は、複数の C M 3 3 を含む。C M 3 3 は、ディスク装置 3 9 の動作制御を行う。C M 3 3 は、ストレージ制御装置として機能する。ストレージシステム 3 1 は、C M 3 3 を 2 以上有することで、冗長性を確保する。

【 0 0 5 5 】

C M 3 3 は、アクセス指示情報として入出力（I / O）コマンドを D E 4 0 に送信し、ディスク装置 3 9 の記憶領域に対するデータの入出力指令を行う。また、入出力指令からアクセス監視時間が経過しても応答が得られないときは、C M 3 3 は、この I / O 処理を中断するアボート指示コマンドを D E 3 3 に送信する。

40

【 0 0 5 6 】

C M 3 3 は、チャンネルアダプタ（C A）3 4、中央演算装置（C P U : Central Processing Unit）3 5、記憶部 3 6、デバイスアダプタ（D A）3 7、C M 間通信インターフェース（I / F）3 8 を含む。チャンネルアダプタ（C A）3 4、C P U 3 5、記憶部 3 6、デバイスアダプタ（D A）3 7 は、内部バスを介して接続されている。

【 0 0 5 7 】

C A 3 4 は、F C スwitch 4 2 を介して、ホスト 4 1 と接続されている。C P U 3 5 は、C M 3 3 全体を制御する。C M 間通信 I / F 3 8 は、他の C M と通信を行うためのインタ

50

ーフェースであり、C M間通信I / F 3 8 間は、通信線で接続されている。

【 0 0 5 8 】

記憶部 3 6 は、例えば、キャッシュメモリ、R O M (Read Only Memory)、R A M (Random Access Memory) 等の情報を記録するデバイスである。記憶部 1 6 には、C M 3 3 を動作させるために用いるデータ、及び本実施形態に係るプログラム及び本実施形態で用いるデータ等が格納される。記憶部 1 6 には、B U D が含まれる。

【 0 0 5 9 】

D A 3 7 は、D E 4 0 に接続される。C P U 3 5 は、D A 3 7 を介して D E 4 0 との間でデータの送受信を行う。

【 0 0 6 0 】

D E 4 0 は、1 以上の D E より形成される。D E 4 0 は、複数のディスク装置 3 9 を含み、冗長性を考慮した R A I D 構成を有する。

【 0 0 6 1 】

図 4 は、本実施形態における C M の機能ブロック図を示す。各 C M 3 3 は、機能制御部 5 1 (5 1 A ~ 5 1 N)、等価制御部 5 2、B U D アクセス制御部 6 2、B U D ドライバ 6 3、B U D 6 4、C M 間通信ドライバ制御部 6 8 を含む。

【 0 0 6 2 】

機能制御部 5 1 (機能 # A 制御部 5 1 A ~ 機能 # N 制御部 5 1 N) は、B U D 6 4 の領域を使用して、それぞれの機能を実行させる制御部である。機能制御部 5 1 は、B U D へのアクセス時にアクセスする範囲に対応するブロックの属性を等価制御部 5 2 に渡す I / F を含む。ブロックとは、後述するように、ユーザ領域 6 6 を所定のサイズで区分されたことで得られる区画のことである。これにより、機能制御部 5 1 が、メタデータ / データを動的にセットする機能を有していても、等価制御部 5 2 はその機能に対応することができる。

【 0 0 6 3 】

等価制御部 5 2 は、各機能制御部 5 1 が使用する B U D 6 4 の領域に関して、B U D 6 4 間でデータの等価状態を維持し、データを管理する制御部であり、C P U 3 5 によって実現される。等価制御部 5 2 は、ライト制御部 5 3、リード制御部 5 4、リカバリ制御部 5 5、ライトサブ制御部 5 6、リードサブ制御部 5 7、ロック / アンロック制御部 5 8、等価チェック制御部 5 9、ブロック属性管理制御部 6 0、ブロック属性解析制御部 6 1 を含む。

【 0 0 6 4 】

ライト (write) 制御部 5 3 は、各機能制御部 5 1 からの書き込み要求に対する処理を制御するメインルーチンである。リード (read) 制御部 5 4 は、各機能制御部 5 1 からの読み出し要求に対する処理を制御するメインルーチンである。

【 0 0 6 5 】

リカバリ (recovery) 制御部 5 5 は、B U D 6 4 間でデータが等価状態にない領域に対して、リカバリを制御するメインルーチンである。リカバリ制御部 5 5 は、正常なデータを有する B U D 6 4 からデータを読み出し、その読み出したデータを、異常なデータを有する B U D 6 4 に書き込む。

【 0 0 6 6 】

ライトサブ (write sub) 制御部 5 6 は、各制御部 (5 3 ~ 5 5 , 5 7 ~ 6 1) から指定された B U D 6 4 の対象領域にデータを書き込むサブルーチンである。リードサブ (read sub) 制御部 5 7 は、各制御部 (5 3 ~ 5 5 , 5 7 ~ 6 1) から指定された B U D 6 4 の対象領域のデータを読み出すサブルーチンである。

【 0 0 6 7 】

ロック / アンロック制御部 5 8 は、機能制御部 5 1 から B U D へのアクセスの排他を制御するメインルーチンである。ロック / アンロック制御部 5 8 は、機能制御部 5 1 から B U D へのアクセス要求があった場合、アクセス要求元の機能制御部によるアクセスを許可し、アクセス要求元以外の機能制御部によるアクセスを制限することにより、アクセスの

10

20

30

40

50

排他制御を行う。

【 0 0 6 8 】

等価チェック制御部 5 9 は、各機能制御部 5 1 で使用する B U D 6 4 の領域に保持されているデータが B U D 6 4 間で等価状態にあるかどうかをチェックする。等価状態にない B U D 6 4 が存在する場合は、等価チェック制御部 5 9 は、等価状態となるようにリカバリ制御部 5 5 にリカバリ処理を実行するように要求する。

【 0 0 6 9 】

ブロック属性管理制御部 6 0 は、各ユーザ領域 6 6 内のブロックの属性を管理するメインルーチンである。ブロック属性管理制御部 6 0 は、等価制御部 5 2 で管理しているユーザ領域 6 5 内で、任意のブロック単位で格納されるデータの種別、有効 / 無効等の属性を含むブロック属性情報を構築する。ブロック属性管理制御部 6 0 は、構築したブロック属性情報を等価管理データとして保持する。ブロックの属性がメタデータなのか実データなのかを判断するために、ブロック属性管理制御部 6 0 は、各機能制御部 5 1 に対して、そのブロックの属性を問い合わせるための I / F を含む。これにより、ブロックの属性が変化しても、B U D にアクセスロックがかけられていない場合には、ブロック属性管理制御部 6 0 は、機能制御部 5 1 からブロック属性を取得することができる。これにより、ブロック属性管理制御部 6 0 は、現在のデータからブロック属性マッピング情報を再構築できる。ブロック属性マッピング情報とは、ユーザ領域の各ブロックに対応させてブロック属性情報を管理する情報を示す。

【 0 0 7 0 】

また、後述するように、一度データの書き込みがされると、その書き込みされた領域のブロック属性が有効に変化する。仮に、各機能制御部 5 1 からブロック属性を取得できないならば、等価制御部 5 2 は、一度書き込みがされた領域がその後、使用しない領域に変更された場合、その領域が使用されなくなったこと、すなわち、その領域が無効であることを判断することができない。しかしながら、本実施形態では、等価制御部は各機能制御部 5 1 からブロック属性を取得することができる。そのため、上記ケースであっても、等価制御部 5 2 は取得したブロック属性に基づいて、等価管理データのブロックの属性を「有効」属性から「無効」属性に変更することができる。

【 0 0 7 1 】

ブロック属性解析制御部 6 1 は、各制御部 (5 3 ~ 6 0) からのアクセス要求があった場合、そのアクセス範囲に対応するブロックの属性を判定する。

【 0 0 7 2 】

B U D アクセス制御部 6 2 は、各制御部から B U D 6 4 へのアクセス要求があった場合、アクセス先のアドレスについて、論理アドレス - 物理アドレスの変換を行う。B U D ドライバ 6 3 は、B U D 6 4 との通信を行うドライバである。B U D 6 4 は、等価管理領域 6 5、ユーザ領域 6 6、バックアップ領域 6 7 を含む。

【 0 0 7 3 】

図 5 は、本実施形態におけるユーザ領域の一例を示す。B U D 6 4 は、各機能制御部 5 1 に対応する領域 (ユーザ領域) を有する。B U D 6 4 のユーザ領域 6 6 には、メタデータ 7 1 と実データ 7 2 が格納される。B U D 6 4 へアクセスを行う C M 3 3 内の機能制御部 5 1 (特に容量が大きい領域を使用する機能) は、機能固有のメタデータ 7 1 (どのアドレスに何のデータが格納されているかを示す情報) を保持する。

【 0 0 7 4 】

書き込んだ実データの有効化 (完全性を保持) または実データの無効化 (破棄、上書き) をメタデータで管理するため、実データ自体の書き込みに付随して、メタデータの更新が行われる。言い換えれば、メタデータ上でその実データが有効とならなければ、その実データは無効なデータとして扱われる。つまり、メタデータを保護することにより、実データの整合性を保持することができる。したがって、バックアップ領域 6 7 への書き込み処理は、メタデータについて行えばよい。

【 0 0 7 5 】

また、ユーザ領域 6 6 に格納されているメタデータの範囲が分かれば、メタデータを読み出し、有効な実データがユーザ領域のどの位置に格納されているかを判断することができる。そのため、C M 3 3 の保守や B U D 6 4 の保守等でのデータコピー処理において、コピー元から必要最小限のデータをコピーすることができる。その結果、C M の組み込み時間の短縮、あるいはコピー元データの読み出し失敗時のデータ破棄の判断が可能になる。

【 0 0 7 6 】

以下では、等価管理データを用いて、等価制御部 5 2 により各ユーザ領域内のデータ属性を管理することにより、各ユーザ領域内のメタデータを保護し、書き込みアクセス処理、及び読み出しリカバリ処理を効率的に行うことについて説明する。

10

【 0 0 7 7 】

図 6 は、本実施形態における等価管理データのデータ構成を示す。等価管理データは、B U D 6 4 間のデータの整合性をユーザ領域単位で管理するためのデータであり、各ユーザ領域 6 6 に格納される情報（管理データ）を管理するために用いられる。等価管理データは、等価管理領域 6 5 に格納されている。

【 0 0 7 8 】

等価管理データは、有効 / 無効フラグ 8 1、書込み判定フラグ 8 2、領域識別番号フラグ 8 3、更新タイムスタンプ 8 4、更新カウンタ 8 5、旧更新タイムスタンプ 8 6、ブロック属性マッピング情報 8 7 を含む。

【 0 0 7 9 】

有効 / 無効フラグ 8 1 は、当該等価管理データが有効であるか無効であることを示すフラグである。

20

【 0 0 8 0 】

書込み判定フラグ 8 2 は、当該等価管理データに対応するユーザ領域へデータが書込み処理中か否かを判定するフラグである。書込み判定フラグ 8 2 は、書込み処理が開始する場合に O N となり、書込み処理が完了すると O F F になる。

【 0 0 8 1 】

領域識別番号 8 3 は、各機能制御部 5 1 が使用しているユーザ領域を識別する番号を示す。

【 0 0 8 2 】

更新タイムスタンプ 8 4 は、当該ユーザ領域 6 6 に格納される情報（管理データ）が更新された時刻を示す情報である。更新タイムスタンプ 8 4 は、書き込み処理により更新される。

30

【 0 0 8 3 】

更新カウンタ 8 5 は、当該管理データが更新された回数を示す情報であり、更新タイムスタンプ 8 4 の補足情報である。更新カウンタ 8 5 は、B U D 6 4 間で、対応するユーザ領域の管理データ同士の更新タイムスタンプ 8 4 が同じ場合にカウントされる。

【 0 0 8 4 】

旧更新タイムスタンプ 8 6 は、更新タイムスタンプ 8 4 の 1 つ前の更新タイムスタンプの情報を示し、B U D 6 4 間のデータの等価に失敗した際に設定される。旧更新タイムスタンプ 8 6 は、旧更新タイムスタンプ 1 (8 6 - 1)、旧更新タイムスタンプ 2 (8 6 - 2)、・・・旧更新タイムスタンプ k (8 6 - k) (k : 任意の整数) を含む。

40

【 0 0 8 5 】

ブロック属性マッピング情報 8 7 は、任意のブロック長を有し、各ブロックの属性を管理するマッピング情報である。具体的には、ブロック属性マッピング情報 8 7 は、当該ユーザ領域を所定のサイズのブロックに区分して、区分したブロック毎に属性情報を付与したマッピング情報である。ブロック属性マッピング情報 8 7 の詳細については、図 7 を用いて説明する。

【 0 0 8 6 】

図 7 は、本実施形態におけるブロック属性マッピング情報を説明するための図である。

50

図 7 (A) は、所定サイズのブロックで区分されたユーザ領域を示す。図 7 (B) は、そのユーザ領域を形成する各ブロックをマッピングしたブロック属性マッピング情報を示す。

【 0 0 8 7 】

図 7 (A) に示すように、ユーザ領域は、メタデータを格納する記憶領域 (メタデータ格納領域) 、実データを格納する記憶領域 (データ格納領域) 、未使用の記憶領域 (フリー領域) を含む。ユーザ領域は、所定のサイズの区画 (ブロック) に区分して管理される。ブロック属性マッピング情報は、図 7 (B) に示すように、そのユーザ領域を形成する各ブロックがマッピングされており、各ブロックの属性情報を有する。

【 0 0 8 8 】

ブロック属性マッピング情報 8 7 において、「Block (Column) 」、「Block (Raw) 」で特定される位置は、当該ユーザ領域において対応するブロックを特定する位置に相当する。図 7 (B) に示すように、ブロック属性マッピング情報 8 7 において「Block (Column) 」、「Block (Raw) 」で特定されるブロックの位置に、そのブロックの属性情報が格納されている。

【 0 0 8 9 】

図 8 は、本実施形態におけるブロック属性情報のデータ構造を示す。ブロック属性情報は、「領域属性情報」9 1、「有効 / 無効属性情報」9 2、「リカバリ属性情報」9 3 を含む。

【 0 0 9 0 】

「領域属性情報」9 1 には、ユーザ領域においてそのブロックに格納されている情報の属性を示す情報が格納される。「領域属性情報」9 1 には、具体的には、「メタ (M) 」、「データ (D) 」、「フリー (F) 」が格納される。「メタ (M) 」は、そのブロックに対応する領域に格納された実データに関するメタデータの格納領域であることを示す。「データ (D) 」は、そのブロックに対応する領域に格納された実データの格納領域であることを示す情報であって、メタデータで管理される情報であることを示す。「フリー (F) 」は、そのブロックに対応する領域が空き領域 (未使用領域) であることを示す。

【 0 0 9 1 】

「有効 / 無効属性情報」9 2 には、「有効 (V) 」、「無効 (I) 」のうち、いずれかが格納される。「有効 (V) 」は、そのブロックが有効な情報を保持していることを示す。「無効 (I) 」は、そのブロックが有効な情報を保持していないことを示す。

【 0 0 9 2 】

「リカバリ属性情報」9 3 には、複数ある B U D 6 4 のうち、等価状態が維持されずにリカバリ処理が必要になった B U D 6 4 を識別する情報が格納される。

【 0 0 9 3 】

次に、等価制御部の動作について説明する。

< ストレージ装置の初期起動時 >

図 9 は、本実施形態におけるストレージ装置の初期起動時の等価制御部の動作フローを示す。ストレージ装置が初期起動でない場合 (S 2 0 で「No」) 、本フローは終了する。ストレージ装置が初期起動である場合 (S 2 0 で「Yes」) 、等価制御部 5 2 は、各ユーザ領域に対応する機能制御部 5 1 に、初期書込みデータの作成を依頼する (S 2 1) 。初期書込みデータとは、初期起動時にユーザ領域に格納するメタデータを示す。

【 0 0 9 4 】

等価制御部 5 2 は、機能制御部 5 1 に問い合わせ、初期起動時のブロック属性マッピング情報を示す初期ブロック属性マッピング情報及び初期起動時の等価管理データを示す初期等価管理データを作成する (S 2 2 , S 2 3) 。

【 0 0 9 5 】

等価制御部 5 2 は、初期書込みデータ (メタデータ) と初期等価管理データを各 B U D に書き込む (S 2 4 , S 2 5) 。各ユーザ領域について、S 2 1 ~ S 2 5 の処理を繰り返す (S 2 6) 。これ以降、各機能制御部による B U D 6 4 へのアクセスが可能になる。

10

20

30

40

50

【 0 0 9 6 】

図 1 0 は、本実施形態におけるストレージ装置の初期起動時の C M 間の動作シーケンスの全体図である。以下では、説明の便宜上、制御する側の C M (制御 C M) の構成に関する要素については添え字 a を付し、それ以外の C M の構成に関する要素については添え字 b を付す。

【 0 0 9 7 】

制御 C M 3 3 a 内のパワー制御部 1 0 1 a は、電源がオンになると、等価制御部 5 2 a へ等価制御を実行させるための等価要求を行う (S 3 1) 。

【 0 0 9 8 】

等価制御部 5 2 a は、等価要求を受信すると、 B U D 6 4 間のアクセスの排他制御を行い、 B U D 6 4 へのアクセスロックを行う (S 3 2) 。アクセスロックとは、等価制御部 5 2 a により所定の機能制御部以外の機能制御部による B U D 6 4 のユーザ領域へのアクセスを制限して B U D 6 4 間のアクセスの排他制御を示す。その後、等価制御部 5 2 a は、各 C M 3 3 b へ B U D 6 4 を初期化するための初期化要求を行う (S 3 3) 。

10

【 0 0 9 9 】

各 C M 3 3 b の等価制御部 5 2 b は、初期化要求を受けると、 B U D 6 4 の初期化処理を行う (S 3 4) 。初期化処理では、初期起動であるかが判定され、初期起動の場合には初期化されたブロック属性マッピング情報及び等価管理データが作成される。また、当該ストレージ装置が搭載している B U D 6 4 のシステム管理データが作成される。 B U D 6 4 の初期化の完了後、等価制御部 5 2 b は、制御 C M 3 3 a へ初期化要求に対する応答を行う (S 3 5) 。

20

【 0 1 0 0 】

初期化要求に対する応答を受信すると、制御 C M 3 3 a の等価制御部 5 2 a は、 B U D 6 4 間のユーザ領域のデータを等価状態にする等価処理を行う (S 3 6) 。等価制御部 5 2 a は、未処理のユーザ領域がある間 (S 3 7 で「 Y e s 」) 、等価処理を行う (S 3 6) 。

【 0 1 0 1 】

B U D 6 4 上のすべてのユーザ領域について等価処理が完了した場合 (S 3 7 で「 N o 」) 、等価制御部 5 2 a は、 B U D 6 4 へのアクセスの排他制御を行い、 B U D 6 4 へのアクセスロックを解除する (S 3 8) 。等価制御部 5 2 a は、パワー制御部 1 0 1 a へ等価制御が完了したことを通知する等価応答を行う (S 3 9) 。

30

【 0 1 0 2 】

次に、図 1 0 で説明した B U D 6 4 の初期化処理 (S 3 4) の詳細について、図 1 1 を用いて説明する。

【 0 1 0 3 】

図 1 1 A 及び図 1 1 B は、本実施形態における B U D の初期化処理 (S 3 4) の詳細なシーケンス図である。 S 3 4 において、制御 C M 3 3 a の等価制御部 5 2 a から初期化要求を受けると、各 C M 3 3 b の等価制御部 5 2 b は、以下の処理を行う。

【 0 1 0 4 】

まず、等価チェック制御部 5 9 b は、リードサブ制御部 5 7 b、 B U D アクセス制御部 6 2 b、 B U D ドライバ 6 3 b を介して、 B U D 6 4 b へ読み出し要求を行い、 B U D 6 4 b からシステム管理データを読み出す (S 4 1、 S 4 2、 S 4 3) 。ここで、システム管理データとは、各ストレージ装置で独立した情報であって、各ストレージ装置が保持する B U D 6 4 を管理する情報である。システム管理データは、ストレージ装置が初期起動であるか否かの情報を含んでいる。

40

【 0 1 0 5 】

等価チェック制御部 5 9 b は、読み出したシステム管理データに基づいて、 B U D の初期化が行われているかを判定する (S 4 4) 。システム管理データを参照して、初期化処理が既に行われていると判定した場合 (S 4 4 で「 Y e s 」) 、等価チェック制御部 5 9 b は、初期化処理 (S 3 4) を終了する。システム管理データを参照して、初期化処理が

50

まだ行われていないと判定した場合（S 4 4で「No」）、等価チェック制御部 5 9 b は、リカバリ制御部 5 5 b へ、デフォルトの等価状態を形成するためにリカバリ要求を行う（S 4 5）。

【0106】

リカバリ制御部 5 5 b は、機能制御部 5 1 b に、機能制御部 5 1 b が使用するユーザ領域に対する初期データの作成要求を行う。機能制御部 5 1 b の初期データ作成部 1 1 1 b は、初期データの作成依頼を受けると、機能制御部 5 1 b が使用するユーザ領域のデフォルトのメタデータと初期データとして作成し、リカバリ制御部 5 5 b へ返信する（S 4 6）。

【0107】

リカバリ制御部 5 5 b は、ブロック属性管理制御部 6 0 b を介して、機能制御部 5 1 b のブロック属性応答部 1 1 2 b へ、初期ブロック属性マッピング情報を要求する（S 4 7、S 4 8）。初期ブロック属性マッピング情報とは、機能制御部 5 1 b が使用するユーザ領域の各ブロックに対応する初期化されたブロック属性情報のことである。

【0108】

ブロック属性応答部 1 1 2 b は、その機能制御部 5 1 b が使用するユーザ領域の各ブロックの属性情報をブロック属性管理制御部 6 0 b へ返信する。ブロック属性管理制御部 6 0 b は、そのブロック属性情報から、その機能制御部 5 1 b が使用するユーザ領域の各ブロックにマッピングさせた形式の情報（ブロック属性マッピング情報）を生成する。ブロック属性管理制御部 6 0 b は、その初期時のブロック属性マッピング情報をリカバリ制御部 5 5 b へ返信する（S 4 9）。

【0109】

リカバリ制御部 5 5 b は、初期ブロック属性マッピング情報を含めた、その機能制御部 5 1 b が使用するユーザ領域に対応する初期化された等価管理データ（初期等価データ）を生成する（S 5 0）。

【0110】

リカバリ制御部 5 5 b は、ライトサブ制御部 5 6 b、BUDアクセス制御部 6 2 b、BUDドライバ 6 3 b を介して、BUD 6 4 b において、その機能制御部 5 1 b が使用するユーザ領域へ初期データ（メタデータ）を書き込む（S 5 1、S 5 2、S 5 3）。

【0111】

リカバリ制御部 5 5 b は、ライトサブ制御部 5 6 b、BUDアクセス制御部 6 2 b、BUDドライバ 6 3 b を介して、BUD 6 4 b の等価管理領域へ初期等価管理データの書き込む（S 5 4、S 5 5、S 5 6）。リカバリ制御部 5 5 b は、等価チェック制御部 5 9 b へ、リカバリ要求に対する応答を行う（S 5 7）。

【0112】

等価チェック制御部 5 9 b は、BUD 6 4 において、未処理のユーザ領域がある場合（S 5 8で「Yes」）、S 4 5 の処理へ戻る。

【0113】

各機能制御部 5 1 b が使用するユーザ領域について、処理が完了した場合（S 5 8で「No」）、等価チェック制御部 5 9 b は、各ストレージ装置が保持するBUDを管理するシステム管理データを更新する（S 5 9）。等価チェック制御部 5 9 b は、ライトサブ制御部 5 6 b、BUDアクセス制御部 6 2 b、BUDドライバ 6 3 b を介して、BUD 6 4 b へシステム管理データの書き込み要求を行う（S 6 0、S 6 1、S 6 2）。

【0114】

次に、図 1 0 で説明したBUD間のユーザ領域のデータが等価状態を維持するための等価処理（S 3 6）の詳細について、図 1 2 A、図 1 2 B 及び図 1 2 C を用いて説明する。

【0115】

図 1 2 A、図 1 2 B 及び図 1 2 C は、本実施形態におけるBUD間の等価処理（S 3 6）の詳細なシーケンス図である。S 3 6 において、各CM 3 3 b の等価制御部 5 2 b から初期化要求に対する応答を受けると、制御CM 3 3 a の等価制御部 5 2 a は、以下の処理

10

20

30

40

50

を行う。

【 0 1 1 6 】

まず、等価チェック制御部 5 9 b は、リードサブ制御部 5 7 b、B U D アクセス制御部 6 2 b、B U D ドライバ 6 3 b を介して、各 B U D 6 4 b へ読み出し要求を行い、全 B U D 6 4 b から等価管理データを読み出す (S 7 1 , S 7 2、S 7 3)。

【 0 1 1 7 】

等価チェック制御部 5 9 a は、各 B U D 6 4 から受信した等価管理データが等価状態を維持しているかを判定する。すなわち、等価チェック制御部 5 9 a は、各 B U D 6 4 から読み出した等価管理データの更新タイムスタンプ及び更新カウンタを比較して、更新タイムスタンプ及び更新カウンタが同一であるかを判定する。

10

【 0 1 1 8 】

各 B U D 6 4 から読み出した等価管理データの更新タイムスタンプ及び更新カウンタが同一である場合 (S 7 4 で「 Y e s 」)、等価チェック制御部 5 9 a は、全等価管理データは正常 (等価状態) であると判定し、本シーケンスの処理を終了する。

【 0 1 1 9 】

各 B U D 6 4 から読み出した等価管理データの更新タイムスタンプまたは更新カウンタが同一でない場合 (S 7 4 で「 N o 」)、リカバリの必要な B U D があるので、等価チェック制御部 5 9 a は、リカバリ制御部 5 5 a へ、リカバリ要求を送信する (S 7 5)。ここで、リカバリ制御部 5 5 a は、S 7 4 で比較した等価管理データのうち、更新タイムスタンプが最新か、または更新タイムスタンプが最新かつ更新カウンタの値が最も多い等価管理データを等価状態の基準とする。以下では、その等価状態の基準とする B U D を等価元 B U D と称する。また、等価元 B U D に合わせて等価状態にする対象の B U D、すなわちリカバリ対象 B U D を等価先 B U D と称する。

20

【 0 1 2 0 】

リカバリ制御部 5 5 a は、等価管理データを作成する (S 7 6)。ここでは、リカバリ制御部 5 5 a は、リカバリの必要な B U D (等価先 B U D) の等価管理データを、等価元 B U D の等価管理データで上書きする。リカバリ制御部 5 5 a は、ライトサブ制御部 5 6 a、B U D アクセス制御部 6 2 b、B U D ドライバ 6 3 b を介して、有効 / 無効フラグ = 「無効」に更新して、等価管理データを無効化する (S 7 7 , S 7 8 , 7 9)。

【 0 1 2 1 】

リカバリ制御部 5 5 a は、等価元 B U D の等価管理データのブロック属性マッピング情報から 1 つのブロック属性情報を抽出する (S 8 0)。リカバリ制御部 5 5 a は、B U D アクセス制御部 6 2 b、B U D ドライバ 6 3 b を介して、等価元 B U D の、対応するユーザ領域から、抽出したブロック属性情報に対応する情報を読み出す (S 8 1 , S 8 2 , S 8 3)。

30

【 0 1 2 2 】

リカバリ制御部 5 5 a は、ライトサブ制御部 5 6 a、B U D アクセス制御部 6 2 b、B U D ドライバ 6 3 b を介して、等価元 B U D から読み出した情報を、等価先 B U D 6 4 へ書込む (S 8 4、S 8 5、S 8 6)。

【 0 1 2 3 】

リカバリ制御部 5 5 a は、等価元 B U D の等価管理データのブロック属性マッピング情報に未処理ブロックがある場合 (S 8 7 で「 Y e s 」)、S 8 0 の処理へ戻る。

40

【 0 1 2 4 】

等価元 B U D の等価管理データのブロック属性マッピング情報の全ブロックについて、処理が完了した場合 (S 8 7 で「 N o 」)、リカバリ制御部 5 5 a は、各 B U D の等価管理データのブロック属性情報に含まれるリカバリ属性情報を、「 0 」でクリアする (S 8 8)。

【 0 1 2 5 】

リカバリ制御部 5 5 a は、ライトサブ制御部 5 6 a、B U D アクセス制御部 6 2 b、B U D ドライバ 6 3 b を介して、S 8 8 で更新したブロック属性マッピング情報を含む等価

50

管理データの有効／無効フラグを「有効」に更新する。それから、リカバリ制御部 55a は、等価管理データを有効化する（S89、S90、S91）。

【0126】

また、リカバリ制御部 55a は、ライトサブ制御部 56a、BUD アクセス制御部 62b、BUD ドライバ 63b を介して、リカバリ対象外の BUD 64b の等価管理データのブロック属性のリカバリ属性を「0」で更新する（S92、S93、S94）。等価チェック制御部 59a は、リカバリ制御部 55a から、リカバリ要求に対する応答を受信する（S95）。

【0127】

このように、ストレージ装置の初期起動時において、機能制御部が使用するユーザ領域を初期化する場合、等価管理部 52 は、等価管理データの初期データを作成し、各 BUD 64 に保持する。本実施形態では、等価管理データにより、ユーザ領域に格納されるデータを管理するので、論理的な情報（等価管理データ）の初期化で済むので、ユーザ領域の全領域を 0 で書き込むことによる初期化の必要がない。

【0128】

<各機能制御部から BUD へ読み出し処理を行う場合>

図 13 は、本実施形態における各機能制御部から BUD へ読み出し処理を行う場合の等価制御部の動作フローを示す。機能制御部 51 からの読み出し要求を受けると、等価制御部 52 は、各 BUD 64 から、等価管理データを読み出す（S101）。

【0129】

等価制御部 52 は、各 BUD 64 から読み出した等価管理データを比較して、等価管理データが一致しない BUD があるかを判定する（S102）。すなわち、等価制御部 52 は、各 BUD 64 から読み出した等価管理データの更新タイムスタンプ及び更新カウンタを比較して、更新タイムスタンプ及び更新カウンタが同一であるかを判定する。

【0130】

各 BUD 64 からリードした等価管理データの更新タイムスタンプまたは更新カウンタが同一でない場合（S102 で「Yes」）、等価制御部 52 は、リカバリ処理を行う（S103）。

【0131】

各 BUD 64 から読み出した等価管理データの更新タイムスタンプ及び更新カウンタが同一である場合（S102 で「No」）、等価制御部 52 は、次の処理を行う。すなわち、等価制御部 52 は、等価管理データのブロック属性マッピング情報を用いて、機能制御部 51 からの読み出し要求があったアクセス範囲に対応するブロックの属性情報の有効性をする（S104）。

【0132】

S104 において、アクセス範囲に対応するブロックの属性情報が「無効（I）」属性または「フリー（F）」属性を示す場合（S104 で「Yes」）、等価制御部 52 は、次の処理を行う。すなわち、等価制御部 52 は、BUD 64 から直接データを読み出さずに、そのアクセス範囲に相当するゼロでパディングしたデータを読み出しデータとして作成し、読み出し要求した機能制御部 61 へ応答する（S105、S107）。

【0133】

S104 において、アクセス範囲に対応するブロックの属性情報に「有効（V）」属性を含むブロック属性がある場合（S104 で「No」）、等価制御部 52 は、アクセス範囲の変更を行う。すなわち、等価制御部 52 は、アクセス範囲のうち「有効（V）」属性を有するブロック（「無効（I）」属性及び「フリー（F）」属性を有するブロックを除いたブロック）のデータを、ユーザ領域から読み出す（S106、S107）。

【0134】

図 14 は、本実施形態における機能制御部から BUD へ読み出し処理を行う場合の CM 間の動作シーケンスの全体図である。

【0135】

10

20

30

40

50

制御CM33a内の機能制御部51aは、等価制御部52aに対して、BUDアクセスロック要求を行う(S111)。

【0136】

等価制御部52aは、BUDアクセスロック要求を受信すると、BUD間のアクセスの排他制御を行い、BUDアクセスロックを行う(S112)。等価制御部52aは、アクセスロック完了を、その要求元の機能制御部に応答する(S113)。

【0137】

機能制御部51aは、等価制御部52aへ、データの読み出し要求を行う(S114)。等価制御部52aは、BUD間のユーザ領域のデータが等価状態を維持させる等価処理を行う(S115)。S115は、図10のS36、図12A～図12Cと同様なので、その説明を省略する。

10

【0138】

その後、等価制御部52aは、機能制御部51aから要求されたデータを読み出し(S116)、読み出し要求元の機能制御部51aへ応答する(S117)。

【0139】

機能制御部51aは、その読み出しデータを受信すると、等価制御部52aへDUDアクセスアンロック要求を行う(S118)。等価制御部52aは、BUDへのアクセスの排他制御を行い、BUDのユーザ領域へのアクセスのロックを解除する(S119)。等価制御部52aは、DUDアクセスアンロック要求に対する応答を行う(S120)。

【0140】

20

図15は、本実施形態における読み出し処理(S116)の詳細なシーケンス図である。S116において、制御CM33aの等価制御部52aは、機能制御部からの読み出し要求に基づいて、以下の処理を行う。

【0141】

まず、リード制御部54は、リードサブ制御部57a、BUDアクセス制御部62b、BUDドライバ63bを介して、BUD64から、書込み要求元の機能制御部が使用するユーザ領域に対応する等価管理データを読み出す(S121、S122、S123)。

【0142】

リード制御部54は、ブロック属性解析制御部61aへ、ユーザ領域から読み出す対象のアクセスの範囲を示すアクセス範囲データを要求する(S124)。ブロック属性解析制御部61aは、アクセス範囲データ要求を受信すると、読み出した等価管理データのブロック属性マッピング情報から、そのアクセス範囲に対応するブロック属性情報を読み出す。アクセス範囲に「有効(V)」属性を有するブロックが含まれる場合、ブロック属性解析制御部61aは、アクセス範囲を変更する(S125)。すなわち、ブロック属性解析制御部61aは、アクセス範囲データ要求によるアクセス範囲のうち、「有効」属性を有するブロックに対応する範囲(「無効」属性または「フリー」属性のブロックを含まない範囲)にアクセス範囲を変更する。または、アクセス範囲に対応するブロックの属性情報が「無効(I)」属性または「フリー(F)」属性を示す場合、等価制御部52は、そのアクセス範囲に相当するゼロでパディングしたデータを読み出しデータとして作成する。ブロック属性解析制御部61aは、その変更したアクセス範囲またはゼロでパディングしたデータをリード制御部54に応答する(S126)。

30

40

【0143】

リード制御部54は、その変更されたアクセス範囲に基づいて、リードサブ制御部57a、BUDアクセス制御部62b、BUDドライバ63bを介して、BUD64からデータを読み出す(S127、S128、S129)。または、リード制御部54は、ゼロでパディングしたデータを読み出す。

【0144】

このように、ユーザ領域に格納されたデータの読み取りにおいて、ブロック属性情報により、読み取りアクセス先のブロックに格納された情報が有効かどうかを判定することができる。そして、読み取りアクセス先に有効な情報が格納されている場合にはそのブロッ

50

クを読み出す（読み取りアクセス先に有効な情報が格納されていない場合にはそのブロックを読み出さない）。これにより、ユーザ領域内の有効な情報を読み出すことができるので（有効でない情報については読み出しを行わない）、読み出し効率を向上させることができる。

【0145】

<各機能制御部からの書き込み時>

図16A及び図16Bは、本実施形態における各機能制御部からBUDへ書き込み処理を行う場合の等価制御部の動作フローを示す。機能制御部51からの書き込み要求を受けると、等価制御部52は、各BUD64の等価管理領域からその機能制御部51が使用するユーザ領域に対応する等価管理データを読み出す（S131）。

10

【0146】

等価制御部52は、各BUD64から読み出した等価管理データを比較して、等価管理データが一致しないBUDがあるかを判定する（S132）。すなわち、等価制御部52は、各BUD64から読み出した等価管理データの更新タイムスタンプ及び更新カウンタを比較して、更新タイムスタンプ及び更新カウンタが同一であるかを判定する。

【0147】

各BUD64から読み出した等価管理データの更新タイムスタンプまたは更新カウンタが同一でない場合、等価制御部52は、リカバリ処理を行う（S133）。

【0148】

各BUD64から読み出した等価管理データの更新タイムスタンプ及び更新カウンタが同一である場合、またはS133の処理後、等価制御部52は、次の処理を行う。すなわち、等価制御部52は、等価管理データを用いて、機能制御部により書き込み要求のあったアクセス先（書き込み先）の範囲のブロックの属性が「メタ（M）」属性であるか否かを判定する（S134）。

20

【0149】

S134において、アクセス先の範囲のブロックの属性が「メタ（M）」属性である場合（S134で「Yes」）、すなわち、書き込み対象データがメタデータである場合、等価制御部52は、書き込み対象データをバックアップ用データとする（S135）。等価制御部52は、バックアップ用データを、BUD64のバックアップ領域67へ書き込む（S136）。

30

【0150】

S134において、アクセス先の範囲のブロックの属性が「データ（D）」属性である場合（S134で「No」）、書き込み対象データはメタデータではないので（この場合、書き込み対象データは実データである）、S137の処理へ進む。

【0151】

等価制御部52は、読み出した等価管理データの有効／無効フラグを「無効」、書き込み判定フラグを「ON」に更新する（S137）。等価制御部52は、更新した等価管理データを、BUD64へ書き込む（S138）。等価制御部52は、機能制御部からの書き込み要求による書き込み対象のデータを、BUD64のユーザ領域へ書き込む（S139）。等価制御部52は、ユーザ領域において、書き込みを行った領域に対応するブロックの属性を「有効（V）」属性に更新する（S140）。

40

【0152】

等価制御部52は、書き込み処理中に、等価対象のBUDのいずれかで書き込み処理が失敗したかどうかを判断する（S141）。等価対象のBUDのいずれかで書き込み処理が失敗した場合、すなわちリカバリが必要なBUDが存在する場合（S141で「Yes」）、等価制御部52は、等価管理データを更新する。

【0153】

すなわち、等価制御部52は、等価管理データにおいて、機能制御部からの書き込み要求があったアクセス範囲に対応するブロック属性情報に含まれるリカバリ属性情報に、アクセスが失敗したBUDの識別情報を設定する（S142）。

50

【 0 1 5 4 】

ここで、等価制御部 5 2 は、リカバリが必要な B U D が存在するかどうかの判断を、B U D の異常か、C M の異常かに基づいて判断する。等価制御部 5 2 は、B U D の異常か、C M の異常かを示すエラーを検知することができる。B U D の異常であれば、B U D の交換が行われるため、等価制御部 5 2 は、リカバリ属性情報への設定を行わない。一方、C M の異常であれば、同じ B U D を使用して、C M の組み込みが行われる可能性があるため、等価制御部 5 2 は、リカバリ属性情報への設定を行う。

【 0 1 5 5 】

また、等価制御部 5 2 は、等価管理データの旧更新タイムスタンプに更新前のタイムスタンプが設定されていない場合 (S 1 4 3 で「 Y e s 」)、次の処理を行う。すなわち、等価制御部 5 2 は、読み出した等価管理データの「旧更新タイムスタンプ」8 6 を、「更新タイムスタンプ」8 4 で更新する (S 1 4 4)。

10

【 0 1 5 6 】

等価対象の B U D のいずれかで書き込み処理が成功した場合 (S 1 4 1 で「 N o 」) または S 1 4 4 の処理が終了した場合、等価制御部 5 2 は、次の処理を行う。等価制御部 5 2 は、等価管理データの有効 / 無効フラグを「有効」、書込み判定フラグを「 O F F 」に、「更新タイムスタンプ」に現在の時刻に更新する (S 1 4 5)。等価制御部 5 2 は、更新した等価管理データを、B D U 6 4 へ書き込む (S 1 4 6)。

【 0 1 5 7 】

図 1 7 は、本実施形態における書き込み処理シーケンスの全体図である。制御 C M 3 3 a 内の機能制御部 5 1 a は、等価制御部 5 2 a に対して、B U D アクセスロック要求を行う (S 1 5 1)。

20

【 0 1 5 8 】

等価制御部 5 2 a は、B U D アクセスロック要求を受信すると、B U D 間のアクセスの排他制御を行い、B U D アクセスロックを行う (S 1 5 2)。等価制御部 5 2 a は、アクセスロック完了を、その要求元の機能制御部に応答する (S 1 5 3)。

【 0 1 5 9 】

機能制御部 5 1 a は、等価制御部 5 2 a へ、データの書き込み要求を行う (S 1 5 4)。等価制御部 5 2 a は、B U D 間のユーザ領域のデータが等価状態を維持するための等価処理を行う (S 1 5 5)。S 1 5 5 は、図 1 0 の S 3 6、図 1 2 A、図 1 2 B と同様なので、その説明を省略する。

30

【 0 1 6 0 】

その後、等価制御部 5 2 a は、データの書き込み要求に基づいて、ユーザ領域へのデータの書き込みを行う (S 1 5 6)。等価制御部 5 2 a は、書き込み要求元の機能制御部 5 1 a へその書き込み要求に対して応答する (S 1 5 7)。

【 0 1 6 1 】

機能制御部 5 1 a は、その応答を受信すると、等価制御部 5 2 a へ B U D アクセスアンロック要求を行う (S 1 5 8)。等価制御部 5 2 a は、B U D へのアクセスの排他制御を行い、B U D へのアクセスロックを解除する (S 1 5 9)。等価制御部 5 2 a は、アンロック要求に対する応答を行う (S 1 6 0)。

40

【 0 1 6 2 】

図 1 8 A 及び図 1 8 B は、本実施形態における書き込み処理 (S 1 5 6) の詳細なシーケンス図である。S 1 5 6 において、制御 C M 3 3 a の等価制御部 5 2 a は、機能制御部からの書き込み要求に基づいて、以下の処理を行う。

【 0 1 6 3 】

ライト制御部 5 3 a は、リードサブ制御部 5 7 a、B U D アクセス制御部 6 2 b、B U D ドライバ 6 3 b を介して、B U D 6 4 から、書き込み要求元の機能制御部 5 1 a が使用するユーザ領域に対応する等価管理データを読み出す (S 1 7 1、S 1 7 2、S 1 7 3)。

【 0 1 6 4 】

ライト制御部 5 3 a は、ブロック属性解析制御部 6 1 a を介して、機能制御部 5 1 a の

50

ブロック属性応答部 112a から、アクセス先の領域（ブロック）の属性情報を取得する（S174、S175、S176）。

【0165】

ライト制御部 53a は、ブロック属性解析制御部 61a へ、アクセス範囲の変更を要求する（S177）。ブロック属性解析制御部 61a は、アクセス範囲データ要求を受信すると、実アクセス範囲の変更を行う（S178）。すなわち、ブロック属性解析制御部 61a は、そのアクセス範囲に対応するブロック属性が「フリー」属性であるか否かを判定する。そのアクセス範囲に対応するブロック属性が「フリー」属性である場合、ブロック属性解析制御部 61a は、ライト制御部 53a に、等価管理データにおいて、対応するブロック属性情報に「フリー（F）」を設定する旨を応答する。そのアクセス範囲に対応するブロック属性が「フリー」以外の属性である場合、ブロック属性解析制御部 61a は、その旨を応答する（S179）。

10

【0166】

ブロック属性解析制御部 61a からの応答結果が「フリー」属性である場合、ライト制御部 53a は、ライトサブ制御部 56a を介して、等価管理データにおいて、対応するブロック属性情報に「フリー（F）」を設定する。これにより、アクセス範囲にフリー領域が含まれている場合、そのフリー領域を除いたアクセス範囲に調整することができる。

【0167】

ライト制御部 53a は、アクセス先（書込み先）の範囲のブロック属性情報が「メタ（M）」属性か否かをチェックする（S180）。ブロック属性情報が「メタ（M）」属性である場合、書込み対象の情報はメタデータであるので、ライト制御部 53a は、書込み対象の情報（メタデータ）をバックアップ用データとする（S181）。この場合、ライト制御部 53a は、ライトサブ制御部 56a を介して、バックアップ用データを、BUD64 のバックアップ領域 67 へ書き込む処理を行う（S182、S183、S184）。

20

【0168】

ブロック属性情報が「メタ（M）」属性でない場合、書込み対象の情報は実データであるので、S185 の処理へ進む。

【0169】

ライト制御部 53a は、取得した等価管理データの有効／無効フラグを「無効」、書込み判定フラグを「ON」に更新する（S185）。

30

【0170】

ライト制御部 53a は、ライトサブ制御部 56a、BUDアクセス制御部 62b、BUDドライバ 63b を介して、その更新した等価管理データを BUD64 へ書き込む（S186、S187、S188）。

【0171】

ライト制御部 53a は、ライトサブ制御部 56a、BUDアクセス制御部 62b、BUDドライバ 63b を介して、書込み対象の情報を BUD64 のユーザ領域へ書き込む（S189、S190、S191）。

【0172】

ライト制御部 53a は、更新した等価管理データのブロック属性マッピング情報について、ユーザ領域のうち書込みを行った領域に対応するブロックの属性を「有効（V）」属性に更新する（S192）。

40

【0173】

ライト制御部 53a は、ブロック属性マッピング情報を更新した等価管理データの有効／無効フラグを「有効」、書込み判定フラグを「OFF」に更新する。ライト制御部 53a は、ライトサブ制御部 56a、BUDアクセス制御部 62b、BUDドライバ 63b を介して、更新した等価管理データを BUD64 へ書き込む（S193、S194、S195）。

【0174】

これにより、メタデータの書き込みの場合にバックアップを行うようにし、実データの

50

書き込みの場合にはバックアップを行わないので、書き込み処理の時間を短縮することができる。

【 0 1 7 5 】

<リカバリ時（ＣＭ組み込み時）>

図 1 9 は、本実施形態におけるリカバリ処理を行う場合の等価制御部の動作フローを示す。ＣＭ組み込み時、等価制御部 5 2 は、リカバリ対象のＢＵＤ及び正常なＢＵＤから等価管理データを読み出す（Ｓ 2 0 1 ）。

【 0 1 7 6 】

等価制御部 5 2 は、リカバリ対象のＢＵＤから読み出した等価管理データと、正常なＢＵＤから読み出した等価管理データが一致するかを判定する（Ｓ 2 0 2 ）。Ｓ 2 0 2 において等価管理データが一致する場合（Ｓ 2 0 2 で「Ｙ ｅ ｓ」）、本フローは終了する。

【 0 1 7 7 】

Ｓ 2 0 2 において等価管理データが一致しない場合（Ｓ 2 0 2 で「Ｎ ｏ」）、等価制御部 5 2 は、リカバリ対象ＢＵＤの更新タイムスタンプと、正常なＢＵＤの旧更新タイムスタンプとが一致するかを判定する（Ｓ 2 0 3 ）。

【 0 1 7 8 】

Ｓ 2 0 3 において、リカバリ対象ＢＵＤの更新タイムスタンプと、正常なＢＵＤの旧更新タイムスタンプとが一致する場合（Ｓ 2 0 3 で「Ｙ ｅ ｓ」）、そこまでの時点の書き込み処理は完了していると判断される。この場合、等価制御部 5 2 は、正常なＢＵＤの等価管理データから、リカバリ属性情報にＢＵＤの識別情報が設定されたブロック属性情報を抽出する（Ｓ 2 0 8 ）。等価制御部 5 2 は、正常なＢＵＤのユーザ領域から、そのリカバリ属性情報を有するブロック属性情報に対応するユーザ領域の情報（リカバリデータ）を読み出す（Ｓ 2 0 9 ）。等価制御部 5 2 は、リカバリ対象のＢＵＤのユーザ領域の対応する領域に、リカバリデータを書き込む（Ｓ 2 1 0 ）。リカバリ属性情報にＢＵＤの識別情報が設定された全ブロックについて、Ｓ 2 0 8 ～Ｓ 2 1 0 の処理を繰り返す（Ｓ 2 1 1 ）。

【 0 1 7 9 】

Ｓ 2 0 3 において、リカバリ対象ＢＵＤの更新タイムスタンプと、正常なＢＵＤの旧更新タイムスタンプとが一致していない場合（Ｓ 2 0 3 で「Ｎ ｏ」）、等価制御部 5 2 は、正常なＢＵＤの等価管理データから「有効（Ｖ）」属性を有するブロック属性情報を抽出する（Ｓ 2 0 4 ）。等価制御部 5 2 は、正常なＢＵＤのユーザ領域から、その「有効（Ｖ）」属性を有するブロック属性情報に対応する領域の情報（リカバリデータ）を読み出す（Ｓ 2 0 5 ）。等価制御部 5 2 は、リカバリ対象のＢＵＤのユーザ領域の対応する領域に、リカバリデータを書き込む（Ｓ 2 0 6 ）。「有効（Ｖ）」属性の全ブロックについて、Ｓ 2 0 4 ～Ｓ 2 0 7 の処理を繰り返す（Ｓ 2 0 7 ）。

【 0 1 8 0 】

「有効（Ｖ）」属性の全ブロックについてＳ 2 0 4 ～Ｓ 2 0 7 の処理を繰り返した場合、またはリカバリ属性情報にＢＵＤの識別情報が設定された全ブロックについてＳ 2 0 8 ～Ｓ 2 1 0 の処理を繰り返した場合、等価制御部 5 2 は、次の処理を行う。すなわち、等価制御部 5 2 は、正常なＢＵＤの等価管理データを、リカバリ対象ＢＵＤへ書き込む（Ｓ 2 1 2 ）。このとき、図 1 2 ＢのＳ 8 8 で説明したように、リカバリ対象ＢＵＤのリカバリ完了後、全ＢＵＤの等価管理データのリカバリ属性情報からリカバリしたＢＵＤを識別する情報をクリアする。

【 0 1 8 1 】

本実施形態におけるリカバリ時のＢＵＤ間のリカバリ処理の詳細なシーケンスは、図 1 2 Ａ - 図 1 2 Ｃと同様なので、その説明を省略する。

【 0 1 8 2 】

このように、リカバリを行う場合、ユーザ領域全域ではなく、部分的な領域（リカバリが必要であることを示す属性情報を有するブロック）について、リカバリを行えばよいので、リカバリに要する時間を短縮することができる。

10

20

30

40

50

【 0 1 8 3 】

また、C M 保守またはB U D 保守において、正常なB U D からリカバリ対象B U D にデータをコピーする際に、リカバリ属性情報を除いて有効なデータについてコピーを行う（無効データについてはコピーは行わない）。そのため、ユーザ領域内の全データのコピーを行う必要がない。その結果、リカバリに要する時間を短縮することができる。

【 0 1 8 4 】

また、コピー処理中にコピー元データの読み出しができなかった場合、その読み出しができない領域が未使用領域であっても、コピー処理していた領域の破棄によるデータロストの心配もない。

【 0 1 8 5 】

なお、本実施形態は、以上に述べた実施の形態に限定されるものではなく、本実施形態の要旨を逸脱しない範囲内で種々の構成または実施形態を取ることができる。

【 0 1 8 6 】

上記実施形態に関し、更に以下の付記を開示する。

(付 記 1)

情報を格納する第 1 格納領域と、前記第 1 格納領域が区分された各区画に格納される情報を示す区画情報の属性に関する属性情報を含む管理情報を格納する第 2 格納領域と、を含む複数の格納部と、

前記複数の格納部から前記管理情報を取得する取得部と、

取得した前記管理情報を比較して異常な前記管理情報があるかを判定する判定部と、

前記異常な管理情報があると判定された場合、正常な管理情報の区画毎の属性情報に基づいて、該正常な管理情報が取得されたいずれかの前記格納部から、前記異常な管理情報が取得された前記格納部の前記第 1 格納領域に格納される情報を、区画単位で再構成するリカバリ部と、

を備えることを特徴とするストレージ制御システム。

(付 記 2)

前記ストレージ制御システムは、さらに、

前記格納部への対象情報の書き込み要求があった場合、前記複数の格納部において、前記第 1 格納領域に該対象情報を書き込み、該対象情報を書き込んだ区画に格納される区画情報に対応する属性情報に、該区画情報が有効であるか否かを示す情報を書き込む書き込み部

を備え、

前記リカバリ部は、前記正常な管理情報が取得されたいずれかの前記格納部から、前記区画情報が有効であることを示す属性情報を有する該区画情報を抽出し、前記異常な管理情報が取得された格納部の前記第 1 格納領域における対応する区画に、抽出した該区画情報を格納する

ことを特徴とする付記 1 に記載のストレージ制御システム。

(付 記 3)

前記書き込み部は、前記書き込みが失敗した格納部がある場合、該書き込みが成功した格納部の管理情報における、前記書き込み要求によるアクセス範囲に対応する前記区画に格納された前記区画情報の前記属性情報に、リカバリが必要であることを示す情報を書き込み、

前記リカバリ部は、前記書き込みが成功した前記格納部から、前記リカバリが必要であることを示す属性情報を有する該区画情報を抽出し、前記書き込みが失敗した格納部の前記第 1 格納領域における対応する区画に、抽出した該区画情報を格納する

ことを特徴とする付記 2 に記載のストレージ制御システム。

(付 記 4)

前記リカバリが必要であることを示す属性情報は、前記書き込みが失敗した格納部を識別する情報である

ことを特徴とする付記 3 に記載のストレージ制御システム。

10

20

30

40

50

(付記 5)

前記属性情報は、前記区画情報の種別情報を含み、

前記書き込み部は、前記書き込み対象情報の書き込み要求があった場合であって、該書き込み要求によるアクセス範囲に対応する前記区画に格納された区画情報の属性情報に含まれる前記種別情報がメタデータを示す場合、前記書き込み対象情報を前記格納部の前記第 1 格納領域とバックアップ領域とに格納し、前記種別情報がメタデータを示さない場合、書き込み対象情報を前記第 1 格納領域に格納する

ことを特徴とする付記 2 ～ 4 のうちいずれか 1 項に記載のストレージ制御システム。

(付記 6)

コンピュータに、

情報を格納する第 1 格納領域と、前記第 1 格納領域が区分された各区画に格納される情報を示す区画情報の属性に関する属性情報を含む管理情報を格納する第 2 格納領域と、を含む、複数の格納部から前記管理情報を取得し、

取得した前記管理情報を比較して異常な前記管理情報があるかを判定し、

前記異常な管理情報があると判定された場合、正常な管理情報の区画毎の属性情報に基づいて、該正常な管理情報が取得されたいずれかの前記格納部から、前記異常な管理情報が取得された前記格納部の前記第 1 格納領域に格納される情報を、区画単位で再構成する処理を実行させることを特徴とするリカバリプログラム。

(付記 7)

前記コンピュータは、さらに、

前記格納部への対象情報の書き込み要求があった場合、前記複数の格納部において、前記第 1 格納領域に該対象情報を書き込み、該対象情報を書き込んだ区画に格納される区画情報に対応する属性情報に、該区画情報が有効であるか否かを示す情報を書き込む

処理を実行させ、

前記情報を再構成する場合、前記正常な管理情報が取得されたいずれかの前記格納部から、前記区画情報が有効であることを示す属性情報を有する該区画情報を抽出し、前記異常な管理情報が取得された格納部の前記第 1 格納領域における対応する区画に、抽出した該区画情報を格納する

ことを特徴とする付記 6 に記載のリカバリプログラム。

(付記 8)

前記書き込みが失敗した格納部がある場合、該書き込みが成功した格納部の管理情報における、前記書き込み要求によるアクセス範囲に対応する前記区画に格納された前記区画情報の前記属性情報に、リカバリが必要であることを示す情報を書き込み、

前記情報を再構成する場合、前記書き込みが成功した前記格納部から、前記リカバリが必要であることを示す属性情報を有する該区画情報を抽出し、前記書き込みが失敗した格納部の前記第 1 格納領域における対応する区画に、抽出した該区画情報を格納する

ことを特徴とする付記 7 に記載のリカバリプログラム。

(付記 9)

前記リカバリが必要であることを示す属性情報は、前記書き込みが失敗した格納部を識別する情報である

ことを特徴とする付記 8 に記載のリカバリプログラム。

(付記 10)

前記属性情報は、前記区画情報の種別情報を含み、

前記書き込み対象情報の書き込み要求があった場合であって、該書き込み要求によるアクセス範囲に対応する前記区画に格納された区画情報の属性情報に含まれる前記種別情報がメタデータを示す場合、前記書き込み対象情報を前記格納部の前記第 1 格納領域とバックアップ領域とに格納し、前記種別情報がメタデータを示さない場合、書き込み対象情報を前記第 1 格納領域に格納する

ことを特徴とする付記 6 ～ 9 のうちいずれか 1 項に記載のリカバリプログラム。

(付記 11)

10

20

30

40

50

コンピュータが、

情報を格納する第 1 格納領域と、前記第 1 格納領域が区分された各区画に格納される情報を示す区画情報の属性に関する属性情報を含む管理情報を格納する第 2 格納領域と、を含む、複数の格納部から前記管理情報を取得し、

取得した前記管理情報を比較して異常な前記管理情報があるかを判定し、

前記異常な管理情報があると判定された場合、正常な管理情報の区画毎の属性情報に基づいて、該正常な管理情報が取得されたいずれかの前記格納部から、前記異常な管理情報が取得された前記格納部の前記第 1 格納領域に格納される情報を、区画単位で再構成する処理を実行することを特徴とするリカバリ方法。

(付記 1 2)

10

前記コンピュータは、さらに、

前記格納部への対象情報の書き込み要求があった場合、前記複数の格納部において、前記第 1 格納領域に該対象情報を書き込み、該対象情報を書き込んだ区画に格納される区画情報に対応する属性情報に、該区画情報が有効であるか否かを示す情報を書き込む

処理を実行させ、

前記情報を再構成する場合、前記正常な管理情報が取得されたいずれかの前記格納部から、前記区画情報が有効であることを示す属性情報を有する該区画情報を抽出し、前記異常な管理情報が取得された格納部の前記第 1 格納領域における対応する区画に、抽出した該区画情報を格納する

ことを特徴とする付記 1 1 に記載のリカバリ方法。

20

(付記 1 3)

前記書き込みが失敗した格納部がある場合、該書き込みが成功した格納部の管理情報における、前記書き込み要求によるアクセス範囲に対応する前記区画に格納された前記区画情報の前記属性情報に、リカバリが必要であることを示す情報を書き込み、

前記情報を再構成する場合、前記書き込みが成功した前記格納部から、前記リカバリが必要であることを示す属性情報を有する該区画情報を抽出し、前記書き込みが失敗した格納部の前記第 1 格納領域における対応する区画に、抽出した該区画情報を格納する

ことを特徴とする付記 1 2 に記載のリカバリ方法。

(付記 1 4)

前記リカバリが必要であることを示す属性情報は、前記書き込みが失敗した格納部を識別する情報である

30

ことを特徴とする付記 1 3 に記載のリカバリ方法。

(付記 1 5)

前記属性情報は、前記区画情報の種別情報を含み、

前記書き込み対象情報の書き込み要求があった場合であって、該書き込み要求によるアクセス範囲に対応する前記区画に格納された区画情報の属性情報に含まれる前記種別情報がメタデータを示す場合、前記書き込み対象情報を前記格納部の前記第 1 格納領域とバックアップ領域とに格納し、前記種別情報がメタデータを示さない場合、書き込み対象情報を前記第 1 格納領域に格納する

ことを特徴とする付記 1 1 ~ 1 4 のうちいずれか 1 項に記載のリカバリ方法。

40

【符号の説明】

【0 1 8 7】

2 1 ストレージ制御システム

2 2 格納部

2 3 取得部

2 4 判定部

2 5 リカバリ部

2 6 書き込み部

3 1 ストレージシステム

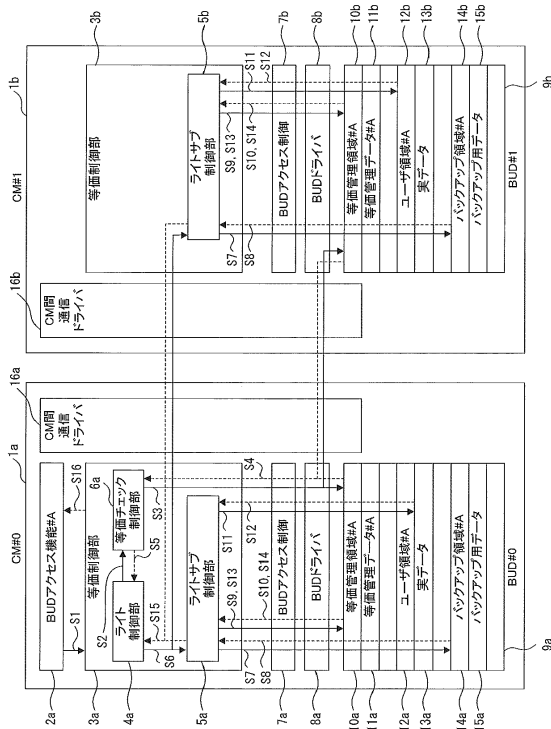
3 2 コントローラエンクロージャ (C E)

50

3 3	コントローラモジュール (C M)	
3 4	チャネルアダプタ (C A)	
3 5	C P U	
3 6	記憶部	
3 7	デバイスアダプタ (D A)	
3 8	C M間通信インターフェース (I / F)	
3 9	ディスク装置	
4 0	ドライブエンクロージャ (D E)	
4 1	ホスト	
4 2	ファイバチャネル (F C) スイッチ	10
5 1 (5 1 A ~ 5 1 N)	機能制御部	
5 2	等価制御部	
5 3	ライト制御部	
5 4	リード制御部	
5 5	リカバリ制御部	
5 6	ライトサブ制御部	
5 7	リードサブ制御部	
5 8	ロック / アンロック制御部	
5 9	等価チェック制御部	
6 0	ブロック属性管理制御部	20
6 1	ブロック属性解析制御部	
6 2	B U D アクセス制御部	
6 3	B U D ドライバ	
6 4	B U D	
6 5	等価管理領域	
6 6	ユーザ領域	
6 7	バックアップ領域	
6 8	C M間通信ドライバ制御部	

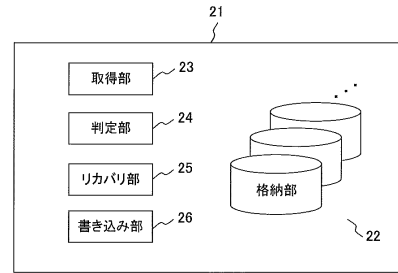
【図 1】

BUD間のデータの整合性を保持する
等価制御について説明するための図



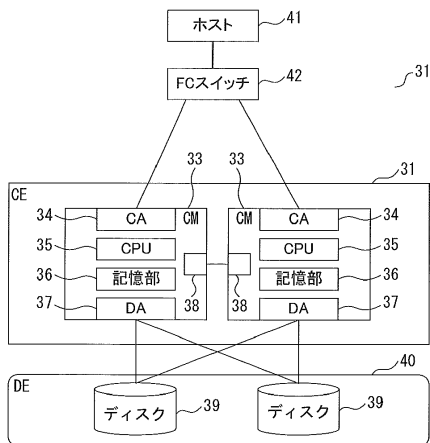
【図 2】

本実施形態における、
ストレージ制御システムの一例を示す図



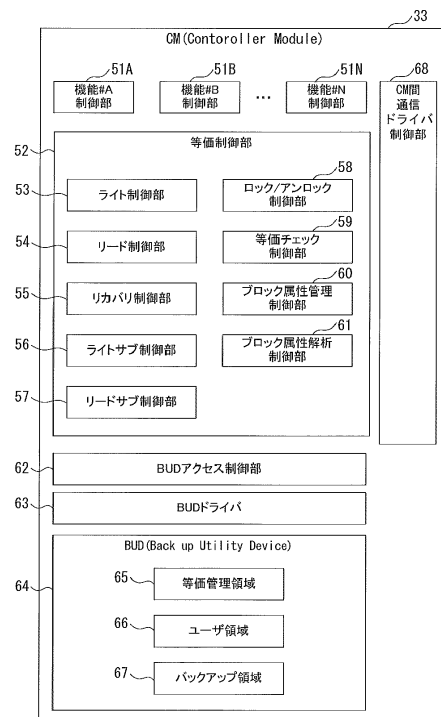
【図 3】

本実施形態におけるストレージシステムの
ハードウェア構成を示す図

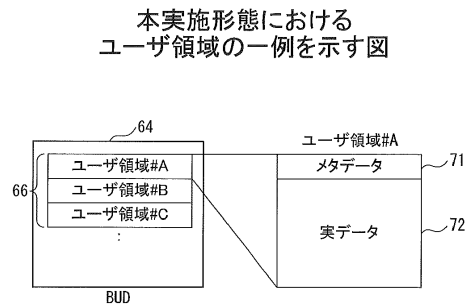


【図 4】

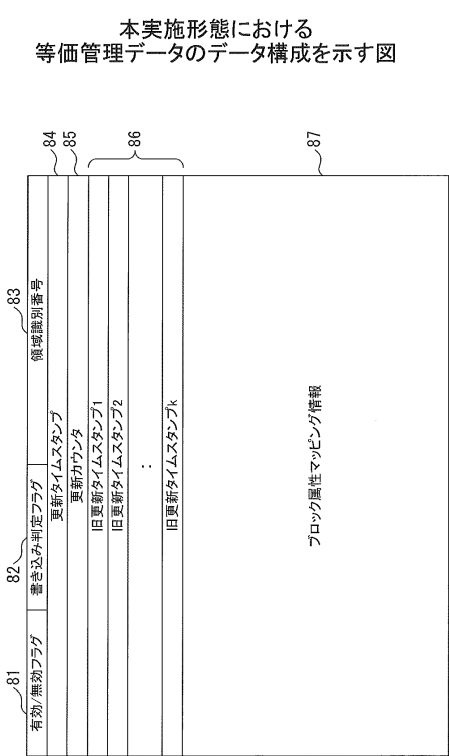
本実施形態におけるCMの機能ブロック図



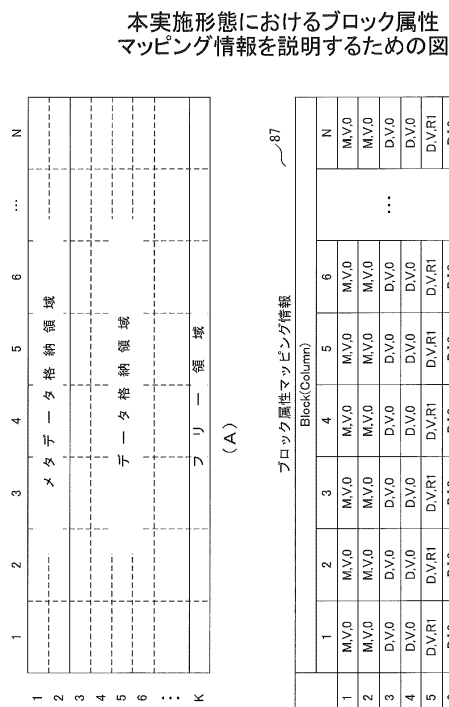
【図 5】



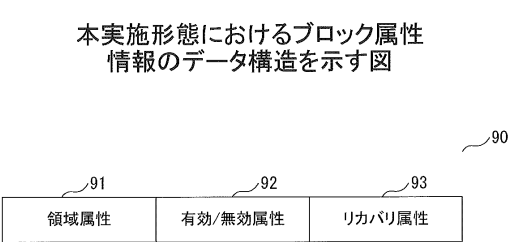
【図 6】



【図 7】

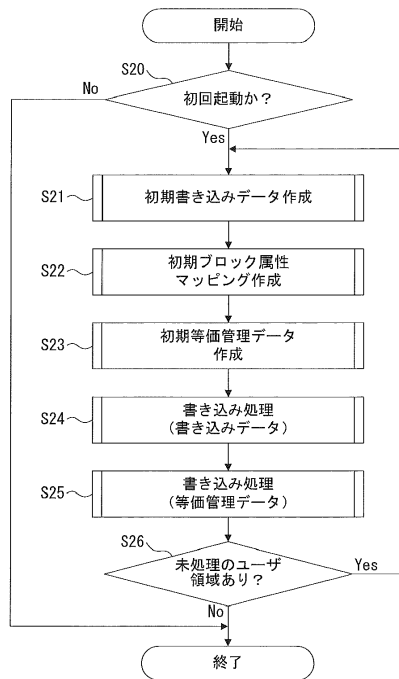


【図 8】



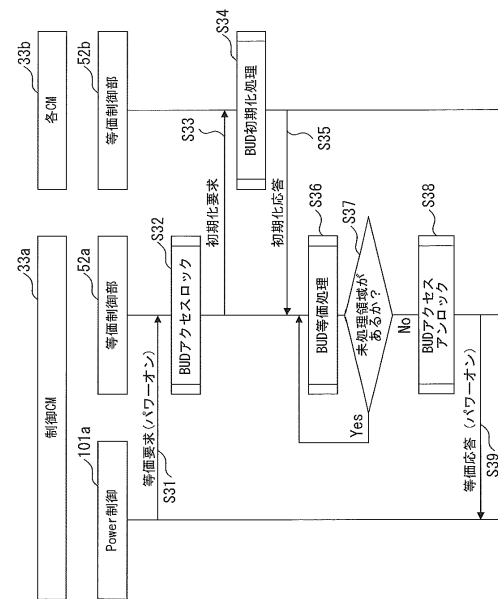
【図 9】

本実施形態におけるストレージ装置の
初期起動時の等価制御部の動作フロー



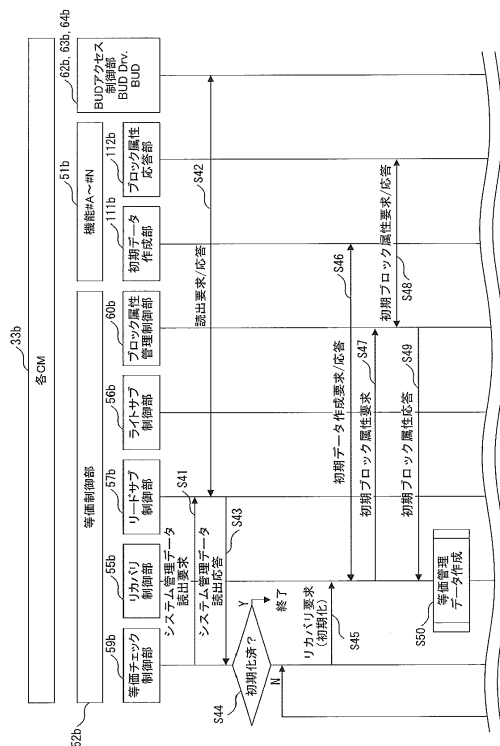
【図 10】

本実施形態におけるストレージ装置の
初期起動時のCM間の動作シーケンスの全体図



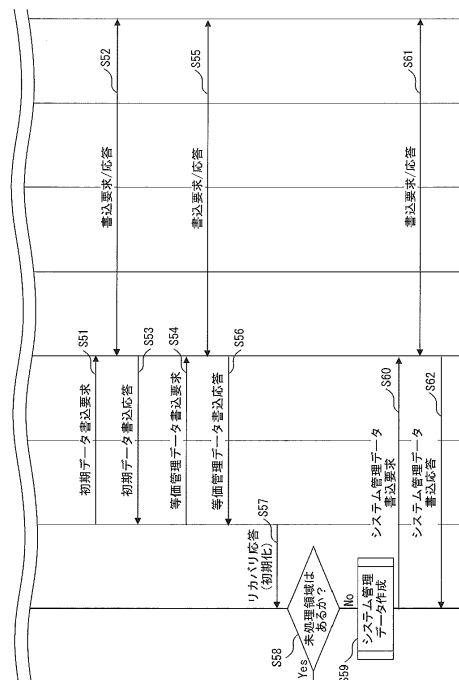
【図 11 A】

本実施形態におけるBUDの初期化処理
(S34)の詳細なシーケンス図(その1)



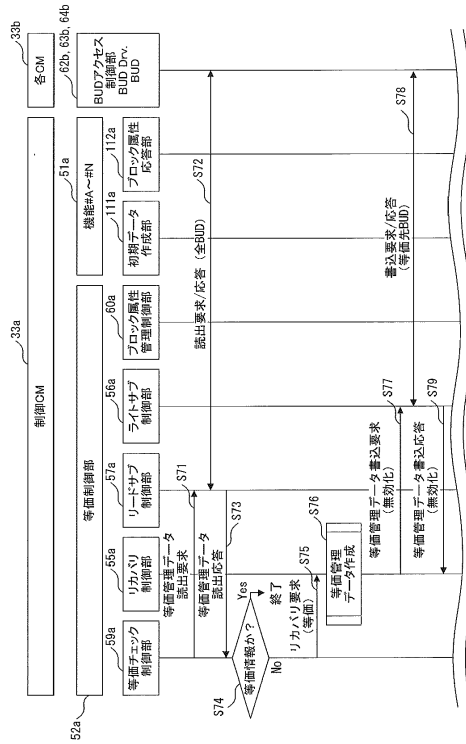
【図 11 B】

本実施形態におけるBUDの初期化処理
(S34)の詳細なシーケンス図(その2)



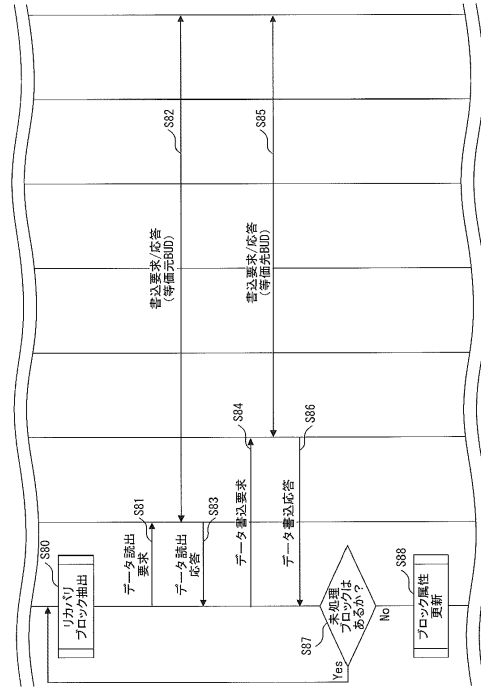
【図 1 2 A】

本実施形態におけるBUD間の等価処理
(S36)の詳細なシーケンス図(その1)



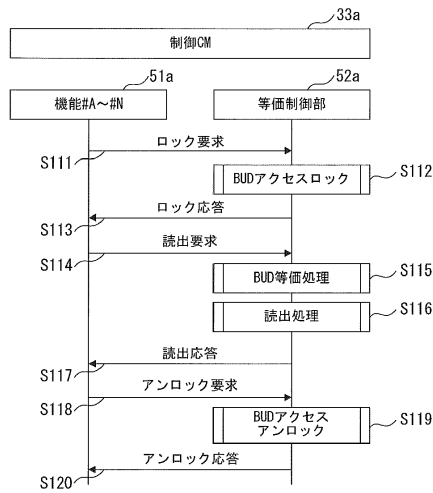
【図 1 2 B】

本実施形態におけるBUD間の等価処理
(S36)の詳細なシーケンス図(その2)



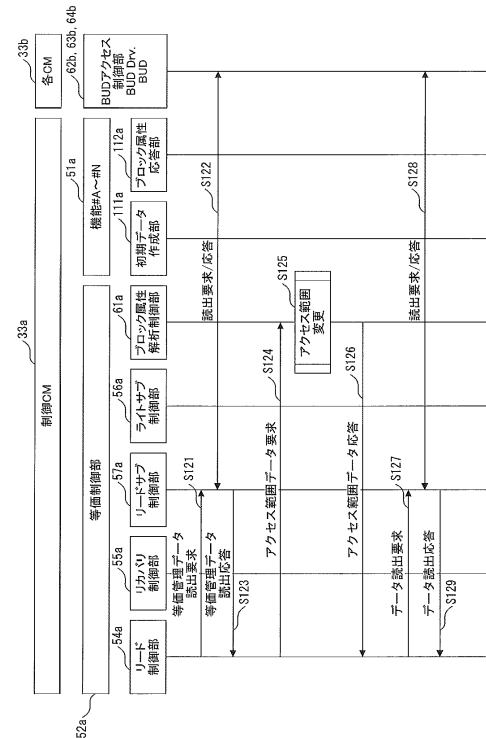
【図 14】

本実施形態における機能制御部から
BUDへ読み出し処理を行う場合の
CM間の動作シーケンスの全体図



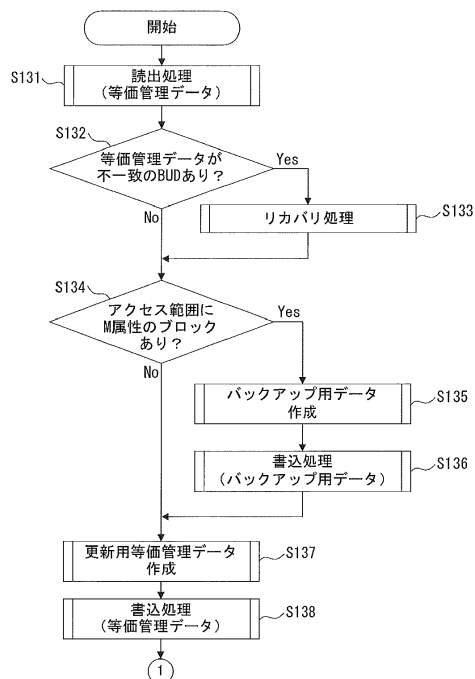
【図 15】

本実施形態における読み出し処理
(S116)の詳細なシーケンス図



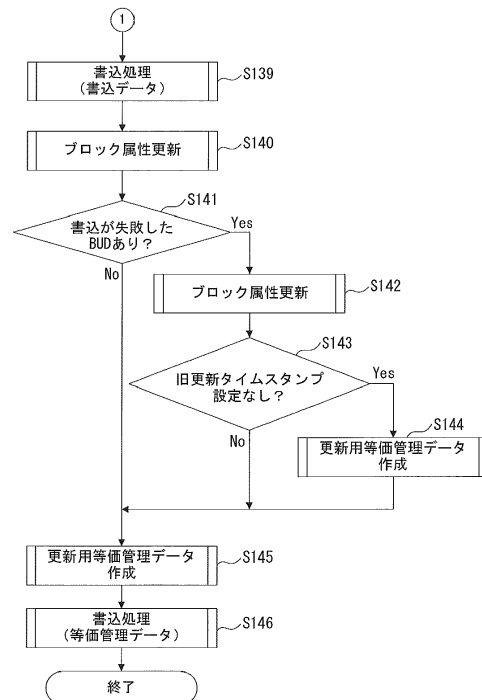
【図 16 A】

本実施形態における各機能制御部から
BUDへ書き込み処理を行う場合の
等価制御部の動作フロー(その1)

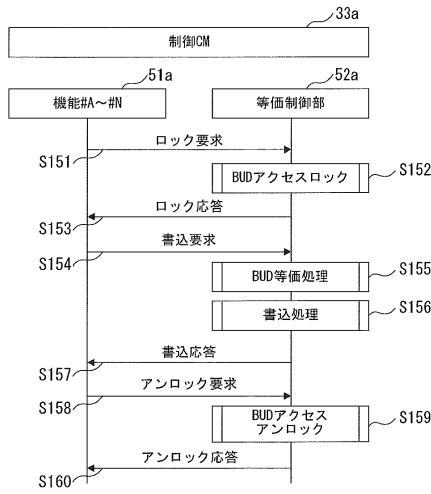


【図 16 B】

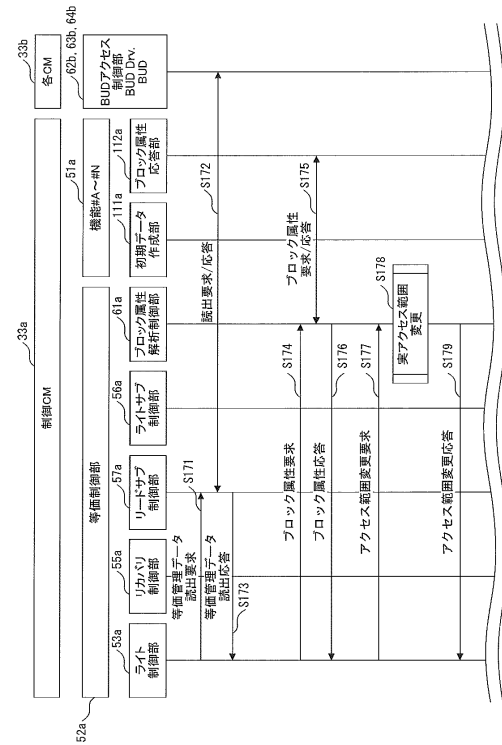
本実施形態における各機能制御部から
BUDへ書き込み処理を行う場合の
等価制御部の動作フロー(その2)



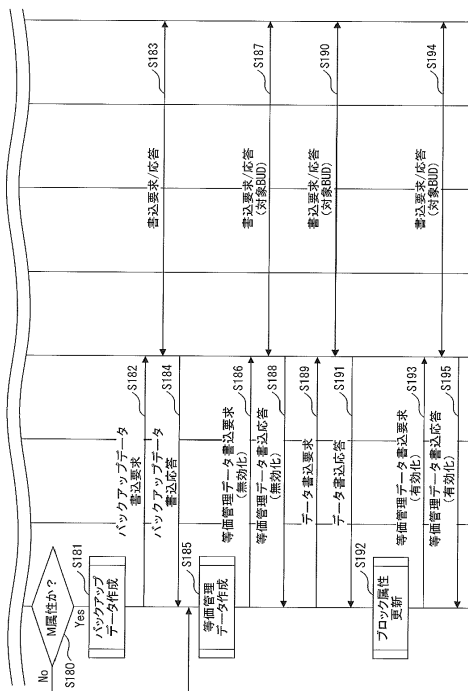
【図 17】

本実施形態における
書き込み処理シーケンスの全体図

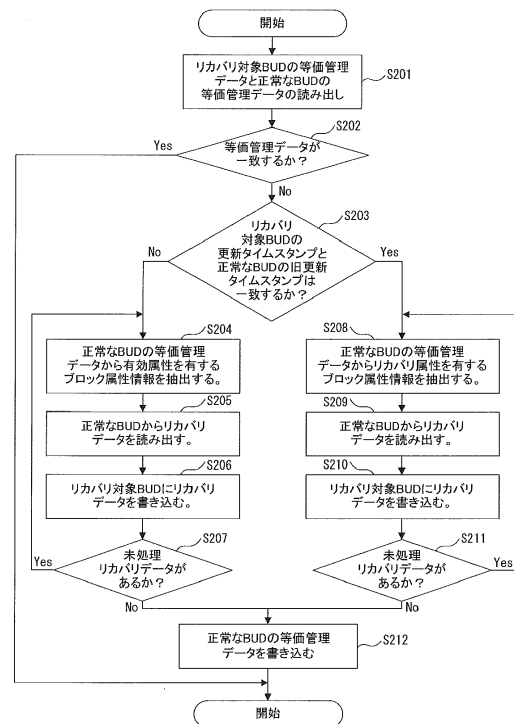
【図 18 A】

本実施形態における書き込み処理
(S156)の詳細なシーケンス図(その1)

【図 18 B】

本実施形態における書き込み処理
(S156)の詳細なシーケンス図(その2)

【図 19】

本実施形態におけるリカバリ処理を
行う場合の等価制御部の動作フロー

フロントページの続き

- (72)発明者 室山 智彦
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
- (72)発明者 曾谷 基紀
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

審査官 桜井 茂行

- (56)参考文献 特開2011-258103(JP,A)
特開2004-094789(JP,A)
特開平02-312068(JP,A)
特開2005-149248(JP,A)

- (58)調査した分野(Int.Cl., DB名)
- | | |
|---------|-------------------|
| G 0 6 F | 3 / 0 6 - 3 / 0 8 |
| G 0 6 F | 1 2 / 0 0 |
| G 0 6 F | 1 2 / 1 6 |
| G 0 6 F | 1 1 / 0 0 |