(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2018-181202
(P2018-181202A)
(43)公開日　平成30年11月15日(2018.11.15)

(51)Int.Cl.		F I			テーマコード（参考）
G06F 3/06	(2006.01)	G06F	3/06	301K	5B060
G06F 3/08	(2006.01)	G06F	3/08	H	5B160
G06F 12/00	(2006.01)	G06F	3/06	301R	
G06F 12/02	(2006.01)	G06F	12/00	597U	
		G06F	12/02	570A	

審査請求 未請求　請求項の数 8　ＯＬ　（全 18 頁）　最終頁に続く

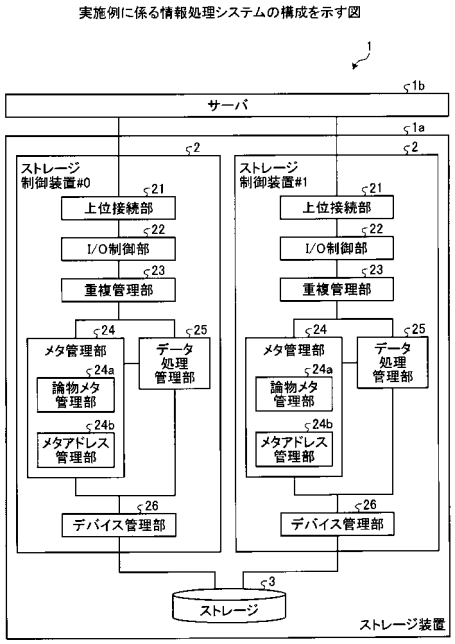
(21)出願番号　特願2017-83857(P2017-83857)
(22)出願日　平成29年4月20日(2017.4.20)

(71)出願人　000005223
富士通株式会社
神奈川県川崎市中原区上小田中４丁目１番１号
(74)代理人　110002147
特許業務法人酒井国際特許事務所
(72)発明者　武田　直浩
神奈川県川崎市中原区上小田中４丁目１番１号　株式会社富士通コンピュータテクノロジーズ内
(72)発明者　久保田　典秀
神奈川県川崎市中原区上小田中４丁目１番１号　株式会社富士通コンピュータテクノロジーズ内

最終頁に続く

(54)【発明の名称】ストレージ制御装置、ストレージ制御方法及びストレージ制御プログラム

(57)【要約】
【課題】ＳＳＤへの書き込み回数を減少させること。
【解決手段】メタ管理部２４は、メタアドレス及び論物メタを用いて、仮想ボリュームの論理アドレスとＳＳＤの物理アドレスの変換処理を行う。メタ管理部２４は、論物メタ管理部２４ａと、メタアドレス管理部２４ｂとを有する。論物メタ管理部２４ａが、論理アドレスと物理アドレスを対応付ける論物メタの情報を管理する。データ処理管理部２５は、論物メタの情報をＲＡＩＤユニットの単位でＳＳＤへ追記及びまとめ書きする。
【選択図】図８



【特許請求の範囲】
【請求項１】
書き込み回数に制限を有する記憶媒体を用いる記憶装置を制御するストレージ制御装置において、
前記記憶装置を使用する情報処理装置がデータの識別に用いる論理アドレスと前記記憶媒体上の該データの記憶位置を示す物理アドレスとを対応付けるアドレス変換情報を管理する変換情報管理部と、
前記変換情報管理部により管理されるアドレス変換情報を前記記憶媒体に追記及びまとめ書きを行う書込処理部と
を有することを特徴とするストレージ制御装置。
【請求項２】
前記変換情報管理部により管理されるアドレス変換情報が前記記憶媒体に追記及びまとめ書きされた位置を示す物理アドレスをメタアドレスとして前記論理アドレスと対応付けて管理する変換情報位置管理部をさらに有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装置。
【請求項３】
前記物理アドレスに記憶されたデータには、該データを参照する論理アドレスを示す参照情報が付加されることを特徴とする請求項１又は２に記載のストレージ制御装置。
【請求項４】
前記書込処理部は、前記データの更新によって新たなデータが追記された場合に、更新前の前記データとともに前記参照情報を前記記憶媒体上に維持することを特徴とする請求項３に記載のストレージ制御装置。
【請求項５】
前記メタアドレスは、前記記憶媒体の所定の位置に記憶されることを特徴とする請求項２に記載のストレージ制御装置。
【請求項６】
前記記憶媒体は、ＳＳＤであることを特徴とする請求項１～５のいずれか１つに記載のストレージ制御装置。
【請求項７】
書き込み回数に制限を有する記憶媒体を用いる記憶装置を制御するストレージ制御装置によるストレージ制御方法において、
前記記憶装置を使用する情報処理装置がデータの識別に用いる論理アドレスと前記記憶媒体上の該データの記憶位置を示す物理アドレスとを対応付けるアドレス変換情報を管理し、
前記アドレス変換情報を前記記憶媒体に追記及びまとめ書きを行う
ことを特徴とするストレージ制御方法。
【請求項８】
書き込み回数に制限を有する記憶媒体を用いる記憶装置を制御するストレージ制御装置が有するコンピュータにより実行されるストレージ制御プログラムにおいて、
前記記憶装置を使用する情報処理装置がデータの識別に用いる論理アドレスと前記記憶媒体上の該データの記憶位置を示す物理アドレスとを対応付けるアドレス変換情報を管理し、
前記アドレス変換情報を前記記憶媒体に追記及びまとめ書きを行う
処理を前記コンピュータに実行させることを特徴とするストレージ制御プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、ストレージ制御装置、ストレージ制御方法及びストレージ制御プログラムに関する。
【背景技術】

【０００２】
昨今、ストレージ装置の記憶媒体は、ＨＤＤ（Hard Disk Drive）からよりアクセススピードの速いＳＳＤ（Solid State Drive）等のフラッシュメモリに移行している。ＳＳＤでは、メモリセルへの上書きを直接行うことはできず、例えば１ＭＢ（メガバイト）の大きさのブロックの単位でデータの消去が行われた後にデータの書き込みが行われる。
【０００３】
このため、ブロック内の一部のデータを更新する場合には、ブロック内の他のデータを退避し、ブロックを消去した後に、退避したデータと更新データを書き込むことが行われるため、ブロックの大きさに較べて小さいデータを更新する処理が遅い。また、ＳＳＤには書き込み回数の上限がある。このため、ＳＳＤでは、ブロックの大きさに較べて小さいデータの更新をできるだけ避けることが望ましい。そこで、ブロック内の一部のデータを更新する場合に、ブロック内の他のデータと更新データを新たなブロックに書き込むことが行われる。
【０００４】
しかし、新たなブロックを用いてデータの更新が行われると、データを記憶する物理アドレスが変更されるので、論理アドレスと物理アドレスを対応付ける管理データ（メタデータ）の更新が必要となる。また、ストレージ装置では、データの書き込み容量を削減するために、重複するデータブロックの排除が行われるが、重複排除（Deduplication）のための管理データの更新も必要となる。
【０００５】
なお、ログ構造化ファイルシステムにおいて、ストレージ装置を第１領域と第２領域とに分けて、第１領域と第２領域を以下のように使用する技術がある。第２領域には、多数のデータと、多数のデータと関連する多数のノードとが格納される。第１領域には、多数のノード各々に対応する多数のノード識別子と、多数のノード識別子各々に対応する多数の物理アドレスとを含むノードアドレステーブルが格納される。この技術によれば、メタデータの修正のための付加的な書き込み動作を減らすことができる。
【０００６】
また、ランダムライトアクセスが行われた時に、不使用ページに基づいて選択されたブロックのページに記録されたデータをバッファに書き込み、ブロックを消去後にバッファに書き込まれたデータをブロックに書き込む技術がある。この技術によれば、ガベージコレクションを行わないので、ＩＯＰＳ（Input Output Per Second）性能を向上させることができる。
【０００７】
また、Ｎ台のディスク装置から構成されるディスク記憶装置において、Ｎ×Ｋ個の論理ブロックに相当する容量を持つ書き込みバッファを備え、書き込みバッファに更新すべきデータの論理ブロックを蓄積し、制御装置が、以下の制御を行う技術がある。すなわち、制御装置は、蓄積した論理ブロックがＮ×Ｋ－１個に達するまで論理ブロックの更新を遅延させ、Ｎ×Ｋ－１個に達すると当該論理ブロックの論理アドレスタグブロックを加えたＮ×Ｋ個の論理ブロックを空領域に連続して順次書き込む。この技術によれば、論理アドレスと物理アドレスのマップを原理的に不要とすることで、安価で高速なディスク記憶装置を構築することができる。
【先行技術文献】
【特許文献】
【０００８】
【特許文献１】特開２０１４－７１９０６号公報
【特許文献２】特開２０１０－２３７９０７号公報
【特許文献３】特開平１１－５３２３５号公報
【発明の概要】
【発明が解決しようとする課題】

【０００９】
論理アドレスと物理アドレスを対応付ける管理データ、重複排除のための管理データ等の更新では、ブロック内の一部のデータが更新されるので、管理データはメインメモリ上に配置することが望まれる。しかしながら、管理データが大きくなるとメインメモリ上に全ての管理データを保持することができない。このため、管理データをＳＳＤに書き込むことになるが、管理データの更新によってＳＳＤへの書き込み回数が多くなるという問題がある。
【００１０】
本発明は、１つの側面では、管理データの更新によるＳＳＤへの書き込み回数を減少させることを目的とする。
【課題を解決するための手段】
【００１１】
１つの態様では、ストレージ制御装置は、書き込み回数に制限を有する記憶媒体を用いる記憶装置を制御し、変換情報管理部と書込処理部とを有する。変換情報管理部は、記憶装置を使用する情報処理装置がデータの識別に用いる論理アドレスと記憶媒体上の該データの記憶位置を示す物理アドレスとを対応付けるアドレス変換情報を管理する。書込処理部は、変換情報管理部により管理されるアドレス変換情報を記憶媒体に追記及びまとめ書きを行う。
【発明の効果】
【００１２】
１つの側面では、本発明は、管理データの更新によるＳＳＤへの書き込み回数を減少させることができる。
【図面の簡単な説明】
【００１３】
【図１】図１は、実施例に係るストレージ装置の記憶構成を示す図である。
【図２】図２は、ＲＡＩＤユニットのフォーマットを示す図である。
【図３】図３は、参照メタのフォーマットを示す図である。
【図４】図４は、論物メタのフォーマットを示す図である。
【図５】図５は、実施例に係るメタメタ方式を説明するための図である。
【図６】図６は、メタアドレスのフォーマットを示す図である。
【図７】図７は、ドライブグループにおけるＲＡＩＤユニットの配置例を示す図である。
【図８】図８は、実施例に係る情報処理システムの構成を示す図である。
【図９】図９は、機能部間の関係を示す図である。
【図１０Ａ】図１０Ａは、重複のないデータの書き込み処理のシーケンスを示す図である。
【図１０Ｂ】図１０Ｂは、重複のあるデータの書き込み処理のシーケンスを示す図である。
【図１１】図１１は、読み出し処理のシーケンスを示す図である。
【図１２Ａ】図１２Ａは、メタメタ方式導入前のスモールライトの数を示す図である。
【図１２Ｂ】図１２Ｂは、メタメタ方式でのスモールライトの数を示す図である。
【図１２Ｃ】図１２Ｃは、メタメタ方式で旧データ無効化を行わない場合のスモールライトの数を示す図である。
【図１３】図１３は、実施例に係るストレージ制御プログラムを実行するストレージ制御装置のハードウェア構成を示す図である。
【発明を実施するための形態】
【００１４】
以下に、本願の開示するストレージ制御装置、ストレージ制御方法及びストレージ制御プログラムの実施例を図面に基づいて詳細に説明する。なお、この実施例は開示の技術を限定するものではない。
【実施例】

【００１５】
まず、実施例に係るストレージ装置のデータ管理方法について図１～図７を用いて説明する。図１は、実施例に係るストレージ装置の記憶構成を示す図である。図１に示すように、実施例に係るストレージ装置は、ＲＡＩＤ（Redundant Arrays of Inexpensive Disks）６ベースのプール３ａとして複数のＳＳＤ３ｄを管理する。また、実施例に係るストレージ装置は、複数のプール３ａを有する。
【００１６】
プール３ａには、仮想化プールと階層化プールがある。仮想化プールは１つのティア３ｂを有し、階層化プールは２つ以上のティア３ｂを有する。ティア３ｂは、１つ以上のドライブグループ３ｃを有する。ドライブグループ３ｃは、ＳＳＤ３ｄのグループであり、６～２４台のＳＳＤ３ｄを有する。例えば、１つのストライプを記憶する６台のＳＳＤ３ｄのうち、３台はデータ記憶用に用いられ、２台はパリティ記憶用に用いられ、１台はホットスペア用に用いられる。なお、ドライブグループ３ｃは、２５台以上のＳＳＤ３ｄを有してよい。
【００１７】
実施例に係るストレージ装置は、ＲＡＩＤユニットの単位でデータを管理する。シン・プロビジョニングの物理割当の単位は、一般に固定サイズのチャンク単位で行われ、１チャンクは１ＲＡＩＤユニットに該当する。以降の説明では、チャンクをＲＡＩＤユニットと呼称する。ＲＡＩＤユニットは、プール３ａから割り当てられる２４ＭＢの連続な物理領域である。実施例に係るストレージ装置は、データをＲＡＩＤユニット単位でメインメモリ上にバッファリングし、追記型でＳＳＤ３ｄに書き込む。
【００１８】
図２は、ＲＡＩＤユニットのフォーマットを示す図である。図２に示すように、ＲＡＩＤユニットには、複数のユーザデータユニット(データログとも呼ばれる。)が含まれる。ユーザデータユニットには、参照メタと圧縮データが含まれる。参照メタは、ＳＳＤ３ｄに書き込まれるデータの管理データである。
【００１９】
圧縮データは、ＳＳＤ３ｄに書き込まれるデータが圧縮されたものである。データの大きさは最大８ＫＢ（キロバイト）である。圧縮率を５０％とすると、実施例に係るストレージ装置は、１つのＲＡＩＤユニットに、例えば２４ＭＢ÷４．５ＫＢ　５４６１個のユーザデータユニットが溜まると、ＲＡＩＤユニットをＳＳＤ３ｄに書き込む。
【００２０】
図３は、参照メタのフォーマットを示す図である。図３（ａ）に示すように、参照メタには、ＳＢ（Super Block）と最大６０個の参照先の参照ＬＵＮ（Logical Unit Number：論理ユニット番号）／ＬＢＡ（Logical Block Address：論理ブロックアドレス）情報を書き込むことができる記憶容量の領域が確保されている。ＳＢのサイズは３２Ｂ（バイト）であり、参照メタのサイズは５１２Ｂ（バイト）である。各参照ＬＵＮ／ＬＢＡ情報のサイズは８Ｂ（バイト）である。参照メタは、重複排除により新しい参照先ができると、参照先が追加となり、参照メタが更新される。ただし、データの更新により参照先がなくなった場合にも参照ＬＵＮ／ＬＢＡ情報は削除されないで保持される。無効になった参照ＬＵＮ／ＬＢＡ情報はガベージコレクションにより回収される。
【００２１】
図３（ｂ）に示すように、ＳＢには、４ＢのＨｅａｄｅｒ　Ｌｅｎｇｔｈと、２０ＢのＨａｓｈ　Ｖａｌｕｅと、２ＢのＮｅｘｔ　Ｏｆｆｓｅｔ　Ｂｌｏｃｋ　Ｃｏｕｎｔが含まれる。Ｈｅａｄｅｒ　Ｌｅｎｇｔｈは、参照メタの長さである。Ｈａｓｈ　Ｖａｌｕｅは、データのハッシュ値であり、重複排除のために用いられる。Ｎｅｘｔ　Ｏｆｆｓｅｔ　Ｂｌｏｃｋ　Ｃｏｕｎｔは、次に格納する参照ＬＵＮ／ＬＢＡ情報の位置である。なお、Ｒｅｓｅｒｖｅｄは、将来の拡張用である。
【００２２】
図３（ｃ）に示すように、参照ＬＵＮ／ＬＢＡ情報には、２ＢのＬＵＮと、６ＢのＬＢ

Ａが含まれる。
【００２３】
また、実施例に係るストレージ装置は、論物変換情報である論物メタを用いてデータの論理アドレスと物理アドレスの対応関係を管理する。図４は、論物メタのフォーマットを示す図である。実施例に係るストレージ装置は、８ＫＢのデータ毎に、図４に示した情報を管理する。
【００２４】
図４に示すように、論物メタの大きさは３２Ｂである。論物メタには、２ＢのＬＵＮと、６ＢのＬＢＡがデータの論理アドレスとして含まれる。また、論物メタには、２ＢのＣｏｍｐｒｅｓｓｉｏｎ Ｂｙｔｅ Ｃｏｕｎｔが、圧縮されたデータのバイト数として含まれる。
【００２５】
また、論物メタには、２ＢのＮｏｄｅ Ｎｏと、１ＢのＳｔｏｒａｇｅ Ｐｏｏｌ Ｎｏと、４ＢのＲＡＩＤ Ｕｎｉｔ Ｎｏと、２ＢのＲＡＩＤ Ｕｎｉｔ Ｏｆｆｓｅｔ ＬＢＡが物理アドレスとして含まれる。
【００２６】
Ｎｏｄｅ Ｎｏは、ユーザデータユニットを記憶するＲＡＩＤユニットが属するプール３ａを担当するストレージ制御装置を識別するための番号である。なお、ストレージ制御装置については後述する。Ｓｔｏｒａｇｅ Ｐｏｏｌ Ｎｏは、ユーザデータユニットを記憶するＲＡＩＤユニットが属するプール３ａを識別するための番号である。ＲＡＩＤ Ｕｎｉｔ Ｎｏは、ユーザデータユニットを記憶するＲＡＩＤユニットを識別するための番号である。ＲＡＩＤ Ｕｎｉｔ Ｏｆｆｓｅｔ ＬＢＡは、ユーザデータユニットのＲＡＩＤユニット内でのアドレスである。
【００２７】
実施例に係るストレージ装置は、ＲＡＩＤユニットの単位で論物メタを管理する。実施例に係るストレージ装置は、論物メタをＲＡＩＤユニット単位でメインメモリ上にバッファリングし、バッファに例えば７８６４３２エントリ溜まると、論物メタを追記型でＳＳＤ３ｄにまとめ書きする。このため、実施例に係るストレージ装置は、論物メタがある場所を示す情報をメタメタ方式で管理する。
【００２８】
図５は、実施例に係るメタメタ方式を説明するための図である。図５（ｄ）に示すように、（１）、（２）、（３）、・・・で表されるユーザデータユニットは、ＲＡＩＤユニットの単位でＳＳＤ３ｄにまとめ書きされる。そして、図５（ｃ）に示すように、ユーザデータユニットの位置を示す論物メタも、ＲＡＩＤユニットの単位でＳＳＤ３ｄにまとめ書きされる。
【００２９】
そして、実施例に係るストレージ装置は、図５（ａ）に示すように、論物メタの位置をＬＵＮ／ＬＢＡ毎にメタアドレスを用いてメインメモリ上で管理する。ただし、図５（ｂ）に示すように、メインメモリから溢れたメタアドレス情報は、外部キャッシュ(２次キャッシュ)される。ここで、外部キャッシュとは、ＳＳＤ３ｄでのキャッシュである。
【００３０】
図６は、メタアドレスのフォーマットを示す図である。図６に示すように、メタアドレスの大きさは８Ｂである。メタアドレスには、Ｓｔｏｒａｇｅ Ｐｏｏｌ Ｎｏと、ＲＡＩＤ Ｕｎｉｔ Ｏｆｆｓｅｔ ＬＢＡと、ＲＡＩＤ Ｕｎｉｔ Ｎｏとが含まれる。メタアドレスは、論物データのＳＳＤ３ｄでの格納位置を示す物理アドレスである。
【００３１】
Ｓｔｏｒａｇｅ Ｐｏｏｌ Ｎｏは、論物メタを記憶するＲＡＩＤユニットが属するプール３ａを識別するための番号である。ＲＡＩＤ Ｕｎｉｔ Ｏｆｆｓｅｔ ＬＢＡは、論物メタのＲＡＩＤユニット内のアドレスである。ＲＡＩＤ Ｕｎｉｔ Ｎｏは、論物メタを記憶するＲＡＩＤユニットを識別するための番号である。

【００３２】
５１２個のメタアドレスがメタアドレスページ（４ＫＢ）として管理され、メタアドレスページの単位でメインメモリ上にキャッシングされる。また、メタアドレス情報は、ＲＡＩＤユニットの単位で例えばＳＳＤ３ｄの先頭から記憶される。
【００３３】
図７は、ドライブグループ３ｃにおけるＲＡＩＤユニットの配置例を示す図である。図７に示すように、メタアドレスを記憶するＲＡＩＤユニットは、先頭に配置される。図７では、番号が「０」～「１２」のＲＡＩＤユニットが、メタアドレスを記憶するＲＡＩＤユニットである。メタアドレスの更新があった場合は、メタアドレスを記憶するＲＡＩＤユニットは上書き保存される。
【００３４】
論物メタを記憶するＲＡＩＤユニット及びユーザデータユニットを記憶するＲＡＩＤユニットは、それぞれのバッファがいっぱいになると順番にドライブグループに書き出される。図７では、ドライブグループにおいて、番号が「１３」、「１７」、「２７」、「４０」、「５１」、「６３」及び「７０」のＲＡＩＤユニットが、論物メタを記憶するＲＡＩＤユニットであり、その他のＲＡＩＤユニットが、ユーザデータユニットを記憶するＲＡＩＤユニットである。
【００３５】
実施例に係るストレージ装置は、メタメタ方式によって最低限の情報をメインメモリに保持し、論物メタとユーザデータユニットをＳＳＤ３ｄに追記及びまとめ書きすることでＳＳＤ３ｄへの書き込み回数を削減することができる。
【００３６】
次に、実施例に係る情報処理システムの構成について説明する。図８は、実施例に係る情報処理システムの構成を示す図である。図８に示すように、実施例に係る情報処理システム１は、ストレージ装置１ａとサーバ１ｂとを有する。ストレージ装置１ａは、サーバ１ｂが使用するデータを記憶する装置である。サーバ１ｂは、情報処理などの業務を行う情報処理装置である。ストレージ装置１ａとサーバ１ｂとの間は、ＦＣ（Fibre Channel）及びｉＳＣＳＩ（Internet Small Computer System Interface）で接続される。
【００３７】
ストレージ装置１ａは、ストレージ装置１ａを制御するストレージ制御装置２とデータを記憶するストレージ（記憶装置）３とを有する。ここで、ストレージ３は、複数台の記憶装置（ＳＳＤ）３ｄの集まりである。
【００３８】
なお、図８では、ストレージ装置１ａは、ストレージ制御装置＃０及びストレージ制御装置＃１で表される２台のストレージ制御装置２を有するが、ストレージ装置１ａは、３台以上のストレージ制御装置２を有してよい。また、図８では、情報処理システム１は、１台のサーバ１ｂを有するが、情報処理システム１は、２台以上のサーバ１ｂを有してよい。
【００３９】
ストレージ制御装置２は、ストレージ３を分担して管理し、１つ以上のプール３ａを担当する。ストレージ制御装置２は、上位接続部２１と、Ｉ／Ｏ制御部２２と、重複管理部２３と、メタ管理部２４と、データ処理管理部２５と、デバイス管理部２６とを有する。
【００４０】
上位接続部２１は、ＦＣドライバ及びｉＳＣＳＩドライバとＩ／Ｏ制御部２２との間の情報の受け渡しを行う。Ｉ／Ｏ制御部２２は、キャッシュメモリ上のデータを管理する。重複管理部２３は、データ重複排除／復元の制御を行うことで、ストレージ装置１ａ内に格納されているユニークなデータを管理する。
【００４１】
メタ管理部２４は、メタアドレス及び論物メタを管理する。また、メタ管理部２４は、メタアドレス及び論物メタを用いて、仮想ボリュームにおけるデータの識別に用いる論理

アドレスとＳＳＤ３ｄにおけるデータが記憶された位置を示す物理アドレスの変換処理を行う。

【００４２】

メタ管理部２４は、論物メタ管理部２４ａとメタアドレス管理部２４ｂとを有する。論物メタ管理部２４ａは、論理アドレスと物理アドレスとを対応付けるアドレス変換情報に関連する論物メタを管理する。論物メタ管理部２４ａは、論物メタのＳＳＤ３ｄへの書き込み、及び、論物メタのＳＳＤ３ｄからの読み出しをデータ処理管理部２５に依頼する。論物メタ管理部２４ａは、メタアドレスを用いて論物メタの記憶場所を特定する。

【００４３】

メタアドレス管理部２４ｂは、メタアドレスを管理する。メタアドレス管理部２４ｂは、メタアドレスの外部キャッシュ(２次キャッシュ)への書き込み、及び、外部キャッシュからのメタアドレスの読み出しをデバイス管理部２６に依頼する。

【００４４】

データ処理管理部２５は、ユーザデータを連続的なユーザデータユニットで管理し、ＲＡＩＤユニットの単位でＳＳＤ３ｄに追記及びまとめ書きを行う。また、データ処理管理部２５は、データの圧縮解凍、参照メタの生成を行う。ただし、データ処理管理部２５は、データが更新された場合に、古いデータに対応するユーザデータユニットに含まれる参照メタの更新は行わない。

【００４５】

また、データ処理管理部２５は、論物メタをＲＡＩＤユニットの単位でＳＳＤ３ｄに追記及びまとめ書きを行う。論物メタの書き込みでは、１小ブロック（５１２Ｂ）に論物メタの１６エントリが追記書きされるため、データ処理管理部２５は、同一小ブロック内にＬＵＮとＬＢＡが同じものが存在しないように管理する。

【００４６】

データ処理管理部２５は、同一小ブロックにＬＵＮとＬＢＡが同じものが存在しないように管理することで、ＲＡＩＤユニット番号とＲＡＩＤユニット内ＬＢＡにより、ＬＵＮとＬＢＡを検索することができる。なお、データの消去単位である１ＭＢのブロックと区別するため、ここでは５１２Ｂのブロックを小ブロックと呼ぶ。

【００４７】

また、メタ管理部２４から論物メタの読み出しを要求されると、データ処理管理部２５は、メタ管理部２４に指定された小ブロックから対象のＬＵＮとＬＢＡを検索して応答する。

【００４８】

データ処理管理部２５は、メインメモリ上のバッファであるライトバッファにライトデータを溜め、一定の閾値を超えるとＳＳＤ３ｄに書き出す。データ処理管理部２５は、プール３ａの物理スペースを管理し、ＲＡＩＤユニットの配置を行う。デバイス管理部２６は、ＲＡＩＤユニットのストレージ３への書き込みを行う。

【００４９】

図９は、機能部間の関係を示す図である。図９に示すように、重複管理部２３とメタ管理部２４との間では、論物メタの取得と更新が行われる。重複管理部２３とデータ処理管理部２５との間では、ユーザデータユニットのライトバック（writeback）とステージング（staging）が行われる。ここで、ライトバックとは、ストレージ３へのデータの書き込みであり、ステージングとは、ストレージ３からのデータの読み出しである。

【００５０】

メタ管理部２４とデータ処理管理部２５との間では、論物メタのライト（Write）とリード（Read）が行われる。データ処理管理部２５とデバイス管理部２６との間では、追記データのストレージリードとストレージライトが行われる。メタ管理部２４とデバイス管理部２６との間では、外部キャッシュのストレージリードとストレージライトが行われる。デバイス管理部２６とストレージ３との間では、ストレージリードとストレージライトが行われる。

【００５１】
次に、書き込み処理のシーケンスについて説明する。図１０Ａは、重複のないデータの書き込み処理のシーケンスを示す図であり、図１０Ｂは、重複のあるデータの書き込み処理のシーケンスを示す図である。
【００５２】
重複のないデータの書き込み処理では、図１０Ａに示すように、Ｉ／Ｏ制御部２２は、重複管理部２３にデータのライトバックを要求する（ステップＳ１）。すると、重複管理部２３は、データの重複判定を行って重複はないので（ステップＳ２）、新規のユーザデータユニットのライトをデータ処理管理部２５に要求する（ステップＳ３）。
【００５３】
すると、データ処理管理部２５は、ライトバッファを獲得し（ステップＳ４）、デバイス管理部２６にＲＵ（ＲＡＩＤユニット）の取得を要求する（ステップＳ５）。なお、ライトバッファを既に獲得している場合には、ライトバッファの獲得は不要である。そして、データ処理管理部２５は、デバイス管理部２６からＤＰ＃（Ｓｔｏｒａｇｅ　Ｐｏｏｌ　Ｎｏ）とＲＵ＃（ＲＡＩＤ　Ｕｎｉｔ　Ｎｏ）を獲得する（ステップＳ６）。
【００５４】
そして、データ処理管理部２５は、データを圧縮し（ステップＳ７）、参照メタを生成する（ステップＳ８）。そして、データ処理管理部２５は、ライトバッファ内でユーザデータユニットの追記書きを行い（ステップＳ９）、まとめ書き判定を行う（ステップＳ１０）。そして、データ処理管理部２５は、まとめ書きが必要と判定した場合には、デバイス管理部２６へライトバッファのまとめ書きを依頼する。そして、データ処理管理部２５は、重複管理部２３にＤＰ＃とＲＵ＃を応答する（ステップＳ１１）。
【００５５】
すると、重複管理部２３は、メタ管理部２４に論物メタの更新を要求し（ステップＳ１２）、メタ管理部２４は、更新した論物メタのライトをデータ処理管理部２５に要求する（ステップＳ１３）。
【００５６】
すると、データ処理管理部２５は、ライトバッファを獲得し（ステップＳ１４）、デバイス管理部２６にＲＵの取得を要求する（ステップＳ１５）。なお、獲得するライトバッファは、ユーザデータユニット用のライトバッファとは異なるバッファである。また、ライトバッファを既に獲得している場合には、ライトバッファの獲得は不要である。そして、データ処理管理部２５は、デバイス管理部２６からＤＰ＃とＲＵ＃を獲得する（ステップＳ１６）。
【００５７】
そして、データ処理管理部２５は、ライトバッファ内で論物メタの追記書きを行い（ステップＳ１７）、まとめ書き判定を行う（ステップＳ１８）。そして、データ処理管理部２５は、まとめ書きが必要と判定した場合には、デバイス管理部２６へライトバッファのまとめ書きを依頼する。そして、データ処理管理部２５は、メタ管理部２４にＤＰ＃とＲＵ＃を応答する（ステップＳ１９）。
【００５８】
すると、メタ管理部２４は、メタアドレス更新のためにメタアドレスの追い出しが必要か否かを判定し（ステップＳ２０）、追い出しが必要と判定した場合には、デバイス管理部２６に追い出しを依頼する。そして、ＤＰ＃とＲＵ＃に基づいてメタアドレスを更新する（ステップＳ２１）。
【００５９】
そして、メタ管理部２４は、重複管理部２３に完了を通知し（ステップＳ２２）、重複管理部２３は、メタ管理部２４から完了を通知されると、Ｉ／Ｏ制御部２２に完了を通知する（ステップＳ２３）。
【００６０】
このように、データ処理管理部２５が、ユーザデータユニットに加えて論物メタも追記

書き及びまとめ書きを行うことで、ＳＳＤ３ｄへの書き込み回数を減らすことができる。
【００６１】
また、重複のあるデータの書き込みの場合には、図１０Ｂに示すように、Ｉ／Ｏ制御部２２は、重複管理部２３にデータのライトバックを要求する（ステップＳ３１）。すると、重複管理部２３は、データの重複判定を行って重複があるので（ステップＳ３２）、重複するユーザデータユニットのライトをデータ処理管理部２５に要求する（ステップＳ３３）。
【００６２】
すると、データ処理管理部２５は、重複するユーザデータユニットを含むＲＡＩＤユニットについて、ストレージ３のリードをデバイス管理部２６に要求する（ステップＳ３４）。そして、デバイス管理部２６が、重複するユーザデータユニットを含むＲＡＩＤユニットを読み出してデータ処理管理部２５に応答する（ステップＳ３５）。そして、データ処理管理部２５は、ハッシュ値を比較し（ステップＳ３６）、データの重複を確認する。
【００６３】
そして、データ処理管理部２５は、重複が確認されれば、重複するユーザデータユニットにおける参照メタに参照先を追加して参照メタを更新する（ステップＳ３７）。データ処理管理部２５は、参照メタが更新されたユーザデータユニットを含むＲＡＩＤユニットのストレージ３へのライトをデバイス管理部２６に要求し（ステップＳ３８）、デバイス管理部２６から応答を受け取る（ステップＳ３９）。そして、データ処理管理部２５は、重複管理部２３にＤＰ＃とＲＵ＃を応答する（ステップＳ４０）。
【００６４】
すると、重複管理部２３は、メタ管理部２４に論物メタの更新を要求し（ステップＳ４１）、メタ管理部２４は、更新した論物メタのライトをデータ処理管理部２５に要求する（ステップＳ４２）。
【００６５】
すると、データ処理管理部２５は、ライトバッファを獲得し（ステップＳ４３）、デバイス管理部２６にＲＵの取得を要求する（ステップＳ４４）。そして、データ処理管理部２５は、デバイス管理部２６からＤＰ＃とＲＵ＃を獲得する（ステップＳ４５）。
【００６６】
そして、データ処理管理部２５は、ライトバッファ内で論物メタの追記書きを行い（ステップＳ４６）、まとめ書き判定を行う（ステップＳ４７）。そして、データ処理管理部２５は、まとめ書きが必要と判定した場合には、デバイス管理部２６へライトバッファのまとめ書きを依頼する。そして、データ処理管理部２５は、メタ管理部２４にＤＰ＃とＲＵ＃を応答する（ステップＳ４８）。
【００６７】
すると、メタ管理部２４は、メタアドレス更新のためにメタアドレスの追い出しが必要か否かを判定し（ステップＳ４９）、追い出しが必要と判定した場合には、デバイス管理部２６に追い出しを依頼する。そして、ＤＰ＃とＲＵ＃に基づいてメタアドレスを更新する（ステップＳ５０）。
【００６８】
そして、メタ管理部２４は、重複管理部２３に完了を通知し（ステップＳ５１）、重複管理部２３は、メタ管理部２４から完了を通知されると、Ｉ／Ｏ制御部２２に完了を通知する（ステップＳ５２）。
【００６９】
このように、データ処理管理部２５が、重複データについて、論物メタを追記書き及びまとめ書きを行うことで、ＳＳＤ３ｄへの書き込み回数を減らすことができる。
【００７０】
次に、読み出し処理のシーケンスについて説明する。図１１は、読み出し処理のシーケンスを示す図である。図１１に示すように、Ｉ／Ｏ制御部２２は、データのステージングを重複管理部２３に要求する（ステップＳ６１）。すると、重複管理部２３は、データの

論物メタの獲得をメタ管理部２４に要求する（ステップＳ６２）。
【００７１】
　すると、メタ管理部２４は、データのメタアドレスがメインメモリ上にあることを確認し（ステップＳ６３）、メタアドレスを指定して、論物メタのリードをデータ処理管理部２５に要求する（ステップＳ６４）。なお、データのメタアドレスがメインメモリ上にない場合には、メタ管理部２４は、デバイス管理部２６にストレージ３からのリードを要求する。
【００７２】
　そして、データ処理管理部２５は、論物メタを含むＲＡＩＤユニットについて、ストレージ３からのリードをデバイス管理部２６に要求し（ステップＳ６５）、デバイス管理部２６からＲＡＩＤユニットを受け取る（ステップＳ６６）。そして、データ処理管理部２５は、ＲＡＩＤユニットから論物メタを検索し（ステップＳ６７）、検索した論物メタをメタ管理部２４に渡す（ステップＳ６８）。
【００７３】
　すると、メタ管理部２４は、論物メタを解析し（ステップＳ６９）、ユーザデータユニットを含むＲＡＩＤユニットのＤＰ＃とＲＵ＃とＯｆｆｓｅｔを重複管理部２３に渡す（ステップＳ７０）。ここで、Ｏｆｆｓｅｔは、ＲＡＩＤユニット内のユーザデータユニットのアドレスである。すると、重複管理部２３は、ＤＰ＃とＲＵ＃とＯｆｆｓｅｔを指定してユーザデータユニットのリードをデータ処理管理部２５に要求する（ステップＳ７１）。
【００７４】
　すると、データ処理管理部２５は、ユーザデータユニットを含むＲＡＩＤユニットのストレージ３からのリードをデバイス管理部２６に要求し（ステップＳ７２）、デバイス管理部２６からＲＡＩＤユニットを受け取る（ステップＳ７３）。そして、データ処理管理部２５は、Ｏｆｆｓｅｔを用いてＲＡＩＤユニットから取り出したユーザデータユニットに含まれる圧縮データを伸張し（ステップＳ７４）、ユーザデータユニットから参照メタを削除する（ステップＳ７５）。
【００７５】
　そして、データ処理管理部２５は、データを重複管理部２３に渡し（ステップＳ７６）、重複管理部２３は、Ｉ／Ｏ制御部２２にデータを渡す（ステップＳ７７）。
【００７６】
　このように、ストレージ制御装置２は、メタアドレスを用いて論物メタを取得し、論物メタを用いてユーザデータユニットを取得することで、ストレージ３からデータを読み出すことができる。
【００７７】
　次に、ストレージ制御装置２による書き込み処理の効果について図１２Ａ～図１２Ｃを用いて説明する。図１２Ａは、メタメタ方式導入前のスモールライトの数を示す図であり、図１２Ｂは、メタメタ方式でのスモールライトの数を示す図であり、図１２Ｃは、メタメタ方式で旧データ無効化を行わない場合のスモールライトの数を示す図である。ここで、スモールライトとは、ブロック（１ＭＢ）に較べて小さな単位（４ＫＢ）のライトである。
【００７８】
　図１２Ａに示すように、メタメタ方式導入前は、サーバ１ｂからの８ＫＢのデータの書き込みに対して、データに関してはまとめ書きが行われるが、論物メタの更新、参照メタの更新に関してスモールライトが行われる。ここで、参照メタの更新としては、旧データ（参照ＬＵＮ／ＬＢＡ情報）の無効化がある。また、ＲＡＩＤ６の場合、データの書き込みに対応して２つのパリティＰとＱの書き込みが行われる。したがって、合計６回のスモールライトが行われる。
【００７９】
　これに対して、メタメタ方式では、図１２Ｂに示すように、論物メタの更新は追記書き

のためスモールライトではなくなり、３回のスモールライトで済む。さらに、旧データ無効化を行わない場合、図１２Ｃに示すように、参照メタの更新もなくなり、スモールライトがなくなる。

【００８０】

このように、メタメタ方式を用いることにより、ストレージ制御装置２は、スモールライトの数を減らし、書き込み処理を高速化することができる。また、ストレージ制御装置２は、旧データ無効化を行わないことで、さらにスモールライトの数を減らすことができる。

【００８１】

上述してきたように、実施例では、論物メタ管理部２４ａが、データの論理アドレスと物理アドレスを対応付ける論物メタの情報を管理し、データ処理管理部２５が、論物メタの情報をＲＡＩＤユニットの単位でＳＳＤ３ｄへ追記及びまとめ書きする。したがって、ストレージ制御装置２は、スモールライトの数を減らし、書き込み処理を高速化することができる。

【００８２】

また、実施例では、メタアドレス管理部２４ｂが、論理アドレスと論物メタのアドレスを対応付けるメタアドレスの情報を管理するので、論物メタ管理部２４ａはメタアドレスを用いて論物メタの位置を特定することができる。

【００８３】

また、実施例では、データが更新された場合に、古いデータに対応するユーザデータユニットの参照メタを更新しない。したがって、ストレージ制御装置２は、スモールライトの数をさらに減らすことができる。

【００８４】

また、実施例では、メタアドレスはメインメモリ上で管理され、溢れたメタアドレスの情報はＳＳＤ３ｄの所定の位置に記憶される。したがって、ストレージ制御装置２は、ＳＳＤ３ｄの所定の位置から読み出すことでメタアドレスの情報を取得することができる。

【００８５】

なお、実施例では、ストレージ制御装置２について説明したが、ストレージ制御装置２が有する構成をソフトウェアによって実現することで、同様の機能を有するストレージ制御プログラムを得ることができる。そこで、ストレージ制御プログラムを実行するストレージ制御装置２のハードウェア構成について説明する。

【００８６】

図１３は、実施例に係るストレージ制御プログラムを実行するストレージ制御装置２のハードウェア構成を示す図である。図１３に示すように、ストレージ制御装置２は、メモリ４１と、プロセッサ４２と、ホストＩ／Ｆ４３と、通信Ｉ／Ｆ４４と、接続Ｉ／Ｆ４５とを有する。

【００８７】

メモリ４１は、プログラムやプログラムの実行途中結果などを記憶するＲＡＭ（Random Access Memory）である。プロセッサ４２は、メモリ４１からプログラムを読み出して実行する処理装置である。

【００８８】

ホストＩ／Ｆ４３は、サーバ１ｂとのインタフェースである。通信Ｉ／Ｆ４４は、他のストレージ制御装置２と通信するためのインタフェースである。接続Ｉ／Ｆ４５は、ディスク３とのインタフェースである。

【００８９】

そして、プロセッサ４２において実行されるストレージ制御プログラムは、可搬記録媒体５１に記憶され、メモリ４１に読み込まれる。あるいは、ストレージ制御プログラムは、通信インタフェース４４を介して接続されたコンピュータシステムのデータベースなどに記憶され、これらのデータベースから読み出されてメモリ４１に読み込まれる。

【００９０】

また、実施例では、ＳＳＤ３ｄを不揮発性記憶媒体として用いる場合について説明したが、本発明はこれに限定されるものではなく、ＳＳＤ３ｄと同様なデバイス特性を有する他の不揮発性記憶媒体を用いる場合にも同様に適用することができる。

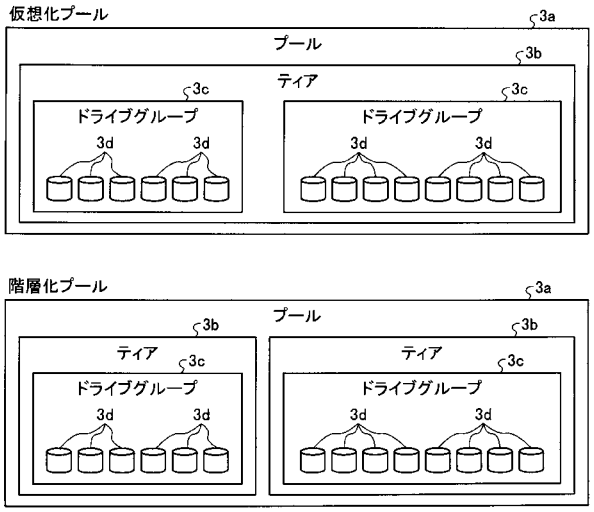
【符号の説明】

【００９１】

１　情報処理システム
１ａ　ストレージ装置
１ｂ　サーバ
２　ストレージ制御装置
３　ストレージ
３ａ　プール
３ｂ　ティア
３ｃ　ドライブグループ
３ｄ　ＳＳＤ
２１　上位接続部
２２　Ｉ／Ｏ制御部
２３　重複管理部
２４　メタ管理部
２４ａ　論物メタ管理部
２４ｂ　メタアドレス管理部
２５　データ処理管理部
２６　デバイス管理部
４１　メモリ
４２　プロセッサ
４３　ホストＩ／Ｆ
４４　通信Ｉ／Ｆ
４５　接続Ｉ／Ｆ
５１　可搬記録媒体

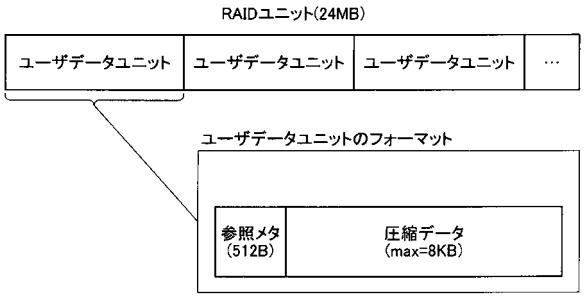
【図１】

実施例に係るストレージ装置の記憶構成を示す図



【図２】

RAIDユニットのフォーマットを示す図



【図３】

参照メタのフォーマットを示す図

(a)参照メタ

(b)SB(Super Block)

オフセット	バイト#0	バイト#1	バイト#2	バイト#3
0x00	Header Length			
0x04	Hash Value(20B)			
:				
0x14				
0x18	Next Offset Block Count		Reserved	
0x1c	Reserved			

(c)参照LUN/LBA情報

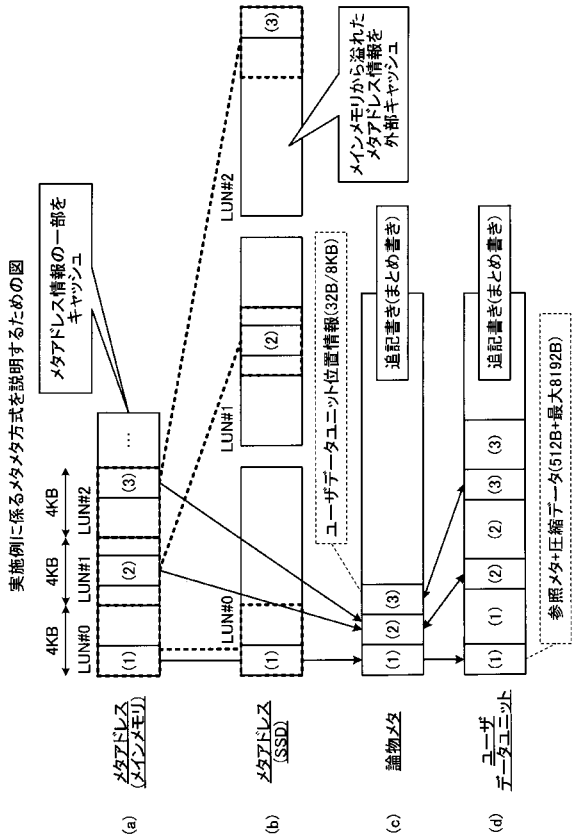
オフセット	バイト#0	バイト#1	バイト#2	バイト#3
0x00	LUN		LBA(上位バイト)	
0x04	LBA(下位バイト)			

【図４】

論物メタのフォーマットを示す図

オフセット	バイト#0	バイト#1	バイト#2	バイト#3
0x00	…	…	Reserved	
0x04	LUN		LBA(上位バイト)	
0x08	LBA(下位バイト)			
0x0c	Compression Byte Count		Reserved	
0x10	Node No		Storage Pool No	Reserved
0x14	RAID Unit No			
0x18	RAID Unit Offset LBA		…	
0x1c	…			

【図５】



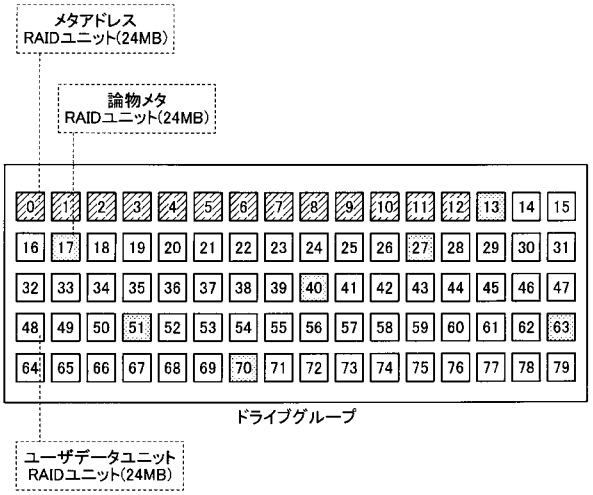
【図６】

メタアドレスのフォーマットを示す図

オフセット	バイト#0	バイト#1	バイト#2	バイト#3
0x00	Storage Pool No	Reserved	RAID Unit Offset LBA	
0x04	RAID Unit No			

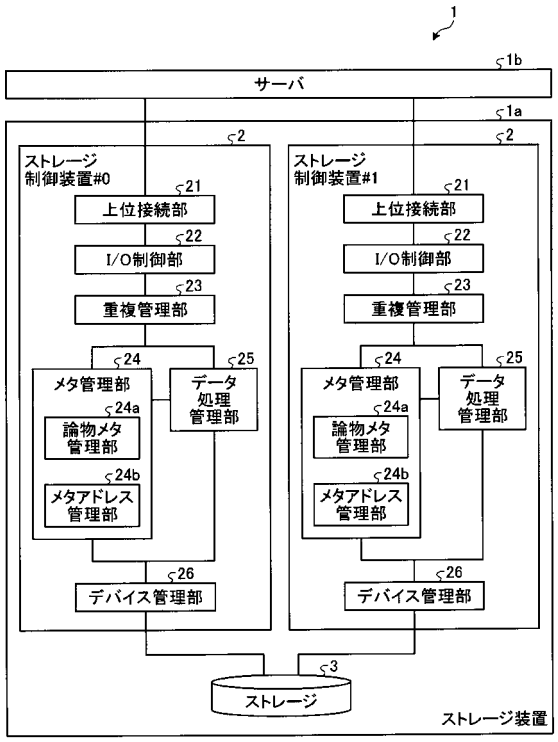
【図７】

ドライブグループにおけるRAIDユニットの配置例を示す図

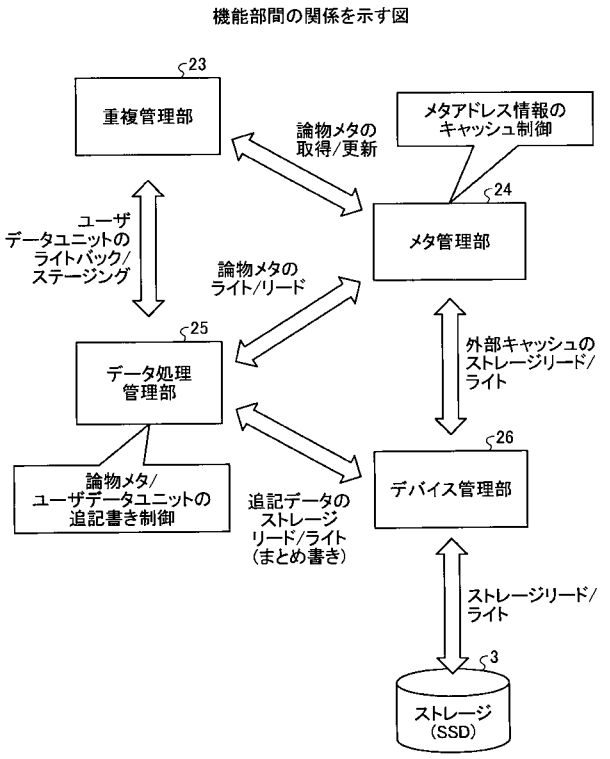


【図８】

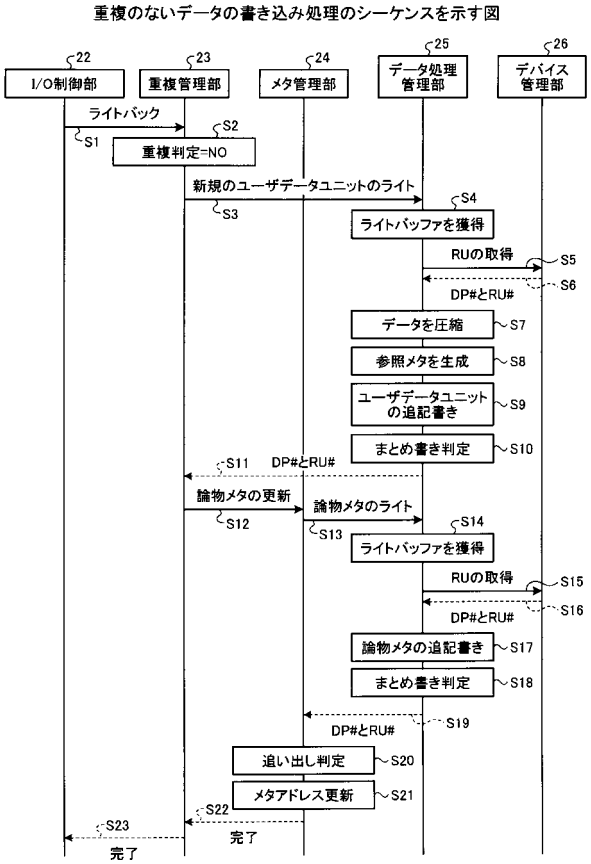
実施例に係る情報処理システムの構成を示す図



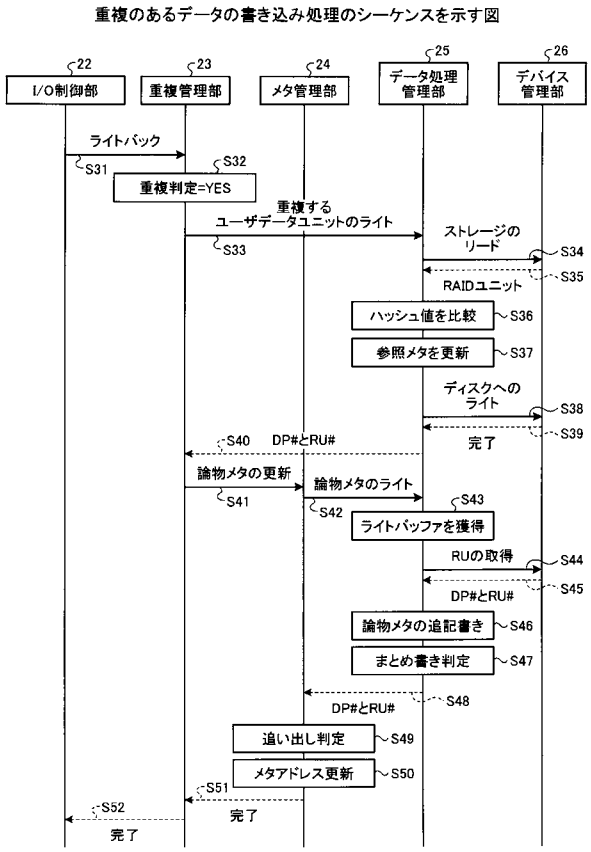
【図９】



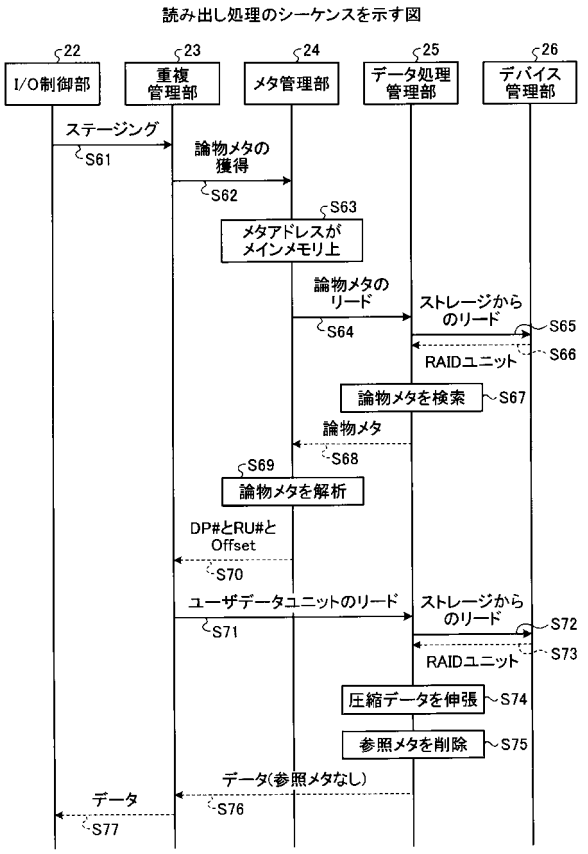
【図１０Ａ】



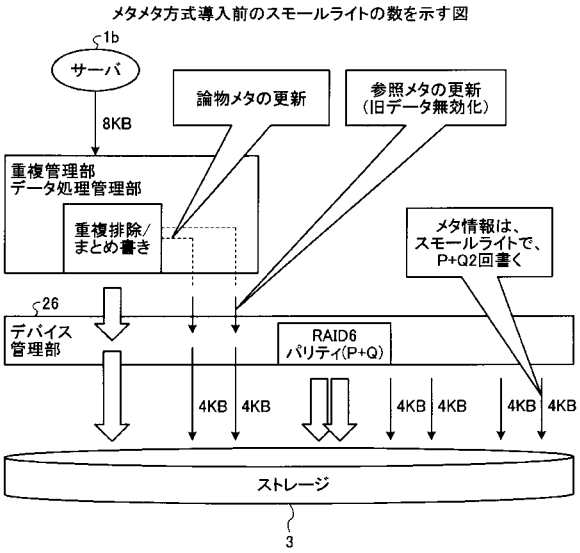
【図１０Ｂ】



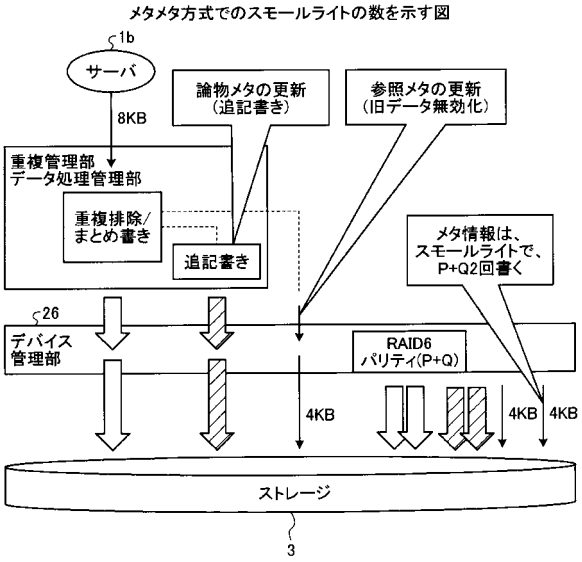
【図１１】



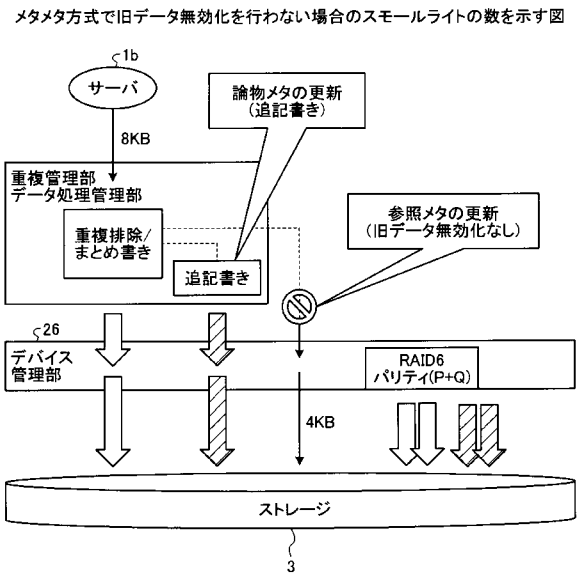
【図１２Ａ】



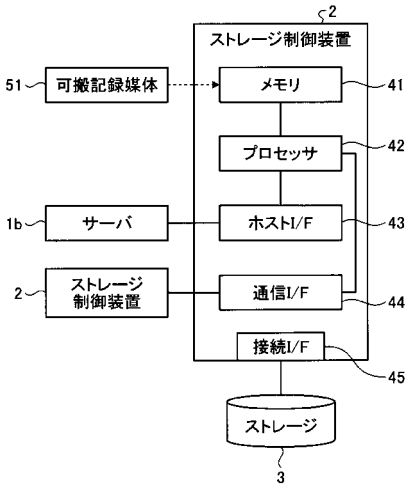
【図１２Ｂ】



【図１２Ｃ】



【図１３】



フロントページの続き

(51)Int.Cl.	FI	テーマコード（参考）
	Ｇ０６Ｆ　12/02　　５１０Ａ	

(72)発明者　紺田　與志仁
神奈川県川崎市中原区上小田中４丁目１番１号　株式会社富士通コンピュータテクノロジーズ内
(72)発明者　倉澤　祐輔
神奈川県川崎市中原区上小田中４丁目１番１号　株式会社富士通コンピュータテクノロジーズ内
(72)発明者　菊池　利夫
神奈川県川崎市中原区上小田中４丁目１番１号　株式会社富士通コンピュータテクノロジーズ内
(72)発明者　田中　勇至
神奈川県川崎市中原区上小田中４丁目１番１号　株式会社富士通コンピュータテクノロジーズ内
(72)発明者　梶山　真理乃
神奈川県川崎市中原区上小田中４丁目１番１号　株式会社富士通コンピュータテクノロジーズ内
(72)発明者　鈴木　悠介
神奈川県川崎市中原区上小田中４丁目１番１号　株式会社富士通コンピュータテクノロジーズ内
(72)発明者　篠▲崎▼　祥成
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　渡辺　岳志
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

Ｆターム(参考)　5B060 AB26
5B160 AB26