

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号
特許第6155769号
(P6155769)

(45) 発行日 平成29年7月5日 (2017.7.5)

(24) 登録日 平成29年6月16日 (2017.6.16)

(51) Int.Cl.

G 0 6 F 3 / 0 6 (2 0 0 6 . 0 1)

F 1

G 0 6 F 3 / 0 6 3 0 4 F

G 0 6 F 3 / 0 6 3 0 1 J

請求項の数 7 (全 23 頁)

(21) 出願番号	特願2013-74825 (P2013-74825)	(73) 特許権者	000005223
(22) 出願日	平成25年3月29日 (2013.3.29)		富士通株式会社
(65) 公開番号	特開2014-199583 (P2014-199583A)		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10月23日 (2014.10.23)	(74) 代理人	100074099
審査請求日	平成27年11月13日 (2015.11.13)		弁理士 大宮 義之
		(74) 代理人	100133570
			弁理士 ▲徳▼永 民雄
		(72) 発明者	山崎 暁人
			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	御宿 義勝
			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】 ストレージ制御装置、制御プログラム及び制御方法

(57) 【特許請求の範囲】

【請求項 1】

ホストにより更新がなされる更新対象の元データを格納する第1の格納領域と、
前記更新対象の元データの更新後のデータを、前記第1の格納領域のアドレスを含む複数のデータに分割し、分割後の複数のデータを、前記第1の格納領域内の連続したアドレスを含む複数の更新データにまとめる制御部と、
前記第1の格納領域内の連続したアドレスにまとめられた前記複数の更新データを格納する第2の格納領域と、
前記更新対象の元データのうち更新対象の前記第1の格納領域のアドレスである更新領域アドレスと、前記複数の更新データを格納する前記第2の格納領域のアドレスである格納先アドレスとを対応付けたマッピング情報が、前記制御部から格納される第3の格納領域を備える、
ことを特徴とするストレージ制御装置。

【請求項 2】

前記マッピング情報は、前記更新領域アドレスが示す数の順に前記格納先アドレスを格納することにより示された情報であり、前記第3の格納領域及びキャッシュに保存されることを特徴とする請求項1記載のストレージ制御装置。

【請求項 3】

前記マッピング情報を前記第3の格納領域に格納する際に、
前記制御部は、前記複数の更新データを、キャッシュページごとに纏める

ことを特徴とする請求項 2 記載のストレージ制御装置。

【請求項 4】

前記第 3 の格納領域は、前記第 2 の格納領域への前記複数の更新データの書き込み処理の進行具合を示す情報と、前記更新領域アドレスと前記格納先アドレスを対応付けた情報である対応情報の書き込み処理の進行具合を示す情報とを、履歴情報として保存しており、

前記制御部は、起動する際に前記履歴情報を参照し、前記履歴情報に、前記制御部が停止する前に前記対応情報の書き込み処理が終了していない更新データに関する情報が含まれていた場合、

前記対応情報の書き込み処理が終了していない更新データよりも先に処理されたデータを正常に実行された更新データであると判定し、前記正常に実行された更新データに関する情報を、前記履歴情報に保存し、

前記対応情報の書き込み処理が終了していない更新データからバックアップを再開することを特徴とする請求項 1 ~ 3 のいずれかに記載のストレージ制御装置。

【請求項 5】

前記制御部は、前記複数の更新データを、シーケンシャルアクセスを行うデータか、ランダムアクセスを行うデータか判定し、

前記複数の更新データがシーケンシャルアクセスを行うデータであった場合、

前記複数の更新データが上書きする領域を指定するアドレスを含むキャッシュページの次のキャッシュページをあらかじめ前記キャッシュに保存する

ことを特徴とする請求項 2 に記載のストレージ制御装置。

【請求項 6】

ホストにより更新がなされる更新対象の元データを第 1 の格納領域に格納し、

前記更新対象の元データの更新後のデータを、前記第 1 の格納領域のアドレスを含む複数のデータに分割し、分割後の複数のデータを、前記第 1 の格納領域内の連続したアドレスを含む複数の更新データにまとめ、

前記第 1 の格納領域内の連続したアドレスにまとめられた前記複数の更新データを第 2 の格納領域に格納し、

前記更新対象の元データのうち更新対象の前記第 1 の格納領域のアドレスである更新領域アドレスと、前記複数の更新データを格納する前記第 2 の格納領域のアドレスである格納先アドレスとを対応付けたマッピング情報を、第 3 の格納領域に格納する処理をストレージ装置に実行させる

ことを特徴とする制御プログラム。

【請求項 7】

ホストにより更新がなされる更新対象の元データを第 1 の格納領域に格納し、

前記更新対象の元データの更新後のデータを、前記第 1 の格納領域のアドレスを含む複数のデータに分割し、分割後の複数のデータを、前記第 1 の格納領域内の連続したアドレスを含む複数の更新データにまとめ、

前記第 1 の格納領域内の連続したアドレスにまとめられた前記複数の更新データを第 2 の格納領域に格納し、

前記更新対象の元データのうち更新対象の前記第 1 の格納領域のアドレスである更新領域アドレスと、前記複数の更新データを格納する前記第 2 の格納領域のアドレスである格納先アドレスとを対応付けたマッピング情報を、第 3 の格納領域に格納する処理をストレージ装置に実行させる

ことを特徴とする制御方法。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は、ストレージ制御装置に関する。

【背景技術】

【0002】

ストレージ装置は、多数のハードディスクドライブと大容量キャッシュメモリを搭載し、例えばサーバで用いられるデータを保存する装置である。ストレージ装置は、ある瞬間のデータのイメージを保持するスナップショットの機能を提供する。スナップショットは、例えばサーバからの要求を契機に実行される。スナップショット機能には、例えばCopy-On-Write（以下、COWと記す）方式がある。スナップショットは、スナップショットのコピー元となるデータをそのままコピーするという処理は行わない。スナップショットは、コピー元のデータに関するメタ情報を取得する処理を行うので、バックアップに費やす時間は短いという利点を持つ。しかし、メタ情報がバックアップされた状態では、コピー元のデータが更新されると、スナップショットを取った時点でのデータが上書きされてしまい、あとから参照することができなくなる。そのため、COW方式のスナップショットを導入しているストレージ装置は、スナップショット取得後にあるデータを更新する際、更新前のデータを含むデータブロックを別のアドレスに退避する。ストレージ装置は、データブロックを退避させた場所に対応するメタ情報を取得し、退避したデータに更新するデータを上書きする。

10

【0003】

COW方式のスナップショットを導入しているストレージ装置は、コピー元となるデータをマスタースライスと称し、チャンクと呼ばれる区切りで管理している。また、ストレージ装置は、スナップショットで使用されるデータを保持するCOWスライスを有する。スナップショットにおいて、あるデータが更新された場合、ストレージ装置は、マスタースライスのデータを更新したデータをCOWスライスのデータ領域に保存する。その際に、ストレージ装置は、マスタースライスのあるデータが格納されていたアドレス（物理）と、退避先であるCOWスライスのアドレス（論理）とを関連づけたマッピング情報をメタ情報として、COWスライスに保存する。

20

【0004】

ストレージ装置の大規模化及びシステムの大規模化に伴い、COWスライスのチャンクを管理するマッピング情報の量は増大する傾向にある。マッピング情報の量が増えると、COWスライスへのアクセス回数が増え、ストレージ装置のメモリ使用量が増えてしまう。それに伴い、スナップショットの性能も低下してしまう。マッピング情報の量が増えても、スナップショットの性能を維持するのが好ましい。

30

【0005】

スナップショットの性能を維持する技術として、ホストと記憶装置とでスナップショット作成の負荷を分散する技術が知られている。（例えば特許文献1参照）。

【先行技術文献】

【特許文献】

【0006】

【特許文献1】特開2004-118413号公報

【発明の概要】

【発明が解決しようとする課題】

【0007】

上述した背景技術では以下のような問題がある。

マッピング情報の量が増えると、スナップショットの性能が低下してしまう。

【課題を解決するための手段】

【0008】

ストレージ制御装置は、第1～第3の格納領域、制御部を備える。第1の格納領域は、ホストにより更新がなされる更新対象の元データを格納する。制御部は、第1の格納領域に格納された更新対象の元データに対して書き込むべきデータを複数の更新データに分割することにより、連続したアドレスごとに複数の更新データを振り分ける。第2の格納領域は、制御部により振り分けられた複数の更新データを格納する。第3の格納領域は、更新対象の元データの複数の更新データによって上書きされるべき第1の格納領域のアドレ

40

50

スである更新領域アドレスを、複数の更新データを格納している第２の格納領域のアドレスである格納先アドレスに対応付けた情報を、格納する。

【発明の効果】

【０００９】

大規模なストレージ装置であっても、スナップショットの性能を維持できる。

【図面の簡単な説明】

【００１０】

【図１】ストレージ装置を説明する例を示す図である。

【図２】マッピング情報の例を示す図である。

【図３】ストレージ装置の構成の例を示す図である。

【図４】メタスライスの例を詳細に説明した図である。

【図５Ａ】同一のキャッシュページを保持する書き込み要求を一括化する例を説明する図である。

【図５Ｂ】同一のキャッシュページを保持する書き込み要求を一括化する例を説明する図である。

【図６Ａ】連続した物理アドレスを保持する書き込み要求を一括化する例を説明する図である。

【図６Ｂ】連続した物理アドレスを保持する書き込み要求を一括化する例を説明する図である。

【図６Ｃ】連続した物理アドレスを保持する書き込み要求を一括化する例を説明する図である。

【図７Ａ】スナップショットの処理の例を説明する図である。

【図７Ｂ】スナップショットの処理の例を説明する図である。

【図８】複数のチャンクに跨るデータの書き込み要求に関する処理の例を示す図である。

【図９Ａ】データスライスへの書き込み要求を一括処理する例を説明する図である。

【図９Ｂ】データスライスへの書き込み要求を一括処理する例を説明する図である。

【図１０】履歴情報のダウンリカバリーの例を示す図である。

【図１１】メタデータのプリフェッチの例を示す図である。

【００１１】

【図１２】ＩＯ要求を受領した制御部の処理の例を説明するフローチャートである。

【図１３】制御部がマッピング情報を参照する処理の例を説明するフローチャートである。

【図１４】データのコピーに関する処理の例を説明するフローチャートである。

【図１５】制御部がメタスライスにメタ情報を書き込む処理を説明するフローチャートである。

【図１６】同一のキャッシュページを保持する書き込み要求を一括化する例を説明するフローチャートである。

【図１７】連続した物理アドレスへの書き込み要求を一括化する例を説明するフローチャートである。

【図１８】データスライスへの書き込み要求を一括処理する例を説明するフローチャートである。

【図１９】履歴情報のダウンリカバリーの例を説明するフローチャートである。

【発明を実施するための形態】

【００１２】

以下、本実施形態について、図面を参照しながら詳細に説明する。

図１は、ストレージ装置を説明する例を示す図である。図１のストレージ装置２００は、制御部２０１、メタスライス２０２、データスライス２０３、マスタースライス２０４を備える。制御部２０１は、ストレージ装置に接続されるホストとなるサーバの要求を契機に、スナップショットの処理を制御する。サーバの要求は、サーバのデータが更新され

10

20

30

40

50

た場合に通知される。マスタースライス203は、ホストにより更新がなされる更新対象の元データを格納する。また、マスタースライス203は、ある時点でのホストが保持する何らかのデータをバックアップしたものであり、。マスタースライス204は、スナップショットのコピー元となるデータ群である。実施形態に係るマスタースライス204の各データは、4KBのチャンクサイズで区切られて管理される。図1のストレージ装置200は、COW方式で用いられるCOWスライスを、メタスライス202とデータスライス203とに分割したものである。メタスライス202は、ヘッダチャンク及びメタチャンクなどスナップショットの処理に用いられるメタ情報を保存する領域である。データスライス203は、スナップショットの更新データを保存するデータ領域である。メタスライス202は、アクセスに使用されるマッピング情報を有する。マッピング情報は、適宜キャッシュに格納され使用される。制御部201は、ホストからの書き込み要求であるI/O要求(Input Output)に応じて、マッピング情報を参照する。キャッシュに格納されたマッピング情報を参照することで、メタスライスからマッピング情報を参照する場合よりも高速に処理を行うことができ、スナップショットの性能を向上できる。I/O要求は、更新するデータ、更新するデータのサイズ、更新するデータを書き込む先のマスタースライスのアドレス番号の情報が含まれる。図1のデータスライス203は、中間にメタスライスなどの割り込みとなるスライスがない。そのため、制御部201は、更新をするデータをデータスライス203のスライスを連続的に書き込むことができる。そのため、制御部201は、データスライス203の書き込み用いるシーク時間を軽減することができる、かつ一括でI/O要求を行うことができる。

10

20

【0013】

スケールアウト型のストレージ装置では、容量が大規模であるため、1MBのチャンクサイズが用いられている。しかし、チャンクサイズが大きいとスナップショットのコピー速度が低下する。そのため、チャンクサイズを小さくすることで、コピー速度を改善することができる。例えば、チャンクサイズを4KBとすることで、スナップショットのコピー速度は改善される。しかし、チャンクサイズを小さくした場合、チャンクを管理するマッピング情報の量が増え、マッピング情報をオンメモリにすることが困難になり、キャッシュなどの方法を取り工夫することになる。チャンクサイズの具体的な数値は、一例であり、何ら限定するものではない。

【0014】

図2は、マッピング情報の例を示す図である。マッピング情報は、マスタースライスのアドレスとCOWスライスのアドレスとを対応づけた情報である。図2(a)のマッピング情報は、COW方式を用いたストレージ装置で用いられるマッピング情報である。図2(a)のマッピング情報では、例えば、マスタースライスのアドレス8番のデータは、COWスライスのアドレス2番に保存されることを示す。なお、マスタースライスのアドレスは、物理アドレスであり、COWスライスのアドレスは、論理アドレスが用いられる。マッピング情報が少ないときはオンメモリとし、ハッシュ探索を利用して高速にアドレス変換できる。しかし、マッピング情報が増えるとオンメモリにすることは困難となるため、実施形態ではキャッシュを利用する方法を取る。キャッシュ上にのせられる情報を増やすため、実施形態ではCOWスライスのアドレスのみを持つ構造とする。この構造であれば、マスタースライスのアドレスから対応するCOWスライスのアドレスを検索することも容易である。

30

40

【0015】

図2(b)のマッピング情報は、実施形態に係るマッピング情報である。図2(b)の例は、図2(a)のマッピング情報を、実施形態に係るマッピング情報に置き換えた図である。図2(b)のマッピング情報は、マスタースライスのアドレスの順に連続して並べられ、COWスライスのアドレス情報を保持している。マスタースライスのアドレスは0から始まるため、例えば、マスタースライスのアドレス8番のデータであるCOWスライスのアドレス2番は、9番目に保持されている。実施形態に係るマッピング情報は、マスタースライスのアドレスとデータスライスのアドレスとを対応づけた情報であり、マスター

50

スライスのアドレスを連続的に羅列した順に、データスライスのアドレスを格納している。図2(b)のようにマッピング情報を作成することによって、マッピング情報に使用されるデータサイズを削減することができる。また、マスタースライスのアドレスの順のマッピング情報を用いることで、マスタースライスのアドレスからCOWスライスへのアドレス変換の時間を短縮することができる。制御部201は、実施形態に係るマッピング情報をキャッシュに読み込んで使用する。更に、キャッシュページを複数作ることによって、キャッシュヒット率を上げることができ、スナップショットの性能は向上する。なお、図2(b)のマッピング情報に格納されている数字0は、何も割り当てられていないことを示す。数字の0が格納されている領域は、実際のメモリは割り当てられない。そのため、マッピング情報のメモリ領域を小さくできる。ストレージ装置は、未書き込みのマッピング情報を実際にはメモリ領域を割り当てないように使用するため、メタスライスにシンプロビジョニング機構を利用するのが好ましい。

10

【0016】

図3は、ストレージ装置の構成の例を示す図である。ストレージ装置300は、サーバ340と接続されており、サーバ340で使用されるデータを保存する装置である。ストレージ装置300は、制御装置310、制御装置320、記憶装置330を有する。制御装置320は、制御部321、CPU(Central Processing Unit)322、メモリ323、VDISK(Virtual Disk)324、HDD(Hard Disk Drive)325を備える。制御部321は、ストレージシステム内に複数あってもよい。制御部321は、制御装置320に関するストレージ装置を制御し、図1の制御部201の一例である。CPU322は、制御部321の処理を実行する。CPU322で使用する各種データは、メモリ323及びHDD325に保存される。VDISK324は、記憶装置330の物理的な記憶領域を統合した仮想的な記憶領域であり、制御部321によって制御される。VDISK324は、メタスライス202、データスライス203、マスタースライス204を有する。制御装置310は、管理部311、CPU312、メモリ313、HDD314、制御部315、VDISK316、管理用DB(Data Base)317を備える。制御装置310は、制御装置320を管理するために、管理部311と管理用DB317を備えている。管理部311は、ストレージシステム内の1つの制御装置に存在し、全制御装置のハードウェアの割り当て情報や状態の情報に関して管理用DB317を用いて管理する。制御装置310は、制御装置320の親となる制御装置である。CPU312は、管理部311と制御部315の処理を実行する。CPU312で使用する各種データはメモリ313、HDD314に保存される。VDISK316は、記憶装置330を用いた仮想的な記憶領域であり、制御部315によって制御される。記憶装置330は、複数のハードディスクを備え、各ハードディスクはLUN(Logical Unit Number)によって識別される。LUNは、各ハードディスクを識別するための数字である。図3のストレージ装置は、LUN331a~LUN331nを備え、LUN331b~LUN331mがVDISK324に割り当てられており、LUN331a及びLUN331nがVDISK316に割り当てられている。VDISKは、シンプロビジョニング機構を用いる。図1の制御部201は、制御部315であってもよい。また、メタスライス202、データスライス203、マスタースライス204は、VDISK316が有してもよい。

20

30

40

【0017】

図4は、メタスライスの例を詳細に説明した図である。メタスライス202は、8KBのチャンクサイズを持つデータ群である。チャンクサイズは、変更することができ、データのサイズを限定するものではない。制御部201は、メタスライス202を参照することで、スナップショットの制御を行う。メタスライス202のチャンク0の領域は、メタスライスに関するヘッダ情報が格納している。メタスライス202のチャンク1の領域は、メタスライスに関するマッピング情報を格納している。メタスライス202のチャンク2の領域は、スナップショットに用いられる後述する(図10)履歴情報、後述する(図11)プリフェッチ情報が格納している。メタスライス202のチャンク3以降の領域は

50

、本実施形態に用いられるマッピング情報であるアドレスが格納される。

【 0 0 1 8 】

<メタスライスへの書き込み要求の一括化>

制御部 2 0 1 は、サーバ 3 4 0 からの書き込み要求を受け取る。書き込み要求は、データを更新するために用いられるデータやメタ情報が含まれる。制御部 2 0 1 は、受領した要求を所定のサイズに分割し、処理を行う。制御部 2 0 1 は、分割された各々の I O 要求に応じて、データスライス、メタスライスにアクセスを行う。制御部 2 0 1 は、例えば、連続した物理アドレスに書き込みを行う場合、I O 要求をまとめて1度で実行したほうが効率的であり、データスライス又はメタスライスへのアクセス回数も減らすことができる。そこで、実施形態に係るストレージ装置は、メタスライスへの I O 要求に対して、同一のキャッシュページを保持する I O 要求を一括化する。更に、実施形態に係るストレージ装置は、メタスライスへの I O 要求に対して、連続した物理アドレスへの I O 要求を一括化する。

10

【 0 0 1 9 】

同一のキャッシュページの情報を保持する I O 要求を一括化し、更に、連続した物理アドレスへの I O 要求を一括化することで、制御部 2 0 1 は、キャッシュへのアクセス回数を減らすことができる。スナップショットの処理に時間がかかる原因には、アクセスごとにアクセス先のアドレスを特定する処理が発生することが挙げられる。本実施形態にかかる方法では、同一のキャッシュページに関する複数の I O 要求が1つの I O 要求にまとめられる、更に、連続した物理アドレスへの I O 要求にまとめられるため、スナップショットの性能を向上させることができる。なお、I O 要求は、書き込みをするデータサイズ、書き込みをする対象となるマスタースライスのアドレス番号、キャッシュページに関する情報などの付加情報を持つので制御部はこれらの情報を用いて一括化を行う。

20

【 0 0 2 0 】

図 5 は、同一のキャッシュページを保持する書き込み要求を一括化する例を説明する図である。図 5 は、書き込み要求を一括化する例を、図 5 A (1) ~ 図 5 B (7) を用いて説明している。図 5 の例を、キュー 4 0 1 a ~ キュー 4 0 1 b、ジョブ 4 0 2 a ~ ジョブ 4 0 2 e、付加情報 4 0 3 a ~ 付加情報 4 0 3 f を用いて説明する。キュー 4 0 1 a は、ジョブ 4 0 2 a、ジョブ 4 0 2 b・・・ジョブ 4 0 2 e の順にジョブを順次処理する。ジョブ及びジョブに関連づけられた付加情報は、制御部によって分割された I O 要求である。図 5 A (1) では、ジョブ 4 0 2 a ~ ジョブ 4 0 2 e が、キュー 4 0 1 a の処理を待機している状態である。各ジョブ 4 0 2 a ~ ジョブ 4 0 2 e は、各ジョブ 4 0 2 a ~ ジョブ 4 0 2 e に対応した付加情報 4 0 3 a ~ 4 0 3 e を備える。図 5 の例では、付加情報 4 0 3 b と付加情報 4 0 3 d とが同一のキャッシュページの情報を持っている。

30

【 0 0 2 1 】

図 5 A (1) 制御部 2 0 1 は、キュー 4 0 1 a の処理を待っている末尾のジョブ 4 0 2 e から順に各ジョブに関連づいている付加情報をチェックする。

図 5 A (2) 制御部 2 0 1 は、同一のキャッシュページの情報を持つジョブを見つける。図 5 の例では、付加情報 4 0 3 b と付加情報 4 0 3 d とが同一のキャッシュページの情報を持っている。

40

図 5 A (3) 制御部 2 0 1 は、同一のキャッシュページの情報を持つジョブを対象にした専用の付加情報を作成する。図 5 A (3) では、付加情報 4 0 3 d を対象にした専用の付加情報 4 0 3 f を作成する。

図 5 A (4) 制御部 2 0 1 は、同一のキャッシュページの情報を持つ後続のジョブが実行される際に、並列して同一のキャッシュページの情報を持つ先頭のジョブが実行されるよう設定する。図 5 A (4) では、キュー 4 0 1 b に、ジョブ 4 0 2 b を割り当てる。付加情報 4 0 3 d は、所定のキューに格納される。

図 5 B (5) 制御部 2 0 1 は、付加情報 4 0 3 f をジョブ 4 0 2 d の付加情報として設定する。

図 5 B (6) 制御部 2 0 1 は、非同期 I O を実行する。

50

図 5 B (7) 制御部 2 0 1 は、終了した各ジョブの付加情報の情報を集める。なお、図 5 A (3) で作成された付加情報 4 0 3 f は、破棄される。

【 0 0 2 2 】

図 6 は、連続した物理アドレスを保持する書き込み要求を一括化する例を説明する図である。図 6 は、連続した物理アドレスへの書き込み要求を一括化する例を、図 6 A (1) ~ 図 6 C (9) を用いて説明している。図 6 の例を、キュー 4 1 1 a ~ キュー 4 1 1 c、ジョブ 4 1 2 a ~ ジョブ 4 1 2 f、付加情報 4 1 3 a ~ 付加情報 4 1 3 h を用いて説明する。キュー 4 1 1 a は、ジョブ 4 1 2 a、ジョブ 4 1 2 b・・・ジョブ 4 1 2 e の順にジョブを順次処理する。ジョブ及びジョブに関連づけられた付加情報は、制御部によって分割された I O 要求である。なお、図 6 の連続した物理アドレスへの書き込み要求を一括化する処理は、図 5 の同一のキャッシュページを保持する書き込み要求を一括化する処理と併せて実施される。連続した物理アドレスへの I O 要求を一括処理することで、制御部 2 0 1 は、キャッシュへのアクセス回数を減らすことができ、スナップショットの性能を向上させることができる。図 6 A (1) に示す図は、図 5 B (5) の処理後の例である。図 6 A (1) では、ジョブ 4 1 2 a ~ ジョブ 4 1 2 e が、キュー 4 1 1 a の処理を待機している状態である。ジョブ 4 1 2 f は、ジョブ 4 1 2 d の処理と並列してキュー 4 1 1 b に実行されるように設定されている。付加情報 4 1 3 d は、図 5 A (3) で作成される付加情報の例である。付加情報 4 1 3 f は、所定のキューに格納された付加情報である。各ジョブ 4 1 2 a ~ ジョブ 4 1 2 e は、各ジョブ 4 1 2 a ~ ジョブ 4 1 2 e に対応した付加情報を備える。図 6 の例では、付加情報 4 1 3 b と付加情報 4 1 3 f とが連続した物理アドレスの情報を持っている。

【 0 0 2 3 】

図 6 A (1) 制御部 2 0 1 は、キュー 4 1 1 a の処理を待っている先頭のジョブ 4 1 2 a から順に各ジョブに関連づいている付加情報をチェックする。

図 6 A (2) 制御部 2 0 1 は、連続した物理アドレスの情報を持つジョブを見つける。図 6 の例では、付加情報 4 1 3 b と付加情報 4 1 3 f とが連続した物理アドレスの情報を持っている。

図 6 A (3) 制御部 2 0 1 は、連続した物理アドレスの情報を持つジョブを対象にした専用の付加情報を作成する。図 5 6 (3) では、付加情報 4 1 3 b と付加情報 4 1 3 d を対象にした専用の付加情報 4 1 3 h を作成する。

図 6 B (4) 制御部 2 0 1 は、連続した物理アドレスの情報を持つ先頭のジョブが実行される際に、並列して連続した物理アドレスの情報を持つ後続のジョブが実行されるよう設定する。図 6 A (4) では、キュー 4 1 1 c に、ジョブ 4 1 2 d を割り当てる。付加情報 4 1 3 b は、所定のキューに格納される。

図 6 B (5) 制御部 2 0 1 は、ジョブ 4 1 2 d の付加情報 4 1 3 d を削除し、ジョブ 4 1 2 f をジョブ 4 1 2 d の後続のジョブと設定する。

図 6 B (6) 制御部 2 0 1 は、付加情報 4 1 3 h をジョブ 4 1 2 b の付加情報として設定する。

図 6 C (7) 制御部 2 0 1 は、非同期 I O を実行する。

図 6 C (8) 制御部 2 0 1 は、終了した各ジョブの付加情報の情報を集める。なお、付加情報 4 1 3 h は、破棄される。

【 0 0 2 4 】

同一のキャッシュページを持つジョブを一括化する処理及び連続した物理アドレスの情報を持つジョブを一括化する処理について、制御部 2 0 1 は、どちらか一方を行うようにしてもよく、両方の処理を行うようにしてもよい。

【 0 0 2 5 】

< スナップショットの書き込み方法 >

図 7 は、スナップショットの処理の例を説明する図である。図 7 A は、C O W 方式のスナップショットの処理の例を説明する図である。図 7 は、サーバから、データ b 及びデータ c を書き込みする要求がきた場合の例である。この例では、データ b 及びデータ c は、

マスタースライス102のデータaのチャンクを更新するためのデータである。COW方式のスナップショットの処理を、図7Aの(1)~(3)を用いて説明する。図7Aは、制御部101、マスタースライス102、COWスライス103を備える。図7Aの(1)に例示するように、サーバから書き込み要求があると、制御部101は、更新用のデータb及びデータcをメモリに保存する。制御部101は、データb及びデータcがデータを書き込む対象のデータチャンクであるデータaを、マスタースライス102から読み取る。制御部101は、データaをメモリに保存する。図7Aの(2)では、制御部101は、読み出したデータaをCOWスライス103に書き込む。図7Aの(3)において、制御部101は、データb及びデータcをCOWスライスのデータaに書き込み、データa'を作成する。

10

【0026】

図7Bは、実施形態に係るスナップショットの処理の例を説明する図である。図7Bにおいて、図1と同じものは、同一の番号を付している。実施形態に係るスナップショットの処理を図7Bの(4)~(6)を用いて説明する。図7Bの(4)に例示するように、サーバから書き込み要求があると、制御部201は、データb及びデータcをメモリに保存する。制御部201は、データb及びデータcがデータを書き込む対象であるデータaを、マスタースライス204から読み取る。制御部201は、データaをメモリに保存する。図7Bの(5)において、制御部201は、データb及びデータcをデータaに書き込み、データa'を作成する。図7Bの(6)において、制御部201は、データa'をデータスライス203に書き込む。図7Aのスナップショットでは、COWスライスへの書き込みをするI/O要求は、データaの書き込み、データbの書き込み、及び、データcの書き込みの3回発生する。一方、図7Bのスナップショットでは、データスライスへの書き込みをするI/O要求は、データa'の一回である。そのため、図7Bに示すスナップショットの方法で書き込みを行うことで、データスライスへのアクセスの回数を減らすことがで、スナップショットの処理速度が向上する。

20

【0027】

<制御部の読み込み処理の削減>

図8は、複数のチャンクに跨るデータの書き込み要求に関する処理の例を示す図である。図8は、チャンク1~チャンク3に跨ったデータに対して書き込み要求があった場合の例を示している。チャンク1及びチャンク3への書き込み要求は、チャンク1、チャンク3の一部分のデータを更新する要求である。チャンク2への書き込み要求は、チャンク2全体のデータを更新する要求である。図7Bに示すスナップショットの方法では、制御部は、チャンク1~チャンク3のデータをマスタースライスから読み取る。次に、制御部は、チャンク1~チャンク3のデータをCOWスライスに書き込む。その後、制御部は、COWスライスに書き込んだチャンク1~チャンク3のデータに対して、更新用のデータを上書きする。

30

【0028】

実施形態に係るスナップショットでは、制御部201は、チャンク全体のデータが更新されるチャンクのデータの読み込みを行わない。ここでは、チャンク2のデータを制御部201は読み込まない。チャンク2のデータは、制御部201でデータを更新される際に全部のデータが更新されてしまう。そのため、制御部201がチャンク2のデータを読み込む処理は不要である。このような、全体が更新されるチャンクの読み込みを行わないことで、無駄な読み込み処理を削減することができ、スナップショットの性能が向上する。

40

【0029】

<データスライスへの書き込み要求の一括化>

制御部201は、サーバ340からの書き込み要求を受け取ると、受領した要求を所定のサイズに分割し、処理を行う。制御部201は、分割された各々のI/O要求に応じて、データスライスにアクセスを行う。制御部201は、例えば、連続した物理アドレスに書き込みを行う場合、I/O要求をまとめて1度で実行したほうが効率的であり、データスライスへのアクセス回数を減らすことができる。そこで、実施形態に係るストレージ装置は

50

、データスライスへのＩＯ要求に対して、連続した物理アドレスへのＩＯ要求を一括化して処理を行う。連続した物理アドレスへのＩＯ要求を一括化することで、制御部２０１は、キャッシュへのアクセス回数を減らすことができ、スナップショットの性能を向上させることができる。なお、ＩＯ要求は、書き込みをするデータサイズ、書き込みをする対象となるマスタースライスのアドレス番号、キャッシュページに関する情報などの付加情報を持つので制御部はこれらの情報を用いて一括化を行う。

【００３０】

図９は、データスライスへの書き込み要求を一括処理する例を説明する図である。図９は、データスライスへの書き込み要求を一括化する例を、図９Ａ（１）～図９Ｂ（７）を用いて、説明する。図９の例を、キュー４２１ａ～キュー４２１ｂ、ジョブ４２２ａ～ジョブ４２２ｅ、付加情報４１３ａ～付加情報４１３ｆを用いて説明する。キュー４２１ａは、ジョブ４２２ａ、ジョブ４２２ｂ・・・ジョブ４２２ｅの順にジョブを順次処理する。ジョブ及びジョブに関連づけられた付加情報は、制御部によって分割されたＩＯ要求である。連続した物理アドレスに書き込みをするデータを一括処理することで、制御部２０１は、キャッシュへのアクセス回数を減らすことができ、スナップショットの性能を向上させることができる。図９Ａ（１）では、ジョブ４２２ａ～ジョブ４２２ｅが、キュー４２１ａの処理を待機している状態である。各ジョブ４２２ａ～ジョブ４２２ｅは、各ジョブ４２２ａ～ジョブ４２２ｅに対応した付加情報４２３ａ～付加情報４２３ｅを備える。図９の例では、付加情報４２３ｂと付加情報４２３ｄとが、同じマスタースライスのアドレスの情報を持っている。

【００３１】

図９Ａ（１） 制御部２０１は、キュー４２１ａの処理を待っている先頭のジョブ４２２ａから順に各ジョブに関連づいている付加情報をチェックする。

図９Ａ（２） 制御部２０１は、同じマスタースライスのアドレスの情報を持つジョブを見つける。図５（２）では、付加情報４２３ｂと付加情報４２３ｄは、同じマスタースライスのアドレス番号の情報を持っている。

図９Ａ（３） 制御部２０１は、同じマスタースライスのアドレスの情報を持つジョブを対象にした専用の付加情報を作成する。図９（３）では、付加情報４２３ｂと付加情報４２３ｄとを対象にした専用の付加情報４２３ｆを作成する。

図９Ａ（４） 制御部２０１は、同じマスタースライスのアドレスの情報を持つジョブのうち、後続にあるジョブを先頭のジョブが実行される際に、並列して処理されるよう設定する。図９（４）では、キュー４２１ｂに、ジョブ４２２ｄを割り当てる。付加情報４２３ｂは、処理が終了した付加情報として、他のキューに保存される。

図９Ｂ（５） 制御部２０１は、付加情報４２３ｆをジョブ４２２ｂの付加情報として設定する。

図９Ｂ（６） 制御部２０１は、非同期ＩＯを実行する。

図９Ｂ（７） 制御部２０１は、終了した各ジョブの付加情報の情報を集める。なお、図９（３）で作成する付加情報は、破棄される。

【００３２】

<ダウンリカバリーについて>

実施形態に係るメタスライス２０２は、データスライスのどこかのアドレスまでデータを書き込みしたかを示す、データ書き込み処理の進行具合を示す情報である履歴情報を保存している。履歴情報は、更に、メタスライスへの書き込み処理の進行具合を示す情報を保存している。制御部２０１は、データスライスの書き込み処理が終了し、その後、メタスライスへの書き込み処理を行う際に、履歴情報を更新する。制御部２０１が、履歴情報を更新中に何らかの障害が発生し、履歴情報が更新できなかった場合、履歴情報と実際の状態にずれが発生する可能性がある。何らかの障害が発生しても、履歴情報のリカバリーができるのが好ましい。なお、履歴情報は、制御部２０１を停止する際にも更新される。

【００３３】

図１０は、履歴情報のダウンリカバリーの例を示す図である。図１０は、図１と同じも

のは、同一の番号を付している。図10は、制御部201がデータ1～データ9をデータスライスに書き込み済みであるにも関わらず、何らかの障害によりデータ7に関するメタ情報のメタスライスへの書き込みが失敗した例である。なお、データ1～データ6、データ8～データ9のメタ情報は、メタスライス202への書き込みが成功している。図10(a)は、障害発生後に、制御部201を起動した際のストレージ装置の例である。図10(a)のデータスライス203は、データ1～データ9を保存している。図10(a)のメタスライス202は、データ1～データ6、データ8～データ9のメタ情報を保存している。メタスライス202のアドレス3が示す領域には、データ1とデータ4のメタ情報が格納されている。メタスライス202のアドレス4が示す領域には、データ2とデータ3のメタ情報が格納されている。メタスライス202のアドレス5が示す領域には、データ8とデータ9のメタ情報が格納されている。メタスライス202のアドレス6が示す領域には、データ5とデータ6のメタ情報が格納されている。メタスライス202のアドレス7が示す領域は、データ7のメタ情報を格納する領域である。

10

【0034】

履歴情報は、既にデータスライスに書き込み済みのデータチャンク数、メタスライスに更新中のメタチャンクの数、メタスライスに更新中のメタチャンクのアドレス情報を有する。履歴情報の既にデータスライスに書き込み済みのデータチャンク数は、データスライスへのデータ書き込み処理の進行具合を示す情報である。履歴情報のメタスライスに更新中のメタチャンクの数と、メタスライスに更新中のメタチャンクのアドレス情報とは、メタスライスへの書き込み処理の進行具合を示す情報である。図10(a)の履歴情報の例は、既にデータスライスに書き込み済みのデータチャンク数が4、メタスライスに更新中のメタチャンク数が3、メタスライスに更新中のメタチャンクのアドレス情報が5, 6, 7を保持している。データスライス203では、データ1～データ9までの書き込みが成功しているにも関わらず、履歴情報には、書き込み済みのデータチャンク数が4となっており、履歴情報と実際の状態にずれが生じている。そのため、制御部201は、履歴情報のリカバリーを行う。なお、ストレージ装置の制御部201が正常に停止された場合、更新中のメタチャンク数は0が格納される。そのため、制御部201は、起動したときに取得した履歴情報の更新中のメタチャンク数が0でなかった場合、リカバリーを行う。

20

【0035】

制御部201は、データスライスに書き込みが成功したデータのメタ情報をリストアップする。ここで、履歴情報は、データ1～データ4までのデータが、データスライスに書き込みが成功していること示している。そのため、履歴情報に含まれる情報は、リストアップの対象から外す。制御部201は、データスライスに書き込みが成功したデータは、データ5、データ6、データ8、データ9と判断する。制御部201は、障害によってメタスライスに書き込みが失敗したデータを、履歴情報が有するメタスライスに更新中のメタチャンクのアドレス情報を用いてメタスライスから探す。制御部201は、書き込みが失敗したデータよりも前に書き込まれたデータ5, データ6を正常に書き込みが成功したデータと判断する。制御部201は、データ6までが正常に処理が行われたことを、履歴情報に書き込む。図10(b)は、履歴情報のリカバリーが行われた後の例を示すストレージ装置の図である。制御部201によって、履歴情報の書き込み済みのデータチャンク数は、6となっている。逆に、制御部201は、データ7よりも後に書き込まれるデータ8, データ9は正常に書き込みがされたとは判断せず、無効なデータと判断する。そのため、メタスライス202からデータ8、データ9に関するメタ情報を削除する。制御部201は、書き込みが失敗したデータからバックアップを再開する。

30

40

【0036】

<メタデータのプリフェッチ>

図11は、メタデータのプリフェッチの例を示す図である。本実施形態では、マッピング情報の一部をキャッシュ上で扱うことによって、スナップショットの性能向上が見込まれる。しかしながら、キャッシュヒットした場合以外は、キャッシュを用いたマッピング

50

情報はスナップショットの性能向上の効果を得られない。シーケンシャルアクセスでは、複数のマッピング情報を取得する。連続した全てのマッピング情報に対してキャッシュヒットをしない限り、データスライスへの書き込み要求のたびに、メタスライスのマッピング情報をキャッシュに読み込む処理をすることになる。書き込み処理及び、読み込む処理が増えると、スナップショットの性能は劣化してしまう。メタ情報のキャッシュで参照の可能性の高いマッピング情報をプリフェッチすることで、キャッシュヒットする確率を向上させ、スナップショットの性能劣化を緩和する。

【 0 0 3 7 】

制御部 201 は、物理アドレスを閾値以上の数備え、連続した物理アドレスへの I/O 要求を一括化したデータをシーケンシャルアクセスと判定する。連続した物理アドレスへの I/O 要求を一括化したデータは、制御部に用いられるキャッシュページに関する情報を有している。制御部は、使用すると予想されるキャッシュページの情報をキャッシュが持っているかを判定する。制御部は、キャッシュが使用すると予想されるキャッシュページを持っていなかった場合、先にキャッシュページをキャッシュに読み込むプリフェッチを実行する。図 11 は、図 1 と同一のものは、同一の番号を付している。図 11 のストレージ装置は、メタスライス 202 が読み込むキャッシュ領域 210 を有する。マスタースライス 204 は、チャンク 0 ~ チャンク n を備えるデータ群であり、図 11 は、チャンク 0 ~ チャンク 8 までのシーケンシャルアクセスがあった場合の例である。図 11 のメタスライス 202 が備えるマッピング情報は、キャッシュページ 00 ~ キャッシュページ 08 を持つ。シーケンシャルアクセスで使用されるチャンク 0 ~ チャンク 5 のデータは、キャッシュページ 04 を使用するもので、チャンク 6 ~ チャンク 8 のデータは、キャッシュページ 05 を使用するものとする。図 11 のキャッシュ領域 210 は、キャッシュページ 02 ~ キャッシュページ 04 までを保持しているとする。制御部 201 は、シーケンシャルアクセスによって、チャンク 0 ~ チャンク 5 にアクセスが発生した場合、次に使用されると予想されるキャッシュページ 05 を予めキャッシュ領域 210 に読み込むプリフェッチを行う。プリフェッチを行うことで、次にキャッシュ領域 210 で参照される可能性の高い情報をキャッシュ領域に読み込むことができ、キャッシュヒットする確率が高くなり、スナップショットの性能が向上する。また、キャッシュ領域のキャッシュページはアクセスの統計情報を取得し、使用頻度の少ないキャッシュページを削除していくようにしてもよい。

【 0 0 3 8 】

図 12 は、I/O 要求を受領した制御部の処理の例を説明するフローチャートである。制御部 201 は、サーバより I/O 要求を受領する（ステップ S101）。制御部 201 は、サーバより受領した I/O 要求のデータサイズが閾値以上かを判断する（ステップ S102）。制御部 201 は、I/O 要求を分割する（ステップ S103、ステップ S102 で YES）。制御部 201 は、分割された I/O 要求が書き込みを行う予定のデータが、現在使用中であるかを判断する（ステップ S104、S103 の処理後及び S102 で NO）。制御部 201 は、分割された I/O 要求を後続のキューとして設定する（ステップ S106）。制御部 201 は、ステップ S106 の処理が終わると、処理を S104 に返す。制御部 201 は、マッピング情報を参照する（ステップ S105、ステップ S104 で NO）。ステップ S105 は、図 13 のフローチャートで詳細を示す。

図 13 は、制御部がマッピング情報を参照する処理の例を説明するフローチャートである。制御部 201 は、I/O 要求が書き込みを行うデータのマスタースライスのアドレスがキャッシュに含まれるかを判断する（ステップ S201）。制御部 201 は、メタスライスからキャッシュ領域にマッピング情報を読み込む（ステップ S202、ステップ S201 で NO）。制御部 201 は、I/O 要求が更新するマスタースライスのアドレスに、データスライスのチャンクのアドレスが割り当てられているかを判断する（ステップ S203、ステップ S202 の後又はステップ S201 で YES）。制御部 201 は、I/O 要求のマスタースライスのアドレスをデータスライスのチャンクアドレスに変換する（ステップ S204、ステップ S203 で YES）。制御部 201 は、I/O 要求が使用するマスタースライスのアドレスを使用中とする（ステップ S205）。制御部 201 は、データの Kopi

10

20

30

40

50

ーに関する処理を行う（ステップS 2 0 6）。ステップS 2 0 6は、図1 4にフローチャートで詳細を示す。

【0 0 3 9】

図1 4は、データのコピーに関する処理の例を説明するフローチャートである。制御部2 0 1は、I O要求で更新するデータが、各チャンクにおいて一部コピーかを判断する（ステップS 3 0 1）。ステップS 3 0 1の処理は、図8の説明に該当する。ステップS 3 0 1でNOと判断されるチャンクは、図8のチャンク2のように全体を更新対象とされるチャンクであり、この場合、制御部2 0 1へのチャンク2のコピーはしなくてよい。一方、一部コピーは、図8のチャンク1, 3であり、制御部2 0 1へチャンクをコピーする。制御部2 0 1は、チャンクを制御部2 0 1に読み込む（ステップS 3 0 2、ステップS 3 0 1でYES）。制御部2 0 1は、図1 2のS 1 0 3で分割したI O要求を一括化する（ステップS 3 0 3、ステップS 3 0 1でNO）。ステップS 3 0 3の一括化処理は、図5, 図6, 図9で説明した処理である。制御部2 0 1は、I O要求の更新用データを制御部内で更新し、更新後のデータをデータスライス2 0 3に書き込む（ステップS 3 0 4）。制御部2 0 1は、マッピング情報の書き込み処理を行う（ステップS 3 0 5）。マッピングの書き込み処理は、図1 5で詳細に説明する。

10

【0 0 4 0】

図1 5は、制御部がメタスライスにメタ情報を書き込む処理を説明するフローチャートである。制御部2 0 1は、履歴情報及びプリフェッチに関する情報をメタスライス2 0 2に書き込む（ステップS 4 0 1）。制御部2 0 1は、マッピング情報をメタスライス2 0 2に書き込む（ステップS 4 0 2）。ステップS 4 0 2の処理が終了すると、制御部2 0 1は処理を終了する。

20

【0 0 4 1】

図1 6は、同一のキャッシュページを保持する書き込み要求を一括化する例を説明するフローチャートである。制御部2 0 1は、キューイングされているジョブの付加情報を末尾にキューイングされているジョブからチェックする（ステップS 5 0 1）。制御部2 0 1は、同一のキャッシュページを持つジョブがあるかを判断する（ステップS 5 0 2）。制御部2 0 1は、同一のキャッシュページを持つジョブのうち、後続側にキューイングされているジョブの付加情報を対象にした専用の付加情報を作成する（ステップS 5 0 3、ステップS 5 0 2でYES）。制御部2 0 1は、同一のキャッシュページを持つジョブのうち、後続にキューイングされているジョブが実行される際に、前方にキューイングされているジョブが実行されるよう設定する（ステップS 5 0 4）。制御部2 0 1は、専用の付加情報を後続にキューイングされているジョブに設定する（ステップS 5 0 5）。制御部2 0 1は、非同期I Oを実行する（ステップS 5 0 6、ステップS 5 0 2でNO）。制御部2 0 1は、全ジョブの付加情報を収集する（ステップS 5 0 7）。制御部2 0 1は、同一のキャッシュページを保持する書き込み要求を一括化する処理を終了する。

30

【0 0 4 2】

図1 7は、連続した物理アドレスへの書き込み要求を一括化する例を説明するフローチャートである。制御部2 0 1は、キューイングされているジョブの付加情報を先頭にキューイングされているジョブからチェックする（ステップS 6 0 1）。制御部2 0 1は、連続した物理アドレスを持つジョブがあるかを判断する（ステップS 6 0 2）。制御部2 0 1は、連続した物理アドレスを持つジョブを対象にした専用の付加情報を作成する（ステップS 6 0 3、ステップS 6 0 2でYES）。制御部2 0 1は、連続した物理アドレスの情報を持つ先頭のジョブが実行される際に、並列して連続した物理アドレスの情報を持つ後続のジョブが実行されるよう設定する（ステップS 6 0 4）。制御部2 0 1は、連続した物理アドレスの情報を持つ後続のジョブが、更に並列して実行されるジョブを保持しているかを判断する（ステップS 6 0 5）。制御部2 0 1は、連続した物理アドレスの情報を持つ後続のジョブの更に後続に、並列して実行されるジョブをキューイングする（ステップS 6 0 6、ステップS 6 0 5でYES）。制御部2 0 1は、専用の付加情報を、連続した物理アドレスの情報を持つ先頭のジョブに設定する（ステップS 6 0 7、ステップS

40

50

605でNO)。制御部201は、非同期IOを実行する(ステップS608、ステップS602でNO)。制御部201は、全ジョブの付加情報を収集する(ステップS609)。制御部201は、連続した物理アドレスへの書き込み要求を一括化する処理を終了する。

【0043】

図18は、データスライスへの書き込み要求を一括処理する例を説明するフローチャートである。制御部201は、キューイングされているジョブの付加情報を先頭にキューイングされているジョブからチェックする(ステップS701)。制御部201は、同じマスタースライスのアドレスの情報を持つジョブがあるかを判断する(ステップS702) 10
制御部201は、同じマスタースライスのアドレスの情報を持つジョブを対象にした専用の付加情報を作成する(ステップS703、ステップS702でYES)。制御部201は、同じマスタースライスのアドレスの情報を持つジョブのうち、前方にキューイングされているジョブが実行される際に、後方にキューイングされているジョブが実行されるよう設定する(ステップS704)。制御部201は、専用の付加情報を前方にキューイングされているジョブに設定する(ステップS705)。制御部201は、非同期IOを実行する(ステップS706、ステップS702でNO)。制御部201は、全ジョブの付加情報を収集する(ステップS707)。制御部201は、データスライスへの書き込み要求を一括化する処理を終了する。

【0044】

図19は、履歴情報のダウンリカバリーの例を説明するフローチャートである。制御部201は、起動するとメタスライスから履歴情報を読み込む(ステップS801)。制御部201は、履歴情報に含まれる既書き込み済みのデータチャンク数と、更新中のメタチャンクの数 20
の情報を取得する(ステップS802)。制御部201は、更新中のメタチャンクの数か0かどうかを判断する(ステップS803)。制御部201は、データスライスに書き込みが成功したデータのメタ情報をリストアップする(ステップS804、ステップS803でNO)。制御部201は、障害によってメタスライスに書き込みが失敗したデータを抽出する(ステップS805)。制御部201は、書き込みが失敗したデータよりも前に書き込まれたデータを、正常に書き込みが成功したデータと判断する(ステップS806)。制御部201は、正常に書き込みが成功したデータを用いて履歴情報を修復する(ステップS807)。制御部201は、S804でリストアップされたメタ情報 30
のうち、書き込みが失敗したデータよりも後に書き込まれたメタ情報を、メタスライスから削除する(ステップS808)。制御部201は、履歴情報をメタスライスに書き込む(ステップS809)。制御部201は、既書き込み済みのデータチャンク数の情報を取得する(ステップS810、ステップS803でYES)。

【0045】

以上、説明したように、実施形態にかかる方法では、ストレージ装置の大規模化及びシステムの大規模化に伴い、マッピング情報の量が増えても、スナップショットの性能を維持することができる。

【0046】

上述の実施形態に関し、さらに以下の付記を開示する。 40

(付記1)

ホストにより更新がなされる更新対象の元データを格納する第1の格納領域と、

前記第1の格納領域に格納された前記更新対象の元データに対して書き込むべきデータを複数の更新データに分割することにより、連続したアドレスごとに前記複数の更新データを振り分ける制御部と、

前記制御部により振り分けられた前記複数の更新データを格納する第2の格納領域と、

前記更新対象の元データの前記複数の更新データによって上書きされるべき前記第1の格納領域のアドレスである更新領域アドレスを、前記複数の更新データを格納している前記第2の格納領域のアドレスである格納先アドレスに対応付けた情報を、格納する第3の格納領域を備える、 50

ことを特徴とするストレージ制御装置。

(付記 2)

前記更新領域アドレスと前記格納先アドレスを対応付けた情報は、前記更新領域アドレスが示す数の順に前記格納先アドレスを格納することにより示された情報であり、前記第 3 の格納領域及びキャッシュに保存される

ことを特徴とする付記 1 記載のストレージ制御装置。

(付記 3)

前記更新領域アドレスと前記格納先アドレスを対応付けた情報を前記第 3 の格納領域に格納する際に、

前記制御部は、前記複数の更新データを、更に、同一のキャッシュページごとに振り分ける

ことを特徴とする付記 1 又は付記 2 記載のストレージ制御装置。

(付記 4)

前記第 3 の格納領域は、前記第 2 の格納領域への前記複数の更新データの書き込み処理の進行具合を示す情報と、前記更新領域アドレスと前記格納アドレスを対応付けた情報である対応情報の書き込み処理の進行具合を示す情報とを、履歴情報として保存しており、

前記制御部は、起動する際に前記履歴情報を参照し、前記履歴情報に、前記制御部が停止する前に前記対応情報の書き込み処理が終了していない更新データに関する情報が含まれていた場合、

前記対応情報の書き込み処理が終了していない更新データよりも先に処理されたデータを正常に実行された更新データであると判定し、前記正常に実行された更新データに関する情報を、前記履歴情報に保存し、

前記対応情報の書き込み処理が終了していない更新データからバックアップを再開することを特徴とする付記 1 ~ 3 のいずれかに記載のストレージ制御装置。

(付記 5)

前記制御部は、前記複数の更新データを、シーケンシャルアクセスを行うデータか、ランダムアクセスを行うデータか判定し、

前記複数の更新データがシーケンシャルアクセスを行うデータであった場合、

前記複数の更新データが上書きする領域を指定するアドレスを含むキャッシュページの次のキャッシュページをあらかじめ前記キャッシュに保存する

ことを特徴とする付記 2 に記載のストレージ制御装置。

(付記 6)

ホストにより更新がなされる更新対象の元データを第 1 の格納領域に格納し、前記第 1 の格納領域に格納された前記更新対象の元データに対して書き込むべきデータを複数の更新データに分割することにより、連続したアドレスごとに前記複数の更新データを振り分け、

振り分けられた前記複数の更新データを第 2 の格納領域に格納し、

前記更新対象の元データの前記複数の更新データによって上書きされるべき前記第 1 の格納領域のアドレスである更新領域アドレスを、前記複数の更新データを格納している前記第 2 の格納領域のアドレスである格納先アドレスに対応付けた情報を、第 3 の格納領域に格納する処理をストレージ装置に実行させる

ことを特徴とする制御プログラム。

(付記 7)

ホストにより更新がなされる更新対象の元データを第 1 の格納領域に格納し、前記第 1 の格納領域に格納された前記更新対象の元データに対して書き込むべきデータを複数の更新データに分割することにより、連続したアドレスごとに前記複数の更新データを振り分け、

振り分けられた前記複数の更新データを第 2 の格納領域に格納し、

前記更新対象の元データの前記複数の更新データによって上書きされるべき前記第 1 の格納領域のアドレスである更新領域アドレスを、前記複数の更新データを格納している前

10

20

30

40

50

記第2の格納領域のアドレスである格納先アドレスに対応付けた情報を、第3の格納領域に格納する処理をストレージ装置に実行させる

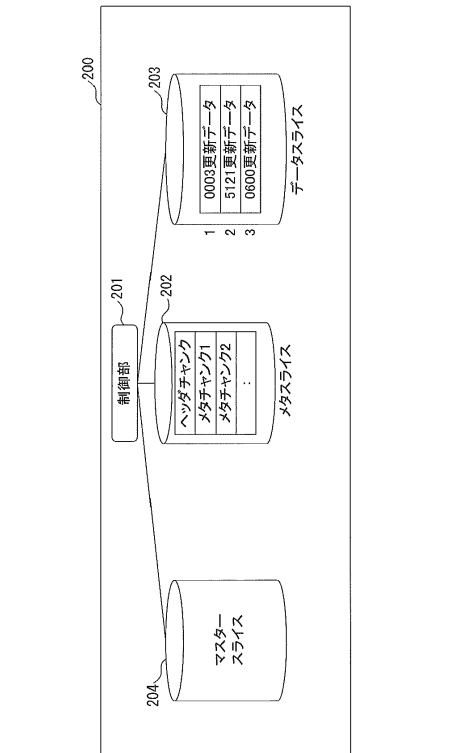
ことを特徴とする制御方法。

【符号の説明】

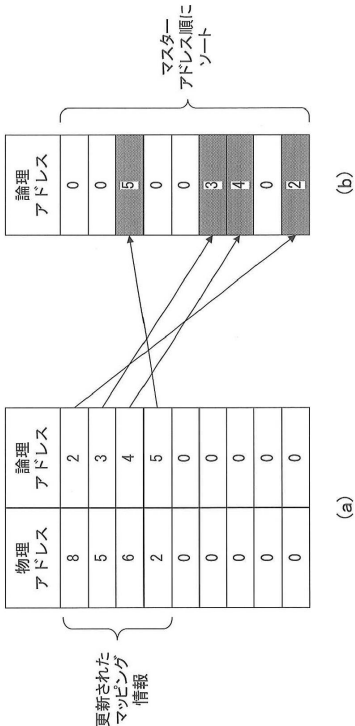
【0047】

101	制御部	
102	マスタースライス	
103	COWスライス	
200	ストレージ装置	
201	制御部	10
202	メタスライス	
203	データスライス	
204	マスタースライス	
210	キャッシュ領域	
300	ストレージ装置	
310	制御装置	
311	管理部	
312	CPU	
313	メモリ	
314	HDD	20
315	制御部	
316	VDISK	
317	管理用DB	
320	制御装置	
321	制御部	
322	CPU	
323	メモリ	
324	VDISK	
325	HDD	
330	記憶装置	30
331a ~ 331n	LUN	
340	サーバ	
401a ~ 401b	キュー	
402a ~ 402e	ジョブ	
403a ~ 403f	付加情報	
411a ~ 411c	キュー	
412a ~ 412f	ジョブ	
413a ~ 413h	付加情報	
421a ~ 421b	キュー	
422a ~ 422e	ジョブ	40
413a ~ 413f	付加情報	

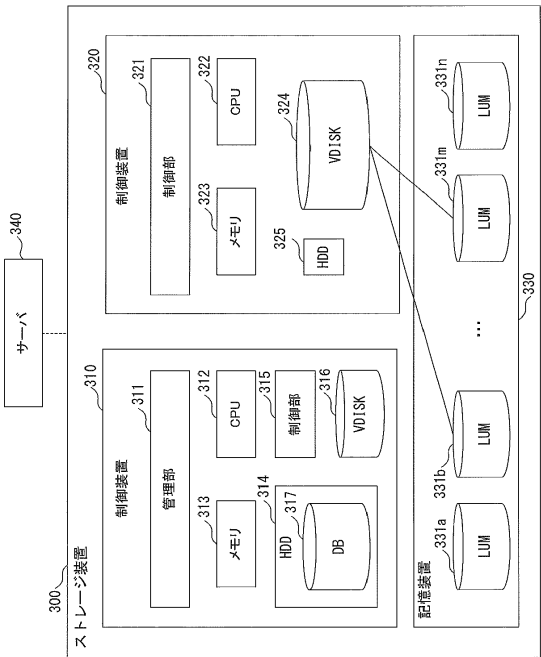
【図 1】
ストレージ装置を説明する例を示す図



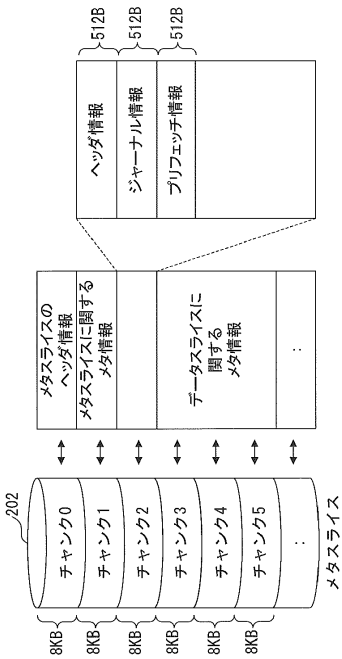
【図 2】
マッピング情報の例を示す図



【図 3】
ストレージ装置の構成の例を示す図

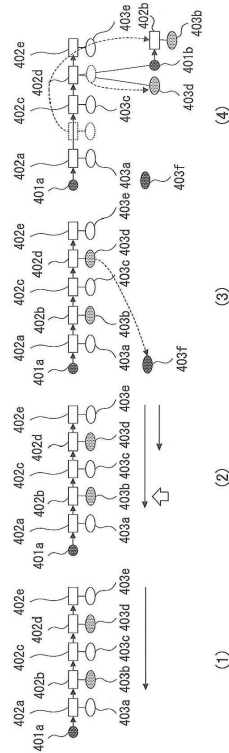


【図 4】
メタスライスの例を詳細に説明した図



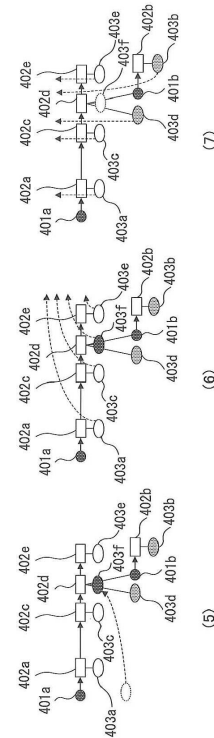
【図 5 A】

同一のキャッシュページを保持する
書き込み要求を一括化する例を説明する図



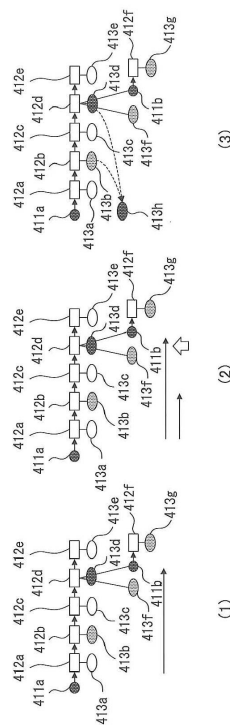
【図 5 B】

同一のキャッシュページを保持する
書き込み要求を一括化する例を説明する図



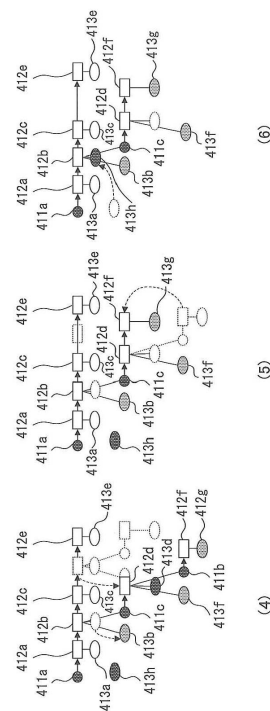
【図 6 A】

連続した物理アドレスを保持する
書き込み要求を一括化する例を説明する図



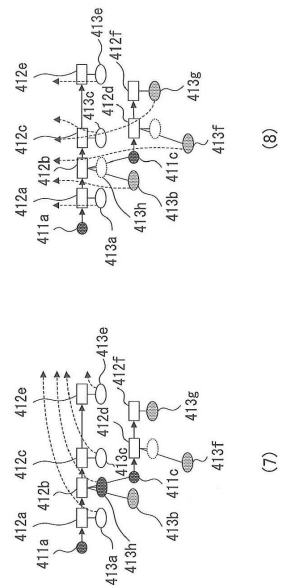
【図 6 B】

連続した物理アドレスを保持する
書き込み要求を一括化する例を説明する図



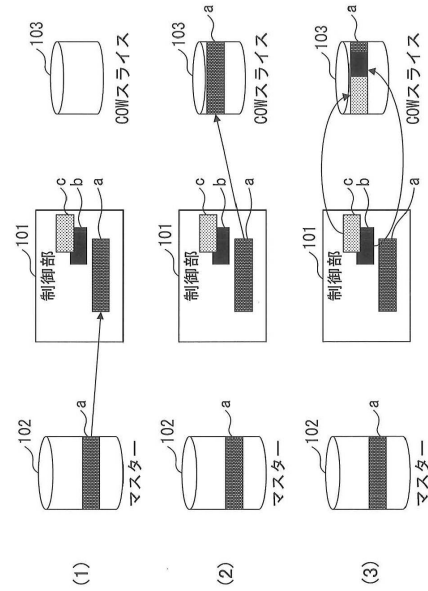
【図 6 C】

連続した物理アドレスを保持する
書き込み要求を一括化する例を説明する図



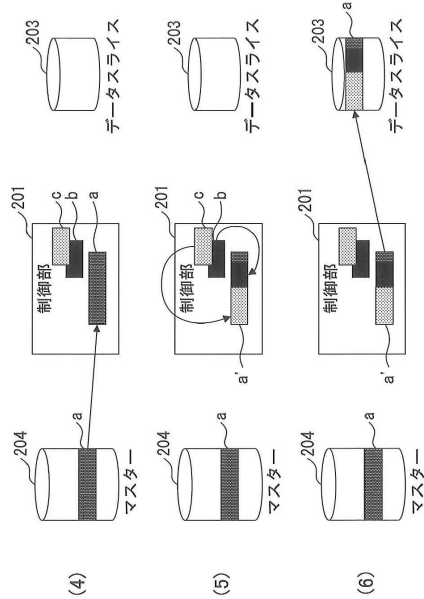
【図 7 A】

スナップショットの処理の例を説明する図



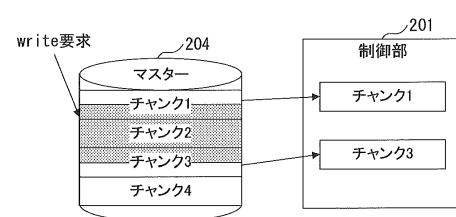
【図 7 B】

スナップショットの処理の例を説明する図



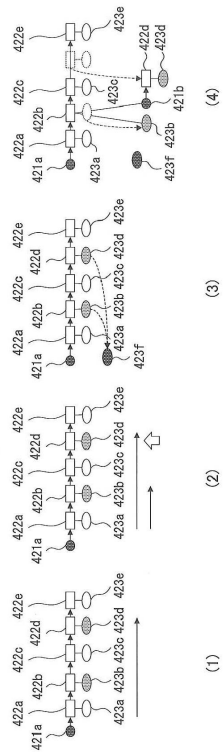
【図 8】

複数のチャンクに跨るデータの
書き込み要求に関する処理の例を示す図



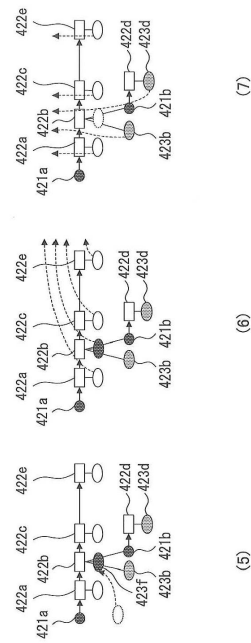
【 図 9 A 】

データスライスへの書き込み要求を一括処理する例を説明する図



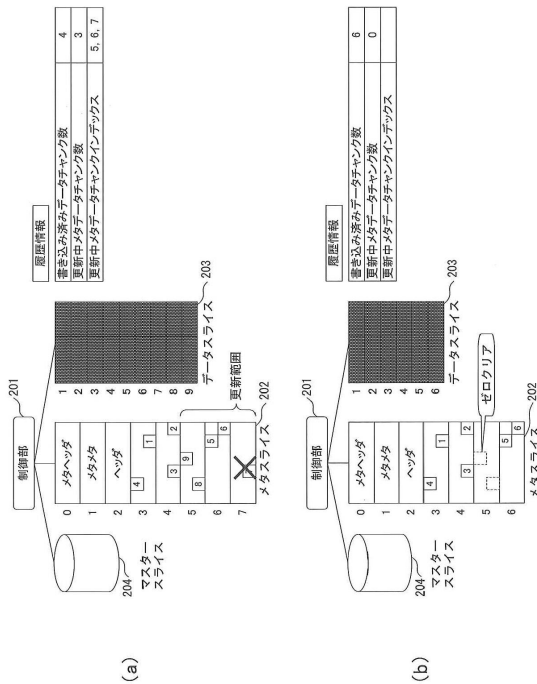
【 図 9 B 】

データスライスへの書き込み要求を一括処理する例を説明する図



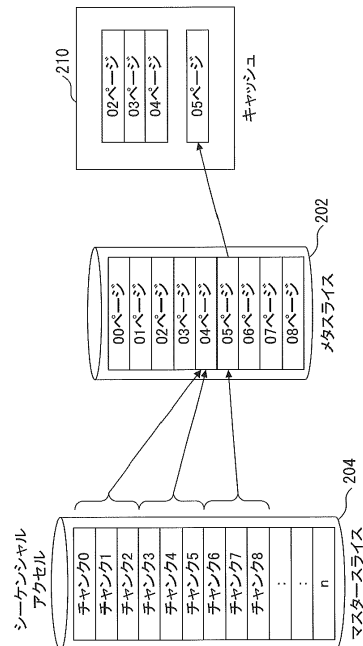
【 ㄨ 1 0 】

履歴情報のダウンリカバリーの例を示す図



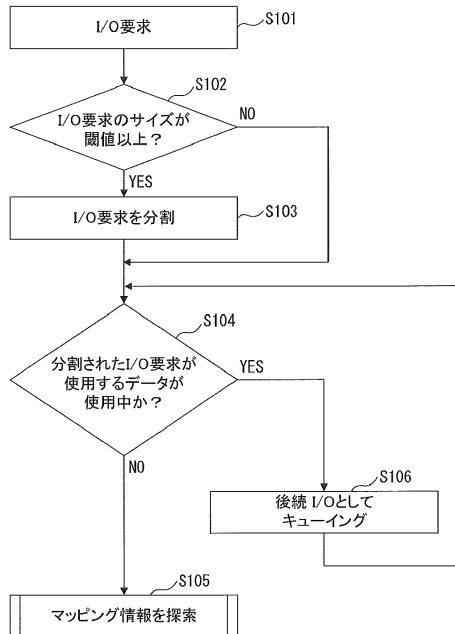
【 図 1 1 】

メタデータのプリフェッチの例を示す図



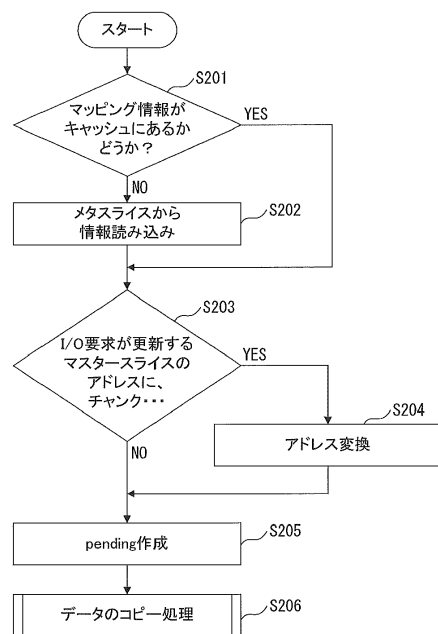
【図 12】

I/O要求を受領した制御部の
処理の例を説明するフローチャート



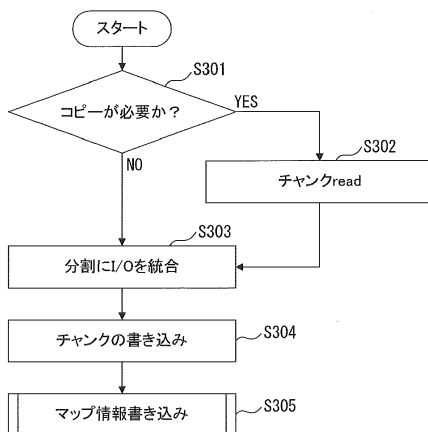
【図 13】

制御部がマッピング情報を参照する
処理の例を説明するフローチャート



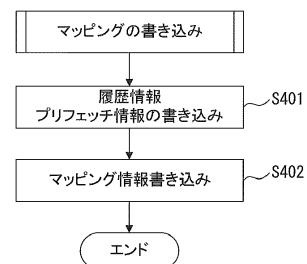
【図 14】

データのコピーに関する
処理の例を説明するフローチャート



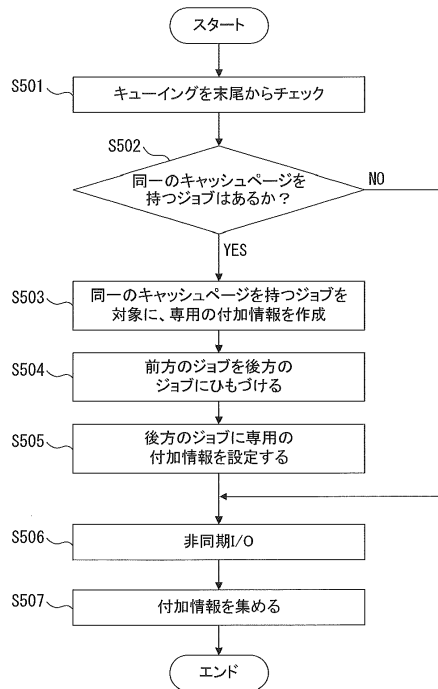
【図 15】

制御部がメタスライスにメタ情報を
書き込む処理を説明するフローチャート



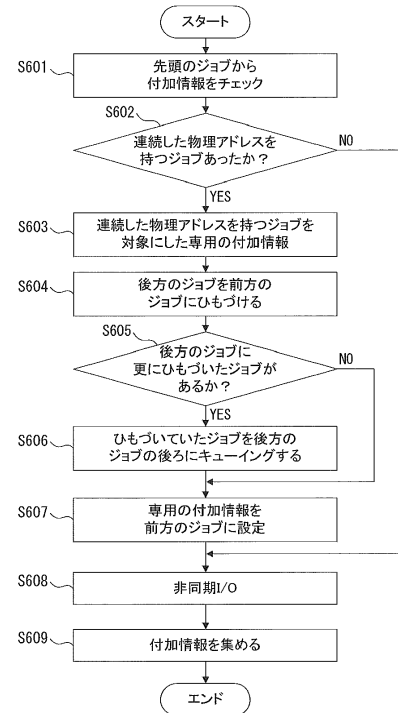
【図 16】

同一のキャッシュページを保持する書き込み要求を一括化する例を説明するフローチャート



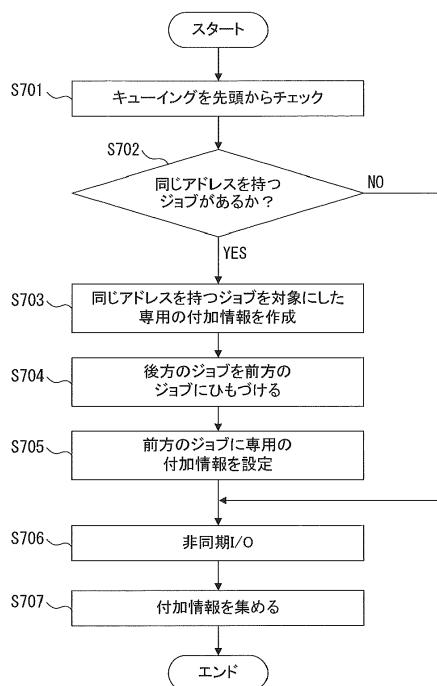
【図 17】

連続した物理アドレスを保持する書き込み要求を一括化する例を説明するフローチャート



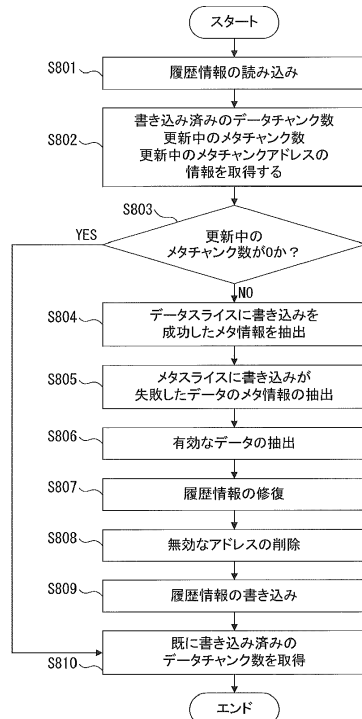
【図 18】

データスライスへの書き込み要求を一括処理する例を説明するフローチャート



【図 19】

履歴情報のダウンリカバリーの例を説明するフローチャート



フロントページの続き

- (72)発明者 村山 浩
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 小橋 一範
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 田中 啓介

- (56)参考文献 特開2010-020570(JP,A)
国際公開第2012/090239(WO,A1)
特開2010-122761(JP,A)
特表2013-504820(JP,A)

- (58)調査した分野(Int.Cl., DB名)
G06F3/06-3/08
12/00、12/08-12/12