(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第4988214号
(P4988214)

(45)発行日　平成24年8月1日(2012.8.1)　(24)登録日　平成24年5月11日(2012.5.11)

(51)Int.Cl.		F I		
G06F 3/06	(2006.01)	G06F	3/06	304U
G06F 12/08	(2006.01)	G06F	12/08	557
		G06F	12/08	543Z
		G06F	12/08	551J

請求項の数 2　(全 14 頁)

(21)出願番号　特願2006-17056(P2006-17056)
(22)出願日　平成18年1月26日(2006.1.26)
(65)公開番号　特開2006-221623(P2006-221623A)
(43)公開日　平成18年8月24日(2006.8.24)
審査請求日　平成20年10月24日(2008.10.24)
(31)優先権主張番号　11/053226
(32)優先日　平成17年2月7日(2005.2.7)
(33)優先権主張国　米国(US)

前置審査

(73)特許権者　390009531
インターナショナル・ビジネス・マシーンズ・コーポレーション
ＩＮＴＥＲＮＡＴＩＯＮＡＬ　ＢＵＳＩＮＥＳＳ　ＭＡＳＣＨＩＮＥＳ　ＣＯＲＰＯＲＡＴＩＯＮ
アメリカ合衆国１０５０４　ニューヨーク州　アーモンク　ニュー　オーチャード　ロード
(74)代理人　100108501
弁理士　上野　剛史
(74)代理人　100112690
弁理士　太佐　種一
(74)代理人　100091568
弁理士　市位　嘉宏

最終頁に続く

(54)【発明の名称】記憶装置における喪失した書き込みの検出および回復

(57)【特許請求の範囲】
【請求項１】
記憶媒体にデータを書き込むための要求を受け取るステップと、
前記記憶媒体に書き込むように要求された前記データを、キャッシュに格納するステップと、
前記記憶媒体への前記データの書き込みを開始するステップと、
前記キャッシュに格納された前記データが前記記憶媒体に書き込まれた前記データと同じであるかどうかを定期的に判定するステップと、
前記キャッシュ内のエントリの数がしきい値を超えるかどうかを判定するステップを有し、
前記エントリはキャッシュに入れられた書き込みに対応しており、
合体された非重複書き込みアドレスのリストを生成するために、前記エントリ内の書き込みアドレスを合体するステップと、
前記合体された非重複書き込みアドレスのリストを順序付けするステップをさらに有し、
低順位の合体された非重複書き込みアドレスより前に、高順位の合体された非重複書き込みアドレスが、前記記憶媒体に対する書き込みエラーについて検証される、方法。
【請求項２】
前記しきい値を超えることは、前記キャッシュが満杯の１０パーセントを上回ることを示す、請求項１に記載の方法。

【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、記憶装置における喪失した書き込みの検出および回復のための方法に係る。
【背景技術】
【０００２】
ディスク・ドライブにデータを書き込む書き込み動作は、断続的または永続的に障害が発生する可能性がある。こうした障害を検出するために、ある種のドライブは、書き込み電流が一定のしきい値よりも低い場合に検出する、ヘッド読み取りおよび書き込み前置増幅（preamplification）回路を使用する。この検出回路に加えられた複雑性が、信頼性の問題を提示する可能性がある。さらに、検出しきい値の設定が、すべての書き込みエラーの検出を保証するものでない場合がある。
【０００３】
ある種の実施では、アクチュエータを定期的にディスクの予約領域へ移動させ、その後書き込んで、各ヘッドによって書き込まれた内容を検証することによる、ディスク・ドライブの書き込み喪失の問題がある。このすべてのヘッドに関する書き込み能力の検証は、永続問題用自己テスト（persistent problem self test：ＰＰＳＴ)と呼ばれることがある。このメカニズムは、書き込み問題が永続的である場合、すなわち、書き込みを喪失しているドライブが不良ヘッドに関する後続のすべての書き込みの喪失を継続する場合にのみ、喪失した書き込みを検出する。ＰＰＳＴ検証メカニズムは、断続的に喪失した書き込みの検出には十分に有効ではない。加えてＰＰＳＴ検証は、連続するＰＰＳＴ書き込み能力検証の間に喪失した書き込みが原因で書き込まれなかったデータを回復させることはできない。さらに、データ破損の量を最小限にするためにＰＰＳＴ検証の頻度が上がった場合、入力／出力（Ｉ／Ｏ）性能は許容不可能なレベルまで低下する可能性がある。
【０００４】
ある種の実施では、連続するＰＰＳＴ検証間にキャッシュ内のすべての書き込みを維持することによって、ＰＰＳＴ検証を増補することができる。誤った書き込みが検出された場合、データはキャッシュから直接回復される。ヘッドの定期的チェックでは断続的に喪失する書き込みが検出されない場合があるため、このメカニズムでは断続的に喪失する書き込みは保護しない。さらに、ＰＰＳＴ検証のオーバヘッドによる性能の低下を緩和するために、かなり大型で潜在的に高価な専用キャッシュが必要となる可能性がある。すべての書き込みを取り込むために必要なキャッシュは、ＰＰＳＴ検証間の時間間隔の増加と共に増大する。
【０００５】
断続的および永続的の両方の喪失した書き込みを保護するある種の実施では、それぞれの書き込み動作に対して書き込み検証を実行することが可能であり、書き込みが実行されるごとにディスク・ドライブが回転を完了し、書き込まれたばかりのデータを読み取って、その書き込まれたばかりのデータと書き込みバッファ内のデータとを比較する。これによってデータが損失しないことは保証されるが、ドライブの待ち時間が追加され、結果として生じるＩ／Ｏ性能が受け入れ不可能なものとなる可能性がある。
【発明の開示】
【発明が解決しようとする課題】
【０００６】
本発明の目的は、記憶媒体における喪失した書き込みの検出および回復のための方法を提供することである。
【課題を解決するための手段】
【０００７】
記憶媒体に書き込むように要求されたデータがキャッシュに格納される。記憶媒体へのデータの書き込みが開始される。キャッシュに格納されたデータと記憶媒体に書き込まれたデータとが同じであるかどうかが、定期的に判別される。

【０００８】
追加の実施形態では、キャッシュに格納されたデータと記憶媒体に書き込まれたデータとが同じであるという判定に応答して、格納されたデータがキャッシュから除去される。
【０００９】
さらに追加の実施形態では、キャッシュに格納されたデータと記憶媒体に書き込まれたデータとが同じでないという判定に応答して、記憶媒体は書き込み保護される。記憶媒体に対して喪失された少なくとも１つの書き込みを示すエラーが生成される。
【００１０】
他の実施形態では、記憶媒体はディスクであり、キャッシュはディスクよりも記憶容量が小さく、ディスクに比べてキャッシュにデータを書き込む方が信頼性が高く、ホスト・アプリケーションは記憶媒体からではなくキャッシュから未検証のデータを読み取る。
【００１１】
さらに他の実施形態では、キャッシュ内のエントリの数がしきい値を超えたかどうかが判別され、エントリはキャッシュに入れられた書き込みに対応する。エントリ内の書き込みアドレスが合体され、合体済み書き込みアドレスのリストを生成する。合体済み書き込みアドレスのリストが順序付けされ、低順位の合体済み書き込みアドレスより前に、記憶媒体への書き込みエラーについて高順位の合体済み書き込みアドレスが検証される。ある種の実施形態では、しきい値を超えることはキャッシュが満杯の１０パーセントを上回ることを示す。
【００１２】
追加の実施形態では、キャッシュ内のエントリの数がしきい値を超えたかどうかが判別され、エントリは書き込みアドレスに対応する。選択されたエントリに対応するデータが、エラーなしで記憶媒体に書き込まれたかどうかが判別される。ある種の実施形態では、しきい値を超えることはキャッシュが満杯の９０パーセントを上回ることを示す。他の実施形態では、検証は、総合検証用自己テストが実行可能であるという判定に応答するものであり、総合検証用自己テストが実行可能でない場合、永続問題用自己テストが実行される。
【００１３】
ある種の追加の実施形態では、記憶媒体への書き込みはある１つの順序で実行され、記憶媒体への書き込みの検証は別の順序で実行される。
【００１４】
さらに追加の実施形態では、キャッシュに格納されたデータが記憶媒体に書き込まれたデータと同じでない場合、断続的書き込みエラーが検証される。
【００１５】
次に図面を参照するが、全体を通じて同じ参照番号は対応する部分を表す。
【発明を実施するための最良の形態】
【００１６】
以下の説明では、本明細書の一部を形成し、いくつかの実施形態を示す、添付の図面を参照する。他の実施形態が使用可能であり、構造および動作の変更が可能であることを理解されよう。
【００１７】
ある種の実施形態では、ディスク・ドライブが全ブロック書き込みを断続的または永続的に喪失した場合、および書き込みエラーの報告に障害が生じた場合に、ディスク・ドライブからデータを回復することができる。
【００１８】
図１は、ある種の実施形態に従ったコンピューティング環境１００を示すブロック図である。ディスク・ドライブ１０２などの記憶装置が、ホスト・コンピュータ・デバイスなどのコンピュータ・デバイス１０４に結合される。ディスク・ドライブ１０２は、直接、またはストレージ・エリア・ネットワーク（ＳＡＮ）、ローカル・エリア・ネットワーク（ＬＡＮ）、イントラネット、インターネットなどの、ネットワークを介して、コンピュ

ータ・デバイスに結合することができる。

【００１９】

コンピュータ・デバイス１０４は、クライアント、ストレージ・サーバ、サーバ、パーソナル・コンピュータ、ワークステーション、メインフレーム、ミッドレンジ・コンピュータ、ネットワーク・アプライアンス、パーム・トップ・コンピュータ、テレフォニ・デバイス、ブレード・コンピュータ、ハンド・ヘルド・コンピュータなどの、当分野で現在知られているものを含む、任意の好適なコンピュータ・デバイスとすることができる。図１は、ディスク・ドライブ１０２を記憶装置として示しているが、代替の実施形態では、記憶装置は当分野で知られた任意の他の好適なデバイスとすることができる。例示的なディスク・ドライブ１０２は、ある種の実施形態では、新磁気ディスク制御機構（ＲＡＩＤ：Redundant Array of Independent Disk）アレイに含めるか、または単純ディスク束（ＪＢＯＤ：Just a Bunch of Disks）に含めることができる。

【００２０】

ディスク・ドライブ１０２は、未検証の書き込みデータ１０６ａを格納するキャッシュ１０６と、キャッシュ１０６に格納されたエントリに対応する非重複論理ブロック・アドレスおよび関連するカウントの順序付けリスト１１０を維持するリスト維持管理（maintainer）アプリケーション１０８と、所定の条件１１４のセットのうちの１つまたは複数が満たされた場合に実行される総合検証用自己テスト（ＣＶＳＴ）アプリケーション１１２と、ＰＰＳＴアプリケーション１１６とを含むことができる。ディスク・ドライブ１０２は、データを書き込むことが可能な１つまたは複数のディスク１１８を備えることができる。

【００２１】

キャッシュ１０６は、任意の好適な不揮発性メモリを含むことができる。ある種の実施形態では、キャッシュへの書き込みはディスク１１８への書き込みよりもかなり高い信頼度で実行される。キャッシュ１０６に格納された未検証書き込みデータ１０６ａは、たとえディスク・ドライブ１０２がリセットされた場合であっても保持される。

【００２２】

リスト維持管理アプリケーション１０８、ＣＶＳＴアプリケーション１１６、およびＰＰＳＴアプリケーション１１６は、ソフトウェア、ファームウェア、ハードウェア、またはそれらの任意の組み合わせで実施可能である。順序付けリスト１１０および所定の条件１１４は、任意の好適なデータ構造に格納することができる。

【００２３】

ある種の実施形態では、ディスク・ドライブ１０２でコンピュータ・デバイス１０４から書き込み要求が受け取られた場合、書き込み要求に対応するデータはキャッシュ１０６に格納される。ディスク１１８へのデータの書き込み時に何らかのエラーが生じた場合、ＣＶＳＴアプリケーション１１２あるいはＰＰＳＴアプリケーション１１６またはその両方は、ディスク１１８に書き込まれたデータとキャッシュ１０６に格納されたデータとを比較することによって、エラーを判別することができる。

【００２４】

図２は、ある種の実施形態に従った、コンピューティング環境１００に含まれるコンポーネントおよびデータ構造を示すブロック図である。

【００２５】

リスト維持管理アプリケーション１０８は順序付けリスト１１０を維持し、順序付けリスト１１０は非重複論理ブロック・アドレス（ＬＢＡ）および関連するカウントを備えたエントリを含む。たとえば、順序付けリスト１１０の例示的エントリは、カウント＝４で論理ブロック・アドレス＝３への書き込みに対応することができる。これは、データが４つの論理ブロック・アドレス３、４、５、および６に書き込まれることを暗黙に示す。論理ブロック・アドレス３、４、５、および６はすでに例示的エントリに含まれているため、順序付けリスト１１０には論理ブロック・アドレス３、４、５、または６を含む他のエントリはない。リスト維持管理アプリケーション１０８は、ディスク・ドライブ１０２か

らの読み取り性能を向上させるために、リスト１１０も順序付けする。

【００２６】

ある種の所定の条件が満たされた場合、ＣＶＳＴアプリケーション１１２は順序付けリスト１１０のサブセット２００を選択し、対応する読み取りコマンド２０２を発行して、ディスク１１８からデータを読み取ることができる。ある種の実施形態では、選択されたサブセット２００は、順序付けリスト１１０内のエントリの上位５％を含むことができる。たとえば、サブセット２００内の選択されたエントリが論理ブロック・アドレス＝３およびカウント＝４に対応する場合、ＣＶＳＴアプリケーション１１２は、ディスク１１８から論理ブロック・アドレス３、４、５、６を読み取るために読み取りコマンド２０２を発行することができる。

【００２７】

ディスク１１８内に格納されたデータ２０４とキャッシュ１０６に格納されたデータとが比較され（参照数２０６）、データ２０４は実行された読み取りコマンド２０２の結果である。比較の結果を示すステータス２０８を、リスト維持管理アプリケーション１０８に送信するために生成することができる。たとえばステータス２０８が、ディスク１１８内のデータ２０４がキャッシュ１０８内のデータと一致することを示す場合、リスト維持管理アプリケーション１０８は順序付けリスト１１０内の対応するエントリを削除することができる。というのは、この一致は、削除されたエントリに対応する未検証書き込みデータ１０６ａがディスク１１８に正しく書き込まれることが検証されたことを示すからである。

【００２８】

したがって図２は、リスト維持管理アプリケーション１０８およびＣＶＳＴアプリケーション１１２が、ディスク１１８に書き込まれたデータとキャッシュ１０６に書き込まれたデータとを比較することによってディスク１１８に書き込まれたデータを検証する、ある種の実施形態を示す。

【００２９】

図３は、ある種の実施形態に従った、ディスク・ドライブ１０２で実施される喪失した書き込みから回復するための動作を示す図である。

【００３０】

制御はブロック３００で開始され、ディスク・ドライブ１０２のディスク１１８にデータを書き込むための要求が受け取られる。ディスク・ドライブ１０２は（ブロック３０２で）書き込み要求に対応するデータをキャッシュ１０６に格納し、データはディスク１１８に書き込まれる。制御は、ブロック３０２からブロック３０４および３０８へ並行して進む。

【００３１】

ディスク・ドライブ１０２は（ブロック３０４で）、キャッシュ１０６がかなりの数のエントリを有するかどうかを判別する。ある種の実施形態では、キャッシュ１０６は、キャッシュが満杯の１０パーセントを上回る場合に、かなりの数のエントリを有するものとみなすことができる。これを上回る場合ディスク・ドライブ１０２は（ブロック３０６で）、リスト維持管理アプリケーション１０８を実行してキャッシュ・エントリを合体する。リスト維持管理アプリケーション１０８は、非重複論理ブロック・アドレスおよび関連するカウントの順序付けリスト１１０を生成することができる。その後、制御はブロック３００に戻る。キャッシュ１０６がかなりの数のエントリを有していない場合、ディスク・ドライブ１０２は制御をブロック３００に戻し、ここで書き込み要求が受け取られる。

【００３２】

ディスク・ドライブ１０２は（ブロック３０８で）、キャッシュ１０６がほぼ満杯であるかどうかを判別する。たとえばある種の実施形態では、キャッシュが満杯の９０％を超えている場合、キャッシュ１０６はほぼ満杯であるものとみなされる。ディスク・ドライブ１０２は（ブロック３１０）で、ＣＶＳＴ処理が実行可能であるかどうかを判別する。実行可能である場合、ディスク・ドライブ１０２は（ブロック３１２で）、順序付けリス

ト１１０内の選択されたエントリ上でＣＶＳＴアプリケーション１１２を実行する。ある種の実施形態では、選択されたエントリは、順序付けリスト１１０内で最高順位を有するエントリ、たとえば順序付けリスト１１０内で上位５％のエントリとすることができる。新しい書き込み要求が受け取られると、制御はブロック３００に戻る。ディスク・ドライブ１０２が（ブロック３０８で）、キャッシュ１０６がほぼ満杯ではないと判別した場合も、制御はブロック３００に戻る。

【００３３】

ディスク・ドライブ１０２が（ブロック３１０で）、ＣＶＳＴ処理が実行不能であると判別した場合、ディスク・ドライブ１０２は（ブロック３１４で）即時にＰＰＳＴアプリケーション１１６を実行し、制御はブロック３００に戻る。ＰＰＳＴアプリケーション１１６は、永続的に喪失した書き込み問題を検出することができる。ある種の実施形態では、書き込み問題が永続的である場合にのみ、喪失した書き込みの検出がある種の環境で適切であることから、ＣＶＳＴ処理は実行不能である。ＣＶＳＴ処理が実行不能なこうした実施形態では、ＣＶＳＴ処理のある種のオーバヘッドを招く可能性がある。

【００３４】

ある種の実施形態では、キャッシュがほぼ満杯であるかどうかの判別（ブロック３０８）あるいはキャッシュがかなりの数のエントリを有するかどうかの判別（ブロック３０４）、またはその両方は、ある種の実施形態でキャッシュの満杯の程度またはキャッシュ内のエントリ数を示すことが可能な、所定のしきい値との比較に基づいて実施することができる。

【００３５】

したがって図３は、キャッシュがほぼ満杯の場合にＣＶＳＴアプリケーション１１２が実行される、ある種の実施形態を示す。ＣＶＳＴアプリケーション１１２は、ディスク１１８への書き込みの検証のために順序付けリスト１１０からエントリを選択し、順序付けリスト１１０は、キャッシュ１０６内のエントリ数が事前に定義されたしきい値を超えた場合に、リスト維持管理アプリケーション１０８によってアクティブに管理可能である。

【００３６】

図４は、ある種の実施形態に従った、リスト維持管理アプリケーション１０８で実施されるある種の動作を示す図である。

【００３７】

制御はブロック４００で開始され、リスト維持管理アプリケーション１０８の実行が開始される。制御は、ブロック４００からブロック４０２および４０６へ並行して進む。

【００３８】

リスト維持管理アプリケーション１０８は（ブロック４０２で）、連続するかまたは重複するキャッシュに入れられた書き込みの論理ブロック・アドレスおよびカウントを含む、書き込みアドレスを合体する。その結果、リスト維持管理アプリケーション１０８は、非重複論理ブロック・アドレスおよび関連するカウントのリスト１１０を作成する。たとえば、カウント＝４でＬＢＡ＝３への書き込み、およびその後のカウント＝８でＬＢＡ＝５への書き込みは、データ・ブロックのうちの２つが重複することから、カウント＝１０でＬＢＡ＝３の単一のアドレスに合体することができる。これにより、論理ブロック・アドレスおよび関連するカウントを含む、連続するかまたは重複する２つまたはそれ以上の書き込みの検証が、単一の読み取りに効果的に削減される。代替の実施形態では、２つまたはそれ以上のほぼ連続する書き込みも合体することができる。ブロック４０２の結果、重複なしの論理ブロック・アドレスおよびカウントの削減されたリストとなる。

【００３９】

リスト維持管理アプリケーション１０８は（ブロック４０４で）、その後のディスク・ドライブ１０２からの読み取り性能を向上させるために順序付けリスト１１０を再度順序付けする。たとえばある種の実施形態では、順序付けリスト１１０は、同じランクにある関連するカウント・サイズおよび論理ブロック・アドレス数に基づいて論理ブロック・アドレスをランク付けすることができる。

【００４０】
ブロック４０６でリスト維持管理アプリケーション１０８は、ＣＶＳＴアプリケーションからステータス２０８の通知を受け取る。受け取ったステータス２０８の通知に基づき、リスト維持管理アプリケーション１０８は（ブロック４０８で）、新しいホスト書き込みの論理ブロック・アドレスおよびカウントを順序付けリスト１１０に追加することができる。リスト維持管理アプリケーション１０８は、キャッシュ内に格納された論理ブロック・アドレスに対応するデータが、ディスク１１８に書き込まれたデータと首尾よく一致した、ＣＶＳＴアプリケーション１１２によって実行された論理ブロック・アドレスのリストを削除することもできる。
【００４１】
したがって図４は、ディスク・ドライブ１０２からのその後の読み取り性能を向上させるために、リスト維持管理アプリケーション１０８が非重複論理ブロック・アドレスおよび関連するカウントの順序付けリストを維持する、ある種の実施形態を示す。
【００４２】
図５は、ある種の実施形態に従った、ＣＶＳＴアプリケーション１１２で実施される動作を示す図である。
【００４３】
制御はブロック５００で開始され、ＣＶＳＴアプリケーション１１２が開始される。ＣＶＳＴアプリケーション１１２は（ブロック５０２で）、所定の条件１１４が満たされたかどうかを判別する。所定の条件は、ＣＶＳＴアプリケーション１１２によって順序付けリスト１１０から選択されるエントリのサイズを決定することができる。
【００４４】
ＣＶＳＴアプリケーション１１２が、所定の条件１１４が満たされていると判定した場合、ＣＶＳＴアプリケーション１１２は（ブロック５０４で）、リスト維持管理アプリケーション１０８によって提供された順序付けリスト１１０のサブセット２００を読み取りコマンド２０２に変換する。そうでない場合、ＣＶＳＴアプリケーション１１２は（ブロック５０２で）、所定の条件が満たされたかどうかの判別を続ける。
【００４５】
ＣＶＳＴアプリケーション１１２は（ブロック５０６で）、実行されたディスク１１８からの読み取りと、キャッシュ１０６内の未検証書き込みデータ１０６ａを示す対応するキャッシュ・エントリとを比較する。
【００４６】
ＣＶＳＴアプリケーション１１２はブロック５０８で、キャッシュ１０８内のデータがディスク１１８に書き込まれたデータと一致するかどうかを判別する。一致する場合、ＣＶＳＴアプリケーション１１２は（ブロック５１０で）、キャッシュ１０６からの対応するデータを削除し、（ブロック５１２で）ステータス２０８の通知をリスト維持管理アプリケーション１０８に送信し、リスト維持管理アプリケーション１０８は順序付けリスト１１０の修正および再順序付けを管理する。
【００４７】
ＣＶＳＴアプリケーション１１２がブロック５０８で、実行された読み取りについてキャッシュ１０８内のデータがディスク１１８に書き込まれたデータと一致しないと判定した場合、ＣＶＳＴアプリケーション１１２は（ブロック５１４で）、ディスク・ドライブ１０２を書き込み保護モードに設定し、エラーを報告する。ディスク・ドライブ１０２が書き込み保護モードの場合、ディスク１１８にデータを書き込むことはできない。ディスク１１８への書き込みエラーは、ディスク・ドライブ１０２が書き込み保護モードにされた後に判定することができる。正しいデータは、キャッシュ１０６に格納された未検証書き込みデータ１０６ａから回復することができる。
【００４８】
ブロック５０４でＣＶＳＴアプリケーション１１２によって実行される読み取りコマンドの数は、複数の所定の条件１１４によって決定することができる。たとえば第１の例示

的条件において、ホストのアイドル時間が２秒の場合、ＣＶＳＴアプリケーション１１２は順序付けリスト１１０の２０％を実行することができる。第２の例示的条件では、順序付けリスト１１０から選択されるエントリの数を、ホストの作業負荷に基づくものとすることができる。第３の例示的条件では、順序付けリスト１１０から選択されるエントリの数を、使用されるキャッシュ１０６の量に基づくものとすることができる。たとえば、キャッシュが９０％満杯である場合、ＣＶＳＴアプリケーション１１２は順序付けリスト１１０の５％を選択することができる。このパーセントの数字を調整して、ディスク１１８の読み取りおよび書き込み性能を最適化することができる。

【００４９】

したがって図５は、ＣＶＳＴアプリケーション１１２が所定の条件１１４を使用して、ディスク１１８への書き込み中に断続的書き込みエラーが生じたかどうかを検証する、ある種の実施形態を示す。

【００５０】

図６は、ある種の実施形態に従った、ディスク・ドライブ１０２で実施される読み取り要求を処理するための動作を示す図である。

【００５１】

制御はブロック６００で開始され、ディスク・ドライブ１０２がコンピュータ・デバイス１０４から読み取り要求を受け取る。ディスク・ドライブ１０２は（ブロック６０２で）、読み取り要求に対応するデータがキャッシュ１０６内に存在する未検証書き込みデータ１０６ａであるかどうかを判別する。存在する未検証書き込みデータ１０６ａである場合、（ブロック６０４で）キャッシュ１０６からのデータを戻すことによって、読み取り要求が満たされる。存在する未検証書き込みデータ１０６ａでない場合、（ブロック６０６で）ディスク１１８からのデータを戻すことによって、読み取り要求が満たされる。

【００５２】

したがって図６は、読み取り要求に対応するデータが未検証書き込みデータ１０６ａである場合、外部ホストからの読み取り要求がキャッシュ１０６から満たされる、ある種の実施形態を示す。結果として、読み取り要求に応答して戻されるデータはエラーではない。

【００５３】

ある種の実施形態は、リスト維持管理アプリケーション１０８およびＣＶＳＴアプリケーション１１２を提供し、ＣＶＳＴアプリケーション１１２は、リスト維持管理アプリケーション１０８によってＣＶＳＴアプリケーション１１２に提供された順序付けリスト１１０を使用することによって、キャッシュ１０６内の書き込みデータの一部とディスク１１８から読み取られたデータとを比較する。ある種の実施形態では、キャッシュ１０６に格納される書き込み数を削減することによって性能を向上させる。さらにある種の実施形態は、書き込みキャッシュ１０６がほぼ満杯になった際にＰＰＳＴを実施するＰＰＳＴアプリケーションがＣＶＳＴアプリケーションに置き換えられた場合、断続的および永続的の両方の喪失した書き込みに対する保護も提供する。

【００５４】

ある種の実施形態は、検出段階間の書き込みキャッシュ１０６のコンテンツを削減するために、リスト維持管理アプリケーション１０８およびＣＶＳＴアプリケーション１１２を使用しながら、喪失した書き込みのチェックとしてＰＰＳＴアプリケーション１１６を使用することにより、永続的に喪失した書き込みに対する完全な保護を提供することができる。このケースでは、書き込みキャッシュ１０６がほぼ満杯になり、エラーが検出されなければ書き込みキャッシュに入れられたエントリがフラッシュされる場合、ＰＰＳＴアプリケーション１１６を呼び出すことができる。

【００５５】

ある種の実施形態では、データの回復は背景モードまたはリアルタイムで実行することができる。ある種の実施形態は、インターフェースから独立した任意のディスク・ドライブに関する任意の種類の喪失した書き込み問題に対する保護を提供する。喪失した書き込

みは断続的または永続的とすることが可能であり、ホスト・アプリケーションへの中断（disruption）の量を削減することが可能である。ディスク・ドライブでの喪失した書き込みを保護するある種の実施形態は、入力／出力（Ｉ／Ｏ）性能に大幅な影響を与えないことが可能である。

【００５６】

リスト維持管理アプリケーション１０８を含めることにより、キャッシュ１０６がほぼ満杯の状態に達するまでの期間を引き延ばし、ＰＰＳＴまたはＣＶＳＴ検証のオーバヘッドによって生じる可能性のある性能の影響を遅らせる。ある種の実施形態では、ＣＶＳＴおよびＰＰＳＴは長いアイドル・サイクル中にキャッシュ１０６全体をフラッシュすることができる。ホストの作業負荷が少ない場合、削減の速度がホストからの書き込み追加速度よりも速い場合があり、その結果としてキャッシュ全体のフラッシュも可能である。したがってこうした状況では、キャッシュ１０６は決して満杯に近づくことはなく、後続のＰＰＳＴまたはＣＶＳＴを実行する必要のない可能性がある。

【００５７】

ある種の実施形態では、論理ブロック・アドレスおよび関連するカウンタを単純化すること、順序付けリスト１１０内のエントリを再順序付けすること、およびディスク１１８に読み取りのキューを送信することにより、ディスク１１８へのあらゆる書き込みが検証される場合の検証のプロセスがより効率良くなる。ある種の実施形態では、重複する書き込みを単一の読み取りに組み合わせ、再順序付けされた読み取りをキューとして送信し、所与の時間についての読み取り数を最大限にするために異なるシリンダへのシークを最適化することにより、ディスク・ドライブの読み取り性能をさらに向上させることが可能である。

【００５８】

追加の詳細な実施形態

説明される技法は、ソフトウェア、ファームウェア、マイクロコード、ハードウェア、あるいはそれらの任意の組み合わせ、またはそれらすべてを含む、方法、装置、または装置として実施することができる。本明細書で使用される「装置」という用語は、回路内（たとえば集積回路チップ、プログラマブル・ゲート・アレイ（ＰＧＡ）、ＡＳＩＣなど）、あるいは、コンピュータ読み取り可能メディア（たとえばハード・ディスク・ドライブ、フロッピィ・ディスク、テープなどの磁気記憶媒体）、光ストレージ（たとえばＣＤ－ＲＯＭ、ＤＶＤ－ＲＯＭ、光ディスクなど）、揮発性および不揮発性メモリ・デバイス（たとえば電気的消去可能プログラマブル読み取り専用メモリ（ＥＥＰＲＯＭ）、読み取り専用メモリ（ＲＯＭ）、プログラマブル読み取り専用メモリ（ＰＲＯＭ）、ランダム・アクセス・メモリ（ＲＡＭ）、動的ランダム・アクセス・メモリ（ＤＲＡＭ）、静的ランダム・アクセス・メモリ（ＳＲＡＭ）、フラッシュ、ファームウェア、プログラマブル論理など）内、またはそれらすべてで実施される、プログラム命令、コード、あるいは論理またはそれらすべてを言い表す。コンピュータ読み取り可能メディア内のコードには、プロセッサなどのマシンによってアクセスし、実行することができる。ある種の実施形態では、実施形態が作成されたコードに、伝送メディアを介して、またはネットワークを介してファイル・サーバからもアクセスすることができる。こうしたケースでは、コードが実施される装置は、ネットワーク伝送回線、無線伝送メディア、空気、電波、赤外線信号を介して伝搬される信号などの、伝送メディアを含むことができる。もちろん当業者であれば、諸実施形態の範囲を逸脱することなく多くの修正が実行可能であること、および、装置が当分野で知られた任意の情報搬送メディアを含むことが可能であることを理解されよう。たとえば装置は、マシンによって実行されると結果として実行される動作を生じる命令を格納した、記憶媒体を含む。

【００５９】

図７は、ある種の実施形態が実施可能なシステム７００を示すブロック図である。ある種の実施形態では、コンピュータ・デバイス１０４および記憶装置１０２をシステム７００に従って実施することができる。システム７００は、ある種の実施形態ではプロセッサ

７０４を含むことができる回路７０２を含むことができる。システム７００は、メモリ７０６（たとえば揮発性メモリ・デバイス）およびストレージ７０８を含むこともできる。システム７００のある種の要素は、コンピュータ・デバイス１０４および記憶装置１０２のうちの一部またはすべてで見られるか、または見られない場合がある。ストレージ７０８は、不揮発性メモリ・デバイス（たとえばＥＥＰＲＯＭ、ＲＯＭ、ＰＲＯＭ、ＲＡＭ、ＤＲＡＭ、ＳＲＡＭ、フラッシュ、ファームウェア、プログラマブル論理など）、磁気ディスク・ドライブ、光ディスク・ドライブ、テープ・ドライブなどを含むことができる。ストレージ７０８は、内部記憶装置、取り付け型記憶装置、あるいはネットワーク・アクセス可能記憶装置、またはそれらすべてを含むことができる。システム７００は、メモリ７０６にロード可能であり、プロセッサ７０４または回路７０２によって実行可能である、コード７１２を含むプログラム論理７１０を含むことができる。ある種の実施形態では、コード７１２を含むプログラム論理７１０をストレージ７０８に格納することができる。ある種の他の実施形態では、プログラム論理７１０を回路７０２内で実施することができる。したがって、図７ではプログラム論理７１０を他の要素とは別に示しているが、プログラム論理７１０はメモリ７０６あるいは回路７０２またはその両方で実施することができる。

【００６０】

ある種の実施形態は、人間またはコンピュータ読み取り可能コードをコンピュータ・システムに統合する自動化された処理によって、コンピュータ命令を展開するための方法を対象とすることが可能であり、コードとコンピュータ・システムと組み合わせることによって、説明された実施形態の動作を実行することができる。

【００６１】

図３～６に示された動作の少なくとも一部は、並行して、ならびに逐次、実行することができる。代替の実施形態では、動作の一部は異なる順序で実行する、修正する、または削除することができる。

【００６２】

さらに、ソフトウェアおよびハードウェア・コンポーネントの多くは、例示の目的で別々のモジュールにあるものとして説明してきた。こうしたコンポーネントを、より少ない数のコンポーネントに統合するか、またはより多くの数のコンポーネントに分割することが可能である。さらに、特定のコンポーネントによって実行されるものとして説明したある種の動作を、他のコンポーネントによって実行することも可能である。

【００６３】

図１～７で示されたかまたは参照されたデータ構造およびコンポーネントは、特定のタイプの情報を有するものとして説明される。代替の実施形態では、データ構造およびコンポーネントは異なる構造とすること、ならびにこれらの図面で示されたかまたは参照されたものよりも少ない、多い、または異なるフィールド、あるいは異なる機能を有することが可能である。したがって、諸実施形態の前述の説明は例示および説明の目的で提示してきた。諸実施形態を網羅するか、または開示された精密な形に限定することを意図するものではない。上記の教示に鑑みて、多くの修正および変形が可能である。

【図面の簡単な説明】

【００６４】

【図１】ある種の実施形態に従ったコンピューティング環境を示すブロック図である。

【図２】ある種の実施形態に従った、コンピューティング環境に含まれるコンポーネントおよびデータ構造を示すブロック図である。

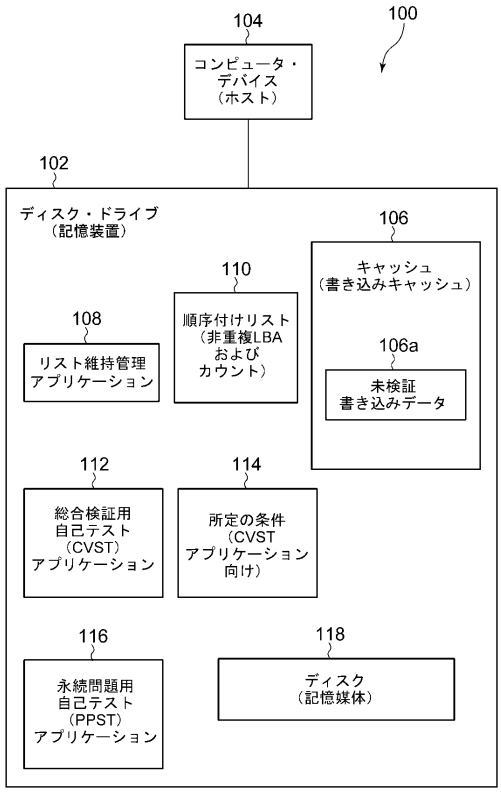
【図３】ある種の実施形態に従った、喪失した書き込みから回復するための動作を示す図である。

【図４】ある種の実施形態に従った、リスト維持管理アプリケーションで実施される動作を示す図である。

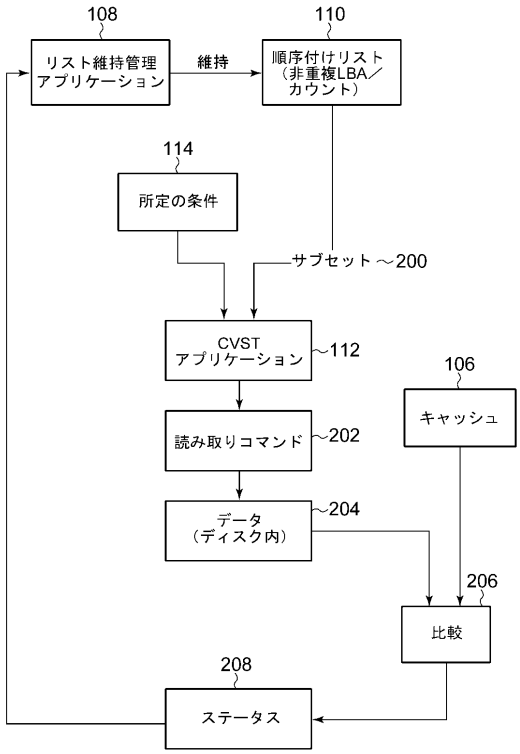
【図５】ある種の実施形態に従った、総合検証用自己テスト・アプリケーションで実施される動作を示す図である。

【図６】ある種の実施形態に従った、読み取り要求を処理するための動作を示す図である。
【図７】ある種の実施形態が実施されるシステムを示す図である。
【符号の説明】
【００６５】
１００　コンピューティング環境
１０２　ディスク・ドライブ（記憶装置）
１０４　コンピュータ・デバイス（ホスト）
１０６　キャッシュ（書き込みキャッシュ）
１０６ａ　未検証書き込みデータ
１０８　リスト維持管理アプリケーション
１１０　順序付けリスト（非重複ＬＢＡおよびカウント）
１１２　総合検証用自己テスト（ＣＶＳＴ）アプリケーション
１１４　所定の条件（ＣＶＳＴアプリケーション向け）
１１６　永続問題用自己テスト（ＰＰＳＴ）アプリケーション
１１８　ディスク（記憶媒体）

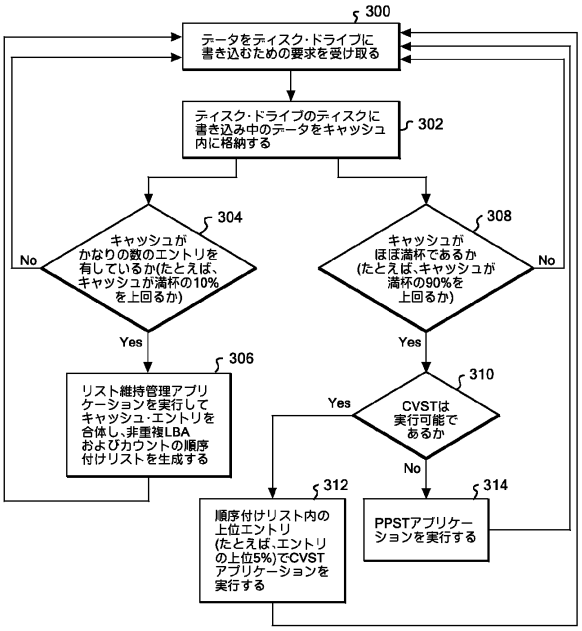
【図１】



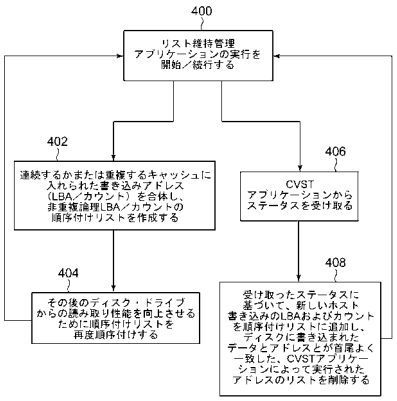
【図２】



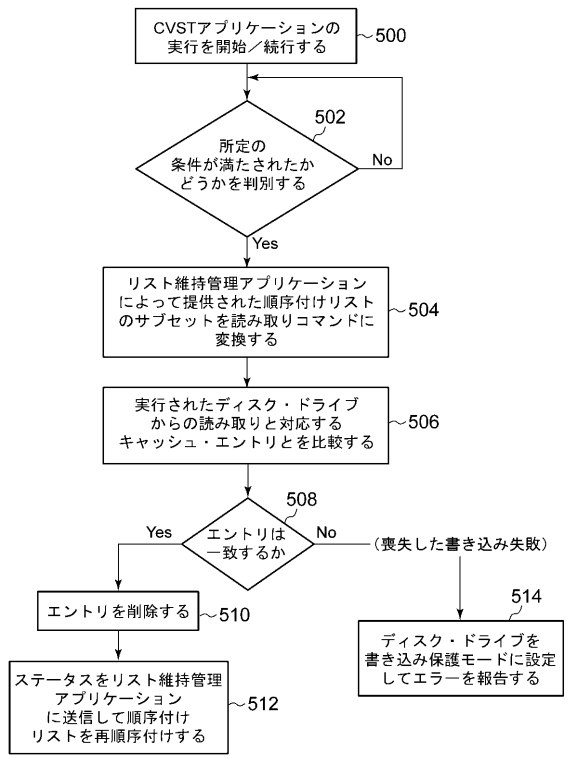
【図３】


300
データをディスク･ドライブに書き込むための要求を受け取る
ディスク･ドライブのディスクに書き込み中のデータをキャッシュ内に格納する
302
304
キャッシュがかなりの数のエントリを有しているか(たとえば、キャッシュが満杯の10%を上回るか)
No
Yes
306
リスト維持管理アプリケーションを実行してキャッシュ･エントリを合体し、非重複LBAおよびカウントの順序付けリストを生成する
308
キャッシュがほぼ満杯であるか(たとえば、キャッシュが満杯の90%を上回るか)
No
Yes
310
CVSTは実行可能であるか
Yes
No
312
順序付けリスト内の上位エントリ(たとえば、エントリの上位5%)でCVSTアプリケーションを実行する
314
PPSTアプリケーションを実行する

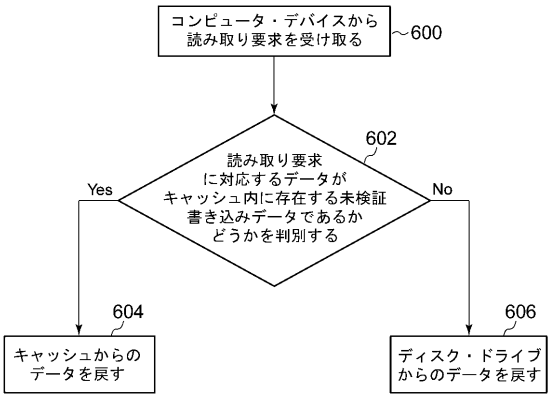
【図４】


400
リスト維持管理アプリケーションの実行を開始／続行する
402
連続するかまたは重複するキャッシュに入れられた書き込みアドレス(LBA／カウント)を合体し、非重複論理LBA／カウントの順序付けリストを作成する
404
その後のディスク・ドライブからの読み取り性能を向上させるために順序付けリストを再度順序付けする
406
CVSTアプリケーションからステータスを受け取る
408
受け取ったステータスに基づいて、新しいホスト書き込みのLBAおよびカウントを順序付けリストに追加し、ディスクに書き込まれたデータとアドレスとが首尾よく一致した、CVSTアプリケーションによって実行されたアドレスのリストを削除する

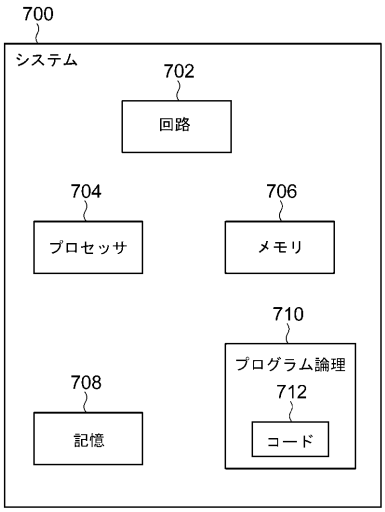
【図５】


CVSTアプリケーションの実行を開始／続行する
500
502
所定の条件が満たされたかどうかを判別する
No
Yes
リスト維持管理アプリケーションによって提供された順序付けリストのサブセットを読み取りコマンドに変換する
504
実行されたディスク・ドライブからの読み取りと対応するキャッシュ・エントリとを比較する
506
508
エントリは一致するか
Yes
No
(喪失した書き込み失敗)
エントリを削除する
510
514
ディスク・ドライブを書き込み保護モードに設定してエラーを報告する
ステータスをリスト維持管理アプリケーションに送信して順序付けリストを再順序付けする
512

【図６】


コンピュータ・デバイスから読み取り要求を受け取る
600
602
読み取り要求に対応するデータがキャッシュ内に存在する未検証書き込みデータであるかどうかを判別する
Yes
No
604
キャッシュからのデータを戻す
606
ディスク・ドライブからのデータを戻す

【図７】



フロントページの続き

(72)発明者　ウィリアム・ジョン・デュリカ
アメリカ合衆国９５０３７　カリフォルニア州モーガン・ヒル　ライヴ・オーク・レーン３８５５
(72)発明者　エム・アミン・ハッジ
アメリカ合衆国９５１２０　カリフォルニア州サンノゼ　クエイル・クリーク・クラーク１２１１
(72)発明者　ジョゼフ・スミス・ハイドＩＩ世
アメリカ合衆国８５７１２　アリゾナ州トゥーソン　イースト・プラシータ・デル・メスキート５３７０
(72)発明者　ロナルド・ジェイ・ヴェンチューリ
イギリス国エス・オー２２　６キュー・エル　ウィンチェスター州リトルトン　ザ・ホール・ウェイ２３　ニューアーク

審査官　菅原　浩二

(56)参考文献　特開平０４－０１７０２０（ＪＰ，Ａ）
特開２００１－１４２６５０（ＪＰ，Ａ）
特開平０７－２１０３２７（ＪＰ，Ａ）
特開２００３－２６３７０３（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　　３／０６
Ｇ０６Ｆ　１２／０８