

(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2005-310159

(P2005-310159A)

(43) 公開日 平成17年11月4日(2005.11.4)

(51) Int. Cl. ⁷	F I	テーマコード (参考)
G06F 12/00	G06F 12/00 531M	5B065
G06F 3/06	G06F 3/06 304F	5B082

審査請求 有 請求項の数 10 O L (全 15 頁)

(21) 出願番号	特願2005-122283 (P2005-122283)	(71) 出願人	503003854 ヒューレット・パカード デベロップメント カンパニー エル. ピー. アメリカ合衆国 テキサス州 77070 ヒューストン 20555 ステイト ハイウェイ 249
(22) 出願日	平成17年4月20日 (2005. 4. 20)	(74) 代理人	110000039 特許業務法人アイ・ピー・エス
(31) 優先権主張番号	10/829, 715	(72) 発明者	ブライアン・パターソン アメリカ合衆国カリフォルニア州 パロアルト ハノーバー・ストリート 3000 ヒューレット・パカード・カンパニー 内
(32) 優先日	平成16年4月21日 (2004. 4. 21)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 スナップショットに分散されたデータを使用するデータ保護

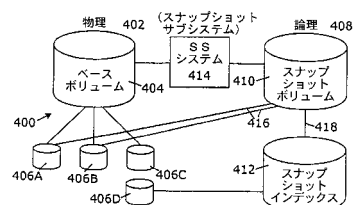
(57) 【要約】

【課題】 スナップショットに分散されたデータを使用するデータ保護を提供する。

【解決手段】 ベースボリューム(404)および少なくとも1つの物理ブロック(406A、406B、406C、および406D)を備える物理ストア(402)と、

スナップショットボリューム(410)およびスナップショットインデックス(412)を備える論理ストア(408)と、前記スナップショットボリューム(410)から前記物理ブロック(406A、406B、406C、および406D)のうちの選択されたものへのポインタをポイントインタイムでサポートでき、親論理ユニット(LUN)を定義でき、かつ、該親LUNのデータが複数のスナップショットから分岐すると、前記親LUNの前記複数のスナップショットにわたってデータを分散させることができるスナップショットサブシステム(414)と、を備えるストレージシステム(400)とする。

【選択図】 図4C



【特許請求の範囲】

【請求項 1】

ストレージシステム(400)であって、
ベースボリューム(404)および少なくとも1つの物理ブロック(406A、406B、406C、および406D)を備える物理ストア(402)と、
スナップショットボリューム(410)およびスナップショットインデックス(412)を備える論理ストア(408)と、
前記スナップショットボリューム(410)から前記物理ブロック(406A、406B、406C、および406D)のうちの選択されたものへのポインタをポイントインタイムでサポートでき、親論理ユニット(LUN)を定義でき、かつ、該親LUNのデータが複数のスナップショットから分岐すると、前記親LUNの前記複数のスナップショットにわたってデータを分散させることができるスナップショットサブシステム(414)と、
を備えるストレージシステム。

【請求項 2】

ストレージシステム(400)であって、
前記物理ストア(402)のデータを指し示す、前記スナップショットボリューム(410)の少なくとも1つのマップポイントと、
前記スナップショットインデックス(412)のデータを指し示すことができる複数のスナップショットポイント(418)と、
分岐したデータを、前記スナップショットインデックス(412)の複数のスナップショット部分に分散させ、該分散させたデータ部分を前記複数のスナップショットに書き込む前記スナップショットサブシステム(414)と、
をさらに備える請求項1に記載のストレージシステム。

【請求項 3】

ストレージシステム(400)であって、
前記スナップショットサブシステム(414)は、前記複数のスナップショットにわたってほぼ等しい割合でデータを分散させる、
請求項2に記載のストレージシステム。

【請求項 4】

ストレージシステム(400)であって、
前記スナップショットサブシステム(414)は、前記親LUNの削除、該親LUNへのデータ書き込みオペレーション、および該親LUNの障害を含む該親LUNのデータの分岐状態を検出することができる、
請求項1に記載のストレージシステム。

【請求項 5】

ストレージシステム(400)であって、
前記スナップショットサブシステム(414)は、時間の経過と共にスナップショットの個数を変更することができ、前記複数のスナップショット間におけるほぼ均一なデータの分散を各変更時に開始して行うことができる、
請求項1に記載のストレージシステム。

【請求項 6】

ストレージシステム(400)であって、
磁気ディスク、光ディスク、コンパクトディスク(CD)、CD-R、CD-RW、ディスクレット、テープ、およびテープカートリッジの中から選択された媒体をさらに備える、
請求項1に記載のストレージシステム。

【請求項 7】

ベースボリューム(506)および少なくとも1つの物理ブロック(508)を含む物理ストア(504)と、スナップショットボリューム(512)およびスナップショット

10

20

30

40

50

インデックス(514)を含む論理ストア(510)とを有するストレージシステム502で使用されるコンピュータシステム(500)であって、

親LUN(518)のデータが複数のスナップショットから分岐すると、前記親LUN(518)の前記複数のスナップショットにわたってデータを分散させるスナップショットサブシステム、

を備えるコンピュータシステム。

【請求項8】

コンピュータシステム(500)であって、

前記物理ストア(504)のデータを指し示す前記スナップショットボリューム(512)の少なくとも1つのマップポイントと、前記スナップショットインデックス(514)のデータを指し示すことができる複数のスナップショットポイントとを生成するマッピングロジックと、

分岐したデータを前記スナップショットインデックスの複数のスナップショット部分に分散させ、該分散させたデータ部分を前記複数のスナップショットに書き込むスナップショットロジックと、

をさらに備える、請求項7に記載のコンピュータシステム。

【請求項9】

コンピュータシステム(500)であって、

前記親LUNの削除、該親LUNへのデータ書き込みオペレーション、および該親LUNの障害を含む該親LUNのデータの分岐状態を検出する、前記スナップショットロジックに関連付けられたロジックと、

時間の経過と共にスナップショットの個数を変更し、前記複数のスナップショット間におけるほぼ均一なデータの分散を各変更時に開始して行う、前記スナップショットロジックに関連付けられたロジックと、

をさらに備える、請求項7に記載のコンピュータシステム。

【請求項10】

コンピュータシステム(500)であって、

磁気ディスク、光ディスク、コンパクトディスク(CD)、CD-R、CD-RW、ディスクレット、テープ、およびテープカートリッジの中から選択された媒体にデータを記憶することができる少なくとも1つのストレージデバイス、

をさらに備える、請求項7に記載のコンピュータシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、スナップショットに分散されたデータを使用するデータ保護に関する。

【背景技術】

【0002】

データストレージアドミニストレータは、長い間、システムバックアップを使用して、貴重なデータの保護を保証してきた。バックアップは、従来、他のアプリケーションのシャットダウン中に行われ、夜または休み時間に実行されることが多いプロセスである。このような従来のバックアップオペレーションでは、連続して利用可能なストレージシステムの非常に望ましい有用性を得ることはできない。

【0003】

他のオペレーションの中断を回避するバックアップオペレーションを容易にするために、スナップショット技法が開発されている。スナップショット画像は、バックアップのソースとして使用することができる。スナップショットは、一般に、アプリケーションを静止させ、ほぼ瞬時に作成されるコピーを、その使用が基本的に遅延を気付かせないように実行することによって取り込まれる。

【0004】

情報を復元する一般的な理由は、ユーザがその後取り消したい不注意の削除またはフ

10

20

30

40

50

イルへの変更等のユーザエラーである。スナップショット技法によって、高速で効率的なデータの位置の特定および修復に利用しやすい状態で、記憶したデータのコピーを保持することが可能になる。

【発明の開示】

【課題を解決するための手段】

【0005】

データストレージシステムの一実施の形態によると、データを保護する方法は、親論理ユニット(LUN)のデータがスナップショットから分岐すると、親LUNの複数のスナップショットにわたってデータを分散させることを含む。

【0006】

オペレーションの構造および方法に関する本発明の実施の形態は、以下の説明および添付図面を参照することによって最も良く理解することができる。

【発明を実施するための最良の形態】

【0007】

性能および信頼性の改善を可能にするようにストレージデバイスを構成することができる。例えば、独立ディスク冗長配列(RAID)の形のストレージデバイスは、2つまたは3つ以上のストレージドライブを組み合わせて使用し、フォールトトレランスおよび性能を達成する。RAIDストレージ配列は、従来型RAIDストレージ配列タイプおよび仮想RAIDストレージ配列タイプを含む2つのタイプに分類することができる。従来型配列は、ホストコンピュータから物理媒体へのアドレス空間の固定マッピングによって定義される。したがって、従来型配列では、ホストコンピュータが特定の1つのデータ要素にアクセスするアドレスが与えられると、RAID配列を構成する実際のストレージドライブにおいて、そのデータの物理位置を特定することができる。仮想配列では、ホストコンピュータがその配列の特定の1つのデータ要素にアクセスするのに使用するアドレスと、RAID配列を構成するストレージドライブのそのデータの実際の物理位置との間に、仮想化とも呼ばれる少なくとも1つのレベルの間接指定が存在する。そのデータ要素の具体的なホストアドレスが既知である仮想配列では、そのデータの実際の物理位置は、配列のドライブのいずれかにすることができる。仮想配列に存在する間接指定レイヤは、ホストベースのアドレスからストレージドライブのデータの物理位置へのマッピングによって生成される。

【0008】

マッピングによって、仮想配列の機能が可能になり、マッピングを使用することによって、記憶されたデータボリュームのスナップショットコピーを行う機能を援助することができる。スナップショットコピーは、選択されたデータセットのポイントインタイムコピー(point-in-time copy)を可能にする情報を保存する。従来型配列では、コピーされるボリュームへの書き込みを一時停止し、そのポイントインタイムにおけるボリュームの状態を保持し、その後、ポイントインタイムコピーを表す新たなアドレス空間にコピーされるボリュームからあらゆるデータ要素をコピーすることによって、スナップショットコピーが得られる。元のボリュームからのあらゆるデータ要素がコピーされた後にのみ、コピーされるボリュームへの書き込みが再開される。例えばコピーされるデータ量といったスナップショットサイズに応じて、プロセスは、許容されない時間量を要する可能性がある。

【0009】

仮想配列では、スナップショットプロセスを非常に簡略化することができる。仮想化の副産物として、アドレスの元のボリュームから関連付けられた物理位置へのマップがすでに存在する。仮想化マッピングを使用すると、ポイントインタイムコピーを表す新たなアドレス空間にそれらマップをコピーすることによって、ポイントインタイムコピーを簡単に作成することができる。一般の実施態様では、元のホストアドレスおよび関連付けられたスナップショットコピーアドレスの双方が、ストレージデバイスの同じ物理位置にマッピングされる。したがって、スナップショットは、マップのコピーに要する時間量が少な

10

20

30

40

50

く完了する。仮想配列によって、或る物理位置から別の物理位置へ全データをコピーする従来型配列のオペレーションよりもはるかに高速に進むスナップショット機能が可能になる。

【0010】

しかしながら、仮想配列スナップショット技法は、その後のデータ変更の問題が未解決のままである。元のデータの変更は、2つの状態下で発生する可能性がある。図1Aおよび図1Bに示す概略ブロック図に図示した第1の状態では、配列のマップの粒度が、スナップショットコピーと元のデータとの間の分岐 (divergence) が配列内に記録される粒度と同じである。粒度は、例えば、1つのマップエントリによって画定される連続した物理エリアのサイズとして定義することができる。したがって、図1Aおよび図1Bに示す状態100では、分岐の粒度は、マップエントリのサイズおよび位置と正確に一致する。

10

【0011】

図1Aは、マップエントリ1 102に新たなデータを書き込む前の状態100を示している。マップエントリ1 102は記憶データ1 104を指し示す。マップエントリ2 106は記憶データ2 108を指し示す。スナップショットエントリ1 110も記憶データ1 104を指し示す。スナップショットエントリ2 112も記憶データ2 108を指し示す。

【0012】

図1Bは、新たなデータがマップエントリ1 102に書き込まれた後の状態120を示している。仮想配列では、スナップショットは、変更された元のデータを新たな位置に書き込むことによって取り込まれる。この新たな位置は、この例示では、記憶データ3 122とラベル付けされている。スナップショットエントリ1 110として示すスナップショットコピーは、記憶データ1 104として示す元の位置の元の物理データを依然として指し示す状態にある。マップエントリ1 102のラベルを有する元のデータは、新たな物理位置122を指し示すように変更される。したがって、スナップショットコピーと元のデータとの分岐は、新たな位置に新たなデータを書き込むことによって得られる。

20

【0013】

この技法は、配列のマップの粒度が分岐の粒度と同じである場合に機能する。図2Aおよび図2Bに示す別の状態では、スナップショット分岐は、分岐の粒度とマップの粒度が異なる場合の状態200に示されている。図2Aは、マップエントリ1 202に新たなデータを書き込む前の状態200を示している。マップエントリ1 202は記憶データ1 204を指し示す。マップエントリ2 206は記憶データ2 208を指し示す。スナップショットエントリ1 210はスナップショットターゲットデータ214を指し示し、スナップショットエントリ2 212はスナップショットターゲットデータ216を指し示す。

30

【0014】

図2Bは、マップエントリ1 202に新たなデータが書き込まれた後の状態220を示している。記憶データ1 204のラベルを有する元のデータを保持する記憶位置が書き込み可能になって、変更記憶データ1 222を形成する前に、元のデータは、記憶データ1 204の元の位置から、元の記憶データ1 224として示す別の物理位置にコピーされて、適切なポイントインタイムコピーが維持される。スナップショットエントリ1 210および2 212は、この時点で、それぞれスナップショットターゲットデータ214およびスナップショットターゲットデータ216を収容する新たな物理位置を指し示す。元のデータである記憶データ1 204だけでなく、スナップショットに関連付けられたすべての元のデータも、新たな物理位置にコピーされ、その結果、スナップショットエントリ1 210およびスナップショットエントリ2 212は、引き続き完全なポイントインタイムコピーを指し示す。図2Bに示す例示では、記憶データ2 208は新たな物理位置226にコピーされる。元のデータを維持するために、物理記憶位置の内容は、元のデータがコピーされるまで変更されない。スナップショットオペレーションの

40

50

1つの難しい点は、1つのデータ要素の変更が、はるかに多くのデータのコピーを引き起こす可能性があり、場合によっては、スナップショットを使用しないシステムと比較して、データ変更により時間を劇的に増加させる可能性があるということである。

【0015】

性能に劇的な影響を有する可能性のあるデータ変更の一例は、元のデータの削除であり、場合によっては、非常に短い期間で多数のデータ要素のコピーを引き起こす可能性がある。この問題は、同じデータの複数のスナップショットが取り込まれている場合に悪化する。従来型システムでは、親論理ユニット(LUN)が削除されると、データは、第1のスナップショットに書き込まれ、それ以外のすべてのスナップショットは、第1のスナップショットに向けられる。その後、第1のスナップショットが削除されると、データはすべて、次のスナップショット、例えば第2のスナップショットに再書き込みされ、続いて、残りのすべてのスナップショットは、第2のスナップショットに向けられる。その後、第2のスナップショットが削除されると、同じオペレーションが再び実行され、すべてのデータは、再び、後続のスナップショットに再書き込みされる。このように、所与のスナップショットのデータのすべてが、特定のスナップショットが削除される場合に応じて多数回コピーされる可能性があり、それによって、非生産的なコピーに時間および資源が浪費される。

10

【0016】

改良されたストレージシステムおよびストレージシステムハンドリング技法のさまざまな実施の形態によると、分岐が発生すると、データのすべてを1つのスナップショットにコピーするのではなく、データの一部が複数のスナップショットにわたって分散される。データの分岐は、例えば、親論理ユニット(LUN)が削除された時に発生する可能性がある。図3Aおよび図3Bを参照して、全体ブロック図が、データを保護する方法の一実施の形態に関連したストレージ構成300を示している。この方法は、親論理ユニット(LUN)306のデータがスナップショットから分岐する時に、親LUNの複数のスナップショット304A、304B、304C、および304Dにわたってデータ302を分散させることを含む。

20

【0017】

親LUNの削除、親LUNへのデータ書き込みオペレーション、および親LUNの障害を含む検出のために、さまざまなタイプの親LUNのデータの分岐状態を選択することができる。親LUNデータの分岐の検出に続いて、分岐したデータを複数の部分に分散させることができ、分散したデータ部分を複数のスナップショットに書き込むことができる。いくつかの実施の形態では、データを、ほぼ等しい割合で複数のスナップショットにわたって分散させることができる。

30

【0018】

スナップショットの個数は、時間の経過と共に変化し得る。スナップショットの個数の変更時に、データを複数のスナップショットにわたってほぼ均一に分散させることができる。データは、スナップショットにわたってほぼ等しく、または、ほぼ均一に分散され、この分散は、数学的に正確なデータ割り当て条件に対して行われるだけでなく、正確な精度が可能でないか、望ましい状態を含む条件に対しても行われる。例えば、データは、特定のデータ粒度で、特定の個数のスナップショットに等しく分割できない場合がある。したがって、データは、概ね等しく分散したデータで、割り当てられたスナップショットに分散される場合がある。

40

【0019】

スナップショットデータは、任意の適切な媒体に記憶することができる。この媒体には、例えば、磁気ディスク、光ディスク、コンパクトディスク(CD)、CD-R、CD-RW、ディスク、テープ、テープカートリッジ等が含まれる。

【0020】

データの分岐に回答し、データの一部を複数のスナップショットの各スナップショットにコピーすることによって、データのいずれかの特定の部分を「所有する」特定のスナッ

50

ブショットが変化し、単一のスナップショットが削除されると、データの一部のみが再コピーされることになる。この技法の利益は、スナップショットの個数が増加するにつれて増加する。例えば、単一のスナップショットで構成されたシステムは利点がないが、スナップショットの個数が2つまたは3つ以上に増加するにつれて、性能は改善される。

【0021】

図3Aおよび図3Bに示す状態は、4つのスナップショット304A、304B、304C、および304Dを有する親LUNを示している。スナップショット機能を有するストレージシステムの従来の実施では、親LUNが削除されると、すべてのデータが第1のスナップショットにコピーされ、残りの3つのスナップショットが第1のスナップショットに向けられる。その後、第1のスナップショットが削除されると、すべてのデータが第2のスナップショットにコピーされ、再度、残りの第3および第4のスナップショットが第2のスナップショットに向けられる。起こり得る最悪の場合には、データは4回コピーされる。

10

【0022】

これとは対照的に、図3Aおよび図3Bに示す技法は、データの分岐の結果として生じるコピーを大幅に削減する。例えば親LUNの削除といった分岐が発生すると、データのすべてを第1のスナップショット304Aにコピーするのではなく、データの4分の1をスナップショット304A、304B、304C、および304Dのそれぞれにコピーすることができる。或るスナップショットが削除されると、データの4分の1のみを再コピーすることによって、データは正確に管理され、再コピーされたデータは、残りの3つのスナップショットにわたって分散される。別のスナップショットが削除されると、データの3分の1のみが再コピーされ、その再コピーされたデータは、残りの2つのスナップショットにわたって均一に分散される。最後に、残りのスナップショットの一方が削除されると、データの半分のみが残りの最後のスナップショットに再コピーされる。したがって、第1の親LUNを削除しその後データをコピーした後、次のスナップショットを削除する時には、最悪の場合でも、データの $1/4 + 1/3 + 1/2$ しかコピーされない。これと対照的に、従来技法は、最初の親LUNの削除に続いて、さらに3回、データのすべてをコピーすることがある。

20

【0023】

この例示は、親LUNの削除としての分岐事象を説明しているが、この技法は、親LUNおよびスナップショットを分岐させる任意の状況にも同様に適用することができる。分岐事象には、親LUNへの分岐書き込み (diverging write) だけでなく、それ以外の事象も含まれる。したがって、この技法は、別の例では、分岐書き込みが親LUNに送られた際にデータを分散させるのに使用することができる。このように、1つの分岐書き込みは、元のデータを1つのスナップショットにコピーするのに使用することができ、次の分岐書き込みは、コピーしたデータを次のスナップショットにコピーすることができ、以下同様である。その結果、個々のスナップショットは、分岐データの基本的に均等な部分を「所有する」。

30

【0024】

どの所与の時刻においても、データは、既存のスナップショットにわたってのみ基本的に均一に分散され、それによって、スナップショット管理システムのコピーの負担が低減される。親LUNが、既存の2つのスナップショットをすでに有する例では、時間の経過と共に、これら2つのスナップショットは、親LUNからの分岐したデータのすべてを均一に共有する。親の第3のスナップショットが取り込まれると、この新たなスナップショットは、元の2つのスナップショットと同じ程度に分岐したデータを最終的に受け取ることが「支持」される必要はない。その理由は、第3のスナップショットが、元の2つのスナップショットが先に受け取った、分岐したデータのいずれも共有しないからである。したがって、第3のスナップショットの発生時に、分岐書き込みは、その時にのみ3つのスナップショットにわたって均一に分散される。第3のスナップショットの生成前に発生した事象は、この現在の分散とは関係しない。分散割り当ては、現在存在するスナップシ

40

50

ットの個数によってのみ決定される。

【0025】

親LUNのデータがスナップショットから分岐した時に、親LUNの複数のスナップショットにわたってほぼ均一にデータを分散するコンセプトは、親LUNとスナップショットとの間の分岐を生じるあらゆる理由、状況、または状態に一般化することができる。この技法は、さらに、スナップショット機能をサポートする任意の適切な記憶方法にも一般化することができる。この適切な記憶方法には、おそらく、RAID0、RAID1、RAID2、RAID3、RAID4、RAID5、RAID6、RAID7、RAID10等の1つまたは2つ以上を含むさまざまな独立ディスク冗長配列(RAID)タイプが含まれる。同様に、この技法は、スナップショット機能をサポートする任意の適切な記憶媒体に使用することができる。この記憶媒体には、おそらく、さまざまな磁気ディスク、光ディスク、コンパクトディスク(CD)、CD-R、CD-RW、ディスケット、テープ、テープカートリッジ等が含まれる。また、この技法は、さらに、任意の適切な分岐の原因、状態、今後開発されるスナップショット機能をサポートする記憶方法または記憶媒体にも一般化することができる。

【0026】

図4A、図4B、および図4Cを参照して、実体ブロック図が、ストレージシステム400の一実施の形態を示している。このストレージシステム400は、ベースボリューム404および少なくとも1つの物理ブロック406A、406B、406C、および406Dを有する物理ストア402と、スナップショットボリューム410およびスナップショットインデックス412を含む論理ストア408とを含む。ストレージシステム400は、さらに、スナップショットボリュームから、物理ブロックのうちの選択されたものへのポインタをポイントインタイムでサポートできるスナップショットサブシステム414も含む。スナップショットサブシステム414は、親論理ユニット(LUN)を定義し、親LUNのデータがスナップショットから分岐すると、親LUNの複数のスナップショットにわたってデータを分散させる。

【0027】

さまざまな実施の形態では、物理ストア402および論理ストア408は、磁気ディスク、光ディスク、コンパクトディスク(CD)、CD-R、CD-RW、ディスケット、テープ、テープカートリッジ等の中から選択された媒体にスナップショットデータを記憶することができる。

【0028】

ストレージシステム400は、物理ストア402のデータを指し示す、スナップショットボリューム410の1つまたは2つ以上のマップポインタ416と、スナップショットインデックス412のデータを指し示すことができる複数のスナップショットポインタ418とを含むこともできる。スナップショットサブシステム414は、分岐したデータを、スナップショットインデックス412の複数のスナップショット部分に分散させ、この分散させたデータ部分を複数のスナップショットに書き込む。

【0029】

いくつかの実施の形態または状態では、スナップショットサブシステム414は、ほぼ等しい割合で複数のスナップショットにわたってデータを分散させる。また、スナップショットサブシステム414は、親LUNの削除、親LUNへのデータ書き込みオペレーション、親LUNの障害等を含むものの中から選択された親LUNデータ分岐状態を検出するように構成することもできる。

【0030】

また、スナップショットサブシステム414は、時間の経過と共にスナップショットの個数を変更し、複数のスナップショット間におけるほぼ均一なデータの分散を各変更時に開始して行うように構成することもできる。

【0031】

スナップショットサブシステム414によって、ストレージコンテナデータのポイント

インタタイムコピーを高速かつ効率的に作成する機能が可能になる。スナップショットは、元のデータを危険にさらすことなく、他のスナップショットから分離できて、かつ、バックアップ、アーカイブ、データ保護、テスト、および他の操作に使用できるコンテナのデータのマップをフリーズする。スナップショットが取り込まれた後、スナップショットコピーが、選択されたポイントインタタイムを維持すると同時に、元のデータは、引き続き更新および使用することができる。

【 0 0 3 2 】

特定のポイントインタタイムの複製が望まれる場合、スナップショットサブシステム 4 1 4 は、選択した時点でデータスナップショットの取得を指示する。通常、スナップショットサブシステム 4 1 4 は、複数のスナップショットを取得することができ、取得を繰り返すことができる。スナップショット機能によって、データのミラーおよびクローンに関連したオーバーヘッドの一部が回避される。

10

【 0 0 3 3 】

図 5 を参照して、概略ブロック図が、ストレージシステム 5 0 2 で使用されるコンピュータシステム 5 0 0 の一実施の形態を示している。コンピュータシステム 5 0 0 は、ベースボリューム 5 0 6 および少なくとも 1 つの物理ブロック 5 0 8 を含む物理ストア 5 0 4 と、スナップショットボリューム 5 1 2 およびスナップショットインデックス 5 1 4 を含む論理ストア 5 1 0 とを有する。このコンピュータシステムは、さらに、プロセッサ 5 1 6 で実行可能なスナップショットサブシステムを含む。このスナップショットサブシステムは、親 L U N 5 1 8 のデータがスナップショットから分岐すると、親 L U N 5 1 8 の複数のスナップショットにわたってデータを分散させる。

20

【 0 0 3 4 】

プロセッサ 5 1 6 は、ベースボリューム 5 0 6 を定義するマッピングロジックを実施でき、物理ブロック 5 0 8 をベースボリューム 5 0 6 に割り当て、スナップショットボリューム 5 1 2 から、選択された物理ブロック 5 0 8 およびスナップショットインデックス 5 1 4 へのポインタを作成する。

【 0 0 3 5 】

スナップショット管理機能を実行するプロセッサ 5 1 6 は、ネットワークの任意の適切なデバイスに配置することができる。図示するように、プロセッサ 5 1 6 は、ストレージコントローラ内に収容することができる。他の実施の形態では、スナップショット機能を実行できるプロセッサは、ホスト、ストレージ配列ネットワーク (S A N) 内の適切な制御デバイス、ネットワークに取り付けられたネットワーク機器、配列ファームウェア、またはポイントインタタイムコピーを行うことができる他の任意のレベルの実行に存在することができる。

30

【 0 0 3 6 】

複数のスナップショットへのデータ分散プロセスは、書き込み前コピーオペレーション、書き込み時コピーオペレーション等を含めて、プログラム可能オペレーション等のさまざまなオペレーションで実施することができる。

【 0 0 3 7 】

プロセッサ 5 1 6 は、さらに、1 つまたは 2 つ以上のマップポインタをスナップショットボリューム 5 1 2 に生成するマッピングロジックを実行することができる。このマップポインタは、物理ストア 5 0 4 のデータと、スナップショットインデックス 5 1 4 のデータを指し示すことができる 1 つまたは 2 つ以上のスナップショットポインタとを指し示す。プロセッサ 5 1 6 は、さらに、スナップショットロジックを実行する。このスナップショットロジックは、分岐したデータを、スナップショットインデックス 5 1 4 の複数のスナップショット部分に分散させ、分散させたデータ部分を複数のスナップショットに書き込む。データを、等しい割合またはおよそ等しい割合で複数のスナップショットにわたって最も効率的に分散させることができる。

40

【 0 0 3 8 】

プロセッサ 5 1 6 は、親 L U N の削除、親 L U N へのデータ書き込みオペレーション、

50

親 L U N の障害等を含む 1 つまたは 2 つ以上の親 L U N のデータ分岐状態を検出することができる。

【 0 0 3 9 】

プロセッサ 5 1 6 は、スナップショットハンドラを実行することができる。このスナップショットハンドラは、時間の経過と共にスナップショットの個数を変更し、複数のスナップショット間におけるほぼ均一なデータの分散を各変更時に開始して行う。通常、システムは、特定の最大数のスナップショットを割り当て、タイミングインターバル、起動信号、監視状態等の選択された事象の発生時にスナップショットの作成を開始する。スナップショットの個数は、通常、選択された個数に制限される。ただし、いくつかの実施態様は、仮想的に無制限のスナップショットをサポートすることができる。

10

【 0 0 4 0 】

物理ストア 5 0 4 は、スナップショット機能に適した任意のストレージデバイスを含むことができ、磁気ディスク、光ディスク、コンパクトディスク (C D)、C D - R、C D - R W、ディスク、テープ、テープカートリッジ等のさまざまな媒体を含むことができる。

【 0 0 4 1 】

システムが遂行または実行するさまざまな機能、プロセス、方法、およびオペレーションを、さまざまなタイプのロジック、プロセッサ、コントローラ、中央処理装置、マイクロプロセッサ、デジタル信号プロセッサ、状態マシン、プログラマブルロジックアレイ等で実行可能なプログラムとして実施することができる。プログラムは、任意のコンピュータ関連システムまたは方法によって、または、これらと共に使用される任意のコンピュータ可読媒体に記憶することができる。コンピュータ可読媒体は、電子デバイス、磁気デバイス、光デバイス、または、他の物理デバイスもしくは物理手段である。他の物理デバイスまたは物理手段は、コンピュータ関連システム、方法、プロセス、もしくはプロシージャによって、または、これらと共に使用されるコンピュータプログラムを収容または記憶できるものである。プログラムは、命令実行システム、デバイス、コンポーネント、エレメント、もしくは装置によって、または、これらと共に使用されるコンピュータ可読媒体に実施することができる。これら命令実行システム等は、コンピュータまたはプロセッサに基づくシステムや、任意の適切なタイプの命令メモリまたはストレージから命令をフェッチできる他のシステム等である。コンピュータ可読媒体は、任意の構造、デバイス、コンポーネント、製品、または、命令実行システム、装置、もしくはデバイスによって、もしくは、これらと共に使用されるプログラムを記憶、通信、伝搬、もしくは転送できる他の手段とすることができる。

20

30

【 0 0 4 2 】

例示のブロック図およびデータ構造図は、プロセスの特定の論理機能またはステップを実施する 1 つまたは 2 つ以上の実行可能命令を含むモジュール、セグメント、またはコード部分を表すことができるプロセスステップまたはブロックを示している。特定の例は、特定のプロセスステップまたは動作を示しているが、多くの代替的な実施態様が可能であり、一般に、簡単な設計選択によって行うことができる。動作およびステップは、機能、目的、標準規格への適合、レガシー、構造等の考慮に基づいて、本明細書の特定の説明とは異なる順序で実行することができる。

40

【 0 0 4 3 】

本開示はさまざまな実施の形態を説明しているが、これらの実施の形態は、例示として理解されるべきであり、特許請求の範囲を限定するものではない。説明した実施の形態の多くの変形、変更、追加、および改良が可能である。例えば、当業者は、本明細書に開示した構造および方法を提供するのに必要なステップを容易に実施し、プロセスパラメータ、材料、および寸法が単なる例として与えられているにすぎないことを理解するであろう。これらパラメータ、材料、および寸法を変更して、所望の構造および変更を得ることができる。これら所望の構造および変更も特許請求の範囲内に含まれる。本明細書に開示した実施の形態の変形および変更も、添付した特許請求の範囲内に含まれることを維持しつ

50

つ行うことができる。例えば、例示のスナップショット技法は、あらゆる適切な媒体を含めて、このような技法に適したあらゆるタイプのストレージシステムで実施することができる。同様に、例示の技法は、あらゆる適切なストレージシステムアーキテクチャで実施することができる。

【図面の簡単な説明】

【0044】

【図1A】配列のマップの粒度が、スナップショットコピーと元のデータとの間の分岐が配列内に記録される粒度と同じであるスナップショット機能を実施するストレージシステムに使用されるデータ構造を示す概略ブロック図である。

【図1B】配列のマップの粒度が、スナップショットコピーと元のデータとの間の分岐が配列内に記録される粒度と同じであるスナップショット機能を実施するストレージシステムに使用されるデータ構造を示す概略ブロック図である。

【図2A】配列のマップの粒度が、スナップショットコピーと元のデータとの間の分岐が配列内に記録される粒度と同じでないスナップショット機能を実施するストレージシステムに使用されるデータ構造を示す概略ブロック図である。

【図2B】配列のマップの粒度が、スナップショットコピーと元のデータとの間の分岐が配列内に記録される粒度と同じでないスナップショット機能を実施するストレージシステムに使用されるデータ構造を示す概略ブロック図である。

【図3A】データが複数のスナップショットにわたって分散されるスナップショット機能を実施するストレージシステムの一実施の形態で使用されるデータ構造を示す概略ブロック図である。

【図3B】データが複数のスナップショットにわたって分散されるスナップショット機能を実施するストレージシステムの一実施の形態で使用されるデータ構造を示す概略ブロック図である。

【図4A】複数のスナップショットにわたってデータを分散させることによってスナップショット機能を管理するストレージシステムの一実施の形態を示す実体ブロック図である。

【図4B】複数のスナップショットにわたってデータを分散させることによってスナップショット機能を管理するストレージシステムの一実施の形態を示す実体ブロック図である。

【図4C】複数のスナップショットにわたってデータを分散させることによってスナップショット機能を管理するストレージシステムの一実施の形態を示す実体ブロック図である。

【図5】複数のスナップショットにわたってデータを分散させることによってスナップショット機能を管理するストレージシステムで使用されるコンピュータシステムの一実施の形態を示す概略ブロック図である。

【符号の説明】

【0045】

102、106 マップエントリ
 104、108、122 記憶データ
 110、112 スナップショットエントリ
 202、206 マップエントリ
 204、208 記憶データ
 210、212 スナップショットエントリ
 222 変更記憶データ
 224 元の記憶データ
 302 記憶データ
 304A～304D スナップショット
 306 親論理ユニット(LUN)
 400 ストレージシステム

10

20

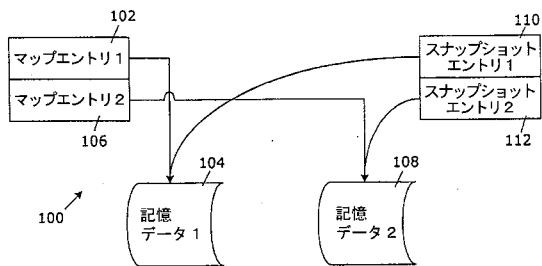
30

40

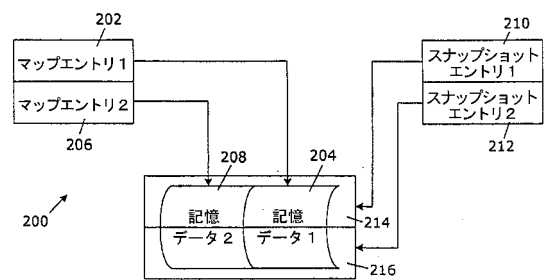
50

- 4 0 2 物理ストア
- 4 0 4 ベースボリューム
- 4 0 6 A ~ 4 0 6 C 物理ブロック
- 4 0 8 論理ストア
- 4 1 0 スナップショットボリューム
- 4 1 2 スナップショットインデックス
- 4 1 4 S Sシステム (スナップショットサブシステム)
- 4 1 6 マップポインタ
- 4 1 8 スナップショットポインタ
- 5 0 0 コンピュータシステム
- 5 0 2 ストレージシステム
- 5 0 4 物理ストア
- 5 0 6 ベースボリューム
- 5 0 8 物理ブロック
- 5 1 0 論理ストア
- 5 1 2 スナップショットボリューム
- 5 1 4 スナップショットインデックス
- 5 1 6 プロセッサ
- 5 1 8 親 L U N

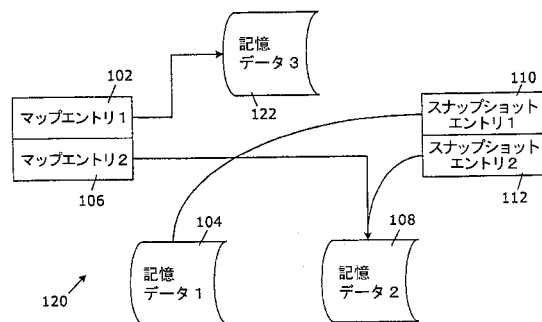
【図 1 A】



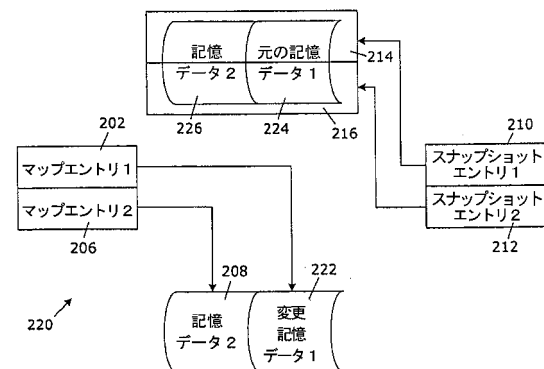
【図 2 A】



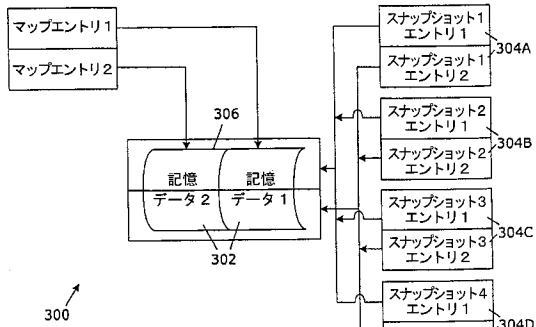
【図 1 B】



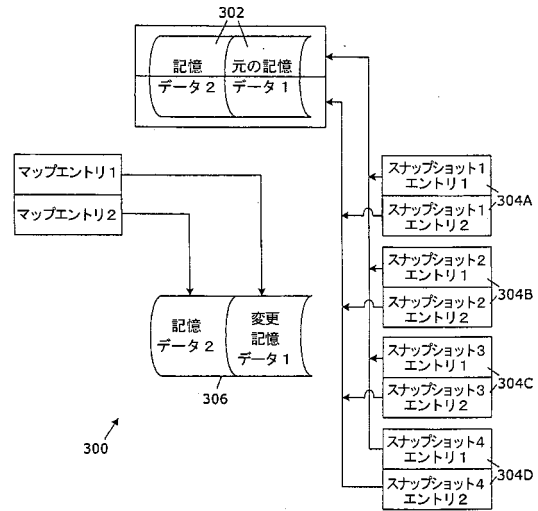
【図 2 B】



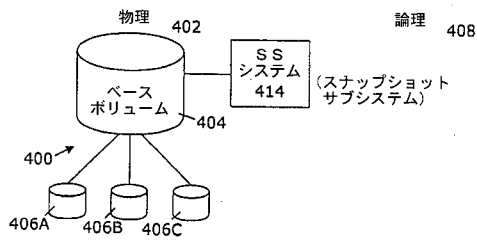
【 図 3 A 】



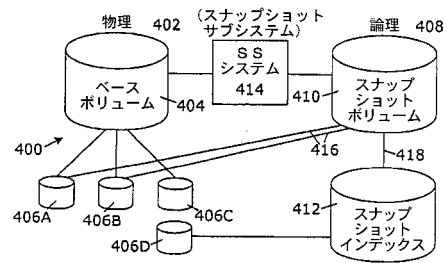
【 図 3 B 】



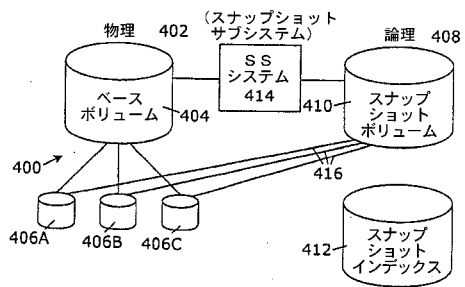
【 図 4 A 】



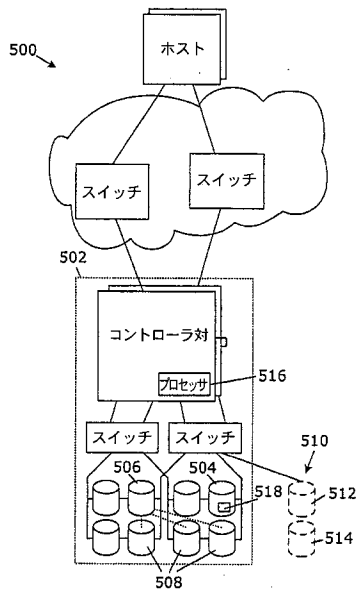
【 図 4 C 】



【 図 4 B 】



【 図 5 】



フロントページの続き

(72)発明者 リー・ネルソン

アメリカ合衆国カリフォルニア州 パロアルト ハノーバー・ストリート 3000 ヒューレッ
ト・パッカード・カンパニー内

Fターム(参考) 5B065 BA01 CA30 EA33
5B082 DE06