

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7097379号

(P7097379)

(45)発行日 令和4年7月7日(2022.7.7)

(24)登録日 令和4年6月29日(2022.6.29)

(51)国際特許分類

F I

G 0 6 F 3/06 (2006.01)

G 0 6 F 3/06 3 0 1 P

G 0 6 F 13/10 (2006.01)

G 0 6 F 3/06 3 0 1 K

G 0 6 F 16/18 (2019.01)

G 0 6 F 13/10 3 4 0 B

G 0 6 F 16/18 1 0 0

請求項の数 14 (全27頁)

(21)出願番号 特願2019-548590(P2019-548590)

(86)(22)出願日 平成30年3月9日(2018.3.9)

(65)公表番号 特表2020-511714(P2020-511714
A)

(43)公表日 令和2年4月16日(2020.4.16)

(86)国際出願番号 PCT/US2018/021659

(87)国際公開番号 WO2018/165502

(87)国際公開日 平成30年9月13日(2018.9.13)

審査請求日 令和3年1月27日(2021.1.27)

(31)優先権主張番号 15/453,949

(32)優先日 平成29年3月9日(2017.3.9)

(33)優先権主張国・地域又は機関
米国(US)

(73)特許権者 303039534

ネットアップ, インコーポレイテッド
アメリカ合衆国, カリフォルニア州 9
4 0 8 9, サニーベール, クロスマン・
アベニュー 1 3 9 5

(74)代理人 100107766

弁理士 伊東 忠重

(74)代理人 100070150

弁理士 伊東 忠彦

(74)代理人 100091214

弁理士 大貫 進介

(72)発明者 ドロナムラジュ, ラビカント

アメリカ合衆国, カリフォルニア州 9
4 0 8 9, サニーベール, イースト・ジ
ャバ・ドライブ 4 9 5

最終頁に続く

(54)【発明の名称】 ストリームを使用するデータの割り振りエリアへの選択的記憶

(57)【特許請求の範囲】

【請求項1】

プロセッサにより、ストレージに書き込むべき第1のデータがユーザデータを含むと決定するステップと、前記プロセッサにより、前記ストレージに書き込むべき第2のデータがメタデータを含むと決定するステップと、前記プロセッサにより、前記第1のデータを前記ストレージの第1の割り振りエリアに選択的に記憶するステップであり、前記第1の割り振りエリアは、前記ユーザデータのための複数の割り振りエリアをソートすることにより生成される前記ユーザデータのための前記複数の割り振りエリアのソートセットから選択された割り振りエリアであり、前記ユーザデータのための前記複数の割り振りエリアは、ポリシーにより規定される、ステップと、
前記プロセッサにより、前記第2のデータを前記ストレージの第2の割り振りエリアに選択的に記憶するステップであり、前記第2の割り振りエリアは、前記メタデータのための複数の割り振りエリアをソートすることにより生成される前記メタデータのための前記複数の割り振りエリアのソートセットから選択された割り振りエリアであり、前記メタデータのための前記複数の割り振りエリアは、前記ポリシーにより規定される、ステップと、
を含み、前記ユーザデータのための前記ソートセット及び前記メタデータのための前記ソートセットは、前記ストレージに書き込むべきデータが受信されたとき、前記ポリシー及び/又は各割り振りエリアの利用可能な空き領域の量に基づいて動的にソートされる、方法。

【請求項 2】

前記第 1 の割り振りエリア及び前記第 2 の割り振りエリアは前記ストレージのストレージ装置内に含まれる、請求項 1 に記載の方法。

【請求項 3】

前記第 1 の割り振りエリアは前記ストレージの第 1 のストレージ装置内に含まれ、前記第 2 の割り振りエリアは前記ストレージの第 2 のストレージ装置内に含まれる、請求項 1 に記載の方法。

【請求項 4】

前記プロセッサにより、前記ユーザデータ及び前記メタデータが異なるアクセスパターンを有することに基づいて、前記ユーザデータが前記第 1 の割り振りエリア内に記憶されるべきであること及び前記メタデータが前記第 2 の割り振りエリア内に記憶されるべきであることを規定するように前記ポリシーを定義するステップを含む請求項 1 乃至 3 のうちいずれか 1 項に記載の方法。

10

【請求項 5】

前記第 1 のデータを選択的に記憶するステップは、
前記プロセッサにより、前記第 1 の割り振りエリアに関連づけられた第 1 のストリーム識別子でタグ付けされた第 1 のストリームを使用して、前記第 1 のデータを前記ストレージに送信するステップを含む、請求項 1 乃至 4 のうちいずれか 1 項に記載の方法。

20

【請求項 6】

前記ポリシーは、前記第 1 のストリーム識別子が前記第 1 の割り振りエリアに関連づけられることを規定する、請求項 5 に記載の方法。

【請求項 7】

前記第 2 のデータを選択的に記憶するステップは、
前記プロセッサにより、前記第 2 の割り振りエリアに関連づけられた第 2 のストリーム識別子でタグ付けされた第 2 のストリームを使用して、前記第 2 のデータを前記ストレージに送信するステップを含む、請求項 1 乃至 6 のうちいずれか 1 項に記載の方法。

【請求項 8】

前記ポリシーは、前記第 2 のストリーム識別子が前記第 2 の割り振りエリアに関連づけられることを規定する、請求項 7 に記載の方法。

30

【請求項 9】

前記プロセッサにより、前記第 1 の割り振りエリアが閾値未満の量の空き領域を有するという決定に基づいて、前記第 1 の割り振りエリアの利用可能なストレージ領域を増加させるように前記第 1 の割り振りエリアを再定義するステップを含む請求項 1 乃至 8 のうちいずれか 1 項に記載の方法。

【請求項 10】

前記プロセッサにより、前記第 1 の割り振りエリア及び前記第 2 の割り振りエリアを、ストレージ装置の消去ブロック単位の整数倍として定義するステップを含む請求項 1 乃至 9 のうちいずれか 1 項に記載の方法。

40

【請求項 11】

マシンに請求項 1 乃至 10 のうちいずれか 1 項に記載の方法を実行させる命令を含むコンピュータプログラム。

【請求項 12】

マシン実行可能コードを含むメモリと、
前記メモリに結合されたプロセッサと、を含み、
前記プロセッサは、前記マシン実行可能コードを実行して前記プロセッサに、
ストレージに書き込むべき第 1 のデータがユーザデータを含むと決定し、
前記ストレージに書き込むべき第 2 のデータがメタデータを含むと決定し、
前記第 1 のデータを前記ストレージの第 1 の割り振りエリアに選択的に記憶し、前記第 1

50

の割り振りエリアは、前記ユーザデータのための複数の割り振りエリアをソートすることにより生成される前記ユーザデータのための前記複数の割り振りエリアのソートセットから選択された割り振りエリアであり、前記ユーザデータのための前記複数の割り振りエリアは、ポリシーにより規定され、

前記第2のデータを前記ストレージの第2の割り振りエリアに選択的に記憶し、前記第2の割り振りエリアは、前記メタデータのための複数の割り振りエリアをソートすることにより生成される前記メタデータのための前記複数の割り振りエリアのソートセットから選択された割り振りエリアであり、前記メタデータのための前記複数の割り振りエリアは、前記ポリシーにより規定される

ことをさせるように構成され、

前記ユーザデータのための前記ソートセット及び前記メタデータのための前記ソートセットは、前記ストレージに書き込むべきデータが受信されたとき、前記ポリシー及び/又は各割り振りエリアの利用可能な空き領域の量に基づいて動的にソートされる、コンピューティング装置。

【請求項13】

前記マシン実行可能コードは前記プロセッサに、

前記第1の割り振りエリアが閾値未満の量の空き領域を有するという決定に基づいて、前記第1の割り振りエリアの利用可能なストレージ領域を増加させるように前記第1の割り振りエリアを再定義する

ことをさせる、請求項12に記載のコンピューティング装置。

【請求項14】

前記マシン実行可能コードは前記プロセッサに、

前記第1の割り振りエリア及び前記第2の割り振りエリアを、ストレージ装置の消去ブロック単位の整数倍として定義する

ことをさせる、請求項12又は13に記載のコンピューティング装置。

【発明の詳細な説明】

【背景技術】

【0001】

ストレージコントローラの多くのファイルシステムは、種々のプロトコルに関連づけられた複数のチャンネルを通じてデータを受信することができる。ファイルシステムは、そのデータをストレージ装置内に又は複数のストレージ装置にわたり記憶することができる。こうしたデータは、例えば、ユーザデータ（例えば、ユーザデータベースファイル）又はメタデータ（例えば、ボリュームのボリュームサイズ、ストレージコントローラのネットワークアドレス、複製ポリシー、及び/又はファイルシステム及び/又はストレージコントローラにより使用される他のデータ）であるなど、種々の特徴を有し得る。別の例において、特性は、ホットデータ（例えば、ファイルシステムにより頻繁に変更されているメタデータなどの、閾頻度を超えてアクセスされているデータ）又はコールドデータ（例えば、閾頻度未満でアクセスされているユーザデータ）に対応することができる。別の例において、特性は、順次アクセスされるデータ（例えば、隣接するブロック内に記憶されるデータ）又はランダムにアクセスされるデータ（例えば、隣接しないブロック内に記憶されるデータ）に対応することができる。

【0002】

仮想化レイヤは、複数のストレージ装置からの物理ストレージをクライアント及びアプリケーションに対して単一のストレージオブジェクトに見えるものに一緒にグループ化する間接レイヤとして使用できる（例えば、ボリューム又は論理ユニット番号（LUN）が、複数の物理ストレージ装置に及ぶ可能性がある）。仮想化レイヤは、ストレージの物理レイアウトを抽象化し、ゆえに、基礎となる物理アドレス空間にマッピングされた論理アドレス空間で動作する。仮想化されたストレージでは、ストレージ装置は、ストレージ装置により選択されたロケーションの中のストレージ装置の物理ブロック内にデータを物理的に記憶する役割を担うことがある。しかしながら、ストレージ装置は、いかなる論理的/

10

20

30

40

50

物理的分離もなしに、ホットデータ、コールドデータ、ユーザデータ、及びメタデータなどの任意タイプのデータを一緒に記憶する可能性がある。不運なことに、異なる特性を有するデータは、異なるアクセス及び上書きパターンを有する可能性があり、ゆえに、このようなデータが一緒に記憶されているとき、フラグメンテーションが結果として生じる可能性がある。

【 0 0 0 3 】

一例において、ソリッドステートドライブ（SSD）は、前に書き込まれたブロックを上書きする能力を有さない場合があり、空の宛先セルにのみ書き込むことができる。ゆえに、宛先セルが再使用されるべきデータを有する場合、データは異なる空セルに移動されることがあり、宛先セルは再プログラム（例えば、消去）されなければならない、それにより、新しいデータが宛先セルに書き込みできる。これは、フラグメンテーションの増加を結果としてもたらし、異なる特性のデータが一緒に記憶されているとき、フラグメンテーションはより著しくなる。ストレージ装置は、領域を予約して、セルを先見的に解放できるバックグラウンドガベージコレクションを提供することができる。しかしながら、ストレージ装置のストレージのうち約 28 % 又は任意の他のパーセンテージなど、相当量のストレージ領域が予約される可能性がある。これは、オーバープロビジョニングに起因するストレージリソースの非効率的な使用及びコストの増加をもたらす。また、書き込み増幅は後続の上書きで問題になり、ストレージ装置における性能及び摩耗の劣化をもたらす可能性がある。

【図面の簡単な説明】

【 0 0 0 4 】

【図 1】本明細書に記載された対策の 1 つ以上に従う、例示的なクラスタ化ネットワークを示すコンポーネントブロック図である。

【図 2】本明細書に記載された対策の 1 つ以上に従う、例示的なデータストレージシステムを示すコンポーネントブロック図である。

【図 3】ストリームを使用してデータを割り振りエリアに選択的に記憶する例示的な方法を示すフローチャートである。

【図 4】ストリームを使用してデータを割り振りエリアに選択的に記憶するための例示的なコンピューティング装置を示すコンポーネントブロック図であり、割り振りエリアが定義され、割り振りエリアにポリシーが割り当てられる。

【図 5】ストリームを使用してデータを割り振りエリアに選択的に記憶するための例示的なコンピューティング装置を示すコンポーネントブロック図であり、割り振りエリアは複数のストレージ装置にわたり定義される。

【図 6】ストリームを使用してデータを割り振りエリアに選択的に記憶する例示的な方法を示すフローチャートである。

【図 7】ストリームを使用してデータを割り振りエリアに選択的に記憶するための例示的なコンピューティング装置を示すコンポーネントブロック図である。

【図 8】本明細書に記載された対策の 1 つ以上に従う、コンピュータ読取可能媒体の例である。

【発明を実施するための形態】

【 0 0 0 5 】

次に、請求される対象事項のいくつかの例が図面を参照して記載される。ここで、同様の参照番号は、全体を通して同様の要素を参照するために一般に使用される。以下の記載において、説明の目的で、請求される対象事項の理解を提供するために、多数の特定の詳細が説明される。しかしながら、請求される対象事項はこれらの特定の詳細なしに実施され得ることが明らかであろう。この詳細な説明の如何なる記載も、従来技術として認められない。

【 0 0 0 6 】

ストリームを使用してデータを割り振りエリア（allocation areas）に選択的に記憶するための 1 つ以上の手法及び / 又はコンピューティング装置が本明細書で提供される。

【 0 0 0 7 】

ストレージ装置は、仮想化されたストレージをクライアントに提供するために仮想化レイヤにより使用されることがある（例えば、仮想化レイヤは、物理ストレージの基礎となる詳細を隠すことがあり、複数の物理装置の物理ストレージをクライアント及びアプリケーションに晒される単一のストレージオブジェクトにグループ化することがある）。ストレージ装置が、論理アドレス空間の物理アドレス空間への十分定義されたマッピングを有さない場合、ストレージ装置は単に、任意タイプのデータを一緒に記憶することになる。ある時間にわたり、異なるタイプのデータ（例えば、異なるアクセス頻度を有するデータ、異なる上書きパターン及び頻度を有するデータ、異なる集合体のデータ、ランダムにアクセスされるデータ、順次アクセスされるデータ、ホットデータ、コールドデータ、ユーザデータ、メタデータ等）を一緒に記憶することは、ストレージ装置のフラグメンテーションを結果としてもたらす可能性がある。書き込み増幅（Write amplification）もまた、その後の上書きで結果として生じることになる。また、ソリッドステートドライブなどの特定タイプのストレージ装置は、ガベージコレクション機能による使用のためにストレージをオーバプロビジョニングし（例えば、あるパーセンテージの、さもないと空いているストレージを予約し）、ソリッドステートドライブのセルを先見的に解放する。こうしたオーバプロビジョニングは、さもないと我々がユーザデータ及び／又はメタデータを記憶するために使用できるストレージ領域を浪費する。ゆえに、フラグメンテーションを低減させ、書き込み増幅を低減させ、ストレージ領域の記憶効率を改善する必要がある。

10

20

【 0 0 0 8 】

したがって、本明細書に提供されるように、書き込みストリームのデータは、こうしたデータの特性に基づいて異なるストリームに割り当てられる。詳細には、異なる特性を有するデータがストレージ装置の異なる割り振りエリアに（例えば、ストレージ装置の異なる物理アドレス範囲又は仮想ブロック番号内に）記憶されるべきであると規定する1つ以上のポリシーに基づいて、頻繁にアクセスされるデータが第1のストリームに割り当てられてもよく、稀にアクセスされるデータが第2のストリームに割り当てられてもよく、ランダムにアクセスされるデータが第3のストリームに割り当てられてもよく、順次アクセスされるデータが第4のストリームに割り当てられてもよい、等である。

【 0 0 0 9 】

ポリシーとストリームへのデータの割り当てとはファイルシステムにより実現されてもよく、それにより、ストレージ装置のストレージが仮想化され、かつ／あるいはストレージ装置が物理アドレス空間への論理アドレス空間の十分定義されたマッピングを維持せず、ゆえにさもないとすべてのデータを一緒に又はいかなる識別もなく単に記憶することになるとしても、データはストレージ装置内の別個の場所に記憶できる。各ストリームは、ポリシーにより対応する割り振りエリアに割り当てられた特定のストリーム識別子でタグ付けされてもよい。例えば、ポリシーは、頻繁にアクセスされるデータが割り振りエリア（C）に記憶されるべきであると規定し、したがって、頻繁にアクセスされるデータの第1のストリームは、第1のストリームのデータが割り振りエリア（C）を使用して処理される（例えば、その中に記憶される）べきであるというストレージ装置に対するインジケータとして使用されるストリーム識別子でタグ付けされる。同様の特性を有するデータを同じ割り振りエリア内に一緒に記憶し、同様でない文字を有するデータを別個の割り振りエリアに記憶することは、全体的ストレージ装置のフラグメンテーション及び書き込み増幅を低減させる（例えば、頻繁に上書きされるデータは、ストレージ装置全体にわたり拡散されるのと反対に単一の割り振りエリア内に含むことができ、それにより、上書きによるフラグメンテーションは、ストレージ装置全体に影響しないことになる。さもないと、特に、任意の書き込みオプションに対してデータを新しい位置に書き込むライトエニウェア（write anywhere）ファイルシステムについて、ストレージ装置全体にわたりフラグメンテーションが結果として生じることになる）。これはまた、バックグラウンドのガベージコレクション処理が必要とされない可能性があり、あるいはガベージコレクション

30

40

50

に対してストレージ装置の一層小さい予約エリアを使用する可能性があるため、記憶効率を改善する。

【 0 0 1 0 】

ストリームを使用するデータの割り振りエリアへの選択的記憶を提供するために、図 1 は、クラスタ化されたネットワーク環境 1 0 0 又はネットワークストレージ環境の一実施形態を示す。しかしながら、本明細書に記載される手法等は、クラスタ化ネットワーク環境 1 0 0、非クラスタネットワーク環境、及び / 又はデスクトップコンピューティング環境などの種々の他のコンピューティング環境内に実現されてもよいことが十分理解され得る。すなわち、別記の特許請求の範囲を含む本開示は、本明細書に提供される実施形態に限定されることを意図しない。同一又は同様のコンポーネント、要素、特徴、アイテム、モジュール等が後の図で図示されているが、前の図に関して前に論じられていた場合、（例えば、理解を簡単及び容易にする目的で）後続の図を説明するとき、上記コンポーネント、要素、特徴、アイテム、モジュール等についての同様の（例えば、冗長な）議論は省略され得ることが十分理解されよう。

10

【 0 0 1 1 】

図 1 は、本明細書に記載される手法及び / 又はシステムの少なくともいくつかの実施形態を実現し得る、クラスタ化ネットワーク環境 1 0 0 を示すブロック図である。クラスタ化ネットワーク環境 1 0 0 は、データストレージシステム 1 0 2 及び 1 0 4 を含み、データストレージシステム 1 0 2 及び 1 0 4 は、該データストレージシステム 1 0 2 及び 1 0 4（及び、例えば、その中の 1 つ以上のモジュール、コンポーネント等、例えばノード 1 1 6 及び 1 1 8 など）の間の通信を容易にするプライベートインフィニバンド（Infiniband）、ファイバチャネル（Fibre Channel、FC）、又はイーサネット（登録商標）ネットワークとして具現化されたコンピューティングネットワークなどのクラスタファブリック 1 0 6 を通じて結合される。2 つのデータストレージシステム 1 0 2 及び 1 0 4 並びに 2 つのノード 1 1 6 及び 1 1 8 が図 1 に示されるが、任意の適切な数のこうしたコンポーネントが企図されることが十分理解されよう。一例において、ノード 1 1 6、1 1 8 は、ストレージコントローラを含み（例えば、ノード 1 1 6 が、プライマリ又はローカルストレージコントローラを含んでもよく、ノード 1 1 8 が、セカンダリ又はリモートストレージコントローラを含んでもよい）、上記ストレージコントローラは、データストレージ装置 1 2 8、1 3 0 内に記憶されたデータへのアクセスを、ホスト装置 1 0 8、1 1 0 などのクライアント装置に提供する。同様に、本明細書において特に別段提供されない限り、本明細書で参照され及び / 又は添付の図面に示される他のモジュール、要素、特徴、アイテム等についても同様である。すなわち、本明細書に開示される特定の数のコンポーネント、モジュール、要素、特徴、アイテム等は、限定的な方法で解釈されることを意図しない。

20

30

【 0 0 1 2 】

さらに、クラスタ化ネットワークは、いかなる特定の地理的エリアにも限定されず、ローカルに及び / 又はリモートにクラスタ化できることが十分理解されよう。ゆえに、一実施形態において、クラスタ化ネットワークは、複数の地理的ロケーションに位置する複数のストレージシステム及び / 又はノードにわたり分散できる。一方、別の実施形態において、クラスタ化ネットワークは、同じ地理的ロケーションに（例えば、データストレージ装置の単一のオンサイトのラックに）存在するデータストレージシステム（例えば、1 0 2、1 0 4）を含むことができる。

40

【 0 0 1 3 】

図示の例において、例えば、クライアント装置、パーソナルコンピュータ（PC）、記憶のために使用されるコンピューティング装置（例えば、ストレージサーバ）、及び他のコンピュータ又は周辺装置（例えば、プリンタ）を含み得る 1 つ以上のホスト装置 1 0 8、1 1 0 は、ストレージネットワーク接続 1 1 2、1 1 4 によりそれぞれのデータストレージシステム 1 0 2、1 0 4 に結合される。ネットワーク接続は、例えば、データパケットを交換するための共通インターネットファイルシステム（Common Internet File System、CIFS）プロトコル又はネットワークファイルシステム（Network File System、NFS）プロトコル等を含むことができる。

50

m、N F S) プロトコルなどのネットワークアタッチトストレージ (Network Attached Storage、N A S) プロトコル、スモールコンピュータシステムインターフェース (Small Computer System Interface、S C S I) 又はファイバチャネルプロトコル (Fiber Channel Protocol、F C P) などのストレージエリアネットワーク (Storage Area Network、S A N) プロトコル、S 3 などのオブジェクトプロトコル等を利用するローカルエリアネットワーク (L A N) 又はワイドエリアネットワーク (W A N) を含んでもよい。例示的に、ホスト装置 1 0 8、1 1 0 は、アプリケーションを実行する汎用コンピュータであってもよく、情報の交換のためにクライアント/サーバモデルを使用してデータストレージシステム 1 0 2、1 0 4 と相互作用してもよい。すなわち、ホスト装置は、データストレージシステムからのデータ (例えば、ホスト装置によりストレージ装置に対して発行された I / O コマンドを処理するように構成されたネットワークストレージ制御により管理される、該ストレージ装置上のデータ) を要求してもよく、データストレージシステムは、1 つ以上のストレージネットワーク接続 1 1 2、1 1 4 を介して、要求の結果をホスト装置に返してもよい。

【 0 0 1 4 】

クラスタ化されたデータストレージシステム 1 0 2、1 0 4 上のノード 1 1 6、1 1 8 は、例えば、リモートのロケーション、クラウドストレージ (例えば、ストレージエンドポイントがデータクラウド内に記憶されてもよい) 等を有する企業などに対してデータストレージ及び管理サービスを提供するためにクラスタとして相互接続されたネットワーク又はホストノードを含むことができる。クラスタ化されたネットワーク環境 1 0 0 内のこうしたノードは、例えば、接続ポイント、再分配ポイント、又は通信エンドポイントとしてネットワークにアタッチされた装置とすることができる。ノードは、ネットワーク通信チャネルを通じて情報を送信、受信、及び/又は転送可能であってもよく、これらの基準のいずれか又は全てを満たす任意の装置を含むことができる。ノードの一例は、ネットワークにアタッチされたデータストレージ及び管理サーバであってもよく、サーバは、データストレージ及び管理システムにおいてサーバとして動作するように特に構成された汎用コンピュータ又はコンピューティング装置を含むことができる。

【 0 0 1 5 】

一例において、ノード 1 1 6、1 1 8 などの第 1 のノードクラスタ (例えば、1 つ以上のストレージ装置の第 1 の論理グループ化を含む第 1 のストレージ集合体 (storage aggregate) へのアクセスを提供するように構成された第 1 のストレージコントローラセット) が、第 1 のストレージサイトに位置してもよい。図示されていない第 2 のノードクラスタが、第 2 のストレージサイトに位置してもよい (例えば、1 つ以上のストレージ装置の第 2 の論理グループ化を含む第 2 のストレージ集合体へのアクセスを提供するように構成された第 2 のストレージコントローラセット)。第 1 のノードクラスタ及び第 2 のノードクラスタは、災害復旧構成に従って構成されてもよく、災害ノードクラスタを含む災害ストレージサイトで災害が発生した事象において、生き残りのノードクラスタが災害ノードクラスタのストレージ装置への切り替えアクセスを提供する (例えば、第 2 のストレージサイトで災害が発生した事象において、第 1 のノードクラスタが、第 2 のストレージ集合体のストレージ装置への切り替えデータアクセスをクライアント装置に提供する)。

【 0 0 1 6 】

クラスタ化ネットワーク環境 1 0 0 に示されるように、ノード 1 1 6、1 1 8 は、クラスタのための分散ストレージアーキテクチャを提供するように協調する様々な機能コンポーネントを含むことができる。例えば、ノードは、ネットワークモジュール 1 2 0、1 2 2 と、ディスクモジュール 1 2 4、1 2 6 とを含むことができる。ネットワークモジュール 1 2 0、1 2 2 は、例えば、ノード 1 1 6、1 1 8 (例えば、ネットワークストレージコントローラ) がネットワーク接続 1 1 2、1 1 4 を通じてホスト装置 1 0 8、1 1 0 と接続することを可能にするように構成でき、ホスト装置 1 0 8、1 1 0 が分散ストレージシステムに記憶されたデータにアクセスすることを可能にする。さらに、ネットワークモジュール 1 2 0、1 2 2 は、クラスタファブリック 1 0 6 を通して 1 つ以上の他のコンポー

10

20

30

40

50

ネットとの接続を提供することができる。例えば、図 1 において、ノード 1 1 6 のネットワークモジュール 1 2 0 は、ノード 1 1 8 のディスクモジュール 1 2 6 を通して要求を送信することにより第 2 のデータストレージ装置にアクセスすることができる。

【 0 0 1 7 】

ディスクモジュール 1 2 4、1 2 6 は、1 つ以上のデータストレージ装置 1 2 8、1 3 0、例えば、ディスク又はディスクアレイ、フラッシュメモリ、又は何らかの他の形式のデータストレージなどを、ノード 1 1 6、1 1 8 に接続するように構成できる。ノード 1 1 6、1 1 8 は、例えば、クラスタファブリック 1 0 6 により相互接続でき、クラスタ内のそれぞれのノードが、クラスタ内の異なるノードに接続されたデータストレージ装置 1 2 8、1 3 0 上のデータにアクセスすることを可能にする。しばしば、ディスクモジュール 1 2 4、1 2 6 は、例えば、S C S I 又は F C P などの S A N プロトコルに従ってデータストレージ装置 1 2 8、1 3 0 と通信する。ゆえに、ノード 1 1 6、1 1 8 上のオペレーティングシステムから見られるとき、データストレージ装置 1 2 8、1 3 0 は、オペレーティングシステムにローカルにアタッチされているように見え得る。こうして、異なるノード 1 1 6、1 1 8 等が、抽象ファイルを明示的に要求するのではなくオペレーティングシステムを通してデータブロックにアクセスすることができる。

10

【 0 0 1 8 】

クラスタ化ネットワーク環境 1 0 0 は等しい数のネットワーク及びディスクモジュールを示すが、他の実施形態が異なる数のこれらモジュールを含んでよいことを十分理解されたい。例えば、複数のネットワーク及びディスクモジュール間で 1 対 1 対応を有さないクラスタ内で相互接続された該ネットワーク及びディスクモジュールがあり得る。すなわち、異なるノードが、異なる数のネットワーク及びディスクモジュールを有することができ、同じノードが、ディスクモジュールとは異なる数のネットワークモジュールを有することができる。

20

【 0 0 1 9 】

さらに、ホスト装置 1 0 8、1 1 0 は、ストレージネットワーキング接続 1 1 2、1 1 4 を通じてクラスタ内のノード 1 1 6、1 1 8 とネットワーク化されることができる。一例として、クラスタにネットワーク化されたそれぞれのホスト装置 1 0 8、1 1 0 は、クラスタ内のノード 1 1 6、1 1 8 のサービス（例えば、データパケットの形式における情報の交換）を要求してもよく、ノード 1 1 6、1 1 8 は、要求されたサービスの結果をホスト装置 1 0 8、1 1 0 に返すことができる。一実施形態において、ホスト装置 1 0 8、1 1 0 は、データストレージシステム 1 0 2、1 0 4 内のノード 1 1 6、1 1 8（例えば、ネットワークホスト）に存在するネットワークモジュール 1 2 0、1 2 2 との間で情報を交換することができる。

30

【 0 0 2 0 】

一実施形態において、データストレージ装置 1 2 8、1 3 0 は、ボリューム（volumes）1 3 2 を含み、ボリューム 1 3 2 は、例えば、データのファイルシステムとしてのディスクドライブ又はディスクアレイ又は他のストレージ（例えば、フラッシュ）への情報の記憶の一実現例である。一例において、ディスクアレイは、全ての従来のハードドライブ、全てのフラッシュドライブ、又は従来のハードドライブとフラッシュドライブの組み合わせを含むことができる。ボリュームは、例えば、ディスクの一部、ディスクの集合、又はディスクの複数部分に及ぶことができ、典型的に、ストレージシステム内のディスク領域上にファイルストレージの全体的論理配置を定義する。一実施形態において、ボリュームは、記憶されたデータを、ボリューム内の階層ディレクトリ構造に存在する 1 つ以上のファイルとして含むことができる。

40

【 0 0 2 1 】

ボリュームは、特定のストレージシステムに関連づけられ得るフォーマットで典型的に構成され、それぞれのボリュームフォーマットは、例えばボリュームがクラスタを形成する能力を提供することなどの、ボリュームに対する機能性を提供する特徴を典型的に含む。例えば、第 1 のストレージシステムがそのボリュームについて第 1 のフォーマットを利用

50

し得る場合に、第2のストレージシステムはそのボリュームについて第2のフォーマットを利用することがある。

【0022】

クラスタ化ネットワーク環境100において、ホスト装置108、110は、データストレージシステム102、104を利用して、データを記憶し、ボリューム132から取り出すことができる。この実施形態において、例えば、ホスト装置108は、データストレージシステム102内のノード116のネットワークモジュール120にデータパケットを送信することができる。ノード116は、ディスクモジュール124を使用してデータストレージ装置128にデータを転送することができ、データストレージ装置128は、ボリューム132Aを含む。こうして、この例において、ホスト装置は、ストレージネットワーク接続112により接続されたデータストレージシステム102を使用して、ボリューム132Aにアクセスしてデータを記憶し、かつ/あるいは取り出すことができる。さらに、この実施形態において、ホスト装置110は、データストレージシステム104（例えば、これはデータストレージシステム102からリモートであり得る）内のノード118のネットワークモジュール122との間でデータを交換することができる。ノード118は、ディスクモジュール126を使用してデータストレージ装置130にデータを転送することができ、これにより、データストレージ装置130に関連づけられたボリューム132Bにアクセスする。

10

【0023】

ストリームを使用してデータを割り振りエリアに選択的に記憶することは、クラスタ化ネットワーク環境100内で実現されてもよいことが十分理解され得る。一例において、割り振りエリアは、データストレージ装置128及び/又はデータストレージ装置130内に定義されてもよい。データは、ストリームを通してデータストレージ装置128及び/又はデータストレージ装置130に選択的に送信されてもよい。ストリームは、こうしたストリームが処理されるべき割り振りエリアに対応するストリーム識別子でタグ付けされてもよい。ストリームを使用してデータを割り振りエリアに選択的に記憶することは、任意タイプのコンピューティング環境に対して及び/又は該コンピューティング環境の間で実現されてもよく、物理装置（例えば、ノード116、ノード118、デスクトップコンピュータ、タブレット、ラップトップ、ウェアラブルデバイス、モバイルデバイス、ストレージ装置、サーバ等）及び/又は（例えば、クラスタ化ネットワーク環境100からリモートの）クラウドコンピューティング環境の間で移転可能（transferrable）であつてもよいことが十分理解され得る。

20

30

【0024】

図2は、データストレージシステム200（例えば、図1の102、104）の一説明例であり、本明細書に記載される手法及び/又はシステムの1つ以上を実現できるコンポーネントの一実施形態のさらなる詳細を提供する。データストレージシステム200は、ノード202（例えば、図1のノード116、118）と、データストレージ装置234（例えば、図1のデータストレージ装置128、130）とを含む。ノード202は、例えば、汎用コンピュータ、又は、ストレージサーバとして動作するように特に構成された何らかの他のコンピューティング装置であってもよい。ホスト装置205（例えば、図1の108、110）は、例えば、ネットワーク216を通じてノード202に接続されて、データストレージ装置234上に記憶されたファイル及び/又は他のデータへのアクセスを提供することができる。一例において、ノード202は、データストレージ装置234内に記憶されたデータへのアクセスをホスト装置205などのクライアント装置に提供するストレージコントローラを含む。

40

【0025】

データストレージ装置234は、大容量ストレージ装置、例えば、ディスクアレイ218、220、222のディスク224、226、228などを含むことができる。本明細書に記載される手法及びシステムは例示的な実施形態により限定されないことが十分理解されよう。例えば、ディスク224、226、228は、任意タイプの大容量ストレージ装

50

置を含んでもよく、これらに限られないが、磁気ディスクドライブ、フラッシュメモリ、及び、例えばデータ（D）及び／又はパリティ（P）情報を含む情報を記憶するように適合された任意の他の同様の媒体が含まれる。

【0026】

ノード202は、システムバス242により相互接続された1つ以上のプロセッサ204、メモリ206、ネットワークアダプタ210、クラスタアクセスアダプタ212、及びストレージアダプタ214を含む。ストレージシステム200は、例えば、リダンダント・アレイ・オブ・インディペンデント（又はインエクスペンシブ）・ディスク（Redundant Array of Independent (or Inexpensive) Disks、RAID）最適化手法を実現してアレイ内の機能しなくなったディスクのデータの再構築処理を最適化することができる、ノード202のメモリ206にインストールされたオペレーティングシステム208をさらに含む。

10

【0027】

オペレーティングシステム208はまた、データストレージシステムの通信と、クラスタファブリック215（例えば、図1の106）にアタッチされているなどでクラスタ化ネットワーク内にあり得る他のデータストレージシステム間の通信とを管理することができる。ゆえに、ノード202、例えばネットワークストレージコントローラなどは、ホスト装置要求に応答して、これらホスト装置要求に従ってデータストレージ装置234（例えば、又はさらなるクラスタ化装置）上のデータを管理することができる。オペレーティングシステム208は、データストレージシステム200上の1つ以上のファイルシステムをしばしば確立することができ、ファイルシステムは、例えば、ファイル及びディレクトリの持続的階層名前空間を実現するソフトウェアコード及びデータ構造を含むことができる。一例として、新しいデータストレージ装置（図示されていない）がクラスタ化ネットワークシステムに追加されるとき、オペレーティングシステム208は、既存のディレクトリツリー内のどこに新しいデータストレージ装置に関連づけられた新しいファイルが記憶されるべきかを知らされる。これは、ファイルシステムを「マウントすること」としてしばしば参照される。

20

【0028】

例示的なデータストレージシステム200において、メモリ206は、関連したソフトウェアアプリケーションコード及びデータ構造を記憶するためにプロセッサ204及びアダプタ210、212、214によりアドレス指定可能なストレージロケーションを含むことができる。プロセッサ204及びアダプタ210、212、214は、例えば、ソフトウェアコードを実行しデータ構造を操作するように構成された処理要素及び／又は論理回路を含んでよい。オペレーティングシステム208は、その一部が典型的にはメモリ206に存在し処理要素により実行され、とりわけストレージシステムにより実現されたファイルサービスをサポートするストレージ動作を呼び出すことにより、ストレージシステムを機能的に編成する。本明細書に記載される手法に属するアプリケーション命令を記憶及び／又は実行するために様々なコンピュータ読取可能媒体を含む他の処理及びメモリ機構が使用されてもよいことが当業者に明らかであろう。例えば、オペレーティングシステムはまた、1つ以上の制御ファイル（図示されていない）を利用して仮想マシンの供給を支援することができる。

30

40

【0029】

ネットワークアダプタ210は、ネットワーク216を通じてデータストレージシステム200をホスト装置205に接続するために必要な機械的、電気的、及びシグナリング回路を含み、ネットワーク216は、とりわけ、ポイントツーポイント接続又は共有媒体、例えばローカルエリアネットワークなどを含んでもよい。ホスト装置205（例えば、図1の108、110）は、アプリケーションを実行するように構成された汎用コンピュータであってもよい。上記されたように、ホスト装置205は、情報配信のクライアント／ホストモデルに従ってデータストレージシステム200と相互作用してもよい。

【0030】

50

ストレージアダプタ 214 は、ノード 202 上で実行しているオペレーティングシステム 208 と協調して、ホスト装置 205 により要求された情報にアクセスする（例えば、ネットワークストレージコントローラにより管理されたストレージ装置上のデータにアクセスする）。情報は、磁気ディスクドライブ、フラッシュメモリ、及び/又は情報を記憶するように適合された任意の他の同様の媒体などの、任意タイプのアタッチされた書き込み可能媒体アレイに記憶されてもよい。例示的なデータストレージシステム 200 において、情報は、ディスク 224、226、228 上のデータブロックに記憶できる。ストレージアダプタ 214 は、ストレージエリアネットワーク（SAN）プロトコル（例えば、スモールコンピュータシステムインターフェース（SCSI）、iSCSI、hyperSCSI、ファイバチャネルプロトコル（FCP））などの入力/出力（I/O）相互接続配置を通じてディスクに結合する I/O インターフェース回路を含むことができる。情報は、ストレージアダプタ 214 により取り出され、必要な場合、システムバス 242 を通じてネットワークアダプタ 210（及び/又は、クラスタ内の別のノードに送信する場合にはクラスタアクセスアダプタ 212）に転送される前に 1 つ以上のプロセッサ 204（又は、ストレージアダプタ 214 自体）により処理され、上記ネットワークアダプタ 210（及び/又はクラスタアクセスアダプタ 212）において、情報はデータパケットにフォーマットされ、ネットワーク 216 を通じてホスト装置 205 に返される（かつ/あるいは、クラスタファブリック 215 を通じてクラスタにアタッチされた別のノードに返される）。

10

【0031】

20

一実施形態において、ディスクアレイ 218、220、222 上の情報のストレージは、ディスク領域の全体的論理配置を定義するディスク 224、226、228 のクラスタを含む 1 つ以上のストレージボリューム 230、232 として実現できる。1 つ以上のボリュームを含むディスク 224、226、228 は、RAID の 1 つ以上のグループとして典型的に編成される。一例として、ボリューム 230 は、ディスクアレイ 218 及び 220 の集合体を含み、ディスクアレイ 218 及び 220 は、ディスク 224 及び 226 のクラスタを含む。

【0032】

一実施形態において、ディスク 224、226、228 へのアクセスを容易にするために、オペレーティングシステム 208 は、情報をディスク上のディレクトリ及びファイルの階層構造として論理的に編成するファイルシステム（例えば、ライトエニウェアファイルシステム）を実現してもよい。この実施形態において、それぞれのファイルは、情報を記憶するように構成されたディスクブロックのセットとして実現されてもよく、これに対し、ディレクトリは、他のファイル及びディレクトリに関する情報が記憶された特別にフォーマットされたファイルとして実現されてもよい。

30

【0033】

このデータストレージシステム 200 内の基礎となる物理構成が何であれ、データが物理及び/又は仮想ボリューム内にファイルとして記憶でき、上記ボリュームは、ファイルシステム識別子（FSID）などのそれぞれのボリューム識別子に関連づけでき、上記ファイルシステム識別子は、一例において長さ 32 ビットとすることができる。

40

【0034】

物理ボリュームは、そのアドレス、アドレス指定可能空間、ロケーション等が変わらない物理ストレージ装置の少なくとも一部、例えば、1 つ以上のデータストレージ装置 234（例えば、リダンダント・アレイ・オブ・インディペンデント（又はインエクスペンシブ）・ディスク（RAID システム））のうち少なくともいくつかなどに対応する。典型的に、物理ボリュームのロケーションは、それにアクセスするために使用されるアドレス（の範囲）が一般に一定のままであるという点で、変わらない。

【0035】

対照的に、仮想ボリュームは、異なる物理ストレージ装置の全く異なる部分の集合体にわたり記憶される。仮想ボリュームは、異なる物理ストレージ装置ロケーションの異なる利

50

用可能部分、例えば、ディスク 2 2 4、2 2 6、及び / 又は 2 2 8 の各々からのいくらかの利用可能な領域などの、集合であってもよい。仮想ボリュームはいずれか 1 つの特定ストレージ装置に「結び付けられ」ないため、仮想ボリュームは抽象化又は仮想化のレイヤを含むと言うことができ、上記抽象化又は仮想化のレイヤは、仮想ボリュームがいくつかの観点でリサイズされ、かつ / あるいは柔軟であることを可能にすることが十分理解されよう。

【 0 0 3 6 】

さらに、仮想ボリュームは、1 つ以上の論理ユニット番号 (L U N) 2 3 8、ディレクトリ 2 3 6、Q t r e e 2 3 5、及びファイル 2 4 0 を含むことができる。とりわけ、これらの特徴、より詳細には L U N は、例えば、データが記憶される全く異なるメモリロケーションが識別され、データストレージユニットとしてグループ化されることを可能にする。そのようなものとして、L U N 2 3 8 は、仮想ボリューム内のデータが集合体の範囲内で記憶される仮想ディスク又はドライブを構成するものとして特徴付けられてもよい。例えば、L U N は、仮想ドライブとしてしばしば参照され、それにより、これらは、ボリュームの様々な部分に記憶されたデータブロックを実際には含むと同時に、汎用コンピュータからのハードドライブをエミュレートする。

10

【 0 0 3 7 】

一実施形態において、1 つ以上のデータストレージ装置 2 3 4 が 1 つ以上の物理ポートを有することができ、各物理ポートはターゲットアドレス (例えば、S C S I ターゲットアドレス) を割り当てできる。データストレージ装置に記憶されたそれぞれのボリュームを表すために、データストレージ装置上のターゲットアドレスが使用されて、1 つ以上の L U N 2 3 8 を識別することができる。ゆえに、例えば、ノード 2 0 2 がストレージアダプタ 2 1 4 を通してボリューム 2 3 0、2 3 2 に接続するとき、ノード 2 0 2 とボリュームの基礎となる 1 つ以上の L U N 2 3 8 との間の接続が作成される。

20

【 0 0 3 8 】

一実施形態において、それぞれのターゲットアドレスは複数の L U N を識別することができ、それにより、ターゲットアドレスは複数のボリュームを表すことができる。例えば、ストレージアダプタ 2 1 4 内の回路及び / 又はソフトウェアとして、又はメモリ 2 0 6 に存在しプロセッサ 2 0 4 により実行される実行可能コードとして実現できる I / O インターフェースは、1 つ以上の L U N 2 3 8 を識別する 1 つ以上のアドレスを使用することによりボリューム 2 3 0 に接続することができる。

30

【 0 0 3 9 】

ストリームを使用してデータを割り振りエリアに選択的に記憶することは、データストレージシステム 2 0 0 に対して実現されてもよいことが十分理解され得る。一例において、割り振りエリアは、1 つ以上のデータストレージ装置 2 3 4 内に定義されてもよい。データは、ストリームを通して 1 つ以上のデータストレージ装置 2 3 4 に選択的に送信されてもよい。ストリームは、こうしたストリームが処理されるべき割り振りエリアに対応するストリーム識別子でタグ付けされてもよい。ストリームを使用してデータを割り振りエリアに選択的に記憶することは、任意タイプのコンピューティング環境に対して及び / 又は該コンピューティング環境の間で実現されてもよく、物理装置 (例えば、ノード 2 0 2、

40

ホスト装置 2 0 5、デスクトップコンピュータ、タブレット、ラップトップ、ウェアラブルデバイス、モバイルデバイス、ストレージ装置、サーバ等) 及び / 又は (例えば、ノード 2 0 2 及び / 又はホスト装置 2 0 5 に対してリモートの) クラウドコンピューティング環境の間で移転可能であってもよいことが十分理解され得る。

【 0 0 4 0 】

ストリームを使用してデータを割り振りエリアに選択的に記憶する一実施形態が、図 3 の例示的な方法 3 0 0 により示される。3 0 2 において、ストレージ装置の第 1 の領域が、第 1 の割り振りエリアとして定義され得る。ストレージ装置の第 2 の領域が、第 2 の割り振りエリアとして定義されてもよい。ストレージ装置に対して任意数の割り振りエリアが定義されてもよく、かつ / あるいは単一の割り振りエリアが任意の数及びタイプのストレ

50

ージ装置にわたり及んでもよい（例えば、第1の記憶媒体の第1の部分、第2の記憶媒体の第2の部分等に及ぶ割り振りエリア、RAID構成の第1のディスク、第2のディスク、及びパリティディスクに及ぶ割り振りエリアなど）ことが十分理解され得る。ストレージ装置は、ソリッドステート装置、フラッシュ装置、区分された（partitioned）ストレージ装置、論理アドレス空間の物理アドレス空間への直接のマッピングがないストレージ装置、ストレージ装置のストレージを仮想化する仮想化レイヤなどの間接レイヤにより使用されるストレージ装置などの、任意タイプのストレージ装置を含んでもよい。一例において、割り振りエリアは、ソリッドステート装置などのストレージ装置の消去ブロック単位の整数倍として定義されてもよい（例えば、データは、複数のセルを含むページ単位でフラッシュメモリに書き込まれ、フラッシュメモリは、複数のページ単位を含むブロック単位と呼ばれるより大きい単位でのみ消去できる）。

10

【0041】

304において、（例えば、ファイルシステムにより）ストレージ装置とのネゴシエーション（negotiation）が容易にされ、第1のストリーム識別子がストレージ装置に対して第1のインジケータとして使用されることを規定して、第1のストリーム識別子でタグ付けされたストリームが第1の割り振りエリアを使用して処理されるべきであると示し得る（例えば、第1のストリーム識別子でタグ付けされたストリームのデータはストレージ装置により第1の割り振りエリア内に記憶されるべきであり、他の割り振りエリア内に記憶されるべきでない）。306において、（例えば、ファイルシステムにより）ストレージ装置とのネゴシエーションが容易にされ、第2のストリーム識別子がストレージ装置に対して第2のインジケータとして使用されることを規定して、第2のストリーム識別子でタグ付けされたストリームが第2の割り振りエリアを使用して処理されるべきであると示し得る。こうして、ストレージ装置は、ファイルシステムによりストリームにタグ付けするために使用されるストリーム識別子に対応する割り振りエリアを使用してこうしたストリームを処理することに同意する。

20

【0042】

308において、特定の特性を有するデータが特定の割り振りエリアを使用して処理されるべきであると規定するポリシーが維持され（maintained）得る。例えば、ポリシーは、第1の特性を有するデータが第1の割り振りエリアを使用して処理されるべきである（例えば、こうしたデータは、第1の割り振りエリアから記憶され、読み出されるべきである）こと、及び第2の特性を有するデータが第2の割り振りエリアを使用して処理されるべきである（例えば、こうしたデータは、第2の割り振りエリアから記憶され、読み出されるべきである）ことを規定してもよい。単一のポリシーが単一の特性に対して又は複数の特性に対して規定されてもよく（例えば、ポリシーは、ホットデータをどこに記憶するか、コールドデータをどこに記憶するか、ランダムにアクセスされるデータをどこに記憶するか、順次アクセスされるデータをどこに記憶するかを規定する）、かつ／あるいは1つ以上のポリシーが個々の特性又は特性のペア（例えば、ホットデータ及びコールドデータに対する第1のポリシー、ユーザデータ及びメタデータに対する第2のポリシー等）に対して規定されてもよいことが十分理解され得る。ポリシーは、こうしたポリシーが適用されるべき割り振りエリアに割り当てられてもよい。ポリシーは、こうしたポリシーが適用されるべき割り振りエリアに対するストリーム識別子を規定してもよい。

30

40

【0043】

一例において、第1の特性はユーザデータ特性に対応してもよく、第2の特性はメタデータ特性に対応してもよい（例えば、メタデータはユーザデータよりさらに頻繁に上書きされることがあり、ゆえに異なるアクセスパターンを有し、別個に記憶されるべきである）。別の例において、第1の特性は第1のデータ頻度アクセス（data frequency access）特性に対応してもよく、第2の特性は第2のデータ頻度アクセス特性に対応してもよい（例えば、ホットデータなどのより頻繁にアクセスされるデータは、コールドデータなどのより頻繁にアクセスされないデータと異なる割り振りエリア内に記憶されてもよい）。別の例において、第1の特性は順次アクセス特性に対応してもよく、第2の特性はランダ

50

ムアクセス特性に対応してもよい（例えば、順次アクセスされるデータは、ランダムにアクセスされるデータと異なる割り振りエリア内に記憶されてもよい）。別の例において、第1の特性は第1のストレージ集合体特性に対応してもよく、第2の特性は第2のストレージ集合体特性に対応してもよい（例えば、第1のクライアントに提供される第1のストレージ集合体のデータは、第2のクライアントに提供される第2のストレージ集合体のデータと異なる割り振りエリア内に記憶されてもよい）。ポリシー内で様々な他の特性が定義されてもよいことが十分理解され得る。

【0044】

一例において、ストレージ装置に対して割り振りエリアのセットが定義される。割り振りエリアのセットのうちの割り振りエリアにポリシーが割り当てられてもよい。割り振りエリアのセットは、ポリシー、各割り振りエリアの利用可能な空き領域の量、及び／又は他のソート基準に基づいて、割り振りエリアのソートされたセットとして動的にソートされる（例えば、書き込みストリームが処理のためにファイルシステムにより受信されたとき、オンザフライで（on-the-fly）ソートされ、かつ／あるいは再ソートされる）（例えば、ユーザデータが第1の割り振りエリア及び第5の割り振りエリア内に記憶できる場合、より多くの利用可能ストレージ領域を有する割り振りエリアがより高くランク付けされ、ゆえに使用されてもよい）。ポリシーのセットは、データの特性に基づいてデータをストリームに割り当てるために使用される。ポリシーのセットは、適切なストリーム識別子でストリームにタグ付けするためにも使用される（例えば、ポリシーは、メタデータが第3の割り振りエリア内に記憶されるべきであると示してもよく、ゆえに、ポリシーは、書き込みストリームのメタデータをストリームに割り当てるために使用され、ポリシーは、該ストリームに第3の割り振りエリアのストリーム識別子でタグ付けするために使用される）。

【0045】

310において、ストレージ装置に書き込むべきデータの書き込みストリームが受信される。一例において、ファイルシステムが書き込みストリームを受信する。データの特性が識別されてもよく、例えば、ユーザデータ、メタデータ、及び／又はランダムにアクセスされるデータなどの他タイプのデータなどである。割り振りエリアのソートセットが評価されて、ユーザデータ及びメタデータを処理するために使用されるべき割り振りエリアを識別してもよい。例えば、ポリシーが、第1の割り振りエリア及び／又は他の割り振りエリアがユーザデータを処理するために使用されるべきであると規定してもよい。上記ポリシー又は異なるポリシーが、第2の割り振りエリア及び／又は他の割り振りエリアがメタデータを処理するために使用されるべきであると規定してもよい。したがって、ターゲット割り振りエリアは、ユーザデータを記憶するための割り振りエリアのソートセットから、ターゲット割り振りエリアがユーザデータに関して閾を超えるソートランクを有すること（例えば、第1の割り振りエリアなどの、ユーザデータを記憶するために使用できる最も高いランクの割り振りエリア）に基づいて選択されてもよい。ターゲット割り振りエリアは、メタデータを記憶するための割り振りエリアのソートセットから、ターゲット割り振りエリアがメタデータに関して閾を超えるソートランクを有すること（例えば、第2の割り振りエリアなどの、メタデータを記憶するために使用できる最も高いランクの割り振りエリア）に基づいて選択されてもよい。

【0046】

312において、書き込みストリームのデータは、対応する割り振りエリアのストリーム識別子でタグ付けされたストリームを通してストレージ装置に提供され得る。例えば、ユーザデータは第1のストリームに割り当てられてもよい。第1のストリームは、ユーザデータを処理するために使用されるべき第1の割り振りエリアに対する第1のストリーム識別子でタグ付けされてもよい。こうして、ストレージ装置が第1のストリームを受信したとき、ストレージ装置は、第1のストリーム識別子に基づいて第1の割り振りエリアを使用して第1のストリームのユーザデータを処理することがわかる。メタデータは第2のストリームに割り当てられてもよい。第2のストリームは、メタデータを処理するために使

10

20

30

40

50

用されるべき第2の割り振りエリアに対する第2のストリーム識別子でタグ付けされてもよい。こうして、ストレージ装置が第2のストリームを受信したとき、ストレージ装置は、第2のストリーム識別子に基づいて第2の割り振りエリアを使用して第2のストリームのメタデータを処理することがわかる。

【0047】

割り振りエリアは、閾未満の量の空き領域を有するように決定されてもよい。一例において、割り振りエリアのポリシーは終了させることができる（例えば、自動的に終了でき、あるいは、ポリシーを終了させることについて示唆がストレージ管理者に提供されてもよい）。別の例において、割り振りエリアは、空き領域の量を増加させるように再定義されてもよく、ポリシーは、割り振りエリアに対して保持されてもよい。

10

【0048】

同様のタイプのデータを一緒に選択的に記憶し、同様でないタイプのデータを別個の割り振りエリアに記憶することは、フラグメンテーションを低減させ、書き込み増幅を低減させ、記憶効率を増加させる。

【0049】

図4は、ストリームを使用してデータを割り振りエリアに選択的に記憶するシステム400の一例を示す。割り振りエリアのセットが、ストレージ装置416に対して、例えばストレージ装置416の記憶媒体418に対して規定され得る。例えば、第1の割り振りエリア420が、仮想ブロック番号(0)から仮想ブロック番号(49)までなどの第1のブロック範囲を包含するものとして定義されてもよい。第2の割り振りエリア422が、仮想ブロック番号(50)から仮想ブロック番号(99)までなどの第2のブロック範囲を包含するものとして定義されてもよい。第3の割り振りエリア424が、仮想ブロック番号(100)から仮想ブロック番号(149)までなどの第3のブロック範囲を包含するものとして定義されてもよい。割り振りエリアは互いに同じ又は異なるサイズを有してもよく、ストレージ装置416のストレージ領域全体又はストレージ領域の一部のみが割り振りエリアを定義するために使用されてもよいことが十分理解され得る。こうして、任意数の割り振りエリアがストレージ装置416に対して定義され得る。

20

【0050】

ポリシー402は、割り振りエリアに割り当てられ得る。一例において、第1のポリシー404が、第1の割り振りエリア420及び/又は他の割り振りエリアに割り当てられてもよい。第1のポリシー404は、ホットデータ（例えば、閾頻度を超えてアクセスされるデータ）が第1の割り振りエリア420及び/又は他の割り振りエリア内に記憶されるべきであると規定してもよい。第2のポリシー406が、第2の割り振りエリア422及び/又は他の割り振りエリアに割り当てられてもよい。第2のポリシー406は、コールドデータ（例えば、閾頻度未満でアクセスされるデータ）が第2の割り振りエリア422及び/又は他の割り振りエリア内に記憶されるべきであると規定してもよい。

30

【0051】

第3のポリシー408が、第3の割り振りエリア424及び/又は他の割り振りエリアに割り当てられてもよい。第3のポリシー408は、ユーザデータ（例えば、ユーザテキストドキュメント）が第3の割り振りエリア424及び/又は他の割り振りエリア内に記憶されるべきであると規定してもよい。第4のポリシー410が、第4の割り振りエリア及び/又は他の割り振りエリアに割り当てられてもよい。第4のポリシー410は、メタデータ（例えば、ポリシーサイズ情報、パートナーストレージコントローラ情報、複製ポリシー情報、バックアップポリシーなどの、ストレージファイルシステムにより維持されるメタデータ）が第4の割り振りエリア及び/又は他の割り振りエリア内に記憶されるべきであると規定してもよい。

40

【0052】

第5のポリシー412が、第5の割り振りエリア及び/又は他の割り振りエリアに割り当てられてもよい。第5のポリシー412は、ランダムアクセスデータが第5の割り振りエリア及び/又は他の割り振りエリア内に記憶されるべきであると規定してもよい。第6の

50

ポリシー 4 1 4 が、第 6 の割り振りエリア及び / 又は他の割り振りエリアに割り当てられてもよい。第 6 のポリシー 4 1 4 は、順次アクセスされるデータが第 6 の割り振りエリア及び / 又は他の割り振りエリア内に記憶されるべきであると規定してもよい。

【 0 0 5 3 】

第 7 のポリシーが、第 7 の割り振りエリア及び / 又は他の割り振りエリアに割り当てられてもよい。第 7 のポリシーは、第 1 の集合体のデータが第 7 の割り振りエリア及び / 又は他の割り振りエリア内に記憶されるべきであると規定してもよい。第 8 のポリシーが、第 8 の割り振りエリア及び / 又は他の割り振りエリアに割り当てられてもよい。第 8 のポリシーは、第 2 の集合体のデータが第 8 の割り振りエリア及び / 又は他の割り振りエリア内に記憶されるべきであると規定してもよい。

10

【 0 0 5 4 】

任意数のポリシーが単一の割り振りエリアに割り当てられてもよく（例えば、第 9 のポリシーが、ホットデータが第 7 の割り振りエリア内に記憶できることを規定し、第 1 0 のポリシーが、メタデータが第 7 の割り振りエリア内に記憶できることを規定する）、一ポリシーが 2 つ以上の割り振りエリアに割り当てられてもよい（例えば、ポリシーが、ホットデータが第 1 の割り振りエリア 4 2 0、第 7 の割り振りエリア、及び第 9 の割り振りエリア内に記憶できることを規定する）ことが十分理解され得る。さらに、一ポリシーが、単一の分類のデータ（例えば、ホットデータ）に適用されてもよく、あるいは複数の分類のデータに適用されてもよい（例えば、ポリシーが、ホットデータをどこに記憶するか、コールドデータをどこに記憶するか、ユーザデータをどこに記憶するか等を規定する）ことが十分理解され得る。

20

【 0 0 5 5 】

図 5 は、ストリームを使用してデータを割り振りエリアに選択的に記憶するシステム 5 0 0 の一例を示す。割り振りエリアは、複数のストレージ装置 5 0 2 にわたり定義され得る。例えば、第 1 の割り振りエリア 5 1 2 が、第 1 のストレージ装置 5 0 4、第 2 のストレージ装置 5 0 6、第 3 のストレージ装置 5 0 8、パリティストレージ装置 5 1 0、及び / 又は他のストレージ装置（例えば、RAID 構成を有するストレージ装置）の第 1 の部分にわたり定義されてもよい。こうして、第 1 の割り振りエリア 5 1 2 に関連づけられたストリーム識別子でタグ付けされたストリーム内のデータは、第 1 の割り振りエリア 5 1 2 内に記憶され、例えば、第 1 のストレージ装置 5 0 4、第 2 のストレージ装置 5 0 6、第 3 のストレージ装置 5 0 8、及び / 又はパリティストレージ装置 5 1 0 のうち 1 つ以上にわたり記憶され得る。

30

【 0 0 5 6 】

第 2 の割り振りエリア 5 1 4 が、第 1 のストレージ装置 5 0 4、第 2 のストレージ装置 5 0 6、第 3 のストレージ装置 5 0 8、パリティストレージ装置 5 1 0、及び / 又は他のストレージ装置の第 2 の部分にわたり定義されてもよい。こうして、第 2 の割り振りエリア 5 1 4 に関連づけられたストリーム識別子でタグ付けされたストリーム内のデータは、第 2 の割り振りエリア 5 1 4 内に記憶され、例えば、第 1 のストレージ装置 5 0 4、第 2 のストレージ装置 5 0 6、第 3 のストレージ装置 5 0 8、及び / 又はパリティストレージ装置 5 1 0 のうち 1 つ以上にわたり記憶され得る。

40

【 0 0 5 7 】

第 3 の割り振りエリア 5 1 6 が、第 1 のストレージ装置 5 0 4、第 2 のストレージ装置 5 0 6、第 3 のストレージ装置 5 0 8、パリティストレージ装置 5 1 0、及び / 又は他のストレージ装置の第 3 の部分にわたり定義されてもよい。こうして、第 3 の割り振りエリア 5 1 6 に関連づけられたストリーム識別子でタグ付けされたストリーム内のデータは、第 3 の割り振りエリア 5 1 6 内に記憶され、例えば、第 1 のストレージ装置 5 0 4、第 2 のストレージ装置 5 0 6、第 3 のストレージ装置 5 0 8、及び / 又はパリティストレージ装置 5 1 0 のうち 1 つ以上にわたり記憶され得る。

【 0 0 5 8 】

任意数の割り振りエリアが、単一のストレージ装置に対して、又は任意数のストレージ装

50

置にわたり定義されてもよいことが十分理解され得る。さらに、割り振りエリアが、単一のストレージ装置内に、又は任意数のストレージ装置にわたり定義されてもよいことが十分理解され得る。

【 0 0 5 9 】

ストリームを使用してデータを割り振りエリアに選択的に記憶する一実施形態が、図 6 の例示的な方法 6 0 0 により示される。6 0 2 において、第 1 のデータ及び第 2 のデータが受信され得る。例えば、ストレージコントローラのファイルシステムが、クライアントアプリケーションから第 1 のデータ及び第 2 のデータを含む書き込みストリームを受信してもよい。一例において、ファイルシステムは、第 1 のデータ及び第 2 のデータが書き込まれるべきストレージ装置のデータを仮想化する仮想化レイヤなどの、間接レイヤに関連づけられてもよい。

10

【 0 0 6 0 】

ポリシーが、第 1 の特性をユーザデータ特性として、第 2 の特性をメタデータ特性として、第 3 の特性をホットデータ特性として、第 4 の特性をコールドデータ特性などで定義してもよい。6 0 4 において、第 1 のデータは、ポリシー内に定義された第 1 の特性を有するものとして識別され得る（例えば、第 1 のデータは、ユーザデータベース内のユーザデータである）。6 0 6 において、第 2 のデータは、ポリシー内に定義された第 2 の特性を有するものとして識別され得る（例えば、第 2 のデータは、ユーザデータベースの複製を管理するためにストレージコントローラにより使用されるメタデータである）。こうして、第 1 のデータはユーザデータとして識別され得、第 2 のデータはメタデータとして識別され得る。

20

【 0 0 6 1 】

6 0 8 において、第 1 のデータは第 1 のストリームに割り当てられる。第 1 のストリームは、ユーザデータの第 1 の特性に対してポリシーにより規定された第 1 のストリーム識別子でタグ付けされる。第 1 のストリーム識別子は、ユーザデータを記憶するためにポリシーにより定義された、ストレージ装置の第 1 の割り振りエリアに関連づけられる（例えば、ファイルシステムとストレージ装置とがネゴシエートし（negotiated）て、第 1 の割り振りエリアを使用して処理されるべきユーザデータのストリームにタグ付けするために第 1 のストリーム識別子が使用されることを決定していてもよい）。

【 0 0 6 2 】

6 1 0 において、第 2 のデータは第 2 のストリームに割り当てられる。第 2 のストリームは、メタデータの第 2 の特性に対してポリシーにより規定された第 2 のストリーム識別子でタグ付けされる。第 2 のストリーム識別子は、メタデータを記憶するためにポリシーにより定義された、ストレージ装置の第 2 の割り振りエリアに関連づけられる（例えば、ファイルシステムとストレージ装置とがネゴシエートして、第 2 の割り振りエリアを使用して処理されるべきメタデータのストリームにタグ付けするために第 2 のストリーム識別子が使用されることを決定していてもよい）。

30

【 0 0 6 3 】

6 1 2 において、第 1 のストリームは、第 1 のストリームが第 1 のストリーム識別子でタグ付けされていることに基づいて、ユーザデータの第 1 のデータを第 1 の割り振りエリアに書き込むためにストレージ装置に送信される。こうして、ユーザデータは、ユーザデータに対して指定されていない他の割り振りエリア内でなく、第 1 の割り振りエリア内に選択的に記憶され得る。

40

【 0 0 6 4 】

6 1 4 において、第 2 のストリームは、第 2 のストリームが第 2 のストリーム識別子でタグ付けされていることに基づいて、メタデータの第 2 のデータを第 2 の割り振りエリアに書き込むためにストレージ装置に送信される。こうして、メタデータは、メタデータに対して指定されていない他の割り振りエリア内でなく、第 2 の割り振りエリア内に選択的に記憶され得る。

【 0 0 6 5 】

50

図 7 は、ストリームを使用してデータを割り振りエリアに選択的に記憶するシステム 7 0 0 の一例を示す。ファイルシステム 7 0 2 又は任意の他のハードウェア若しくはソフトウェアモジュールが、ストレージ装置 7 1 2 内の 1 つ以上の割り振りエリアを定義し得る。例えば、第 1 の割り振りエリア 7 1 4 が、ストレージ装置 7 1 2 の第 1 のブロック範囲（例えば、仮想ブロック番号の第 1 の範囲）に対して定義されてもよい。第 2 の割り振りエリア 7 1 6 が、ストレージ装置 7 1 2 の第 2 のブロック範囲（例えば、仮想ブロック番号の第 2 の範囲）に対して定義されてもよい。第 3 の割り振りエリア 7 1 8 が、ストレージ装置 7 1 2 の第 3 のブロック範囲（例えば、仮想ブロック番号の第 3 の範囲）に対して定義されてもよい。こうして、複数の割り振りエリアが、ストレージ装置 7 1 2 内に、及び / 又は他のストレージ装置にわたり定義されてもよい。

10

【 0 0 6 6 】

ファイルシステム 7 0 2 は、ストレージ装置 7 1 2 とネゴシエートして、データのストリームにタグ付けするためにファイルシステム 7 0 2 が使用することになるストリーム識別子を決定することができる。ストリーム識別子は、ストリーム識別子でタグ付けされたストリームのデータが対応する割り振りエリア内で処理される（例えば、記憶される）べきであるという、ストレージ装置 7 1 2 へのインジケータになる。例えば、第 1 のストリーム識別子が第 1 の割り振りエリア 7 1 4 に対して規定されてもよい。第 2 のストリーム識別子 7 2 4 が第 2 の割り振りエリア 7 1 6 に対して規定されてもよい。第 3 のストリーム識別子 7 2 6 が第 3 の割り振りエリア 7 1 8 に対して規定されてもよい。

【 0 0 6 7 】

ファイルシステム 7 0 2 は、ポリシー 7 0 4 を割り振りエリアに割り当てることができる。例えば、ポリシーが、第 2 の割り振りエリア 7 1 6 及び / 又は第 3 の割り振りエリア 7 1 8 に割り当てられてもよい。ポリシーは、ホットデータ（例えば、閾より高い頻度でアクセスされるデータ）が第 2 の割り振りエリア 7 1 6 内に記憶されるべきであり、ホットデータのストリームが第 2 の割り振りエリア 7 1 6 に対して規定された第 2 のストリーム識別子 7 2 4 でタグ付けされべきであると規定してもよい。こうして、ホットデータは、他の割り振りエリア内でなく、第 2 の割り振りエリア 7 1 6 内に記憶される / 含まれることになる。ゆえに、ホットデータへの頻繁なアクセスから結果として生じるフラグメンテーションが第 2 の割り振りエリア 7 1 6 内に含まれ得、他の割り振りエリアにさらなるフラグメンテーションをもたらさないことになる。

20

【 0 0 6 8 】

ポリシーは、コールドデータ（例えば、閾未満の頻度でアクセスされるデータ）が第 3 の割り振りエリア 7 1 8 内に記憶されるべきであり、コールドデータのストリームが第 3 の割り振りエリア 7 1 8 に対して規定された第 3 のストリーム識別子 7 2 6 でタグ付けされるべきであると規定してもよい。こうして、コールドデータは、他の割り振りエリア内でなく、第 3 の割り振りエリア 7 1 8 内に記憶される / 含まれることになる。ゆえに、有効なデータを宛先セルから空きセルに移動させ、それにより新しいデータをこれら宛先セルに書き込みできるガベージコレクション手法及び / 又は他の手法が、コールドデータを不必要にあちこちに移動させない。

30

【 0 0 6 9 】

割り振りエリアは、割り振りエリアのソートセットへソートされ得る。割り振りエリアは、ポリシー 7 0 4、利用可能な空き領域、及び / 又は他のソート基準に基づいてソートされてもよい。例えば、ポリシーが、ランダムにアクセスされるデータが第 2 の割り振りエリア 7 1 6 及び第 3 の割り振りエリア 7 1 8 内に記憶できることを規定してもよい。ランダムにアクセスされるデータの書き込みストリームが受信されたとき、割り振りエリアは、ランダムにアクセスされるデータをストレージ装置 7 1 2 内に記憶する現在のシナリオに関連して動的にソートされてもよい。第 1 の割り振りエリア 7 1 4 は、第 1 の割り振りエリア 7 1 4 でなく第 2 の割り振りエリア 7 1 6 及び第 3 の割り振りエリア 7 1 8 がランダムにアクセスされるデータを記憶するために使用されるべきであるとポリシーが規定することに基づいて、第 2 の割り振りエリア 7 1 6 より低く、及び第 3 の割り振りエリア 7

40

50

１８より低くランク付けされてもよい。第２の割り振りエリア７１６は、第２の割り振りエリア７１６が第３の割り振りエリア７１８より多くの利用可能な空き領域を有することに基づいて、第３の割り振りエリア７１８より高くランク付けされてもよい。

【００７０】

ファイルシステム７０２は、書き込みストリーム７０６を受信することができる。書き込みストリーム７０６は、ホットデータ７０８（例えば、閾より高い頻度でアクセスされるデータ）、コールドデータ７１０（例えば、閾未満の頻度でアクセスされるデータ）、及び／又は他の特性を有するデータを含んでもよい。ファイルシステム７０２は、ポリシー７０４を利用して、第２の割り振りエリア７１６がホットデータ７０８を記憶するために使用されるべきであると決定する（例えば、割り振りエリアをソートし、最も高いランク付けの割り振りエリアを選択する）ことができる。ファイルシステム７０２は、ホットデータ７０８を第１のストリーム７２０に割り当ててもよい。ファイルシステム７０２は、第１のストリーム７２０に第２の割り振りエリア７１６の第２のストリーム識別子７２４でタグ付けしてもよい。こうして、第１のストリーム７２０は、ストレージ装置７１２に提供される。ストレージ装置７１２は、第１のストリーム７２０が第２のストリーム識別子７２４でタグ付けされていることに基づいて、第２の割り振りエリア７１６を使用して第１のストリーム７２０のホットデータ７０８を処理することになる。

10

【００７１】

ファイルシステム７０２は、ポリシー７０４を利用して、第３の割り振りエリア７１８がコールドデータ７１０を記憶するために使用されるべきであると決定する（例えば、割り振りエリアをソートし、最も高いランク付けの割り振りエリアを選択する）ことができる。ファイルシステム７０２は、コールドデータ７１０を第２のストリーム７２２に割り当ててもよい。ファイルシステム７０２は、第２のストリーム７２２に第３の割り振りエリア７１８の第３のストリーム識別子７２６でタグ付けしてもよい。こうして、第２のストリーム７２２は、ストレージ装置７１２に提供される。ストレージ装置７１２は、第２のストリーム７２２が第３のストリーム識別子７２６でタグ付けされていることに基づいて、第３の割り振りエリア７１８を使用して第２のストリーム７２２のコールドデータ７１０を処理することになる。

20

【００７２】

さらに別の実施形態には、本明細書に提示される手法の１つ以上を実現するように構成されたプロセッサ実行可能命令を含むコンピュータ読取可能媒体が含まれる。これら方法で考えられるコンピュータ読取可能媒体又はコンピュータ読取可能装置の一例示的な実施形態が図８に示され、実現例８００は、コンピュータ読取可能データ８０６が符号化されたコンピュータ読取可能媒体８０８、例えば、コンパクトディスクレコーダブル（ＣＤ Ｒ）、デジタル多用途ディスクレコーダブル（ＤＶＤ Ｒ）、フラッシュドライブ、ハードディスクドライブのプラッタ等を含む。このコンピュータ読取可能データ８０６、例えばゼロ又はイチのうち少なくとも一方を含むバイナリデータなどは、本明細書に説明される原理の１つ以上に従い動作するように構成されたプロセッサ実行可能コンピュータ命令８０４を同様に含む。いくつかの実施形態において、プロセッサ実行可能コンピュータ命令８０４は、方法８０２、例えば、図３の例示的な方法３００のうち少なくともいくつか及び／又は図６の例示的な方法６００のうち少なくともいくつかなどを実行するように構成される。いくつかの実施形態において、プロセッサ実行可能コンピュータ命令８０４は、システム、例えば、図４の例示的なシステム４００のうち少なくともいくつか、図５の例示的なシステム５００のうち少なくともいくつか、及び／又は図７の例示的なシステム７００のうち少なくともいくつかなどを実現するように構成される。多くのこうしたコンピュータ読取可能媒体が、本明細書に提示される手法に従い動作するように企図される。

30

40

【００７３】

本明細書に記載される処理、アーキテクチャ、及び／又は手順は、ハードウェア、ファームウェア、及び／又はソフトウェアにおいて実現できることが十分理解されよう。さらに、本明細書に記載される対策は、ストレージシステムとして具現化され又はストレージシ

50

システムを含む、スタンドアロンコンピュータ又はその一部を含む、任意タイプの専用コンピュータ（例えば、ファイルホスト、ストレージサーバ、及び／又はストレージ供給アプリケーション）及び／又は汎用コンピュータに適用されてもよいことが十分理解されよう。さらに、本明細書における教示は、様々なストレージシステムアーキテクチャに対して構成でき、これらに限られないが、ネットワークアタッチトストレージ環境及び／又はストレージエリアネットワーク、並びにクライアント又はホストコンピュータに直接アタッチされたディスクアセンブリが含まれる。ストレージシステムは、したがって、ストレージ機能を実行するように構成され他の機器又はシステムに関連づけられた任意のサブシステムに追加でこうした配置を含むように広く取られるべきである。

【0074】

いくつかの実施形態において、本開示において記載及び／又は例示された方法は、全体的又は部分的にコンピュータ読取可能媒体上に実現されてもよい。コンピュータ読取可能媒体は、本明細書に提示される方法の1つ以上を実現するように構成されたプロセッサ実行可能命令を含むことができ、コンピュータシステムによりその後読み出せるこのデータを記憶する任意の機構を含んでもよい。コンピュータ読取可能媒体の例には、（ハード）ドライブ（例えば、ネットワークアタッチトストレージ（NAS）を介してアクセス可能）、ストレージエリアネットワーク（SAN）、揮発及び不揮発メモリ、例えば、読取専用メモリ（ROM）、ランダムアクセスメモリ（RAM）、電氣的消去可能プログラマブル読取専用メモリ（EEPROM）、及び／又はフラッシュメモリ、コンパクトディスク読取専用メモリ（CD ROM）、CD R、コンパクトディスクリライタブル（CD RW）、DVD、カセット、磁気テープ、磁気ディスクストレージ、光学又は非光学データストレージ装置、及び／又はデータを記憶するために使用できる任意の他の媒体が含まれる。

【0075】

対象事項が構造的特徴又は方法論動作に固有の言語で記載されたが、別記の特許請求の範囲に定義される対象事項は必ずしも上記された特定の特徴又は動作に限定されないことが理解されるべきである。むしろ、上記された特定の特徴及び動作は、請求項のうち少なくともいくつかを実現する例示的な形式として開示される。

【0076】

様々な実施形態動作が本明細書において提供される。動作のうちいくつか又はすべてが記載された順序は、これら動作が必ず順序依存であることを示すようにみなされるべきではない。代替的な順序付けが、本説明の恩恵を考慮して十分理解されよう。さらに、すべての動作が必ずしも本明細書に提供された各実施形態に存在するわけではないことが理解されよう。また、いくつかの実施形態においてすべての動作が必要なわけではないことが理解されよう。

【0077】

さらに、請求される対象事項は、コンピュータを制御して開示された対象事項を実現するためのソフトウェア、ファームウェア、ハードウェア、又はこれらの任意の組み合わせを生み出すように標準のアプリケーション又はエンジニアリング手法を用いて方法、装置、又は製造品として実現される。本明細書において用いられるときの用語「製造品」は、任意のコンピュータ読取可能装置、キャリア、又は媒体からアクセス可能なコンピュータアプリケーションを包含することが意図される。当然ながら、請求される対象事項の範囲又は主旨から逸脱することなくこの構成に対して多くの変更が行われてもよい。

【0078】

本出願において用いられるとき、用語「コンポーネント」、「モジュール」、「システム」、「インターフェース」などは一般に、コンピュータ関連エンティティ、ハードウェア、ハードウェア及びソフトウェアの組み合わせ、ソフトウェア、又は実行中のソフトウェアを参照することが意図される。例えば、コンポーネントが、プロセッサ上で実行するプロセス、プロセッサ、オブジェクト、実行ファイル、実行のスレッド、アプリケーション、又はコンピュータを含む。例示として、コントローラ上で実行するアプリケーションと

10

20

30

40

50

コントローラとの双方がコンポーネントであり得る。実行のプロセス又はスレッド内に存在する1つ以上のコンポーネントと一コンポーネントとが、1つのコンピュータ上に局所化されてもよく、あるいは2つ以上のコンピュータ間に分散されてもよい。

【0079】

さらに、「例示的」は本明細書において、必ずしも利点としてでなく、例、事例、例示等の役割を果たすことを意味するために用いられる。本出願において用いられるとき、「又は」は、排他的な「又は」でなく包含的な「又は」を意味することが意図される。さらに、本出願において用いられる一の(“a”、“an”)は一般に、単数形に向けられるように別段指定されるか又は文脈から明らかでない限り、「1つ以上」を意味するようにみなされるべきである。また、A及びBのうち少なくとも1つ及び/又は同様のものは一般に、A、又はB、及び/又はA及びBの双方を意味する。さらに、「含める」、「有している」、「有する」、「備える」、又はこれらの変形が用いられる範囲で、こうした用語は、用語「含む」と同様に包含的であることが意図される。

10

【0080】

請求される対象事項の範囲又は主旨から逸脱することなく多くの変更が本開示に対して行われてもよい。別段指定されない限り、「第1の」、「第2の」などは、時間的態様、空間的態様、順序付け等を示すことは意図されない。むしろ、こうした用語は、特徴、要素、アイテム等のための、識別子、名前等として専ら用いられる。例えば、第1の情報セット及び第2の情報セットは一般に、情報セットA及び情報セットB、又は2つの異なる若しくは2つの同一の情報セット、又は一の同じ情報セットに対応する。

20

【0081】

また、本開示は1つ以上の実現例に関して図示され記載されたが、本明細書及び添付図面を読んで理解した当業者は同等の修正及び変更に基づくであろう。本開示は、すべてのこうした変更及び修正を含み、下記の特許請求の範囲によってのみ限定される。特に、上記されたコンポーネント(例えば、要素、リソース等)により実行される様々な機能に関し、こうしたコンポーネントを記載するために用いられる用語は、別段示されない限り、開示された構造に対して構造的に同等でないとしても、(例えば、機能的に同等である)記載されたコンポーネントの指定された機能を実行する任意のコンポーネントに対応することが意図される。さらに、開示の特定の特徴がいくつかの実現例のうち1つのみに関して開示されている可能性があるが、こうした特徴は、所望され得る他の実現例の1つ以上の他の特徴、及び任意の所与の又は特定の適用のための利点と組み合わせられてもよい。

30

上記の実施形態につき以下の付記を残しておく。

〔付記1〕

データを記憶する方法であって、

ストレージ装置に書き込むべき第1のデータ及び第2のデータを受信するステップと、

前記第1のデータを、ポリシー内に定義された第1の特性を有するものとして識別するステップと、

前記第2のデータを、前記ポリシー内に定義された第2の特性を有するものとして識別するステップと、

前記第1のデータを、前記第1の特性に対して前記ポリシーにより規定された第1のストリーム識別子でタグ付けされた第1のストリームに割り当てるステップであり、前記第1のストリーム識別子は、前記ストレージ装置内の第1の割り振りエリアに関連づけられるステップと、

40

前記第2のデータを、前記第2の特性に対して前記ポリシーにより規定された第2のストリーム識別子でタグ付けされた第2のストリームに割り当てるステップであり、前記第2のストリーム識別子は、前記ストレージ装置内の第2の割り振りエリアに関連づけられるステップと、

前記第1のストリームが前記第1のストリーム識別子でタグ付けされていることに基づいて、前記第1のデータを前記第1の割り振りエリアに書き込むために前記第1のストリームを前記ストレージ装置に送信するステップと、

50

前記第 2 のストリームが前記第 2 のストリーム識別子でタグ付けされていることに基づいて、前記第 2 のデータを前記第 2 の割り振りエリアに書き込むために前記第 2 のストリームを前記ストレージ装置に送信するステップと、
を含む方法。

[付記 2]

前記ポリシーは、前記第 1 の特性をユーザデータ特性として、及び前記第 2 の特性をメタデータ特性として規定する、付記 1 に記載の方法。

[付記 3]

前記ポリシーは、前記第 1 の特性を第 1 のデータ頻度アクセス特性として、及び前記第 2 の特性を第 2 のデータ頻度アクセス特性として規定する、付記 1 に記載の方法。

10

[付記 4]

前記ポリシーは、前記第 1 の特性を順次アクセス特性として、及び前記第 2 の特性をランダムアクセス特性として規定する、付記 1 に記載の方法。

[付記 5]

前記ポリシーは、前記第 1 の特性を、前記ストレージ装置に関連づけられた第 1 の集合体の第 1 のストレージ集合体特性として、及び前記第 2 の特性を、前記ストレージ装置に関連づけられた第 2 の集合体の第 2 のストレージ集合体特性として規定する、付記 1 に記載の方法。

[付記 6]

前記第 1 のデータを前記第 1 のストリームに、及び前記第 2 のデータを前記第 2 のストリームに割り当てるファイルシステム、をさらに含む付記 1 乃至 5 のうちいずれか 1 項に記載の方法。

20

[付記 7]

前記ストレージ装置を使用してデータへのアクセスを提供する仮想化レイヤ、をさらに含む付記 1 に記載の方法。

[付記 8]

前記仮想化レイヤは、前記ストレージ装置が論理アドレス対物理アドレスマッピングがない場合、前記データへのアクセスを提供する、付記 7 に記載の方法。

[付記 9]

マシンにより実行されると前記マシンに付記 1 乃至 8 のうちいずれか 1 項に記載の方法を実行させる命令を含む非一時的マシン読取可能媒体。

30

[付記 10]

方法を実行する命令を含む非一時的マシン読取可能媒体であって、前記命令はマシンにより実行されると前記マシンに、

ストレージ装置の記憶特性に基づいて、前記ストレージ装置の第 1 の領域を第 1 の割り振りエリアとして、及び前記ストレージ装置の第 2 の領域を第 2 の割り振りエリアとして定義し、

前記ストレージ装置とネゴシエートして、前記ストレージ装置に対して第 1 のインジケータとして使用される第 1 のストリーム識別子を規定して、前記第 1 のストリーム識別子でタグ付けされたストリームが前記第 1 の割り振りエリアを使用して処理されるべきであると示し、

40

前記ストレージ装置とネゴシエートして、前記ストレージ装置に対して第 2 のインジケータとして使用される第 2 のストリーム識別子を規定して、前記第 2 のストリーム識別子でタグ付けされたストリームが前記第 2 の割り振りエリアを使用して処理されるべきであると示し、

第 1 の特性を有するデータが前記第 1 の割り振りエリアを使用して処理されるべきであり、第 2 の特性を有するデータが前記第 2 の割り振りエリアを使用して処理されるべきであると規定するポリシーを維持し、

前記ストレージ装置に書き込むべきデータの書き込みストリームを受信し、

前記書き込みストリームのデータを、前記データが前記第 2 の特性でなく前記第 1 の特性

50

を有することに基づいて前記第 1 の割り振りエリアを使用して前記ストレージ装置により処理するために、前記第 2 のストリーム識別子でなく前記第 1 のストリーム識別子でタグ付けされたストリームを通して前記ストレージ装置に提供することをさせる非一時的マシン読取可能媒体。

[付記 1 1]

前記命令は前記マシンに、

前記第 1 の割り振りエリア及び前記第 2 の割り振りエリアを、前記ストレージ装置の消去ブロック単位の整数倍として定義する

ことをさせる付記 1 0 に記載の非一時的マシン読取可能媒体。

[付記 1 2]

前記命令は前記マシンに、

前記ストレージ装置に対して割り振りエリアのセットを定義し、

前記割り振りエリアのセットのうちの割り振りエリアにポリシーを割り当て、

前記ポリシーを利用して、前記データの特性に基づいてデータをストリームに割り当て、

前記ストリームにストリーム識別子でタグ付けする

ことをさせる付記 1 0 に記載の非一時的マシン読取可能媒体。

[付記 1 3]

前記命令は前記マシンに、

前記割り振りエリアのセットを、前記ポリシーと前記割り振りエリアの利用可能な空き領域とに基づいて、割り振りエリアのソートセットとしてソートする

ことをさせる付記 1 2 に記載の非一時的マシン読取可能媒体。

[付記 1 4]

前記命令は前記マシンに、

前記割り振りエリアのソートセットからターゲット割り振りエリアを、前記ターゲット割り振りエリアが閾を超えるソートランクを有することに基づいて選択し、

ストリーム識別子を、前記ターゲット割り振りエリアに関連づけられるものとして識別し、

ターゲットストリームを前記ストリーム識別子でタグ付けする

ことをさせる付記 1 3 に記載の非一時的マシン読取可能媒体。

[付記 1 5]

前記命令は前記マシンに、

前記第 1 の割り振りエリアが閾未満の量の空き領域を有すると決定し、

この決定に応答して、前記第 1 の割り振りエリアに関連づけられた第 1 のポリシーを終了させる

ことをさせる付記 1 0 乃至 1 4 のうちいずれか 1 項に記載の非一時的マシン読取可能媒体。

[付記 1 6]

前記命令は前記マシンに、

前記第 1 の割り振りエリアが閾未満の量の空き領域を有すると決定し、

この決定に応答して、前記第 1 の割り振りエリアを再定義して前記空き領域の量を増加させる

ことをさせる付記 1 0 乃至 1 5 のうちいずれか 1 項に記載の非一時的マシン読取可能媒体。

[付記 1 7]

データ記憶のためのコンピューティング装置であって、

命令を記憶させたマシン実行可能コードを含むマシン読取可能媒体を含むメモリと、

前記メモリに結合されたプロセッサと、を含み、

前記プロセッサは、前記マシン実行可能コードを実行して前記プロセッサに付記 1 乃至 8 のうちいずれか 1 項に記載の方法を実行させるように構成される、コンピューティング装置。

[付記 1 8]

データ記憶のためのコンピューティング装置であって、

方法を実行する命令を記憶させたマシン実行可能コードを含むマシン読取可能媒体を含む

10

20

30

40

50

メモリと、

前記メモリに結合されたプロセッサと、を含み、

前記プロセッサは、前記マシン実行可能コードを実行して前記プロセッサに、

ストレージ装置を使用して仮想化レイヤにより提供される仮想化ストレージに書き込むべき第 1 のデータ及び第 2 のデータを受信し、

前記第 1 のデータを、ポリシー内に定義された第 1 の特性を有するものとして識別し、

前記第 2 のデータを、前記ポリシー内に定義された第 2 の特性を有するものとして識別し、

前記第 1 のデータを、前記第 1 の特性に対して前記ポリシーにより規定された第 1 のストリーム識別子でタグ付けされた第 1 のストリームに割り当て、前記第 1 のストリーム識別子は、前記ストレージ装置内の第 1 の割り振りエリアに関連づけられ、

10

前記第 2 のデータを、前記第 2 の特性に対して前記ポリシーにより規定された第 2 のストリーム識別子でタグ付けされた第 2 のストリームに割り当て、前記第 2 のストリーム識別子は、前記ストレージ装置内の第 2 の割り振りエリアに関連づけられ、

前記第 1 のストリームが前記第 1 のストリーム識別子でタグ付けされていることに基づいて、前記第 1 のデータを前記第 1 の割り振りエリアに書き込むために前記第 1 のストリームを前記ストレージ装置に送信し、

前記第 2 のストリームが前記第 2 のストリーム識別子でタグ付けされていることに基づいて、前記第 2 のデータを前記第 2 の割り振りエリアに書き込むために前記第 2 のストリームを前記ストレージ装置に送信する

ことをさせるように構成される、コンピューティング装置。

20

[付記 1 9]

前記ストレージ装置は、第 1 のパーティション及び第 2 のパーティションを有する区分されたストレージ装置である、付記 1 8 に記載のコンピューティング装置。

[付記 2 0]

前記ストレージ装置は、第 1 の記憶媒体及び第 2 の記憶媒体を含む、付記 1 8 に記載のコンピューティング装置。

[付記 2 1]

前記第 1 の割り振りエリアは、前記第 1 の記憶媒体の第 1 の部分と前記第 2 の記憶媒体の第 2 の部分とを含む、付記 2 0 に記載のコンピューティング装置。

[付記 2 2]

30

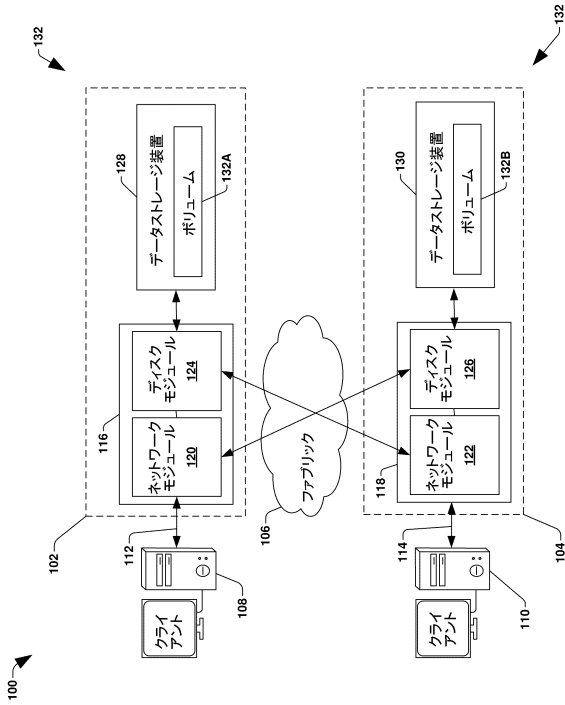
前記第 1 の記憶媒体及び前記第 2 の記憶媒体は R A I D 構成を有し、前記第 1 の割り振りエリアはパリティディスクの第 3 の部分を含む、付記 2 1 に記載のコンピューティング装置。

40

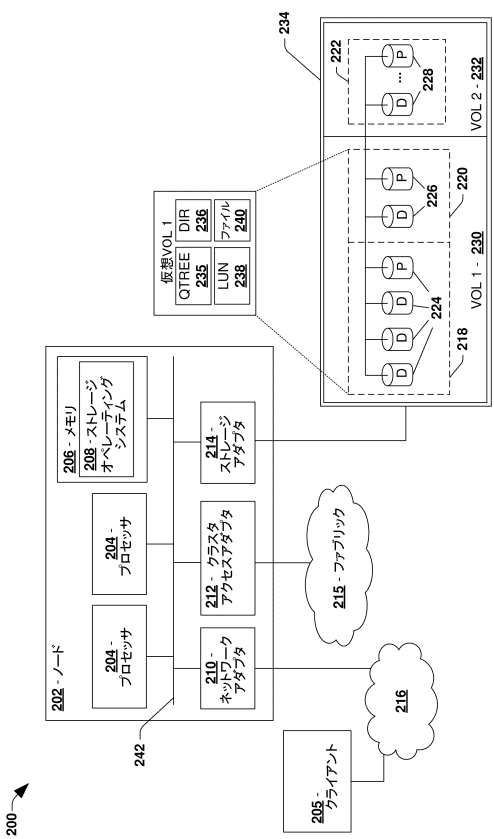
50

【図面】

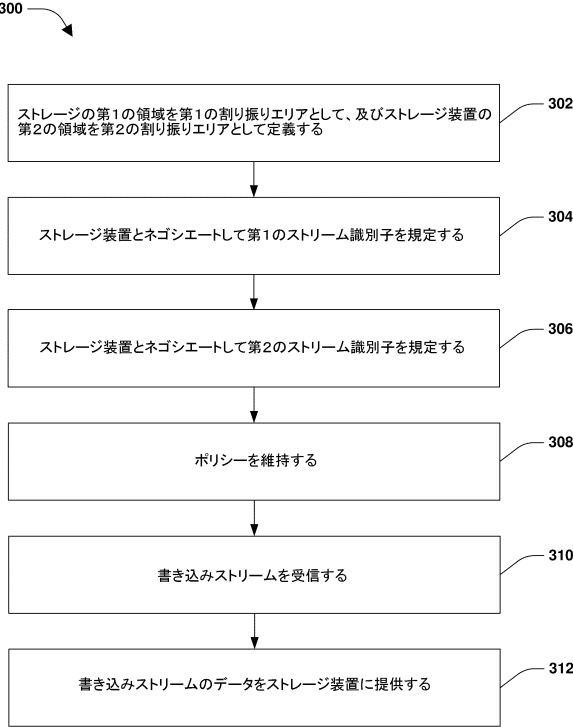
【図 1】



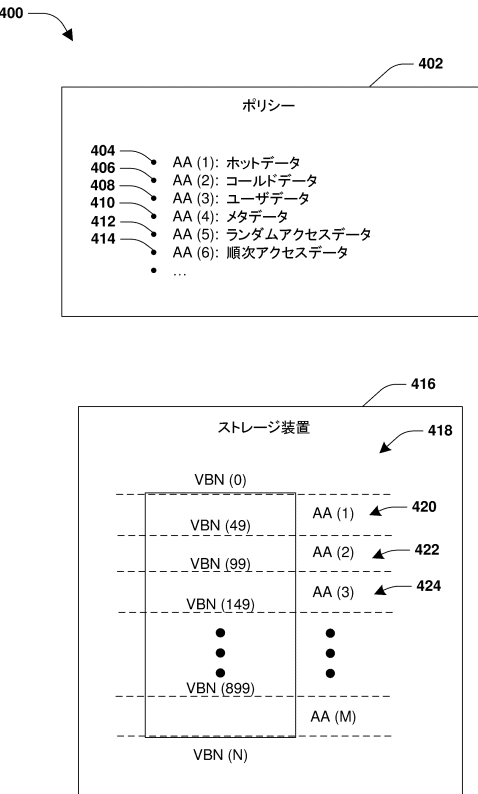
【図 2】



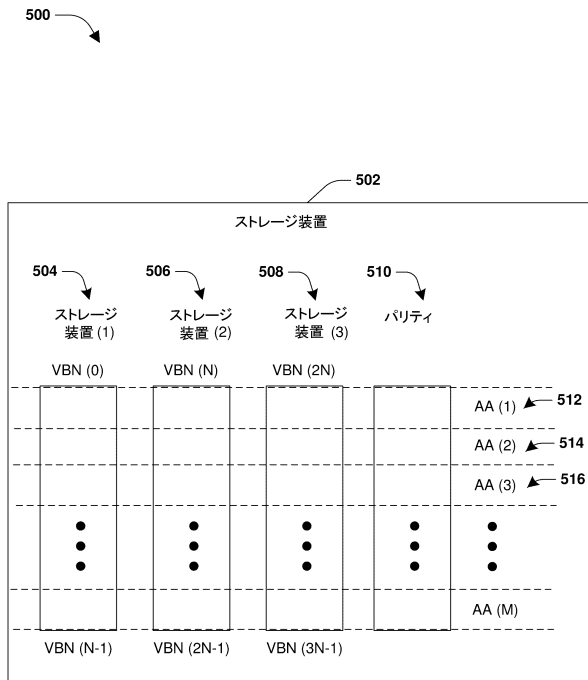
【図 3】



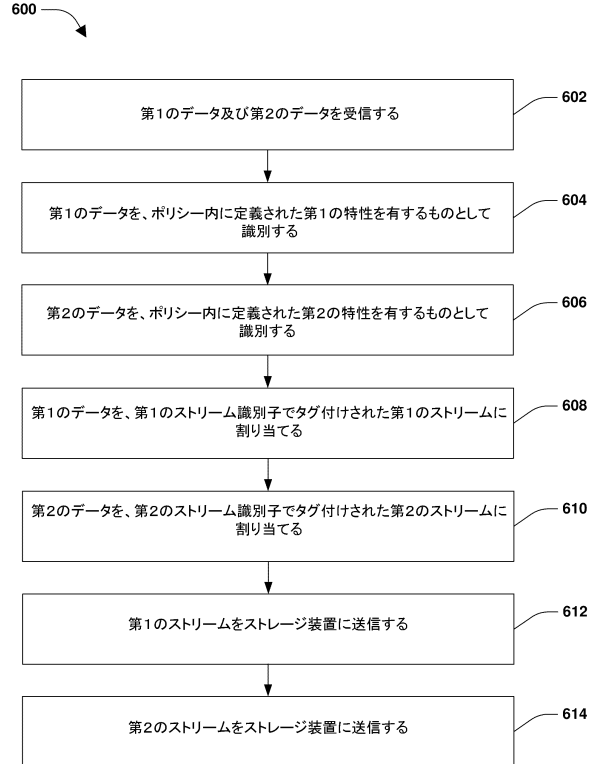
【図 4】



【 図 5 】



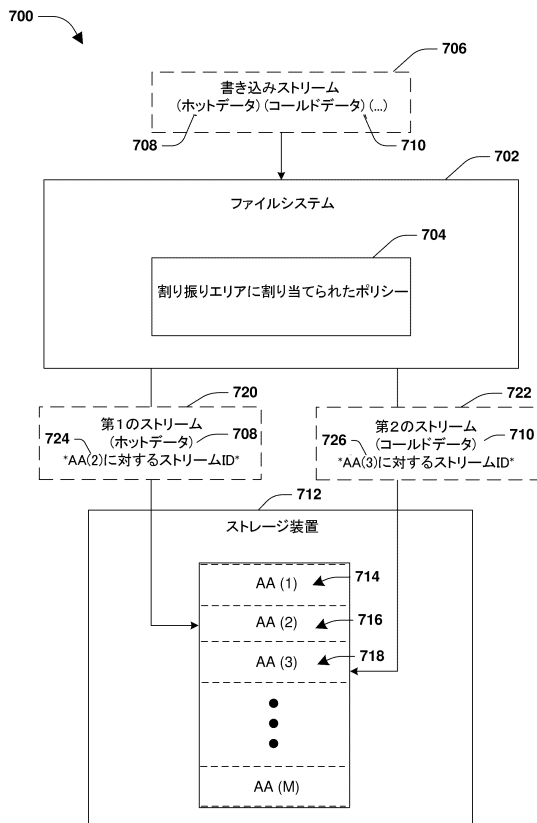
【 図 6 】



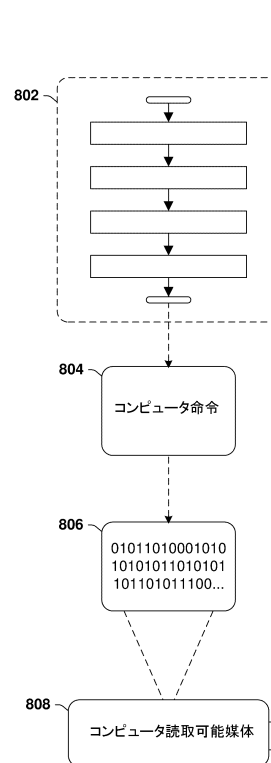
10

20

【 図 7 】



【 図 8 】



30

40

フロントページの続き

- (72)発明者 スターリング, カイル ディグズ
 アメリカ合衆国, カリフォルニア州 9 4 0 8 9 , サニーベール, イースト・ジャバ・ドライブ 4
 9 5
- (72)発明者 バッタチャヤ, ムリナール ケイ .
 アメリカ合衆国, カリフォルニア州 9 4 0 8 9 , サニーベール, イースト・ジャバ・ドライブ 4
 9 5
- (72)発明者 グプタ, モヒト
 アメリカ合衆国, カリフォルニア州 9 4 0 8 9 , サニーベール, イースト・ジャバ・ドライブ 4
 9 5
- 審査官 松平 英
- (56)参考文献 特開 2 0 0 8 - 0 8 4 3 1 7 (J P , A)
 米国特許出願公開第 2 0 1 1 / 0 1 0 7 0 4 2 (U S , A 1)
 特開 2 0 0 8 - 1 7 6 6 0 6 (J P , A)
 特開 2 0 0 9 - 0 4 8 6 1 3 (J P , A)
 特表 2 0 1 1 - 5 1 5 7 2 7 (J P , A)
- (58)調査した分野 (Int.Cl., D B 名)
- G 0 6 F 3 / 0 6 - 3 / 0 8
 1 2 / 0 0
 1 3 / 1 0 - 1 3 / 1 4
 1 5 / 7 8
 1 6 / 0 0 - 1 6 / 9 5 8