(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)特許出願公表番号
特表2016-510458
(P2016-510458A)
(43)公表日　平成28年4月7日(2016.4.7)

(51)Int.Cl.		F I			テーマコード（参考）
G06F 3/06	***(2006.01)***	G06F	3/06	304F	5B005
G06F 12/10	***(2016.01)***	G06F	3/06	301Z	
		G06F	3/06	301F	
		G06F	12/10	501Z	

審査請求 未請求　予備審査請求 未請求　（全 21 頁）

(21)出願番号　特願2015-552744 (P2015-552744)
(86)(22)出願日　平成26年1月8日 (2014.1.8)
(85)翻訳文提出日　平成27年8月18日 (2015.8.18)
(86)国際出願番号　PCT/US2014/010690
(87)国際公開番号　WO2014/110137
(87)国際公開日　平成26年7月17日 (2014.7.17)
(31)優先権主張番号　61/751,142
(32)優先日　平成25年1月10日 (2013.1.10)
(33)優先権主張国　米国 (US)
(31)優先権主張番号　14/046,872
(32)優先日　平成25年10月4日 (2013.10.4)
(33)優先権主張国　米国 (US)

(71)出願人　513076589
ピュア・ストレージ・インコーポレイテッド
アメリカ合衆国・９４０４１・カリフォルニア州・マウンテン　ビュー・カストロ　ストリート・６５０・スイート・２２０
(74)代理人　100064621
弁理士　山川　政樹
(74)代理人　100098394
弁理士　山川　茂樹
(72)発明者　コルグローヴ，ジョン
アメリカ合衆国・９４０２４・カリフォルニア州・ロス　アルトス・ヴィスタ　グランデ　アヴェニュ・７２２

最終頁に続く

(54)【発明の名称】ストレージシステム内でのコピーの実行

(57)【要約】
コピーオフロード動作を実行するためのシステム及び方法。（第１の媒体を指す）第１のボリュームから（第２の媒体を指す）第２のボリューム）へのコピーオフロード動作を要求されると、このコピーオフロード動作は、コピーされるデータにアクセスすることなく実行される。第３の媒体が生成され、第１の媒体はこの第３の媒体の下層媒体として記録される。第１のボリュームは第３の媒体を再び指す。また第４の媒体が生成され、第２のボリュームは第４の媒体を再び指し、第２の媒体は第４の媒体の標的範囲の下層媒体として記録される。第４の媒体の他の全ての範囲はその下層媒体として第２の媒体を有する。
【選択図】図４

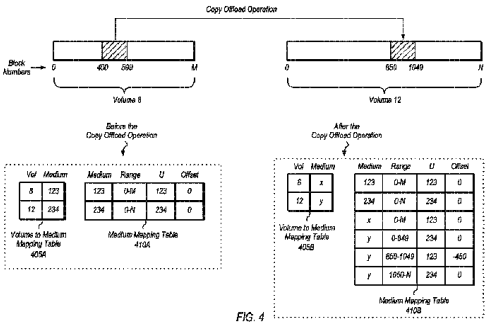

FIG. 4

【特許請求の範囲】
【請求項１】
１つ又は複数のストレージデバイス；及び
前記１つ又は複数のストレージデバイスに接続されたストレージコントローラ
を備える、コンピュータシステムであって、
前記ストレージコントローラは、第１の媒体に関連する第１のボリュームからの第１の範囲のデータを、第２の媒体に関連する第２のボリュームの第２の範囲へコピーするための要求の受信に応答して：
第３の媒体を生成し、前記第１の媒体が前記第３の媒体の下層にあることのインジケータを記憶し、前記第１のボリュームを前記第３の媒体に関連付けるよう；
第４の媒体を生成し、前記第２の媒体が前記第４の媒体の下層にあることのインジケータを記憶し、前記第２のボリュームを前記第４の媒体に関連付けるよう；及び
前記第１の媒体内の前記第１の範囲のデータが、前記第４の媒体の前記第２の範囲の下層にあることのインジケータを記憶する
よう構成される、コンピュータシステム。
【請求項２】
前記第１の範囲のデータをコピーするための前記要求は、前記第１の範囲のデータにアクセスする必要なしに、前記ストレージコントローラによって満たされる、請求項１に記載のコンピュータシステム。
【請求項３】
前記ストレージコントローラは更に、前記第３の媒体の生成を、前記第１のボリュームの前記第１の範囲を標的とする書き込み要求を受信するまで遅延させるよう構成される、請求項１に記載のコンピュータシステム。
【請求項４】
前記ストレージコントローラは更に、前記第４の媒体の生成を、前記第２のボリュームの前記第２の範囲を標的とする書き込み要求を受信するまで遅延させるよう構成される、請求項３に記載のコンピュータシステム。
【請求項５】
前記第１の範囲は、前記第１のボリューム内に第１のオフセットで配置され、
前記第２の範囲は、前記第２のボリューム内に第２のオフセットで配置され、
前記ストレージコントローラは、前記第２のオフセットと前記第１のオフセットとの差を、媒体マッピングテーブル内のエントリに記憶するよう構成され、
前記エントリは前記第４の媒体の前記第２の範囲に対応する、請求項１に記載のコンピュータシステム。
【請求項６】
前記ストレージコントローラは更に：
前記第１の媒体が前記第３の媒体の下層にあることのインジケータを記憶するのに応答して、前記第１の媒体が読み出し専用であることのインジケータを記憶するよう；及び
前記第２の媒体が前記第４の媒体の下層にあることのインジケータを記憶するのに応答して、前記第２の媒体が読み出し専用であることのインジケータを記憶するよう構成される、請求項１に記載のコンピュータシステム。
【請求項７】
前記第１の範囲のデータは第５の媒体内に記憶され、
前記第５の媒体は前記第１の媒体の下層にある、請求項１に記載のコンピュータシステム。
【請求項８】
第１のボリュームからの第１の範囲のデータを、第２のボリュームの第２の範囲へコピーするための要求の受信に応答して：
第３の媒体を生成するステップ；
第１の媒体が前記第３の媒体の下層にあることのインジケータを記憶し、ここで前記第

１のボリュームの前記第１の範囲は前記第１の媒体に関連する、ステップ；
前記第１のボリュームを前記第３の媒体と関連付けるステップ；
第４の媒体を生成するステップ；
前記第１の媒体内の前記第１の範囲のデータが前記第４の媒体の下層にあることのインジケータを記憶するステップ；
第２の媒体が、前記第４の媒体の他の全てのデータの下層にあることの１つ又は複数のインジケータを記憶し、ここで前記第２のボリュームは前記第２の媒体に関連する、ステップ；及び
前記第２のボリュームを前記第４の媒体と関連付けるステップ
を含む、ストレージシステムにおいて使用するための方法。
【請求項９】
前記第１の範囲のデータをコピーするための前記要求を、前記第１の範囲のデータにアクセスすることなく満たすステップを更に含む、請求項８に記載の方法。
【請求項１０】
前記第３の媒体の生成を、前記第１のボリュームの前記第１の範囲を標的とする書き込み要求を受信するまで遅延させるステップを更に含む、請求項８に記載の方法。
【請求項１１】
前記第４の媒体の生成を、前記第２のボリュームの前記第２の範囲を標的とする書き込み要求を受信するまで遅延させるステップを更に含む、請求項１０に記載の方法。
【請求項１２】
前記第１の範囲は、前記第１のボリューム内に第１のオフセットで配置され、
前記第２の範囲は、前記第２のボリューム内に第２のオフセットで配置され、
前記方法は、前記第２のオフセットと前記第１のオフセットとの差を、媒体マッピングテーブル内のエントリに記憶するステップを更に含み
前記エントリは前記第４の媒体の前記第２の範囲に対応する、請求項８に記載の方法。
【請求項１３】
前記第１の媒体が前記第３の媒体の下層にあることのインジケータを記憶するのに応答して、前記第１の媒体が読み出し専用であることのインジケータを記憶するステップ；及び
前記第２の媒体が前記第４の媒体の下層にあることのインジケータを記憶するのに応答して、前記第２の媒体が読み出し専用であることのインジケータを記憶するステップ
を更に含む、請求項８に記載の方法。
【請求項１４】
前記第１の範囲のデータは第５の媒体内に記憶され、
前記第５の媒体は前記第１の媒体の下層にある、請求項１３に記載の方法。
【請求項１５】
プログラム命令を記憶する非一時的コンピュータ可読ストレージ媒体であって、
第１のボリュームからの第１の範囲のデータを、第２のボリュームの第２の範囲へコピーするための要求の受信に応答して：
第３の媒体を生成し；
第１の媒体が前記第３の媒体の下層にあることのインジケータを記憶し、ここで前記第１のボリュームは前記第１の媒体に関連し；
前記第１のボリュームを前記第３の媒体と関連付け；
第４の媒体を生成し；
前記第１の媒体内の前記第１の範囲のデータが前記第４の媒体の下層にあることのインジケータを記憶し；
第２の媒体が、前記第４の媒体の他の全てのデータの下層にあることの１つ又は複数のインジケータを記憶し、ここで前記第２のボリュームは前記第２の媒体に関連し；及び
前記第２のボリュームを前記第３の媒体と関連付ける
ために、前記プログラム命令をプロセッサによって実行できる、コンピュータ可読ストレ

ージ媒体。
【請求項１６】
前記第１の範囲のデータをコピーするための前記要求を、前記第１の範囲のデータにアクセスすることなく満たすために、前記プログラム命令をプロセッサによって更に実行できる、請求項１５に記載のコンピュータ可読ストレージ媒体。
【請求項１７】
前記第３の媒体の生成を、前記第１のボリュームの前記第１の範囲を標的とする書き込み要求を受信するまで遅延させるために、前記プログラム命令をプロセッサによって更に実行できる、請求項１５に記載のコンピュータ可読ストレージ媒体。
【請求項１８】
前記第４の媒体の生成を、前記第２のボリュームの前記第２の範囲を標的とするデータ要求を受信するまで遅延させるために、前記プログラム命令をプロセッサによって更に実行できる、請求項１７に記載のコンピュータ可読ストレージ媒体。
【請求項１９】
前記第１の範囲は、前記第１のボリューム内に第１のオフセットで配置され、
前記第２の範囲は、前記第２のボリューム内に第２のオフセットで配置され、
前記第２のオフセットと前記第１のオフセットとの差を、媒体マッピングテーブル内のエントリに記憶するために、前記プログラム命令をプロセッサによって更に実行でき、
前記エントリは前記第４の媒体の前記第２の範囲に対応する、請求項１５に記載のコンピュータ可読ストレージ媒体。
【請求項２０】
前記第１の媒体が前記第３の媒体の下層にあることのインジケータを記憶するのに応答して、前記第１の媒体が読み出し専用であることのインジケータを記憶し；及び
前記第２の媒体が前記第４の媒体の下層にあることのインジケータを記憶するのに応答して、前記第２の媒体が読み出し専用であることのインジケータを記憶する
ために、前記プログラム命令をプロセッサによって更に実行できる、請求項１５に記載のコンピュータ可読ストレージ媒体。
【請求項２１】
１つ又は複数のストレージデバイス；及び
前記１つ又は複数のストレージデバイスに接続されたストレージコントローラ
を備える、コンピュータシステムであって、
前記ストレージコントローラは、第１のボリューム内の第１の既存の媒体に関連する第１の位置からの第１のデータを、第２の位置へコピーするための要求の受信に応答して：
第１の新規の媒体を生成し；
前記第１の既存の媒体が前記第１の新規の媒体の下層にあることの第１のインジケータを記憶し；
前記第１のボリュームを前記第１の新規の媒体と関連付け；
前記第２の位置が前記第１のボリューム内にある場合、前記第１の媒体内の前記第１の位置が前記第１の新規の媒体内の前記第２の位置の下層にあることの第２のインジケータを記憶し；
前記第２の位置が、第２の既存の媒体に関連する第２のボリューム内にある場合：
第２の新規の媒体を生成し、
前記第２の既存の媒体が前記第２の新規の媒体の下層にあることの第３のインジケータを記憶し；
前記第１のデータが前記第２の新規の媒体の下層にあることの第４のインジケータを記憶し；
前記第２のボリュームを前記第２の新規の媒体と関連付ける
よう構成される、コンピュータシステム。
【請求項２２】
１つ又は複数のストレージデバイス；及び

前記１つ又は複数のストレージデバイスに接続されたストレージコントローラ
を備える、コンピュータシステムであって、
前記ストレージコントローラは、第１のボリューム内の第１の既存の媒体に関連する第１の位置からの第１のデータを、第２の位置へコピーするための要求の受信に応答して：
第１の新規の媒体を生成し；
前記第１の既存の媒体が前記第１の新規の媒体の下層にあることの第１のインジケータを記憶し；
前記第１のボリュームを前記第１の新規の媒体と関連付け；
前記第１の媒体内の前記第１の位置が前記第１の新規の媒体内の前記第２の位置の下層にあることの第２のインジケータを記憶する
よう構成される、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、ストレージシステム内でのコピーオフロード動作の実行に関する。
【背景技術】
【０００２】
データ保守は、製品開発、人的資源、財務又はその他のいずれに関する業務であっても、その操作の重要な部分である。このようなデータ保守の一部として、多様な複数の理由のうちのいずれかにより、データのコピーを生成する必要が生じる場合がある。コピーを生成するための１つのアプローチは、（例えばネットワークを介してストレージシステムに接続された）クライアントがコピープロセスにアクティブに従事することを伴う。このアプローチでは、１つ又は複数のクライアントプロセッサが、コピープロセス全体を通してストレージシステムに搬送されるトランザクションを開始させ、管理する。従ってクライアントプロセッサはコピー動作によってビジー状態となり、一般に他の作業に従事できなくなる。データをコピーするための別のアプローチは、コピープロセスの管理を別の１つ又は複数のプロセッサにオフロードすることを伴う。例えばコピープロセスをストレージシステムにオフロードし、このストレージシステムがコピープロセスを管理してよい。このようなアプローチは「コピーオフロード」と呼ばれることがある。この方法では、クライアントは一般に、コピーを実行している間に他の作業を実行できるよう自由である。コピーオフロード動作は、ストレージシステムがデータ（場合によっては大量のデータ）を、単一のボリューム内又は２つのボリューム間において、ある位置から別の位置へとコピーする動作を意味する。
【０００３】
以上に鑑みて、コピーオフロード動作を効率的に実行するためのシステム及び方法に対する需要が存在する。

【発明の概要】
【発明が解決しようとする課題】
【０００４】
コピーオフロード動作を実行するためのシステム及び方法の様々な実施形態を考える。
【課題を解決するための手段】
【０００５】
ストレージシステムは、ストレージコントローラ及び１つ又は複数のストレージデバイスを含んでよい。このストレージシステムは、１つ又は複数のホストクライアントシステムに接続してよい。一実施形態では、ストレージコントローラはボリューム及び媒体を利用して、ストレージシステム内に記憶されているクライアントデータを追跡する。媒体はデータの論理グループ化として定義され、各媒体は、データの論理グループ化を識別するための識別子を有する。ストレージコントローラは、ボリューム－媒体マッピングテーブルを保守することによって各ボリュームを単一の媒体にマッピングでき、この媒体をそのボリュームのアンカー媒体と呼ぶ。各媒体は、いずれの個数の他の媒体に対してマッピン

グでき、ストレージコントローラはまた、媒体マッピングテーブルを保守することによってマッピングを媒体間で追跡できる。

【０００６】

ホストシステムが、データを第１のボリュームから第２のボリュームへ（又は第１のボリューム内の第１の位置から第１のボリューム内の第２の位置へ）データをコピーするためのコピーオフロード動作を要求すると、ストレージコントローラは、ボリューム－媒体マッピングテーブルを操作するだけでコピーオフロード動作を実行でき、この媒体マッピングテーブルはコピー中のデータにアクセスすることはない。その結果、コピーオフロード動作を迅速に、かつストレージシステムのリソースの利用を最低限とした状態で実行できる。

【０００７】

これらの及びその他の実施形態は、以下の説明及び添付の図面を考察することによって明らかになるであろう。

【図面の簡単な説明】

【０００８】

【図１】図１は、ストレージシステムの一実施形態を示す概略ブロック図である。

【図２】図２は、媒体の有向非巡回グラフ（ｄｉｒｅｃｔｅｄ　ａｃｙｃｌｉｃ　ｇｒａｐｈ：ＤＡＧ）の一実施形態の概略ブロック図である。

【図３】図３は、媒体マッピングテーブルの一実施形態を示す。

【図４】図４は、コピーオフロード動作の一実施形態を示す。

【図５】図５は、コピーオフロード動作の別の実施形態の概略ブロック図である。

【図６】図６は、コピーオフロード動作を実行するための方法の一実施形態を示す概略フロー図である。

【図７】図７は、コピーオフロード動作を実行するための別の方法の一実施形態を示す概略フロー図である。

【発明を実施するための形態】

【０００９】

本発明は様々な修正例及び代替形態を許容するものであるが、例として複数の具体的実施形態を図面に示し、また本明細書において詳細に説明する。しかしながら、図面及びこれに対する詳細な説明は、本出願において開示されている特定の形態に本発明を限定することを意図したものではなく、その反対に、本発明は、添付の請求項によって定義される本発明の精神及び範囲内のあらゆる修正例、均等物、代替例を包含できることを理解されたい。

【００１０】

以下の説明では、本発明の完全な理解を提供するために、多数の具体的な詳細を挙げる。しかしながら、これらの具体的な詳細を用いずに本発明を実施してもよいことを、当業者は認識するべきである。例えば公知の回路、構造、信号、コンピュータプログラム命令、技術については、本発明を不明瞭にするのを回避するために、詳細には示していない。

【００１１】

ここで図１を参照すると、ストレージシステム１００の一実施形態の概略ブロック図が示されている。ストレージシステム１００は、ストレージコントローラ１１０及びストレージデバイス群１３０、１４０を含んでよく、上記ストレージデバイス群１３０、１４０はいずれの個数のストレージデバイス群（又はデータストレージアレイ）の代表例である。図示したように、ストレージデバイス群１３０はストレージデバイス１３５Ａ～Ｎを含み、これらはいずれの個数及びタイプのストレージデバイス（例えばソリッドステートドライブ（ｓｏｌｉｄ－ｓｔａｔｅ　ｄｒｉｖｅ：ＳＤＤ））の代表例である。ストレージコントローラ１１０はクライアントコンピュータシステム１２５に直接接続してよく、またストレージコントローラ１１０は、ネットワーク１２０を介して、クライアントコンピュータシステム１１５から離間した状態でクライアントコンピュータシステム１１５に接続してよい。クライアント１１５、１２５は、システム１００においてデータを記憶しデ

ータにアクセスするためにストレージコントローラ１１０を利用できるいずれの個数のクライアントの代表例である。なお、いくつかのシステムは、ストレージコントローラ１１０に直接又は離間した状態で接続された単一のクライアントしか含まなくてもよい。

【００１２】

ストレージコントローラ１１０は、ストレージデバイス１３５Ａ～Ｎへのアクセスを提供するよう構成されたソフトウェア及び／又はハードウェアを含んでよい。ストレージデバイス群１３０、１４０から離間したものとしてストレージコントローラ１１０を図示しているが、いくつかの実施形態では、ストレージコントローラ１１０をストレージデバイス群１３０、１４０のうちの一方又はそれぞれの中に配置してよい。ストレージコントローラ１１０は、ベースオペレーティングシステム（ｏｐｅｒａｔｉｎｇ　ｓｙｓｔｅｍ：ＯＳ）、ボリュームマネージャ、本明細書で開示される様々な技術を実装するための追加の制御論理を含んでよく、又はこれらに接続してよい。

【００１３】

ストレージコントローラ１１０は実施形態に応じて、いずれの個数のプロセッサを含んでよく、及び／又はいずれの個数のプロセッサ上で実行されてよく、また、単一のホストコンピューティングデバイスを含んでよく、及び／若しくは単一のホストコンピューティングデバイス上で実行されてよく、又は複数のホストコンピューティングデバイスに亘って分散されていてよい。いくつかの実施形態では、ストレージコントローラ１１０は一般に、１つ若しくは複数のファイルサーバ及び／若しくはブロックサーバを含むか、又は１つ若しくは複数のファイルサーバ及び／若しくはブロックサーバ上で実行されてよい。ストレージコントローラ１１０は、デバイスの不良又はデバイス内のストレージ位置の不良によるデータの損失を防止するために、デバイス１３５Ａ～Ｎに亘ってデータを複製するための様々な技術のうちのいずれを使用してよい。ストレージコントローラ１１０は、共通のデータを重複排除することによって、デバイス１３５Ａ～Ｎに記憶されるデータ量を削減するための、様々な重複排除、圧縮又はその他の技術のうちのいずれを利用してよい。

【００１４】

ストレージコントローラ１１０はまた、システム１００内でスナップショットを生成してこれを管理するよう構成してよい。媒体のセットは、ストレージコントローラ１１０によって記録され、保守される。媒体のほとんどは、特定のボリュームが使用中の最新の媒体等の１つ又は複数の選択された媒体を除いて、読み出し専用であってよい。各媒体は論理的に、媒体内の全てのブロックを含む。しかしながら、媒体が生成された時点から媒体が閉鎖された時点までに変化したブロックのみが保存され、これらのブロックに対するマッピングを上記媒体を用いて保守してもよい。

【００１５】

様々な実施形態では、ストレージコントローラ１１０が複数のマッピングテーブルを保守してよい。これらのマッピングテーブルは、媒体マッピングテーブル及びボリューム－媒体間マッピングテーブルを含んでよい。これらのテーブルを利用して、媒体と下層媒体との間のマッピング及びボリュームと媒体との間のマッピングを記録及び保守してよい。ストレージコントローラ１１０はまた、複数のエントリを有するアドレス翻訳テーブルを含んでよく、各エントリは、対応するデータ成分に関する仮想－物理間マッピングを保持する。このマッピングテーブルを用いて、クライアントコンピュータシステム１１５、１２５からストレージデバイス１３５Ａ～Ｎ内の物理的位置へ、論理読み書き要求をマッピングできる。受信した読み書き要求に対応する探索動作の間に、所定の媒体に関連するマッピングから「物理」ポインタ値を読み出してよい。用語「マッピング（ｍａｐｐｉｎｇ）」は、所定の媒体ＩＤ及びブロック番号を物理ポインタ値に変換するアドレス翻訳マッピングテーブルの１つ又は複数のエントリとして定義される。続いてこの物理ポインタ値を用いて、ストレージデバイス１３５Ａ～Ｎ内の物理的位置を配置してよい。なお、物理ポインタ値を用いて、ストレージデバイス１３５Ａ～Ｎのうちの所定のストレージデバイス内の別のマッピングテーブルにアクセスしてもよい。その結果、物理ポインタ値と標的

ストレージ位置との間に１つ又は複数のレベル間接参照を存在させることができる。

【００１６】

なお代替実施形態では、クライアントコンピュータ、ストレージコントローラ、ネットワーク、ストレージデバイス群、データストレージデバイスの個数及びタイプは、図１に示したものに限定されない。様々な時点において、１つ又は複数のクライアントがオフラインで動作できる。更に動作中、ユーザがシステム１００に対して接続、接続解除、再接続を行うのに従って、個々のクライアントコンピュータ接続タイプを変更してよい。更に本明細書に記載のシステム及び方法は、直接接続型ストレージシステム又はネットワーク接続型ストレージシステムに適用してよく、また本明細書に記載の方法の１つ又は複数の態様を実行するよう構成されたホストオペレーティングシステムを含んでよい。多数のこのような代替例が可能であり、考察の対象となる。

【００１７】

ネットワーク１２０は：無線接続；直接ローカルエリアネットワーク（ｌｏｃａｌ ａｒｅａ ｎｅｔｗｏｒｋ：ＬＡＮ）接続；インターネット、ルータ、ストレージエリアネットワーク、イーサネット（登録商標）等の広域ネットワーク（ｗｉｄｅ ａｒｅａ ｎｅｔｗｏｒｋ：ＷＡＮ）接続を含む多様な技術を利用してよい。ネットワーク１２０は、無線式であってもよい１つ又は複数のＬＡＮを備えてよい。ネットワーク１２０は更に、リモートダイレクトメモリアクセス（ｒｅｍｏｔｅ ｄｉｒｅｃｔ ｍｅｍｏｒｙ ａｃｃｅｓｓ：ＲＤＭＡ）ハードウェア及び／若しくはソフトウェア、伝送制御プロトコル／インターネット・プロトコル（ｔｒａｎｓｍｉｓｓｉｏｎ ｃｏｎｔｒｏｌ ｐｒｏｔｏｃｏｌ／ｉｎｔｅｒｎｅｔ ｐｒｏｔｏｃｏｌ：ＴＣＰ／ＩＰ）ハードウェア及び／若しくはソフトウェア、ルータ、リピータ、スイッチ並びに／又はグリッド並びに／又はその他のものを含んでよい。ファイバチャネル、ファイバチャネルオーバイーサネット（Ｆｉｂｒｅ Ｃｈａｎｎｅｌ ｏｖｅｒ Ｅｔｈｅｒｎｅｔ：ＦＣｏＥ）、ｉＳＣＳＩ等のプロトコルをネットワーク１２０において使用してよい。ネットワーク１２０は、伝送制御プロトコル（Ｔｒａｎｓｍｉｓｓｉｏｎ Ｃｏｎｔｒｏｌ Ｐｒｏｔｏｃｏｌ：ＴＣＰ）及びインターネット・プロトコル（Ｉｎｔｅｒｎｅｔ Ｐｒｏｔｏｃｏｌ：ＩＰ）、即ちＴＣＰ／ＩＰ等の、インターネット用に使用される一連の通信プロトコルとインタフェース接続してよい。

【００１８】

クライアントコンピュータ１１５、１２５は、デスクトップパーソナルコンピュータ（ｐｅｒｓｏｎａｌ ｃｏｍｐｕｔｅｒ：ＰＣ）、サーバ、サーバファーム、ワークステーション、ラップトップ、ハンドヘルドコンピュータ、サーバ、パーソナルデジタルアシスタント（ｐｅｒｓｏｎａｌ ｄｉｇｉｔａｌ ａｓｓｉｓｔａｎｔ：ＰＤＡ）、スマートフォン等の、いずれの個数の据置型又は移動体コンピュータの代表例である。一般にクライアントコンピュータシステム１１５、１２５は、１つ又は複数のプロセッサコアを備える１つ又は複数のプロセッサを含む。各プロセッサコアは、事前に定義された汎用命令セットに従って命令を実行するための回路構成を含む。例えばｘ８６命令セットアーキテクチャを選択してよい。あるいはＡＲＭ（登録商標）、Ａｌｐｈａ（登録商標）、ＰｏｗｅｒＰＣ（登録商標）、ＳＰＡＲＣ（登録商標）又は他のいずれの汎用命令セットアーキテクチャを選択してよい。プロセッサコアは、データ及びコンピュータプログラム命令のためのキャッシュメモリサブシステムにアクセスしてよい。上記キャッシュサブシステムは、ランダムアクセスメモリ（ｒａｎｄｏｍ ａｃｃｅｓｓ ｍｅｍｏｒｙ：ＲＡＭ）及びストレージデバイスを備えるメモリ階層に接続してよい。

【００１９】

ここで図２を参照すると、媒体の有向非巡回グラフ（ＤＡＧ）２００を示すブロック図が示されている。また図２は、各ボリュームに関して、ストレージシステムによって使用されている間にボリュームがどの媒体にマッピングされるかを示す、ボリューム－媒体間マッピングテーブル２０５も示す。ボリュームはグラフ２００への考えられるポインタであり得る。

【００２０】
本明細書で使用される用語「媒体（ｍｅｄｉｕｍ）」は、データの論理グループ化として定義される。媒体は、データの論理グループ化を識別するための対応する識別子を有してよい。各媒体はまた、コンテンツ位置、重複排除エントリ及びその他の情報に対する論理ブロック番号のマッピングを含むか、又はこれに関連してよい。一実施形態では、媒体の識別子をストレージコントローラが使用してよいが、媒体の識別子はユーザ管理下でなくてよい。ユーザ（又はクライアント）は、ボリュームＩＤに付随するデータ要求を送信して、この要求がどのデータを標的とするかを特定してよく、ストレージコントローラはボリュームＩＤを媒体ＩＤに対してマッピングして、上記要求を処理する際に媒体ＩＤを使用してよい。

【００２１】
用語「媒体」を用語「ストレージ媒体（ｓｔｏｒａｇｅ　ｍｅｄｉｕｍ）」又は「コンピュータ可読ストレージ媒体（ｃｏｍｐｕｔｅｒ　ｒｅａｄａｂｌｅ　ｓｔｏｒａｇｅ　ｍｅｄｉｕｍ）」と混同してはならない。ストレージ媒体は、データを記憶するために利用される実際の物理デバイス（例えばＳＳＤ、ＨＤＤ）として定義される。コンピュータ可読ストレージ媒体（又は非一時的コンピュータ可読ストレージ媒体）は、プロセッサ又はその他のハードウェアデバイスが実行できるプログラム命令を記憶するよう構成された物理ストレージ媒体として定義される。本明細書に記載の方法及び／又はメカニズムを実装する様々なタイプのプログラム命令を、コンピュータ可読媒体上で搬送又は記憶してよい。プログラム命令を記憶するよう構成された多数のタイプの媒体利用手可能であり、これにはハードディスク、フロッピー（登録商標）ディスク、ＣＤ－ＲＯＭ、ＤＶＤ、フラッシュメモリ、プログラマブルＲＯＭ（Ｐｒｏｇｒａｍｍａｂｌｅ　ＲＯＭ：ＰＲＯＭ）、ランダムアクセスメモリ（ＲＡＭ）、他の様々な形態の揮発性又は不揮発性ストレージが含まれる。

【００２２】
また、用語「ボリューム－媒体間マッピングテーブル（ｖｏｌｕｍｅ　ｔｏ　ｍｅｄｉｕｍ　ｍａｐｐｉｎｇ　ｔａｂｌｅ）」は、単なる単一のテーブルではなく複数のテーブルを指してよいことにも留意されたい。同様に用語「媒体マッピングテーブル（ｍｅｄｉｕｍ　ｍａｐｐｉｎｇ　ｔａｂｌｅ）」も、単なる単一のテーブルではなく複数のテーブルを指してよい。更に、ボリューム－媒体間マッピングテーブル２０５は、ボリューム－媒体間マッピングテーブルの一例にすぎないことにも留意されたい。他のボリューム－媒体間マッピングテーブルは、異なる個数のボリュームのための異なる個数のエントリを有してよい。

【００２３】
各媒体を、３つの結合されたボックスとしてグラフ２００に示し、左のボックスは媒体ＩＤを示し、中央のボックスは下層媒体を示し、右のボックスは媒体のステータス（ＲＯ：読み出し専用（ｒｅａｄ－ｏｎｌｙ））又は（ＲＷ：読み書き（ｒｅａｄ－ｗｒｉｔｅ））を示す。グラフ２００内では、媒体はその下層媒体を指す。例えば媒体２０は媒体１２を指し、これによって媒体１２が媒体２０の下層媒体であることを表す。また媒体１２は媒体１０を指し、媒体１０は媒体５を指し、媒体５は媒体１を指す。いくつかの媒体は、２つ以上の高次の媒体に対する下層媒体となる。例えば３つの別個の媒体（１２、１７、１１）は媒体１０を指し、２つの別個の媒体（１８、１０）は媒体５を指し、２つの別個の媒体（６、５）は媒体１を指す。少なくとも１つの高次媒体に対する下層媒体となる媒体はそれぞれ、読み出し専用のステータスを有する。

【００２４】
グラフ２００の左下の媒体のセットは、線形セットの例である。グラフ２００に示すように、まず媒体３が生成され、続いてスナップショットが実行され、その結果媒体３は安定する（即ちこの時点以降、媒体３内の所定のブロックの探索の結果が常に同一の値を返すことになる）。媒体３を下層媒体として用いて、媒体７が生成される。媒体３が安定した後に書かれるいずれのブロックは、媒体７内にあるものとして標識される。媒体７に対

する探索は、媒体７内に値が見つかる場合は媒体７から値を返すが、媒体７内にブロックが見つからない場合は媒体３を探索することになる。その後媒体７のスナップショットを実行し、媒体７は安定となり、媒体１４が生成される。媒体１４内のブロックの探索により、媒体７、続いて媒体３を検査して、標的論理ブロックを見つける。最後に媒体１４のスナップショットを実行し、媒体１４は安定となり、媒体１５が生成される。グラフ２００のこの時点において媒体１４は、媒体１５へと向かうボリューム１０２への書き込みによって安定となる。

【００２５】

ボリューム－媒体間マッピングテーブル２０５は、ユーザ管理下のボリュームを媒体に対してマッピングする。各ボリュームは、アンカー媒体としても知られる単一の媒体に対してマッピングしてよい。このアンカー媒体は、他の全ての媒体と同様に、それ固有の探索を処理してよい。複数のボリュームが依存する媒体（例えば媒体１０）は、それ自体の複数のブロックを、これらブロックが依存するボリュームとは別個に追跡する。各媒体をブロックの範囲に分解してもよく、各範囲を媒体のＤＡＧ２００内で別個に処理してよい。

【００２６】

ここで図３を参照すると、媒体マッピングテーブル３００の一実施形態を示す。媒体マッピングテーブル３００のいずれの部分又は全体は、ストレージコントローラ１１０及び／又はストレージデバイス１３５Ａ～Ｎのうちの１つ若しくは複数に記憶してよい。ボリューム識別子（ＩＤ）を使用してボリューム－媒体間マッピングテーブル２０５にアクセスし、このボリュームＩＤに対応する媒体ＩＤを決定してよい。次にこの媒体ＩＤを使用して媒体マッピングテーブル３００にアクセスしてよい。なおテーブル３００は媒体マッピングテーブルの単なる一例であり、他の実施形態では、異なる個数のエントリを有する異なる媒体マッピングテーブルを利用してよい。更に他の実施形態では、媒体マッピングテーブルは他の属性を含んでよく、図３に示したものとは異なる方法で整理されてよい。

【００２７】

テーブル３００の左列に示すように、媒体ＩＤによって各媒体を識別してよい。テーブル３００の各エントリに範囲属性も含んでよく、この範囲はデータブロックに関するものであってよい。データのブロックのサイズ（例えば４ＫＢ、８ＫＢ）は、実施形態に応じて変化し得る。媒体は複数の範囲に分解してよく、媒体の各範囲を、固有の属性及びマッピングを有する独立した媒体であるかのように処理してよい。例えば媒体ＩＤ２は２つの別個の範囲を有する。媒体ＩＤ２の範囲０～９９は、テーブル３００内で媒体ＩＤ２の範囲１００～９９９のためのエントリから離間したエントリを有する。

【００２８】

媒体ＩＤ２のこれらの範囲を共に下層媒体ＩＤ１に対してマッピングするが、同一のソース媒体の別個の範囲を異なる下層媒体に対してマッピングすることもできる。例えば媒体ＩＤ３５の別個の範囲を別個の下層媒体に対してマッピングする。例えば媒体ＩＤ３５の範囲０～２９９を、オフセット４００で下層媒体ＩＤ１８に対してマッピングする。これは、媒体ＩＤ３５のブロック０～２９９を媒体ＩＤ１８のブロック４００～６９９に対してマッピングすることを示す。更に媒体ＩＤ３５の範囲３００～４９９を、～３００のオフセットで下層媒体ＩＤ３３に対してマッピングし、媒体ＩＤ３５の範囲５００～８９９を、～４００のオフセットで下層媒体ＩＤ５に対してマッピングする。これらのエントリは、媒体ＩＤ３５のブロック３００～４９９が媒体ＩＤ３３のブロック０～１９９に対してマッピングされ、媒体ＩＤ３５のブロック５００～８９９が媒体ＩＤ５のブロック１００～４９９に対してマッピングされることを示す。なお他の実施形態では、媒体は４つ以上の範囲に分解してよい。

【００２９】

テーブル３００の「状態」列は、ブロックのための探索をより効率的に実行できるようにする情報を記録する。状態「Ｑ」は、媒体が静止状態（ｑｕｉｅｓｃｅｎｔ）であることを示し、「Ｒ」は媒体が登録状態（ｒｅｇｉｓｔｅｒｅｄ）であることを示し、「Ｕ」

は媒体が非マスク状態（ｕｎｍａｓｋｅｄ）であることを示す。静止状態では、テーブル３００内で特定された１つ又は２つの媒体だけに対して探索を実行する。登録状態では、探索を再帰的に実行する。非マスク状態は、探索を基底媒体において実行するべきか、又は探索を下層媒体においてのみ実行するべきかを決定する。いずれのエントリに関してテーブル３００には示されていないが、別の状態「Ｘ」を用いて、ソース媒体が非マッピング状態であることを特定してよい。この非マッピング状態は、ソース媒体が到達可能なデータを全く含まず、破棄できることを示す。この非マッピング状態はソース媒体の範囲に適用してよい。媒体全体が非マッピング状態である場合、媒体ＩＤをシーケンス無効化テーブルに入力して最終的に破棄してよい。

【００３０】

一実施形態では、媒体が生成されると、この媒体が下層媒体を有する場合はこの媒体は登録状態となり、又はこの媒体が既存の状態を有さない新規のボリュームである場合はこの媒体は静止状態となる。媒体に書き込みが行われると、この媒体の一部は非マスク状態となり得、媒体自体及び下層媒体内にマッピングが存在する。これは、単一の範囲を複数の範囲のエントリに分割することによって実行でき、上記複数の範囲のエントリのうちのいくつかは、初期のマスク状態に保持され、その他は非マスク状態として記録される。

【００３１】

更にテーブル３００内の各エントリは、基底属性を含んでよく、これは媒体の基底を示し、これはこの場合ソース媒体自体を指す。各エントリはまた、オフセットフィールドを含んでよく、これは下層媒体に対してソース媒体にマッピングする際にブロックのアドレスに適用するべきオフセットを特定する。これにより媒体は、下層媒体の開始ブロックから下層媒体の頂部に構成されるだけではなく、下層媒体内の他の位置に対してマッピングできる。テーブル３００に示すように、媒体８は５００のオフセットを有し、これは媒体８のブロック０を下層媒体（媒体１）のブロック５００に対してマッピングすることを示す。従って媒体８による媒体１の探索は、５００のオフセットを、その要求の初期ブロック番号に付加することになる。「オフセット」の行により、媒体を複数の媒体で構成できる。例えば一実施形態では、媒体は「ゴールドマスター（ｇｏｌｄ　ｍａｓｔｅｒ）」オペレーティングシステム画像及びＶＭ（仮想マシン）毎のスクラッチスペースからなる。その他の柔軟なマッピングも可能であり、考察の対象となる。

【００３２】

各エントリは下層媒体属性も含み、これはソース媒体の下層媒体を示す。（媒体１と同様に）下層媒体がソース媒体を指す場合、これはソース媒体が下層媒体を有さず、全ての探索はソース媒体のみにおいて実行されることを示す。各エントリは安定属性も含み、「Ｙ」（ｙｅｓ：はい）は媒体が安定であること（又は読み出し専用であること）を示し、「Ｎ」（ｎｏ：いいえ）は媒体が読み書き可能であることを示す。安定である媒体では、媒体内の所定のブロックに対応するデータは変化しないが、このデータを生成するマッピングは変化し得る。例えば媒体２は安定であるが、媒体２のブロック５０は媒体２又は媒体１に記録してよく、媒体２及び媒体１はこの順に論理的に探索されるが、この探索は必要に応じて並列に実行してよい。一実施形態では、媒体がこれ自体以外のいずれの媒体によって下層媒体として使用される場合、この媒体は安定となる。

【００３３】

ここで図４を参照すると、コピーオフロード動作の一実施形態を示す。図４に示すコピーオフロード動作は、ボリューム８のブロック４００～５９９はボリューム１２のブロック８５０～１０４９にコピーされることになることを特定する。ボリューム８及びボリューム１２の論理表現を、連続データ構造として図４に示しているが、ボリューム８及びボリューム１２に対応するデータは、ホストストレージシステムのストレージデバイス全体に亘って、多数の別個の非連続的な位置に配置され得ることを理解されたい。

【００３４】

ストレージコントローラは、このコピーオフロード動作を実行するための要求の受信に応答して、ボリューム８及びボリューム１２がどの媒体を指すかを決定できる。一実施形

態では、ストレージコントローラはボリューム－媒体間マッピングテーブルに問い合わせを行うことによって、ボリューム８及びボリューム１２のアンカー媒体を決定してよい。この議論の目的に関して、（コピーオフロード動作実行前のボリューム－媒体間マッピングテーブルの２つのエントリを表すボリューム－媒体間マッピングテーブル４０５Ａに示すように）ボリューム８のアンカー媒体が媒体１２３である、及びボリューム１２のアンカー媒体が媒体２３４であると考えられる。また媒体マッピングテーブル４１０Ａは、コピーオフロード動作実行前の媒体マッピングテーブルの２つのエントリを表す。

【００３５】

要求されたコピーオフロード動作を実行するために、ストレージコントローラは媒体「ｘ」を生成してよく、媒体ｘの下層媒体として媒体１２３を記録してよい。この時点において、媒体１２３は安定となる。値「ｘ」は現在使用中でないいずれの媒体ＩＤ値であってよい。また、（コピーオフロード動作実行後のボリューム－媒体間マッピングテーブルの２つのエントリを表す）ボリューム－媒体間マッピングテーブル４０５Ｂに示すように、ボリューム８は媒体ｘを指すように更新してよい。更に媒体「ｙ」を生成してよく、媒体ｙの下層媒体として媒体２３４を記録してよい。またボリューム１２は媒体ｙを指すように更新してよい。コピーされる媒体ｙの範囲に関してエントリを生成してよく、このエントリは、その下層（ｕｎｄｅｒｌｙｉｎｇ：Ｕ）媒体として媒体１２３を有してよい。他の範囲の媒体ｙのための他の全てのエントリは、下層媒体として媒体２３４を有してよい。これらのエントリを媒体マッピングテーブル４１０Ｂに示し、これはコピーオフロード動作実行後の媒体マッピングテーブルの一部分を表す。なお媒体マッピングテーブル４１０Ａ－Ｂのエントリは、図４に示しているものに対する追加の情報を含んでよい。

【００３６】

上述のステップに従ってコピーオフロード動作を実行することにより、ストレージコントローラは、実際に問題のデータブロックをコピーすることなく、要求されたコピーオフロード動作を実行できる。その代わりにコピーオフロード動作は、ボリューム－媒体マッピングテーブル及び媒体マッピングテーブルに対して変更を行うだけで遂行される。その結果として、コピーオフロード動作を実装するにあたって即時データ書き込みは実行されず、コピーオフロード動作を最小のリソース利用によって迅速に実行できる。

【００３７】

コピーオフロード動作の上述の説明はまた、他のタイプのコピーオフロード動作を実行する異なる実施形態にも応用できる。例えば別の実施形態では、第１のボリュームの第１の範囲から複数の別個のボリュームに対してコピーオフロード動作を要求してよい。このコピーオフロード動作に関して、媒体「ｘ」及びボリューム８に対応する上述の複数のステップを、ソースボリュームに対して１回実行してよい。媒体「ｙ」及びボリューム１２に対応する複数のステップを、このコピーオフロード動作の標的である各対象ボリュームに関して反復してよく、各対象ボリュームに関して新規の媒体を生成する。

【００３８】

いくつかの実施形態では、コピーオフロード動作を実行するための上述の複数のステップは、コピーオフロード動作を実行するための要求の受信後すぐに実装するのではなく、バッファリングしてよい。複数のコピーオフロード動作をバッファリングして、これを後にバッチ・モードで実行してよい。更にいくつかの実施形態では、ソースボリュームに対応するコピーオフロード動作の複数のステップを、ソースボリュームを標的とする要求をストレージコントローラが受信するまで遅延させてよい。この時点において、ソースボリュームに対応するコピーオフロード動作の一部分を実行してよい。同様に対象ボリュームに対応するコピーオフロード動作の複数のステップを、対象ボリュームを標的とする要求をストレージコントローラが受信するまで遅延させてよい。

【００３９】

ここで図５を参照すると、コピーオフロード動作の別の実施形態のブロック図が示されている。このコピーオフロード動作では、複数のデータブロックのセットが、あるボリューム内の第１の位置から、同じボリューム内の第２の位置へとコピーされている。ボリュ

ーム３５の論理表現を図５に示しており、データは位置２００～４９９から位置１８００～２０９９へとコピーされる。この議論の目的に関して、媒体３５５はテーブル５０５Ａに示すように、ボリューム３５のアンカー媒体であると考えられる。媒体３５５に関するエントリを媒体マッピングテーブル５１０Ａに示しており、コピーオフロード動作実行前の媒体マッピングテーブルの１つのエントリの代表例である。

【００４０】

一実施形態では、このコピーオフロード動作を実行するための要求の受信に応答して、ストレージコントローラは新規の媒体「ｚ」を生成してよい。テーブル５０５Ｂに示すように、媒体ｚはボリューム３５のアンカー媒体として記録してよい。また、媒体ｚに関する３つの別個のエントリを、媒体マッピングテーブル（その一部分をテーブル５１０Ｂに示す）に追加してよい。媒体ｚに関する第１のエントリは、データブロックの範囲０～１７９９のためのものであり、これに対する下層（Ｕ）媒体は媒体３５５として記録される。第１のエントリのためのオフセットは０に設定される。同様に、データブロックの範囲２１００～Ｎのための、媒体ｚに関する第３のエントリは、第１のエントリと同一の属性を有する。第１のエントリ及び第３のエントリはそれぞれ０のオフセットを有し、これは下層媒体（媒体３５５）に対するマッピング時に使用される。媒体ｚに関する第２のエントリは、コピーされたデータが標的とする範囲（１８００～２０９９）に対応する。第２のエントリはまた、その下層媒体として記録された媒体３５５を有する。しかしながらこの第２のエントリは～１６００のオフセットを有し、これにより、コピーオフロード動作において特定されたデータに対応する媒体３５５内の正しい位置に対してマッピングできる。

【００４１】

上述の技術を用いることによって、ストレージコントローラは、問題のデータブロックを物理的にコピーすることなく、要求されたコピーオフロード動作を達成できる。寧ろコピーオフロード動作は、データブロックにアクセスせずにボリューム及び媒体マッピングテーブルを操作するだけで実行される。コピーのために要求されるデータが記憶されている実際の物理ストレージ位置は、このコピーオフロード動作中にアクセスされることはない。

【００４２】

ここで図６に移ると、コピーオフロード動作を実行するための方法６００の一実施形態が示されている。上述のシステム１００に統合されている構成部品（例えばストレージコントローラ１１０）は一般に、方法６００に従って動作し得る。更にこの実施形態のステップを順番に示す。しかしながら別の実施形態では、いくつかのステップは図示したものと異なる順序で実行してよく、いくつかのステップは同時に実行してよく、いくつかのステップは他のステップと組み合わせてよく、またいくつかのステップが存在しなくてもよい。

【００４３】

第１のボリュームの第１の範囲から第２のボリュームの第２の範囲へのコピーオフロード動作を実行するための要求を、ストレージコントローラが受信してよい（ブロック６０５）。この議論の目的に関して、第１のボリュームは第１の媒体に関連する（即ち第１の媒体を指す）と考えられる。また第２のボリュームは第２の媒体を指すと考えられる。一実施形態では、これらの関連はボリューム－媒体マッピングテーブルに問い合わせを行うことによって決定してよい。

【００４４】

この要求の受信に応答して、第３の媒体を生成してよく、媒体マッピングテーブルにおいてこの第３の媒体に関する新規のエントリを生成してよい（ブロック６１０）。様々な実施形態では、新規の媒体の生成プロセスは、この媒体に関する新規のＩＤを生成して、媒体マッピングテーブルにおいて上記媒体に関する新規のエントリを生成することを伴う。第３の媒体の下層媒体として第１の媒体を指定する指示を記憶してよい（ブロック６１５）。一実施形態では、ブロック６１５は、媒体マッピングテーブルの新規のエントリに

第３の媒体の下層媒体として第１の媒体を記録することによって実装してよい。
【００４５】
続いてボリューム－媒体間マッピングテーブルを更新して、第１のボリュームを第３の媒体に関連させてよい（ブロック６２０）。換言すると、第３の媒体を第１のボリュームのアンカー媒体として特定してよい。また、第１の媒体を読み出し専用（即ち安定）として特定するインジケータを記憶してよい（ブロック６２５）。一実施形態では、このインジケータを媒体マッピングテーブル内の対応するエントリに記憶してよい。
【００４６】
更に第４の媒体を生成してよく、媒体マッピングテーブル内でこの第４の媒体に関する新規のエントリを生成してよい（ブロック６３０）。第２の媒体を、第４の媒体の下層媒体として指定してよい（ブロック６３５）。また、第１の媒体の第１の範囲が第４の媒体の第２の範囲の下層となるよう指定するインジケータを記憶してよい（ブロック６４０）。第４の媒体の第２の範囲に関して、別個の範囲のエントリを媒体マッピングテーブル内に生成してよく、この別個の範囲のエントリは、第４の媒体の第２の範囲を第１の媒体の第１の範囲に対してマッピングできる。一実施形態では、オフセット値をこの範囲のエントリに含むことにより、第４の媒体の第２の範囲を第１の媒体の第１の範囲に対して整列させる方法を特定してよい。第２の媒体は、第４の媒体の他の全ての範囲に関して下層媒体として保持され得る。
【００４７】
更に第４の媒体を、第２のボリュームのアンカー媒体として特定してよい（ブロック６４５）。また更に、第２の媒体を読み出し専用と特定するインジケータを記憶してよい（ブロック６５０）。ブロック６５０の後、方法６００を終了してよい。上述の方法を用いて、いずれの数のコピーオフロード動作を並行して実行してよい。なお上述のステップのうちのいくつかを、コピーオフセット動作が作用するボリュームの領域を標的とする読み出し又は書き込み動作が受信されるまで遅延させてよい。
【００４８】
ここで図７を参照すると、コピーオフロード動作のための別の方法７００の一実施形態が示されている。上述のシステム１００に統合されている構成部品（例えばストレージコントローラ１１０）は一般に、方法７００に従って動作し得る。更にこの実施形態のステップを順番に示す。しかしながら別の実施形態では、いくつかのステップは図示したものと異なる順序で実行してよく、いくつかのステップは同時に実行してよく、いくつかのステップは他のステップと組み合わせてよく、またいくつかのステップが存在しなくてもよい。
【００４９】
コピーオフロード動作を実行するための要求をストレージコントローラが受信してよい（ブロック７０５）。この議論の目的に関して、コピーオフロード動作の目標は、第１のボリュームの第１の範囲から第２のボリュームの第２の範囲にコピーされることになるデータであると考えられる。しかしながら他の動作は、異なるソース及び／又は対象を有してよい。要求の受信に応答して、受信したコピーオフロード要求に対応する動作は、即座に実行するのではなくバッファリングしてよい（ブロック７１０）。様々な実施形態では、複数の動作を別個にバッファリングしてよい。例えば第１のボリュームの第１の範囲を後続の要求が標的とした時に、第１のボリュームに対応する動作を生成し、バッファリングし、実行準備状態としてよい。同様に、第２のボリュームの第２の範囲が標的とされた時に、第２のボリュームに対応する動作を生成し、バッファリングし、実行準備状態とする。
【００５０】
ブロック７１０の後、ストレージコントローラが、過去に受信したコピーオフロード動作が作用する領域を標的とする読み出し又は書き込み要求を受信したかどうかを決定してよい（条件ブロック７１５）。作用される位置を標的とする要求が受信されていなかった場合（条件ブロック７１５の「いいえ（Ｎｏ）」分岐）、ストレージコントローラは、１

つ又は複数のバッファリングされたコピーオフロード動作の実行を防止できる（ブロック７２０）。ブロック７２０の後、別のコピーオフロード要求を受信したかどうかを決定してよい（ブロック７２５）。コピーオフロード要求が受信されていた場合（条件ブロック７２５の「はい（Ｙｅｓ）」分岐）、方法７００はブロック７１０に戻り、受信したコピーオフロード要求を実行できる。なお、ストレージコントローラは、コピーオフロード要求を（ブロック７２５が表す時点だけではなく）いずれの時点において受信してよく、この場合方法７００はこれに従ってブロック７１０にジャンプして、受信したコピーオフロード要求をバッファリングしてよい。コピーオフロード要求が受信されていなかった場合（条件７２５の「いいえ」分岐）、方法７００はブロック７１５に戻り、バッファリングされたコピーオフロード要求に対応する領域を標的とする要求を受信したかどうかを決定してよい。

【００５１】

作用される位置を標的とする要求が受信されていた場合（条件ブロック７１５の「はい」分岐）、ストレージコントローラによって対応するコピーオフロード動作を実行してよい（ブロック７３０）。バッファリングされたコピーオフロード動作は、ソースボリューム及び対象ボリュームを標的としてよく、受信された要求がこれらのボリュームのうちの１つのみを標的とする場合、上記作用されるボリュームを標的とするコピーオフロード動作のみを実行してよい。コピーオフロード動作のその他の部分はバッファリングされたままであってよく、（対応するボリューム内における上記部分の位置が標的とされた場合、又は処理リソースがアイドル状態であり、使用できる状態である場合に）後の時点で実行してよい。ブロック７３０の後、方法７００はブロック７２５に戻り、別のコピーオフロード要求を受信したかどうかを決定してよい。

【００５２】

作用される領域（又はボリューム）を標的とする後続の要求が受信されるまで、コピーオフロード動作の実行を待機することによって、ストレージシステムの処理リソースを、他のタスクを実行できるよう自由にすることができる。コピーオフロード動作のバッファリングはまた、追加の媒体の生成を、それが実際に必要となるまで防止することによって補助できる。ストレージシステムに不要な負荷を与えることなく、複数のコピーオフロード動作を受信してバッファリングできる。また、ストレージシステムのリソースが利用可能な期間中、ストレージコントローラはアイドル状態の処理性能を用いて、多数のバッファリングされたコピーオフロード動作を実行してよい。このようにして、ストレージシステムが実行している他のタスクに干渉することなく、コピーオフロード動作を実行できる。従っていくつかの実施形態では、ブロック７１０が指示するように、全ての受信したコピーオフロード動作をバッファリングするのではなく、ストレージコントローラは、受信したコピーオフロード動作がケースバイケースでバッファリングされているかどうかを決定してよい。この決定は、少なくともストレージシステムの現在の実行条件（例えば処理負荷、ストレージ利用、保留中の要求の個数）に基づいてよい。他の実施形態では、受信したコピーオフロード動作を自動的にバッファリングしてよく、またバッファリングされるコピーオフロード動作の個数が閾値を超えると、ストレージコントローラは、バッチ・モードで複数のコピーオフロード動作を実行してよい。これらの実施形態では、作用される領域を標的とするデータ要求を受信すると、対応するバッファリングされたコピーオフロード動作を、その他のコピーオフロード動作がバッファリングされたままの状態で実行してよい。

【００５３】

なお上述の実施形態はソフトウェアを備えてよい。このような実施形態では、本方法及び／又は機構を実装するプログラム命令を、コンピュータ可読媒体上で搬送又は記憶してよい。プログラム命令を記憶するよう構成された多数のタイプの媒体が利用可能であり、これにはハードディスク、フロッピーディスク、ＣＤ－ＲＯＭ、ＤＶＤ、フラッシュメモリ、プログラマブルＲＯＭ（ＰＲＯＭ）、ランダムアクセスメモリ（ＲＡＭ）、他の様々な形態の揮発性又は不揮発性ストレージが含まれる。

【００５４】

様々な実施形態において、本明細書に記載の方法及び機構の１つ又は複数の部分は、クラウドコンピューティング環境の一部を形成してよい。このような実施形態では、１つ又は複数の様々なモデルに従って、必要に応じてインターネットを介してリソースを提供してよい。このようなモデルは、インフラストラクチャ・アズ・ア・サービス（Ｉｎｆｒａｓｔｒｕｃｔｕｒｅ ａｓ ａ Ｓｅｒｖｉｃｅ：ＩａａＳ）、プラットフォーム・アズ・ア・サービス（Ｐｌａｔｆｏｒｍ ａｓ ａ Ｓｅｒｖｉｃｅ：ＰａａＳ）、ソフトウェア・アズ・ア・サービス（Ｓｏｆｔｗａｒｅ ａｓ ａ Ｓｅｒｖｉｃｅ：ＳａａＳ）を含んでよい。ＩａａＳでは、コンピュータインフラストラクチャを必要に応じて提供する。このような場合、一般にサービスプロバイダがコンピュータ設備を所有し、動作させる。ＰａａＳモデルでは、サービスプロバイダが、開発者がソフトウェアによる解決法を開発するために使用するソフトウェアツール及びその下層設備を必要に応じて提供し、そのホストとなってよい。ＳａａＳは典型的には、サービスに応じてソフトウェアを認可するサービスプロバイダを含む。このサービスプロバイダはソフトウェアのホストとなってよく、又はソフトウェアを所定の期間だけカスタマに対して配備してよい。上述のモデルの多数の組み合わせが可能であり、考察の対象となる。

【００５５】

以上の実施形態についてかなり詳細に説明したが、上述の開示を完全に理解すれば、当業者には多数の変形例及び修正例が明らかとなるであろう。以下の請求項は、これらの変形例及び修正例の全てを包含するものと解釈されることを意図したものである。

【図１】

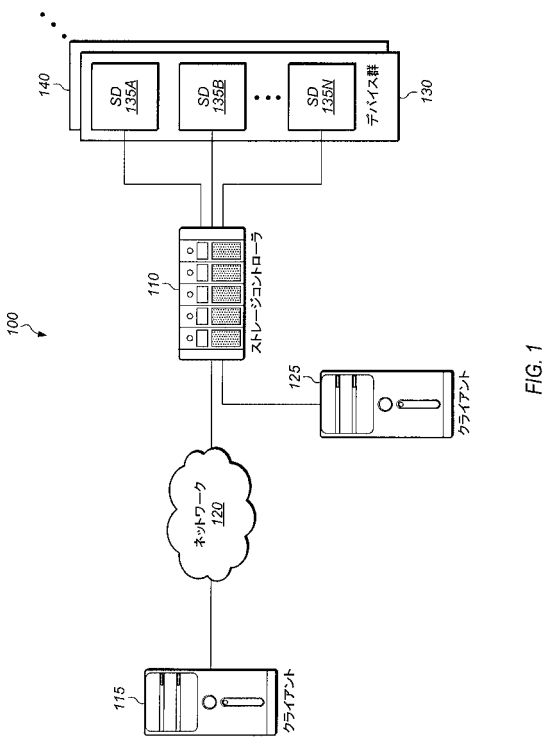

FIG. 1

【図２】

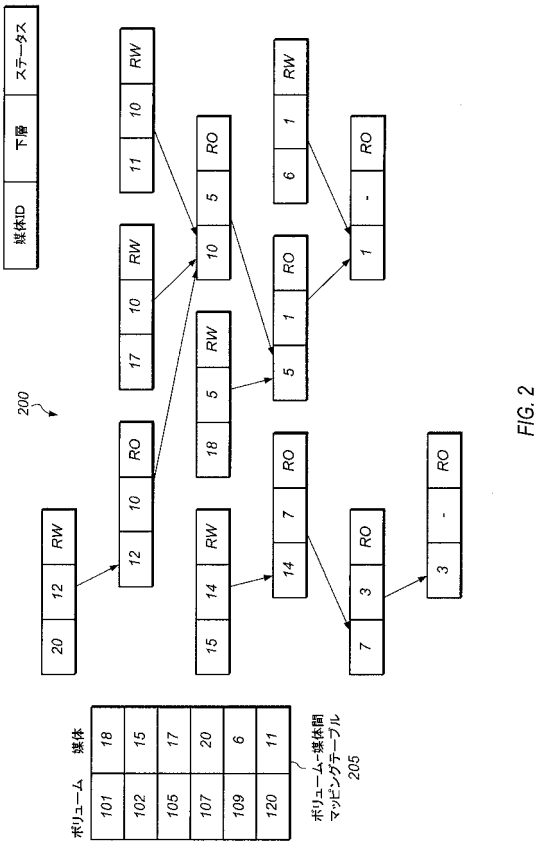

FIG. 2

【図３】

300

媒体ID	範囲	状態	基底	オフセット	下層	安定
1	0-999	Q	1	0	1	Y
2	0-99	QU	2	0	1	Y
2	100-999	Q	2	0	1	Y
5	0-999	RU	5	0	2	N
8	0-499	R	8	500	1	N
10	0-999	QU	10	0	1	Y
14	0-999	RU	14	0	10	Y
18	0-999	RU	18	0	14	N
25	0-999	RU	25	0	14	Y
33	0-999	RU	33	0	25	N
35	0-299	RU	35	400	18	N
35	300-499	RU	35	-300	33	Y
35	500-899	RU	35	-400	5	N

Q – 静止状態;　R – 登録状態;　U - 非マスク状態

FIG. 3

【図４】

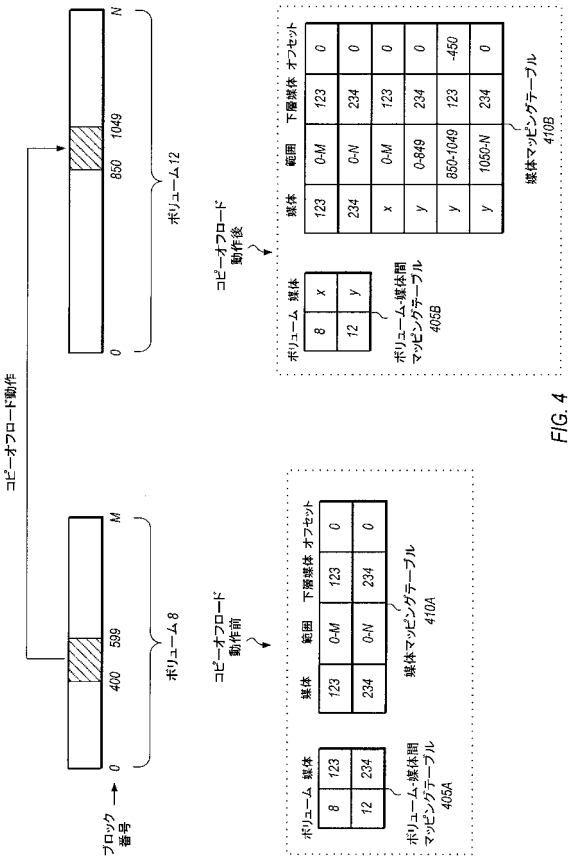

FIG. 4

【図５】

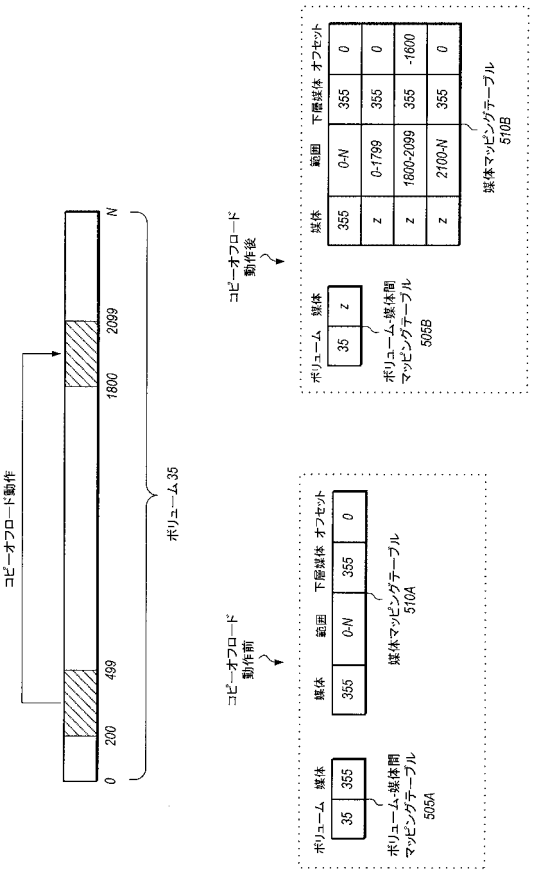

FIG. 5

【図６】

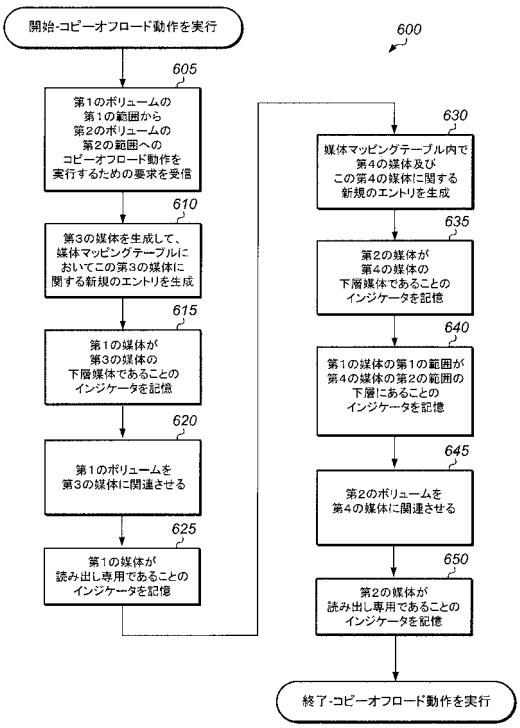

FIG. 6

【図７】

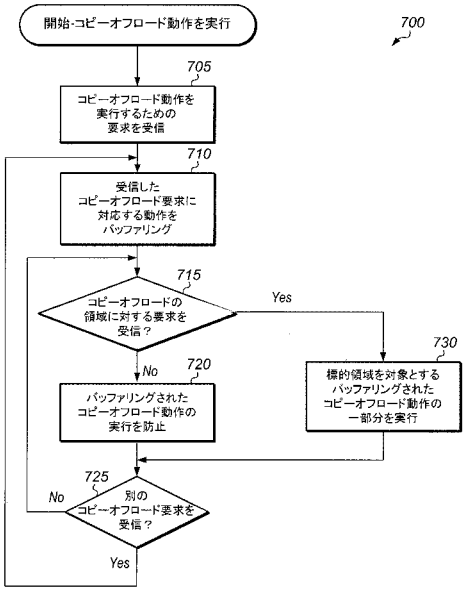
開始-コピーオフロード動作を実行
700
705
コピーオフロード動作を
実行するための
要求を受信
710
受信した
コピーオフロード要求に
対応する動作を
バッファリング
715
コピーオフロードの
領域に対する要求を
受信？
Yes
730
標的領域を対象とする
バッファリングされた
コピーオフロード動作の
一部分を実行
No
720
バッファリングされた
コピーオフロード動作の
実行を防止
725
No
別の
コピーオフロード要求を
受信？
Yes

FIG. 7

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2014/010690

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F3/06
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 8 290 911 B1 (JANAKIRAMAN VISWESVARAN [US] ET AL) 16 October 2012 (2012-10-16) figures 3-6 column 7, line 17 - column 10, line 33	1-22
Y	US 2006/174074 A1 (BANIKAZEMI MOHAMMAD [US] ET AL) 3 August 2006 (2006-08-03) figures 2,4 paragraph [0024] paragraph [0026]	1-22
A	US 2012/330903 A1 (PERIYAGARAM SUBRAMANIAM [US] ET AL) 27 December 2012 (2012-12-27) paragraph [0075] - paragraph [0083] figures 8-10	1-22

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
11 March 2014	24/03/2014

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Andlauer, J

1

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2014/010690

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 8290911 B1	16-10-2012	NONE	
US 2006174074 A1	03-08-2006	CA 2593289 A1	10-08-2006
		CN 101120305 A	06-02-2008
		EP 1853992 A2	14-11-2007
		JP 2008529187 A	31-07-2008
		US 2006174074 A1	03-08-2006
		WO 2006083327 A2	10-08-2006
US 2012330903 A1	27-12-2012	US 2012330903 A1	27-12-2012
		WO 2012177318 A1	27-12-2012

フロントページの続き

(81)指定国 AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,CZ,DE,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,JP,KE,KG,KN,KP,KR,KZ,LA,LC,LK,LR,LS,LT,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,SM,ST,SV,SY,TH,TJ,TM,TN,TR,TT,TZ,UA,UG,US

（特許庁注：以下のものは登録商標）
１．ＥＴＨＥＲＮＥＴ

(72)発明者 ミラー，イーサン
アメリカ合衆国・９５０６０・カリフォルニア州・サンタクルーズ・カルカー ドライブ・２０３
(72)発明者 ヘイズ，ジョン
アメリカ合衆国・９４０４１・カリフォルニア州・マウンテン ビュー・ハイスクール ウェイ・８００・ナンバー・３３０
(72)発明者 サンドヴィグ，ケイリー
アメリカ合衆国・９４３０３・カリフォルニア州・パロ アルト・ドノホー ストリート・２８４
(72)発明者 ゴールデン，クリストファー
アメリカ合衆国・９４０４１・カリフォルニア州・マウンテン ビュー・ハイスクール ウェイ・８００・ナンバー・２２９
(72)発明者 カオ，ジャンティン
アメリカ合衆国・９４０４３・カリフォルニア州・マウンテン ビュー・サン カルロス アヴェニュ・７９１
(72)発明者 イノゼムツェフ，グリゴリ
アメリカ合衆国・９４０８５・カリフォルニア州・サニーベイル・エスカロン アヴェニュ・１０００・ナンバー３０２０

Ｆターム(参考) 5B005 MM51