

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6767115号  
(P6767115)

(45) 発行日 令和2年10月14日 (2020. 10. 14)

(24) 登録日 令和2年9月23日 (2020. 9. 23)

(51) Int. Cl.	F I
<b>G06F 3/06 (2006.01)</b>	G06F 3/06 301Z
<b>G06F 16/13 (2019.01)</b>	G06F 3/06 301F
	G06F 16/13 110

請求項の数 14 (全 19 頁)

(21) 出願番号	特願2015-552796 (P2015-552796)	(73) 特許権者	511175211
(86) (22) 出願日	平成26年1月10日 (2014. 1. 10)		ピュア ストレージ, インコーポレイテッド
(65) 公表番号	特表2016-508275 (P2016-508275A)		アメリカ合衆国 カリフォルニア 940
(43) 公表日	平成28年3月17日 (2016. 3. 17)		41-2055, マウンテン ビュー,
(86) 国際出願番号	PCT/US2014/011003		カストロ ストリート 650, スイート 400
(87) 国際公開番号	W02014/110343	(74) 代理人	100079108
(87) 国際公開日	平成26年7月17日 (2014. 7. 17)		弁理士 稲葉 良幸
審査請求日	平成29年1月10日 (2017. 1. 10)	(74) 代理人	100109346
審判番号	不服2019-2478 (P2019-2478/J1)		弁理士 大貫 敏史
審判請求日	平成31年2月25日 (2019. 2. 25)	(74) 代理人	100117189
(31) 優先権主張番号	61/751, 142		弁理士 江口 昭彦
(32) 優先日	平成25年1月10日 (2013. 1. 10)	(74) 代理人	100134120
(33) 優先権主張国・地域又は機関	米国 (US)		弁理士 内藤 和彦

最終頁に続く

(54) 【発明の名称】 ボリューム動作のための安全装置

(57) 【特許請求の範囲】

【請求項 1】

ストレージシステムにおいて使用するための方法であって、前記方法は、  
 ストレージコントローラによって、  
 前記ストレージシステム上の論理ストレージを表す論理ボリュームを復元するための要求を受信することと、

前記論理ボリュームを復元するための前記要求を受信することに応答して、

前記ストレージシステムに記憶されたデータに現在マッピングしている論理データ表現に、過去にマッピングされていた前記論理ボリューム、を決定することと、

前記ストレージシステムに記憶されたデータに現在マッピングしている論理データ表現に、過去にマッピングされていた前記論理ボリューム、を決定することに応じて、

前記論理ボリュームの第2の論理データ表現を生成し、

前記第2の論理データ表現が前記論理データ表現にマッピングされていることのインジケータを記憶し、

ボリューム - 論理データ表現マッピングテーブル内に新規のエントリを生成すること、を含むことであって、前記新規のエントリは、前記論理ボリュームが前記第2の論理データ表現にマッピングされていることの第2のインジケータを含む、ことによって、

前記論理ボリュームを前記論理データ表現に論理的に再マッピングすることにより、前記論理ボリュームと、前記ストレージシステムに記憶された前記データとの間の接続を復元することと、

10

20

を備える方法。

【請求項 2】

さらに、削除ボリュームテーブルの中で、前記論理ボリュームが前記論理データ表現に過去にマッピングされていたことのインジケータを記憶することを備え、前記論理データ表現は、ある時点における前記ストレージシステムに記憶されたデータを指す、請求項 1 に記載の方法。

【請求項 3】

さらに、

前記論理ボリュームを削除するための要求を受信することと、

前記論理ボリュームを削除するための前記要求を受信することに応答して、

前記論理ボリュームがマッピングされている特定の論理データ表現を決定することと、

前記論理ボリュームが、前記論理ボリュームの前記特定の論理データ表現に過去にマッピングされていたことのインジケータを記憶することと、

前記論理ボリュームが有効であることのいずれのインジケータを除去することと、  
を備える、請求項 1 に記載の方法。

【請求項 4】

さらに、

論理ボリュームと、前記論理ボリュームの論理データ表現との関係を追跡するために、ボリューム - 論理データ表現マッピングテーブルを維持することを備え、前記論理ボリュームが有効であることのいずれのインジケータを除去することは、前記ボリューム - 論理データ表現マッピングテーブルの中で、前記論理ボリュームに対応するエントリを除去することを備える、請求項 3 に記載の方法。

【請求項 5】

さらに、前記論理ボリュームを削除するための前記要求を受信することに応答して、前記削除の動作後、第 1 の期間だけ前記特定の論理データ表現を保持することを備える、請求項 3 に記載の方法。

【請求項 6】

1 つ以上のストレージデバイスと、

前記 1 つ以上のストレージデバイスに結合されたストレージコントローラと、  
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、

ストレージシステム上の論理ストレージを表す論理ボリュームを復元するための要求を受信し、

前記論理ボリュームを復元するための前記要求を受信することに応じて、

前記ストレージシステムに記憶されたデータに現在マッピングしている論理データ表現に、過去にマッピングされていた前記論理ボリューム、を決定し、

前記ストレージシステムに記憶されたデータに現在マッピングしている論理データ表現に、過去にマッピングされていた前記論理ボリューム、を決定することに応答して、

前記論理ボリュームの第 2 の論理データ表現を生成し、

前記第 2 の論理データ表現が前記論理データ表現にマッピングされていることのインジケータを記憶し、

ボリューム - 論理データ表現マッピングテーブル内に新規のエントリを生成することを含むことであって、前記新規のエントリは、前記論理ボリュームが前記第 2 の論理データ表現にマッピングされていることの第 2 のインジケータを含む、ことによって、

前記論理ボリュームを前記論理データ表現に論理的に再マッピングすることにより、前記論理ボリュームと、前記ストレージシステムに記憶された前記データとの間の接続を復元する、

ように構成されている、コンピュータシステム。

【請求項 7】

前記ストレージコントローラはさらに、削除ボリュームテーブルの中で、前記論理ボリ

10

20

30

40

50

ュームが前記論理データ表現に過去にマッピングされていたことのインジケータを記憶することを備え、前記論理データ表現は、ある時点における前記ストレージシステムに記憶されたデータを指すように構成されている、請求項 6 に記載のコンピュータシステム。

【請求項 8】

前記ストレージコントローラはさらに、  
前記論理ボリュームを削除するための要求を受信し、  
前記論理ボリュームを削除するための前記要求を受信することに応答して、  
前記論理ボリュームがマッピングされている特定の論理データ表現を決定し、  
前記論理ボリュームが前記論理ボリュームの前記特定の論理データ表現に過去にマッピングされていたことのインジケータを記憶し、  
前記論理ボリュームが有効であることのいずれのインジケータを除去する、  
ように構成されている、請求項 6 に記載のコンピュータシステム。

10

【請求項 9】

前記ストレージコントローラは、さらに、  
論理ボリュームと、前記論理ボリュームの論理データ表現との関係を追跡するために、  
ボリューム - 論理データ表現マッピングテーブルを維持するように構成され、前記論理  
ボリュームが有効であることのいずれのインジケータを除去することは、前記ボリューム  
- 論理データ表現マッピングテーブルの中で、前記論理ボリュームに対応するエントリを  
除去することを備える、請求項 8 に記載のコンピュータシステム。

20

【請求項 10】

前記ストレージコントローラはさらに、  
前記論理ボリュームを削除するための前記要求を受信することに応答して、前記削除  
の動作後、第 1 の期間だけ前記特定の論理データ表現を保持するように構成される、請求  
項 8 に記載のコンピュータシステム。

【請求項 11】

プログラム命令を記録したコンピュータ読み取り可能な記録媒体であって、前記プログラ  
ム命令は、プロセッサによって、

論理ボリュームを復元するための要求を受信し、  
ストレージシステム上の論理ストレージを表す前記論理ボリュームを復元するための  
前記要求を受信することに応答して、

30

前記ストレージシステムに記憶されたデータに現在マッピングしている論理データ  
表現に、過去にマッピングされていた前記論理ボリューム、を決定し、

前記ストレージシステムに記憶されたデータに現在マッピングしている論理データ  
表現に、過去にマッピングされた前記論理ボリューム、を決定することに応じて、

前記論理ボリュームの第 2 の論理データ表現を生成し、

前記第 2 の論理データ表現が前記論理データ表現にマッピングされていることのイン  
ジケータを記憶し、

ボリューム - 論理データ表現マッピングテーブル内に新規のエントリを生成するこ  
とを含むことであって、前記新規のエントリは、前記論理ボリュームが前記第 2 の論理デ  
ータ表現にマッピングされていることの第 2 のインジケータを含む、ことによって、

40

前記論理ボリュームを、前記論理データ表現に論理的に再マッピングすることにより  
、前記論理ボリュームと前記ストレージシステムに記憶された前記データとの間の接続を  
復元する、

ために実行できる、コンピュータ読み取り可能な記録媒体。

【請求項 12】

前記プログラム命令は、削除ボリュームテーブルの中で、前記論理ボリュームが前記論  
理データ表現に過去にマッピングされていたことのインジケータを記憶することを、プロ  
セッサによって実行でき、前記論理データ表現は、ある時点における前記ストレージシ  
ステムに記憶されたデータを指す、請求項 11 に記載のコンピュータ読み取り可能な記録媒  
体。

50

## 【請求項 1 3】

前記プログラム命令は、プロセッサによって、  
前記論理ボリュームを消去するための要求を受信し、  
前記論理ボリュームを消去するための前記要求を受信したことに応答して、  
前記論理ボリュームがマッピングされている特定の論理データ表現を決定し、  
前記論理ボリュームが、前記論理ボリュームの前記特定の論理データ表現に過去にマッピングされていたことのインジケータを記憶し、  
前記論理ボリュームが有効であることのいずれのインジケータを除去するために実行できる、請求項 1 1 に記載のコンピュータ読み取り可能な記録媒体。

## 【請求項 1 4】

前記プログラム命令は、プロセッサによって、  
論理ボリュームと、前記論理ボリュームの論理データ表現との関係を追跡するために、  
ボリューム - 論理データ表現マッピングテーブルを維持するために実行でき、前記論理ボリュームが有効であることのいずれのインジケータを除去することは、前記ボリューム - 論理データ表現マッピングテーブルの中で、前記論理ボリュームに対応するエントリを除去することを備える、請求項 1 3 に記載のコンピュータ読み取り可能な記録媒体。

## 【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削除されたボリュームを復元することに関する。

## 【背景技術】

## 【0002】

コンピュータメモリストレージ及びデータ帯域幅が増大するに従って、業務において日常的に管理されるデータの量及び複雑性も増大する。データセンタ等の大規模分散型ストレージシステムは典型的には、多数の業務動作を実行する。サーバルームと呼ばれることもあるデータセンタは、1つ又は複数の業務に付随するデータの記憶、管理及び配布のための、物理又は仮想中央レポジトリである。分散型ストレージシステムは、1つ又は複数のネットワークによって相互接続されたクライアントコンピュータに接続してよい。分散型ストレージシステムのいずれの部分の性能が不良である場合、会社の運営を害し得る。従って分散型ストレージシステム、データ可用性及び高性能機能化に関する高い基準を維持する。

## 【0003】

論理ボリュームマネージャ又はディスクアレイマネージャ等のソフトウェアアプリケーションは、マストレージアレイ上にスペースを割り当てる手段を提供する。更にこのソフトウェアにより、システム管理者は、論理ボリュームを含むストレージグループのユニットを生成できるようになる。ストレージ仮想化により、物理ストレージから論理ストレージを抽出し、これによりエンドユーザが物理ストレージを識別することなく論理ストレージにアクセスできる。

## 【0004】

ストレージ仮想化を支援するために、ボリュームマネージャは、エンドユーザからの論理アドレスを用いた入ってくる入力/出力(I/O)要求を、ストレージデバイス内の物理位置に関連するアドレスを用いた新規の要求に翻訳することによって、I/Oリダイレクションを実行する。いくつかのストレージデバイスは、ソリッドステートストレージデバイス内で使用してよいアドレス翻訳層といった、追加のアドレス翻訳機構を含んでよい。ため、論理アドレスから上述の別のアドレスへの翻訳は、唯一の又は最終的なアドレス翻訳を表さない場合がある。

## 【0005】

多くのストレージシステムに関して、ボリューム動作は頻繁に実行され得る。例えば多数のボリュームは毎日生成及び削除され得る。ユーザは時にボリュームを不注意に削除してしまい、そのボリュームを復元したいことに気付くことがある。しかしながら従来のス

10

20

30

40

50

ストレージシステムでは、ボリュームを削除してしまうと、削除したボリュームを復元することは不可能である。

【0006】

以上に鑑みて、削除されたボリュームを復元できる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】

【0007】

ボリューム反転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】

【0008】

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。

【0009】

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、アンカー媒体と下層媒体との間の関係を追跡でき、各下層媒体はスナップショットを表す。

【0010】

一実施形態では、ストレージコントローラを、ボリューム動作を反転できるよう構成してよい。ストレージコントローラは第1のボリュームを削除するための第1の要求を受信してよい。第1のボリュームはそのアンカー媒体として第1の媒体を有してよい。第1の要求の受信に応答して、ストレージコントローラは、第1のボリュームと第1の媒体との間のリンクを除去してよい。ストレージコントローラは第1の期間に亘って、又は第1の媒体が利用しているストレージ空間が、他のデータを記憶するために必要となるまで、第1の媒体の削除を遅らせてよい。上記第1の期間はプログラム可能であってよく、実施形態によって変化し得る。ストレージコントローラはまた、第1のボリュームと第1の媒体との間の過去のリンクのインジケータを、ユーザがこのリンクを復元したい場合に備えて一時的に記憶してよい。

【0011】

後の時点において、ユーザが第1のボリュームを復元したい場合がある。ユーザは第1のボリュームを復元するための第2の要求を生成して、この第2の要求をストレージコントローラに送信してよい。一実施形態では、第2の要求の受信に応答して、ストレージコントローラは第1のボリュームと第1の媒体との間のリンクを復元してよい。別の実施形態では、第2の要求の受信に応答して、ストレージコントローラは第2の媒体を生成し、第1の媒体が第2の媒体の下層にあることのインジケータを記憶し、その後第1のボリュームと第2のボリュームとの間のリンクを生成してよい。この実施形態では、再接続されるボリュームに対する変更を、接続解除に先行する変更と区別できる。

【0012】

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】

【0013】

【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)

10

20

30

40

50

p h : D A G ) 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、ボリュームを削除及び復元するための動作の一実施形態を示す。

【図 5】図 5 は、削除されたボリュームを復元するための代替動作の一実施形態を示す。

【図 6】図 6 は、ボリューム動作を取り消す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、削除されたボリュームを復元するための代替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】

【 0 0 1 4 】

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。

【 0 0 1 5 】

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。

【 0 0 1 6 】

ここで図 1 を参照すると、ストレージシステム 1 0 0 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 1 0 0 は、ストレージコントローラ 1 1 0 及びストレージデバイス群 1 3 0、1 4 0 を含んでよく、上記ストレージデバイス群 1 3 0、1 4 0 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 1 3 0 はストレージデバイス 1 3 5 A ~ 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 : SSD））の代表例である。ストレージコントローラ 1 1 0 はクライアントコンピュータシステム 1 2 5 に直接接続してよく、またストレージコントローラ 1 1 0 は、ネットワーク 1 2 0 を介して、クライアントコンピュータシステム 1 1 5 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 1 1 5 に接続してよい。クライアント 1 1 5、1 2 5 は、システム 1 0 0 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 1 1 0 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 1 1 0 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。

【 0 0 1 7 】

ストレージコントローラ 1 1 0 は、ストレージデバイス 1 3 5 A ~ 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 1 3 0、1 4 0 から離間したものとしてストレージコントローラ 1 1 0 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 1 1 0 をストレージデバイス群 1 3 0、1 4 0 のうち的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 1 1 0 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。

【 0 0 1 8 】

ストレージコントローラ 1 1 0 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

10

20

30

40

50

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 135 A ~ 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 はまた、共通のデータセグメントを重複排除することによって、デバイス 135 A ~ N に記憶されるデータ量を削減するための、様々な微粒度重複排除技術のうちのいずれを利用してよい。

#### 【0019】

10

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。従って媒体のセットは、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。読み出し専用媒体は既に生成されたスナップショットを表す。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变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。

#### 【0020】

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、ボリューム - 媒体間マッピングテーブル、アドレス翻訳テーブル、重複排除テーブル、オーバレイテーブル及び / 又はその他のテーブルを含んでよい。いくつかの実施形態では、これらのテーブルのうちの 2 つ以上に記憶された情報を、単一のテーブルへと結合してよい。媒体マッピングテーブルを利用して、媒体と下層の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く、ボリューム - 媒体間マッピングテーブルを利用して、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。

20

#### 【0021】

アドレス翻訳テーブルは、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 135 A ~ 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。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 135 A ~ 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 135 A ~ 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。

30

#### 【0022】

様々な実施形態において、媒体 ID、論理又は仮想アドレス、セクタ番号等を含むキーを用いて、アドレス翻訳テーブルにアクセスしてよい。受信された読み書きストレージアクセス要求は、特定のボリューム、セクタ及び長さを識別してよい。ボリューム ID は、ボリューム - 媒体間マッピングテーブルを用いて媒体 ID に対してマッピングしてよい。セクタは媒体内に記憶されたデータの論理ブロックであってよい。セクタは異なる媒体において異なるサイズを有してよい。アドレス翻訳テーブルは媒体を、セクタサイズのユニットとしてマッピングしてよい。一実施形態では、アドレス翻訳テーブルにアクセスするためのキー値は、媒体 ID 及び受信したセクタ番号の組み合わせであってよい。キーは、データのある行を別の行と区別するマッピングテーブル内のエンティティである。他の実施形態では、他のタイプのアドレス翻訳テーブルを利用してよい。

40

#### 【0023】

50

一実施形態では、アドレス翻訳テーブルは、媒体を物理ポインタ値に対してマッピングしてよい。実施形態に応じて、物理ポインタ値は、ストレージデバイスがデバイス内の物理位置に対してマッピングする物理アドレス又は論理アドレスであってよい。一実施形態では、アドレス翻訳テーブルにアクセスするためにインデックスを利用してよい。このインデックスは、アドレス翻訳テーブル内のマッピングの位置を識別してよい。媒体ID及びセクタ番号から生成されたキー値を用いてインデックスを問い合わせてよく、また、キー値に適合するか又はそれ以外の様式で対応する1つ又は複数のエントリについてインデックスを検索してよい。次に適合エントリからの情報を用いて、受信した読み出し又は書き込み要求の標的であるストレージ位置を識別するマッピングを配置及び取得してよい。一実施形態では、インデックス内でのヒットにより、ストレージシステムのストレージデバイス内のページを識別する、対応する仮想ページIDが提供され、ページはキー値と、対応する物理ポインタ値との両方を記憶する。次にキー値を用いてこのページを検索して、物理ポインタ値を見つけることができる。

#### 【0024】

重複排除テーブルは、微粒度レベルでデータを重複排除するために使用される情報を含んでよい。重複排除テーブルに記憶された情報は、所定のデータ成分に関して計算された1つ又は複数のハッシュ値と、上記所定のデータ成分を保持するストレージデバイス135A~Nのうちの1つ内の物理位置に対する物理ポインタとの間のマッピングを含んでよい。更に重複排除テーブルには、上記所定のデータ成分の長さ、対応するエントリに関するステータス情報とを記憶してよい。なお、いくつかの実施形態では、物理ポインタ値と対応する物理ストレージ位置との間には1つ又は複数のレベルの間接参照が存在してよい。従ってこれらの実施形態では、物理ポインタを用いて、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できる。

#### 【0025】

ストレージコントローラ110は、ユーザがボリュームを誤って削除してしまったと決定した場合に、削除されたボリュームを復元できるよう構成してよい。一実施形態では、ストレージコントローラ110は、ボリューム-媒体間マッピングテーブル内の、あるボリュームに対応するエントリを削除することによって、そのボリュームを削除してよい。これはボリュームとそのアンカー媒体との間のリンクを除去することになる。またストレージコントローラは、削除されたボリュームとそのアンカー媒体との間の過去のマッピングのインジケータを記憶してよい。このマッピングは実施形態に応じて、ログ、テーブル又は別の位置に記憶してよい。次に後の時点において、ユーザが削除されたボリュームの復元を要求すると、ストレージコントローラは上記マッピングを取得して、これを用いて削除されたボリュームとそのアンカー媒体との間のリンクを再生成してよい。このリンクが再生成されると、これによって削除されたボリュームが過去の状態に効果的に復元される。

#### 【0026】

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。

#### 【0027】

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 n

10

20

30

40

50

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ーイーサネット(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。

10

#### 【0028】

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。

20

#### 【0029】

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トであり得る。

30

#### 【0030】

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。

40

#### 【0031】

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

50

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。

#### 【0032】

10

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。

#### 【0033】

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の媒体を示し、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。読み書き型媒体をアクティブ媒体と呼んでよく、読み出し専用媒体は既に生成されたスナップショットを表してよい。グラフ200内では、媒体はその下層の媒体を指す。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の下層の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の媒体となる。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。

20

30

#### 【0034】

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を下層の媒体として用いて、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。

40

#### 【0035】

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

50

ブロックの範囲に分解してもよく、各範囲を媒体の D A G 2 0 0 内で別個に処理してよい。

#### 【 0 0 3 6 】

ここで図 3 を参照すると、媒体マッピングテーブル 3 0 0 の一実施形態を示す。媒体マッピングテーブル 3 0 0 のいずれの部分又は全体は、ストレージコントローラ 1 1 0 及び / 又はストレージデバイス 1 3 5 A ~ 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 ( I D ) を使用してボリューム - 媒体間マッピングテーブル 2 0 5 にアクセスし、このボリューム I D に対応する媒体 I D を決定してよい。次にこの媒体 I D を使用して媒体マッピングテーブル 3 0 0 にアクセスしてよい。なおテーブル 3 0 0 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 に示したものと異なる方法で整理されてよい。また、効率的な検索を提供するために、いずれの適切なデータ構造 ( 例えば B ツリー、バイナリツリー、ハッシュテーブル等 ) を用いて、マッピングテーブル情報を記憶してよいことに留意されたい。これらのデータ構造全てが考察の対象となる。

#### 【 0 0 3 7 】

テーブル 3 0 0 の左列に示すように、媒体 I D によって各媒体を識別してよい。テーブル 3 0 0 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 4 K B 、 8 K B ) 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 I D 2 は 2 つの別個の範囲を有する。媒体 I D 2 の範囲 0 ~ 9 9 は、テーブル 3 0 0 内で媒体 I D 2 の範囲 1 0 0 ~ 9 9 9 のためのエントリから離間したエントリを有する。

#### 【 0 0 3 8 】

媒体 I D 2 のこれらの範囲を共に下層媒体 I D 1 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 I D 3 5 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 I D 3 5 の範囲 0 ~ 2 9 9 を、オフセット 4 0 0 で下層媒体 I D 1 8 に対してマッピングする。これは、媒体 I D 3 5 のブロック 0 ~ 2 9 9 を媒体 I D 1 8 のブロック 4 0 0 ~ 6 9 9 に対してマッピングすることを示す。更に媒体 I D 3 5 の範囲 3 0 0 ~ 4 9 9 を、 ~ 3 0 0 のオフセットで下層媒体 I D 3 3 に対してマッピングし、媒体 I D 3 5 の範囲 5 0 0 ~ 8 9 9 を、 ~ 4 0 0 のオフセットで下層媒体 I D 5 に対してマッピングする。これらのエントリは、媒体 I D 3 5 のブロック 3 0 0 ~ 4 9 9 が媒体 I D 3 3 のブロック 0 ~ 1 9 9 に対してマッピングされ、媒体 I D 3 5 のブロック 5 0 0 ~ 8 9 9 が媒体 I D 5 のブロック 1 0 0 ~ 4 9 9 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。

#### 【 0 0 3 9 】

テーブル 3 0 0 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 ( q u i e s c e n t ) であることを示し、「R」は媒体が登録状態 ( r e g i s t e r e d ) であることを示し、「U」は媒体が非マスク状態 ( u n m a s k e d ) であることを示す。静止状態では、テーブル 3 0 0 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 3 0 0 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 I D 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。

#### 【 0 0 4 0 】

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、媒体自体及び下層の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。

【 0 0 4 1 】

更にテーブル 3 0 0 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 3 0 0 に示すように、媒体 8 は 5 0 0 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 5 0 0 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、5 0 0 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。

【 0 0 4 2 】

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（又は読み取り専用であること）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 5 0 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てよいが、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。

【 0 0 4 3 】

図 4 に移ると、ボリュームを削除及び復元する動作の一実施形態が示されている。ユーザは所定のボリュームを削除したい場合があり、この所定のボリュームを削除するための要求は、ストレージコントローラへと搬送され得る。削除されるボリュームは、図 4 においてボリューム 4 1 0 で表されている。この削除動作を実行する前に、ボリューム - 媒体間マッピングテーブル 4 2 5 A 内にはボリューム 4 1 0 に対応するエントリが存在し得、このエントリは、媒体 4 1 5 がボリューム 4 1 0 のアンカー媒体であることを記録してよい。この議論の目的のために、媒体 4 1 5 は下層媒体を有すると考えられ、この下層媒体は媒体 4 2 0 として示されている。媒体 4 2 0 は状況に応じて下層媒体を有していても有しなくてもよい。なお、場合によっては媒体 4 1 5 は下層媒体を有しなくてもよい。

【 0 0 4 4 】

ストレージコントローラは、削除済みボリュームテーブル 4 3 0 内にアンカー媒体 4 1 5 に対するボリューム 4 1 0 の過去のマッピングを記録して、ボリューム - 媒体間マッピングテーブル 4 2 5 B 内のボリューム 4 1 0 に対応するエントリを削除することにより、ボリューム 4 1 0 の削除動作を実行してよい。ボリューム及びアンカー媒体の番号は、ボリューム - 媒体マッピングテーブル 4 2 5 B 内で消去され、これはこのエントリがテーブル 4 2 5 B から削除されたことを表す。またこのブロック図は、ボリューム 4 1 0 が削除されていること、及びボリューム 4 1 0 と媒体 4 1 5 との間のリンクがもはや存在しないことを表すために、ボリューム 4 1 0 を消去されたものとして示している。

【 0 0 4 5 】

ボリューム 4 1 0 に対応するエントリがボリューム - 媒体間マッピングテーブル 4 2 5 B から削除されると、これによりボリューム 4 1 0 はストレージシステムから効果的に削除される。ボリューム 4 1 0 が削除される結果として、そのアンカー媒体（媒体 4 1 5）は安定となる。削除ボリュームテーブル 4 3 0 は、削除されたボリューム及び削除されたボリュームとアンカー媒体との過去の接続に関する情報を記憶するために利用できる、あるタイプのストレージ構造である。他の実施形態では、この情報を記憶するその他のタイプの構造及びその他の方法を利用してよい。この情報を記憶することにより、後の時点において、削除されたボリュームの復元動作を実行できる。更に、媒体 4 1 5、媒体 4 2 0 及び他のいずれの下層媒体に対応する媒体マッピングテーブル（図示せず）中のエントリを、削除動作後に保持してよい。ストレージコントローラは、削除動作が実行された後、ある期間に亘って媒体 4 1 5 を保持してよい。場合によっては、他のボリュームが媒体 4 1 5 を利用する場合又は他の媒体が下層媒体として媒体 4 1 5 を利用する場合、媒体 4 1 5 は無期限に保持されることになる。

10

#### 【 0 0 4 6 】

ボリューム 4 1 0 の削除動作実行後のいずれの時点において、ユーザはボリューム 4 1 0 の復元動作を要求してよい。ストレージコントローラは、削除ボリュームテーブル 4 3 0 に記憶された情報に基づいて、ボリューム 4 1 0 を復元してよい。図 4 の下部の点線のボックスは、復元動作実行後のボリューム - 媒体マッピングテーブル 4 2 5 C のステータスを示す。復元動作により、ボリューム 4 1 0 は媒体 4 1 5 に再接続される。換言すると、復元動作はボリューム 4 1 0 を、削除動作実行前の状態に復元している。媒体 4 1 5 の

20

#### 【 0 0 4 7 】

一般にストレージコントローラは、所定のボリュームの削除動作の後、最短の期間だけ、削除ボリューム連携情報を保持してよい。あるいはストレージコントローラは、他のデータの記憶のために空間が必要になるまで、削除ボリューム連携情報を保持してよい。ストレージコントローラがこの特定の期間だけ待機したら、上記所定のボリュームに対応する連携情報を削除してよい。この時点において、ストレージコントローラはこの削除されたボリュームの所定のアンカー媒体を即座に再利用してよく、又はストレージコントローラは、上記所定のアンカー媒体を再利用するのに更に便利な後の時点まで待機してよい。場合によっては、ストレージコントローラは、他のデータの記憶のために空間が必要になるまで、上記所定のアンカー媒体の再利用を待機してよい。この所定のアンカー媒体のいずれの下層媒体もまた、他のボリューム又は媒体によって利用されていなければ再利用してよい。

30

#### 【 0 0 4 8 】

ここで図 5 を参照すると、削除されたボリュームを復元するための代替動作の一実施形態が示されている。この実施形態では、過去に削除したボリューム 4 1 0 を復元するための動作は、ボリューム 4 1 0 を復元するための要求を受信した後、新規の媒体（媒体 5 0 5）を生成することを伴う。そして媒体 5 0 5 に関して、媒体マッピングテーブル 5 1 5 内に新規のエントリを生成してよく、媒体 4 1 5 が媒体 5 0 5 の下層にあることのインジケータをこの新規のエントリに記憶してよい。媒体 4 1 5 に関する既存のエントリも媒体

40

#### 【 0 0 4 9 】

また、ボリューム - 媒体間マッピングテーブル 5 2 5 に新規のエントリを生成してよく、媒体 5 0 5 を媒体 4 1 0 のアンカー媒体としてこのエントリに記録してよい。図 5 からわかるように、復元動作の後、ボリューム 4 1 0 は媒体 5 0 5 を指しており、媒体 5 0 5 は媒体 4 1 5 を指している。媒体 5 0 5 を生成して、媒体 5 0 5 をボリューム 4 1 0 とその過去のアンカー媒体（媒体 4 1 5）との間に配置することにより、ストレージコントローラは、ボリューム 4 1 0 の復元後にボリューム 4 1 0 に発生するいずれの変更を、ボリューム 4 1 0 が削除される前に発生した変更と確実に区別できる。

50

## 【 0 0 5 0 】

ここで図 6 を参照すると、ボリューム動作を取り消すための方法 6 0 0 の一実施形態が示されている。上述のシステム 1 0 0 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 1 1 0 ）は一般に、方法 6 0 0 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。

## 【 0 0 5 1 】

ストレージコントローラは、第 1 のボリュームを削除するための要求を受信してよい（ブロック 6 0 5 ）。この要求の受信に 응답して、ストレージコントローラは第 1 のボリュームのアンカー媒体を決定してよい（ブロック 6 1 0 ）。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内の、第 1 のボリュームに対応するエントリから、アンカー媒体を決定してよい。この議論の目的のために、第 1 のボリュームのアンカー媒体を「第 1 の媒体」と呼んでよい。次にストレージコントローラは、第 1 のボリュームと第 1 の媒体との間のマッピングのインジケータを記憶してよい（ブロック 6 1 5 ）。一実施形態では、このマッピングを削除ボリュームテーブルに記憶してよい。他の実施形態では、このマッピングを他のいずれの適切な構造（例えばログファイル）に記憶してよい。一実施形態では、このマッピングは、第 1 のボリュームが削除された後一定の期間だけ保持してよい。ストレージコントローラは、第 1 のボリュームの復元をユーザが望む場合に第 1 のボリュームを復元できるよう、適切な期間だけこのマッピングを保持するよう構成される。この期間が経過すると、マッピングを削除してよい。別の実施形態では、マッピングを無期限に保持してよい。なお、マッピングを削除すると、ストレージコントローラは第 1 のボリュームをその過去の状態に復元できなくなる。

## 【 0 0 5 2 】

次にストレージコントローラは、第 1 のボリュームが有効であることのいずれのインジケータを除去してよい（ブロック 6 2 0 ）。第 1 のボリュームが有効であることのいずれのインジケータを除去することによって、ストレージコントローラは第 1 のボリュームを、ユーザにとって利用不可能とする。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内の、第 1 のボリュームに対応するエントリを削除することによって、第 1 のボリュームが有効であることのいずれのインジケータを除去してよい。他の実施形態では、この情報をテーブルに記憶する以外の、ボリュームと媒体との間のマッピングを維持するための他の技術を用いてよい。これらの実施形態では、ストレージコントローラは第 1 のボリュームのステータスをいずれの適切な様式で更新してよい。また、一定の期間が経過した後、ストレージコントローラは第 1 の媒体を、再利用の準備ができた状態としてマークしてよく、続いて第 1 の媒体を、次回ゴミ集め動作が実施される際に削除してよい。場合によっては、ストレージコントローラは、削除されたボリュームより前に、他のタイプの空き空間の再利用を優先してよい。例えばストレージコントローラは、削除されたボリュームから空間を再利用する前に、（上書きによって）到達できない空間を要求するよう試みてよい。なお、第 1 の媒体が削除されると、ストレージコントローラは第 1 のボリュームをその過去の状態に復元できなくなる。しかしながらストレージコントローラは、第 1 のボリュームを過去の状態に復元するための復元動作をユーザが要求した場合にこの復元動作を可能とするよう、相当な期間に亘って第 1 の媒体を保持してよい。

## 【 0 0 5 3 】

そしてブロック 6 2 0 の後、後の時点において、ストレージコントローラは第 1 のボリュームを復元するための要求を受信してよい（ブロック 6 2 5 ）。例えばユーザは、第 1 のボリュームを誤って削除してしまったことに気付く場合があり、従ってこの第 1 のボリュームをその過去の状態に復元したい場合がある。第 1 のボリュームを復元するためのこの要求は、第 1 の媒体に対する第 1 のボリュームのマッピングが削除される前、かつ第 1

10

20

30

40

50

の媒体が削除される前に、ストレージコントローラが受信すると想定され得る。第1のボリュームを復元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、削除ボリュームテーブル内の、第1のボリュームに対応するマッピングを取得してよい(ブロック630)。次にこの取得したマッピングを用いて、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内に第1のボリュームに関するエントリを生成してよく、このエントリは第1の媒体を、第1のボリュームのアンカー媒体として記録してよい(ブロック635)。ボリューム - 媒体間マッピングテーブル内に第1のボリュームに関する上記エントリを生成することによって、ストレージコントローラは第1のボリュームをその過去の状態に復元する。ブロック635の後、方法600を終了してよい。

#### 【0054】

ここで図7を参照すると、削除されたボリュームを復元するための代替方法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。

#### 【0055】

ストレージコントローラは、過去に削除された第1のボリュームを復元するための要求を受信してよい(ブロック705)。この要求の受信に応答して、ストレージコントローラは、削除ボリュームテーブル内の、第1のボリュームに対応するマッピングを取得してよい(ブロック710)。取得したマッピングは、第1のボリュームが削除される前の第1のボリュームのアンカー媒体の識別を記憶してよく、この議論の目的のために、このアンカー媒体を「第1の媒体」と呼んでよい。またこ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を生成してよい(ブロック715)。この議論の目的のために、この新規の媒体を「第2の媒体」と呼んでよい。次にストレージコントローラは、第2の媒体に関して媒体マッピングテーブル内に新規のエントリを生成してよい(ブロック720)。ストレージコントローラは、媒体マッピングテーブル内のこの新規のエントリにおいて第1の媒体が第2の媒体の下層にあることのインジケータを記憶してよい(ブロック725)。

#### 【0056】

またストレージコントローラは、第1のボリュームに関してボリューム - 媒体間マッピングテーブル内に新規のエントリを生成してよい(ブロック730)。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内のこの新規のエントリにおいて第2の媒体を第1のボリュームのアンカー媒体として記録してよい(ブロック735)。方法700は、過去に削除されたボリュームを復元するための代替実装形態である。この代替実装形態では、新規の媒体を生成し、復元されるボリュームがこの新規のボリュームを指すようにすることによって、ストレージコントローラは、復元されたボリュームに対する変更を、ボリュームが削除される前に発生したいずれの変化と確実に区別できる。

#### 【0057】

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。

#### 【0058】

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

10

20

30

40

50

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つてよく、又はソフトウェアを所定の期間だけカスタム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。

【0059】

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。

10

【図1】

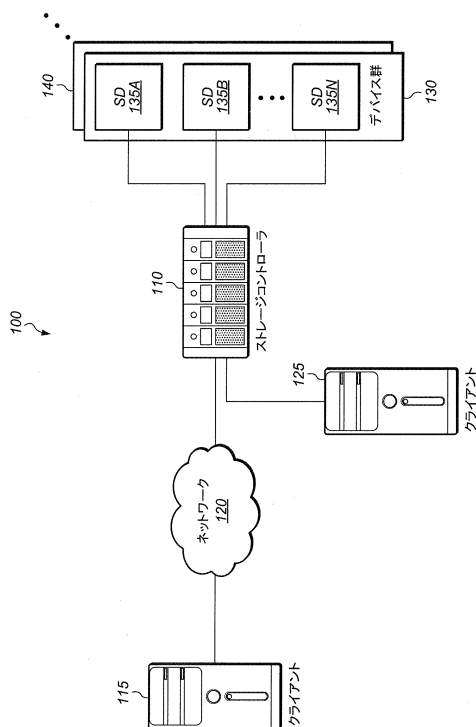


FIG. 1

【図2】

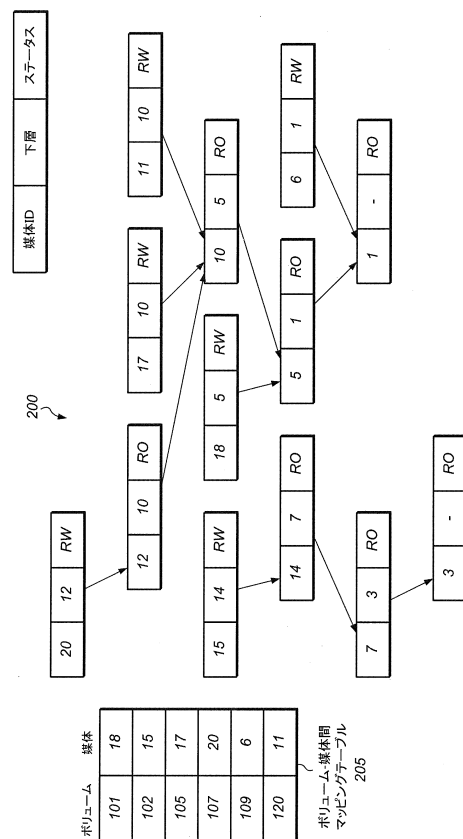


FIG. 2

【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	Y
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

【図 4】

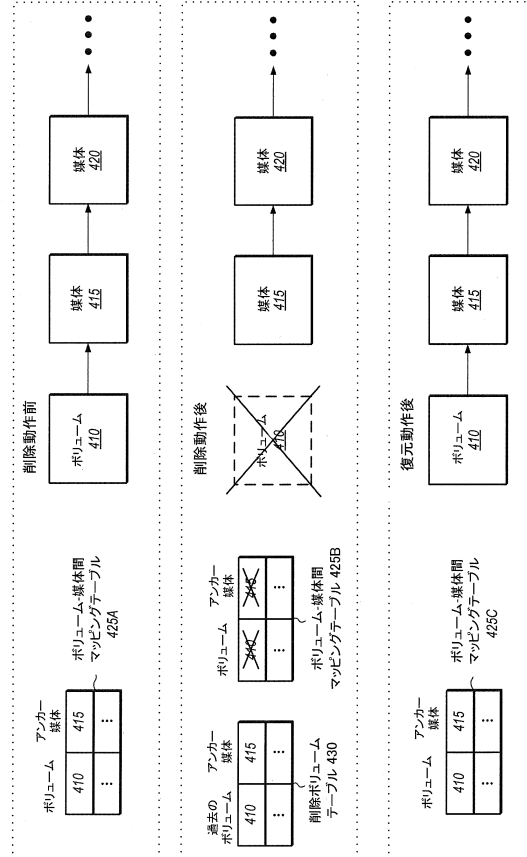


FIG. 3

FIG. 4

【図 5】

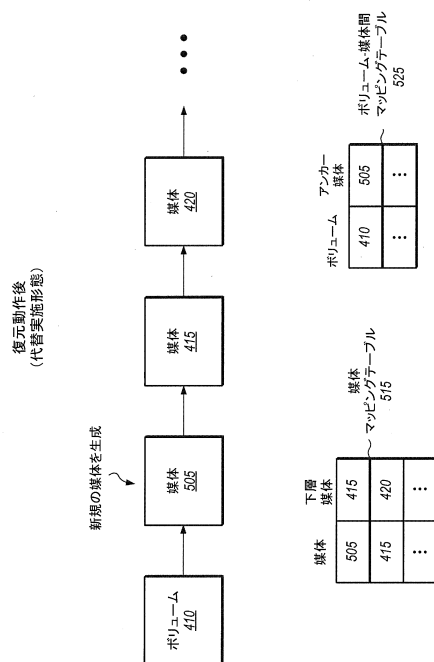


FIG. 5

【図 6】

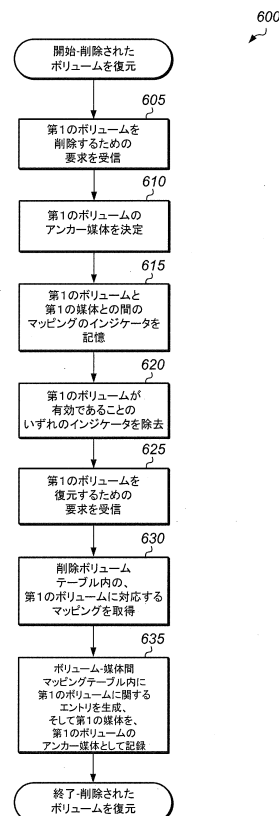


FIG. 6

【図 7】

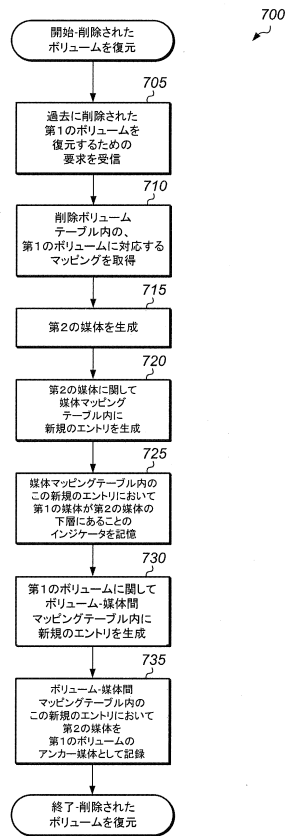


FIG. 7

## フロントページの続き

(31)優先権主張番号 14/046,878

(32)優先日 平成25年10月4日(2013.10.4)

(33)優先権主張国・地域又は機関  
米国(US)

(74)代理人 100126480

弁理士 佐藤 睦

(74)代理人 100121843

弁理士 村井 賢郎

(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  
ニュー・722

(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  
ニュー・791

## 合議体

審判長 田中 秀人

審判官 はま 中 信行

審判官 山崎 慎一

(56)参考文献 米国特許出願公開第2010/0042802(US, A1)  
特開2007-058264(JP, A)

(58)調査した分野(Int.Cl., DB名)

G06F 3/06 - 3/08, 12/00, 13/10 -13/14