(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5798258号
(P5798258)
(45)発行日　平成27年10月21日(2015.10.21)　(24)登録日　平成27年8月28日(2015.8.28)

(51)Int.Cl.　　F I
G06F 3/06 (2006.01)　　G06F 3/06 301Z
G06F 12/00 (2006.01)　　G06F 3/06
G06F 3/06 301X
G06F 3/06 304N
G06F 3/06 304F
請求項の数 20　(全 25 頁)　最終頁に続く

(21)出願番号　特願2014-550282 (P2014-550282)
(86)(22)出願日　平成24年3月29日 (2012.3.29)
(65)公表番号　特表2015-509235 (P2015-509235A)
(43)公表日　平成27年3月26日 (2015.3.26)
(86)国際出願番号　PCT/US2012/031102
(87)国際公開番号　WO2013/147783
(87)国際公開日　平成25年10月3日 (2013.10.3)
審査請求日　平成26年6月30日 (2014.6.30)

(73)特許権者　510071448
ヒタチ　データ　システムズ　コーポレーション
ＨＩＴＡＣＨＩ　ＤＡＴＡ　ＳＹＳＴＥＭＳ　ＣＯＲＰＯＲＡＴＩＯＮ
アメリカ合衆国　カリフォルニア州　９５０５０　サンタ・クララ　ラファイエット　ストリート　２８４５
(74)代理人　110000279
特許業務法人ウィルフォート国際特許事務所
(72)発明者　ロジャース，リチャード
アメリカ合衆国　９８００５　ワシントン州　ベルビュー，サウスイースト　４番地　１００２番，１１８７０
最終頁に続く

(54)【発明の名称】記憶階層化のためのコンテンツ選択

(57)【特許請求の範囲】
【請求項１】
１つ以上のノードを有し、各ノードが、プロセッサを有し、且つ、第１の階層としての複数の第１の論理記憶ユニットおよび第２の階層としての複数の第２の論理記憶ユニットの基になる複数の記憶媒体を有する記憶システムに接続され、
各ノードのプロセッサが、
前記複数の第１の論理記憶ユニットに複数の第１のコンテンツオブジェクトを記憶し、前記記憶システムに記憶すべき前記第１のコンテンツオブジェクトの数が前記第１のコンテンツオブジェクトのデータ保護レベルに基づいて規定されており、
前記複数の第１の論理記憶ユニットの使用された容量が閾値を上回る場合、前記第１のコンテンツオブジェクトのうちの１つの第１のコンテンツオブジェクト以外の全ての第１のコンテンツオブジェクトを前記複数の第２の論理記憶ユニットに移動し、且つ、前記複数の第１の論理記憶ユニットに前記１つの第１のコンテンツオブジェクトを維持する、
ことを特徴とするシステム。
【請求項２】
前記複数の第１の論理記憶ユニットは、複数の稼働ユニットであり、前記複数の稼働ユニットは、それらの稼働ユニットに対応する１つ以上の第１のＲＡＩＤグループに対してスピンダウン機能の実行を可能にせず、
前記複数の第２の論理記憶ユニットは、複数のスピンダウンユニットであり、複数のスピンダウンユニットが、それらのスピンダウンユニットに対応する１つ以上の第２のＲＡ

ＩＤグループに対してスピンダウン機能の実行を可能にする、
ことを特徴とする請求項１に記載のシステム。
【請求項３】
前記プロセッサが、前記１つ以上のノードの論理パーティションであるネームスペースを管理し、
前記複数の第１の論理記憶ユニットにおけるどのコンテンツオブジェクトが前記複数の第２の論理記憶ユニットへ移動されるべき資格があるかを示す階層化ルールが、前記ネームスペースに設定され、
前記複数の第１の論理記憶ユニットの前記使用された容量が前記閾値を上回る場合、前記プロセッサが、前記階層化ルールに従って、前記複数の第２の論理記憶ユニットに前記ネームスペース内のコンテンツオブジェクトを移動する、
ことを特徴とする請求項１に記載のシステム。
【請求項４】
或るコンテンツオブジェクトのデータ保護レベルが１より大きい場合、前記階層化ルールが、前記ネームスペースにおける前記或るコンテンツオブジェクトのうちの１つの或るコンテンツオブジェクト以外の全ての或るコンテンツオブジェクトが前記複数の第２の論理記憶ユニットに移動されるべき資格があることを示す、
ことを特徴とする請求項３記載のシステム。
【請求項５】
前記階層化ルールが、或る時間前に前記ネームスペースに取込まれたコンテンツオブジェクトが前記複数の第２の論理記憶ユニットへ移動されるべき資格があることを示す、
ことを特徴とする請求項３に記載のシステム。
【請求項６】
前記階層化ルールが、前記ネームスペース内のいずれのコンテンツオブジェクトも前記複数の第２の論理記憶ユニットへ移動されるべき資格がないことを示す、
ことを特徴とする請求項３に記載のシステム。
【請求項７】
階層化ルールを変更することによって前記第１のコンテンツオブジェクトのうちの１つの第１のコンテンツオブジェクト以外の全ての第１のコンテンツオブジェクトが、前記複数の第２の論理記憶ユニットに移動されるべき資格を有さないようになる場合、前記第１のコンテンツオブジェクトのうちの１つの第１のコンテンツオブジェクト以外の全ての第１のコンテンツオブジェクトが前記複数の第１の論理記憶ユニットへ戻される、
ことを特徴とする請求項１に記載のシステム。
【請求項８】
第１の階層としての複数の第１の論理記憶ユニットおよび第２の階層としての複数の第２の論理記憶ユニットの基になる複数の記憶媒体を有する記憶システムに接続された１つ以上のノードのための方法であって、
前記複数の第１の論理記憶ユニットに複数の第１のコンテンツオブジェクトを格納し、前記記憶システムに記憶すべき前記第１のコンテンツオブジェクトの数が前記第１のコンテンツオブジェクトのデータ保護レベルに基づいて規定されており、
前記複数の第１の論理ユニットの使用された容量が閾値を上回る場合、
前記第１のコンテンツオブジェクトのうちの１つの第１のコンテンツオブジェクト以外の全ての第１のコンテンツオブジェクトを前記複数の第２の論理記憶ユニットに移動し、且つ、
前記複数の第１の論理記憶ユニットに前記１つの第１のコンテンツオブジェクトを維持する、
方法。
【請求項９】
前記複数の第１の論理記憶ユニットは、複数の稼働ユニットであり、前記複数の稼働ユニットは、それらの稼働ユニットに対応する１つ以上の第１のＲＡＩＤグループに対して

スピンダウン機能の実行を可能にせず、
前記複数の第２の論理記憶ユニットは、複数のスピンダウンユニットであり、複数のスピンダウンユニットが、それらのスピンダウンユニットに対応する１つ以上の第２のＲＡＩＤグループに対してスピンダウン機能の実行を可能にする、
ことを特徴とする請求項８に記載の方法。
【請求項１０】
前記１つ以上のノードの論理パーティションであるネームスペースを管理し、
前記複数の第１の論理記憶ユニットにおけるどのコンテンツオブジェクトが前記複数の第２の論理記憶ユニットに移動されるべき資格があるかを示す階層化ルールを前記ネームスペースに設定する、
ことを更に実行し、
前記複数の第１の論理ユニットの前記使用された容量が前記閾値を上回る場合には、前記階層化ルールに従って、前記複数の第２の論理記憶ユニットに、前記ネームスペース内のコンテンツオブジェクトを移動する、
ことを特徴とする請求項８に記載の方法。
【請求項１１】
或るコンテンツオブジェクトのデータ保護レベルが１より大きい場合、前記階層化ルールが、前記ネームスペースにおける前記或るコンテンツオブジェクトのうちの１つの或るコンテンツオブジェクト以外の全ての或るコンテンツオブジェクトが前記複数の第２の論理記憶ユニットに移動されるべき資格があることを示す、
ことを特徴とする請求項１０に記載の方法。
【請求項１２】
前記階層化ルールが、或る時間前に前記ネームスペースに取込まれたコンテンツオブジェクトが前記複数の第２の論理記憶ユニットへ移動されるべき資格があることを示す、
ことを特徴とする請求項１０に記載の方法。
【請求項１３】
前記階層化ルールが、前記ネームスペース内のいずれのコンテンツオブジェクトも前記複数の第２の論理記憶ユニットへ移動されるべき資格がないことを示す、
ことを特徴とする請求項１０に記載の方法。
【請求項１４】
前記階層化ルールを変更し、
階層化ルールを変更することによって前記第１のコンテンツオブジェクトのうちの１つの第１のコンテンツオブジェクト以外の全ての第１のコンテンツオブジェクトが、前記複数の第２の論理記憶ユニットに移動されるべき資格を有さないようになる場合、前記第１のコンテンツオブジェクトのうちの１つの第１のコンテンツオブジェクト以外の全ての第１のコンテンツオブジェクトを前記複数の第１の論理記憶ユニットへ戻す、
ことを特徴とする請求項８に記載の方法。
【請求項１５】
第１の階層としての複数の第１の論理記憶ユニットおよび第２の階層としての複数の第２の論理記憶ユニットの基になる複数の記憶媒体を有する記憶システムに接続されたノードで実行されるコンピュータプログラムであって、
前記複数の第１の論理記憶ユニットに複数の第１のコンテンツオブジェクトを格納し、前記記憶システムに記憶すべき前記第１のコンテンツオブジェクトの数が前記第１のコンテンツオブジェクトのデータ保護レベルに基づいて規定されており、
前記複数の第１の論理ユニットの使用された容量が閾値を上回る場合、
前記第１のコンテンツオブジェクトのうちの１つの第１のコンテンツオブジェクト以外の全ての第１のコンテンツオブジェクトを前記複数の第２の論理記憶ユニットへ移動し、且つ、
前記複数の第１の論理記憶ユニットに前記１つの第１のコンテンツオブジェクトを維持する

ことを前記ノードに実行させるコンピュータプログラム。
【請求項１６】
前記１つ以上のノードの論理パーティションであるネームスペースを管理し、
前記複数の第１の論理記憶ユニットにおけるどのコンテンツオブジェクトが前記複数の第２の論理記憶ユニットに移動されるべき資格があるかを示す階層化ルールを前記ネームスペースに設定する、
ことを更に前記ノードに実行させ、
前記複数の第１の論理ユニットの前記使用された容量が前記閾値を上回る場合には、前記階層化ルールに従って、前記複数の第２の論理記憶ユニットに、前記ネームスペース内のコンテンツオブジェクトを移動する、
ことを特徴とする請求項１５に記載のコンピュータプログラム。
【請求項１７】
或るコンテンツオブジェクトのデータ保護レベルが１より大きい場合、前記階層化ルールが、前記ネームスペースにおける前記或るコンテンツオブジェクトのうちの１つの或るコンテンツオブジェクト以外の全ての或るコンテンツオブジェクトが前記複数の第２の論理記憶ユニットに移動されるべき資格があることを示す、
ことを特徴とする請求項１６に記載のコンピュータプログラム。
【請求項１８】
前記階層化ルールが、或る時間前に前記ネームスペースに取込まれたコンテンツオブジェクトが前記複数の第２の論理記憶ユニットへ移動されるべき資格があることを示す、
ことを特徴とする請求項１６に記載のコンピュータプログラム。
【請求項１９】
前記階層化ルールが、前記ネームスペース内のいずれのコンテンツオブジェクトも前記複数の第２の論理記憶ユニットへ移動されるべき資格がないことを示す、
ことを特徴とする請求項１６に記載のコンピュータプログラム。
【請求項２０】
前記階層化ルールを変更し、
階層化ルールを変更することによって前記第１のコンテンツオブジェクトのうちの１つの第１のコンテンツオブジェクト以外の全ての第１のコンテンツオブジェクトが、前記複数の第２の論理記憶ユニットに移動されるべき資格を有さないようになる場合、前記第１のコンテンツオブジェクトのうちの１つの第１のコンテンツオブジェクト以外の全ての第１のコンテンツオブジェクトを前記複数の第１の論理記憶ユニットへ戻す、
ことを更に前記ノードに実行させることを特徴とする請求項１５に記載のコンピュータプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、一般に、記憶システムに、及びより詳しくは、記憶階層化のためのコンテンツ選択に基づいてコンテンツシステム内のコンテンツの記憶を管理するためのシステム及び方法に関する。
【背景技術】
【０００２】
固定コンテンツシステム（ＦＣＳ）は、変化しないデータを収容している。コンテンツの僅かな割合だけが、実際にその後アクセスされる；しかしながら、データはパワー資源を継続的に消費している記憶媒体上になおとどまる。たとえ媒体上のコンテンツの多くがアクセスされなくても、パワーは媒体をスピンさせ続けて無駄にされる。
【発明の概要】
【発明が解決しようとする課題】
【０００３】
米国特許第８，００６，１１１号明細書が、スピンダウンの概念を開示する。このアプ

ローチの下で、ファイル移動及びディスクドライブのパワー状態が、ファイルの活動に基づいて決定される。活性／不活性ディスクドライブへのアクセスが活性／不活性ディスクドライブをパワーダウンするための不活性閾値に到達すると、各活性／不活性ディスクドライブが類似した確率のアクセスを有するファイルをロードされるように、能動記憶域内に不活性になっていたファイルの群が、それぞれの空にされた活性／不活性ディスクドライブに連続して移行される。焦点は、個々のディスクドライブにある。不活性閾値が到達された時だけ、ファイルが移動される。

【課題を解決するための手段】

【０００４】

本発明の例示的な実施態様が、複製されたオブジェクト記憶システムまたはコンテンツプラットホームであることができ、かつ、記憶媒体を２つのタイプ、すなわち、稼働（Ｒｕｎ）ユニット（ＲＵ）及びスピンダウンユニット（ＳＤＵ））に分類する、固定コンテンツシステム（ＦＣＳ）を提供する。ＲＵはおそらく使用されるコンテンツに対して指定され、及びＳＤＵはおそらく使用されないコンテンツを収容する。定期的に、ＦＣＳはＲＵ媒体とＳＤＵ媒体との間を移動させるための候補を識別する設定可能なサービスプラン及び記憶階層化ポリシーに基づいてコンテンツを評価する。記憶階層化は、オブジェクト移動及び装置操作を最適化するためにシステム状態に基づいてシステム挙動を決定し、それで、システム状態がニーズを決定する時だけ、データ移動が実行される。監視されるシステム状態は、オブジェクトが存在するＲＵが消費利用閾値（すなわちスペース／記憶消費閾値）に到達したかどうか、かつ、オブジェクトのためにＳＤＵ上で利用可能なスペースがあるかどうかである。消費利用閾値の一例は、消費された記憶域である（例えば、７０％消費された記憶域がデフォルトとして用いられることができる）。記憶階層化ルール（ＳＴＲ）は、ＳＤＵまたはＲＵ上に存在するオブジェクト立候補のためのポリシーを特定する。特定の実施態様において、設定は「決してない」「保護コピーだけ」または、「取込後Ｎ日」である。ＳＴＲは、ＦＣＳ内の全てのオブジェクトまたは一まとまりのオブジェクトに適用されることができる。

【０００５】

特定の実施態様において、階層化のためのコンテンツ識別が、バックグラウンドで実行される。コンテンツが識別される時、ＳＤＵ媒体が（必要に応じて）スピンアップされ、及びオブジェクト（複数オブジェクト）が移動される。不活性タイムアウトの後、ＳＤＵ媒体はスピンダウンされる。コンテンツがＦＣＳから要求され、及びそれがＳＤＵ上に存在する場合、ＳＤＵがスピンアップされ、及びコンテンツが読み出されて要求側に返される。不活性タイムアウトの後、ＳＤＵ媒体は再びスピンダウンされる。

【０００６】

本発明は、米国特許第８，００６，１１１号明細書と種々の点において異なる。例えば、ファイルを移動させる判定は、個々のファイルの活動に基づかない。その代わりに、ファイル移動を確立するための基準が、ファイルの経過時間、冗長バックアップコピーの存在、などを含む。米国特許第８，００６，１１１号明細書が個々のディスクドライブに焦点を合わせるとはいえ、本発明の特定の実施態様に従う解決策はディスク配列内に実現される機能を包含し、及び、焦点は一まとまりのディスクドライブの形のＲＡＩＤ群にある。さらに、それが存在する記憶域で、スペース／記憶消費閾値が到達されない限り、本発明の特定の実施態様に従う解決策は能動記憶域からコンテンツを移動させない。対照的に、米国特許第８，００６，１１１号明細書は不活性閾値が到達された時にだけファイルを移動させる。

【０００７】

固定コンテンツシステムは、スピンダウン機能を可能にしない稼働ユニット及びスピンダウン機能を可能にするスピンダウンユニットを含む複数の記憶ユニット内にコンテンツを記憶する複数の独立ノードを有する。本発明の一態様に従って、コンテンツの記憶を管理する方法が：どんなコンテンツが稼働ユニット上に記憶されるべき資格があるか、及びどんなコンテンツがスピンダウンユニット上に記憶されるべき資格があるかを示すポリシ

ーを設定する記憶階層化ルールであって、記憶ユニット内の記憶及び記憶ユニット間の移行のためにそれの記憶されたコンテンツの適格性を判定するためにコンテンツシステム内の少なくとも一群のコンテンツに適用可能である記憶階層化ルールを確立するステップ；コンテンツシステムの状態を監視するステップ；ならびに、記憶階層化ルール、コンテンツシステムの状態及び少なくとも一群のコンテンツの記憶されたコンテンツの適格性に基づいて、稼働ユニットとスピンダウンユニットとの間の移行を含む記憶ユニット間を移行するコンテンツの候補を識別するステップを含む。

【０００８】

いくつかの実施態様において、記憶階層化ルールが、期待される使用、ライフサイクル及びコンテンツの経過時間及びコンテンツの１つ以上の冗長バックアップコピーの存在を含む一組の基準に基づいて記憶されたコンテンツの適格性を判定するためにコンテンツを評価するように確立される。記憶階層化ルールが、決してスピンダウンされない記憶ユニット上に記憶されるべきコンテンツにあてはまる「けっしてない」ルール、スピンダウンユニット内の記憶の候補であるバックアップコピーのためのコンテンツにあてはまる「保護コピーだけ」ルール、及びスピンダウンユニットに記憶されるべき候補になるのに十分である一定の時間Ｘの間コンテンツシステム内に存在したコンテンツにあてはまる「取込後Ｘ時間」ルールを含む。記憶階層化ルールは、資格のあるコンテンツが存在する稼働ユニットが設定可能な消費利用閾値に到達した時にだけ、資格のあるコンテンツをスピンダウンユニットに移行するために確立される。コンテンツシステムの状態を監視するステップが、コンテンツが存在する稼働ユニットが、コンテンツをスピンダウンユニットに移行するパーミッションを示す設定可能な消費利用閾値に到達したかどうかを判定するステップ、及び稼働ユニットからコンテンツを収容するためにスピンダウンユニット上に利用可能なスペースがあるかどうかを判定するステップを含む。

【０００９】

特定の実施態様において、この方法が識別に基づいてコンテンツを移行するステップ；及びスピンダウンユニットの各々の状態を、コンテンツを記憶するためのシステムニーズに基づいてパワーアップまたはパワーダウンの適切な状態に管理するステップ、ならびに監視、識別及び移行の結果として、コンテンツを移行するステップを更に含む。この方法が、記憶サブシステム内の稼働ユニット及びスピンダウンユニットを含む記憶ユニットの異なる記憶クラスの組合せを利用する記憶階層化戦略、及び記憶ユニットの異なる記憶クラス内のコンテンツの記憶のための記憶されたコンテンツの適格性を判定するポリシーを設定する記憶階層化ルールを各々特定する複数のサービスプランからサービスプランを選ぶステップを更に含む。

【００１０】

本発明の別の態様が、スピンダウン機能を可能にしない稼働ユニット及びスピンダウン機能を可能にするスピンダウンユニットを含む複数の記憶ユニット内にコンテンツを記憶する複数の独立ノードを有するコンテンツシステム内のコンテンツの記憶を管理するための器具を目的とする。器具は、プロセッサ、メモリ及び記憶階層化サービスモジュールを備える。記憶階層化サービスモジュールが：どんなコンテンツが稼働ユニット上に記憶されるべき資格があるか、及びどんなコンテンツがスピンダウンユニット上に記憶されるべき資格があるかを示すポリシーを設定する記憶階層化ルールであって、記憶ユニット内の記憶及び記憶ユニット間の移行のためにそれの記憶されたコンテンツの適格性を判定するためにコンテンツシステム内の少なくとも一群のコンテンツに適用可能な記憶階層化ルールを確立し；コンテンツシステムの状態を監視し；かつ、記憶階層化ルール、コンテンツシステムの状態及び少なくとも一群のコンテンツの記憶されたコンテンツの適格性に基づいて、稼働ユニットとスピンダウンユニットとの間の移行を含む記憶ユニット間を移行するコンテンツの候補を識別するように構成される。

【００１１】

実施態様によっては、この器具が識別に基づいてコンテンツを移行するように構成される移行モジュール、及びコンテンツを記憶し、かつ監視、識別及び移行の結果として、コ

ンテンツを移行するためのシステムニーズに基づいて、スピンダウンユニットの各々の状態をパワーアップまたはパワーダウンの適切な状態に管理するように構成される記憶ユニット状態管理モジュールを更に備える。この器具は、記憶サブシステム内の稼働ユニット及びスピンダウンユニットを含む記憶ユニットの異なる記憶クラスの組合せを利用する記憶階層化戦略、ならびに記憶ユニットの異なる記憶クラス内のコンテンツの記憶のための記憶されたコンテンツの適格性を判定するポリシーを設定する記憶階層化ルールを各々特定する複数のサービスプランからサービスプランを選ぶためのユーザインタフェースを提供するように構成されるサービスプラン選択モジュールを更に備える。スピンダウン機能を可能にするように構成されないＲＡＩＤ群上に稼働ユニットが収容され、そして、スピンダウンを可能にするように構成されるＲＡＩＤ群上にスピンダウンユニットが収容されている。記憶ユニットは、異なる信頼性、性能またはコスト特性の少なくとも１つを有する異なるクラスの記憶ユニットを含む。この器具は、識別に基づいてコンテンツを移行し、かつコンテンツのメタデータ基準に基づいて異なるクラスの稼働ユニットとスピンダウンユニットとの間でコンテンツを移動させるように構成される移行モジュールを更に備え、このオブジェクトメタデータ基準が「データのタイプ」、「最後のアクセスからの時間」、「取込からの時間」及び「コンテンツのバージョン」の１つ以上を含む。

【００１２】

本発明の別の態様が、スピンダウン機能を可能にしない稼働ユニット及びスピンダウン機能を可能にするスピンダウンユニットを含む複数の記憶ユニット内にコンテンツを記憶する複数の独立ノードを有するコンテンツシステム内のコンテンツの記憶を管理するデータプロセッサを制御するための複数の命令を記憶するコンピュータ可読の記憶媒体を目的とする。この複数の命令が：どんなコンテンツが稼働ユニット上に記憶されるべき資格があるか及びどんなコンテンツがスピンダウンユニット上に記憶されるべき資格があるかを示すポリシーを設定する記憶階層化ルールであって、記憶ユニット内の記憶及び記憶ユニット間の移行のためにそれの記憶されたコンテンツの適格性を判定するようにコンテンツシステム内の少なくとも一群のコンテンツに適用可能な記憶階層化ルールをデータプロセッサに確立させる命令；コンテンツシステムの状態をデータプロセッサに監視させる命令；ならびに記憶階層化ルール、コンテンツシステムの状態及び少なくとも一群のコンテンツの記憶されたコンテンツの適格性に基づいて、稼働ユニットとスピンダウンユニットとの間の移行を含む記憶ユニット間を移行するコンテンツの候補をデータプロセッサに識別させる命令を含む。

【００１３】

本発明のこれらの、そしてまた他の、特徴及び効果が、特定の実施態様の以下の詳細な説明を考慮して当業者に明白になるであろう。

【図面の簡単な説明】

【００１４】

【図１】本発明の方法と器具が適用されることができる固定コンテンツ記憶アーカイブの簡略ブロック図である。

【図２】その各々が対称性でかつアーカイブクラスタアプリケーションをサポートする独立ノードの冗長配列の簡単にされた表現である。

【図３】所定のノード上で実行するアーカイブクラスタアプリケーションの種々のコンポーネントのハイレベル表現である。

【図４】固定コンテンツを記憶するためにブロックベースの記憶サブシステムに連結される一群のノードを有する固定コンテンツシステムの簡略図である。

【図５】記憶階層化サービス（ＳＴＳ）モジュール、ノード内の記憶媒体ユニット及びそれらの状態をリストする記憶媒体状態テーブルならびに記憶マネージャの複数のインスタンスを有するノードに関して略図で例示する簡略図である。

【図６】記憶階層化のためのコンテンツ選択のプロセスを例示する一例を示す。

【図７】ＳＴＳモジュールによって実行されるＳＴＳを例示する流れ図の一例である。

【図８】（ａ）サービスプランを作り出してかつ階層化ポリシーを選ぶためのスクリーン

ショット、（ｂ）作り出されるべきネームスペースのためのサービスプランを選ぶ時何が起こるかについて示すスクリーンショット及び（ｃ）既存のネームスペース上のサービスプランをどのように変更するべきかについて示すスクリーンショットを含む記憶階層化戦略を各々特定する複数のサービスプランからサービスプランを選ぶためのユーザインタフェースの一例を示す。

【発明を実施するための形態】

【００１５】

本発明の以下の詳細な説明では、参照が開示の一部を形成する添付の図面になされ、及び本発明が実施されることができる例示的な実施態様を例証としてかつ限定でなく示す。図において、同様な数字がいくつかの図の全体にわたって実質的に類似したコンポーネントを記述する。更に、詳細な説明が種々の例示的な実施態様を提供するとはいえ、後述するようにかつ図面内に図示するように、本発明は、本願明細書に図と共に説明される実施態様に限定されないが、しかし、公知であるように、または当業者にとって公知になるように他の実施態様に拡張することができる。「一実施態様」、「この実施態様」、または「これらの実施態様」への明細書における参照は、実施態様と関連して記述される特定の特徴、構造または特性が本発明の少なくとも１つの実施態様内に含まれることを意味し、及び、明細書内の種々の場所におけるこれらの句の出現が、必ずしも全て同じ実施態様を参照するというわけではない。加えて、以下の詳細な説明において、数多くの具体的な詳細が本発明の徹底的な理解を提供するために記載される。しかしながら、これらの具体的な詳細が、本発明を実践するために全てが必要であるというわけではないことは、当業者に明白である。他の状況では、本発明を不必要に不明瞭にしないために、周知の構造、材料、回路、プロセス及びインタフェースは詳述せず、及び／またはブロック図形式で例示されることができる。

【００１６】

さらに、あとに続く詳細な説明のいくつかの部分が、コンピュータ内の操作のアルゴリズム及び記号表示に関して提示される。これらのアルゴリズム的記述及び記号表示は、データ処理技術の当業者によってそれらの革新の本質を最も効果的に他の当業者に伝えるために使用される手段である。アルゴリズムは、所望の終了状態または結果に至る一連の規定されたステップである。本発明において、実施されるステップは、具体的な結果を達成するための具体的な数量の物理操作を必要とする。通常、必然的にではないとはいえ、これらの数量は、記憶され、伝達され、組み合わせられ、比較され、かつさもなければ操作されることが可能な電気もしくは磁気信号または命令の形式をとる。これらの信号をビット、値、要素、シンボル、文字、用語、数、命令、などとして参照することは、主に一般的な使用の理由で折に触れて都合がいいと立証されている。しかしながら、これら及び類似の用語の全てが適切な物理量と関連しているべきであり、かつこれらの数量に適用される単に都合がいいラベルであるだけであることが念頭に置おかれるべきである。とりわけ別の方法で述べられない限り、以下の説明から明白に認識されることは、記述の全体にわたって、「処理」、「計算」、「算出」、「判定」、「表示」などのような用語を利用する考察は、コンピュータシステムのレジスタ及びメモリ内で物理（電子）数量として表されるデータを、コンピュータシステムのメモリまたはレジスタまたは他の情報記憶、送信もしくはディスプレイ装置内で物理量として同じように表される他のデータに、操作してかつ変換するコンピュータシステムまたは他の情報処理装置のアクション及びプロセスを含むことができるということである。

【００１７】

本発明は、さらにここで操作を実行するための器具に関する。この器具は必要とされた目的のために特別に構成されることができるかまたは、それが１つ以上のコンピュータプログラムによって選択的に活性化されるかまたは再構成される１台以上の汎用コンピュータを含むことができる。この種のコンピュータプログラムは、限定されないが、光ディスク、磁気ディスク、読取り専用メモリ、ランダムアクセスメモリ、固体素子装置及びドライブまたは電子情報を記憶することに適しているその他のタイプの媒体のようなコンピュ

ータ可読の記憶媒体内に記憶されることができる。本願明細書に提示されるアルゴリズム及び表示は、いかなる特定のコンピュータまたは他の器具にも本質的に関連がない。種々の汎用システムが、本願明細書における教示に従うプログラム及びモジュールとともに使用されることができるか、または、所望の方法ステップを実行するためにより専門の器具を構成することが、都合がいいと立証されることができる。加えて、本発明は任意の特定のプログラミング言語を参照して記述されない。理解されるであろうことは、本願明細書に記述されるように、種々のプログラミング言語が本発明の教示を実現するために使用されることができるということである。プログラミング言語（複数言語）の命令は、１台以上の処理装置、例えば中央処理ユニット（ＣＰＵ）、プロセッサまたはコントローラによって実行されることができる。

【００１８】

後で詳しく述べるように、本発明の例示的な実施態様は、記憶階層化のためのコンテンツ選択に基づいて固定コンテンツシステム内の固定コンテンツの記憶を管理するための器具、方法及びコンピュータプログラムを提供する。

【００１９】

Ｉ．固定コンテンツ分散データ記憶

【００２０】

従来のテープ及び光記憶解法を置換するかまたは補充する高度に利用可能な、信頼性が高いかつ持続的な方法での、「固定コンテンツ」のアーカイブ記憶に対するニーズが出現した。用語「固定コンテンツ」は、一般的に参照または他の目的のために変更なしで保持されるのを期待される任意のタイプのデジタル情報を指す。この種の固定コンテンツの例は多くの他の中に、電子メール、文書、診断画像、チェック画像、音声録音、フィルム及びビデオ、などを含む。従来の独立ノードの冗長配列（ＲＡＩＮ）記憶アプローチは、この種の固定コンテンツ情報資産の記憶用の大きなオンラインアーカイブを作り出すための選り抜きのアーキテクチャとして現れた。ノードを必要に応じてクラスタに接続してかつそれから出ることを可能にすることによって、ＲＡＩＮアーキテクチャは１つ以上のノードの故障から記憶クラスタを遮断する。複数ノード上にデータを複製することによって、ＲＡＩＮタイプアーカイブはノード故障または除去を自動的に補正することができる。一般的に、ＲＡＩＮシステムは主として閉システム内の同一のコンポーネントから設計されたハードウェア器具として供給される。

【００２１】

図１は、そのような拡張可能なディスクベースのアーカイブ記憶管理システムを例示する。ノードは、異なるハードウェアを備えることができてしたがって、「異種」とみなされることができる。ノードは一般的に、実際の物理記憶ディスクまたは記憶領域ネットワーク（ＳＡＮ）内における仮想記憶ディスクであることができる１台以上の記憶ディスクにアクセスする。各ノード上でサポートされるアーカイブクラスタアプリケーション（及び、任意選択で、そのアプリケーションが実行する基本オペレーティングシステム）は同じであるかまたは実質的に同じであることができる。各ノード上のソフトウェアスタック（それはオペレーティングシステムを含むことができる）は、対称であるが、一方、ハードウェアは異質であることができる。このシステムを使用して、図１にて図示したように、とりわけ、企業は文書、電子メール、衛星画像、診断画像、チェック画像、音声録音、ビデオ、などのような多くの異なるタイプの固定コンテンツ情報用のパーマネント記憶を作り出すことができる。もちろん、これらのタイプは、単に例証となるだけである。高レベルの信頼性が、独立サーバまたはいわゆる記憶ノード上にデータを複製することによって達成される。好ましくは、各ノードはそのピアと対称である。したがって、好ましくは、任意の所定のノードが全ての機能を実行することができるので、任意の１つのノードの故障はアーカイブの利用可能度にほとんど影響を及ぼさない。

【００２２】

自己の米国特許第７，１５５，４６６号明細書にて説明したように、デジタル資産を収集して、保存して、管理してかつ読み出す各ノード上で実行される分散ソフトウェアアプ

リケーションを組み込むことは、ＲＡＩＮベースのアーカイブのシステムにおいて公知である。図２は、そのようなシステムを例示する。個々のアーカイブの物理境界は、クラスタ（またはシステム）と称する。一般的に、クラスタは単一装置でなく、しかし、むしろ一まとまりの装置である。装置は、均質であるかまたは異質であることができる。典型的装置は、Ｌｉｎｕｘのようなオペレーティングシステムを実行するコンピュータまたは機械である。コモディティーハードウェア上でホストされるＬｉｎｕｘベースのシステムのクラスタは、２、３の記憶ノードサーバから何千ものテラバイト単位のデータを記憶する多くのノードまでスケールアップされることができるアーカイブを提供する。このアーキテクチャは、記憶容量が組織の増大するアーカイブ要件と常に足並みをそろえることができることを確実にする。

【００２３】

上記したような記憶システムにおいて、データは一般的にアーカイブが装置故障から常に保護されるようにクラスタ全体にランダムに分散される。ディスクまたはノードが故障する場合、クラスタは、同じデータの複製を維持するクラスタ内の他のノードまで自動的に故障する。このアプローチがデータ保護の見地からよく機能するとはいえ、クラスタのデータロスに対する算出された平均時間（ＭＴＤＬ）が要望するほど高くないかもしれない。特に、ＭＴＤＬは一般的にアーカイブがデータを失う前に算出された時間を表す。デジタルアーカイブにおいて、いかなるデータロスも望ましくないが、しかし、ハードウェアコンポーネント及びソフトウェアコンポーネントの性質に起因して、（いかにかすかでも）この種の発生の可能性が常にある。アーカイブクラスタ内のオブジェクト及びそれらのコピーのランダム分布の理由でたとえば、所定のノード内の所定のディスク（そこではミラーコピーが記憶される）が予想外に故障する場合、オブジェクトの必要とされたコピーが利用できないかもしれないので、ＭＴＤＬは結局必要とされるより低いことになるかもしれない。

【００２４】

図２に示すように、本発明が実現される例証となるクラスタが好ましくは、以下の全般的なカテゴリのコンポーネントを備える：ノード２０２、一対のネットワークスイッチ２０４、配電器（ＰＤＵ）２０６及び無停電電源（ＵＰＳ）２０８。ノード２０２は、一般的に１台以上のコモディティーサーバを備えてかつＣＰＵ（例えばインテルｘ８６）、適切なランダムアクセスメモリ（ＲＡＭ）、１台以上のハードディスク（例えば標準ＩＤＥ／ＳＡＴＡ、ＳＣＳＩ、など）及び２枚以上のネットワークインタフェースカード（ＮＩＣ）を収容している。典型的ノードは、２．４ＧＨｚのチップ、５１２ＭＢ　ＲＡＭ及び６台の（６台の）２００ＧＢハードディスクを備えた２Ｕラックマウントユニットである。しかしながら、これは限定でない。ネットワークスイッチ２０４は、一般的にノード間のピアツーピア通信を可能にする内部スイッチ２０５及び各ノードへの追加クラスタアクセスを可能にする外部スイッチ２０７を備える。各スイッチは、クラスタ内の全ての可能性があるノードを取り扱うのに十分なポートを必要とする。イーサネットまたはＧｉｇＥスイッチが、このために使用されることができる。ＰＤＵ　２０６は全てのノード及びスイッチを動かすために使用され、全てのノード及びスイッチを保護するＵＰＳ２０８が使用される。限定するつもりではないとはいえ、一般的に、クラスタは公共インターネット、企業イントラネットまたは他の広域もしくはローカルエリアネットワークのようなネットワークに接続可能である。例示の実施態様では、クラスタは企業環境内に実現される。たとえばサイトの会社ＤＮＳネームサーバを通してナビゲートすることによって、それが到達されることができる。したがって、例えば、クラスタのドメインは既存のドメインの新規のサブドメインであることができる。代表実現において、サブドメインは会社ＤＮＳサーバにおいてクラスタ自体のネームサーバに委任される。エンドユーザは、任意の従来インタフェースまたはアクセスツールを使用してクラスタにアクセスする。したがって、例えば、クラスタへのアクセスは、任意のＩＰベースのプロトコル（ＨＴＴＰ、ＦＴＰ、ＮＦＳ、ＡＦＳ、ＳＭＢ、ウェブサービス、など）を介して、ＡＰＩ経由で、またはその他の周知のもしくは後で開発されたアクセス方式、サービス、プログラムもしくはツール

を通して実施されることができる。

【００２５】

クライアントアプリケーションは、標準ＵｎｉｘファイルプロトコルまたはＨＴＴＰＡＰＩのような１つ以上のタイプの外部ゲートウェイを通してクラスタにアクセスする。アーカイブは好ましくは任意の標準Ｕｎｉｘファイルプロトコル志向ファシリティの下に任意選択で位置することができる仮想ファイルシステムを通して露出される。これらは、ＮＦＳ、ＦＴＰ、ＳＭＢ／ＣＩＦＳ、などを含む。

【００２６】

一実施態様において、アーカイブクラスタアプリケーションがクラスタとして共にネットワーク化される（例えばイーサネット経由で）独立ノードの冗長配列（Ｈ－ＲＡＩＮ）上で動作する。所定のノードのハードウェアは、異質であることができる。しかしながら最大信頼性のために好ましくは各ノードが、次に図３内に例示されるようにいくつかのランタイムコンポーネントを備える、分散アプリケーションのインスタンス３００（それは同じインスタンスまたは実質的に同じインスタンスであることができる）を実行する。したがって、ハードウェアが異質であることができるとはいえ、（少なくともそれが本発明に関する）ノード上のソフトウェアスタックは同じである。これらのソフトウェアコンポーネントは、ゲートウェイプロトコルレイヤ３０２、アクセスレイヤ３０４、ファイルトランザクション及び管理レイヤ３０６及びコアコンポーネントレイヤ３０８を備える。機能が他の意味がある方法で特徴づけられることができると当業者が認識するように、「レイヤ」名称は説明的な目的のために提供される。レイヤ（またはその中のコンポーネント）の１つ以上が、一体化されるかまたはその逆であることができる。いくつかのコンポーネントが、レイヤ全体に共有されることができる。

【００２７】

ゲートウェイプロトコルレイヤ３０２内のゲートウェイプロトコルが、既存のアプリケーションへの透明性を提供する。特に、ゲートウェイはＮＦＳ　３１０及びＳＭＢ／ＣＩＦＳ　３１２のようなネイティブファイルサービス、同じくカスタムアプリケーションを構築するウェブサービスＡＰＩを提供する。ＨＴＴＰサポート３１４が、さらに提供される。アクセスレイヤ３０４が、アーカイブへのアクセスを提供する。特に、本発明によれば、固定コンテンツファイルシステム（ＦＣＦＳ）３１６がアーカイブオブジェクトへの完全なアクセスを提供するネイティブファイルシステムをエミュレートする。ＦＣＦＳは、あたかもそれらが通常ファイルであるかのように、アーカイブコンテンツへのアプリケーションの直接アクセスを与える。好ましくは、アーカイブされたコンテンツがその本来の形態で表示され、一方メタデータがファイルとして露出される。彼らによく知られている方法で、管理者が固定コンテンツデータを供給することができるように、ＦＣＦＳ３１６はディレクトリ及びパーミッション及びルーチンファイルレベル呼出しの従来のビューを提供する。ファイルアクセス呼出しは好ましくはユーザスペースデーモンによってインターセプトされてかつ呼出しアプリケーションに対する適切なビューを動的に作り出す（レイヤ３０８内の）適切なコアコンポーネントに経由される。ＦＣＦＳ呼出しは好ましくは、自律性のアーカイブ管理を容易にするためにアーカイブポリシーによって限定される。したがって、一例では、管理者またはアプリケーションはその保持期間（所定のポリシー）がなお有効であるアーカイブオブジェクトを削除することができない。

【００２８】

アクセスレイヤ３０４は、好ましくはさらにウェブユーザインタフェース（ＵＩ）３１８及びＳＮＭＰゲートウェイ３２０を含む。ファイルトランザクション及び管理レイヤ３０６内の管理エンジン３２２への対話型アクセスを提供する管理者コンソールとして、ウェブユーザインタフェース３１８が好ましくは実現される。管理コンソール３１８は好ましくは、アーカイブオブジェクト及び個々のノードを含むアーカイブの動的ビューを提供するパスワードで保護されたウェブベースのＧＵＩである。ＳＮＭＰゲートウェイ３２０は管理エンジン３２２への記憶管理アプリケーションの簡単なアクセスを提供し、それらがクラスタ活動をしっかりと監視して制御することを可能にする。管理エンジンはシステ

ム及びポリシーイベントを含むクラスタ活動を監視する。ファイルトランザクション及び管理レイヤ３０６は、さらにリクエストマネージャプロセス３２４を含む。リクエストマネージャ３２４は、（アクセスレイヤ３０４を通して）外界からの全てのリクエスト、同じくコアコンポーネントレイヤ３０８内のポリシーマネージャ３２６からの内部リクエストを組織化する。

【００２９】

ポリシーマネージャ３２６に加えて、コアコンポーネントはさらにメタデータマネージャ３２８及び記憶マネージャ３３０の１つ以上のインスタンスを含む。メタデータマネージャ３２８は、好ましくは各ノード上にインストールされる。集合的に、クラスタ内のメタデータマネージャが分散型データベースとして働き、全てのアーカイブオブジェクトを管理する。所定のノード上で、メタデータマネージャ３２８がアーカイブオブジェクトのサブセットを管理し、そこで好ましくは、各オブジェクトが、外部ファイル（「ＥＦ」、記憶用のアーカイブに入れられたデータ）とアーカイブデータが物理的に位置する一組の内部ファイル（各「ＩＦ」）との間をマップする。同じメタデータマネージャ３２８が、さらに他のノードから複製された一組のアーカイブオブジェクトを管理する。したがって、すべての外部ファイルの現在の状態がいくつかのノード上の複数のメタデータマネージャに常に利用可能である。ノード故障の場合には、他のノード上のメタデータマネージャが、故障したノードによって以前に管理されていたデータへのアクセスを提供し続ける。記憶マネージャ３３０は、分散アプリケーション内の全ての他のコンポーネントに利用可能なファイルシステムレイヤを提供する。好ましくは、それがノードのローカルファイルシステム内にデータオブジェクトを記憶する。所定のノード内の各ドライブが、好ましくはそれ自体の記憶マネージャを有する。これは、ノードが個別ドライブを除去してスループットを最適化することを可能にする。記憶マネージャ３３０は、さらにシステム情報、データに関する完全性チェック及びローカル構造を直接横断する能力を提供する。

【００３０】

さらに、図３内に例示されるように、クラスタは通信ミドルウェアレイヤ３３２及びＤＮＳマネージャ３３４を通して内部及び外部通信を管理する。インフラストラクチャ３３２は、アーカイブコンポーネントの間の通信を可能にする効率的なかつ信頼性が高いメッセージベースのミドルウェアレイヤである。例示の実施態様では、レイヤは、マルチキャスト及びポントツーポイント通信をサポートする。ＤＮＳマネージャ３３４は、全てのノードを企業サーバに接続する分散型のネームサービスを実行する。好ましくは、ＤＮＳマネージャ（単独でまたはＤＮＳサービスと共に）負荷バランスが、全てのノードにわたって最大クラスタスループット及び利用可能度を確実にすることを要求する。

【００３１】

例示の実施態様では、ＨＣＰ（日立コンテンツプラットホーム）アプリケーションインスタンスのようなアプリケーションが、Ｒｅｄ Ｈａｔ Ｌｉｎｕｘ ９．０、フェドーラコア６、などのような、基本オペレーティングシステム３３６上で実行する。通信ミドルウェアは、任意の都合がいい分散通信機構である。他のコンポーネントがＦＵＳＥ（ユーザスペース内のファイルシステム）を含むことができ、それが固定コンテンツファイルシステム（ＦＣＦＳ）３１６に対して使用されることができる。ＮＦＳゲートウェイ３１０は、標準ｎｆｓｄ ＬｉｎｕｘカーネルＮＦＳドライバによって実現されることができる。各ノード内のデータベースは、例えばＰｏｓｔｇｒｅＳＱＬ（また、本願明細書でＰｏｓｔｇｒｅｓと称される）を実現されることができ、それはオブジェクトリレーショナルデータベース管理システム（ＯＲＤＢＭＳ）である。ノードは、Ｊａｖａ ＨＴＴＰサーバ及びサーブレットコンテナであるＪｅｔｔｙのようなウェブサーバを含むことができる。もちろん、上記の機構は単に例証となるだけである。

【００３２】

所定のノード上の記憶マネージャ３３０は、物理記憶装置を管理する役割を果たす。好ましくは、各記憶マネージャインスタンスが、全てのファイルがそのプレースメントアルゴリズムに従って配置される単一ルートディレクトリの原因となる。複数の記憶マネージ

ャインスタンスが同時にノード上で実行することができ、各々が通常、システム内の異なる物理ディスクを代表する。記憶マネージャは、残りのシステムから使用されるドライブ及びインタフェース技術を分離する。記憶マネージャインスタンスがファイルを書き込むよう依頼される時、それはそれが原因である表現に対するフルパス及びファイルネームを生成する。代表実施態様において、記憶マネージャ上に記憶されるべき各オブジェクトは、次いで異なるタイプの情報を得続けるためにデータを記憶するにつれてファイルにそれ自体のメタデータを追加する記憶マネージャによって記憶されるべき生データとして収容される。例証として、このメタデータは以下を含む：ＥＦ長（バイトでの外部ファイルの長さ）、ＩＦセグメントサイズ（内部ファイルのこの部分のサイズ）、ＥＦ保護表現（ＥＦ保護モード）、ＩＦ保護役割（この内部ファイルの表現）、ＥＦ作成タイムスタンプ（外部ファイルタイムスタンプ）、シグネチャ（シグネチャタイプを含む書込みの時間（ＰＵＴ）での内部ファイルのシグネチャ）及びＥＦファイルネーム（外部ファイルファイルネーム）。内部ファイルデータと共にこの追加的なメタデータを記憶することは、追加的なレベルの保護を提供する。特に、スカベンジングは内部ファイル内に記憶されるメタデータからデータベース内に外部ファイルレコードを作り出すことができる。他のポリシーは、内部ファイルが損なわれていないままであることを確認するために内部ファイルに対して内部ファイルハッシュを確認することができる。

【００３３】

内部ファイルはアーカイブオブジェクト内の本来の「ファイル」の一部を表すデータの「チャンク」であることができ、及び、それらはストライピング及び保護ブロックを達成する異なるノード上に配置されることができる。より小さいチャンクユニットへの外部ファイルのこの分裂は、要件でないが、しかしながら；代替案では、内部ファイルが外部ファイルの完全コピーであることができる。一般的に、各外部ファイルエントリに対して多くの内部ファイルエントリがあることができる一方、１つの外部ファイルエントリが各アーカイブオブジェクトに対してメタデータマネージャ内にある。一般的に、内部ファイルレイアウトはシステムに依存する。所定の実現において、ディスク上のこのデータの実際の物理フォーマットが一連の可変長レコード内に記憶される。

【００３４】

リクエストマネージャ３２４は、システム内の他のコンポーネントと対話することによってアーカイブアクションを実行するために必要な操作の組を実行する役割を果たす。リクエストマネージャは、異なるタイプの多くの同時アクションをサポートして、任意の失敗したトランザクションをロールバックすることが可能であり、かつ実行するのに長い間かかる可能性があるトランザクションをサポートする。リクエストマネージャは、更にアーカイブ内の読出し／書込み操作が適切に取り扱われることを確実にしてかつ全てのリクエストがいつでも既知の状態にあることを保証する。それはさらに、所定のクライアントリクエストを満足させるためにノードにわたって複数の読出し／書込み操作を調整するためのトランザクション制御を提供する。加えて、リクエストマネージャは最近使用されたファイルに対するメタデータマネージャエントリをキャッシュにいれてかつセッション、同じくデータブロックに対するバッファリングを提供する。

【００３５】

クラスタの主要な応答性は、ディスク上に無制限の数のファイルを確実に記憶することである。所定のノードは、それがどんな理由にせよ到達できないかまたはさもなければ利用できなくなるかもしれないという意味で、「信頼できない」と考えられることができる。一まとまりのこの種の潜在的に信頼できないノードが、信頼性が高いかつ高度に利用可能な記憶域を作り出すために共同する。概ね、記憶される必要がある情報の２つのタイプ：ファイルそれ自体及びファイルについてのメタデータがある。固定コンテンツ分散データ記憶の追加的な詳細が、米国特許出願公開第２００７／０１８９１５３号明細書及び米国特許第７，６５７，５８１号明細書内に見つけることができ、それらを本願明細書に引用したものとする。

【００３６】

ここで使用しているように、ネームスペースはクラスタの論理パーティションであってかつ少なくとも１つの規定されたアプリケーションに特定の一まとまりのオブジェクトとして基本的に機能する。各ネームスペースは、他のネームスペースに対して私的ファイルシステムを有する。さらに、１つのネームスペースへのアクセスは別のネームスペースへのユーザアクセスを許可しない。ノードのクラスタ／システムは、物理アーカイブインスタンスである。テナントは、ネームスペース（複数スペース）及びおそらく他のサブテナントのグループ化である。クラスタ／システムは、物理アーカイブインスタンスである。同一譲受人の米国特許出願公開第２０１１／０１０６８０２号明細書を参照されたい、それを全体として本願明細書に引用したものとする。

【００３７】

ＩＩ　．記憶階層化のためのコンテンツ選択

【００３８】

例示的な実施態様によれば、固定コンテンツシステム（ＦＣＳ）４００は、一般的に、図４に示すように固定コンテンツを記憶する複数の記憶媒体ユニット４３０を有するブロックベースの記憶サブシステム４２０にネットワーク経由で連結される一群のノード４１０を有する。好ましい実施態様において、記憶サブシステムは、一群のディスクがＲＡＩＤ群に共に構成されてかつ記憶ニーズを満足させるために個々の論理ディスクユニットに切り分けられることを可能にするバックエンドディスク配列である。より高度な記憶サブシステムが、どのディスクまたはディスクの群が、ディスクのパワーアップ及びダウンを可能にするためにＲＡＩＤ群に制御されるパワーを有することを可能にされるかを示す能力を有する。

【００３９】

本発明の実施態様は、稼働ユニット（ＲＵ）及びスピンダウンユニット（ＳＤＵ）と呼ばれる２つのクラスの論理ディスクユニットを作り出すためにこれらの記憶サブシステム特徴を利用する。パワーが常にディスクにあり、かつデータアクセスに利用可能なように、ディスクスピンダウン機能を可能にするように構成されないＲＡＩＤ群上に、ＲＵが収容される。ディスクがパワーダウンされることができ、及び、それゆえに、ディスクが再起動されてかつスピンアップされるまでそれらのディスク上のデータが直ちに利用可能でないように、ディスクスピンダウンを可能にするように構成されるＲＡＩＤ群上に、ＳＤＵが収容される。

【００４０】

ＲＵ及びＳＤＵがどのように使用されるかを規定するために、記憶階層化ルール（ＳＴＲ）が確立される。適切なＳＴＲは、ＦＣＳ上に記憶される予想される使用及びデータのライフサイクルから決定される。ＳＴＲは、どんなコンテンツがＲＵまたはＳＤＵ上に記憶されるべき資格があるかを示す。一実施態様によれば、ＦＣＳ上に記憶される全てのコンテンツまたは一群のコンテンツのどちらかに適用される規定される３つのＳＴＲがある。第１のＳＴＲは、「けっしてない」であり、ＲＵ上にまたはスピンダウンされないＳＤＵ指示された記憶域上に記憶されるべき全てのデータを指す。第２のＳＴＲは、「保護コピーだけ」であってかつ保護コピーを指す。固定コンテンツシステムでは、コンテンツはコンテンツの複数コピーを記憶することによって保護されることができる。一般的に、バックアップコピーまたは複数コピーはまずアクセスされない。これらのバックアップコピーは、保護コピーとしてのＳＤＵの候補である。第３のＳＴＲは、「取込後Ｘ時間」であってかつＳＤＵ上に記憶されるべき候補になるのに十分な一定の時間の間ＦＣＳ内に存在したコンテンツを指す。

【００４１】

割り当てられるＳＴＲに関係なく、ＲＵが設定可能な使用されたスペース／記憶容量しきい値に到達した時にだけ、ＳＤＵに資格のあるコンテンツを移行するように、ＦＣＳがさらに構成されることができる。任意の資格のあるコンテンツが、追加的なコストを伴わずに資格のないコンテンツと共にＲＵ上にとどまることができる；さらに、それをＲＵ上に残すより、資格のあるコンテンツを移行するためにＳＤＵをパワーアップすることはあ

まり有益でないかもしれない。

【００４２】

ＳＴＲ構成とともにＲＵ及びＳＤＵを提供することによって、ＦＣＳは記憶階層化サービス（ＳＴＳ）を使用してＦＣＳのコンテンツ上で操作することが可能である。ＳＴＳは、定期的に実行してかつ以下の活動の原因となる：（１）システムニーズに基づいてＳＤＵの状態をパワーアップまたはパワーダウンの適切な状態に管理する；及び（２）ＳＴＲ構成、ＦＣＳの状態及び記憶されたコンテンツの適格性に基づいてＲＵとＳＤＵとの間でコンテンツを移行する。記憶媒体ユニット（稼働ユニット及びスピンダウンユニット）の状態を含む、ＦＣＳの状態が監視される。特定の実施態様において、ＳＴＳがＦＣＳ内の各ノード内に提供される。

【００４３】

図５は、ＳＴＳモジュール３４０、ノードに対する記憶媒体ユニット及びそれらの状態（例えば容量）をリストする記憶媒体状態テーブル３５０及び記憶マネージャ３３０の複数のインスタンス（図３を参照のこと）を有するノード（図３の３００または図４の４１０）に関して略図で例示する簡略図である。例えば、記憶マネージャ３３０の各インスタンスが、全てのファイルがそのプレースメントアルゴリズムに従って配置される単一ルートディレクトリの原因となる。各記憶マネージャインスタンスは、通常ＦＣＳ内の異なる論理ボリュームを表す。図５は、記憶マネージャ３３０のインスタンス（一般的にネットワークを通しての接続）経由でノード３００と記憶ユニット４３０（ＲＵ及びＳＤＵ）との間の接続を示す。好ましい一実施態様では、全てのノードが、全てのノードに対する記憶媒体ユニット及びそれらの状態をリストする同じ記憶媒体状態テーブル３５０を記憶する。記憶媒体ユニットの状態は別々に監視されることができ、及びその情報が次いで同じ記憶媒体状態テーブル３５０を形成するために相互接続されたノードの間で共有される。ＳＴＳモジュール３４０は、以下の論理アクションを有することができる：

【００４４】

ＳＤＵへの移動に適格なコンテンツのＭｏｖｅＴｏＳＤＵリストを構築する。

【００４５】

ＳＤＵに適格でない以外のＳＤＵ上のコンテンツのＭｏｖｅＦｒｏｍＳＤＵリストを構築する。

【００４６】

構成される容量しきい値に到達したＲＵ上に存在するコンテンツだけを収容するためにＭｏｖｅＴｏＳＤＵリストにフィルターをかける。

【００４７】

コンテンツがＭｏｖｅＴｏＳＤＵ及びＭｏｖｅＦｒｏｍＳＤＵリスト内に存在する全てのＳＤＵのＳｐｉｎＵｐＳＤＵリストを構築する。

【００４８】

ＳｐｉｎＵｐＳＤＵリスト内のＳＤＵをスピンアップする。

【００４９】

ＭｏｖｅＴｏＳＤＵ及びＭｏｖｅＦｒｏｍＳＤＵリストを処理するために、適切な宛先への移行処理を開始する。

【００５０】

図６は、記憶階層化のためのコンテンツ選択のプロセスを例示する一例を示す。図６では、ＦＣＳはクラスタにわたりノード全体に広がるＲＵ用の３台のディスク及びＳＤＵ用の２台のディスクを有する。保護コピーがＳＤＵ上に配置されるべきことを示すサービスプランによって、ＦＣＳが構成される。コンテンツはＦＣＳ用の２のデータ保護レベル（ＤＰＬ）を有し、及びＦＣＳ上に２つのコピーを備えた３つのオブジェクト（Ａ、Ｂ、Ｃ）がある（Ａ１、Ａ２、Ｂ１、Ｂ２、Ｃ１、Ｃ２）。第１の２台のＲＵ（Ａ１Ｂ２Ｃ２及びＢ１Ｃ１）は、それらの消費閾値より上にある。ＳＴＳは、ＳＤＵへの移動の候補を処理するためにＦＣＳを走査する。走査に基づいて、それらが二次保護コピーであってかつ容量しきい値より上にある第１のＲＵ上に存在するので、コピーＢ２及びＣ２が移動され

る。コンテンツの一次コピー（Ｂ１及びＣ１）だけがそこに存在するので、何のコンテンツも第２のＲＵから移動されない。それが存在する第３のＲＵが容量しきい値より上にならないので、第２のコピーのＡ２は移動されない。

【００５１】

一般に、ＳＤＵへ移動されるオブジェクトは、サービスプランがオブジェクトに対して変更され、かつそれがサービスプランの下でもはや基準を満たさない限り、オブジェクトの寿命の間そこにとどまる。それから、オブジェクトがＲＵへ戻される。

【００５２】

本発明の代替実施態様は、さらに種々の信頼性、性能及びコスト特性を有する種々のクラスの記憶域にコンテンツを移行するために使用されることができる。異なるクラスの記憶域は、限定されるものではないが固体ディスク（ＳＳＤ）、Ｆｉｂｒｅ Ｃｈａｎｎｅｌハードディスクドライブ（ＦＣ－ＨＤＤ）または他のネットワーク記憶ユニットを含むことができる。

【００５３】

コンテンツ移動に対する判断は、システムによって自動的に生成されるかまたはコンテンツの所有者によって提供されるオブジェクトメタデータに基づく記憶または位置のクラスの高機能な選択を提供するように拡張されることができる。以下は、いくつかの例基準である。第一は、「データのタイプ」である。例えば、怪我が回復したあと、Ｘ線画像はまず観察されない、したがって、この種のタイプのデータはスピンダウンの良い候補であり、一方、全般的な医療記録は多用され、及びこの種のタイプのデータはスピンダウンの良い候補でないかもしれない。第２の基準は、「最後のアクセスからの時間」である。そのアクセス履歴に基づいてデータを移動することが望ましくなることができる。しばらく（例えば６ヵ月以上）アクセスされなかったコンテンツは、スピンダウンの良い候補である、一方よりしばしばアクセスされるコンテンツが装置を動かすために移動されることができる。第３の基準は、「取込からの時間」である。いくつかのタイプのデータを取込ですぐに（例えば、データがバックアップだけのためにある時）または取込の６ヵ月後にスピンダウンに移すことなどは意味をなすことができる。第４の基準は、「オブジェクトのバージョン」である。例えば、最新版以外のオブジェクトの全てのバージョンはスピンダウン上にあるべきである。

【００５４】

長期目標は、クラスタ管理者がテナント管理者に利用可能にするかまたは販売することができる異なるサービスプランをＦＣＳがサポートするためにある。サービスプランは、記憶クラス（例えば常にスピンアップされたディスク及びスピンダウンされたディスク）ならびにいつデータが異なるクラスの記憶域上で記憶されるかの一組のルールの組合せである記憶階層化戦略を有する。例えば、クラスタ管理者はその戦略内のデータが常にスピンアップされたディスク上にあるプレミアムサービスプランをサポートしたいかもしれない。「標準」サービスプランは６ヵ月の間スピンアップされたディスク上にデータを含むことができるが、しかし、６ヵ月後にアクセスされない場合、データはスピンダウンディスクへ移動される。「アーカイブ」サービスプランが、データアクセスがおそらくなく、かつ、コンテンツへのアクセスに対する遅延が受け入れられるという予想によってコンテンツをスピンダウンディスクの候補にすぐにすることができる。最終的に、目標は、クラスタ管理者が、サービスプランを規定してかつテナント管理者にそれらのプランを異なるレートで利用可能にするかまたは販売することが可能であるようにすることである。

【００５５】

図７は、ＳＴＳモジュール３４０によって実行される記憶階層化サービス（ＳＴＳ）を例示する流れ図の一例である。ステップ７０２において、ＳＴＳモジュール３４０がどんな固定コンテンツが稼働ユニット上に記憶されるべき資格があるか、及び、どんな固定コンテンツがスピンダウンユニット上に記憶されるべき資格があるかを示すポリシーを設定する記憶階層化ルールであって、記憶ユニット内の記憶及び記憶ユニット間の移行のためにそれの記憶されたコンテンツの適格性を判定するために固定コンテンツシステム内の少

なくとも一群の固定コンテンツに適用可能な記憶階層化ルールを確立する。これは、一般的に管理者からの入力によって実行される。特定の実施態様において、記憶階層化ルールが予想される使用、ライフサイクル及びコンテンツオブジェクトの経過時間及びコンテンツオブジェクトの１つ以上の冗長バックアップコピーの存在を含む一組の基準に基づいて記憶されたコンテンツの適格性を判定するためにコンテンツオブジェクトを評価するように確立される。さらに、記憶階層化ルールは、資格のあるコンテンツが存在する稼働ユニットが設定可能な消費利用閾値に到達した時にだけ資格のあるコンテンツをスピンダウンユニットに移行するために確立されることができる。

【００５６】

ステップ７０４では、ＳＴＳモジュール３４０が固定コンテンツシステムの状態を監視する。例えば、オブジェクトが存在する稼働ユニットがスピンダウンユニットにオブジェクトを移行するパーミッションを示す設定可能な消費利用閾値に到達したかどうか、及び稼働ユニットからオブジェクトを収容するためのスピンダウンユニット上で利用可能なスペースがあるかどうかを判定することをこれが含むことができる。

【００５７】

ステップ７０６では、ＳＴＳモジュール３４０が記憶階層化ルール、固定コンテンツシステムの状態及びこの少なくとも一群の固定コンテンツの記憶されたコンテンツの適格性に基づいて記憶媒体ユニット間を（例えば、稼働ユニットとスピンダウンユニットの間を）移行するコンテンツオブジェクトの候補を識別する。

【００５８】

ステップ７０８では、移行モジュール３４２（図５を参照のこと）が識別に基づいてコンテンツオブジェクトを移行する。ステップ７１０では、記憶ユニット状態管理モジュール３４４（図５を参照のこと）が、コンテンツオブジェクトを記憶してかつ監視、識別及び移行の結果として、コンテンツオブジェクトを移行するためのシステムニーズに基づいて、スピンダウンユニットの各々の状態をパワーアップまたはパワーダウンの適切な状態に管理する。

【００５９】

特定の実施態様では、記憶ユニットは異なる信頼性、性能またはコスト特性の少なくとも１つを有する異なるクラスの記憶ユニットを含む。移行モジュール３４２は、コンテンツオブジェクトのオブジェクトメタデータ基準であって、「データのタイプ」、「最後のアクセスからの時間」、「取込からの時間」及び「オブジェクトのバージョン」の１つ以上を含むオブジェクトメタデータ基準に基づいて異なるクラスの稼働ユニットとスピンダウンユニットとの間でコンテンツオブジェクトを移動するように構成される。

【００６０】

特定の実施態様に従って図７のＳＴＳプロセスを開始するために、サービスプラン選択モジュール３４６（図５を参照のこと）が、記憶サブシステム内の稼働ユニット及びスピンダウンユニットを含む記憶ユニットの異なる記憶クラスの組合せを利用する記憶階層化戦略、及び記憶ユニットの異なる記憶クラス内のコンテンツオブジェクトの記憶のための記憶されたコンテンツの適格性を判定するポリシーを設定する記憶階層化ルールを各々特定する複数のサービスプランからサービスプランを選ぶためにグラフィカルユーザインタフェースを提供するように構成される。

【００６１】

図８は、ユーザ／管理者が、サービスプランを作り出す、及び／または割り当てることを可能にする表示の形の、この種のユーザインタフェース８００の一例を示す。ここに示した例では、図８ａはサービスプランを作り出して階層化ポリシーを選ぶためのスクリーンショットを示す。規定されたポリシーを閲覧して編集するオプションが、スクリーンの底部にある。図８ｂでは、スクリーンショットが、作り出されるべきネームスペースに対するサービスプランを選ぶ時何が起こるかを示す。図８ｃでは、スクリーンショットが既存のネームスペース上のサービスプランを変更する方法を示す。

【００６２】

もちろん、図１内に例示されるシステム構成は、本発明が実現されることができるコンテンツプラットホームまたは複製オブジェクト記憶システムの純粋に例示的なものであり、及び、本発明は特定のハードウェア構成に限定されない。本発明を実現するコンピュータ及び記憶システムは、さらに上記の本発明を実現するために使用されるモジュール、プログラム及びデータ構造を記憶して読み出すことができる周知の入出力装置（例えばＣＤ及びＤＶＤドライブ、フロッピーディスクドライブ、ハードディスク、その他）を有することができる。これらのモジュール、プログラム及びデータ構造は、この種のコンピュータ可読の媒体上でコード化されることができる。例えば、本発明のデータ構造が、本発明内に使用されるプログラムがその上に存在する１つ以上のコンピュータ可読の媒体から独立にコンピュータ可読の媒体上に記憶されることができる。システムのコンポーネントは、デジタルデータ通信の任意の形式または媒体、例えば通信ネットワークによって相互接続されることができる。通信ネットワークの例は、ローカルエリアネットワーク、広域ネットワーク、例えばインターネット、無線ネットワーク、記憶領域ネットワーク、などを含む。

【００６３】

記述では、数多くの詳細が本発明の徹底的な理解を提供するために説明のために記載されている。しかしながら、これらの具体的な詳細の全てが本発明を実践するために必要とはされないことが当業者にとって明らかである。さらに注意されることは、本発明が通常フローチャート、流れ図、構造線図またはブロック図として表されるプロセスとして記述されることができることである。フローチャートがシーケンシャルプロセスとして操作を記述することができるとはいえ、操作の多くは並列にまたは同時に実行されることができる。加えて、操作の順序は再配置されることができる。

【００６４】

周知のように、上記した操作はハードウェア、ソフトウェアまたはソフトウェア及びハードウェアのいくつかの組合せによって実行されることができる。本発明の実施態様の種々の態様が、回路及び論理装置（ハードウェア）を使用して実現されることができ、一方他の態様が、プロセッサによって実行されると、本発明の実施態様を実施する方法をプロセッサに実行させる機械可読の媒体上に記憶される命令（ソフトウェア）を使用して実現されることができる。さらに、本発明のいくつかの実施態様はハードウェアでもっぱら実行されることができるが、一方、他の実施態様はソフトウェアでもっぱら実行されることができる。さらに、記述される種々の機能が、単一ユニット内に実行されることができるかまたは任意の数の方法で複数のコンポーネントにわたって広げられることができる。ソフトウェアによって実行される時、この方法は、コンピュータ可読媒体上に記憶される命令に基づいて汎用コンピュータのようなプロセッサによって実行されることができる。必要に応じて、命令は圧縮された及び／または暗号化された形態で媒体上に記憶されることができる。

【００６５】

上記のことから、本発明が、記憶階層化のためのコンテンツ選択に基づいて固定コンテンツシステム内の固定コンテンツの記憶を管理するための方法、器具及びコンピュータ可読の媒体上に記憶されるプログラムを提供することが明白である。加えて、特定の実施態様がこの明細書内に例示されて記述されたとはいえ、当業者は、同じ目的を達成するために算出される任意の配置が開示される特定の実施態様と置換されることができると認識する。この開示は本発明のありとあらゆる適応または変形を包含することを意図され、及び、以下の請求項内に使用される用語が、本発明を明細書内に開示される特定の実施態様に限定するように解釈されるべきでないことが理解されるべきである。むしろ、本発明の範囲は以下の請求項によって完全に決定されるべきであり、この種の請求項が権利を与えられる等価物の完全な範囲とともに、それが請求項解釈の確立された原則に従って解釈されるべきである。

【００６６】

「先行技術文献のリスト」

「特許文献」
「ＰＴＬ１」米国特許第８，００６，１１１号明細書
「ＰＴＬ２」米国特許第７，１５５，４６６号明細書
「ＰＴＬ３」米国特許出願公開第２００７／０１８９１５３号明細書
「ＰＴＬ４」米国特許第７，６５７，５８１号明細書
「ＰＴＬ５」米国特許出願公開第２０１１／０１０６８０２号明細書

【図１】

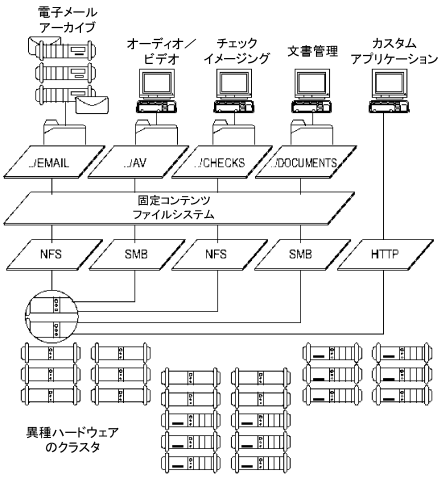

FIG. 1

【図２】

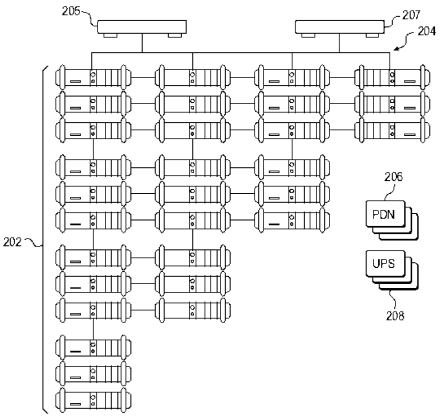

FIG. 2

【図３】

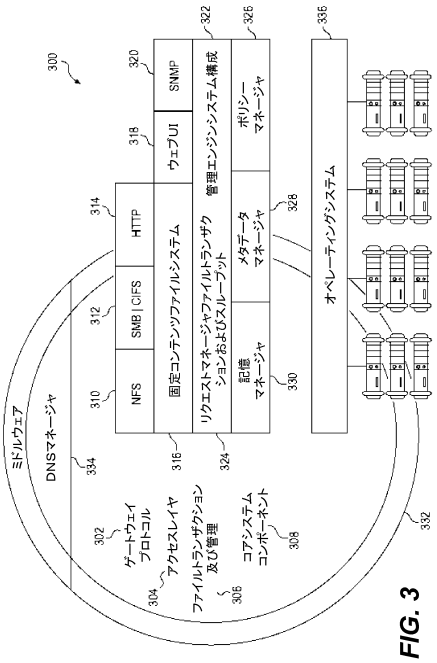

FIG. 3

【図４】

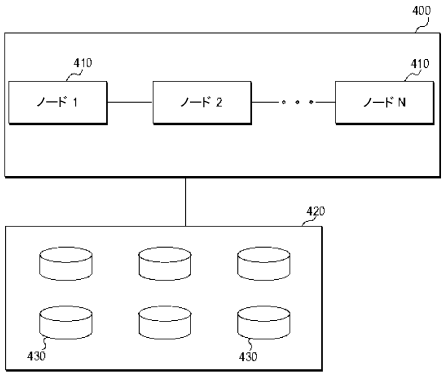

FIG. 4

【図５】

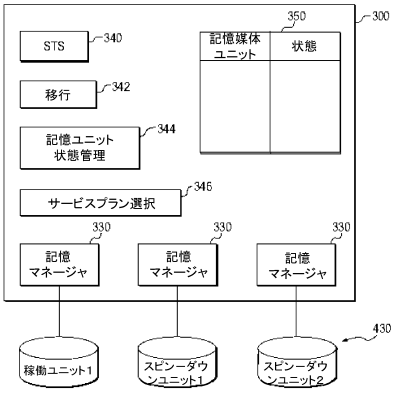

FIG. 5

【図６】

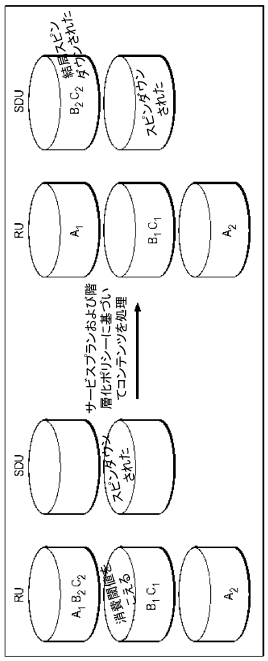

FIG. 6

【図７】

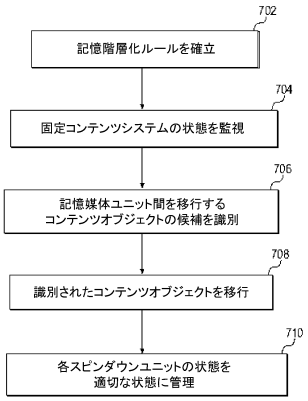
702
記憶階層化ルールを確立
704
固定コンテンツシステムの状態を監視
706
記憶媒体ユニット間を移行する
コンテンツオブジェクトの候補を識別
708
識別されたコンテンツオブジェクトを移行
710
各スピンダウンユニットの状態を
適切な状態に管理

FIG. 7

【図８Ａ】

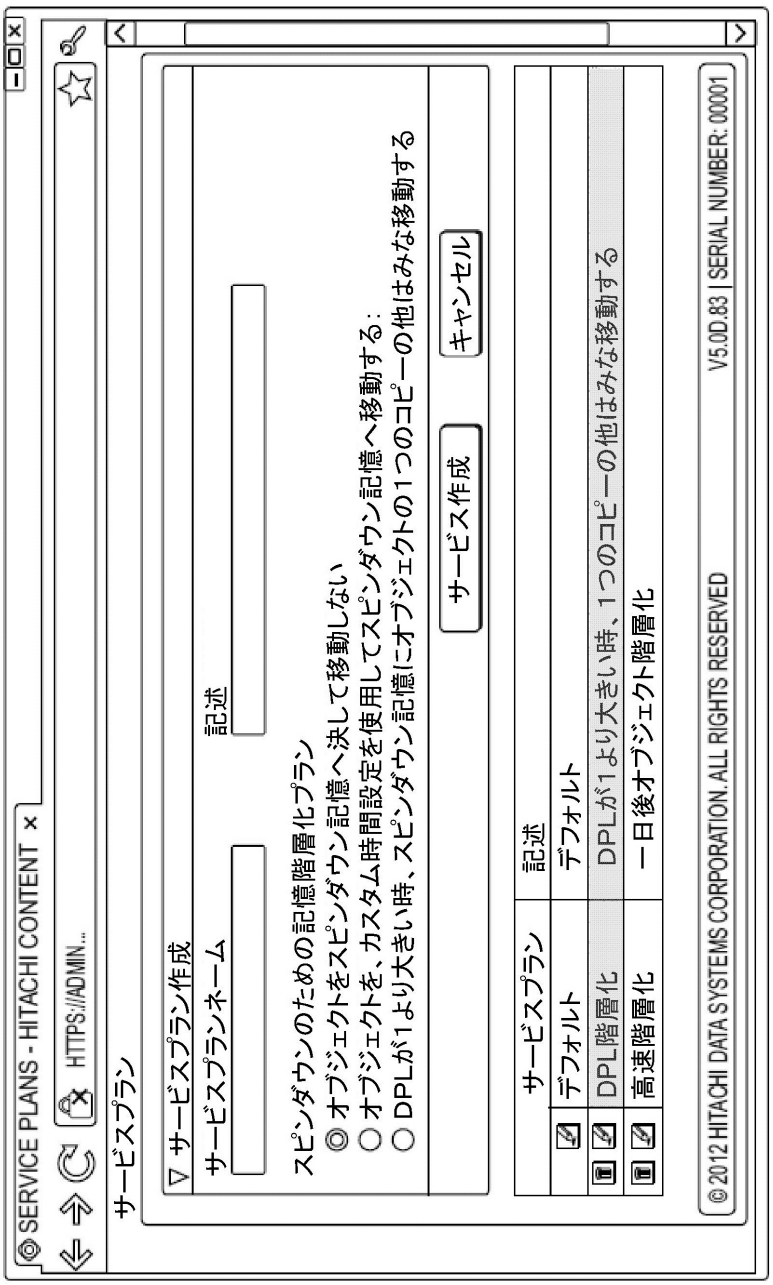
SERVICE PLANS - HITACHI CONTENT
HTTPS://ADMIN...
サービスプラン
▽ サービスプラン作成
サービスプランネーム
記述
スピンダウンのための記憶階層化プラン
◎ オブジェクトをスピンダウン記憶へ決して移動しない
○ オブジェクトを、カスタム時間設定を使用してスピンダウン記憶へ移動する:
○ DPLが1より大きい時、スピンダウン記憶にオブジェクトの1つのコピーの他はみな移動する
サービス作成
キャンセル
サービスプラン
記述
デフォルト
デフォルト
DPL階層化
DPLが1より大きい時、1つのコピーの他はみな移動する
高速階層化
一日後オブジェクト階層化
© 2012 HITACHI DATA SYSTEMS CORPORATION. ALL RIGHTS RESERVED
V5.0D.83 | SERIAL NUMBER: 00001

FIG. 8A

【図８Ｂ】


NAMESPACES - HITACHI CONTENT
HTTPS://TEN..
ネームスペース作成
ネームスペースネーム
記述
サービスプラン
サービスプラン
記述
デフォルト
DPL階層化
DPLが1より大きい時、1つのコピーの他はみな移動する
高速階層化
一日後オブジェクト階層化
サービスプラン適用
キャンセル
© 2012 HITACHI DATA SYSTEMS CORPORATION. ALL RIGHTS RESERVED
V5.0D.83 | SERIAL NUMBER: 00001

FIG. 8B

【図８Ｃ】

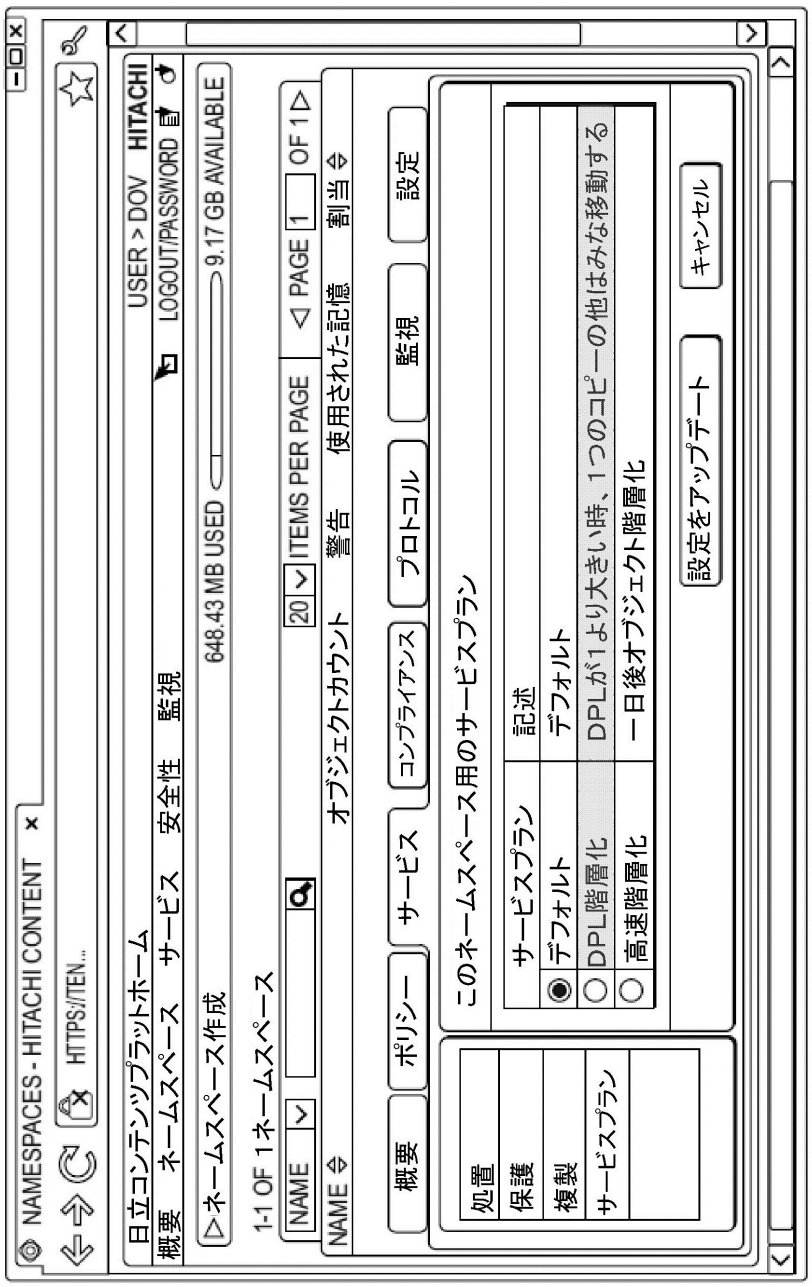
NAMESPACES - HITACHI CONTENT
HTTPS://TEN...
日立コンテンツプラットホーム
USER > DOV HITACHI
概要 ネームスペース サービス 安全性 監視
LOGOUT/PASSWORD
▷ネームスペース作成
648.43 MB USED
9.17 GB AVAILABLE
1-1 OF 1ネームスペース
NAME
20 ITEMS PER PAGE
◁ PAGE 1 OF 1▷
NAME
オブジェクトカウント 警告 使用された記憶 割当
概要
ポリシー
サービス
コンプライアンス
プロトコル
監視
設定
処置
保護
複製
サービスプラン
このネームスペース用のサービスプラン
サービスプラン
記述
デフォルト
DPL階層化
DPLが1より大きい時、1つのコピーの他はみな移動する
高速階層化
一日後オブジェクト階層化
設定をアップデート
キャンセル

FIG. 8C

フロントページの続き

(51)Int.Cl. FI
G06F 12/00 501B

(72)発明者 ゾロツスキー, ヴィタリー
アメリカ合衆国 02474 マサチュセッツ州 アーリントン, カレッジ アベニュー 107
(72)発明者 ブライアント, アラン, ジー.
アメリカ合衆国 02032 マサチュセッツ州 イースト ウォルポール, ユニオン ストリート 228

審査官 木村 雅也

(56)参考文献 特開平07－262058（JP, A）
特開2012－027933（JP, A）
特開2011－076294（JP, A）
国際公開第2011／108021（WO, A1）
国際公開第2010／131292（WO, A1）
米国特許出願公開第2004／0243761（US, A1）
米国特許第07567188（US, B1）

(58)調査した分野(Int.Cl., DB名)
G06F 3／06
G06F 12／00