

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5060485号  
(P5060485)

(45) 発行日 平成24年10月31日(2012.10.31)

(24) 登録日 平成24年8月10日(2012.8.10)

(51) Int.Cl.		F I			
<b>G06F 12/00</b>	<b>(2006.01)</b>		G06F 12/00	533J	
<b>G06F 3/06</b>	<b>(2006.01)</b>		G06F 12/00	531D	
			G06F 3/06	304F	

請求項の数 14 (全 17 頁)

(21) 出願番号	特願2008-533573 (P2008-533573)	(73) 特許権者	508093322
(86) (22) 出願日	平成18年9月27日 (2006. 9. 27)		オナロ インコーポレイテッド
(65) 公表番号	特表2009-510624 (P2009-510624A)		ONARO, INC.
(43) 公表日	平成21年3月12日 (2009. 3. 12)		アメリカ合衆国 02118 マサチュー
(86) 国際出願番号	PCT/US2006/037711		セッツ、ボストン、ハリソン アベニュー
(87) 国際公開番号	W02007/038617		500、スイート 4エフ
(87) 国際公開日	平成19年4月5日 (2007. 4. 5)		500 Harrison Avenue
審査請求日	平成21年9月28日 (2009. 9. 28)		, Suite 4F, Boston, MA
(31) 優先権主張番号	60/720, 977	(74) 代理人	110000523
(32) 優先日	平成17年9月27日 (2005. 9. 27)		アクシス国際特許業務法人
(33) 優先権主張国	米国 (US)	(72) 発明者	ラファエル ヤハロム
			アメリカ合衆国 02494 マサチュー
			セッツ、ニーダム、フーバー ロード 1
			12

最終頁に続く

(54) 【発明の名称】 複製データの可用性及び最新性を検証するための方法及びシステム。

(57) 【特許請求の範囲】

【請求項1】

ネットワーク内で接続されたコンポーネンツと組み合わせた、レプリカデータの検証管理システムであって、

ネットワークのためのレプリカデータポリシーにして、少なくとも、ネットワーク内のストレージコンポーネントに記憶されたレプリカボリュームのためのリカバリーポイント時間 ( a g e ) 条件及びアクセス経路条件、を指定するレプリカデータポリシーを記憶するように構成されたメモリーサブコンポーネントと、

コミュニケーション可能な状態下にネットワークに連結されたネットワークコンポーネントにして、ネットワーク内の前記コンポーネンツが、1つ以上のホストコンポーネントと、当該ホストコンポーネント上で実行されるアプリケーションとを含み得るネットワークコンポーネントと、

メモリーサブコンポーネント及びネットワークコンポーネントに電子的に連結したレプリカデータ検証マネージャにして、該レプリカデータ検証マネージャが、

ネットワークコンポーネントを使用して、レプリカボリュームと関連するアクセス経路セットにして、該セットにおけるアクセス経路が、レプリカボリュームと、ネットワーク内のコンポーネンツの1つとの間のデータフローパスを表すアクセス経路セットを監視し、

ネットワークコンポーネントを使用してネットワーク内のデータ複製を監視し、該データ複製の監視が、レプリカボリュームがネットワーク内のストレージコンポーネントに

記憶されたソースボリュームを示したスナップショット時間をレプリカデータ検証マネージャにより確認することを含み、

レプリカボリュームの、該スナップショット時間からの経過時間を表すリカバリーポイント時間を算出し、

算出したリカバリーポイント時間をリカバリーポイント時間条件と比較し、

レプリカボリュームに関連するアクセス経路セットをアクセス経路条件と比較し、

レプリカデータポリシーの条件が違反のものであった場合に違反を通知する、

如く構成されるレプリカデータの検証管理システム。

【請求項 2】

レプリカデータポリシーが、ホストとレプリカボリュームとの間の要求アクセス経路、アプリケーションとレプリカボリュームとの間の要求アクセス経路、ソースボリュームを表すレプリカボリュームの要求数、レプリカボリュームの要求位置、ソースボリュームを表す複数のレプリカボリューム間の1つ以上の要求アクセス経路、の少なくとも1つを規定することを更に含む請求項1のレプリカデータの検証管理システム。

10

【請求項 3】

各レプリカボリュームが、同期、スナップショット時間、不一致、から選択したコピー状況 ( s t a t u s ) に関連する請求項1のレプリカデータの検証管理システム。

【請求項 4】

アクセス経路条件が、ローカルホストとレプリカボリュームとの間のアクセス経路、リモートホストとレプリカボリュームとの間のアクセス経路、レプリカボリュームと関連する、破局的障害からの復旧用のアクセス経路、データコピー機とレプリカボリュームとの間のアクセス経路、の少なくとも1つを含む請求項1のレプリカデータの検証管理システム。

20

【請求項 5】

アクセス経路条件が、アクセス経路の冗長性、アクセスパスの中間コンポーネント数、アクセスパスのパフォーマンスのレベル、の1つ以上が含まれる請求項1のレプリカデータの検証管理システム。

【請求項 6】

レプリカデータ検証マネージャが、レプリカボリュームに関するアクセス経路セットでの、ネットワークにおける変化の効果を予測する構成を更に有する請求項1のレプリカデータの検証管理システム。

30

【請求項 7】

アクセス経路セットを監視することが、ネットワークをグラフとして略示することを含み、グラフにおける各ノードがネットワーク内のホストコンポーネントを表し、グラフにおける2つのノード間の各縁部が、2つのノードにより表されるホストコンポーネント間のデータフロー能力を表し、データフロー能力が、2つのノードにより表されるホストコンポーネントの構成によって決定される請求項1のレプリカデータの検証管理システム。

【請求項 8】

コンピューターシステムにおいてネットワークにおけるコンポーネントのデータ複製を監視するための方法であって、

40

ネットワークにおけるコンポーネントが、1つ以上のホストコンポーネントと、該ホストコンポーネント上で実行されるアプリケーションとを含み得、

前記方法が、レプリカデータ検証マネージャにより、

ネットワークのためのレプリカデータポリシーにして、ネットワーク内のストレージコンポーネントに記憶させたレプリカボリュームのためのリカバリーポイント時間条件及びアクセス経路条件を規定するレプリカデータポリシーをメモリーサブコンポーネントに記憶すること、

レプリカボリュームに関するアクセス経路セットにして、該セットにおけるアクセス経路が、レプリカボリュームと、ネットワーク内のコンポーネントとの間のデータフローパスを表すアクセス経路セットを監視すること、

50

ネットワーク内のデータ複製を監視し、該監視が、レプリカボリュームがネットワーク内のストレージコンポーネントに記憶されたソースボリュームを示したスナップショット時間をレプリカデータ検証マネージャにより確認することを含み、

レプリカボリュームの、該スナップショット時間からの経過時間を表すリカバリーポイントの時間を算出すること、

算出したリカバリーポイント時間をリカバリーポイント時間条件と比較すること、

レプリカボリュームに関連するアクセス経路セットをアクセス経路条件と比較すること

、  
レプリカデータポリシーの条件が違反のものであった場合に違反を通知すること、  
を含む方法。

10

#### 【請求項 9】

レプリカデータポリシーが、ホストコンポーネントとレプリカボリュームとの間の要求アクセス経路、アプリケーションとレプリカボリュームとの間の要求アクセス経路、ソースボリュームを表すレプリカボリュームの要求数、レプリカボリュームの要求位置、ソースボリュームを表す複数のレプリカボリューム間の1つ以上の要求アクセス経路、の少なくとも1つを規定することを更に含む請求項 8 の方法。

#### 【請求項 10】

各レプリカボリュームが、同期、スナップショット時間、不一致、から選択したコピー状況に関連する請求項 8 の方法。

#### 【請求項 11】

アクセス経路条件が、ローカルホストとレプリカボリュームとの間のアクセス経路、リモートホストとレプリカボリュームとの間のアクセス経路、レプリカボリュームと関連する、破局的障害からの復旧用のアクセス経路、データコピー機とレプリカボリュームとの間のアクセス経路、の少なくとも1つを含む請求項 8 の方法。

20

#### 【請求項 12】

アクセス経路条件が、アクセス経路の冗長性、アクセスパスの中間コンポーネント数、アクセスパスのパフォーマンスのレベル、の1つ以上が含まれる請求項 8 の方法。

#### 【請求項 13】

レプリカデータ検証マネージャが、レプリカボリュームに関するアクセス経路セットでの、ネットワークにおける変化の効果を予測する構成を更に有する請求項 8 の方法。

30

#### 【請求項 14】

アクセス経路セットを監視することが、ネットワークをグラフとして略示することを含み、グラフにおける各ノードがネットワーク内のコンポーネントを表し、グラフにおける2つのノード間の各縁部が、2つのノードにより表されるホストコンポーネント間のデータフロー能力を表し、データフロー能力が、2つのノードにより表されるホストコンポーネントの構成によって決定される請求項 8 の方法。

#### 【発明の詳細な説明】

#### 【技術分野】

#### 【0001】

本件出願は、ここに引用することによりその全てを本明細書に含むものとする2005年9月27日付で提出した米国仮特許出願番号第60/720,977号の利益を主張するものである。

40

本発明は、ネットワーク環境内で複製（以下、レプリケーションとも称する）されたデータの可用性及び最新性、即ち、実態を判定及び検証するための方法及びシステムに関する。

#### 【背景技術】

#### 【0002】

データレプリケーションはデータの可用性を高めるために通常用いられる手法である。データセットを多重に複製して別々の場所に保存すれば、例え幾つかのコンポーネント障害が発生する又はデータセットの幾つかが破損した場合でも、クライアントアプリケーシ

50

ョン用の複製データ（以下、レプリカデータとも称する）をずっと入手し易くなる。

コンピューティングシステムには、データをコピーし、多数のレプリカデータを管理するための数多くの手法がある。複製手法は2つの主要カテゴリー、即ち、同期複製法と非同期複製法とに分類される。同期複製法ではソースデータセットとレプリカデータとは連続的に完全同期されるのでトランザクションが高度に保証され、またソースデータセットのアップデートは一貫して且つ直ちに全部の同期レプリカデータに反映される。然し乍ら同期複製法は、それがコンピューティングリソースに課す諸経費上、同法を実現するための費用が法外に高額になったり、全く実現不能になる場合（例えば、環境内の幾つかのコンポーネントの一時的破損による）もある。

他方、非同期複製法ではデータは定期的に複製されるだけなので、各レプリカデータ間の時間整合性の厳密度はずっと低くなり、レプリカデータは現在のデータソースではなくむしろ、ある程度前のデータソース状態を表すものとなり得る。参照ポイントがどの程度前かにもよるが、ある例外的状況（例えば、破局的障害からの復旧）下のあるクライアントアプリケーションにとってはそうした不整合は尚、許容し得るものである。非同期複製法は、コンピューティングリソースに課す諸経費はずっと低いので、障害復旧（DR）用にアプリケーションデータのリモートコピーを維持する等の多くの環境で一般に使用されている。

#### 【0003】

然し乍ら、データセット及びその複製をアプリケーション条件と確実に一致させ続けるのは数多くの理由から困難な課題となっており、そうした理由には、あるアプリケーションのレプリカデータの最新性に関する最低条件は別のアプリケーションのそれとは異なり得る（つまり、コストを取るか最新性を取るかについて相違がある）こと、代表的な環境では同時に実行可能な多数のデータコピー機が存在し得ること、コピー動作がオリジナルのデータセットではなくむしろレプリカデータセット（完全に最新では無い可能性がある）に基づくものであり得るので依存関係連鎖が生じること、個々のコピー動作が完全に失敗し、例えば、ネットワーク又はコポーネント設定上の問題が生じ、リモートサイトに置いたレプリカデータをホスト側で利用できなくなること、等がある。

結局、アプリケーションは必要時にリモートサイトの十分最新のレプリカデータを利用することができない可能性があるが、現段階ではそうした欠陥はアプリケーションがレプリカデータを実際に必要とするまで分からない。現在の複製手法は、多数のレプリカデータの最新性や可用性を連続的にEnd-to-End検証することではなく、個々のコピーメカニズムの実用精度に関心が向けられている。

#### 【0004】

【特許文献1】米国仮特許出願番号第60/720,977号

【特許文献2】米国特許出願番号第10/693632号

【特許文献3】米国特許出願番号第11/112,942号

【特許文献4】米国特許出願番号第11/112,624号

【発明の開示】

【発明が解決しようとする課題】

#### 【0005】

ネットワーク上のレプリカデータセットがアプリケーションの最新性及び可用性に関する指定条件と一致しているかを連続的に検証し、また、不一致を特定し、望ましからざる結果が生じる前に正しい取り扱いがなされるようにユーザーにこの不一致を知らせるシステム及び方法を提供することである。

【課題を解決するための手段】

#### 【0006】

本発明によれば、ネットワーク上のレプリカデータセットがレプリカデータポリシーと合致していることを連続的に検証するシステム及び方法が提供される。

本発明の1様相によれば、ネットワーク上のネットワーク機器に置いたレプリカデータを検証するための方法であって、ネットワーク上でデータを複製するためのレプリカデー

10

20

30

40

50

タポリシーを定義する段階と、各ネットワーク機器間又はネットワーク機器上で実行される各アプリケーション間のアクセス経路を監視すること、ネットワーク上でのデータ複製動作を監視すること、レプリカデータの最新性と可用性とをレプリカデータポリシーの条件と比較してレプリカデータポリシーとの不一致を識別すること、を含む方法が提供される。

【0007】

本発明の他の様相によれば、ネットワーク上のネットワーク機器に置いたレプリカデータを検証するためのレプリカデータ検証マネージャであって、ネットワーク上のレプリカデータ用のレプリカデータポリシーを記憶したポリシーエンジンと、各ネットワーク機器又はネットワーク機器上で実行される各アプリケーション間のアクセス経路を監視する検証エンジンと、を含むレプリカデータ検証マネージャが提供される。検証エンジンはネットワーク上での複製動作をも監視し、レプリカデータの最新性と可用性とをレプリカデータポリシーの条件と比較してレプリカデータポリシーとの不一致を識別する。レプリカデータ検証マネージャは、レプリカデータを検証できない時に違反報告を提供する通知エンジンをも含む。

10

【0008】

本発明のある実施例には以下に説明する特徴部分の1つ以上が含まれ得る。レプリカデータポリシーは、ホスト又はアプリケーションとレプリカデータとの間のアクセス経路、及び又はネットワーク上のレプリカデータ数、レプリカデータ位置、各レプリカデータ間の1つ以上のアクセス経路、及び又は各レプリカデータ位置での相当するレプリカデータの最長有効時間(maximum-age)を規定し得る。データ複製動作には、同期、分割、コピー開始、コピー完了、ソースデータボリューム、ターゲットデータボリューム、レプリケーション時間、の少なくとも1つを監視することが含まれ得る。各レプリカデータは、ソースデータセットネーム、コピー状態、タイムスタンプの少なくとも1つを含み得るタグと関連付けされ得る。レプリカデータのタグはコピー機イベントにตอบสนองしてアップデートされ得る。

20

【0009】

アクセス経路属性には、冗長性及び又は中間コンポーネント数及び又はパフォーマンス及び又はセキュリティ及び又はインターオペラビリティ、共有性及び又はキャパシティが含まれ得る。不一致は、最新性違反及び又はアクセス経路欠失及び又は不正アクセス経路、及び又は経路属性違反、によって識別され得る。本発明の別の実施例では、レプリケーションリポートには、アプリケーションのプロパティ、レプリケーション違反及びその修正時間、データ保管期間におけるレプリケーションリソース利用、又はそれらの組み合わせ、が含まれ得る。

30

【発明の効果】

【0010】

ネットワークのレプリカデータセットがアプリケーションの最新性及び可用性に関する規定条件と一致しているかを連続的に検証し、不一致を特定し、望ましからざる結果が生じる前に正しい取り扱いがなされるようにユーザーにこの不一致を知らせるシステム及び方法が提供される。

40

【発明を実施するための最良の形態】

【0011】

本発明に関し、レプリカデータ環境の各コンポーネントを以下の用語を用いて分類する。

各用語は以下のように定義する。

“ホストコンポーネント”とは、アプリケーションソフトウェアプログラムがある有益な目的を達成するために実行され得るプラットフォームを言う。各ホストコンポーネントでは任意の時点で1つ以上のアプリケーションが実行され得、各アプリケーションは1つ以上のホストコンポーネント(以下、単にホストとも称する)上で実行され得る。また各ホストはホストリソースへのアクセスや外部コンポーネントへのアクセスを制御する制御

50

プログラムを含み得る。

【 0 0 1 2 】

“ストレージコンポーネント”又は“ストレージデバイス”とは、データを読み書きできるプラットフォームを言う。各ストレージデバイスは、各々が1つ以上のビットを記憶できる複数のアドレスを持つメモリーサブコンポーネントを収納し、データが、ボリュームとして参照される単位でストレージデバイスに関して読み書きされる。1つのボリュームはビット数で表す任意量のデータを含み得、各ボリュームは、特定ストレージデバイスの特定開始アドレスを付けてストレージデバイスに記憶させる。また、各ストレージデバイスにはメモリーサブコンポーネント内のデータへのアクセスを制御するコントローラサブコンポーネントも格納される。

10

【 0 0 1 3 】

“ネットワークコンポーネント”とは、データをそこを介して転送し得る、また任意のソースコンポーネントから任意のターゲットコンポーネントに経由させ得るプラットフォームを言う。各ネットワークコンポーネントは、ソース、宛先及びステータス状態、に基づいてデータフローを制御することができる。

説明した各コンポーネントは、関連するユニーク識別子(名前)を持っている他、データのやり取りを可能にする1つ以上のポートをも有する他に、コンポーネントの現在の“制御構成”を表すローカル状態をも有し得る。この現在の“制御構成”では、ソース及びターゲットの各コンポーネントに基づいてどのデータフローを有効化させるべきかといった特定の情報フロー特性が定義される。

20

【 0 0 1 4 】

一方のコンポーネントから他方のコンポーネントにデータを流し得る“コミュニケーションリンク”を使用して、異なるコンポーネント同士を相互に連結することができ、そうしたコミュニケーションリンクは、1つのサイト位置又はリモート位置で極めて接近して位置付けた各コンポーネントを連結させ得る。あるコミュニケーションチャンネル例では、ケーブル、ポイント間連結部、ローカルエリアネットワーク、ワイドエリアネットワークその他が含まれ得る。

“アクセス経路”とは、コミュニケーションリンクが接続性を持っており、各中間コンポーネントのみならずエンドポイント自体が当該エンドポイント同士間でデータを流せるような構成を持つ場合に、2つのエンドポイント(コンポーネント、データセット等)間に存在するものを言う。

30

【 0 0 1 5 】

“環境構成イベント”とは、環境内で生じ得るものであって、異なるクラス、中でも、コンポーネント構成の変化、コンポーネントの追加及び削除、コンポーネントの故障と復帰、データの送受信、データボリュームの読み書き及びその他、を含む異なるクラスに分類され得るものを言う。

“アプリケーション”とは、ホスト上で実行され、新規の“データボリューム”を発生し得、発生したデータボリュームをストレージデバイスに記憶させるべく送信するのみならず、ストレージデバイス上に記憶させた既存のデータボリュームをアップデート又は読み込むものを言う。

40

【 0 0 1 6 】

“データコピー機”とは、プログラムであって、ある時点にストレージデバイスからデータボリュームを読み取り、読み取ったデータボリュームの同一コピーを同じストレージデバイスの別の位置(アドレス)又は別のストレージデバイスに書き込むものを言う。あるデータコピー機は任意のコンポーネント(ホストコンポーネント、ストレージコンポーネント)上で実行され得る。アプリケーションによりアップデートされた初期ソースボリュームはソースボリュームとして参照され、コピー機から発生する任意のボリュームはレプリカボリュームとして参照される。

【 0 0 1 7 】

図1には、レプリカデータ環境の1環境例が示され、3つのホストコンポーネント10

50

2、108、110と、多数のネットワークコンポーネント、例えば、ストレージエリアネットワーク内のスイッチ又はルーター112、114、116と、ワイドエリアネットワーク内のネットワークコンポーネント118と、2つのストレージコンポーネント104及び106と、幾つかのデータボリューム及びその複製と、を含んでいる。例えばソースボリュームは128、そのローカルレプリカは130であり、132はソースボリューム128のリモートレプリカであり、134はリモートレプリカ132のローカルレプリカ、つまりソースボリューム128のリモートコピーである。各レプリカのためのコピー動作は異なる時点に別個に実施される。ホストコンポーネント102、スイッチ又はルーター114、ソースボリューム128、ローカルレプリカ130、ネットワークコンポーネント118、リモートレプリカ132、ローカルレプリカ134、ホストコンポーネント108、のシーケンス的構成は、全てのローカルコンポーネントが障害復旧(DR)アクセス経路に沿ってデータを転送させ得るべく正しく構成されていることを条件として、障害復旧(DR)アクセス経路の一例である。

10

#### 【0018】

ネットワークには、矢印172、174、176、178で示すようにレプリカデータ検証マネージャ101を連結する。レプリカデータ検証マネージャ101は、以下に説明するプロセスを実行し、規定のアプリケーションデータ複製ポリシーに従い、全てのレプリカデータの最新性と可用性とを検証する。

図2には、レプリカデータ検証プロセスのハイレベルでのプロセスデータフロー200を示す。プロセスデータフロー200は、ターゲット環境(コンポーネント、リンク、データコピー機等)を先ず発見するステップ202から開始される。

20

#### 【0019】

ステップ204ではレプリカデータアプリケーションポリシーを定義する。レプリカデータアプリケーションポリシーは、各データボリューム用に、当該データボリュームへのアプリケーションアクセス経路(及び属性)と、当該ボリュームのレプリカ間のアクセス経路と、属性と、各レプリカから別のホストへのアクセス経路と、異なる場所(目標復旧時点(RPO)としても参照される)で各レプリカに要求される最低限の最新性(又は許容できる最大有効時間)条件、を定義する。

ステップ206では環境内での、各レプリカの最新性又は可用性に悪影響を及ぼすイベントについての情報が入手される。例えば、レプリカ130からレプリカ132(図1)までのコピー動作が、レプリカ132の最新状態判定に関わる時間Tに成功裡に終了したことを示す情報が入手され得る。レプリカ134の可用性判定に関わる、ホストコンポーネント108とレプリカ134との間のリンク切れに関する情報が入手され得る。

30

ステップ208では、各ボリューム及びレプリカの最新性及び可用性の状態(ステータス)に関するイベント情報を引き出すための分析を実行する。

#### 【0020】

ステップ210では、ステップ208で得た分析結果をポリシーの規定の最新性及び可用性条件と比較し、ステップ212で違反が検出されると当該違反が報告される。そうした違反は状態又は最新性のレベルに関わり得るものであり、例えば、遠隔地に位置付けた、ソースボリューム128の全レプリカの現時点での最新性は、最新性及び可用性条件に規定する最低限の最新性よりも劣る。またそうした違反は可用性にも関わり得るものであり、例えば、リモートホストコンポーネントが、要求されてもソースボリューム128のリモートレプリカを現在利用できないといった場合である。

40

#### 【0021】

図3はアプリケーションデータレプリケーションポリシーの部分例示図である。こうしたポリシーは上述したステップ202、即ち、データレプリケーション検証プロセスステップで設定され、各データボリューム及び各レプリカ用の、アクセス経路301に関連付けすべきアクセス経路のタイプ、例えば経路の冗長性302のレベル、レプリカ数及びコピー動作のタイプ303(例えば、同期、非同期)のみならずその他の属性、を含む。

一般に、アプリケーションデータレプリケーションポリシーは、環境内の各ボリューム

50

に対し、(1)当該ボリュームへのアクセス経路を有すべきホスト及びアプリケーション、(2)持つべきレプリカ数、レプリカを置くべき場所、各レプリカ間で使用するべきアクセス経路、(3)各位置のレプリカが持つべき最大有効時間(RPOはどこか)、(4)リモートホストが任意のレプリカへのアクセス経路を有すべきか、を指定し得る。

#### 【0022】

かくして、アプリケーションデータレプリケーションポリシーは、例えば以下に例示するような条件、即ち、

\*アプリケーションデータボリュームが少なくともある数(例えば、K)のレプリカを有すべきであること。

\*アプリケーションデータボリュームの少なくとも1つが、アプリケーションホストコンポーネント位置から遠隔させたストレージコンポーネントに存在(リモートレプリカ)すべきであること。

\*リモートレプリカの少なくとも1つが、現在のアプリケーションホストコンポーネント位置から遠隔させた別のホストコンポーネントから利用可能とされるべきであること。

\*アプリケーションデータボリュームの少なくとも1つが、現在時間から時間単位Tよりは長くない以前のスナップショット(復旧時点)を表すべきであること。

\*アプリケーションホストコンポーネントからリモートレプリカまでのアクセス経路が特定のアクセス経路属性(例えば、デュアルファブリック冗長構成)を有すべきであること。

\*所定のアプリケーションの全データボリュームが、現在時間から時間単位Tよりは長くない以前のスナップショットを使用して遠隔複製されるべきであること。

\*所定のアプリケーションの全データボリュームが、同じ時点であって且つ現在時間から時間単位Tよりは長くない以前の時点に関わるスナップショットを使用して遠隔複製されるべきであること。

を含む多様なアプリケーション条件を表し得る。

#### 【0023】

レプリカデータ検証プロセスの次のステップ(図2でステップ206として示される)には、レプリケーションイベント情報及び構成イベント情報を収集することが含まれる。

データコピー機レプリケーションイベント情報には、例えば以下の如きものが含まれる。

\*ソースボリューム(ストレージデバイス及びメモリアドレス)、宛先ボリューム(ストレージデバイス及びメモリアドレス)及びイベント時間、を伴う、データコピー機による“同期”イベント。

\*同期イベントは、宛先ボリュームへの状態ソースのコピーを開始し、次いで、宛先ボリュームに重要な各ソースアップデートを連続的にコピーすることを表す。

\*ソースボリューム(ストレージデバイス及びメモリアドレス)、宛先ボリューム(ストレージデバイス及びメモリアドレス)及びイベント時間、を伴う、データコピー機による“コピー完了”イベント。データコピー機による“コピー完了”イベントは、ソースから宛先ボリュームへのコピーが正常に完了したことを表す。当該イベント時点後は、ソースボリュームの任意のエントリが実行する任意の重要なアップデートが、宛先ボリュームへの当該アップデートの必然的なコピーをトリガさせる。

\*ソースボリューム(ストレージデバイス及びメモリアドレス)、宛先ボリューム(ストレージデバイス及びメモリアドレス)及びイベント時間、を伴う、データコピー機による“スプリット(以下、分割とも称する)”イベント。分割イベントは、各ボリューム間の同期関係が終了したことを表す。当該分割イベント後はソースボリュームと宛先ボリュームとの間でのそれ以上のコピー(状態又はアップデートの)は実施されない(次の同期イベントまで)。

#### 【0024】

あるボリュームを別のボリュームに従来通り非同期コピーすることも、同期イベントと当該同期イベント後に同時に実行されるコピー完了イベント及び分割イベントとにより表

10

20

30

40

50

される。

図4にはレプリケーションイベントを視覚化した一例を示す。

時点401ではボリューム1と2との間で同期コピーイベントが生じ、時点402ではコピーが完了し、時点403では分割イベントが発生する。時点404でボリューム2及び3の間でのコピーが開始されると時点405でコピーが完了され、且つ分割が行なわれる。時点405以後のボリューム3の最新性を判定するには以前のコピーイベント履歴を考慮する必要がある。コピー開始時点(例えば、時点401)からコピーが成功裡に完了した時点(例えば時点402)までの間はターゲットレプリカは非整合状態にある点にも注意すべきである。

#### 【0025】

環境構成イベントもアクセス経路に悪影響を及ぼす。収集されるイベント情報には以下の如きものが含まれる。

\* 任意のホストコンポーネント、ネットワークコンポーネント、又はストレージコンポーネント、における任意のアクセス制御構成の変更。そうしたイベントは、例えば、LUN(論理ユニット番号)マスキングの変更、又はゾーニングの変更を含むが、コンポーネントが特定のソースコンポーネントから特定の宛先コンポーネントへのデータフローを有効化するかどうかに関する悪影響を及ぼす。

\* ホストコンポーネント、ネットワークコンポーネント、ストレージコンポーネントの任意の追加、削除、又は移行。

\* 各コンポーネント間でのコミュニケーションリンクの任意の追加又は削除。

\* ホストコンポーネントでのアプリケーションの任意の追加、削除、又は移行。

\* ストレージデバイスでのボリュームの任意の追加、削除、又は移行。

\* 任意のコンポーネントでのデータコピー機の任意の追加、削除、又は移行。

#### 【0026】

レプリケーション検証プロセスの次のステップ(図2ではステップ208として表す)には、収集したイベント情報を分析して最新性及び可用性の判定結果を得ることが含まれる。

最新性は、レプリケーションイベント情報から以下のようにして導出される。各レプリカボリュームとレプリカヘッダタグとの特定の関連付けが維持される。そうした関連付けには、中でも以下の属性が含まれる。

\* オリジナルソースボリューム---ネーム(ストレージデバイス及びアドレス)。

\* 状態-----同期、スナップショット、又は不整合。

\* タイムスタンプ-----スナップショット時間。

レプリカのレプリカヘッダタグは先に説明したように、異なるデータコピー機イベントにตอบสนองしてアップデートされる。こうしたタグアップデートは、例えば、以下の原理に従うものである。

#### 【0027】

\* 当該レプリカを宛先とする同期イベントが発生し、ソースボリュームがオリジナルソースボリューム(レプリカではなく)である時は、

\* \* タグの“オリジナルソースボリューム”属性が同期イベントのソースネームのネームに設定される。

\* \* 宛先のタグの“状態”属性が“不整合”に設定される。

\* 当該レプリカを宛先とする同期イベントが発生し、ソースボリュームがレプリカである時は、

\* \* タグの“オリジナルソースボリューム”属性がソースレプリカのタグのオリジナルソースネームのネームに設定される。

\* \* 宛先のタグの“状態”属性が“不整合”に設定される。

\* 当該レプリカを宛先とするコピー完了イベントが発生し、ソースボリュームがオリジナルソースボリューム(レプリカではなく)である時は、

\* \* 宛先タグの“状態”属性が“非同期”に設定される。

10

20

30

40

50

\* 当該レプリカを宛先とするコピー完了イベントが発生し、ソースボリュームがレプリカである時は、

\* \* 宛先タグの“状態”属性がソースタグの“状態”属性と同じになるように設定される。

\* \* 宛先タグの“タイムスタンプ”がソースタグの“タイムスタンプ”属性と同じになるように設定される。

\* 当該レプリカを宛先とする分割イベントが発生し、ソースボリュームがオリジナルソースボリューム（レプリカではなく）であり、宛先タグ状態が“非同期”ではない時は、

\* \* 宛先タグの“状態”属性が“スナップショット”に設定される。

\* \* 宛先タグの“タイムスタンプ”属性が現在時間に設定される。

\* 当該レプリカを宛先とする分割イベントが発生し、ソースが又は宛先タグの状態が“不整合”である時は、

\* \* 宛先タグの“状態”属性が“不整合”に設定される。

\* 当該レプリカを宛先とする分割イベントが発生し、ソースボリュームがレプリカである時は、

\* \* 宛先タグの“状態”属性が“スナップショット”に設定される。

\* \* 宛先タグの“タイムスタンプ”属性がソースタグの“タイムスタンプ”属性と同じになるように設定される。

#### 【0028】

可用性は、環境内の全アクセス経路を判定することにより、環境構成イベント情報から導出する。つまり、各イニシエーターコンポーネントがデータを取得又は供給可能な全ストレージデバイスの全ボリューム及びレプリカが導出される。各イベント後、以下の原理を使用して分析を反復して実施する。

仮にイニシエーターコンポーネントとターゲットコンポーネントとの間にコミュニケーションコネクティビティ（少なくとも1つのコミュニケーションリンク及び、コミュニケーションリンクを用いて相互連結した、追加され得る中間コンポーネント）があり、しかも、シーケンス（ホストから開始され、ストレージデバイスで終了する）上の各コンポーネントのアクセス制御構成が、シーケンス上で前方のソースと同シーケンス上で後方の宛先との間のデータフローが有効化されるように設定されている場合、イニシエーターコンポーネントとターゲットコンポーネントの間にはアクセス経路が存在する。

#### 【0029】

更には、本プロセスの当該ステップでは、識別した各アクセス経路に関する様々な End-to-End 属性を、当該アクセス経路上の各コンポーネントに関わる全ての状態情報及びイベント情報を累積的に分析して判定したものととして導出する。こうして導出したアクセス経路属性は、中でも、アクセス経路の冗長性のレベル、アクセス経路内のアクセスパフォーマンスのレベル、ホスト及びターゲットボリューム間の中間コンポーネント数、ストレージキャパシティ、のような色々のサービスディメンションを表す。

この分析によれば、同様に、データコピー機及びデータボリューム間（ソース及びターゲットボリューム間）の適宜のアクセス経路の存在も確認される。

#### 【0030】

ホストコンポーネントと、一連の多数のレプリカボリューム（ローカル及びリモート）との間のアクセス関係を表す障害復旧アクセス経路（DRアクセス経路）は、アクセス経路の拡張形式である。

DRアクセス経路のそうした概念はホストからボリュームへの要求アクセス属性のみならず、ボリュームからレプリカへの、またレプリカからレプリカへの要求アクセス属性の何れをも指定及び強制する上で有益な方法である。

そうしたDRアクセス経路は形式的には、例えば以下の如く多様に定義され得る。

\*  $v_1 \sim v_n$  をデータセットとした場合、各データセット  $v_i$  はタイプセット（ $t_1 \sim t_m$ ）に属する特定タイプ  $t_k$  のものである。

\*  $v_i$  と  $v_{i+1}$  とが物理的に連結され且つ論理的にも連結されている場合、 $v_i$  と  $v_{i+1}$  と

10

20

30

40

50

の間に経路セグメントが存在する。 $v_i$ と $v_{i+1}$ とは、それらの間の物理的な単数又は複数の経路が、 $v_i$ 及び $v_{i+1}$ のタイプの関数である特定のルールセットに適合する場合は論理的に連結されている。

\* 仮に $k \leq i < l$ である全ての場合に $v_i$ から $v_{i+1}$ までの経路セグメントが存在する場合、 $v_i \sim v_{i+1}$ には経路が存在する。

#### 【 0 0 3 1 】

その他の定義例には以下のようなものがある。

\*  $v_i \sim v_n$ を計算用ノードとした場合、各ノード $v_i$ はタイプセット $\{t_1 \sim t_m\}$ に属するタイプ $t_k$ のものである。

\*  $v_k$ から $v_l$ までが物理的に連結され、各ノード $v_i$ が $k \leq i < l$ のシーケンスの論理状態に設定され、当該状態により、長期記憶型ノードへの当該シーケンスに沿ったデータフロー（シーケンス上のその他のノードの1つによる動作開始に回答しての）が有効化される場合、 $v_k$ から $v_l$ までの記憶データアクセス経路 $v_k$ から $v_l$ が存在する。

その他の定義例には以下のようなものがある。

\*  $H_i$ をホストとし、 $D_j$ をストレージエリアネットワーク（SAN）デバイス（スイッチ、ルーター、HBA、等のローカルエリア又はワイドエリアネットワーク）とし、 $V_k$ を記憶データセット（ストレージデバイスのボリューム）とした場合、DRアクセス経路のシーケンスは以下の如く $H_a \sim V_m \sim V_n [- H_b]$ となる。

\*  $V_m \sim V_n$ は同じデータセットの同一コピーであるか又は導出したコピー（古いスナップショット、プロセス処理したデータマイニングバージョン等）である。

\* シーケンス上で連続する各部材間は物理的に連結される（ローカルケーブル、リモートリンク、中間SANデバイス等）。

\* 各中間SANデバイスはシーケンスに沿ったデータフローを許容するべく正しく構成される（例えば、ゾーニング、lunマスキング等）。

\* （随意的に） $H_b$ 及び $V_m \sim V_n$ （又はこれらのサブセット）間での情報フローが同様に有効化される（物理的及び論理的に）。

#### 【 0 0 3 2 】

DRアクセス経路は多数の形式を既定しておくことができる。各形式は、レプリカタイプの特定の順列を表すが、場合によっては特定のシーケンスに関わるリモートホストを表す。

例えば、米国マサチューセッツ州HopkintonのEMC社は、インフラストラクチャ用のストレージコンポーネントを供給するが、EMC社のテクノロジーではレプリカタイプを固有名で参照し（例えば、ローカル及びリモートの同期及び非同期型の各レプリカを夫々表すBCVs、R1s、R2s）、レプリカタイプコピーのシーケンス関係に特定の制約を加えている。

#### 【 0 0 3 3 】

EMC社のインフラストラクチャに関しては、以下に示す既定のDRアクセス経路タイプ例、即ち、

\*\* ホスト-R1-BCV

\*\* ホスト-R1-R2

\*\* ホスト-R1-R2-リモートホスト

\*\* ホスト-R1-R2-リモートBCV-リモートホスト

\*\* ホスト-R1-BCV-R2

\*\* ホスト-R1-BCV-R2-リモートBCV

\*\* ホスト-R1-BCV-R2-リモートBCV-リモートホスト

を使用及び実施可能である。各対プレイは、ホスト及びボリューム、ボリューム及びレプリカ、恐らくはレプリカ及びレプリカ、そして、レプリカ及びホスト、の夫々間のアクセス経路関係を表している。

#### 【 0 0 3 4 】

レプリカデータ検証プロセスの次のステップ（図2ではステップ210として表す）に

10

20

30

40

50

は、可用性と最新性の導出結果をポリシーのそれと比較することが含まれる。つまりこの分析により、各アプリケーションのレプリカデータを当該アプリケーションの適宜のレプリカデータ条件に常に完全準拠させることが連続的に可能となる。

こうした条件その他は、先に説明したレプリケーション分析及びアクセス経路分析メカニズムによって以下の如く確認され得る。

\* \* オリジナルソースボリュームが所定ボリュームであるタグを持つ全レプリカの識別に対応して、所定アプリケーションの所定データボリュームのレプリカ数を確認する。

\* \* ホストコンポーネント上のアプリケーションと、ストレージコンポーネント上の各レプリカボリュームとの間のアクセス経路の導出に対応して、所定のアプリケーションでどのレプリカを現在利用可能であるかを確認する。

\* \* ストレージコンポーネントがホストコンポーネントに関して十分に長距離の位置（又は、オリジナルソースボリュームに対するストレージデバイスの位置）にあるとの判定に対応して、遠隔位置のレプリカを確認。

\* \* 当該ホストコンポーネントと、当該レプリカボリュームとの間のアクセス経路の識別及び、ホストコンポーネントとストレージコンポーネントとがオリジナルアプリケーションホストコンポーネント（又はオリジナルソースボリュームストレージコンポーネント）から十分に遠隔位置にあるとの識別に対応して、リモートホストが利用可能な所定のリモートレプリカを確認。

\* \* オリジナルソースボリューム属性が要求属性に相当すること、状態属性が同期していること、又は状態属性がスナップショットであり且つ現在時刻マイナスタイムスタンプ属性が時間単位Tより大きくないこと、の判定に対応して、アプリケーションボリュームの前記リモートレプリカの少なくとも1つが、現在時間以前の時間単位Tよりも長くないスナップショット時間（当該アプリケーションに対する復旧時点）を表すことを確認。

\* \* アプリケーションの各ボリュームに対する上記プロセスの実施と、同期状態にない全タグの最小タイムスタンプの判定と、現在時間マイナス当該最小タイムスタンプが時間単位Tよりも大きくないことの確認、に対応して、所定アプリケーションの全レプリカが過去の大抵の時間単位T（アプリケーション用の復旧時点）の状態を反映すべきである点を確認。

\* \* アプリケーションの各ボリュームに対する上記プロセスの実施と、全レプリカが同期状態か又はスナップショット状態にあってタイムスタンプが同じであり、当該タイムスタンプが現在時間以前の時間単位Tよりも長くないことの判定に対応して、所定アプリケーションの全レプリカが、過去の大抵の時間単位T（当該アプリケーションに対する復旧時点）である同時点での状態を反映すべきであることを確認。

#### 【 0 0 3 5 】

導出結果とポリシー条件との間の不一致は違反と考えられることから、レプリカデータの違反は、ボリュームがホスト又は別のレプリカでポリシーのDRアクセス経路規定と同じ様式では利用できない場合、つまり、既存のアクセス経路又はDRアクセス経路の属性の1つがポリシーの規定の1つとは異なる場合、又は規定距離よりも遠くにあるボリュームの最新レプリカの殆どが、ポリシーの規定する目標復旧時点よりも以前である場合、を表し得る。

#### 【 0 0 3 6 】

図5は、列501を違反例として表したレプリケーション違反の概略例示ダイヤグラム図である。この場合の違反は“DRアクセス経路停止(outage)”形式のものであり、ストレージコンポーネント502上のボリュームからストレージコンポーネント505におけるそのレプリカとの間の可用性断絶(disruption)を含んでいる。これら2つのストレージコンポーネント間のアクセス経路にはネットワークコンポーネント503及び504も含まれる。このアクセス経路に沿ったどこかにイベントが発生してストレージコンポーネント502及び505間の情報フローが妨害され、それによってこれら2つのストレージコンポーネント間でのボリューム及びレプリカに関わるコピー動作が妨害されると、ストレージコンポーネント505のレプリカはおそらく十分に最新化され

10

20

30

40

50

ず、又は整合しなくなる（コピー動作中に障害が生じた場合）。

【 0 0 3 7 】

レプリケーション検証プロセスの次のステップ（図 2 ではステップ 2 1 2 ）には、各違反及び当該違反に関わる関連情報の適宜通知が含まれる。関連情報には、例えば、違反タイプ、悪影響を受けるボリューム及びアプリケーション、違反時間、違反の根本原因、等が含まれ得る。違反の原因イベントは、当該イベントの発生又は非発生の何れかの事実により、違反の根本原因として参照される。違反はそうした関連情報に基づいて分類され、どの違反通知をどのパーティーに送るかが適宜のフィルタリングメカニズムにより判定され得る。

【 0 0 3 8 】

図 6 にはレプリケーション違反の根本原因分析の概略例示ダイヤグラム図が示される。列 6 0 2 は違反 6 0 1 （図 5 では 5 0 1 とも表示される）の根本原因としてのイベントを識別している。イベントは ' 0 5 年 4 月 2 4 日午後 2 : 5 0 分の時点で発生し、ストレージコンポーネント 5 0 2 及び 5 0 5 間でのそれ以降のデータトラフィックを妨害する、ネットワークコンポーネント 5 0 3 上でのゾーニング構成のアップデートを含んでいた。この関連情報を入手したシステム管理者は違反を修正するべく、ストレージコンポーネント 5 0 3 に直ちに別のゾーニング構成を送ることができる。

【 0 0 3 9 】

違反と関連情報とはリスト及び記憶され、違反情報に適用するフィルタリングルールに基づく適宜の通知メッセージが発生され、適宜のパーティーに送られ得る。違反の修正もまた、説明したメカニズムによって自動検出され、記憶した情報がアップデートされて適宜の通知メッセージが送られ得る。

本プロセスでは、受けたイベント情報（コンポーネントやデータコピーからの）もデータベースのヒストリー構造内に記憶され、アクセス経路の導出結果が各イベントに関連付けされる。このヒストリーデータベースは、環境内のアクセス経路状態の完全エポリューションを表し、それ以降の分析及びサマリーレポート用に使用可能である。

【 0 0 4 0 】

本発明の別の実施例によれば、将来的な変更、イベント、障害がレプリケーション条件に与え得る影響を判定し、そうした影響を事前に防止するべく、そうした変更、イベント、障害をプランニング及びシミュレーションすることができる。

具体的なイベントとしては、将来生じ得る、コンポーネント変更、データコピー機動作、又はコンポーネント故障又はサイト障害、がある。

分析は最新状態の環境コンポーネント及びレプリカを用いて先に説明した 1 つに類似して実施され、シミュレートしたイベント入力を、特定動作の累積的影響をシミュレートする環境（図 2 ではステップ 2 0 4 ）から収集したイベント情報として検討し、こうした動作が環境内で実際に発生した場合に違反が発生するかどうかを判定する。

【 0 0 4 1 】

そうした将来的な違反の夫々についての関連情報が、通常の違反（上記説明参照）に対して提供されると同じ方法で発生され提供される。検証後、イノベーションは相当する変更イベントの実施を実際に追跡し、その実施の進捗を追跡し、成功裡に完了したことをレポート（又は違反発生を通知）する。

本発明の更に他の実施例では、レプリケーション違反プロセス中に累積し処理された情報のサマリーレポートが可能とされ得る。最新イノベーションにより収集及び分析した情報は、広範で有益なデータレプリケーション検証のサマリーレポート及びトレンドレポートを発生させ得る。

【 0 0 4 2 】

発生され得るレポート形式の例には、中でも以下のものが含まれる。

\* \* 選択したアプリケーション及びその属性のボリュームの全てのレプリカ。

\* \* 条件（多過ぎる場合はキャピタルリソースが廃棄され得ることを、少な過ぎればリスクの可能性のあることを夫々表す）に関する、各アプリケーションに対するレプリカ

10

20

30

40

50

ボリューム数。

\*\* 最新の全レプリケーション違反と関連情報パラメータ。

\*\* 所定の時間ウィンドウ内で生じた全レプリケーション違反。

\*\* レプリケーション違反の平均修正時間（時間トレンド）。

\*\* 障害シナリオ（破滅的なサイト障害）

\*\* アプリケーションのレプリカ数の時間トレンド、レプリカが要するストレージ容量、その他。

#### 【 0 0 4 3 】

以上、本発明を実施例を参照して説明したが、本発明の中で種々の変更をなし得ることを理解されたい。アクセス経路の種々の様相及びそれらの違反及び取り扱い上のそうした変更は、例えば、2003年10月23日に提出された米国特許出願番号第10/693、632号、2005年4月22日に提出された米国特許出願番号第11/112,942号及び同米国特許出願番号第11/112,624号に記載される。従って、本発明の精神及び範囲は添付する請求の範囲によってのみ限定されるべきである。

10

#### 【 図面の簡単な説明 】

#### 【 0 0 4 4 】

【 図 1 】 本発明に従うレプリカデータ環境及びレプリカデータ検証マネージャの例示ダイヤグラム図である。

【 図 2 】 レプリカデータ検証用のハイレベルの概略フローダイヤグラム図である。

【 図 3 】 レプリカデータポリシーの例示図である。

20

【 図 4 】 レプリケーションイベントのタイムラインを視覚化した例示図である。

【 図 5 】 レプリケーション違反の概略例示ダイヤグラム図である。

【 図 6 】 レプリケーション違反の根本原因分析の概略例示ダイヤグラム図である。

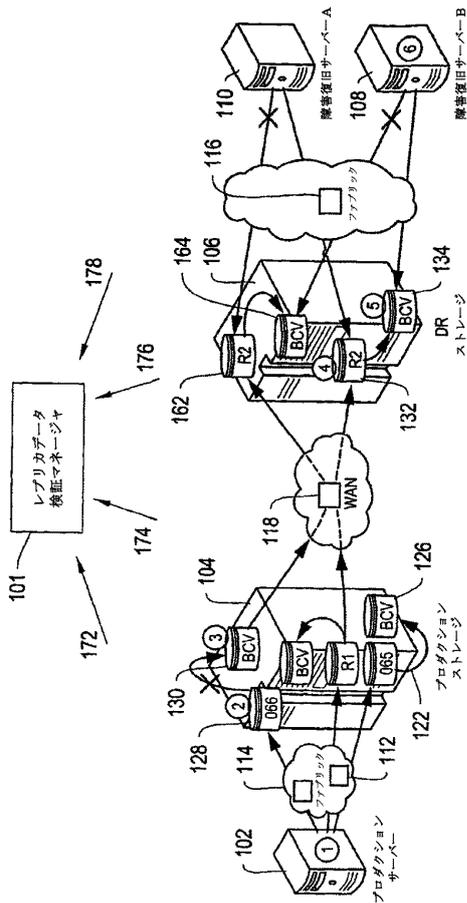
#### 【 符号の説明 】

#### 【 0 0 4 5 】

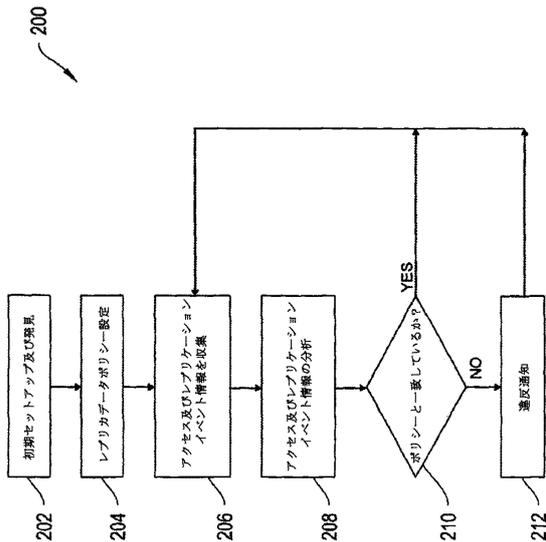
- 101 レプリカデータ検証マネージャ
- 102、108、110 ホストコンポーネント
- 104、106 ストレージコンポーネント
- 112、114、116 スイッチ又はルーター
- 118 ネットワークコンポーネント
- 128 ソースボリューム
- 130、132、134 ローカルレプリカ
- 200 プロセスデータフロー
- 301 アクセス経路
- 302 経路の冗長性
- 303 レプリカ数及びコピー動作のタイプ
- 401、402、403、404、405 時点

30

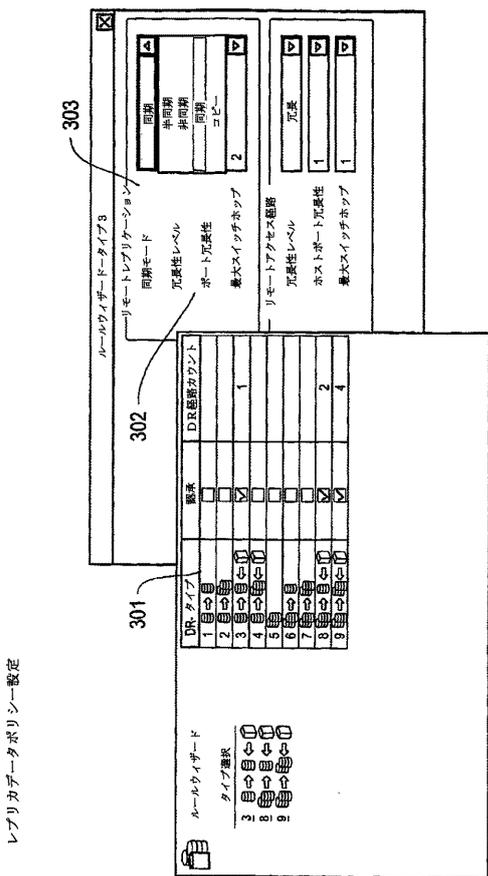
【図1】



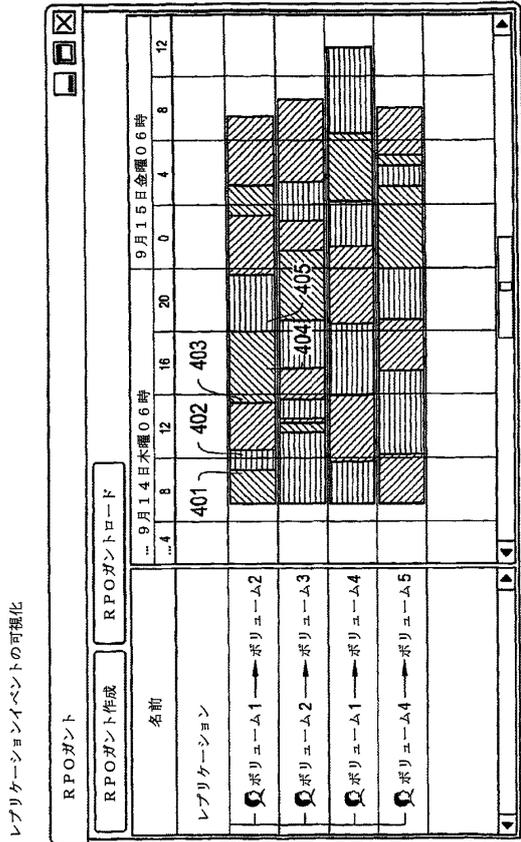
【図2】



【図3】



【図4】





## フロントページの続き

- (72)発明者 ロー アロン  
アメリカ合衆国 02115 マサチューセッツ、ボストン、カンバーランド ストリート 30
- (72)発明者 アサフ レヴィ  
アメリカ合衆国 02446 マサチューセッツ、ブルックライン、マリオン ストリート 77  
、アパートメント 106
- (72)発明者 オムリ ケッセル  
アメリカ合衆国 02118 マサチューセッツ、ボストン、ウォルサム ストリート 15、ピ  
ー401

審査官 桜井 茂行

- (56)参考文献 米国特許出願公開第2004/0205089(US, A1)  
米国特許出願公開第2005/0114403(US, A1)  
米国特許第5452448(US, A)  
米国特許出願公開第2004/030739(US, A1)

## (58)調査した分野(Int.Cl., DB名)

G06F 12/00  
G06F 17/30  
G06F 3/06